Co-evolution Transformer for Protein Contact Prediction

He Zhang[*1], **Fusong Ju**[*2], Jianwei Zhu[2], Liang He[2],
Bin Shao[2], Nanning Zheng[1], Tie-Yan Liu[2]

[1]Xi'an Jiaotong University. [2]Microsoft Research Asia.
*Equal contribution

Contents

- Background and Motivation

- Co-evolution Transformer

- Experimental Results

- Conclusion and Future work

Background and Motivation

➢ **Goal:** Inter-residue contact/distance prediction

  ✓ Input: Protein sequence or multiple sequence alignment (MSA)

  ✓ Output: Contact/distance matrix

➢ **Application:** The essential block of structure-related applications

  ✓ **Protein structure prediction**

 ✓ Protein design

  ✓ …

Background and Motivation

➢ **Goal:** Inter-residue contact/distance prediction

   ✓ Input: Protein sequence or multiple sequence alignment (MSA)

   ✓ Output: Contact/distance matrix

➢ **Application:** The essential block of structure-related applications

   ✓ **Protein structure prediction**

   ✓ Protein design

   ✓ ...
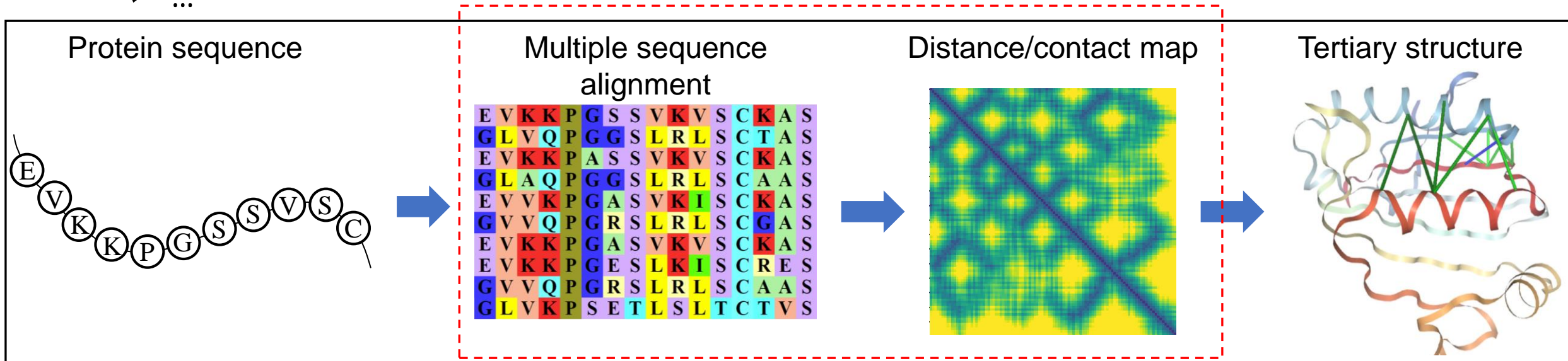
A popular framework for protein structure prediction
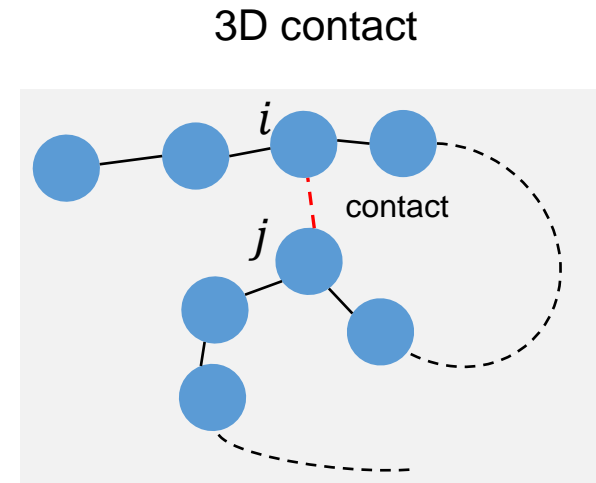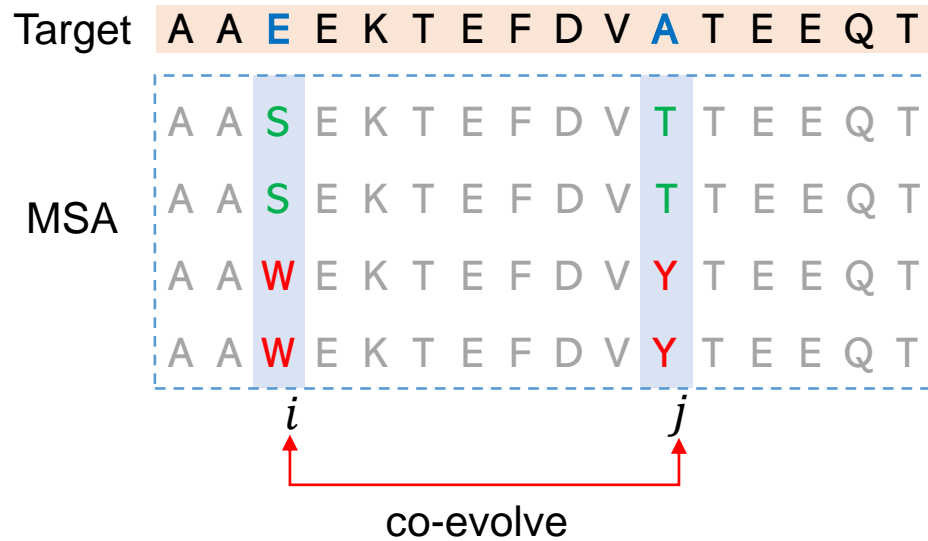
Background and Motivation
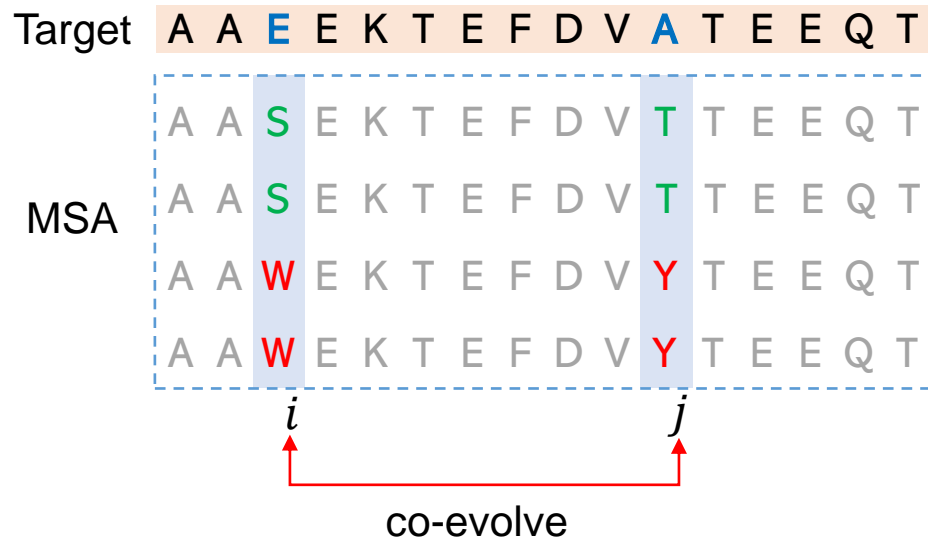
➢ **Key prior:** the co-evolution principle

  ✓ Spatially proximate residues tend to co-evolve

➢ **Previous works:**

  ✓ Unsupervised methods

  ✓ **Supervised methods**

  ✓ Pre-training based methods

Background and Motivation

➤ **Key prior:** the co-evolution principle

  ✓ Spatially proximate residues tend to co-evolve

➤ **Previous works:**

  ✓ Unsupervised methods

  ✓ Supervised methods

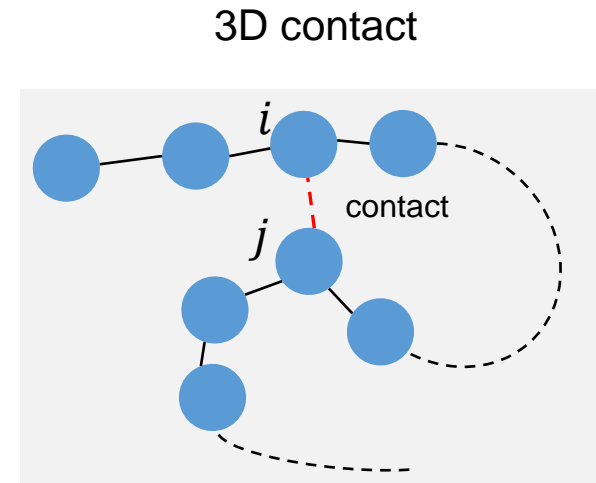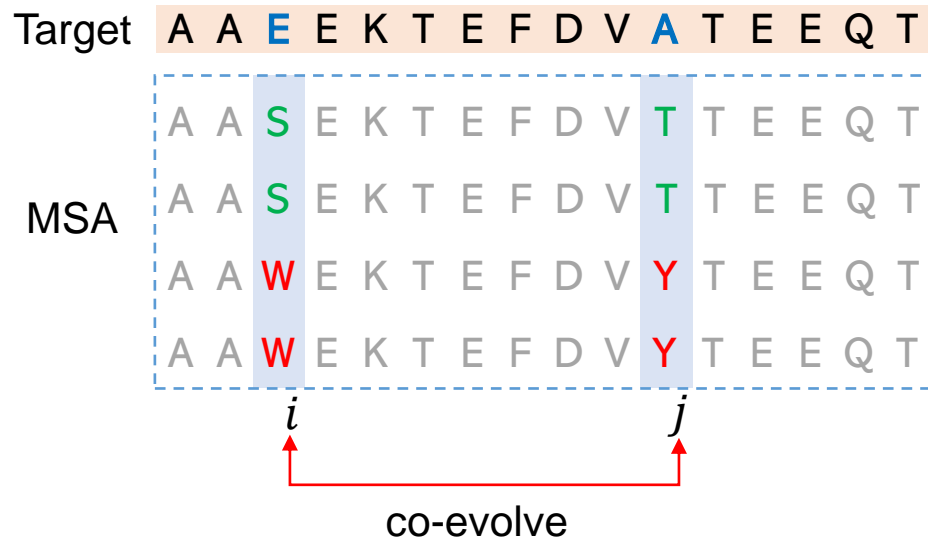 ✓ Pre-training based methods

3D contact

co-evolve

Background and Motivation

➢ **Key prior:** the co-evolution principle

   ✓ Spatially proximate residues tend to co-evolve

➢ **Previous works:**

   ✓ Unsupervised methods

   ✓ Supervised methods

   ✓ Pre-training based methods
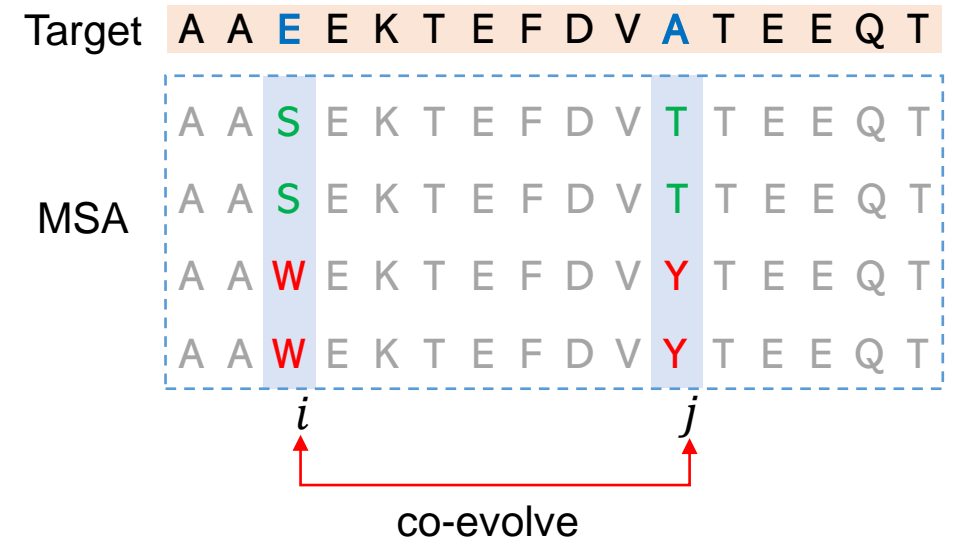
How to extract and leverage co-evolutionary patterns?

Background and Motivation

Inferring co-evolution information from MSA

- Direct coupling analysis (DCA)

  ➢ RaptorX

  ➢ trRosetta

  ➢ AlphaFold

$$P(\boldsymbol{\sigma}) = \frac{1}{Z} \exp \left(\sum_{i=1}^{N} h_i(\sigma_i) + \sum_{1 \le i < j \le N} J_{ij}(\sigma_i, \sigma_j) \right)$$

Yang J, Anishchenko I, Park H, et al. Improved protein structure prediction using predicted interresidue orientations[J]. Proceedings of the National Academy of Sciences, 2020.
*Wang S, Sun S, Li Z, et al. Accurate de novo prediction of protein contact map by ultra-deep learning model[J]. PLoS computational biology, 2017.*
Senior A W, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning[J]. Nature, 2020.

Background and Motivation

Inferring co-evolution information from MSA

- Direct coupling analysis (DCA)

  ➢ RaptorX

  ➢ trRosetta

  ➢ AlphaFold
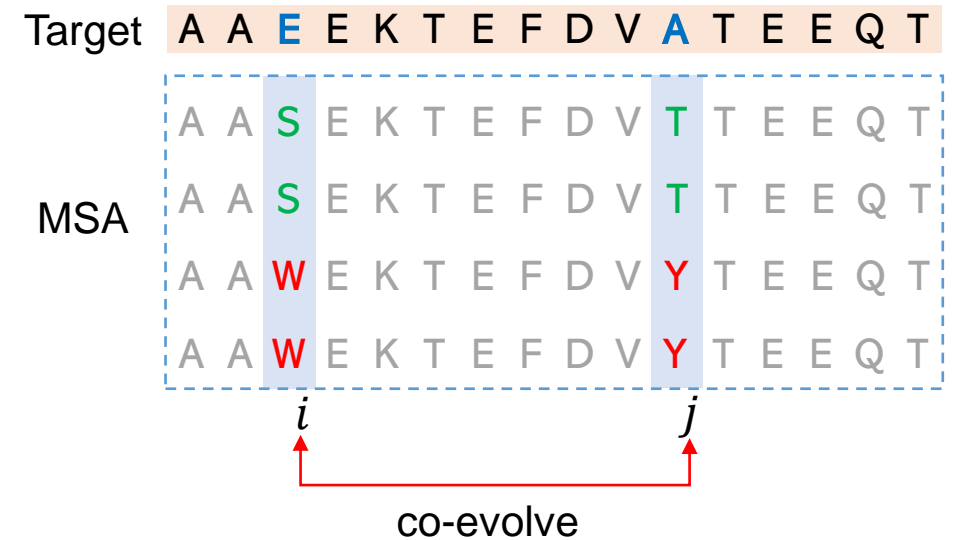
$$P(\boldsymbol{\sigma}) = \frac{1}{Z} \exp\left(\sum_{i=1}^{N} h_i(\sigma_i) + \sum_{1 \leq i < j \leq N} J_{ij}(\sigma_i, \sigma_j) \right)$$
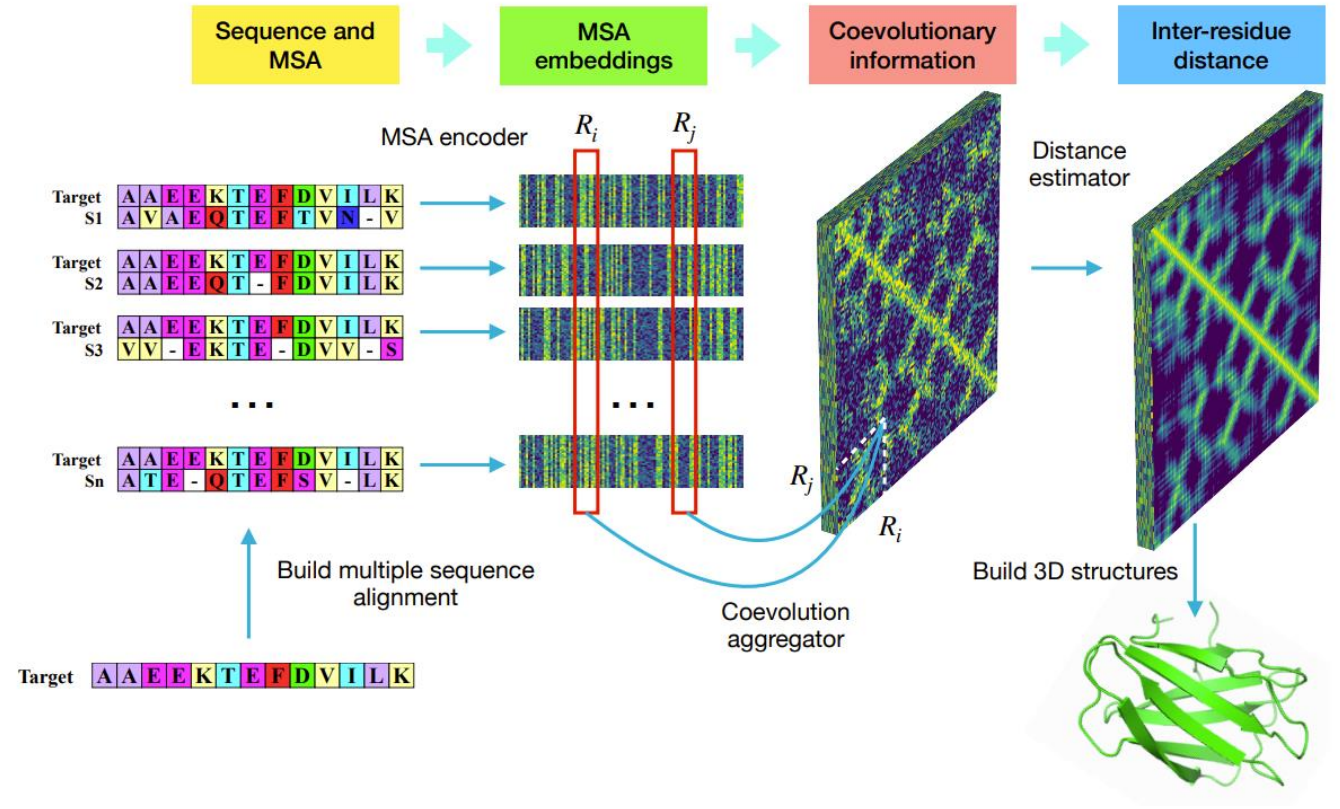
Single-residue and pairwise statistics are well considered

High-order interactions are ignored

Background and Motivation

Inferring co-evolution information from MSA

- Learning directly from MSA

  ➢ **CopulaNet (SOTA)**

  ➢ RawMSA

Ju F, Zhu J, Shao B, et al. CopulaNet: Learning residue co-evolution directly from multiple sequence alignment for protein structure prediction[J]. Nature communications, 2021

Mirabello C, Wallner B. RAWMSA: End-to-end deep learning using raw multiple sequence alignments[J]. PloS one, 2019

Background and Motivation

Inferring co-evolution information from MSA

- Learning directly from MSA

  ➢ **CopulaNet (SOTA)**

  ➢ RawMSA

  Better "end-to-end" learning frameworks

  But some important priors are missed

Ju F, Zhu J, Shao B, et al. CopulaNet: Learning residue co-evolution directly from multiple sequence alignment for protein structure prediction[J]. Nature communications, 2021
Mirabello C, Wallner B. RAWMSA: End-to-end deep learning using raw multiple sequence alignments[J]. PloS one, 2019
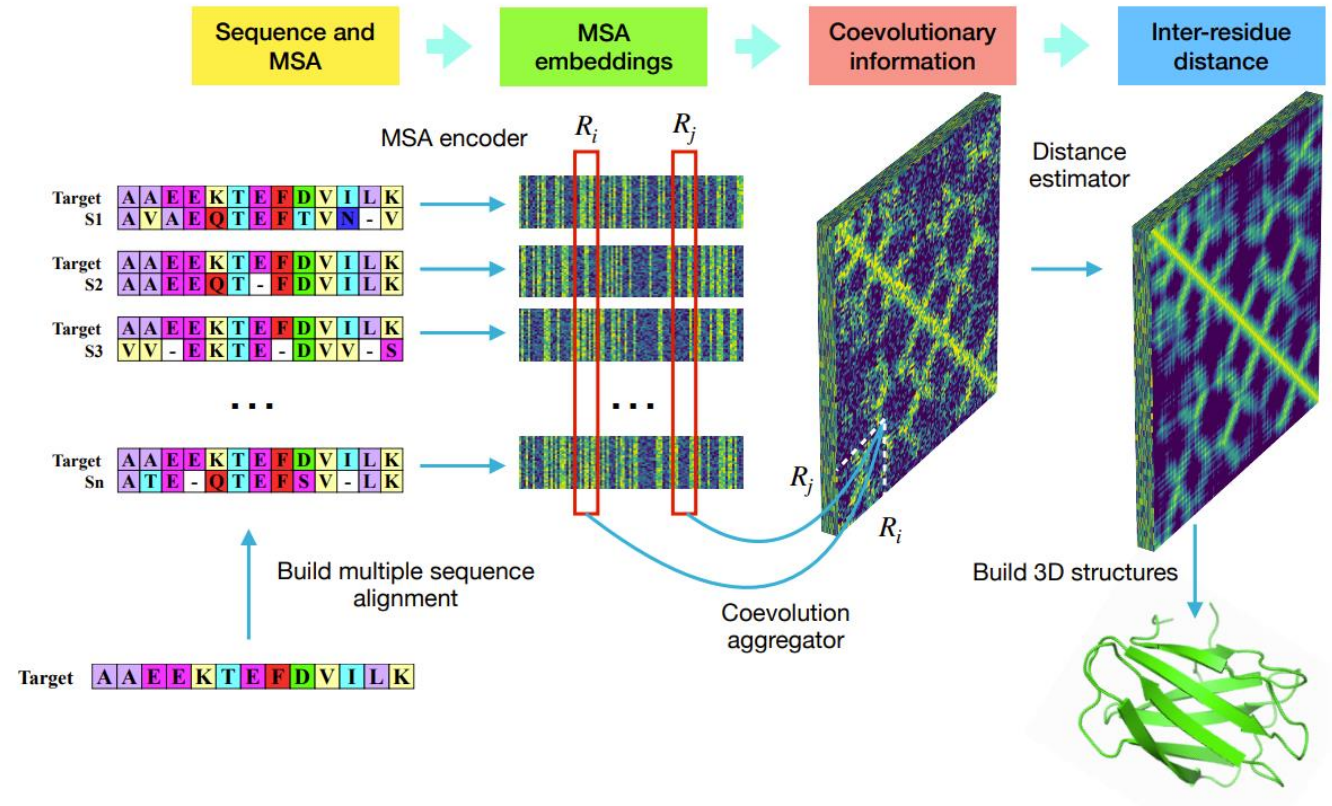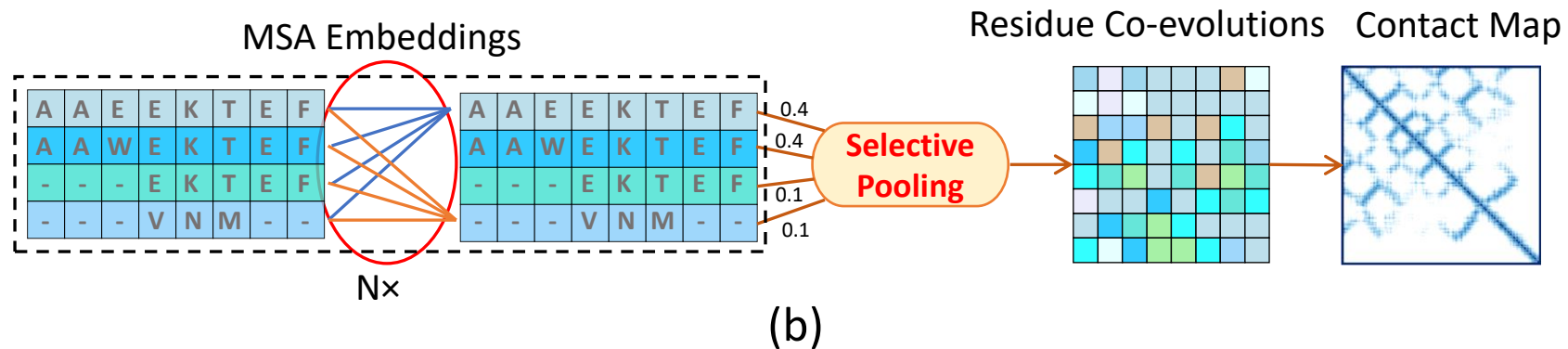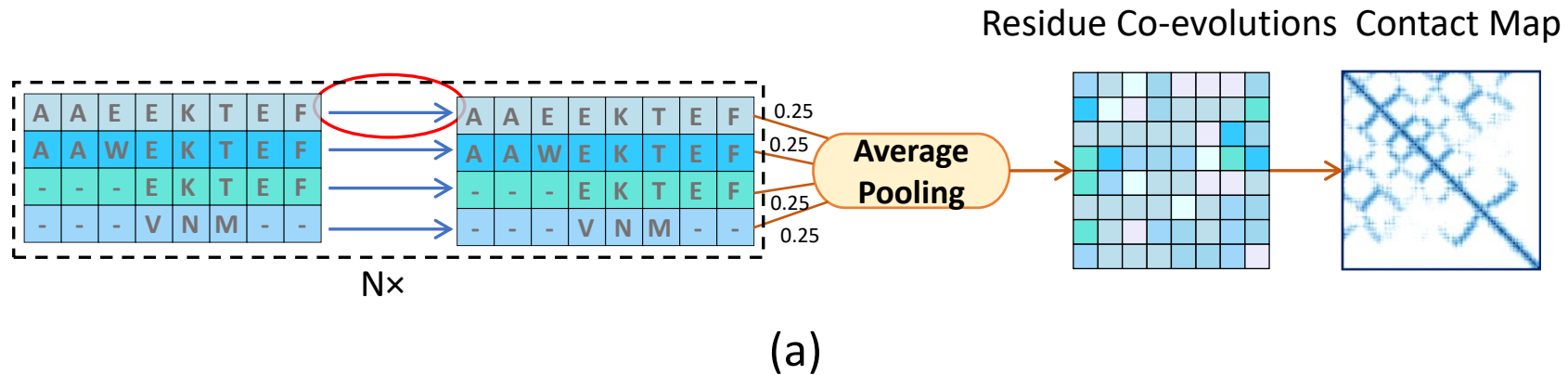
11

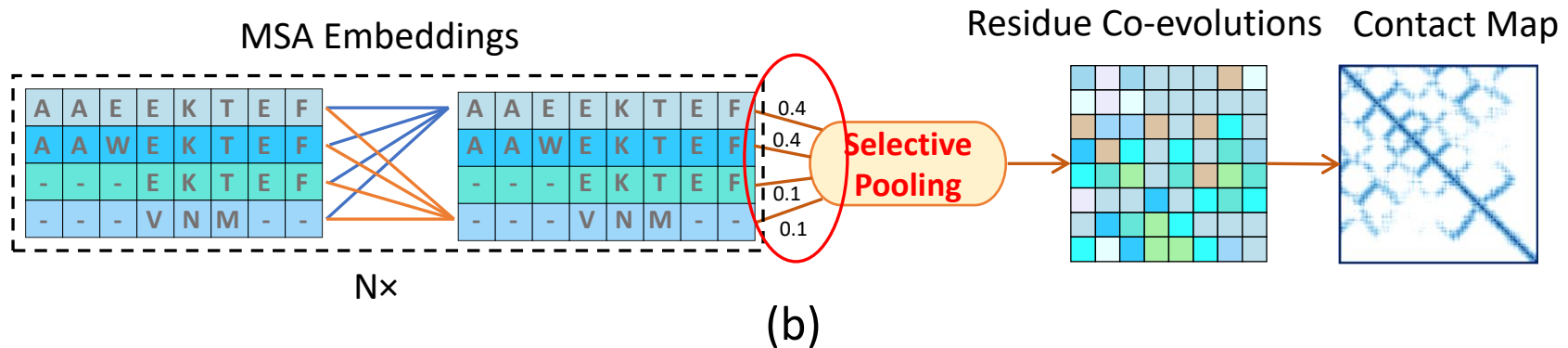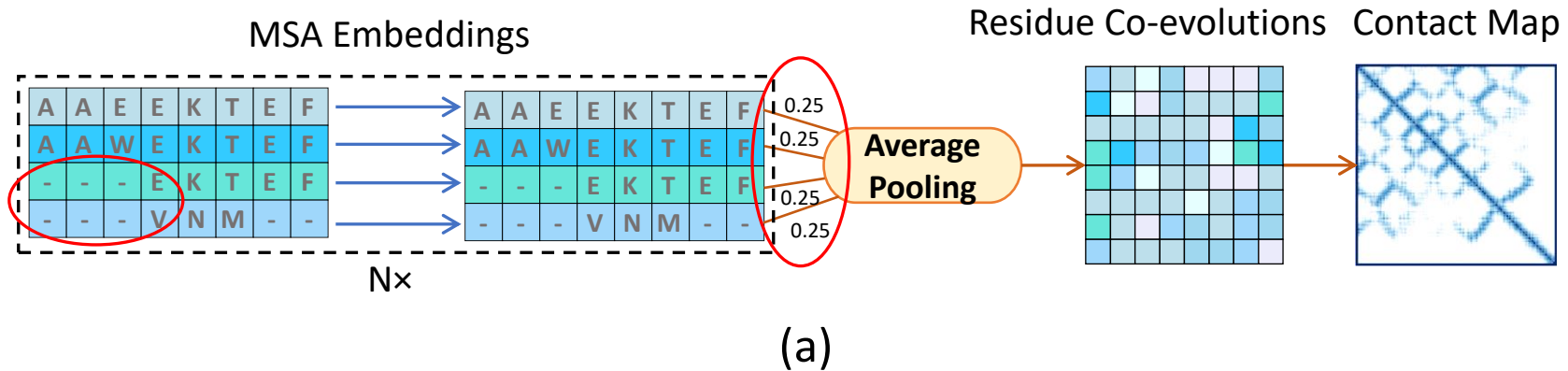# Background and Motivation

Motivation

- Modeling individual sequences independently vs. jointly

- Assigning equal vs. unequal weights to different homologs

(a)

(b)

Background and Motivation

Motivation

- Modeling individual sequences independently vs. jointly

- Assigning equal vs. unequal weights to different homologs

(a)

(b)

Co-evolution Transformer

Leveraging these two insights, we propose the Co-evolution Transformer (CoT)

Compatibility Function of Co-evolution Attention
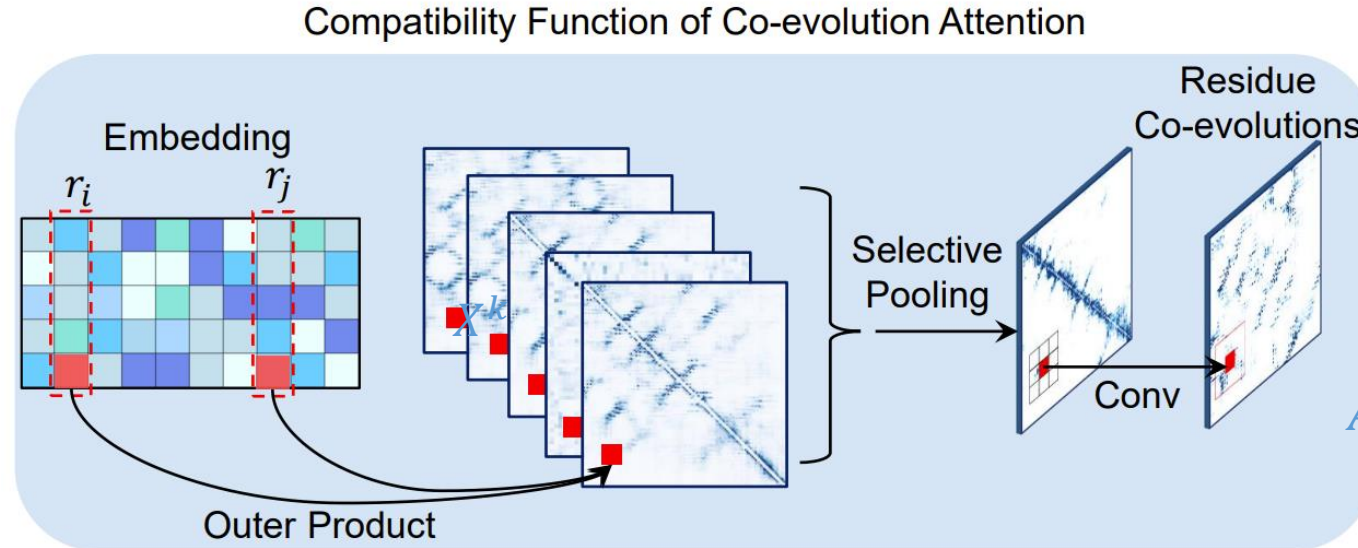
For the k-th sequence in the MSA, the CoA module is defined as:

$$X^k = \text{LAYERNORM}(X^k + \text{COATTN}(X)),$$
$$X^k = \text{LAYERNORM}(X^k + \text{FFN}(X^k))$$

$$\text{COATTN}(X) = \text{CONCAT}(\text{head}_1, \ldots, \text{head}_H),$$
$$\text{head}_h = \text{ATTN}_h(X, A)\, X_h^k\, W_h,$$

Co-evolution Transformer

Leverage these two insights, we propose the Co-evolution Attention module …

Compatibility Function of Co-evolution Attention
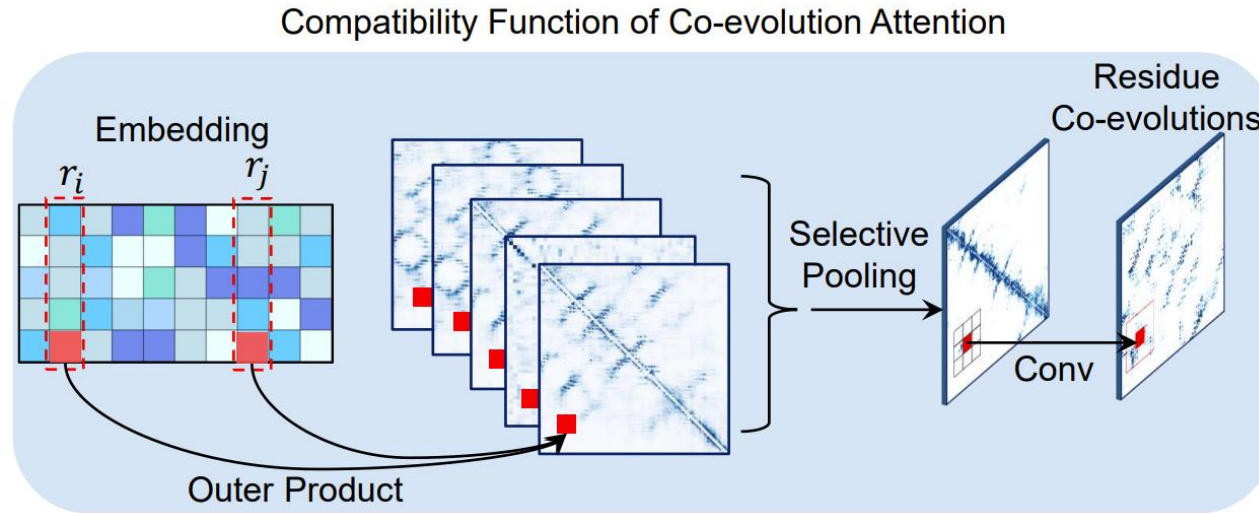
Overall learning framework
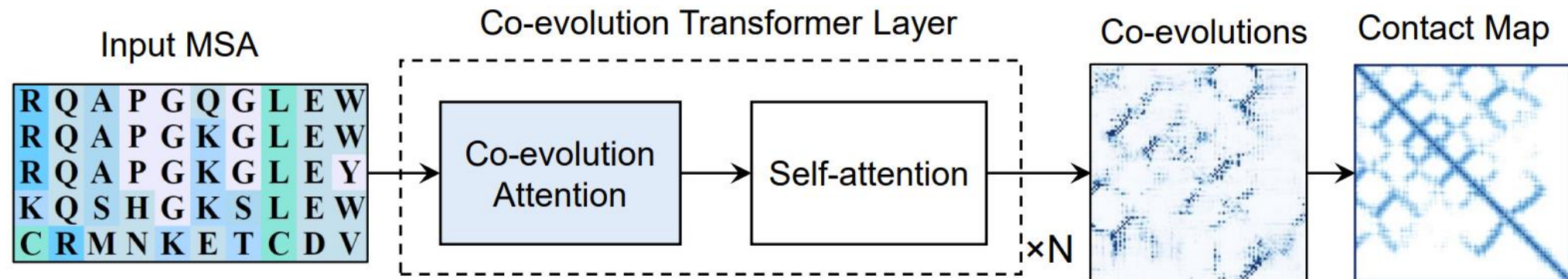
Co-evolution Transformer
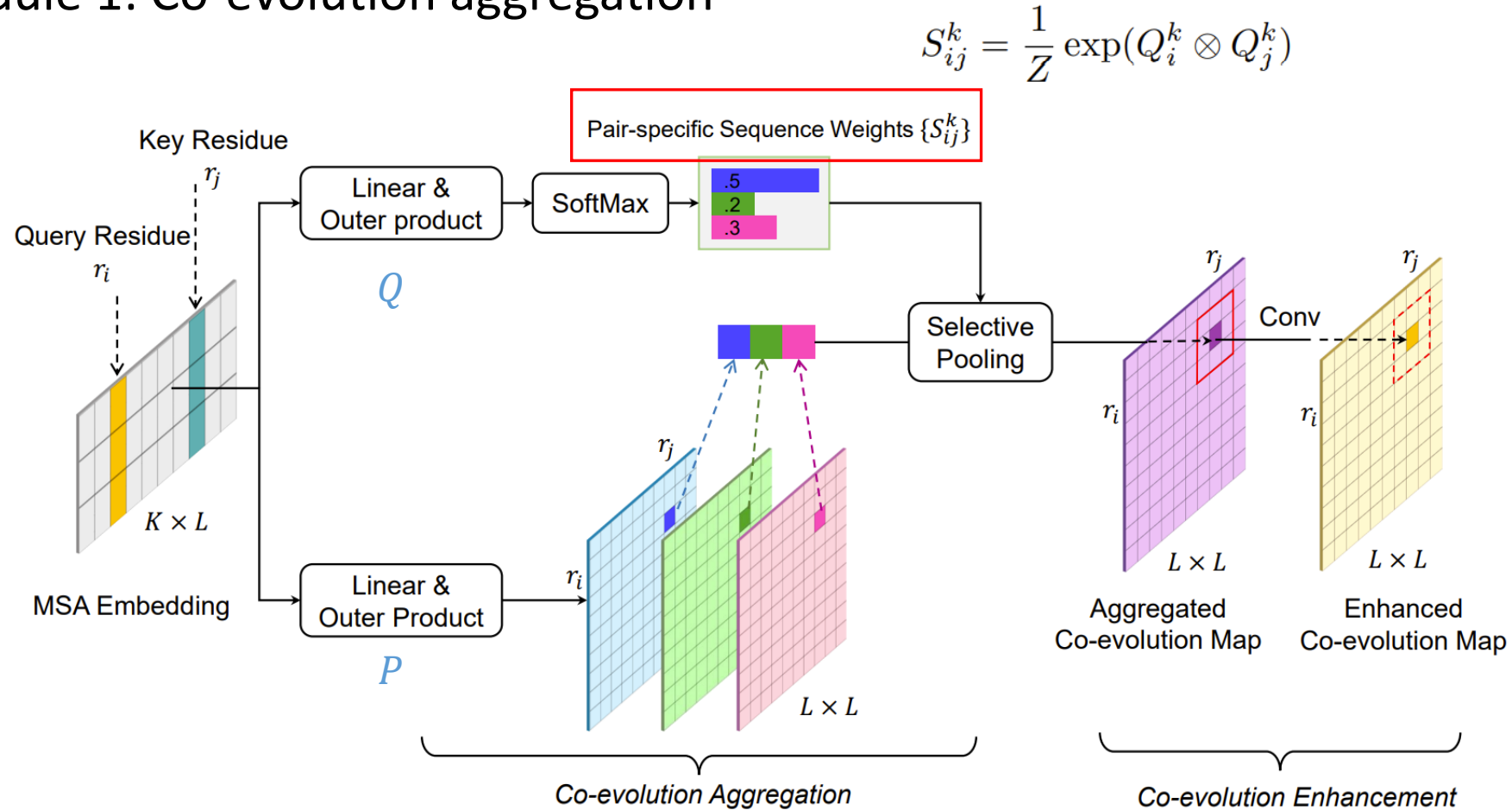
❖Submodule 1: Co-evolution aggregation

$$S_{ij}^k = \frac{1}{Z} \exp(Q_i^k \otimes Q_j^k)$$

$$A_{ij} = \text{Proj}\left(\sum_{k=1}^{K} S_{ij}^k \odot (P_i^k \otimes P_j^k)\right)$$

Co-evolution Transformer

❖Submodule 1: Co-evolution aggregation

$$S_{ij}^k = \frac{1}{Z} \exp(Q_i^k \otimes Q_j^k)$$

$$A_{ij} = \text{PROJ}\left(\sum_{k=1}^{K} S_{ij}^k \odot (P_i^k \otimes P_j^k)\right)$$

Co-evolution Transformer

❖Submodule 2: Co-evolution enhancement

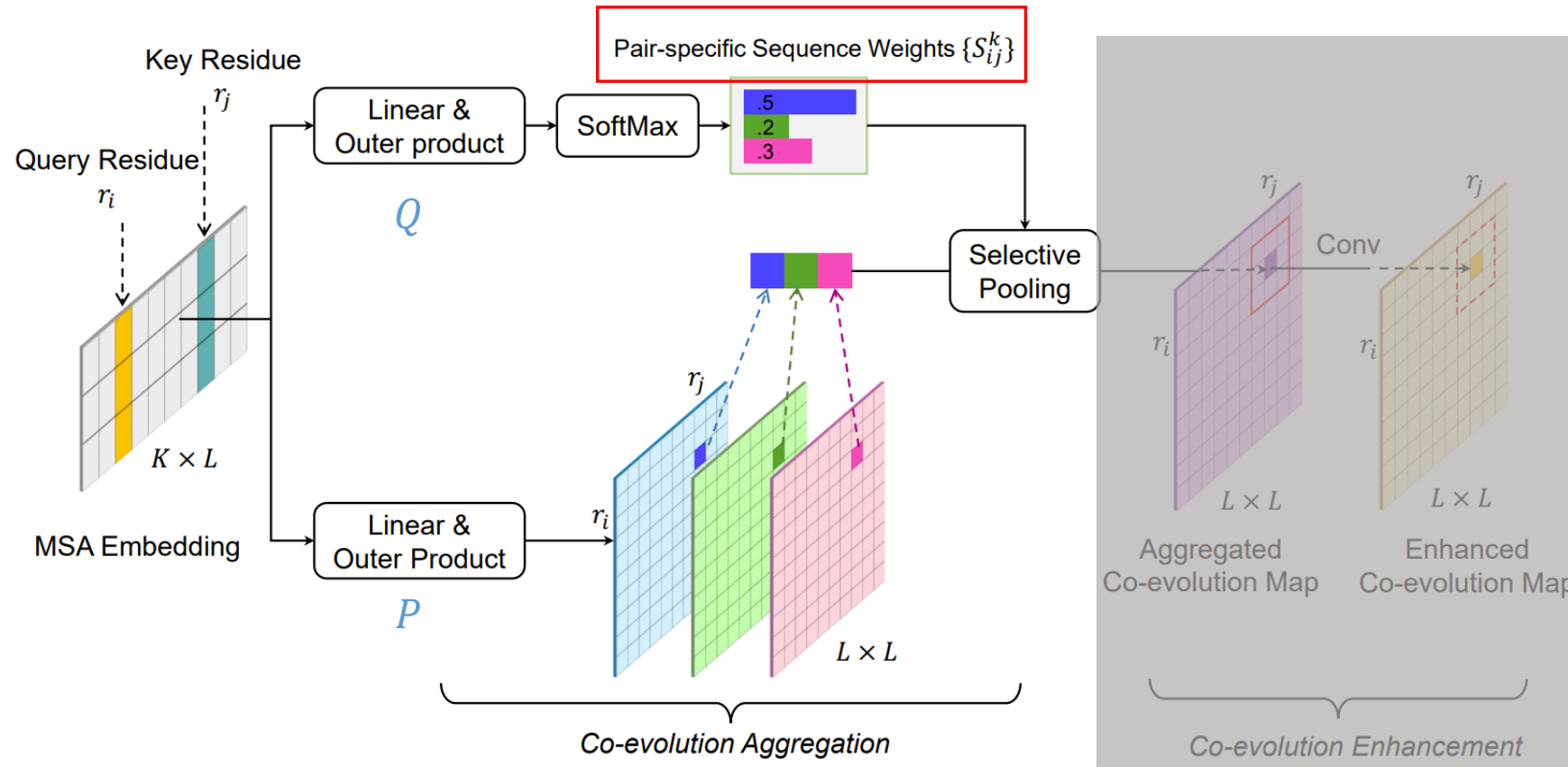# Experimental results

- Quantitative results

Table 1: Comparison on CASP14 and CAMEO (*Precision@L*)

Methods	CASP14			CAMEO
	FM (22)	FM/TBM (14)	TBM (50)	Hard (176)
RaptorX [14]	33.9	58.1	63.1	53.2
trRosetta [15]	31.3	57.6	61.1	50.1
CopulaNet [2]	38.5	62.2	65.5	56.5
CoT-SA (ours)	41.8	59.2	67.9	59.8
CoT (ours)	**47.5**	**63.0**	**76.1**	**66.4**

CoT outperforms CopulaNet, the best of the SOTAs, by 9.0%, 0.8%, 10.6% and 9.9% for Precision@L scores on four kinds of targets, respectively

Experimental results

- Quantitative results

Table 1: Comparison on CASP14 and CAMEO (*Precision@L*)

Methods	CASP14			CAMEO
	FM (22)	FM/TBM (14)	TBM (50)	Hard (176)
RaptorX [14]	33.9	58.1	63.1	53.2
trRosetta [15]	31.3	57.6	61.1	50.1
CopulaNet [2]	38.5	62.2	65.5	56.5
CoT-SA (ours)	41.8	59.2	67.9	59.8
CoT (ours)	**47.5**	**63.0**	**76.1**	**66.4**

A statistical test is conducted, CoT is better than CopulaNet significantly with the p-value 0.003.

Experimental results

- Quantitative results

Table 1: Comparison on CASP14 and CAMEO (*Precision@L*)

	CASP14			CAMEO
Methods	FM (22)	FM/TBM (14)	TBM (50)	Hard (176)
RaptorX [14]	33.9	58.1	63.1	53.2
trRosetta [15]	31.3	57.6	61.1	50.1
CopulaNet [2]	38.5	62.2	65.5	56.5
CoT-SA (ours)	41.8	59.2	67.9	59.8
CoT (ours)	**47.5**	**63.0**	**76.1**	**66.4**

Table 2: Comparison on CASP14. Gr. 368, Gr. 488, and Gr. 010 are the results of the top-3 groups in the CASP14 challenge. CoT† refers to the results of CoT with MSA selection.

|	FM (22)			FM/TBM (14)			TBM (50)		
Method	L	$L/2$	$L/5$	L	$L/2$	$L/5$	L	$L/2$	$L/5$
Gr. 368	41.8	55.7	66.6	64.5	78.6	87.4	73.1	87.1	94.5
Gr. 488	40.4	52.9	65.0	63.6	78.8	88.5	72.0	86.9	93.7
Gr. 010	39.6	53.4	63.8	61.5	77.0	86.8	66.1	80.9	89.5
CoT† (ours)	**50.4**	**65.5**	**76.0**	**66.7**	**81.5**	**90.6**	**78.9**	**92.0**	**98.2**

Official SOTA

Experimental results

- Ablative results

Table 3: Ablations for CoA on CASP14 (*Precision@L*). AGGRE., ENHAN. and SA refer to the co-evolution aggregation submodule, the co-evolution enhancement submodule, and the self-attention module, respectively.

			CASP14			CAMEO
AGGRE.	ENHAN.	SA	FM (22)	FM/TBM (14)	TBM (50)	Hard (176)
Selective Pooling	✓	✓	47.5	63.0	76.1	66.4
Average Pooling	✓	✓	42.7	62.2	73.6	64.2
Selective Pooling		✓	41.6	61.2	70.3	61.8
Selective Pooling	✓		46.4	66.9	74.8	66.3

Why is co-evolution attention better than self-attention?

(a) Self-attention   (b) CoA w/o Enhan.   (c) CoA   (d) Ground-truth Contacts
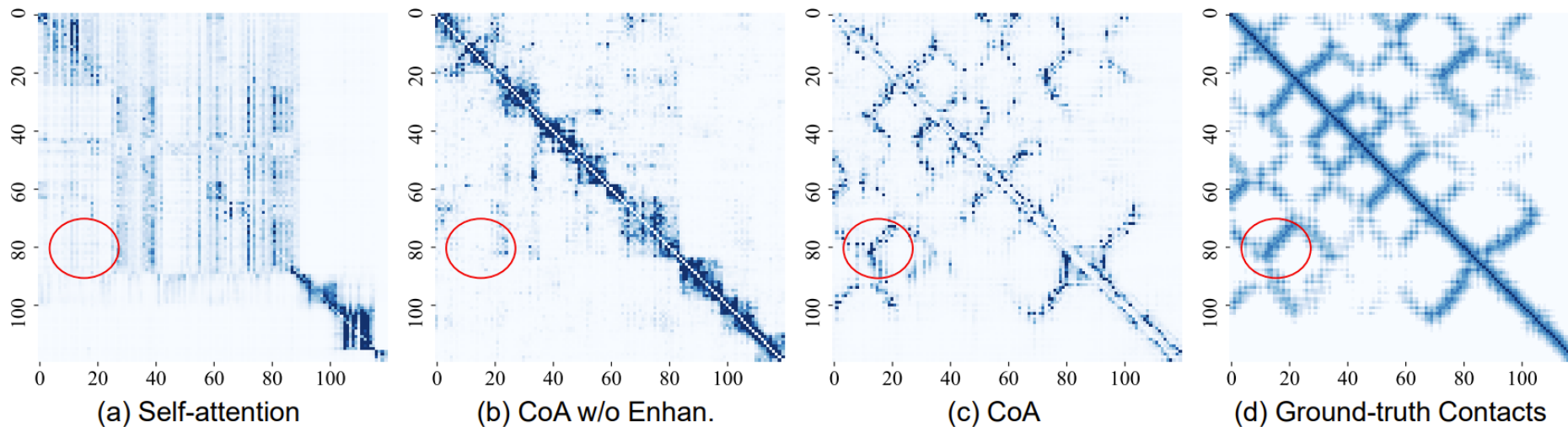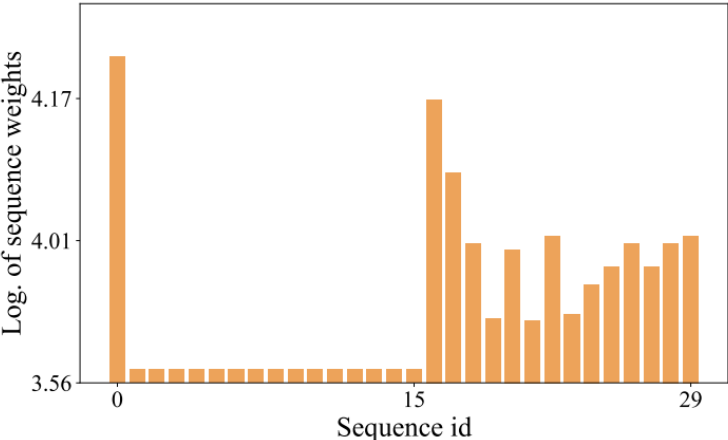
Experimental results
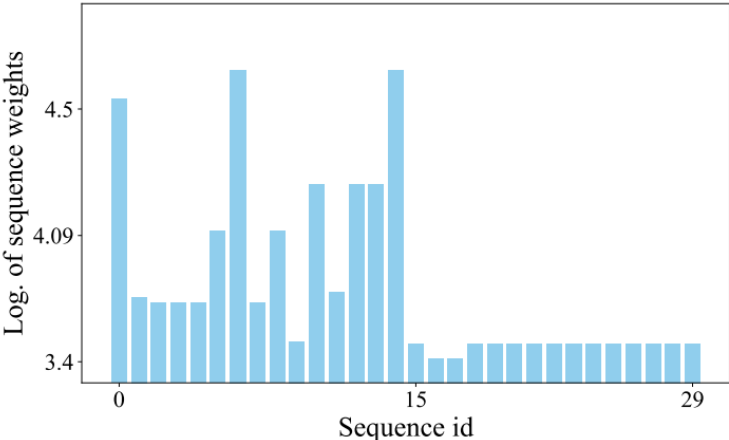
- Qualitative results

Whether selective pooling strategy works or not?
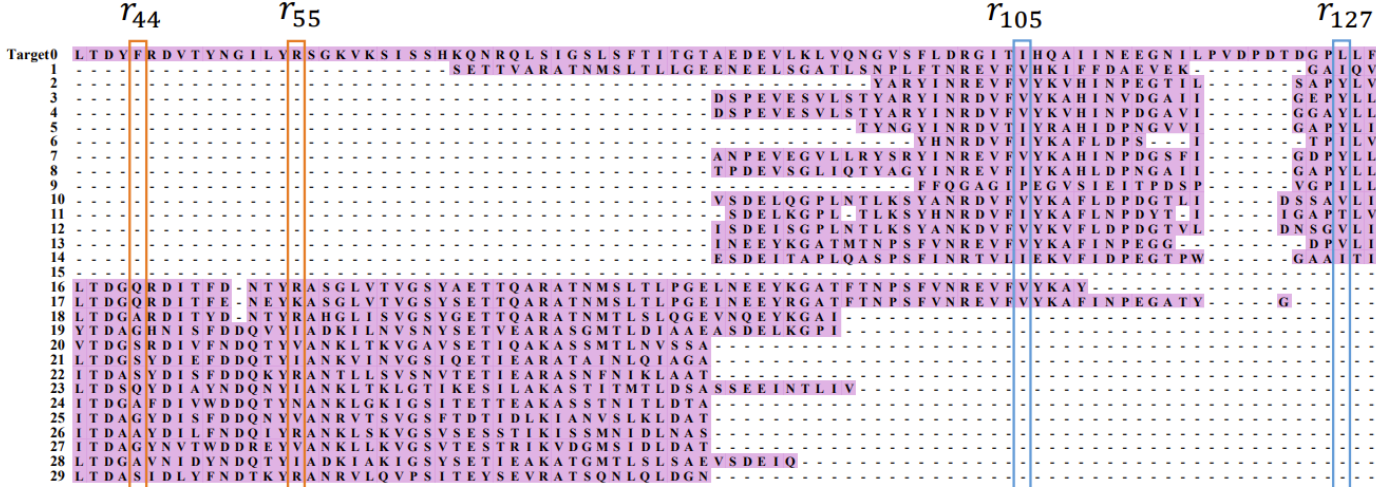
(a)

(b)

(c)

Conclusion and Future work

Concluding remarks

- Jointly modeling multiple homologs

- Selectively aggregating features from different homologs

- High-order interactions are important

Future work

- Proteins with low-depth MSAs are still hard

- Pretraining-based models may be a potential solution

Conclusion and Future work

Concluding remarks

- Jointly modeling multiple homologs

- Selectively aggregating features from different homologs

- High-order interactions are important

Future work

- Proteins with low-depth MSAs are still challenging

- Pretraining-based models may be a potential solution

Thanks

Contact: *mao736488798@stu.xjtu.edu.cn*