DISTILLING AND ADAPTING: A TOPOLOGY-AWARE FRAME-WORK FOR ZERO-SHOT INTERACTION PREDICTION IN MUL-TIPLEX BIOLOGICAL NETWORKS

Anonymous authors

Paper under double-blind review

ABSTRACT

Multiplex Biological Networks (MBNs), which represent multiple interaction types between entities, are crucial for understanding complex biological systems. Yet, existing methods often inadequately model multiplexity, struggle to integrate structural and sequence information, and face difficulties in zero-shot prediction for unseen entities with no prior neighbourhood information. To address these limitations, we propose a novel framework for zero-shot interaction prediction in MBNs by leveraging context-aware representation learning and knowledge distillation. Our approach leverages domain-specific foundation models to generate enriched embeddings, introduces a topology-aware graph tokenizer to capture multiplexity and higher-order connectivity, and employs contrastive learning to align embeddings across modalities. A teacher–student distillation strategy further enables robust zero-shot generalization. Experimental results demonstrate that our framework outperforms state-of-the-art methods in interaction prediction for MBNs, providing a powerful tool for exploring various biological interactions and advancing personalized therapeutics.

1 Introduction

Exploring complex interactions within biological networks is vital for advancing personalized medicine and drug discovery. These interactions enable targeted therapies and the study of disease pathways, calling for predictive models that are both precise and reliable. Moving beyond traditional single-layer analyses, multiplex networks provide a powerful framework to capture the multi-dimensional nature of biological systems, with each layer representing a distinct interaction type, providing a holistic picture of biological connectivity across diverse contexts.

Current predictive models (Kipf & Welling, 2016; Hamilton et al., 2017; Shi et al., 2020; Zhu et al., 2018; Rao et al., 2022; Tao et al., 2023; Dang & Vu, 2024), often rely on single-layer network analyses. Their limitations can be summarized in three areas: (1) Handling Multiplex Data: Single-layer representations discard relational heterogeneity and semantic distinctions between interaction types, overlooking context-dependent reasoning where interactions may rely on both relation type and broader network context. (2) Integrating Multimodal Information: These models have limited capacity to combine biological or chemical sequence features with network topology, which is essential for capturing complex biological interactions. (3) Zero-Shot Prediction: They struggle to predict interactions for unseen entities without prior neighborhood data, restricting generalization to novel biochemical entities.

To address these shortcomings, we propose CAZI-MBN (Context-Aware and Zero-shot Interaction prediction in Multiplex Biological Networks), a framework tailored for multiplex biological interaction prediction. Its four key contributions include: (1) Multiplex-oriented representation learning, capturing interaction-type-specific, context-aware patterns across layered networks. (2) Unified graph-sequence

learning, leveraging domain-specific LLMs for biochemically informed representations. (3) **Strategic knowledge distillation**, transferring latent knowledge from observed to unobserved regions to enable zero-shot prediction. (4) **A Mixture of Experts** (MoE), adaptively modeling diverse interaction types and label dependencies for multi-label prediction.

2 RELATED WORK

Domain-Specific LLMs. Pre-trained LLMs are advancing sequence analysis by applying NLP techniques to biochemical data. ProtTrans (Elnaggar et al., 2021) and ESM-2 (Lin et al., 2023) leverage transformer models for protein function and structure prediction, while ChemBERTa (Chithrananda et al., 2020) and ChemBERTa-2 (Ahmad et al., 2022), trained on up to 77 million molecules (Kim et al., 2023), adapt BERT to chemical SMILES. DNABERT (Ji et al., 2021) and DNABERT-2 (Zhou et al., 2023) extend these approaches to genomic sequences.

Interaction Prediction and Graph Representation Learning. Interaction prediction in networks has advanced through graph representation learning, which embeds nodes or subgraphs into vector spaces preserving topology. Random-walk methods (Perozzi et al., 2014; Ribeiro et al., 2017; Dong et al., 2017; Gao et al., 2018; Grover & Leskovec, 2016) gave way to Graph Neural Networks (GNNs) (Scarselli et al., 2008; Wu et al., 2020; Zhang & Chen, 2018; Xia et al., 2024), including GCN (Kipf & Welling, 2016), Graph Transformer (Shi et al., 2020), and GraphSAGE (Hamilton et al., 2017) for neighborhood aggregation. Contrastive approaches like DGI (Veličković et al., 2018) further improve representations by maximizing mutual information.

Multiplex Network Modeling. Multiplex networks can be modeled as Knowledge Graphs (KGs), which enable semantic reasoning (Walsh et al., 2020; Mohamed et al., 2021; Chen et al., 2023; Chandak et al., 2023) but flatten layered structures and lose interaction-type specificity, limiting zero-shot performance. Multi-layer network models (Park et al., 2020; Jing et al., 2021; Gallo et al., 2023; Gu et al., 2024) instead preserve intra- and inter-layer dependencies, supporting relation-aware learning and multiplex link prediction. Yet, in MBNs (Xiang et al., 2021; Yu et al., 2022; Valdeolivas et al., 2019), integrating sequence features and capturing complex biomedical dependencies across layers remains an open challenge.

Knowledge Distillation. Knowledge distillation (Hinton, 2015) trains a smaller "student" to emulate a larger "teacher," enabling efficient deployment with minimal performance loss. It has proven effective in computer vision (Chen et al., 2017) and NLP (Sanh, 2019). However, its potential to MBNs and zero-shot generalization remains largely unexplored.

To summarize, existing approaches lack a systematic framework that unifies sequence embeddings with multiplex structures. Single-layer methods overlook relational heterogeneity, multiplex models underuse multimodal features, and most approaches rely on neighborhood data, with little effort devoted to exploring zero-shot generalization.

3 PROBLEM STATEMENT

Definition: Multiplex Networks. A multiplex network is a graph $\mathcal{G} = \{V, P, E_{intra}\} = \{\mathcal{G}_1, \dots, \mathcal{G}_L\}$ where V is a set of nodes across layers, P is a set of entities, and E_{intra} represents the set of intra-layer edges. Each layer $\mathcal{G}_i = \{P_i, V_i, E_i\}$ represents the graph of an interaction type in \mathcal{G} , where P_i , V_i , and E_i are respectively the entity set, node set, and edge set for layer i, for $i = 1, \dots, L$. In a multiplex network, an entity (e.g. gene, compound, protein) can be present as nodes in multiple layers, while node sets and the connectivity (edges) can differ from one layer to another. That is, $P_1, \dots, P_L \in P$ are intersecting while

 $V_1,\ldots,V_L\in V$ and $E_1,\ldots,E_L\in E$ are disjoint. Each node $v\in V_i$ corresponds to an entity $p\in P$. This allows the representation of different types of relationships or interactions between the same set of entities. An inter-layer graph $G_{inter}=(V,E_{inter})$ can be formed by connecting the same entities in different layers of \mathcal{G} .

Supra-Adjacency Matrix. Given a multiplex network \mathcal{G} , there is a set of adjacency matrices $A = \{A_1, A_2, \dots, A_L\}$ where $A_i \in \mathcal{R}^{|V_i| \times |V_i|}$ is the layer-specific adjacency matrix for layer i. A encodes the intra-layer links in \mathcal{G} . G_{inter} can be represented by an inter-layer adjacency matrix $C \in \mathcal{R}^{|V| \times |V|}$. The supra-adjacency matrix, which encodes both intra-layer and inter-layer connections, is defined as

$$\hat{A} = \bigoplus_{i \in \{1, \dots, L\}} A_i + C,\tag{1}$$

where \bigoplus denotes the direct sum. The illustration of an example multiplex network and its supra-adjacency matrix can be found in Figure 1.

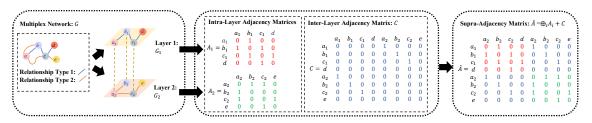


Figure 1: Illustration of an example multiplex network and its supra-adjacency matrix. A_1 and A_2 are the layer-specific adjacency matrices for layer 1 and 2, C is the inter-layer adjacency matrix, and \hat{A} is the supra-adjacency matrix that represents the example multiplex network \mathcal{G} .

Generalizing Interaction Prediction in Multiplex Networks: From Seen to Unseen Entities. Interaction prediction in multiplex networks is a multi-label classification task, since entity pairs can share multiple interaction types. Transductive settings predict missing links among seen entities, while zero-shot settings require predicting interactions for entirely unseen entities with no prior known neighborhood.

4 METHODS

The proposed CAZI-MBN framework addresses multiplex biological interaction prediction in both transductive and zero-shot settings. It integrates **domain-specific LLM embeddings** (see Appendix A) with **topology-aware network representations**. A **Context-Aware Enhancement (CAE)** module (see Appendix C) captures inter-layer dependencies, while a **MoE** model enables multi-label prediction and knowledge distillation supports zero-shot inference. The workflow is shown in Figure 2.

Training of the teacher model in CAZI-MBN is guided by a hybrid loss function:

$$\mathcal{L} = \mathcal{L}_{disc} + \mathcal{L}_{reg} + \mathcal{L}_{cls} + \beta \|\Theta\|^2, \tag{2}$$

where \mathcal{L}_{disc} and \mathcal{L}_{reg} are the discriminator's loss and the consensus regularizer's loss from the CAE, \mathcal{L}_{cls} is a multi-label soft margin loss (see Appendix B), and β is the coefficient for l2 regularization applied to the trainable parameters Θ . The student model is trained with the sum of a Mean Square Error (MSE) distillation loss $\mathcal{L}distill$ and \mathcal{L}_{cls} .

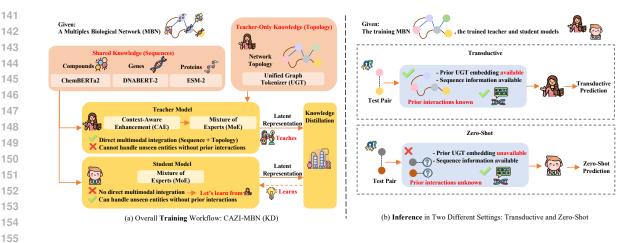


Figure 2: (a) The overall training workflow of CAZI-MBN. (b) Two inference settings in CAZI-MBN: Transductive and zero-shot.

4.1 FEATURE REPRESENTATION

To represent the multimodal nature of MBN data, our approach integrates two complementary sources of information: sequence-based embeddings derived from **domain-specific pre-trained LLMs** and topology-based representations generated by a **Unified Graph Tokenizer (UGT)**.

4.1.1 SEQUENCE EMBEDDING: DOMAIN-SPECIFIC LLMS

We use three pre-trained LLMs for sequence embeddings: ChemBERTa-2 (Ahmad et al., 2022) for drug/metabolite SMILES, DNABERT-2 (Zhou et al., 2023) for gene sequences, and ESM-2 (Lin et al., 2023) for proteins.

4.1.2 Unified Graph Tokenizer (UGT)

The UGT generates generates topology, multiplexity, and high-order connectivity-aware node embedding from the supra-adjacency matrix \hat{A} by (1) constructing a smoothed high-order matrix $\tilde{A}=\bar{A}+\bar{A}^2+\cdots+\bar{A}^O$, with $\bar{A}=D^{-1/2}\hat{A}D^{-1/2}$, and (2) projecting topology-aware embeddings. For node $v\in V$:

$$e_v = \tilde{A}_{v,:}, \text{LN}(U\sqrt{\Sigma} \parallel V\sqrt{\Sigma}), \quad e_v \in \mathbb{R}^{d_u},$$
 (3)

where $U, V \in \mathbb{R}^{|V| \times d_u}$ and $\Sigma \in \mathbb{R}^{d_u \times d_u}$ are from the SVD of \tilde{A} .

4.2 CONTEXT-AWARE ENHANCEMENT (CAE)

The CAE module refines multiplex embeddings through node- and layer-level inter-layer attention combined with contrastive learning (Figure 3 and Appendix C.). Each layer is encoded with a Graph Transformer, and inter-layer attention enables nodes to adaptively attend to counterparts across interaction types. A contrastive learning framework aligns embeddings by training a discriminator to distinguish real from perturbed edges, while a consensus regularizer learns a shared embedding that maximizes agreement across true layers and minimizes alignment with negative views. This design enhances generalization in complex biological systems

by improving representation consistency and contextual relevance across the multiplex structure. A CAE forward pass is outlined in Algorithm 1 in Appendix C.

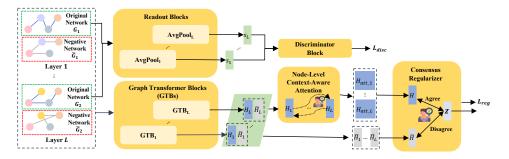


Figure 3: Illustration of the CAE module.

4.2.1 Node-Level Context-Aware Inter-Layer Attention

Node-level context-aware inter-layer attention enables CAZI-MBN to adaptively weight an entity's representations across layers. The weight $a_n^{(p \leftarrow q)}$ quantifies how much node n in layer p attends to its counterpart in layer q.

$$a_n^{(p \leftarrow q)} = \operatorname{softmax} \left(\frac{\sigma \left(\theta^{(p)} \cdot \left(H_n^{(p)} \otimes H_n^{(q)} \right) \right)}{\sum_{l=1, l \neq p}^{|L|} \sigma \left(\theta^{(p)} \cdot \left(H_n^{(p)} \otimes H_n^{(l)} \right) \right)} \right), \tag{4}$$

where $\theta^{(p)}$ is a learnable weight vector, and $\sigma(\cdot)$ is a sigmoid activation function.

4.2.2 LAYER-LEVEL CONTEXT-AWARE CONSENSUS EMBEDDING

To capture the varying importance of interaction types, CAZI-MBN applies an attention module that aggregates layer-specific node features into H and its negative counterpart \tilde{H} . A context-aware contrastive learning-based consensus regularizer then learns a consensus embedding $Z \in \mathbb{R}^{|P| \times d}$ by maximizing agreement with H and minimizing disagreement with \tilde{H} , using a cosine-similarity loss.

4.3 MIXTURE OF EXPERTS (MOE)

In MBNs, interaction prediction is a multi-label classification task since entity pairs can share multiple interaction types (Figure 2). To handle label sparsity and interdependence, we use a MoE framework where each expert captures distinct interaction patterns, and a gating network assigns weights based on the input. For embedding \mathbf{h} , experts $f_k(\mathbf{h})k = 1^K$ produce logits, with weights $\mathbf{a} = \operatorname{softmax}(W_g\mathbf{h} + b_g)$; the final prediction is $\hat{\mathbf{Y}} = \sum k = 1^K \mathbf{a}_k f_k(\mathbf{h})$.

4.4 KNOWLEDGE DISTILLATION

We apply knowledge distillation in CAZI-MBN to enable zero-shot interaction prediction (Figure 2). The teacher leverages sequence and topology embeddings but depends on neighborhood context, while the

student is topology-agnostic and relies only on sequence data. By minimizing the MSE between their latent representations, the student approximates the teacher's richer features, enabling effective zero-shot inference.

5 EXPERIMENTS

5.1 EXPERIMENT SETUP

5.1.1 Datasets

To our knowledge, no standardized benchmark datasets currently exist for evaluating models in the context of MBNs. To address this gap and provide reliable, well-validated benchmarks, we curated five MBNs from high-quality data sources (Cannon et al., 2024; Zdrazil et al., 2024; Li et al., 2024; Peng et al., 2025; Han et al., 2018) published in high-profile journals. A summary of the MBNs is provided in Table 1, with details on data collection and preprocessing available in Appendix D.

Table 1: Dataset summary (ITs: Interaction Types)

	Context	# of Nodes	ITs	# of ITs
DGIdb	Drug-Gene	1846	Interaction mechanisms	5
ChEMBL	Compound-Bacteria	9368	Response phenotypes	3
PINNACLE	Protein-Protein	7044	Interaction in different cell types	12
MetaConserve	Protein-Metabolite	329	Conservation in bacteria strains	4
TRRUST	Gene-Gene	2,862	Gene regulartory roles	3

5.1.2 BASELINES AND EVALUATION

We benchmark CAZI-MBN against 11 models across four categories: **sequence-based** (Chen & Guestrin, 2016), **single-graph** (Kipf & Welling, 2016; Hamilton et al., 2017; Shi et al., 2020; Veličković et al., 2018), **multiplex graph-based** (Boccaletti et al., 2014; Park et al., 2020; Jing et al., 2021), and **domain-specific** (Rao et al., 2022; Tao et al., 2023; Chen et al., 2019; Dang & Vu, 2024; Bai et al., 2023; Huang et al., 2021; Cui et al., 2022; Fakhry et al., 2023) (Table 2). Since most are not designed for multiplex networks, we train a separate classifier per interaction type. For zero-shot comparison, we also include knowledge-distilled variants of models lacking native zero-shot capability.

Table 2: Benchmark models

Type	Benchmark Models
Sequence-based Single graph-based	MultiLayer Perceptron (MLP), eXtreme Gradient Boosting (XGB) GCN, GraphSAGE, Graph Transformer, DGI
Multiplex graph-based	MultiplexSAGE, DMGI, HDMI
Domain-Specific	 (1) Drug-Gene & Compoud-Bacteria: CoSMIG, DGCL (2) Protein-Protein: PIPR, xCAPT5 (3) Protein-Metabolite: DrugBAN, MolTrans (4) Gene-Gene: DL-GGI, GENER

We evaluate performance with Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision-Recall Curve (AUPRC), and multi-label metrics including Hamming Score (HS) and Subset Accuracy (SA). Each experiment is conducted over three repeated trials, and we report the mean and

Appendix E.1 and Appendix E.2

5.2 EMPIRICAL ANALYSIS

We conduct a comprehensive evaluation of the proposed CAZI-MBN across five benchmark MBNs, under both transductive (shown as "T" in the tables) and zero-shot (shown as "ZS" in the tables) settings.

Standard Deviation (SD) of each metric. For details of the evaluation metrics and experimental settings, see

For each of the 11 models, Tables 3, 4, 5, 6, and 7 report performance comparisons between CAZI-MBN and the strongest single graph-based, multiplex graph-based, and domain-specific benchmarks, selected based on AUROC. For detailed results of all benchmark models, refer to Appendix E.3.1. A minority interaction type-specific analysis is provided in Appendix E.3.2, with detailed ablation studies discussed in Appendix E.3.3. Figure 4 further illustrates the average performance drop across all metrics and datasets when each module (LLMs, UGT, CAE, MoE, discriminator in CAE, consensus regularizer in CAE) is removed, in both transductive and zero-shot settings. Additionally, a case study on IBD-related genes is presented in Appendix F 3 4

Table 3: Evaluation of transductive and zero-shot multiplex interaction prediction on DGIdb (top model per category).

Setting	Model	AUROC	AUPRC	HS	SA
Т	Graph Transformer HDMI DGCL	0.505 ± 0.007 0.551 ± 0.006 0.519 ± 0.005	0.514±0.005 0.557±0.007 0.531±0.004	0.493±0.004 0.540±0.012 0.527±0.007	0.508±0.003 0.511±0.006 0.512±0.006
	CAZI-MBN	0.715 ± 0.007	0.729 ± 0.009	0.687 ± 0.011	$0.684{\pm}0.015$
	Graph Transformer DMGI	0.498 ± 0.011 0.524 ± 0.008	0.502 ± 0.004 0.528 ± 0.016	0.486 ± 0.004 0.529 ± 0.004	0.504 ± 0.007 0.502 ± 0.006
ZS	DGCL	0.513 ± 0.003	0.515 ± 0.007	0.509 ± 0.011	0.502 ± 0.004
	CAZI-MBN	0.671 ± 0.008	0.709 ± 0.011	0.688±0.009	0.663±0.014

^{*}See Table 9 in Appendix E.3.1 for full results of all evaluated models.

Table 4: Evaluation of transductive and zero-shot multiplex interaction prediction on ChEMBL (top model per category).

Setting	Model	AUROC	AUPRC	HS	SA
	DGI	0.651±0.019	0.719 ± 0.005	0.738 ± 0.012	0.673±0.011
	HDMI	0.663 ± 0.012	0.762 ± 0.014	0.789 ± 0.012	0.730 ± 0.006
T	CoSMIG	0.661 ± 0.021	0.732 ± 0.011	0.755 ± 0.009	0.702 ± 0.025
	CAZI-MBN	0.812±0.008	0.863±0.006	0.889±0.014	0.757±0.011
	Graph Transformer	0.628 ± 0.011	0.720 ± 0.008	0.725 ± 0.013	0.628 ± 0.014
	DMGI	0.652 ± 0.021	0.745 ± 0.015	0.756 ± 0.007	0.711 ± 0.009
ZS	CoSMIG	0.641 ± 0.011	0.733 ± 0.007	0.729 ± 0.010	0.683 ± 0.005
	CAZI-MBN	0.791±0.018	0.839±0.015	0.857±0.011	0.723±0.009

^{*}See Table 10 in Appendix E.3.1 for full results of all evaluated models.

We note the following observations: (1) **CAZI-MBN consistently outperforms all benchmarks in AUROC** and **AUPRC**, achieving gains of 3.1–20.4% across five MBN datasets in both transductive and zero-shot

Table 5: Evaluation of transductive and zero-shot multiplex interaction prediction on PINNACLE (top model per category).

Setting	Model	AUROC	AUPRC	HS	SA
	Graph Transformer	0.701 ± 0.007	0.704 ± 0.015	0.700 ± 0.005	0.726 ± 0.024
	HDMI	0.773 ± 0.022	0.798 ± 0.023	0.776 ± 0.017	0.766 ± 0.016
T	xCAPT5	0.781 ± 0.011	$0.804 {\pm} 0.014$	0.726 ± 0.012	0.752 ± 0.015
	CAZI-MBN	0.831 ± 0.018	$0.845 {\pm} 0.011$	0.751 ± 0.009	0.772 ± 0.013
	Graph Transformer	0.687 ± 0.019	0.690 ± 0.017	0.680 ± 0.018	0.705 ± 0.019
	HDMI	0.748 ± 0.025	0.776 ± 0.021	0.757 ± 0.024	0.742 ± 0.014
ZS	xCAPT5	0.785 ± 0.012	0.791 ± 0.015	0.733 ± 0.013	0.739 ± 0.016
	CAZI-MBN	0.812 ± 0.013	0.820 ± 0.008	0.748 ± 0.010	0.763 ± 0.011

^{*}See Table 11 in Appendix E.3.1 for full results of all evaluated models.

Table 6: Evaluation of transductive and zero-shot multiplex interaction prediction on MetaConserve (top model per category).

Setting	Model	AUROC	AUPRC	HS	SA
	DGI	0.668 ± 0.020	0.673 ± 0.019	0.665 ± 0.010	0.669 ± 0.010
	HDMI	0.726 ± 0.015	0.729 ± 0.021	0.715 ± 0.006	0.718 ± 0.015
T	DrugBAN	0.737 ± 0.010	0.714 ± 0.012	0.710 ± 0.019	0.711 ± 0.008
	CAZI-MBN	0.752 ± 0.020	$0.744 {\pm} 0.012$	0.779 ± 0.008	$0.656 {\pm} 0.014$
	Graph Transformer	0.651 ± 0.007	0.660 ± 0.020	0.652 ± 0.015	0.648 ± 0.015
	HDMI	0.709 ± 0.017	0.710 ± 0.014	0.704 ± 0.014	0.707 ± 0.009
ZS	DrugBAN	0.692 ± 0.017	0.705 ± 0.020	0.707 ± 0.015	0.708 ± 0.011
	CAZI-MBN	0.738±0.015	0.722±0.010	0.749±0.020	0.653±0.009

^{*}See Table 12 in Appendix E.3.1 for full results of all evaluated models.

Table 7: Evaluation of transductive and zero-shot multiplex interaction prediction on TRRUST (top model per category).

Setting	Model	AUROC	AUPRC	HS	SA
	DGI	0.841 ± 0.014	0.844 ± 0.011	0.845 ± 0.009	0.847±0.006
	HDMI	0.878 ± 0.008	0.879 ± 0.014	0.883 ± 0.005	0.881 ± 0.004
T	GENER	0.867 ± 0.012	$0.868 {\pm} 0.006$	$0.865 {\pm} 0.010$	0.869 ± 0.006
	CAZI-MBN	0.905±0.013	0.872±0.008	0.791±0.023	0.784±0.015
	DGI	0.832 ± 0.011	0.835 ± 0.014	0.836 ± 0.007	0.838 ± 0.006
	HDMI	0.868 ± 0.006	0.869 ± 0.010	0.872 ± 0.006	0.870 ± 0.004
ZS	GENER	0.857 ± 0.007	$0.858 {\pm} 0.008$	$0.855 {\pm} 0.006$	0.859 ± 0.007
	CAZI-MBN	0.899±0.017	0.869±0.013	0.775±0.019	0.764±0.007

^{*}See Table 13 in Appendix E.3.1 for full results of all evaluated models.

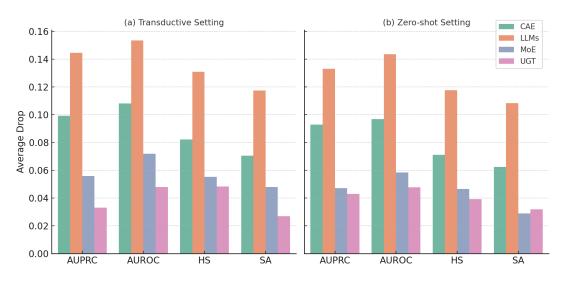


Figure 4: Average metric-wise performance drop across all five datasets when each module is ablated.

settings, highlighting strong generalization. (2) **Multiplex graph-based and domain-specific models** (e.g., DMGI, HDMI, CoSMIG, xCAPT5, GENER) generally surpass single-graph baselines, emphasizing the value of structural and biochemical context. (3) **Interaction type-specific analysis** (Appendix E.3.2) shows CAZI-MBN excels on minority interaction types in zero-shot settings, demonstrating effective generalization in data-scarce cases. (4) **Ablation studies** (Appendix E.3.3 and Figure 4) reveal that LLMs contribute the largest gains (15–20% AUROC), followed by CAE (7–10%) and UGT/MoE (5–8%), validating the modular design. (5) **Case study** (Table 17 in Appendix E.3.4) highlights CAZI-MBN's biological validity, recovering a high proportion of known interactions (82.7% DGIdb, 85.7% TRRUST, 75.1% PINNACLE, 62.5% MetaConserve) and literature-supported IBD-related links such as GAPDH–Omigapil (Foley et al., 2024; Zhou et al., 2015; Özsoy et al., 2022), thereby demonstrating biologically coherent predictions in a zero-shot setting.

6 CONCLUSION, LIMITATIONS, AND FUTURE WORK

We present CAZI-MBN, a framework for interaction prediction in multiplex biological networks (MBNs). It integrates domain-specific LLMs, multiplex-aware embedding, context-aware attention with contrastive learning, and knowledge distillation to enable zero-shot prediction. Empirically, CAZI-MBN achieves strong performance on multiplex interactions and demonstrates robust generalization to unseen entities.

To our knowledge, this is the first framework specifically designed for zero-shot multiplex interaction prediction. To support reliable evaluation, we also curated five high-quality MBN datasets, addressing the lack of standardized benchmarks. CAZI-MBN shows promise for accelerating drug discovery and antibiotic research by predicting interactions involving novel biochemical entities, and its flexible design is readily applicable to other biomedical domains.

Nonetheless, the framework has limitations. In particular, it does not yet incorporate structural data (e.g., genomic, protein, or compound 3D information), which constrains fine-grained biological modeling. Future work will integrate such structural information and extend the framework to more complex settings such as cross-species networks.

ETHICS STATEMENT

This work introduces a framework for predicting interactions in multiplex biological networks, with potential to accelerate discovery in areas such as drug development and precision medicine. We identify no direct ethical risks and are committed to responsible dissemination, supporting the ethical use of biological data and AI technologies.

REPRODUCIBILITY STATEMENT

To ensure reproducibility, the source code of CAZI-MBN is provided in the supplementary materials and will be made public upon publication. Details on dataset processing are given in Appendix C, and the experimental and evaluation setup is described in Appendix E.1 and Appendix E.2.

REFERENCES

- Walid Ahmad, Elana Simon, Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta-2: Towards chemical foundation models. *arXiv* preprint arXiv:2209.01712, 2022.
- F Algieri, Al Rodriguez-Nogales, Jose Garrido-Mesa, Deseada Camuesco, Teresa Vezza, Natividad Garrido-Mesa, Pilar Utrilla, María Elena Rodriguez-Cabezas, Ivo Pischel, and Julio Galvez. Intestinal anti-inflammatory activity of calcium pyruvate in the tnbs model of rat colitis: Comparison with ethyl pyruvate. *Biochemical Pharmacology*, 103:53–63, 2016.
- Peizhen Bai, Filip Miljković, Bino John, and Haiping Lu. Interpretable bilinear attention network with domain adaptation improves drug-target prediction. *Nature Machine Intelligence*, 5(2):126–136, 2023.
- Robert D Barber, Dan W Harmer, Robert A Coleman, and Brian J Clark. Gapdh as a housekeeping gene: analysis of gapdh mrna expression in a panel of 72 human tissues. *Physiological genomics*, 21(3):389–395, 2005.
- Anna Bas, Göte Forsberg, Sten Hammarström, and M-L Hammarström. Utility of the housekeeping genes 18s rrna, β -actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human t lymphocytes. *Scandinavian journal of immunology*, 59(6):566–573, 2004.
- Stefano Boccaletti, Ginestra Bianconi, Regino Criado, Charo I Del Genio, Jesús Gómez-Gardenes, Miguel Romance, Irene Sendina-Nadal, Zhen Wang, and Massimiliano Zanin. The structure and dynamics of multilayer networks. *Physics reports*, 544(1):1–122, 2014.
- Simon Bos and Debby Laukens. Metabolic modulation during intestinal fibrosis. *Journal of Digestive Diseases*, 21(6):319–325, 2020.
- Matthew Cannon, James Stevenson, Kathryn Stahl, Rohit Basu, Adam Coffman, Susanna Kiwala, Joshua F McMichael, Kori Kuzma, Dorian Morrissey, Kelsy Cotto, et al. Dgidb 5.0: rebuilding the drug–gene interaction database for precision medicine and drug discovery platforms. *Nucleic acids research*, 52(D1): D1227–D1235, 2024.
- Seow-Neng Chan, Eden Ngah Den Low, Raja Affendi Raja Ali, and Norfilza Mohd Mokhtar. Delineating inflammatory bowel disease through transcriptomic studies: current review of progress and evidence. *Intestinal Research*, 16(3):374, 2018.

Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1):67, 2023.

 Guobin Chen, Wongun Choi, Xiang Yu, Tony Han, and Manmohan Chandraker. Learning efficient object detection models with knowledge distillation. *Advances in neural information processing systems*, 30, 2017.

Jiaoyan Chen, Hang Dong, Janna Hastings, Ernesto Jiménez-Ruiz, Vanessa López, Pierre Monnin, Catia Pesquita, Petr Škoda, and Valentina Tamma. Knowledge graphs for the life sciences: Recent developments, challenges and opportunities. *arXiv preprint arXiv:2309.17255*, 2023.

Muhao Chen, Chelsea J-T Ju, Guangyu Zhou, Xuelu Chen, Tianran Zhang, Kai-Wei Chang, Carlo Zaniolo, and Wei Wang. Multifaceted protein–protein interaction prediction based on siamese residual rcnn. *Bioinformatics*, 35(14):i305–i314, 2019.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta: large-scale self-supervised pretraining for molecular property prediction. arXiv preprint arXiv:2010.00885, 2020.

pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885*, 2020.

Tianyu Cui, Khaoula El Mekkaoui, Jaakko Reinvall, Aki S Havulinna, Pekka Marttinen, and Samuel Kaski.

Gene–gene interaction detection with deep learning. Communications Biology, 5(1):1238, 2022.

Thanh Hai Dang and Tien Anh Vu. xcapt5: protein–protein interaction prediction using deep and wide multi-kernel pooling convolutional neural networks with protein language model. *BMC bioinformatics*, 25 (1):106, 2024.

Anindhya Sundar Das, Anandita Basu, and Rupak Mukhopadhyay. Ribosomal proteins: the missing piece in the inflammation puzzle? *Molecular and Cellular Biochemistry*, pp. 1–13, 2024.

Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 135–144, 2017.

Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, et al. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):7112–7127, 2021.

Ahmed Fakhry, Raneem Khafagy, and Adriaan-Alexander Ludl. Gener: A parallel layer deep learning network to detect gene-gene interactions from gene expression data. *arXiv preprint arXiv:2310.03611*, 2023.

Elena Ferrer-Picón, Isabella Dotti, Ana M Corraliza, Aida Mayorgas, Miriam Esteller, José Carlos Perales, Elena Ricart, Maria C Masamunt, Anna Carrasco, Eva Tristán, et al. Intestinal inflammation modulates the epithelial response to butyrate in patients with inflammatory bowel disease. *Inflammatory bowel diseases*, 26(1):43–55, 2020.

A Reghan Foley, Pomi Yun, Meganne E Leach, Sarah B Neuhaus, Gilberto V Averion, Ying Hu, Leslie H Hayes, Sandra Donkervoort, Minal S Jain, Melissa Waite, et al. Phase 1 open-label study of omigapil in patients with lama2-or col6-related dystrophy. *Neurology: Genetics*, 10(3):e200148, 2024.

Luca Gallo, Vito Latora, and Alfredo Pulvirenti. Multiplexsage: A multiplex embedding algorithm for inter-layer link prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

Ming Gao, Leihui Chen, Xiangnan He, and Aoying Zhou. Bine: Bipartite network embedding. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pp. 715–724, 2018.

- Alain P Gobert, Yulan Cheng, Mahmood Akhtar, Benjamin D Mersey, Darren R Blumberg, Raymond K Cross, Rupesh Chaturvedi, Cinthia B Drachenberg, Jean-Luc Boucher, Amy Hacker, et al. Protective role of arginase in a mouse model of colitis. *The Journal of Immunology*, 173(3):2109–2117, 2004.
- Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, 2016.
- Weiwei Gu, Linbi Lv, Gang Lu, and Ruiqi Li. Pay attention to weak ties: A heterogeneous multiplex representation learning framework for link prediction. *arXiv preprint arXiv:2406.11904*, 2024.
- Thomas Guerbette, Gaëlle Boudry, and Annaïg Lan. Mitochondrial function in intestinal epithelium homeostasis and modulation in diet-induced obesity. *Molecular metabolism*, 63:101546, 2022.
- Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.
- Heonjong Han, Jae-Won Cho, Sangyoung Lee, Ayoung Yun, Hyojin Kim, Dasom Bae, Sunmo Yang, Chan Yeong Kim, Muyoung Lee, Eunbeen Kim, et al. Trrust v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic acids research*, 46(D1):D380–D386, 2018.
- Geoffrey Hinton. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015.
- Kexin Huang, Cao Xiao, Lucas M Glass, and Jimeng Sun. Moltrans: molecular interaction transformer for drug-target interaction prediction. *Bioinformatics*, 37(6):830–836, 2021.
- Yanrong Ji, Zhihan Zhou, Han Liu, and Ramana V Davuluri. Dnabert: pre-trained bidirectional encoder representations from transformers model for dna-language in genome. *Bioinformatics*, 37(15):2112–2120, 2021.
- Yun Ji, Ying Yang, Shiqiang Sun, Zhaolai Dai, Fazheng Ren, and Zhenlong Wu. Insights into diet-associated oxidative pathomechanisms in inflammatory bowel disease and protective effects of functional amino acids. *Nutrition Reviews*, 81(1):95–113, 2023.
- Jie Jia, Abul Arif, Belinda Willard, Jonathan D Smith, Dennis J Stuehr, Stanley L Hazen, and Paul L Fox. Protection of extraribosomal rpl13a by gapdh and dysregulation by s-nitrosylation. *Molecular cell*, 47(4): 656–663, 2012.
- Baoyu Jing, Chanyoung Park, and Hanghang Tong. Hdmi: High-order deep multiplex infomax. In *Proceedings* of the Web Conference 2021, pp. 2414–2424, 2021.
- Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, et al. Pubchem 2023 update. *Nucleic acids research*, 51(D1): D1373–D1380, 2023.
- Sup Kim, June-Young Lee, Seul Gi Shin, Jin Kyung Kim, Prashanta Silwal, Young Jae Kim, Na-Ri Shin, Pil Soo Kim, Minho Won, Sang-Hee Lee, et al. Esrra (estrogen related receptor alpha) is a critical regulator of intestinal homeostasis through activation of autophagic flux via gut microbiota. *Autophagy*, 17(10): 2856–2875, 2021.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv* preprint arXiv:1609.02907, 2016.

- Annaïg Lan, Thomas Guerbette, Mireille Andriamihaja, Benjamin Magnin, Martin Bordet, Pierre-Jean Ferron, Agnès Burel, Roselyne Viel, Bernard Fromenty, Anne Corlu, et al. Mitochondrial remodeling and energy metabolism adaptations in colonic crypts during spontaneous epithelial repair after colitis induction in mice. *Free Radical Biology and Medicine*, 205:224–233, 2023.
- Michelle M Li, Yepeng Huang, Marissa Sumathipala, Man Qing Liang, Alberto Valdeolivas, Ashwin N Ananthakrishnan, Katherine Liao, Daniel Marbach, and Marinka Zitnik. Contextual ai models for single-cell protein biology. *Nature Methods*, 21(8):1546–1557, 2024.
- Yilong Lin, Wanna Xiong, Simeng Xiao, Fang Li, Zhuo Lu, Jiongyi Yan, Xuewan Fang, Xiaojuan Cui, Yilei Wen, Jianqin Liang, et al. Pharmacoproteomics reveals the mechanism of chinese dragon's blood in regulating the rsk/tsc2/mtor/ribosome pathway in alleviation of dss-induced acute ulcerative colitis. *Journal of ethnopharmacology*, 263:113221, 2020.
- Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.
- Sameh K Mohamed, Aayah Nounu, and Vít Nováček. Biological applications of knowledge graph embedding models. *Briefings in bioinformatics*, 22(2):1679–1693, 2021.
- Mihriban Özsoy, Nathalie Stummer, Franz A Zimmermann, René G Feichtinger, Wolfgang Sperl, Daniel Weghuber, and Anna M Schneider. Role of energy metabolism and mitochondrial function in inflammatory bowel disease. *Inflammatory Bowel Diseases*, 28(9):1443–1450, 2022.
- Chanyoung Park, Donghyun Kim, Jiawei Han, and Hwanjo Yu. Unsupervised attributed multiplex network embedding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 5371–5378, 2020.
- Hui Peng, Sergei Kotelnikov, Megan E Egbert, Shany Ofaim, Grant C Stevens, Sadhna Phanse, Tatiana Saccon, Mikhail Ignatov, Shubham Dutta, Zoe Istace, et al. Ligand interaction landscape of transcription factors and essential enzymes in e. coli. *Cell*, 2025.
- Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710, 2014.
- Jiahua Rao, Shuangjia Zheng, Sijie Mai, and Yuedong Yang. Communicative subgraph representation learning for multi-relational inductive drug-gene interaction prediction. *arXiv* preprint arXiv:2205.05957, 2022.
- Eva Rath, Antonio Moschetta, and Dirk Haller. Mitochondrial function—gatekeeper of intestinal epithelial cell homeostasis. *Nature reviews Gastroenterology & hepatology*, 15(8):497–516, 2018.
- Leonardo FR Ribeiro, Pedro HP Saverese, and Daniel R Figueiredo. struc2vec: Learning node representations from structural identity. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 385–394, 2017.
- David A Sallman and Alan List. The central role of inflammatory signaling in the pathogenesis of myelodysplastic syndromes. *Blood, The Journal of the American Society of Hematology*, 133(10):1039–1048, 2019.

- V Sanh. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.
 - Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.
 - Wei-Yun Sheng and Tzu-Chien V Wang. Proteomic analysis of the differential protein expression reveals nuclear gapdh in activated t lymphocytes. *PLoS One*, 4(7):e6322, 2009.
 - Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjin Wang, and Yu Sun. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv* preprint arXiv:2009.03509, 2020.
 - Sarah A Smith, Sayaka A Ogawa, Lillian Chau, Kelly A Whelan, Kathryn E Hamilton, Jie Chen, Lu Tan, Eric Z Chen, Sue Keilbaugh, Franz Fogt, et al. Mitochondrial dysfunction in inflammatory bowel disease alters intestinal epithelial metabolism of hepatic acylcarnitines. *The Journal of clinical investigation*, 131 (1), 2021.
 - Qiyun Tang, Qianqian Ji, Weiwei Xia, Liren Li, Jian'an Bai, Runzhou Ni, and Yongwei Qin. Pyruvate kinase m2 regulates apoptosis of intestinal epithelial cells in crohn's disease. *Digestive Diseases and Sciences*, 60: 393–404, 2015.
 - Wen Tao, Yuansheng Liu, Xuan Lin, Bosheng Song, and Xiangxiang Zeng. Prediction of multi-relational drug—gene interaction via dynamic hypergraph contrastive learning. *Briefings in Bioinformatics*, 24(6): bbad371, 2023.
 - Alberto Valdeolivas, Laurent Tichit, Claire Navarro, Sophie Perrin, Gaelle Odelin, Nicolas Levy, Pierre Cau, Elisabeth Remy, and Anaïs Baudot. Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics*, 35(3):497–505, 2019.
 - Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. *arXiv preprint arXiv:1809.10341*, 2018.
 - Brian Walsh, Sameh K Mohamed, and Vít Nováček. Biokg: A knowledge graph for relational learning on biological data. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 3173–3180, 2020.
 - Gaoying Wang, Jintao Yuan, Ji Luo, Dickson Kofi Wiredu Ocansey, Xu Zhang, Hui Qian, Wenrong Xu, and Fei Mao. Emerging role of protein modification in inflammatory bowel disease. *Journal of Zhejiang University-SCIENCE B*, 23(3):173–188, 2022.
 - Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.
 - Lianghao Xia, Ben Kao, and Chao Huang. Opengraph: Towards open graph foundation models. *arXiv* preprint arXiv:2403.01121, 2024.
 - Ju Xiang, Jiashuai Zhang, Ruiqing Zheng, Xingyi Li, and Min Li. Nidm: network impulsive dynamics on multiplex biological network for disease-gene prediction. *Briefings in Bioinformatics*, 22(5):bbab080, 2021.
 - Liang Yu, Mingfei Xia, and Qi An. A network embedding framework based on integrating multiplex network for drug combination prediction. *Briefings in bioinformatics*, 23(1):bbab364, 2022.

Barbara Zdrazil, Eloy Felix, Fiona Hunter, Emma J Manners, James Blackshaw, Sybilla Corbett, Marleen de Veij, Harris Ioannidis, David Mendez Lopez, Juan F Mosquera, et al. The chembl database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic acids research, 52(D1):D1180-D1192, 2024. Muhan Zhang and Yixin Chen. Link prediction based on graph neural networks. Advances in neural information processing systems, 31, 2018. Xiang Zhou, Wen-Juan Liao, Jun-Ming Liao, Peng Liao, and Hua Lu. Ribosomal proteins: functions beyond the ribosome. *Journal of molecular cell biology*, 7(2):92–104, 2015. Zhihan Zhou, Yanrong Ji, Weijian Li, Pratik Dutta, Ramana Davuluri, and Han Liu. Dnabert-2: Efficient foundation model and benchmark for multi-species genome. arXiv preprint arXiv:2306.15006, 2023. Siyi Zhu, Jiaxin Bing, Xiaoping Min, Chen Lin, and Xiangxiang Zeng. Prediction of drug-gene interaction by using metapath2vec. Frontiers in genetics, 9:248, 2018. Tao Zou and Jiyan Zhang. Diverse and pivotal roles of neddylation in metabolism and immunity. The FEBS Journal, 288(13):3884-3912, 2021.

DISTILLING AND ADAPTING: A TOPOLOGY-AWARE FRAMEWORK FOR ZERO-SHOT INTERACTION PREDICTION IN MULTIPLEX BIOLOGICAL NETWORKS A Domain-Specific Pre-trained LLMs **B** Multi-Label Soft Margin Loss (MLSML) The CAE Module **D** Datasets and Preprocessing **E** Experiments and Evaluation E.3.2 E.3.3 E.3.4 Case Study: Interactions Involving IBD-Related Genes and Proteins DOMAIN-SPECIFIC PRE-TRAINED LLMS DNABERT-2 (Zhou et al., 2023). DNABERT-2 is a domain-adapted language model designed for nu-cleotide sequences, extending the capabilities of the original DNABERT by incorporating a deeper architecture and pretraining on a larger corpus of genomic data. It tokenizes DNA sequences using k-mers and leverages the transformer architecture to capture regulatory patterns and sequence dependencies across long genomic regions. DNABERT-2 has demonstrated improved performance on a range of genomics tasks such as promoter recognition, splice site prediction, and enhancer identification. ChemBERTa-2 (Ahmad et al., 2022). ChemBERTa-2 is a transformer-based language model pre-trained on SMILES strings representing chemical structures. It extends ChEMBERTa by employing a larger parameter space and more comprehensive chemical corpora, enabling it to capture intricate molecular substructures, functional groups, and reaction patterns. By treating SMILES as a specialized language, ChEMBERTa-2 facilitates downstream tasks such as molecular property prediction, reaction classification, and drug-likeness

assessment. Its embeddings serve as chemically informed representations that integrate both syntactic and structural aspects of small molecules.

ESM-2 (**Lin et al., 2023**). ESM-2 is a protein language model developed by Meta AI, pre-trained on hundreds of millions of protein sequences using a masked language modeling objective. It models the grammar of amino acid sequences and captures biologically relevant features such as structural motifs, binding sites, and evolutionary conservation without explicit supervision. Compared to its predecessor, ESM-2 offers deeper transformer layers and larger-scale training, enabling superior performance across protein structure prediction, variant effect analysis, and protein-protein interaction tasks. Its embeddings encode rich evolutionary and functional signals useful for downstream applications in proteomics and systems biology.

B MULTI-LABEL SOFT MARGIN LOSS (MLSML)

The MLSML is defined as:

$$\mathcal{L}_{cls} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N_C} y_{ij} \cdot \log(\sigma(x_{ij})) + (1 - y_{ij}) \cdot \log(1 - \sigma(x_{ij})), \tag{5}$$

where N is the number of samples, N_C is the number of classes, y_{ij} is a binary indicator for the presence of class j in sample i, x_{ij} is the raw model output for class j of sample i, and σ is the sigmoid function.

C THE CAE MODULE

A CAE forward pass is outlined in Algorithm 1. For each multiplex layer G_i , a Graph Transformer encodes node features X_i into embeddings H_i . Negative sampling generates perturbed graphs \tilde{G}_i and corresponding embeddings \tilde{H}_i . Context-aware node-level inter-layer attention refines H_i into H_{att_i} , from which edge features are extracted. A discriminator differentiates real edges from corrupted ones. Embeddings are then aggregated across layers with attention, and a consensus regularizer produces updated embeddings Z.

The CAE module starts with layer-specific node embedding. For each multiplex layer $G_i (i \in \{1, ..., L\})$ in the multiplex network, a Graph Transformer module first updates the layer-specific node embeddings according to:

$$H_{i}^{(l+1)} = \sigma \left(\text{LN} \left(\text{MHA} \left(H_{i}^{(l)} \right) + H_{i}^{(l)} \right) \cdot W_{i}^{(l)} \right), \tag{6}$$

where LN represents layer normalization, $H_i^{(l)} \in \mathbb{R}^{|V_i| \times d}$ denotes the feature matrix of multiplex layer G_i at Graph Transformer layer l, $W_i^{(l)}$ is the layer-specific weight matrix, MHA performs multi-head self-attention over the node features, and d represents the output dimension. To generate the layer-specific negative node embedding \tilde{H}_i , negative sampling is applied to G_i , and the same Graph Transformer module as defined in Equation 6 processes the negative network \tilde{G}_i .

The edge feature matrix $H_{edge_i} \in \mathbb{R}^{|E_i| \times 2d}$ and the negative edge feature matrix $\tilde{H}_{edge_i} \in \mathbb{R}^{|E_i| \times 2d}$ are constructed by concatenating the node features of adjacent nodes within G_i and \tilde{G}_i , respectively. A high-level graph summary vector s_i is computed to encapsulate the layer-level edge information for each G_i using average pooling:

$$s_i = \sigma(\text{AvgPool}(H_{\text{edge}_i})), \quad (s_i \in \mathbb{R}^{2d})$$
 (7)

where σ is the logistic sigmoid nonlinearity.

Algorithm 1 CAE Forward Pass

799

800

801

802

803

804

805

806 807

809

810

811 812

814

815

816

817 818 819

820

821

822

823 824

825 826

827 828

829

830 831

832 833 834

835

836

837 838

839

840

841 842

843

844

845

```
Input: Feature matrices X_i, graphs G_i, size L
              for i = 1 to L do
                 \hat{G}_i \leftarrow NegSampling(G_i) {Negative sampling}
                 H_i \leftarrow GraphTransformer(X_i, G_i) {Feature matrix X_i}
                 H_{att\_i} \leftarrow ContextAwareAttention(H_i) {Context-aware inter-layer attention}
                 \tilde{H}_i \leftarrow GraphTransformer(X_i, \tilde{G}_i)
                 H_{edge_i} \leftarrow GetEdgeFeature(H_{att\_i}, G_i) {Edge embedding}
                 \tilde{H}_{edge_i} \leftarrow GetEdgeFeature(\tilde{H}_i, \tilde{G}_i) {Negative edge embedding}
                 s_i \leftarrow \sigma(AvgPool(H_{edge_i}))
                 logit_{disc_i} \leftarrow Discriminator(H_{edge_i}, H_{edge_i}, s_i)
                 \mathcal{L}_{disc_i} \leftarrow BCE(Y_{disc_i}, \text{logit}_{disc_i})
              H \leftarrow ATTENTION(H_{att\_1}, \dots, H_{att\_L})
813
              H \leftarrow ATTENTION(H_1, \dots, H_L)
              \mathcal{L}_{disc} \leftarrow SUM(\mathcal{L}_{disc_1}, \dots, \mathcal{L}_{disc_L}) {Discriminator's loss}
              Z \leftarrow ConsensusRegularizer(H, H) {Updated embeddings}
              \mathcal{L}_{reg} \leftarrow 1 + CosineSim(H, Z) - CosineSim(\tilde{H}, Z) {Regularizer's loss}
```

The layer-specific BCE loss $\mathcal{L}_{\text{disc}_i}$ can be calculated using the edge embedding matrix H_{edge_i} , its summary representation s_i , and its corresponding negative matrix H_{edge_i} :

$$\mathcal{L}_{\text{disc}_i} = \sum_{e_j \in E_i} \log \mathcal{D}(H_{\text{edge}_i, j}, s_i) + \sum_{e_q \in \hat{E}_i} \log (1 - \mathcal{D}(\tilde{H}_{\text{edge}_i, q}, s_i))$$
(8)

where $H_{\text{edge}_i,j}$ specifies the feature vector for edge e_j , \hat{E}_i represents the edge set of \tilde{G}_i , and \mathcal{D} is a discriminator function that evaluates patch-summary representation pairs. The discriminator \mathcal{D} is defined

$$\mathcal{D}(H_{\text{edge}_i,j}, s_i) = \sigma(H_{\text{edge}_i,j}M_i s_i) \tag{9}$$

where σ signifies the logistic sigmoid function, and $M_i \in \mathbb{R}^{2d \times 2d}$ is a trainable scoring matrix.

D DATASETS AND PREPROCESSING

D.1 DGIDB

The Drug-Gene Interaction Database (DGIdb) (Cannon et al., 2024) aggregates comprehensive information on drug-gene interactions and gene druggability from diverse curated sources.

The multiplex nature of DGIdb reflects the biological complexity wherein a single drug-gene pair may exhibit multiple modes of interaction. For example, a drug may function as an agonist under certain conditions, but act as an inhibitor, modulator, or activator in others. These variations can arise due to factors such as dosage, cellular context, or the presence of cofactors.

During preprocessing, entries lacking interaction type annotations were removed, and semantically similar interaction types were merged to reduce redundancy. To emphasize higher-confidence associations, we excluded interactions with scores falling in the lowest 15% quantile. Furthermore, to address class imbalance, the four least prevalent interaction categories were combined into a single class, simplifying downstream

modeling and evaluation. We retained five broad mechanism categories after preprocessing: agonist, activator, modulator, inhibitor/blocker, and other.

D.2 CHEMBL

ChEMBL (Zdrazil et al., 2024) is a manually curated database of bioactive molecules with drug-like properties. In ChEMBL, different Minimum Inhibitory Concentration (MIC) values for the same compound-bacteria pair indicate that the bacterium may exhibit varying levels of susceptibility to the compound under different experimental conditions or at different times. A compound bacteria pair with multiple MIC values might show different response phenotype - Susceptible (S), Intermediate (I), or Resistant (R), based on environmental or genetic factors. This variability can inform clinicians of potential shifts in bacterial resistance over time, the need for adjusted drug dosages, or the importance of considering strain-specific responses when choosing treatment strategies.

We preprocess ChEMBL into a multiplex network by (1) standardizing the standard units to "ug.mL-1"; (2) defining cutoffs for S, I, and R:

- Susceptible (S): MIC ≤ 4 ;
- Intermediate (I): $4 < MIC \le 16$;
- Resistant (R): MIC < 16.

D.3 PINNACLE

We curated the Protein Network-based Algorithm for Contextual Learning (PINNACLE) dataset from (Li et al., 2024). The original dataset captures context-aware protein-protein interactions across 156 cell type contexts spanning 62 tissues, derived from the Tabula Sapiens single-cell transcriptomic atlas. Protein-protein interactions can differ across cell types because protein expression levels, post-translational modifications, subcellular localizations, and interacting partners vary depending on the cellular environment. A protein may be active in one cell type but absent or inactive in another, or it may participate in different pathways depending on the cell's functional role. These context-dependent variations are critical for understanding cell-specific mechanisms of health and disease.

For our study, we selected 12 cell types specifically associated with Inflammatory Bowel Disease (IBD) to construct a cell type-specific multiplex protein-protein interaction network. This enables us to investigate cell-type-dependent protein interactions relevant to IBD pathogenesis and therapeutic targeting, capturing the cellular heterogeneity of intestinal inflammation.

We selected the following cell types based on their established roles in IBD:

- Enterocytes (Enterocyte of epithelium of large intestine and Enterocyte of epithelium of small intestine): These epithelial cells form the intestinal barrier and regulate nutrient absorption and immune signaling, both of which are disrupted in IBD.
- Goblet cells (Small intestine goblet cell and Large intestine goblet cell): Goblet cells produce mucins that maintain the protective mucus layer; their depletion is linked to barrier dysfunction and increased inflammation in IBD.
- Paneth cells (Paneth cell of epithelium of small intestine and Paneth cell of epithelium of large intestine): These cells secrete antimicrobial peptides and are often dysregulated or dysfunctional in inflamed or chronically affected IBD tissue.
- CD4⁺ helper T cells (CD4-positive helper T cell): Key mediators of adaptive immunity and cytokine signaling, CD4⁺ T cells are known to drive pathogenic responses in IBD.

- CD8+ cytotoxic T cells (CD8-positive alpha-beta cytotoxic T cell): These cells contribute to epithelial damage by directly killing intestinal epithelial cells in active IBD.
- **Regulatory T cells** (**Tregs**) (*Regulatory T cell*): Tregs suppress immune responses and maintain tolerance; impaired Treg function is frequently observed in IBD patients.
- **Memory T cells** (*CD4-positive alpha-beta memory T cell*): These cells retain antigen-specific memory and are implicated in the persistent immune activation characteristic of chronic IBD.
- **B cells** (*B cell*): B cells contribute to antigen presentation and cytokine production and can influence both protective and pathogenic pathways in IBD.
- **Memory B cells** (*Memory B cell*): These cells are involved in long-term humoral responses and are elevated in IBD-associated lymphoid aggregates and inflamed mucosa.

D.4 METACONSERVE

We curated a cell type-specific protein—metabolite MBN dataset, which we named MetaConserve, from (Peng et al., 2025), which systematically mapped high-confidence ligand interactions in *Escherichia coli* using affinity purification mass spectrometry and structural modeling. This dataset provides a valuable foundation for exploring metabolite-mediated regulatory mechanisms at the host–microbe interface.

To investigate the evolutionary conservation of these protein–metabolite interactions in the context of IBD, we selected four representative bacterial orders: *Campylobacterales*, *Desulfovibrionales*, *Fusobacteriales*, and *Veillonellales*, all of which have been implicated in IBD pathogenesis. *Campylobacterales* (e.g., *Campylobacter concisus*) are frequently enriched in inflamed mucosa and are associated with epithelial barrier disruption and proinflammatory signaling. *Desulfovibrionales* are sulfate-reducing bacteria that produce hydrogen sulfide, a genotoxic metabolite elevated in IBD patients and implicated in mucosal injury. *Fusobacteriales*, particularly *Fusobacterium nucleatum*, are linked to mucosal inflammation and have been shown to exacerbate disease through immune modulation and epithelial adhesion. *Veillonellales* are enriched in dysbiotic gut microbiota and are associated with IBD disease activity and disrupted bile acid metabolism.

We leveraged conservation scores provided in the original dataset to identify protein–metabolite interactions that are preserved across these four orders, retaining only those with a conservation score ≥ 2 to ensure high-confidence evolutionary retention. These conserved interactions define four distinct interaction types in our multiplex network, enabling a strain-aware investigation of microbial metabolic activity relevant to IBD onset and progression.

D.5 TRRUST

The Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining (TRRUST) (Han et al., 2018) dataset is a curated database of human and mouse transcriptional regulatory interactions. It compiles experimentally validated regulatory relationships between transcription factors (TFs) and their target genes, extracted through literature mining and manual curation. Each entry in TRRUST specifies the regulator (TF), the target gene, the mode of regulation (activation or repression), and the supporting evidence. TRRUST is widely used in systems biology and regulatory network analysis, offering a high-confidence resource for studying gene regulatory mechanisms, constructing transcriptional networks, and evaluating computational predictions of TF-target interactions.

In our experiments, we use the human transcriptional regulatory data from TRRUST. As a gene can exert multiple regulatory effects on another gene, such as activation, repression, or an undefined role, and the resulting network is modeled as a multiplex graph. This structure captures the complexity of transcriptional regulation, where interactions may not be limited to a single regulatory type and can reflect more nuanced or context-dependent mechanisms.

E EXPERIMENTS AND EVALUATION

E.1 EVALUATION METRICS

Let there be N samples and N_c possible labels. We denote the ground-truth labels for sample i by a binary vector $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,N_C}) \in \{0,1\}^{N_C}$, and the predicted labels by $\hat{\mathbf{y}}_i = (\hat{y}_{i,1}, \hat{y}_{i,2}, \dots, \hat{y}_{i,N_C})$. Below, we briefly define several common evaluation metrics.

Area Under the ROC Curve (AUROC). For each label j, we can compute the Receiver Operating Characteristic (ROC) curve by plotting the True Positive Rate (TPR) versus the False Positive Rate (FPR) at various threshold settings. The area under this curve for label j is denoted $AUROC_j$. The AUROC in multi-label classification is calculated by macro-average:

$$AUROC = \frac{1}{N_C} \sum_{j=1}^{N_C} AUROC_j.$$
 (10)

Area Under the Precision-Recall Curve (AUPRC). For each label j, the Precision-Recall curve is constructed by plotting Precision versus Recall at different thresholds. The area under this curve for label j is $AUPRC_j$. The macro-average of AUPRC for multi-label classification is:

$$AUPRC = \frac{1}{N_C} \sum_{j=1}^{N_C} AUPRC_j.$$
 (11)

Hamming Score (HS). The HS measures the fraction of correctly predicted labels across all samples and all labels. Often, the Hamming Loss (HL) is defined first:

$$HL = \frac{1}{N \cdot N_C} \sum_{i=1}^{N} \sum_{j=1}^{N_C} \mathbf{1} (y_{i,j} \neq \hat{y}_{i,j}),$$
 (12)

where $\mathbf{1}(\cdot)$ is the indicator function, which is 1 if the condition is true and 0 otherwise. The Hamming Score is then

$$HS = 1 - HL = 1 - \frac{1}{N \cdot N_C} \sum_{i=1}^{N} \sum_{j=1}^{N_C} \mathbf{1}(y_{i,j} \neq \hat{y}_{i,j}).$$
 (13)

Subset Accuracy (SA). SA, also known as Exact Match Ratio, measures the fraction of samples for which all labels match exactly. Formally,

$$SA = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\mathbf{y}_i = \hat{\mathbf{y}}_i). \tag{14}$$

Here, $\mathbf{1}(\mathbf{y}_i = \hat{\mathbf{y}}_i)$ is 1 if and only if every label is predicted correctly for sample i, and 0 otherwise.

E.2 EXPERIMENTAL SETTINGS.

We divide our dataset into training, validation, and test sets with proportions of 75%, 15%, and 15%, respectively. Each experiment is repeated three times, and we report the mean and SD of each evaluation metric.

In the transductive setting, entities in the training, validation, and test sets overlap, meaning that all nodes are present during training but the edges are partitioned. As a result, models can exploit topological context from the training graph when making predictions, and the task is to infer unseen interactions among already observed entities. This setting reflects scenarios where the biological entities are known, but their interaction space is incomplete.

In contrast, the zero-shot setting enforces a strict disjoint entity-based split, where entities in the validation and test sets are entirely excluded from the training graph. These entities do not appear as endpoints or even as neighbors during training, ensuring that the model cannot rely on structural information about them. Instead, predictions must be made solely from their intrinsic features (e.g., sequence embeddings), without any prior connectivity. The task therefore involves predicting interactions among completely novel entities, simulating realistic discovery scenarios where new drugs, genes, or proteins enter the system with no known interactions. This strict separation guarantees that no information leakage occurs between training and test sets in either setting.

Hyperparameters are selected via grid search, and models are trained using the Adam optimizer. All experiments are conducted on a single NVIDIA A100 Tensor Core GPU with 80GB of RAM.

In zero-shot setting, where traditional graph-based models cannot inherently predict interactions involving entities with no observed training interactions, we adopt knowledge-distilled variants of these models. Specifically, the original graph-based models serve as teacher models, while a two-layer MLP acts as the student model. This setup facilitates performance comparisons in zero-shot learning.

For benchmark models using node attributes, compounds are encoded with Morgan fingerprints from SMILES, and genes/genomes/proteins with k-mer features (k=6) from sequence data. The training deatils of CAZI-MBN are presented in Table 8.

Table 8: Training details of CAZI-MBN.

Component	Value/Detail
Optimizer	Adam
Learning rate	1e-4
Batch size	256
Number of epoches	10000
Early stopping	Yes
Order of UGT	5
UGT output dimension	128
Readout (CAE)	SAGPool
Latent dimension (knowledge distillation)	32

E.3 SUPPLEMENTAL EXPERIMENTAL RESULTS

E.3.1 EMPIRICAL ANALYSIS

Tables 9, 10, 11, 12, and 13 present the performance comparisons between CAZI-MBN and all benchmark models under both transductive ("T") and zero-shot ("ZS") settings across five MBNs. All metrics are reported in Mean \pm SD over three repeated trails.

Table 9: Evaluation of transductive and zero-shot multiplex interaction prediction on DGIdb.

Setting	Model	AUROC	AUPRC	HS	SA
	XGB	0.350 ± 0.010	0.389 ± 0.011	0.579 ± 0.011	0.421±0.011
	MLP	$0.284{\pm}0.013$	0.375 ± 0.005	0.609 ± 0.009	0.407 ± 0.011
	GCN	0.482 ± 0.010	0.493 ± 0.005	0.511 ± 0.003	0.491±0.003
	GraphSAGE	0.447 ± 0.005	0.467 ± 0.003	0.527 ± 0.006	0.473 ± 0.006
	Graph Transformer	0.505 ± 0.007	0.514 ± 0.005	0.493 ± 0.004	0.508 ± 0.003
T	DGI	0.503 ± 0.002	0.507 ± 0.004	0.510 ± 0.003	0.500 ± 0.003
•	MultiplexSAGE	0.532 ± 0.007	0.539 ± 0.004	0.549 ± 0.005	0.507 ± 0.005
	DMGI	0.547 ± 0.002	0.542 ± 0.003	0.533 ± 0.005	0.505 ± 0.002
	HDMI	0.551 ± 0.006	0.557 ± 0.007	0.540 ± 0.012	0.511 ± 0.006
	CoSMIG	0.517±0.003	0.523 ± 0.002	0.522 ± 0.003	0.509 ± 0.001
	DGCL	0.519 ± 0.005	0.531 ± 0.004	0.527 ± 0.007	0.512 ± 0.006
	CAZI-MBN	0.715 ± 0.007	0.729 ± 0.009	0.687 ± 0.011	$0.684 {\pm} 0.015$
	XGB	0.332 ± 0.013	0.361±0.008	0.513±0.013	0.401±0.005
	MLP	0.258 ± 0.012	0.307 ± 0.007	0.590 ± 0.011	0.347 ± 0.005
	GCN	0.448 ± 0.008	0.479 ± 0.004	0.502 ± 0.005	0.473 ± 0.007
	GraphSAGE	0.419 ± 0.006	0.434 ± 0.004	0.518 ± 0.011	0.464 ± 0.005
	Graph Transformer	0.498 ± 0.011	0.502 ± 0.004	0.486 ± 0.004	0.504 ± 0.007
ZS	DGI	0.489 ± 0.004	0.504 ± 0.011	0.502 ± 0.003	0.477 ± 0.006
	MultiplexSAGE	$0.518 {\pm} 0.008$	$0.520 {\pm} 0.013$	$0.537 {\pm} 0.009$	0.504 ± 0.006
	DMGI	0.524 ± 0.008	0.528 ± 0.016	0.529 ± 0.004	0.502 ± 0.006
	HDMI	0.507 ± 0.004	0.536 ± 0.012	0.530 ± 0.011	0.507 ± 0.008
	CoSMIG	0.509 ± 0.012	0.514 ± 0.005	0.516 ± 0.013	0.500 ± 0.020
	DGCL	0.513 ± 0.003	0.515 ± 0.007	0.509 ± 0.011	0.502 ± 0.004
	CAZI-MBN	0.671 ± 0.008	0.709 ± 0.011	0.688±0.009	0.663±0.014

^{*}Benchmarks are grouped into four categories: (1) Sequence-based (XGB, MLP);(2) Single graph-based (GCN, GraphSAGE, Graph Transformer, DGI); (3) Multiplex graph-based (MultiplexSAGE, DMGI, HDMI); and (4) Domain-specific (others).

1	0	8	4
1	0	8	5
1	0	8	6
1	0	8	7
1	0	8	8
1	0	8	9
1	0	9	0
	0		
1	0	9	2
1	0	9	3
1	0	9	4
1	0	9	5
1	0	9	6
	0		
1	0	9	8
	0		
	1		
	1		
	1		
	1		
	1		
1	1	0	5
1	1	0	6
	1		
	1		
1	1	0	9
	1		
	1		
	1		
	1		
		1	
	1		
	1		
		1	
1			8
			9
1	1	2	0
1	1	2	1
	1		
	1		
	1		
	1		
1	1	2	6

Table 10:	Table 10: Evaluation of transductive and zero-shot multiplex interaction prediction on ChEMBL.				
Setting	Model	AUROC	AUPRC	HS	SA
	XGB	0.515±0.013	0.607 ± 0.018	0.699 ± 0.007	0.587 ± 0.012
	MLP	0.531 ± 0.005	0.629 ± 0.012	0.657 ± 0.009	0.613 ± 0.010
	GCN	$0.632 {\pm} 0.014$	0.700 ± 0.008	0.709 ± 0.004	0.641 ± 0.012
	GraphSAGE	0.631 ± 0.005	0.698 ± 0.012	0.734 ± 0.006	0.663 ± 0.005
	Graph Transformer	0.644 ± 0.015	0.715 ± 0.010	0.741 ± 0.005	0.657 ± 0.012
T	DGI	0.651 ± 0.019	0.719 ± 0.005	0.738 ± 0.012	0.673 ± 0.011
-	MultiplexSAGE	0.653 ± 0.007	0.709 ± 0.011	0.712 ± 0.005	0.705 ± 0.019
	DMGI	0.661 ± 0.017	0.749 ± 0.008	0.773 ± 0.021	0.712 ± 0.014
	HDMI	0.663 ± 0.012	0.762 ± 0.014	0.789 ± 0.012	0.730 ± 0.006
	CoSMIG	0.661 ± 0.021	$0.732 {\pm} 0.011$	0.755 ± 0.009	0.702 ± 0.025
	DGCL	0.659 ± 0.014	0.712 ± 0.005	0.739 ± 0.007	0.698 ± 0.0709
	CAZI-MBN	0.812 ± 0.008	$0.863 {\pm} 0.006$	$0.889 {\pm} 0.014$	0.757 ± 0.011
	XGB	0.481 ± 0.023	0.542 ± 0.012	0.608 ± 0.014	0.574 ± 0.009
	MLP	0.487 ± 0.012	0.539 ± 0.006	0.621 ± 0.012	0.581 ± 0.018
	GCN	0.612 ± 0.013	0.702 ± 0.007	0.633 ± 0.006	0.607 ± 0.011
	GraphSAGE	0.624 ± 0.009	0.713 ± 0.018	0.640 ± 0.006	0.611 ± 0.008
	Graph Transformer	0.628 ± 0.011	0.720 ± 0.008	0.725 ± 0.013	0.628 ± 0.014
ZS	DGI	0.619 ± 0.010	0.709 ± 0.022	0.708 ± 0.007	0.634 ± 0.019
	MultiplexSAGE	$0.638 {\pm} 0.007$	0.718 ± 0.012	0.711 ± 0.008	0.699 ± 0.009
	DMGI	0.652 ± 0.021	0.745 ± 0.015	0.756 ± 0.007	0.711 ± 0.009
	HDMI	0.648 ± 0.015	0.735 ± 0.011	0.726 ± 0.004	0.716 ± 0.011
	CoSMIG	0.641 ± 0.011	0.733 ± 0.007	0.729 ± 0.010	0.683 ± 0.005
	DGCL	0.631 ± 0.019	0.702 ± 0.018	0.693 ± 0.007	0.675 ± 0.012
	CAZI-MBN	0.791±0.018	0.839±0.015	0.857±0.011	0.723±0.009

Setting	Model	AUROC	AUPRC	HS	SA
	XGB	0.650±0.007	0.644 ± 0.007	0.651±0.005	0.660 ± 0.023
	MLP	0.676 ± 0.008	$0.668 {\pm} 0.008$	0.656 ± 0.011	0.666 ± 0.01
	GCN	0.659 ± 0.010	0.670 ± 0.012	0.666 ± 0.014	0.679 ± 0.02
	GraphSAGE	0.692 ± 0.017	0.677 ± 0.008	$0.686 {\pm} 0.005$	0.693 ± 0.02
	Graph Transformer	0.701 ± 0.007	0.704 ± 0.015	0.700 ± 0.005	0.726 ± 0.02
Т	DGI	0.688 ± 0.009	0.695 ± 0.010	0.698 ± 0.006	0.684 ± 0.02
•	MultiplexSAGE	0.752±0.020	0.775±0.022	0.763±0.009	0.749 ± 0.01
	DMGI	$0.768 {\pm} 0.021$	$0.784 {\pm} 0.025$	0.759 ± 0.014	0.762 ± 0.02
	HDMI	0.773 ± 0.022	0.798 ± 0.023	0.776 ± 0.017	0.766 ± 0.01
	PIPR	0.747±0.025	0.759±0.018	0.741±0.026	0.732 ± 0.02
	xCAPT5	$0.781 {\pm} 0.011$	0.804 ± 0.014	0.726 ± 0.012	0.752 ± 0.01
	CAZI-MBN	0.831±0.018	0.845 ± 0.011	0.751±0.009	0.772±0.01
	XGB	0.626 ± 0.023	0.628 ± 0.010	0.636 ± 0.015	0.639 ± 0.02
	MLP	0.652 ± 0.015	0.650 ± 0.017	0.626 ± 0.007	0.642 ± 0.02
	GCN	0.639 ± 0.017	0.652 ± 0.016	0.649 ± 0.012	0.654 ± 0.01
	GraphSAGE	0.673 ± 0.013	0.666 ± 0.019	0.668 ± 0.023	0.668 ± 0.01
	Graph Transformer	0.687 ± 0.019	0.690 ± 0.017	0.680 ± 0.018	0.705 ± 0.01
DGI					
75	DGI	0.674 ± 0.011	0.678 ± 0.022	0.679 ± 0.015	0.663 ± 0.02
ZS		0.674 ± 0.011 0.728 ± 0.020	0.678 ± 0.022 0.753 ± 0.019	0.679 ± 0.015 0.747 ± 0.021	
ZS	MultiplexSAGE DMGI				0.725±0.02
ZS	MultiplexSAGE	0.728±0.020	0.753±0.019	0.747±0.021	0.725 ± 0.02 0.738 ± 0.02
ZS	MultiplexSAGE DMGI	0.728±0.020 0.744±0.020	0.753±0.019 0.762±0.024	0.747±0.021 0.738±0.025	0.725 ± 0.02 0.738 ± 0.02 0.742 ± 0.01
ZS	MultiplexSAGE DMGI HDMI	0.728 ± 0.020 0.744 ± 0.020 0.748 ± 0.025	0.753±0.019 0.762±0.024 0.776±0.021	0.747±0.021 0.738±0.025 0.757±0.024	0.663±0.02 0.725±0.02 0.738±0.02 0.742±0.01 0.711±0.02 0.739±0.01

Table 12: Evaluation of transductive and zero-shot multiplex interaction prediction on MetaConserve	Table 12: Evaluation	of transductive and	l zero-shot multiple	ex interaction i	prediction on MetaConserve
---	----------------------	---------------------	----------------------	------------------	----------------------------

Setting	Model	AUROC	AUPRC	HS	SA
	XGB	0.611 ± 0.006	0.612 ± 0.023	0.617 ± 0.021	0.613 ± 0.010
	MLP	0.617 ± 0.010	0.614 ± 0.015	0.620 ± 0.017	0.628 ± 0.007
	GCN	0.630 ± 0.007	0.642 ± 0.022	$0.628 {\pm} 0.011$	0.633 ± 0.018
	GraphSAGE	0.649 ± 0.013	0.646 ± 0.011	0.644 ± 0.021	0.645 ± 0.012
T	Graph Transformer	0.663 ± 0.022	0.662 ± 0.017	0.653 ± 0.012	0.656 ± 0.007
	DGI	0.668 ± 0.020	0.673 ± 0.019	0.665 ± 0.010	0.669 ± 0.010
-	MultiplexSAGE	0.693 ± 0.012	0.705 ± 0.021	$0.678 {\pm} 0.018$	0.672 ± 0.014
	DMGI	0.717 ± 0.008	0.716 ± 0.017	0.719 ± 0.015	0.720 ± 0.013
	HDMI	0.726 ± 0.015	0.729 ± 0.021	0.715 ± 0.006	0.718 ± 0.015
	MolTrans	0.710 ± 0.014	0.720 ± 0.008	0.719 ± 0.011	0.716 ± 0.020
	DrugBAN	0.737 ± 0.010	0.714 ± 0.012	0.710 ± 0.019	0.711 ± 0.008
	CAZI-MBN	0.752 ± 0.020	$0.744 {\pm} 0.012$	0.779 ± 0.008	$0.656 {\pm} 0.014$
	XGB	0.598 ± 0.013	0.591 ± 0.015	0.600 ± 0.007	0.598 ± 0.010
	MLP	0.607 ± 0.016	0.600 ± 0.019	0.612 ± 0.010	0.621 ± 0.015
	GCN	0.617 ± 0.015	0.621 ± 0.019	0.610 ± 0.012	0.622 ± 0.013
	GraphSAGE	0.630 ± 0.010	0.633 ± 0.016	0.627 ± 0.010	0.634 ± 0.006
	Graph Transformer	0.651 ± 0.007	0.660 ± 0.020	0.652 ± 0.015	0.648 ± 0.015
ZS	DGI	0.646 ± 0.018	0.647 ± 0.018	0.638 ± 0.014	0.640 ± 0.008
	MultiplexSAGE	0.697 ± 0.015	0.692 ± 0.023	$0.685{\pm}0.011$	$0.688 {\pm} 0.012$
	DMGI	0.702 ± 0.012	0.701 ± 0.015	0.698 ± 0.017	0.695 ± 0.009
	HDMI	0.709 ± 0.017	0.710 ± 0.014	0.704 ± 0.014	0.707 ± 0.009
	MolTrans	$0.688 {\pm} 0.007$	0.700 ± 0.017	0.690 ± 0.007	0.693 ± 0.012
	DrugBAN	0.692 ± 0.017	0.705 ± 0.020	0.707 ± 0.015	0.708 ± 0.011
	CAZI-MBN	0.738±0.015	0.722 ± 0.010	0.749 ± 0.020	0.653±0.009

Table 13: Evaluation of transductive and zero-shot multiplex interaction prediction on TRRUST.						
Setting	Model	AUROC	AUPRC	HS	SA	
	XGB MLP	$0.786\pm0.006 \\ 0.796\pm0.008$	0.791 ± 0.016 0.799 ± 0.019	0.794 ± 0.005 0.801 ± 0.011	0.788 ± 0.007 0.803 ± 0.012	
	GCN	0.810 ± 0.008	0.815±0.022	0.812±0.008	0.817±0.005	
	GraphSAGE	0.822 ± 0.010	0.826 ± 0.007	0.819 ± 0.008	0.824 ± 0.010	
	Graph Transformer	0.832 ± 0.005	0.835 ± 0.015	0.837 ± 0.004	0.830 ± 0.006	
Т	DGI	0.841 ± 0.014	$0.844 {\pm} 0.011$	0.845 ± 0.009	0.847 ± 0.006	
•	MultiplexSAGE	0.860 ± 0.018	0.864 ± 0.010	0.858 ± 0.009	0.866 ± 0.005	
	DMGI	0.871 ± 0.013	0.874 ± 0.007	0.869 ± 0.008	0.872 ± 0.008	
	HDMI	0.878 ± 0.008	0.879 ± 0.014	0.883 ± 0.005	0.881 ± 0.004	
	DL-GGI	$0.857 {\pm} 0.006$	$0.862 {\pm} 0.012$	$0.854 {\pm} 0.007$	0.860 ± 0.005	
	GENER	0.867 ± 0.012	0.868 ± 0.006	0.865 ± 0.010	0.869 ± 0.006	
	CAZI-MBN	0.905 ± 0.013	0.872 ± 0.008	0.791 ± 0.023	0.784 ± 0.015	
	XGB	0.779 ± 0.012	$0.781 {\pm} 0.006$	$0.783 {\pm} 0.015$	0.777 ± 0.014	
	MLP	0.789 ± 0.010	0.790 ± 0.014	0.793 ± 0.010	0.794 ± 0.006	
	GCN	$0.801 {\pm} 0.011$	$0.804 {\pm} 0.010$	0.800 ± 0.006	$0.805 {\pm} 0.004$	
	GraphSAGE	0.812 ± 0.006	0.817 ± 0.012	0.809 ± 0.009	0.815 ± 0.007	
	Graph Transformer	0.823 ± 0.008	0.826 ± 0.016	0.828 ± 0.006	0.820 ± 0.005	
ZS	DGI	0.832 ± 0.011	0.835 ± 0.014	0.836 ± 0.007	0.838 ± 0.006	
20	MultiplexSAGE	0.850 ± 0.007	$0.854{\pm}0.010$	0.848 ± 0.004	$0.855 {\pm} 0.005$	
	DMGI	0.861 ± 0.008	0.864 ± 0.012	0.858 ± 0.005	0.860 ± 0.008	
	HDMI	0.868 ± 0.006	0.869 ± 0.010	0.872 ± 0.006	0.870 ± 0.004	
	DL-GGI	$0.848 {\pm} 0.006$	$0.852 {\pm} 0.006$	$0.844 {\pm} 0.008$	0.851 ± 0.006	
	GENER	0.857 ± 0.007	0.858 ± 0.008	0.855 ± 0.006	0.859 ± 0.007	
	CAZI-MBN	$0.899 {\pm} 0.017$	$0.869 {\pm} 0.013$	0.775 ± 0.019	0.764 ± 0.007	

E.3.2 Interaction Type-Specific Analysis.

 To assess performance at a more detailed level, we focus on minority interaction types across the five MBNs. Accuracy is used as the evaluation metric, and CAZI-MBN is compared against the top-performing benchmark in each category (sequence-based, single-graph-based, multiplex-graph-based, and domain-specific), based on their accuracy for each minority interaction type.

Figure 5 presents the interaction type-specific prediction accuracy across five datasets in the zero-shot setting. CAZI-MBN consistently outperforms the best benchmark models, demonstrating strong capability in capturing subtle class distinctions and reinforcing its suitability for zero-shot prediction in complex biological networks. This is particularly important in practical applications where rare or underrepresented interaction types may be biologically significant.

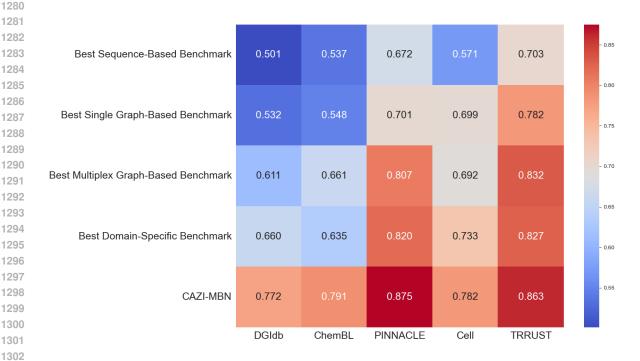


Figure 5: Interaction type-specific performance analysis across five datasets, evaluated by prediction accuracy.

E.3.3 ABLATION STUDIES

To rigorously evaluate the individual contributions of each core component in the CAZI-MBN framework, we conducted systematic ablation studies by selectively removing one module at a time while keeping the rest of the architecture intact. This approach allows us to isolate the functional impact of each component on overall model performance and to better understand how each contributes to the model's capacity to learn from multiplex biological interaction data.

A summary of the ablation configurations is provided in Table 14. The detailed results are reported in Tables 15 and 16.

Ablation studies further validate the necessity of CAZI-MBN's modular architecture. Across all datasets, CAZI-MBN consistently outperforms every ablated variant, and the most significant performance drops occur when the pre-trained LLMs or the CAE module are removed. The LLMs capture fine-grained biochemical semantics across drugs, genes, and proteins, while the CAE module introduces multiplex-aware, context-sensitive attention and contrastive alignment that enhance cross-layer consistency. Together, these modules form the backbone of CAZI-MBN's representational power, enabling it to jointly model both sequence information and heterogeneous interaction patterns.

CAZI-MBN was specifically designed to integrate high-quality biochemical sequence embeddings with multiplex graph structure and to distill this joint information for generalization. This design choice is particularly critical in the zero-shot setting, where models must predict interactions for entities with no prior neighborhood context. The strong contribution of LLM-based embeddings to overall performance is therefore not surprising, as they provide rich sequence-level semantics that directly support generalization. Indeed, when compared against baselines that lack domain-specific LLMs, CAZI-MBN shows substantial improvements in zero-shot prediction, confirming the importance of leveraging modality-specific pre-trained models.

Taken together, these findings demonstrate that CAZI-MBN's strength lies in its capacity to integrate complementary modalities, sequence-level semantics and relational structure, into coherent representations. This synergy allows the framework to capture both fine-grained molecular details and higher-order dependencies across interaction layers, yielding robust performance in transductive tasks and strong generalization in zero-shot scenarios.

Table 14: Summary of ablation variants.

Variant	Description
w/o UGT	Remove UGT; use only LLM-derived features as input
w/o LLMs	Replace LLM features with 6-mer and Morgan fingerprints
w/o CAE	Replace CAE with a single GCN per interaction type
w/o MoE	Replace MoE with a plain MLP per interaction type
w/o L_{dis}	Discard discriminator
w/o L_{reg}	Discard consensus regularizer and replace with attention-only fusion for Z

Table 15: Ablation results in transductive setting across five MBNs.

Dataset	Model	AUROC	AUPRC	HS	SA
	w/o UGT	0.666 ± 0.015	0.705±0.008	0.639 ± 0.012	0.634 ± 0.017
	w/o LLMs	0.563 ± 0.021	0.579 ± 0.023	0.591 ± 0.014	0.582 ± 0.016
	w/o CAE	0.601 ± 0.014	0.635 ± 0.018	0.613 ± 0.016	0.602 ± 0.012
DGIdb	w/o MoE	0.672 ± 0.019	0.707 ± 0.011	0.645 ± 0.020	0.637 ± 0.018
	w/o L_{dis}	0.692 ± 0.024	0.701±0.017	0.673±0.019	0.655 ± 0.022
	w/o L_{reg}	0.679 ± 0.012	0.696 ± 0.013	0.660 ± 0.010	0.644 ± 0.014
	CAZI-MBN	0.715 ± 0.007	0.729 ± 0.009	$0.687 {\pm} 0.011$	$0.684 {\pm} 0.015$
	w/o UGT	0.768 ± 0.010	0.815 ± 0.006	$0.844 {\pm} 0.010$	0.740 ± 0.013
	w/o LLMs	0.655 ± 0.016	0.691 ± 0.020	0.710 ± 0.013	0.612 ± 0.019
	w/o CAE	0.702 ± 0.008	0.727 ± 0.012	0.778 ± 0.011	0.691 ± 0.016
ChEMBL	w/o MoE	0.735 ± 0.014	0.781 ± 0.017	0.797 ± 0.009	0.701 ± 0.009
	w/o L_{dis}	0.790 ± 0.023	0.824 ± 0.017	0.859 ± 0.011	0.728 ± 0.012
	w/o L_{reg}	0.731 ± 0.014	0.811 ± 0.015	0.826 ± 0.012	0.715 ± 0.016
	CAZI-MBN	0.812 ± 0.008	$0.863 {\pm} 0.006$	$0.889 {\pm} 0.014$	0.757 ± 0.011
	w/o UGT	0.801±0.009	0.831±0.013	0.735 ± 0.011	0.749 ± 0.012
	w/o LLMs	0.683 ± 0.019	0.699 ± 0.017	0.629 ± 0.019	0.609 ± 0.018
	w/o CAE	0.719 ± 0.014	0.738 ± 0.015	0.685 ± 0.008	0.654 ± 0.014
PINNACLE	w/o MoE	0.748 ± 0.015	0.765 ± 0.016	0.707 ± 0.010	0.693 ± 0.013
	w/o L_{dis}	0.811 ± 0.016	0.832 ± 0.25	0.740 ± 0.013	0.742 ± 0.018
	w/o L_{reg}	0.795 ± 0.011	0.818 ± 0.013	0.737 ± 0.013	0.722 ± 0.013
	CAZI-MBN	$0.831 {\pm} 0.018$	$0.845{\pm}0.011$	0.751 ± 0.009	0.772 ± 0.013
	w/o UGT	0.721 ± 0.015	0.705 ± 0.010	0.699 ± 0.012	0.661 ± 0.011
	w/o LLMs	0.610 ± 0.018	0.620 ± 0.022	0.641 ± 0.020	0.601 ± 0.021
	w/o CAE	0.678 ± 0.019	0.679 ± 0.018	0.693 ± 0.010	0.635 ± 0.014
MetaConserve	w/o MoE	0.702 ± 0.016	0.719 ± 0.020	0.729 ± 0.008	0.642 ± 0.010
	w/o L_{dis}	0.731 ± 0.011	0.720 ± 0.009	0.737 ± 0.014	0.625 ± 0.018
	w/o L_{reg}	0.727 ± 0.015	0.719 ± 0.014	0.743 ± 0.011	0.637 ± 0.014
	CAZI-MBN	$0.752 {\pm} 0.020$	$0.744 {\pm} 0.012$	0.779 ± 0.008	$0.656 {\pm} 0.014$
	w/o UGT	0.813±0.013	0.829 ± 0.011	0.722±0.010	0.714 ± 0.015
	w/o LLMs	0.731 ± 0.024	0.738 ± 0.015	0.655 ± 0.012	0.642 ± 0.018
	w/o CAE	0.768 ± 0.017	0.775 ± 0.020	0.701 ± 0.013	0.698 ± 0.011
TRRUST	w/o MoE	0.792 ± 0.014	0.799 ± 0.016	0.727 ± 0.019	0.720 ± 0.012
	w/o L_{dis}	0.857 ± 0.009	0.812 ± 0.014	0.733 ± 0.013	0.730 ± 0.020
	w/o L_{reg}	0.831 ± 0.012	0.840 ± 0.012	0.728 ± 0.011	0.722 ± 0.013
	CAZI-MBN	0.899±0.017	0.869±0.013	0.775±0.019	0.764 ± 0.007

Table 16: Ablation results in zero-shot setting across five MBNs for CAZI-MBN(KD).

			ng across five M		
Dataset	Model	AUROC	AUPRC	HS	SA
	w/o UGT	0.633 ± 0.015	0.664 ± 0.010	0.637 ± 0.012	0.625 ± 0.014
	w/o LLMs	0.527 ± 0.017	0.553 ± 0.025	0.599 ± 0.018	0.573 ± 0.015
	w/o CAE	0.589 ± 0.019	0.607 ± 0.020	0.618 ± 0.009	0.604 ± 0.012
DGIdb	w/o MoE	0.641 ± 0.012	0.662 ± 0.015	0.639 ± 0.011	0.631 ± 0.011
	w/o L_{dis}	0.643 ± 0.017	0.677 ± 0.009	0.819 ± 0.022	0.690 ± 0.015
	w/o L_{reg}	0.593 ± 0.012	0.616 ± 0.013	0.621±0.011	0.601±0.012
	CAZI-MBN	0.671 ± 0.008	0.709±0.011	0.688±0.009	0.663±0.014
	w/o UGT	0.744 ± 0.016	0.787 ± 0.014	0.811 ± 0.010	0.699 ± 0.017
	w/o LLMs	0.644 ± 0.021	0.673 ± 0.027	0.689 ± 0.014	0.591 ± 0.015
	w/o CAE	0.686 ± 0.017	0.709 ± 0.016	0.756 ± 0.008	0.667 ± 0.011
ChEMBL	w/o MoE	0.718 ± 0.010	0.765 ± 0.012	0.791 ± 0.011	0.707 ± 0.013
	w/o L_{dis}	0.772 ± 0.023	0.810 ± 0.019	0.815 ± 0.017	0.693 ± 0.025
	w/o L_{reg}	0.718 ± 0.013	0.746 ± 0.014	0.775 ± 0.010	0.686 ± 0.014
	CAZI-MBN	0.791±0.018	0.839±0.015	0.857±0.011	0.723±0.009
	w/o UGT	0.755 ± 0.012	0.768 ± 0.018	0.713 ± 0.014	0.725 ± 0.012
	w/o LLMs	0.671 ± 0.019	0.685 ± 0.021	0.628 ± 0.012	0.614 ± 0.014
	w/o CAE	0.708 ± 0.016	0.715 ± 0.014	0.691 ± 0.010	0.667 ± 0.016
PINNACLE	w/o MoE	0.733 ± 0.009	0.745 ± 0.011	0.717 ± 0.007	0.701 ± 0.010
	w/o L_{dis}	0.775 ± 0.015	0.801 ± 0.008	0.715 ± 0.021	0.730 ± 0.019
	w/o L_{reg}	0.758±0.011	0.794±0.014	0.732±0.012	0.712±0.013
	CAZI-MBN	0.812 ± 0.013	0.820 ± 0.008	0.748 ± 0.010	0.763 ± 0.011
	w/o UGT	0.692 ± 0.011	0.699 ± 0.013	0.719 ± 0.012	0.631 ± 0.014
	w/o LLMs	0.609 ± 0.017	0.634 ± 0.015	0.651 ± 0.016	0.591 ± 0.012
	w/o CAE	0.656 ± 0.014	0.662 ± 0.016	0.687 ± 0.009	0.615 ± 0.011
MetaConserve	w/o MoE	0.703 ± 0.015	0.717 ± 0.018	0.708 ± 0.011	0.642 ± 0.009
	w/o L_{dis}	0.719 ± 0.022	0.693 ± 0.025	0.716 ± 0.014	0.633 ± 0.016
	w/o L_{reg}	0.713±0.014	0.700 ± 0.015	0.735±0.011	0.629±0.013
	CAZI-MBN	0.738 ± 0.015	0.722 ± 0.010	0.749 ± 0.020	$0.653 {\pm} 0.009$
	w/o UGT	0.848 ± 0.013	0.826 ± 0.016	0.741 ± 0.015	0.726 ± 0.012
	w/o LLMs	0.742 ± 0.018	0.749 ± 0.021	0.662 ± 0.010	0.655 ± 0.014
	w/o CAE	$0.788 {\pm} 0.019$	0.802 ± 0.013	0.709 ± 0.014	0.701 ± 0.013
TRRUST	w/o MoE	$0.824 {\pm} 0.014$	$0.834 {\pm} 0.015$	0.729 ± 0.013	0.740 ± 0.015
	w/o L_{dis}	0.867 ± 0.019	0.831 ± 0.029	0.732 ± 0.020	0.735 ± 0.011
	w/o L_{reg}	0.851 ± 0.012	0.818±0.011	0.729 ± 0.013	0.716±0.014
	CAZI-MBN	0.899±0.017	0.869±0.013	0.775±0.019	0.764 ± 0.007

E.3.4 CASE STUDY: INTERACTIONS INVOLVING IBD-RELATED GENES AND PROTEINS

To evaluate the effectiveness of our approach in real-world biological scenarios, we conducted a case study focused on zero-shot multiplex interaction prediction involving IBD-related genes and their protein products. The selected genes, *GAPDH*, *PC*, *RPS14*, *OAT*, and *NDUFAB1*, were chosen based on their established functional relevance to IBD and their evolutionary conservation from *E. coli* orthologs, including *gapA*, *accC*, *rpsK*, *orn*, and *fabD*. These genes represent a diverse set of biological functions, such as energy metabolism, redox balance, immune regulation, and mitochondrial activity, all of which are closely associated with IBD pathogenesis. This focused selection enables a robust evaluation of the model's ability to predict novel, biologically meaningful interactions involving entirely unseen entities.

The genes and their IBD-relevant roles are summarized as follows:

- *GAPDH* (ortholog: *gapA*): Known for its central role in glycolysis, GAPDH also regulates T-cell function and cytokine production, highlighting its dual role in metabolism and immune modulation (Sheng & Wang, 2009; Bas et al., 2004; Barber et al., 2005).
- *PC* (ortholog: *accC*): Pyruvate carboxylase plays a key role in gluconeogenesis and epithelial energy metabolism, both of which are affected in IBD (Algieri et al., 2016; Ferrer-Picón et al., 2020; Tang et al., 2015; Bos & Laukens, 2020).
- *RPS14* (ortholog: *rpsK*): A ribosomal protein with emerging evidence linking it to immune and stress response regulation in intestinal inflammation (Sallman & List, 2019; Zou & Zhang, 2021; Lin et al., 2020; Das et al., 2024).
- *OAT* (ortholog: *orn*): Ornithine aminotransferase regulates amino acid metabolism and redox balance, with links to epithelial injury in IBD (Ji et al., 2023; Smith et al., 2021; Lan et al., 2023; Gobert et al., 2004).
- *NDUFAB1* (ortholog: *fabD*): This mitochondrial protein supports fatty acid synthesis and electron transport and is associated with epithelial barrier integrity (Guerbette et al., 2022; Rath et al., 2018; Kim et al., 2021).

For each dataset-specific case study, a separate model was trained using a standard zero-shot setting. All biological entities and interactions involving the five specified genes and their protein products were held out as the test set. The remaining interactions and associated entities were randomly split into training and validation sets using an 5:1 ratio. This setup ensures that the model does not encounter the held-out genes or their associated links during training, enabling a clear evaluation of its zero-shot generalization ability.

Table 17 presents the zero-shot prediction performance of **CAZI-MBN** across four benchmark datasets, reporting an overall, as well as per-gene and gene product-level interaction statistics. These results evaluate the model's ability to generalize to previously unseen entities, specifically those involving five held-out human genes and their bacterial orthologs. None of these genes or their products were included in the training set, allowing for a rigorous assessment of the model's zero-shot capabilities in the context of sparse and novel biological interactions.

CAZI-MBN recovered a substantial proportion of known interactions across all evaluated datasets, achieving 82.7% accuracy in DGIdb, 85.7% in TRRUST, 75.1% in PINNACLE, and 62.5% in MetaConserve. These datasets collectively span a range of biological interaction types, including drug-gene interactions (DGIdb), transcriptional regulatory links (TRRUST), cell-type-specific protein-protein interactions (PINNACLE), and evolutionarily conserved co-occurrence patterns (MetaConserve). The model's consistently strong performance across these diverse settings underscores its robustness and flexibility in handling heterogeneous biological data.

Closer inspection of individual genes reveals that CAZI-MBN maintains strong predictive performance even for held-out genes with relatively few known interactions in the test sets. For example, *PC* (*accC*) and

1	5	0	4
-1	5	Λ	5

1515 1516 1517

1518 1519

1520 1521 1522

1523

1524

1525

1526 1527 1528

1531 1532 1533

1534

1530

1540 1541 1542

1543

1544

1539

1545 1546 1547

1550

1548 1549 Table 17: Per-Gene/Protein Product Statistics Across Datasets

	DGIdb	TRRUST	PINNACLE	MetaConserve
GAPDH(gapA)	34 / 41 (82.9%)	2 / 3 (66.7%)	103 / 140 (73.6%)	_
PC(accC)	2 / 2 (100%)	2 / 2 (100%)	_	_
RPS14(rpsK)	3 / 5 (60%)	_	340 / 454 (74.9%)	_
OAT(orn)	1 / 1 (100%)	2 / 2 (100%)	11 / 16 (68.7%)	_
NDUFAB1(fabD)	3 / 3 (100%)	-	46 / 57 (80.7%)	5 / 8 (62.5%)
Total	43 / 52 (82.7%)	6 / 7 (85.7%)	500 / 666 (75.1%)	5 / 8 (62.5%)

Each cell shows: Predicted / Actual (Accuracy), where Predicted indicates the number of interactions correctly predicted by the model, and **Actual** indicates the total number of known interactions involving the given gene or its product. "-" indicates no active interaction with the gene or its product was found in that dataset.

OAT (orn) each exhibited low interaction frequency within DGIdb and TRRUST, yet the model successfully recovered all known interactions for both cases in these datasets. This is noteworthy given that these genes were entirely excluded from training, highlighting the model's ability to generalize to novel genes and their interaction profiles. Although these results reflect only test-time performance, they demonstrate that CAZI-MBN can effectively capture biological relationships across different interaction types, including drug-gene interactions, transcriptional regulation, cell-type specific protein-protein interactions, without relying on prior exposure to the specific biological entities involved. Such flexibility is critical for modeling MBNs, where each biological entity may participate in diverse biological processes and be subject to distinct regulatory mechanisms. Overall, these findings highlight the strength of CAZI-MBN in generalizing across both interaction types and gene contexts. Its capacity to make accurate predictions in a zero-shot setting suggests its potential as a valuable tool for expanding known interaction networks, prioritizing novel gene candidates, and supporting downstream biological discovery.

Several high-confidence predictions made by CAZI-MBN are further supported by existing biomedical literature, reinforcing the biological plausibility of the model's outputs. For example, the model predicts an interaction between GAPDH and Omigapil (DGIdb, pred: 0.912), which is validated by evidence showing that Omigapil inhibits the GAPDH-Siah1 apoptotic signaling pathway in neuromuscular disorders (Foley et al., 2024; Zhou et al., 2015). Additionally, GAPDH has been shown to chaperone ribosomal protein L13a, stabilizing the GAIT complex and modulating inflammation-related translation, which aligns with predicted interactions between GAPDH and ribosomal proteins in PINNACLE (Jia et al., 2012). The model also identifies numerous interactions involving RPS14 with other ribosomal proteins (e.g., RPL19, RPS17), which are consistent with its role in ribosome assembly and immune-related translational control (Zhou et al., 2015; Chan et al., 2018; Wang et al., 2022). These literature-backed findings confirm that CAZI-MBN is not only capable of recovering known links but also makes biologically coherent predictions in a fully zero-shot setting.

This case study highlights the effectiveness of multiplex-aware zero-shot learning frameworks in predicting biologically meaningful interactions for novel or under-characterized entities. It reinforces CAZI-MBN's potential as a tool for accelerating hypothesis generation and advancing discovery across complex, multilayered biological systems.

^b Total interaction counts may not equal the sum of individual gene or gene product counts, as some interactions involve multiple listed genes or their products.