

REVIEW

Antibody diversification by somatic mutation: from Burnet onwards

Michael S Neuberger

The clonal selection theory proposed by Burnet required a genetic process, for which there was then no precedent, which randomizes the region of the gene(s) responsible for the specification of γ -globulin molecules. Work over the subsequent half-century substantiated Burnet's speculation, revealing two distinct novel genetic processes. During early development (when Burnet first thought the randomization took place) programmed gene segment rearrangement catalysed by the RAG1/RAG2 recombinase generates a substantial diversity of immunoglobulin molecules (the primary repertoire). Somatic hypermutation (triggered by the activation-induced deaminase (AID) DNA deaminase) then occurs following antigen encounter in man and mouse, yielding a secondary repertoire. This hypermutation allows both limitless diversification as well as maturation of the antibody response by a process of somatic evolution akin to that envisioned by Burnet in later formulations of the clonal selection theory. AID-triggered antigen receptor diversification probably arose earlier in evolution than RAG-mediated repertoire generation. Here I trace our insights into the molecular mechanism antibody somatic mutation from when it was first proposed through to our current understanding of how it is triggered by targeted deamination of deoxycytidine residues in immunoglobulin gene DNA.

Immunology and Cell Biology (2008) **86**, 124–132; doi:10.1038/sj.icb.7100160; published online 8 January 2008

Keywords: clonal selection theory; somatic mutation; DNA deamination; antibody diversification

'The theory requires at some stage in early embryonic development a genetic process for which there is no available precedent. In some way we have to picture a 'randomization' of the coding responsible for part of the specification of gamma globulin molecules.' (Burnet, 1957).

ANTIBODY DIVERSIFICATION AS ORIGINALLY VIEWED BY BURNET

In his October 1957 formulation of the clonal selection theory,¹ Burnet assumed that the production of a repertoire of γ -globulin molecules required a genetic process for the randomization of their specification that took place at an early stage in embryonic development. Interestingly, he seems to have assumed that a somatic gene diversification process was needed, presumably excluding the possibility that the germline could provide sufficient γ -globulin diversity (Figure 1). The article gives no indication that Burnet had any great interest in speculating about the molecular mechanism responsible for the randomization of the coding responsible for part of the specification of γ -globulin molecules, other than noting that it required a genetic process for which there was no available precedent. Indeed, especially in light of later evidence of his limited affection for molecular biology,² one might well have thought that the molecular mechanism of γ -globulin randomization was an aspect of the clonal selection theory which Burnet considered of limited interest as

compared to the other aspects and implications of the theory. This may indeed have been the case. However, as I first learned at the fiftieth anniversary meeting, Burnet had indeed given some thought to the possible mechanism of coding randomization. This is evidenced by some text and a diagram in a first draft of his 1957 article (Figure 2a), material that did not make it into the final published version (see article by Baxter in this volume). This aspect of his speculation is no less prescient than the many others discussed in his short but wide-ranging and farseeing paper.

DEVELOPMENT BY LEDERBERG AND SOMATIC EVOLUTION

Lederberg³ arrived at the WEHI in August 1957, staying there for 3 months on a trimester's Fulbright fellowship. He came to the WEHI in order to learn from Burnet about influenza virus and its recombinational processes, only to learn on arrival that Burnet had closed down his flu work in order to work full blast on the mechanism of antibody synthesis. It was clearly a period of exciting discussions with Lederberg subsequently noting³ that their 'discussions became intense, although somewhat clouded by Burnet's tendency to resist the 'simplistic' mechanisms of DNA-based molecular genetics that are today's foundation stone. I would receive his exciting ideas, and then have to translate them into a contemporary idiom to get the full benefit of his marvellous biological intuition.' During the following year, Lederberg came to the conclusion that Burnet's ideas would

receive greater due if they were retranslated into DNA language.⁴ Lederberg used the occasion of giving the Howard J Mueller memorial lecture at Harvard in November 1958 to frame a reformulation of the clonal selection theory with this lecture forming the basis of a manuscript that he submitted to Science and which was published in June 1959.⁵ It appeared back-to-back with an article by Talmadge⁶

(as requested by the two authors), in which Talmadge developed his own ideas on clonal selection that he had first briefly alluded to in his 1957 article in the Annual Reviews of Medicine⁷ (and which had been acknowledged by Burnet¹).

Lederberg's 1959 paper⁵ takes the discussion of the possible molecular mechanism of antibody diversification somewhat further than that was developed in Burnet's short 1957 article¹ (Figure 2b). While Burnet had originally envisioned that the genetic process driving γ -globulin randomization would only occur at an early stage in embryonic development, Lederberg proposed that this randomization was a consequence of a high frequency of somatic mutation that took place throughout lymphocyte proliferation. This was a fundamental distinction. The scenario originally envisioned by Burnet had the advantage that it allowed antibody diversification to occur prior to the period of induced self tolerance. The scenario envisioned by Lederberg allowed for a continuous neo-Darwinian maturation and evolution of antibody specificities by an iterative alternation of gene diversification and antigen-mediated selection. This appealed to Burnet, who incorporated such thinking into his later formulations of the clonal selection theory.⁸

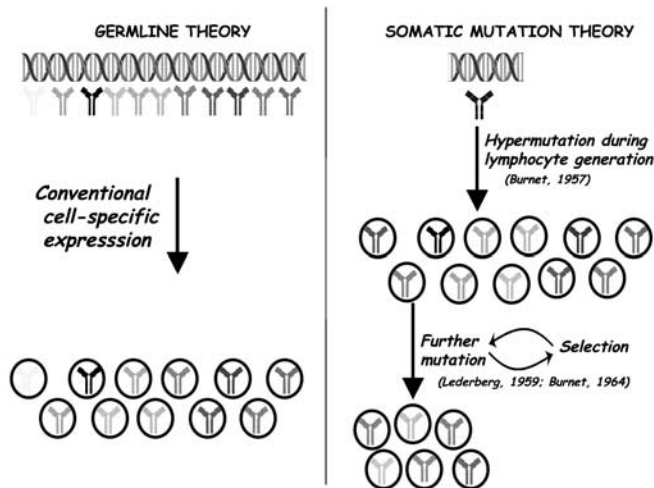


Figure 1 Comparison between germline and somatic mutation theories of antibody diversification. Somatic mutation (illustrated in an extreme form with the antibody repertoire arising through diversification of a single germline gene) was originally envisioned by Burnet¹ to occur only during early embryonic development but was later speculated to occur throughout (lymphocyte) proliferation,^{5,8} thereby allowing continuous maturation of the antibody response through a process of Darwinian somatic evolution.

IMMUNOGLOBULIN POLYPEPTIDE SEQUENCES TRIGGER NEW MODELS

However, the application of protein sequencing methodology to myeloma proteins during the early 1960s provided information about the structural nature of antibody diversity which was hard to reconcile with the idea of antibody diversity simply being a result of a high frequency of genome-wide mutation during lymphocyte expansion.^{9–11} Antibody diversity was found to be largely restricted to the N-terminal domains of the immunoglobulin (Ig) heavy and light chains with the remainder of the polypeptide chain being composed of one of a small number of constant regions.

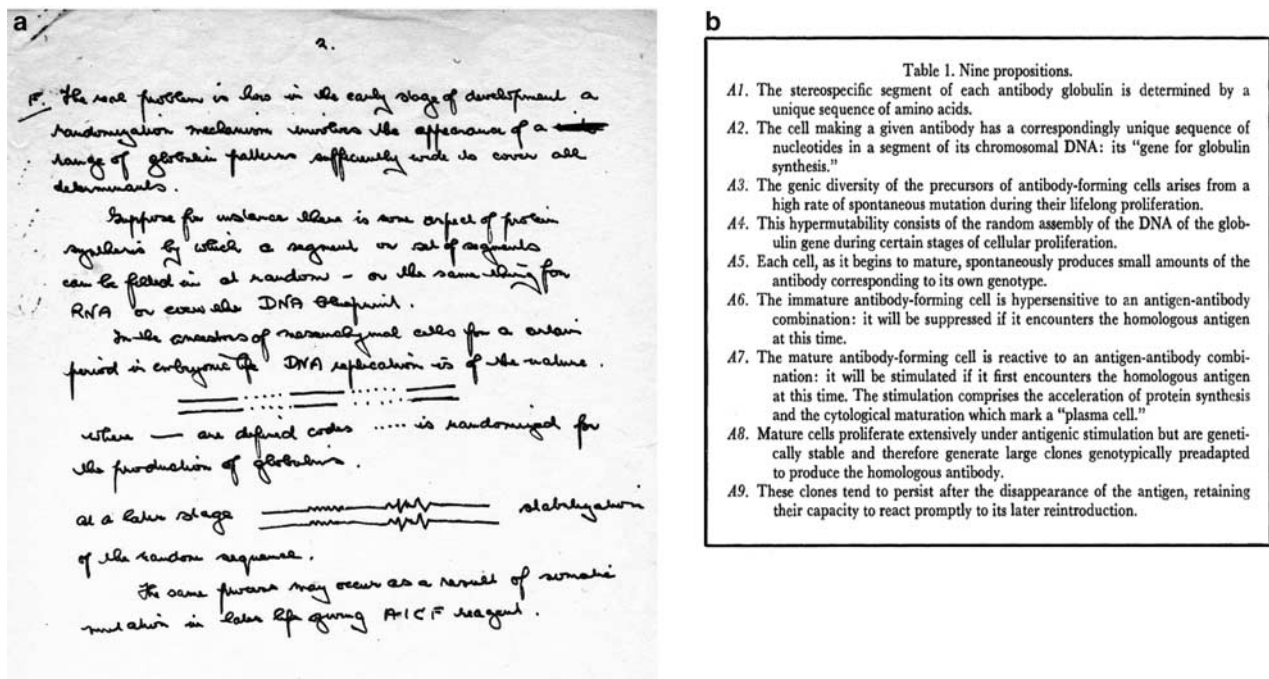


Figure 2 Early thoughts about somatic hypermutation. (a) Reflection on possible mechanisms was included in a draft of Burnet's first paper on the clonal selection theory but this was not included in the final, published paper.¹ (Reproduced with kind permission of Elizabeth Dexter.) (b) Lederberg⁵ developed the idea that the genic diversity of antibodies reflects a high rate of spontaneous mutation during the lifelong proliferation of the antibody-forming cell precursors. (Reproduced from ref.⁵)

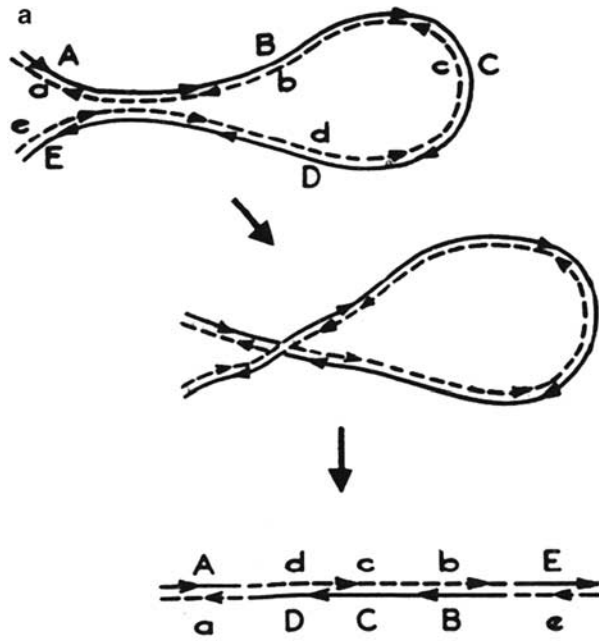


Fig. 1. A diagram representing the consequences of an intra-chromatid inversion. The complementary strands of DNA are represented by solid and dotted lines; their directional senses are shown by arrows. The inversion causes a portion of the 'dotted' strand to become continuous with the 'solid' strand. The capital and small letters are used to represent complementary sequences of bases

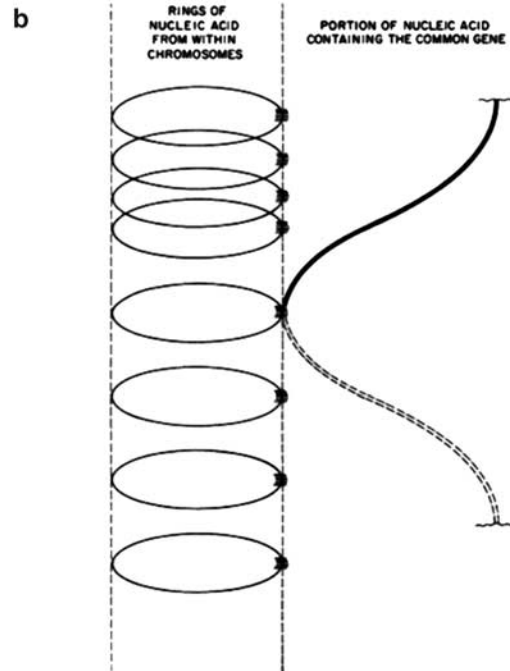


Fig. 3.—Diagram of the proposed genetic mechanism which accounts for the amino acid sequence variations found in L-chains. Genetic material which codes for the "variable" portion of L-chain molecules is inserted into that which codes for the "common" region of amino acid sequence by a mechanism similar to the insertion of the λ -virus into a bacterial chromosome (see text).

Figure 3 Some early models envisioning a role for DNA recombination/rearrangement in the generation of antibody diversity. (a) Smithies¹² put forward a model involving inversional recombination between homologous repeated sequences. (Reproduced from ref.¹²) (b) Dreyer and Bennett¹³ proposed that DNA encoding the 'variable' portion of immunoglobulin (Ig) chains is inserted by DNA rearrangement into the region coding for the 'common' (that is, constant) region. (Reproduced from ref.¹³)

Alternative explanations needed to be formulated in order to explain antibody diversity in light of the accumulating amino-acid sequence information. Several possible mechanisms were put forward during the 1960s. In 1963, Smithies¹² proposed a mechanism of gene diversification based on multiple recombinations between homologous sequences—a mechanism somewhat akin to the gene conversion process now known to contribute to IgV diversification in chicken as well as some other species (Figure 3a). Dreyer and Bennett,¹³ while acknowledging that such recombination, possibly allied to a hypermutation process, could provide a plausible account for antibody diversity, rated it unlikely. Instead, they speculated that a DNA rearrangement process might be involved, citing the incorporation of bacteriophage- λ into the *Escherichia coli* genome as providing a precedent for DNA rearrangement (Figure 3b). In contrast, Brenner and Milstein¹⁴ proposed that a hypermutation process could still be invoked to account for antibody diversity if the proposals of Lederberg were suitably modified. Thus, they resurrected Lederberg's earlier ideas but modified the model by envisioning that, at some stage during the differentiation of antibody-producing cells, an unspecified cleaving enzyme would cut the DNA in the region of the genome, which specified the Ig variable region. Subsequent DNA degradation (for example, by an enzyme akin to bacterial exonuclease III) and repair by an error-prone DNA polymerase (as had been described for polymerases encoded by bacteriophage T4 variants) would then lead to localized mutagenesis (Figure 4a). So again, it was knowledge of the genetic processes in bacteria and bacteriophage, which provided the precedent for the proposed mechanism.

These were not the only possible mechanisms being considered that might account for antibody diversity. Others considered a role for joining at the protein or RNA level. Many of the proponents of the different models were brought together at the 1967 Cold Spring Harbor meeting on antibodies. In his summing up of this meeting, Jerne¹⁵ congratulated Burnet on his stimulating ideas, noted the vindication of the clonal selection theory witnessed by the presentations at that meeting and went on famously to conclude, 'As this younger generation of professionals is pressing rapidly towards the definitive solution of the antibody problem, we older amateurs had perhaps better sit back, waiting for the End'. In fact, the end did not come that rapidly. The tools necessary to achieve the solution were still wanting. It was the development of recombinant DNA technology that was to facilitate the unveiling of the molecular mechanisms of antibody gene diversification. Perhaps, it was the absence of the techniques needed to identify the antibody diversification mechanism that resulted in the high level of theoretical discussion and speculation. It is notable that many of the key post-Burnet (1957) papers discussing possible diversification mechanisms (for example, those by Lederberg, Smithies, Dreyer and Bennet, Brenner and Milstein) were published in major journals but without linking the speculation to the provision of any fresh experimental data. Indeed, Brenner and Milstein specifically noted in their letter to *Nature*,¹⁴ 'We cannot propose a definitive test of this theory apart from the direct demonstration of a cleaving enzyme and a mutagenic polymerase in precursor cells, which may not be easy to do.' They were right; it was to be more than three decades before activation-induced deaminase (AID) and the relevant DNA polymerases were identified.

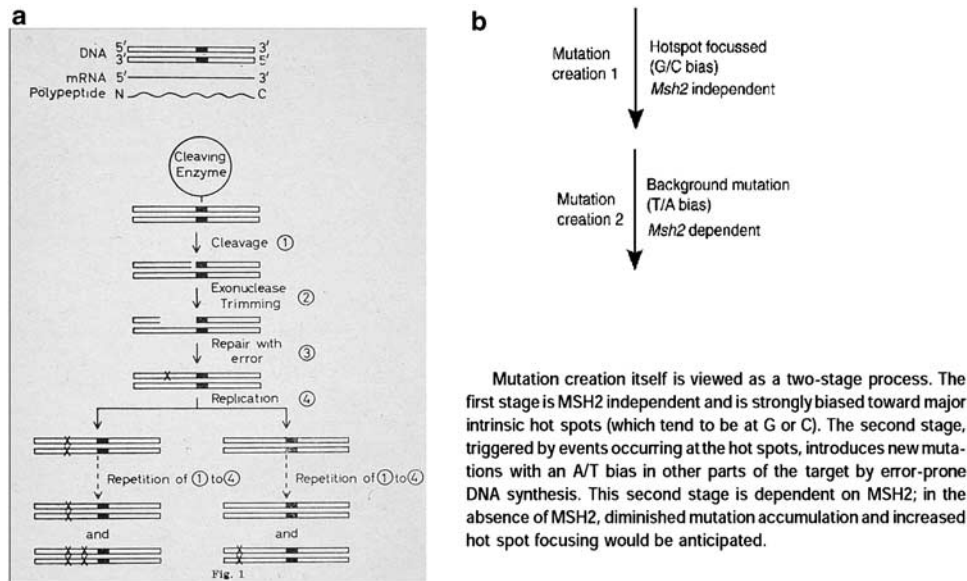


Figure 4 Development of thinking about somatic mutation mechanisms. **(a)** Brenner and Milstein¹⁴ proposed that localized genetic diversity could be achieved through repair (using an error-prone polymerase) of a DNA break generated by targeted endonucleolytic cleavage. (Reproduced from ref.¹⁴) **(b)** Rada *et al.*³⁶ proposed that the mutation spectra in repair-deficient mice and human B-cell lines could best be accounted for by mutation generation occurring in two phases; the first preferentially at G:C-biased hotspots with mutations at A:T pairs being generated in a second phase that is triggered by MSH2-mediated recognition of lesions created in the first phase. (Reproduced from ref.³⁶)

GENE REARRANGEMENT

Following the development of recombinant DNA technology, rapid progress was achieved with the characterization of the Ig loci and the identification of the central contribution of gene rearrangement to antibody diversity.¹⁶ The multiplicity of V, D and J gene segments, the process of their assortment and the incorporation of nucleotide insertions and deletions at the borders of gene segment integration, all contributed to a substantial antibody repertoire, accounting (at least in man and mouse) for the early diversification originally envisioned by Burnet. Subsequently, the lymphocyte-specific enzymes critical for the process were identified: RAG1/RAG2 for the gene rearrangement^{17,18} and terminal deoxynucleotidyl transferase¹⁹ for the non-templated nucleotide insertions at the sites of gene segment fusion.

SOMATIC MUTATION AND GENE CONVERSION REMAIN IMPLICATED

It turned out, however, that the amount of diversity generated by gene rearrangement varied considerably between species as well as between loci. Thus, whereas gene rearrangement is responsible for substantial diversity among IgH and IgK chains of man and mouse, the mouse Igλ locus has only few gene segments with rearrangement yielding a very small repertoire of polypeptide sequences.²⁰ Even more striking is the situation in the chicken. Here, rearrangement at both heavy and light chain loci yields practically no diversity, with the dominant contribution to diversity being achieved by post-rearrangement gene conversion in which a panel of IgV pseudogenes that flank the single functional IgV gene are used as donors for a patchwork of gene conversion following V_λ-J_λ or V_{DH}-J_H integration.²¹

However, analysis of the mouse Igλ locus²⁰ not only provided the first direct evidence for antibody gene rearrangement: it also gave a substantial boost to the proposal that diversity was also at least in part achieved through somatic hypermutation. Thus back in 1970, Weigert and colleagues²² had used a comparison of amino-acid sequences of

mouse Igλ chains to argue powerfully for a major role for somatic mutation in antibody diversification. This proposal was strongly supported when the limited diversity of the mouse germline IgVλ segments became apparent from gene cloning.

It was application of hybridoma technology that allowed the very major contribution of somatic mutation to the maturation of the immune response to be revealed. The completion of organismal genome sequences has enabled somatic mutations in individual antibodies to be unambiguously assigned, confirming its dominant contribution to antibody diversity. In mouse, it is clear that the primary Ig repertoire is solely generated by gene rearrangement with somatic mutation being responsible for the diversification that occurs subsequent to encounter with foreign antigen (Figure 5). However, this divvying up of roles with rearrangement being responsible for primary repertoire generation and somatic mutation acting to drive subsequent maturation of specific responses is not a ubiquitous pattern. Somatic mutation has been shown to make a major contribution to the generate of the primary Igλ repertoire in sheep and a role for somatic mutation in primary repertoire generation has also been proposed for a subset of human B cells.^{23,24}

FEATURES AND GENETIC REQUIREMENTS OF SOMATIC MUTATION

Despite this advancing knowledge revealing the importance of somatic mutation in antibody diversification, understanding of the molecular mechanism of somatic mutation was slow to develop. The idea that it was due to some form of localized error-prone DNA synthesis (along the lines proposed by Brenner and Milstein in 1966¹⁴) persisted for many years, though without much supportive evidence. Major insights nevertheless emerged as the amount of DNA sequence information increased. Comparison of the Ig sequences in dynastically related hybridomas showed that mutation occurred in a stepwise manner with roughly one mutation being incorporated over the IgV domain in each cellular generation following antigen triggering.²⁵

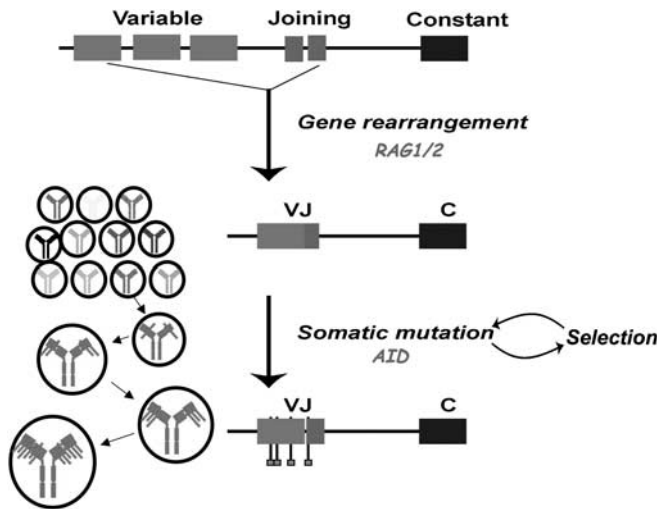


Figure 5 Repertoire generation in man and mouse occurs in two main stages. A primary, antigen-independent repertoire is generated by RAG-catalysed V-(D)-J joining; the secondary (antigen-dependent repertoire) occurs by AID-dependent somatic mutation, underpinning affinity maturation of the response. For simplicity, diversity (D) segments are not depicted.

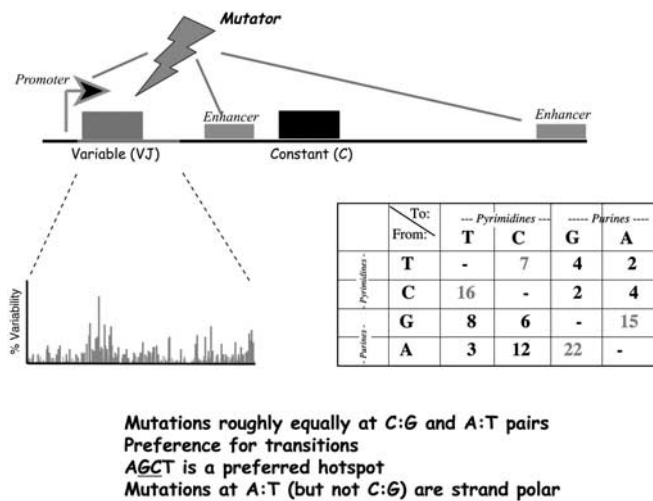


Figure 6 Features of antibody hypermutation.

Early studies on genomic DNA in myelomas and hybridomas revealed that mutation was not restricted to the IgV coding region but that the mutation domain extended over some 2 kb, starting shortly downstream of the IgV promoter.^{26,27} The discovery that somatic mutation could be recapitulated on Ig transgenes²⁸ coupled to the development of PCR enabled information about mutation patterns as well as the requirements for *cis*-acting DNA sequences to be rapidly gleaned²⁹ (Figure 6). It became clear that the transcription regulatory elements (but not the IgV coding sequence itself) played a major role in mutation recruitment with the location of the domain being determined by the position of the IgV promoter.³⁰ The nucleotide substitutions themselves were shown to occur roughly equally at C:G and A:T pairs, though with a bias towards transitions over transversions. Within the IgV, the distribution of nucleotide substitutions was shown to exhibit strong intrinsic biases. This could be seen even using target sequences that were not a focus for antigen-mediated selection. Certain positions within the V were scarcely ever targeted for

mutation whereas others constituted favoured hotspots, with major hotspots usually conforming to the sequence AGCT (with substitutions occurring at the G or C).^{31–33} (While major hotspots conformed to the consensus, not all sequences within the IgV that conform to the consensus constitute hypermutation hotspots.) The major hotspots are not randomly distributed with analysis of IgV sequences strongly suggesting that germline IgV sequences have evolved so as to conserve strategic targeting of somatic mutation, presumably so as to optimize affinity maturation of the immune response.³⁴

All this information, while illuminating, did not in itself identify the molecular mechanism of the hypermutation process. Attempts were made to gain insight by ascertaining whether mutation was perturbed in mice or humans deficient in various DNA repair proteins. By and large, such repair deficiencies either had no effect on mutation spectra or, alternatively, impeded productive V-D-J joining and were therefore similarly uninformative regarding hypermutation. However, there were exceptions. In particular, deficiency in the mismatch recognition protein MSH2 was shown to lead to a diminished proportion of mutations at A:T pairs,^{35,36} a phenomenon similarly observed in individuals deficient in DNA polymerase- η .³⁷

DISCOVERY OF AID

In order to facilitate mechanistic investigation, renewed attempts were made to obtain cell lines in which antibody hypermutation was ongoing. Success was eventually achieved with the most convincing examples being provided by several Burkitt lymphoma lines.^{38,39} An intriguing feature of this mutation was, again, that it was found to be heavily biased towards C:G pairs.³⁹

Nevertheless, it was an analysis of a cell line that performed class-switch recombination (rather than hypermutation) in culture that ultimately led to molecular cloning of the mutator. Honjo and colleagues^{40,41} identified mRNAs that were specifically induced on stimulating a mouse B-cell line to undergo switch recombination *in vitro* and went on to show by use of gene targeting that one of them (which encoded a B-cell-specific protein that he designated AID) was necessary not only for switch recombination but also for somatic hypermutation. Durandy and colleagues⁴² revealed that several patients with hyper-IgM syndrome (that is, humans who were deficient in Ig class switching) carried mutations in AID. The ability to perform directed gene targeting in the chicken DT40 cell line facilitated the demonstration that AID was also necessary for gene conversion.^{43,44}

The predicted amino-acid sequence of AID revealed that it belonged to the superfamily of zinc-dependent deaminases (hence its original designation as activation-induced deaminase). In particular, its closest homologue was APOBEC1 (apolipoprotein B-editing catalytic subunit 1), an enzyme that forms part of a protein complex which edits apolipoprotein B RNA transforming the glutamine codon CAA at position 6666 into a termination codon UAA, thereby triggering tissue-specific production of a truncated form of apolipoprotein B. With this paradigm in mind, it was initially suggested that AID also acted as an RNA-editing enzyme with the possibility that this editing would lead to the production of a DNA cleaving enzyme that could act as a trigger not only for class-switch recombination but also (in keeping with the 1966 Brenner–Milstein model) for somatic mutation.⁴¹

A DNA DEAMINATION MECHANISM EXPLAINS PREVIOUS FINDINGS

It was, however, the integration of the discovery of AID into the large amount of information that had already been accumulated on

post-rearrangement antibody diversification that led to unveiling of the DNA deamination mechanism of antibody diversification (Figures 7 and 8). Indeed, the discovery of AID provided the missing piece of the jigsaw, which when interpreted in terms of a DNA deamination mechanism, made sense of several earlier disparate findings.⁴⁵

Most salient is the fact that the C:G-biased mutation in MSH2-deficient mice had previously been taken to indicate that somatic mutation is initiated by diversification at C:G pairs with MSH2-mediated recognition then triggering a subsequent stage of diversification at A:T pairs.³⁶ Such a two-stage model was supported by the observation that C:G-biased mutation was also a characteristic of somatic mutation in B-cell lines³⁹ and had also been noted in the frog⁴⁶ as well as in a mutating mouse pre-B-cell line⁴⁷ (although the relevance of the mutation in the pre-B line to physiological antibody hypermutation had been controversial; Figure 4b). While a process that initially targets mutations to C:G pairs and only subsequently mutates A:T pairs is not readily compatible with a simple error-prone polymerase mechanism in the mould of the Brenner–Milstein model, I am not aware that any significant thought had been given prior to the discovery of AID of how such C:G-targeted mutation might be achieved. However, once AID was in the picture, direct deamination of cytidine in DNA should have provided a ready explanation.

Although the phenotype of AID-deficient mice and humans provided a dramatic pointer towards switch recombination and somatic mutation exhibiting some commonality in their early steps, there were also earlier hints in that direction coming from other sources. Thus, not only were the intrinsic hotspots of somatic hypermutation known to share a consensus sequence that was essentially identical to one of the constituent motifs of S regions; it had also been noted that MSH2-deficiency caused a refocusing of switch breakpoints in a manner analogous to the refocusing of somatic hypermutation.⁴⁸ Similarly, it had been speculated several years previously that Ig gene conversion and somatic mutation might constitute alternative ways of processing a common initiating DNA lesion,^{49,50} a suggestion that obtained strong support from the demonstration that the overall pattern of IgV gene diversification in chicken DT40 cells could be shifted from one of gene conversion to one of C:G-hypermutation by inactivating genes involved in recombinational repair.⁵¹ Interestingly, however, the preference for nucleotide transitions that was evident in *in vivo* antibody in man and mouse was not retained in the C:G-biased IgV hypermutation that occurred in recombination-deficient DT40 cells.

Although there was no precedent for any physiological programme in vertebrates that depended on targeted deamination of DNA, invoking such a mechanism of AID action immediately suggested explanations for these disparate observations that had previously remained as isolated, unexplained experimental findings.

TESTING THE DNA DEAMINATION MECHANISM

The DNA deamination scheme⁴⁵ made many predictions, several of which were readily testable. It did, of course, predict that AID could act to deaminate cytidine in DNA. This was confirmed using both bacterial genetic and biochemical assays.^{45,52–57} The enzyme has, however, been relatively refractory to detailed biochemical assay and structural determination since intact active recombinant protein is not readily produced in high yield. Nevertheless, it is evident that recombinant AID as well as AID rescued from B cells will deaminate cytidine in the context of a single-stranded DNA substrate. This obviously raises questions about how the Ig target is rendered single-stranded *in vivo* in order for AID to gain access; an association

SOMATIC HYPERMUTATION

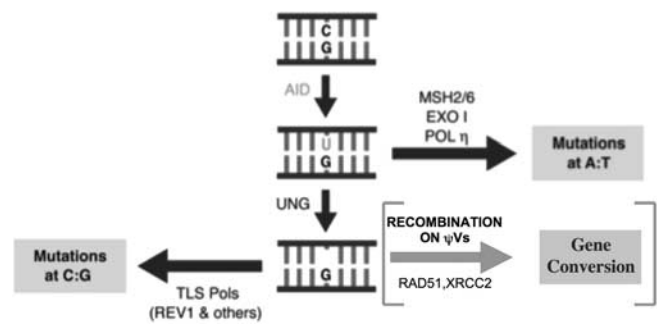


Figure 7 Mechanism of somatic mutation. AID-mediated cytosine deamination in the IgV creates U:G mismatches. Replication across the resultant UNG-generated abasic site leads to nucleotide substitutions at C:G pairs. Recognition of the U:G mismatch by MSH2/MSH6 yields to an error-generating patch repair, which generates mutations predominantly at A:T pairs. In chicken, gene conversion results from recombinational repair (templated on the IgV pseudogenes) of the UNG-generated abasic site.

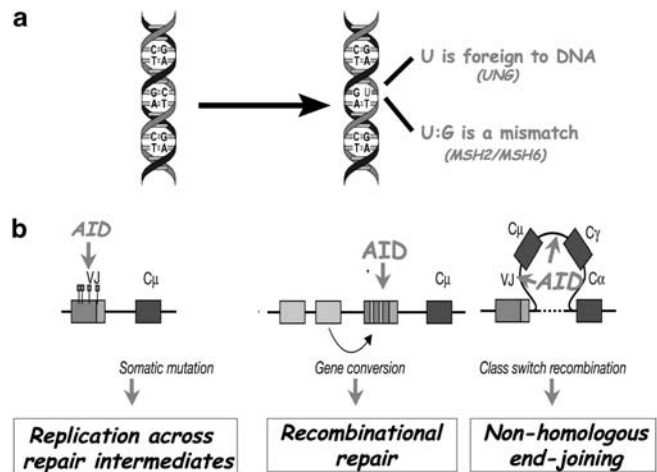


Figure 8 AID-mediated DNA deamination triggers multiple pathways of antibody diversification. (a) AID acts by deaminating cytosine in DNA to uracil, generating a U:G mismatch that is recognized by components of both base-excision (UNG, uracil-DNA glycosylase) and mismatch (MSH, homologue of bacterial mutS) pathways. (b) Targeted deamination in the IgV leads to lesions which are resolved by forms of lesion bypass (yielding somatic hypermutation) or recombinational repair (yielded IgV gene conversion). (see Figure 7 for more detail.) Targeted deamination in the vicinity of the switch regions appears to lead (following UNG action) to double-strand DNA breaks, which result in class-switch recombination following non-homologous (or, in some cases, microhomology-mediated) end-joining.

with the single-stranded protein RPA appears to be at least a significant part of the story.⁵⁸ AID does not target all cytidines equally but rather shows preferences that accord with a preference for a 5'-WRC substrate where R=purine and W=A/T. The intrinsic hotspots of antibody hypermutation (WRCY, where Y=pyrimidine with AGCT being the preferred exemplification) might then be explained by different subunits of an AID oligomer acting on a substrate that contains overlapping WRC motifs on opposite DNA strands.⁵⁹

While biochemistry revealed that AID could act on cytidine in DNA, genetics provided strong support for such targeted DNA deamination constituting the trigger for antibody diversification. A clear prediction of the DNA deamination scheme is that deficiency in

uracil excision should perturb the diversification pathways. This is indeed the case. Inhibition or deficiency of uracil-DNA glycosylase (UNG) in DT40 cells caused a shift in IgV diversification from gene conversion to transition mutations at C:G pairs.⁶⁰ UNG-deficiency in mice also leads to a predictable perturbation of antibody diversification although the block on switch recombination is incomplete since a pathway that appears to be based on MSH2/MSH6-mediated recognition of the AID-generated initiating U:G lesion appears to provide a back up.^{61,62} (Interestingly, this MSH2/MSH6-triggered pathway seems to provide less backup switching in man since UNG-deficiency leads to a robust hyper-IgM phenotype.⁶³ However, it is possible that the apparent difference between mouse and man might lie less in a species difference and more with polymorphisms within species coupled to an ascertainment bias.) The DNA deamination scheme envisages that mutations at A:T pairs are dependent on MSH2-mediated recognition of the U:G mismatch and, supporting this, mutations at A:T pairs are diminished in mice deficient in MSH6 but not in those deficient in MSH3, consistent with the expectation that a single-base mismatch would be recognized by the MSH2/MSH6 but not MSH2/MSH3 heterodimer.

SOME OPEN QUESTIONS ABOUT THE MECHANISM

While the broad outlines of the molecular mechanisms of somatic mutation are now apparent, major aspects remain to be elucidated. How is AID targeted to the IgV region for somatic mutation/gene conversion but to the S regions for switch recombination. The targeting is not perfect and, indeed, mistargeted action of AID is clearly associated with proto-oncogene mutations and translocations in various B-cell malignancies.^{64–66} Nevertheless, there is clearly substantial targeting of the deamination. There are abundant pointers that this targeting is somehow linked to transcription of the Ig gene substrate, but the details of this linkage remain obscure.

The molecular basis of the mutations at A:T pairs remains to be elucidated. The diminished abundance of substitutions at A:T pairs in various mutant mice and humans strongly point to a mechanism in which MSH2/MSH6 recognition of the AID-generated U:G lesion leads to a mutagenic patch repair that involves degradation by exonuclease 1 and mutagenic patch resynthesis by DNA polymerase- η .^{37,67–70} But it is unclear why MSH2/MSH6-initiated patch repair should be mutagenic; why is polymerase- η recruited and are the mutations generated simply a consequence of misincorporation errors by this polymerase?

A further issue relates to the issue of how is it that the mutations occur at all; why is it that the AID-generated U:G lesions are not simply repaired in a manner analogous to the repair of the products of spontaneous cytidine deamination. A possible, but speculative, answer is that the answer lies in the timing of the deamination, such that if it occurred around the time of passage of the replication fork, fixation could be preferred over repair.⁷¹

There are also a multitude of questions that relate to the regulation of AID gene expression, its three-dimensional structure, its nuclear shuttling, its post-translational modifications and the differential requirement of somatic mutation and class-switch recombination for the AID C-terminus.^{72–75}

EVOLUTION OF AID

AID is a member of the AID/APOBEC subgrouping of the zinc-dependent deaminase superfamily, which contains deaminases active on free base/nucleotides (cytosine, cytidine and dCTP) and which are spread throughout the biological world⁷⁶ (Figure 9). The superfamily also includes RNA-editing enzymes, which deaminate adenosine in the

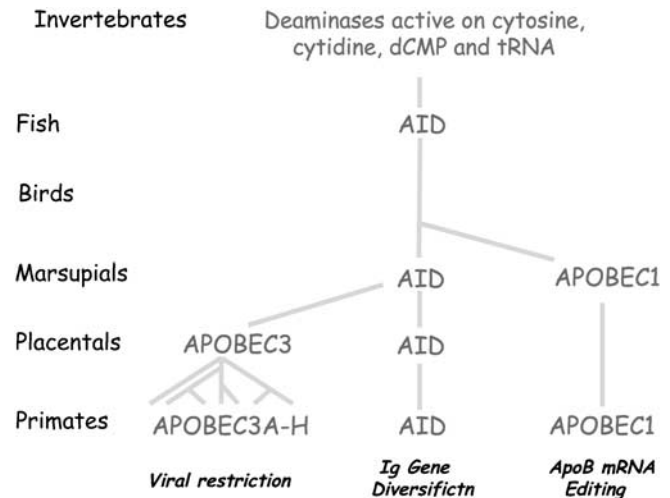


Figure 9 The AID/APOBEC family of polynucleotide deaminases.

context of tRNA (to form the wobble-position inosine), rRNA or the occasional mRNA (for example, that encoding glutamate receptor). Whereas the RNA-editing adenosine deaminases as well as the cytidine deaminases active on free base/nucleotide are found widely distributed among both prokaryotes and eukaryotes, the AID/APOBEC family appear to be restricted to vertebrates.

Whereas the RAG enzymes seem only to be found among jawed vertebrate (although a related protein without evident transposition activity is found in sea urchin⁷⁷), an AID homologue is found in hagfish⁷⁸ where its DNA deaminating activity and tissue-specific expression pattern suggest that it likely contributes to diversification of the VLR antigen receptors (receptors that are members of the leucine-rich repeat family which includes TLRs⁷⁹). Thus AID-mediated antigen receptor diversification appears to go back to the earliest vertebrate, likely predating RAG-mediated antigen receptor assembly (Figure 10).

TARGETED DNA DEAMINATION IN BOTH ADAPTIVE AND INNATE IMMUNITY

Investigations into the mechanism of the γ -globulin gene randomization process anticipated by Burnet revealed the first example of a physiological programme of targeted DNA deamination. However, it then became apparent that several other members of the AID/APOBEC family were also capable of deaminating deoxycytidine in DNA.⁸⁰ It transpired⁸¹ that APOBEC3G (and also APOBEC3F) acts as host restriction factors for HIV-1. In the absence of the viral vif gene product (which targets them for degradation), APOBEC3F and APOBEC3G become incorporated into the viral particle, they then deaminate cytidines in the viral replication intermediates.^{82–84} This can lead to the production of hypermutated viral genomes and may underpin APOBEC3-mediated viral restriction. Other members of the APOBEC3 family may also play a role in the deamination and restriction of other retroelements (reviewed in Conticello *et al.*⁸⁵). While APOBEC3 family members are restricted to higher vertebrates, targeted DNA deamination may also occur in lower eukaryotes—accounting for the repeat-induced point mutation observed in neurospora.⁸⁶ However, in this case it is likely that the deaminations are catalysed as a side reaction of a cytidine methyl transferase rather than by a member of the superfamily of zinc-dependent deaminases.⁸⁷ Thus, DNA deamination appears to be targeted to the cellular genome as a strategy for antibody diversification in adaptive immunity and to

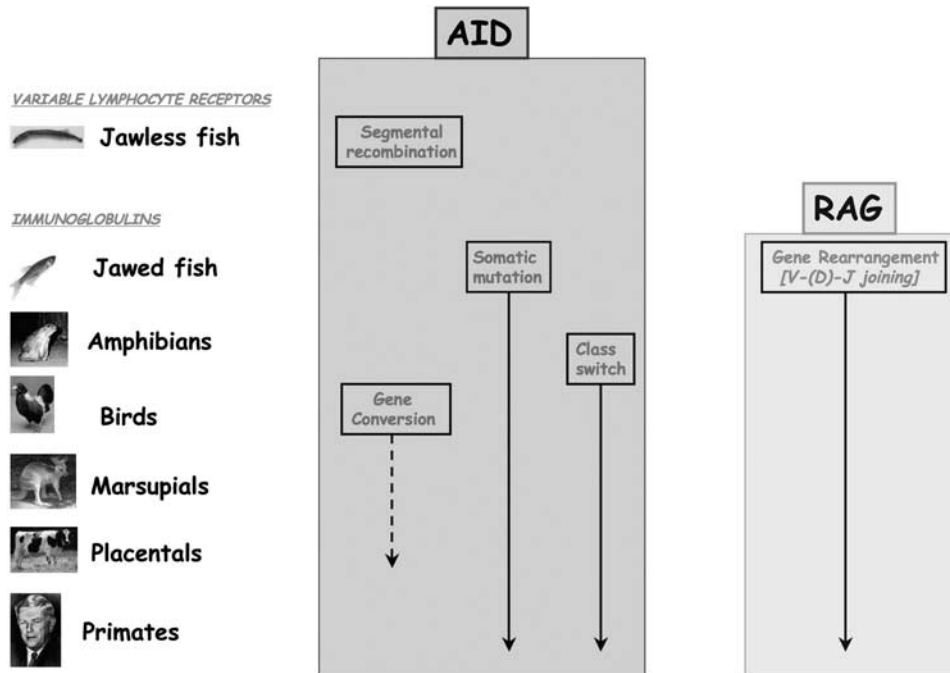


Figure 10 Evolution of the mechanisms of antigen receptor diversification. The suggestion⁷⁸ that AID is implicated in VLR assembly in jawless fish is very reasonable but not definitively established.

invading genomes as a strategy for their destruction in innate immunity.

CONCLUSION

The process of antibody gene randomization anticipated by Burnet in fact comprises two broad components. Gene rearrangement is triggered by the RAG1/RAG2 recombinase. Somatic hypermutation, IgV gene conversion and IgC switch recombination (possibly as well as assembly of VLR receptors in jawless fish) are all triggered by AID-catalysed targeted deamination of deoxycytidine. Burnet's prescient assumption that a previously undescribed process might underpin antibody diversification in early embryonic life and his later incorporation of the idea of ongoing mutation to allow somatic evolution of the antibody response have both been fully vindicated. Indeed, in addition to revealing previously undisclosed mechanisms of programmed genomic modification, his ideas have provided a stimulating trigger for ongoing research with many questions still requiring an answer before we fully understand the molecular processes of gene diversification that he anticipated.

- 1 Burnet FM. A modification of Jerne's theory of antibody formation using the concept of clonal selection. *Aust J Sci* 1957; **20**: 67–69.
- 2 Burnet FM. Men or molecules? A tilt at molecular biology. *Lancet* 1966; **1**: 37–39.
- 3 Lederberg J. Ontogeny of the clonal selection theory of antibody formation: reflections on Darwin and Ehrlich. *Ann NY Acad Sci* 1988; **546**: 175–187.
- 4 Lederberg J. Genetic approaches to somatic cell variation: summary comment. *J Cell Comp Physiol* 1958; **52** (Suppl. 1): 383–402.
- 5 Lederberg J. Genes and antibodies. Do antigens bear instructions for antibody specificity or do they select cell lines that arise by mutation? *Science* 1959; **129**: 1649–1653.
- 6 Talmadge DW. Immunological specificity: unique combinations of selected natural globulins provide an alternative to the classical concept. *Science* 1959; **129**: 1643–1648.
- 7 Talmadge DW. Allergy and immunology. *Annu Rev Med* 1957; **8**: 239–256.
- 8 Burnet FM. The clonal selection theory of acquired immunity—a Darwinian modification. *Aust J Sci* 1964; **27**: 6–7.

- 9 Hilschmann N, Craig LC. Amino acid sequence studies with Bence-Jones proteins. *Proc Natl Acad Sci USA* 1965; **53**: 1403–1409.
- 10 Titani K, Whitley E, Avogadro L, Putnam FW. Immunoglobulin structure: partial amino acid sequence of a Bence Jones protein. *Science* 1966; **152**: 1513–1516.
- 11 Milstein C. Variations in amino acid sequence near the disulphide bridges of Bence-Jones proteins. *Nature* 1966; **209**: 370–373.
- 12 Smithies O. Gamma-globulin variability: a genetic hypothesis. *Nature* 1963; **199**: 1231–1236.
- 13 Dreyer WJ, Bennett JC. The molecular basis of antibody formation: a paradox. *Proc Natl Acad Sci USA* 1965; **54**: 864–869.
- 14 Brenner S, Milstein C. Origin of antibody variation. *Nature* 1966; **211**: 242–243.
- 15 Jerne NK. Summary: waiting for the end. *Cold Spring Harbor Symp Quant Biol* 1967; **32**: 591–603.
- 16 Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983; **302**: 575–581.
- 17 Schatz DG, Oettinger MA, Baltimore D. The V(D)J recombination activating gene, RAG-1. *Cell* 1989; **59**: 1035–1048.
- 18 Gellert M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 2002; **71**: 101–132.
- 19 Alt F, Baltimore D. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proc Natl Acad Sci USA* 1982; **79**: 4118–4122.
- 20 Hozumi N, Tonegawa S. Evidence for somatic rearrangements of immunoglobulin genes coding for variable and constant regions. *Proc Natl Acad Sci USA* 1976; **73**: 3628–3632.
- 21 Reynaud CA, Bertocci B, Dahan A, Weill JC. Formation of the chicken B cell repertoire: ontogenesis, regulation of Ig rearrangement and diversification by gene conversion. *Adv Immunol* 1994; **57**: 353–378.
- 22 Weigert MG, Cesarini IM, Yonkovich SJ, Cohn M. Variability in the λ light chain sequences of mouse antibody. *Nature* 1970; **228**: 1045–1047.
- 23 Reynaud CA, Garcia C, Hein W, Weill JC. Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell* 1995; **80**: 115–125.
- 24 Weller S, Braun MC, Tan BK, Rosenwald A, Cordier C, Conley ME *et al.* Human blood IgM 'memory' B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire. *Blood* 2004; **104**: 3647–3654.
- 25 McKean D, Huppi K, Bell M, Staudt L, Gerhard W, Weigert M. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc Natl Acad Sci USA* 1984; **81**: 3180–3184.
- 26 Gearhart PJ, Bogenhagen DF. Clusters of point mutations are found exclusively around rearranged antibody variable genes. *Proc Natl Acad Sci USA* 198; **380**: 3439–3443.
- 27 Lebecque SG, Gearhart PJ. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. *J Exp Med* 1980; **172**: 1717–1727.
- 28 O'Brien RL, Brinster RL, Storb U. Somatic hypermutation of an immunoglobulin transgene in κ transgenic mice. *Nature* 1987; **326**: 405–409.

- 29 Neuberger MS, Milstein C. Somatic hypermutation. *Curr Opin Immunol* 1995; **7**: 248–254.
- 30 Betz AG, Milstein C, González-Fernández A, Pannell R, Larson T, Neuberger MS. Elements regulating somatic hypermutation of an immunoglobulin κ gene: critical role for the intron enhancer/matrix attachment region. *Cell* 1994; **77**: 239–248.
- 31 Sharpe M, Milstein C, Jarvis JM, Neuberger MS. Somatic hypermutation of immunoglobulin κ may depend on sequences 3' of C κ and occurs on passenger transgenes. *EMBO J* 1991; **10**: 2139–2145.
- 32 Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim Biophys Acta* 1992; **1171**: 11–18.
- 33 Betz A, Rada C, Pannell R, Milstein C, Neuberger MS. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. *Proc Natl Acad Sci USA* 1993; **90**: 2385–2388.
- 34 Wagner SD, Milstein C, Neuberger MS. Codon bias targets mutation. *Nature* 1995; **376**: 732.
- 35 Phung QH, Winter DB, Cranston A, Tarone RE, Bohr VA, Fishel R *et al.* Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J Exp Med* 1998; **187**: 1745–1751.
- 36 Rada C, Ehrenstein MR, Neuberger MS, Milstein C. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* 1998; **9**: 135–141.
- 37 Zeng X, Winter DB, Kasmer C, Kraemer KH, Lehmann AR, Gearhart PJ. DNA polymerase η is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol* 2001; **2**: 537–541.
- 38 Denepoux S, Razanajao D, Blanchard D, Meffre G, Capra JD, Banchereau J *et al.* Induction of somatic mutation in a human B cell line *in vitro*. *Immunity* 1997; **6**: 35–46.
- 39 Sale JE, Neuberger MS. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity* 1998; **9**: 859–869.
- 40 Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO *et al.* Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* 1999; **274**: 18470–18476.
- 41 Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 2000; **102**: 553–563.
- 42 Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell* 2000; **102**: 565–575.
- 43 Arakawa H, Hauschild J, Buerstedde JM. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* 2002; **295**: 1301–1306.
- 44 Harris RS, Sale JE, Petersen-Mahrt SK, Neuberger MS. AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Curr Biol* 2002; **12**: 435–438.
- 45 Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* 2002; **418**: 99–104.
- 46 Wilson M, Hsu E, Marcuz A, Courtet M, Du Pasquier L, Steinberg C. What limits affinity maturation of antibodies in *Xenopus*—the rate of somatic mutation or the ability to select mutants? *EMBO J* 1992; **11**: 4337–4347.
- 47 Bachl J, Wabl M. An immunoglobulin mutator that targets GC base pairs. *Proc Natl Acad Sci USA* 1996; **93**: 851–855.
- 48 Ehrenstein MR, Neuberger MS. Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination. *EMBO J* 1999; **18**: 3484–3490.
- 49 Maizels N. Somatic hypermutation: how many mechanisms diversify V region sequences? *Cell* 1995; **83**: 9–12.
- 50 Weill JC, Reynaud CA. Rearrangement-hypermutation-gene conversion: when, where and why? *Immunol Today* 1996; **17**: 92–97.
- 51 Sale JE, Calandrin DM, Takata M, Takeda S, Neuberger MS. Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. *Nature* 2001; **412**: 921–926.
- 52 Pham P, Branstetter R, Petruska J, Goodman MF. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 2003; **424**: 103–107.
- 53 Branstetter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci USA* 2003; **100**: 4102–4107.
- 54 Chaudhuri J, Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E *et al.* Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 2003; **422**: 726–730.
- 55 Dickerson SK, Market E, Besmer E, Papavasiliou FN. AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* 2003; **197**: 1291–1296.
- 56 Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat Immunol* 2003; **4**: 452–456.
- 57 Sohail A, Klapacz J, Samaranyake M, Ullah A, Bhagwat AS. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res* 2003; **31**: 2990–2994.
- 58 Chaudhuri J, Khuong C, Alt FW. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* 2004; **430**: 992–998.
- 59 Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra *in vivo*. *J Mol Biol* 2004; **337**: 585–596.
- 60 Di Noia J, Neuberger MS. Altering the pathway of Ig V gene hypermutation by inhibiting uracil DNA glycosylase. *Nature* 2002; **419**: 43–48.
- 61 Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* 2002; **12**: 1748–1755.
- 62 Rada C, Di Noia JM, Neuberger MS. Mismatch recognition and uracil-excision provide complementary paths to both Ig switching and the A-T-focused phase of somatic mutation. *Mol Cell* 2004; **16**: 163–171.
- 63 Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N *et al.* Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* 2003; **4**: 1023–1028.
- 64 Pasqualucci L, Migliazza A, Fracchiolla N, William C, Neri A, Baldini L *et al.* BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc Natl Acad Sci USA* 1998; **95**: 11816–11821.
- 65 Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Kuppers R *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 2001; **412**: 341–346.
- 66 Shen HM, Peters A, Baron B, Zhu X, Storb U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* 1998; **280**: 1750–1752.
- 67 Delbos F, Aoufouchi S, Faili A, Weill JC, Reynaud CA. DNA polymerase η is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J Exp Med* 2007; **204**: 17–23.
- 68 Wiesendanger M, Kneitz B, Edelmann W, Scharff MD. Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J Exp Med* 2000; **191**: 579–584.
- 69 Bardwell PD, Woo CJ, Wei K, Li Z, Martin A, Sack SZ *et al.* Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat Immunol* 2004; **5**: 224–229.
- 70 Neuberger MS, Di Noia JM, Beale RCL, Williams GT, Yang Z, Rada C. Somatic hypermutations at A-T pairs: polymerase error versus dUTP incorporation. *Nat Rev Immunol* 2005; **5**: 171–178.
- 71 Di Noia JM, Rada CA, Neuberger MS. SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J* 2006; **25**: 585–595.
- 72 Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell* 2003; **12**: 501–508.
- 73 Ta VT, Nagaoka H, Catalan N, Durandy A, Fischer A, Imai K *et al.* AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat Immunol* 2003; **4**: 843–848.
- 74 Ito S, Nagaoka H, Shinkura R, Begum N, Muramatsu M, Nakata M *et al.* Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci USA* 2004; **101**: 1975–1980.
- 75 McBride KM, Barreto V, Ramiro AR, Stavropoulos P, Nussenzweig MC. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J Exp Med* 2004; **199**: 1235–1244.
- 76 Conticello SG, Thomas CF, Petersen-Mahrt S, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 2005; **22**: 367–377.
- 77 Fugmann SD, Messier C, Novack LA, Cameron RA, Rast JP. An ancient evolutionary origin of the Rag1/2 gene locus. *Proc Natl Acad Sci USA* 2006; **103**: 3728–3733.
- 78 Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI *et al.* Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat Immunol* 2007; **8**: 647–655.
- 79 Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 2004; **430**: 174–180.
- 80 Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* 2002; **10**: 1247–1253.
- 81 Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 2002; **418**: 646–650.
- 82 Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell* 2003; **113**: 803–809.
- 83 Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 2003; **424**: 99–103.
- 84 Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* 2003; **424**: 94–98.
- 85 Conticello SG, Langlois MA, Yang Z, Neuberger MS. DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv Immunol* 2007; **94**: 37–73.
- 86 Selker EU. Premiotic instability of repeated sequences in *Neurospora crassa*. *Annu Rev Genet* 1990; **24**: 579–613.
- 87 Freitag M, Williams RL, Kothe GO, Selker EU. A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proc Natl Acad Sci USA* 2002; **99**: 8802–8807.