STRUCTURAL
BIOLOGY

Volume 76 (2020)

Supporting information for article:

Structural basis for designing an array of engrailed homeodomains

Tomoko Sunami, Yu Hirano, Taro Tamada and Hidetoshi Kono

S1. Sequence

S1.1. Protein sequences

Crystallization and EMSA

- $(EHD[R53A])_2$

MGSSHHHHHHAIEDLYFQSPG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKAAKIKKS-
GGGGG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKAAKIKKSGT

B1H assay

- EHD[wt]-library

MARVTXQDAVEKIGNRFDLVLVAARRARQMQVGGKDPLVPEENDKTTVIALREIEEGLINN
QILDVRERQEQQEQEAAELQAVTAIAEGRAAADYKDDDDKFRTGSKTPPHGTG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIK**X**WF**X**NKR**X**KIKKSGS

- $(EHD[R53A])_2$-library

MARVTXQDAVEKIGNRFDLVLVAARRARQMQVGGKDPLVPEENDKTTVIALREIEEGLINN
QILDVRERQEQQEQEAAELQAVTAIAEGRAAADYKDDDDKFRTGSKTPPHGTG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKAAKIKKS-
GGGGG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIK**X**WF**X**NKA**X**KIKKSGS

Amino acid sequences corresponding to *Sac*I and *Bam*HI sites that were used to insert DNA oligo shown in S1.2 are underlined and doubly-underlined, respectively. The symbol "**X**" represents randomized amino acid.

EMSA

- $(EHD[R53A])_2$

MGSSHHHHHHAIEDLYFQSPG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKAAKIKKS-

GGGGG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKAAKIKKSGT

- EHD[R53A, R5E]-EHD[R53A]

MGSSHHHHHHAIEDLYFQSPG-
DEKRPETAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKAAKIKKS-
GGGGG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKAAKIKKSGT

- EHD[R53A]-EHD[R5E, R53A]

MGSSHHHHHHAIEDLYFQSPG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKAAKIKKS-
GGGGG-
DEKRPETAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKAAKIKKSGT

- EHD[wt]

MGSSHHHHHHAIEDLYFQSPG-
DEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKRAKIKKSGT

- EHD[R5E]

MGSSHHHHHHAIEDLYFQSPG-
DEKRPETAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFKNKRAKIKKSGT

S1.2. Oligonucleotide sequences

B1H assay

Library sequence

- EHD[wt]-library

5’-C GGC CTG AAC GAG GCA CAG ATC AAA NNS TGG TTC NNS AAC AAG CGG NNS AAG ATC AAG AAG TCG GG-3’

5’-GATCCCGACTTCTTGATCTT-3’

5’-CTCGTTCAGGCCGAGCT-3’

- (EHD[R53A])$_2$-library

5’-C GGC CTG AAC GAG GCA CAG ATC AAA NNS TGG TTC NNS AAC AAG GCG NNS AAG ATC AAG AAG TCG GG-3’

5’-GATCCCGACTTCTTGATCTT-3’

5’-CTCGTTCAGGCCGAGCT-3’

Insert in pH3U3 plasmid for B1H assay

- Seq1

5’-[*Eco*RI]-GCACTACTAATCCGGCGGCTGTGATCGAA-[*Not*I]-3’

- Seq2

5’-[*Eco*RI]-GCTAATCCATTTGCTGAGGCGTAGCCCGC-[*Not*I]-3’

- Seq3

5’-[*Eco*RI]-GGTTATAATCCTAATCCCCCGGCCCACAC-[*Not*I]-3’

Sequencing primer for B1H assay

5’-CTGCGCGTAACCACCACA-3’

EMSA

- TAATCC

5’-<CY5>-CGCAGTG-TAATCC-CCTCGAC-3’

3’-GCGTCAC-ATTAGG-GGAGCTG-5’

- TAATCC-TAATCC

5’-<CY5>-CGCAGTG-TAATCC-TAATCC-CCTCGAC-3’

3’-GCGTCAC-ATTAGG-ATTAGG-GGAGCTG-5’

S2. EMSA

S2.1. Protein preparation

We constructed plasmids as shown in 2.1. All proteins were expressed in *E. coli* BL21(DE3) cells with the autoinduction method (Studier, 2005). The harvested cells were lysed in binding buffer (10 mM Tris pH 7.5, 500 mM NaCl) and loaded onto a HisTrap FF Crude column (GE healthcare). The bound proteins were eluted using a linear imidazole gradient, after which the imidazole was removed with centrifugal filtration. SDS-PAGE of the final products is shown Figure S2. The protein sequences used in this study are listed in Supporting information S1.1.

S2.2. Electrophoretic mobility shift assay

In vitro DNA-binding affinity of EHD was measured using electrophoretic mobility shift assays (EMSAs). CY5-labeled and unlabeled oligonucleotides were purchased from Eurofins Genomics and annealed to form dsDNA as shown in section 2.2. Reactions were run in a 20-µl volume. The reaction mixture consisted of 0.5 nM CY5-labeled DNAs, 20 mM HEPES (pH 7.6), 100 mM KCl, 2 mM $MgCl_2$, 1 mM EDTA, 0.02 % (v/v) NP40 substitute, 5% (v/v) glycerol, 0.1 µg/ml BSA, 20 ng/µl calf thymus DNA, and various concentrations of the proteins. The mixtures were incubated for 30 min at room temperature, after which 5-µl aliquots were loaded onto 15% polyacrylamide gels in 10 mM Tris-acetate and 0.25 mM EDTA. The gels were run for 30-40 min. Fluorescence was acquired in Pharos FX (Bio-Rad). To derive the dissociation constant (K_d), the bound DNA fraction (ϑ) was fitted to the following equation,

$$\vartheta = \frac{\mathrm{D} + \mathrm{P} + K_d - \sqrt{(\mathrm{D} + \mathrm{P} + K_d)^2 - 4K_d\mathrm{D}}}{2\mathrm{D}}$$

where D and P are the total DNA and protein concentrations, respectively. The fitting was accomplished using KaleidaGraph 4.0J. The DNA oligo used for EMSAs is shown in Supporting information S1.2.

Table S1 Apparent dissociation constant determined from EMSAs

Name	Kd (nM) *
EHD[wt]	19.2 ± 0.6
EHD[R5E]	ND
$(EHD[R53A])_2$	22.9 ± 2.9
EHD[R5E, R53A]-EHD[R53A]	143 ± 47
EHD[R53A]-EHD[R5E, R53A]	182 ± 72

* Target sequences are TAATCC for EHD[wt] and EHD[R5E], and TAATCC-TAATCC for $(EHD[R53A])_2$, EHD[R5E, R53A]-EHD[R53A] and EHD[R53A]-EHD[R5E, R53A], respectively. The means and standard deviations were obtained from three repeated experiments.

Table S2 BI/BII conformations in $(EHD[R53A])_2$-DNA complex

Chain: Residue	**BI/BII (ε-ζ) ***		**Chain: Residue**	**BI/BII (ε-ζ) ***	
	Alt A	**Alt B**		**Alt A**	**Alt B**
A: T1-A2	BI (-66.6)	BI (-87.1)	**B: A12-T11**	BI (-73.4)	BI (-44.3)
A: A2-A3	BI (-62.6)	BI (-80.5)	**B: T11-T10**	BI (-85.4)	BI (-102.3)
A: A3-T4	BI (-69.1)	BI (-92)	**B: T10-A9**	BI (-100.6)	BI (-60.5)
A: T4-C5	BI (-77.1)	BI (-93.2)	**B: A9-G8**	BII (111.1)	BI (-86.6)
A: C5-C6	BI (-57.8)	BI (-83.7)	**B: G8-G7**	BI (-79.0)	BII (67.9)
A: C6-T7	BI (-92.7)	BI (-70.2)	**B: G7-A6**	BI (15.7)	BI (-66.6)
A: T7-A8	BII (70.9)	BII (51.5)	**B: A6-T5**	BI (-66.2)	BI (-72.8)
A: A8-A9	BI (-101.9)	BI (-86.9)	**B: T5-T4**	BI (-126.1)	BI (-81.5)
A: A9-T10	BI (-79.6)	BI (-76)	**B: T4-A3**	BI (-77.7)	BI (-97.4)
A: T10-C11	BI (-70.7)	BI (-52.7)	**B: A3-G2**	BI (-83.9)	BI (-94.8)
A: C11-C12	BI (-81.7)	BI (-92.3)	**B: G2-G1**	BI (-50.0)	BI (-18.5)

* ε and ζ are torsional angles of C4'(i-1)-C3'(i-1)-O3'(i-1)-P and C3'(i-1)-O3'(i-1)-P-O5', respectively. BI/BII is defined as BI: ε-ζ= [-160, +20], BII: ε-ζ= [+20, +200].

(a)

Apparent Kd of EHD and EHD_2 using EMSA

Name	Kd (nM)
EHD[wt]	22.7 ± 0.6
$(EHD[wt])_2$	0.93 ± 0.13
$(EHD[R53A])_2$	16.9 ± 1.6

(b)

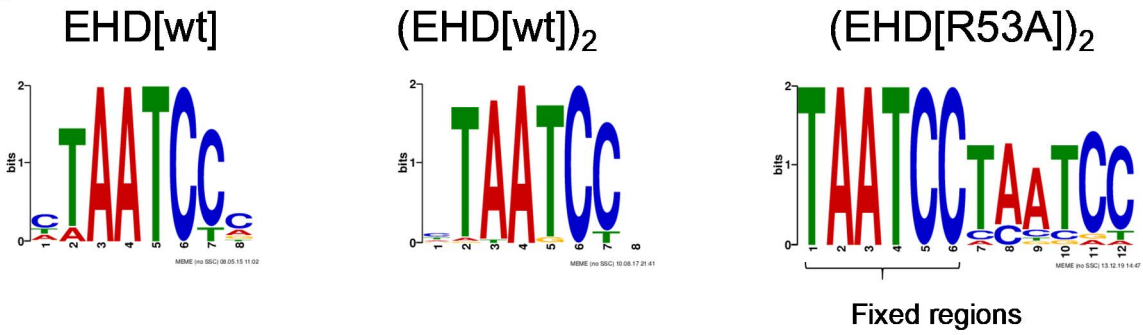

Figure S1 Affinity and DNA specificity of EHD and EHD_2. (a) Apparent Kd of EHD[wt], $(EHD[wt])_2$ and $(EHD[R53A])_2$ obtained from electrophoretic mobility shift assays. TAATCC and TAATCC-TAATCC were used as target sequences for EHD[wt] and for $(EHD[wt])_2$, respectively, and TAATCC-TAATCC for $(EHD[R53A])_2$, (b) Binding sequence profiles for EHD[wt], $(EHD[wt])_2$ and $(EHD[R53A])_2$ obtained by B1H assays. The table and the figures were generated from the data in our previous report (Sunami & Kono, 2019).

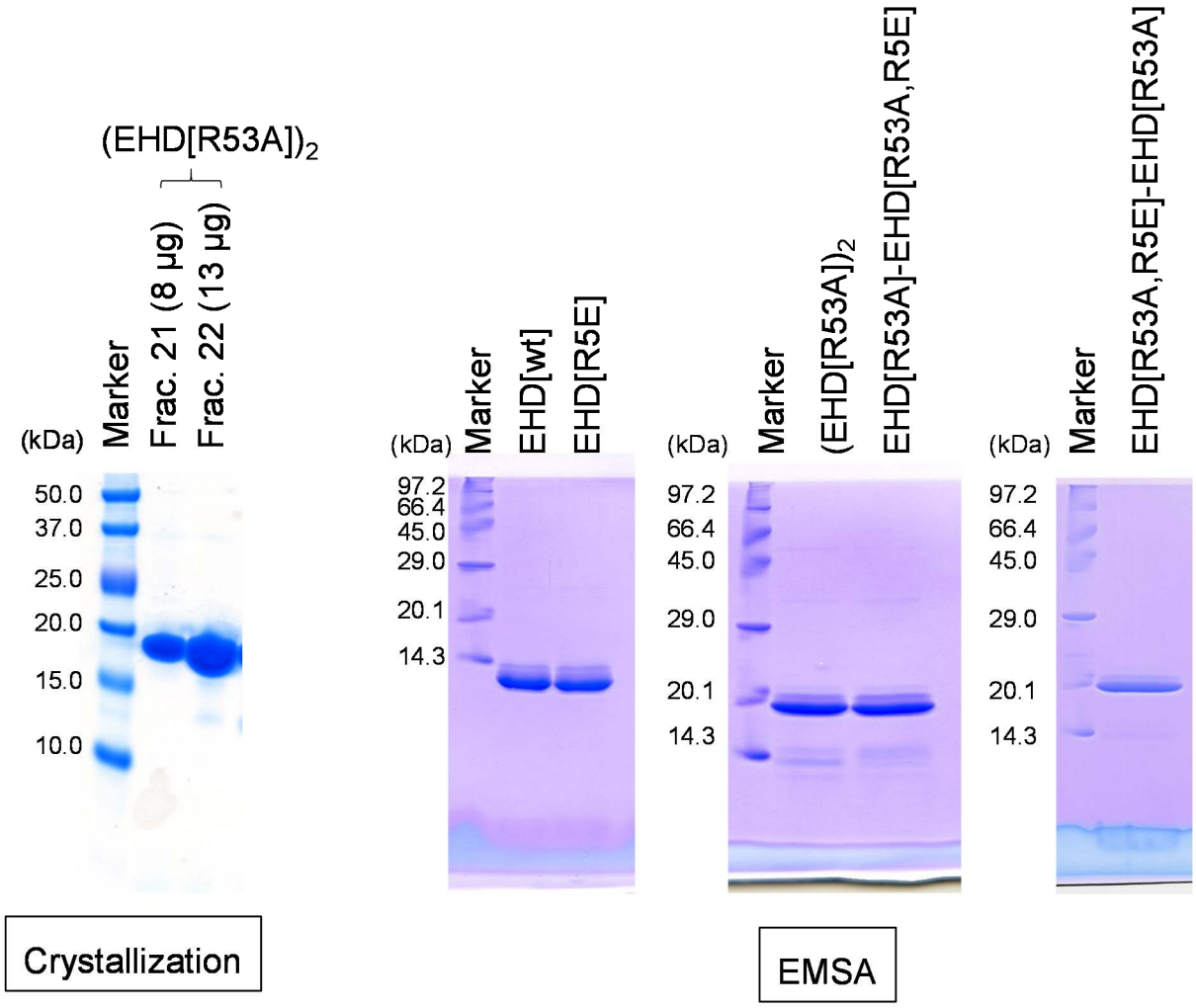

Figure S2 SDS-PAGE of purified proteins for crystallization and EMSAs. Protein samples for crystallization were loaded onto 10% acrylamide gel. For EMSAs, 5-µg protein samples were loaded into each well of the 18% or 15% acrylamide gels. The gels were stained with CBB. The purity of all proteins was estimated to be greater than 90%.

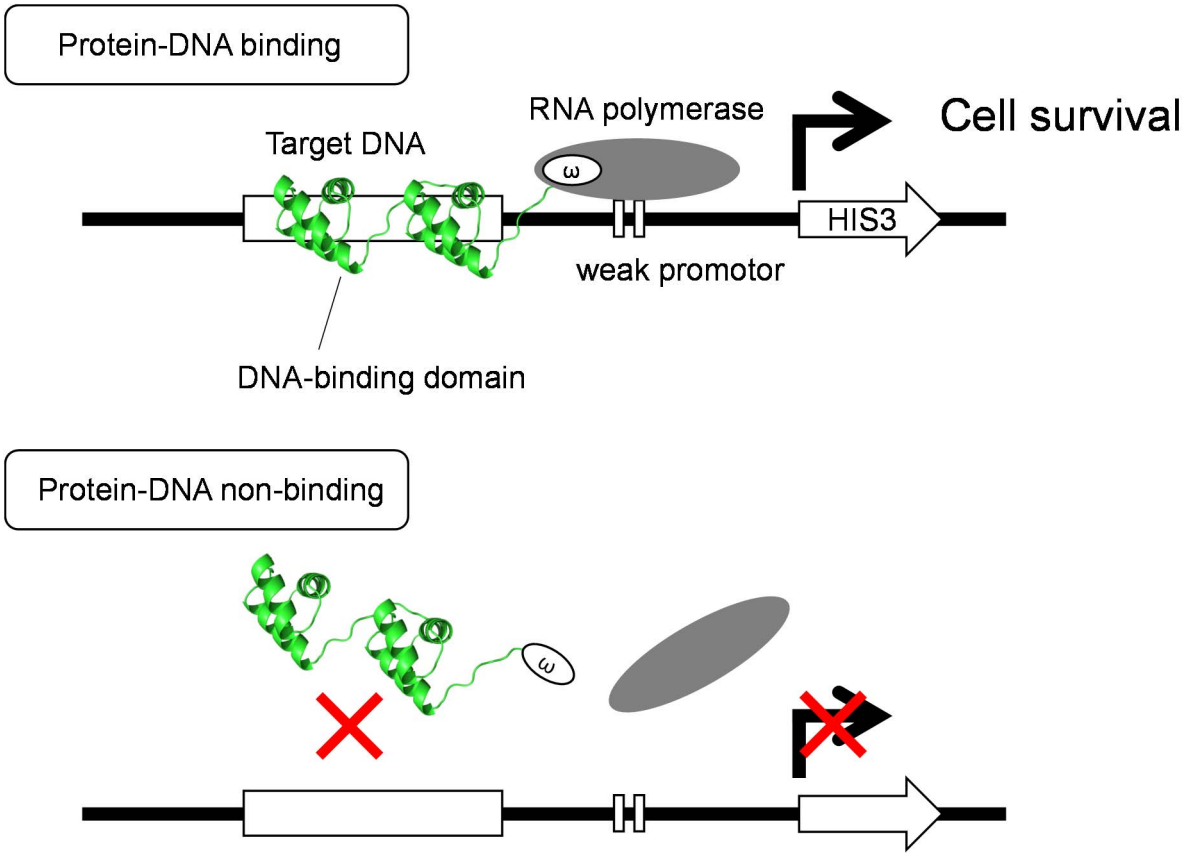

Figure S3 Schematic diagram of the B1H assay. If a protein-DNA complex is formed at the region of the target DNA, transcription of the downstream reporter gene, *HIS3*, is activated. The gene products enable cells to survive, which leads to colony formation after a few days. In our experiment, three residues within DNA-binding domains were randomized, and the target DNA sequence was fixed.

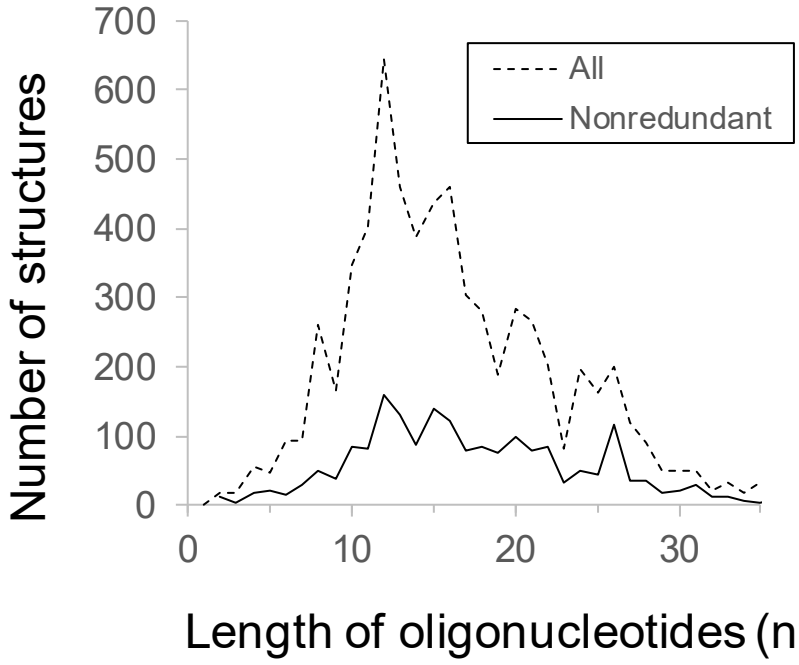

Figure S4 Distribution of the lengths of oligonucleotides in protein-DNA complexes in the PDB. All and Nonredundant in the graph respectively denote the distributions of all protein chains in the PDB and the representative set of protein chains in the PDB whose sequences shared <30% identity. This figure shows that oligonucleotides with a variety of lengths have been used for the crystallization.

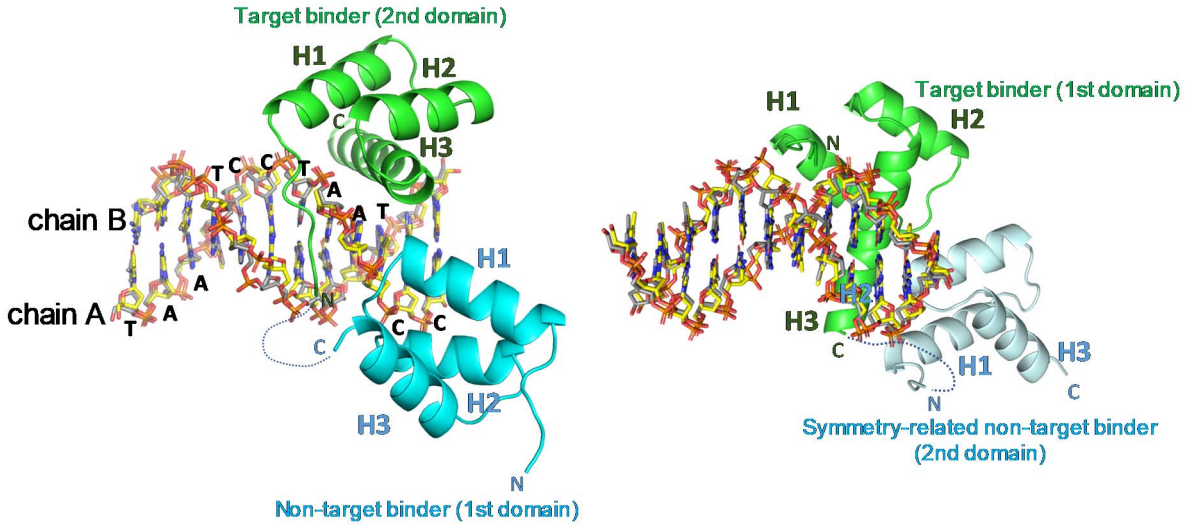

Figure S5 Possible orders of the two EHD domains. The left panel shows that the first and second EHDs are assigned as the non-target binder and target binder, respectively, and *vice versa* in the right panel. In the left panel, their terminal Cα atoms are 17 Å apart. Within the crystal, there is space enough for 11 disordered amino acids in the linker. The right panel shows the other interpretation, where the target binder is the first domain and the non-target binder is the second domain. In this case, the C-terminal Cα atom of the target binder is only 14 Å from the N-terminal Cα atom of the symmetry-related non-target binder. From the electron density map, we determined that neither of these interpretations is more appropriate than the other.

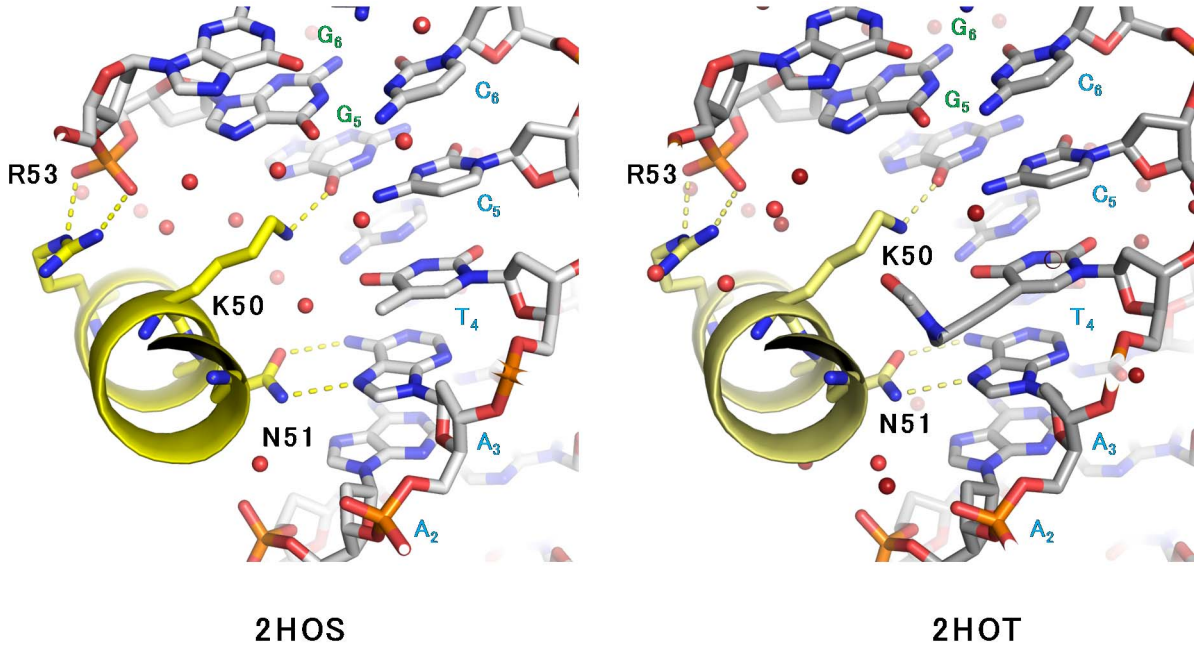

Figure S6 Crystal structures of EHDs without multiple conformations of K50. Shown are close up views of the recognition helix from the crystal structures of EHD-DNA complexes with K50 (PDB code: 2HOS and 2HOT). 2HOS and 2HOT were solved at resolutions of 1.9 Å and 2.19 Å, respectively. The Nζ atom of K50 interacts with O6 of the guanine in the 5th base pair in both structures, as in our structure.

References

Studier, F.W. (2005). *Protein Expr. Purif.* **41**, 207-234.

Sunami, T. & Kono, H. (2019). *Protein Sci.* **28**, 1630–1639.