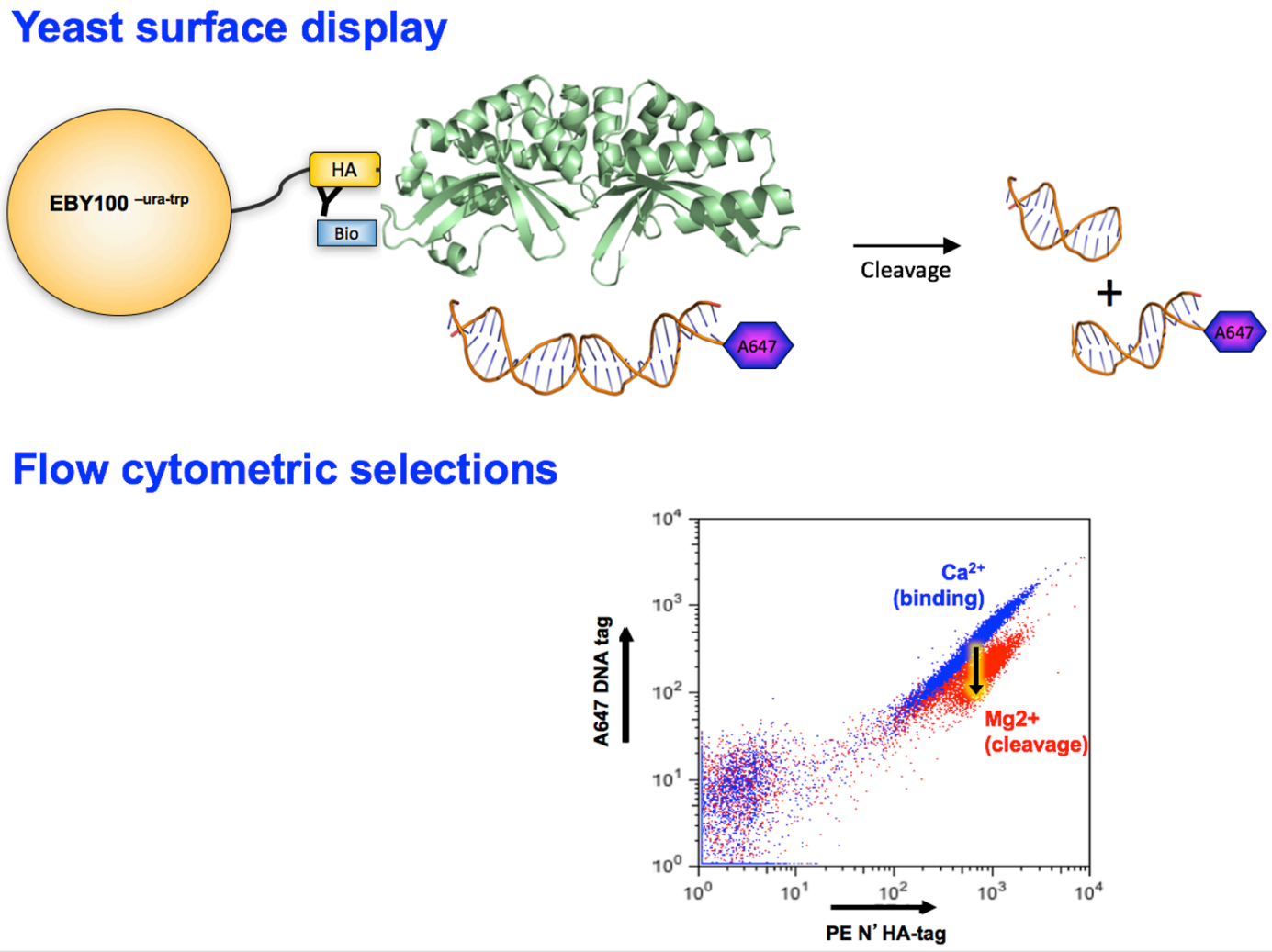
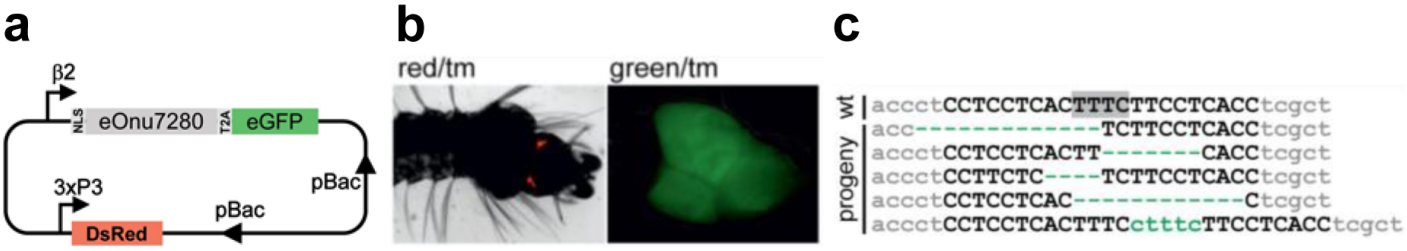
SUPPLEMENTARY DATA

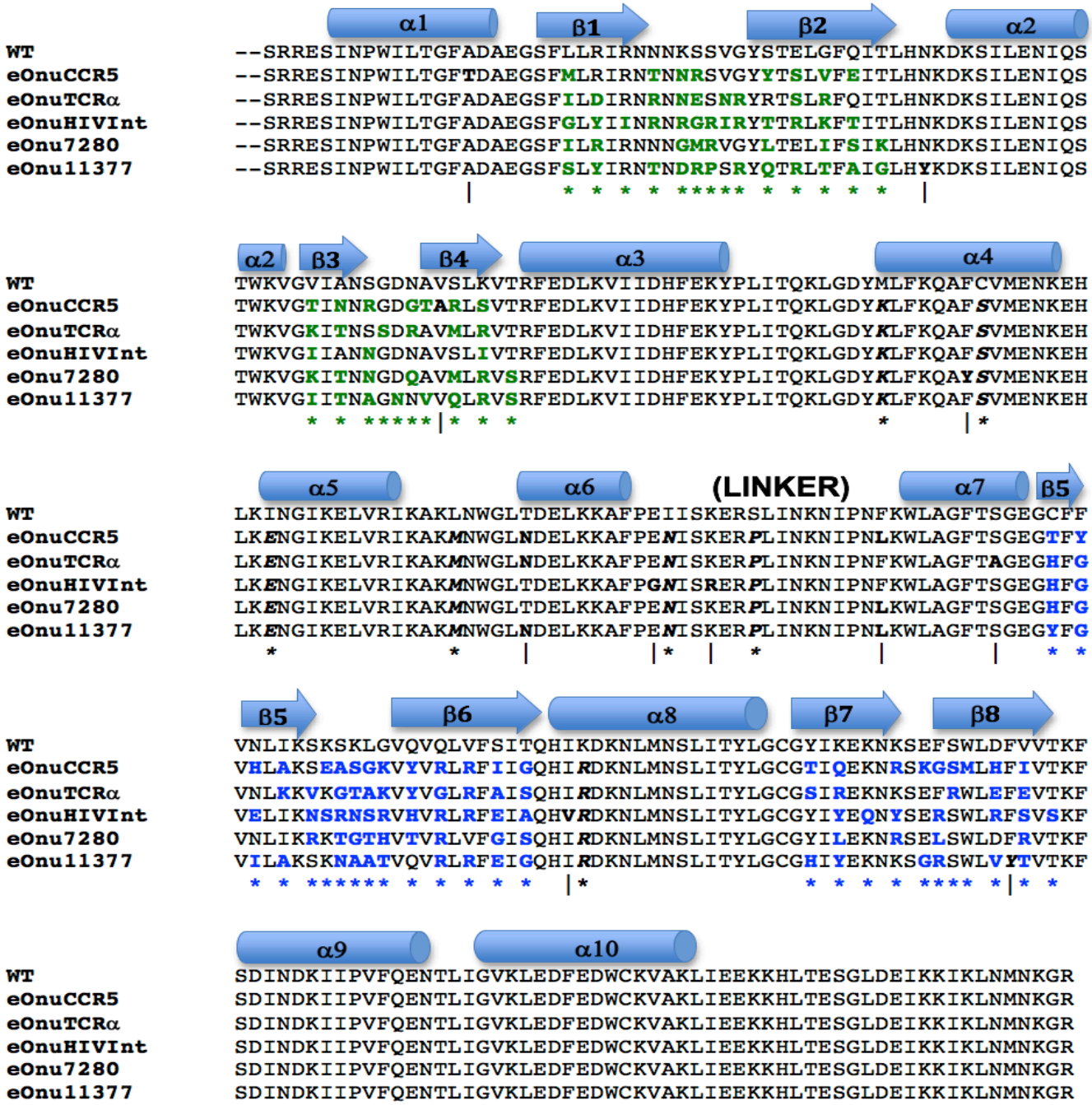
Supplementary Figure S1. Method of yeast surface display and flow cytometry to measure enzyme specificity. The enzyme being assayed is displayed on the surface of yeast, and its ability of cleave each substrate (both the unmodified, intended target, and a panel of 66 separate DNA targets that each differ from the intended target by a single base pair substitution as one position) is measured by cell staining and flow cytometric analyses. In each separate cleavage measurement, a fluorescently labeled DNA substrate is provided in the presence of Ca^{2+} (which allows binding but fully inhibits cleavage) versus Mg^{2+} (which allows cleavage, and results in a reduction in staining fluorescence intensity). See *Methods* and references (31,33) for further experimental details.



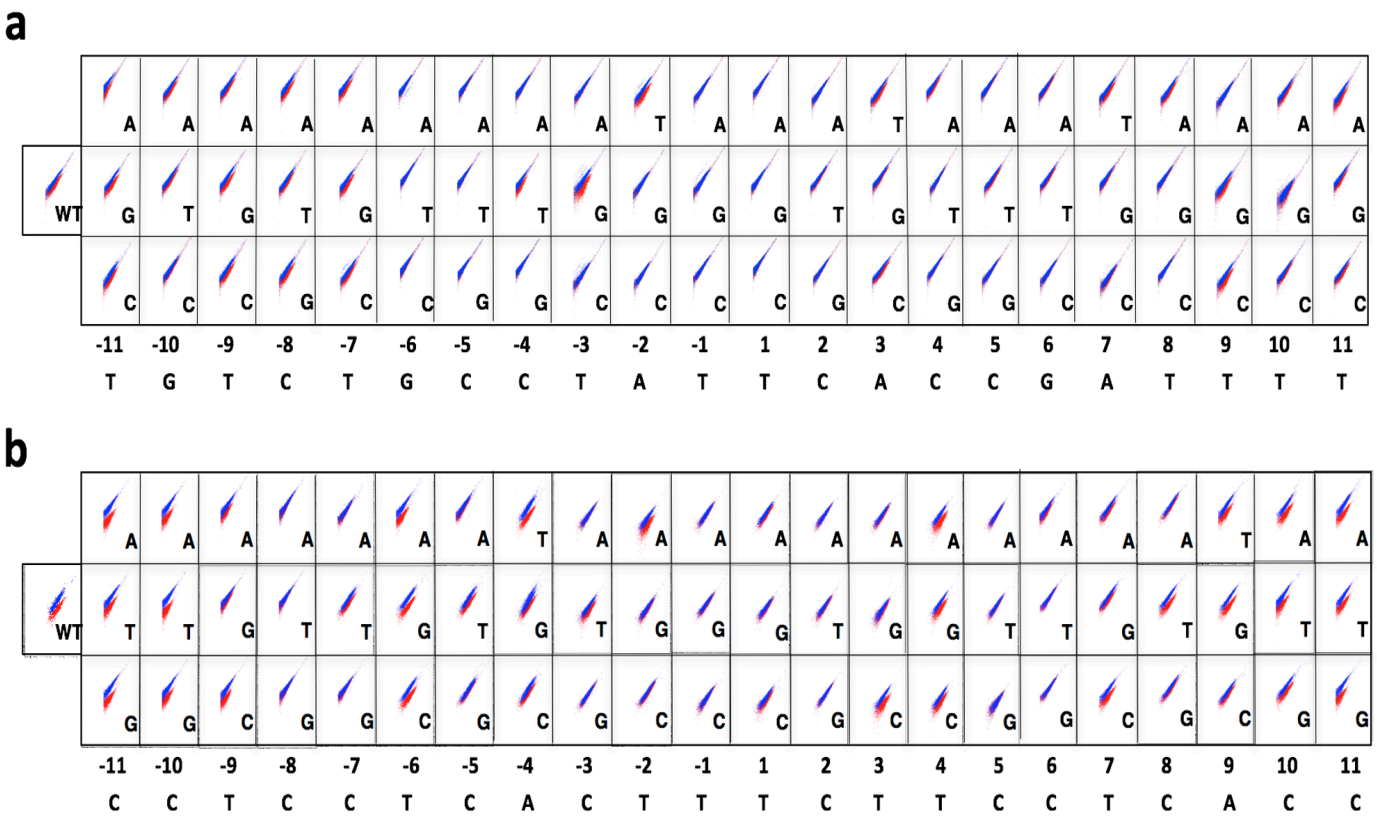
Supplementary Figure S2. Generation of transgenic mosquitos harboring and expressing eOnu7280. ***Panel a:*** Schematic diagram of the transformation vector used to generate the I1-H7280A1 transgenic mosquito strain. Contains PiggyBac inverted repeats (pBac), a Pax promoter (3xP3), a marker gene (DsRed), the mail spermatogenesis specific b2 tubulin promoter (b2), the eOnu7280 endonuclease reading frame (eOnu7280) self-cleaving peptide signals (T2A) and a SV40 nuclear localization signal (NLS). ***Panel b:*** Overlays of transmission (tm) and fluorescent images of the larval head and dissected adult testes of male mosquitos. ***Panel c:*** Native target site in the AGAP007280 gene and representative disruptions identified in F1 progeny of transgenic I1-H7280A1 male mosquitos.



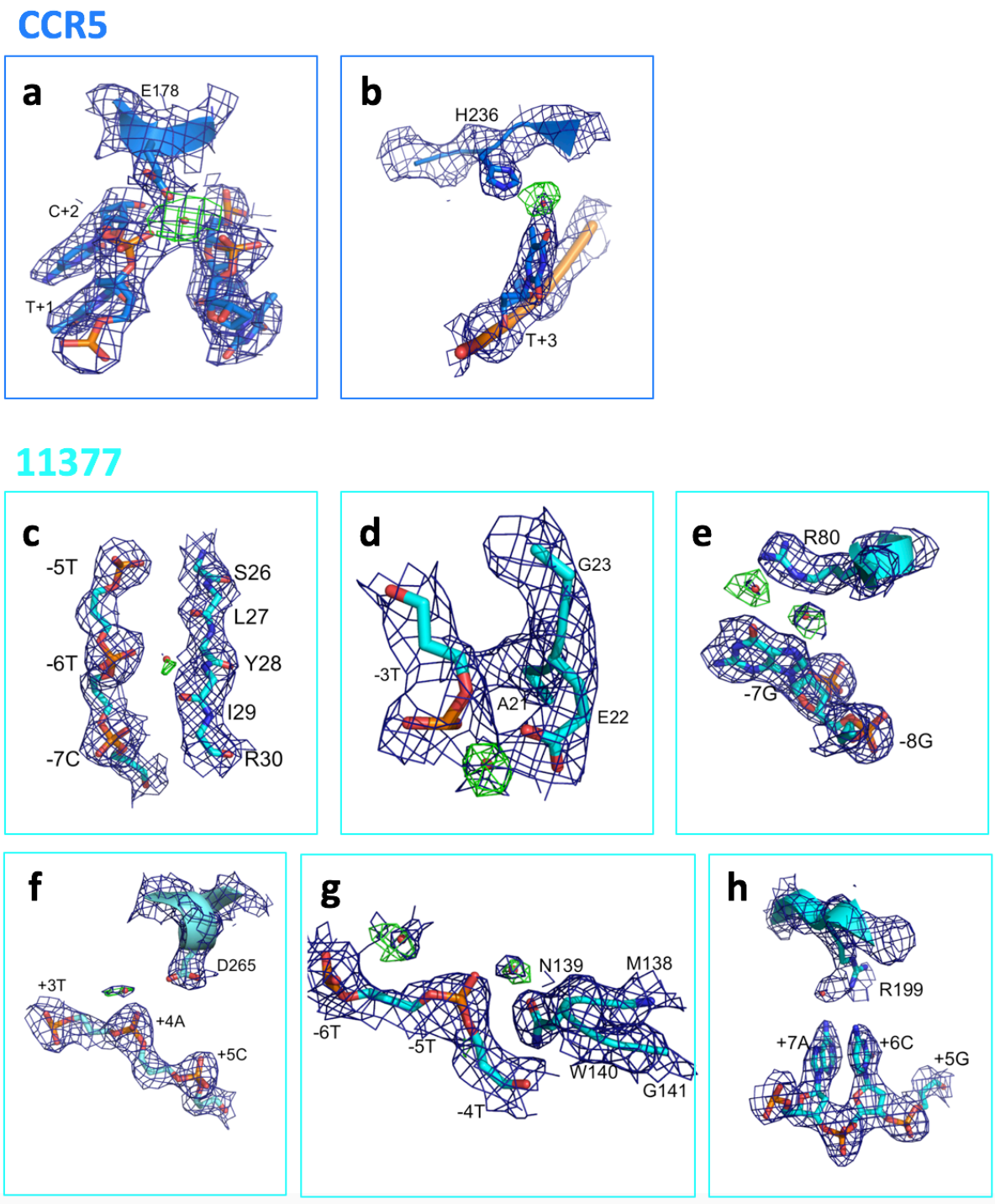
Supplementary Figure S3. Sequence alignment of eOnu constructs. Related to Figure 1. Secondary structural elements, and the linker that connects the N- and C-terminal protein domains, are indicated above the sequence alignments. Positions that have been mutated on the protein surface and inter-domain linker to improve protein solubility, and that are present in all 5 engineered constructs are indicated with bold font and underlying black asterisks. Positions in the DNA-binding surface that were subjected to randomization and selection, and that are altered relative to the same position in the wild-type enzyme, are indicated with bold green (N-terminal domain) or bold blue (C-terminal domain) fonts and similarly colored underlying asterisks. Positions that were found to be mutated at other scaffold positions in individual engineered constructs (due to random PCR-induced mutations) are indicated with bold font and vertical lines. The sequence of the first two N-terminal residues differ depending on whether the construct was being studied with activity assays using yeast surface display or being used for crystallization experiments, and the positions are therefore indicated with dashes.



Supplementary Figure S4. 'One-off' cleavage specificity analyses of eOnuTCRα and eOnu7280. Related to Supplementary Figure S1 and Figure 5. ***Panel a:*** Cleavage data for eOnuTCRα (presented in tabulated form in **Figure 5)** for a panel of substrates that each differ from the enzyme's intended target by a single basepair substitution at one position**.** ***Panel b:*** Similar cleavage data for eOnu7280 (also presented in tabulated form in **Figure 5).**



Supplementary Figure S5. Evidence for waters mediating direct contacts between DNA and protein in low resolution structures. eOnuCCR5 (slate blue; a and b) and eOnu11377 (cyan; c through h) have waters modeled and refined at positions in the protein-DNA interface that appear to participate in target recognition. Unbiased simulated annealing omit maps were generate with Phenix for all waters in the eOnuCCR5 and eOnu11377 structures. The waters mediating direct contacts between the DNA and protein are shown in their DNA/protein context with 2Fo-Fc density maps (dark blue) contoured at a sigma cut-off 1.0, and Fo-Fc difference maps (green) contoured at a sigma cut-off 3.0. Green density, indicating the presence of 3s peak in the S.A. omit Fo-Fc difference map, is visible for each water except in panel h, where the sidechain of R199 moved into the place of the water; at that position the 2Fo-Fc density (dark blue) around the water is presented. The remaining waters in the two structures were modeled according to the same criteria, and are not located in the protein-DNA interface.



Supplementary Table S1. Crystallographic data and refinement statistics

Structure	WT	eOnuCCR5	eOnuTCRα	eOnuHIVInt	eOnu7280	eOnu11377
PDB Code	3QQY	5THG	5T2H	5T8D	5T2N	5T2O
Data Collection						
Space group	P2(1)2(1)2(1)	P2(1)2(1)2(1)	P2(1)2(1)2(1)	P2(1)2(1)2(1)	P2(1)2(1)2(1)	P2(1)2(1)2(1)
Cell dimensions						
a, b, c (Å)	37.95, 73.93, 166.93	75.96, 112.35, 122.88	39.70, 68.93, 157.53	39.70, 75.01, 165.16	40.83, 67.42, 166.88	40.18, 64.84, 167.01
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	44.46 - 2.40	47.77-3.11 (3.2 - 3.11)	50.0 - 2.52 (2.61 – 2.52)	50.0 - 2.15 (2.23 - 2.15)	50.0 - 2.08 (2.15 - 2.08)	50.0 – 2.80 (2.90 – 2.80)
Rmerge	0.080 (0.140)	0.157 (0.649)	0.102 (0.805)	0.093 (0.618)	0.094 (0.793)	0.155 (0.877)
I/σI	18.6 (9.62)	12.7 (2.2)	25.7 (2.2)	29.3 (3.2)	19.1 (2.3)	15.4 (2.3)
Completeness (%)	96.8 (91.0)	99.9 (100.0)	99.2 (95.3)	99.9 (99.2)	99.6 (99.0)	98.4 (96.9)
Redundancy	6.7 (4.8)	7.2 (6.8)	13.0 (10.7)	12.6 (11.0)	11.9 (8.9)	12.6 (10.9)
CC 1/2		0.924	0.853	0.904	0.883	0.859
Refinement						
No. Reflections	18568	19313	15161	27694	28473	11138
Rwork (Rfree) (%)	18.9 (24.0)	22.64 (27.88)	21.16 (26.53)	18.22 (22.95)	21.99 (25.78)	22.43 (28.08)
No. Complex in ASU	1	2	1	1	1	1
No. Atoms						
Protein	2419	4414	2292	2335	2269	2301
DNA	1060	2318	1025	1066	1063	1066
Ca2+	0	4	2	3	4	2
Mg2+	1	0	0	0	0	0
Water	73	13	25	261	59	13
B-factor	30.56	59.3	54.8	34.3	49.8	47.4
R.m.s deviations						
Bond lengths (Å)	0.018	0.006	0.003	0.007	0.003	0.003
Bond angles (°)	2.217	0.690	0.517	0.998	0.697	0.545
Ramachandran						
Preferred (%)	90.40	92.29	96.26	95.83	95.12	96.19
Allowed (%)	8.50	5.43	3.74	4.17	4.53	3.11
Outliers (%)	1.10	2.28	0.00	0.00	0.35	0.69

Supplementary Table S2. Pairwise structural RMSD values for superposed proteins, DNA targets, and protein-DNA complexes, respectively. The number of alpha carbons overlapped in protein alignments is shown in parentheses.

PROTEIN	I-OnuI	eOnuCCR5	eOnuTCRa	eOnuHIVInt	eOnu7280	eOnu11377
I-OnuI	0 (299)	0.70 (253)	0.52 (252)	0.50 (256)	0.59 (246)	0.66 (255)
eOnuCCR5		0 (294)	0.57 (242)	0.55 (239)	0.65 (238)	0.61 (249)
eOnuTCRa			0 (296)	0.30 (242)	0.41 (241)	0.39 (234)
eOnuHIVInt				0 (292)	0.31 (215)	0.37 (230)
eOnu7280					0 (292)	0.38 (226)
eOnu11377						0 (293)

DNA	I-OnuI	eOnuCCR5	eOnuTCRa	eOnuHIVInt	eOnu7280	eOnu11377
I-OnuI	0	1.53	1.34	1.77	1.42	1.74
eOnuCCR5		0	1.45	1.66	1.41	1.51
eOnuTCRa			0	1.59	1.32	1.48
eOnuHIVInt				0	1.51	1.45
eOnu7280					0	1.26
eOnu11377						0

COMPLEX	I-OnuI	eOnuCCR5	eOnuTCRa	eOnuHIVInt	eOnu7280	eOnu11377
I-OnuI	0	1.71	1.77	1.58	1.42	1.81
eOnuCCR5		0	1.57	1.45	1.56	1.69
eOnuTCRa			0	1.07	1.36	1.68
eOnuHIVInt				0	1.11	1.55
eOnu7280					0	1.46
eOnu11377						0