



# 蛋白质功能预测及 蛋白质-配体绑定位点预测

南京农业大学 人工智能学院

汇报人：朱一亨

2024年04月12日

# 目录

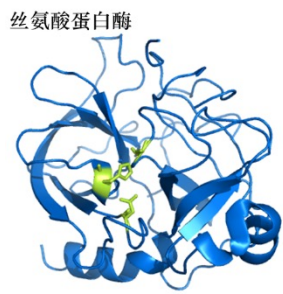
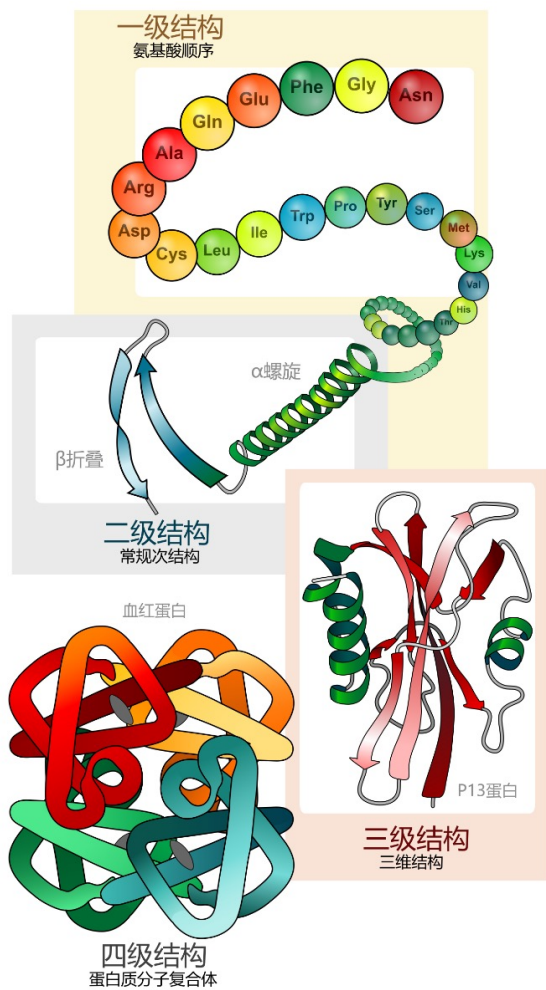
01

蛋白质功能预测

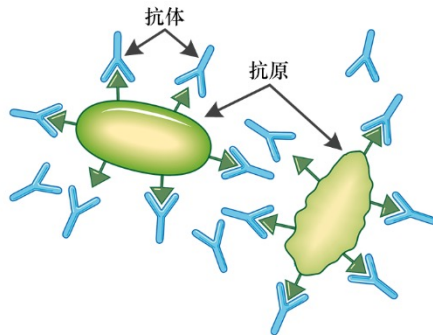
02

蛋白质-配体绑定定位点预测

# 1-1 蛋白质的生物功能



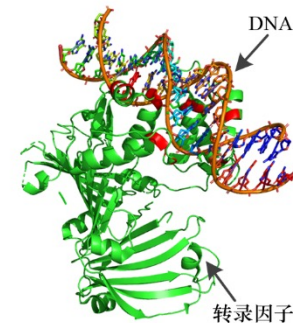
(a) 催化反应



(b) 免疫保护



(c) 运输载体



(d) 基因调控

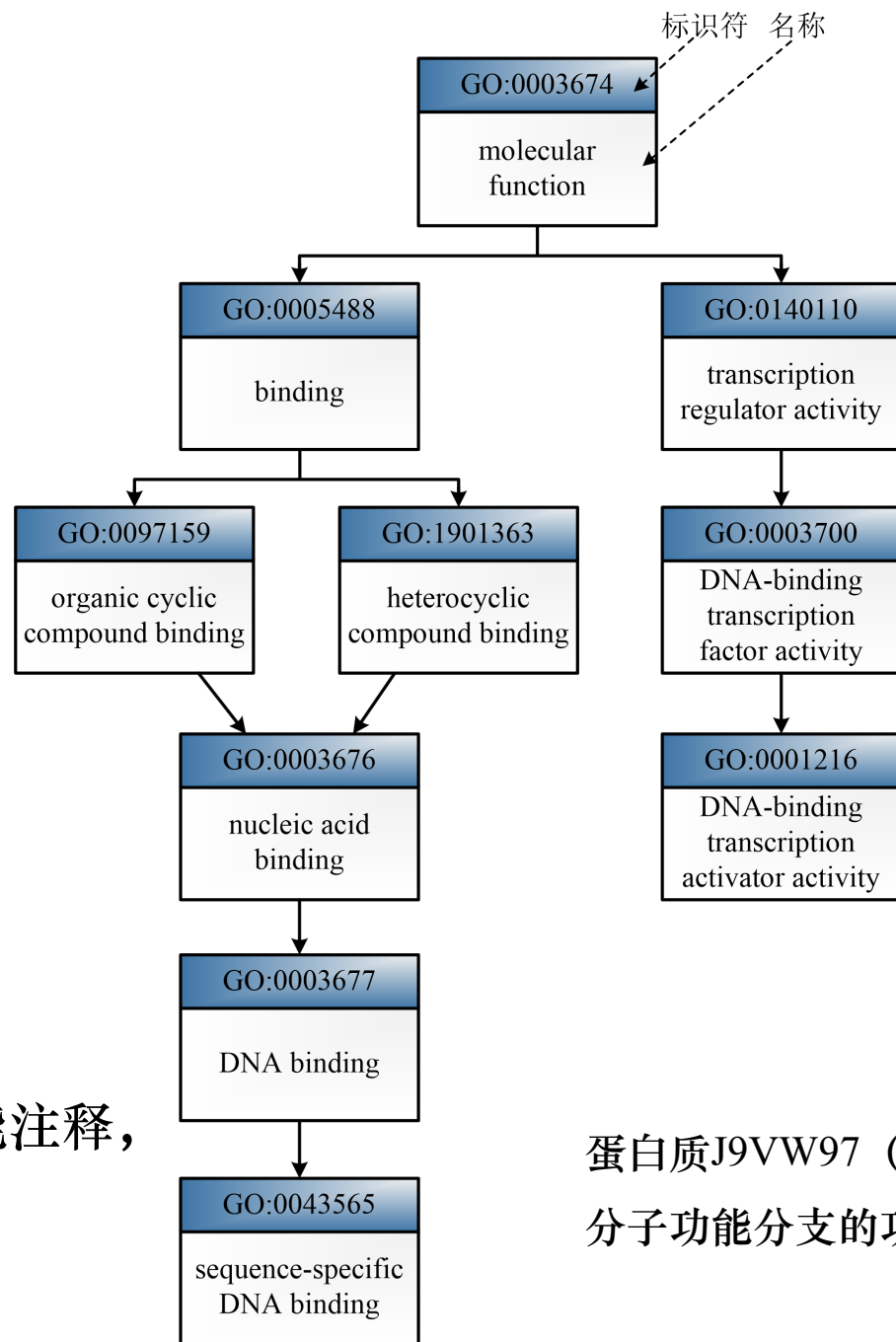
- 识别和分析蛋白质的功能有助于解释各种生命活动现象，并阐明相关疾病的发病机理，进而指导相应的药物设计，以期推动智能医疗的发展。
- 蛋白质功能注释是后基因时代的首要任务之一。

## 1-2 蛋白质的功能注释方法

- 基因本体论 (Gene Ontology, GO)
  - 分子功能 (Molecular Function, MF)
  - 生物过程 (Biological Process, BP)
  - 细胞组件 (Cellular Component, CC)

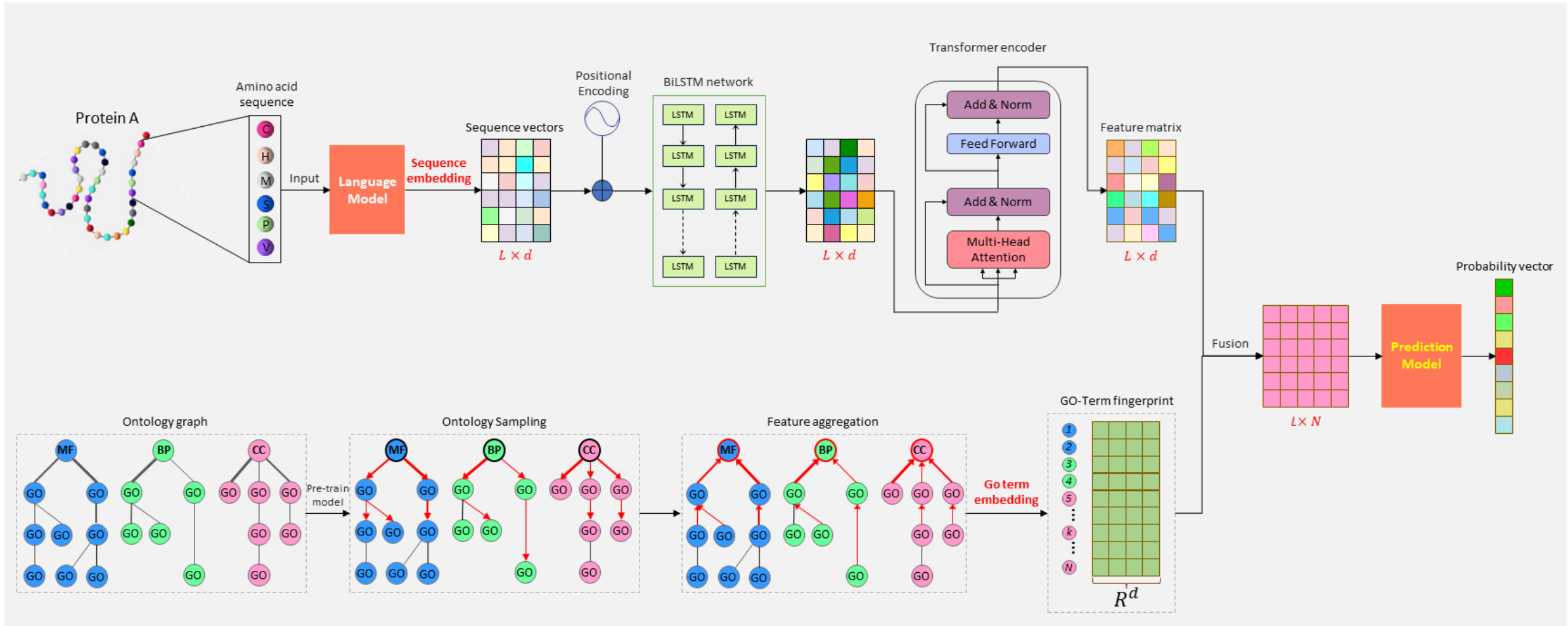
### ➤ 蛋白质功能注释目标

用GO术语对蛋白质在三个分支下分别进行功能注释，形成三张有向无环图。



蛋白质J9VW97 (UniProt ID) 在分子功能分支的功能注释图

# 1-3 DALTGO



The flowchart of DALTGO

## 1-4 论文代码复现

DALTGO 在 CAFA3 数据集上的预测性能

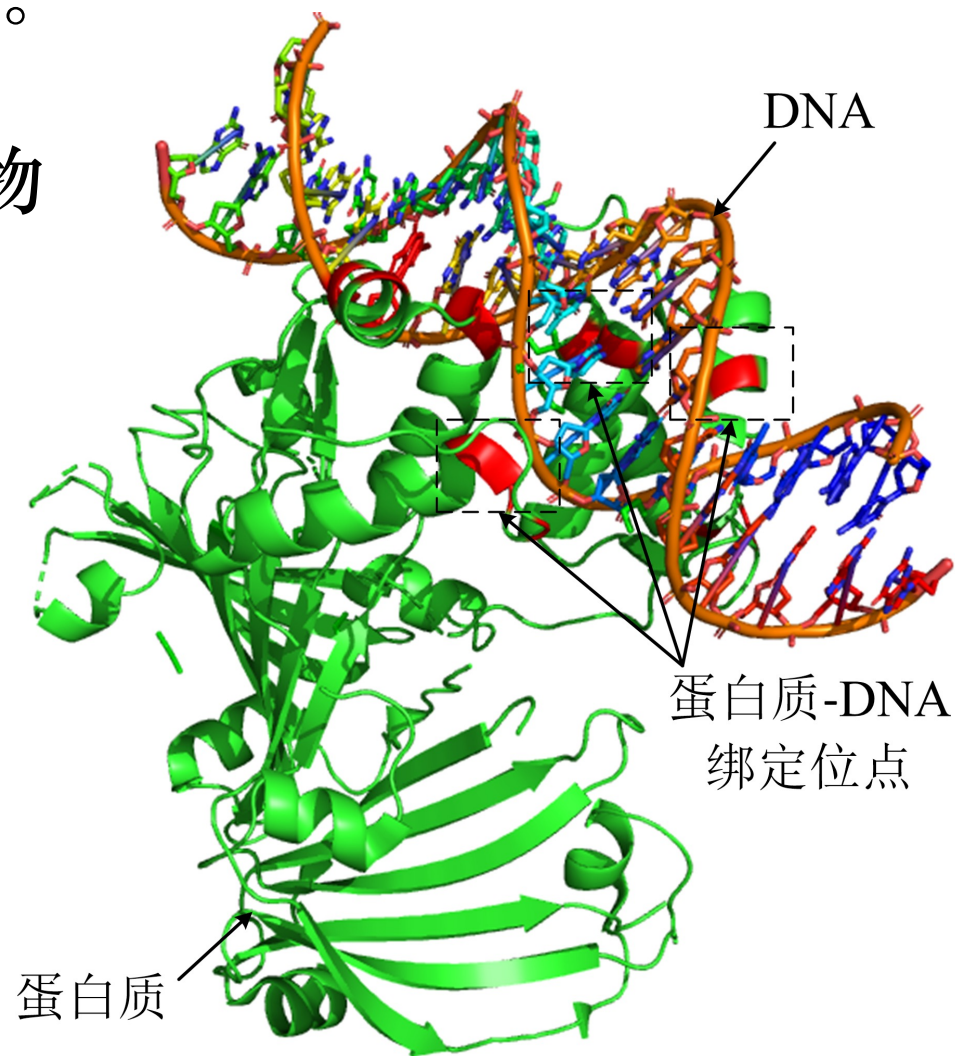
	$F_{\max}$			$S_{\min}$			AUPR			AUROC		
	MF	BP	CC	MF	BP	CC	MF	BP	CC	MF	BP	CC
Xie	0.544	0.441	0.659	16.938	30.997	23.168	0.494	0.318	0.648	0.981	0.980	0.983
Me	0.554	0.381	0.659	16.961	33.215	23.337	0.485	0.266	0.634	0.978	0.967	0.982

## 1-5 后续的改进

- 检查代码和数据集
- 增加后处理步骤
- 特征表示：将 Word2Vec 替换成大语言模型 (ESM2、ProtTrans、Anhk)  
(副作用：性能虽然能提高，但是可能会导致消融实验不明显)

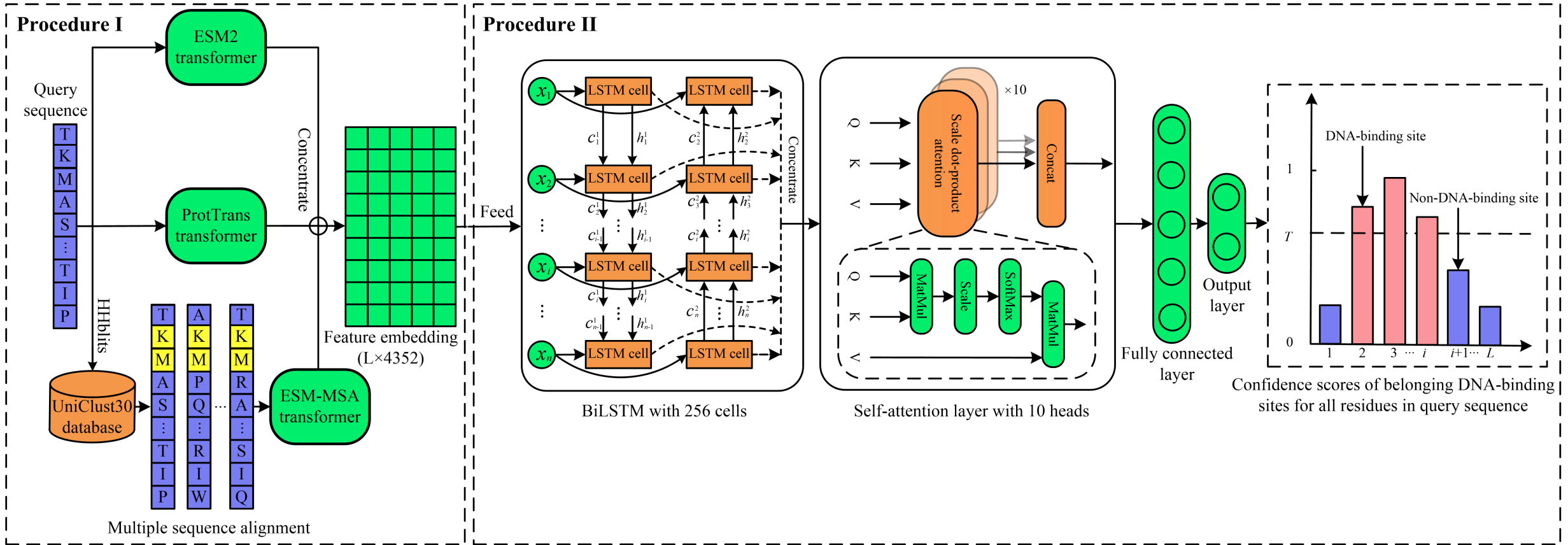
## 2-1 蛋白质-配体绑定位点

- 蛋白质在发挥功能时，并不是孤立存在的。
- 蛋白质 + 配体  $\longrightarrow$  蛋白质-配体复合物



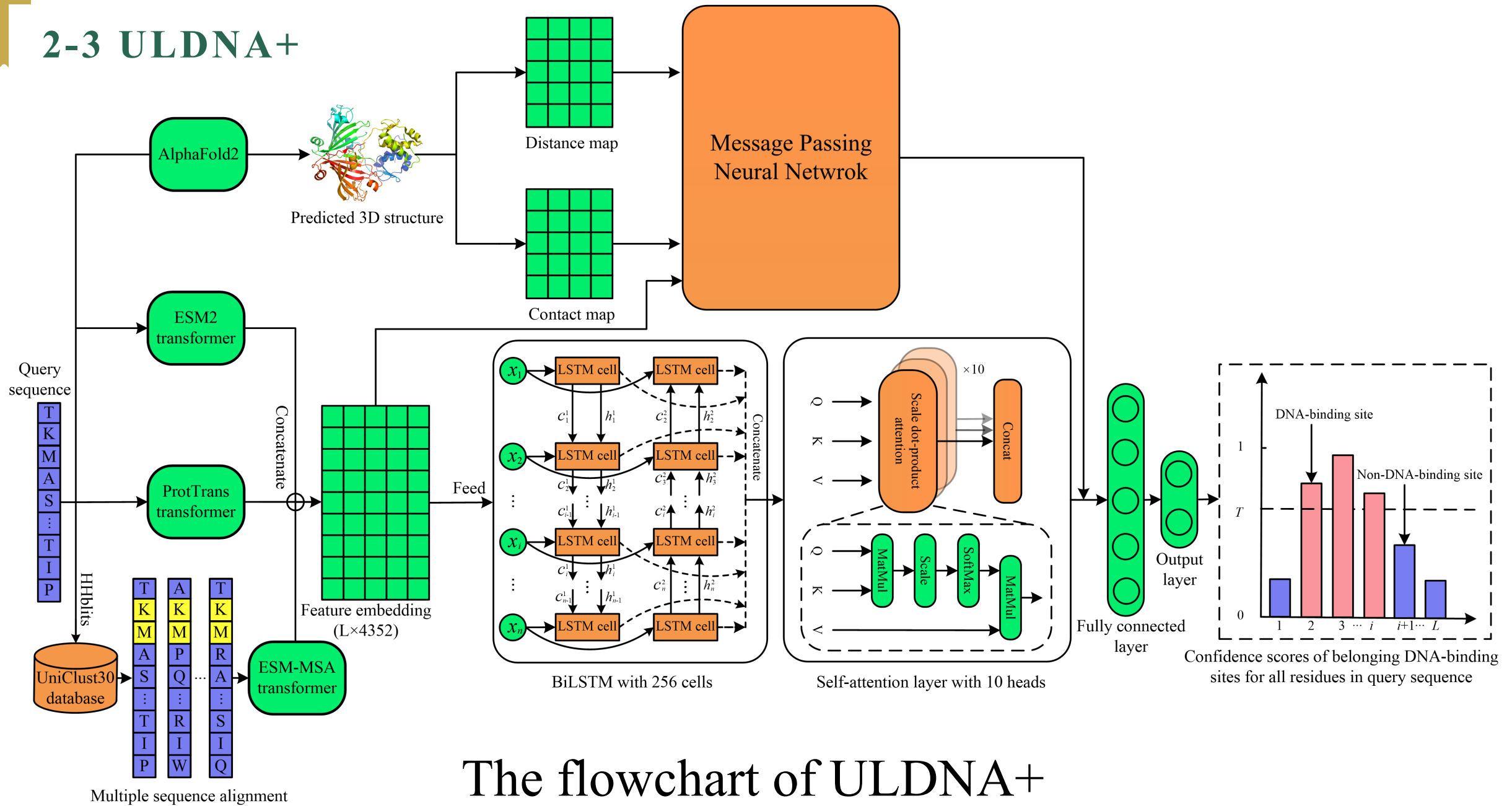


## 2-2 ULDNA



The flowchart of ULDNA

# 2-3 ULDNA+



The flowchart of ULDNA+

## 2-4 性能比较

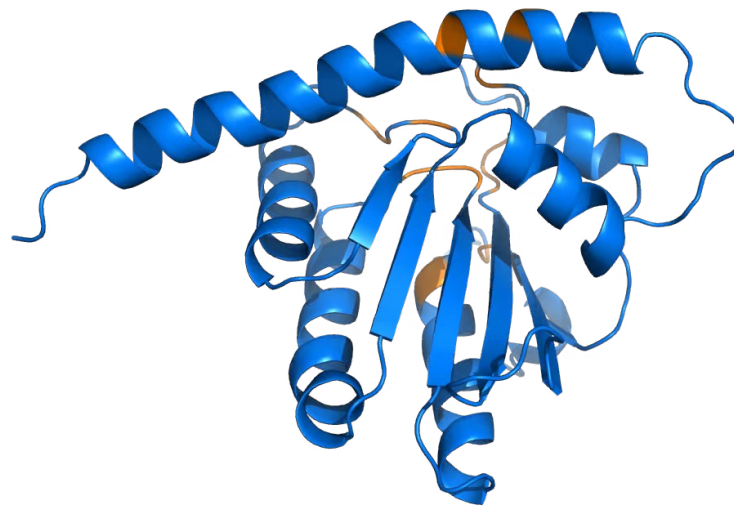
不同模型在 PDNA181 数据集上的测试结果  
(模型训练在 PDNA573 数据集上)

Method	Sen	Spe	Acc	MCC	AUROC
GraphSite (BIB 2022)	0.517	0.958	-	0.397	0.917
ULDNA (BIB 2024)	0.526	0.959	0.940	0.406	0.926
ULDNA+	<b>0.559</b>	<b>0.959</b>	<b>0.942</b>	<b>0.430</b>	<b>0.929</b>

## 2-5 后续的改进

- 缺陷：创新性有限，目前的结果已经很好了，精度很难继续提升。
- 构建Web Server，尝试Web Server论文。

T
K
M
A
S
⋮
T
I
P



谢谢各位老师和同学观看  
请批评指正！