Seminars in Immunology 24 (2012) 287–292

Contents lists available at SciVerse ScienceDirect

Seminars in Immunology

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

Review

Does DNA repair occur during somatic hypermutation?

Huseyin Saribasak, Patricia J. Gearhart*

Laboratory of Molecular Biology and Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, United States

ARTICLE INFO

Keywords:
Activation-induced deaminase
Base excision repair
Mismatch repair
DNA polymerases

ABSTRACT

Activation-induced deaminase (AID) initiates a flood of DNA damage in the immunoglobulin loci, leading to abasic sites, single-strand breaks and mismatches. It is compelling that some proteins in the canonical base excision and mismatch repair pathways have been hijacked to increase mutagenesis during somatic hypermutation. Thus, the AID-induced mutagenic pathways involve a mix of DNA repair proteins and low fidelity DNA polymerases to create antibody diversity. In this review, we analyze the roles of base excision repair, mismatch repair, and mutagenesis during somatic hypermutation of rearranged variable genes. The emerging view is that faithful base excision repair occurs simultaneously with mutagenesis, whereas faithful mismatch repair is mostly absent.

Published by Elsevier Ltd.

1. Introduction

After it was shown that AID can initiate all antibody diversity mechanisms in mature B cells [1,2], an extensive search was undertaken to identify its mode of action. Initially, AID was thought to be an RNA editing enzyme, due to homology with the RNA deaminase protein APOBEC1 [3], but recently it has been clearly shown that it deaminates DNA in B cells [4]. Thus, AID deaminates cytosine (C) to uracil (U) and thereby introduces an unwanted base at a high frequency into the immunoglobulin (*Ig*) loci, and at a low frequency elsewhere in the genome [5]. Proteins in the base excision repair (BER) and mismatch repair (MMR) pathways then deal with the resulting U:G mismatches. Normally, the uracils would be repaired without error, but this is not the case during antibody diversification. In the *Ig* loci, AID-induced uracils produce DNA lesions and strand breaks to generate somatic hypermutation (SHM), class switch recombination (CSR), and gene conversion (GC) to ensure diversity in antibodies against pathogens. During this flurry of DNA damage, does error-free BER and MMR occur in the *Ig* loci? In this review, we will describe the balance between DNA repair and mutagenesis during the processing of AID-induced damage.

2. Canonical DNA repair pathways

2.1. BER pathway

BER recognizes small base modifications, abasic sites, and single strand breaks [6]. Repair of uracils occurs by the following steps: (1) excision of uracil by uracil DNA glycosylase (UNG), (2) incision at the resulting abasic site by apurinic/apyrimidinic endonuclease 1 (APE1), (3) replacement of the excised nucleotide by DNA polymerase (pol) β, (4) removal of the 5′-deoxyribose phosphate group by pol β, and (5) sealing of the final nick by DNA ligase 3 (Lig3) (Fig. 1). X-ray cross complementing 1 (XRCC1) is a scaffold protein that coordinates pol β and Lig3 activity [7], and is involved in steps 3–5. As pol β has relatively high fidelity, C is usually inserted opposite template guanine (G) to generate error-free repair of the deaminated cytosine.

2.2. MMR pathway

MMR repairs mismatches and other types of damage made during DNA replication and recombination [8]. Repair of mismatches occurs by the following steps: (1) recognition of the mismatch by the MSH2–MSH6 heterodimer, (2) recruitment of MLH1 and PMS2 to introduce a single strand nick near the mismatch, (3) excision of the mismatch and adjacent bases by exonuclease 1 (Exo1) to generate a gap, (4) synthesis in the gap by pol δ bound to the proliferating cell nuclear antigen (PCNA) clamp, and (5) ligation of the ends by DNA ligase 1 (Fig. 1). Synthesis by the high fidelity pol δ ensures that the correct bases are inserted opposite their complementary bases to produce error-free repair of the mismatch.

* Corresponding author at: Laboratory of Molecular Biology and Immunology, National Institute on Aging, National Institutes of Health, Biomedical Research Center, 251 Bayview Boulevard, Suite 100, Room 6C220, Baltimore, MD 21224, United States. Tel.: +1 410 558 8561; fax: +1 410 558 8386.
E-mail address: gearhartp@mail.nih.gov (P.J. Gearhart).

1044-5323/$ – see front matter. Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.smim.2012.05.002

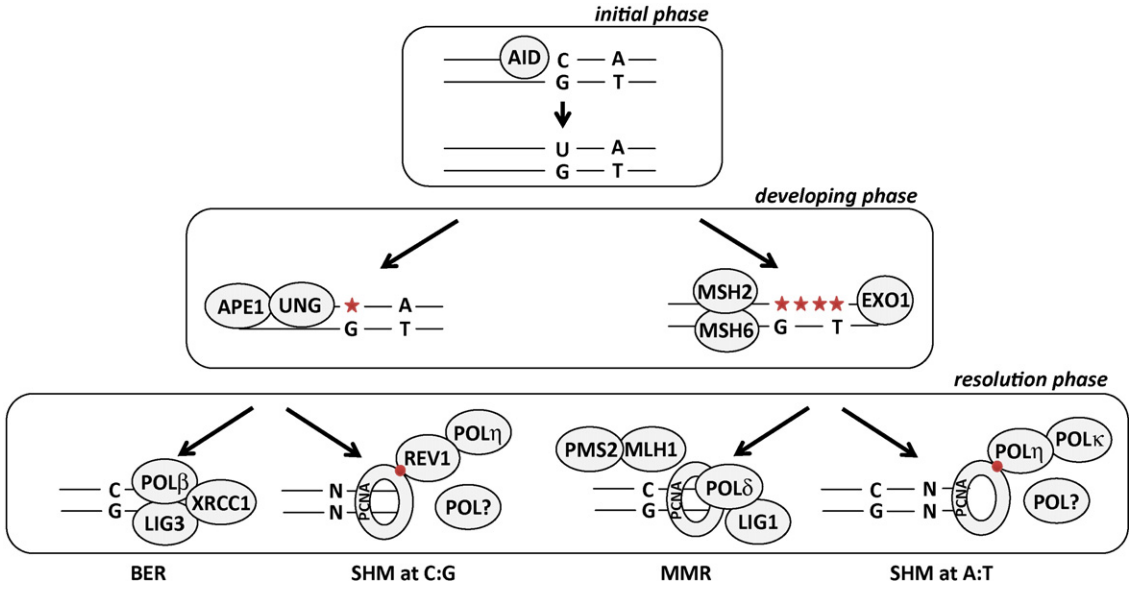

Fig. 1. AID-induced uracils are processed through either DNA repair or mutagenesis. In the initial phase, AID deaminates cytosine to uracil. In the developing phase, the rogue uracils are recognized by two sets of proteins: UNG or MSH2–MSH6. The DNA undergoes incision by APE1 or excision by EXO1, with stars representing the deleted bases. In the resolution phase, the UNG processed substrates can either be faithfully repaired by components in BER pathway, or mutagenically managed by low fidelity DNA polymerases to produce SHM. Similarly, the MSH2–MSH6 processed substrates can either be faithfully repaired by proteins in the MMR pathway, or mutagenically handled by low fidelity polymerases. Polymerases are recruited to the breaks by monoubiquitinated (red circle) PCNA to generate SHM.

3. SHM mutagenesis

During SHM, mutations accumulate in rearranged variable (V), diversity (D), and joining (J) genes on the heavy (H), kappa (κ) and lambda (λ) *Ig* loci. The mutations are mostly single base substitutions, along with occasional tandem double base substitutions, deletions, and insertions. Mutations start 100–200 bp downstream of the transcription initiation site and extend for 1.5–2.0 kb [9]. The frequency of mutation, which is highest in the V(D)J coding exon and the downstream J intron [10–14], occurs at 10^{-2} to 10^{-3} mutations per bp, which is a million times higher than mutation levels in the rest of the genome. As recorded from the nontranscribed strand, C and G nucleotides are mutated equally, implying that AID deaminates C on both DNA strands [15]. However, adenine (A) bases are mutated twice as frequently as the complementary thymine (T) bases, which is likely due to synthesis on the nontranscribed strand by DNA pol η, a low fidelity polymerase that preferentially synthesizes mispairs when copying T bases located on the transcribed strand [16–21]. In terms of the nature of the mutations, transitions are more frequent than transversions, and many mutations occur at C within WGCW (W = A/T), an *in vivo* hot-spot motif for AID [22–25]. Transcription is required for SHM, and the rate of transcription is related to the frequency of mutations [26–28].

3.1. Initial phase

AID is a master catalyst which regulates SHM, CSR, and GC [1,2,29–31]. Early biochemical studies shed light on the catalytic activity of AID (Fig. 1). AID deaminates C to U on single strand DNA, but not on double strand DNA, DNA:RNA hybrids, or RNA in any form [32–35]. To achieve single strand substrates, DNA could be exposed in transcription bubbles or in RNA-DNA loops. A major question is how AID is targeted to *Ig* genes at a higher frequency than to other genes. There are three levels of targeting to consider. (1) Global targeting, to the H, κ, and λ loci. This is the most enigmatic level and may involve *cis* DNA sequences located at the 3′ region of the constant genes to recruit AID [36–38]. (2) Regional targeting to V and switch regions. Mutation occurs for 2–4 kb downstream of promoters located before rearranged V genes and switch regions. It is believed that AID is brought to stalled RNA polymerases, which increases its activity on single strand DNA exposed during transcription [39,40]. (3) Local targeting to hotspots. *In vitro*, AID has preferential activity at C within the WRC (R = A/G) motif, which is a subset of the WGCW motif defined *in vivo*. We and others have recently shown that a loop of 11 amino acids located in the C-terminal region of AID recognizes the WRC sequence [41–44]. If this loop is replaced by a different loop from the APOBEC3 family of DNA deaminases, there is less CSR and SHM.

3.2. Developing phase

Once AID transforms cytosines into uracils in *Ig* genes, proteins in two distinct repair pathways, BER and MMR, can recognize the uracil. It is still not completely understood why some of the canonical repair proteins in the BER and MMR repair pathways are shared by the mutagenic pathways during the developing phase (Fig. 1) [45].

Shared proteins in the BER/mutagenic pathway include UNG and APE1. UNG excises the uracil from U:G to create an abasic site opposite G. Other uracil glycosylases, such as MBD4 and SMUG, do not significantly contribute to SHM [46,47]. APE1 then converts the abasic site to a single strand break. There is some evidence that the MRN (MRE11–RAD50–NBS1) complex, normally involved in double strand break repair, can also participate as an endonuclease in this step [48,49]. Single strand breaks then provide entry for DNA polymerases to synthesize either faithfully or mutagenically, and multiple breaks initiate CSR.

Shared proteins in the MMR/mutagenic pathway involve MSH2, MSH6, and Exo1. The heterodimer MSH2–MSH6 binds to the U:G mismatch [50], and Exo1 creates a gap [51]. Chromatin immunoprecipitation studies also showed that Exo1 is associated with mutating V regions in the BL2 cell line which undergoes SHM [52]. However, a major unanswered question is what causes a DNA nick to allow entry of Exo1 to create a gap? It has been suggested that APE1 might make the nick [53], but the high frequency of A:T mutations in *Ung*$^{-/-}$ mice [54], which would have very