Supplementary Materials

1. Proteins for Multiple Structures Alignment

Entity ID	Resolution (Å)	Sequence	Auth Asym ID	Entity ID	Polymer Entity ID
1AAP_1	1.5	VREVCSEQAETGPCRAMISRWYFDVTE GKCAPFFYGGCGGNRNNFDTEEYCMA VCGSA	A	1	1AAP
1BUN_2	2.45	RKRHPDCDKPPDTKICQTVVRAFYYKP SAKRCVQFRYGGCNGNGNHFKSDHLC RCECLEYR	B	2	1BUN
1DTX_1	2.2	QPRRKLCILHRNPGRCYDKIPAFYYNQ KKKQCERFDWSGCGGNSNRFKTIEECR RTCIG	A	1	1DTX
1FAK_4	2.1	APDFCLEPPYDGPCRALHLRYFYNAKAG LCQTFYYGGCLAKRNNFESAEDCMRTC	I	4	1FAK
1KTH_1	0.95	ETDICKLPKDEGTCRDFILKWYYDPNT KSCARFWYGGCGGNENKFGSQKECEK VCAPV	A	1	1KTH
1YC0_2	2.6	MKHQHQHQHQHQHQMHQTEDYCLA SNKVGRCRGSFPRWYYDPTEQICKSFV YGGCLGNKNNYLREEECILACRGV	I	2	1YC0
1ZR0_2	1.802	PTGNNAEICLLPLDYGPCRALLLRYYYD RYTQSCRQFLYGGCEGNANNFYTWEA CDDACWRIE	B	2	1ZR0
3BYB_1	1.63	KDRPDFCELPADTGPCRVRFPSFYYNPD EKKCLEFIYGGCEGNANNFITKEECEST CAA	A	1	3BYB
3M7Q_2	1.7	EAEASICSEPKKVGRCKGYFPRFYFDSET GKCTPFIYGGCGGNGNNFETLHQCRAI CRALG	B	2	3M7Q
4BQD_1	2.48	DSEEDEEHTIITDTELPPLKLMHSFCAFK ADDGPCKAIMKRFFFNIFTRQCEEFIYG GCEGNQNRFESLEECKKMCTRD	A	1	4BQD
4DTG_3	1.8	QEKPDFCFLEEDPGICRGYITRYFYNNQ TKQCERFKYGGCLGNMNNFETLEECK NICEDGHHHHHH	K	3	4DTG
4ISO_2	2.01	QTEDYCLASNKVGRCRGSFPRWYYDPT EQICKSFVYGGCLGNKNNYLREEECILA CRGVQ	B	2	4ISO
4NTW_2	2.07	QIRPAFCYEDPPFFQKCGAFVDSYYFNR SRITCVHFFYGQCDVNQNHFTTMSECN RVCHG	B	2	4NTW
4U30_2	2.5	TVAACANLPIVRGPCRAFIQLWAFDAVK GKCVLFPYGGCQGNGNKFYSEKECREY CGVP	X	2	4U30

4U32_1	1.65	IHDFCLVSKVVGRCRASMPRWWYNVT DGSCQLFVYGGCDGNSNNYLTKEECLK KC	X	1	4U32
5M4V_1	1.06	RPSFCNLPVKPGPCKAFFSAFYYSQKTN KCHSFTYGGCKGNANRFSTLEKCRRTC VG	A	1	5M4V
5NX1_3	1.853	AMISRWYFDVTEGKCAPFFYGGCGGNR NNFDTEEYCMAVCGSAIPRHHHHHHA AAN	D	3	5NX1
5PTI_1	1	RPDFCLEPPYTGPCKARIIRYFYNAKAGL CQTFVYGGCRAKRNNFKSAEDCMRTC GGA	A	1	5PTI
5YV7_1	2.395	WQPPWYCKEPVRIGSCKKQFSSFYFKW TAKKCLPFLFSGCGGNANRFQTIGECR KKCLGK	A	1	5YV7
6Q61_1	1.3	KDRPSLCDLPADSGSGTKAEKRIYYNSA RKQCLRFDYTGQGGNENNFRRTYDCA RTCLYTA	A	1	6Q61
6YHY_1	1.55	DRPSYCNLPADSGSGTKSEQRIYYNSAR KQCLTFTYNGQGGNENNFRRTYDCRR TCQYP	A	1	6YHY

Table S1. Customized tabular report from PDB Advanced Search for the 21 selected protein structures necessary for the Multiple Structure Alignment. There are reported: Entry ID, Resolution (Å), Sequence, Auth Asym ID (chain), Entity ID and Entry Id (Polymer Entity Identifiers).

2. Performance Evaluation

TH	Accuracy	MCC	F1	TP	TN	FP	FN
0,1	0,95856	0,11930	0,02926	125	191706	8294	0
0,01	0,99652	0,38952	0,26427	125	199304	696	0
0,001	0,99967	0,80885	0,79114	125	199934	66	0
0,0001	0,99995	0,96223	0,96154	125	199990	10	0
1,00E-05	0,99998	0,98057	0,98039	125	199995	5	0
1,00E-06	0,99999	0,98820	0,98814	125	199997	3	0
1,00E-07	1,00000	0,99602	0,99602	125	199999	1	0
1,00E-08	1,00000	0,99599	0,99598	124	200000	0	1
1,00E-09	1,00000	0,99599	0,99598	124	200000	0	1
0	1,00000	0,99599	0,99598	124	200000	0	1
1,00E-11	1,00000	0,99599	0,99598	124	200000	0	1

1,00E-12	1,00000	0,99599	0,99598	124	200000	0	1
1,00E-13	1,00000	0,99599	0,99598	124	200000	0	1
1,00E-14	0,99998	0,98386	0,98374	121	200000	0	4
1,00E-15	0,99998	0,98386	0,98374	121	200000	0	4

Table S2. Summary table of the testing to select the best E-Value threshold resulting from Set 1. There are reported: E-Value threshold, Accuracy, Matthews Correlation Coefficient, F1 score, True Positives, True Negatives, False Positives and False Negatives.

TH	Accuracy	MCC	F1	TP	TN	FP	FN
0,1	0,95829	0,11890	0,02908	125	191652	8348	0
0,01	0,99652	0,38928	0,26399	125	199303	697	0
0,001	0,99969	0,81528	0,79872	125	199937	63	0
0,0001	0,99992	0,94152	0,93985	125	199984	16	0
1,00E-05	0,99999	0,98820	0,98814	125	199997	3	0
1,00E-06	0,99999	0,99209	0,99206	125	199998	2	0
1,00E-07	0,99999	0,99200	0,99200	124	199999	1	1
1,00E-08	1,00000	0,99599	0,99598	124	200000	0	1
1,00E-09	1,00000	0,99599	0,99598	124	200000	0	1
0	0,99999	0,99196	0,99194	123	200000	0	2
1,00E-11	0,99999	0,98792	0,98785	122	200000	0	3
1,00E-12	0,99999	0,98792	0,98785	122	200000	0	3
1,00E-13	0,99998	0,97978	0,97959	120	200000	0	5
1,00E-14	0,99996	0,96745	0,96694	117	200000	0	8
1,00E-15	0,99996	0,96745	0,96694	117	200000	0	8

Table S3. Summary table of the testing to select the best E-Value threshold resulting from Set 2. There are reported: E-Value threshold, Accuracy, Matthews Correlation Coefficient, F1 score, True Positives, True Negatives, False Positives and False Negatives.

	Actual Positive	Actual Negative
Predicted Positive	124	0
Predicted Negative	1	200000

TH = 1,00E-08
Accuracy = 0,99999
MCC = 0,99536
F1 = 0,99535

Table S4. Confusion Matrix of the final performance testing on the Test Set. It results in 124 True Positives, 200000 True Negatives, 1 False Negative and 0 False Positive. There are reported also the metrics: E-Value threshold, Accuracy, Matthews Correlation Coefficient, F1 score.