Supplementary data to

Restriction endonuclease AgeI is a monomer which dimerizes to cleave DNA

Giedre Tamulaitiene*, Virginija Jovaisaite[¥], Gintautas Tamulaitis, Inga Songailiene, Elena Manakova, Mindaugas Zaremba, Saulius Grazulis, Shuang-yong Xu[#] and Virginijus Siksnys*

Institute of Biotechnology, Vilnius University, Sauletekio al. 7, LT-10257, Vilnius, Lithuania
[¥] present address: Laboratory for Integrative and Systems Physiology, École Polytechnique Fédérale de Lausanne, EPFL, CH-1015 Lausanne, Switzerland
[#] New England Biolabs, Inc. 240 County Road, Ipswich, MA 01938, USA

* To whom correspondence should be addressed. Tel: +370-5-2234357; Fax: +370-5-2234367; Email: eigie@ibt.lt
Correspondence may also be addressed to Virginijus Siksnys. Tel. +370-5-2234359; Fax: +370-5-2234367; Email: siksnys@ibt.lt

Supplementary Table S1. Oligoduplexes used in this study

Substrate	Used for	Sequence
SP11 (specific)	crystallization	5'-TCG**ACCGGT**CG -3' 3'- GC**TGGCCA**GCT-5'
SP13 (specific)	crystallization	5'-TTCG**ACCGGT**CGA -3' 3'- AGC**TGGCCA**GCTT-5'
SP (specific)	DNA binding, hydrolysis and gelfiltration	5'-AGACCCACGCTC**ACCGGT**TCCAGATTTATC-3' 3'-TCTGGGTGCGAG**TGGCCA**AGGTCTAAATAG-5'
NC (non-canonical)	DNA binding	5'-AGACCCACGCTC**ACCGG**ATCCAGATTTATC-3' 3'-TCTGGGTGCGAG**TGGCC**TAGGTCTAAATAG-5'
NSP (non-specific)	DNA binding	5'-CAGCACAGTTCAGCAGCCCAGTGCTACGCT-3' 3'-GTCGTGTCAAGTCGTCGGGTCACGATGCGA-5'
Mutant		
Q86A	mutagenesis	5'-CGACCCAAGCGCTAGCTGGGTTCGCACGTCGGGCG-3' 5'-GCGAACCCAGCTAGCGCTTGGGTCGGTTCCGATCTTC-3'
S138A	mutagenesis	5'-CGCGTCGGCGCCCGCAAGGTTGACGTGATGATCG-3' 5'-CAACCTTGCGGGCGCCGACGCGGTCGGCCACAC-3'
D142A	mutagenesis	5'-CTCGCAAGGTTGCCGTCATGATCGAGAAGCAGGGAGGCGG-3' 5'-CTTCTCGATCATGACGGCAACCTTGCGAGAGCCGACGCGG-3'
D177A	mutagenesis	5'-GGGTCTCGGCCGACATACCCGCCAGCAGGATC-3' 5'-CGGGTATGTCGGCCGAGACCCTCTCGGCTAG-3'
D178A	mutagenesis	5'-GTCTCGGACGCCATACCCGCCAGCAGGATCATG-3' 5'-CTGGCGGGTATGGCGTCCGAGACCCTTTCGGCTAGGCTC-3'
D223A	mutagenesis	5'-CCGGCCCTCGGCGAAGAGGAATTACATTGAGGGACAC-3' 5'-GTAATTCCTCTTCGCCGAGGGCCGGTCGGGCGTGCC-3'

AgeI recognition sequence (5'-ACCGGT-3') is underlined. In NC only nucleotides, which coincide with the recognition sequence, are underlined.

Supplementary Table S2. Crystallization and cryo-protection conditions

Crystal form	Space group	Crystallization conditions	Cryo-protection conditions
AgeI+SP11	Form I $P6_122$	Na Malonate7.0 0.2M, PEG3500 19% protein concentration 8.3 mg/ml, AgeI:DNA ratio 1:1.2	Mother liquor supplemented with 20% PEG400
AgeI+SP11	Form II $P2_12_12_1$	0.2 M NaCl, 0.1 M BisTRIS 5.5, 17-19% PEG3500 (protein concentration 8.3 mg/ml, AgeI:DNA ratio 1:1.2)	stepwise transfer into mother liquor containing 2-5-10-15% ethylene glycol
AgeI+SP13	Form III $P2_1$	0.2M NaCl, 0.1M BisTRIS 5.5, 17-19% PEG3500 protein concentration 7.2 mg/ml, AgeI:DNA ratio 1:1.2	stepwise transfer into mother liquor containing 2-5-10-15% ethylene glycol
AgeI	$P2_12_12_1$	0.1 M Na-Hepes (pH 7.5), 0.2 M NaCl, 25% PEG3350, 3% ethylene glycol (protein concentration 5.7 mg/ml)	stepwise transfer into mother liquor containing 5-10-15% ethylene glycol

Supplementary Table S3. SAXS data collection parameters and shape models description.

Data collection parameters	
X-ray wave length, nm	0.124
Sample to detector distance, m	3.1
Detector	Pilatus 2M
s range, nm^{-1}	0.0290642-4.488440
Exposure time, s	0.05
Collected frames	20
Sample storage temperature, °C	10
Data collection temperature, °C	20

Structural parameters of models		
Parameter	**AgeI**	**AgeI+DNA**
Protein concentration, mg/ml	1.3 and 3.1	2.8
s range of P(r) calculated by GNOM, $Å^{-1}$	0.0229 – 0.2986	0.0111 – 0.3435
Real space Rg calculated by GNOM, Å	21.29 ± 0.050	24.00 ± 0.022
D_{max} used by GNOM, Å	67	74
Porod volume calculated by GNOM, $Å^3$	65894	53576
Averaged volume of DAMMIN models, $Å^3$	56078 ± 1247	73930 ± 2012

Supplementary Table S4. Comparison DNA recognition and active site residues of SP- and preSP-complexes

	SP-complex **residue subunits A/B**	**preSP-complex** **residue subunits A/B**
DNA	13 bp	11 bp
Dimer *buried surface area*, Å^2	630	400
DNA binding *buried surface area*, Å^2	5823	5215
Major groove contacts	E173 +/+ R174 +/+ K200 +/+ E214 +/+ K224 +/+	**E173 –/–** **R174 +/–** K200 +/+ E214 +/+ **K224 +/–**
Minor groove contacts	Q86 +/+ R90 +/+	Q86 +/+ R90 +/+
Active site	E97 +/+ D142 +/+ K168 +/+ D178 +/+	E97 +/+ D142 +/+ **K168 +/–** **D178 –/–**

+/+ Identical contacts to DNA are made in both subunits of both complexes or the conformation of residues in both subunits are similar to the SP-complex

+/– Only one subunit of a dimer makes contacts similar to the SP-complex

–/– None of the subunits of preSP-complex makes contacts similar to the SP-complex

Residues, making different contacts to DNA in the SP-complex and at least in one subunit of the preSP-complex are shown in boldface

Supplementary Table S5. AgeI contacts with phosphates in SP- and preSP-complexes

Phosphate # (base of AgeI recognition site)	Interacting residues from SP-complex*	Interacting residues from preSP-complex		
		both subunits	only A	only B
-2	–			
-1	–			
0	K224 D223(H_2O)			
1 (A)	K224(H_2O)			R139
2 (C)		K140		
3 (C)	V169 R174		V169(H_2O)	V169
4 (G)	H68 S170 W88	H68 S170 W88		
5 (G)	Y72 R73(H_2O) S201(H_2O)	Y72 R73(H_2O) S201(H_2O)		
6 (T)	D83 T82	D83 T82	H206	
7	T82 S87 K79(H_2O)	T82 S87 K79(H_2O)		
8	S43(H_2O) R44(H_2O)	S43(H_2O) R44(H_2O)		
9	R44	R44**		
	K42			

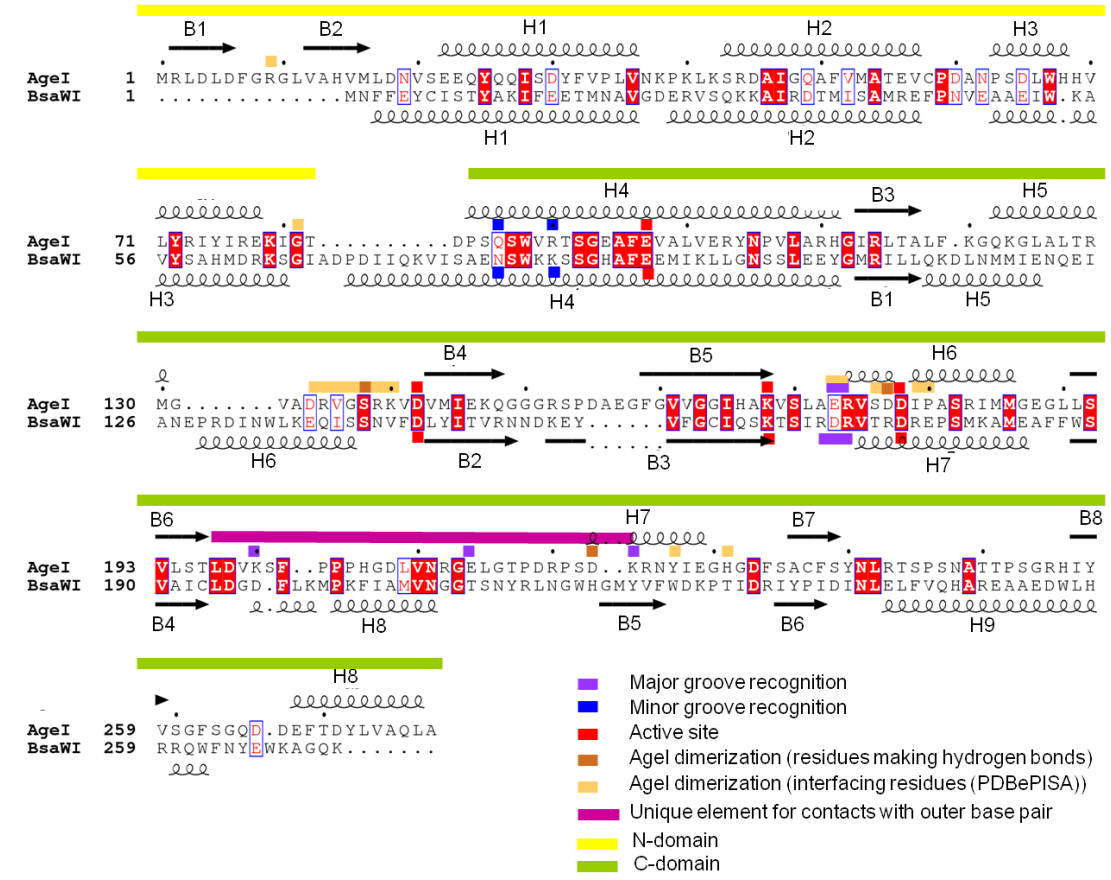
*(H_2O) – the residue makes water-mediated hydrogen bond. Similar contacts are shaded in grey.
**R44 makes hydrogen bond to O3* atom of G8.

Supplementary Table S6. Gel filtration of AgeI.

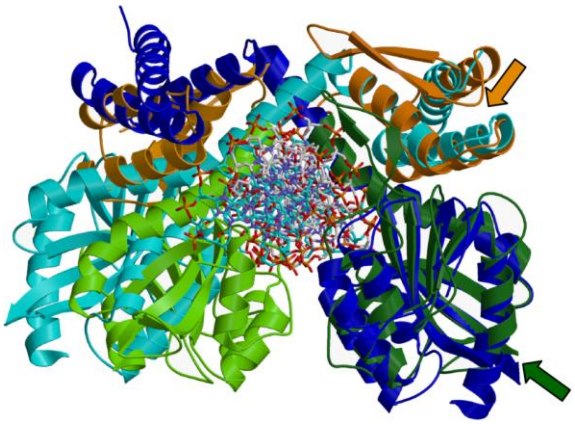
Apo AgeI concentration, μM	Estimated M_w, kDa
35	24.9
20	28.8
10	24.9
7	26.3
5	26.0
AgeI:DNA ratio	**Estimated M_w, kDa**
1:0.5	65.3
1:1	64.3
1:5	67.1
1:10	66.3
1:20	67.2
DNA	**Estimated M_w, kDa**
SP	44

Calculated theoretical M_w for AgeI monomer 32.5 kDa, dimer 65 kDa, DNA 18.4 kDa.

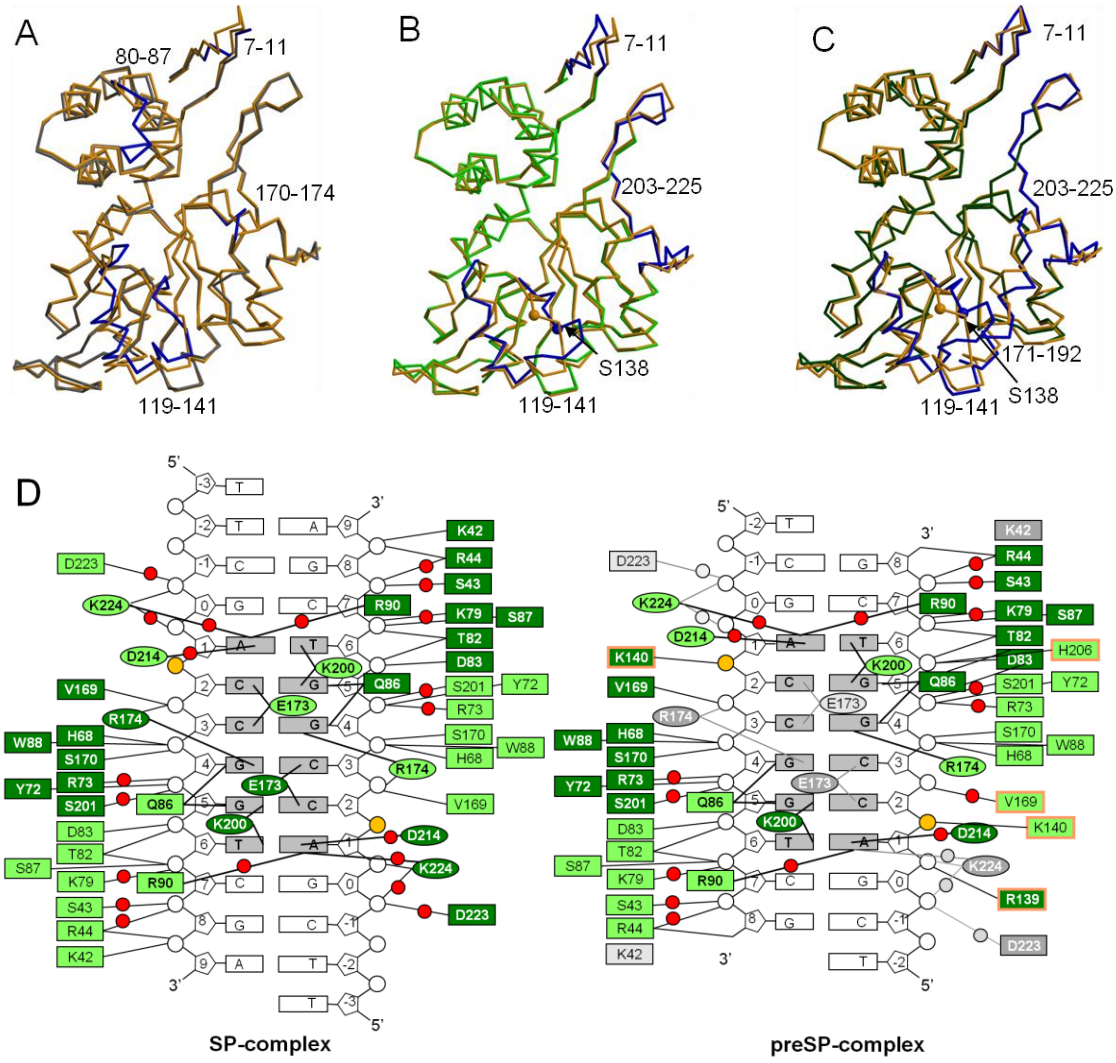
A



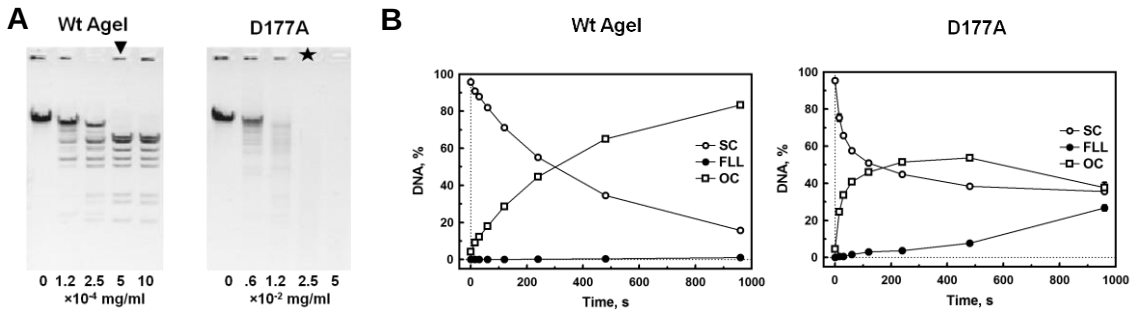
B



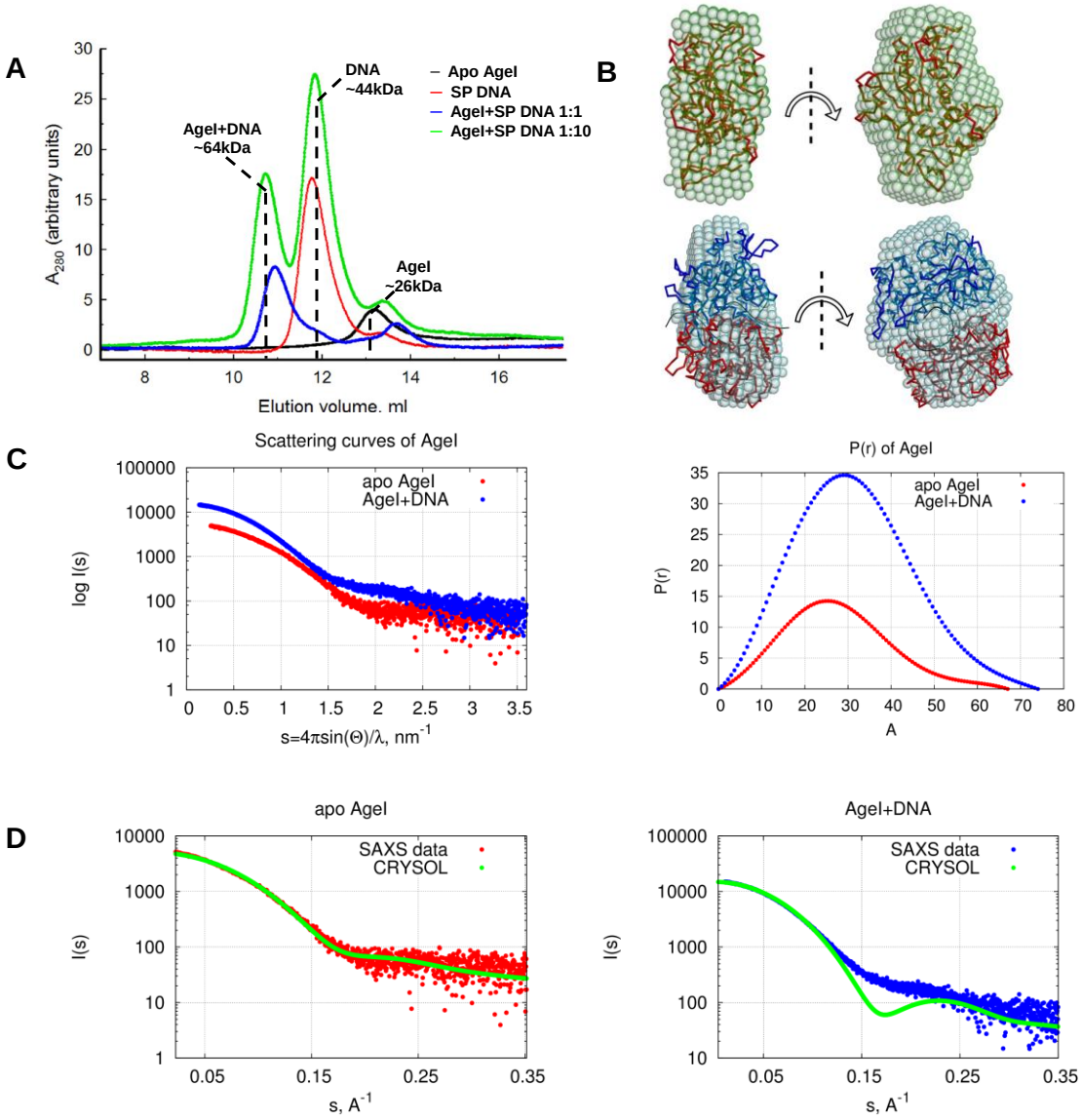
Supplementary Figure S1. Similarity between AgeI and BsaWI. (A) Pairwise alignment of AgeI and BsaWI sequences. Secondary structure elements of AgeI and BsaWI are shown above and below the alignment, respectively. DNA recognition, active site and AgeI dimerization residues are marked. The figure was prepared with ESPript (http://espript.ibcp.fr). (B) Structural overlay of AgeI and BsaWI dimers. BsaWI subunits are coloured blue, N-domains of AgeI are orange, C-domains are green. Orange and green arrows show overlay of the same AgeI subunit with the domains of BsaWI from different subunits.



Supplementary Figure S2. Comparison of AgeI structures. (A-C) Conformational differences of AgeI in different structures. (A) Overlay of apo-AgeI (grey) and A subunit of SP-complex (orange). Parts of the apo-subunit which do not overlay with the subunit of SP-complex are blue. Numbers indicate non-overlapping residues. (B-C) Overlay of A subunit of SP-complex (orange) and A subunit of preSP-complex (green, B) or B subunit of preSP-complex (dark green, C). Parts of the preSP-complex subunit which do not overlay with the subunit of SP-complex are blue. Cα atoms of S138 from the SP-complex subunit (orange) and the preSP-complex subunit (blue) are shown. (D) Comparison of AgeI contacts with DNA in the SP- and preSP-complexes. Schematic representation of the AgeI-DNA contacts: residues from the subunit A are coloured light green, from the subunit B - dark green; residues making contacts in the major groove are shown in ellipses; red spheres represent water molecules. Scissile phosphates are coloured orange. Missing contacts in the preSP-complex are shaded in grey, new contacts are boxed in orange.



Supplementary Figure S3. Relaxed specificity of the D177A mutant analyzed using (A) phage λ DNA (13 AgeI sites) and (B) plasmid pUC18 DNA (0 AgeI site). The DNA cleavage reactions were performed and analyzed as described in 'Materials and Methods'. In (A) the final protein concentrations are indicated below the gels; a downward-facing triangle indicates the minimal enzyme concentration needed for complete digestion at the cognate recognition sites; a star indicates the minimal enzyme concentration needed for complete DNA digestion. In (B) the rate constant values ($2.6\pm0.1\times10^{-3}$ s^{-1} for Wt AgeI and $2.0\pm0.1\times10^{-2}$ s^{-1} for D177A) were obtained by fitting a single exponential to the time-courses of supercoiled form depletion.

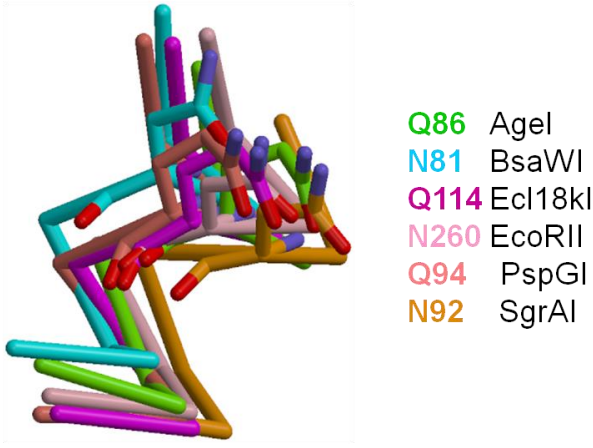


Supplementary Figure S4. AgeI structure in solution. (A) Gel filtration analysis of AgeI. Apo AgeI (black line) elutes from the column as a 26 kDa protein with a molecular weight close to the calculated Mw of the monomer (calculated Mw of apo AgeI 32kDa). SP DNA (red) elutes from the column at a volume corresponding to 44 kDa possibly due to the elongated form of the molecule (calculated Mw is 18.4 kDa). AgeI-DNA complex elutes from the column as a 64 kDa species (blue and green). Experimentally determined Mw's are shown on the graph. Gel filtration was carried out as described in 'Materials and Methods'. (B) Averaged DAMMIN models of apo-AgeI (transparent green spheres) and AgeI-SP13 complex (light blue transparent spheres) complex were superimposed with the apo (monomer) and SP-complex (dimer) crystal structures, respectively (shown as backbone with different monomer colours). (C). SAXS data of apo-AgeI (red curves) and AgeI-SP13 complex (blue curves). (D). Theoretical scattering curves (green) calculated from crystal structures of apo-AgeI (left panel) and SP-complex (right panel) by CRYSOL (35) compared with the experimental scattering curves.

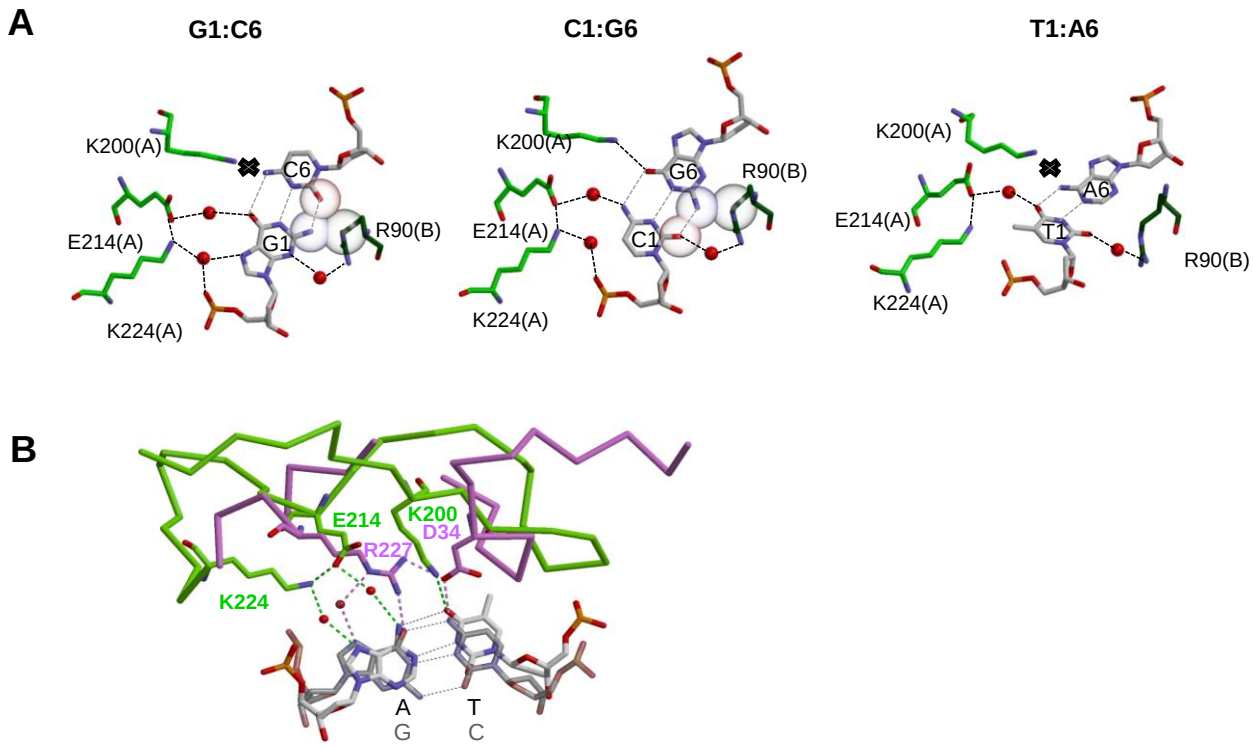
```
                N/Q  ▼      *
Bse634I    68   SGAFSNTRGTWFEVMIAIQSWNYRVKRELN  97
Cfr10I     58   DGSFNKCNGDWYEWLIGIRAIEFFLESETN  87
NgoMIV     58   ERLPGQTSGNAFEAICSEFVQSAFEKLQHI  87
SgrAI      91   SNAKAKVAGDIFEIVSSAVMWNCAARWNSL 120
BsaWI      80   ENSWKKSSGHAFEEMIKLLGNSSLEEYGMR 109
AgeI       85   SQSWVRTSGEAFEVALVERYNPVLARHGIR 114
Ec118kI   113   TQSRRSRAGKEFESILELLMMGAGIPVDVQ 142
PspGI      93   QQARFSRGGKAFEIIFTKLLNKFGIRYEHD 122
EcoRII    259   SNRRKSRAGKSLELHLEHLFIEHGLRHFAT 288

                *                                                    *   D
                                                                         R-ER   *  *
Bse634I   145   PDLLIIRQKDLIKSEYNLPINKLTHENIDVALTLFKDIE.............GKCKWDSLVAGVGLKTSLRPDRRLQLVHEGN 214
Cfr10I    133   PDFSIIDIRGRREELKSMLKDISFSNISLSTISEIDNLY.........KNFIDYAELEHIKSFLSVKTTFRPDRRLQLAHEGS 206
NgoMIV    139   PDIIVTRNLIADAEINRNEFLVDENIATYASLRA...................GNGNMPLLHASISCKWTIRSDRAQNARSEGL 209
SgrAI     187   KDGLGLPTSTPDLAVVVLPEEFQNDEMWREEIAGLTRPNQILLSGAYQRLQGRVQPGEISLAVAFKRSLRSDRLYQPLYEAN 268
BsaWI     144   FDLYITVRNNDKEY.........................................VFGCIQSKTSIR-DRVTRDREPSM 181
AgeI      141   VDVMIEKQGGGRSPDAEGFG.....................................VVGGIHAKVSLA-ERVSDDIPASR 184
Ec118kI   159   VDLVMPGVVQYTSNKR.........................................NTMLISAKTTLR-ERWQEVPEEVN 198
PspGI     137   PDFIIPSVRAFLNDPS.........................................SAILITVKRKVR-ERWREAVGEAQ 176
EcoRII    298   PDFLFPSAGAYHDTEFPVE.....................................NLRMLAVKTTCK-DRWRQILNEAD 340
```

Supplementary Figure S5. Conserved sequence motifs of the CCGG-family restriction endonucleases containing residues belonging to the active site (coloured red marked by red stars), CCGG recognition (blue) and minor groove contact with the outer base pair (coloured magenta and marked by magenta triangle).



Supplementary Figure S6. Conserved N/Q residues in the superimposed AgeI (5DWB), BsaWI (4FSZ), Ecl18kI (2FQZ), EcoRII-C (3HQG), PspGI (3BM3) and SgrAI (3DW9) structures.



Supplementary Figure S7. Outer bp recognition by AgeI. (A) *In silico* models of AgeI interaction with the non-cognate C1:G6, G1:C6 and T1:A6 base pairs. X represents no possibility to form hydrogen bonds between two donors. (B) Outer bp recognition elements of AgeI (green) and NgoMIV (violet) in the overlaid structures. DNA from NgoMIV complex is transparent.