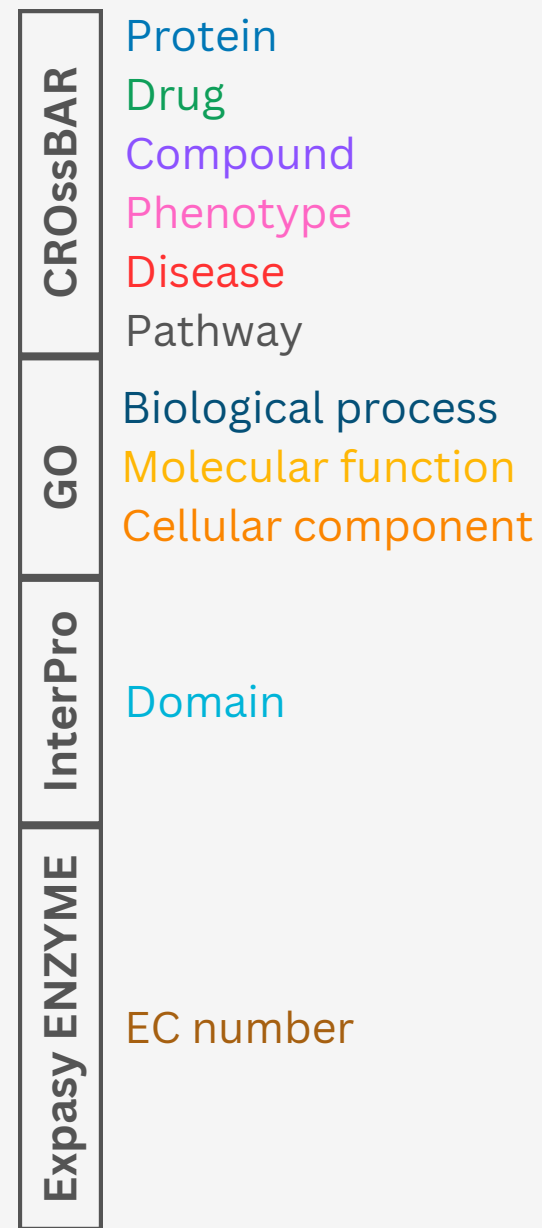
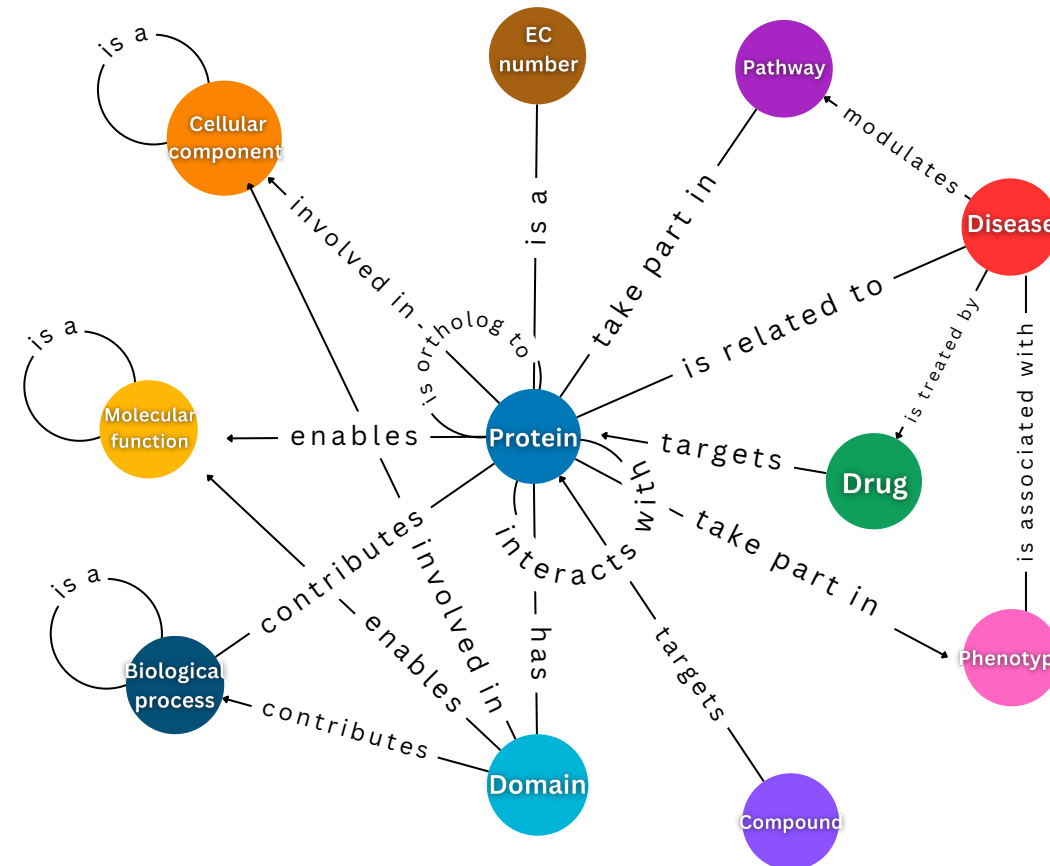


a) Dataset



Data
Integration

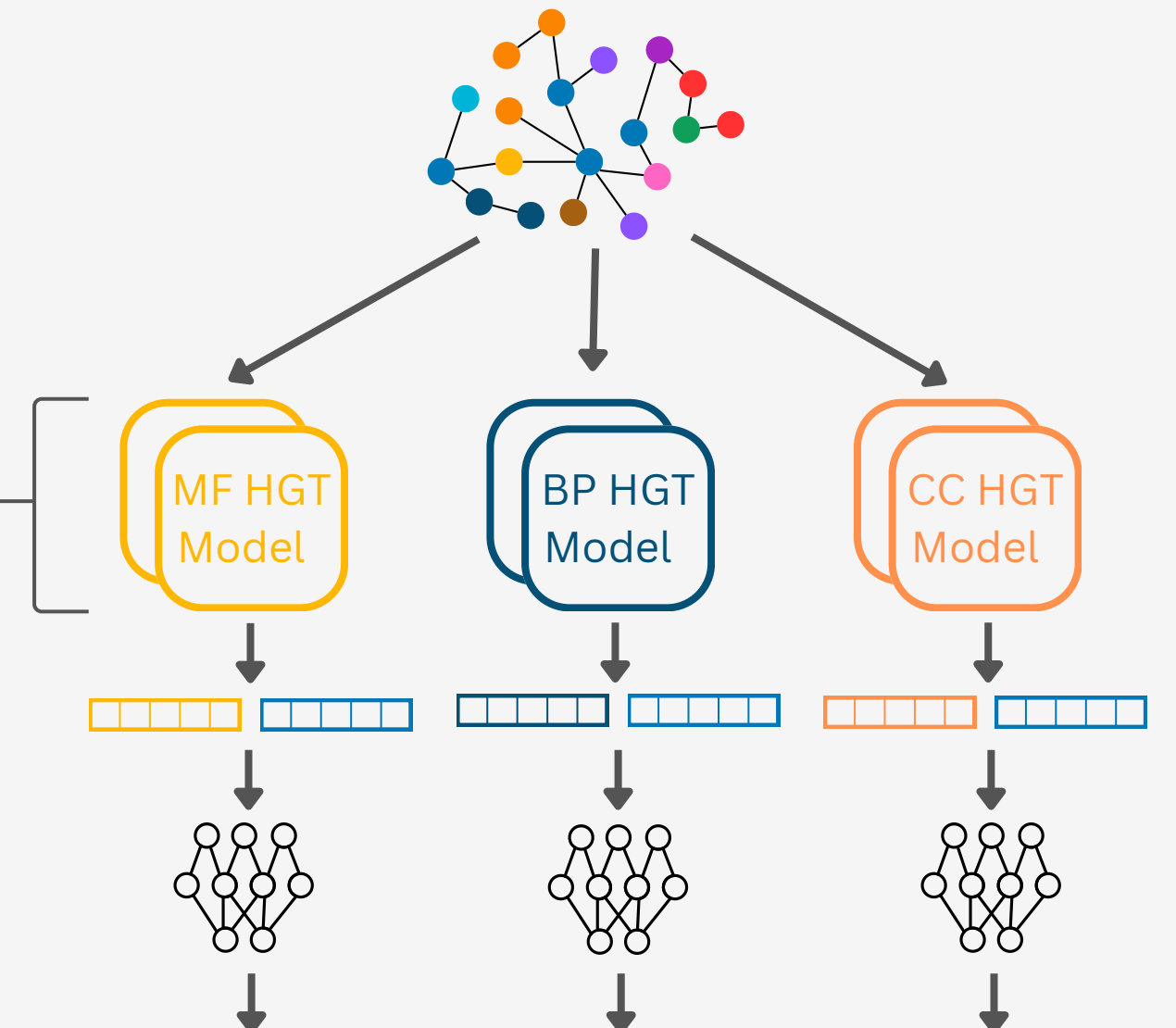


b) Heterogeneous Graph

c) Feature Vectors

Embedding method	(node count x vector size)
TAPE (Protein)	326783 x 768
SELformer (Compound)	133104 x 768
anc2Vec (Biological process)	27855 x 200
anc2Vec (Molecular function)	10955 x 200
dom2vec (Domain)	10229 x 50
CADA (Phenotype)	8971 x 160
TAPE (Drug)	6179 x 768
doc2vec (Disease)	5694 x 100
RXNFP (EC number)	4627 x 256
anc2Vec (Cellular component)	4075 x 200
BioKEEN (Pathway)	3993 x 200

d) Prediction Models



P = 0.78

P = 0.42

P = 0.94

Predicted Probabilities

e) Heterogeneous Graph Transformer Layer

