Published in final edited form as:

Nat Immunol. 2004 December ; 5(12): 1203–1204. doi:10.1038/ni1204-1203.

AID: a very old motif newly recognized

Amy L Kenter and Palash Bhattacharya

Department of Microbiology and Immunology, University of Illinois College of Medicine, Chicago, Illinois 60612-7344, USA

Amy L Kenter: star1@uic.edu

Abstract

Immunoglobulin class-switch recombination occurs in both frogs and mammals. A new study shows that the recognition mechanism used for the targeting of switch sequences might be evolutionary conserved.

The generation of high-affinity antibodies encoded by immunoglobulin genes is crucial for fending off pathogenic challenges. In B cells, immunoglobulin genes undergo additional refinements to allow the emergence of high-affinity antibody with full protective potential. These refinements are mediated by distinct genetic alterations known as somatic hypermutation (SHM), gene conversion and class-switch recombination (CSR; Fig. 1). The identification of activation-induced deaminase (AID) as the unique regulator of SHM and CSR led to the remarkable insight that these genetic alterations are mechanistically linked. In this issue of *Nature Immunology*, an elegant study by Zarrin *et al.* designed to assess the contribution of specialized transcription-dependent structures to CSR provides new evidence that short, evolutionarily conserved DNA motifs are the main target of AID¹.

Although SHM and CSR seem to be unique activities, they share two properties: the requirement for AID and transcription. SHM is a mutational process focused on variable (V) segments of the immunoglobulin genes that encode the antigen-binding portion of the antibody molecule. Gene conversion acts to substitute V-gene sequence from pseudogenes to the expressed V gene. CSR underlies immunoglobulin isotype switching, a mechanism in which rearranged V genes are swapped from one immunoglobulin heavy-chain constant region (C_H) gene to another through targeted recombination to immunoglobulin switch (S) regions, thereby determining antibody effector functions and tissue distribution. This work has also led to a mystery: how does AID differentially target V genes or S regions? A flurry of investigative activity has indicated that single-stranded DNA (ssDNA) is the main substrate for AID. This hints that the requirement for transcription in both SHM and CSR might be related to generating ssDNA by promoting the formation of special secondary structures in AID target sequences.

How does AID work to create DNA breaks in S regions and mutations in V genes? Insight has come from sequence comparisons showing that AID is related to a family of cytidine deaminases. Although there is some debate about whether AID is an RNA-editing enzyme², the preponderance of evidence indicates that AID deaminates cytidine residues in DNA that are then processed by uracil DNA glycosylase and targeted by repair pathways³. It has been

perplexing that AID is targeted exclusively to V genes and S regions but not to other transcribed genes. How can AID focus so specifically on its target substrates?

S regions extend 1–10 kilobases and consist of highly repetitive sequences. Mouse and human S regions are particularly G rich on the nontemplate strand. Deletion of the endogenous S_{μ} region reduces CSR to downstream C_H genes and targeted deletion of $S_{\gamma 1}$ abolishes switching to immunoglobulin G1, demonstrating that S regions are essential targets of CSR^{4,5}. Based on the correlation of S-region germline transcript expression with specific targeting of downstream S regions for CSR, it was proposed that recombination is initiated by the modulation of access of S-region DNA to AID through transcription. A series of papers demonstrated that AID acts directly on ssDNA or on double-stranded DNA altered by transcription but not on RNA substrates³. It has been speculated that transcription leads to the formation of R-loops in S DNA. When S DNA is transcribed *in vitro*, the newly formed RNA can remain associated with the template strand of DNA, thereby stabilizing the non-template strand as ssDNA for attack by AID. R-loops do not form when the sequence is inverted and the G-rich strand is transcribed. This proposal gained considerable currency when R-loops were specifically detected on S DNA *in vivo* with a sensitive PCR assay⁶.

The requirement of R-loops as the recognition structure for AID has remained controversial. When the endogenous $S_{\gamma 1}$ region is inverted, CSR is reduced somewhat (one fourth)⁵. That result is in keeping with the R-loop model, but the extent of the reduction is less than expected if R-loops provide the recognition structure for AID. When the endogenous $S_{\gamma 1}$ region is replaced with a random G-rich sequence capable of R-loop formation *in vitro*, very low but detectable CSR is found⁵. Inversion of the synthetic sequence leads to an absence of R-loop formation *in vitro* and complete loss of switching *in vivo*⁵. Those findings indicate that although R-loops may form on S DNA undergoing CSR, they are not intrinsic to the mechanism of recombination and not required for AID recognition. Notably, the artificial G-rich S region used in those experiments is devoid of the DNA motifs commonly found in S-region DNA. It is also salient that R-loops do not form on V genes undergoing SHM. It is plausible that there is sufficient ssDNA formed at transcription ‘bubbles’ during germline transcription to serve as substrate for AID (although this raises the question of how AID recognizes its substrate).

Now, Zarrin *et al.* have taken an evolutionary approach to examine whether R-loops are the sole recruiting element for AID in CSR¹. Xenopus is one of the most primitive organisms known to undergo CSR. Its immunoglobulin H locus comprises three constant region genes, C_{μ} , C_{γ} and C_{χ} , each associated with a corresponding S region (Fig. 1a). The xenopus S_{μ} region is A-T rich, containing repetitive sequences, and is not prone to form R-loops during transcription. The absence of R-loop formation in transcribed xenopus S_{μ} DNA was confirmed *in vitro* with a nuclease assay capable of R-loop detection. Zarrin *et al.* substituted the 4.1-kilobase xenopus S_{μ} region for the $S_{\gamma 1}$ sequence in mouse and found that xenopus S_{μ} DNA is an efficient substrate for CSR, recombining at a frequency 25% that of intact $S_{\gamma 1}$ DNA¹. Furthermore, no alteration of CSR capability was noted when the xenopus S_{μ} was inverted. The introduction into the $S_{\gamma 1}$ locus of a 4-kilobase fragment of the xeroderma pigmentosum intron gene, containing average base composition and no extensive repetitive sequence, did not support endogenous CSR. These studies indicate that R-loops are not

crucial for switching and that xenopus S_{μ} DNA contains some essential feature that promotes the recombination reaction.

Examination of S junctions in a xenopus S_{μ} insert located in the mouse $S_{\gamma 1}$ locus demonstrated a preferential localization of recombination breakpoints toward the 5' end, a subregion containing a high density of AGCT motifs¹. This same region is a focus for S junctions in xenopus⁷. CSR breakpoints segregated to the same sub-area even in the inverted xenopus S_{μ} construct, confirming the association of CSR breakpoints with the AGCT-rich region. The AGCT motif is a conserved sequence found in all chicken, mouse and human S regions and is a variant of the DGYW motif, which serves as a 'hotspot' for mutation during SHM. AGCT is a short palindromic sequence that in high density might contribute to the formation of complex secondary structures in DNA, including stem-loop configurations with short ssDNA stretches. However, using the same analysis tools, two groups have reached opposite conclusions regarding the likelihood of secondary structures in S DNA^{7,8}. The simplest interpretation suggests that Zarrin *et al.* have identified an evolutionarily conserved AID target motif used for CSR and that has also been associated with SHM.

CSR and SHM can be initiated independently, leading to the differential targeting of AID to its respective substrates. The C-terminal 8–17 amino acids of AID are vital for CSR, whereas the N terminus is important for SHM^{9,10}. Those studies^{9,10} suggest that unique cofactors 'collaborate' with AID to independently facilitate CSR and SHM. The identities of previously unknown factors required for AID function are beginning to be determined¹¹. Replication protein A has been identified as a critical cofactor that together with AID deaminates DNA in a transcription-dependent way *in vitro* focusing on RGYW motifs¹¹. The facility with which non-R-loop-forming xenopus S_{μ} DNA can serve as an efficient substrate for CSR prompted Zarrin *et al.* to determine whether this sequence is a substrate for deamination¹. They found that xenopus S_{μ} DNA is a substrate for the AID–replication protein A complex but is not acted on by AID alone. These findings are similar to conclusions from studies using substrates with high densities of SHM motifs but lacking R-loop-forming capacity¹¹.

The identification of the DGYW motif as the unifying principle specifying AID targets in CSR as well as SHM prompts both relief and gratification. Conservation of this motif through evolution together with the new genetic and biochemical data provided by Zarrin *et al.* strongly argue that AID specifically targets these sequences. This is an important new insight that will propel investigative efforts. Analysis of a xenopus S region will not necessarily elucidate the DNA motifs that modulate S region function in the mouse. Additional motifs in mouse S regions have been characterized and seem to contribute to isotype specificity to CSR¹². These motifs may have evolved both positive and negative regulatory functions to protect against the promiscuous AID mutator activity in a locus with multiple targets of AID action. How AID focuses on S regions during CSR when V_H genes are actively transcribed will be an area of intense inquiry in the near future. It is probable that additional AID cofactors specific for the N and C termini of AID remain to be found, and they are likely candidates for this function.

References

1. Zarrin AA, et al. *Nat Immunol.* 2004; 5:1275–1281. [PubMed: 15531884]
2. Honjo T, Kinoshita K, Muramatsu M. *Annu Rev Immunol.* 2002; 20:165–196. [PubMed: 11861601]
3. Chaudhuri J, Alt FW. *Nat Rev Immunol.* 2004; 4:541–552. [PubMed: 15229473]
4. Luby TM, Schrader CE, Stavnezer J, Selsing E. *J Exp Med.* 2001; 193:159–168. [PubMed: 11148220]
5. Shinkura R, et al. *Nat Immunol.* 2003; 4:435–441. [PubMed: 12679811]
6. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber M. *Nat Immunol.* 2003; 4:442–451. [PubMed: 12679812]
7. Mussmann R, Courtet M, Schwager J, Du Pasquier L. *Eur J Immunol.* 1997; 27:2610–2619. [PubMed: 9368617]
8. Pan-Hammarstrom Q, Zhao Y, Hammarstrom L. *J Immunol.* 2004; 172:2727. [PubMed: 14978068]
9. Ta VT, et al. *Nat Immunol.* 2003; 4:843–848. [PubMed: 12910268]
10. Shinkura R, et al. *Nat Immunol.* 2004; 5:707–712. [PubMed: 15195091]
11. Chaudhuri J, Khuong C, Alt FW. *Nature.* 2004; 430:992–998. [PubMed: 15273694]
12. Kenter AL, Wuerffel R, Dominguez C, Shanmugam A, Zhang H. *J Exp Med.* 2004; 199:617–627. [PubMed: 14993249]

CSR, SHM and gene conversion. **(a)** The immunoglobulin H (*Igh*) locus in xenopus. A-T-rich S regions have been identified preceding each of the three C_H genes. The relative positions of the C_H genes is not known. VDJ, variable, diversity and joining. **(b)** The steps involved in CSR, with the mouse *Igh* locus as a 'guide'. IgM and IgD are transcribed from the same V_H promoter and are expressed by alternative RNA processing. The germline transcript for IgG1 is below the C_{γ1} gene. CSR exchanges the initially expressed C_μ gene for downstream C_H region genes, including C_γ, C_ε and C_α. CSR occurs by an intrachromosomal deletional rearrangement that focuses on regions of repetitive S DNA. A CSR event leads to the formation of an extrachromosomal excision circle and a new hybrid S_μ-S_{γ1} configuration in the *Igh* locus. The double-strand breaks (DSBs; arrows) in the respective S regions are mediated by AID. Bottom, *Igh* locus modified by CSR. **(c)** In SHM, the mutations induced by AID focus only on the V regions but not on the C_H gene. **(d)** Gene conversion, initiated by AID-induced DNA breaks and facilitated by homologous recombination, results in the transfer of V-gene information from upstream pseudogenes to the expressed V gene. Gene conversion is the dominant mechanism for the diversification of V genes in chicken and rabbits.