



Figure S6: Cosine Similarity distributions for OneProt ablations. The plot shows the similarity of a given protein to three groups: the 50 most evolutionarily similar proteins, the 50 most evolutionarily divergent sequences, and 1000 unrelated sequences. ST corresponds to Structure Token modality, SG corresponds to Structure Graph modality, '+' indicates a combination of multiple modalities.