Supplementary Data

Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease

John R. Horton[1], Megumu Yamada Mabuchi[2,†], Devora Cohen-Karni[2,†], Xing Zhang[1], Rose Griggs[1], Mala Samaranayake[2], Richard J. Roberts[2], Yu Zheng[2,*] and Xiaodong Cheng[1,*]

[1] Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, Georgia 30322, USA

[2] New England Biolabs, 240 County Road, Ipswich, MA 01938, USA

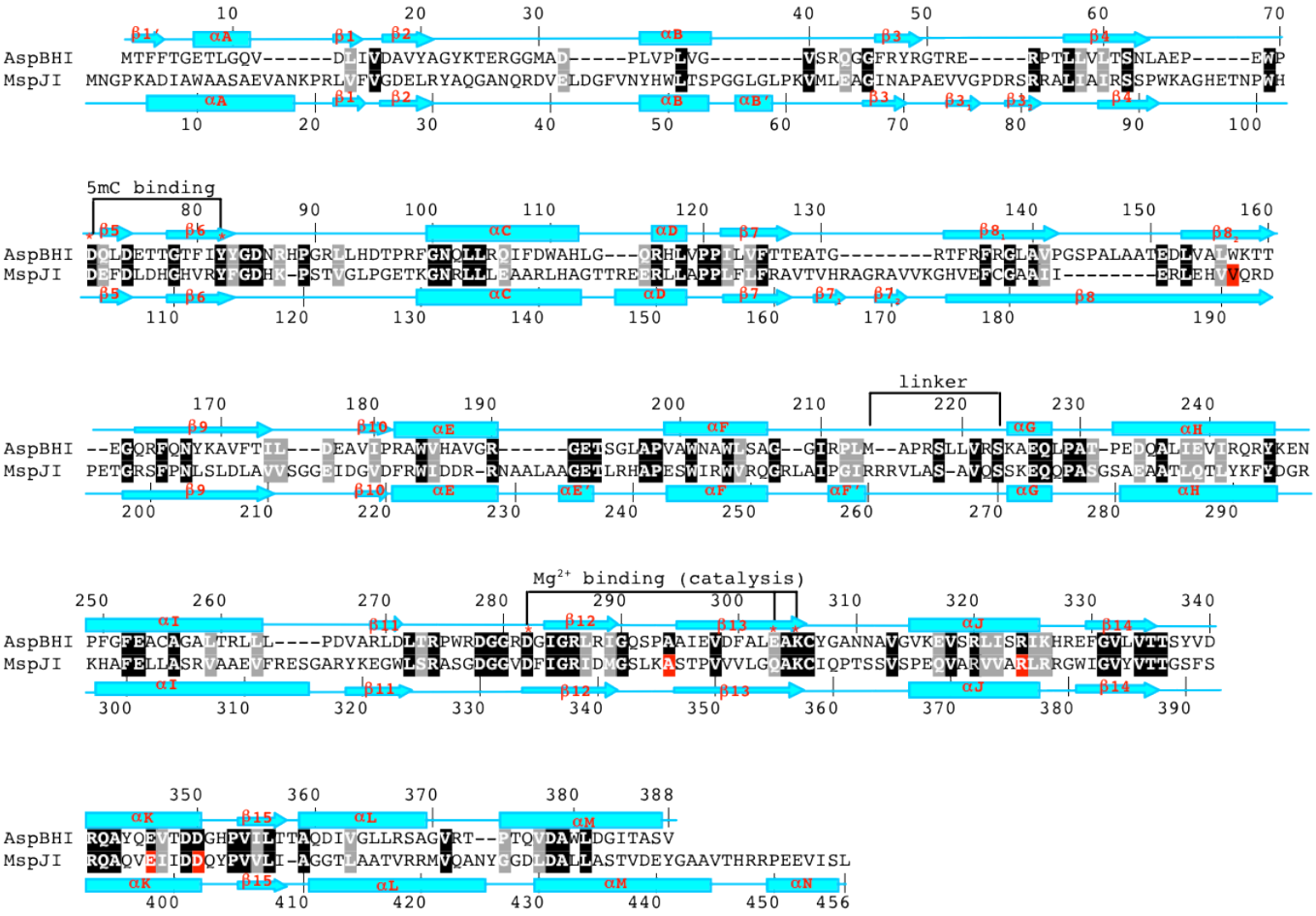
† Both authors contributed equally to this work.

* Correspondence should be addressed to:

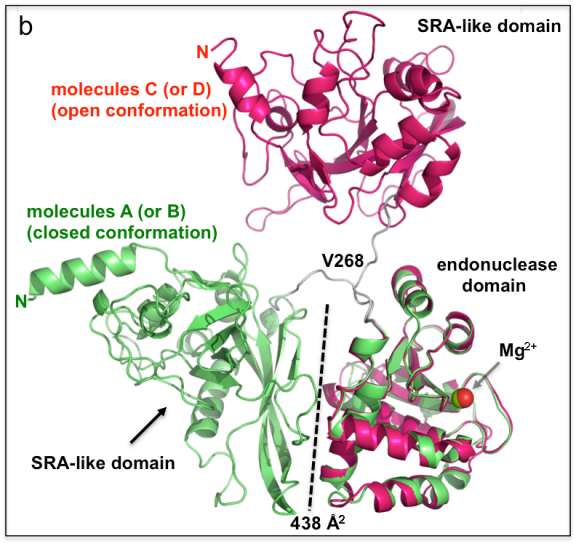
Xiaodong Cheng, Phone: 404-727-8491, Fax: 404-727-3746, Email: xcheng@emory.edu

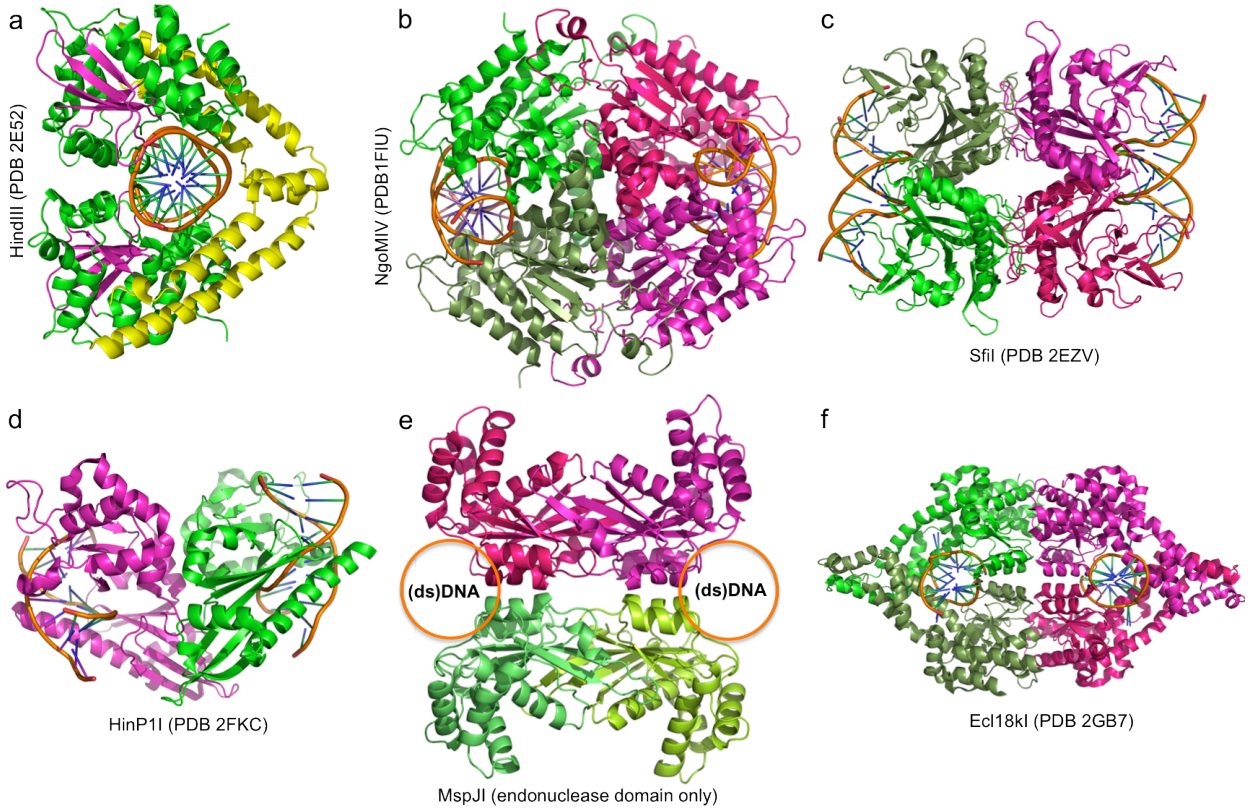
Yu Zheng, Phone: 978-380-7441, Email: zhengy@neb.com

a



Supplementary Figure S1. (a) Structure-based sequence alignment between MspJI (bottom line) and AspBHI (top line). Secondary structural elements are shown above or below the aligned sequences. White-on-black residues are invariant between the two sequences examined, while gray- highlighted positions are conserved (R and K, E and D, T and S, F and Y, V, I, L and M, and G and P). The structure of AspBHI will be described separately. Positions highlighted by * are responsible for 5-methylcytosine recognition or for catalysis. Residues in red were mutated in this study. (**b**) Superimposition of molecules A (in green) and C (in red) with their respective C-terminal endonuclease domains.





Supplementary Figure S2. Structural comparison of various restriction endonucleases

(**a**) Structure of HindIII in complex with DNA (PDB 2E52). The helices in yellow are involved in dimerization.

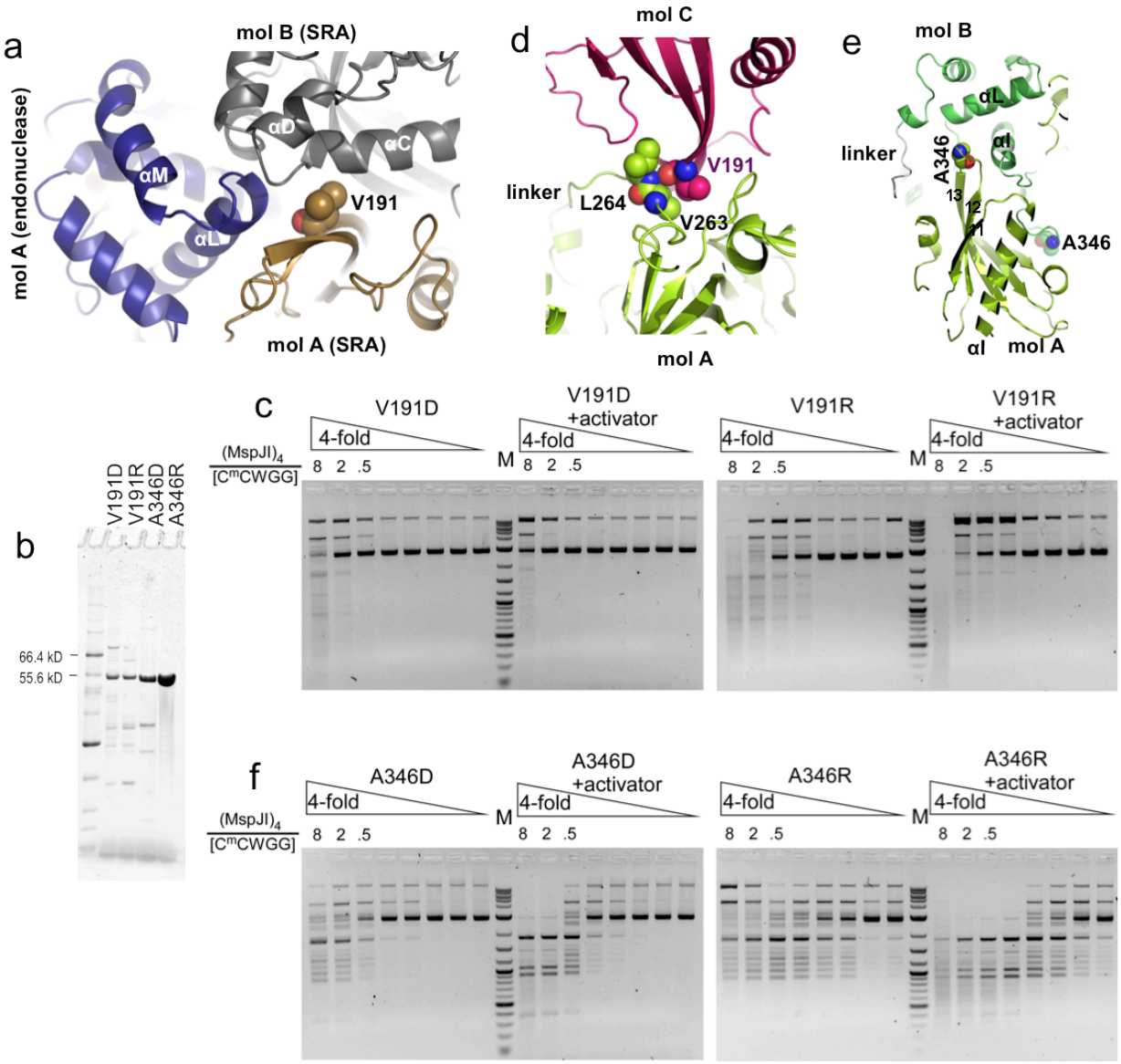
(**b**) Structure of NgoMIV (PDB 1FIU), a tetrameric Type IIF restriction enzyme. Two face-to-face dimers (in green and red) are arranged back-to-back to form a tetramer, resulting in two ds DNA binding and cleavage modules (left and right).

(**c**) Structure of SfiI (PDB 2EZV), two DNA molecules are bound to opposite sides of the SfiI tetramer, in a back-to-back arrangement, with each dimer bind one DNA molecule.

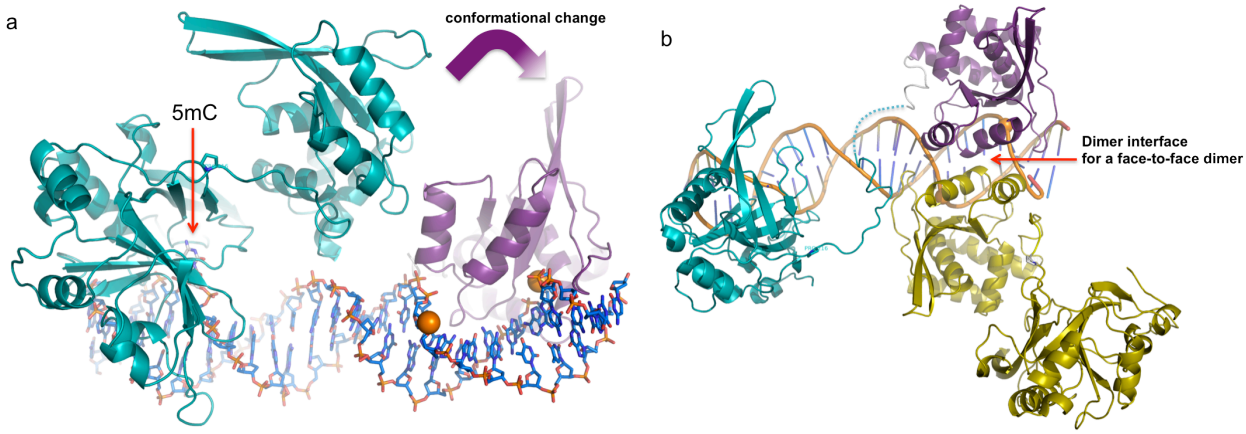
(**d**) Structure of HinP1I in complex with DNA (PDB 2FKC). Two monomers (in green and magenta) are arranged back-to-back, resulting in two DNA binding and nicking sites (left and right).

(**e**) The C-terminal endonuclease domains of MspJI are arranged as dimers of two back-to-back dimers (in green and red) positioned face-to-face, resulting in two ds DNA binding and cleavage modules (left and right).

(**f**) Structure of EcI18kI (PDB 2GB7), which existed as dimers in solution but associated to form tetramers via DNA looping.



Supplementary Figure S3. Mutagenesis of residues involved in dimer/tetramer formation (**a**) Val191 of molecule A sits in the three way junction of the N- and C-domains of molecule A as well as the N-terminal domain of molecule B. (**b**) V191D and V191R have decreased protein yield, whereas A346R has the equivalent level of wild type protein. A346D has intermediate level of protein yield. (**c**) V191D (left panel) and V191R (right panel) exhibited lower specific activity by a factor of approximately 1000 than that of wild type enzyme (Figure 5b), in the absence and presence of activator oligonucleotide containing a 5-methylcytosine (12). (**d**) The corresponding Val191 of molecule C (or D) is involved in tetramer formation via hydrophobic interactions with Val263 and Leu264 of molecule A (or B), but not involved in the dimer formation. (**e**) Ala346, located in the loop between strands β12 and β13 of the C-terminal domain, could accommodate a larger side chain by pointing to solvent. (**f**) Mutations of Ala346 to Asp (A346D) (left panel) and Arg (A346R) (right panel) have no effect on specific activity (approximately the same as that of wide type enzyme, Figure 5b).



Supplementary Figure S4. A modeling study of MspJI monomer bound with ds DNA

(a) We initially performed a modeling study of MspJI monomer bound with ds DNA, which suggested that the C-terminal endonuclease domain of the same monomer bound to the modified cytosine would make the distal N_{16} cut in the opposite strand after a conformational change upon DNA binding. (b) Like FokI, a second molecule would be needed to dimerize at the cleavage site to make the proximal N_{12} cut.

Supplementary Table S1. Primers used in mutagenesis

Mutant	Primers
V191D	5′-GCCTGGAGCACGTCGATCAGCGTGATCCAGAAA-3′ 5′-TTTCTGGATCACGCTGATCGACGTGCTCCAGGC-3′
V191R	5′-GCCTGGAGCACGTCCGCCAGCGTGATCCAGAAA-3′ 5′-TTTCTGGATCACGCTGGCGGACGTGCTCCAGGC-3′
A346D	5′-CATGGGTTCATTGAAAGACTCAACGCCGGTTGTTG-3′ 5′-CAACAACCGGCGTTGAGTCTTTCTAATGAACCCATG-3′
A346R	5′-CATGGGTTCATTGAAACGGTCAACGCCGGTTGTTG-3′ 5′-CAACAACCGGCGTTGACCGTTTCTAATGAACCCATG-3′
R376A	5′-GTGGCGCGCGTGGTCGCCGCGTTGCGCCGCGGTTGGATC-3′ 5′-GATCCAACCGCGGCGCAACGCGGCGACCACGCGCGCCAC-3′
E398A	5′-TCACGCCAAGCCCAAGTGGCGATTATCGATGACCAATAC-3′ 5′-GTATTGGTCATCGATAATCGCCACTTGGGCTTGGCGTGA-3′
D402A	5′-CAAGTGGAAATTATCGATGCGCAATACCCGGTGGTTTTA-3′ 5′-TAAAACCACCGGGTATTGCGCATCGATAATTTCCACTTG-3′

Supplementary Table S2. Summary of diffraction and refinement statistics of MspJI crystals

Data collection	**Native MspJI**	**Native MspJI**	**Hg MspJI**
PDB	4F0Q	4F0P	
Space group	**$P2_1$**	**$P3_1$**	**$P3_1$**
Cell dimensions	$\alpha=\gamma=90°$	$\alpha=\beta=90°$ and $\gamma=120°$	
(Å)	a=87.76 b=144.28 c= 87.84 β=116.3°	a=b=144.95 c=101.45	a=b=146.04 c=101.63
Beamline (SERCAT)	APS 22-BM	APS 22-BM	APS 22-ID
Wavelength (Å)	1.00000	1.00000	1.00000
Resolution (Å)*	34.57-2.05 (2.12-2.05)	34.65-2.79 (2.89-2.79)	34.82-2.99 (3.10-2.99)
R_{sym} or R_{merge*}	0.138 (0.651)	0.129 (0.740)	0.102 (0.615)
I/σI *	14.6 (3.8)	28.2 (5.7)	24.9 (4.8)
Completeness (%)*	99.5 (100.0)	100.0 (100.0)	100.0 (99.9)
Redundancy*	8.5 (8.3)	19.2 (18.9)	11.8 (11.8)
Observed reflections	1,040,363	1,137,792	578,428
Unique reflections*	122,248 (12,227)	59,294 (5,926)	48,970 (4,931) (48,922 have both I^+ and I^-)
Mean FOM (SAD) after refinement:			0.31
Density Modification (SAD), R-factor:			0.2518
Refinement			
Resolution (Å)	2.05	2.79	
No. reflections	122,216	59,272	
R_{work} / R_{free}	0.213/0.246	0.164/0.221	
Twinning Fraction	0.393	0.223	
(Operator)	(l, -k, h)	(h, -h-k, -l)	
No. Atoms			
Protein (closed dimer)	6826	6828	
Protein (open dimer)	6533	6513	
Metal Ion	3	2	
Water	989	296	
B Factors ($Å^2$)			
Protein (closed dimer)	24.0	49.4	
Protein (open dimer)	33.5	76.3	
Metal Ion	24.9	60.6	
Water	27.4	42.7	
R.m.s. deviations			
Bond lengths (Å)	0.002	0.006	
Bond angles (°)	0.59	0.60	

*Values in parenthesis correspond to highest resolution shell

Supplementary Table S3. Structural similarity of C-terminal endonuclease domain of MspJI

	PDB C D	Ali. Len.	SCORE	P-VAL	RMSD	%Id	Description
☐	1Y88 A 1	95	13.3	10e-10.4	1.8	12.6	Hypothetical Protein Af1548
☐	1Y88 A	91	13.3	10e-8.6	1.7	12.1	Hypothetical Protein Af1548
☐	2PH7 A 2	89	11.1	10e-8.4	3.5	13.5	Crystal Structure Of Af2093 From Archaeoglobus Fulgidus
☐	2E52 A 1	84	11.6	10e-6.5	2.2	14.3	Crystal Structural Analysis Of Hindiii Restriction Endonuclease In Complex With Cognate Dna At 2.0 Angstrom Resolution
☐	2VLD A 2	73	10.7	10e-6.0	1.6	15.1	Crystal Structure Of A Repair Endonuclease From Pyrococcus Abyssiÿ
☐	3LWL B 2	70	10.7	10e-5.7	2.5	2.9	Crystal Structure Of Staphylococcus Aureus Protein Sa1388ÿ
☐	2PH7 A	89	11.1	10e-5.4	3.3	13.5	Crystal Structure Of Af2093 From Archaeoglobus Fulgidus
☐	1NM0 A 2	50	9.8	10e-4.9	1.7	10.0	Structural Genomics, Protein Ybgi, Unknown Function
☐	2EWF A 4	107	11.6	10e-4.9	3.0	11.2	Crystal Structure Of The Site-Specific Dna Nickase N.Bspd6i
☐	2E52 A	83	11.6	10e-4.9	2.2	14.5	Crystal Structural Analysis Of Hindiii Restriction Endonuclease In Complex With Cognate Dna At 2.0 Angstrom Resolution
☐	2OT9 A 2	76	9.2	10e-4.8	2.7	10.5	Crystal Structure Of Yaeq Protein From Pseudomonas Syringae
☐	2CZR A 1	60	8.5	10e-4.6	2.2	8.3	Crystal Structure Of Tbp-Interacting Protein (Tk-Tip26) And Implications For Its Inhibition Mechanism Of The Interaction Between Tbp And Tata-Dna
☐	2YYE B 1	55	8.3	10e-4.6	3.3	10.9	Crystal Structure Of Selenophosphate Synthetase
☐	1DFM A 1	50	8.7	10e-4.3	2.0	18.0	Crystal Structure Of Restriction Endonuclease Bglii Complexed With Dna 16-Merÿ
☐	3LUK A	79	9.5	10e-4.1	3.5	8.9	Crystal Structure Of Mid Domain From Hago2ÿ
☐	2JPD A	61	9.0	10e-4.1	3.0	8.2	Solution Structure Of The Ercc1 Central Domain
☐	1FOK A 4	113	10.0	0.0001	3.3	8.8	Structure Of Restriction Endonuclease Foki Bound To Dnaÿ
☐	1XMX A 2	83	10.2	0.0002	1.7	10.8	Crystal Structure Of Hypothetical Protein Vc1899, Mcsg Apc26666
☐	3CAE A 1	38	7.1	0.0002	2.0	13.2	Structure Of Nnqqny As An Insert In T7 Endonuclease I
☐	3BM3 A	78	10.7	0.0003	3.1	15.4	Restriction Endonuclease Pspgi-Substrate Dna Complex