Seminars in Immunology 24 (2012) 255–263

Contents lists available at SciVerse ScienceDirect

Seminars in Immunology

journal homepage: www.elsevier.com/locate/ysmim

Review

The biochemistry of activation-induced deaminase and its physiological functions

Mani Larijani[a,*], Alberto Martin[b,**]

[a] Immunology and Infectious Diseases Program, Division of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3V6, Canada
[b] Department of Immunology, University of Toronto, Toronto, Ontario M5S 1A8, Canada

ARTICLE INFO

Keywords:
Activation-induced deaminase
Somatic hypermutation, Class switch recombination, Gene conversion
Antibody
B cell

ABSTRACT

Activation-induced deaminase (AID) initiates somatic hypermutation (SHM) and class switch recombination (CSR) by inducing mutations and double-strand breaks at the immunoglobulin (Ig) locus in B cells. AID converts deoxycytidine (dC) to deoxyuridine (dU) in single-stranded DNA (ssDNA). This deamination reaction is enzymatically straightforward, but ultimately results in diverse biological consequences. Here, we review the enzymatic features of AID, such as the parameters that govern substrate binding and catalysis. We discuss how these properties of AID relate to secondary antibody diversification processes and the manners in which they may regulate the targeting of AID to various loci. Based on the current data on AID and other related deaminases, we propose a 3-dimensional structure for AID and how this model provides clues into AID's catalytic mechanism.

1. AID: a DNA-specific mutator governing adaptive immunity, immune physiology and lymphomagenesis

Due to the random nature of V(D)J recombination, the primary antibody repertoire of naïve B cells is broad but relatively low affinity to specific antigens. In order to increase the affinity and alter the effector function of antibodies, B cells that have been activated by antigen recognition undergo further genetic alterations at the immunoglobulin (Ig) locus [1]. These processes include somatic hypermutation (SHM) and class switch recombination (CSR), followed by cellular selection in the germinal center resulting in affinity maturation of antibodies. These events, collectively known as secondary antibody diversification processes, are essential for an effective adaptive immune response. SHM is characterized by mutations in the variable exon that occur at the rate of $\sim 10^{-3}$ muts/bp/generation, which is approximately a million fold higher than the background genome mutation rate [2]. CSR involves double strand breaks (DSB) in switch (S) regions that lie upstream of constant exons [3]. Both processes are initiated by the DNA-mutating enzyme AID [4], which deaminates dC to dU at Ig loci.

The gene for AID was identified 12 years ago by Honjo and colleagues, using a subtractive hybridization strategy between a CSR-stimulated and unstimulated murine B lymphoma cell line [5]. It was later shown that genetically inherited defects in AID result in a hyper IgM syndrome and a lack of SHM in the V-regions of antigen experienced B cells both in humans and mice [4,6]. Patients deficient in AID are characterized by recurrent bacterial infections [6,7]. AID deficiency has also been found to result in a range of other abnormalities, such as autoimmunity and lymphoid hyperplasia [8,9]. In contrast, aberrant and even normal expression of AID have been shown to induce some disease states, such as autoimmunity [10,11] and cancer. The role of AID in transformation of B cells has been shown to be due to its DNA damaging activities [12–15].

AID was initially proposed to act as a RNA editing enzyme, resulting in the downstream production of a DNA mutating agent. This postulate was largely based on the homology of AID to the RNA-editing family member APOBEC1; however, subsequent evidence indicates that AID directly mutates ssDNA. First, expression of AID in hybridomas, fibroblasts and even bacteria is sufficient to cause hypermutation on transcribed episomal substrates and endogenous genes [16–18]. Second, AID is physically localized at the S-region DNA in B cells undergoing CSR [3,19,20]. Third, *in vitro* studies have shown that AID acts on ssDNA and not RNA or dsDNA [21–24]. Fourth, interfering with DNA-specific uracil repair processes diminishes the levels of SHM and CSR [25,26], indicating that these processes result from the generation of uracil in DNA, an activity that AID is known to carry out. Fifth, the mutational spectrum of purified AID *in vitro* is nearly identical to that found at the Ig loci in mice [24,27–30]. While it is established that AID acts on ssDNA, the location(s) and manner(s) in which ssDNA substrates are generated *in vivo* have not yet been fully elucidated (see below).

* Corresponding author at: Division of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, Room 1809A, Health Sciences Center, 300 Prince Phillip dr., St. John's, NL, Canada, A1B 3V6. Tel.: +1 709 777 2515.
** Corresponding author at: Department of Immunology, University of Toronto, Room 7302, Medical Sciences Building, 1 King's College Circle, Toronto, ON, Canada, M5S 1A8. Tel.: +1 416 978 4230.
E-mail addresses: mlarijani@mun.ca (M. Larijani), alberto.martin@utoronto.ca (A. Martin).

1044-5323/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.smim.2012.05.003

2. AID in the context of the APOBEC family

AID is a relatively small enzyme of 198–210 amino acids (depending on the species) with expression primarily limited to centroblasts [5]. AID is a member of the apolipoprotein B mRNA-editing catalytic component (APOBEC) family, a conglomerate of Zn-dependent cytidine/cytosine deaminases with substrates restricted to polynucleotides [31,32]. The APOBEC family itself is a sub-family within the larger Zn-dependent cytidine/cytosine deaminase superfamily that includes bacterial and yeast enzymes acting on polynucleotides, mono-nucleotides and free bases. In humans, the APOBEC family consists of 11 members, including APOBEC1, APOBEC2, APOBEC3(A–H), APOBEC4 and AID. This family of enzymes appeared first in jawed vertebrates, the earliest of which are cartilaginous fish (sharks) [33]. The APOBEC3 family members diverged most recently from one member in non-primate mammals to eight in primates [34]. This recent expansion has been proposed to have occurred due to selective pressures based on the roles of several APOBEC3 members (most notably A3G and A3F) in anti-retroviral immunity [35–37]. A3G and A3F restrict HIV propagation in large part by causing hypermutation of the nascent reverse transcribed minus strand DNA of its genome [38,39].

APOBEC family members contain either one or two deaminase domains. In the members that have two deaminase domains (*i.e.* APOBEC3B, D–G), usually only one domain is catalytically active [34]. Nevertheless, the domains are similar in amino acid sequence, indicating that the second domain arose as a result of gene duplication and that an APOBEC with a single domain was the founding member of the family. Amongst the family members, AID, APOBEC2 and APOBEC4 gene segments have been identified in bony fish, and partial-sequences of AID and APOBEC4 orthologs have also been identified in cartilaginous fish, which are the earliest jawed vertebrates [33,34,40], suggesting that one of these enzymes is the founding member of the APOBEC family. The emergence of AID correlates with the appearance of classical primary and secondary antibody diversification processes in the jawed vertebrates [41,42]. Recently, a primordial form of adaptive immune receptor modification utilizing a gene conversion-like process has been described in lamprey, a non-jawed vertebrate. Two putative cytidine deaminases (pmcda1 and 2) have been identified in the lamprey genome [43–45], although their role in immune receptor diversification has not yet been carefully examined. These findings suggest that the ancestor of the APOBEC family may have arisen in non-jawed vertebrates and hence may predate AID, APOBEC2 and APOBEC4.

3. Tertiary and quaternary structures of AID

A common signature of cytidine/cytosine deaminase enzymes is the architecture of their active site, consisting of two cysteines and a histidine residue acting as the Zn-coordinators and a glutamic acid serving as the proton donor in the deamination reaction [31]. These catalytic residues are present within the monomer in the vast majority of the family members [46–50], though there are instances where it is formed in the multimerization interface of two or four subunits [51]. Thus far only the 3-dimensional structures of APOBEC2 and of the catalytic domain of APOBEC3G have been determined [48–50]. Notably, catalytic activity for APOBEC2 has not yet been demonstrated. Although the structure of AID has not been solved, based on homology with its family members and through mutational studies, several domains have been proposed. A catalytic domain has been defined based on the aforementioned two cysteines (C87, C90 for the human AID), histidine (H56) and glutamic acid (E58) [31,33]. Mutational studies have delineated N- and C-termini motifs that contain segments essential for SHM and CSR, respectively [52]. The basis of the differential requirement of these motifs for SHM and CSR is unclear, but it might be due to the interaction of these domains with targeting co-factors that recruit AID to the Ig variable and switch gene segments, respectively. Alternatively, it may reflect distinct requirements of SHM and CSR for the cellular localization of AID. A nuclear localization signal (NLS) and a nuclear export signal (NES) sequence have been identified in the N- and C-terminal domains of AID respectively, and appear to be important for the shuttling of AID between the nucleus and the cytoplasm [53–55]. Whether these motifs function solely as export or import sequences or have additional roles that are differentially required by AID to catalyze SHM or CSR is unclear. For instance, it is possible that the differential requirement of the N- and C- terminal motifs of AID for SHM and CSR lie more directly in modulating the enzymatic activity of AID itself with respect to its interaction with ssDNA substrates, or assembly into the proper (and possibly differential) stoichiometry required for SHM *vs.* CSR.

As shown in Fig. 1, we have modeled a human AID monomer based on the recently solved structure of the catalytic domain of APOBEC3G [49,50]. As with APOBEC3G, the model shows AID as having a core of 5 beta strands surrounded by 6 alpha helices (Fig. 1, left panel). This model has several notable features. First, the model recapitulates the canonical position of the Zn-coordinating residues (*i.e.* H56, C87 and C90), which in all other cytidine/cytosine deaminases are located at the ends of opposing alpha helices [31]. In this AID model, the H56, C87 and C90 are located close to a putative proton-donor, glutamic acid (E58). Thus, the model generates a viable catalytic site akin to that of other crystallized deaminases (Fig. 1, middle panel) [31,49,50,56,57]. In support of this notion, mutagenesis of H56, C87, C90 and E58 lead to an inactive enzyme [22] (M. Larijani, unpublished observations). Second, the model recapitulates secondary structure predictions of specific regions forming alpha helices or beta sheets [46,47,49,50]. Third, analysis of the AID surface topography reveals a highly charged surface with a large proportion of positive residues, consistent with our previous finding of high affinity binding of AID to negatively charged ssDNA (Fig. 1, right panel). In fact, a putative DNA-binding groove is observed passing over the catalytic pocket. This groove is ~10 Å in width at its widest points and is lined with a high proportion of positively-charged residues. Given that double-helical DNA is ~20 Å wide, its dimensions are consistent with the restriction of AID activity to ssDNA [24,29,58], and its high affinity binding to ssDNA, but not dsDNA [58]. Fourth, a recent random mutagenesis study identified dozens of AID single, double, triple and multiple mutants with altered catalytic rates [59], and we note that the position of many (but not all – see below) of these residues are predicted to be close to the catalytic site.

The quaternary structure of enzymatically active AID has not yet been conclusively determined, although most evidence suggests that native AID is likely a dimer or tetramer. Thus, purified AID migrates as a tetramer [22], and when two differently tagged versions of AID were co-expressed in the same cell, immunoprecipitation of either tag yielded both forms, indicating that AID can form a multimer [60]. By separating purified GST- or His-tagged AID on a size-exclusion column and measuring the activity of various fractions, we found maximum activity in fractions consistent with the size of a tetramer [58], with no activity from fractions consistent with the size of monomers. It is critical to distinguish between isolating multimerized forms of AID *in vitro* and measurement of the enzymatic activity of differentially-sized fractions. Furthermore, our results did not rule out that a small fraction of AID does indeed exist as monomers, since the addition of a GST tag is likely to eliminate this fraction due to inherent propensity of GST itself to self-associate. APOBEC2, which is 75% identical to AID, was crystallized as a tetramer formed by head-to-head interaction of two dimers [48]. Mutational studies disrupting putative dimerization domains of AID (based on homology of primary sequence with other