

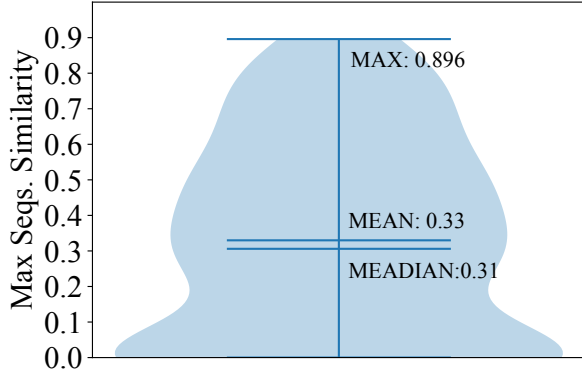


Figure 1: The highest similarity distribution of sequences in the test set, as retrieved from the pre-training set using HHblits.

Table 1: Performance of structure prediction on CASP datasets using additional metrics: pTM, GDT-TS, and LDDT (higher values indicate better performance).

Model	CASP (avg. Depth = 4.6)					
	Zero-Shot			Few-Shot		
	pTM	GDT-TS	LDDT	pTM	GDT-TS	LDDT
AF2 MSA	31.2	29.5	40.5	40.8	41.6	51.9
EvoDiff	40.2	27.3	38.4	44.1	35.6	46.7
MSA-Aug.	39.1	33.2	42.7	44.3	43.2	51.9
EvoGen	36.0	34.8	44.9	42.8	45.1	54.6
MSAGPT	40.0	36.3	45.5	47.7	47.8	56.0
+ RFT	41.2	36.1	45.5	49.8	49.8	56.0
+ DPO	40.0	39.5	48.1	49.8	53.1	60.0

Table 2: Performance of structure prediction on PDB datasets using additional metrics: pTM, GDT-TS, and LDDT (higher values indicate better performance).

Model	PDB (avg. Depth = 2.6)					
	Zero-Shot			Few-Shot		
	pTM	GDT-TS	LDDT	pTM	GDT-TS	LDDT
AF2 MSA	40.3	43.7	56.6	46.2	50.4	62.8
EvoDiff	47.4	43.3	55.9	45.5	45.0	57.1
MSA-Aug.	44.8	47.6	59.8	48.8	52.8	65.2
EvoGen	44.3	47.0	59.4	49.1	53.0	65.4
MSAGPT	48.6	50.6	63.3	53.7	57.0	69.2
+ RFT	48.6	50.5	62.4	53.6	57.2	69.0
+ DPO	47.6	50.9	63.2	52.7	57.8	69.5

Table 3: Performance of structure prediction on CASP datasets using additional metrics: pTM, GDT-TS, and LDDT (higher values indicate better performance).

Model	CAMEO (avg. Depth = 8.5)					
	Zero-Shot			Few-Shot		
	pTM	GDT-TS	LDDT	pTM	GDT-TS	LDDT
AF2 MSA	41.3	60.8	66.7	58.2	75.5	80.3
EvoDiff	54.2	62.4	68.1	48.1	63.0	67.7
MSA-Aug.	47.2	62.3	68.3	58.4	75.1	80.6
EvoGen	45.0	63.5	68.8	61.8	78.8	82.8
MSAGPT	51.4	64.3	69.4	63.6	78.8	83.2
+ RFT	48.1	63.5	68.1	62.4	79.0	83.1
+ DPO	49.4	64.4	68.7	63.0	79.4	83.5

Table 4: Zero-shot evaluation on approximate 8k PDB Datasets (Released before 2024-01-22).

Model	pTM	pLDDT	TM	GDT-TS	LDDT
AF MSA	28.9	69.6	42.0	62.4	67.0
EvoDiff	34.7	70.8	44.0	64.1	69.9
MSA-Aug.	33.2	74.3	45.8	65.9	71.7
EvoGen	31.7	71.6	44.6	64.6	70.5
MSAGPT	37.9	80.4	50.7	70.7	76.1
+RFT & DPO	37.9	79.7	51.6	71.4	76.9

Table 5: Comparison of structure prediction methods: single sequence, augmentation with random MSA, and augmentation with MSAGPT-generated MSA from a sampled 1k PDB dataset.

Model	pTM	pLDDT	TM	GDT-TS	LDDT
Single	39.8	54.8	45.5	43.5	55.9
+ Random	43.6	66.7	45.4	43.3	55.7
+ MSAGPT	53.7	71.3	59.7	57.3	68.9

Table 6: Performance comparison in MSA-abundant scenarios across all 194 cases in the CAMEO datasets.

# Natural MSA	# Virtual MSA	TM	GDT-TS
0	0	39.1	28.5
128	0	85.1	82.0
128	128	85.3	82.2
256	0	86.2	83.8
256	256	86.0	83.5
512	0	86.5	84.2
512	512	86.4	84.0