Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2023

Supplementary Information:

Evaluation of DNA-protein complex structures using the deep learning method

Chengwei Zeng[1+], Yiren Jian[2+], Chen Zhuo[1], Anbang Li[1], Chen Zeng[3], and Yunjie Zhao[1]*

[1]Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan, 430079, China

[2]Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA

[3]Department of Physics, The George Washington University, Washington, DC 20052, USA

[+] These authors contributed equally.

*Correspondence:

Yunjie Zhao. Email: yjzhaowh@ccnu.edu.cn

Supplementary Table 1. The atom types and charge on DNA nucleotides.

Nucleotide	Atom	Charge	Nucleotide	Atom	Charge
DA	P	1.165900	DG	P	1.165900
	OP1	-0.776100		OP1	-0.776100
	OP2	-0.776100		OP2	-0.776100
	O5'	-0.495400		O5'	-0.495400
	C5'	-0.006900		C5'	-0.006900
	C4'	0.162900		C4'	0.162900
	O4'	-0.369100		O4'	-0.369100
	C1'	0.043100		C1'	0.035800
	N9	-0.026800		N9	0.057700
	C8	0.160700		C8	0.073600
	N7	-0.617500		N7	-0.572500
	C5	0.072500		C5	0.199100
	C6	0.689700		C6	0.491800
	N6	-0.912300		O6	-0.569900
	N1	-0.762400		N1	-0.505300
	C2	0.571600		C2	0.743200
	N3	-0.741700		N2	-0.923000
	C4	0.380000		N3	-0.663600
	C3'	0.071300		C4	0.181400
	C2'	-0.085400		C3'	0.071300
	O3'	-0.523200		C2'	-0.085400
				O3'	-0.523200
DC	P	1.165900	DT	P	1.165900
	OP1	-0.776100		OP1	-0.776100
	OP2	-0.776100		OP2	-0.776100
	O5'	-0.495400		O5'	-0.495400

C5'	-0.006900	C5'	-0.006900
C4'	0.162900	C4'	0.162900
O4'	-0.369100	O4'	-0.369100
C1'	-0.011600	C1'	0.068000
N1	-0.033900	N1	-0.023900
C6	-0.018300	C6	-0.220900
C5	-0.522200	C5	0.002500
C4	0.843900	C7	-0.226900
N4	-0.977300	C4	0.519400
N3	-0.774800	O4	-0.556300
C2	0.795900	N3	-0.434000
O2	-0.654800	C2	0.567700
C3'	0.071300	O2	-0.588100
C2'	-0.085400	C3'	0.071300
O3'	-0.523200	C2'	-0.085400
		O3'	-0.523200

Supplementary Table 2. The atom types and charge on protein residues.

Residue	Atom	Charge	Residue	Atom	Charge
ALA	N	-0.415700	HIP	N	-0.347900
	CA	0.033700		CA	-0.135400
	CB	-0.182500		CB	-0.041400
	C	0.597300		CG	-0.001200
	O	-0.567900		ND1	-0.151300
ARG	N	-0.347900		CE1	-0.017000
	CA	-0.263700		NE2	-0.171800
	CB	-0.000700		CD2	-0.114100
	CG	0.039000		C	0.734100
	CD	0.048600		O	-0.589400
	NE	-0.529500	ILE	N	-0. 415700
	CZ	0.807600		CA	-0.059700
	NH1	-0.862700		CB	0.130300
	NH2	-0.862700		CG2	-0.320400
	C	0.734100		CG1	-0.043000
	O	-0.589400		CD1	-0.066000
ASH	N	-0.415700		C	0.597300
	CA	0.034100		O	-0.567900
	CB	-0.031600	LEU	N	-0.415700
	CG	0.646200		CA	-0.051800
	OD1	-0.555400		CB	-0.110200
	OD2	-0.637600		CG	0.353100
	C	0.597300		CD1	-0.412100
	O	-0.567900		CD2	-0.412100
ASN	N	-0.415700		C	0.597300
	CA	0.014300		O	-0.567900
	CB	-0.204100	LYN	N	-0.415700
	CG	0.713000		CA	-0.072060
	OD1	-0.593100		CB	-0.048450
	ND2	-0.919100		CG	0.066120
	C	0.597300		CD	-0.037680
	O	-0.567900		CE	0.326040
ASP	N	-0.516300		NZ	-1.035810
	CA	0.038100		C	0.597300
	CB	-0.030300		O	-0.567900
	CG	0.799400	LYS	N	-0.347900

	OD1	-0.801400
	OD2	-0.801400
	C	0.536600
	O	-0.581900
CYM	N	-0.415700
	CA	-0.035100
	CB	-0.241300
	SG	-0.884400
	C	0.597300
	O	-0.567900
CYS	N	-0.415700
	CA	0.021300
	CB	-0.123100
	SG	-0.311900
	C	0.597300
	O	-0.567900
CYX	N	-0.415700
	CA	0.042900
	CB	-0.079000
	SG	-0.108100
	C	0.597300
	O	-0.567900
GLH	N	-0.415700
	CA	0.014500
	CB	-0.007100
	CG	-0.017400
	CD	0.680100
	OE1	-0.583800
	OE2	-0.651100
	C	0.597300
	O	-0.567900
GLN	N	-0.415700
	CA	-0.003100
	CB	-0.003600
	CG	-0.064500
	CD	0.695100
	OE1	-0.608600
	NE2	-0.940700
	C	0.597300
	O	-0.567900

	CA	-0.240000
	CB	-0.009400
	CG	0.018700
	CD	-0.047900
	CE	-0.014300
	NZ	-0.385400
	C	0.734100
	O	-0.589400
MET	N	-0.415700
	CA	-0.023700
	CB	0.034200
	CG	0.001800
	SD	-0.273700
	CE	-0.053600
	C	0.597300
	O	-0.567900
PHE	N	-0.415700
	CA	-0.002400
	CB	-0.034300
	CG	0.011800
	CD1	-0.125600
	CE1	-0.170400
	CZ	-0.107200
	CE2	-0.170400
	CD2	-0.125600
	C	0.597300
	O	-0.567900
PRO	N	-0.254800
	CD	0.019200
	CG	0.018900
	CB	-0.007000
	CA	-0.026600
	C	0.589600
	O	-0.574800
SER	N	-0.415700
	CA	-0.024900
	CB	0.211700
	OG	-0.654600
	C	0.597300
	O	-0.567900

GLU	N	-0.516300
	CA	0.039700
	CB	0.056000
	CG	0.013600
	CD	0.805400
	OE1	-0.818800
	OE2	-0.818800
	C	0.536600
	O	-0.581900
GLY	N	-0.415700
	CA	-0.025200
	C	0.597300
	O	-0.567900
HID	N	-0.415700
	CA	0.018800
	CB	-0.046200
	CG	-0.026600
	ND1	-0.381100
	CE1	0.205700
	NE2	-0.572700
	CD2	0.129200
	C	0.597300
	O	-0.567900
HIE	N	-0.415700
	CA	-0.058100
	CB	-0.007400
	CG	0.186800
	ND1	-0.543200
	CE1	0.163500
	NE2	-0.279500
	CD2	-0.220700
	C	0.597300
	O	-0.567900

THR	N	-0.415700
	CA	-0.038900
	CB	0.365400
	CG2	-0.243800
	OG1	-0.676100
	C	0.597300
	O	-0.567900
TRP	N	-0.415700
	CA	-0.027500
	CB	-0.005000
	CG	-0.141500
	CD1	-0.163800
	NE1	-0.341800
	CE2	0.138000
	CZ2	-0.260100
	CH2	-0.113400
	CZ3	-0.197200
	CE3	-0.238700
	CD2	0.124300
	C	0.597300
	O	-0.567900
TYR	N	-0.415700
	CA	-0.001400
	CB	-0.015200
	CG	-0.001100
	CD1	-0.190600
	CE1	-0.234100
	CZ	0.322600
	OH	-0.557900
	CE2	-0.234100
	CD2	-0.190600
	C	0.597300
	O	-0.567900
VAL	N	-0.415700
	CA	-0.087500
	CB	0.298500
	CG1	-0.319200
	CG2	-0.319200
	C	0.597300
	O	-0.567900

Supplementary Table 3. The atom types and mass on the RNA nucleotides and protein residues.

Atom type	Mass
C	12.0107
O	15.9994
N	14.0067
P	30.9738
S	32.0655

Supplementary Table 4. The PDB IDs of the 585 DNA-protein complexes in the training set for the DDPScore model.

1A6Y, 1AIS, 1AM9, 1B72, 1B8I, 1BF5, 1BL0, 1C9B, 1CGP, 1CIT, 1CKT, 1CQT, 1D5Y, 1DE8, 1DEW, 1DH3, 1E3M, 1ECR, 1EFA, 1EGW, 1EJ9, 1EWQ, 1FJL, 1FOS, 1GT0, 1GU4, 1GU5, 1GXP, 1H88, 1HLV, 1HWT, 1I3J, 1IF1, 1IG7, 1IMH, 1IXY, 1JE8, 1JFI, 1JJ8, 1JKO, 1KB6, 1L1T, 1L3L, 1LAT, 1LB2, 1LLI, 1LLM, 1MDY, 1MEY, 1MJ2, 1MJQ, 1MTL, 1N56, 1NGM, 1NH2, 1NK5, 1NKP, 1NWQ, 1ORP, 1P71, 1P7D, 1PDN, 1PUE, 1PUF, 1PYI, 1R0N, 1R4I, 1R71, 1R7M, 1R8D, 1RAM, 1REP, 1RIO, 1RZR, 1S9K, 1SAX, 1SKM, 1SKN, 1SKR, 1SRS, 1T2T, 1TSR, 1U4B, 1U78, 1UBD, 1X9M, 1YF3, 1YFH, 1YFL, 1YO5, 1YRN, 1YTB, 1ZLK, 1ZX4, 2ACJ, 2AS5, 2BNW, 2BNZ, 2BSQ, 2C9L, 2D5V, 2DDG, 2DRP, 2DY4, 2EFW, 2ERE, 2GLI, 2H1K, 2H27, 2H7G, 2HAN, 2HHQ, 2HZV, 2I13, 2IVH, 2JEJ, 2NNY, 2NRA, 2NTC, 2O8B, 2O8F, 2ODI, 2OYQ, 2P5L, 2P6R, 2PI4, 2Q2U, 2QBY, 2QL2, 2QSG, 2RBF, 2RDJ, 2V1U, 2VLA, 2VWJ, 2W7N, 2WB2, 2WJ0, 2WWY, 2XRZ, 2YVH, 2Z9O, 3A01, 3BI3, 3BJY, 3BRG, 3BS1, 3BTX, 3BTY, 3BTZ, 3C2I, 3C58, 3CBB, 3CO6, 3COQ, 3D1N, 3DFV, 3DFX, 3DSD, 3EI1, 3F27, 3F2B, 3FHZ, 3FYL, 3G0R, 3G73, 3G8U, 3GLG, 3GZ6, 3H0D, 3H25, 3IRQ, 3IV5, 3JRA, 3JSO, 3JTG, 3JX7, 3JXB, 3K5L, 3KHL, 3L2Q, 3LDS, 3LNQ, 3M4A, 3M9O, 3MKW, 3MLN, 3MVA, 3MX4, 3MZH, 3N6S, 3N97, 3NBN, 3O3F, 3O9X, 3OQM, 3OQN, 3OQO, 3OSF, 3POV, 3PT6, 3PVP, 3Q05, 3Q5F, 3QMG, 3QMH, 3QSV, 3QYN, 3RI4, 3RKQ, 3RN2, 3RN5, 3SM4, 3SQ2, 3TAR, 3THW, 3TMM, 3TQ6, 3TWM, 3U3W, 3U60, 3UFD, 3UK3, 3UKG, 3VD0, 3VD6, 3VW3, 3VW4, 3VXV, 3W2A, 3W3C, 3WAZ, 3WGI, 3ZH2, 3ZHM, 3ZKC, 3ZQC, 3ZQL, 4A09, 4AIK, 4AUW, 4AWL, 4B9N, 4BHK, 4CN2, 4CN3, 4CN5, 4DKJ, 4DL6, 4DL7, 4DM0, 4E54, 4E9F, 4EGZ, 4ER8, 4F41, 4F4Y, 4F4Z, 4F6M, 4FBT, 4FGN, 4FNC, 4FX4, 4FZX, 4FZY, 4G3I, 4G92, 4GCT, 4GDF, 4H10, 4HCA, 4HF1, 4HT4, 4I2O, 4I6Z, 4I7Y, 4IHS, 4IX7, 4J00, 4J01, 4J9M, 4JCX, 4K8X, 4KFC, 4KIS, 4L62, 4LDX, 4LEY, 4LG0, 4LLL, 4LT5, 4M3W, 4MG2, 4MHG, 4N47, 4NHJ, 4NNU, 4NOD, 4NQA, 4O3M, 4OI7, 4OI8, 4OMY, 4P9U, 4PAR, 4PBA, 4PU3, 4PW5, 4Q0W, 4Q10, 4QTK, 4QTR, 4R28, 4RB2, 4RBO, 4RD5, 4REC, 4RKG, 4ROD, 4ROE, 4RUA, 4RZR, 4S04, 4S2Q, 4TMU, 4UZB, 4WU4, 4WUL, 4WWC, 4WZW, 4XQQ, 4XVI, 4XVK, 4XXE, 4Y60, 4YNQ, 4Z5C, 4ZM0, 4ZTU, 5A39, 5BMZ, 5BOX, 5C51, 5CL3, 5CLV, 5CNQ, 5CQQ, 5CY1, 5D2S, 5D39, 5D4R,

5D4S, 5D5X, 5D8C, 5D8L, 5D9Y, 5DDG, 5DEU, 5DFI, 5DUI, 5E3M, 5E3O, 5E69, 5E6B, 5E6C, 5E6D, 5ED4, 5EDW, 5EGB, 5EXH, 5FLV, 5FMP, 5GKI, 5GPC, 5GZB, 5HLG, 5HOD, 5HR4, 5HSO, 5I3U, 5I50, 5ITR, 5J2Y, 5JUB, 5K5H, 5K7Z, 5KBJ, 5KRB, 5L0M, 5L6L, 5LEJ, 5MHK, 5MPF, 5NJ8, 5NM9, 5NNX, 5NO6, 5T01, 5TRD, 5U1J, 5UK7, 5UND, 5UUF, 5V0E, 5V3G, 5VA7, 5VFX, 5VPE, 5X5L, 5X6E, 5X6M, 5XM8, 5XXP, 5YEF, 5YEH, 5YEJ, 5YI2, 5YUX, 5YWS, 5YX2, 5Z00, 5Z2T, 5Z6Z, 5ZGN, 5ZMC, 6AKO, 6AKP, 6AMK, 6B0Q, 6BLW, 6BQU, 6BYY, 6C1U, 6C2S, 6CHV, 6CT9, 6DCF, 6DFY, 6E33, 6EDB, 6EL8, 6EMY, 6EN8, 6ES2, 6F59, 6FAS, 6FBG, 6G1T, 6GCD, 6HB4, 6HC3, 6HT5, 6IDE, 6IFM, 6IMJ, 6IR8, 6IS7, 6J5B, 6J9B, 6JBX, 6JG8, 6JGW, 6JW2, 6KHY, 6KKS, 6L6L, 6L6Q, 6L8E, 6LC1, 6LQF, 6LTY, 6LXN, 6ML2, 6MRJ, 6NCE, 6NCM, 6NSM, 6O8E, 6O8H, 6OD3, 6ON0, 6P0U, 6PAX, 6PBD, 6Q2B, 6QFD, 6QH0, 6QHD, 6QIL, 6QUA, 6RYD, 6RYI, 6RYL, 6SDG, 6SEI, 6SY0, 6T1F, 6U82, 6UKE, 6W1A, 6W1F, 6W5X, 6WC5, 6WIG, 6WMI, 6WYA, 6X71, 6XWG, 6XWH, 6Y35, 6Y36, 6YL2, 6YWW, 6ZIX, 7AMN, 7AMT, 7B0C, 7BCA, 7BZG, 7C9O, 7CE1, 7CLI, 7CVQ, 7D3T, 7D8T, 7DCJ, 7DCT, 7DV2, 7DVV, 7EDS, 7EL3, 7ET6, 7EYI, 7F2F, 7F4Y, 7F9H, 7JK1, 7JM4, 7JNP, 7JU3, 7K30, 7KUF, 7L4C, 7L4F, 7L4K, 7LNI, 7M5W, 7N5S, 7N5U, 7N8S, 7N94, 7NX5, 7O0H, 7O56, 7OGS, 7PRV, 7PZB, 7Q0N, 7R6R, 7S03, 7SR6, 7T8K, 7TDX, 7TXC, 7UBL, 7UPZ, 7V5N, 7V6W, 7VE5, 7VJM, 7VJQ, 7VN8, 7VP2, 7VP5, 7VP7, 7VUQ, 7W27, 7WB3, 7WQ5, 7Y3I, 7Y43, 7YZC, 8AMU, 8CXT, 8E3D

Supplementary Table 5. The PDB IDs of the 47 DNA-protein complexes in the testing set I for the DDPScore model.

1A74, 1AZP, 1B3T, 1BDT, 1BY4, 1CMA, 1DDN, 1DFM, 1DIZ, 1EA4, 1EMH, 1EYU, 1F4K, 1FOK, 1G9Z, 1H9T, 1HJC, 1JJ4, 1JT0, 1K79, 1KC6, 1KSY, 1MNN, 1O3T, 1PT3, 1QNE, 1QRV, 1R4O, 1RPE, 1RVA, 1TRO, 1VAS, 1VRR, 1W0T, 1Z63, 1Z9C, 1ZME, 1ZS4, 2C5R, 2FIO, 2FL3, 2IRF, 2OAA, 3BAM, 3CRO, 4KTQ, 7MHT

Supplementary Table 6. The PDB IDs of the 10 DNA-protein complexes in the testing set II for the DDPScore model.

1O3R, 2ITL, 2NTC, 2PI0, 2X6V, 3MFK, 3MLO, 3POV, 4UUV, 5JTL

Supplementary Table 7. The ranking and ligand RMSDs (Å) of the independent semi-flexible testing set I for DRPScore and HDOCK.

Semi-flexible Testing Set I	DDPScore		HDOCK	
Case	Rank	L_{rmsd}	Rank	L_{rmsd}
1A74	1	2.86	3	2.02
1AZP	5	9.49	70	9.14
1B3T	66	4.39	76	9.09
1BDT	2	3.74	8	2.46
1BY4	1	5.93	36	5.83
1CMA	18	5.49	2	6.58
1DDN	1	6.62	13	8.47
1DFM*	\	\	\	\
1DIZ	168	8.07	470	8.09
1EA4	2	0.83	1	0.55
1EMH	3	6.41	1	3.85
1EYU	77	4.97	244	8.79
1F4K	4	7.65	461	9.32
1FOK	4	2.04	82	2.68
1G9Z	2	9.17	1	9.49
1H9T	2	1.43	2	1.29
1HJC	46	6.60	1	3.33
1JJ4	2	9.35	8	1.10
1JT0*	\	\	\	\
1K79	1	3.07	1	1.97
1KC6	2	3.01	1	1.45
1KSY	6	8.19	1	0.84
1MNN	1	9.87	1	0.65

1O3T	3	9.78	1	9.78
1PT3	22	9.86	14	7.26
1QNE	7	4.81	1	4.14
1QRV	94	8.46	450	9.63
1R4O	1	6.07	147	6.12
1RPE	5	5.89	5	2.33
1RVA*	\	\	\	\
1TRO	8	5.28	30	3.12
1VAS	1	1.36	1	0.74
1VRR	2	4.78	2	5.03
1W0T	3	5.83	3	8.60
1Z9C	1	1.32	917	1.32
1Z63	378	7.33	889	7.47
1ZME*	\	\	\	\
1ZS4*	\	\	\	\
2C5R	1	9.85	85	6.75
2FIO	15	9.14	6	9.14
2FL3	22	2.66	1	3.33
2IRF*	\	\	\	\
2OAA	10	4.58	11	4.40
3BAM*	\	\	\	\
3CRO	42	6.93	4	6.61
4KTQ	39	5.01	1	3.46
7MHT	20	9.27	13	8.98

*These DNA-protein complexes have no structure with ligand RMSD less than 10 Å in 1000 decoys generated by HDOCK. Thus, they have no Rank and L_{rmsd}.

Supplementary Table 8. The ranking and ligand RMSDs (Å) of the independent semi-flexible testing set II for DRPScore and HDOCK.

Semi-flexible Testing Set II	DDPScore		HDOCK	
Case	Rank	L_{rmsd}	Rank	L_{rmsd}
1O3R	72	8.81	103	9.24
2ITL	5	1.35	1	0.26
2NTC	55	6.33	27	5.44
2PI0	1	2.14	1	0.56
2X6V	110	1.49	61	2.11
3MFK	3	3.68	1	1.50
3MLO	4	1.17	1	0.56
3POV	187	7.12	181	7.12
4UUV	2	5.06	109	5.32
5JLT	17	6.94	23	6.85

Supplementary Table 9. The ranking and ligand RMSDs (Å) of the independent flexible testing set I for DRPScore and HDOCK.

Flexible Testing Set I	DDPScore		HDOCK	
Case	Rank	L_{rmsd}	Rank	L_{rmsd}
1A74	177	9.34	256	9.34
1AZP	173	9.81	239	9.37
1B3T	81	8.90	83	5.82
1BDT	1	4.64	44	6.63
1BY4	4	7.03	13	9.36
1CMA	4	2.26	4	4.10
1DDN*	\	\	\	\
1DFM*	\	\	\	\
1DIZ*	\	\	\	\
1EA4	7	8.32	31	9.98
1EMH	7	7.48	9	8.57
1EYU	81	7.24	25	7.90
1F4K	2	7.27	497	6.91
1FOK	14	5.10	47	7.85
1G9Z	11	6.55	6	7.32
1H9T	1	9.05	275	8.71
1HJC	23	4.95	141	3.96
1JJ4	66	8.39	7	8.39
1JT0*	\	\	\	\
1K79	20	6.22	109	7.04
1KC6*	\	\	\	\
1KSY	185	9.26	4	9.38
1MNN	15	3.07	21	8.21

1O3T*	\	\	\	\
1PT3	29	9.28	16	9.08
1QNE	5	4.85	89	4.23
1QRV*	\	\	\	\
1R4O*	\	\	\	\
1RPE	2	7.84	13	9.91
1RVA*	\	\	\	\
1TRO	78	5.13	299	3.76
1VAS	423	9.62	543	9.90
1VRR	11	7.50	61	8.92
1W0T	263	9.04	227	9.04
1Z9C*	\	\	\	\
1Z63*	\	\	\	\
1ZME*	\	\	\	\
1ZS4*	\	\	\	\
2C5R	9	8.31	199	9.67
2FIO*	\	\	\	\
2FL3	6	9.56	2	4.71
2IRF	104	9.76	349	9.31
2OAA	15	5.66	126	7.35
3BAM*	\	\	\	\
3CRO	63	5.25	34	6.12
4KTQ	97	9.77	2	6.73
7MHT	3	9.62	20	6.47

*These DNA-protein complexes have no structure with ligand RMSD less than 10 Å in 1000 decoys generated by HDOCK. Thus, they have no Rank and L_{rmsd}.

Supplementary Table 10. The ranking and ligand RMSDs (Å) of the independent flexible testing set II for DRPScore and HDOCK.

Flexible Testing Set II	DDPScore		HDOCK	
Case	Rank	L_{rmsd}	Rank	L_{rmsd}
1O3R	86	4.03	546	4.03
2ITL	1	5.69	1	5.67
2NTC	308	7.35	2	7.35
2PI0*	\	\	\	\
2X6V	42	7.44	56	5.29
3MFK	5	5.56	9	8.28
3MLO	3	2.17	22	4.31
3POV	115	9.88	60	9.99
4UUV	10	7.77	263	7.35
5JLT	111	6.98	328	9.80

*These DNA-protein complexes have no structure with ligand RMSD less than 10 Å in 1000 decoys generated by HDOCK. Thus, they have no Rank and L_{rmsd}.

Supplementary Table 11. Data of detailed analysis on structural categories of flexible DNA-protein testing set I.

Case	Flexible categories	I_{rmsd}	DDPScore		HDOCK		structural categories
			Rank	L_{rmsd}	Rank	L_{rmsd}	
2C5R	Easy	0.49	9	8.31	199	9. 67	helix-turn-helix
1PT3	Easy	1.35	29	9.28	16	9.08	enzyme
1MNN	Easy	1.48	15	3.07	21	8.21	helix-turn-helix
1FOK	Easy	1.53	14	5.10	47	7.85	helix-turn-helix
1KSY	Easy	1.58	185	9.26	4	9.38	α-helix
3CRO	Easy	1.58	63	5.25	34	6.12	helix-turn-helix
1EMH	Easy	1.62	7	7.48	9	8.57	enzyme
1H9T	Easy	1.68	1	9.05	275	8.71	helix-turn-helix
1TRO	Easy	1.70	78	5.13	299	3.76	helix-turn-helix
1BY4	Easy	1.77	4	7.03	13	9.36	zinc-coordinating
1HJC	Easy	1.80	23	4.95	141	3.96	β-sheet
1DIZ*	Easy	1.82	\	\	\	\	enzyme
1RPE	Easy	1.87	2	7.84	13	9.91	helix-turn-helix
1VRR	Intermediate	2.08	11	7.50	61	8.92	enzyme
1F4K	Intermediate	2.26	2	7.27	497	6.91	helix-turn-helix
1K79	Intermediate	2.37	20	6.22	109	7.04	helix-turn-helix
1KC6*	Intermediate	2.38	\	\	\	\	enzyme
1EA4	Intermediate	2.43	7	8.32	31	9.98	β-hairpin/ribbon
1Z63*	Intermediate	2.51	\	\	\	\	enzyme
1R4O*	Intermediate	2.61	\	\	\	\	zinc-coordinating
1AZP	Intermediate	2.70	173	9.81	239	9.37	β-hairpin/ribbon
1W0T	Intermediate	2.78	263	9.04	227	9.04	helix-turn-helix
1CMA	Intermediate	2.81	4	2.26	4	4.1	β-hairpin/ribbon
1JJ4	Intermediate	2.83	66	8.39	7	8.39	α-helix

1VAS	Intermediate	3.04	423	9.62	543	9.90	enzyme
4KTQ	Intermediate	3.23	97	9.77	2	6.73	enzyme
1Z9C*	Intermediate	3.24	\	\	\	\	helix-turn-helix
1DDN*	Intermediate	3.26	\	\	\	\	helix-turn-helix
2IRF	Intermediate	3.35	104	9.76	349	9.31	helix-turn-helix
1JT0*	Intermediate	3.49	\	\	\	\	helix-turn-helix
1G9Z	Intermediate	3.67	11	6.55	6	7.32	enzyme
1A74	Intermediate	4.26	177	9.34	256	9.34	enzyme
2FIO*	Intermediate	4.41	\	\	\	\	α-helix
1QNE	Intermediate	4.57	5	4.85	89	4.23	β-sheet
1ZS4*	Intermediate	4.71	\	\	\	\	helix-turn-helix
1QRV*	Difficult	5.19	\	\	\	\	α-helix
1O3T*	Difficult	5.20	\	\	\	\	helix-turn-helix
1B3T	Difficult	5.32	81	8.90	83	5.82	α-helix
3BAM*	Difficult	5.55	\	\	\	\	enzyme
1RVA*	Difficult	5.68	\	\	\	\	enzyme
1ZME*	Difficult	5.76	\	\	\	\	zinc-coordinating
1DFM*	Difficult	6.31	\	\	\	\	enzyme
1BDT	Difficult	6.45	1	4.64	44	6.63	β-hairpin/ribbon
7MHT	Difficult	6.71	3	9.62	20	6.47	enzyme
2FL3	Difficult	6.71	6	9.56	2	4.71	enzyme
1EYU	Difficult	6.82	81	7.24	25	7.90	enzyme
2OAA	Difficult	8.95	15	5.66	126	7.35	enzyme

*These DNA-protein complexes have no structure with ligand RMSD less than 10 Å in 1000 decoys generated by HDOCK. Thus, they have no Rank and L_{rmsd}.

Supplementary Table 12. The interface hydrogen bond interactions of the lowest RMSD model of top 5 DNA-protein complexes (PDB ID:1H9T) ranking by DDPScore and HDOCK. The 'Native' refers to the hydrogen bond occurs in the native DNA-protein complex structure.

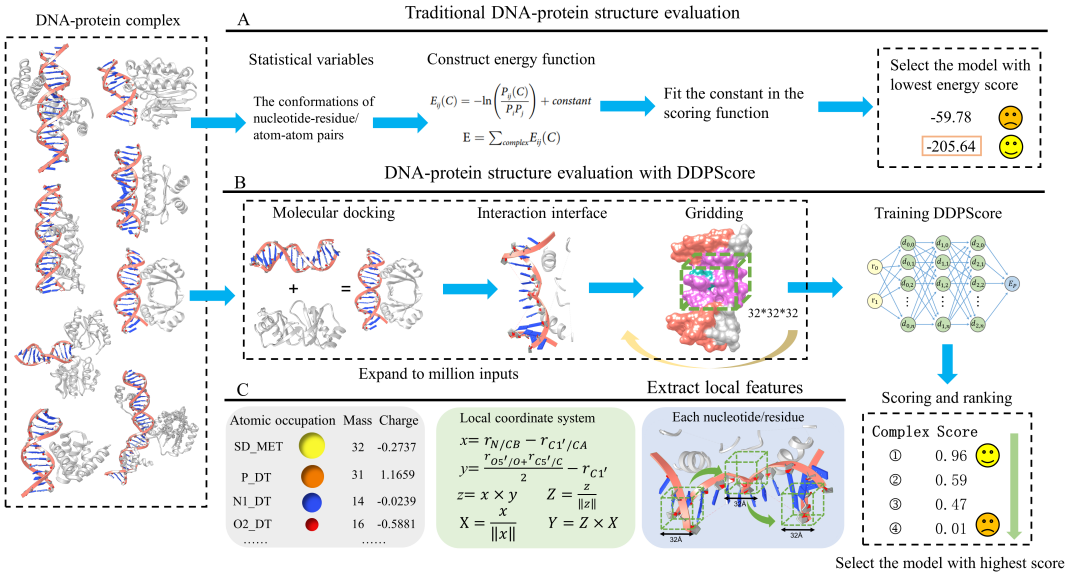
Scoring function	Number/Type	Distance	Native
DDPScore	A45ARG--C25DG	2.73	×
	A45ARG--C25DG	3.19	×
	A49ARG--C26DG	2.61	√
	A49ARG--C26DG	2.41	√
	C10DC--A46THR	3.44	√
HDOCK	A15TYR--C24DT	3.11	×
	A91HIS--C12DA	2.57	×
	A137HIS--C3DT	3.22	×
	A174ARG--C32DC	2.49	×
	A177ARG--C32DC	2.96	×
	A177ARG--C32DC	3.09	×
	B65HIS--C9DA	2.41	×
	C11DG--A14GLU	3.17	×
	C12DA--A6GLN	2.91	×

Supplementary Table 13: Evaluation time of DDPScore compared with 3D CNN on a single NVIDIA 2080Ti GPU for DNA-protein complex structures on testing set I.

Evaluation Time for DNA-protein Complex Structures (s)	Number of predictions (DDPScore)				3DCNN
	5	20	100	1000	1000
1A74	4.81	9.69	32.83	310.24	423.83
1AZP	4.64	8.81	30.60	282.25	300.24
1B3T	4.72	9.33	37.29	297.59	397.06
1BDT	4.74	9.37	33.59	313.19	459.12
1BY4	4.77	9.61	34.08	314.99	490.94
1CMA	4.65	9.18	32.01	292.67	362.90
1DDN	4.81	9.56	34.11	306.45	507.85
1DFM	4.56	8.70	29.81	626.34	765.99
1DIZ	4.65	9.12	31.64	285.67	365.67
1EA4	4.69	9.01	32.45	294.75	397.46
1EMH	4.65	8.80	31.43	291.17	342.75
1EYU	4.67	9.08	32.25	294.10	425.11
1F4K	4.71	9.40	32.25	293.41	407.81
1FOK	4.86	9.84	34.44	351.89	615.43
1G9Z	4.77	9.52	33.28	305.68	545.58
1H9T	4.65	9.28	32.45	297.83	481.28
1HJC	4.68	8.97	31.38	290.16	414.78
1JJ4	4.69	9.20	31.51	297.22	284.56
1JT0	4.78	9.49	32.84	305.52	364.61
1K79	4.64	9.17	31.28	290.13	286.27
1KC6	4.72	9.57	33.91	302.79	385.31
1KSY	4.73	9.58	32.54	322.70	345.95
1MNN	4.71	9.43	32.35	301.56	315.66
1O3T	4.68	9.40	33.41	293.96	332.15

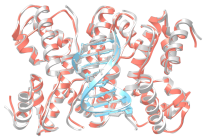
1PT3	4.64	9.25	31.67	296.40	296.15
1QNE	4.72	9.11	32.62	293.83	373.50
1QRV	4.67	9.28	33.55	292.04	327.40
1R4O	4.74	9.53	33.15	299.40	376.08
1RPE	4.67	9.44	32.73	301.71	365.86
1RVA	4.73	9.33	33.94	305.45	413.58
1TRO	4.67	8.93	32.04	303.54	359.51
1VAS	4.67	8.92	31.95	299.96	343.14
1VRR	4.81	9.62	43.66	431.21	614.68
1W0T	4.73	9.26	33.71	303.74	470.76
1Z63	4.68	9.04	34.37	302.48	443.84
1Z9C	4.77	9.32	34.00	304.36	474.00
1ZME	4.68	8.96	32.14	300.73	360.17
1ZS4	4.72	9.10	32.37	302.59	408.38
2C5R	4.69	9.01	32.18	298.81	402.84
2FIO	4.65	9.05	32.70	299.32	391.26
2FL3	4.71	9.06	33.13	298.72	390.04
2IRF	4.66	8.96	32.45	291.98	336.86
2OAA	4.74	9.15	33.63	299.99	386.74
3BAM	4.84	9.66	43.21	507.65	669.01
3CRO	4.70	9.12	33.33	290.97	372.94
4KTQ	4.76	9.25	34.28	308.73	479.42
7MHT	4.59	8.71	31.16	283.75	354.23
Average	4.71	9.24	33.23	314.46	413.38
Average for a structure	0.94	0.46	0.33	0.31	0.41

Supplementary Figure 1: Flow charts comparing DDPScore and traditional scoring functions on DNA-protein structure evaluation. (A) The steps of traditional scoring function based on statistical potential energy to evaluate the DNA-protein complex decoys. (B) The steps of DDPScore to evaluate the DNA-protein complex decoys. (C) extracting local features through local coordinate system on each nucleotide/residue.

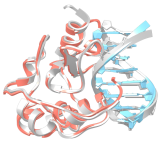


Supplementary Figure 2: The flexible conformational change (interface RMSD) between bound (DNA in gray, and protein in gray) and unbound (DNA in blue, and protein in salmon) DNA-protein complexes in DNA-protein testing set I.

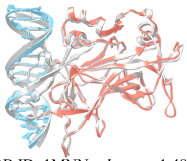
Easy targets:



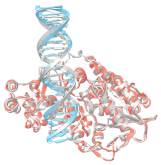
PDB ID: 2C5R I_{rmsd} = 0.49Å



PDB ID: 1PT3 I_{rmsd} = 1.35Å

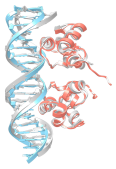


PDB ID: 1MNN I_{rmsd} = 1.48Å

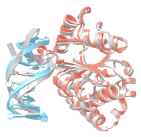


PDB ID: 1FOK I_{rmsd} = 1.53Å

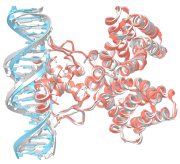
PDB ID: 1KSY I_{rmsd} = 1.58Å



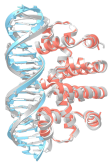
PDB ID: 3CRO I_{rmsd} = 1.58Å



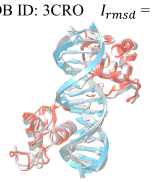
PDB ID: 1EMH I_{rmsd} = 1.62Å



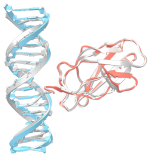
PDB ID: 1H9T I_{rmsd} = 1.68Å



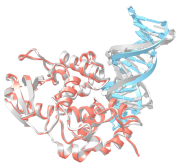
PDB ID: 1TRO I_{rmsd} = 1.70Å



PDB ID: 1BY4 I_{rmsd} = 1.77Å



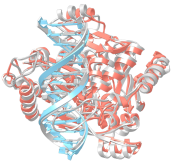
PDB ID: 1HJC I_{rmsd} = 1.80Å



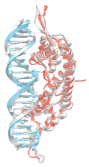
PDB ID: 1DIZ I_{rmsd} = 1.82Å

PDB ID: 1RPE I_{rmsd} = 1.87Å

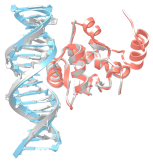
Intermediate targets:



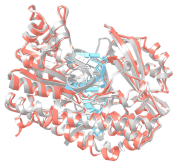
PDB ID: 1VRR I_{rmsd} = 2.08Å



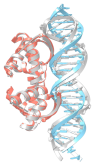
PDB ID: 1F4K I_{rmsd} = 2.26Å



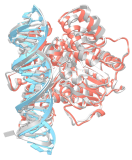
PDB ID: 1K79 I_{rmsd} = 2.37Å



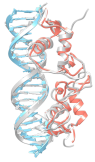
PDB ID: 1KC6 I_{rmsd} = 2.38Å



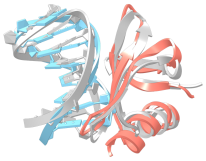
PDB ID: 1EA4 I_{rmsd} = 2.43Å



PDB ID: 1Z63 I_{rmsd} = 2.51Å

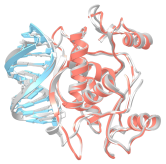


PDB ID: 1R4O I_{rmsd} = 2.61Å

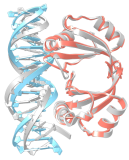


PDB ID: 1AZP I_{rmsd} = 2.70Å

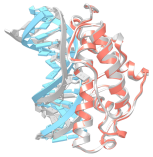
PDB ID: 1W0T I_{rmsd} = 2.78Å



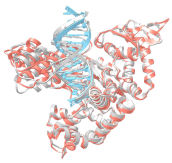
PDB ID: 1CMA I_{rmsd} = 2.81Å



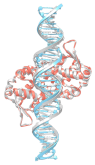
PDB ID: 1JJ4 I_{rmsd} = 2.83Å



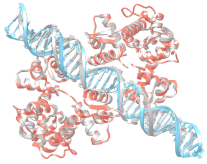
PDB ID: 1VAS I_{rmsd} = 3.04Å



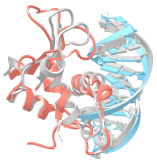
PDB ID: 4KTQ I_{rmsd} = 3.23Å



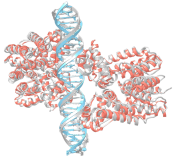
PDB ID: 1Z9C I_{rmsd} = 3.24Å



PDB ID: 1DDN I_{rmsd} = 3.26Å

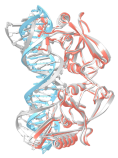


PDB ID: 2IRF I_{rmsd} = 3.35Å



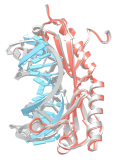
PDB ID: 1JT0 I_{rmsd} = 3.49Å

PDB ID: 1G9Z I_{rmsd} = 3.67Å

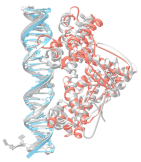


PDB ID: 1A74 I_{rmsd} = 4.26Å

PDB ID: 2FIO I_{rmsd} = 4.41Å

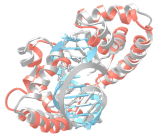


PDB ID: 1QNE I_{rmsd} = 4.57Å

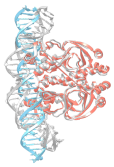


PDB ID: 1ZS4 I_{rmsd} = 4.71Å

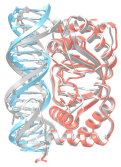
Difficult targets:



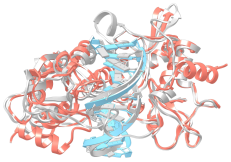
PDB ID: 1QRV I_{rmsd} = 5.19Å



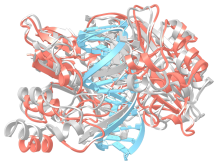
PDB ID: 1O3T I_{rmsd} = 5.20Å



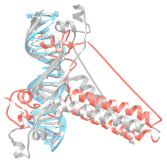
PDB ID: 1B3T I_{rmsd} = 5.32Å



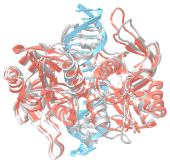
PDB ID: 3BAM I_{rmsd} = 5.55Å



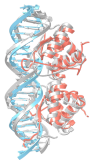
PDB ID: 1RVA I_{rmsd} = 5.68Å



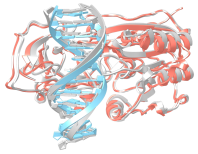
PDB ID: 1ZME I_{rmsd} = 5.76Å



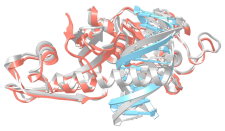
PDB ID: 1DFM I_{rmsd} = 6.31Å



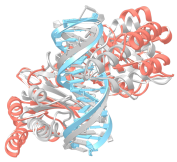
PDB ID: 1BDT I_{rmsd} = 6.45Å



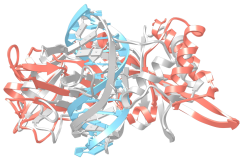
PDB ID: 7MHT l_{rmsd} = 6.71Å



PDB ID: 2FL3 l_{rmsd} = 6.71Å

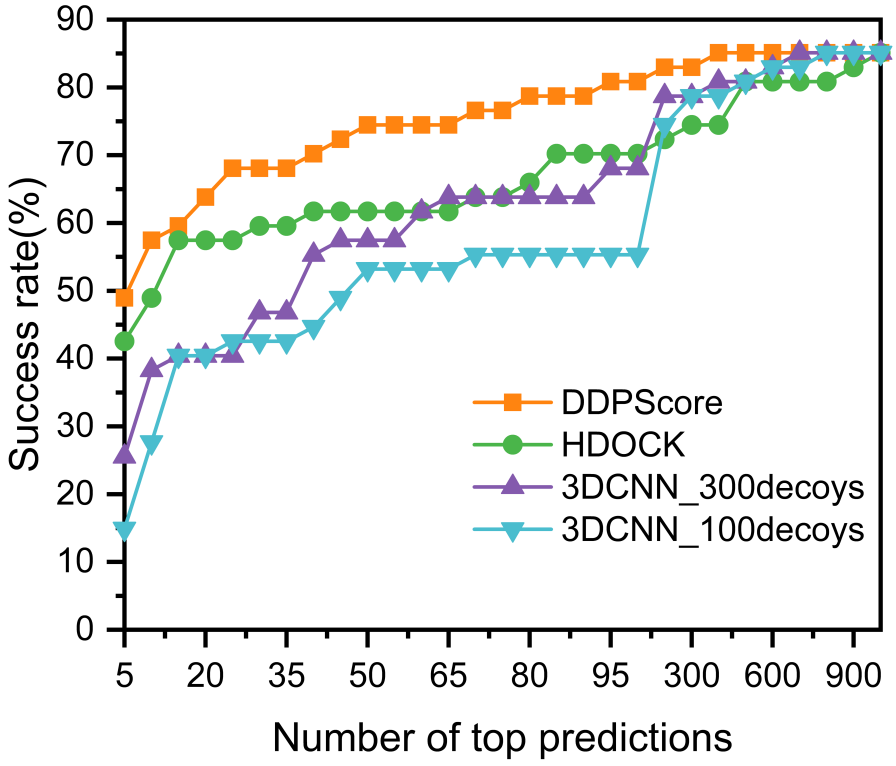


PDB ID: 1EYU l_{rmsd} = 6.82Å



PDB ID: 2OAA l_{rmsd} = 8.95Å

Supplementary Figure 3: The performance of DDPScore, HDOCK and 3DCNN on the semi-flexible DNA-protein testing set I. The success rate of DDPScore (orange square), HDOCK (green circle), 3DCNN (training on 300 decoys per DNA-protein complex: purple triangle), 3DCNN (training on 100 decoys per DNA-protein complex: blue inverted triangle) on the semi-flexible DNA-protein testing set I.



Supplementary Figure 4: Evaluation time for DNA-protein complex structures. The histogram represents the number of predictions relative to the computation time of DDPScore on three examples of DNA-protein flexible docking (PDB ID:1H9T, 1QNE, and 7MHT). DDPScore typically requires less than one second, on average, to evaluate a DNA-protein complex.