

# Identifying protein–protein interactions in somatic hypermutation

Myron F. Goodman and Matthew D. Scharff

**Somatic hypermutation (SHM) in immunoglobulin genes is required for high affinity antibody–antigen binding. Cultured cell systems, mouse model systems, and human genetic deficiencies have been the key players in identifying likely SHM pathways, whereas “pure” biochemical approaches have been far less prominent, but change appears imminent. Here we comment on how, when, and why biochemistry is likely to emerge from the shadows and into the spotlight to elucidate how the somatic mutation of antibody variable (V) regions is generated.**

Wilson et al. (1) report in this issue that the error-prone DNA polymerase  $\eta$  is stimulated by the heterodimeric MSH2–MSH6 mismatch repair recognition complex. Humans deficient for polymerase (pol)  $\eta$  (2) or mice deficient in MSH2–MSH6 (3–6) show a significant reduction in hypermutation at A:T sites. This study, which shows a functional interaction between error-prone pol  $\eta$  and MSH2–MSH6, suggests that this low fidelity polymerase might be working hand-in-hand with the mismatch repair proteins to generate A:T mutations. At the very least, these data provide an impetus for more extensive biochemical analysis.

## Properties and pathways of SHM

SHM is characterized by specific types of base substitutions in the variable portion of immunoglobulin genes which occur at a million-fold higher frequency than normal somatic mutations ( $10^{-3}/\text{bp}$  versus  $10^{-9}/\text{bp}$ ). V region mutations at G:C and A:T sites occur on both strands of DNA in vivo. Mutations at G:C sites tend to be favored in DNA WRC hot spot motifs (WR: W = A or T, R = purine). These mutations are attributable to the C  $\rightarrow$  U deamination specificity of activation-induced cyti-

dine deaminase (AID; reference 7) acting either on single-stranded DNA (ssDNA) or on the nontranscribed strand within a moving transcription bubble (7–9). Mutations at A:T sites show a preference for WA hot spot motifs reminiscent of the in vitro behavior of pol  $\eta$ . Pol  $\eta$  is highly error prone and tends to generate AT  $\rightarrow$  GC mutations preferentially in TA motifs in vitro (10). The expression of AID, which occurs during a short time span in B cells, is essential for SHM as is active transcription of the V gene (11, 12). In fact, biochemical studies with semipurified AID provided the evidence that ssDNA was the substrate for AID, thus explaining the perplexing need for transcription (7–9, 13).

A synopsis of potential SHM pathways responsible for the V region mutations is shown in Fig. 1. C  $\rightarrow$  T mutations might be initiated by AID tracking along a moving transcription bubble and generating U residues on the nontranscribed DNA strand or, less often, on the transcribed DNA strand (7–9). Faithful copying of U by a high fidelity polymerase, such as pol  $\delta$  or pol  $\epsilon$ , would generate C to T transitions, whereas aberrant copying of U by a low fidelity polymerase, such as pol  $\eta$ , could cause both transitions (the replacement of a purine with a different purine and pyrimidine with a different pyrimidine) and transversions (the replacement of a purine with a pyrimidine or vice versa) (Fig. 1, left). However, this simplified picture cannot account for the equal numbers of mutations at C on the transcribed

strand, nor can it explain mutations at A:T sites.

Further SHM diversity could arise if the U residue created by AID is excised rather than copied. The removal of U residues by the enzyme uracil N-glycosylase (UNG) results in an abasic site that can then be removed by base excision repair. Alternatively, a mismatched U:G base pair can be repaired by mismatch repair. In this process, single base mismatches are recognized by an MSH2–MSH6 dimer which then recruits other proteins to excise the mismatch and replace the excised DNA. In either case, the repair patches—short in the case of base excision repair and much longer in the case of mismatch repair—would expose the transcribed strand to the action of pol  $\eta$  to generate mutations at A:T sites (Fig. 1, right). AID might also attack C residues on the transcribed strand.

## Biochemical issues and challenges

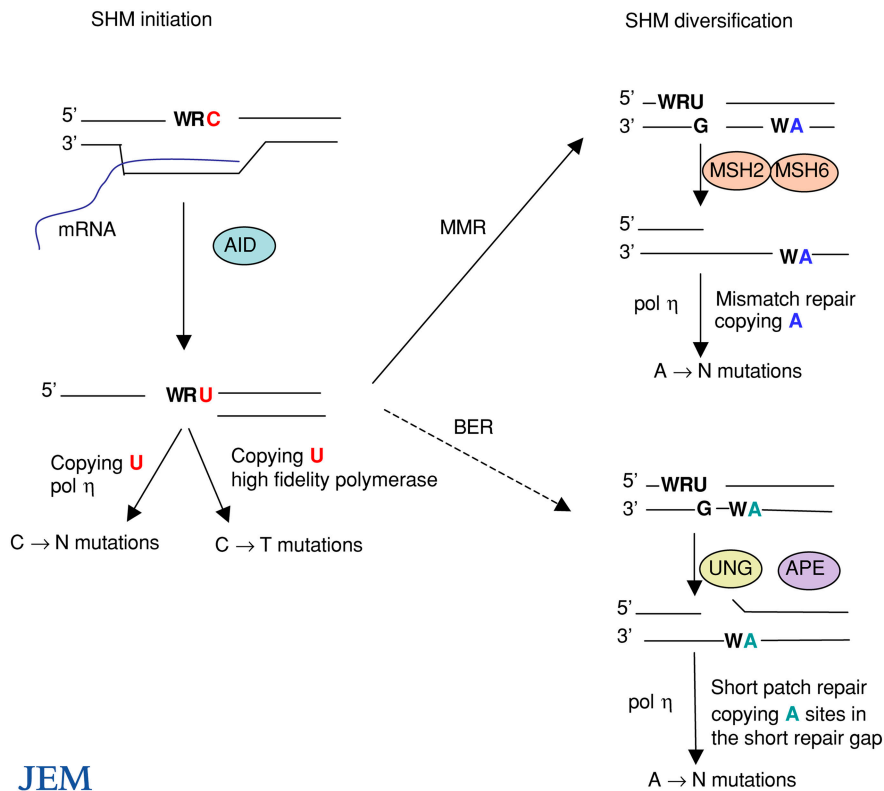
The “holy grail” from a biochemical perspective would be to reconstitute SHM entirely in a cell-free system using purified proteins. This possibility might not be all that remote given the availability of eukaryotic base excision repair, mismatch repair, and RNA pol II transcription systems in vitro. However, defining where pol  $\eta$  and AID might fit into any one of these systems is itself a difficult problem, and full reconstitution of SHM would appear to require integration of all of the components of the base excision repair, mismatch repair, and transcription systems—a daunting task. The reconstitution of the system is further complicated by the finding that deletions and mutations of the COOH-terminal end of AID result in the loss of class switch recombination (which requires AID-induced mutations in switch regions upstream from the constant regions, double-stranded DNA breaks, and recombination), but the pres-

M.F.G. is at University of Southern California, Los Angeles, CA 90089.

M.D.S. is at Albert Einstein College of Medicine, Bronx, NY 10461.

## CORRESPONDENCE

M.F.G.: mgoodman@usc.edu



## JEM

**Figure 1. AID-initiated SHM pathways.** AID acting on ssDNA exposed in the transcription bubble in an immunoglobulin *V* region deaminates C preferentially at WRC hot spot motifs. Subsequent copying of U by a high-fidelity DNA polymerase results in C → T transitions, or copying by error-prone pol  $\eta$  could generate C → N transition or transversion mutations. A second stage of diversification leading to mutations at A:T sites requires the action of either mismatch repair proteins, including MSH2–MSH6, or base excision repair proteins, including UNG and apurinic endonuclease (APE). Mutations at WA hot spot motifs could be made by pol  $\eta$  during long (mismatch repair) or short (base excision repair) gap-filling synthesis. The mismatch repair pathway (solid arrow) has been suggested to play a greater role than the base excision repair pathway (dashed arrow) in introducing mutations at A:T sites (20), although the relative importance of each pathway remains to be determined.

ervation of *V* region SHM (14, 15), whereas mutations at the NH<sub>2</sub>-terminal end of AID affect SHM but not class switch recombination (16). This suggested that proteins such as RNA pol II and replication protein A, which interact with AID (17, 18), and possibly other unidentified proteins, are required for the selective targeting of AID to certain parts of the immunoglobulin gene. A sensible point of departure, therefore, would be to take a step back and investigate individual protein–protein and protein–nucleic acid interactions.

#### MSH2–MSH6 stimulates pol $\eta$ activity

The paper by Wilson et al. (1) identifies a potentially relevant functional interaction between MSH2–MSH6 and pol  $\eta$ , proteins that have been shown

to be involved in SHM diversification (Fig. 1, right). The authors show that endogenous MSH2 binds to pol  $\eta$  in cell extracts, and that MSH2–MSH6 binds U:G mismatched base pairs, as has also been noted previously (19). The principal observation, however, is that the rate of pol  $\eta$ -catalyzed DNA synthesis is increased 2.4-fold in the presence of MSH2–MSH6. This increase reflects an enhancement in the catalytic efficiency ( $V_{\max}/K_m$ ) of pol  $\eta$  with no discernible change in polymerase processivity. Control experiments showed that DNA synthesis was unaffected in the presence of MSH2–MSH3 which recognizes larger mismatches, nor did MSH2–MSH6 stimulate pol  $\epsilon$ , an error-prone polymerase of unknown function. Admittedly, although the increase in DNA synthesis

is a modest 2.4-fold, it nonetheless may imply a coupling of mismatch repair and DNA synthetic functions—which could not be established solely by detecting protein–protein binding interactions using analyses such as chromatin immunoprecipitation.

A possible coupling of error-prone DNA synthesis with mismatch repair should be viewed in the broader context of determining the temporal fate of a U:G base misrepair. MSH2–MSH6 can bind to U:G (19) (Fig. 1, right), but competition for access to the U residue might also include high and low fidelity DNA polymerases vying with proteins involved in base excision repair. Each process, depending on when it occurs, could leave its own mutagenic signature. Because a heavy enzymatic traffic jam is likely to converge at the U residue once AID has acted, the hypermutagenic outcome depends on the enzyme queue. Do replication and repair enzymes encounter U residues in a temporally ordered process or in a random manner governed by mass action?

A recent paper by Rada et al. (20) addresses the temporal access to U by analyzing SHM and class switch recombination in mice deficient for both mismatch repair and base excision repair machinery. The data showed that A:T hypermutation in immunoglobulin *V* regions was eliminated entirely in mice lacking both MSH2 and UNG. It had been shown previously that the elimination of either mismatch repair (MSH2<sup>-/-</sup>) or base excision repair (UNG<sup>-/-</sup>) caused a reduction in mutations at A:T sites, with the loss of mismatch repair having a far greater effect (4, 21). It was concluded from that study that mismatch repair is primarily responsible for SHM at A:T sites with base excision repair used as a backup pathway (20). This suggests that AID-generated U residues do lead to a process in which MSH2–MSH6 competes with UNG for access to the U residue (Fig. 1, right).

Studies using the *Msh2*<sup>-/-</sup> *Ung*<sup>-/-</sup> double knock-out mice nicely defined the biochemical realities *in vivo* by, for example, eliminating a role for uracil

DNA glycosylases other than UNG and focusing our attention on the origins of A:T mutation (20). However, many questions cannot be addressed *in vivo*, in short-term culture, or in cell lines because molecules such as proliferating cell nuclear antigen and RNA pol II are critical for all cell processes. For example, a possible queuing mechanism directing protein access to U:G might involve the proliferating cell nuclear antigen. Along with its role in increasing the processivity of pol  $\delta$  during DNA replication, proliferating cell nuclear antigen is known to bind pol  $\eta$  (22) and is required for mismatch repair (23). It is possible that proliferating cell nuclear antigen facilitates binding of pol  $\eta$  with MSH2–MSH6 proximal to a U:G mismatch, thus coordinating the initiation of MMR with subsequent error-prone gap-filling synthesis. Now that we have learned that MSH2–MSH6 appears to stimulate pol  $\eta$  activity directly (1), we can begin to imagine how pol  $\eta$  might be selectively recruited to generate mutations at A:T sites as part of the mismatch repair machinery.

#### Help from biochemistry

A biochemical approach to reconstitute SHM using purified proteins should help us to advance beyond speculating on the molecular mechanisms. It might also make it possible to address the differences that govern the targeting of each of the repair mechanisms to *V* and switch region, since the different DNA substrates that characterize these regions could be used.

An attempt to model G:C and A:T hypermutation raises different questions and engenders different challenges. If we assume that both mutational processes originate from the action of AID on ssDNA, perhaps by tracking along the nontranscribed strand of a transcription bubble, as suggested from *in vitro* experiments using a bacterial T7 transcription assay (7–9), then how do mutations originate at the same frequency at C sites on the transcribed strand *in vivo* (24)? The chance that AID might attack a transcribed strand gap during mismatch repair or base excision repair seems unlikely because

*Msh2*<sup>-/-</sup>*Ung*<sup>-/-</sup> mutant and normal mice show a similar distribution of G:C-targeted mutations, except for the absence of transversions in mutant mice, which is presumably caused by the absence of UNG-generated abasic lesions (20). Instead the mammalian RNA pol II transcription apparatus, in contrast to T7, may provide greater access to the transcribed strand.

Thus, a mechanistic understanding of G:C and A:T hypermutation would benefit from studying AID in a mammalian transcription system for G:C mutations, and by studying pol  $\eta$  in conjunction with MSH2–MSH6-based mismatch repair for A:T mutations. The rationale for splitting the two classes of mutations into two distinct systems is based on the mouse data showing that *Msh2*<sup>-/-</sup>*Ung*<sup>-/-</sup> mice appear to mutate G:C but not A:T sites (20). However, if AID acting on ssDNA during transcription is responsible for triggering SHM (and class switch recombination), then what is responsible for confining these mutations to immunoglobulin *V* regions? Perhaps specialized transcription factors or chromatin accessibility restrict the targeting of AID to Ig*V* genes.

A biochemical approach may also help in resolving several extant controversies. One involves the possibility that AID may be acting on RNA *in vivo*, not on DNA (25). A second posits that antibody gene diversification requires the presence of UNG during class switch recombination, not for its ability to excise U, but rather to stabilize a protein complex needed for the mutational process (26). Arguments for (27) and against (28) this hypothesis have recently appeared in the literature. A third controversy involves the suggestion that endonuclease-catalyzed blunt end double-stranded DNA breaks, not AID, are responsible for initiating SHM (29, 30). It would seem that the burden of proof in each of these examples rests on showing that AID can use RNA as a substrate, that catalytically inactive UNG is an essential subunit of a complex whose activity is currently unknown, and that an endonuclease can be identified having a specificity com-

mensurate with the properties of class switching and mutagenesis.

As a means of addressing the targeting and trafficking of mutator proteins to immunoglobulin *V* but not *C* regions, and in investigating why one repair pathway or error-prone polymerase is chosen in preference to another, biochemistry should play a decisive role in deciphering the mechanisms of antibody diversification. To paraphrase Arthur Kornberg (31), we cannot be fully confident in our fundamental understanding of somatic hypermutation until the process has been reconstituted successfully using purified proteins with the ultimate goal to “capture it alive.”

#### REFERENCES

1. Wilson, T.M., A. Vaisman, S.A. Martomo, P. Sullivan, L. Lan, F. Hanoaka, A. Yasui, R. Woodgate, and P.J. Gearhart. 2005. MSH2–MSH6 stimulates DNA polymerase  $\eta$ , suggesting a role for A:T mutations in antibody genes. *J. Exp. Med.* 201:637–645.
2. Zeng, X., D.B. Winter, C. Kasmer, K.H. Kraemer, A.R. Lehmann, and P.J. Gearhart. 2001. DNA polymerase eta is an A–T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat. Immunol.* 2:537–541.
3. Phung, Q., D. Winter, A. Cranston, R. Tarone, V. Bohr, R. Fishel, and P. Gearhart. 1998. Increased hypermutation to G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J. Exp. Med.* 187: 1745–1751.
4. Rada, C., M.R. Ehrenstein, M.S. Neuberger, and C. Milstein. 1998. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity.* 9:135–141.
5. Li, Z., S.J. Scherer, D. Ronai, M.D. Iglesias-Ussel, J.U. Peled, P.D. Bardwell, M. Zhuang, K. Lee, A. Martin, W. Edelmann, and M.D. Scharff. 2004. Examination of Msh6- and Msh3-deficient mice in class switching reveals overlapping and distinct roles of mutS homologues in antibody diversification. *J. Exp. Med.* 200:47–59.
6. Martomo, S.A., W.W. Yang, and P.J. Gearhart. 2004. A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. *J. Exp. Med.* 200:61–68.
7. Pham, P., R. Bransteitter, J. Petruska, and M.F. Goodman. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature.* 424:103–107.
8. Ramiro, A.R., P. Stavropoulos, M. Janjovic, and M.C. Nussenzweig. 2003. Tran-

- scription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat. Immunol.* 4:452–456.
9. Bransteitter, R., P. Pham, P. Calabrese, and M.F. Goodman. 2004. Biochemical analysis of hyper-mutational targeting by wild type and mutant AID. *J. Biol. Chem.* 279:51612–51621.
  10. Rogozin, I.B., Y.I. Pavlov, K. Bebenek, T. Matsuda, and T.A. Kunkel. 2001. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. *Nat. Immunol.* 2:530–536.
  11. Maizels, N. 1995. Somatic hypermutation: how many mechanisms diversify V region sequences? *Cell.* 83:9–12.
  12. Peters, A., and U. Storb. 1996. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity.* 4:57–65.
  13. Chaudhuri, J., M. Tian, C. Khuong, K. Chua, E. Pinaud, and F.W. Alt. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature.* 422:726–730.
  14. Barreto, V., B. Reina-San-Martin, A.R. Ramiro, K.M. McBride, and M.C. Nussenzweig. 2003. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol. Cell.* 12:501–508.
  15. Ta, V.T., H. Nagaoka, N. Catalan, A. Durandy, A. Fischer, K. Imai, S. Nonoyama, J. Tashiro, M. Ikegawa, S. Ito, et al. 2003. AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat. Immunol.* 4:843–848.
  16. Shinkura, R., S. Ito, N.A. Begum, H. Nagaoka, M. Muramatsu, K. Kinoshita, Y. Sakakibara, H. Hijikata, and T. Honjo. 2004. Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat. Immunol.* 5:707–712.
  17. Nambu, Y., M. Sugai, H. Gonda, C. Lee, T. Katakai, Y. Agata, Y. Yokota, and A. Shimizu. 2003. Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science.* 302:2137–2140.
  18. Chaudhuri, J., C. Khuong, and F.W. Alt. 2004. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature.* 430:992–998.
  19. Wang, H., C.W. Lawrence, G.-M. Li, and J.B. Hays. 1999. Specific binding of human MSH2-MSH6 mismatch-repair protein heterodimers to DNA incorporating thymine- or uracil-containing UV light photoproducts opposite mismatched bases. *J. Biol. Chem.* 274:16894–16900.
  20. Rada, C., J.M. Di Noia, and M.S. Neuberger. 2004. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell.* 16:163–171.
  21. Rada, C., G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, and M.S. Neuberger. 2002. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.* 12:1748–1755.
  22. Haracska, L., R.E. Johnson, I. Unk, B. Phillips, J. Hurwitz, L. Prakash, and S. Prakash. 2001. Physical and functional interaction of human DNA polymerase  $\eta$  with PCNA. *Mol. Cell. Biol.* 21:7199–7206.
  23. Umar, A., A.B. Buermeier, J.A. Simon, D.C. Thomas, A.B. Clark, R.M. Liskay, and T.A. Kunkel. 1996. Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell.* 87:65–73.
  24. Milstein, C., M.S. Neuberger, and R. Staden. 1998. Both DNA strands of antibody genes are hypermutation targets. *Proc. Natl. Acad. Sci. USA.* 95:8791–8794.
  25. Honjo, T., M. Muramatsu, and S. Fagarasan. 2004. AID: how does it aid antibody diversity? *Immunity.* 20:659–668.
  26. Begum, N.A., K. Kinoshita, N. Kakazu, M. Muramatsu, H. Nagaoka, R. Shinkura, D. Biniszkiwicz, L.A. Boyer, R. Jaenisch, and T. Honjo. 2004. Uracil DNA glycosylase activity is dispensable for immunoglobulin class switch. *Science.* 305:1160–1163.
  27. Begum, N., and T. Honjo. 2004. Uracil DNA Glycosylase activity is dispensable for immunoglobulin class switch. *Science.* 305:1160–1163.
  28. Stivers, J.T. 2004. Comment on “Uracil DNA glycosylase activity is dispensable for immunoglobulin class switch”. *Science.* 306:2042.
  29. Papavasiliou, F.N., and D.G. Schatz. 2000. Cell-cycle-regulated DNA double-stranded breaks in somatic hypermutation of immunoglobulin genes. *Nature.* 408:216–221.
  30. Zan, H., X. Wu, A. Komori, W.K. Holloman, and P. Casali. 2003. AID-dependent generation of resected double-strand DNA breaks and recruitment of Rad52/Rad51 in somatic hypermutation. *Immunity.* 18:727–738.
  31. Kornberg, A. 1989. For the Love of Enzymes: The Odyssey of a Biochemist. Harvard University Press, Cambridge, MA. 336 pp.