Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Steps along the Pathway of V(D)J Recombination

Martin Gellert and J. Fraser McBlane

Phil. Trans. R. Soc. Lond. B 1995 **347**, 43-47
doi: 10.1098/rstb.1995.0007

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

This journal is © 1995 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

Steps along the pathway of V(D)J recombination

MARTIN GELLERT AND J. FRASER McBLANE

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Building 5, Room 241, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

SUMMARY

The mechanism of lymphoid-specific gene rearrangement (V(D)J recombination) is discussed, with a focus on the existence of broken DNA intermediates. Older evidence in support of this idea includes the sequence alterations at the recombined junctions and the presence of aberrant recombinants. More recently, broken DNA molecules have been directly detected in recombinationally active cells. The signal sequence ends have normal blunt-ended DNA breaks, but the coding ends have a hairpin (self-joined) structure that provides an explanation for the self-complementary P nucleotide insertions often found after V(D)J joining in the antigen receptor genes.

1. INTRODUCTION

During the development of lymphoid cells, a series of DNA rearrangements puts functional immunoglobulin and T-cell receptor genes together from the gene fragments existing in germline DNA. Because of the many possible choices of the V, D, and J segments that are joined, these DNA rearrangements are responsible for generating much of the diversity of the immune system (Tonegawa 1983; Blackwell & Alt 1988; Gellert 1992; Kallenbach *et al.* 1992). This article will compare the reaction, which is generally called V(D)J recombination, with other recombination systems for the purpose of gaining some insights about its mechanism.

V(D)J joining always occurs at the boundaries between a pair of coding segments and the recombination signal sequences that adjoin them. These signal sequences are made up of conserved heptamer and nonamer motifs (their consensus sequences are CACAGTG and ACAAAAACC), separated by a spacer of 12 or 23 (± 1) base pairs (b.p.) of non-conserved sequence. Joining is efficient only when the two signals have different spacer lengths (the 12/23 rule). The coding segments become fused in a structure called a coding joint, and the signals are joined by fusing heptamer to heptamer (a signal joint). Because the coding side of the recombination site is not specifically recognized (almost any sequence can be used), a coding joint looks much like random-sequence DNA, and could not be identified as resulting from V(D)J recombination unless one knew its history.

2. SIGNIFICANCE OF INEXACT JOINING

Some clues about the reaction come from the structures of the recombined junctions. Although signal joints are normally precise fusions of the two signal sequences, V(D)J coding joints are remarkably imprecise. Often a few bases from one or both ends have been lost, and several bases of non-germline DNA have been inserted. It is well known that these alterations are responsible for a large share of immune system diversity because the sequence around the junction forms part of the antigen binding site in both immunoglobulins and T-cell receptors. By contrast, most types of site-specific recombination (such as those catalysed by the integrase and resolvase enzymes) exchange DNA strands precisely, without loss or gain of nucleotides.

The imprecision of coding joints gives indications about the reaction mechanism. It means that there cannot be close coupling (by protein–DNA bonds or other means) of the beginning and end of the process. At some time in between, the ends of the coding sequence must be available for modification. Furthermore, as a result of the increasing number of sequenced junctions, it is now possible to distinguish two types of insertions, one non-templated and one whose sequence depends on the attached coding end. Non-templated sequences up to 15 nucleotides long (called N regions) are added by the enzyme terminal deoxynucleotidyl transferase (TdT), which is expressed mainly in the early lymphoid cells where V(D)J recombination occurs. In situations where TdT is not expressed, such as fibroblasts where V(D)J recombination is activated by expression of the *RAG1* and *RAG2* genes, coding joints are formed without N regions, but the recombination process is otherwise normal (Kallenbach *et al.* 1992). Similarly, in mice with a disruption of the *TdT* gene, V(D)J joining proceeds normally except that N regions are absent (Gilfillan *et al.* 1993; Komori *et al.* 1993).

A second class of templated insertions is more enlightening; it adds DNA with a sequence related to the attached coding end. It was noted that commonly one or two nucleotides complementary to the last base(s) of the coding end are added but occasionally longer, complementary inserts of 3 to 5 nucleotides are found (Lafaille *et al.* 1989; McCormack *et al.* 1989).

Phil. Trans. R. Soc. Lond. B (1995) **347**, 43–47
Printed in Great Britain

© U.S. Government

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions: Biological Sciences.* www.jstor.org

Downloaded from rstb.royalsocietypublishing.org

These 'P nucleotide' insertions (P for palindromic) have been useful in clarifying the reaction mechanism. Most or all of them are attached to coding ends that have not lost any nucleotides, implying that the addition occurred early in the recombination process, before any removal of bases. Furthermore, complementary insertions are most likely to result from the resolution of a DNA hairpin precursor; as is discussed below, these DNA structures have actually been observed.

3. DNA WITH DOUBLE-STRAND BREAKS

The variability of coding joints implies that DNA with double-strand breaks might well be an intermediate in the reaction. Support for this idea also comes from a class of aberrant products of V(D)J recombination, in which a signal sequence becomes attached to the coding segment that originally flanked its partner signal. It seems that the recombining pieces are at the same time available for joining in either a correct or an incorrect configuration. These 'hybrid joints' (Lewis *et al.* 1988; Morzycka-Wroblewska *et al.* 1988) can represent more than 20% of all the recombination events in some artificial substrates.

A search for broken DNA associated with V(D)J recombination was thus plausible, and recently such broken molecules have been detected in mouse lymphoid cells, at the TCRδ locus (Roth *et al.* 1992), at the immunoglobulin loci (Schlissel *et al.* 1993), and at the TCRβ locus (J. F. McBlane, unpublished results). An initially puzzling result was that all of the broken DNA at TCRδ in cells of normal mice corresponded to signal ends only, although each primary breakage should liberate both a signal and a coding end. The puzzle of the missing coding ends was solved when DNA from *scid* mice was looked at. Mice with the *scid* defect lack mature B or T cells, and only occasionally have correct V(D)J rearrangements in their antigen receptor genes (Schuler *et al.* 1986; Bosma & Carroll 1991). In *scid* lymphoid cell lines, a defect in V(D)J recombination can readily be shown, and it is apparent that the two types of junctions are very differently affected. Although signal joints are made at a normal rate, very few coding joints are completed (Lieber *et al.* 1988; Blackwell *et al.* 1989). Thus it must still be possible to initiate V(D)J recombination in *scid* cells, but a later step in the pathway to coding joints is apparently blocked.

In DNA from *scid* cells, it is found that both the broken coding end and signal end resulting from cleavage at the same site are detectable (Roth *et al.* 1992). One presumes this is because the coding ends in *scid* cells are prevented from joining and thus have a longer life. In fact, coding ends and signal ends are present in comparable amounts.

4. HAIRPIN DNA IN V(D)J RECOMBINATION

The striking result is that the coding ends have the hairpin DNA structure anticipated from the presence of P nucleotides; they are closed back on themselves, apparently with the 5′ end at the break joined to the 3′ end of the complementary strand (Roth *et al.* 1992) (see figure 1). Broken coding ends have been detected at three sites in the TCRδ locus, and in each case they are covalently sealed in this way. Hairpins are strange structures to be recombination intermediates, because they cannot take part in joining without first being rebroken, but the presence of P nucleotides provides compelling evidence for this pathway. Figure 1 shows in more detail how the off-centre nicking of a hairpin, followed by fill-in of the shorter strand, would naturally result in adding complementary bases to the coding end. Studies of the structure of hairpin DNA have shown that a hairpin cannot be double-stranded all the way to its tip, because of the constraint of turning the chain back on itself. In a fully self-complementary DNA, at least two and more probably four bases have to be unpaired, and may then be recognized and nicked, perhaps off-centre, by a single-strand-specific nuclease. This model was originally suggested to explain the self-complementary insertions (quite simi-

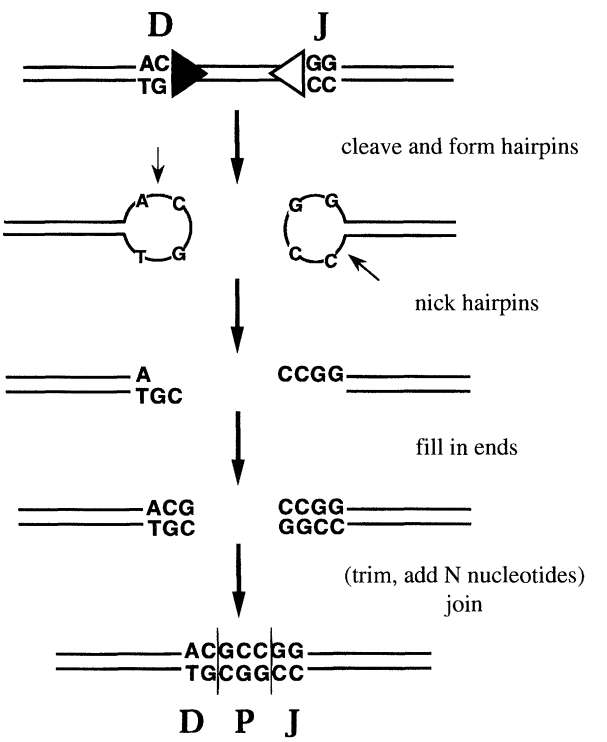

Figure 1. A scheme for the processing of hairpin intermediates to generate self-complementary insertions. The drawing shows the joining of two coding segments labelled D and J. In the top panel, the last two base pairs of D and the first two base pairs of J are shown. The DNA is cleaved at both signal/coding borders, leaving a hairpin structure with four unpaired bases on each coding end. A nick on one side of each hairpin, one base from the terminus on the left and two bases on the right (at the sites marked by the small arrows), leaves single-stranded tails whose last one or two nucleotides are complementary to the original end. If the normal convention is used that the top strand reads 5′ to 3′ from left to right, the ends as drawn can be filled in by a DNA polymerase. Later nucleotide removal by an exonuclease or addition by TdT could occur before joining, but is not shown here. To simplify the figure, the signal ends are not shown after breakage. The self-complementary (P nucleotide) insertion is indicated in the bottom panel.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

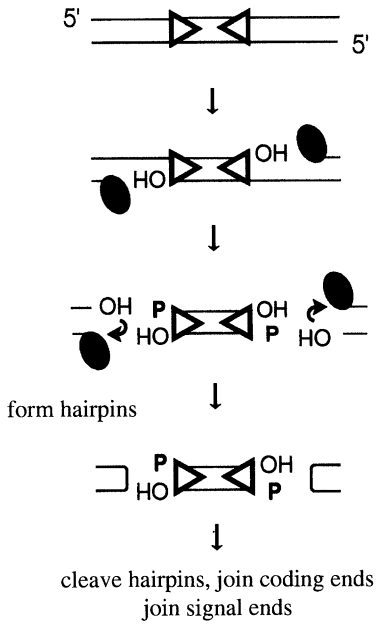

Figure 2. A possible scheme for V(D)J recombination involving a covalent protein–DNA intermediate. The triangles represent the signal sequences, and the dark ellipses indicate bound protein molecules. As drawn, the first nick at each signal/coding border attaches a protein to the 5′ end of the coding DNA and leaves a free 3′ end on the signal. A nick on the other strand then results in a blunt end on the signal and yields a 3′-hydroxyl end on the coding DNA that is capable of attacking the protein–DNA bond, displacing the protein and generating a hairpin end.

lar to P nucleotides) found after the excision of transposable elements in plants (Coen *et al.* 1986). The hairpin processing may not always leave a P nucleotide trace, because the nick can also be in the exact centre of the hairpin and leave no overhang. Alternatively, a P nucleotide tract initially formed could be trimmed off during later processing of the end.

How might hairpins be formed? A number of known biochemical reactions can couple cleavage of a DNA bond with rejoining to the same DNA or a new DNA partner. This is the well-known activity of topoisomerases, and a speculative model based on this class of enzymes is presented below. All topoisomerases share the same basic mode of operation. While breaking DNA, the protein attaches to one broken end, usually through a phosphoryl-tyrosine bond, and then this protein–DNA bond activates the DNA end for rejoining to either the same or a different partner DNA. Some of these enzymes and their close relatives (for instance the nicking-joining enzyme of vaccinia virus (Reddy & Bauer 1989) and the A* protein of phage ϕX174 (van der Ende *et al.* 1981)) are known to be capable of generating DNA hairpins. Hairpins can also result from the reaction of the topoisomerase-like phage λ Int or yeast Flp recombinases, when they act at damaged recombination sites (Nash & Robertson 1989; Chen *et al.* 1992).

In the hypothetical scheme (Roth *et al.* 1993) shown in figure 2, the protein that nicks at the signal-coding border ends up attached to the coding DNA. A subsequent nick (without protein attachment) on the other strand leaves a DNA end capable of attacking the protein–DNA bond, making a hairpin and liberating the bound protein. This scheme is preferred to an alternative involving a second protein–DNA bond that, by chemical symmetry, would leave a bound protein on one strand of the signal end. The preference derives from recent experiments probing the structure of broken signal ends in more detail. These experiments have shown that the signal ends have the simplest possible structure: almost all of them are cut at the border of the signal sequence and at the same position on both strands, leaving blunt ends. Furthermore, the ends are available for ligation to (non-phosphorylated) adapter fragments of DNA, and so must be free of any interfering protein fragments, the exposed 5′ ends carrying a phosphoryl group (Roth *et al.* 1993; Schlissel *et al.* 1993).

This is by no means the only possible mechanism. Another class of DNA strand–transfer enzymes, typified by the phage Mu transposase, can activate and transfer a DNA end without covalent linkage at any stage (Mizuuchi 1992). Such a process would avoid the need for an unsymmetrical reaction posed by the scheme shown in figure 2.

Until now we have discussed the formation of hairpin ends, but we also have to inquire how they are processed once made. One starting point is to ask why hairpins accumulate in *scid* cells. It is possible that the *scid* defect arrests the normal processing of hairpin DNA, blocking the pathway leading to coding joints at this stage. An observation consistent with this idea is that among the rare coding joints found in *scid* cells, some contain P nucleotide tracts much longer than those in normal cells, with up to 15 added nucleotides perfectly complementary to the adjoining coding sequence (Kienker *et al.* 1991; Schuler *et al.* 1991). In rare cases, this hairpin processing defect might be bypassed by a random nick further from the tip, leading to a junction with a long stretch of P nucleotides. However, the observation that hairpin-ended DNA transfected into *scid* cells is rejoined with normal efficiency (Lewis 1994) does not agree with this model, leaving only the possible reservation that transfected DNA may be subject to different processing from DNA resident in the nucleus.

The exact role of the *scid* factor in V(D)J recombination is still unknown. However, it has become clear that the *scid* factor operates more generally in DNA repair. It is widely expressed, as shown by the fact that non-lymphoid *scid* cell lines are hypersensitive to X-rays and defective in DNA double-strand break repair (Fulop & Phillips 1990; Biedermann *et al.* 1991; Hendrickson *et al.* 1991). The role of a general repair function in V(D)J joining is made more plausible by recent observations that mutations in several other DNA repair genes also induce a defect in V(D)J recombination (Pergola *et al.* 1993; Taccioli *et al.* 1993; Taccioli *et al.* 1994). One class of these mutations has now been shown to affect the Ku protein that is known to preferentially bind to DNA ends (Getts & Stamato 1994; Rathmell & Chu 1994).

Double-strand breaks can be initiating events in homologous recombination. In those cases, the broken ends invade the unbroken DNA partner leading,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

eventually, to the formation of a Holliday junction. In V(D)J recombination there is no partner capable of donating a homologous sequence bridging the final junction, so that the usage of the broken molecules must be quite different. Because they carry no recognition sites, the coding ends in particular may plausibly be joined by a process with many features common to the repair of double-strand breaks resulting from DNA damage. This could account for the linkage to multiple repair functions. In this context, the DNA hairpin intermediates remain enigmatic. It would be surprising if damaged DNA commonly passed through a hairpin stage on its way to repair, so it is unclear how the hairpins generated by the V(D)J process become engaged with the repair machinery. However, the possibility that hairpin structures are used in the repair of some forms of DNA damage has not been systematically tested, and may be worth pursuing.

We are grateful to the members of our laboratory for many valuable discussions.

REFERENCES

Biedermann, K.A., Sun, J.R., Giaccia, A.J., Tosto, L.M. & Brown, J.M. 1991 *scid* mutation in mice confers hypersensitivity to ionizing radiation and a deficiency in DNA double-strand break repair. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1394–1397.

Blackwell, K., Malynn, B., Pollock, R., *et al.* 1989 Isolation of scid pre-B cells that rearrange kappa light chain genes: formation of normal signal and abnormal coding joins. *EMBO J.* **8**, 735–742.

Blackwell, T.K. & Alt, F.W. 1988 Immunoglobulin genes. In *Molecular immunology* (ed. B. D. Hames & D. M. Glover), pp. 1–60. Oxford: IRL Press.

Bosma, M.J. & Carroll, A.M. 1991 The scid mouse mutant: definition, characterization, and potential uses. *A. Rev. Immunol.* **9**, 323–344.

Chen, J.-W., Lee, J. & Jayaram, M. 1992 DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell* **69**, 647–658.

Coen, E.S., Carpenter, R. & Martin, C. 1986 Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. *Cell* **47**, 285–296.

Fulop, G.M. & Phillips, R.A. 1990 The *scid* mutation in mice causes a general defect in DNA repair. *Nature, Lond.* **347**, 479–482.

Gellert, M. 1992 Molecular analysis of V(D)J recombination. *A. Rev. Genet.* **22**, 425–446.

Getts, R.C. & Stamato, T.D. 1994 Absence of a Ku-like DNA end binding activity in the *xrs* double-strand DNA repair-deficient mutant. *J. biol. Chem.* **269**, 15981–15984.

Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C. & Mathis, D. 1993 Mice lacking TdT: Mature animals with an immature lymphocyte repertoire. *Science, Wash.* **261**, 1175–1178.

Hendrickson, E.A., Qin, X.-Q., Bump, E.A., Schatz, D.G., Oettinger, M.A. & Weaver, D.T. 1991 A link between double-strand break-related repair and V(D)J recombination: the *scid* mutation. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4061–4065.

Kallenbach, S., Doyen, N., Fanton d'Andon, M. & Rougeon, F. 1992 Three lymphoid-specific factors account for all junctional diversity characteristic of somatic assembly of T-cell receptor and immunoglobulin genes. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2799–2803.

Kienker, L.J., Kuziel, W.A. & Tucker, P.W. 1991 T cell receptor γ and δ gene junctional sequences in SCID mice: excessive P nucleotide insertion. *J. exp. Med.* **174**, 769–773.

Komori, T., Okada, A., Stewart, V. & Alt, F.W. 1993 Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science, Wash.* **261**, 1171–1175.

Lafaille, J.J., DeCloux, A., Bonneville, M., Takagaki, Y. & Tonegawa, S. 1989 Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining. *Cell* **59**, 859–870.

Lewis, S.M. 1994 P nucleotide insertions and the resolution of hairpin DNA structures in mammalian cells. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1332–1336.

Lewis, S.M., Hesse, J.E., Mizuuchi, K. & Gellert, M. 1988 Novel strand exchanges in V(D)J recombination. *Cell* **55**, 1099–1107.

Lieber, M.R., Hesse, J.E., Lewis, S., *et al.* 1988 The defect in murine severe combined immune deficiency: Joining of signal sequences but not coding segments in V(D)J recombination. *Cell* **55**, 7–16.

McCormack, W.T., Tjoelker, L.W., Carlson, L.M., *et al.* 1989 Chicken IgL gene rearrangement involves deletion of a circular episome and addition of single nonrandom nucleotides to both coding segments. *Cell* **56**, 785–791.

Mizuuchi, K. 1992 Transpositional recombination: Mechanistic insights from studies of Mu and other elements. *A. Rev. Biochem.* **61**, 1011–1051.

Morzycka-Wroblewska, E., Lee, F.E.H. & Desiderio, S.V. 1988 Unusual immunoglobulin gene rearrangement leads to replacement of recombinational signal sequences. *Science, Wash.* **242**, 261–263.

Nash, H.A. & Robertson, C.A. 1989 Heteroduplex substrates for bacteriophage lambda site-specific recombination: cleavage and strand transfer products. *EMBO J.* **8**, 3523–3533.

Pergola, F., Zdzienicka, M.Z. & Lieber, M.R. 1993 V(D)J recombination in mammalian cell mutants defective in DNA double-strand break repair. *Molec. Cell. Biol.* **13**, 3464–3471.

Rathmell, W.K. & Chu, G. 1994 A DNA end-binding factor involved in double-strand break repair and V(D)J recombination. *Molec. Cell. Biol.* **14**, 4741–4748.

Reddy, M.K. & Bauer, W.R. 1989 Activation of the vaccinia virus nicking-joining enzyme by trypsinization. *J. biol. Chem.* **264**, 443–449.

Roth, D.B., Menetski, J.P., Nakajima, P.B., Bosma, M.J. & Gellert, M. 1992 V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell* **70**, 983–991.

Roth, D.B., Nakajima, P.B., Menetski, J.P., Bosma, M.J. & Gellert, M. 1992 V(D)J recombination in mouse thymocytes: double-strand breaks near T cell receptor δ rearrangement signals. *Cell* **69**, 41–53.

Roth, D.B., Zhu, C. & Gellert, M. 1993 Characterization of broken DNA molecules associated with V(D)J recombination. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10788–10792.

Schlissel, M., Constantinescu, A., Morrow, T., Baxter, M. & Peng, A. 1993 Double-strand signal sequence breaks in V(D)J recombination are blunt, 5′-phosphorylated, RAG-dependent, and cell-cycle-regulated. *Genes Dev.* **7**, 2520–2532.

Schuler, W., Ruetsch, N.R., Amsler, M. & Bosma, M.J. 1991 Coding joint formation of endogenous T cell receptor genes in lymphoid cells from scid mice: unusual P-nucleotide additions in VJ-coding joints. *Eur. J. Immun.* **21**, 589–596.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

Schuler, W., Weiler, I.J., Schuler, A., *et al.* 1986 Rearrangement of antigen receptor genes is defective in mice with severe combined immune deficiency. *Cell* **46**, 963–972.

Taccioli, G.E., Cheng, H.L., Varghese, A.J., Whitmore, G. & Alt, F.W. 1994 A DNA repair defect in Chinese hamster ovary cells affects V(D)J recombination similarly to the murine scid mutation. *J. biol. Chem.* **269**, 7439–7442.

Taccioli, G.E., Rathbun, G., Oltz, E., Stamato, T., Jeggo, P.A. & Alt, F.W. 1993 Impairment of V(D)J recombination in double-strand repair mutants. *Science, Wash.* **260**, 207–210.

Tonegawa, S. 1983 Somatic generation of antibody diversity. *Nature, Lond.* **302**, 575–581.

van der Ende, A., Langeveld, S., Teertstra, R., van Arkel, G. & Weisbeek, P.J. 1981 Enzymatic properties of the bacteriophage ϕX174 A* protein on superhelical ϕX174 DNA. *Nucl. Acids Res.* **9**, 2037–2053.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES