BERT for Proteins

Protein Sequence

| M | K | [M] | L | [M] | V |

↓

Multi-Head Self-Attention

Feed Forward Network

Multi-Head Self-Attention

Feed Forward Network

× 12 Encoder Layers

↓

| M | K | A | L | I | V |

Predicted: [MASK] → Amino Acid

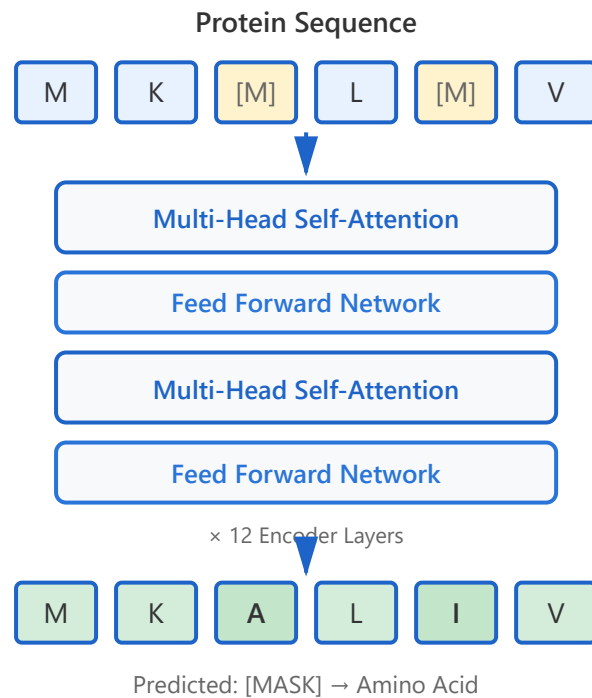## ProtBERT architecture
12-layer bidirectional encoder

Masked language modeling
15% random masking strategy

Attention patterns
Learns residue interactions

Structural insights
Captures 3D contact maps

Function prediction
GO terms, EC numbers