



SAGEPHOS: SAGE BIO-COUPLED AND AUGMENTED FUSION FOR PHOSPHORYLATION SITE DETECTION

Authors: **Jingjie Zhang**, Hanqun Cao, Zijun Gao, Xiaorui Wang, Chunbin Gu*

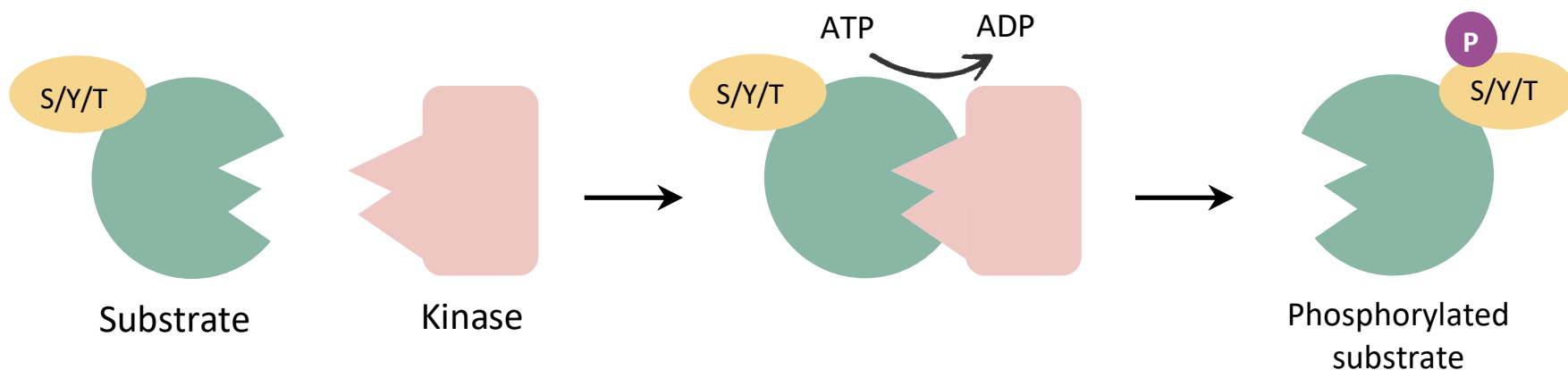
Affiliation: Department of Computer Science and Engineering, **The Chinese University of Hong Kong**

ICLR 2025

Speaker: Jingjie Zhang

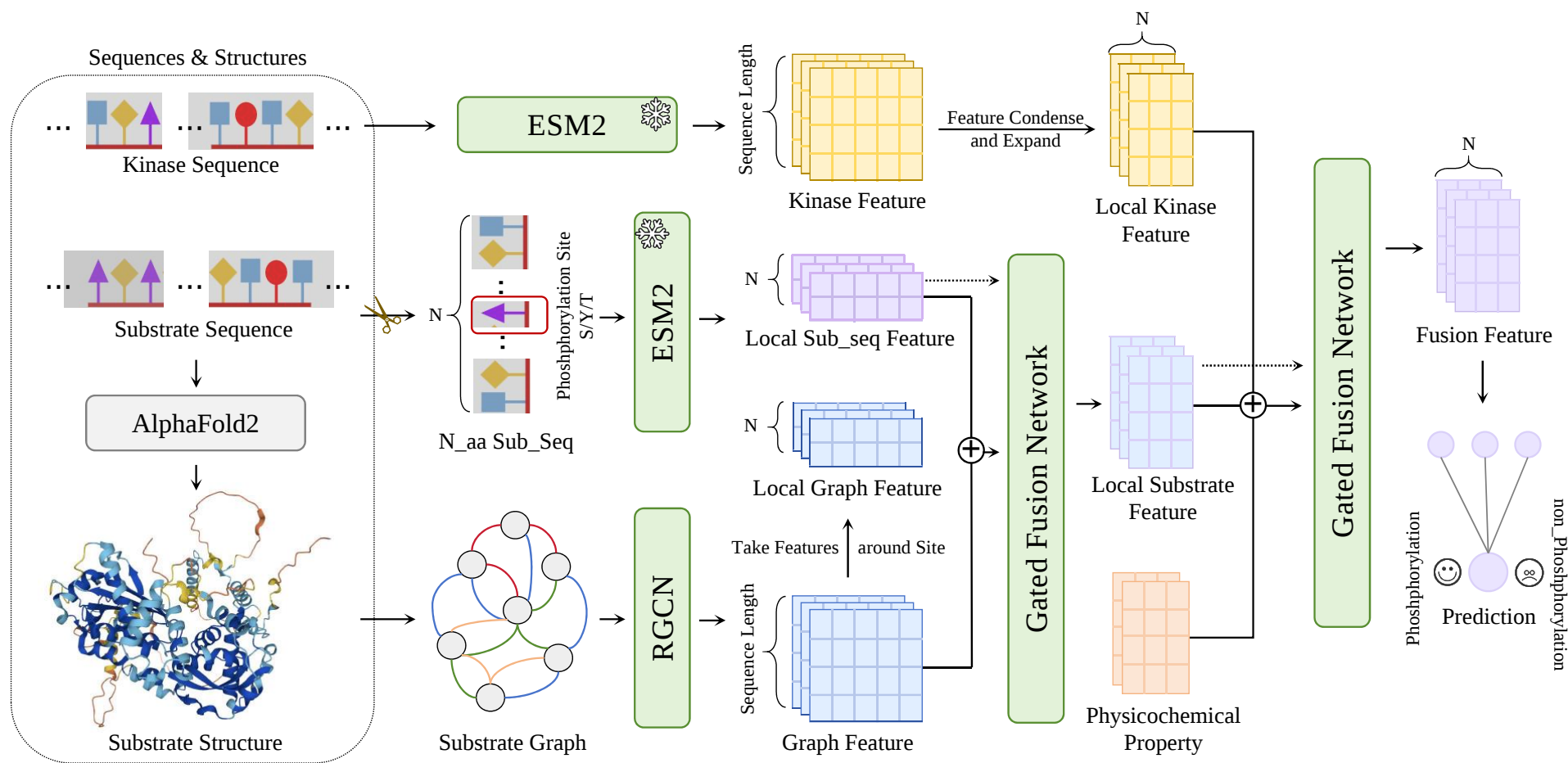
Introduction

- **Phosphorylation** is a crucial post-translational modification.



- Computational methods for phosphorylation site prediction can be categorized into kinase-family-focused and individual kinase-targeted approaches.
- Existing individual kinase-based approaches focus solely on sequence inputs, **neglecting structural information**.

SAGEPhos Framework



- **Multi-Modal:** Sequence & Structure, Inter modality & Intra modality.
- **Key components:** Bio-Coupled Modal Fusion and Bio-Augmented Fusion.

Dataset and Methodology

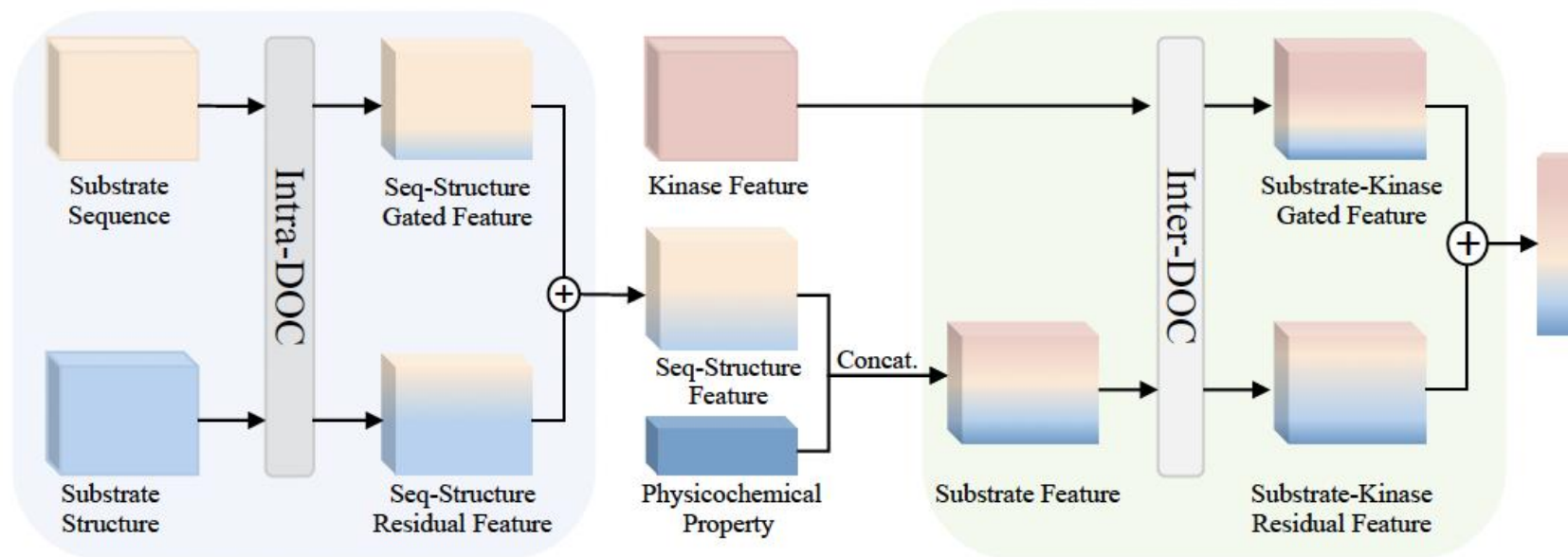
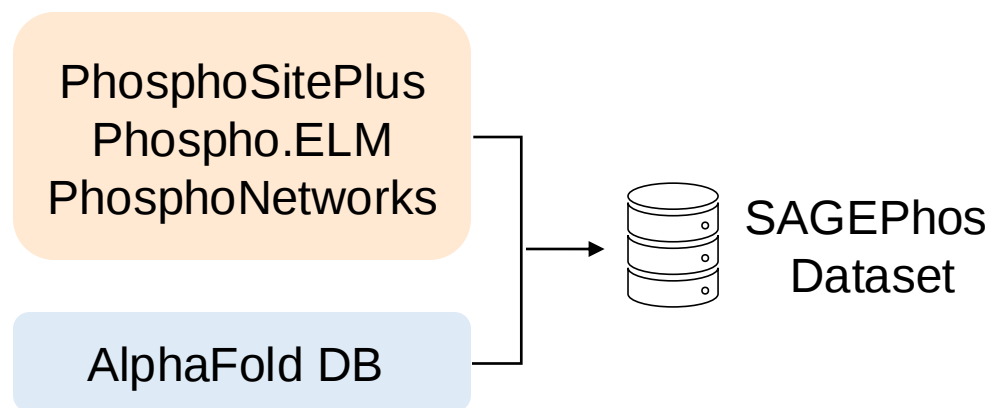


Figure 3: A two-step multi-modal feature fusion process. DOC is a dynamic fusion mechanism.



Experimental Results

Table 1: Comparison of our model SAGEPhos with existing phosphorylation predictors on warm-start dataset. Results in **bold** and underlined are the top-1 and top-2 performances, respectively.³

Method	Warm-start Partition			
	Acc $\uparrow$	AUC-ROC $\uparrow$	AUC-PRC $\uparrow$	FPR $\downarrow$
MusiteDeep	<u>73.5</u> $\pm$ 0.3	78.5 $\pm$ 0.1	7.3 $\pm$ 0.1	26.4 $\pm$ 0.3
DeepPhos	58.5 $\pm$ 0.4	61.9 $\pm$ 0.7	62.0 $\pm$ 0.2	44.2 $\pm$ 3.0
PhosIDNSeq	51.8 $\pm$ 0.0	50.4 $\pm$ 0.2	51.8 $\pm$ 1.8	89.1 $\pm$ 0.0
PhosIDN	51.6 $\pm$ 0.1	50.9 $\pm$ 0.1	51.2 $\pm$ 0.1	93.7 $\pm$ 0.5
EMBER	53.4 $\pm$ 2.1	52.5 $\pm$ 0.4	5.9 $\pm$ 0.1	46.4 $\pm$ 2.3
Phosformer	54.6	72.7	73.6	1.4
Phosformer-ST	64.7	<u>78.9</u>	<u>79.0</u>	95.6
SAGEPhos (Ours)	80.6 $\pm$ 0.2	88.3 $\pm$ 0.1	86.2 $\pm$ 0.2	<u>21.7</u> $\pm$ 0.4

- *Experiment Setting*

- ✓ **Warm-start:** standard train-validation-test splits.
- ✓ **Cold-start:** new kinase or substrate sequences in test set.

- *Performance Comparison*

- ✓ In warm - start, SAGEPhos outperforms existing models with 10% higher accuracy and 12% better AUC-ROC, thanks to its effective integration of structural and interaction data.

Experimental Results

Table 2: Comparison of our model SAGEPhos with existing kinase-specific phosphorylation predictors on cold-start datasets.

Method	Kinase-cold-start Partition				Substrate-cold-start partition			
	Acc $\uparrow$	AUC-ROC $\uparrow$	AUC-PRC $\uparrow$	FPR $\downarrow$	Acc $\uparrow$	AUC-ROC $\uparrow$	AUC-PRC $\uparrow$	FPR $\downarrow$
MusiteDeep	68.3 $\pm$ 3.5	73.8 $\pm$ 0.1	6.4 $\pm$ 0.1	31.7 $\pm$ 3.6	67.5 $\pm$ 1.7	22.2 $\pm$ 0.4	0.8 $\pm$ 0.1	31.8 $\pm$ 1.7
DeepPhos	55.7 $\pm$ 0.5	57.8 $\pm$ 0.6	58.5 $\pm$ 0.6	51.9 $\pm$ 5.3	59.5 $\pm$ 0.4	64.0 $\pm$ 0.3	60.6 $\pm$ 0.3	33.9 $\pm$ 1.9
PhosIDNSeq	55.0 $\pm$ 0.0	48.5 $\pm$ 2.0	49.4 $\pm$ 1.8	89.1 $\pm$ 0.0	52.5 $\pm$ 0.0	52.2 $\pm$ 1.5	52.8 $\pm$ 1.3	89.1 $\pm$ 0.0
PhosIDN	55.0 $\pm$ 0.1	51.1 $\pm$ 0.1	50.7 $\pm$ 0.2	89.1 $\pm$ 0.1	52.3 $\pm$ 0.1	51.0 $\pm$ 0.1	51.4 $\pm$ 0.1	92.2 $\pm$ 0.4
EMBER	52.0 $\pm$ 3.0	48.6 $\pm$ 2.2	4.8 $\pm$ 0.4	4.8 $\pm$ 3.6	49.4 $\pm$ 4.1	49.5 $\pm$ 1.8	6.2 $\pm$ 0.2	50.6 $\pm$ 5.0
Phosformer	55.7	72.3	73.2	1.5	55.2	74.9	75.7	1.6
Phosformer-ST	61.0	72.7	73.3	94.9	63.5	79.3	79.7	97.0
SAGEPhos (Ours)	68.6$\pm$0.8	76.9$\pm$0.3	75.9$\pm$0.3	22.4$\pm$0.6	79.1$\pm$0.4	87.2$\pm$0.2	85.3$\pm$0.4	17.1$\pm$0.4

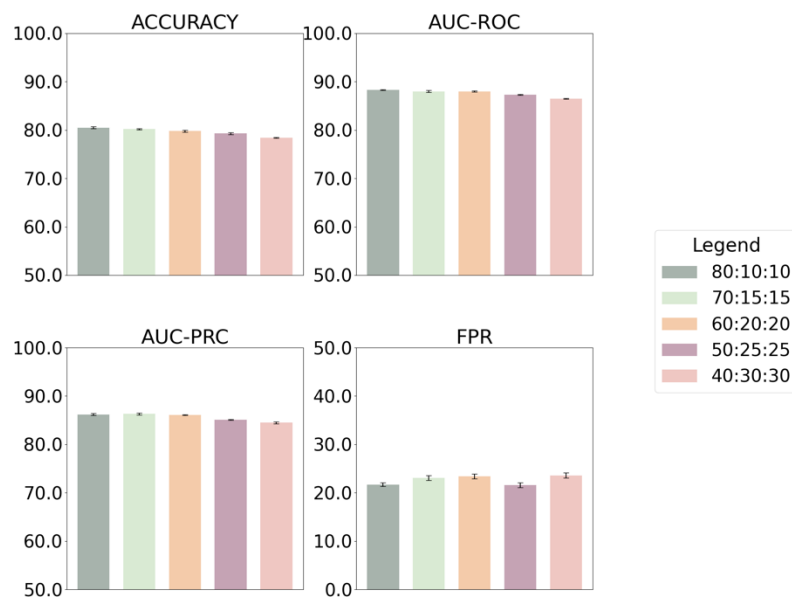


Table A2: Comparison of our model SAGEPhos with MusiteDeep on zero-shot datasets.

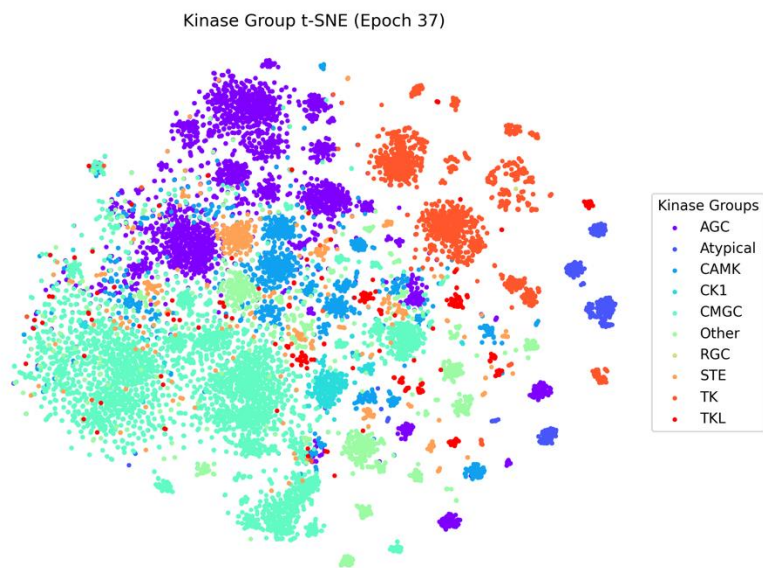
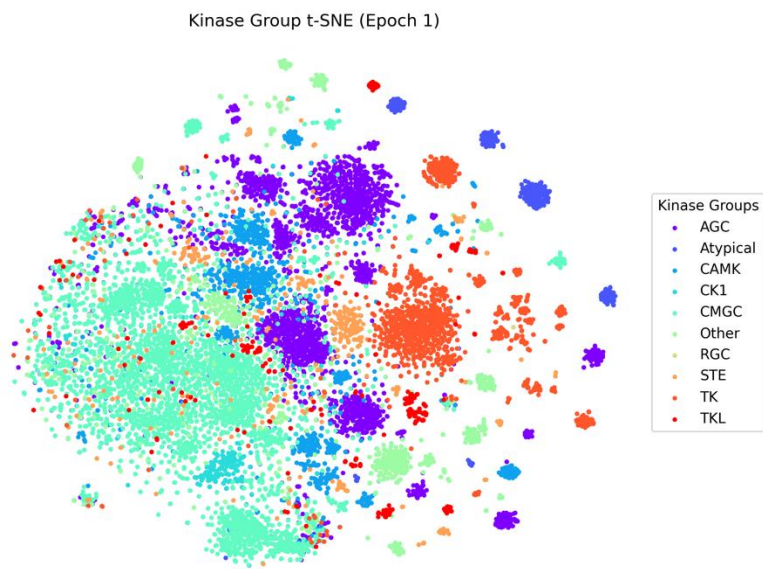
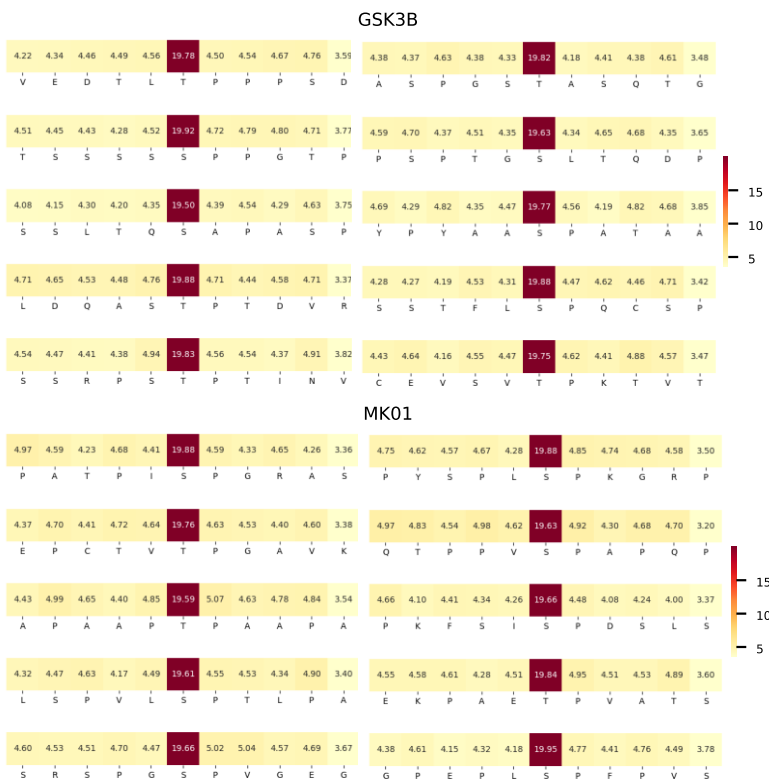
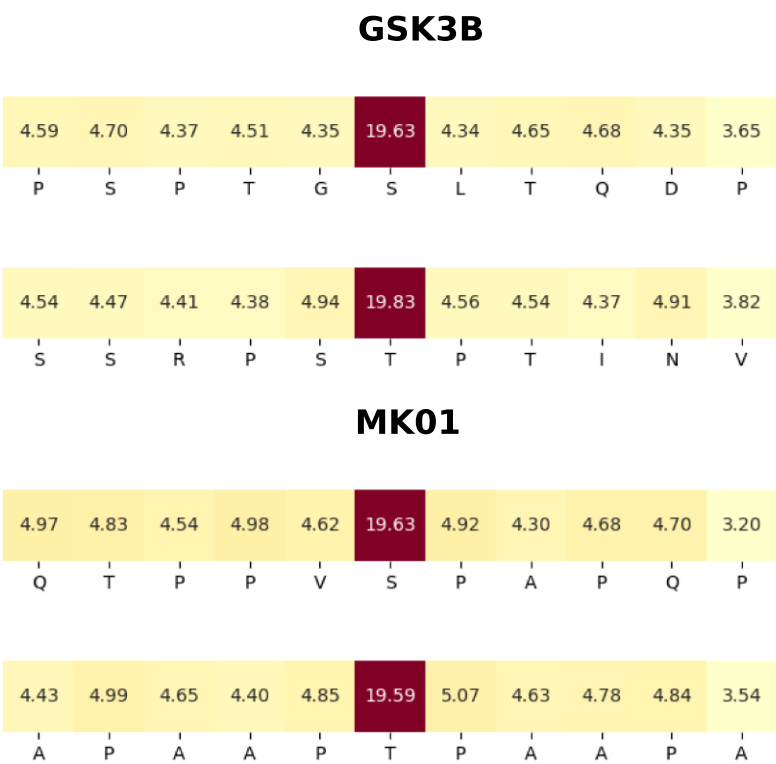
Method	SAGEPhos			MusiteDeep		
	Acc $\uparrow$	AUC-ROC $\uparrow$	AUC-PRC $\uparrow$	Acc $\uparrow$	AUC-ROC $\uparrow$	AUC-PRC $\uparrow$
MYLK4	84.4	97.9	98.1	71.1	94.5	23.8
CDKL1	82.6	90.0	91.0	71.1	91.1	4.0
PHKG2	88.2	95.8	95.3	71.8	90.3	15.2
SRPK3	81.3	88.9	88.2	70.6	95.6	8.7

- MYLK4, CDKL1, PHKG2, SRPK3: kinases absent from our training set

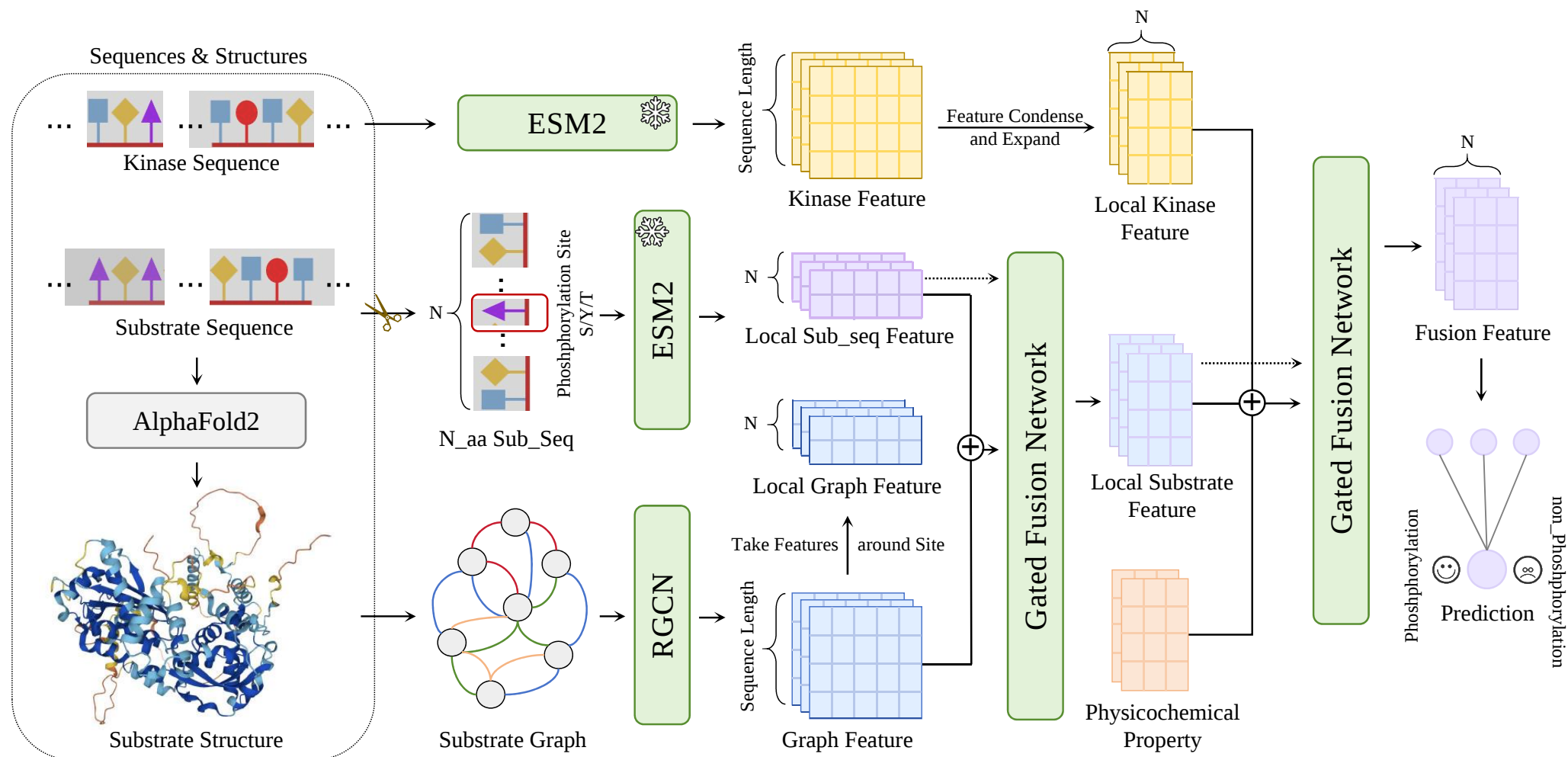
- ✓ Robustness
- ✓ Generalization

Case Study and Visualization

Case study:



Conclusion and Future Work



Thank you for your attention!

- Contact information: jjzhang24@link.cuhk.edu.hk
- GitHub repository: <https://github.com/ZhangJJ26/SAGEPhos>