

DNA polymerases in adaptive immunity

Jean-Claude Weill and Claude-Agnès Reynaud

Institut National de la Santé et de la Recherche Médicale (INSERM) U783 « Développement du système immunitaire », Faculté de Médecine Paris Descartes, Site Necker-Enfants Malades, 156 rue de Vaugirard, 75730 Paris Cedex 15, France.

Correspondence to J.-C.W. or C.-A.R. e-mails: weill@necker.fr; reynaud@necker.fr

“Prediction is very difficult, especially if it’s about the future”

Niels Bohr

Preface

To cope with an unpredictable variety of potential pathogenic insults, the immune system must generate an enormous diversity of recognition structures, and it does so by making stepwise modifications of key genetic loci in each lymphoid cell. This proceeds through the action of lymphoid-specific proteins acting together with the general DNA-repair machinery of the cell. Strikingly, these general mechanisms are usually diverted from their normal functions, being used in rather atypical ways in order to privilege diversity over accuracy. In this Review, we focus on the contribution of a set of DNA polymerases discovered in the last decade to these unique DNA transactions.

Introduction

Starting from their development in the bone marrow, B cells undergo a series of complex molecular events to generate mature B cells that express a single functional B-cell receptor (BCR). After antigen encounter, B cells can remain for the life of the host in the memory B-cell compartment to provide protection from infection (Fig.1). For this to occur, B cells must first generate a diverse pre-immune repertoire of BCRs for recognition of the diverse antigenic world. In mice and humans, this repertoire is generated by ongoing rearrangement of a large pool of variable (V) gene segments (around 50–100) and a small pool of diversity (D) and joining (J) units (around 10) that combine to form the BCR by a process known as V(D)J recombination¹. Activation of B cells by foreign antigen induces a new cascade of genomic modifications in the responding B cells. These involve the introduction of point mutations in the V regions of the rearranged antibody to improve its affinity for the antigen (a process known as somatic hypermutation (SHM)), as well as a switch in the DNA that encodes the constant (C) region of the immunoglobulin heavy chain, from C μ to another heavy chain, C γ , C α or C ϵ , to provide appropriate effector function² (a process known as class-switch recombination (CSR)). SHM and/or a third type of DNA modification known as gene conversion, are also used in some species that have more restricted numbers of germline V gene segments to generate the pre-immune BCR repertoire³ (Fig.1).

In this Review, we focus on a specific set of enzymes — DNA polymerases — that are mobilized during these different DNA transactions in B cells. New DNA polymerases have

been described in the last decade, and several of them have now been shown to have a distinct role in the adaptive immune system. We describe the variety of enzymatic activities carried out by DNA polymerases that contribute to immunoglobulin gene diversity either through the processing of DNA junctions or through mutagenesis.

DNA polymerases

DNA polymerases are enzymes that synthesize new DNA strands using a DNA template. DNA polymerases act in semi-conservative DNA replication and in various DNA-repair pathways that require *de novo* DNA synthesis. *Escherichia coli* encodes five polymerases (PolI to PolV), two of which are involved in replication, with the other three being involved in DNA repair, DNA damage tolerance and stress-induced mutagenesis⁴. By contrast, mammals have 14 such enzymes (or 15, if one includes terminal deoxynucleotidyl transferase (TdT), a template-independent polymerase). Four are replicative polymerases (Pol α , Pol γ , Pol δ and Pol ϵ) and 11 perform various non-replicative tasks^{5,6} (Table 1). These non-replicative functions concern the removal and repair of damaged bases (for example, Pol β in base-excision repair), the rejoining of broken DNA ends (Pol λ and Pol μ in non-homologous end-joining (NHEJ)), the by-pass of DNA lesions that block the progression of replication forks (Y family polymerases, as well as Pol ζ , in translesion DNA synthesis (TLS)) and template-independent insertion of nucleotides (TdT during V(D)J recombination). One hypothesis proposed for the striking increase in the number of non-replicative polymerases in higher vertebrates compared with prokaryotes is that each polymerase would display a preferential handling of specific DNA lesions, for which it would harbour both higher efficiency and accuracy, thereby reducing the deleterious consequences of the corresponding damage⁷.

V(D)J recombination and DNA polymerases

A complete V gene sequence of the BCR is assembled from dispersed V, D and J coding segments for the heavy chain, and V and J elements for the light chain. This recombination process is mediated by the lymphoid-specific genes recombination-activating gene 1 (RAG1) and RAG2, which ensure that the process is targeted to antigen-receptor loci, together with the general NHEJ machinery of the cell. The ubiquitous NHEJ machinery mediates several essential steps: the binding of DNA ends, their processing and their proper religation. Extensive modifications that occur at DNA ends before the continuity of the immunoglobulin

coding sequence is restored are the hallmark of V(D)J recombination, and the enzymatic activities involved are discussed below.

The rearrangement process. The assembly of a complete immunoglobulin V gene involves the introduction of double-strand breaks (DSBs) in the DNA at the border of each coding element followed by rejoining of the ends^{8, 9} (Fig. 2). For this to occur, a synapsis forms between the two distantly located genomic regions and the DNA is cleaved by a process involving the two RAG proteins. The RAG1/RAG2 complex binds to conserved elements in the DNA known as recombination signal sequences (RSSs) that flank each coding segment (V, D and J segments) and introduces a nick into one strand of the DNA between the RSS and the coding sequence. The nick creates a free 3' hydroxyl group, which attacks the phosphodiester bond on the opposite strand, forming the DSB. This cleavage generates two coding ends that terminate in a covalently sealed hairpin and two blunt signal ends. The signal ends remain associated with the RAG complex, which serves as a scaffold to recruit factors that religate the ends — by the NHEJ pathway — to generate a circular episome that will eventually be eliminated. The hairpin coding ends, also synapsed by the RAG complex, are encircled by the Ku proteins (Ku70 and Ku86), which recruit DNA-PKcs (DNA-dependent protein kinase catalytic subunit), which in turn recruits and activates Artemis, enabling it to open the hairpin ends.

In most places in the genome, the DSBs are then resolved by rendering the DNA ends compatible for their religation. However, at the antigen-receptor loci, the coding ends are subjected to extensive 'nibbling' (nucleotide trimming) by as-yet-undefined exonucleases followed by DNA resynthesis by template-dependent polymerases, and random nucleotide insertions (N-additions) by the template-independent polymerase TdT. These different enzymatic activities create diversity by varying the number of amino acids in the complementarity-determining region 3 (CDR3) of the antigen receptor, a region that makes a major contribution to the specificity of the antibody produced. Finally, the strands are religated and the coding joint formed by the activities of XRCC4 (X-ray repair cross-complementing protein 4) in complex with Cernunnos (also known as XLF) and DNA ligase IV¹⁰.

Random pairing of the rearranged heavy and light chains then occurs during assembly of the antibody molecule to create different specificities. It should be emphasized that the junctional variability generated during V(D)J recombination at the heavy chain locus exceeds

by orders of magnitude the diversity created by both the assortment of V, D and J gene segments and the combination of heavy-chain and light-chain pairing.

Here we focus on the role of DNA polymerases during the rearrangement process of the BCR and we start by summarizing our current knowledge of the role played by the PolX family (which comprises Pol β , Pol λ , Pol μ and TdT in mammals). Owing to their capacity to interact with DNA-PKcs complexed to DNA, Pol μ and Pol λ have emerged as possible partners of both general NHEJ and V(D)J recombination^{11, 12}. Pol β in contrast seems to be mainly involved in BER¹³.

PolX polymerases in NHEJ. Polymerases are required during the repair of DSBs by NHEJ to fill in gaps in the DNA or to extend recessed 3' ends. In the yeast *Saccharomyces cerevisiae*, the Pol4 enzyme, which belongs to the PolX family, has been shown to be involved in fill-in DNA synthesis during NHEJ¹⁴. Among PolX polymerases, Pol λ is closer to yeast Pol4 while Pol μ and TdT, which are highly homologous to each other, are more distantly related¹⁵⁻¹⁷. Pol μ and Pol λ are expressed in most tissues, with enhanced expression of Pol μ found in lymphoid cells, whereas TdT is exclusively expressed in developing B and T cells.

Several laboratories have studied the role of PolX polymerases during DSB repair. Pol μ and Pol λ have been shown to associate with the protein complex Ku–DNA-PKcs once they are bound to DNA and can perform gap-filling synthesis of paired DNA ends *in vitro*^{11, 12, 18}. The possible use of these enzymes during NHEJ seems to be due to their capacity to synthesize DNA from less stable primer structures compared to replicative polymerases, such as primers that encompass a DNA template with a gap¹⁹ (Fig.3c). Lieber and colleagues, using purified polymerase proteins *in vitro*, showed that a competitive hierarchy exists in the recruitment of polymerases to the DNA during NHEJ, with Pol μ being preferentially recruited over Pol λ in the absence of TdT, whereas the reverse was true when TdT was present¹². Ramsden and colleagues proposed the existence of a range of increasing template dependence among mammalian PolX family members, with TdT showing the least template dependency, then Pol μ and then Pol λ , which shows higher template dependence²⁰. Accordingly, in their assay, Pol μ has the unique capacity to prime DNA synthesis across a junction without any base pairing (Fig.3c).

Paradoxically, the absence of these polymerases induces no sensitivity to DSB-inducing agents, as embryonic fibroblasts derived from mice lacking either or both Pol μ and Pol λ were not sensitive to ionizing radiation²¹. This implies that other processing activities

can substitute and provide efficient DSB repair, although the outcome of DSB rejoining may be qualitatively different.

The nibbling process during V(D)J recombination. The CDR3 that is generated during the rearrangement process displays a size heterogeneity that is more marked for the heavy chain (15 to 55 base pairs) than for the light chain (20 to 30 base pairs). Such size differences imply a variable trimming of the coding ends before religation, but the enzyme responsible for this activity has not yet been formally characterized. It has been proposed that a splicing variant of TdT may have such exonuclease activity²², but biochemical and functional assays have challenged this idea^{23, 24}. During NHEJ, Artemis forms a complex with DNA-PKcs, which phosphorylates Artemis and allows it to acquire an endonucleolytic activity acting on 3' or 5' overhanging ends and on various branched DNA structures. This role for Artemis explains the general radiosensitivity observed cells that lack this protein^{25, 26}. Artemis cleaves RAG-induced hairpin ends preferentially 3' from the tip *in vitro*²⁵. *In vivo*, this cleavage results in a small palindrome sequence (referred to as P addition) at the coding end of the immunoglobulin gene segments. Once the hairpin has been cleaved, Artemis could be responsible for nucleotide trimming at the 5' and 3' ends of the coding ends, but the involvement of additional exonucleases has not been ruled out. We found no effect of inactivation of the exonuclease Exo1 on the size of the CDR3 of heavy-chain V regions (C.-A. Reynaud and J.-C. Weill, unpublished observations, in collaboration with W. Edelmann), but other enzymes, such as TREX1 (three prime repair exonuclease 1)²⁷, remain to be assessed.

PolX polymerases during V(D)J recombination. A role for Pol μ and Pol λ polymerases in the V(D)J recombination process has been assessed by the analysis of immunoglobulin rearrangement in mice deficient for these enzymes, alone or in combination. We reported that light-chain junctions were shorter by about 6 base pairs in the absence of Pol μ , whereas the size of heavy-chain junctions was not affected²⁸. These altered light-chain junctions resulted in an impaired pre-B-cell to B-cell transition, at which stage light-chain genes rearrange, as well as a moderate and more variable peripheral B-cell deficiency. Surprisingly, Pol λ -deficient mice had normal light-chain joints but heavy-chain junctions that were about 5 base pairs shorter, although no quantitative effect on the B-cell population was observed²¹. Nucleotide additions were unaffected in the absence of Pol λ , suggesting that Pol λ was acting

before TdT-mediated nucleotide addition and was therefore not involved in filling the gaps generated by 3' TdT-mediated extensions. Mice deficient in both polymerases displayed the additive effects of both phenotypes, without any further decrease in junction size or B-cell deficiency²¹. Together, these studies suggest that Pol μ and Pol λ act on different immunoglobulin chains: Pol μ on the light chain and Pol λ on the heavy chain. During B-cell ontogeny, Pol λ and Pol μ are expressed at similar levels at the pro-B-cell stage, when heavy-chain gene rearrangement takes place, whereas Pol μ expression increases 25-fold at the stage when the light chain genes rearrange. By comparison, there is a several hundred-fold higher expression of TdT at the pro-B-cell stage, when N-additions are made, compared with later differentiation steps, indicating that the recruitment of these different enzymes may be dictated by both their relative affinity for DNA and their respective expression level.

Exactly how Pol λ and Pol μ participate in V(D)J recombination remains to be elucidated (Fig. 3). Their role could involve protecting the DNA ends from an exonuclease or, as proposed for Pol μ , filling in gaps in the DNA at junctions with minimal homology, or possibly both^{20, 21}. Whatever their exact roles, the fact that the immune system has evolved two different polymerases for each chain of the BCR is an unexpected finding. We have proposed that their precise interventions may be linked to the particular property of each CDR3²¹. For light chains, a restricted heterogeneity of CDR3 sizes is indeed required to maintain a semi-invariant residue encoded at the 3' end of V κ genes²⁹, whereas CDR3 heterogeneity is key for the heavy-chain repertoire. Minimal processing of light-chain coding ends would be tightly controlled by Pol μ , whereas a less potent activity of Pol λ would allow for the large size diversity that is characteristic of CDR3s of immunoglobulin heavy chains.

Somatic hypermutation and DNA polymerases

In the 1950s, SHM was proposed as a central player in the immune system either to improve the primary repertoire or to mature a specific immune response^{30, 31}. It was not until decades later that mutations were actually observed, first in the immunoglobulin protein³² and then at the DNA level³³⁻³⁶, providing proof that a process of adaptive mutation targeted to a single locus, not observed in any other living organism, was co-opted by the immune system.

One of the central problems researchers faced when trying to understand SHM was the nature of the enzymatic activity involved³⁷. The discovery of activation-induced cytidine deaminase (AID) provided a clue to this problem³⁸⁻⁴⁰. It is now generally accepted that the deamination of cytidines into uracils by AID is the initiating event that allows the specific

recruitment of factors involved in SHM, CSR or gene conversion⁴¹. Uracils can normally be excised by several uracil glycosylases and replaced by the base-excision repair pathway, which faithfully restores the original DNA sequence. By contrast, during SHM, the uracils generated by AID can be recognized either by uracil-DNA glycosylase (UNG) or, as U-G mispairs, by the mismatch repair complex. However, both repair pathways appear to function abnormally. Effectively, the action of UNG is restricted to uracil excision, giving rise to abasic sites, and, for mismatch repair, the mismatch-binding moiety (MSH2–MSH6, together with Exo1) proceeds alone without recruiting the effector part of the complex (MLH1–PMS2)⁴²⁻⁴⁴. These two repair pathways therefore trigger further DNA alterations through mutagenic DNA synthesis, rather than their normal role in correcting the error to match the original sequence (see model proposed below). Evidence for the involvement of mutagenic polymerases in these atypical repair processes is reviewed in the following sections, and a model is proposed based on our knowledge of the different partners involved.

Candidate polymerases for SHM. The first candidate polymerase known to lack proofreading activity and to act outside the S phase of the cell cycle was Pol β . As Pol β -deficient mice die after birth, reconstitution of the immune system of irradiated animals with Pol β -deficient fetal liver cells was carried out, but this failed to reveal any deficiency in the SHM process⁴⁵.

Another group of polymerases — namely the TLS polymerases⁴⁶⁻⁴⁸ — were considered good candidates for being involved SHM. These enzymes can mediate polymerase activity opposite DNA lesions that induce a stalling (pausing) of the replication fork. Because of their relaxed catalytic sites that allow them to accommodate distorted DNA structures, TLS polymerases are highly error-prone when copying normal undamaged DNA. Although studies of TLS have been largely focused on polymerases of the Y family (together with Pol ζ), other enzymes, such as Pol μ , Pol λ and Pol θ , have shown robust translesion activity *in vitro*⁴⁹⁻⁵¹. Each non-replicative polymerase (except Pol ν) has been tested for its role in SHM. Since most of these enzymes have a specific mutagenic signature *in vitro*, when copying normal DNA or when by-passing abasic sites, one should expect that inactivation of a polymerase involved in SHM would result in a modification of the immunoglobulin gene mutation pattern in a manner that reflects its enzymatic preferences.

Pol η . The study of patients affected by the xeroderma pigmentosum variant (XP-V) syndrome, a genetic disease associated with the inactivation of pol η ^{52, 53}, provided the first

experimental evidence for the involvement of a TLS polymerase in SHM. A reduced mutation frequency at A/T bases was observed in these patients^{54, 55}, an observation that was further confirmed in Pol η -deficient mice⁵⁶⁻⁵⁸ (Fig.4).

The preferred mutation of Pol η when copying normal DNA *in vitro* is the incorporation of Gs opposite Ts, thereby generating T to C and A to G transition mutations, the complementary mutation (that is, the incorporation of Cs opposite As) being 10 times less frequent⁵⁹. Altogether, the nucleotide misincorporation frequency of Pol η is 2–3%, with Ts mutated 2–2.5 times more frequently than As, and Cs plus Gs about 3–4 times less frequently than both As and Ts. Pol η is not a processive enzyme and therefore favours the synthesis of a few nucleotides (short-patch DNA synthesis). Mutagenicity of Pol η was estimated using a V κ template sequence when synthesizing either the transcribed or the non-transcribed strand *in vitro*⁶⁰. Interestingly, despite the fact that this assay monitored the error rate of the polymerase during the synthesis of a few hundred bases that may not occur *in vivo*, a strong correlation with the *in vivo* pattern of mutations at A/T base pairs was observed, in particular during synthesis of the coding strand (the non-transcribed strand), although hotspots on both coding and non-coding strands were clearly observed. This suggested that both strands could be copied by Pol η *in vivo*, possibly with some bias for the coding strand.

A residual A/T mutagenesis (10–15%) is observed in Pol η -deficient mice, suggesting that other polymerase activities might contribute to the formation of mutations at A/T bases⁵⁶. Half of these mutations consist in A to C and T to G transversion mutations, a pattern that evokes the signature of another TLS enzyme, Pol κ ^{61, 62}(Fig.4). As Pol κ does not seem to play a role in the physiological situation^{63, 64}, it suggests that this polymerase acts as a back-up when the normal pathway is compromised. Accordingly, the reduction in A/T mutagenesis corresponds well with the lower mutagenicity of Pol κ ^{61, 62}. Altogether these data support the proposition that Pol η is the sole contributor of A/T mutations during physiological SHM in mice, with other TLS polymerases playing a backup role⁶⁵.

Rev1. Rev1 has a deoxycytidyltransferase activity that is directed opposite various damaged bases, as well as abasic sites and normal DNA^{66, 67}. The Rev1 protein contains an N-terminal domain that is unique among other Y family enzymes⁴⁶ (known as BRCA1 C-terminal (BRCT), which is involved in protein–protein interactions) and a C-terminal region that mediates interaction with other TLS polymerases⁶⁸. Deletion of the BRCT domain of Rev1 has been shown to be required for TLS activity of certain DNA lesions, although it does not

seem to be required for SHM⁶⁹. By contrast, the complete deficiency of Rev1 has a clear effect on SHM⁷⁰. G/C mutations represent around 50% of all mutations during SHM, among which approximately 25% are G to C and C to G transversion mutations that could correspond to the mutation signature of Rev1 when bypassing abasic sites. Surprisingly, in Rev1-deficient mice, C to G mutations were almost absent from the coding strand, whereas G to C mutations were only reduced (by around 50%), suggesting a different use of Rev1 depending on the strand synthesized and the implication of another polymerase for generating the remaining C to G mutations. As these mutation data were obtained using functional V λ 1 sequences, analysis of unselected mutations would be useful to confirm this strand bias. These data therefore indicate that other polymerases, in addition to Rev1, are responsible for generating transversions mutations at G/C bases.

Pol ζ . Pol ζ is composed of two subunits –Rev3, the catalytic subunit, and Rev7– and is a major partner in the bypass of DNA lesions. It has been proposed that pol ζ could work as a general DNA extender in lesion bypass performed by other TLS enzymes, being able to synthesize a few nucleotides before the replication fork restarts its progression⁷¹. Moreover, Pol ζ is the only TLS polymerase that when inactivated in mice is embryonic lethal, indicating that this enzyme has other functions beyond TLS activity⁷²⁻⁷⁴. It was recently reported that the activity of Rev3 is error-prone when copying undamaged DNA, albeit to a lesser extent than Pol η and Pol ι ⁷⁵. However, it should be noted that all biochemical assays on Rev3 are performed using the yeast enzyme, as no active recombinant mammalian Rev3 has been produced so far. In transgenic mice expressing an antisense transcript against Rev3 that reduces its expression in B cells, a lower mutation frequency was observed, with no specific impact on the mutation pattern⁷⁶. A similar observation was made in a Burkitt's lymphoma cell line (CL-01) that is induced to undergo SHM in culture, using antisense oligonucleotides against Rev3⁷⁷. Mice with conditional Rev3 inactivation, allowing its deletion in mouse B cells, were recently generated. Rev3 deficiency impaired B-cell proliferation, presumably due to a defect in the maintenance of genome stability, as Rev3-deficient B cells frequently harboured chromosomal aberrations. A decreased SHM frequency was observed, but again with no obvious change in the SHM pattern⁷⁸.

Taken together, these results can be interpreted in two ways. First, Pol ζ could be a major partner in all the mutations produced during SHM. This is a feasible scenario because it could participate in extending the short-patch DNA synthesis performed by Pol η and the

bypass of abasic sites. Second, the effect of the Rev3 deficiency observed could be indirect, being due instead to the altered cell viability caused by the absence of Pol ζ .

Pol ι . Like Pol η , Pol ι tends to incorporate a T opposite a G when copying normal DNA⁷⁹. Pol ι is ten-fold more potent than Pol η in generating this error, although recent data suggest that its fidelity may vary considerably depending on the nature of the divalent ionic cofactor present in the reaction⁸⁰.

Human Burkitt's lymphomas undergo SHM in culture either constitutively (Ramos cell line) or after induction (BL2 or CL-01 cell lines)⁸¹⁻⁸⁴. The pattern of mutations does not mimic the SHM pattern generated *in vivo* as it displays more mutations at G/C base pairs (80% versus 20% at A/T base pairs). It seems therefore that the A/T pathway does not function properly in these cell lines despite the presence of a functional mismatch repair pathway and normal Pol η expression⁸⁵. Strikingly, SHM could no longer be induced when Pol ι was inactivated in the BL2 cell line⁸⁶. In contrast, SHM was found to be normal in mice deficient for Pol ι (a deficiency generated by a spontaneous stop codon mutation occurring in the 129 mouse strain), thus excluding Pol ι as a physiological partner of hypermutation in the mouse⁸⁷. Therefore, in BL2 cells, Pol ι may be involved in an error-prone repair pathway mediated by UNG, similar to what occurs for Pol η in MSH2-deficient mice (see model below). Accordingly, BL2 subclones displaying increased Pol β expression fail to induce SHM, suggesting that Pol β could displace Pol ι to perform its normal error-free repair function⁸⁸.

Pol θ . Two groups have reported contrasting results on the impact of a complete inactivation of Pol θ on SHM in mice. In one report, deletion of *Polq* (the gene encoding Pol θ) leads to a large decrease in the frequency of mutations, with a pattern biased towards transition mutations⁸⁹. In another report, deletion of this gene only induces a moderate reduction in mutation frequency, without alteration of the pattern of mutations⁹⁰. Moreover, deletion of both Pol θ and Pol η did not further modify the mutation profile seen in Pol η -deficient mice⁹¹. Deletion of the sole polymerase domain of Pol θ also resulted in a mild phenotype, with a small increase of A/T mutagenesis but no alteration of the G/C mutation profile⁹², a result that, again, does not fit well with the robust translesional activity of Pol θ at abasic sites *in vitro* and its preference for insertion of As⁵¹. Additional experiments are therefore needed to clarify whether there is a direct role of Pol θ during SHM.

PCNA. Proliferating cell nuclear antigen (PCNA) is a sliding clamp that tethers DNA polymerases to the DNA template during replication. PCNA is monoubiquitinated at lysine residue 164 in human cells after UV-irradiation, and it has been proposed that this modification may mediate the exchange between replicative and TLS polymerases when the replication fork arrests at a lesion⁹³. Accordingly, TLS polymerases bind to monoubiquitinated PCNA through two types of motif: a PCNA interacting peptide (PIP box) that binds directly to PCNA⁹⁴ and an ubiquitin-binding motif that is required for their bypass function⁹⁵. The role of PCNA in SHM was analysed using mice bearing a PCNA mutation (Lys164Arg) that is known to inhibit its monoubiquitination and therefore its interaction with TLS polymerases. Mutations at A/T bases were found to be practically absent⁹⁶. Mutations at G/C bases were mildly affected, with only a 50% reduction of C to G transversion mutations and no alteration of other types of G/C mutations, including G to C transversions. When these results are compared with those observed for *Rev1*^{-/-} mice, they imply that, for C to G mutations, bypass of half of the abasic sites on the coding strand can be accomplished by Rev1 independently of a monoubiquitinated form of PCNA. MSH2 and UNG also have a PIP box, and it has been shown that PCNA can be recruited in specific DNA processes, such as mismatch repair or base-excision repair which takes place outside DNA chromosomal replication⁹⁷. However it is not known whether PCNA is monoubiquitinated during these processes. The contribution of the monoubiquitination of PCNA during SHM appears therefore somewhat paradoxical, being required during the short-patch DNA synthesis performed by Polη, but almost dispensable during the bypass of abasic sites that represents a classical TLS process.

UNG, MSH2 and the A/T mutagenesis pathway

The uracils generated by AID are processed by only two pathways, UNG and mismatch repair, as clearly established by Neuberger and colleagues. These authors used UNG/MSH2 double-deficient mice⁹⁸ to show that uracils generated by AID are carried over into replication without processing, to generate only G to A and C to T transition mutations (Fig.4). These mutations therefore represent the footprint of AID deamination at the immunoglobulin locus.

UNG deficiency alone shifts mutagenesis at G/C bases towards transition mutations, with a minor impact on A/T mutations⁹⁹ (Fig. 4). By contrast, MSH2 deficiency results in a decreased mutagenesis at A/T bases, a phenotype that is also associated with MSH6 and Exo1

deficiency¹⁰⁰⁻¹⁰³ (but which is not observed in the absence of PMS2 or MLH1, the effector partners of the mismatch repair complex)^{42, 104}. These contrasting phenotypes led to the proposition of two alternative phases of SHM: one mediated by UNG and generating mutations at G/C bases after uracil excision, and one mediated by MSH2–MSH6, introducing A/T mutations⁹⁸. However, alteration of the mutation pattern in MSH2-deficient animals appears to be more complex: in addition to a reduced A/T mutagenesis, it shows an overall reduced mutation frequency, an increase in transition mutations at G/C bases, and an altered mutation profile showing an increased targeting at a few hotspot positions (Fig.4). For a long time, no comprehensive explanation was provided for this complex phenotype, which was described before the discovery of AID. However, when compared to the footprint of AID targeting revealed in *Ung*^{-/-}*Msh2*^{-/-} mice, this profile can be simply explained by an increased error-free repair mediated by UNG in the absence of MSH2–MSH6, thus reducing the overall number of mutations⁶⁵. Inefficient displacement of AID from its target DNA would explain the increase in G/C transition mutations (uracils being carried over to replication), especially at a few hotspot positions showing a symmetrical WGCW structure (W standing for A or T). The residual A/T mutagenesis appears to be generated by an error-prone repair mobilizing UNG and Polη, as clearly shown by the complete absence of A/T mutations in MSH2/Polη double-deficient mice⁶⁵ (Fig.4), but whether this UNG-mediated error-prone repair is artificially caused by the absence of MSH2 is an open question.

A model for SHM: two competitive rather than alternative pathways

The following model attempts to bring together data obtained from mice deficient in TLS polymerases, mismatch repair, Exo1 and UNG (Fig.5).

In this model, AID attacks the V gene segment undergoing SHM, introducing uracils on both strands of the DNA most probably during a few rounds of replication. Because AID is expressed throughout the cell cycle¹⁰⁵, this attack can occur in the different phases of the cell cycle. Uracils generated during the G1 phase are preferentially recognized as U-G mismatches by the MSH2–MSH6 complex. If this complex is absent, UNG will process most of the uracils and repair them error free. In wild-type cells, the MSH2–MSH6 complex recruits Exo1¹⁰³, to excise a patch of DNA from a single-strand nick, the origin of which is yet unknown, as well as Polη, which results in mutagenic synthesis of the strand targeted by AID. AID deaminates 60% of Cs on the coding strand and 40% on the non-coding strand⁹⁸, whereas the G:C mutation ratio is reequilibrated after Polη intervention (Fig.5). Together with As

being mutated about twice as frequently as Ts, this suggests that Pol η exerts its activity more frequently on the coding strand, and possibly extends a longer patch of DNA on this strand. The equivalent mutation frequency observed at the V gene segment for wild-type and MSH2/UNG-deficient mice suggests that, in general, every uracil that is repaired by the mismatch-repair pathway gives rise to a distant mutation. It has been proposed recently, using a transgenic model, that A/T mutations are only produced on targeting of the coding strand¹⁰⁶. Whereas such a strand-specific repair would be compatible with the equalization of mutations between G and C bases, and the A over T bias observed in V gene mutations (Fig.4), the proposition deserves additional studies, as the mutations reported do not fit with the pattern expected from Pol η -mediated repair⁶⁰.

Uracils that remain during replication or those that are generated in S phase, while present in single-stranded DNA, will be processed by UNG. As the replication fork proceeds, it will stall opposite the abasic site generated by UNG and will recruit TLS polymerases in a classical TLS bypass process, although it is surprisingly largely independent of PCNA monoubiquitination. So the question remains: which polymerases, besides Rev1, might create mutations at these abasic sites? Transition mutations could be produced opposite uracils by the replicative polymerases, but it is also possible that UNG might excise most of them and that several TLS polymerases could follow the “A rule” opposite an abasic site (a default pathway leading to insertion of A opposite a lesion⁴), and thus also produce G to A and C to T transition mutations. Pol ζ could also be involved in the bypass of these abasic sites, and a role for replicative polymerases has been proposed as well¹⁰⁷.

Rather than representing two alternative pathways for the handling of AID-mediated deaminations, we therefore propose that the mismatch repair and UNG-mediated pathways are competitive, with the MSH2–MSH6 complex restricting the role of UNG in the S phase to the single-stranded configuration of uracils.

CSR, gene conversion and DNA polymerases

CSR occurs in activated B cells and allows the exchange of heavy chain constant region genes, from C μ to other downstream C regions (Fig.1). Specific sequences (switch regions) that are located upstream of each constant region are involved in this recombination event that requires the deamination of cytidine residues in these switch regions by AID. Uracils are further processed by UNG and by mismatch-repair factors, the former having a more

important role in CSR than in SHM, suggesting that the competition between these two activities might differ during both processes.

Pol η -dependent mutations have been observed upstream of switch junctions in human and mouse B cells, but there does not seem to be any quantitative alteration in CSR in the absence of this enzyme^{55-57, 108}. As DNA sequences surrounding switch junctions have no function once CSR has occurred, these mutations are likely to have no physiological role. It was recently reported that, during CSR, error-free repair by Pol β after uracil excision could compete with the generation of the DSBs that are required to initiate recombination¹⁰⁹. Pol η (as well as Pol κ if used as a back-up enzyme) lacks deoxyribose-phosphate-lyase (dRP-lyase) activity to remove the sugar moiety left at an abasic site after incision of the sugar-phosphate DNA backbone, whereas Pol β has such an activity, which allows it to contribute efficiently to the base-excision repair process. Absence of this dRP-lyase activity could slow down the ligation step after DNA synthesis by Pol η and thus increase the occurrence of two nearby nicks in DNA, generating DSBs.

Gene conversion is a non-reciprocal recombination event used by some species, such as chickens and rabbits, to generate their pre-immune repertoire (Fig.1). V(D)J recombination remains the initial event in these species, but it involves either a single or a restricted number of V genes. Rearranged heavy- and light-chain genes are thereafter diversified by templated gene conversion during the extensive proliferation of B cells in gut-associated lymphoid tissues (bursa of Fabricius, Peyer's patches and appendix)³. A virally transformed bursal B-cell line, DT40, was reported to undergo ongoing gene conversion at its light chain locus in culture^{110, 111}, a process strictly dependent on AID activity^{112, 113}. The DT40 cell line has also become a popular model for studying hypermutation, notably upon removal of the pool of donor genes, thus forcing the outcome of AID-induced deamination towards untemplated mutagenesis¹¹⁴. However, the characteristics of this mutation process appear to diverge substantially from physiological SHM (Box 1).

Gene conversion was significantly reduced in the chicken DT40 cell line after deletion of Pol η ¹¹⁵. Considering the error rate of this polymerase during the copying of a normal DNA template and the low frequency of untemplated mutations within gene conversion tracts, this result appears rather paradoxical. One could imagine that the mutations introduced by Pol η will be corrected by mismatch repair, but the rationale for selecting an error-prone polymerase for this process remains difficult to assess. It is tempting to speculate that Pol η is recruited by AID together with MSH2–MSH6 at U/G mismatches, this complex being somehow mobilized

as a whole, whatever the outcome of the repair process¹¹⁶. Alternatively, this could reflect a more general role of Polη in homologous recombination, as recently described *in vitro*¹¹⁷.

Conclusion

Bacteria use TLS polymerases to avoid irreversible replication blocks when other damage avoiding strategies have failed⁷. By so doing, mutagenesis is favoured over cell death, and it has been widely discussed that mutator phenotypes have an adaptive value in prokaryotes^{118, 119}. The paradigm of the XPV syndrome, in which a TLS polymerase deficiency results in sunlight sensitivity and an increased risk of skin cancers, suggests an alternative scenario in higher organisms. Amplification of polymerases in higher vertebrates may have evolved to ensure a specialization towards defined lesions for which each polymerase would be selected for both efficiency and accuracy, thus avoiding deleterious mutations that could lead to cancer in multicellular organisms. Strikingly, the adaptive immune system mobilizes TLS polymerases in their ancient mutagenic role, mirroring during its fight against pathogens the stress-induced adaptability of the prokaryotic world.

BOX 1 | SHM in the DT40 cell line: “passive” versus “active” mutagenesis

The DT40 cell line is a chicken bursal lymphoma that undergoes high rate of activation-induced cytidine deaminase (AID)-dependent gene conversion in culture¹¹⁰⁻¹¹³. Deletion of the all pseudogenes upstream of the single functional V λ gene deprives this cell line of all donor sequences required in *cis* for gene conversion, and shifts the V-targeted modifications towards mutations¹¹⁴. Mutagenesis observed in DT40 cells appears essentially “passive”, occurring by saturation of the uracil-DNA glycosylase (UNG) repair pathway. Accordingly, deletion of *Ung* in DT40 cells results in a sevenfold increase in mutagenesis¹²⁰, implying that a large fraction of the uracils generated by AID are repaired error-free, by base-excision repair or by sister chromatid exchange. All mutations observed are targeted to G or C bases, being almost exclusively transition mutations or G to C and C to G transversion mutations: they thus appear to be generated during replication by copy of unrepaired uracils or by bypass of abasic sites, performed by Rev1 in a translesion process that appears strictly dependent on PCNA monoubiquitination^{121, 122}. This cell line therefore does not mimic the “active” SHM process — in which the MSH2–MSH6 complex recruits Pol η to generate mutations at A/T bases and the bypass of abasic sites involves TLS polymerases other than Rev1 — but it nevertheless reproduces the initial steps of SHM consisting in the targeting of AID to the V region.

Glossary

BCR

B cell receptor. Refers to the Ig molecule present at the B cell surface, irrespectively of its heavy chain isotype.

V(D)J recombination

Somatic rearrangement of variable (V), diversity (D) and joining (J) regions of the genes that encode antigen receptors, leading to repertoire diversity of both T-cell and B-cell receptors.

gene conversion

A non-reciprocal homologous recombination event in which the donor gene(s) remains unmodified and an acceptor gene acquires the recombined segment. In chickens, variable (V) pseudogenes are donors that modify the functional, rearranged V gene in follicles of the bursa of Fabricius to generate a diverse pre-immune repertoire.

somatic hypermutation

A process by which antigen-activated B cells in germinal centres acquire point mutations targeted to the variable regions of rearranged immunoglobulin gene segments. The B cells are subsequently selected for those expressing the 'best' mutations on the basis of the ability of the surface immunoglobulin to bind antigen. This process occurs in activated B cells in all jawed vertebrates, but it also occurs in immature B cells in sheep to generate the pre-immune repertoire.

class-switch recombination

(Class or isotype switching). A region-specific recombination process that occurs in antigen-activated B cells. This takes place between 'switch region' DNA sequences located upstream from each immunoglobulin heavy chain constant region genes and results in a change from the IgM to one of the IgG, IgA or IgE immunoglobulin isotypes. This imparts flexibility to the humoral immune response and allows it to exploit the different capacities of the immunoglobulins to activate the appropriate downstream effector mechanisms.

translesion DNA synthesis

(TLS). Process that allows the bypass of DNA damage that otherwise blocks the progression of the replication fork, through the replacement of the replicative polymerase with specialized polymerases (known as TLS polymerases) that can copy non-instructive DNA lesions. TLS polymerases have low fidelity on undamaged templates. Most TLS polymerases belong to the Y family polymerases.

recombination-activating genes

(*RAG1* and *RAG2*). *RAG1* and *RAG2* are involved in creating the synapsis and the double-strand DNA breaks required for the assembly of the dispersed gene segments that encode the complete protein chains of B-cell and T-cell receptors.

synapsis

Non-covalent juxtaposition of two non-adjacent stretches of DNA.

recombination signal sequences

(RSSs). Conserved elements that constitute recognition sites for the V(D)J recombinase proteins, which are encoded by the recombination-activating gene 1 (*RAG1*) and *RAG2*. They consist of a palindromic heptamer that is immediately adjacent to the coding gene segments — V (variable), D (diversity) or J (joining) — and is separated by a 12- or 23-base-pair spacer from a conserved nonamer sequence.

non-homologous end joining

(NHEJ). The process that joins broken DNA ends without depending on extended homology. Components of this pathway include the proteins Ku70, Ku80, Artemis, X-ray repair cross-complementing protein 4 (XRCC4), DNA ligase IV, Cernunnos/XLF and the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs).

complementarity-determining regions

(CDRs). The three hypervariable antigen-receptor regions (in both chains of the T-cell and B-cell receptors) that interact with the antigen. The third CDR (CDR3) is partly encoded by the germline variable (V), diversity (D) and joining (J) regions of each receptor chain. Extensive diversity is generated in CDR3 during gene rearrangement by nucleotide trimming and/or template-independent nucleotide additions by terminal deoxynucleotidyltransferase.

base-excision repair

A DNA-repair pathway that removes single non-canonical bases from the DNA, such as deaminated or oxidized bases, and replaces them with an appropriate base templated on the opposite strand. Repair is initiated by a DNA glycosylase that is specialized for a particular type of damage, and nucleotide replacement is performed by DNA polymerase β .

deamination

Removal of an amine group from a pyrimidine or purine nucleic-acid base. Deamination of cytosine and adenosine yields uracil and inosine, respectively.

mismatch repair

A repair pathway that recognizes mismatched bases that arise in DNA because of errors made by replicative DNA polymerases. These are then repaired by an excision system that removes a tract of DNA including the mismatch, and re-copies the original strand. The mismatch repair complex includes a mismatch-binding moiety (such as MSH2–MSH6), and an effector part (such as MLH1–PMS2) that triggers excision of the mismatch-containing DNA sequence.

hotspot motif

A short DNA motif (DGYW or WRCH; where D denotes adenosine (A), guanosine (G) or thymidine (T); Y denotes cytidine (C) or T; W denotes A or T; R denotes A or G; and H denotes T, C or A) at which mutations are preferentially targeted during somatic hypermutation.

replication fork

Site in double-stranded DNA at which the template strands are separated, allowing a newly formed copy of the DNA to be synthesized, with the fork moving in the direction of leading strand synthesis.

abasic site

A common form of DNA damage in which a base is lost from a strand of DNA, spontaneously or by the action of DNA repair enzymes such as uracil glycosylase, while leaving the phosphodiester bond intact.

xeroderma pigmentosum

(XP). A rare inherited human disorder, in which patients are sensitive to the DNA-damaging effects of sunlight. XP can be caused by disabling any of eight different genes. Seven of the genes, denoted XPA–XPG, encode components of the nucleotide excision-repair pathway. A variant form of the syndrome, XPV, corresponds to the inactivation of the gene encoding DNA polymerase η .

transition mutations

Base changes in DNA in which a pyrimidine (C or T) is replaced by another pyrimidine, or a purine (A or G) is replaced by another purine.

transversion mutations

Base changes in DNA in which a pyrimidine (C or T) is replaced by a purine (A or G), or a purine is replaced by a pyrimidine.

proliferating-cell nuclear antigen

(PCNA). PCNA was first identified as a DNA sliding clamp for replicative polymerases, but is now known to coordinate the organization of protein partners that are involved in many processes, including DNA replication, DNA repair and cell-cycle control.

Acknowledgements

We thank our colleagues at INSERM U783 for contributing to this review by many discussions and for sharing unpublished data. We apologize to those whose work could not be cited because of space limitations. C.-A. R. is Director of Research of the Centre National de la Recherche Scientifique.

Figure legends.**Figure 1. DNA transactions during the formation of the B cell repertoire.**

The pre-immune B-cell repertoire in mice and humans is generated by site-specific recombination of V (variable), D (diversity) and J (joining) coding segments that make up the antigen receptor. A second wave of DNA modifications occurs in B cells participating in an

immune response. These consist of locus-specific point mutagenesis (known as somatic hypermutation), which modifies the rearranged V region, and a recombination process between non-homologous repeated sequences (switch (S) regions) located upstream constant (C) region genes (known as class-switch recombination). Somatic hypermutation and/or gene conversion, which involves the exchange of gene segments between a rearranged V gene and donor sequences (often non-functional (pseudogenes, ψV)), contribute to the formation of the pre-immune repertoire in species such as chickens, sheep and rabbits.

Figure 2. Junctional diversification during V(D)J recombination at the immunoglobulin locus.

During V(D)J recombination at the immunoglobulin locus, two pairs of coding segments with defined signal sequences (orange and green triangles) come together to form a synapsis. Double-strand breaks are introduced into the DNA, resulting in the generation of coding ends that terminate in a hairpin and two signal ends that have blunt ends. The hairpin coding ends are then opened by the Artemis-DNA-PKcs complex and, before the ends are religated by the XRCC4-Cernunnos-ligase IV complex, extensive processing of DNA ends takes place, creating diversity of the coding joint. The coding ends are subjected to nucleotide trimming, end-filling and/or protection by DNA polymerases Pol λ and N-nucleotide addition by terminal deoxynucleotidyltransferase (TdT) (heavy chain) or Pol μ (light chain). Modifications at the coding junction are represented as a black bar. Little modification occurs during signal end joining.

Figure 3. Possible roles of DNA polymerases at V(D)J junctions.

Three possible functions of DNA polymerases during V(D)J recombination are shown, arbitrarily represented as symmetrical configurations: a role in protecting DNA ends, by competition with exonucleolytic degradation; a fill-in synthesis function, generating blunt DNA ends; a role in the bridging of 3' protruding ends. Pol μ and pol λ have been shown to be able to extend DNA ends in such a primer-template configuration that shows lower stability, since the primer encompasses a nicked DNA template.

Figure 4. The C over G targeting bias mediated by AID (60 : 40) is equalized by the mismatch-repair pathway.

The pattern of nucleotide substitutions observed in mutated intronic J_{H4} sequences from

PNA^{high} Peyer's patch B cells is shown for the genetic backgrounds indicated, and expressed as percent of total mutations observed. Values are corrected for base composition, i.e. normalized for a sequence with an equal content of the four bases. Data from *Ung*^{-/-}*Msh2*^{-/-} and *Ung*^{-/-} cells were obtained from Rada *et al.*⁹⁸, all other data are taken from Delbos *et al.*⁶⁵. Deamination by activation-induced cytidine deaminase (AID) generates a C over G mutation bias (60:40, as observed in *Ung*^{-/-}*Msh2*^{-/-} mice). This bias is equalized in MSH2-proficient animals, and result in a two-fold higher targeting of A versus T mutations, whereas it is conserved in MSH2-deficient mice.

Figure 5. Two competitive pathways for hypermutation: the MSH2-MSH6 complex prevents UNG from performing error-free repair.

The uracils produced by activation-induced cytidine deaminase (AID)-mediated cytidine deamination are preferentially handled by the mismatch repair complex MSH2–MSH6, when present as a U-G mismatch (for example in the G1 phase). This prevents uracil-DNA glycosylase (UNG) from performing error-free repair. A patch of DNA is excised by exonuclease 1 (Exo1), and is resynthesized with low fidelity by Polη. Uracils that are carried over into replication or that are generated in the S phase, and present in single-stranded DNA, are processed by UNG, to generate abasic sites in the absence of a repair template. These abasic sites are bypassed by several translesion polymerases, such as Rev1 and others that remain to be identified. Rather than splitting hypermutation in phase I and phase II according to Neuberger and colleagues⁴¹, this model proposes an A/T vs. G/C mutation model that is essentially cell-cycle and/or DNA structure dependent.

References

1. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575-81 (1983).
2. Rajewsky, K. Clonal selection and learning in the antibody system. *Nature* **381**, 751-8 (1996).
3. Weill, J.C. & Reynaud, C.A. Rearrangement/hypermutation/gene conversion: when, where and why? *Immunol Today* **17**, 92-7 (1996).
4. Goodman, M.F. & Tiffin, B. The expanding polymerase universe. *Nat Rev Mol Cell Biol* **1**, 101-9 (2000).
5. Burgers, P.M. et al. Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J Biol Chem* **276**, 43487-90 (2001).
6. Hubscher, U., Maga, G. & Spadari, S. Eukaryotic DNA polymerases. *Annu Rev Biochem* **71**, 133-63 (2002).
7. Friedberg, E.C., Wagner, R. & Radman, M. Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science* **296**, 1627-30 (2002).
8. Gellert, M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* **71**, 101-32 (2002).
9. Schatz, D.G. & Spanopoulou, E. Biochemistry of V(D)J recombination. *Curr Top Microbiol Immunol* **290**, 49-85 (2005).
10. Revy, P., Malivert, L. & de Villartay, J.P. Cernunnos-XLF, a recently identified non-homologous end-joining factor required for the development of the immune system. *Curr Opin Allergy Clin Immunol* **6**, 416-20 (2006).
11. Mahajan, K.N., Nick McElhinny, S.A., Mitchell, B.S. & Ramsden, D.A. Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. *Mol Cell Biol* **22**, 5194-202 (2002).
12. Ma, Y. et al. A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol Cell* **16**, 701-13 (2004).
13. Sobol, R.W. & Wilson, S.H. Mammalian DNA beta-polymerase in base excision repair of alkylation damage. *Prog Nucleic Acid Res Mol Biol* **68**, 57-74 (2001).
14. Wilson, T.E. & Lieber, M.R. Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymerase beta (Pol4)-dependent pathway. *J Biol Chem* **274**, 23599-609 (1999).
15. Aoufouchi, S. et al. Two novel human and mouse DNA polymerases of the polX family. *Nucleic Acids Res* **28**, 3684-93 (2000).
16. Dominguez, O. et al. DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *Embo J* **19**, 1731-42 (2000).
17. Garcia-Diaz, M. et al. DNA polymerase lambda (Pol lambda), a novel eukaryotic DNA polymerase with a potential role in meiosis. *J Mol Biol* **301**, 851-67 (2000).
18. Lee, J.W. et al. Implication of DNA polymerase lambda in alignment-based gap filling for nonhomologous DNA end joining in human nuclear extracts. *J Biol Chem* **279**, 805-11 (2004).
19. Daley, J.M., Laan, R.L., Suresh, A. & Wilson, T.E. DNA joint dependence of pol X family polymerase action in nonhomologous end joining. *J Biol Chem* **280**, 29030-7 (2005).
20. Nick McElhinny, S.A. et al. A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol Cell* **19**, 357-66 (2005).

21. Bertocci, B., De Smet, A., Weill, J.C. & Reynaud, C.A. Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo. *Immunity* **25**, 31-41 (2006).
22. Thai, T.H., Purugganan, M.M., Roth, D.B. & Kearney, J.F. Distinct and opposite diversifying activities of terminal transferase splice variants. *Nat Immunol* **3**, 457-62 (2002).
23. Repasky, J.A., Corbett, E., Boboila, C. & Schatz, D.G. Mutational analysis of terminal deoxynucleotidyltransferase-mediated N-nucleotide addition in V(D)J recombination. *J Immunol* **172**, 5478-88 (2004).
24. Doyen, N., Boule, J.B., Rougeon, F. & Papanicolaou, C. Evidence that the long murine terminal deoxynucleotidyltransferase isoform plays no role in the control of V(D)J junctional diversity. *J Immunol* **172**, 6764-7 (2004).
25. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M.R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781-94 (2002).
26. Moshous, D. et al. Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **105**, 177-86 (2001).
27. Yang, Y.G., Lindahl, T. & Barnes, D.E. Trex1 exonuclease degrades ssDNA to prevent chronic checkpoint activation and autoimmune disease. *Cell* **131**, 873-86 (2007).
28. Bertocci, B., De Smet, A., Berek, C., Weill, J.C. & Reynaud, C.A. Immunoglobulin kappa light chain gene rearrangement is impaired in mice deficient for DNA polymerase mu. *Immunity* **19**, 203-11 (2003).
29. Tomlinson, I.M., Cox, J.P., Gherardi, E., Lesk, A.M. & Chothia, C. The structural repertoire of the human V kappa domain. *Embo J* **14**, 4628-38 (1995).
30. Lederberg, J. Genes and antibodies. *Science* **129**, 1649-53 (1959).
31. Burnett, F. The clonal selection theory of acquired immunity (The University Press, Cambridge, 1959).
32. Weigert, M.G., Cesari, I.M., Yonkovich, S.J. & Cohn, M. Variability in the lambda light chain sequences of mouse antibody. *Nature* **228**, 1045-7 (1970).
33. Bothwell, A.L. et al. Heavy chain variable region contribution to the NPb family of antibodies: somatic mutation evident in a gamma 2a variable region. *Cell* **24**, 625-37 (1981).
34. Gearhart, P.J., Johnson, N.D., Douglas, R. & Hood, L. IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature* **291**, 29-34 (1981).
35. Selsing, E. & Storb, U. Somatic mutation of immunoglobulin light-chain variable-region genes. *Cell* **25**, 47-58 (1981).
36. Griffiths, G.M., Berek, C., Kaartinen, M. & Milstein, C. Somatic mutation and the maturation of immune response to 2-phenyl oxazolone. *Nature* **312**, 271-5 (1984).
37. Brenner, S. & Milstein, C. Origin of antibody variation. *Nature* **211**, 242-3 (1966).
38. Muramatsu, M. 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* **274**, 18470-6 (1999).
39. Muramatsu, M. et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553-63 (2000).

40. Revy, P. et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**, 565-75 (2000).
41. Di Noia, J.M. & Neuberger, M.S. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* **76**, 1-22 (2007).
42. Reynaud, C.A. et al. Mismatch repair and immunoglobulin gene hypermutation: did we learn something? *Immunol Today* **20**, 522-7 (1999).
43. Martin, A. & Scharff, M.D. AID and mismatch repair in antibody diversification. *Nat Rev Immunol* **2**, 605-14 (2002).
44. Martomo, S.A. & Gearhart, P.J. Somatic hypermutation: subverted DNA repair. *Curr Opin Immunol* **18**, 243-8 (2006).
45. Esposito, G. et al. Mice reconstituted with DNA polymerase beta-deficient fetal liver cells are able to mount a T cell-dependent immune response and mutate their Ig genes normally. *Proc Natl Acad Sci U S A* **97**, 1166-71 (2000).
46. Prakash, S., Johnson, R.E. & Prakash, L. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem* **74**, 317-53 (2005).
47. Goodman, M.F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* **71**, 17-50 (2002).
48. Friedberg, E.C. Suffering in silence: the tolerance of DNA damage. *Nat Rev Mol Cell Biol* **6**, 943-53 (2005).
49. Zhang, Y. et al. Lesion bypass activities of human DNA polymerase mu. *J Biol Chem* **277**, 44582-7 (2002).
50. Maga, G. et al. 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. *Nature* **447**, 606-8 (2007).
51. Seki, M. et al. High-efficiency bypass of DNA damage by human DNA polymerase Q. *Embo J* **23**, 4484-94 (2004).
52. Masutani, C. et al. The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature* **399**, 700-4 (1999).
53. Johnson, R.E., Kondratik, C.M., Prakash, S. & Prakash, L. hRAD30 mutations in the variant form of xeroderma pigmentosum. *Science* **285**, 263-5 (1999).
54. Zeng, X. et al. DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol* **2**, 537-41 (2001).
55. Faili, A. et al. DNA polymerase eta is involved in hypermutation occurring during immunoglobulin class switch recombination. *J Exp Med* **199**, 265-70 (2004).
56. Delbos, F. et al. Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. *J Exp Med* **201**, 1191-6 (2005).
57. Martomo, S.A. et al. Different mutation signatures in DNA polymerase eta- and MSH6-deficient mice suggest separate roles in antibody diversification. *Proc Natl Acad Sci U S A* **102**, 8656-61 (2005).
58. Martomo, S.A. et al. Normal hypermutation in antibody genes from congenic mice defective for DNA polymerase iota. *DNA Repair (Amst)* **5**, 392-8 (2006).
59. Matsuda, T. et al. Error rate and specificity of human and murine DNA polymerase eta. *J Mol Biol* **312**, 335-46 (2001).
60. Pavlov, Y.I. et al. Correlation of somatic hypermutation specificity and A-T base pair substitution errors by DNA polymerase eta during copying of a mouse immunoglobulin kappa light chain transgene. *Proc Natl Acad Sci U S A* **99**, 9954-9 (2002).
61. Ohashi, E. et al. Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene. *J Biol Chem* **275**, 39678-84 (2000).

62. Zhang, Y. et al. Error-free and error-prone lesion bypass by human DNA polymerase kappa in vitro. *Nucleic Acids Res* **28**, 4138-46 (2000).
63. Schenten, D. et al. DNA polymerase kappa deficiency does not affect somatic hypermutation in mice. *Eur J Immunol* **32**, 3152-60 (2002).
64. Shimizu, T., Shinkai, Y., Ogi, T., Ohmori, H. & Azuma, T. The absence of DNA polymerase kappa does not affect somatic hypermutation of the mouse immunoglobulin heavy chain gene. *Immunol Lett* **86**, 265-70 (2003).
65. Delbos, F., Aoufouchi, S., Faili, A., Weill, J.C. & Reynaud, C.A. DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J Exp Med* **204**, 17-23 (2007).
66. Nelson, J.R., Lawrence, C.W. & Hinkle, D.C. Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* **382**, 729-31 (1996).
67. Zhang, Y. et al. Response of human REV1 to different DNA damage: preferential dCMP insertion opposite the lesion. *Nucleic Acids Res* **30**, 1630-8 (2002).
68. Guo, C. et al. Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. *Embo J* **22**, 6621-30 (2003).
69. Jansen, J.G. et al. The BRCT domain of mammalian Rev1 is involved in regulating DNA translesion synthesis. *Nucleic Acids Res* **33**, 356-65 (2005).
70. Jansen, J.G. et al. Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J Exp Med* **203**, 319-23 (2006).
71. Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S. & Prakash, L. Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature* **406**, 1015-9 (2000).
72. Esposito, G. et al. Disruption of the Rev3l-encoded catalytic subunit of polymerase zeta in mice results in early embryonic lethality. *Curr Biol* **10**, 1221-4 (2000).
73. Bemark, M., Khamlichi, A.A., Davies, S.L. & Neuberger, M.S. Disruption of mouse polymerase zeta (Rev3) leads to embryonic lethality and impairs blastocyst development in vitro. *Curr Biol* **10**, 1213-6 (2000).
74. Wittschieben, J. et al. Disruption of the developmentally regulated Rev3l gene causes embryonic lethality. *Curr Biol* **10**, 1217-20 (2000).
75. Zhong, X. et al. The fidelity of DNA synthesis by yeast DNA polymerase zeta alone and with accessory proteins. *Nucleic Acids Res* **34**, 4731-42 (2006).
76. Diaz, M., Verkoczy, L.K., Flajnik, M.F. & Klinman, N.R. Decreased frequency of somatic hypermutation and impaired affinity maturation but intact germinal center formation in mice expressing antisense RNA to DNA polymerase zeta. *J Immunol* **167**, 327-35 (2001).
77. Zan, H. et al. The translesion DNA polymerase zeta plays a major role in Ig and bcl-6 somatic hypermutation. *Immunity* **14**, 643-53 (2001).
78. Schenten, D., Rajewsky, K., personal communication.
79. Tissier, A., McDonald, J.P., Frank, E.G. & Woodgate, R. poliota, a remarkably error-prone human DNA polymerase. *Genes Dev* **14**, 1642-50 (2000).
80. Frank, E.G. & Woodgate, R. Increased catalytic activity and altered fidelity of human DNA polymerase iota in the presence of manganese. *J Biol Chem* **282**, 24689-96 (2007).
81. Sale, J.E. & Neuberger, M.S. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity* **9**, 859-69 (1998).
82. Denepoux, S. et al. Induction of somatic mutation in a human B cell line in vitro. *Immunity* **6**, 35-46 (1997).

83. Zan, H. et al. Induction of Ig somatic hypermutation and class switching in a human monoclonal IgM+ IgD+ B cell line in vitro: definition of the requirements and modalities of hypermutation. *J Immunol* **162**, 3437-47 (1999).
84. Faili, A. et al. AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat Immunol* **3**, 815-21 (2002).
85. Xiao, Z. et al. Known components of the immunoglobulin A:T mutational machinery are intact in Burkitt lymphoma cell lines with G:C bias. *Mol Immunol* **44**, 2659-66 (2007).
86. Faili, A. et al. Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase iota. *Nature* **419**, 944-7 (2002).
87. McDonald, J.P. et al. 129-derived strains of mice are deficient in DNA polymerase iota and have normal immunoglobulin hypermutation. *J Exp Med* **198**, 635-43 (2003).
88. Poltoratsky, V., Prasad, R., Horton, J.K. & Wilson, S.H. Down-regulation of DNA polymerase beta accompanies somatic hypermutation in human BL2 cell lines. *DNA Repair (Amst)* **6**, 244-53 (2007).
89. Zan, H. et al. The translesion DNA polymerase theta plays a dominant role in immunoglobulin gene somatic hypermutation. *Embo J* **24**, 3757-69 (2005).
90. Masuda, K. et al. Absence of DNA polymerase theta results in decreased somatic hypermutation frequency and altered mutation patterns in Ig genes. *DNA Repair (Amst)* **5**, 1384-91 (2006).
91. Masuda, K. et al. DNA polymerases eta and theta function in the same genetic pathway to generate mutations at A/T during somatic hypermutation of Ig genes. *J Biol Chem* **282**, 17387-94 (2007).
92. Masuda, K. et al. DNA polymerase theta contributes to the generation of C/G mutations during somatic hypermutation of Ig genes. *Proc Natl Acad Sci U S A* **102**, 13986-91 (2005).
93. Kannouche, P.L., Wing, J. & Lehmann, A.R. Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Mol Cell* **14**, 491-500 (2004).
94. Warbrick, E. PCNA binding through a conserved motif. *Bioessays* **20**, 195-9 (1998).
95. Bienko, M. et al. Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* **310**, 1821-4 (2005).
96. Langerak, P., Nygren, A.O., Krijger, P.H., van den Berk, P.C. & Jacobs, H. A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J Exp Med* **204**, 1989-98 (2007).
97. Moldovan, G.L., Pfander, B. & Jentsch, S. PCNA, the maestro of the replication fork. *Cell* **129**, 665-79 (2007).
98. Rada, C., Di Noia, J.M. & Neuberger, M.S. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* **16**, 163-71 (2004).
99. Rada, C. et al. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* **12**, 1748-55 (2002).
100. Wiesendanger, M., Kneitz, B., Edelmann, W. & Scharff, M.D. 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* **191**, 579-84 (2000).
101. Martomo, S.A., Yang, W.W. & Gearhart, P.J. A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. *J Exp Med* **200**, 61-8 (2004).

102. Shen, H.M., Tanaka, A., Bozek, G., Nicolae, D. & Storb, U. Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice. *J Immunol* **177**, 5386-92 (2006).
103. Bardwell, P.D. et al. Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat Immunol* **5**, 224-9 (2004).
104. Phung, Q.H., Winter, D.B., Alrefai, R. & Gearhart, P.J. Hypermutation in Ig V genes from mice deficient in the MLH1 mismatch repair protein. *J Immunol* **162**, 3121-4 (1999).
105. Aoufouchi, A., Faili, A., Zober, C., D'Orlando, O., Weller, S., Weill, J.-C. and Reynaud, C.-A. Ubiquitin-mediated degradation restricts the nuclear lifespan of AID. *submitted*.
106. Unniraman, S. & Schatz, D.G. Strand-biased spreading of mutations during somatic hypermutation. *Science* **317**, 1227-30 (2007).
107. Avkin, S., Adar, S., Blander, G. & Livneh, Z. Quantitative measurement of translesion replication in human cells: evidence for bypass of abasic sites by a replicative DNA polymerase. *Proc Natl Acad Sci U S A* **99**, 3764-9 (2002).
108. Zeng, X., Negrete, G.A., Kasmer, C., Yang, W.W. & Gearhart, P.J. Absence of DNA polymerase eta reveals targeting of C mutations on the nontranscribed strand in immunoglobulin switch regions. *J Exp Med* **199**, 917-24 (2004).
109. Wu, X. & Stavnezer, J. DNA polymerase beta is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination. *J Exp Med* **204**, 1677-89 (2007).
110. Buerstedde, J.M. et al. Light chain gene conversion continues at high rate in an ALV-induced cell line. *Embo J* **9**, 921-7 (1990).
111. Kim, S., Humphries, E.H., Tjoelker, L., Carlson, L. & Thompson, C.B. Ongoing diversification of the rearranged immunoglobulin light-chain gene in a bursal lymphoma cell line. *Mol Cell Biol* **10**, 3224-31 (1990).
112. Harris, R.S., Sale, J.E., Petersen-Mahrt, S.K. & Neuberger, M.S. AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Curr Biol* **12**, 435-8 (2002).
113. Arakawa, H., Hauschild, J. & Buerstedde, J.M. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* **295**, 1301-6 (2002).
114. Arakawa, H., Saribasak, H. & Buerstedde, J.M. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biol* **2**, E179 (2004).
115. Kawamoto, T. et al. Dual roles for DNA polymerase eta in homologous DNA recombination and translesion DNA synthesis. *Mol Cell* **20**, 793-9 (2005).
116. Reynaud, C.A., Aoufouchi, S., Faili, A. & Weill, J.C. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nat Immunol* **4**, 631-8 (2003).
117. McIlwraith, M.J. et al. Human DNA polymerase eta promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Mol Cell* **20**, 783-92 (2005).
118. Taddei, F. et al. Role of mutator alleles in adaptive evolution. *Nature* **387**, 700-2 (1997).
119. Yeiser, B., Pepper, E.D., Goodman, M.F. & Finkel, S.E. SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. *Proc Natl Acad Sci U S A* **99**, 8737-41 (2002).
120. Saribasak, H. et al. Uracil DNA glycosylase disruption blocks Ig gene conversion and induces transition mutations. *J Immunol* **176**, 365-71 (2006).

121. Simpson, L.J. & Sale, J.E. Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line. *Embo J* **22**, 1654-64 (2003).
122. Arakawa, H. et al. A role for PCNA ubiquitination in immunoglobulin hypermutation. *PLoS Biol* **4**, e366 (2006).

Table 1: Amplification of DNA polymerases in higher vertebrates

Polymerase family	Polymerase name (gene name)	Yeast (<i>S. cerevisiae</i>)	Nematode (<i>C. elegans</i>)	Fishes (<i>Danio rerio</i>)	Birds (<i>Gallus gallus</i>)	Mammals (<i>Homo sapiens</i>)	Relevant DNA transaction ^b (birds or mammals)
A	Pol gamma (PolG)	+	+	+	+	+	mitochondrial DNA replication TLS ? BER ? TLS ?
	Pol theta (PolQ)	-	+	+	+	+	
	Pol nu (PolN)	-	-	+	+	+	
B	Pol alpha (PolA)	+	+	+	+	+	replication priming replication replication TLS , SHM?
	Pol delta (PolD)	+	+	+	+	+	
	Pol epsilon (PolE)	+	+	+	+	+	
	Pol zeta (PolZ)	S ^a	S	L	L	L	
X	Pol beta (PolB)	-	-	+	+	+	base excision repair (BER) BER, NHEJ (V-D-J recombination), TLS ? NHEJ (V-J recombination), TLS ? V-D-J recombination
	Pol lambda (PolL)	Pol4	-	+	+	+	
	Pol mu (PolM)	-	-	+	-	+	
	TdT	-	-	+	+	+	
Y	Pol eta (PolH)	+	+	+	+	+	TLS, SHM, GC ?, homologous recombination ? TLS TLS, NER ? TLS, SHM
	Pol iota (PolI)	-	-	+	-	+	
	Pol kappa (PolK)	-	+	+	+	+	
	Rev1	+	+	+	+	+	

^a S = small Rev3, protein size similar to yeast, approximately 1500 amino acids

L = large Rev3, protein size similar to humans, approximately 3000 amino acids

^b “?” refers to processes mainly documented in vitro

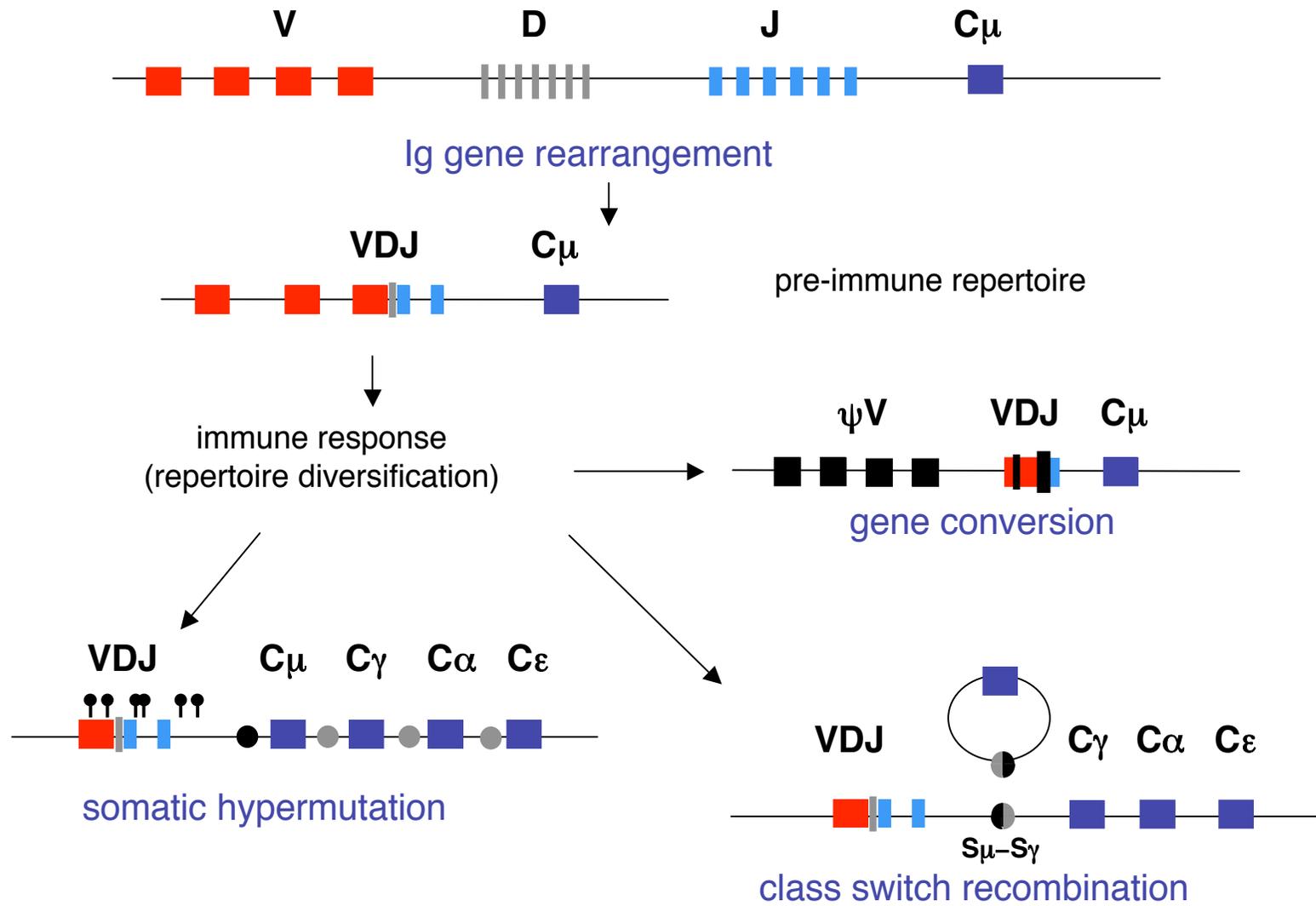


Figure 1

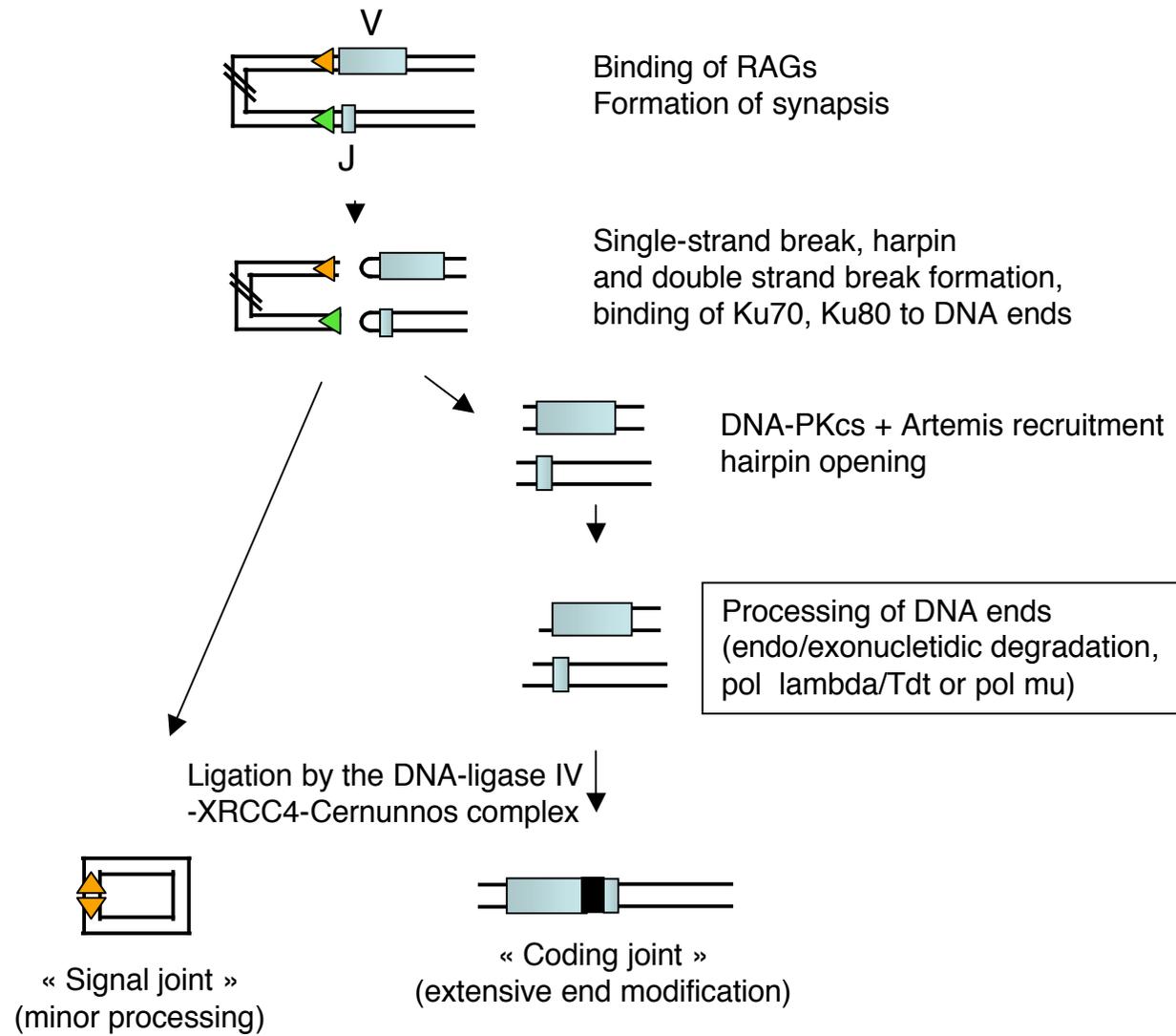


Figure 2

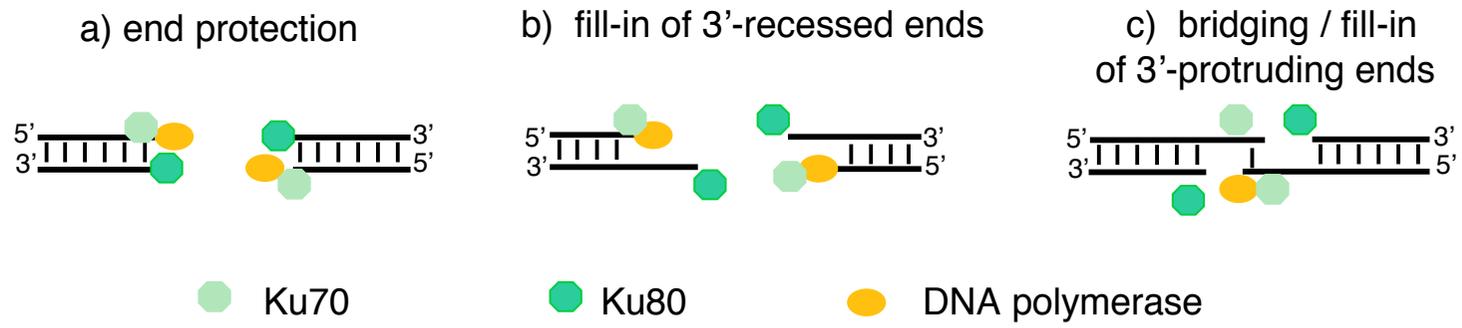


Figure 3

No C over G bias

		wild-type					Ung ^{-/-}					Polη ^{-/-}				
From	To	A	C	G	T	Total	A	C	G	T	Total	A	C	G	T	Total
		A	-	8	18	8	34	-	7	14	8	29	-	6	2	3
	C	3	-	5	17	25	2	-	0	27	29	6	-	11	24	41
	G	13	7	-	5	25	27	1	-	1	29	26	13	-	4	43
	T	5	7	4	-	16	4	7	2	-	13	1	1	3	-	5
		50 % G/C					58 % G/C					84 % G/C				

C over G bias (60 : 40)

		Ung ^{-/-} x Msh2 ^{-/-}					Msh2 ^{-/-}					Polη ^{-/-} x Msh2 ^{-/-}				
From	To	A	C	G	T	Total	A	C	G	T	Total	A	C	G	T	Total
		A	-	0	1	0	1	-	2	2	2	6	-	0	0	0
	C	0	-	0	60	61	6	-	4	44	54	4	-	4	54	62
	G	38	0	-	0	38	28	7	-	2	36	31	5	-	1	37
	T	0	0	0	-	0	2	1	1	-	4	1	0	0	-	1
		99 % G/C					90 % G/C					99 % G/C				

Figure 4

