The C-terminal region of activation-induced cytidine deaminase (AID) is required for efficient class switch recombination and gene conversion.

Author(s): Sabouri, Somayeh; Kobayashi, Maki; Begum, Nasim A; Xu, Jianliang; Hirota, Kouji; Honjo, Tasuku

Citation: Proceedings of the National Academy of Sciences of the United States of America (2014), 111(6): 2253-2258

Issue Date: 2014-02-11

URL: http://hdl.handle.net/2433/183988

© 2014 National Academy of Sciences.; This is not the published version. Please cite only the published version.
The C-terminal region of AID is required for efficient class switch recombination and gene conversion

Somayeh Sabouri\textsuperscript{a}, Maki Kobayashi\textsuperscript{a}, Nasim A. Begum\textsuperscript{a}, Jianliang Xu\textsuperscript{a}, Kouji Hirota\textsuperscript{b,1}, and Tasuku Honjo\textsuperscript{a,2}

\textsuperscript{a}Department of Immunology and Genomic Medicine, Graduate School of Medicine, Kyoto University, Yoshida Sakyo-ku, Kyoto 606-8501, Japan

\textsuperscript{b}Department of Radiation Genetics, Graduate School of Medicine, Kyoto University, Yoshida Sakyo-ku, Kyoto 606-8501, Japan

\textsuperscript{1}Present address: Department of Chemistry, Tokyo Metropolitan University 1-1 Minami-Ohsawa, Hachi-Ohji, Tokyo 192-0397, Japan

\textsuperscript{2}To whom correspondence should be addressed. Tel: +81-75-753-4371; Fax: +81-75-753-4388; Email: honjo@mfour.med.kyoto-u.ac.jp

Key words: AID C-terminal mutants, DNA repair, CSR synapse
Abstract

Activation induced cytidine deaminase (AID) introduces single strand break (SSB) to initiate class switch recombination (CSR), gene conversion (GC) and somatic hypermutation (SHM). CSR is mediated by double strand breaks (DSBs) at donor and acceptor switch (S) regions, followed by pairing of DSB ends in two S regions and their joining. Since AID mutations at its C-terminal region drastically impair CSR, while retaining its DNA cleavage and SHM activity, the C-terminal region of AID is likely required for the recombination step after the DNA cleavage. To test this hypothesis, we analyzed the recombination junctions generated by AID C-terminal mutants and found that 0-3bp microhomology junctions are relatively less abundant which, may reflects the defects of the classical non-homologous end joining (C-NHEJ). Consistently, the accumulation of C-NHEJ factors such as Ku80 and XRCC4 was decreased at the cleaved S region. By contrast, a SSB binding protein PARP1 was recruited more abundantly, suggesting a defect of conversion from SSB to DSB. In addition, recruitment of critical DNA synapase factors such as 53BP1, DNA PKcs and UNG at the S region was reduces during CSR. Furthermore, the chromosome conformation capture (3C) assay revealed that DNA synapase formation is drastically impaired in the AID C-terminal mutants. Interestingly, these mutants showed relative reduction in GC compared to SHM in chicken DT40 cells. Collectively, our data indicate that the C-terminal region of AID is required for efficient generation of DSB in CSR and GC and thus subsequent pairing of cleaved DNA ends during recombination in CSR.
**Significance statement**

AID, the key enzyme for antigen-induced immunoglobulin gene diversification, initiates CSR by inducing Ig locus specific single strand break (SSB). AID C-terminal mutants (C-mt) generate SSB efficiently but fail to support CSR. We found that residual CSR junctions in AID C-mt predominantly repaired by A-EJ repair and consistently the recruitment of C-NHEJ factors such as Ku80 to S region were reduced. Conversely, the accumulation of PARP1 in AID C-mt JP8Bdel was observed. AID C-mt also showed the relative gene conversion reduction. Moreover, AID C-mt did not support synapse formation in the S regions indicating that the C-terminal region of AID is essential for efficient generation of double strand break in CSR and GC and possibly for synapse formation during CSR.
Introduction

AID is essential for three different genetic events including CSR, GC and SHM which contribute to immunoglobulin (Ig) gene diversification (1-5). Although AID generates SSB in the Ig genes, subsequent repair steps for CSR and GC are similar to each other, but distinct from SHM in their mechanistic properties; (i) generation of the DSBs (ii) recombination and (iii) Uracil-DNA-glycosylase (UNG) requirement for the pairing of the DSB ends (6-10). Despite similarities between GC and CSR, their repair mechanisms have distinct features; CSR recombination requires non-homologous end joining (NHEJ) and GC depends on homologous recombination (HR). During CSR, DSB ends are normally joined by C-NHEJ which requires specific repair proteins such as Ku80, XRCC4 or DNA ligase IV (11, 12). Although in the absence of C-NHEJ, another back up end joining pathway called alternative end joining (A-EJ), joins the broken DSBs ends, which is reported to be slower and also more error prone than C-NHEJ (13). On the other hand, HR which is known as the most common form of homology-directed repair, requires long sequence homology between donor and acceptor DNA for the recombination step by recruiting distinct set of repair proteins such as RAD54, RAD52 and RAD51 to the break sites (14, 15).

Various studies on AID mutations, either in its N-terminal or C-terminal region (4, 8, 9, 16-19), have shown that N-terminal AID mutants are compromised for CSR and defective in SHM indicating that the N-terminal region of AID is required for DNA cleavage (9, 16,
19). On the other hand, the C-terminal region of AID, which contains a nuclear-export signal (NES) and is responsible for AID’s shuttling activity between the nucleus and cytoplasm, is required for CSR specific activity but not for DNA cleavage activity and SHM (8, 16). Among series of AID C-terminal mutants examined, two mutants show characteristic features; P20, which has 34 amino acids insertion at residue 182 and normal nuclear-cytoplasmic shuttling activity, and JP8Bdel which has 16 amino acid truncation at residue 183, accumulates in the nucleus and shows higher DNA break activity at the S region (16, 17). Although there are several reports to suggest the involvement of the C-terminal region of AID in protein stability (20, 21), C-terminal mutants of AID stabilized by fusing hormone-binding domain of estrogen receptor (ER) also show similar CSR-defective phenotypes (8). Taken together, it is likely that the DNA cleavage activity and CSR-specific activity depend on different regions of AID (8, 19). In addition, the C-terminal region of AID is essential for the interaction of AID with poly (A)⁺ RNA via specific co-factor (22). Since CSR requires de novo protein synthesis, we proposed that the C-terminal region of AID may be involved in the regulation of the recombination step, after DNA cleavage, through generation of a new protein (8, 16, 22).

DSBs induced by AID during CSR, are ultimately joined by the efficient DNA repair pathway that needs C-NHEJ factors such as Ku70/80 (12, 23). However, in the absence of C-NHEJ, the A-EJ pathway that relies on microhomology can join the broken DNA ends, while this pathway is also associated with chromosomal translocations (11, 24). Previously, we reported that JP8Bdel enhances aberrant c-myc/IgH translocations, whereas it fails to carry out the efficient recombination between donor and acceptor S regions in the IgH locus.
Therefore, it is important to examine whether the AID C-terminal mutants affect the S-S joining in CSR.

In the current work, we examined whether the C-terminal region of AID is involved in the DNA synapse formation and recombination during CSR in CH12F3-2 and spleen B cells. We also examined the effect of AID C-terminal mutations on GC in chicken DT40 cells, which depends on HR between pseudo V genes and downstream IgVλ region. Utilizing these CSR and GC monitoring systems, we demonstrate that the C-terminal region of AID is required for efficient CSR and GC by formation of DSB from SSB and subsequent end synapse. Collectively, we conclude that in addition to DNA cleavage, AID has a novel function in DSB generation which is required for S-S synapse formation and joining in CSR and recombination in GC.
Results and Discussion

**AID C-terminal mutants fail to support NHEJ in CSR.** Since the DNA cleavage in the Sµ region in AID C-terminal mutants remains unaltered during CSR (8), we examined whether the next step after DNA cleavage, namely recombination is affected by these mutants. We analyzed junction microhomology of Sµ-Sγ1 recombination in AID deficient splenic B cells which retrovirally transfected with wild type AID (wtAID) or P20 fused with GFP to monitor their expression (Fig. 1A and B). Prior to transfection, cells were pre-stimulated with lipopolysaccharide (LPS) for two days and their IgG1 switching was induced by incubation with LPS and interleukin-4 (IL4) for another three days. Consistent with the previous report, IgG1 switching was impaired in P20 expressing cells (8, 16) (Fig. 1C). DNA was isolated from IgG1 switched cells and analyzed for Sµ-Sγ1 junction. Interestingly, we found that the Sµ-Sγ1 junctions in P20 expressing cells contain less C-NHEJ repair signature (0-3bp) and more frequent insertions which were presumably introduced during repair by A-EJ as compared to wtAID expressing cells (Fig. 1D and E). The junctions with 0-3bp microhomology occupied 59.2% and 28.9% in wtAID and P20 expressing cells, respectively ($p<0.05$). Our data showed the longer average microhomology in P20 transfectant and the average microhomology length was 5.65 bp and 3.01 bp for P20 and wtAID, respectively ($p<0.05$).

We also examined the microhomology and insertions at the Sµ-Sα recombination in CH12F3-2 cells which almost exclusively switch from IgM to IgA. CH12F3-2 cells expressing JP8Bdel-ER were activated by the addition of 4-hydroxytamoxifen (4-OHT) for
two days and IgA+ cells were sorted to analyze their DNA for Sµ-Sα junction sequences. The phenotypes of the Sµ-Sα recombination junctions generated by JP8Bdel were essentially identical to those described above for P20; less C-NHEJ and more frequent insertions compared with wtAID (Fig. 1F and G). Our findings agree with the report that the Sµ-Sα junctions in hyper IgM syndrome type II patients, AID+ΔC-termΔ carrying AID C-terminal mutation (R190X) contained the longer microhomology compared to those of healthy individuals (25). These results suggest that the C-terminal mutations of AID drastically affect C-NHEJ and relatively less severely A-EJ.

To confirm the defect of NHEJ in P20 expressing cells, we examined the accumulation of bona fide repair proteins involved in C-NHEJ such as Ku80 and XRCC4 in the Sµ region (23). SSBs were introduced by stimulation of wtAID-ER or JP8Bdel-ER expressing CH12F3-2 cells with 1µM 4-OHT for 3 hours (8). Consistent with the microhomology analyses of the Sµ-γ1 and Sµ-α junction, we observed drastic reduction of Ku80 and XRCC4 in the Sµ region by chromatin immunoprecipitation (ChIP) analysis in CH12F3-2 cells expressing JP8Bdel-ER compared to those expressing wtAID-ER (Fig. 2A). These results indicate that the CSR abolishment by P20 and JP8Bdel mutants is attributed to the impairment of end repairs, especially severe defect of the C-NHEJ pathway.

By contrast, we found the more abundant recruitment of a SSB recognizing protein poly(ADP)-ribose polymerase1 (PARP1) in the Sµ region in JP8Bdel-ER expressing CH12 cells (Fig. 2A). This result is consistent with the previous reports that PARP1 facilitates the A-EJ pathway (26) and preferentially binds to SSB (27, 28). Similarly, our ChIP data
exhibited more abundant accumulation of MSH2 (Fig. 2B), a critical mismatch protein at
the $S_\mu$ region. This observation is also in agreement with the report that MSH2 plays an
important role in the absence of C-NHEJ factors during CSR reaction (29). These data
demonstrate that the AID C-terminal mutants fail to generate DSBs from SSBs and
consequently to recruit proteins required for C-NHEJ. Collectively, the C-terminal mutants
P20 and JP8Bdel are inefficient at the recombination step, characterized by more severe
defect of C-NHEJ pathway.

**AID C-terminal mutants impair the $S_\mu$-$S_\alpha$ synapse formation.** Since the defect of CSR
by the C-terminal mutants of AID appears to affect the recombination step, we examined
whether the interaction between $S_\mu$ and downstream S regions in the IgH locus is also
affected. To this end, we examined the DNA synapse formation using chromosome
conformation capture (3C) assay in CH12F3-2 cells expressing wtAID-ER, G23S-ER (N-
terminal mutant), P20-ER or JP8Bdel-ER. CSR was induced by 1µM 4-OHT for 48 hours.

We first confirmed that CSR is compromised by mutation at N-terminal region of AID and
more severely by mutations those in C-terminal region while their protein expression was
not affected (Fig. 3A and B). Subsequently, the stimulated cells were cross-linked in situ
and subjected to 3C assay (Fig. 3C), followed by restriction enzyme digestion and re-
ligation of the chromatin as described (SI materials and methods). The products were
reverse cross-linked and purified DNA was amplified by the specific primer pairs of two
distant S regions (Fig. 3D). While the $S_\mu$-$S_\alpha$ interaction in the cells expressing endogenous
AID was similar amongst all AID transfectants under the CD40 ligand, IL4 and TGF-β
(CIT) stimulated condition, cells expressing the AID C-terminal mutants, P20 and JP8Bdel
showed significant reduction in Sµ-Sα synapse formation upon 4-OHT addition (without CIT stimulation) (Fig. 3E). On the other hand, G23S which has defect in DNA cleavage but not in CSR (8, 9), demonstrated a similar level of Sµ-Sα interaction to wtAID. We sequenced the amplified bands to confirm that amplified DNA fragments represent the joining of the two appropriate S regions by the designed 3C-PCR primer pairs (Fig. S1).

To confirm the defect in the long-range interaction in splenic B cells in the IgH locus, we expressed wtAID and P20 in AID deficient spleen cells by retroviral infection and stimulated the infected cells with IL4 and LPS for three days. wtAID or P20 were fused to GFP to monitor the protein expression level of the infected cells. IgG1 switching by P20 was drastically reduced compared to wtAID in the GFP positive populations (Fig. 1C). The 3C assay in spleen B cells confirmed the reduction of the Sµ-Sγ1 interaction in P20 expressing cells compared to cells expressing wtAID (Fig. 3F). The bands amplified by PCR and were isolated and sequenced to exclude PCR artifacts (Fig. S1). These results clearly indicate that the long-range interactions in the IgH locus to form the S region synapse, is severely impaired by the C-terminal mutants of AID. Taken together, these data demonstrate that the C-terminal region of AID is required for the S-S synapse formation during CSR. Wuerffel et al., (30) also reported that the reduced level of S-S synapsis in AID−/− spleen B cells. The authors speculated that the reduction of the interactions could be due to the C-terminal scaffolding activity of the AID protein to mediate the long-range interaction in the IgH locus although they did not provide any molecular mechanism.
Reduction of the critical DNA synapse factors in the Sµ region by JP8Bdel. To confirm further the defect of synapse formation by AID C-terminal mutants, we tested the occupancy of the critical DNA synapse factors such as DNA PKcs and 53BP1 at the Sµ region (31-34). By utilizing the JP8Bdel-ER expressing CH12F3-2 cells, we found drastic reduction of DNA synapse forming proteins at the break sites compared to wtAID expressing cells (Fig. 2B). The results are consistent with the report that in the absence of 53BP1, long-range microhomology is favored and 53BP1 protects DNA ends from resection (35, 36). We also examined the binding of UNG at the Sµ region in JP8Bdel expressing CH12F3-2 cells (Fig. 2B). Our UNG ChIP data demonstrated that the accumulation of this protein is clearly AID dependent in agreement with Ranjit *et al* (37), and its accumulation is obviously reduced in JP8Bdel expressing cells compared to wtAID expressing cells. In contrast, the accumulation of the MSH2, an essential mismatch protein was enhanced in the Sµ region. Taken together, these results indicate that the C-terminal mutants of AID are defective in recruitment of critical synapse factors such as 53BP1, DNA PKcs and UNG to the DSBs site in order to support Sµ-Sα synapse formation.

C-terminal mutants of AID reduced GC efficiency. Since GC and CSR share similar mechanistic features, namely recombination, we next examined whether mutations in the C-terminal region of AID also affects HR during GC events. wtAID as well as P20 and JP8Bdel fused to ER were retrovirally introduced into AID−/− DT40 cells. The infected cells were selected with puromycin and subcloning was performed. In this experiment, we particularly used c118 DT40 chicken cell line which carries a frameshift mutation in its complementary determining region 1 (CDR1) of the IgVλ gene and expresses no surface
IgM (sIgM). During GC, if this frameshift is repaired by HR with the pseudo V genes located upstream of the functional variable gene, then the sIgM− DT40 cells revert to sIgM positive (sIgM+) cells (3). AID proteins were activated by 4-OHT for four weeks and GC was assessed by sIgM expression and IgV_\lambda region sequencing. We found more than 3-fold reduction of the sIgM level in P20 infected cells whereas JP8Bdel cells showed 2-fold higher sIgM compared to the wtAID (Fig. 4A). Western blot analysis in DT40 cells showed similar levels of AID expression among transfectants (Fig. 4B). The sIgM increment in JP8Bdel cells could be due to the high rate of point mutations, which was demonstrated for the S region (8).

In order to verify the sIgM reversion observed by flow cytometry, we cloned and sequenced the IgV_\lambda region from sIgM+ cells. To enrich the sIgM+ population in P20 infected cells, the reverted population from four weeks stimulated cells by 4-OHT were single or bulk sorted and IgV_\lambda region was analyzed (Fig. S2 and Table S1). As expected the majority of the mutations were accumulated within CDR1 region (Fig. S2). Interestingly, we observed the GC percentage relative to point mutations is reduced in both P20 and JP8Bdel compared to wtAID, although this reduction is not as drastic as their CSR abolishment (Fig. 4C-D, Table S1). The ratios of point mutations (PM) combined with insertion/deletion (Ins/Del) events to the GC combined with ambiguous mutations (Amb.) were significantly augmented in mutants compared to wtAID (p<0.01 for wtAID vs P20 and p<0.001 for wtAID vs JP8Bdel) (Fig. 4E). This partial defect of GC by C-terminal mutants of AID may be due to the defect of DSB generation from SSB as GC is known to be initiated by either DSB or SSB, but probably more efficiently by DSB (38-40). Collectively, these data
indicate that in order to fulfill the efficient recombination during CSR and GC, the C-terminal region of AID is required for efficient processing of cleaved DNA ends.

Since we have previously shown the CSR depends on de novo protein synthesis, we proposed that AID edits an unknown mRNA to produce a new protein. We have shown that AID interacts with poly (A) containing RNA at the C-terminal region of the AID protein (22). Taken together with present results, we propose that mRNA edited by AID may encode a novel factor that is essential for processing SSBs to DSBs, which subsequently recruits synapse forming proteins such as 53BP1, DNA PKcs and UNG.

The biochemical step that requires the product of the C-terminal region of AID. It is tempting to speculate the mechanism for generation of DSBs from AID-induced SSBs. We have provided a series of evidence that topoisomerase 1 (Top1) is responsible for SSB generation immediately after AID activation (41, 42). Top1 generates SSBs only when it irreversibly forms a complex with the 3’ end of DNA. Top1 is known to covalently associate with non-B structure forming DNA which is abundant at repetitive sequences in the S region. For DNA repair, DNA bound Top1 has to be removed by proteasome, followed by TDP1/TPP digestion to expose 3’ end of SSB (43-45). Since the S region is very rich in repetitive sequences, it can be assumed that SSB is generated on both strands of DNA within a reasonable distance that allows conversion of SSB to DSB by either digestion by exonuclease or endonuclease, extension by DNA polymerase or unwinding by DNA helicase. There are several exonucleases and endonucleases which may be involved in CSR. The MRN complex, Ape1, FEN1 and XPF-ERCC1 are such candidates. Since
SHM is not affected by the C-terminal mutation of AID (7, 8, 16), it is unlikely to regulate specific recruitment of any one of the above mentioned enzymes that can convert SSB to DSB. Collectively, the AID C-terminal region product may affect all the enzymes involved in DSB processing.

The AID C-terminal product may also involved in the synapse formation after DSB formation. Consistently, DSBs generated by I-SceI (46) were not so efficient as those by AID, suggesting that AID may have another function to facilitate CSR. It is likely that the AID C-terminal product may share some functions with UNG as both are required for efficient CSR and GC. However, UNG deficiency differs from the AID C-terminal mutation in much severer phenotypes on GC (6), indicating that the two proteins have different roles after AID-induced DNA break. Elucidation of the precise molecular function of the AID C-terminal production is critical for understanding the regulation of CSR.
Materials and Methods

Constructs and 293T Cells Transfection and Retrovirus Infection. All the AID protein fused to GFP (used in spleen cell experiments) (17) or estrogen receptor (ER)-puromycin resistant gene (puro) (in DT40 experiment) were described previously (8).

Cell Culture. Spleen B cells were obtained from AID−/− mice with C57/BL6 background, aged between 8-12 weeks old. LPS pre-treated B cells were retrovirally infected with wtAID and P20-GFP and stimulated with LPS and IL4 for 3 days. The AID−/− DT40 cell line cl18 (47) was retrovirally infected by AID or various AID C-terminal mutants and cultured with RPMI 1640 medium containing L-glutamine (Invitrogen), Penicillin-Streptomycin Mixed Solution (Nacalai tesque) supplemented with 10% fetal calf serum (GibcoR™) and 1% chicken serum (GibcoR) at 39°C. 24 hours after infection, cells were treated with puromycin (0.5µg/ml) for three days and then applied to limited dilution in 96 well plates for one week. The established single clones were stimulated with 50 nM 4-OHT and incubated for four weeks. Cell medium was refreshed every two days.

Sµ-Sγ1 and Sµ-Sα Junction Analysis. Sµ-Sγ1 region was amplified using high fidelity PrimeSTAR DNA polymerase (TaKaRa) as previously described (48). After sorting the switched cells and DNA extraction, nested PCR was performed. The primers used for the amplification of Sµ-Sα region are listed in Table S2. PCR products were purified (WizardR SV Gel and PCR Clean-Up System) and cloned into pGEM-T Easy Vector (all from Promega). Using T7 and SP6 primers, sequencing was performed with ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems).
**Flow Cytometry and Cell Sorting.** Stimulated AID deficient spleen B cells infected by wtAID or P20 were stained with biotinylated anti-mouse IgG1 (Becton Dickinson) followed by APC-conjugated streptavidine (eBioscience). Anti-mouse IgA-PE (Southern Biotech) was used for staining of IgA-switched CH12 cells. Dead cells were excluded by staining with propidium iodide. For Sμ-Sγ1 junction analysis in P20 cells, IgG1+ population was enriched from stimulated cells by using magnetic beads (Anti-Mouse IgG1 Magnetic Particles-DM, BD IMag™ - BD Biosciences). All the FACS analyses were carried out using FACSCalibur (Becton Dickinson). For Sµ-Sα junction analysis, IgA+ cells were sorted by FACSARia™ (BD Biosciences). Chicken IgM antibody (Bethyl) was labeled with Alexa Flour 647 kit (A20173) according to the provided protocol. Subclones from all DT40 transfectants, after puromycin selection and limited dilution, were stimulated by 4-OHT for four weeks and stained using chicken anti-IgM labeled with Alexa Flour 647. In order to enrich the IgM+ population in P20 cells for sequencing, bulk or single sorting was performed by FACSARia™ (BD Biosciences).

**IgVλ Gene Conversion and Point Mutation Analysis.** After four weeks 4-OHT treatments, P20 transfectant was sorted. The bulk-sorted IgM+ cells were cultured for five days and followed by genomic DNA (gDNA) extraction, whereas gDNA from the single-sorted cells were immediately extracted and subsequently used for nested PCR. IgVλ region was amplified and cloned according to the sorted methods: (i) For bulk sorted cells, gDNA was extracted using phenol-chloroform protocol. IgVλ region was amplified (47), cloned and analyzed as described above. For single sorted cells, DNA was extracted using tween20 and proteinase K. The rearranged IgVλ fragments were amplified by nested PCR using the
5’-GGTAT AAAAG GGCAT CGAGG TCCC -3’ (forward) and 5’-TAACC CTAAG TCCTC CATGG CGCA-3’ (reverse) as an outer primer (35 cycles). The inner primer was the same as primers we used for IgVλ amplification in bulk-sorted cells (47). The PCR product was purified by the kit (Exo-SAP-IT, #78200 Affymetrix) and directly applied for sequencing as described above. All of the modification within entire rearranged IgVλ fragments was categorized into four groups: GC, Amb. mutation, PM or Ins/Del. This classification is based on the accessibility of the pseudo V donor genes, searched in database, (http://blast.ddbj.nig.ac.jp/blast/blastn?lang=ja3), which could be considered as a template. The mutation tract which harbors >9bp string homologous to the pseudo V donor is considered as GC. If there was merely one hit in the sequence and such a mutation was found in data base, it was categorized as an Amb. mutation. The mutation which was not found in donor pseudo genes is identified as a PM. In this categorization, any additional or deletional base along IgVλ region was considered as an Ins. or Del., respectively.

**Chromatin Immuno-precipitation (ChIP).** Active Motif ChIP-IT™ Express Kit was used according to the manufacturer’s instructions with slight modification reported elsewhere (49). Briefly, CH12 AID/JP8Bdel ER cells (5×10^6) were fixed for 5 minutes in the presence of 1% formaldehyde at room temperature. To quench the reaction, glycine to a final concentration of 0.125 M was added and rotated for 5 minutes. After washing the cross-linking reaction, cell lysis and sonication was performed. The sheared samples were subjected to the reverse cross linking, treated with proteinase K and RNaseA. Following electrophoresis, the sheared chromatin which yielded smear between 200-1000bp were used for immunoprecipitation. 2-3 μg antibody was added to the lysate and rotated at 4°C
overnight. After washing the beads, the immunoprecipitated DNA was used for real-time PCR using SYBR green Master Mix (Applied Biosystems). The qPCR signal is normalized by the input. Data set was re-calculated as the maximum value is considered as 1.0. The list of the used primers and antibodies are shown in Table S2 and Table S3, respectively.

**Chromosome Confirmation Capture (3C) assay.** The 3C assay is performed as described previously (30) with minor modification in CH12 cells. Additional details of 3C assay is described in SI Materials and Methods.

**ACKNOWLEDGMENTS**

We thank Ms. Y. Shiraki for the manuscript preparation. This work was supported by Grant-in-Aid for Specially Promoted Research 17002015 (to T.H.), Grant-in-Aid for Scientific Research (C) 254400 (to M.K) and Grant-in-Aid for Scientific Research (C) 24590352 (to N.A.B.).
REFERENCES


Figure legends

**Fig. 1. Expression of AID C-terminal mutants suppress C-NHEJ in CSR junctions.** (A) Schematic representation of AID structure showing locations of nuclear localization signal (NLS), cytidine deaminase domain (CDD) and nuclear export signal (NES). Arrowheads indicate the N and C-terminal mutants (G23S, P20 and JP8Bdel) with their respective mutations. (B) Schematic representation of retroviral expression constructs of GFP vector control and C-terminally GFP fused wtAID and P20 mutant. (C) FACS analysis of IgG1 switching in AID<sup>−</sup> spleen B cells following retroviral transduction of indicated constructs and LPS and IL4 stimulation for 3 days. Percent CSR is indicated in each plot, which was calculated based on GFP<sup>+</sup>IgG1<sup>+</sup> cells in GFP gate. (D,E) Analysis of Sµ-Sγ1 junctions of DNA isolated from GFP<sup>+</sup>IgG1<sup>+</sup> sorted cells following AID expression in AID deficient spleen B cells as indicated. (F,G) Analysis of Sµ-Sα junctions of DNA isolated from IgA<sup>+</sup> sorted from CH12F3-2 cells expressing wtAID or JP8Bdel as indicated. (D, F) C-NHEJ versus A-EJ type of microhomology distributions and (E,G) average % microhomology plots are shown. Statistical significance was performed by Fisher’s exact test.

**Fig. 2. WT but not C-terminally defective AID mutant supports DNA end repair and synapse factor recruitment in the Sµ region.** (A) ChIP assay of known C-NHEJ and A-EJ repair proteins in CH12F3-2 cells expressing either AID-ER or JP8Bdel-ER and (B) ChIP assay for S-S synapse associated factors and mismatch repair enzyme MSH2. Primers were specifically designed to examine DNA break prone Sµ region, and Cµ region was selected as a DNA break negative zone control. ChIP analysis was performed following AID
activation by 4-OHT for 3 hours. Rabbit IgG used as a ChIP control antibody and the background value was subtracted. Values were normalized to the DNA input. In each data set, the maximum value is considered as 1.0. Similar results were obtained in repeated independent experiments.

**Fig. 3. Impaired $\mu$-S$\alpha$ synapse formation in CH12 cells expressing AID C-terminal mutants.** (A) FACS profiles of IgA switching in CH12 cells expressing ER fused AID and its mutants as indicated. Cells were harvested after 48hr of 4-OHT (1µM) and 24 hr CIT stimulation. NS represents cells not treated either for 4-OHT or CIT. The numbers in each panel indicate the percentage of IgA$^+$ cells. (B) Examination of comparable expression level for each AID constructs in CH12F3-2 cells. AID expression and loading amount was monitored by anti-ER and anti-Actin antibodies, respectively. (C) Schematic illustration of long–range interactions between $\mu$-S$\alpha$ elements in IgH locus which can be detected by 3C method. (D) Schematic representation of long-range interactions examined in CH12F3-2 (E, $\mu$-S$\alpha$) stimulated by 4-OHT or CIT, and (F, $\mu$-S$\gamma$1) spleen B cells stimulated by LPS and IL4. Primers positions are indicated below the scheme by the gray rectangles (E) 3C assay for $\mu$-S$\alpha$ and GAPDH control in CH12 cells expressing indicated constructs. NS represents cells not treated either for OHT or CIT. E$\mu$- GAPDH did not show any product and omitted. (F) Representative gel picture of 3C assay in AID deficient spleen B cells complemented by either wtAID or P20 mutant. Treatment conditions are as follows; non-stimulated (NS), 2days LPS stimulation (LPS), and 2 days LPS stimulation followed by LPS and IL4 stimulation for 3days (LPS+IL4).
Fig. 4. Relative reduction of gene conversion by AID C-terminal mutants in chicken B cell line DT40. (A) IgM reversion assay of retrovirally infected AID+/− DT40 cells with ER fused wtAID or AID C-terminal mutant constructs. GFP expressing retroviral vector was used as control. Numbers in the FACS profile indicate the percentage of reverted IgM cells treated with or without 4-OHT (1µM) after 4 weeks. (B) WB below the FACS profile, shows the protein expression of DT40 transfectants with indicated constructs. Cells were harvested 3 days after puromycin selection. (C) IgVλ sequence analysis for gene conversion percentage in AID+/− DT40 cells expressing various AID C-terminal mutants indicated. Pie charts represent relative distribution or proportion of various events that includes gene conversion (GC), ambiguous (Amb) or point mutation (PM) and insertion/deletion (Ins/Del). The total number of the events indicated in the center of each chart. (D) Pie charts represent the proportion of the events when GC and PM are combined with the Amb. mutation and Ins/Del, respectively. (E) Plot in the left represents the proportion of PM to GC and the plot at the right represents the combined proportion of PM and Ins/Del versus GC and Amb. mutations in the sequenced IgVλ region (%). Data are collected from three independent subclones isolated for each AID mutant. Statistical significance was determined by Fisher’s exact test.
Fig. 2

A

Ku80

XRCC4

PARP1

\[ \text{Relative enrichment} \]

OHT(-) \hspace{1cm} OHT(+) 

S\& C\& AID-ER JPBdel-ER S\& C\& AID-ER JPBdel-ER S\& C\& AID-ER JPBdel-ER

B

DNA PKcs

53BP1

UNG

MSH2

\[ \text{Relative enrichment} \]

S\& C\& AID-ER JPBdel-ER S\& C\& AID-ER JPBdel-ER S\& C\& AID-ER JPBdel-ER S\& C\& AID-ER JPBdel-ER
Fig. 3
**Supplementary Figures legends**

**Fig. S1.** Sequence confirmation of the specific amplicons produced by 3C assay for Sµ-Sγ1, Sµ-Sα and GAPDH in spleen and CH12F3-2 cells, respectively. The PCR product was loaded on 2% agarose gel. The specific 3C bands were excised out and cloned into pGEM-T Easy Vector. Sequencing was performed by using SP6 or T7 primers.

**Fig. S2.** Distribution of all events within the sequenced IgVλ region. AID C-terminal mutant transfectants are compared in reference to wtAID tranfectant. AID-/- DT40 cells were retrovirally infected to express wtAID or various AID C-terminal mutants followed by puromycin selection and plated in 96 well by limited dilution. Subclones were stimulated by 4-OHT (1µM) for 4 weeks. After FACS analysis, the representative subclones were selected and analyzed for the IgVλ diversification. Each horizontal and vertical line represents the gene conversion (GC) or ambiguous mutation (Amb), respectively. Mutations with lollipop shape represents the point mutations (PM). Reverse triangle describes the insertion (ins.) and vertical filled bar represents the deletion (del.) events within IgVλ region.

**Table. S1.** Summary of IgVλ analysis of AID-/- DT40 transfectants by wtAID or its C-terminal mutants. Details of the sequencing analysis in IgVλ region correspond to those shown in (Fig. 4 C,D). Only unique GC, Amb, PM or Ins/Del events were considered in the analysis. Event’s percentage was determined by dividing the unique events number to the total events number (i.e. GC% in wtAID: 17/68=25).
Table. S2. List of the primers.

Table. S3. List of antibodies.
Supplementary Materials and Methods

Chromosome conformation capture (3C) assay. This method was developed to detect the long-range interactions between the regulatory elements in the constant region of the IgH locus (1). Briefly, CH12-AID/G23S/P20/JP8Bdel-ER cells were stimulated with 4-OHT for 48 hours or 24 hours with CD40 ligand, interleukin-4 and TGFβ (CIT). 5-7×10^6 cells were cross-linked with 1% formaldehyde for 5 minutes at room temperature. The reaction was quenched by adding glycine to a final concentration of 0.125 M and rotated for 5 minutes at room temperature. After PBS washing, the cross-linked cells were lysed by lysis buffer provided by the Active Motif ChIP-IT™ Express Kit. In order to remove the remaining non cross-linked proteins from the DNA, 10% SDS (at final concentration of 0.3%) was added and incubated for 1 hour at 37°C. Triton X-100 was added to the reaction to sequester the SDS in 37°C for 1 hour. HindIII (TaKaRa) was added to digest the DNA for 2 hour at 37°C. With additional HindIII, the reaction was incubated at 37°C for overnight. The restriction enzyme was deactivated by 10% SDS in 65°C for 20 minutes followed by ligase buffer and Triton X-100 incubation for 1 hour in room temperature. Ligation was performed by T4 DNA ligase (TaKaRa) for 4.5 hours in 16°C and incubated for another 30 minutes at room temperature. The cross-links were reversed by proteinase K incubation for overnight. The DNA was purified using phenol/chloroform extraction and PCR was performed by using the specific primers located at the end of each restriction fragments (Table S2). The PCR products were run on 2% agarose gel.
Fig. S1

Sμ-Sγ1 Seq.

Sμ
GCTGACATGGATTATGTGAGGCTGAGCTAGCATGGGCTGGCCTAGCTGATGAGCTAAGCTT(HindIII)TATTGTACATGGGGATTCAGTTC
ATTATCACCAGAAAAACATTTTTTTCCAAAAATTTCGTACTTTAAATGTACC

Sγ1

Sμ-Sα Seq.

Sμ
GCTGACATGGATTATGTGAGGCTGAGCTAGCATGGGCTGGCCTAGCTGATGAGCTAAGCTT(HindIII)GGCTGAGATGAATCAATAATGAGC
TGGGATGAGCTGACCTACTCTGGCATGGTCTGGGCTAGGC

Sα

GAPDH Seq.

CAGTACATCCACGACATACTCAGCACGGCGCTCACCCCATTTTATGTGAATGGGGCTCCTGATCGCTGTCGAGATGGGATGGGGCTTCGTTGGA
TGAC AAGCTT(HindIII)TCAGCTCATCATGCCCCATCCACACCTGCGTACGCGACCACCATGTAGGTAAGGATCTACAGAAGGCACCT
Table S1. Summary of IgVλ analysis of AID<sup>−/−</sup> DT40 transfectants by wtAID or its C-terminal mutants

<table>
<thead>
<tr>
<th>Mutants</th>
<th>IgM</th>
<th>Total Clone #(bp)</th>
<th>GC # (%)</th>
<th>Amb # (%)</th>
<th>PM #(%)</th>
<th>ins.or del. #(%)</th>
<th>Total events</th>
<th>GC%+</th>
<th>Amb%</th>
<th>PM% +ins.or del.%</th>
<th>PM% /GC%</th>
<th>(PM+ins. or del.%)% / (GC+Amb)%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector</td>
<td>0.2</td>
<td>43(17329)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>wtAID</td>
<td>7.1</td>
<td>90(36270)</td>
<td>17(25.0)</td>
<td>11(16.2)</td>
<td>35(51.5)</td>
<td>4/1(7.3)</td>
<td>68</td>
<td>41.2</td>
<td>58.8</td>
<td>2.1</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>P20</td>
<td>2.2</td>
<td>280(112840)</td>
<td>23(17.7)</td>
<td>8(6.1)</td>
<td>86(66.2)</td>
<td>3/10(10.0)</td>
<td>130</td>
<td>23.9</td>
<td>76.1</td>
<td>3.7</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>JP8Bdel</td>
<td>17.6</td>
<td>338(136214)</td>
<td>26(14.9)</td>
<td>9(5.2)</td>
<td>117(67.3)</td>
<td>7/15(12.6)</td>
<td>174</td>
<td>20.1</td>
<td>79.9</td>
<td>4.5</td>
<td>3.9</td>
<td></td>
</tr>
</tbody>
</table>
Table S2. List of the primers

<table>
<thead>
<tr>
<th><strong>ChIP primers</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S(\mu)1F</td>
<td>AAAGAGACATTTGTGTCCTTGAGTACCG</td>
</tr>
<tr>
<td>S(\mu)1R</td>
<td>ATTGGTTAACAGGCAACATTTTTCTTTTAC</td>
</tr>
<tr>
<td>S(\mu)16F</td>
<td>CAATGTGGTTTAATGAAATTTGAGTTGCCA</td>
</tr>
<tr>
<td>S(\mu)16R</td>
<td>TCTCACAACCTTGGATCTAAGCACTGT</td>
</tr>
<tr>
<td>S(\mu)18F</td>
<td>GTATCAAAGGACAGTGCTTAGATCCAAGGT</td>
</tr>
<tr>
<td>S(\mu)18R</td>
<td>TTTCTCAATTTCTGACAGCTGTCGGCCTTCC</td>
</tr>
<tr>
<td>S(\mu)28F</td>
<td>GCTAAACTGAGGTGATTACTCTGAGGTAAG</td>
</tr>
<tr>
<td>S(\mu)28R</td>
<td>GTTTAGCTTACGGGCCCAGCTCATTCAGT</td>
</tr>
<tr>
<td>C(\mu)F</td>
<td>GTC AGT CCT TCC CAA ATG TCT TCC</td>
</tr>
<tr>
<td>C(\mu)R</td>
<td>CTG GAA TGG GCA CAT GCA GAT CTT T</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>3C primers</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>CAGTAGACTCCACGACATA</td>
</tr>
<tr>
<td>GAPDH</td>
<td>AGTAGTGGTCCTGAGATTCC</td>
</tr>
<tr>
<td>S(\mu)</td>
<td>GCTGACATGGATTGTGAGG</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>GCCTAGCCCAGACCATGCCA</td>
</tr>
<tr>
<td>(\gamma_1)</td>
<td>CGACACTGCGGCAGTTTCCAG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>S(\mu)-(\alpha) junction primers</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S(\mu)-Fw.1st</td>
<td>ATCCACACAAAGACTGCCAAG</td>
</tr>
<tr>
<td>S(\mu)-Fw. 2nd</td>
<td>GTAAGGAGGAGCTGCTAAG</td>
</tr>
<tr>
<td>S(\gamma_1)-Rev. 1st</td>
<td>AGCGCTCCAGATTTCTAGCCCTCCTG</td>
</tr>
<tr>
<td>S(\alpha)-Rev. 2nd</td>
<td>TGGGGCGATGAGCTATGTTTCAG</td>
</tr>
</tbody>
</table>
Table S3. List of antibodies

<table>
<thead>
<tr>
<th>ChIP</th>
<th>Antibody</th>
<th>Cat.number</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Ku80</td>
<td>sc-1484</td>
<td>Santa Cruz</td>
<td></td>
</tr>
<tr>
<td>α-XRCC4</td>
<td>sc-365118</td>
<td>Santa Cruz</td>
<td></td>
</tr>
<tr>
<td>α-PARP1</td>
<td>ab6079</td>
<td>Abcam</td>
<td></td>
</tr>
<tr>
<td>α-UNG</td>
<td>sc-28719</td>
<td>Santa Cruz</td>
<td></td>
</tr>
<tr>
<td>α-Msh2</td>
<td>sc-494</td>
<td>Santa Cruz</td>
<td></td>
</tr>
<tr>
<td>α- DNA PKcs</td>
<td>sc-9051</td>
<td>Santa Cruz</td>
<td></td>
</tr>
<tr>
<td>α- 53BP1</td>
<td>NB100-304</td>
<td>NOVUS</td>
<td></td>
</tr>
<tr>
<td>Normal mouse IgG</td>
<td>Sc-2025</td>
<td>Santa Cruz</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Western Blot</th>
<th>Antibody</th>
<th>Cat.number</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-ER</td>
<td>sc-787</td>
<td>Santa Cruz</td>
<td></td>
</tr>
<tr>
<td>α-actin</td>
<td>A1978</td>
<td>Sigma</td>
<td></td>
</tr>
</tbody>
</table>