DNA Repair 45 (2016) 1–24

Contents lists available at ScienceDirect

DNA Repair

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

Somatic hypermutation in immunity and cancer: Critical analysis of strand-biased and codon-context mutation signatures

Edward J. Steele

CYO'Connor ERADE Village Foundation Inc., 24 Genomics Rise, Piara Waters, WA 6112, Australia

ARTICLE INFO

Article history:
Received 28 June 2016
Accepted 1 July 2016
Available online 10 July 2016

Keywords:
Somatic hypermutation
Strand-biased mutations
DNA polymerase-η
A-to-I RNA editing
Codon-context mutations
AID/APOBEC-deaminase oncogenesis

ABSTRACT

For 30 years two general mechanisms have competed to explain somatic hypermutation of immunoglobulin (Ig) genes. The first, the DNA-based model, is focused *only* on DNA substrates. The modern form is the Neuberger "DNA Deamination Model" based on activation-induced cytidine deaminase (AID) and short-patch error-prone DNA repair by DNA Polymerase-η operating around AID C-to-U lesions. The other is an RNA-based mechanism or the "Reverse Transcriptase Model" of SHM which produces strand-biased mutations at A:T and G:C base pairs. This involves error-prone cDNA synthesis via an RNA-dependent DNA polymerase copying the Ig pre-mRNA template and integrating the now error-filled cDNA copy back into the normal chromosomal site. The modern form of this mechanism depends on AID dC-to-dU lesions and long tract error-prone cDNA synthesis of the transcribed strand by DNA Polymerase-η acting as a reverse transcriptase. The evidence for and against each mechanism is critically evaluated. The conclusion is that all the SHM molecular data gathered since 1980 supports directly or indirectly the RNA/RT-based mechanism. All the data and critical analyses are systematically laid out so the reader can evaluate this conclusion for themselves.

Recently we have investigated whether similar RNA/RT-based mutator mechanisms explain how *de novo* mutations arise in somatic tissues (cancer genomes). The data analyses indeed suggest that cancers arise via dysregulated "Ig-like SHM responses" involving rogue DNA and RNA deaminations coupled to genome-wide RT events. Further, Robyn Lindley has recently shown that the strand-biased mutations in cancer genome genes are also in "codon-context." This has been termed Targeted Somatic Mutation (TSM) to highlight that mutations are far more targeted than previously thought in somatic tissues associated with disease. The TSM process implies an "in-frame DNA reader" whereby DNA and RNA deaminases at transcribed regions are guided in their mutagenic action, by the codon reading frame of the DNA.

Abbreviations: AID, activation induced cytidine deaminase a APOBEC family member of DNA/RNA C-to-U editors and viral/transposon restriction factors; AP, apurinic or apyrimidinic site as a result of Base Excision Repair (BER); APE, AP endonuclease causes cleavage of the phosphodiester bond 5′ of AP- sites producing 3′-OH and 5′-deoxyribose phosphate (5′-dRP) residues; APOBEC, generic abbreviation for the dC-to-dU deaminase family (APOBEC3 A, B, C, DE, F, G, H) similar in DNA sequence to the "apolipoprotein B RNA editor" APOBEC1- most common dC-to-dU deaminase role is in targeting/restricting viral genomes and restricting the mobility of potentially mutagenic retrotranspositions; A-to-I, adenosine-to-inosine RNA editing at WA-sites in double stranded (ds) RNA in stem-loops; BCR, B cell antigen receptor cell surface bound Ig molecule with the same V[D]J regions as the secreted form of the Ig from that B cell; BER, base excision repair; CSR, immunoglobulin class switch recombination which switches the same heavy chain VDJ rearrangement from IgM to IgG/IgA/IgE heavy chain isotypes; dC-to-dU, cytosine to uracil deamination at single stranded (ss)DNA regions during transcription; A> >T, generic symbol for A/T strand bias indicating mutations of A-sites exceeds (by 2–3 fold) mutations of T-sites on the NTS; G> >C, generic symbol for G/C strand bias indicating mutations of G-sites exceeds (by 1.7 fold) mutations of C-sites on the NTS; C> >G, generic symbol for C/G strand bias ("AID Deamination Footprint") in genetically deficient UNG$^{-/-}$ MSH 2/6$^{-/-}$ mice indicating mutations of C-sites exceeds (by ∼ 1.7 fold) mutations of G-sites on the NTS; GC, post-antigenic Germinal Center formed and populated with rapidly dividing Centroblasts and smaller differentiated largely non-dividing Centrocytes; Ig-like SHM response, refers to genome-wide strand-biased somatic mutation patterns similar to that observed in Ig SHM; MSH2-MSH6 heterodimer, also known as MutSα complexcomponent of the Mismatch Repair (MMR) pathway which recognizes and binds mismatched bases in duplex DNA; NTS, the non-transcribed, or "Top", strand, and TS, the transcribed, or "Bottom", strand;; Pol-η, or DNA polymerase-η (eta) a Y-family DNA translesion repair polymerase (low fidelity synthesis with characteristic mis-incorporation specificities when copying undamaged DNA) which normally allows bypass of UV-induced cis–syn thymine dimers but also displays efficient reverse transcriptase activity (RT-Pol-η); R-Loops, extended ssDNA regions (in Kb range) which form following germline transcription of both the NTS and TS in Ig CSR switch (S) regions; TSM, targeted somatic mutations the targeting of mutations as they occur in the context of the position of the base within a mutated codon (MC) either at the first (MC1) second (MC2) or third (MC3) position; UNG, uracyl DNA glycosylase involved in BER at dU sites in DNA resulting in either an Abasic site or APE-mediated ssDNA nicks (above); V(D)J, generic symbol for a rearranged variable region gene mediated by the proteins of Recombination Activating Genes RAG1/RAG2 (VDJ heavy chain and VJ for light chains).

E-mail address: ejsteele@cyo.edu.au

http://dx.doi.org/10.1016/j.dnarep.2016.07.001
1568-7864/

Contents

1. DNA-based versus RNA-based models of SHM

The chromosomal substrate for the somatic hypermutation (SHM) process in immunoglobulin (Ig) genes is the somatic V[D]J rearrangement which brings a randomly selected heavy or light chain germline V element into a position just upstream of the Constant (C) region exons (Fig. 1). Since a different germline V is rearranged in each developing B lymphocyte in the bone marrow a large clonal pre-immune repertoire of productive heavy and light chain Ig rearrangements becomes possible arising from a limited number of germline V, D and J elements.

SHM in mammals (mice, humans) is an antigen-driven post-rearrangement diversification process which introduces largely point mutations (and some insertion-deletion events) into the V[D]J coding region on which further antigen-mediated clonal selection operates leading to the affinity maturation of B lymphocyte Ig BCRs over time. The evidence supporting the RNA/RT-based model of SHM (Fig. 2) will be evaluated here against the mainstream DNA-based DNA Deamination model (Fig. 3). There is now overwhelming evidence for the RNA/RT-model that is not necessarily reflected in the current body of literature normally reviewed by the SHM field. The purpose of this paper is to rectify this situation by comparing the two competing models in their ability to explain the SHM data published since about 1980. This process of review of Ig SHM and CSR in concert with the somatic mutation data in genome-wide "Ig-like SHM responses" in cancer cells, leads us to a new era in understanding of the broad field of "hypermutation." It leads also to the highly non-random Targeted Somatic Mutation (TSM) process focused on G:C and A:T base pairs in their "codon-context" in the protein-coding regions of the human genome.

Our interest in antibody diversity and the origins of the variable (V) gene segment repertoire in the vertebrate and mammalian germline began in the 1970s [1–4] and has continued to this day [5–21]. However, for the present work on somatic hypermutation the narrative necessarily begins with the genesis of our RNA/RT-based approach to the SHM mechanism [22] which, contrary to other claims [78], was the first publication in the SHM field to invoke the idea that transcription plays a direct and necessary role in SHM. This then led to the assembly of the relevant molecular evidence [23–46] from both our laboratory [23,29,33,37,41] and that available from public databases [44,46]. However, our work has depended in equal part on the SHM field. Here we have provided our interpretations and analyses [24–27,30–32,34–36,38–43] of the relevant data produced by other laboratories publishing on SHM and CSR [47–158] – the main laboratories are those of Patricia Gearhart [47–69], Ursula Storb [70–80], Cesar Milstein, Cristina Rada and Michael Neuberger [81–93], Latham Claflin [94], Al Bothwell [95–97], Klaus Rajewsky and Garnett Kelsoe [98–103], Tasuku Honjo [104–110], Tom Kunkel, Igor B. Rogozin and Mike Resnick [111–115], Michael Neuberger, Sved Petersen-Mahrt, Reuben Harris, Javier Di Noia [116–131], Fred Alt [132–136], Myron Goodman and Matthew Scharff [137–142], Michael Lieber [143–148], Nina Papavasiliou [149–151], and the Paris group of Reynaud and Weill [152–158].

The principle conclusion is this: the complete spectrum of somatic mutations at V[D]J loci at A:T and G:C base pairs following full blown SHM *in vivo* is best explained by reverse transcription of the Ig RNA primary transcript and then site-directed integration of the error-filled V[D]J cDNA copy of the transcribed strand (TS). More specifically, all the current data are consistent with a scheme where AID initiates SHM (and CSR) by direct C-to-U DNA deamination in the ssDNA of transcribed Ig regions. Transcription over these AID-mediated lesions (C-to-U, Abasic sites) incorporates base modifications into the Ig V[D]J pre-mRNA followed by copying into cDNA then integration and ligation of this error-filled cDNA copy of the transcribed strand (TS) at the V[D]J site. The subsequent sequelae of heteroduplex resolution, replication and antigen-mediated positive (affinity) selection then decide the fate of the mutated B cell bearing the mutated V[D]J regions in the heavy-light chain Ig heterodimers [30,43]. The main molecular steps are shown in Fig. 2.

With such an explicit introduction the author is also aware that the RNA/RT-based model is still not widely accepted by the SHM field. First however, an erroneous citation needs addressing. The claim that "AID had originally been postulated to act on Ig mRNA rather than DNA" (the citation is to Refs. [35,105]) as claimed [78] is incorrect. No such published claim has ever been made by Steele and colleagues [22–46]. Second, the other aim of this review is to clearly explain how all the extant evidence is indeed consistent with a pre-mRNA template intermediate coupled to reverse tran-