Review

CellPress

AID targeting: old mysteries and new challenges

Vivek Chandra, Alexandra Bortnick, and Cornelis Murre

Department of Molecular Biology, Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093-0377, USA

Activation-induced cytidine deaminase (AID) mediates cytosine deamination and underlies two central processes in antibody diversification: somatic hypermutation and class-switch recombination. AID deamination is not exclusive to immunoglobulin loci; it can instigate DNA lesions in non-immunoglobulin genes and thus stringent checks are in place to constrain and restrict its activity. Recent findings have provided new insights into the mechanisms that target AID activity to specific genomic regions, revealing an involvement for noncoding RNAs associated with polymerase pausing and with enhancer transcription as well as genomic architecture. We review these findings and integrate them into a model for multilevel regulation of AID expression and targeting in immunoglobulin and non-immunoglobulin loci. Within this framework we discuss gaps in understanding, and outline important areas of further research.

AID and antibody maturation

Millions of antigens, some innocuous others infectious, challenge the mammalian immune system daily. Adaptive immune responses to these challenges depend on B and T lymphocytes, which express specialized cell surface receptors. Membrane-bound antigen receptors on B lymphocytes recognize specific, pathogen-derived determinants. The B cell repertoire contains a tremendous diversity of antibody molecules, exceeding 10^9 specificities, first achieved during B cell development via V(D)J recombination. This process assembles unique combinations of exons encoding the amino-terminal variable regions of immunoglobulin heavy (IgH) and light chains (IgL) from component variable (V), diversity (D), and joining (J) segments [1–3]. Recombination yields mature but antigen-inexperienced naïve IgM^+ B cells, which exit the bone marrow to secondary lymphoid organs, such as the spleen and lymph nodes. Although V(D)J recombination generates a diverse primary repertoire of antibodies, Ig genes undergo two additional DNA alteration events to enhance the specificity and functionality of antibodies: somatic hypermutation (SHM) and class-switch recombination (CSR). Together, these processes generate antigen-specific, high-affinity, switched B cells that secrete antibodies with unique effector functions [4–6].

Corresponding author: Murre, C. (cmurre@ucsd.edu).
Keywords: AID; SHM; CSR; immunoglobulin loci.

1471-4906/
 http://dx.doi.org/10.1016/j.it.2015.07.003

Both SHM and CSR require the enzyme AID. In addition to immunity-enhancing antibody-diversification processes, AID promotes cancer-causing chromosomal translocations within both Ig and non-Ig loci [7,8]. AID deaminates cytosine (C) in DNA and converts it to uracil (U), resulting in U:G mismatch lesions. These DNA lesions are then converted into point mutations during SHM and into DNA double-stranded breaks (DSBs) during CSR or aberrant chromosomal translocations. AID shows selectivity and deaminates single-stranded DNA (ssDNA) or supercoiled DNA that creates single-stranded patches of DNA [9,10], but does not deaminate double-stranded DNA (dsDNA) or DNA:RNA hybrids [11,12]. During SHM, point mutations and sometimes insertions and deletions, are introduced at a high rate into the recombined variable region exons encoding IgH and IgL, which in some instances will generate B cells with increased antigen affinity. SHM requires transcription and occurs primarily, but not exclusively, at RGYW 'hot-spot' motifs present in variable region exons (where R = purine base, Y = pyrimidine base, and W = A or T nucleotide). It has been proposed that AID-target ssDNA is generated as transcription bubbles or due to supercoilicity during transcription at the immunoglobulin locus [6,13] (Figure 1B).

CSR preferentially occurs in germinal centers, which are microanatomical structures within secondary lymphoid organs where mature B cells encounter antigen presented on helper T cells [14]. The mouse IgH locus comprises eight constant region (CH) exons, with Cμ most proximal to the variable region segments. Newly generated naïve B cells initially express IgM encoded by Cμ exons. Upon CSR, the default Cμ exchanges for an alternative set of downstream CH exons. During this process, B cells switch from expressing IgM to producing different classes of antibody (e.g., IgG, IgE, or IgA), which are encoded by unique CH genes (e.g., Cγ, Cε, and Cα) (Figure 1A). CSR occurs between highly repetitive and evolutionarily conserved sequences termed switch (S) regions, which precede each set of CH exons. Transcription through S regions appears to promote formation of DNA:RNA hybrid structures, such as R loops. These structures create ssDNA substrates for AID-mediated cytidine deamination followed by repair of DNA DSBs by components of the base-excision repair (BER) and mismatch repair (MMR) pathways. End-joining of DSBs between two S regions via nonhomologous end-joining (NHEJ) excises the intervening DNA sequence. This process juxtaposes the rearranged V(D)J segment with a new set of constant region exons,

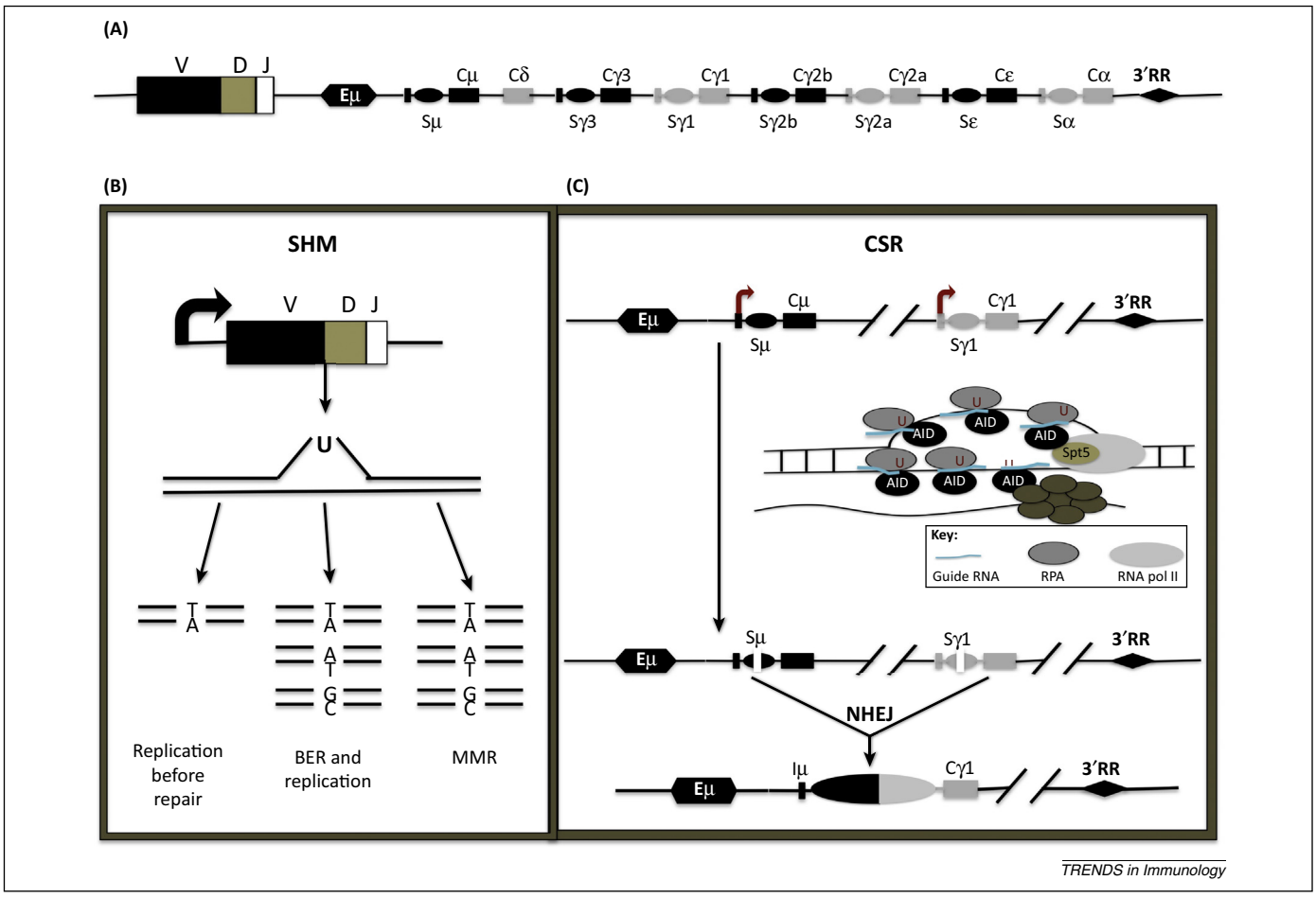

Figure 1. Targeting of activation-induced cytidine deaminase (AID) to the immunoglobulin locus (Ig). **(A)** Structure of the Ig heavy-chain locus after V(D)J recombination. The variable region V(D)J is followed by a series of constant regions, each specifying a different antibody isotype. Each constant-region gene comprises a transcription unit with a cytokine-inducible promoter (P), an intervening (I) exon, S region, and CH exons. The C regions are preceded by corresponding switch (S) repetitive sequences. Distal enhancers are present at the 3′ end of the IgH locus. **(B)** During somatic hypermutation (SHM), mutations are introduced into the rearranged variable region genes. SHM begins with transcription through the variable region. Single-stranded (ss)DNA is created. If replication begins before replacement of the U, a transition mutation (Ts) appears at the original G:C to T:A. If base-excision repair (BER) removes the U before replication, Ts or Tv (transversion) mutations will replace the G:C. Finally, if the U is removed by the mismatch repair (MMR) pathway, both Ts and Tv mutations will occur at G:C and A:T base pairs. **(C)** Class-switch recombination (CSR) is initiated by transcription through the S regions, which leads to the formation of stable R loops. RNA polymerase II (Pol II) stalls at various regions due to the formation of secondary DNA structures, such as R loops. Following stalling, AID interacts with its cofactor, Suppressor of Ty5 homolog (Spt5), and guides RNAs that target AID to complementary S region DNA based on sequence information provided by the guide RNAs. The ssDNA is stabilized by the ssDNA-binding protein complex replication protein A (RPA). Binding of the exosome degrades or displaces the nascent transcript to generate ssDNA, an AID substrate on the template strand.

generating different classes of antibody that retain their original specificity but acquire new effector functions (Figure 1C).

The discovery of AID by subtractive cDNA hybridization from the mouse CH12F3 B lymphoma cells and its essential role in SHM and CSR has propelled efforts to characterize its function for over a decade [15–17]. Despite considerable progress in our understanding of AID activity, how AID selects its genomic targets remains unclear. Earlier models suggesting specific and exclusive AID recruitment to the Ig locus have been revised in light of several findings demonstrating the broad nature of AID targeting. Still, mutation rates at the Ig locus are orders of magnitude greater than at non-Ig loci [18,19]. This bias, which appears even in mice overexpressing AID, suggests that AID is targeted by tightly regulated mechanisms [20]. Recent findings point to a role for genomic architecture and noncoding (nc)RNA in AID targeting, involving mechanisms that include switch RNA, divergent transcription, polymerase pausing, and the action of the RNA exosome complex at certain loci, notably at enhancers [21–24]. We review these findings here, integrating these new insights into earlier findings on AID targeting. From this discussion, we present a framework summarizing the current understanding of the multilevel mechanisms that regulate AID expression and targeting in Ig and non-Ig loci.

Regulation of AID

While the primary and physiological role of AID is to drive antibody diversification by introducing DNA lesions at the Ig loci, AID also poses a threat to genomic integrity. Mistargeted AID activity is known to be a major underlying cause of oncogenic translocations, which are hallmarks of many B cell malignancies [7,8]. Ectopic expression of AID leads to mutation even in non-B cells [25,26]; therefore, regulation of AID expression and function is important for the development of an efficient immune system and maintenance of genomic integrity (Figure 2).

The *Aicda* gene encodes AID and is located on chromosome 6 and 12 in mice and humans, respectively. Early