



Supplementary Fig. S9. t-SNE analysis comparing sequence representations from pre-trained ProTrans and ESM2-3B, evaluated using the Calinski-Harabasz Index (CHI), which assigns higher values to better clustering.