Mechanism of Somatic Hypermutation in Immunity and Cancer:


Edward J Steele¹ and Robyn A Lindley²,³

¹ CYO’Connor ERADE Village Foundation Inc. Piara Waters, WA, AUSTRALIA
² GMDxCo Pty Ltd, Hawthorn Vic, AUSTRALIA; and ³ Department of Pathology Faculty of Medicine, Dentistry & Health Sciences, University of Melbourne Vic, AUSTRALIA

Running head: Mechanism A-to-I at Transcription Bubbles

Keywords:
Somatic Hypermutation at Transcription Bubbles
Strand-Biased Mutations
DNA Polymerase -η
A-to-I RNA editing
Targeted Somatic Mutation (TSM) and Codon-Context Mutations
AID/APOBEC-Deaminase Oncogenesis

Correspondence:
Associate Professor Edward J Steele PhD
Honorary Research Associate
CYO’Connor ERADE Village Foundation Inc.
24 Genomics Rise, Piara Waters, WA 6112, AUSTRALIA

Tel (+61) (0) 420 863 551
Email: ejsteele@cyo.edu.au
Abstract

This paper simply links the findings of Zheng, Lorenzo and Beal (2017) to our previous work on strand biased and codon-context mutation signatures in B lymphocytes (Ig SHM) and codon-contexted exome-wide point mutation patterns in cancer genomes. We conclude that in vivo the A-to-I DNA editing component at RNA:DNA hybrids occurring in Transcription Bubbles, while important, is of far lower A-to-I editing efficiency than in dsRNA substrates (as shown in Zheng et al 2017). The extreme strand biased mutation patterns documented by us in vivo should be logically rationalized by the predicted sequential steps of the RNA/RT-based mechanism.

Abbreviations used in this paper:
Aag, alkyladenine DNA glycosylase; ADAR, Adenosine Deaminase that acts on RNA; ADAT, Adenosine Deaminase that acts on tRNA; AID, activation induced cytidine deaminase, a APOBEC family member, initiating via C-to-U lesions in ssDNA of class switch recombination (CSR) and somatic hypermutation (SHM) processes at somatically rearranged Ig V(D)J gene loci, and known to activate cytidine mutagenic deamination during transcription in other somatic tissues, particularly in cancer; APOBEC family, generic abbreviation for the deoxyribonucleic acid, or dC-to-dU, deaminase family (APOBEC3 A, B, C, D, F, G, H) similar in DNA sequence to the “apolipoprotein B RNA editor” APOBEC1, and known to activate mutagenic cytidine deamination during transcription in somatic tissues, particularly in cancer; AP, an Abasic, or apurinic/apyrimidinic, site; APE, AP endonuclease; A-to-I, adenosine-to-inosine RNA editing; BER, base excision repair; Deaminase, catalytic domain in ADAR and AID/APOBEC enzymes; DSB, double strand DNA breaks; Ig-SHM-like response, strand-biased somatic mutation patterns similar to that observed in Ig SHM; MMR, mismatch repair; Motif, 4 to 6 nucleotide (N) sequence defining specificity of deaminase targeting; MSH2-MSH6, MutSα heterodimer recognising mispaired bases in DNA duplex; N, any nucleotide; NTS, the non-transcribed, or “Top”, strand; NGS, Next Generation Sequencing; Pol-η or DNA polymerase-η (eta); R, Adenosine (A) or Guanine (G), purines; RNA Pol II, RNA Polymerase II; RT, reverse transcriptase; RT-Pol-η, reverse transcriptase activity displayed by Pol-η; S, strong base pair involving Cytosine (C) or Guanine (G); SHM, somatic hypermutation; T, Thymine; TS, the transcribed, or “Bottom”, strand, in context of a Transcription Bubble; TSM, targeted somatic mutations : the process of targeting C and A nucleotides for deamination in actively transcribed genes that results in a dominant type of mutation caused by a DBD or Inf-DBD at a particular codon position; TSRT, target site reverse transcription; U, uracil; UNG, uracil DNA glycosylase involved in BER at dU sites in DNA resulting in either an Abasic site (AP) or APE-mediated ssDNA nicks (above); UTR, untranslated regions in the upstream (5’) and downstream (3’) regulatory regions of protein coding genes; V(D), generic symbol for a rearranged immunoglobulin (or T cell receptor, TCR) variable region gene in the Adaptive Immune System; W, weak base pair involving A or U/T; X, C or A ; Y, pyrimidines T/U or C.
The arresting and very important paper just sent to us as a PDF (by Professor Liam Keegan of CEITEC – Central European Institute of Technology - Masaryk University in Brno) on January 31st 2017 published by Zheng, Lorenzo and Beal (2017) has just appeared In Press in Nucleic Acid Research. These beautiful and striking biochemical data are very relevant to previous data and analyses published by us on the in vivo deamination mechanisms (C-to-U, A-to-I) which generate the characteristic somatic point mutation patterns in rearranged immunoglobulin (Ig) variable genes (B lymphocyte, V[D]Js, Steele et al 2006; Steele 2009; Steele 2016) and in the exomic regions of the human cancer genome (Steele & Lindley 2010; Lindley & Steele 2013; Lindley 2013; Lindley et al 2016; Lindley 2017 In Preparation, about to be submitted to Molec. Genet. Genomic Med.).

Our molecular explanation of the extreme strand biased mutation patterns in both Ig SHM in B lymphocytes and in the Ig-SHM-Like responses in human cancer genomes is summarised in Figure 1 and the legend. In Figure 2, is a summary of the molecular events at "stalled" Transcription Bubbles that we believe play a key role in vivo during Ig SHM and tissue-wide oncogenesis. The data summarised in these Figures are covered in detail in our previous publications (reviewed in Lindley and Steele 2013; Steele 2016; Lindley 2017). Our purpose here is to show the consistency between our work on these mechanisms, and the findings just published by the group of Peter Beal at Department of Chemistry, University of California, at Davis (Zhen et al 2017).

**Figure 1. Pattern of Somatic Point Mutations in Ig Somatic Hypermutation in B Lymphocytes.** Likely molecular explanation for the extreme strand biased somatic mutations in Ig SHM. But very similar data and explanation apply to exome-wide point mutations in cancer genomes (based on the hypothesised dysregulated Ig-like SHM process operating across the cancer genome involving DNA and RNA deaminations coupled to reverse transcription). Adapted from Figure 1 in Lindley and Steele (2013) and Figure 7 in Lindley (2017), and reviewed again in Steele (2016). This is a variant of the target site reverse transcription (TSRT) process originally hypothesized by Luan et al (1993) and first applied to the Ig SHM process in Steele et al (1997). Shown for the generation of the main A-site and G-site strand biased mutation components is a Transcription Bubble and sequelae showing some hypothesised DNA and RNA intermediates highlighted for the generation of the main strand- biased mutation signatures involving A-to-G, G-to-A, G-to-T and G-to-C. Black lines are DNA strands, red lines are mRNA, blue lines are cDNA strands copied off mRNA by a cellular reverse transcriptase such as DNA polymerase η. Steps on the right show various mutated DNA and RNA intermediates and substrate complexes for both
deamination reactions, 8oxoG modifications in RNA (Wu and Li 2008), and cDNA synthesis (it is not known if 8oxoG sites generated by reactive oxygen species are preferred in unpaired loops or dsRNA regions). In over view, mutations are first introduced at the DNA level by AID/APOBEC family-mediated C-to-U deaminations and then uracil DNA glycosylase (UNG)-generated abasic sites in the TS (which can further mature into single strand nicks via the action of AP endonuclease (APE). These template sites are transcribed into mRNA by RNA Pol II generating G-to-A and G-to-C modifications respectively in the pre-mRNA Kuraoka et al, (2003) which on TSRT-mediated reverse transcription, integration and DNA replication result in G-to-A and G-to-C mutations in the NTS. Separately, adenosine-to-inosine (A-to-I) RNA editing events at WA targets in the nascent and Transcription Bubble-proximal dsRNA stem loops may be copied back into DNA by reverse transcription via Pol-I (Franklin et al, (2004; Steele et al 2006). Also shown in green are 8oxoG modifications in mRNA which on reverse transcription, integration and DNA replication would result in strand-biased G-to-T transversions on the NTS. The strand invasion (?) and integration of newly synthesised cDNA TS (?) are hypothesized necessary steps (not shown here). In more detail: RNA Pol II introduces mutations in mRNA as it copies the AID/APOBEC lesions in TS DNA, concurrently A-to-I RNA edited sites appear in RNA stem(-loops) forming in nascent mRNA near the transcription bubble [Steele et al., 2006] or 8oxoG modifications via reactive oxygen species. Next formation of RT-priming substrates (for Y Family translesion DNA polymerase-η, Goodman 2002 now acting in it reverse transcriptase mode, Franklin et al 2004) by annealing of nicked TS strand with an exposed 3′-OH end. This could arise due to excision at a previous AID-mediated abasic site or an excision introduced by endonuclease activity associated with the MSH2-MSH6 heterodimer engaging a U:G mispaired lesion. This allows extension of new TS by cDNA synthesis from the 3′-OH end copying the already base modified mRNA template (with I base pairing preferentially, like G, with C; and 8oxoG mispairing with A). Then an unknown and indeterminant number of steps involving strand invasion(?), heteroduplex formation and/or resolution of heteroduplex (?), full length copying of newly synthesized transcribed strand (?) cDNA is locked into the genomic DNA at the the V(DJ) site as envisaged by Luan et al (1993).

The primary purpose then of the RNA/RT-based mechanism of SHM (Steele and Pollard 1987; Steele and Blanden 2001; Lindley & Steele 2013; Steele 2016) is to explain the extreme strand biased mutations at A:T and G:C base pairs. This model does just that. It does not focus on trying to explain the dominance of A-to-G over T-to-C, but the totality of A-site strand dominance over T-site mutations (A>>>T) and the same type of extreme excess of G-site mutations over those at C-sites (G>>>C, as shown in Figure 1).

The need to explain these two extreme strand biases has led us to the molecular model shown in Figure 1. This model actually does explain the minutaie of all the known major mutation strand biases (Steele 2009). For this reason, we evaluate all new data in the field (e.g. Zheng et al 2017) as to whether or not their molecular implications flow through to real in vivo mutation phenomena - as well as helping us understand the generation strand biased and codon-contexted mutation signatures (Lindley & Steele 2013; Lindley 2013; Lindley et al 2016; Steele 2016).

It is clear from the quantitative A-to-I editing data in Zheng et al (2017), that the editing efficiency of ADAR1/2 on the DNA component (in the RNA:DNA hybrid) is far less (an order of magnitude lower on a molar/time kinetic basis) than normal A-to-I editing of dsRNA substrates. Further, in vivo the relevant configurations in a Transcription Bubble would be as shown in Figure 2, and discussed at great length in Lindley (2017). Since the evidence shows that A-to-I editing does not occur in ssDNA e.g displaced NTS in Figure 2 (Lindahl 1993; Longerich 2007; Alseth et al 2014; and analysed in detail in Lindley 2017) how
then do we explain A>>T strand bias in the context of the Zheng et al 2017 biochemical data? It seems to us, that since A-sites in the RNA:DNA hybrid (in both orientations 'top' v 'bottom') are equally less efficient as editing targets (compared to RNA: RNA duplexes). It is our view that the strand biases at A-sites and G-sites (C-site on TS) must logically occur as predicted in the RNA/RT-based model shown in Figure 1.

We conclude that in vivo, the A-to-I DNA editing component at RNA: DNA hybrids in the context of a Transcription Bubble, while important, it exhibits far lower A-to-I editing efficiency than in dsRNA substrates (as shown in Zheng et al 2017). The extreme strand biased mutation patterns documented by us in vivo, can only be understood in the context of the RNA/RT-based model proposed by us. This is because, in vivo the edited dA would be on the TS strand in the Transcription Bubble see Figure 2; the editing of that dA cannot generate the A>>>T strand biases evident in all the extant in vivo datasets as described in Figure 1. We look forward to further experimentation on somatic hypermutation of Ig V(D)J genes in ADAR1 deficient Aicardi-Goutières Syndrome (AGS) patients (O’Connell et al 2015; Rice et al 2012), and possibly using the VDJ somatic hypermutation read-out of the VH6 gene employed by Patricia J Gearhart and associates (Zeng et al 2001).
Acknowledgement
We thank Liam Keegan for his open minded generosity on the role of A-to-I RNA editing in the generation of the somatic hypermutation pattern, and for bringing the paper by Zheng et al 2017 to our immediate attention.

Conflict of Interest
The authors have no conflict of scientific interest that harms the objectivity of this paper.

References


