

## S4 Abbreviations glossary

<b>AI</b>	Artificial Intelligence
<b>AUC</b>	Area Under Receiver Operating Characteristic curve
<b>AUPR</b>	Area Under Precision-Recall curve
<b>BP</b>	Biological Process
<b>CC</b>	Cellular Component
<b>CLIP</b>	Contrastive Language-Image Pre-training
<b>DDP</b>	Distributed Data Parallel
<b>EC</b>	Enzyme Commission
<b>ESM</b>	Evolutionary Scale Modeling
<b>IF</b>	Inverse Fold
<b>IQR</b>	Inter-Quantile Range
<b>GNN</b>	Graph Neural Network
<b>GO</b>	Gene Ontology
<b>GPT</b>	Generative Pre-trained Transformer
<b>GPU</b>	Graphics Processing Unit
<b>HumanPPI</b>	Human Protein Protein Interactions
<b>InfoNCE</b>	Information Noise-Contrastive Estimation
<b>LoRA</b>	Low-rank Adaptation
<b>MF</b>	Molecular Function
<b>MLM</b>	Masked Language Modeling
<b>MLP</b>	Multi-Layer Perceptron
<b>MR</b>	Median Rank
<b>MSA</b>	Multiple Sequence Alignment
<b>NMR</b>	Nuclear Magnetic Resonance
<b>ProSPECTs</b>	Protein Site Pairs for the Evaluation of Cavity Comparison Tools
<b>SG</b>	Structure Graph
<b>ST</b>	Structure Token