Protein Representation Learning by Capturing Protein Sequence-Structure-Function Relationship

Eunji Ko*[1], Seul Lee*[1], Minseon Kim*[1], Sung Ju Hwang[1]
[1]KAIST
kosu7071@kaist.ac.kr

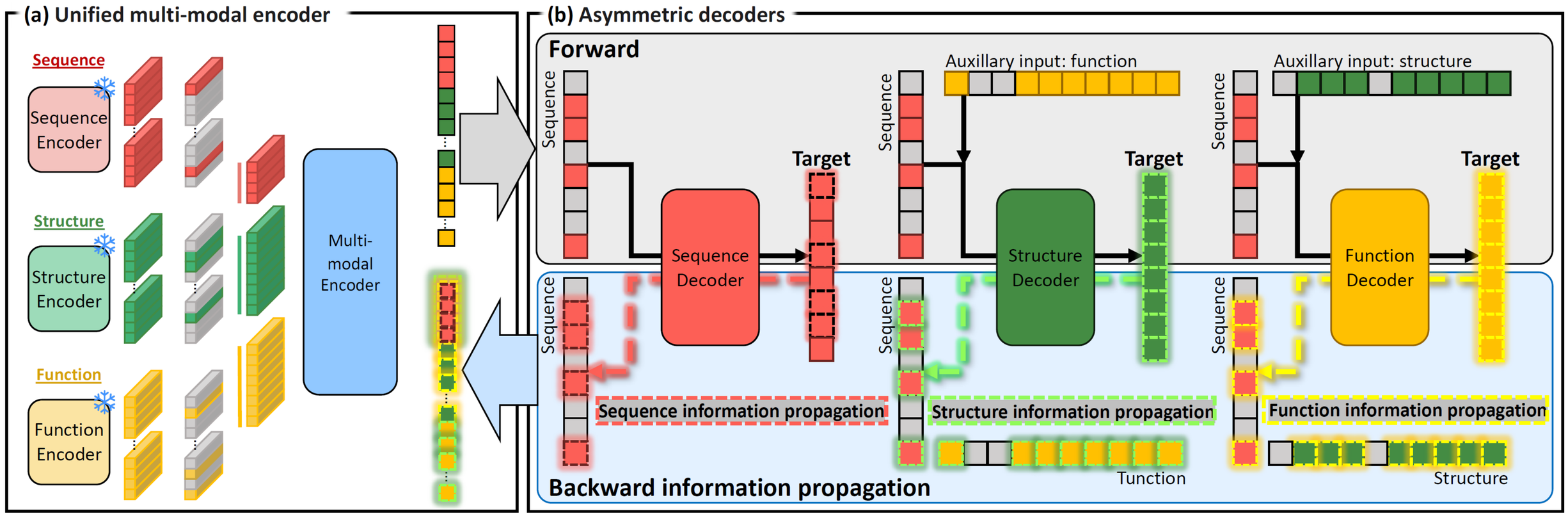

***Figure 2*. The overall framework of AMMA.**

Introduction

- We are the first to propose utilizing the three core modalities for protein representation learning: **sequence, structure, and function**.
- We point out the **asymmetric relationship** between sequence, structure, and function of proteins and propose **AMMA**, a masked autoencoder framework that adopts a **unified multi-modal encoder** and **asymmetric decoders** to account for the asymmetric relationship.
- We experimentally demonstrate that AMMA is **highly effective** in learning protein representations and **benefits performance** on a variety of downstream protein-related tasks.

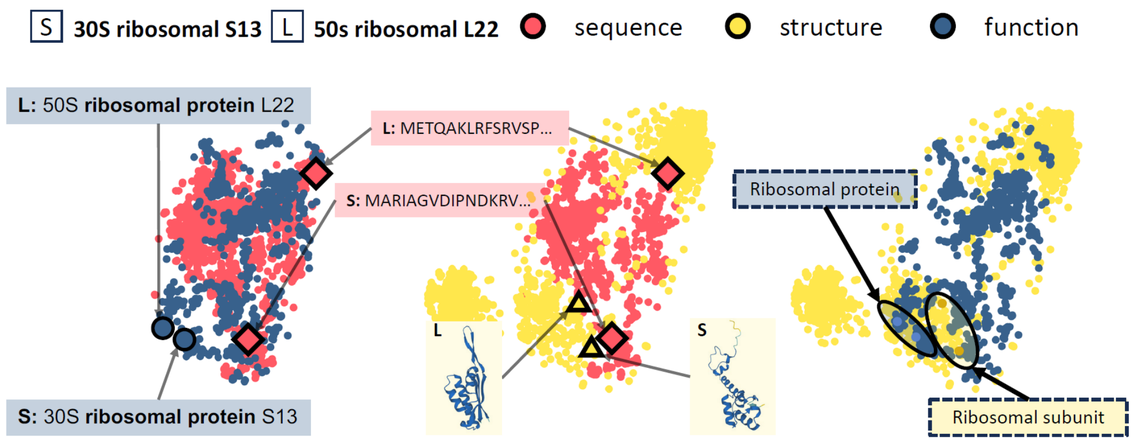

***Figure 1*. t-SNE visualization of the three modalities of proteins.**

Methodology: Asymmetric Multimodal Masked Autoencoder (AMMA)

- We introduce **AMMA**, asymmetric multi-modal masked autoencoder.
  1. Each modality data is encoded by **a frozen pretrained uni-modal encoder**.
  2. The encoded features are then masked and integrated into a unified representations by a **multi-modal encoder**.
  3. Each **asymmetric decoder** reconstructs the original feature for each modality.
- During decoding, the input latent features are **asymmetrically passed** to the decoders.
- This requires AMMA to encode structural and function information into sequence latent features, which allows AMMA to capture **unique asymmetric sequence-structure-function relationships**.

$$\mathcal{L}_{\text{seq}} = \text{MSE}(\hat{X}_{\text{seq}}, X'_{\text{seq}}), \mathcal{L}_{\text{str}} = \text{MSE}(\hat{X}_{\text{str}}, X'_{\text{str}}), \mathcal{L}_{\text{func}} = \text{MSE}(\hat{X}_{\text{func}}, X'_{\text{func}}),$$
$$\mathcal{L} = \mathcal{L}_{\text{seq}} + \mathcal{L}_{\text{str}} + \mathcal{L}_{\text{func}}.$$

Experiments: Qualitative Results

- Evaluate performance of our model **on two standard downstream tasks**.

***Table 1*. Performance on protein function annotation tasks.**

Method	Modality			EC		GO-MF		GO-CC		GO-BP		Avg.$_{\text{Fmax}}$	Avg.$_{\text{AUPR}}$
	Seq.	Str.	Func.	F_{max}	AUPR	F_{max}	AUPR	F_{max}	AUPR	F_{max}	AUPR		
ESM-1b (Rives et al., 2021)	✓			86.9	88.4	65.9	63.0	47.7	32.4	45.2	33.2	61.4	54.3
OntoProtein (Zhang et al., 2022)	✓		✓	84.1	85.4	63.1	60.3	44.1	30.0	43.6	28.4	58.7	51.0
GearNet (Zhang et al., 2023)		✓		87.4	89.2	65.4	59.6	48.8	33.6	**49.0**	29.2	62.7	52.9
SaProt (Su et al., 2023)	✓	✓		**88.8**	85.5	**68.8**	58.2	41.2	20.6	45.1	23.8	61.0	47.0
ProtST (Xu et al., 2023)	✓		✓	87.8	89.4	66.1	64.4	48.8	36.4	48.0	32.8	62.7	55.8
AMMA-symmetric (ours)	✓	✓	✓	71.6	74.9	52.0	52.3	48.8	35.1	35.5	24.3	52.0	46.7
AMMA-contrastive (ours)	✓	✓	✓	87.7	89.5	65.2	61.3	44.3	28.2	28.2	17.3	56.4	49.1
AMMA (ours)	✓	✓	✓	88.7	**89.8**	67.3	**65.5**	**49.8**	**36.9**	46.9	**33.6**	**63.2**	**56.5**

Experiments: Improving performance with unpaired data

Data		EC		GO-MF		Average
Paired	Unpaired	F_{max}	AUPR	F_{max}	AUPR	
120k	0k	88.1	89.7	66.4	**64.6**	77.2
120k	50k	**88.2**	**90.4**	**66.9**	**64.6**	**77.5**

***Table 2*. EC/GO results with extra unpaired data.**

Experiments: Ablation studies and qualitative analysis

- Effect of the **asymmetric decoders**

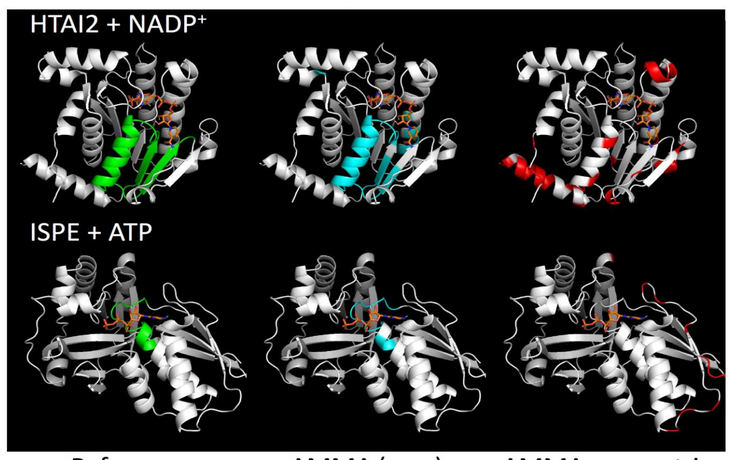

***Figure 3*. Visualization of highly attended residues in a functional context.**

- Comparison with **contrastive learning**

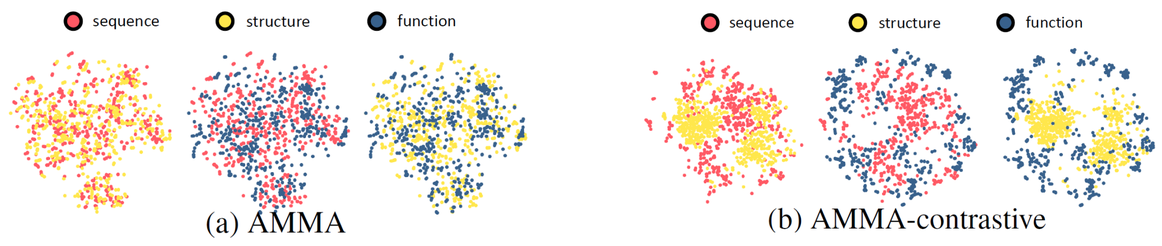

***Figure 4*. t-SNE visualization of three protein modalities**

- Effect of **masking ratio**

Ratio			EC		GO-MF		Average
α_{seq}	α_{str}	α_{func}	F_{max}	AUPR	F_{max}	AUPR	
1	1	1	84.6	87.2	66.4	64.8	75.8
1	2	2	87.7	89.8	66.4	64.2	**77.0**
2	1	1	73.0	75.9	65.1	63.2	69.3
2	1	2	86.7	89.2	65.9	64.7	76.6
2	2	1	87.9	89.5	52.4	53.6	70.9

***Table 3*. Experimental results with different α.**