

# Supplementary Materials

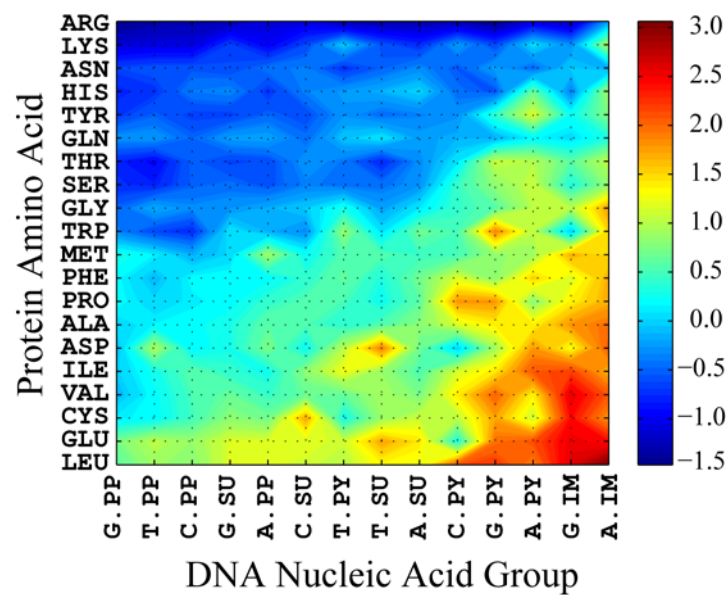
---

For

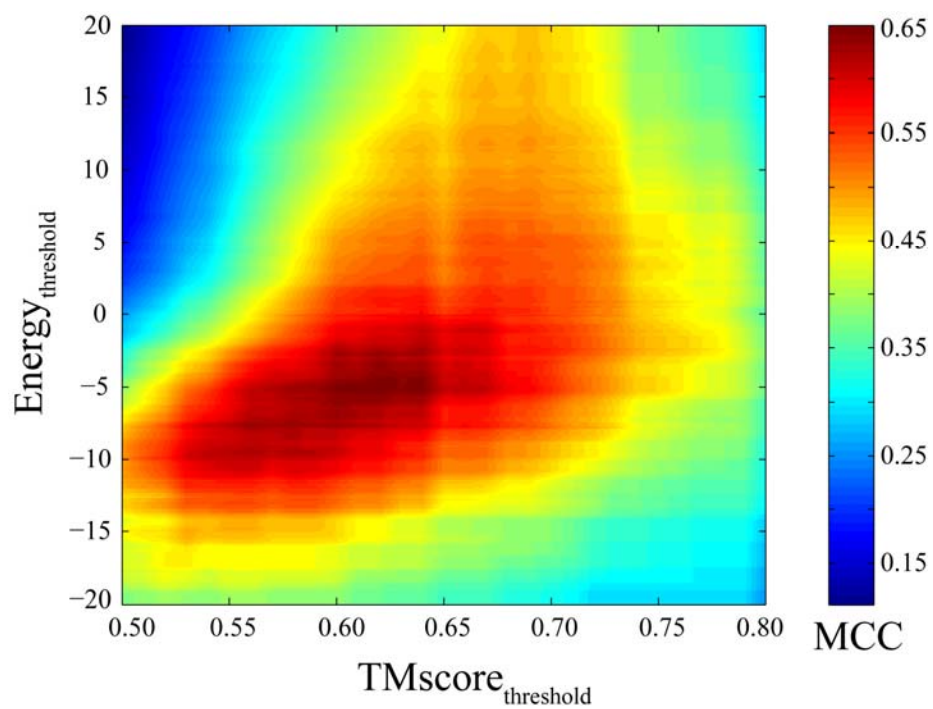
DBD-Hunter: A Knowledge-based Method for the  
Prediction of DNA-Protein Interactions

MU GAO AND JEFFREY SKOLNICK

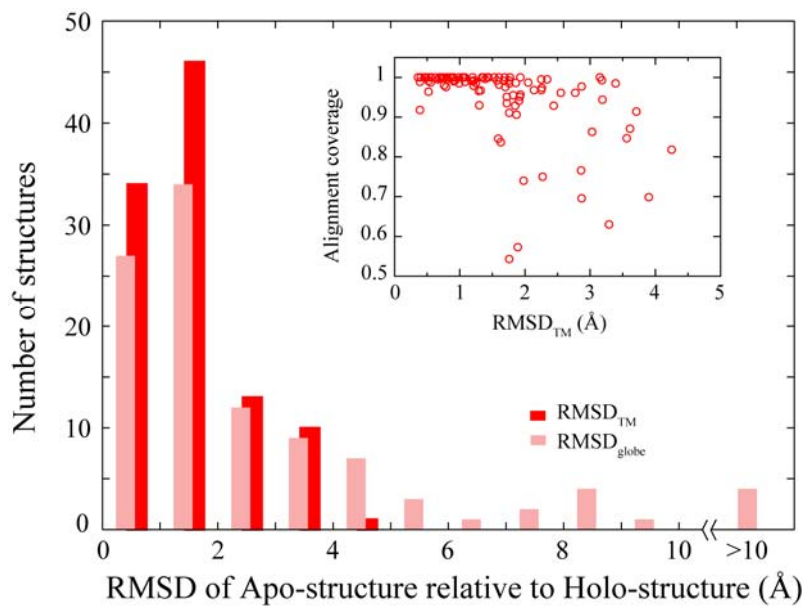
---



**Figure 1.** Statistical pair potential for describing DNA-protein interactions. The potential was derived from 179 crystal structures of DNA-protein complexes. Contour representation of pair potential parameters for individual DNA functional groups and protein amino acids.



**Figure 2.** Contour representation of the MCC for DNA-binding protein prediction tests on DB179/NB3797. A grid size of 0.01 from 0.50 to 0.80 was employed for the TM-score threshold, and a grid size of 0.1 from -20 to 20 was used for the energy threshold.



**Figure 3.** RMSDs of apo-structure relative to holo-structure for 104 apo-holo pairs (APO104/HOLO104). Scattered plot for structural alignment coverage *versus* RMSD<sub>TM</sub> are shown in the insert.

**Table 1.** List of the DNA-protein complex structure set DB179. Each entry is provided with the four-digit PDB code, the protein chain identifier, the residue range of the DNA-binding domain, chain identifiers of dsDNA to which the protein is bound, the group name, and the description of the protein. DNA-binding domain comprising the whole protein chain is indicated by “All” in the residue range column. Proteins are classified into three groups: transcription factor (TF), enzyme (EZ), and the others (OT).

<b>PDB code</b>	<b>Protein chain</b>	<b>Residue Range</b>	<b>DNA Chains</b>	<b>Group</b>	<b>Protein Description</b>
1a0a	B	All	DC	TF	PHO4
1a36	A	All	CB	EZ	Topoisomerase I
1a74	A	All	DC	EZ	Endonuclease I-PpoI
1akh	A	All	CC	TF	MATa1
1am9	B	All	GF	TF	SREBP
1b3t	A	All	DC	OT	EBNA1
1bf5	A	317-568	CB	TF	STAT-1
1bl0	A	All	CB	TF	MarA
1bnk	A	All	ED	EZ	DNA N-glycosylase
1c9b	E	All	DC HG PO	TF	Transcription factor IIB
1c9b	J	All	LK	TF	TATA-box binding protein
1cf7	A	All	DC	TF	E2F-4
1cf7	B	All	DC	TF	DP-2
1cgp	A	138-205	DD	TF	Catabolite gene activator
1ckt	A	All	CB	OT	High-mobility-group protein
1cw0	A	All	OM	EZ	Vsr endonuclease
1d02	A	All	DC	EZ	Endonuclease MunI
1d2i	A	All	DC	EZ	Endonuclease BglII
1d3u	B	All	DC	TF	Transcription factor IIB
1dc1	A	All	CW	EZ	Endonuclease BsoBI
1dct	B	All	MF NG	EZ	Methyltransferase HaeIII
1dew	B	All	YX	EZ	Excision repair enzyme APE1
1diz	A	100-282	FE	EZ	DNA glycosylase AlkA
1ea4	A	All	YY	TF	CopG
1ecr	A	All	CB	OT	Replication-terminator Tus
1f4k	B	All	ED	OT	Replication-terminator RTP
1fjl	A	All	ED	TF	Pax
1fok	A	All	CB	EZ	Endonuclease FokI
1gdt	A	141-183	EC FC	EZ	Gamma delta resolvase
1gt0	D	All	BA	OT	High-mobility-group protein
1gxp	A	All	DC	TF	PhoB

1h0m	D	170-234	HG	TF	Quorum sensing protein TraR
1h38	D	All	ON PN	EZ	Bacteriophage T7 RNA-polymerase
1h89	C	All	ED	TF	c-Myb
1h8a	A	All	ED	TF	C/EBP beta
1hjb	C	All	HG	TF	AML1 RUNT domain
1hlv	A	All	CB	OT	Centromere protein B
1i3j	A	All	CB	EZ	Endonuclease I-TevI
1iaw	A	All	DC FE	EZ	Endonuclease NaeI
1if1	A	All	CC	TF	Interferon regulatory factor
1ig9	A	All	PT	EZ	DNA polymerase I
1ign	B	All	FE	OT	Telomere protein RAP1
1iu3	C	All	BA	TF	SeqA
1j1v	A	All	CB	OT	DnaA
1jey	A	254-534	DC	EZ	Ku70
1jey	B	242-545	DC	EZ	Ku80
1jfi	A	All	ED	TF	Negative Cofactor 2 alpha
1jfi	B	All	ED	TF	Negative Cofactor 2 beta
1jj4	A	All	DC	EZ	Papillomavirus E2
1jkq	C	All	BA	EZ	Recombinase Hin
1jt0	A	2-72	FE	TF	QacR
1k3w	A	All	CB	EZ	Endonuclease VIII
1kb4	A	All	DC	TF	Vitamin D receptor
1kc6	D	All	HG	EZ	Endonuclease HincII
1kx5	A	All	JI	OT	Histone H3
1kx5	B	All	JI	OT	Histone H4
1kx5	C	All	JI	OT	Histone H2A
1kx5	D	All	JI	OT	Histone H2B
1lq1	D	All	FE	TF	Spo0A
1m6x	C	All	GI	EZ	Recombinase Flp
1mdy	B	All	FE	TF	MyoD
1mj0	A	All	GF	TF	MetJ mutant
1mnm	A	All	FE	TF	MCM1
1mow	D	All	FE	EZ	Endonuclease I-DmoI/I-Crel
1mtl	A	All	DC	EZ	DNA glycosylase
1mur	A	All	CB	EZ	Transposase Tn5
1nlw	B	All	GF	TF	Max
1nlw	D	All	JH	TF	Mad
1oct	C	All	BA	TF	Oct-1 POU
1odh	A	All	DC	TF	GCM
1oe6	B	All	FE	EZ	Uracil-DNA glycosylase SMUG1
1oh5	B	270-566	FE	EZ	DNA mismatch repair enzyme MutS
1orn	A	All	CB	EZ	Endonuclease III
1osb	A	All	BB	EZ	Relaxase TrwC
1ozj	A	All	DC	TF	SMAD3 MH1

1p78	A	All	DC	TF	HU
1p7d	B	All	FE FF	EZ	Lambda integrase
1p7h	N	393-575	DC	TF	NFAT1
1p8k	Z	All	DA CB	EZ	Endonuclease I-AniI
1pp7	U	All	FE	TF	Initiator binding protein IBP39
1pt3	A	All	DC	EZ	Endonuclease ColE7
1pue	E	All	BA	TF	ETS
1puf	A	All	ED	TF	HoxA9
1qbj	B	All	ED	EZ	Deaminase ADAR1
1qp9	D	All	HG	TF	HAP1-PC7
1qum	A	All	DB	EZ	Endonuclease IV
1r0a	A	1-429	PT	EZ	HIV-1 reverse transcriptase
1r7l	B	All	FE	TF	KorB
1r7m	A	All	DC	EZ	Endonuclease I-SceI
1r8d	B	All	DC	TF	MerR
1rep	C	All	BA	TF	Initiator protein RepE
1rh6	B	All	DC	EZ	Excisionase Xis
1rrq	A	All	CB	EZ	DNA glycosylase MutY
1rxw	A	All	CB	EZ	Flap Endonuclease-1 FEN-1
1rzz	A	1-60	BE	TF	CcpA
1sax	A	All	DC	TF	Methicillin repressor MecI
1sfu	A	All	DC	OT	Yatapoxvirus E3L-like protein
1skn	P	All	BA	TF	Skn-1
1sxp	B	All	DC	EZ	Beta-glucosyltransferase
1t2k	D	All	FE	TF	ATF-2
1t8e	A	All	DC	EZ	DNA polymerase
1tez	A	All	JI	EZ	DNA photolyase
1tqe	Q	All	DC	TF	Myocyte enhancer factor-2
1trr	G	All	IC	TF	Trp repressor
1tsr	B	All	FE	TF	p53 tumor suppressor
1ttu	A	196-541	CB	TF	Nuclear effector CSL
1u3e	M	All	BA CA	EZ	Endonuclease I-HmuI
1u78	A	All	CB	EZ	Transposase Tc3A
1u8b	A	All	ED	EZ	DNA Methyltransferase Ada
1u8r	J	1-64	LK	TF	Iron-dependent regulator IdeR
1vas	A	All	CB	EZ	Endonuclease V
1w0t	A	All	DC	OT	Telomeric repeat binding protein TRF1
1wte	B	All	YX	EZ	Endonuclease Eco0109I
1wto	A	All	CB	OT	Sac7d
1xbr	A	All	DC	TF	Brachyury transcription factor
1xpx	A	All	CD	TF	Prospero
1y6f	A	All	DC	EZ	Alpha-glucosyltransferase
1yfi	A	All	DC FE	EZ	Endonuclease MspI
1yfi	D	All	65 09	EZ	DNA-adenine Methyltransferase

1ym	B	All	CC	TF	MAT alpha2
1ytf	C	All	FE	TF	Transcription factor IIA
1z63	A	All	DC	EZ	SWI2/SNF2 ATPase
1z9c	F	All	LK	TF	OhrR
1zbl	B	All	DC	EZ	RNase H
1zg5	B	All	DC	TF	NarL
1zme	C	All	BA	TF	PUT3
1zs4	C	All	TU	TF	Lambda cII
1zx4	A	1-244	ST	EZ	ParB
2a3v	B	All	FE	EZ	Integron integrase
2a66	A	All	CB	TF	hLRH-1
2a6o	B	All	CC DD	EZ	Transposase TnpA
2ago	A	All	CB	EZ	DNA polymerase Dpo4
2aor	B	All	DC	EZ	MutH
2aq4	A	All	TP	EZ	DNA polymerase Rev1
2as5	F	All	BA DC	TF	FOXP2
2bam	B	All	DC	EZ	Endonuclease BamHI
2bdp	A	469-876	TP	EZ	DNA polymerase I
2bgw	A	All	DC	EZ	XPF endonuclease
2bzf	A	All	CB	OT	Barrier to autointegration factor
2c7r	A	All	DC	EZ	DNA Methyltransferase HhaI
2c9l	Z	All	BA	TF	Epstein-Barr virus ZEBRA protein
2cax	C	All	YU	TF	Repressor omega
2dnj	A	All	CB	EZ	DNase I
2drp	D	All	FE	TF	Tramtrack protein
2ere	B	All	DC	TF	Leu3
2evg	A	All	CB	TF	Ndt80
2ex5	A	All	YX	EZ	Endonuclease I-CeuI
2ezv	A	All	GF	EZ	Endonuclease SfiI
2f5p	A	All	DC	EZ	DNA glycosylase MutM
2fio	B	All	DC	TF	Protein p4
2fl3	A	All	DC	EZ	Endonuclease HinPII
2fld	B	All	DC	EZ	Endonuclease I-MsoI
2fmq	A	All	CB	EZ	DNA polymerase
2g1p	A	All	GF	EZ	DNA methyltransferase
2h27	A	All	CB	TF	Group IV sigma factor
2h7g	X	All	ZY	EZ	Topoisomerase
2i13	A	All	DC	TF	Designed zinc finger protein
2ibs	A	All	CB	EZ	DNA methyltransferase M.TaqI
2ihm	A	All	PT	EZ	DNA polymerase mu
2iie	A	All	DC EC	OT	Integration host factor
2is2	B	All	DC	EZ	UvrD helicase
2ivk	B	All	FE JI	EZ	Endonuclease ColE7
2ntc	A	All	CW	OT	SV40 T-antigen



2o8b	A	457-553	FE	EZ	MutSalpha MSH2
2o8d	B	362-518	FE	EZ	MutSalpha MSH6
2oaa	A	All	DC	EZ	Endonuclease MavI
2odi	B	All	FE	EZ	Endonuclease BcnI
2ofi	A	All	BC	EZ	DNA glycosylase I
2owo	A	All	CB DB	EZ	DNA ligase LigA
2p0j	B	All	DC	EZ	Endonuclease BstYI
2p6r	A	All	YX	EZ	Helicase Hel308
2pvi	B	All	DC	EZ	Endonuclease PvuII
2pyj	B	All	YX RQ JK	EZ	Phi29 DNA polymerase
3cro	L	All	BA	TF	Phage 434 Cro
4crx	B	All	DC	EZ	Cre recombinase
4rve	B	All	ED	EZ	Endonuclease EcoRV
4skn	E	All	BA	EZ	uracil-DNA glycosylase
6cro	A	All	SR	TF	Lamda Cro protein
6pax	A	All	CB	TF	Pax6