Supplementary data

Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA

John R. Horton[1,†], Janine G. Borgaro[2,†], Rose M. Griggs[1], Aine Quimby[2], Shengxi Guan[2], Xing Zhang[1], Geoffrey G. Wilson[2], Yu Zheng[2], Zhenyu Zhu[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, Inc., 240 County Road, Ipswich, MA 01938, USA

† Both authors contributed equally to this work.

* Correspondence should be addressed to:

Xiaodong Cheng, Phone: 1-404-727-8491, Fax: 1-404-727-3746, Email: xcheng@emory.edu

Zhenyu Zhu, Phone: 1-978-380-7237, Fax: 1-978-380-7518, Email: zhuz@neb.com

Authors email addresses:

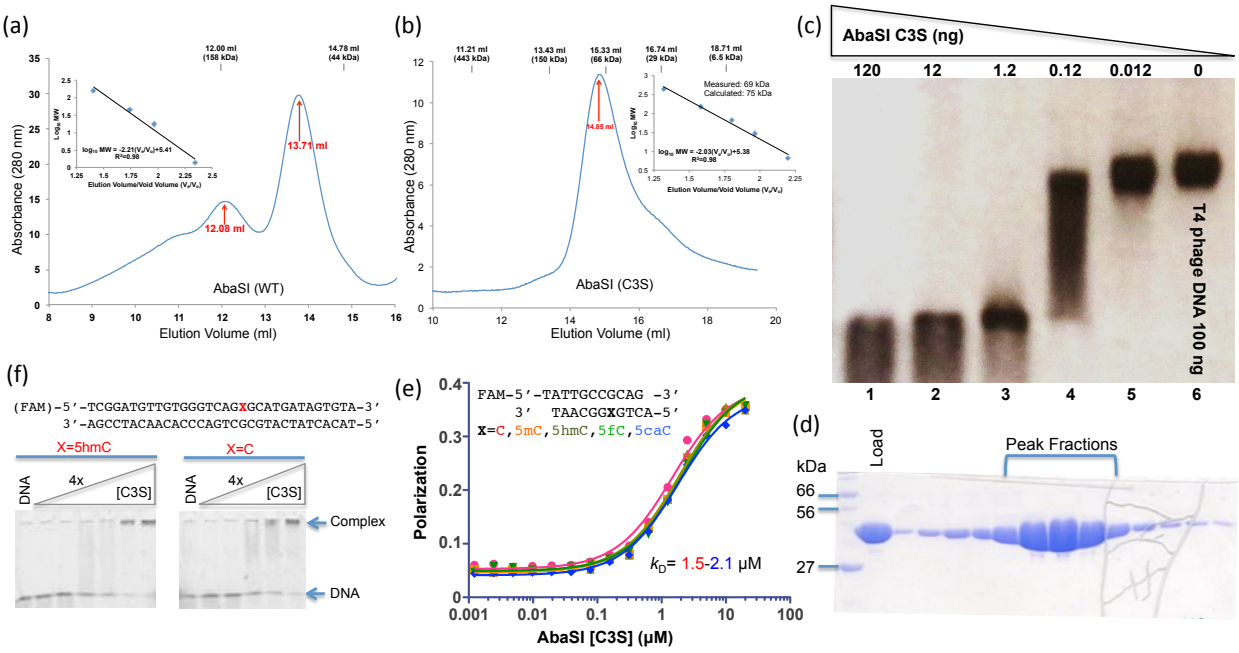
From Emory:

JRH (jrhorto@emory.edu), RMG (rmg02d@gmail.com), XZ (xzhan02@emory.edu), XC (xcheng@emory.edu)

From NEB:

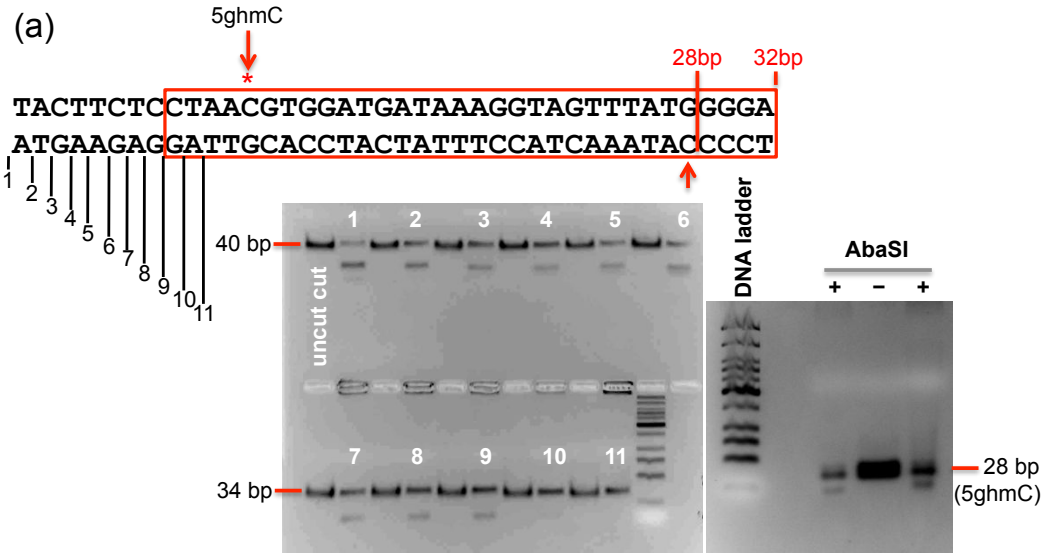
JGB (Borgaro@neb.com), AQ (quimbya@neb.com), SG (guans@neb.com), GGW (wilson@neb.com), YZ (zhengy@neb.com), ZZ (zhuz@neb.com)

Supplementary Figure S1. AbaSI C3S variant



(**a**) Elution profile of AbaSI WT on a Superdex 200 (10/300 GL) column (GE Healthcare). The column buffer was 20 mM tris acetate (pH 8.0) and 150 mM potassium acetate and ~60 µg of AbaSI was loaded onto the column. The inset shows the standardization of the size exclusion column using a protein marker kit (Biorad) at the time AbaSI was profiled using the same buffer. (**b**) Elution profile of AbaSI-C3S on Superdex 200 (10/300 GL) column (GE Healthcare). The column buffer was 20 mM tris acetate (pH 8.0) and 150 mM potassium acetate and ~25 µg of AbaSI was loaded onto the column. The inset shows the standardization of the size exclusion column using a collection of protein markers of various molecular weights (blue dextran for void volume, 443 kDa apoferritin, 150 kDa alcohol dehydrogenase, 66 kDa bovine serum albumin, 29 kDa carbonic anhydrase, and 6.5 kDa aprotinin) at the time AbaSI was profiled using the same buffer. (**c**) Activity of AbaSI-C3S variant on glucosylated T4 phage DNA. We note that 1.2 ng AbaSI 3CS is 0.1 units (lane 3), with 83,000 units/mg, similar to that of WT AbaSI (76,000 units/mg). (**d**) Shown is a sodium dodecyl sulfate polyacrylamide gel electrophoresis (13% SDS-

PAGE) of the purified AbaSI-C3S proteins eluted from a HiLoad 16/60 Superdex 200 column (GE Healthcare). The peak fractions were concentrated and used for crystallization. (**e**) Fluorescence-based DNA binding assay were performed by measuring fluorescence polarization, in 20 mM Tris-Acetate (pH 8.0), and 150 mM potassium acetate, 5 mM calcium acetate and 5 mM DTT at room temperature using a Synergy 4 Microplate Reader (BioTek) at an excitation wavelength of 485 nm and emission wavelength of 528 nm. Fluorescently labeled (6-carboxy-fluorescein or FAM) DNA probe (10 nM) and various amounts of AbaSI-C3S protein with a final volume of 40 μL were incubated at room temperature in a 384-well plate for 0.5 h before measurements were taken. Curves were fitted with Origin version 7.5 (OriginLab). K_D values were calculated as [mP] = [maximal mP] × [C]/(K_D + [C]) + [baseline mP], where [mP] is millipolarization and [C] is protein concentration. (**f**) Gel mobility shift DNA binding assay. Various amounts of AbaSI-C3S protein and 10 nM of 32-bp FAM labeled DNA were mixed in a final volume of 20-μL in 20 mM Tris-Acetate, pH 8.0, 150 mM potassium acetate, 5 mM calcium acetate, 5mM DTT and incubated at room temperature for 45 minutes. Samples were loaded onto a 10% native polyacrylamide gel (10×10 cm^2) in 1×Tris-Borate-EDTA (TBE) buffer and ran at 100 V for 40 min. Typhoon Trio (GE Healthcare) was used to visualize FAM fluorescence.

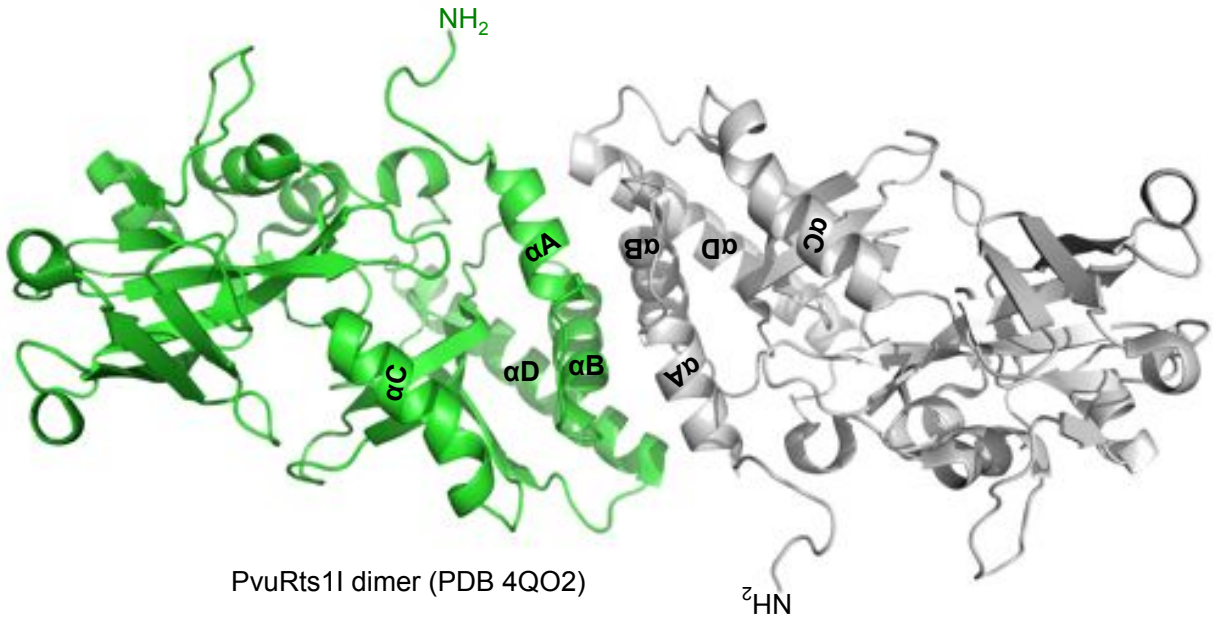


(b)

```
Substrate DNA used for crystallization
bp       1    5    10    15    20    25    30
         |    |    |     |     |     |     |
28    5′-CTAAHGTGGATGATAAAGGTAGTTTATG-3′
      3′-GATTGCACCTACTATTTCCATCAAATAC-5′

29    5′-CTAAHGTGGATGATAAAGGTAGTTTATGT-3′
      3′-GATTGCACCTACTATTTCCATCAAATACA-5′

30    5′-CTAAHGTGGATGATAAAGGTAGTTTATGTT-3′     -self annealing
      3′-GATTGCACCTACTATTTCCATCAAATACAA-5′

32    5′-CTAAHGTGGATGATAATTATCATCCACGTTAG-3′   -self annealing
      3′-GATTGCACCTACTATTAATAGTAGGTGHAATC-5′

30+1  5′-TGCTHGATACAATAGGCCTATTGTATCGAGC -3′   -self annealing
      3′- CGAGCTATGTTATCCGGATAACATAGHTCGT-5′

32    5′-TGCTHGATACAATAGGCCTATTGTATCGAGCA-3′   -self annealing
      3′-ACGAGCTATGTTATCCGGATAACATAGHTCGT-5′

32+1  5′-TGCTHGATACAATAGGCCTATTGTATCGAGCAT-3′  -self annealing
    3′-TACGAGCTATGTTATCCGGATAACATAGHTCGT -5′

Product DNA
14+4  5′-TGCTHGATACAATAGGCC-3′      5′-TGCTHGATACAATAGGCCTATTGTATCGAGCA-3′
      3′-ACGAGCTATGTTAT    -5′  ↔  3′-ACGAGCTATGTTATCCGGATAACATAGHTCGT-5′
```

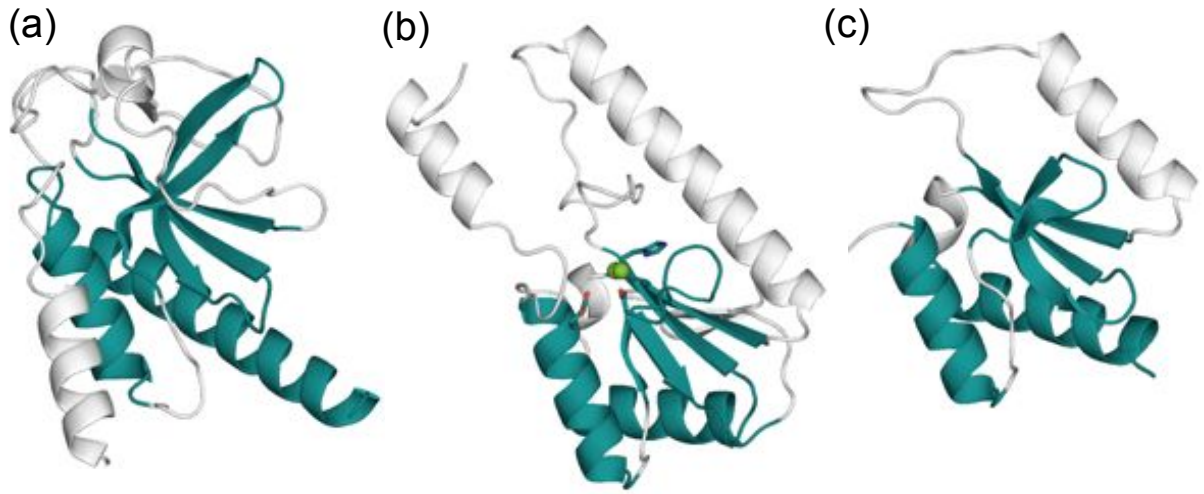
Supplementary Figure S2. Minimum length of DNA substrate for AbaSI
(**a**) We started from a 40-bp DNA duplex and reduced one-bp at a time from the 5' end of the modified cytosine. The AbaSI cleavage activity was significantly reduced between 9-bp and 10-bp deletions. (right panel) AbaSI is active on the 28-bp substrate DNA with a modified cytosine at position 5 from the 5' end and a second cytosine at position 28 of the 5' end of the opposite strand. (**b**) DNA duplexes used in the crystallization. The two product DNA molecules (14 bp plus 4 nt overhang at 3' end) were annealed together by the 4-nt 3' overhangs (probably during co-crystallization) forming a continuous 32-bp DNA with one phosphate backbone break in each strand (Supplementary Fig. 6a).

Supplementary Figure S3. PvuRts1I dimer (PDB 4QO2) mediated by crystallography symmetry



Supplementary Figure S4. AbaSI has an N-terminal Vsr-like DNA cleavage domain

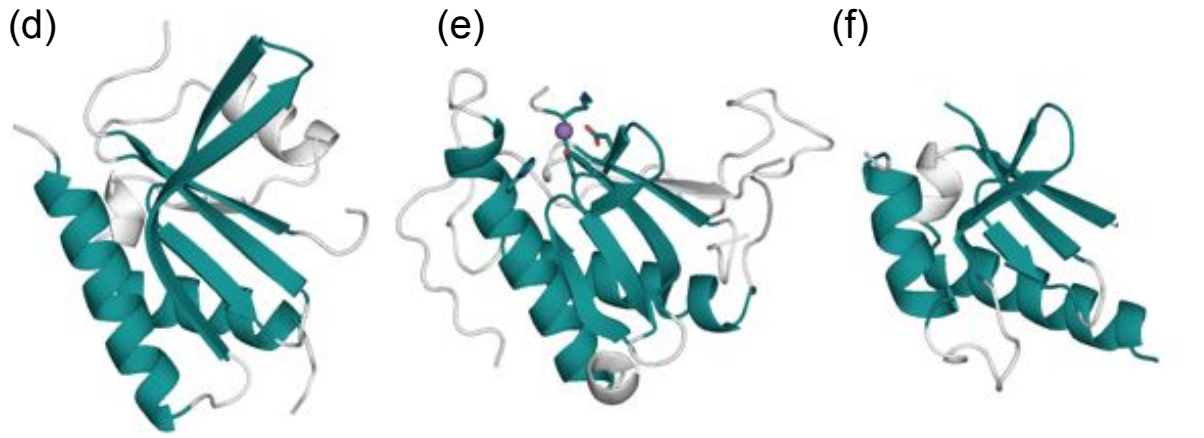
(a) AbaSI, (**b**) Vsr endonuclease, (**c**) Homing endonuclease I-Bth0305I, (**d**) *Vibrio cholerae* 1899 protein, (**e**) Type IIG restriction endonuclease BpuSI, and (**f**) *Neisseria Gonorrhoeae* Endonuclease-like protein.



AbsSI N-terminal Domain

Vsr Endonuclease (PDB 1CW0)
RMSD of 2.9Å across 88 residues

Homing Endonuclease I-Bth0305I (PDB 3R3P)
RMSD of 2.9Å across 86 residues



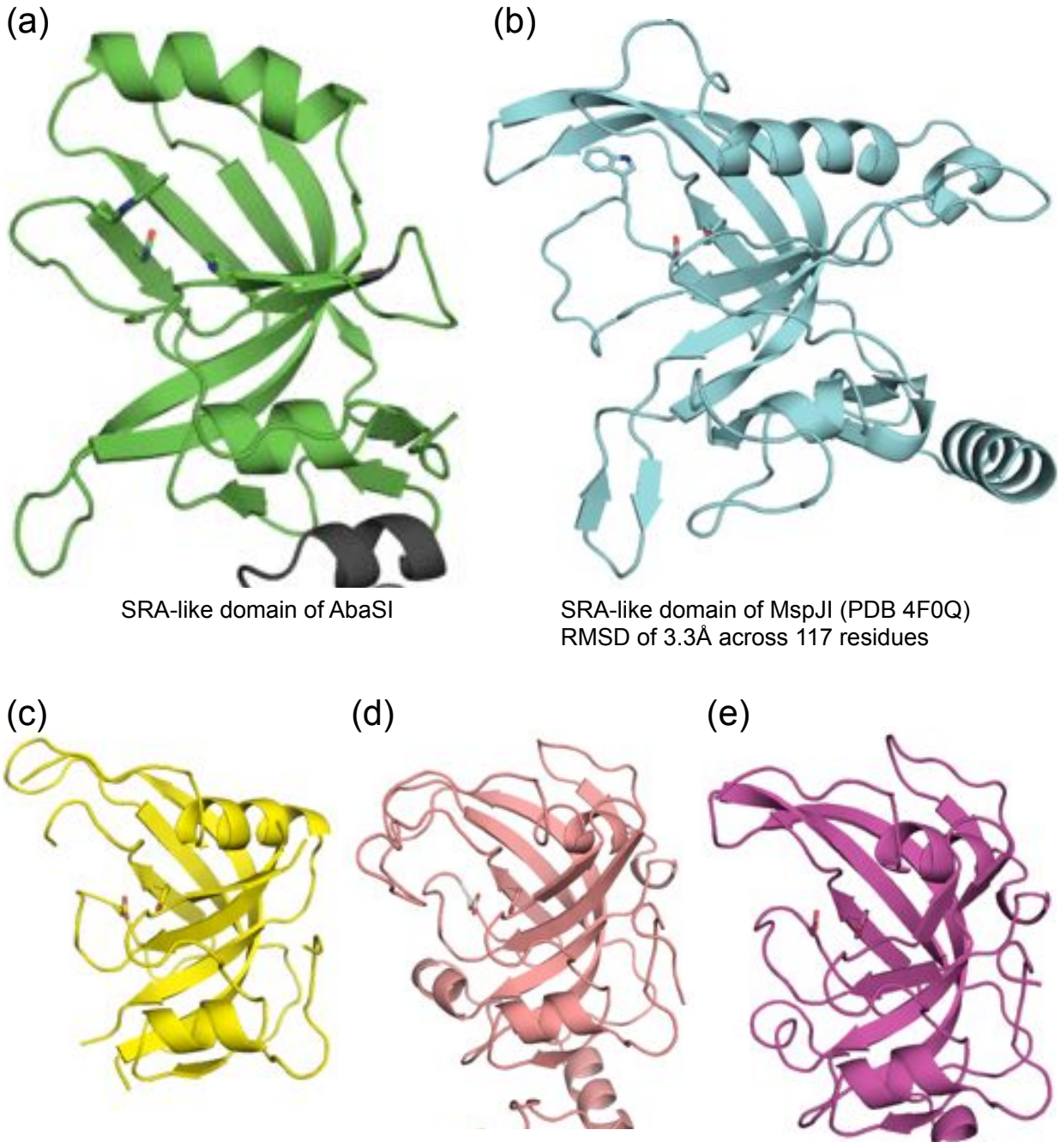
Vibrio cholerae 1899 (PDB 1XMX)
RMSD of 3.3Å across 96 residues

Restriction BpuSI (PDB 3S1S)
RMSD of 4.0Å across 98 residues

Neisseria Gonorrhoeae Endonuclease-like (PDB 3HRL) RMSD of 2.5Å across 86 residues

Supplementary Figure S5. Examples of SRA-like domain

(**a**) AbaSI, (**b**) MspJI, (**c**) Arabidopsis SUVH5, (**d**) mouse UHRF1, and (**e**) human UHRF1.



SRA-like domain of AbaSI

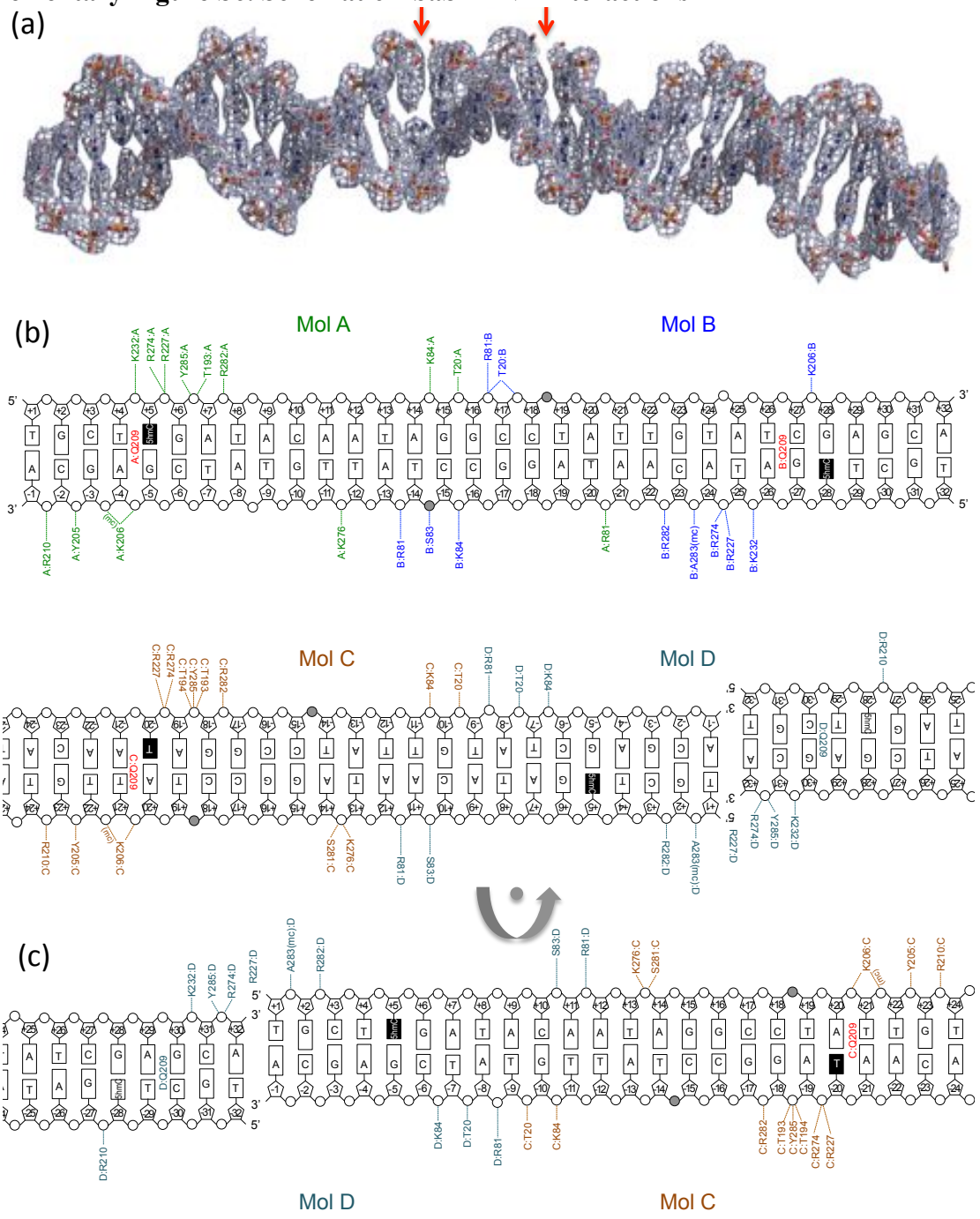
SRA-like domain of MspJI (PDB 4F0Q)
RMSD of 3.3Å across 117 residues

SRA Domain of Arabidopsis SUVH5 (PDB 1Q0D)
RMSD of 3.1Å across 97 residues

SRA Domain of Mouse UHRF1 (PDB 3FDE)
RMSD of 3.4Å across 102 residues

SRA Domain of Human UHRF1 (PDB 3BI7)
RMSD of 3.4Å across 102 residues

Supplementary Figure S6. Schematic AbaSI-DNA interactions



(**a**) Electron density map of 2Fo-Fc, contoured at 1σ above the mean, showing the missing phosphate groups after annealing together two DNA molecules (14 bp plus 4 nt overhang at 3' end) by the 4-nt 3' overhangs, forming a continuous 32-bp DNA with one phosphate backbone break in each strand. (**b**) Schematic AbaSI-DNA interactions via the A-B dimer (top panel) and the C-D dimer (bottom panel). The bottom panel is rotated 180° from that of panel c. (**c**) Schematic AbaSI-DNA interactions via the C-D dimer. The DNA molecules are shown in the same orientation as that of Fig. 3h.