



实现目标：预测酶与底物是否发生反应（酶是否能催化反应），以及后续进一步预测催化效果 KM。

原始数据集：酶 Uniprot ID、底物分子结构图、是否催化、KM

数据准备：由 Uniprot ID 爬取酶 pdb、将酶分子图划分为基序图

核心模块：由邻接矩阵、距离矩阵、动态可学习矩阵三个矩阵组成，动态可学习矩阵可以结合底物特征。将分子图的注意力输出再融入基序图的注意力输出，得到酶特征嵌入。

图神经网络：读取底物分子图，由物理性质组建图，送入图神经网络（有向消息传递图神经网络）输出底物特征指纹。

交叉注意力：首先将酶和底物特征维度对齐，然后分别计算酶-底物特征和底物-酶特征，经过不同 MLP 可输出不同的目标任务。

当前实现方式：由于时间限制，目前蛋白酶采用的是 esm2+GNN 提取特征方案，底物采用的是 transformer 自注意力机制提取特征，即序列特征。特征融合采用的是门控特征融合+交叉注意力，目前架构与上述初始预想架构有差异，后续会逐渐探索修改。项目在实验室已完成 2000 轮训练，在 1600 轮时基本达到拟合，与基线差距约有百分之 4 的准确度，提升空间较大。