Annu. Rev. Biochem. 2002. 71:101–132
DOI: 10.1146/annurev.biochem.71.090501.150203
First published as a Review in Advance on February 12, 2002

V(D)J RECOMBINATION: RAG PROTEINS, REPAIR FACTORS, AND REGULATION*

Martin Gellert
Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 9000 Rockville Pike, Bethesda, Maryland 20892-0540; e-mail: gellert@helix.nih.gov

Key Words immunoglobulin, T-cell receptor, DNA breakage, hairpin, transposition

■ **Abstract** V(D)J recombination is the specialized DNA rearrangement used by cells of the immune system to assemble immunoglobulin and T-cell receptor genes from the preexisting gene segments. Because there is a large choice of segments to join, this process accounts for much of the diversity of the immune response. Recombination is initiated by the lymphoid-specific RAG1 and RAG2 proteins, which cooperate to make double-strand breaks at specific recognition sequences (recombination signal sequences, RSSs). The neighboring coding DNA is converted to a hairpin during breakage. Broken ends are then processed and joined with the help of several factors also involved in repair of radiation-damaged DNA, including the DNA-dependent protein kinase (DNA-PK) and the Ku, Artemis, DNA ligase IV, and Xrcc4 proteins, and possibly histone H2AX and the Mre11/Rad50/Nbs1 complex. There may be other factors not yet known. V(D)J recombination is strongly regulated by limiting access to RSS sites within chromatin, so that particular sites are available only in certain cell types and developmental stages. The roles of enhancers, histone acetylation, and chromatin remodeling factors in controlling accessibility are discussed. The RAG proteins are also capable of transposing RSS-ended fragments into new DNA sites. This transposition helps to explain the mechanism of RAG action and supports earlier proposals that V(D)J recombination evolved from an ancient mobile DNA element.

CONTENTS

INTRODUCTION

V(D)J recombination—recombination of variable (V), diversity (D), and joining (J) gene segments—is an essential step in the development of the vertebrate immune system. The completion of DNA rearrangements that generate functional immunoglobulin and T-cell receptor proteins is indispensable for the progression, and even the survival, of B- and T-cell precursors. V(D)J recombination is also the only known site-specific recombination in higher eukaryotes, and differs from other site-specific rearrangements in bacteria and yeast. It is initiated by breakage at precisely defined locations in DNA, but is then completed by a repair process related to the repair of breaks caused by ionizing radiation or other genotoxic agents. This repair often puts in local sequence changes, which are highly significant for the immune system. As a hybrid of two reaction types, V(D)J recombination thus has unique features, and close study has put them into sharper focus. The site-specific breakage by the RAG1/2 protein complex turns out to be related to the reactions of transposition enzymes, and indeed RAG1/2 can transpose a DNA fragment into a new backbone. The introduction of DNA hairpin ends by RAG cleavage is readily interpreted as a transposase-like reaction.

The pathway of rejoining the breaks is not so well understood, but an increasing number of general DNA repair factors are known to be involved. Studies on V(D)J recombination and on the more general repair of DNA double-strand breaks mutually reinforce each other, because it is easier to screen for radiation-resistance mutations, but V(D)J-induced breaks can be introduced in a controlled manner and are better defined chemically. The regulation of V(D)J recombination also yields new perspectives. The complexity of lymphoid cell

development requires the RAG1/2 recombinase to avoid most of the possible recombination sites in any given cell. The rules for gaining access to these sites may have parallels to transcriptional opening of a locus, and may similarly involve remodeling of the chromatin structure.

Other reviews of V(D)J recombination (e.g., 1–4) have treated the subject from different viewpoints and should be consulted for topics not fully covered here.

SOME ASSEMBLY REQUIRED: HOW ANTIGEN RECEPTOR GENES ARE CONSTRUCTED

The genes for the immunoglobulins and T-cell receptors of vertebrates have a unique structure. They begin as linear arrays of gene segments that require recombinational joining to form functional coding sequences. These variable (V), diversity (D), and joining (J) gene fragments typically are found as multiple copies that lie transcriptionally upstream of a constant (C) region. Mammals have seven antigen receptor loci: the immunoglobulin (Ig) H, κ, and λ loci, and the T-cell receptor (TCR) α, β, γ, and δ loci. Each locus has sets of V and J segments, and the IgH and TCRβ and δ loci also have D segments located between the Vs and Js. While the lymphoid cells are developing, segments of each type are joined by V(D)J recombination so as to make a V-J or V-D-J array that codes for a variable region exon, which becomes linked by RNA splicing to the C region. Because an Ig or TCR locus may contain tens or hundreds of segments of one type, the combinatorial possibilities are great. The variability of Ig and TCR chains is largely responsible for the ability of the immune system to respond to many different infectious agents.

Furthermore, the Ig and TCR proteins are heterodimers. An IgH chain combines with either a κ or a λ light chain, and the TCR proteins are $\alpha\beta$ or $\gamma\delta$ dimers. The diversity of the antigen receptors is increased even more by this pairing. The number of possible Ig or TCR molecules has been estimated to be well above 10^7.

Although this review discusses V(D)J recombination, other contributions to antigen receptor diversity are significant. After V(D)J recombination, the joined Ig genes in a B cell can undergo a locus-specific somatic mutation that alters the affinity of the immunoglobulin for its antigen (5). In some species, such as sheep, somatic mutation is the major source of immunoglobulin diversity.

In chicken Ig loci, the origin of most diversity is yet again different. It is caused by gene conversion using a large set of V pseudogenes as DNA donors to recombine by homology with a single expressed gene copy (6, 7). V(D)J recombination is still essential in chickens to generate this expressed copy.

GENERAL PROPERTIES OF V(D)J RECOMBINATION

Recombination Sites

V(D)J recombination takes place at recombination signal sequences (RSSs) adjacent to each V, D, and J segment. An RSS contains moderately well

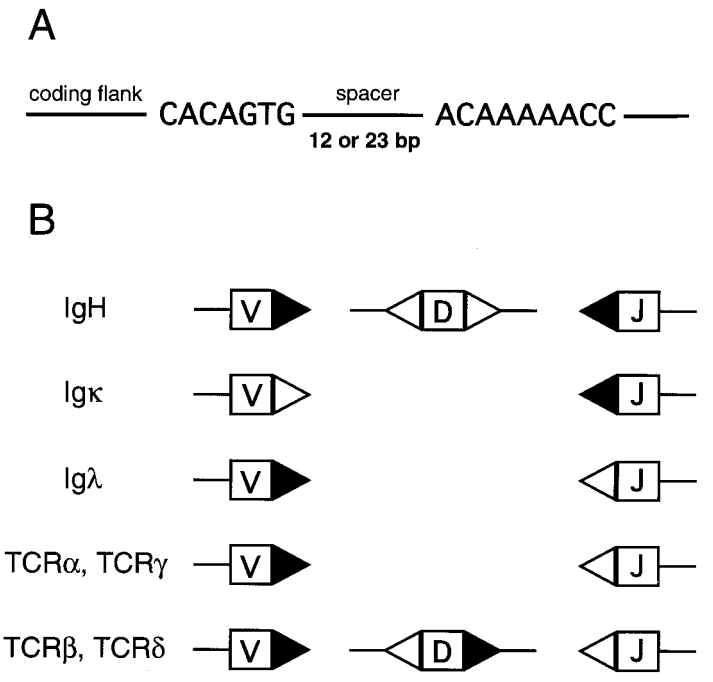

Figure 1 Recombination signal sequences and their arrangements at the antigen receptor loci. (*A*) The consensus heptamer and nonamer sequences of an RSS are shown with the alternative spacer lengths of 12 or 23 base pairs. (*B*) The various arrangements of RSSs at immunoglobulin and T-cell receptor loci. At each locus, all elements of one type (V, D, or J) have the same RSS arrangement. A 12-spacer RSS is indicated by an open triangle, a 23-spacer RSS by a black triangle.

conserved heptamer and nonamer sequences (Figure 1*A*) separated by 12 or 23 base pairs (bp) ($\pm$1 bp) of nonconserved DNA that is called a spacer. Breakage and rejoining of DNA occur between the heptamer and the neighboring coding segment. The length of the spacer is important in determining the functionality of the RSS; efficient recombination occurs only between RSSs with 12- and 23-bp spacers (8). The RSS spacer lengths at each antigen receptor locus are positioned so that recombination is directed to products that could be functional. As an example, the Igκ locus has all its V segments attached to 12-spacer RSSs and all J segments to 23-spacer RSSs, so that V-to-J joining is much more efficient than V-to-V or J-to-J joining, which would not lead to possible open reading frames (Figure 1*B*). At the IgH and TCRβ and δ loci, the rules are similar, with the elaboration that D segments must join to Vs on one side and Js on the other, so they are flanked by RSSs of appropriate spacer lengths on each side (see Figure 1*B*).

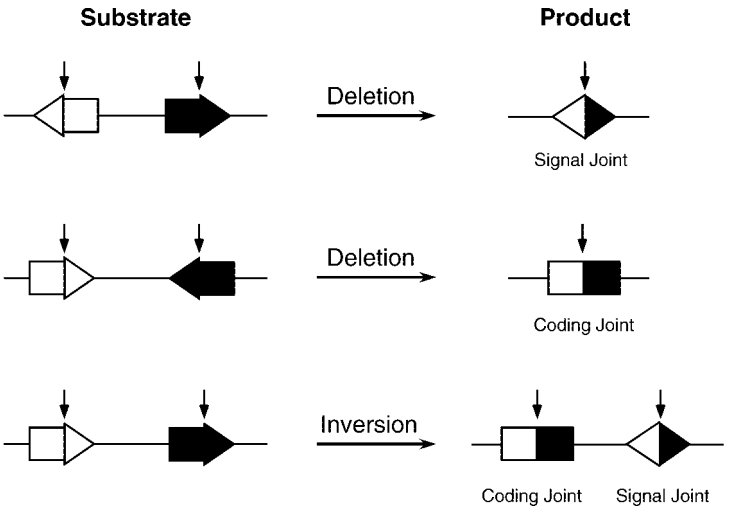

Figure 2 Recombination substrates and products. RSSs are denoted by triangles as in Figure 1, and their coding flanks are denoted by rectangles. Only the products that are retained in the substrate backbone after recombination are shown.

Recombination joins the RSSs as well as the coding regions. This is most easily shown with synthetic recombination substrates in which the V(D)J rearrangement retains either the joined coding segments or joined RSSs in the substrate backbone, or leads to inversion of the DNA between the RSSs (Figure 2). Either chromosomally integrated or extrachromosomal substrates have been used (9, 10).

When compared to many other types of site-specific recombination, V(D)J rearrangement differs in that the essential sequence is almost exclusively on one side of the junction. The heptamer/coding border is the site of recombination, and the coding sequence (or its replacement) can be varied almost at will. [A few flanking sequences, such as a run of Ts reading 5′ to 3′ into the heptamer, are particularly unfavorable to recombination (11–13).]

RSSs are highly conserved among vertebrates; the same recognition motifs are used in all species from sharks to humans. Some flexibility of RSS sequence is tolerated, as shown by variations at the antigen receptor loci and by deliberate mutation of substrates (10). The identity of the three heptamer nucleotides closest to the recombination site is most important, whereas mutations at the other heptamer positions still allow recombination. In the nonamer, changes in positions 5, 6, and 7 decrease rearrangement, but the nonamer is generally more variable than the heptamer. When the same nucleotide changes are put into the 12-spacer or 23-spacer RSS their effects are similar, implying a similar mode of recognition.

RSS variations in the antigen receptor loci may well influence the usage of gene segments. For example, the RSSs of mouse Igκ are closer to the consensus sequence than those of the alternative Igλ locus, possibly explaining the greater

usage of Igκ than Igλ light chains in mouse immunoglobulins. This preference can be modeled in synthetic substrates, which show that a κ RSS pair is used at least 100-fold more frequently than a typical λ RSS pair (14, 15).

The usual arrangement of RSSs at the antigen receptor loci is such that the joined coding segments remain in the chromosome and the junction of the RSSs (a signal joint) is excised on a circular DNA (16–18), which is later lost from the cells. However, some loci contain segments in inverted orientation so that both the coding joint and signal joint are retained in the chromosome. A notable example is the human Igκ locus, with roughly half its V segments lying in inverted orientation relative to J and C (19).

The components of Ig and TCR loci are arranged in many different ways among vertebrates. In sharks, many IgH genes occur in single V-D-J-C (or V-D-D-J-C) clusters scattered in the genome, and recombination usually occurs within one cluster (20). In mammals, a typical arrangement has multiple V segments transcriptionally upstream of several Js (perhaps with Ds between Vs and Js), all upstream of a single C region. This complex arrangement presumably evolved from the more primitive organization in sharks. Later simplifications are also evident; for example, the chicken Ig loci have been reduced to single functional V and J (or V, D, and J) elements, as mentioned above. The number of gene segments at a locus also differs greatly among vertebrates, perhaps as a result of germ line reshuffling made possible by the repetition of similar sequences.

Some loci have many segments and cover a long span of DNA. For example, the human Igκ and IgH loci, with 76 and 123 V segments respectively, extend over 2000–3000 kilobases. Distal V segments are still utilized reasonably often, so distance between RSSs does not seem to greatly affect the efficiency of V(D)J recombination. How RSS pairs locate each other over such large distances is an unsolved problem. However, recombination within a single DNA molecule is strongly preferred. For example, recombination between the mouse κ and λ loci, which are on different chromosomes, is ~1/1000 as frequent as V-J joining within one of these loci (21). Although interchromosomal translocations using the V(D)J machinery have been found in lymphoid tumors, they are rare events. When two antigen receptor loci are on the same chromosome, "illegitimate" V(D)J recombination between them is more frequent (22). In plasmid substrates, intermolecular recombination produces signal joints quite efficiently, but there are 10- to 1000-fold fewer coding joints (23, 24), for reasons that are not yet clear.

Structures of Signal and Coding Joints

The detailed structures of signal joints and coding joints provide clues to the recombination process, because the processing of coding and signal sequences before joining is quite different. Signal joints are simple, usually precise end-to-end fusions of two heptamer sequences. A small fraction of joints have nucleotides inserted between the heptamers, but loss of nucleotides is rare (9, 25). Coding joints are much more variable. They have frequently lost several

nucleotides from one or both ends, and may also have acquired nucleotides that were not present in the starting DNA (8).

These variations are highly significant for the antigen receptors. The junctional sequence is within the antigen binding site, so local alterations in coding joints multiply the diversity of Ig and TCR molecules even beyond that generated by combinatorial joining of the gene segments. But this variability also means that many junctions are wasted, because the length of DNA added or lost is essentially random, so that two-thirds of coding joints change the reading frame and cause premature termination of the protein chain. If the rearrangement is unsuccessful, a second attempt is possible on the other allele. Or, in loci with a V-J array, recombination can be tried again on the same allele, by the use of a V region upstream and a J region downstream of the erroneous junction (26).

Two types of nucleotide insertions are found in coding joints, nontemplated and templated. Nontemplated tracts up to 15 nucleotides in length (so-called N regions) are added by the enzyme called terminal deoxynucleotidyl transferase (TdT), as evidenced by their absence in mice with a disruption of the TdT gene (27, 28). Expression of TdT is normally limited to early lymphoid cells where V(D)J recombination is active, so these insertions are relatively specific to this type of recombination. TdT adds deoxynucleotides without a template to the ends of DNA chains, but with a preference for G residues that results in N regions being generally GC-rich.

A templated type of nucleotide addition is also found in coding joints (29, 30); these additions are more significant for the basic recombination mechanism. These P nucleotide insertions (P for palindromic) add a few nucleotides complementary to the last bases of the coding end next to the RSS. They are now known to result from off-center nicking of the hairpin DNA intermediates (Figure 3) that are formed at coding ends by the action of the RAG proteins (see below). Not all coding junctions have P nucleotide insertions; if the hairpin is nicked exactly at its center, there is no self-complementary overhang. Or such an overhang may be resected before the ends are joined. P nucleotides are usually found on full-length coding ends, but may also rarely exist at resected ends (31–33), presumably due to hairpins occasionally made within the coding flank rather than at the RSS-flank boundary.

Much less is known about the causes of nucleotide loss in coding junctions. Removal of a few nucleotides occurs in all cell types that perform V(D)J recombination, including nonlymphoid cells that ectopically express RAG1 and RAG2 (see below). One or more exonucleases may be responsible, but endonuclease action some distance inside the coding end has also been suggested (9, 34).

Unusual Types of V(D)J Junctions

V(D)J recombination is not restricted to making coding joints and signal joints. Hybrid joints, in which an RSS becomes joined to the coding flank of its partner RSS (Figure 4), have been found quite frequently in synthetic substrates (35), and less often also in the antigen receptor loci (36, 37). In some substrates, hybrid

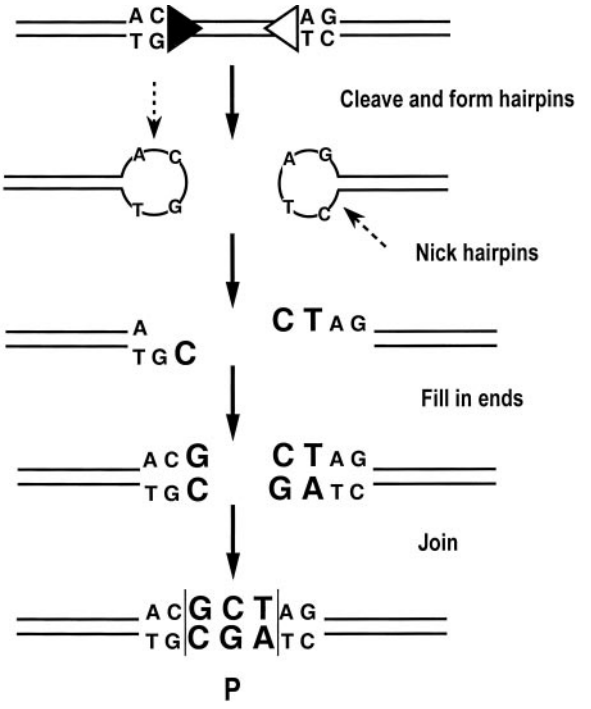

Figure 3 How self-complementary "P nucleotide" insertions arise in coding joints. During cleavage of DNA at the RSS-coding border, the ends of coding DNA are converted to hairpins. These hairpins can be nicked a few bases off-center (shown here as one base off-center on the left, two bases off-center on the right). This nicking leaves self-complementary single-strand extensions (large letters). After fill in and joining, these extensions (marked P) can be incorporated in the junction.

joints can account for 10% or more of the total recombinants (35). Sometimes the same pair of coding and signal ends is broken and rejoined to form an open-and-shut joint, as also shown in Figure 4 (35, 38). These events have been detected when they result in base loss and/or addition.

The formation of these junctions suggests that broken signal ends and coding ends must be present in a complex that possibly contains all four ends together. Physical evidence for such complexes is described below.

V(D)J RECOMBINATION INTERMEDIATES IN CELLS

The sequence alterations in coding joints imply that broken DNA ends are available for processing before joining, rather than being broken and rejoined in a coupled reaction such as those of site-specific recombinases in bacteria and

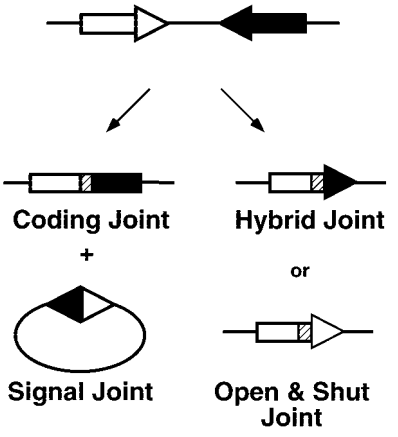

Figure 4 Nonstandard products of V(D)J recombination. Joining of one RSS to the coding flank of its partner generates a hybrid joint. Breakage and rejoining of an RSS to the same coding flank produces an open-and-shut joint, which can be recognized only if the junctional sequence has been changed. Local sequence changes in coding, hybrid, and open-and-shut joints are shown as hatched boxes.

yeast. Broken DNA molecules associated with V(D)J recombination can indeed be detected, which have been useful in dissecting the process. Broken ends have now been identified at most of the antigen receptor loci (18, 39–42) and also in plasmid substrates (43). The ends on the two sides of a break have quite different structures. Signal ends are cut exactly at the border between the RSS heptamer and coding sequence, and are blunt-ended with a 5′-phosphoryl and a 3′-hydroxyl group (39, 44). Coding ends are almost exclusively DNA hairpins, with the 5′ and 3′ termini of the coding flank covalently joined (45). Because hairpins contain the unchanged germline coding sequence on both strands (41), they are most likely the primary products of V(D)J cleavage, made in the same event that generates the broken signal ends. This surmise has been confirmed by the biochemical studies described below.

In some situations, a fraction of open (nonhairpin) coding ends has been found, thought to result from the nicking of hairpins. There is a preponderance of 3′ overhangs in these ends (46), and some of them have lost nucleotides, presumably reflecting a later stage of processing (40, 46).

Cells undergoing V(D)J recombination normally contain far more broken signal ends than coding ends, which suggests that the coding ends are joined more quickly. For example, mouse thymocytes have at least 1000-fold more signal ends than coding ends at the TCRδ locus (41). However, in mouse cells carrying the *scid* mutation, which blocks formation of coding joints (as discussed below), the level of coding ends becomes comparable to that of signal ends (45).

Most V(D)J-induced breaks occur as coupled cuts at both a 12- and a 23-spacer RSS. This is true both at antigen receptor loci and in plasmid substrates, where coupled cutting was found to be 30-fold higher than cutting at a single signal (43). This coordination may help to protect cells from the hazards of single cuts that could not be joined by V(D)J recombination, and would leave a double-strand break in the DNA.

The pathway of V(D)J recombination suggested by these results is that the initial coding ends are hairpins, which are then opened, processed, and quickly joined, possibly within a complex that holds all four cleaved ends together. The blunt-cut signal ends probably persist much longer before they are joined.

BIOCHEMISTRY OF THE RAG PROTEINS

In the picture of V(D)J recombination that has developed during the last several years, the process has two distinct stages. In the first stage, the RAG1 and RAG2 proteins cooperate to recognize the RSSs and to ensure their correct 12/23 pairing, and to break the DNA between each heptamer and the neighboring coding sequence. In the later stage, factors that are also used in other types of nonhomologous end joining act to process and link the ends into coding joints and signal joints. The biochemistry of the first stage is now fairly clear. The second stage still has uncertain aspects, although many of the required factors are known. The two parts of the process are discussed separately.

The RAG Genes and Proteins

The RAG1 and RAG2 proteins are the only lymphoid-specific factors needed for V(D)J recombination, and they carry out the enzymatic first step of the process. Coexpression of RAG1 and RAG2 leads to recombination of test substrates in nonlymphoid mammalian cells, where it would not normally occur (47, 48). Thus all other required factors must be generally available in any cell type. Conversely, mice with disruptions of either the RAG1 or RAG2 gene are completely defective in V(D)J recombination (49, 50), and therefore contain no mature B or T cells. These mice have no other defects, which implies that the RAG genes function only in the immune system.

The RAG locus has an unusual structure. In all species tested, the RAG1 and RAG2 genes are nearest neighbors, convergently transcribed, and in most genomes (such as *Xenopus*, chicken, mouse, and human) lack introns in either structural gene. Only the RAG1 genes of zebrafish (51) and rainbow trout (52) are known to contain introns. The arrangement of the locus led to the conjecture that both RAG genes might have arrived in the vertebrate lineage at the same time, by the insertion of a mobile genetic element (47, 53). This question is discussed in more detail below.

All jawed vertebrates contain a diversified immune system, and correspondingly have closely similar RAG1 and RAG2 genes. The level of conservation of both protein sequences is between 50 and 90% among sharks, fishes, amphibians, birds, and mammals. Below the evolutionary level of the sharks there is a discontinuity; in the lower eukaryotes there is no V(D)J recombination and no close homolog of either RAG gene.

Recombination requires the cooperation of RAG1 and RAG2, but large parts of both RAG genes can be deleted without losing recombination activity

(54–58). The mouse RAG1 protein, whose full length is 1040 amino acids, is still active for plasmid recombination after removal of the N-terminal 383 and C-terminal 32 residues (54). A large section of mouse RAG2 can also be ablated without destroying activity. Activity requires only the first 383 amino acids out of the full-length sequence of 527 (56, 57), even though the dispensable region is highly conserved. It has been noted that the truncated proteins can initiate recombination efficiently, but are not so good at completing it (59). A necessary postcleavage complex (see below) may be less stable with the truncated proteins. The dispensable parts of both proteins may also have regulatory functions, as mentioned at the end of this chapter.

DNA Cleavage by the RAG Proteins

Biochemical studies of RAG1 and RAG2 have been very useful both in emphasizing the central role of the proteins in V(D)J recombination, and in expanding the picture of their functional capacities. The enzymatic function of the RAG proteins is to cut the DNA between the RSS heptamer and the flanking sequence. This reaction and all others described below require both RAG1 and RAG2; no enzymatic action of either protein by itself has been described.

The purified RAG1 and RAG2 proteins by themselves are sufficient for cleavage (60). Biochemical work has almost exclusively used the truncated forms of the proteins described above because the full-length RAG1 and RAG2 proteins have generally been found to be insoluble and/or inactive. These shorter proteins (mouse RAG1 amino acids 384–1008 and RAG2 amino acids 1–383 or 1–387) have been purified from various systems: insect cells infected with baculovirus vectors, or HeLa cells infected with recombinant vaccinia virus (60), or mouse cells transfected with an expression vector (61). The protein sequence is usually linked to a fusion partner (maltose-binding protein or glutathione *S*-transferase) and/or a polyhistidine tail, for ease of purification.

In the presence of Mn^{2+}, RAG1/2 efficiently cuts an RSS in a DNA fragment to yield blunt 5′-phosphorylated signal ends and hairpin coding ends that retain the full coding sequence (60). These are the same cleavage products found in vivo. Cleavage by RAG1/2 happens in two steps. A nick is made at the 5′ end of the signal heptamer, leaving a 5′-phosphoryl group on the RSS and a 3′-hydroxyl on the coding end (this is shown for coupled cleavage at a pair of RSSs in Figure 5). The second step joins this 3′-hydroxyl to the phosphoryl group at the same nucleotide position on the opposite strand, resulting in the DNA hairpin coding end and blunt signal end. Both steps require the RAG1 and RAG2 proteins and the specific sequence of the RSS.

Cleavage by the RAG proteins alone is more efficient at a 12-RSS than a 23-RSS. However, cleavage is increased, particularly at a 23-RSS, by adding one of the chromosomal high-mobility-group proteins, HMG1 or HMG2 (61, 62). HMG1 and 2 are nonspecific DNA-binding and -bending proteins, so it is possible that they deform a 23-RSS to allow better RAG binding. These HMG proteins also affect other RAG reactions described below, and it is possible they

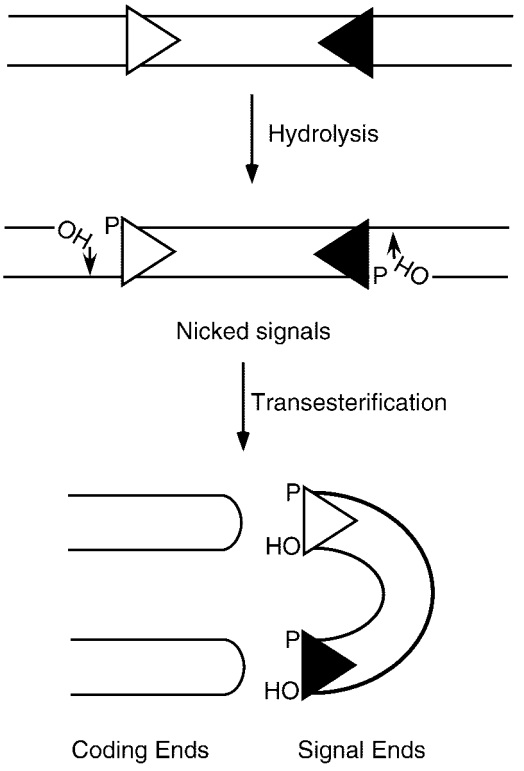

Figure 5 DNA cleavage by the RAG proteins. In the first step, a nick is made at the 5′ end of the RSS heptamer, leaving a 3′-OH on the coding flank. In the second step, this hydroxyl group attacks the opposite strand to produce a hairpin coding end and a blunt signal end. In this figure, the reaction is shown as a coupled process at a pair of RSSs, as it would be in the presence of Mg^{2+} (see text).

are significant cofactors of V(D)J recombination in vivo. However, their effects on the cell may be too widespread to allow a gene disruption test.

Coupled Cleavage

The 12/23 RSS coupling required for V(D)J recombination can also be displayed in DNA cutting by RAG1/2. Although these proteins cut a single RSS in Mn^{2+}, a pair of RSSs is necessary if the divalent ion is changed to Mg^{2+}. A 12/23 pair is most effective, and both RSSs are cut, yielding blunt signal ends and hairpin coding ends at both sites (Figure 5). Coupled cleavage is produced either with crude extracts (63) or the purified RAG proteins (64). The preference of the RAG proteins for a 12/23 pair (over a 12/12 pair, for example) is raised by HMG1 or 2 (62), and further increased by adding nonspecific DNA (65), to a level that

approximates the overall ~50-fold preference for 12/23 pairs in V(D)J recombination. Thus the RSS coupling of V(D)J recombination is determined by RAG protein cleavage.

Even in Mg^{2+}, the RAG proteins nick DNA at a single RSS (64, 67). The requirement for a 12/23 pair comes in only at the hairpinning step, reflecting the use of a synaptic complex in this second step. It is not known whether nicking at RSSs is similarly unregulated in vivo.

RSS Recognition

The RSS positions most conserved in V(D)J recombination are also important for cleavage of DNA by RAG1/2 (68, 69). Cleavage is greatly decreased by mutations in the three bases of the heptamer nearest the cutting site. But only the hairpinning step is so sensitive; the prior nicking step is not so restricted. If the heptamer is totally ablated, no hairpins are made, but some imprecise nicking occurs near the position where the heptamer border would be in a 12-spacer RSS (i.e., 19 bp from the nonamer). In general, complete cleavage has essentially the same requirements as recombination, but nicking is more tolerant.

The length of the spacer is also significant. The normal 12- and 23-bp spacer lengths differ by almost exactly one turn of DNA, so that proteins bound to the heptamer and nonamer would be in the same rotational phase on a 12-RSS or a 23-RSS. Experiments in which the spacer length is altered agree with this view. If the spacer length is changed by half a turn, to 18 or 29 bp, cleavage is inhibited, but it is partly restored at the next integral number of turns, 33 or 34 bp (68, 69).

This pattern implies that the heptamer and nonamer act somewhat separately. An isolated heptamer directs some cleavage to the normal site, and even a nonamer by itself induces the RAG proteins to nick where the heptamer border would be. When the two motifs are in the right helical phase, they increase cleavage synergistically, and in the wrong spacing they conflict.

Modifications of the DNA structure are also helpful in clarifying RAG1/2 action. If the RSS in an oligonucleotide is made single stranded, and only the coding flank remains double stranded, the RAG proteins can still efficiently form a DNA hairpin (68, 69); the only unusual feature is that the coding end can now attack either end of the single-stranded heptamer, to generate two alternative hairpins. The nonamer has no effect in this substrate; only the heptamer is recognized.This reaction is so efficient as to imply that the heptamer may become partly unpaired in normal cleavage. In fact, it is known that the CACA/GTGT sequence that is part of the heptamer has a very distorted structure, both in solution (70, 71) and in crystals (72). It is also possible that the RAG proteins contribute to unwinding the RSS. Thus cleavage at the heptamer may involve both specific sequence recognition and DNA unpairing. Unpairing of the first few coding nucleotides of coding sequence has also been shown to enhance cleavage in some substrates, where the coding sequence is otherwise unfavorable (68, 69, 73).

DNA Transposition by RAG1/2

Up to this point, we have described RAG1/2 as a DNA endonuclease, with the single peculiarity of creating a hairpin on one cut end. However, closer study of the biochemical reactions has revealed that the RAG proteins act very similarly to DNA transposases. In fact, under suitable conditions RAG1/2 can perform transpositional attack in vitro, inserting RSS ends into a second DNA.

The link to transposases was first suggested by stereochemical studies. Hairpin formation is a conservative reaction; the new bond is made at the expense of the bottom-strand bond that is broken. This provides the opportunity to distinguish between two types of conservative DNA strand transfer. One type is topoisomerase-like, using a covalent protein-DNA intermediate, as is found in site-specific recombinases such as Cre, Hin, Gin, lambda Int, and Flp. The other type is a direct transesterification without covalent intermediates, as found with bacteriophage Mu transposase and with HIV integrase, which is also a transposase (74, 75). A test of the chemical chirality of hairpin formation by RAG1/2 showed this to be a direct transesterification, indicating a similarity of the RAG system to the transposase/HIV-integrase family (66).

Biochemical reactions that make hybrid joints or open-and-shut joints (see above) also indicate a link between RAG1/2 and transposases. It was known that some transposases such as HIV integrase can reverse a transpositional strand transfer, cutting the attacking DNA end from a target DNA and in the same reaction resealing the target (76). Because hybrid joints were still produced in cell lines deficient in DNA repair that were unable to make coding joints or signal joints, Bogue et al. (77) suggested that a similar reversal of RAG cleavage could be responsible. Purified RAG1/2 was indeed shown to be capable of making either hybrid or open-and-shut joints (78), and thus able to rejoin an RSS to another DNA end.

This activity of the RAG proteins suggested that they might also be able to use RSSs to attack an exogenous DNA, in a transpositional attack. Quite surprisingly, RAG1/2 is able to catalyze this reaction (79–81), which is entirely distinct from V(D)J recombination. Transposition requires a 12/23 pair of RSSs, but does not always insert both ends in a coupled reaction; either a double-ended insertion (Figure 6) or a single-ended attack (Figure 7) is possible. Transposition can be coupled to RSS cleavage, or can use precut RSS ends. In either case, the RSS attaches to target DNA exactly at the heptamer end. No particular specificity of target DNA sites has been found, but GC-rich regions are somewhat preferred (80). Coupled attack by two RSSs leads to insertion into opposite DNA strands, at positions staggered by 5 bp (80) or 3–5 bp (79). Transposases typically attack the two strands with a defined offset, and RAG1/2 evidently follows this rule.

Transpositional strand transfer by the RAG proteins is quite efficient, but no RAG-driven transposition in cells has yet been observed. This is an unusual case of a well-defined biochemical activity with no obvious cellular correlate. It seems that the action of the RAG proteins has been diverted into the V(D)J recombi-

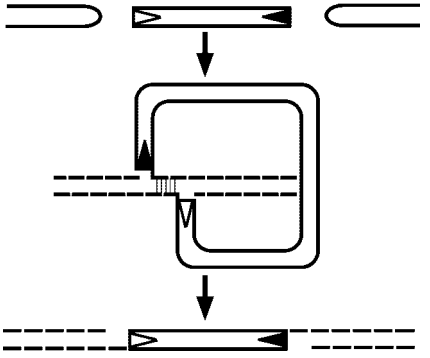

Figure 6 Two-ended transposition by RAG1/2. In the top line, cleavage has liberated the two RSS ends that will be used to attack another DNA (dashed line). The reaction requires a 12/23 RSS pair, but may perform the coupled reaction shown here, or insert one end as shown in Figure 7.

nation pathway, which implies that the initial cleavage products are handed off to the cellular repair machinery rather than being used for further RAG-mediated strand transfer. Transposition in cells would be in competition with the use of RSSs to form signal joints, but it may be inhibited in other ways as well. Several ways in which the activity of transposons is down-regulated have been described (82). As discussed below, ongoing transposition in lymphoid cells could be harmful, possibly leading to tumors.

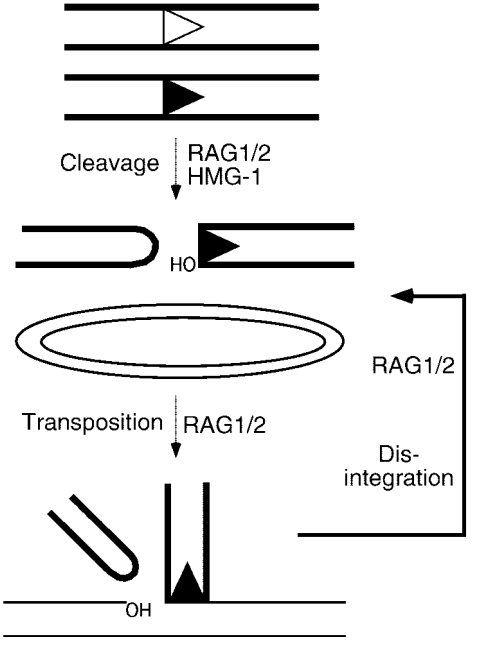

Figure 7 One-ended RAG1/2 transposition, and its reversal by disintegration. A cleaved signal end can attack a target DNA (the double ellipse), and this reaction can also be reversed by the RAG proteins.

The linkage of RAG1/2 to transposition helps to explain why hairpins are made during cleavage. Hairpin DNA has been detected or inferred as an intermediate in several transposition reactions. With the bacterial Tn10 and Tn5 transposases, hairpins have been demonstrated biochemically (83, 84). These transposases make a DNA hairpin on the transposon end and then again cut it open prior to strand transfer. The distinction between hairpins on the mobile element end for Tn10 and Tn5, and hairpins on the flanking sequence for RAG1/2, is correlated with the first nick being made on the "bottom" strand for Tn10 and Tn5 and the "top" strand for RAG1/2. This leaves the exposed 3′-OH in position to put the hairpin on the transposon or the coding flank, respectively.

Several other transposons are likely to generate hairpins on the flanking DNA, in a way more closely similar to RAG1/2. These include the *Ascobolus* transposon Ascot-1 (85), maize Ac/Ds and *Antirrhinum* Tam3 (86), and *Drosophila* Hobo (87). In all these cases, self-complementary insertions, like the P nucleotides of V(D)J recombination, are found in a large fraction of rejoined chromosomal sites after transposon excision. A hairpin precursor is the most plausible explanation, although it has not yet been isolated in these cases.

Evolutionary Implications of RAG-Mediated Transposition

Transposition by the RAG proteins has made it easier to understand the evolutionary origin of V(D)J recombination, as well as its mechanism. It was suggested some time ago that V(D)J recombination might derive from a mobile genetic element, because the usual arrangement of RSSs in the Ig and TCR loci resembles the inverted repeats at transposon ends (88). When the RAG genes were later identified, it was noted that the compactness of the RAG locus, with the two intronless structural genes as nearest neighbors, is reminiscent of transposons (47). The distribution of the RAG genes among animal species also fits with a transpositional origin. Functional RAG genes are found in all jawed vertebrates, starting at the level of the sharks. Lower eukaryotes do not have RAG genes, nor any closely related sequences, which suggests that the RAG genes might have been introduced into vertebrates by horizontal transfer from an unidentified donor organism (47, 53).

These two lines of evidence concerning the RSSs and the RAG genes can now be unified. RAG-mediated transposition makes it plausible that the RAG genes were responsible for their own transfer into vertebrates, using the RSSs as recognition sites (79, 80). One can envisage a primordial transposon that had two RSS ends surrounding the RAG1/2 genes. After arrival in the vertebrate lineage, the RSSs and the RAG genes may have become separated, rendering further transposition between cells impossible. An arrangement of gene segments separated by RSSs could have arisen if a transposition event landed in a precursor of an Ig or TCR gene and split it. These evolutionary considerations are of course purely speculative, and it will not be possible even to do model experiments until a system for RAG-mediated transposition in cells is developed.

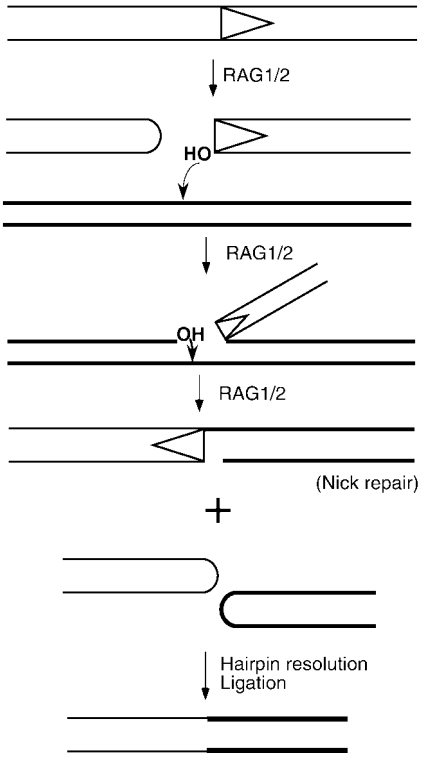

Figure 8 A possible mode of chromosomal translocation by RAG-promoted transposition. A cleaved signal end at an Ig or TCR locus can insert into another chromosome (heavy line) by transpositional strand transfer. In the resulting branched DNA structure, the 3′-OH of the target DNA can be further processed to generate a hairpin end and an interchromosomal junction containing the RSS. As this reaction is likely to occur within a complex that also contains the hairpin coding end from the original cleavage, joining of the two hairpin ends would then generate the reciprocal chromosomal translocation.

Processing of Transposition Intermediates: A Path to Chromosomal Translocations?

The products of strand transfer by RAG1/2 can be resolved in several different ways. Coupled attack by a pair of RSSs inserts donor DNA with RSS ends into the target DNA (79, 80). Attack by a single RSS leads to a branched structure, which can be further processed by the RAG proteins in at least two ways. As was shown for HIV integrase (76), the RAG proteins can perform disintegration by reversing the strand transfer to cut off the RSS end and reseal the target DNA at the same time (Figure 7) (81). In this reaction, the exposed 3′-OH in the target DNA attacks the phosphodiester bond linking the RSS to target DNA.

An alternative mode of resolution, made possible by the ability of the RAG proteins to generate hairpins, could lead to new and unusual products. It has been shown that after strand transfer, the 3′-OH end in the target DNA can attack its opposite strand to form a hairpin just as in the normal course of RAG cleavage (Figure 8) (81). If this reaction happened in cells, it could lead to chromosomal translocations, because this hairpin could link to the one produced by the initial cleavage at the Ig or TCR locus to make an interchromosomal junction (80, 81). The reciprocal translocation would be composed of the signal-ended chromo-

some fragment linked to the complementary part of the target chromosome (Figure 8). These translocations would link an RSS-ended chromosome fragment to a non-RSS site on the partner chromosome. Such translocations have been identified in a few cases (discussed in 81).

Because translocations initiated by V(D)J cleavage are found in a number of lymphoid tumors, such a pathway would have considerable medical significance.

Hairpin Opening and Other Nuclease Activities of RAG1/2

The hairpin ends produced by RAG cleavage must be reopened before the nonhomologous end-joining pathway can process and join them. In order to explain the self-complementary (P nucleotide) additions often present in coding junctions, the nick must sometimes be a few nucleotides away from the center of the hairpin (Figure 3). It has been reported that RAG1/2 can produce such breaks, either on a hairpin substrate or on hairpins made in the same reaction by RAG cleavage (34, 89). The positions of nicks are different in the two cases. In preformed hairpins, some nicks are near the hairpin tip (usually on its 5′ side), but some are further back along the double-helical stem, as much as 10 bp away. P nucleotide tracts are normally much shorter than this (typically one to four nucleotides), so the significance of the more distant cuts is unclear. If hairpin opening is coupled to RSS cleavage, most coding ends are blunt, with a few cut one or two nucleotides off-center.

The biological relevance of this reaction is not yet known. In vivo, hairpin opening depends on the presence of active DNA-dependent protein kinase (DNA-PK), as described below. It is unclear how this requirement would fit with intrinsic hairpin-nicking by the RAG proteins.

Other nuclease activities of RAG1/2 have been described. The RAG proteins can endonucleolytically cut off a 5′-ended overhang on duplex DNA (34, 90), and can also remove a 3′-terminated single-stranded flap (91). It was suggested that this latter activity could be used in the processing of recombination intermediates after two resected coding ends are brought together.

The existence of these many activities, some involving specific DNA sites and others not, implies that the active site (or sites) of RAG1/2 must have considerable flexibility. Structural information on the RAG proteins would be helpful in explaining the relations among these activities.

DNA Binding by the RAG Proteins

Several types of complex between the RAG proteins and DNA have been described. RAG1 alone binds DNA with a moderate preference for RSS sequences, estimated as three- to ninefold by various groups (92). This binding is primarily to the nonamer. RAG2 also binds DNA, but has no specificity for an RSS. RAG1 and RAG2 together bind an RSS much more tightly and specifically (93). In competition experiments, RAG1/2 prefers a 12-RSS by 50-fold over nonspecific DNA, and a preformed RSS-RAG1/2 complex is stable against even

higher levels of competitor DNA. By DNA footprinting, RAG1/2 contacts both the heptamer and nonamer (94–96); by photo-cross-linking, both RAG1 and RAG2 touch the heptamer (96–98). Footprinting of RAG1 by itself reveals contacts only in the nonamer (94, 96, 98). It should be noted that all these studies have been done with the truncated RAG1 and RAG2 proteins described above.

Moving to the next higher level of organization, the synaptic complex of 12- and 23-spacer RSSs also requires both the RAG1 and RAG2 proteins, and only forms efficiently in the presence of HMG1 (99). This complex is highly resistant to nonspecific DNA, and is stable enough that it can be formed with the 12-spacer RSS and 23-spacer RSS on separate DNA fragments. In assembling the synaptic complex, a 12/23 RSS pair is greatly preferred over 12/12 or 23/23 pairs. The synaptic complex accounts for most of the 12/23 specificity of RAG cleavage and V(D)J recombination, although a contribution at the cleavage step itself has also been suggested (67, 100).

An experimental convenience is that complexes of RAG1/2 with either a single RSS or a 12/23 pair can be formed in Ca^{2+}, which does not support either nicking or hairpinning. Cleavage can then be initiated by the addition of Mn^{2+} or Mg^{2+}.

RAG1/2 remains bound to DNA after cleavage. RAG1/2 complexes either with signal ends (101) or with both the signal ends and coding ends (99) have been described. This four-ended postcleavage complex has a role in the later joining process. In a cell-free system that carries out complete V(D)J recombination (although at a low level), the RAG proteins must stay bound after cleavage to allow coding joints to be made (102). It is also likely that the postcleavage complex acts in the formation of hybrid and open-and-shut joints, which require attack by a signal end on a hairpin coding end (78, 90). In transposition, a RAG1/2 complex with the cleaved signal ends alone [analogous to the cleaved donor complex of Mu transposition (103)] can bind a target DNA prior to chemical attack.

Sequence Motifs and Mutational Studies of the RAG Proteins

Sequence comparison to proteins of known structure has been useful in identifying possible folding patterns, particularly for RAG2. The RAG2 core is proposed to have six repeats of a kelch motif, of 50 residues each (104, 105). (The kelch motif is named after a *Drosophila* regulatory protein.) Each repeat would contain a four-stranded twisted antiparallel beta sheet, and the whole array would be displayed circularly, like a six-bladed propeller. Such structures have been found to take part in protein-protein binding, so in this case the RAG2 propeller structure may be involved in binding RAG1, and perhaps DNA as well (104).

RAG2 also has a highly acidic portion [amino acids (aa) 352–410]. The C-terminal half of this region, from aa 383 on, has been deleted, together with the rest of the protein's C terminus, in the recombinationally active "core" protein.

The role of this region is thus not known. A Cys-His rich PHD motif from aa 420 to 480, similar to homeodomain zinc fingers, has also been noted (104). Such domains in other proteins bind to chromatin components, an interesting possibility in light of the proposed role of the RAG2 C-terminal region in controlling recombinase access to RSSs within chromatin (see below).

The N-terminal part of RAG1 contains a possible zinc finger of the RING family (residues 290–328) (106) and two additional zinc finger motifs (107). These segments could be involved in dimerization or protein-DNA interactions. The N-terminal domain containing these motifs is not required for recombination of artificial substrates, but may be important for controlling site access in a chromatin context (see below).

All known RAG activities require both RAG1 and RAG2, so mutational studies have been especially useful in identifying the functional role of each protein. It has already been noted that the first 383 residues of RAG1 and the last 144 residues of RAG2 are dispensable for recombination. Residues involved in the catalytic activity have now been identified by point mutations. One approach has been to search for acidic residues involved in catalysis, because most transposases use an Asp/Glu triad, often called a DDE motif, for binding of divalent metal ion in the active site (108). Mutations in RAG1 of either D600, D708, or E962 abolish recombination in vivo and block cleavage by the purified protein, while permitting RSS binding (109–111). D600 and D708 are most probably involved in metal binding, because mutation at both of these sites abolishes iron-induced hydroxyl cleavage of RAG1 (109), and substitutions of Asp by Cys (in the presence of Mn^{2+}) restore some catalytic activity (109, 110), as has been shown for other transposases (112).

The function of E962 is less clear. Mutation of this site does not affect iron-mediated cleavage, and RAG1 E962C is inactive. This residue is also much more distant in the sequence than is typical for such acidic triads, and may be in a separate domain from D600 and E708 (W. Yang, unpublished information). It is not clear that E962 is directly involved in metal ion binding.

The metal-binding function of the active site is thus supplied by RAG1. Systematic mutagenesis of RAG2 did not reveal any acidic residues required for activity (110). There are other types of mutation in RAG2 that do block catalytic activity (114), implying that the full active site probably involves parts of both RAG1 and RAG2.

Some mutations (RAG1 E423Q and E547Q and several RAG2 mutations) permit RSS cleavage (and transposition) but fail to complete joining in vivo, implying once again that the RAG proteins have a role in the postcleavage phase of the process (90, 114). In addition, some of these proteins (RAG1 E423Q, RAG2 R73A and K118A/K119A) do not open hairpins efficiently in vitro, indicating that this step may normally be performed by the RAG proteins during V(D)J recombination. However, such a block at the hairpin-opening step has not yet been shown in vivo. A complication is that most mutations of this type also

block joining of signal ends, and it is not clear why a defect in hairpin opening should have this consequence.

Mutations of RAG1 or RAG2 are responsible for some cases of human severe combined immune deficiency (SCID), in which both B cells and T cells are depleted (115–117). Mutations that would abolish RAG activity, such as frameshifts, lead to complete immunodeficiency, whereas less complete defects cause a partial SCID phenotype or the related condition known as Omenn's syndrome (116, 117). In a set of RAG2 *scid* mutations, six out of seven sites were found to lie on one face of the proposed six-bladed propeller structure mentioned above, in such a way that they might interact with RAG1 (118).

Certain RAG mutations inhibit binding to DNA, or association of the RAG proteins with each other (110, 119). Some DNA-binding mutations of RAG1 are in a region (aa 389 to 436) similar to part of the bacterial Hin recombinase (120, 121).

LATER STAGES OF V(D)J RECOMBINATION

The later steps of V(D)J recombination have many aspects in common with general DNA double-strand break repair, and the two processes share a number of factors. Cellular mutants sensitive to ionizing radiation are often impaired in V(D)J joining, and vice versa (for reviews, see 122–124). The pathway of rejoining and repair has many unclear aspects, and the list of essential factors may not even be close to complete. This section focuses on factors that are known, or suggested, to be involved. We discuss the DNA-dependent protein kinase (DNA-PK), Ku protein, DNA ligase IV, Xrcc4, Artemis, histone H2AX, and the Mre11/Rad50/Nbs1 complex.

DNA-PK is a multiprotein complex made up of the catalytic subunit (DNA-PK_{CS}) and the Ku protein heterodimer (Ku70 and Ku80). DNA-PK_{CS} is a very large protein (~465,000 kilodaltons) with some homology to the mammalian ATM and ATR kinases and the yeast TEL1 and MEC1 kinases, among others (125). All these proteins belong to the family of serine/threonine protein kinases that are related by sequence to lipid kinases, and all have functions in sensing DNA damage. The kinase activity of DNA-PK_{CS} depends on the presence of both DNA and Ku protein, although a low level of DNA stimulation in the absence of Ku has been reported (126, 127). The role of DNA-PK_{CS} first became apparent from the properties of the mouse *scid* mutation. This mutation (128) causes immunodeficiency by blocking V(D)J recombination, and also leads to X-ray sensitivity and a defect in repair of DNA double-strand breaks. The mutation in *scid* cells is in the kinase domain of DNA-PK_{CS} (129–131). Gene disruptions of DNA-PK_{CS} lead to defects quite similar to those caused by the *scid* mutation (132–134).

V(D)J recombination in *scid* or DNA-$PK^{-/-}$ rodent cells is much more defective for coding joints than signal joints, but a DNA-PK_{CS} mutation in

Arabian horses greatly decreases the formation of signal joints as well as coding joints (135). This difference has not yet been explained.

Hairpin intermediates accumulate in *scid* cells (45), and some of the rare coding joints have large deletions, up to several kilobases, or unusually long P-nucleotide tracts (128). All these abnormalities could be explained if the major defect was in hairpin opening, with occasional bypass by nicks or breaks further away. However, it is not clear why a DNA-PK_{CS} defect should cause a block in hairpin cutting if this step is carried out by the RAG proteins. And DNA-PK_{CS} is also involved in the repair of X-ray damage, which presumably does not involve DNA hairpins. Thus it may also affect a step in V(D)J joining beyond the hairpin-opening stage.

The biologically significant targets of DNA-PK_{CS} are not yet known, in part because the minimal consensus site (SQ or TQ) is so frequent. Purified DNA-PK phosphorylates a wide range of proteins (124, 136).

Defects in the DNA-binding Ku protein also interfere with V(D)J recombination (137–139). Ku is a heterodimer of Ku80 and Ku70 subunits that binds to interruptions in DNA such as broken ends, single-strand gaps, or DNA hairpins. Although Ku is best known as the DNA-binding component of DNA-PK in vertebrates, it probably acts in other ways as well. Ku is also present in lower eukaryotes, such as yeast, worms, and flies, that do not contain DNA-PK_{CS}. In *Saccharomyces cerevisiae*, Ku mutations affect double-strand break repair and the stability of telomeres (140), but no interaction of Ku with any yeast protein kinase has been described. Biochemical experiments suggest a possible second function for Ku in repair and V(D)J joining: Ku helps ligation of blunt or nearly blunt DNA ends, especially by the DNA ligase IV/Xrcc4 complex that is discussed next (141, 142). The recently determined crystal structure of Ku bound to DNA (143) shows it to surround the double helix, suggesting why it prefers DNA ends and other irregularities, and possibly explaining its ability to bridge DNA ends.

DNA ligase IV and Xrcc4 are the two other repair factors in V(D)J joining that have been identified by genetic means. DNA ligase IV is one of three genetically distinct ligases in mammalian cells (144, 145). Xrcc4 complexes with ligase IV, stabilizes it in vivo, and increases its activity in vitro (146, 147). Cell lines defective in ligase IV or Xrcc4 are viable but radiation sensitive and do not make either coding joints or signal joints in V(D)J recombination (138, 148–150). Mice with one of these genes disrupted die late in fetal development, with primary defects in the nervous system (149, 151, 152). At first, it seemed possible that neuronal cells might perform site-specific recombination analogous to the V(D)J reaction, with cell death resulting from unsealed breaks. However, viable mice were obtained by combining disruptions of either ligase IV or Xrcc4 with a disruption of p53 (149, 152) or ATM (153). Thus developing neuronal cells may simply be unusually sensitive to apoptosis if there is accidental DNA damage, and this apoptosis requires signaling via p53 and ATM.

It is not yet clear whether the role of Xrcc4 is limited to its interaction with ligase IV, because Xrcc4 also binds to DNA nonspecifically, and by mutational tests its biological activity may correlate better with DNA binding than with ligase IV stimulation (154). A truncated Xrcc4 protein (residues 1–200 out of 336) is sufficient to support V(D)J recombination in cells (150, 154), and the crystal structure of this fragment has been solved (155), but not yet in complexes with ligase IV or DNA.

A gene defect that causes one form of human severe combined immune deficiency, makes cells radiosensitive, and impairs V(D)J joining has recently been associated with a factor named Artemis (156). The molecular activity of this factor is still unknown.

The factors described so far are dispensable for the viability of cells. Factors essential both for repair and cell survival would not be so easily identified by genetic screens. However, several other factors have been implicated on the basis of biochemical and other results.

One such protein is histone H2AX, an H2A subtype that accounts for 2–25% of the H2A content of various mammalian cells and tissues (157). A very early event after X-irradiation is the phosphorylation of H2AX on a serine residue in its C-terminal tail (158), and the accumulation of this species in nuclear foci at the sites of DNA breaks (159). Later, other repair factors such as Brca1, Mre11/Rad50/Nbs1 (see below), or Rad51 are drawn into these foci (160). In *S. cerevisiae*, mutation of the phosphorylation site in the homologous histone leads to radiosensitivity (160a). The H2AX phosphorylation site is in an SQ motif that could be a target for DNA-PK, ATM, or ATR, but it is not yet clear which kinase is mainly responsible. Wortmannin inhibits phosphorylation at high concentrations that would block all three enzymes, and a human cell line (MO59J) deficient in both DNA-PK and ATM has a greatly depressed level of H2AX phosphorylation (160). The other repair factors do not accumulate in foci in cells where H2AX phosphorylation is deficient. H2AX phosphorylation may thus act as an early signal for repair of radiation damage. Foci of phosphorylated H2AX also form at breaks associated with V(D)J recombination, suggesting that H2AX is involved in repair of these breaks as well (161).

In *S. cerevisiae*, another set of factors, the Mre11/Rad50/Xrs2 complex, is required for DNA repair by nonhomologous end joining (162), in addition to its important roles in meiotic recombination, telomere stability, and recombinational repair (163). By analogy, mammalian Mre11/Rad50/Nbs1 (Nbs1 is the equivalent in mammalian cells of Xrs2) may also be important for radiation repair and V(D)J joining. Genetic dissection of this complex has not progressed far, because all three factors are required for viability of mammalian cells (164–166), but the complex is likely to have a role in DNA damage sensing, as indicated by the properties of some human diseases. A partially defective mutation in Nbs1 causes Nijmegen breakage syndrome, associated with radiation sensitivity, chromosome instability, and a high incidence of tumors (167, 168). Several mutations in

Mre11 lead to the ataxia-telangiectasia-like disorder (ATLD), which displays a similar pattern of symptoms (169).

In addition, the triple complex accumulates in foci at sites of radiation damage (170, 171) after phosphorylation and focus formation of histone H2AX, as mentioned above. Nbs1 is associated with phosphorylated H2AX at V(D)J-induced breaks (161), which also suggests a role for Mre11/Rad50/Nbs1 in the processing of these breaks.

The biochemical activities in this complex are consistent with repair functions. Mre11 (either yeast or human) has a single-strand endonuclease and a 3′ to 5′ DNA exonuclease activity (172–175), which promotes the joining of nonhomologous DNA ends by a DNA ligase, at sites of short homologies (1–4 bp) (176). Mre11 is able to bridge DNA ends, and it appears likely that the exonuclease pauses at sites of homology. Mre11 may well be responsible for the frequent use of microhomologies in V(D)J coding joints. Recent crystal structures show that Rad50 is an SMC-like protein, with globular regions forming ATP- and DNA-binding sites, and separated by a long helical stem (177). SMC proteins have numerous roles in DNA transactions. Mre11 is found to have plausible binding sites for Rad50 and DNA, and sites for two metal ions involved in catalysis (178).

Mre11 and Rad50 are also highly homologous to the *Escherichia coli* proteins SbcD and SbcC, which are known to have a role in the processing of DNA hairpins (179). In correlation, the Mre11/Rad50/Nbs1 complex can nick DNA hairpins in vitro (180), but suitable mutations for testing the biological role of this activity are not yet available.

CONTROL OF V(D)J RECOMBINATION

V(D)J recombination is regulated in at least two ways, first by the expression pattern of RAG1 and RAG2, and second by the limited access of the recombination machinery to particular DNA sites. The first topic is largely outside the scope of this review [for discussion see (180a) and references therein], and only a few points are mentioned here. RAG1 and RAG2 are normally expressed together, and only in early lymphoid cells. The amount of RAG2 protein has been found to vary greatly through the cell cycle, being high in G1 and then decreasing by a factor of 20 or more in the S, G2, and M phases, without much change in RAG2 mRNA (181). The level of RAG2 protein may be controlled by phosphorylation followed by degradation, because mutation of the T490 phosphorylation site in RAG2 increases the steady-state level of RAG2 protein in cells (182). In vitro, T490 is a target for phosphorylation by the cell cycle kinase Cdc2, implying that this event may target RAG2 for destruction in cells (182). Variations in RAG2 could explain why V(D)J recombination and V(D)J breaks occur chiefly in the G1 phase (39, 181).

Isolated expression of one RAG gene without the other has been found in two situations, RAG2 in the bursa of Fabricius in chickens, and very low-level expression of RAG1 in mouse neuronal cells (183, 184). Gene disruption has not caused a functional defect in either case (49, 185). There is still no evidence that one RAG protein has a biological function in the absence of the other.

Regulation of Site Accessibility

Even in lymphoid cells expressing RAG1 and RAG2, only a small fraction of RSS sites is available for recombination at any one time. There are constraints on both the cell type and the developmental stage. Although the same recombination machinery is at work in B and T cells, Ig genes become fully rearranged only in B cells, and TCR genes only in T cells. Each of these lineages also has a preferred order of rearrangement: The IgH locus is rearranged before the Ig light chain loci, and TCRβ before TCRα. Furthermore, at some loci a successful recombination (that produces a protein chain) on one chromosome prevents rearrangement on the other allele, a restriction that is known as allelic exclusion (186).

In the basal state of chromosomal DNA, RSS sites are presumably unavailable. As one example, fibroblasts in which RAG1/2 is expressed will rearrange artificial substrates (which escape normal regulation) but fail to recombine any of their antigen receptor loci. Quite apart from the Ig and TCR loci, many sequences at random in chromosomes are enough like an RSS to be substrates for RAG1/2, and their use must be avoided if cells are to survive. The accessibility hypothesis (187, 188), based on these facts, proposes that RSS sites are normally blocked by chromatin or proteins bound to it, or by DNA modification, and have to be actively opened before recombination can take place. A related type of control has been described in yeast homologous recombination initiated by HO endonuclease breaks (189), and may be more widespread in eukaryotes than is yet known.

The exact means required to open a locus are under active study and are the main topic of this section. Accessibility has, for example, been correlated with germ line transcription of a locus (2) or with DNA demethylation (190–192), but it is not known whether these changes are the cause or the result of locus opening. In some cases there is no correlation of germ line transcription with recombination (193). Studies of model versions of Ig or TCR loci have shown the one common element required for locus opening to be the enhancer region (2). A locus might then become accessible by a process analogous to transcriptional activation, although not necessarily dependent on exactly the same set of factors or on transcription itself (2).

Pursuing the transcriptional analogy, recent attention has been focused on the role of histone acetylation and deacetylation in controlling accessibility. In mice with a TCRδ minilocus transgene, acetylation of histone H3 increases over a large region in parallel with V(D)J recombination (193a). This hyperacetylation occurs even in RAG2 knockout mice, showing that it precedes recombination. Also, if histone deacetylases are inhibited in pre-B cells, thus increasing the

overall level of acetylation, V(D)J recombination at the Igκ locus is increased (194).

In some circumstances, a few *trans*-acting factors are sufficient to activate chromosomal V(D)J recombination. In human kidney cells, which would normally not rearrange their antigen receptor loci, expression of RAG1/2 and suitable transcription factors activates recombination at particular loci. Expression of either of the E2A proteins (E12 or E47) activates Igκ recombination, EBF activates recombination at Igλ, and either one of these factors permits IgH D to J recombination, but not V to DJ (195). Furthermore, expression of E12, E47, or the related protein HEB activates rearrangement at TCRγ and TCRδ (196). This simple model system should be very useful in the dissection of accessibility control.

Accessibility can to some extent be controlled at the level of the RSSs themselves. At the TCRβ locus, the 12/23 rule would allow a V segment to join either to a D or directly to a J segment (see Figure 1*B*). However, V to J joining is rare, and the V to D preference is retained if the RSS from the D segment is moved into the J cluster (197). The RSSs at D and J are slightly different, and whether this preference operates only if the RSSs remain in the TCRβ environment is not yet known.

Control of recombination may also occur at the joining stage, after cleavage. Removal of the TCRβ enhancer has been reported to reduce cleavage moderately, but to diminish completed recombination much more, suggesting that a second level of regulation might be operative (198).

Modulation of the first stage of V(D)J recombination has been investigated in cell-free model systems. In permeabilized nuclei isolated from pre-B or pre-T cells and incubated in a cell extract enriched with RAG1/2, RSSs are cleaved only at the antigen receptor loci appropriate to the cell type and stage (199). It may prove possible to isolate a subnuclear fraction that still retains the same level of in vivo regulation.

The effect of factors that modify or remodel chromatin has been studied in cell-free recombination models. V(D)J cleavage can be inhibited up to 100-fold if the RSS is within a mononucleosome (194, 200, 201). Inhibition is relieved if the nucleosome is treated with the Swi/Snf remodeling complex, or assembled from hyperacetylated histones, and if these two alterations are combined, cleavage can be restored to almost that of naked DNA (202).

There is also evidence that the RAG proteins themselves may play a part in controlling access to certain chromatin structures. The same N-terminally truncated RAG1 and C-terminally truncated RAG2 proteins used in cell-free experiments support recombination of test substrates in vivo. In a pre-B cell line, the truncated form of RAG2 (together with RAG1) supports Igκ and IgH D to J rearrangement, but IgH V to DJ recombination does not occur (202a). Similarly, a naturally occurring RAG1 N-terminal truncation permits TCR but not Ig gene rearrangement (203). Thus portions of the RAG proteins that are dispensable for

their basal catalytic activity may interact specifically with certain remodeled forms of chromatin or with remodeling factors.

ACKNOWLEDGMENTS

I am indebted to Jessica Jones, Meni Melek, and I-hung Shih for valuable suggestions, and to Meni Melek for making the figures.

The *Annual Review of Biochemistry* is online at http://biochem.annualreviews.org

LITERATURE CITED

1. Lewis SM. 1994. *Adv. Immunol.* 56:27–150
2. Sleckman BP, Gorman JR, Alt FW. 1996. *Annu. Rev. Immunol.* 14:459–81
3. Gellert M. 1997. *Adv. Immunol.* 64:39–64
4. Fugmann SD, Lee AI, Shockett PE, Villey IJ, Schatz DG. 2000. *Annu. Rev. Immunol.* 18:495–527
5. Wagner SD, Neuberger MS. 1996. *Annu. Rev. Immunol.* 14:441–57
6. Reynaud CA, Bertocci B, Dahan A, Weill JC. 1994. *Adv. Immunol.* 57:353–78
7. Thompson CB. 1992. *Trends Genet.* 8:416–22
8. Tonegawa S. 1983. *Nature* 302:575–81
9. Lewis S, Gifford A, Baltimore D. 1985. *Science* 228:677–85
10. Hesse JE, Lieber MR, Gellert M, Mizuuchi K. 1987. *Cell* 49:775–83
11. Gerstein RM, Lieber MR. 1993. *Genes Dev.* 7:1459–69
12. Boubnov NV, Wills ZP, Weaver DT. 1995. *Nucleic Acids Res.* 23:1060–67
13. Ezekiel UR, Engler P, Stern D, Storb U. 1995. *Immunity* 2:381–89
14. Ramsden DA, Wu GE. 1991. *Proc. Natl. Acad. Sci. USA* 88:10721–25
15. Feeney AJ, Tang A, Ogwaro KM. 2000. *Immunol. Rev.* 175:59–69
16. Fujimoto S, Yamagishi H. 1987. *Nature* 327:242–43
17. Okazaki K, Davis DD, Sakano H. 1987. *Cell* 49:477–85
18. Roth DB, Nakajima PB, Menetski JP, Bosma MJ, Gellert M. 1992. *Cell* 69:41–53
19. Zachau HG. 1993. *Gene* 135:167–73
20. Yoder JA, Litman GW. 2000. *Curr. Top. Microbiol. Immunol.* 248:271–82
21. Bailey SN, Rosenberg N. 1997. *Mol. Cell. Biol.* 17:887–94
22. Lipkowitz S, Stern MH, Kirsch IR. 1990. *J. Exp. Med.* 172:409–18
23. Han JO, Steen SB, Roth DB. 1999. *Mol. Cell* 3:331–38
24. Tevelev A, Schatz DG. 2000. *J. Biol. Chem.* 275:8341–48
25. Lieber MR, Hesse JE, Mizuuchi K, Gellert M. 1988. *Proc. Natl. Acad. Sci. USA* 85:8588–92
26. Casellas R, Shih TA, Kleinewietfeld M, Rakonjac J, Nemazee D, et al. 2001. *Science* 291:1541–44
27. Gilfillan S, Dierich A, Lemeur M, Benoist C, Mathis D. 1993. *Science* 261:1175–78
28. Komori T, Okada A, Stewart V, Alt FW. 1993. *Science* 261:1171–75
29. Lafaille JJ, DeCloux A, Bonneville M, Takagaki Y, Tonegawa S. 1989. *Cell* 59:859–70
30. McCormack WT, Tjoelker LW, Carlson LM, Petryniak B, Barth CF, et al. 1989. *Cell* 56:785–91
31. Aguilar LK, Belmont JW. 1991. *J. Immunol.* 146:1348–52
32. Engler P, Klotz E, Storb U. 1992. *J. Exp. Med.* 176:1399–404

33. Gauss GH, Lieber MR. 1996. *Mol. Cell. Biol.* 16:258–69
34. Besmer E, Mansilla-Soto J, Cassard S, Sawchuk DJ, Brown G, et al. 1998. *Mol. Cell* 2:817–28
35. Lewis SM, Hesse JE, Mizuuchi K, Gellert M. 1988. *Cell* 55:1099–107
36. Morzycka-Wroblewska E, Lee FEH, Desiderio SV. 1988. *Science* 242: 261–63
37. Alexandre D, Chuchana P, Roncarolo M-G, Yssel H, Spits H, et al. 1991. *Int. Immunol.* 3:973–82
38. Lewis SM, Hesse JE. 1991. *EMBO J.* 10:3631–39
39. Schlissel M, Constantinescu A, Morrow T, Baxter M, Peng A. 1993. *Genes Dev.* 7:2520–32
40. Ramsden DA, Gellert M. 1995. *Genes Dev.* 9:2409–20
41. Zhu C, Roth DB. 1995. *Immunity* 2:101–12
42. Livak F, Schatz DG. 1996. *Mol. Cell. Biol.* 16:609–18
43. Steen SB, Gomelsky L, Roth DB. 1996. *Genes Cells* 1:543–53
44. Roth DB, Zhu C, Gellert M. 1993. *Proc. Natl. Acad. Sci. USA* 90:10788–92
45. Roth DB, Menetski JP, Nakajima PB, Bosma MJ, Gellert M. 1992. *Cell* 70:983–91
46. Schlissel MS. 1998. *Mol. Cell. Biol.* 18:2029–37
47. Oettinger MA, Schatz DG, Gorka C, Baltimore D. 1990. *Science* 248: 1517–23
48. Schatz DG, Oettinger MA, Schlissel MS. 1992. *Annu. Rev. Immunol.* 10:359–83
49. Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tonegawa S, Papaioannou VE. 1992. *Cell* 68:869–77
50. Shinkai Y, Rathbun G, Lam K-P, Oltz EM, Stewart V, et al. 1992. *Cell* 68: 855–67
51. Willett CE, Cherry JJ, Steiner LA. 1997. *Immunogenetics* 45:394–404
52. Hansen JD, Kaattari SL. 1995. *Immunogenetics* 42:188–95
53. Thompson CB. 1995. *Immunity* 3:531–39
54. Sadofsky MJ, Hesse JE, McBlane JF, Gellert M. 1993. *Nucleic Acids Res.* 21:5644–50
55. Silver DP, Spanopoulou E, Mulligan RC, Baltimore D. 1993. *Proc. Natl. Acad. Sci. USA* 90:6100–4
56. Cuomo CA, Oettinger MA. 1994. *Nucleic Acids Res.* 22:1810–14
57. Sadofsky MJ, Hesse JE, Gellert M. 1994. *Nucleic Acids Res.* 22:1805–9
58. Kirch SA, Sudarsanam P, Oettinger MA. 1996. *Eur. J. Immunol.* 26:886–91
59. Steen SB, Han JO, Mundy C, Oettinger MA, Roth DB. 1999. *Mol. Cell. Biol.* 19:3010–17
60. McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, et al. 1995. *Cell* 83:387–95
61. Sawchuk DJ, Weis-Garcia F, Malik S, Besmer E, Bustin M, et al. 1997. *J. Exp. Med.* 185:2025–32
62. van Gent DC, Hiom K, Paull TT, Gellert M. 1997. *EMBO J.* 16:2265–670
63. Eastman QM, Leu TM, Schatz DG. 1996. *Nature* 380:85–88
64. van Gent DC, Ramsden DA, Gellert M. 1996. *Cell* 85:107–13
65. Kim DR, Oettinger MA. 1998. *Mol. Cell. Biol.* 18:4679–88
66. van Gent DC, Mizuuchi K, Gellert M. 1996. *Science* 271:1592–94
67. Yu K, Lieber MR. 2000. *Mol. Cell. Biol.* 20:7914–21
68. Cuomo CA, Mundy CL, Oettinger MA. 1996. *Mol. Cell. Biol.* 16:5683–90
69. Ramsden DA, McBlane JF, van Gent DC, Gellert M. 1996. *EMBO J.* 15:3197–206
70. Cheung S, Arndt K, Lu P. 1984. *Proc. Natl. Acad. Sci. USA* 81:3665–69
71. Patel DJ, Shapiro L, Hare D. 1987. In *Unusual DNA Structures*, ed. RD Wells, SC Harvey, pp. 115–61. New York: Springer

72. Timsit Y, Vilbois E, Moras D. 1991. *Nature* 354:167–70
73. Kale SB, Landree MA, Roth DB. 2001. *Mol. Cell. Biol.* 21:459–66
74. Engelman A, Mizuuchi K, Craigie R. 1991. *Cell* 67:1211–21
75. Mizuuchi K, Adzuma K. 1991. *Cell* 66:129–40
76. Chow SA, Vincent KA, Ellison V, Brown PO. 1992. *Science* 255:723–26
77. Bogue MA, Wang C, Zhu C, Roth DB. 1997. *Immunity* 7:37–47
78. Melek M, Gellert M, van Gent DC. 1998. *Science* 280:301–3
79. Agrawal A, Eastman QM, Schatz DG. 1998. *Nature* 394:744–51
80. Hiom K, Melek M, Gellert M. 1998. *Cell* 94:463–70
81. Melek M, Gellert M. 2000. *Cell* 101: 625–33
82. Hartl DL, Lohe AR, Lozovskaya ER. 1997. *Annu. Rev. Genet.* 31:337–58
83. Kennedy AK, Guhathakurta A, Kleckner N, Haniford DB. 1998. *Cell* 95:125–34
84. Bhasin A, Goryshin IY, Reznikoff WS. 1999. *J. Biol. Chem.* 274:37021–29
85. Colot V, Haedens V, Rossignol JL. 1998. *Mol. Cell. Biol.* 18:4337–46
86. Coen ES, Carpenter R, Martin C. 1986. *Cell* 47:285–96
87. Atkinson PW, Warren WD, O'Brochta DA. 1993. *Proc. Natl. Acad Sci. USA* 90:9693–97
88. Sakano H, Huppi K, Heinrich G, Tonegawa S. 1979. *Nature* 280:288–94
89. Shockett PE, Schatz DG. 1999. *Mol. Cell. Biol.* 19:4159–66
90. Schultz HY, Landree MA, Qiu JX, Kale SB, Roth DB. 2001. *Mol. Cell* 7:65–75
91. Santagata S, Besmer E, Villa A, Bozzi F, Allingham JS, et al. 1999. *Mol. Cell* 4:935–47
92. Sadofsky MJ. 2001. *Nucleic Acids Res.* 29:1399–409
93. Hiom K, Gellert M. 1997. *Cell* 88:65–72
94. Akamatsu Y, Oettinger MA. 1998. *Mol. Cell. Biol.* 18:4670–78
95. Nagawa F, Ishiguro K, Tsuboi A, Yoshida T, Ishikawa A, et al. 1998. *Mol. Cell. Biol.* 18:655–63
96. Swanson PC, Desiderio S. 1998. *Immunity* 9:115–25
97. Eastman QM, Villey IJ, Schatz DG. 1999. *Mol. Cell. Biol.* 19:3788–97
98. Mo XM, Bailin T, Sadofsky MJ. 1999. *J. Biol. Chem.* 274:7025–31
99. Hiom K, Gellert M. 1998. *Mol. Cell* 1:1011–19
100. West RB, Lieber MR. 1998. *Mol. Cell. Biol.* 18:6408–15
101. Agrawal A, Schatz DG. 1997. *Cell* 89:43–53
102. Ramsden DA, Paull TT, Gellert M. 1997. *Nature* 388:488–91
103. Mizuuchi K. 1992. *Annu. Rev. Biochem.* 61:1011–51
104. Callebaut I, Mornon JP. 1998. *Cell. Mol. Life Sci.* 54:880–91
105. Aravind L, Koonin EV. 1999. *J. Mol. Biol.* 287:1023–40
106. Freemont PS, Hanson IM, Trowsdale J. 1991. *Cell* 64:483–84
107. Rodgers KK, Bu Z, Fleming KG, Schatz DG, Engelman DM, Coleman JE. 1996. *J. Mol. Biol.* 260:70–84
108. Rice P, Craigie R, Davies DR. 1996. *Curr. Opin. Struct. Biol.* 6:76–83
109. Kim DR, Dai Y, Mundy CL, Yang W, Oettinger MA. 1999. *Genes Dev.* 13:3070–80
110. Landree MA, Wibbenmeyer JA, Roth DB. 1999. *Genes Dev.* 13:3059–69
111. Fugmann SD, Villey IJ, Ptaszek LM, Schatz DG. 2000. *Mol. Cell* 5:97–107
112. Sarnovsky RJ, May EW, Craig NL. 1996. *EMBO J.* 15:6348–61
113. Deleted in proof
114. Qiu JX, Kale SB, Schultz HY, Roth DB. 2001. *Mol. Cell* 7:77–87
115. Schwarz K, Gauss GH, Ludwig L, Pannicke U, Li Z, et al. 1996. *Science* 274:97–99
116. Villa A, Santagata S, Bozzi F, Giliani

S, Frattini A, et al. 1998. *Cell* 93:885–96
117. Villa A, Sobacchi C, Notarangelo LD, Bozzi F, Abinun M, et al. 2001. *Blood* 97:81–88
118. Corneo B, Moshous D, Callebaut I, de Chasseval R, Fischer A, de Villartay JP. 2000. *J. Biol. Chem.* 275:12672–75
119. Gomez CA, Ptaszek LM, Villa A, Bozzi F, Sobacchi C, et al. 2000. *Mol. Cell. Biol.* 20:5653–64
120. Difilippantonio MJ, McMahan CJ, Eastman QM, Spanopoulou E, Schatz DG. 1996. *Cell* 87:253–62
121. Spanopoulou E, Zaitseva F, Wang F, Santagata S, Baltimore D, Panayotou G. 1996. *Cell* 87:263–76
122. Steen SB, Zhu C, Roth DB. 1996. *Curr. Top. Microbiol. Immunol.* 217:61–77
123. Jeggo PA. 1998. *Adv. Genet.* 38:185–218
124. Smith GCM, Jackson SP. 1999. *Genes Dev.* 13:916–34
125. Hartley KO, Gell D, Smith GC, Zhang H, Divecha N, et al. 1995. *Cell* 82:849–56
126. Yaneva M, Kowalewski T, Lieber MR. 1997. *EMBO J.* 16:5098–112
127. Hammarsten O, Chu G. 1998. *Proc. Natl. Acad. Sci. USA* 95:525–30
128. Bosma MJ, Carroll AM. 1991. *Annu. Rev. Immunol.* 9:323–44
129. Blunt T, Finnie NJ, Taccioli GE, Smith GCM, Demengeot J, et al. 1995. *Cell* 80:813–23
130. Kirchgessner CU, Patil CK, Evans JW, Cuomo CA, Fried LM, et al. 1995. *Science* 267:1178–83
131. Peterson SR, Kurimasa A, Oshimura M, Dynan WS, Bradbury EM, Chen DJ. 1995. *Proc. Natl. Acad. Sci. USA* 92:3171–74
132. Bogue MA, Jhappan C, Roth DB. 1998. *Proc. Natl. Acad Sci. USA* 95:15559–64
133. Gao YJ, Chaudhuri J, Zhu CM, Davidson L, Weaver DT, Alt FW. 1998. *Immunity* 9:367–76
134. Taccioli GE, Amatucci AG, Beamish HJ, Gell D, Xiang XH, et al. 1998. *Immunity* 9:355–66
135. Shin EK, Perryman LE, Meek K. 1997. *J. Immunol.* 158:3565–69
136. Anderson CW, Lees-Miller SP. 1992. *Crit. Rev. Eukaryot. Gene Expr.* 2:283–314
137. Pergola F, Zdzienicka MZ, Lieber MR. 1993. *Mol. Cell. Biol.* 13:3464–71
138. Taccioli GE, Rathbun G, Oltz E, Stamato T, Jeggo PA, Alt FW. 1993. *Science* 260:207–10
139. Gu YS, Jin SF, Gao YJ, Weaver DT, Alt FW. 1997. *Proc. Natl. Acad. Sci. USA* 94:8076–81
140. Haber JE. 1999. *Cell* 97:829–32
141. Ramsden DA, Gellert M. 1998. *EMBO J.* 17:609–14
142. McElhinny SAN, Snowden CM, McCarville J, Ramsden DA. 2000. *Mol. Cell. Biol.* 20:2996–3003
143. Walker JR, Corpina RA, Goldberg J. 2001. *Nature* 412:607–14
144. Wei YF, Robins P, Carter K, Caldecott K, Pappin DJC, et al. 1995. *Mol. Cell. Biol.* 15:3206–16
145. Robins P, Lindahl T. 1996. *J. Biol. Chem.* 271:24257–61
146. Critchlow SE, Bowater RP, Jackson SP. 1997. *Curr. Biol.* 7:588–98
147. Grawunder U, Wilm M, Wu X, Kulesza P, Wilson TE, et al. 1997. *Nature* 388: 492–95
148. Li ZY, Otevrel T, Gao YJ, Cheng H-L, Seed B, et al. 1995. *Cell* 83:1079–89
149. Gao YJ, Sun Y, Frank KM, Dikkes P, Fujiwara Y, et al. 1998. *Cell* 95:891–902
150. Grawunder U, Zimmer D, Fugmann S, Schwarz K, Lieber MR. 1998. *Mol. Cell* 2:477–84
151. Barnes DE, Stamp G, Rosewell I, Denzel A, Lindahl T. 1998. *Curr. Biol.* 8:1395–98
152. Frank KM, Sekiguchi JM, Seidl KJ, Swat W, Rathbun GA, et al. 1998. *Nature* 396:173–77

153. Lee Y, Barnes DE, Lindahl T, McKinnon PJ. 2000. *Genes Dev.* 14:2576–80
154. Modesti M, Hesse JE, Gellert M. 1999. *EMBO J.* 18:2008–18
155. Junop MS, Modesti M, Guarne A, Ghirlando R, Gellert M, Yang W. 2000. *EMBO J.* 19:5962–70
156. Moshous D, Callebaut I, de Chasseval R, Corneo B, Cavazzana-Calvo M, et al. 2001. *Cell* 105:177–86
157. Modesti M, Kanaar R. 2001. *Curr. Biol.* 11:R229–32
158. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. 1998. *J. Biol. Chem.* 273:5858–68
159. Rogakou EP, Boon C, Redon C, Bonner WM. 1999. *J. Cell Biol.* 146:905–16
160. Paull TT, Rogakou EP, Yamazaki V, Kirchgessner CU, Gellert M, Bonner WM. 2000. *Curr. Biol.* 10:886–95
160a. Downs JA, Lowndes NF, Jackson SP. 2000. *Nature* 408:1001–4
161. Chen HT, Bhandoola A, Difilippantonio MJ, Zhu J, Brown MJ, et al. 2000. *Science* 290:1962–65
162. Boulton SJ, Jackson SP. 1998. *EMBO J.* 17:1819–28
163. Haber JE. 1998. *Cell* 95:583–86
164. Xiao YH, Weaver DT. 1997. *Nucleic Acids Res.* 25:2985–91
165. Luo G, Yao MS, Bender CF, Mills M, Bladl AR, et al. 1999. *Proc. Natl. Acad. Sci. USA* 96:7376–81
166. Zhu J, Petersen S, Tessarollo L, Nussenzweig A. 2001. *Curr. Biol.* 11:105–9
167. Carney JP, Maser RS, Olivares H, Davis EM, Le Beau M, et al. 1998. *Cell* 93:477–86
168. Varon R, Vissinga C, Platzer M, Cerosaletti KM, Chrzanowska KH, et al. 1998. *Cell* 93:467–76
169. Stewart GS, Maser RS, Stankovic T, Bressan DA, Kaplan MI, et al. 1999. *Cell* 99:577–87
170. Maser RS, Monsen KJ, Nelms BE, Petrini JH. 1997. *Mol. Cell. Biol.* 17:6087–96
171. Nelms BE, Maser RS, MacKay JF, Lagally MG, Petrini JH. 1998. *Science* 280:590–92
172. Furuse M, Nagase Y, Tsubouchi H, Murakami-Murofushi K, Shibata T, Ohta K. 1998. *EMBO J.* 17:6412–25
173. Paull TT, Gellert M. 1998. *Mol. Cell* 1:969–79
174. Trujillo KM, Yuan SS, Lee EY, Sung P. 1998. *J. Biol. Chem.* 273:21447–50
175. Usui T, Ohta T, Oshiumi H, Tomizawa J, Ogawa H, Ogawa T. 1998. *Cell* 95:705–16
176. Paull TT, Gellert M. 2000. *Proc. Natl. Acad. Sci. USA* 97:6409–14
177. Hopfner KP, Karcher A, Shin DS, Craig L, Arthur LM, et al. 2000. *Cell* 101:789–800
178. Hopfner KP, Karcher A, Craig L, Woo TT, Carney JP, Tainer JA. 2001. *Cell* 105:473–85
179. Connelly JC, de Leau ES, Leach DR. 1999. *Nucleic Acids Res.* 27:1039–46
180. Paull TT, Gellert M. 1999. *Genes Dev.* 13:1276–88
180a. Nagaoka H, Yu W, Nussenzweig MC. 2000. *Curr. Opin. Immunol.* 12:187–90
181. Lin W-C, Desiderio S. 1995. *Immunol. Today* 16:279–89
182. Lin W-C, Desiderio S. 1993. *Science* 260:953–59
183. Carlson LM, Oettinger MA, Schatz DG, Masteller EL, Hurley EA, et al. 1991. *Cell* 64:201–8
184. Chun JJ, Schatz DG, Oettinger MA, Jaenisch R, Baltimore D. 1991. *Cell* 64:189–200
185. Takeda S, Masteller EL, Thompson CB, Buerstedde J-M. 1992. *Proc. Natl. Acad. Sci. USA* 89:4023–27
186. Blackwell TK, Alt FW. 1989. *Annu. Rev. Genet.* 23:605–36
187. Yancopoulos GD, Alt FW. 1986. *Annu. Rev. Immunol.* 4:339–68
188. Roth DB, Roth SY. 2000. *Cell* 103: 699–702
189. Haber JE. 1998. *Annu. Rev. Genet.* 32:561–99

190. Hsieh C-L, Lieber MR. 1992. *EMBO J.* 11:315–25
191. Mostoslavsky R, Singh N, Kirillov A, Pelanda R, Cedar H, et al. 1998. *Genes Dev.* 12:1801–11
192. Cherry SR, Baltimore D. 1999. *Proc. Natl. Acad. Sci. USA* 96:10788–93
193. Angelin-Duclos C, Calame K. 1998. *Mol. Cell. Biol.* 18:6253–64
193a. McMurry MT, Krangel MS. 2000. *Science* 287:495–98
194. McBlane F, Boyes J. 2000. *Curr. Biol.* 10:483–86
195. Romanow WJ, Langerak AW, Goebel P, Wolvers-Tettero IL, van Dongen JJ, et al. 2000. *Mol. Cell* 5:343–53
196. Ghosh JK, Romanow WJ, Murre C. 2001. *J. Exp. Med.* 193:769–76
197. Bassing CH, Alt FW, Hughes MM, D'Auteuil M, Wehrly TD, et al. 2000. *Nature* 405:583–86
198. Hempel WM, Stanhope-Baker P, Mathieu N, Huang F, Schlissel MS, Ferrier P. 1998. *Genes Dev.* 12:2305–17
199. Stanhope-Baker P, Hudson KM, Shaffer AL, Constantinescu A, Schlissel MS. 1996. *Cell* 85:887–97
200. Kwon J, Imbalzano AN, Matthews A, Oettinger MA. 1998. *Mol. Cell* 2:829–39
201. Golding A, Chandler S, Ballestar E, Wolffe AP, Schlissel MS. 1999. *EMBO J.* 18:3712–23
202. Kwon J, Morshead KB, Guyon JR, Kingston RE, Oettinger MA. 2000. *Mol. Cell* 6:1037–48
202a. Kirch SA, Rathbun GA, Oettinger MA. 1998. *EMBO J.* 17:4881–86
203. Noordzij JG, Verkaik NS, Hartwig NG, de Groot R, van Gent DC, van Dongen JJ. 2000. *Blood* 96:203–9

Annual Review of Biochemistry
Volume 71, 2002

CONTENTS