Check for updates

RESEARCH HIGHLIGHT Dancing with DNA: AID embraces flexible partners

Jianshu Wang¹ and David G. Schatz[™]

© The Author(s) under exclusive licence to Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences 2023

Cell Research (2023) 33:743-744; https://doi.org/10.1038/s41422-023-00823-1

Wang et al. reveal that mesoscale DNA sequences containing flexible pyrimidine-pyrimidine dimers determine preferential hypermutability by facilitating binding to positively-charged surface patches of AID.

The generation of diverse antibody repertoires is crucial for the adaptive immune system to recognize a wide range of pathogens. Two mechanisms are central to this process: *V(D)J* recombination, which occurs during the early stages of B cell development and generates the initial diversity of *immunoglobulin* (*Ig*) genes; and somatic hypermutation (SHM), which drives affinity maturation, leading to the production of high-affinity antibodies. SHM occurs in activated germinal center B cells engaging in an immune response and is initiated by activation-induced deaminase (AID), which catalyzes the deamination of cytosine in a single-stranded DNA (ssDNA) template, yielding a uracil residing within DNA.¹ During SHM, AID is thought to associate with and travel along with RNA polymerase 2 (Pol2), and Pol2-mediated transcription is believed to be the source of the ssDNA template, which is a strict requirement for SHM.²

SHM mainly introduces mutations in Ig heavy and light chain gene variable regions (*IqV*) at a rate of $\sim 10^{-3}$ mutations per base pair per cell division, which is 10⁶-fold higher than the spontaneous mutation rate in most genes.³ Moreover, the complementarity determining regions (CDRs) of IqV undergo a higher frequency of mutations than the intervening framework regions (FRs).^{4,5} This process is thought to contribute to the efficiency of affinity maturation, as mutations in the CDRs can alter the binding specificity and affinity of the antibody for antigens. However, AID has also been shown to deaminate non-lg loci, resulting in off-target mutations and chromosomal translocations that can contribute to B cell tumorigenesis.⁶ Nonetheless, the mutation frequencies in non-lg loci are typically much lower than those of the hypermutated IgV regions.⁶ In recent years, substantial progress has been made to identify the factors involved in SHM and define the molecular pathways that result in the introduction of mutations. By contrast, relatively little is understood about the mechanisms that specifically target AID and SHM to Iq loci, despite considerable efforts to address this issue.

Mutations are introduced by AID in *IgV* in a window of ~150–1500 bp downstream of the *Ig* transcription start site (TSS).⁶ The mechanisms responsible for targeting this window are still unknown. Although mutations can occur throughout *IgV* regions and their immediate flanking sequences, it has been shown that there is preferential targeting to RGYW and its reverse complement WRCY (where R denotes adenosine (A) or guanosine (G); Y denotes cytidine (C) or thymidine (T); and W denotes A or T), sequences that are directly targeted by AID and are referred to as

hotspot motifs.⁷ In in vitro assays, recombinant AID exhibits a strong preference for binding and deaminating cytosines within 5'-WRC-3' motifs present in ssDNA substrates. Curiously, some WRC motifs in *IgV* regions are strongly targeted by AID while others are not and the preferentially targeted motifs are almost always in the CDRs and not in the FRs⁵ (Fig. 1a). These findings led Wang et al. to hypothesize that other local sequences or higher-order structures, operating on a mesoscale (5–50 bp), also play a role in influencing the targeting of AID and SHM.⁸ In a recent paper published in *Cell*, Wang et al. report a significant advance in this area by demonstrating that flexible pyrimidine-pyrimidine dinucleotides surrounding AID deamination hotspot motifs facilitate binding to and deamination by AID, helping to explain preferential SHM in CDRs.⁸

Taking advantage of a large database of IqV sequences derived from out-of-frame Ig heavy chain gene rearrangements, Wang et al. established that CDRs have a higher density of WRC motifs than FRs and, importantly, that cytidines in WRC motifs mutate at a higher frequency in CDRs than FRs (Fig. 1a). They then discovered that this in vivo CDR > FR preference could be recapitulated using a simple in vitro deaminase assay containing purified AID and ssDNA substrates. This surprising observation demonstrated that AID and the ssDNA sequence are sufficient to dictate preferential mutability of WRC in CDRs vs FRs. Through clever high-throughput application of the in vitro assay, they established that the CDR > FR preference is broadly conserved in tetrapod species, with the interesting exceptions of horses and GALT (gut-associated lymphoid tissue) species, which use AID-mediated reactions to help generate their primary Ig repertoire.

What then is the special feature of CDR sequences that dictates robust AID activity? Through systematic in vivo and in vitro sequence mapping, Wang et al. provide evidence that the key feature is the intrinsic flexibility of the ~12 nucleotides upstream and ~6 nucleotides downstream of the WRC motif.⁸ High flexibility, conferred by enrichment for pyrimidine-pyrimidine (Py-Py) dinucleotides (which have a low base stacking free energy), results in high mutability, whereas low flexibility, conferred by a high proportion of Purine-Purine (Pu-Pu) pairs (with high base stacking free energies), yields low mutability. The dinucleotide composition of the six nucleotides immediately 5' of the WRC motif appears to be particularly important in this regard. Through mutagenesis and structural modeling of ssDNA binding to AID, Wang et al. provide evidence for a model in which substrate flexibility is important to allow the ssDNA flanks to engage in electrostatic interactions with two positively charged patches on the surface of AID, thereby stabilizing the WRC motif in

¹Department of Immunobiology, Yale School of Medicine, New Haven, CT, USA. ²²email: david.schatz@yale.edu



744

Fig. 1 Mesoscale sequence features promote AID activity. a Plot to illustrate the higher mutation frequency in WRC motifs in CDRs compared to FRs (artificial data created for illustration purposes only). **b** Model for mesoscale features in promoting AID deamination. WRC hotspot motifs in CDRs are often flanked by flexible Py-Py dinucleotides and are more likely to undergo hypermutation by AID. This is because Py-Py dinucleotides have a weak base stacking strength and facilitate the bending of ssDNA needed for binding to two positively charged patches on the surface of AID through electrostatic interactions. FR WRC motifs are typically flanked by stiff dinucleotides containing purines (Pu) and are mutated poorly.

close proximity to the active site (Fig. 1b). This model is consistent with and builds on findings from a landmark structural study of

AID bound to structured DNA substrates.⁹ Notably, insertion of a synthetic mesoscale Py-Py dinucleotide-rich motif adjacent to one of several poorly targeted WRC motifs was sufficient to strongly increase mutation of the motif.⁸

The findings of Wang et al. solve an important mystery in the field of SHM and have broad implications for the evolution of *IgV* sequences, the off-target sites of action of AID, and our ability to engineer sequences with desired mutational profiles, for example, to create broadly neutralizing antibodies for treatment or prevention of viral infections, as noted by the authors.⁸ In addition, their results have the intriguing mechanistic implication that AID's in vivo DNA substrate is a long (> 20 nt) stretch of ssDNA, which is notable given that the ssDNA region in the elongating Pol2 complex has been reported in many studies to be only 10–20 nt long.^{10,11} Future studies will be needed to validate and extend the model of Wang et al., particularly structural analyses of ssDNA–AID complexes and further studies of how mesoscale motifs affect AID targeting and SHM in the context of germinal center B cells and B cell lymphoma.

REFERENCES

- 1. Dickerson, S. K., Market, E., Besmer, E. & Papavasiliou, F. N. J. Exp. Med. 197, 1291–1296 (2003).
- 2. Pham, P. et al. Nucleic Acids Res. 47, 10815-10829 (2019).
- 3. Di Noia, J. M. & Neuberger, M. S. Annu. Rev. Biochem. 76, 1-22 (2007).
- 4. Wagner, S. D., Milstein, C. & Neuberger, M. S. Nature 376, 732 (1995).
- 5. Yeap, L. S. et al. Cell 163, 1124–1137 (2015).
- 6. Casellas, R. et al. Nat. Rev. Immunol. 16, 164-176 (2016).
- Kolchanov, N. A., Solov'ev, V. V. & Rogozin, I. B. Dokl. Akad. Nauk SSSR 281, 994–999 (1985).
- 8. Wang, Y. et al. Cell https://doi.org/10.1016/j.cell.2023.03.030 (2023).
- 9. Qiao, Q. et al. Mol. Cell 67, 361-373.e4 (2017).
- 10. Bernecky, C., Plitzko, J. M. & Cramer, P. Nat. Struct. Mol. Biol. 24, 809-815 (2017).
- 11. Barnes, C. O. et al. *Mol. Cell* **59**, 258–269 (2015).

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to David G. Schatz.

Reprints and permission information is available at http://www.nature.com/ reprints