SUPPLEMENTARY METHODS

Preparation of oligonucleotide DNA substrates

The oligomeric DNA containing 6 nt bubble with $(A/A)_6$ as bubble sequence (Fig. 1A) was created by annealing a [γ-^{32}P] ATP end-labeled 36 nt oligomer AKN 45 with unlabeled complementary oligomer AKN 18 (1:2 ratio). Similarly, substrates with $(C/C)_6$, $(T/T)_6$ or $(G/G)_6$ as bubble sequences were prepared by mixing [γ-^{32}P] ATP end-labeled AKN 46, AKN 11 and AKN47, respectively with unlabeled complementary oligomers (AKN 20, SCR252 and AKN 48, respectively). 12-RSS substrate was prepared by annealing AKN1/2. To generate substrates with longer flanking sequences (25 bp each) (Suppl. Fig. 3) having $(A/A)_6$, $(C/C)_6$, $(T/T)_6$ or $(G/G)_6$ as bubble sequence, AKN67/68, AKN69/70, AKN71/72 and AKN73/74 were paired. Substrates having 6 nt gaps with AAAAAA, CCCCCC, TTTTTT, GGGGGG were generated by annealing AKN18/7/49, AKN 20/7/49, SCR 252/AKN 7/49 and AKN48/7/49, respectively. Similarly AKN45/8, AKN46/8, AKN11/8 and AKN47/8 were annealed to generate substrates with 3' overhangs. AKN87, AKN88, AKN89 and AKN90 were self annealed to form hairpin loop substrates. Substrate with $(C/G)_6$ duplex was generated by pairing. Substrates with C/C, CC/CC, CCC/CCC, CCCC/CCCC, CCCCC/CCCCC bubbles were generated using, AKN85/20, AKN84/20, AKN83/20, AKN82/20 and AKN81/20, respectively. Substrates containing CCCCCG/TTTTTT, CCCCGG/TTTTTT, CCCGGG/TTTTTT, CCGGGG/TTTTTT, CGGGGG/TTTTTT and GGGGG/TTTTTT as bubble sequences were generated by annealing AKN81/SCR252, AKN82/SCR252, AKN83/SCR252, AKN84/SCR252, AKN85/SCR252 and AKN47/ SCR252, respectively. CCGTTG/TTTTTT bubble with 'C' at 5' of CCGTTG was prepared by annealing SCR251/252. Similarly CCGTTG/TTTTTT bubbles with either 'T', 'A', 'G' or 'U' at 5' of CCGTTG (Fig.6A) were prepared by annealing AKN52/53, AKN55/56, AKN57/58 and AKN59/AKN53, respectively. The markers for double-stranded break experiments were prepared by annealing AKN 5/6, AKN 5/8, AKN 7/6, AKN 7/8 and AKN 7/10. The substrates containing 1 nucleotide C/C or C/A mismatch or C/G duplex was prepared by annealing AKN 96/97 and AKN 96/100 and AKN96/102 respectively. The substrates containing 1 nucleotide C/C or C/A mismatch with longer flanking duplex was prepared by annealing AKN 103/104 and AKN 103/105.

REFERENCE

1. Tsai, A. G., H. Lu, S. C. Raghavan, M. Muschen, C. L. Hsieh, and M. R. Lieber. 2008. Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell 135:1130-42.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Western blot and SDS- Polyacrylamide gel profile showing purified RAG proteins. A. Western blot showing purified GST core RAG1 and core RAG2 using anti-GST antibody. Two different elutions of protein preparations are shown as elution1 and elution 2. **B.** SDS PAGE showing MBP tagged RAG1 and RAG2 proteins. Lanes 1-3 show different combinations of copurified core RAG1/core RAG2, core RAG1/full length RAG2, and full length RAG1/core RAG2. M is molecular weight ladder.

Supplementary Figure 2. Nuclease activity observed is associated with RAG1 and RAG2 proteins. Copurified MBP core RAG1/core RAG2 proteins were fractionated on a Biogel P-100 column and fractions were resolved on an 8% PAGE. Fractions of interest were then used for

RAG cleavage assay. **A.** Silver stained polyacrylamide gel showing selected fractions containing cRAG1/cRAG2 proteins. Fraction numbers are indicated on top of the gel. **B.** Polyacrylamide gel profile showing RAG cleavage on top strands of bubbles with poly C. RAG specific cleavage products are indicated by arrow. **C.** PAGE profile showing RAG cleavage on 12-RSS using different fractions. In both panels B and C, M is 1 nt ladder, while M' shows 15 and 21 nt positions. For details refer Figure 1 legend.

Supplementary Figure 3. Comparison of RAG binding efficiency on bubble substrates containing $(A/A)_6$, $(C/C)_6$, $(T/T)_6$ or $(G/G)_6$. Sequences of bubble regions are indicated above respective lanes. The complete DNA sequence of the heteroduplex structures are shown in Figure 1A. RAG bound substrates and free substrates are indicated by arrows. In the case of $(G/G)_6$ substrates a smear was seen at high molecular weight region, irrespective of addition of RAGs.

Supplementary Figure 4. Comparison of RAG cleavage efficiency on bubbles containing poly A, C, T or G's with longer double-stranded arms. A. Diagrammatic representation of oligomeric substrates containing 6 nt bubbles with sequences $(A/A)_6$, $(C/C)_6$, $(T/T)_6$ or $(G/G)_6$ (denoted as I, II, III, and IV) and 25 bp double-stranded arms. **B.** Polyacrylamide gel profile showing RAG cleavage on top and bottom strands of bubbles with poly A, C, T or G. RAG specific cleavage products are indicated by arrow. For details see Figure 1 legend. **C.** Histogram showing quantification of RAG cleavage products shown in panel "B". **D.** Detection of RAG induced double-strand breaks on substrates used in panel A. The bands due to DSB products are indicated by arrow.

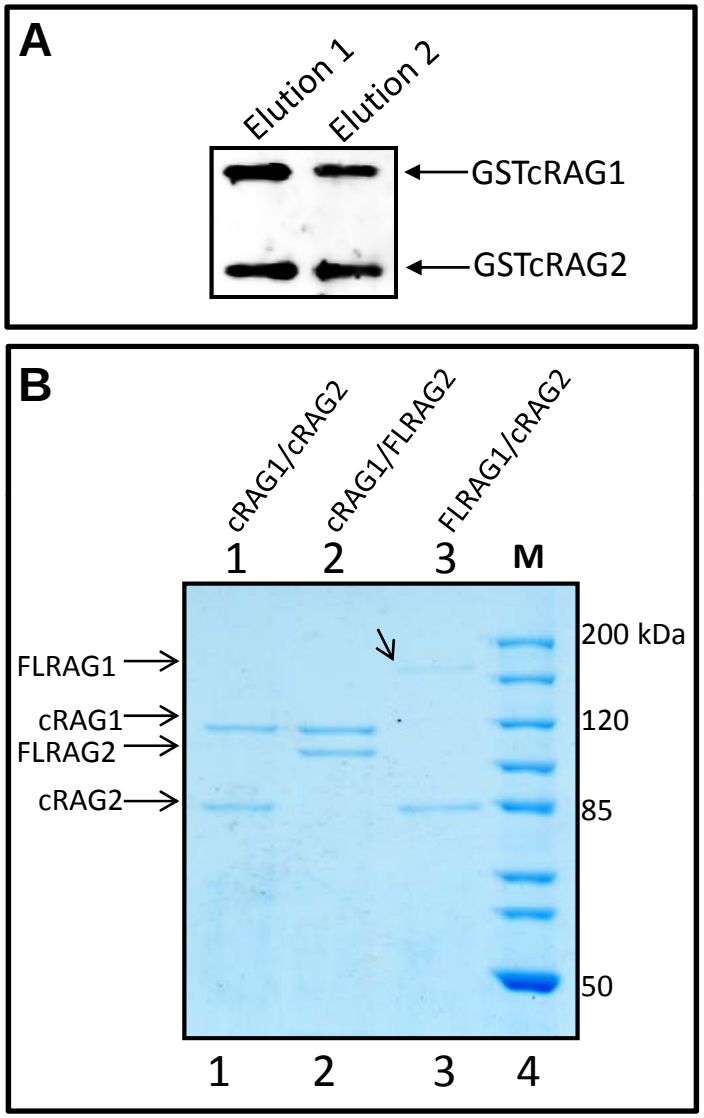
Supplementary Figure 5. P1 nuclease assay on heteroduplex DNA structures containing different number of cytosines. A. Diagrammatic representation of oligomeric substrates containing either 6 nt bubbles with varying number of cytosines (denoted as I-VII) or with varying lengths of bubble region, 0 to 6 (denoted as VIII-XIV). In each case top strand oligomer was radiolabeled with [γ^{32}P] ATP and used. For other details refer Figure 5 legend. **B.** P1 nuclease cleavage on substrates I-VII. **C.** P1 nuclease cleavage on substrates VIII-XIV. In both panels B and C, the radiolabelled substrates were incubated in presence of increasing amount of P1 nuclease (0.0, 0.001, 0.01 and 0.1) for 30 min at 37^{o}C. The products were resolved on 15% gel and presented. M is a 1 nt molecular weight ladder.

Supplementary Figure 6. RAG cleavage on heteroduplex DNA containing 1 nt bubble region. The 1 nt bubble substrate used in the study was synthesized based on our earlier study (1). **A.** The sequence of the bubble region, C/A and its RAG cleavage products are shown (lanes 3, 4). Substrate with C/G duplex at the same position was also used as a control (lanes 1, 2). **B.** RAG cleavage on 1 nt mismatch of C/C. RAG cleavage on a standard RSS was used as a positive control (lanes 1,2). RAG cleavage on top (lanes 3,4) and bottom (lanes 5,6) strands of C/C mismatch is shown. In both panels, the length of the double-stranded arm of the substrate was 15 bp each. **C.** RAG cleavage on 70 mer DNA substrates containing 1 nucleotide mismatches. The 5' labeled oligomeric substrate containing 1 nucleotide mismatch (C/A or C/C) was used for RAG cleavage. M is 1 nt ladder.

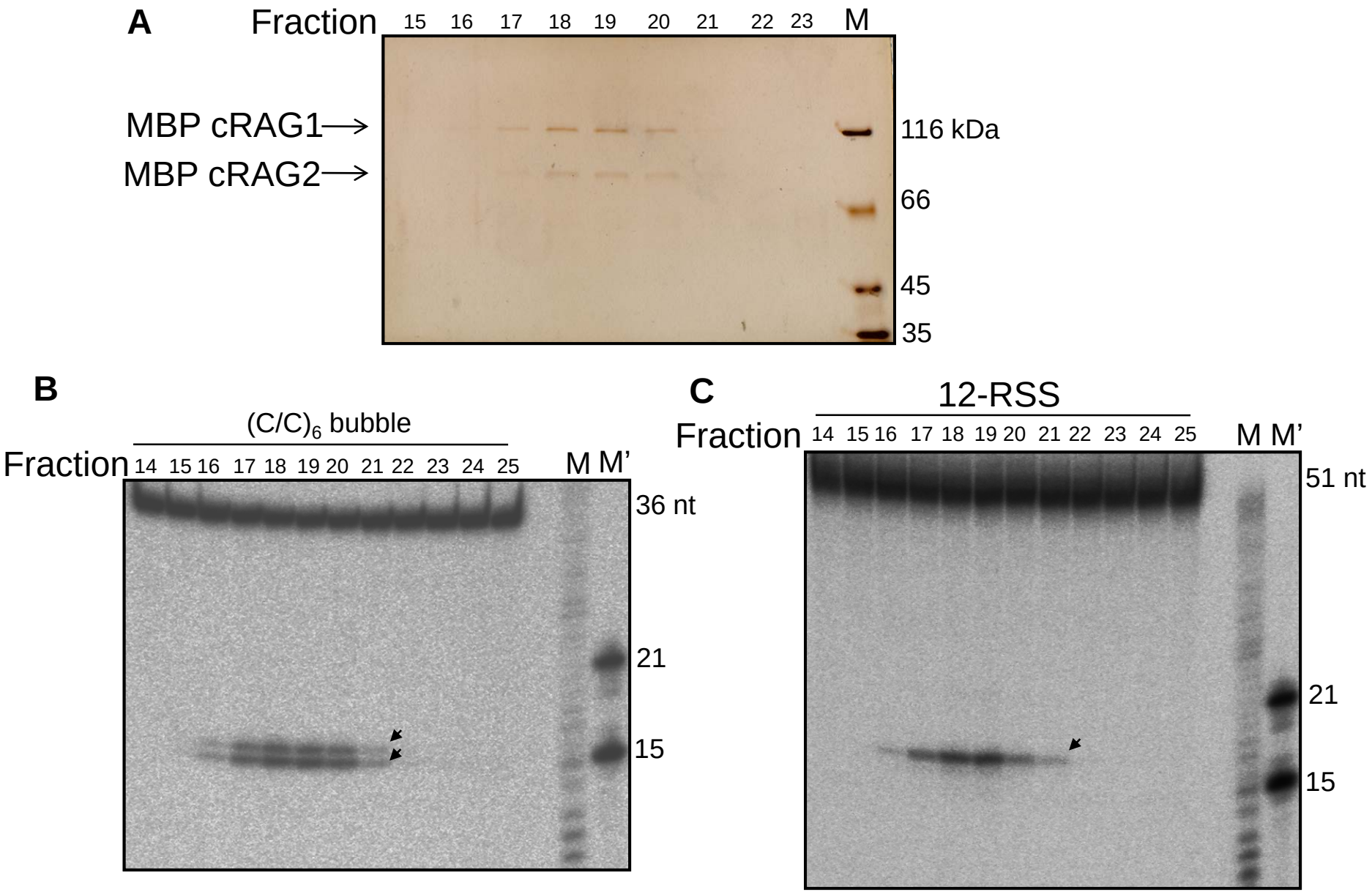
Supplementary Figure 7. Detection of RAG induced single- and double-strand breaks on a 6 nt bubble substrate. DNA substrate possessing 6 nt bubble with $(C/C)_6$ was incubated with MBP core RAGs and the products were resolved on a 15% native PAGE. Different RAG cleavage intermediates used as markers are indicated on the right side. Asterisks indicate radiolabelled strand. RAG specific cleavage products are indicated by arrow.

Supplementary Table 1. Oligomers used in the present study

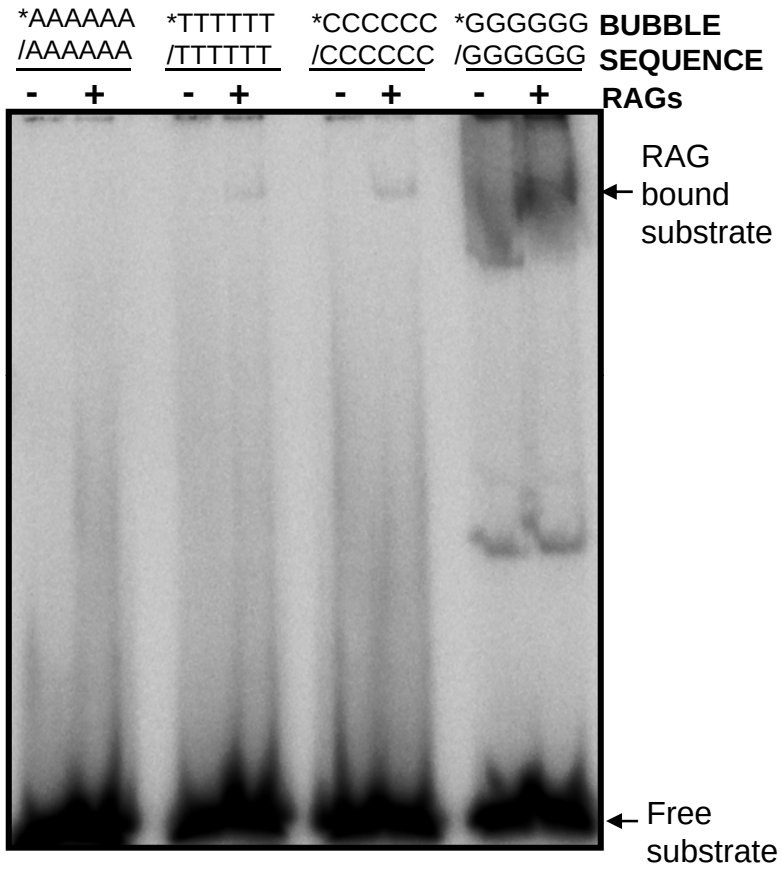
Name	Oligo sequence
AKN1	5′-GATCAGCTGATAGCTACCACAGTGCTACAGACTGGAACAAAAACCCTGCT-3′
AKN2	5′-TAGCAGGGTTTTTGTTCCAGTCTGTAGCACTGTGGTAGCTATCAGCTGAT-3′
AKN5	5′-GACCTGAGGGCGAGCCCGTTG -3′
AKN6	5′-TTTTTTGCTCGCCCTCAGGTC-3′
AKN7	5′-GACCTGAGGGCGAGC-3′
AKN8	5′-GCTCGCCCTCAGGTC-3′
AKN10	5′-TTTGCTCGCCCTCAGGTC-3′
AKN11	5′-GACCTGAGGGCGAGCTTTTTTCGAGTAACTTAACAG-3′
AKN18	5′-CTGTTAAGTTACTCGAAAAAAGCTCGCCCTCAGGTC-3′
AKN20	5′-CTGTTAAGTTACTCGCCCCCCGCTCGCCCTCAGGTC-3′
AKN45	5′-GACCTGAGGGCGAGCAAAAAACGAGTAACTTAACAG-3′
AKN46	5′-GACCTGAGGGCGAGCCCCCCCCGAGTAACTTAACAG-3′
AKN47	5′-GACCTGAGGGCGAGCGGGGGGGAGTAACTTAACAG-3′
AKN48	5′-CTGTTAAGTTACTCGGGGGGGGCTCGCCCTCAGGTC-3′
AKN49	5′-CGAGTAACTTAACAG-3′
AKN52	5′-GACCTGAGGGCGAGTCCGTTGCGAGTAACTT AACAG-3′
AKN53	5′-CTGTTAAGTTACTCGTTTTTTACTCGCCCTCAGGTC-3′
AKN54	5′- CCGTTGCGAGTAACTTAACAG-3′
AKN55	5′-GACCTGAGGGCGAGACCGTTGCGAGTAACTTAACAG-3′
AKN56	5′-CTGTTAAGTTACTCGTTTTTTTCTCGCCCTCAGGTC-3′
AKN57	5′-GACCTGAGGGCGAGGCCGTTGCGAGTAACTTAACAG-3′
AKN58	5′-CTGTTAAGTTACTCGTTTTTTCCTCGCCCTCAGGTC-3′
AKN59	5′-GACCTGAGGGCGAGUCCGTTGCGAGTAACTTAACAG-3′
AKN67	5′–TATCTAATAGTAATCTGAGGGCGAGCAAAAAACGAGTAACTTAACAGGCTTCTATGGA-3′
AKN68	5′-TCCATAGAAGCCTGTTAAGTTACTCGAAAAAAGCTCGCCCTCAGATTACAGTTAGATA-3'
AKN69	5′-TATCTAACTGTAATCTGAGGGCGAGCCCCCCCCGAGTAACTTAACAGGCTTCTATGGA-3′
AKN70	5′-TCCATAGAAGCCTGTTAAGTTACTCGCCCCCCGCTCGCCCTCAGATCACAGTTAGATA-3'
AKN71	5′-TATCTAACTGTAATCTGAGGGCGAGCTTTTTTCGAGTAACTTAACAGGCTTCTATGGA -3′
AKN72	5′-TCCATAGAAGCCTGTTAAGTTACTCGTTTTTTGCTCGCCCTCAGATCACAGTTAGATA -3'
AKN73	5′-TATCTAACTGTAATCTGAGGGCGAGCGGGGGGCGAGTAACTTAACAGGCTTCTATGGA-3′
AKN74	5′-TCCATAGAAGCCTGTTAAGTTACTCGGGGGGGGCTCGCCCTCAGATCACAGTTAGATA-3'
AKN81	5′-GACCTGAGGGCGAGCCCCCCGCGAGTAACTTAACAG-3′
AKN82	5′-GACCTGAGGGCGAGCCCCCGGCGAGTAACTTAACAG-3′
AKN83	5′-GACCTGAGGGCGAGCCCCGGGCGAGTAACTTAACAG-3′
AKN84	5′-GACCTGAGGGCGAGCCCGGGGCGAGTAACTTAACAG-3′
AKN85	5′-GACCTGAGGGCGAGCCGGGGGCGAGTAACTTAACAG-3′
AKN87	5′-CGAGATCGGAGAGATAGAAAAAACTATCTCTCCGATCTCG-3′
AKN88	5′-CGAGATCGGAGAGATAGCCCCCCCTATCTCTCCGATCTCG-3′
AKN89	5′-CGAGATCGGAGAGATAGTTTTTTCTATCTCTCCGATCTCG-3′
AKN90	5′-CGAGATCGGA-GAGATAGGGGGGGCTATCTCTCCGATCTCG -3′
AKN96	5′-GACTGGTTATTATACCGTCAGTCTCTCAGAG-3′
AKN97	5′-CTCTGAGAGACTGACCGTATAATAACCAGTC-3′
AKN100	5′-CTCTGAGAGACTGACAGTATAATAACCAGTC-3′
AKN102	5′-CTCTGAGAGACTGACGGTATAATAACCAGTC-3′
AKN103	5′-GGTATTACGATGCTTGGACTGGTTATTATACCGTCAGTCTCTCAGAGTCCAAGATTTACGCATCCCTGGG-3′
AKN104	5 -CCCAGGGATGCGTAAATCTTGGACTCTGAGAGACTGACAGTATAATAACCAGTCCAAGCATCGTAATACC-3′
AKN105	5 -CCCAGGGATGCGTAAATCTTGGACTCTGAGAGACTGACCGTATAATAACCAGTCCAAGCATCGTAATACC-3′
SCR251	5′-GACCTGAGGGCGAGCCCGTTGCGAGTAACTTAACAG-3′
SCR252	5′-CTGTTAAGTTACTCGTTTTTTGCTCGCCCTCAGGTC-3′



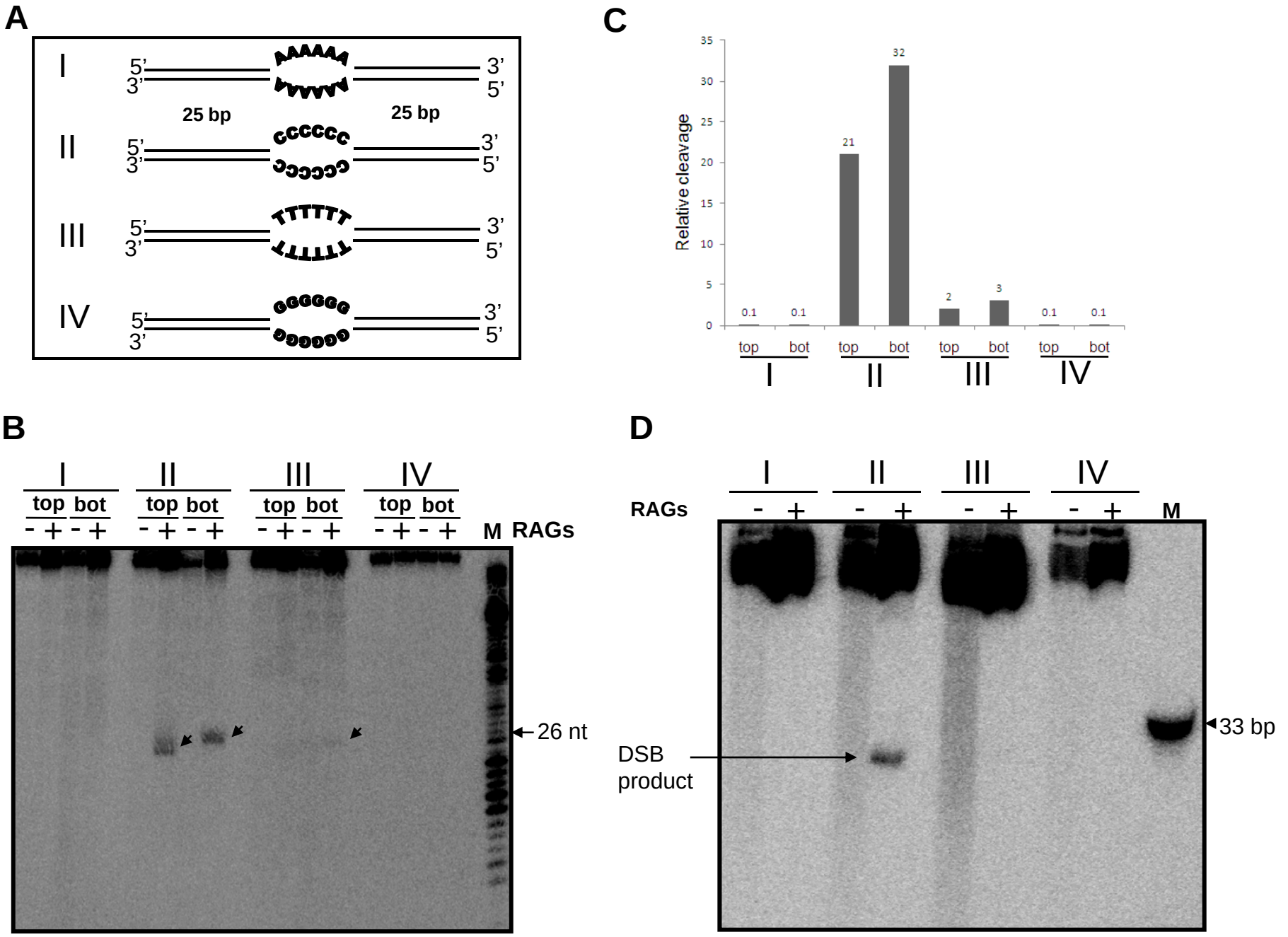
Suppl. Figure 1



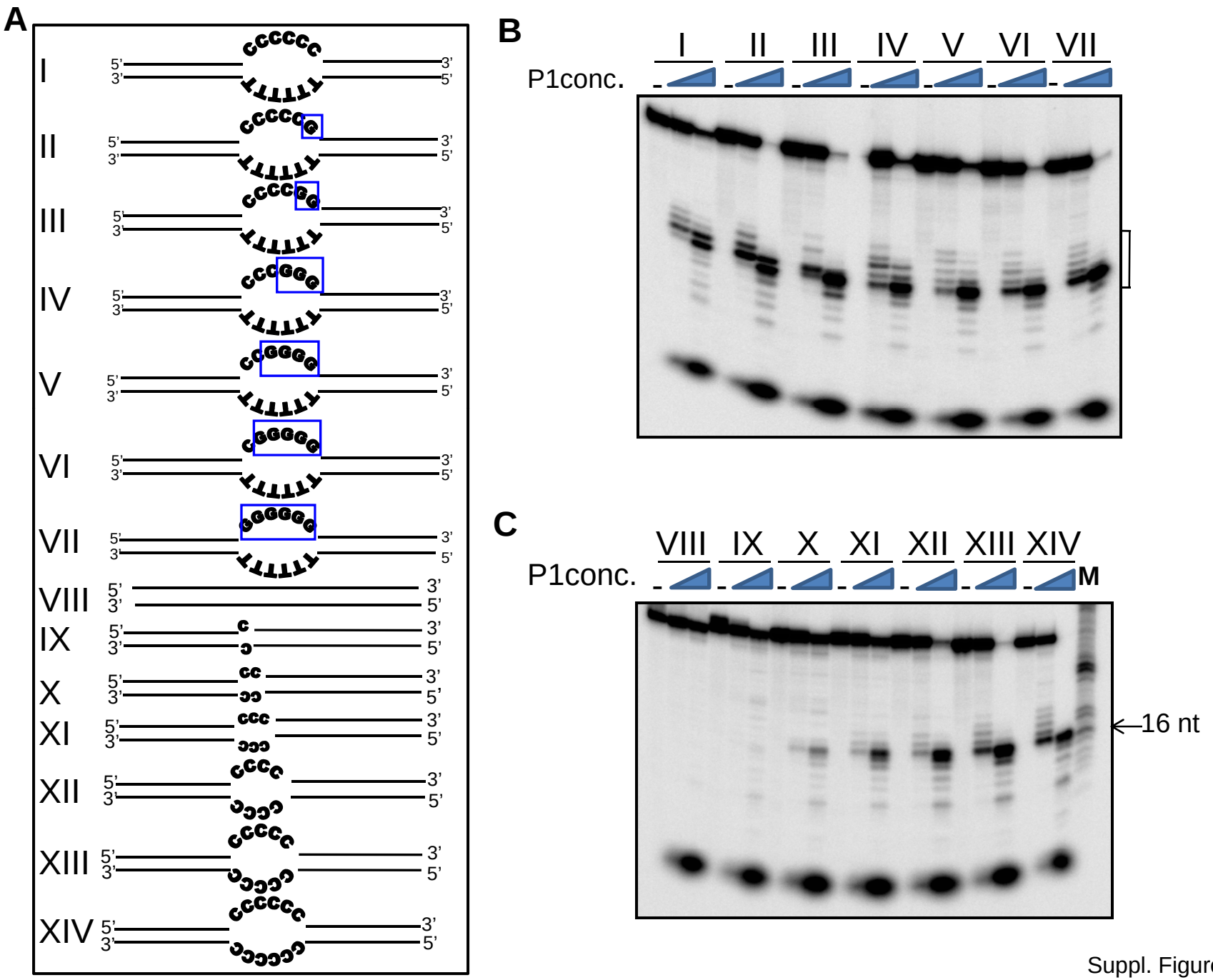
Suppl. Figure 2



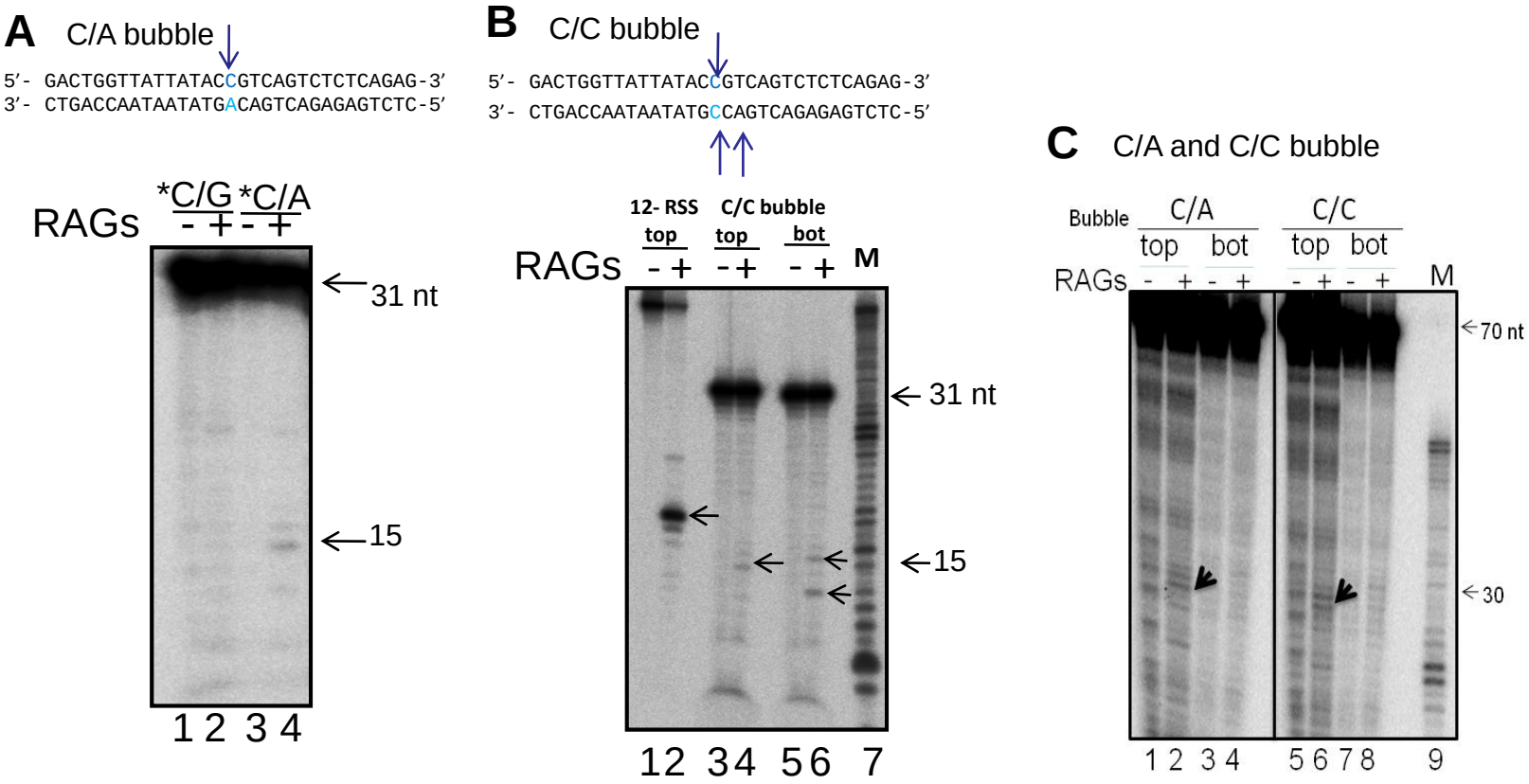
Suppl. Figure 3



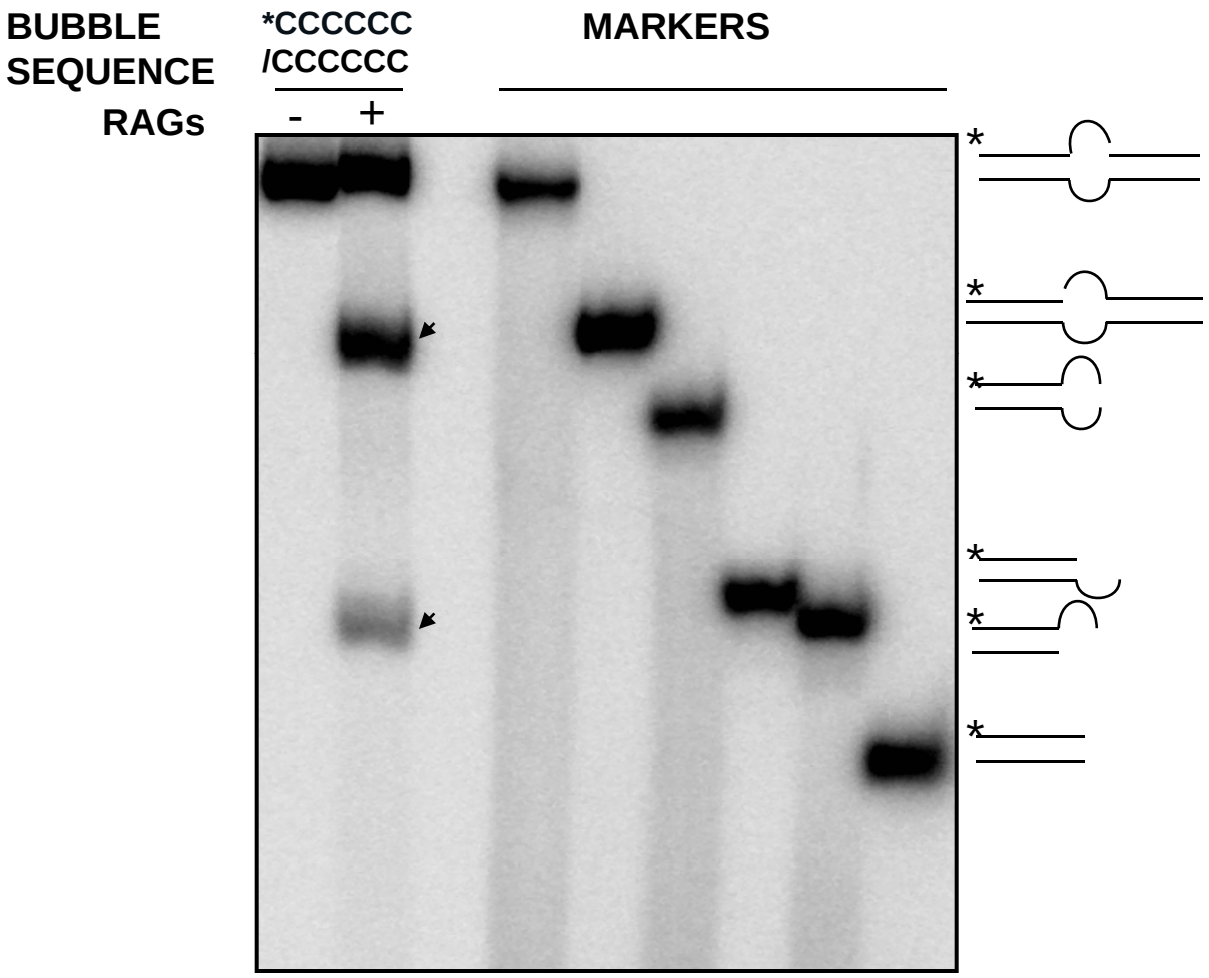
Suppl. Figure 4A-D



Suppl. Figure 5A-C



Suppl. Figure 6



Suppl. Figure 7