Cell, Vol. 59, 1035–1048, December 22, 1989, Copyright © 1989 by Cell Press

The V(D)J Recombination Activating Gene, RAG-1

David G. Schatz, Marjorie A. Oettinger, and David Baltimore
Whitehead Institute for Biomedical Research
Nine Cambridge Center
Cambridge, Massachusetts 02142
and Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Summary

The RAG-1 (recombination activating gene-1) genomic locus, which activates V(D)J recombination when introduced into NIH 3T3 fibroblasts, was isolated by serial genomic transfections of oligonucleotide-tagged DNA. A genomic walk spanning 55 kb yielded a RAG-1 genomic probe that detects a single 6.6–7.0 kb mRNA species in transfectants and pre-B and pre-T cells. RAG-1 genomic and cDNA clones were biologically active when introduced into NIH 3T3 cells. Nucleotide sequencing of human and mouse RAG-1 cDNA clones predicts 119 kd proteins of 1043 and 1040 amino acids, respectively, with 90% sequence identity. RAG-1 has been conserved between species that carry out V(D)J recombination, and its pattern of expression correlates exactly with the pattern of expression of V(D)J recombinase activity. RAG-1 may activate V(D)J recombination indirectly, or it may encode the V(D)J recombinase itself.

Introduction

The enormous number of genes required to encode the subunits of the immunoglobulin (Ig) and T cell receptor (TCR) molecules of B and T cells are generated combinatorially in a process known as V(D)J recombination, so called for the variable (V), diversity (D), and joining (J) gene segments used in the recombination process. V(D)J recombination is known to assemble seven different loci in developing lymphocytes: μ, κ, and λ in B cells, and α, β, γ, and δ in T cells (for reviews see Blackwell and Alt, 1988; Davis and Bjorkman, 1988; Raulet, 1989). The B and T cell lineages appear to share a common V(D)J recombinase (see below) whose expression is tightly regulated during development; recombination of endogenous loci or exogenously introduced substrates occurs only in cells representing the early stages of B and T cell development and does not occur in nonlymphoid cells (Lieber et al., 1987; Schatz and Baltimore, 1988). V(D)J recombination is also regulated at the level of substrate availability, since loci normally recombined in one cell type are never fully recombined in the other. In some instances, it has been possible to correlate the recombination of a locus with its transcription and increased sensitivity to DNAase I, suggesting that the "accessibility" of a locus to the recombinase may be modulated by its transcription (Yancopoulos and Alt, 1985; Blackwell et al., 1986; Yancopoulos et al., 1986; Schlissel and Baltimore, 1989).

The *cis*-acting DNA sequences that are necessary and sufficient for V(D)J recombination consist of a palindromic heptamer and AT-rich nonamer separated by a spacer of either 12 or 23 bp (Tonegawa, 1983; Hesse et al., 1989). Spacer length defines two classes of such recombination signal sequences (RSSs); efficient recombination occurs only when one RSS of each class is involved (the "12–23 joining rule"). RSSs lie directly adjacent to all recombinationally competent V, D, and J gene segments, and their sequences are conserved between the recombining loci and between all species known to carry out V(D)J recombination (Litman et al., 1985a, 1985b; Reynaud et al., 1985; Schwager et al., 1988). RSSs from different loci and species appear to be functionally interchangeable because they direct proper V(D)J recombination in a variety of host pre-B and pre-T cells (Yancopoulos et al., 1986; Bucchini et al., 1987; Goodhardt et al., 1987; Lieber et al., 1987). The sequence conservation and functional equivalence of RSSs together provide strong evidence for a single, evolutionarily conserved V(D)J recombinase.

Remarkably little is known, however, about the *trans*-acting factors that participate in V(D)J recombination; no factor has yet been demonstrated to be an essential component of the enzyme. V(D)J recombination could be expected to require several distinct enzymatic activities: sequence-specific DNA recognition, endonucleolytic cleavage between the RSS and gene segment, and ligation of the cleaved ends. In addition, because nucleotides are frequently lost and/or added at coding segment junctions, both exonucleolytic and DNA polymerase activities should be associated with the recombinational machinery. It is not known how many factors participate in V(D)J recombination, nor is it known how many are lymphoid specific. We will use the term V(D)J recombinase to refer only to the essential component(s) of the enzyme whose activity is restricted to recombinationally active B and T cells. Thus far, only the enzyme terminal deoxynucleotidyl transferase (TdT) has been implicated in V(D)J recombination, where it is likely to add non-germline-encoded nucleotides (N regions) to the coding junction (Landau et al., 1987). TdT is not, however, required for V(D)J recombination. Some of the other activities expected to be associated with the recombinase have been detected in nuclear extracts of lymphoid cells—heptamer binding (Aguilera et al., 1987), nonamer binding (Halligan and Desiderio, 1987), and endonucleolytic cleavage (Desiderio and Baltimore, 1984; Hope et al., 1986; Kataoka et al., 1984)—but the relationship of these factors to V(D)J recombination remains unclear. Finally, a factor critical in V(D)J recombination may be contributed by the gene that is disrupted in the severe combined immunodeficient (*scid*) mouse (Bosma et al., 1983), which is characterized primarily by a defect in recombination of Ig and TCR gene segments (Schuler et al., 1986).

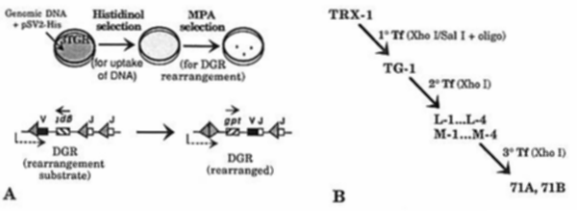


Figure 1. Genomic Transfection Protocol, Recombination Substrate, and Lineage of Genomic Transfectants

(A) Genomic transfection protocol. Genomic DNA and marker plasmid pSV2-His (which directs expression of the HIS gene) were cotransfected into 3TGR cells, an NIH 3T3 fibroblast line that contains two copies of the recombination substrate DGR (DGR and 3TGR have been described in detail in Schatz and Baltimore, 1988). The transfected cells were selected first in histidinol, for cells that had taken up DNA and were expressing the HIS gene, and then in MPA, which selects for expression of the *gpt* gene of pDGR. At the bottom is a schematic representation of the relevant portion of pDGR: in its unrearranged configuration (left), the *gpt* gene cannot be expressed because it lies in opposite transcriptional orientation to the promoter (dashed line); when V(D)J recombination occurs between the V region and one of the two J regions (structure for rearrangement of V to the proximal J is shown at right), the *gpt* gene is inverted and can then be expressed. Recombination signal sequences (RSSs) are represented as triangles; the V_κ21-C gene segment (V) is flanked by an RSS with a 12 bp spacer, and each J_κ1 gene segment (J) is flanked by an RSS with a 23 bp spacer. (B) Lineage of genomic transfectants derived in the oligonucleotide tagging of RAG-1. Genomic DNA from TRX-1 was digested with XhoI and SalI, ligated to the oligonucleotide, and transfected into 3TGR cells to generate the primary transfectant TG-1. TG-1 DNA was digested with XhoI and transfected into 3TGR cells to generate the L and M series of secondary transfectants. XhoI-digested L-4 DNA was used to generate pools of tertiary transfectants 71A and 71B.

Previously, we demonstrated that transfer of genomic DNA could induce stable expression of V(D)J recombinase activity in NIH 3T3 fibroblasts, a cell type that normally expresses no such activity (Schatz and Baltimore, 1988). The results suggested that transfer of a single genetic locus was sufficient for this phenomenon and that this locus might be isolated using serial transfections of genomic DNA.

Using an oligonucleotide marker in conjunction with genomic transfection, we have isolated a gene that stably activates V(D)J recombination in NIH 3T3 cells. The V(D)J recombination activating gene (RAG-1) accomplishes this either by encoding a component of the V(D)J recombinase itself, or by activating other factors that in turn carry out V(D)J recombination.

Results

Oligonucleotide Tagging of RAG-1

Based on our previous results (Schatz and Baltimore, 1988), we postulated that expression of a single lymphoid-specific gene could stably activate the V(D)J recombinase in NIH 3T3 fibroblasts. Isolation of this gene using genomic transfection required that the gene be marked or "tagged" in a way that allowed it to be followed through successive rounds of transfection. After several attempts to isolate the human RAG-1 gene failed (with human repetitive sequences as the marker), we set out to isolate the mouse gene by tagging it with an oligonucleotide.

Our genomic transfection procedure makes use of the recipient cell line 3TGR, an NIH 3T3 subclone containing two copies of the recombination substrate DGR. DGR allows for the identification of cells that express V(D)J recombinase activity based on their rearrangement and subsequent expression of the xanthine-guanine phosphoribosyltransferase (*gpt*) gene of DGR and resulting resistance to the drug mycophenolic acid (MPA; procedure depicted schematically in Figure 1A). In brief, genomic DNA and a drug resistance gene, histidinol dehydrogenase (HIS), are cotransfected into 3TGR cells, and the cells are first selected in histidinol (for cells that have taken up DNA and express the HIS gene) and then in MPA. 3TGR cells are killed by MPA because the *gpt* gene of DGR is not expressed (it lacks a promoter). However, if V(D)J recombinase activity is expressed in a transfected 3TGR cell, it can recognize the recombination signal sequences present in DGR (which lie adjacent to the V and J segments) and carry out an inversional V-to-J rearrangement event. The inverted *gpt* gene can now be expressed from the upstream promoter, and the cell becomes resistant to MPA (see Figure 1A). V-to-J recombination has never been observed in untransfected or control transfected 3TGR cells (rate less than 10^{-8}).

As a source of RAG-1 for oligonucleotide tagging, we chose DNA from the genomic transfectant TRX-1, because it had a roughly 10-fold higher specific activity of transfer of V(D)J recombinase activity than any primary DNA tested (S. Latt, D. G. S., and D. B., unpublished data; for the origin of TRX-1, see Schatz and Baltimore, 1988). This suggested that TRX-1 contained RAG-1 in an amplified and/or activated form. TRX-1 was derived by using genomic DNA from a previous transfectant (TR-1) that potentially contained both human (a transfected gene from a human B cell) and mouse (the endogenous, germline alleles from 3TGR cells) RAG-1. Analysis of DNA from TR-1 and TRX-1 strongly suggested that TRX-1 was generated by transfer of mouse RAG-1 (Schatz and Baltimore, 1988). We have confirmed that TRX-1 contains amplified, and perhaps activated, copies of mouse RAG-1 (see below).

Transfection experiments had shown that the restriction enzymes XhoI and SalI did not inactivate mouse RAG-1 (unpublished data). TRX-1 genomic DNA to be tagged with oligonucleotide was first digested to completion with these enzymes, which leave identical 4 nucleotide overhangs at the sites of cleavage, and was then ligated to a large excess of a 30 bp oligonucleotide, which was designed to have XhoI- and SalI-compatible overhangs at either end. Ligation of the oligonucleotide to itself or to the genomic DNA could regenerate SalI sites but not XhoI sites. The ligated mixture of genomic DNA and oligonucleotide was transfected into 3TGR cells, and the cells were subjected to the selection procedure shown in Figure 1A. In the process of ligation and transfection, we hoped that the oligonucleotide would become closely linked to the RAG-1 gene (shown schematically in Figure 2).

Of the numerous MPA-resistant colonies that resulted from the primary transfection, we chose two independent clones (TG-1 and TG-2; Figure 1B) for further analysis.

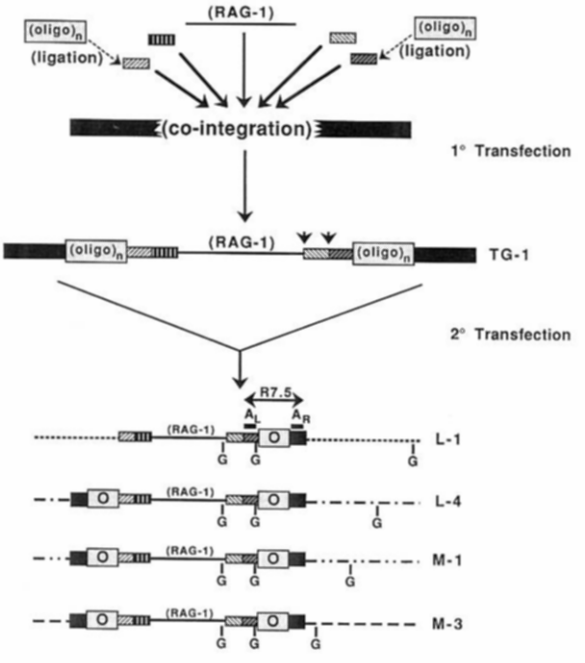


Figure 2. Schematic Representation of the Primary and Secondary Transfections

XhoI- and SalI-digested genomic DNA from TRX-1 was ligated to an excess of the oligonucleotide (top; XhoI–SalI genomic fragments from TRX-1 represented as striped boxes). The oligonucleotide arrays ("oligo$_n$" or "O") did not ligate to the fragment of DNA containing RAG-1 (germline RAG-1 sequences represented as a thin line); instead, it ligated to other genomic fragments (dashed arrows) and became linked to RAG-1 in the primary transfection by cointegration into the chromosome to yield TG-1 (center). The oligonucleotide arrays thereby came to be separated from RAG-1 by a number of irrelevant genomic fragments (striped boxes). Small vertical arrows above the TG-1 chromosome indicate some of the chromosomal breakpoints, where two fragments of DNA, not normally contiguous in the germline, are now juxtaposed. During secondary transfection, the RAG-1 locus, including the flanking blocks of oligonucleotide, was integrated into four independent chromosomal sites (indicated by the flanking dashed lines) in the secondary transfectants L-1, L-4, M-1, and M-3 (bottom). Relevant sites for the restriction enzyme BglII (G) as well as the positions of probes A_L and A_R are shown relative to the right-hand oligonucleotide array (representative of the array in fragment R7.5; see text and Figure 5). Between this array and RAG-1, the locus is identical in the four secondary transfectants (identical BglII fragments detected with probe A_L), while to the right of this oligonucleotide array, the locus differs (different BglII fragments detected by probe A_R). The conserved pattern of hybridization seen with probe A_L and divergent pattern seen with probe A_R suggested that RAG-1 lay to the left of this oligonucleotide array. The second array of oligonucleotide in transfectants L-4, M-1, and M-3 (to the left of RAG-1) was not present in L-1 (see Figure 4). Figure not drawn to scale.

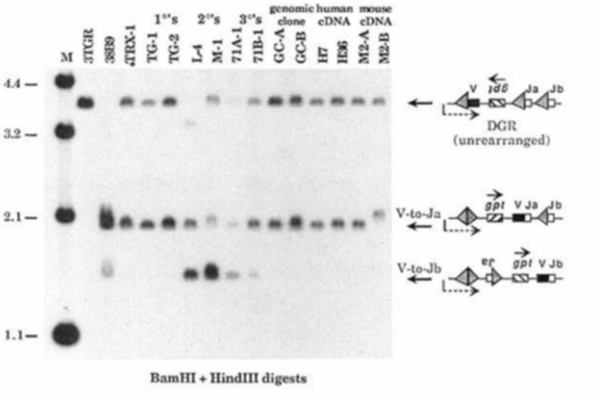


Figure 3. RAG-1 Transfectants Contain Normal V-to-J Rearrangements

Genomic DNA was digested with BamHI and HindIII, electrophoresed through a 0.8% agarose gel, transferred to a Zetabind membrane, and hybridized with radioactive probe 1 (a 1.1 kb fragment of pBR322 that lies just to the right of Jb in DGR; see Schatz and Baltimore, 1988, for additional details). The diagrams show the structures of the DGR DNA in each of the bands, with the symbols used in Figure 1. The upper band is derived from unrearranged DGR (3.8 kb), while V-to-J rearrangement yields bands of either 1.9 kb or 1.5 kb. The single band from 3TGR cells is generated by its two unrearranged copies of DGR, while the recombinationally active pre-B cell 38B9, infected with DGR and selected in MPA, shows the two bands expected for V-to-J rearrangement. Transfectants show one or both of the bands corresponding to correct V-to-J rearrangement; the slightly different position of the lower band from M2-B is also characteristic of normal V-to-J rearrangement of DGR (Landau et al., 1987). GC-A and GC-B are MPA^r colonies derived from the same transfection of phage clone 12C.2 into 3TGR cells. Human cDNA clones H7 and H36 in expression vector pJ3Ω and mouse cDNA clone M2 in expression vector CDM8 (see Experimental Procedures) were individually transfected into 3TGR cells to generate the colonies whose DNA is shown in the last four lanes (M2-A and M2-B represent MPA^r colonies from independent transfections). Marker sizes are shown in kb; exposure = 23 hr.

Both clones contained normal V-to-J rearrangements of the DGR construct, as shown by Southern blot analysis (Figure 3). Furthermore, using a second recombination substrate described previously (Schatz and Baltimore, 1988), both TG-1 and TG-2 were shown to express V(D)J recombinase activity stably (data not shown). When genomic DNA from TG-1 and TG-2 was digested with restriction enzymes, electrophoresed in an agarose gel, and the gel partially dried and then probed with the oligonucleotide, we observed that both clones contained a large number of bands that hybridized to the oligonucleotide (Figure 4; data for TG-2 not shown).

To determine if any of the oligonucleotide in TG-1 was closely linked to RAG-1, we performed secondary transfections of XhoI-digested TG-1 DNA into 3TGR cells. XhoI digestion was required to inactivate the rearranged copy of DGR in TG-1, which would otherwise transfer MPA resistance with high efficiency (discussed in Schatz and Baltimore, 1988). From each of four independent secondary transfections, two MPA-resistant clones were chosen for detailed analysis (L-1, L-2; L-3, L-4; M-1, M-2; M-3, M-4; Figure 1B). Each of the eight secondary transfectants showed multiple, normal V-to-J rearrangements of DGR, indicative of stable expression of V(D)J recombinase activity (representative data shown in Figure 3). Four of these transfectants (L-1, L-4, M-1, and M-3) contained DNA that hybridized to the oligonucleotide, with the pattern of hybridization being particularly significant (Figure 4). Each oligonucleotide-positive transfectant contained a comigrating 7.5 kb EcoRI fragment that hybridized to the oligonucleotide (fragment R7.5). Since each transfectant was independently derived and was selected only for V(D)J recombinase activity, this result strongly suggested that

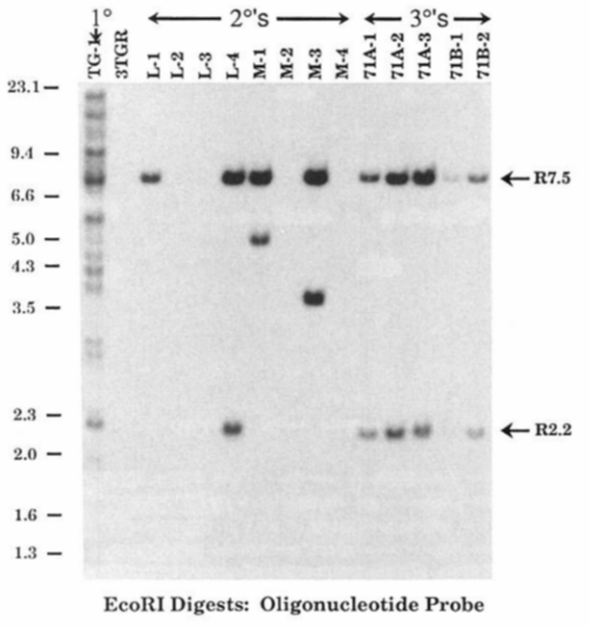


Figure 4. The Oligonucleotide Tag Cosegregates with V(D)J Recombinase Activity through Three Rounds of Transfection

Genomic DNA was digested with EcoRI, electrophoresed through a 0.8% agarose gel, and then probed in situ with one radiolabeled strand of the oligonucleotide used in the ligation to TRX-1 DNA (see Experimental Procedures). Each oligonucleotide-positive secondary transfectant (L-1, L-4, M-1, and M-3) was derived as a MPA^r clone from an independent transfection of TG-1 DNA into 3TGR cells. The tertiary transfectants are pools of MPA^r clones derived from two independent transfections (71A and 71B) of L-4 DNA into 3TGR cells. Marker sizes shown in kb; exposure = 48 hr.

fragment R7.5 represented the same segment of DNA in each transfectant and that, in these four transfectants, this segment of DNA was closely linked to the gene responsible for the activation of the recombinase.

To confirm the tight linkage between R7.5 and RAG-1, XhoI-digested L-4 DNA was transfected into 3TGR cells, and tertiary transfectants were derived (Figure 1B). Analysis of pools of tertiary transfectants (71A and 71B) showed that they contained normal V-to-J rearrangements of DGR (Figure 3) and that they contained fragment R7.5 as well as the second oligonucleotide-hybridizing fragment from L-4 (R2.2; Figure 4). Thus both R7.5 and R2.2 cosegregated with V(D)J recombinase activity. Since our preparations of genomic DNA typically contained few, if any, fragments larger than 150 kb (data not shown), R7.5 and R2.2 were likely to lie within 150 kb of RAG-1.

The Genomic Walk to RAG-1

As the first step in isolating the genomic sequences encoding RAG-1, we attempted to clone fragment R7.5 by using the oligonucleotide to screen genomic libraries of EcoRI-digested, size-selected M-1 DNA (constructed in the bacteriophage vectors λgt10 and λZAP). When the libraries were grown on standard bacterial host strains, we failed to obtain any positives after screening more than 7 × 10^5 recombinants. However, we obtained a single R7.5-containing clone from 1 × 10^5 λgt10 recombinants grown

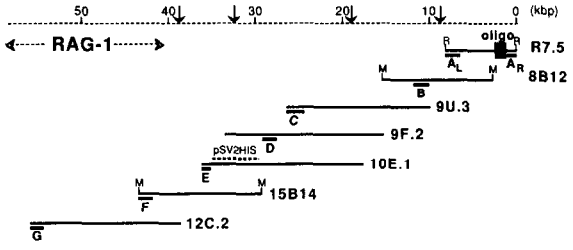


Figure 5. Genomic Walk to RAG-1

The oligonucleotide was used as a probe to obtain the genomic clone R7.5 (see Experimental Procedures). As described in the text and legend to Figure 2, RAG-1 was shown to lie to the left of the oligonucleotide in this clone. Probe A_L was used to obtain genomic clone 8B12, from which probe B was obtained. Repeating this procedure with probes B, C, D, E, and F yielded the additional genomic clones 9U.3, 9F.2, 10E.1, 15B14, and 12C.2 (see Experimental Procedures). Vertical arrows show chromosomal breakpoints encountered during the walk (schematically represented in Figure 2). The procedure for detecting the location of sequences linked to RAG-1 in its germline configuration depended on isolating probes from the genomic clones that would detect non-germline-sized restriction fragments in all genomic transfectants (as did probes F and G; see text). The necessary condition for generation of a non-germline-sized restriction fragment is that the restriction enzyme sites span a chromosomal breakpoint. Genomic clones 9F.2, 10E.1, and 15B14 were found to contain a single copy of the marker plasmid pSV2-His (dashed line), which, like the oligonucleotide, became closely linked to RAG-1 by cointegration during the primary transfection (see Figure 2).

on the host strain MB408 (kindly provided by W. Dove), which contains mutations in several bacterial recombination enzymes (Nader et al., 1986). We subsequently demonstrated that R7.5 and R2.2 each contain a tandem array of 15–30 copies of the oligonucleotide (data not shown); we have been unable to propagate any bacteriophage or plasmid containing these arrays in any strain other than MB408 (in contrast, the DNA sequences from fragment R7.5 that flank this array could be propagated in any bacterial strain tested).

To determine which end of R7.5 lay closest to RAG-1, probes were derived from the left (A_L) and right (A_R) ends of R7.5 (see Figures 2 and 5) and hybridized to Southern blots containing DNA from the secondary transfectants. Since the transfected RAG-1 locus should be integrated at a different chromosomal site in each independent secondary transfectant (depicted schematically in Figure 2, bottom), we predicted that, with the RAG-1–distal probe and appropriate restriction enzyme digest, the oligonucleotide-positive secondary transfectants would show different hybridization patterns (e.g., as shown for restriction enzyme BglII in Figure 2), whereas with the RAG-1–proximal probe, these transfectants would have identical patterns of hybridization (i.e., these transfectants should be identical in the direction of RAG-1; see Figure 2). In agreement with our predictions, the pattern of bands from these transfectants was identical with probe A_L and different with probe A_R (data not shown), suggesting that RAG-1 lay to the left of the oligonucleotide in R7.5.

Using probe A_L and subsequent probes B, C, D, E, and F, we obtained six additional, overlapping genomic clones that provided more than 50 kb of new sequence to the

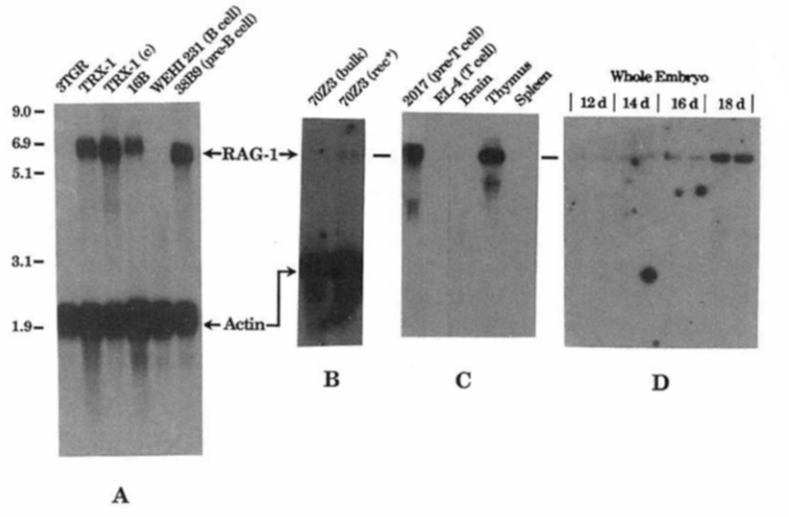


Figure 6. RAG-1 Is Expressed Only in Immature B and T Cells and the Genomic Transfectants (Northern Blot Analysis)

(A) Probe G from genomic clone 12C.2 and a probe specific for actin were hybridized to 4 μg of total cellular poly(A)$^+$ RNA from 3TGR cells, the genomic transfectants TRX-1 and 16B, the mature B cell line WEHI 231, and the pre-B cell line 38B9 (16 hr exposure; lane TRX-1 [c] contains 4 μg of cytoplasmic poly[A]$^+$ RNA). Previous hybridization of this Northern blot to probe G demonstrated that the only RNA species detected, even after prolonged exposure, was the approximately 6.6–7.0 kb mRNA indicated as RAG-1.

(B) Probe G and an actin probe were hybridized to 4 μg of poly(A)$^+$ RNA from 70Z/3 cells, or from 70Z/3 cells that had been previously selected for V(D)J recombinase activity with the recombination substrate pDHR and the drug hygromycin-B (Schatz and Baltimore, 1988); 11 day exposure.

(C) Probe G was hybridized to 30 μg of total RNA from the pre-T cell line 2017, the mature T cell line EL-4, and the brain, thymus, and spleen (2017, 40 hr exposure; EL-4, 96 hr exposure; brain, thymus, and spleen, 18 hr exposure). Prolonged exposure did not reveal any additional hybridization. Hybridization with actin (2017 and EL-4) or tubulin (brain, thymus, and spleen; A. Bernards, personal communication) probes indicated approximately equal amounts of RNA in the lanes (2017 and EL-4 not directly comparable to brain, thymus, and spleen).

(D) Probe G was hybridized to poly(A)$^+$ RNA from mouse embryos at 12, 14, 16, and 18 days of gestation (14 day exposure). Hybridization with a tubulin probe indicated approximately equal amounts of RNA in the lanes (G. Mardon, personal communication).

left of R7.5 (Figure 5). We investigated each successive genomic clone to discover whether it contained sequences contiguous to RAG-1 in its germline configuration (represented by the thin line in Figure 2), or whether it contained sequences that only became linked to RAG-1 as a result of ligation or cointegration during the process of transfection (represented as heavy lines and shaded boxes in Figure 2). To answer this question, probes from each genomic clone were hybridized to a series of Southern blots containing DNA from 3TGR cells and a panel of genomic transfectants, which in addition to those described above included several independent primary transfectants derived by transfer of mouse genomic DNA into 3TGR cells (P transfectants). Each P transfectant should contain a different site of integration of germline RAG-1 sequences from that in TRX-1 and all of its derivatives. We reasoned that a probe derived from germline RAG-1 sequences (thin line in Figure 2) would, with the appropriate restriction enzyme digest (see legend to Figure 5 for details), detect both a germline band and a transfectant-specific band in TRX-1, its derivatives, and the P transfectants. In contrast, probes derived from outside this region (heavy lines and shaded boxes in Figure 2) would detect two bands in some or all of the TRX-1 derivatives, but only one band (germline) in the P transfectants. Probes A_L, B, C, D, and E yielded this latter pattern, while probes F and G showed novel, nongermline bands in TRX-1 and all P transfectants tested (data not shown). This implied that clones 15B14 and 12C.2 contained sequences linked to RAG-1 in its germline configuration and potentially contained some or all of RAG-1 itself. This procedure also demonstrated that the transfected RAG-1 locus was amplified 2- to 5-fold in TRX-1, and most lines derived by transfection of TRX-1 genomic DNA (data not shown). In addition, four chromosomal breakpoints (vertical arrows in Figure 5; schematically represented in Figure 2) were detected during the genomic walk, indicating that four normally unlinked fragments of DNA had become integrated between the oligonucleotide in R7.5 and the RAG-1 locus during the primary transfection. A single copy of the pSV2-His marker plasmid was found integrated into one of these breakpoints (Figure 5).

Detection of the RAG-1 mRNA and Isolation of RAG-1 cDNAs

As an initial screen for RAG-1 sequences in genomic clones 15B14 and 12C.2, probes F and G (Figure 5) were hybridized to filters containing poly(A)$^+$ RNA from V(D)J recombinase–positive (rec$^+$) and –negative (rec$^-$) cell lines. While probe F gave no detectable hybridization (data not shown), probe G hybridized to a single RNA species, 6.6–7.0 kb in length, present in a rec$^+$ pre-B cell (38B9) and in two genomic transfectants (TRX-1 and 16B), but not in a rec$^-$ mature B cell (WEHI231) or 3TGR (Figure 6A). Based on this, we tentatively concluded that this RNA represented the mRNA for RAG-1 and that genomic clone 12C.2 contained a portion of RAG-1.

Probe G was used to screen a mouse pre-B cell (22D6) cDNA library (size selected; 40 positives from 10^6 recombinants) and a human pre-B cell (Nalm 6) library (generously provided by D. Kaul and A. Bernards; 11 positives from 2×10^6 recombinants). Of ten mouse cDNA clones analyzed, the two largest, M6 and M2, contained 6.9 kb and 5.8 kb inserts, respectively, while the largest human cDNA clones (H7 and H36) contained 6.6 kb inserts (Figure 7).

RAG-1 cDNA and Genomic Clones Are Biologically Active

To confirm the identity of the genomic and cDNA clones,

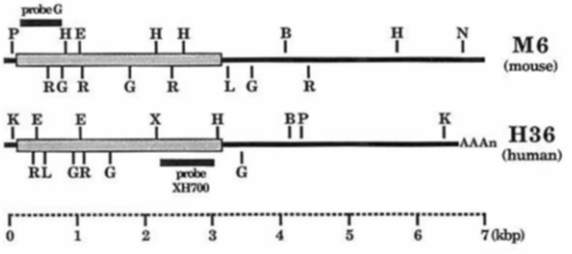


Figure 7. Schematic Diagram of Mouse and Human RAG-1 cDNA Clones

Indicated are the single long open reading frame in each cDNA (shaded box), the locations of probes G and XH700, and all sites for the following restriction enzymes: B, BclI; E, EcoRV; G, BglII; H, HindIII; K, KpnI; L, ClaI; N, NotI, P, HpaI; R, EcoRI; X, XhoI. Probe G, originally derived from genomic clone 12C.2, was exactly colinear with the indicated region of the mouse cDNA, as determined by nucleotide sequencing.

it was essential to demonstrate that they could activate the V(D)J recombinase when transfected into 3TGR cells, thereby recapitulating the results obtained with genomic DNA. When DNA from phage clone 12C.2 was transfected into 3TGR cells, MPA-resistant colonies were obtained that had carried out normal V-to-J recombination events (Figure 3). Similar results were obtained when human cDNAs H7 and H36 and mouse cDNA M2, subcloned into mammalian expression vectors, were transfected into 3TGR cells (Figure 3). In contrast, no colonies resulted from control transfections of the expression vectors themselves. This proved that the genomic clone 12C.2, the human cDNA clones H7 and H36, and the mouse cDNA clone M2 were sufficient to activate the V(D)J recombinase and thus that they represented biologically active clones of RAG-1.

In all cases, however, cloned DNA was no more active in this assay (number of MPA-resistant colonies) than was an efficient genomic DNA (e.g., from TRX-1). Since 0.1% to 1.0% of cells that took up DNA typically went on to express V(D)J recombinase activity (data not shown), the cloned DNAs were roughly 100- to 1000-fold less active than the theoretical maximum. Partial sequencing of genomic clone 12C.2 confirmed that it was mouse in origin and demonstrated that it lacked the 5′ end of the gene, including the first 10^6 coding nucleotides; its low activity can thus be explained by its need to acquire a promoter (and potentially an in-frame start codon) through fortuitous integration into the host chromosome. We are currently investigating factors that influence expression of RAG-1 from the cDNAs.

Pattern of Expression of RAG-1

Using Northern blot analysis, we extended our investigation of the pattern of RAG-1 expression to a number of other cell lines and tissues (representative data shown in Figure 6 and summarized in Table 1). Comparison of this data with the known pattern of expression of V(D)J recombinase activity (Lieber et al., 1987) revealed a striking correlation between RAG-1 expression and recombinase activity. The RAG-1 mRNA could be detected only in pre-B and pre-T cell lines and not in more mature cells of these

Table 1. Pattern of Expression of RAG-1

Cell Lines			
Positive		**Negative**	
Pre-B:	22D6	Pre-B/B:	70Z/3 (bulk)
	38B9	Mature B:	WEHI 231
	1-8		Namalwa
	2M3		BJAB
	Nalm 6	Mature T:	EL-4
scid pre-B:	8D		RLM-11
	A2		Jurkat
	2A	Erythroid:	MEL
		Macrophage:	P388D1
Pre-B/B:	70Z/3 (rec^+)		
Pre-T:	2017	Pheochromocytoma:	PC12
	2052D	Fibroblast:	NIH 3T3
			BALB/c 3T3
Genomic transfectants:	TRX-1		L Cell
	16B		
	L-1	Cervical carcinoma:	HeLa
	L-4		
	M-3		
Tissues			
Positive		**Negative**	
Thymus		Spleen	
Bone marrow culture		Brain	
Whole embryo (12–18 day)			

Expression of RAG-1 was assayed by Northern blot analysis using probe G (see Experimental Procedures). A description of many of the cell lines is provided by Alt et al. (1981), Alt et al. (1984), and Kabat et al. (1987). The *scid* pre-B cell lines are described in Hendrickson et al. (1988), the pre-T lines in Risser et al. (1985), and P388D1 in Lachman et al. (1977).

lineages (Figures 6A and 6C) or in any nonlymphoid cell line tested. Furthermore, RAG-1 mRNA was found in RNA from the thymus, but not in RNA from the brain or spleen (Figure 6C), and it could be detected in RNA from a long-term bone marrow culture (kindly provided by K. Denis). Thus, RAG-1 appeared to be expressed only in cells and tissues that carry out V(D)J recombination.

A further strong correlation between RAG-1 expression and V(D)J recombinase activity came from the late pre-B cell line 70Z/3. We reported previously that in a population of 70Z/3 cells containing a rearrangement substrate similar to DGR, rare rec$^+$ cells could be isolated that had carried out a V-to-J recombination event (Schatz and Baltimore, 1988). While RAG-1 mRNA could not be detected in poly(A)$^+$ RNA from the bulk population of 70Z/3 cells, it could be detected in poly(A)$^+$ RNA from rec$^+$ 70Z/3 cells (Figure 6B).

Finally, the patterns of expression of RAG-1 and of the V(D)J recombinase during mouse embryonic development appear to be similar. RAG-1 mRNA could be detected in poly(A)$^+$ RNA from embryos of all stages of development examined, increasing from day 12 to day 18 of gestation (Figure 6D). This pattern is consistent with the observation that pre-B cells are first evident in fetal liver at day 11–12 of gestation (Andrew and Owen, 1978; Raff et al., 1976) and that lymphoid cells increase in number in liver and thymus thereafter.

Sequence of the RAG-1 cDNAs

We determined the nucleotide sequence of the human cDNA clone H36 (6545 bp, followed by a poly[A] tail) and of the mouse cDNA clone M6 (approximately 7000 nucleotides; no poly[A] tail, two small noncoding regions not sequenced). Both sequences contain a single long open reading frame, which could encode a putative RAG-1 polypeptide of 1043 (human) or 1040 (mouse) amino acids, with a predicted molecular mass of 119 kd (the coding nucleotide sequences and deduced amino acid sequences are shown in Figure 8). The nucleotides surrounding the ATG initiator codon at the beginning of each long open reading frame have most of the features of the optimal sequence for eukaryotic translation initiation (ACCATGG), including the highly conserved A at position −3, the C at position −1, and the G at position +4 (Kozak, 1984, 1986). This ATG triplet is the first found in either sequence and in each case is preceded by an in-frame termination codon. Flanking the long open reading frame in the human and mouse clones are 5′ regions of 112 and 94 nucleotides and AT-rich 3′ regions of 3.4 and 3.8 kb, respectively (Figure 7). The mouse and human 3′ untranslated regions contain 12 and 13 copies, respectively, of the motif AUUUA, which is commonly found in the 3′ untranslated regions of rapid-turnover mRNAs and which may contribute to mRNA instability (Caput et al., 1986; Shaw and Kamen, 1986).

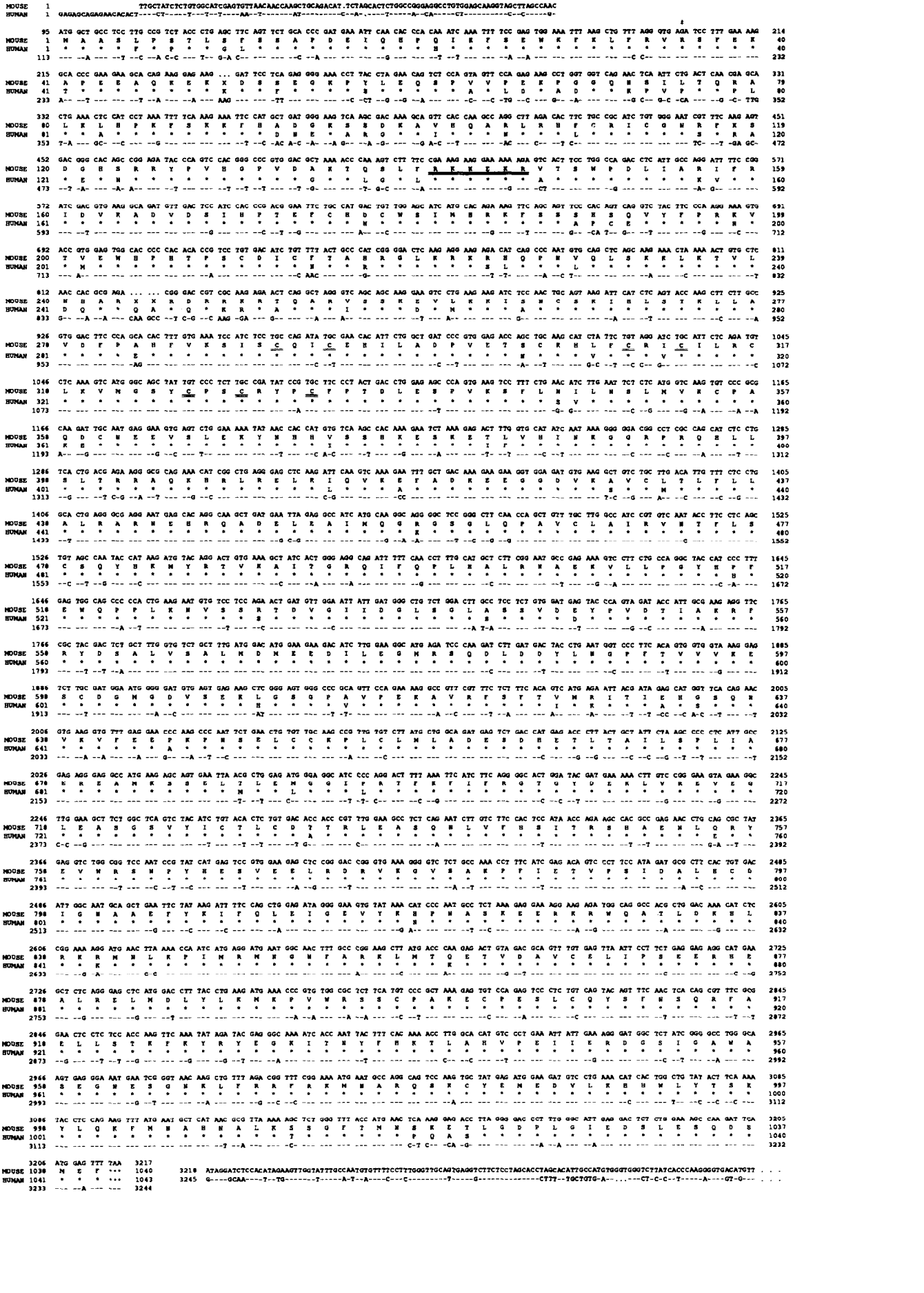
While the human and mouse nucleotide sequences were not similar to any sequence in the GenBank data base, they are extremely similar to each other: in the putative coding region they are 86% identical, while in the flanking 5′ and 3′ regions this decreases somewhat to 79% and 68%, respectively. The human and mouse deduced amino acid sequences are 90% identical, and if conservative amino acid replacements are considered, the similarity rises to 95%. The putative proteins contain a high percentage of charged residues (approximately 32%) and contain no lengthy hydrophobic regions. Each sequence contains the positively charged motif, Arg-(140–141)-Lys-Lys-Glu-Lys-Arg, which is similar to the sequences that direct nuclear localization in a number of other proteins (Lanford et al., 1988; for review see Gerace and Burke, 1988). Finally, when the National Biomedical Research Foundation protein sequence library was searched with the deduced amino acid sequences, the only suggestive homology discovered was to the DNA binding domain of the human, rat, and mouse glucocorticoid receptors (for review see Beato, 1989). When gaps were introduced to maximize alignment, these regions (mouse RAG-1 residues 283–332, human alpha glucocorticoid receptor residues 432–481) are almost 37% identical, including identities at all seven of the cysteine residues in this region of the human glucocorticoid receptor; however, the spacing between cysteine residues is not conserved between the two proteins. This degree of relatedness is statistically significant (Needleman-Wunsch similarity value greater than 6 standard deviation units above the mean of 100 shuffled sequences; Dayhoff et al., 1978; Doolittle, 1981) and suggests that RAG-1 may encode a metal binding protein.

RAG-1 Is Evolutionarily Conserved

The high degree of similarity of the human and mouse RAG-1 nucleotide sequences led us to examine if RAG-1 might be conserved in other species known to carry out V(D)J recombination. Probe XH700, derived from the human cDNA (Figure 7), detected a single band in rabbit, dog, goat, and horse DNA of roughly equal intensity to the band detected in mouse DNA (Figure 9A) and, at a somewhat reduced stringency, also detected a small number of bands in chicken and frog DNA (Figure 9B). We conclude that RAG-1 coding sequences are conserved in a wide range of species that carry out V(D)J recombination.

Discussion

We describe in this report the isolation and initial characterization of a novel participant in the lymphoid-specific activation of V(D)J recombination, the recombination activating gene-1 (RAG-1). Genomic sequences containing RAG-1 were isolated by virtue of their ability to activate V(D)J recombination when introduced into NIH 3T3 fibroblasts. An oligonucleotide marker allowed us to identify the genomic locus containing RAG-1 by virtue of the cosegregation of V(D)J recombinase activity and arrays of the oligonucleotide. The oligonucleotide subsequently served as an initial hybridization probe in a genomic walk that ended within RAG-1 sequences. RAG-1 genomic sequences were used to isolate human and mouse RAG-1 cDNAs; proof of the identity of the genomic and cDNA clones came with the demonstration that each was sufficient to activate expression of the V(D)J recombinase.



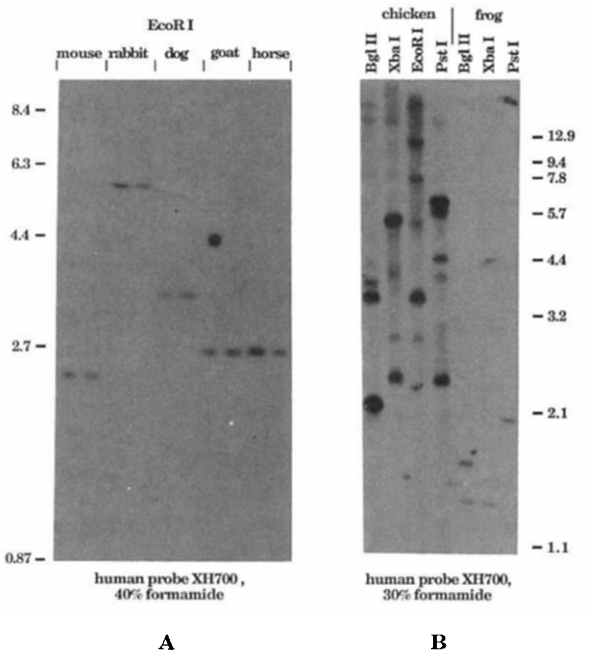


Figure 9. RAG-1 Is Conserved through Evolution (Southern Blot Analysis)

(A) Hybridization with radioactive human probe XH700 was in 40% formamide, 4× SSPE (1× SSPE = 0.15 M NaCl, 10 mM NaH_2PO_4, 1 mM EDTA [pH 7.4]), 2.5× Denhardt's solution, 250 μg/ml sonicated salmon sperm DNA, 1% SDS, at 42°C. Final wash conditions were 2× SSC, 0.1% SDS, 68°C. Marker sizes are shown in kb. Southern blot kindly provided by R. Mosher and D. Page; exposure = 60 hr.

(B) Genomic DNA prepared either from chicken liver or Xenopus oocytes was digested with the indicated restriction enzymes, electrophoresed on a 0.8% agarose gel, and transferred to a Zetabind membrane. Hybridization was carried out as in (A), but with the formamide concentration being 30%. Final wash conditions were 3× SSC, 0.1% SDS, 50°C. Marker sizes are shown in kb; exposure = 20 hr.

RAG-1 mRNA could not be detected in nonlymphoid cell lines or tissues, but was detected only in cell lines representing the early stages of B and T cell development and in thymus and bone marrow. Thus, tight regulation of its expression would appear to confine RAG-1 activity to cell lines and tissues that express V(D)J recombinase activity. Nucleotide sequencing revealed that the mouse and human RAG-1 cDNAs are extremely similar, each with a long open reading frame in the 5′ half that could encode a 119 kd polypeptide. We discuss below the potential roles of this putative protein in the process of lymphoid development, and also comment on the unusual aspects of the means by which the gene was isolated.

Isolation of RAG-1 by Genomic Transfection

Our results arising from transfection of exogenously tagged genomic DNA are comparable to those of others who have isolated genes by this means, with the notable exception that the oligonucleotide marker was found to lie more than 40 kb from our target gene and to be separated from it by several noncontiguous genomic fragments (see Figures 2 and 5). Ligation of plasmid or tRNA markers to restriction enzyme–digested genomic DNA, followed by transfection, has yielded direct juxtapostion of the marker and the germline fragment carrying the target gene (Goldfarb et al., 1982; Lowy et al., 1980; Perucho et al., 1980). The most likely explanation for this difference lies in the size of the genomic restriction fragment that contains the target gene and is available for ligation to the marker. While in the examples cited above the target genes were ligated to the marker as small (less than 8 kb) HindIII fragments, RAG-1 from TRX-1 was contained on a XhoI–SalI fragment of at least 50 kb, and perhaps much larger (data not shown). If the RAG-1–containing XhoI–SalI fragment were larger than 100–150 kb, the restriction sites would almost certainly have been sheared away from the RAG-1 gene during preparation of the DNA, making subsequent ligation to the oligonucleotide unlikely. Successful tagging of RAG-1 would then require cointegration of the oligonucleotide and RAG-1, as depicted in Figure 2, generating an unpredictable amount of intervening DNA.

Use of an oligonucleotide as a marker has the advantages that it can be used in vast molar excess in the ligation to genomic DNA and can be engineered to contain convenient cohesive ends and restriction sites. This procedure, however, has one potential disadvantage: if multimerization of the oligonucleotide is allowed during the ligation, large tandem arrays can be generated that are subsequently difficult to propagate in bacteria, making isolation of sequences flanking the arrays arduous. The bacterial strain MB408, one of a series of strains designed to tolerate DNA repeat systems and function efficiently with the phage vector λgt10 (Nader et al., 1986), allowed cloning of two oligonucleotide arrays (R7.5 and R2.2; data not shown for R2.2) that could not be propagated in a number of other strains.

Characterization of human RAG-1 cDNA clones provided a simple explanation for our failure to isolate human RAG-1 genomic sequences by genomic transfection: all

Figure 8. Nucleotide Sequences of Mouse and Human RAG-1 cDNAs and Predicted Amino Acid Sequences of the Encoded Proteins

The nucleotide sequences of the 5′ half of the mouse cDNA clone M6 and the human cDNA clone H36 are shown in the top and bottom rows of each line (identical residues indicated by a dash in the human sequence). The mouse and human sequences shown include the 5′ untranslated regions (94 and 112 nucleotides, respectively), the entire coding regions, and a small portion of the 3′ untranslated regions. Nucleotides are numbered beginning with the first nucleotide of the cDNA. The predicted mouse and human RAG-1 amino acid sequences (1040 and 1043 amino acids, respectively) are given between the nucleotide sequences (identities indicated by asterisks in the human sequence). Two gaps have been introduced into the mouse coding sequence (nucleotide positions 242 and 824) to maximize alignment. The putative nuclear localization signal and the seven cysteine residues in the region with homology to the glucocorticoid receptor are underlined. The first nucleotide of the genomic clone 12C.2 (nucleotide 201) is indicated by ‡; probe G is colinear with the mouse cDNA and extends from nucleotide 201 to the BglII site at position 879 (see Figure 7). While the human cDNA clone H36 has been shown to be biologically active (see text and Figure 3), the mouse cDNA clone M6 has not.

human cDNA clones examined contained a XhoI restriction site in the putative coding region of the gene (while the mouse cDNAs do not), thus making transfer of an active human gene using XhoI-digested genomic DNA extremely unlikely.

RAG-1: Recombinase or Developmental Switch?

We propose two alternative models for the mechanism of action of the factor encoded by RAG-1: it might participate directly in enzymatic steps of V(D)J recombination, or it might function as a regulatory molecule in a pathway leading to the activation of the recombinase. If the former model is correct, then the factor encoded by RAG-1 will represent the first component of the V(D)J recombinase to be isolated. Analysis of RAG-1 should then provide insights into the molecular mechanism of V(D)J recombination and its regulation during lymphoid development, and expression of RAG-1 might allow development of an in vitro V(D)J recombination system. If the latter model is correct, then RAG-1 should play an important role in the early development of B and T cells, and study of this gene might provide direct insight into the regulation of expression of V(D)J recombinase activity.

Some properties of the putative RAG-1 polypeptide are equally compatible with either model. It is a large polypeptide with a predicted unmodified molecular mass of 119 kd, and contains the potential nuclear localization signal Arg-Lys-Lys-Glu-Lys-Arg and a cysteine-rich domain with similarity to the zinc finger DNA binding domain of glucocorticoid receptors. Thus RAG-1 may encode a nuclear protein with metal and DNA binding activities.

Other properties of RAG-1 provide weak and circumstantial evidence in favor of the recombinase model. The high degree of conservation of RSSs through evolution and the functional equivalence of RSSs from a variety of species and recombining loci have together supplied strong support for the notion of a single, evolutionarily conserved V(D)J recombinase. Consistent with this, RAG-1 appears to be conserved between a number of species known to carry out V(D)J recombination, including rabbit, dog, goat, horse, chicken, and frog (Figure 9). In addition, RAG-1 expression correlates precisely with V(D)J recombinase activity, as examined in cell lines, tissues, the developing embryo, and the unusual, RAG-1–"negative" pre-B cell line 70Z/3, where the rare cells expressing detectable recombinase activity also express RAG-1 (Figure 6 and Table 1).

Other results argue weakly against the model of RAG-1 as a regulator of a developmental pathway leading to the expression of V(D)J recombinase activity. In addition to activating the recombinase, such a regulator might have more pleiotropic effects, and NIH 3T3 fibroblasts expressing RAG-1 might then exhibit other features characteristic of early lymphocytes. We have been unable to detect such lymphoid-specific properties: transfectants stably expressing RAG-1 have the morphology and growth requirements typical of adherent, fibroblastoid cells (data not shown); they do not express the B lineage surface marker B220 (Coffman and Weissman, 1981; data not shown), they do not express detectable levels of TdT mRNA (S. Smale, personal communication), nor do they contain detectable amounts of RNA or protein corresponding to the B lymphocyte–specific octamer binding factor NF-A2 (Staudt et al., 1988; L. Corcoran and L. Staudt, personal communication); finally, they fail to rearrange their endogenous immunoglobulin μ or κ loci, as assayed by a highly sensitive polymerase chain reaction assay (M. Schlissel, personal communication), and also do not rearrange their endogenous T cell receptor β locus, as assayed by Southern blotting (data not shown). Thus, RAG-1 does not appear to be functionally analogous to the MyoD1 or *myd* genes, which in triggering the myogenic conversion of C3H 10T1/2 cells stimulate the coordinate activation of many muscle-specific genes and induce a series of gross morphological alterations (Lassar et al., 1986; Pinney et al., 1988). If RAG-1 were to function in a developmental pathway leading to the expression of the V(D)J recombinase, it would be predicted to occupy a position in the pathway just "upstream" of the recombinase in both the B and T cell lineages. In addition, with the exception of the weak homology to the DNA binding domain of the glucocorticoid receptor, the putative RAG-1 protein contains no significant sequence similarity to any transcription factor for which sequence data is available (for review see Mitchell and Tjian, 1989).

In sum, while neither the recombinase model nor the developmental regulator model is contradicted by the available data, the indirect evidence leads us to favor the hypothesis that RAG-1 encodes some or all of the V(D)J recombinase (where "V(D)J recombinase" refers to the essential, lymphoid-specific component(s) of the enzymatic machinery of recombination).

RAG-1 and the Enzymology of V(D)J Recombination

Detailed mechanistic studies of prokaryotic site-specific recombination systems have revealed that a number of recombinases recruit accessory host factors as obligatory or stimulatory participants in the recombination reaction (Weisberg and Landy, 1983; Huber et al., 1985; Johnson and Simon, 1985; Kahmann et al., 1985). This may also be true for the V(D)J recombinase, which carries out a highly complex site-specific recombination reaction. The reaction requires coordinate recognition of RSSs that may be hundreds of kilobase pairs apart, cleavage between the RSSs and their flanking gene segments, and ligation of the cleaved ends. Before the ends of the coding segments are joined, they may be subjected to exonucleolytic digestion and/or addition of new residues. A number of factors with various binding (Aguilera et al., 1987; Halligan and Desiderio, 1987) or endonucleolytic (Desiderio and Baltimore, 1984; Kataoka et al., 1984; Hope et al., 1986) activities have been detected in nuclear extracts from lymphoid cells, but their relationship to the V(D)J recombinase, or to RAG-1, remains obscure. The V(D)J recombinase appears to recruit the enzyme TdT as a nonessential accessory factor, where it is likely to function in the addition of N regions to the coding junction.

V(D)J recombination may also involve the product of the gene that is mutant in the *scid* mouse. In the *scid* mouse, a recessive mutation on chromosome 16 (Bosma et al.,

1989) causes a severe deficiency in mature B and T cells by disrupting normal V(D)J recombination (Schuler et al., 1986). The V(D)J recombinase of *scid* lymphocytes cannot form proper coding junctions (Lieber et al., 1988), but instead introduces large deletions (Hendrickson et al., 1988; Schuler et al., 1986). These results have prompted the hypothesis that the mutant *scid* gene encodes a central component of the recombinase. It was thus particularly important to determine the relationship between RAG-1 and the *scid* gene. Hybridization of a RAG-1 probe to Southern blots containing DNA from a variety of somatic cell hybrids (kindly provided by P. D'Eustachio) demonstrated that RAG-1 does not map to mouse chromosome 16 (data not shown), establishing that RAG-1 and the *scid* gene are different genes. However, *scid* myeloid and fibroblastoid cells may have an increased sensitivity to radiation-induced damage (Phillips and Fulop, 1989), implying that the *scid* defect may not be confined to lymphoid cells. The simplest interpretation of this result is that the *scid* gene does not encode the V(D)J recombinase, but rather encodes a ubiquitously expressed factor involved both in V(D)J recombination and in the repair of chromosomal damage.

The prokaryotic recombination systems cited above are examples of recombinases with multiple enzymatic activities that also recruit accessory factors in order to carry out efficient recombination. While V(D)J recombination is a complex process requiring at least three distinct enzymatic activities—RSS recognition, endonucleolytic cleavage, and ligation—there is ample precedent for the supposition that RAG-1 also encodes such a recombinase.

Experimental Procedures

Genomic Transfection

The recombination substrate pDGR and cell lines 3TGR and TRX-1 have been described previously (Schatz and Baltimore, 1988). Our previous procedure for genomic transfection (Schatz and Baltimore, 1988) was modified to allow for efficient transfection of unsheared, high molecular weight genomic DNA. Genomic DNA (80 μg) and linearized pSV2-His plasmid (2.5 μg) were combined in a final volume of 0.5 ml water, and an equal volume of 0.5 M $CaCl_2$ was added. The DNA–Ca mixture was then immediately added in a dropwise fashion to 1 ml of HNP buffer with continuous gentle vortexing. HNP buffer (10 ml) was prepared just before use from 9.8 ml of 50 mM HEPES (pH 7.10), 280 mM NaCl, and 0.2 ml of 70 mM Na_2HPO_4 (pH 7.10). The precipitate was allowed to form for 3–8 min and was then added to two 10 cm dishes of 3TGR cells that had previously been refed with 5 ml of medium (DMEM supplemented with 10% calf serum and penicillin–streptomycin sulfate). L-histidinol and MPA drug selections were carried out as described previously (Schatz and Baltimore, 1988), with the exception that MPA selection was carried out in the presence of 500 μM L-histidinol in medium lacking histidine. All transfectants were routinely propogated under continuous MPA–histidinol double selection.

Ligation of Oligonucleotide to Genomic DNA

Two 30 base oligonucleotides (T1: 5′-TCGACATCGCGAACGCATCGTGTTCTAGAG-3′ and T2: 5′-TCGACTCTAGAACACGATGCGTTCGCGATG-3′), were synthesized, annealed, and phosphorylated with polynucleotide kinase. TRX-1 genomic DNA (80 μg), digested to completion with XhoI and SalI, was ligated to 12.5 μg of the annealed, phosphorylated oligonucleotide in a 250 μl volume with 1600 U of T4 DNA ligase (New England Biolabs) for 13 hr at 15°C.

Oligonucleotide Hybridization in Dried Agarose Gels

Our procedure for hybridization of an oligonucleotide to DNA in a partially dried agarose gel has been reported in detail elsewhere (Schatz, 1989). In brief, restriction enzyme–digested genomic DNA was electrophoresed in a 0.8% agarose gel, and the gel was soaked for 45 min in denaturation solution (1.5 M NaCl, 0.5 M NaOH) and then for 30 min each in neutralization solution (1.5 M NaCl, 0.5 M Tris–HCl [pH 7.5]) and water. The gel was partially dried on a conventional gel dryer and prehybridized and hybridized in 1× SSC (0.15 M NaCl, 15 mM sodium citrate [pH 7.4]), 20 mM NaH_2PO_4, 5× Denhardt's solution [1× Denhardt's = 0.02% Ficoll 400, 0.02% polyvinyl pyrrolidone, 0.02% bovine serum albumin]), 250 μg/ml sonicated salmon sperm DNA, 0.1% SDS at 56°C. Single-stranded oligonucleotide T1, phosphorylated to high specific activity with polynucleotide kinase (Boehringer Mannheim) and [γ-^{32}P]ATP, was used in the hybridization at 5 × 10^6 cpm/ml. The gel was washed first in 6× SSC at room temperature and then increasingly stringently, with final wash conditions being 0.5× SSC, 0.1% SDS at 56°C. The gel was again partially dried and exposed at −80°C to X-ray film backed by an intensifying screen.

Southern and Northern Gel Transfer Hybridization

Probes were labeled with [α-^{32}P]dCTP using the hexamer labeling method (Feinberg and Vogelstein, 1983). Southern blotting was carried out as described previously (Schatz and Baltimore, 1988). Probe 1 was a 1.1 kb NruI–PvuII fragment of pBR322. Probe G was a 678 bp fragment from the left end of genomic clone 12C.2 (colinear with mouse cDNA nucleotides 201 to 879; Figure 8). Probe XH700 was a 908 bp XhoI–HindIII fragment corresponding to human cDNA nucleotides 2273 to 3181.

RNA samples were electrophoresed in 2.2 M formaldehyde–1% agarose gels (Maniatis et al., 1982) and transferred to nitrocellulose or Hybond-C extra (Amersham) membranes. RNA loading was evaluated by hybridization with a γ-actin probe derived as a BamHI–HindIII fragment from plasmid pSP6 γ-actin (Enoch et al., 1986).

All blots were exposed at −80°C to X-ray film backed by an intensifying screen.

Genomic and cDNA Libraries

Genomic clones were obtained from five different libraries: clone R7.5 from EcoRI-digested M-1 genomic DNA in a 6.0–9.0 kb size-selected λgt10 (Stratagene) library; clone 8B12 from BamHI-digested M-1 genomic DNA in a 10–14 kb size-selected λEMBL-3 (Promega) library; clones 9U.3, 9F.2, and 10E.1 from partially Sau3a-digested L-4 genomic DNA in a λEMBL-3 library; clone 15B14 from BamHI-digested L-4 genomic DNA in a 12–20 kb size-selected λFix (Stratagene) library; and clone 12C.2 from partially Sau3a-digested L-4 genomic DNA in a λFix library. Libraries were constructed according to the protocols provided by the suppliers of the bacteriophage arms and were propagated in standard bacterial strains, with the exception of the λgt10 library that yielded clone R7.5, which was grown in bacterial strain MB408 (*recF*$^-$, *recB*21, *recC*22, *sbcB*15, *hflA*, *hflB*, *hsdR*$^-$, and tet^r; Nader et al., 1986).

A cDNA library was prepared from poly(A)$^+$ RNA from the Abelson murine leukemia virus–transformed cell line 22D6 (Alt et al., 1981) by the method of Gubler and Hoffman (1983), except that first- and second-strand reactions were carried out in a single tube, without intervening extractions or precipitations (Sartoris et al., 1987). cDNA was ligated to EcoRI–NotI adaptors (Invitrogen) and size fractionated by agarose gel electrophoresis. After glass powder purification (Vogelstein and Gillespie, 1979), the 2.0–8.0 kb size fraction was ligated to λgt10. Probe G was used to screen 1 × 10^6 unamplified recombinants, and 10 of the 40 positives were analyzed further. The two largest cDNA inserts, M6 (7.0 kb) and M2 (5.8 kb), were excised with NotI and subcloned into the NotI site of pBSK$^+$ (Stratagene) and into the eukaryotic expression vector CDM8 using BstXI adaptors (Seed, 1987).

Recombinants (2 × 10^6) from a human pre-B cell (Nalm 6) cDNA library, constructed in the bacteriophage vector λZAP (Stratagene) by D. Kaul and A. Bernards, were screened with probe G to yield 11 positive plaques. pBSK$^-$ plasmids containing the three largest cDNA inserts, H7 and H36 (6.6 kb) and H34 (6.0 kb), were isolated directly from bacteria containing λZAP bacteriophage using helper phage R408, following the protocol supplied by Stratagene for in vivo excision. The H7

and H36 cDNA inserts were excised as BamHI–ClaI fragments and subcloned into the BamHI–ClaI sites of the polylinker of the eukaryotic expression vector pJ3Ω, provided by Jay Morgenstern (Whitehead Institute, MIT), a vector that directs expression under the control of the SV40 early-region promoter.

Ten microgram samples of DNA from genomic clone 12C.2, human cDNA clones H7 and H36 in pJ3Ω, or mouse cDNA clone M2 in pCDM8 were separately cotransfected with 2 μg of linearized pSV2-His into 3TGR cells using the genomic transfection protocol outlined above.

Nucleotide Sequence Analysis

The Erase-a-Base (exonuclease III) system of Promega was used to generate nested deletions of both strands of mouse cDNA clone M6 and of one strand each of human cDNA clones H36 and H34 (which lay in opposite orientation in the pBSK$^-$ vector). Deleted templates were transformed into JM109 (Stratagene), and single-stranded DNA was isolated using helper phage VCS-M13 (Stratagene). Single-stranded templates were sequenced by dideoxy chain termination (Sanger et al., 1977) using Sequenase DNA polymerase and the protocol supplied by the manufacturer (United States Biochemical Corp.). The complete nucleotide sequence of both strands was determined in the 5′ half of each cDNA, which included the region containing the long open reading frame. The complete sequence of the human 3′ region was determined, although in several small regions sequence was obtained from only one strand. The mouse 3′ region was entirely sequenced except for two small gaps, approximately 200–250 nucleotides in total length. Nucleotide and predicted amino acid sequences have been submitted to GenBank and Protein Identification Resource/NBRF, respectively.

Using the FASTA algorithm (Pearson and Lipman, 1988), the human and mouse RAG-1 nucleotide sequences were compared with the GenBank nucleic acid data base, and the predicted amino acid sequences were compared with the National Biomedical Research Foundation protein sequence data base. In addition, the mouse amino acid sequence was compared with the potential translation products of the GenBank nucleic acid data base, using a modification the FASTA algorithm. The degree of statistical significance of the similarities detected was investigated using the random shuffling algorithm (Dayhoff et al., 1978; Doolittle, 1981) of the program IALIGN (Dayhoff et al., 1983).

Acknowledgments

We are deeply indebted to Carolyn Gorka for her superb technical assistance with many of the aspects of the work described here, and particularly for the construction of the cDNA library from 22D6. We thank P. D'Eustachio for somatic cell hybrid DNAs, W. Dove for the bacterial strain MB408, G. Mardon for the Northern blot containing embryo RNA, R. Mosher and D. Page for the Southern blot containing mammalian DNA, A. Winoto for Northern blots and RNA samples, L. Corcoran for RNA samples, D. Weaver for *scid* pre-B RNA samples, D. Kaul and A. Bernards for the Nalm 6 cDNA library, R. Risser for the 2017 and 2052D cell lines, W. Gilbert for assistance with the computer analysis of the RAG-1 sequences, and K. Struhl for many helpful discussions. We are also grateful to S. Smale, D. Black, D. Silver, M. Schlissel, and A. Winoto for critical reading of and many helpful comments on the manuscript. Finally, we gratefully acknowledge the important experimental contributions of the late Dr. Samuel Latt. This work was supported by a grant to D. B. from the American Cancer Society. D. G. S. was supported by fellowships from the Life and Health Insurance Medical Research Fund and the Whitaker Health Sciences Fund. M. A. O. was supported by NIH training grant CA09541-04 and fellowships from the Johnson and Johnson Research Fund and the Whitaker Health Sciences Fund.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 11, 1989.

References

Aguilera, R. J., Akira, S., Okazaki, K., and Sakano, H. (1987). A pre-B cell nuclear protein that specifically interacts with the immunoglobulin V-J recombination sequences. Cell *51*, 909–917.

Alt, F., Rosenberg, N., Lewis, S., Thomas, E., and Baltimore, D. (1981). Organization and reorganization of immunoglobulin genes in A-MuLV-transformed cells: rearrangement of heavy but not light chain genes. Cell *27*, 381–390.

Alt, F. W., Yancopoulos, G. D., Blackwell, T. K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., and Baltimore, D. (1984). Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO J. *3*, 1209–1219.

Andrew, T. A., and Owen, J. J. T. (1978). Studies on the earliest sites of B cell differentiation in the mouse embryo. Dev. Comp. Immunol. *2*, 339–346.

Beato, M. (1989). Gene regulation by steroid hormones. Cell *56*, 335–344.

Blackwell, T. K., and Alt, F. W. (1988). Immunoglobulin genes. In Molecular Immunology, B. D. Hames and D. M. Glover, eds. (Washington D.C.: IRL Press), pp. 1–60.

Blackwell, T. K., Moore, M. W., Yancopoulos, G. D., Suh, H., Lutzker, S., Selsing, E., and Alt, F. W. (1986). Recombination between immunoglobulin variable region gene segments is enhanced by transcription. Nature *324*, 585–589.

Bosma, G. C., Custer, R. P., and Bosma, M. J. (1983). A severe combined immunodeficiency mutation in the mouse. Nature *301*, 527–530.

Bosma, G. C., Davisson, M. T., Ruetsch, N. R., Sweet, H. O., Shultz, L. D., and Bosma, M. J. (1989). The mouse mutation severe combined immune deficiency (scid) is on chromosome 16. Immunogenetics *29*, 54–57.

Bucchini, D., Reynaud, C., Ripoche, M., Grimal, H., Jami, J., and Weill, J. (1987). Rearrangement of a chicken immunoglobulin gene occurs in the lymphoid lineage of transgenic mice. Nature *326*, 409–411.

Caput, D., Buetler, B., Hartog, K., Thayer, R., Brown-Shimer, S., and Cerami, A. (1986). Identification of a common nucleotide sequence in the 3′-untranslated region of mRNA molecules specifying inflammatory mediators. Proc. Natl. Acad. Sci. USA *83*, 1670–1674.

Coffman, R., and Weissman, I. (1981). B220: a B cell specific member of the T200 glycoprotein family. Nature *289*, 681–683.

Davis, M. M., and Bjorkman, P. J. (1988). T-cell antigen receptor genes and T-cell recognition. Nature *334*, 395–402.

Dayhoff, M. O., Barker, W. C., Hunt, L. T., and Schwartz, R. M. (1978). Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, (Washington D.C.: National Biomedical Research Foundation), pp. 1–8.

Dayhoff, M. O., Barker, W. C., and Hunt, L. T. (1983). Establishing homologies in protein sequences. Meth. Enzymol. *91*, 524–545.

Desiderio, S., and Baltimore, D. (1984). Double-stranded cleavage by cell extracts near recombinational signal sequences of immunoglobulin genes. Nature *308*, 860–862.

Doolittle, R. F. (1981). Similar amino acid sequences: chance or common ancestry? Science *214*, 149–159.

Enoch, T., Zinn, K., and Maniatis, T. (1986). Activation of the human β-interferon gene requires an interferon-inducible factor. Mol. Cell. Biol. *6*, 801–810.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction fragments to high specific activity. Anal. Biochem. *132*, 6–13.

Gerace, L., and Burke, B. (1988). Functional organization of the nuclear envelope. Annu. Rev. Cell Biol. *4*, 335–374.

Goldfarb, M., Shimizu, K., Perucho, M., and Wigler, M. (1982). Isolation and preliminary characterization of a human transforming gene from T24 bladder carcinoma cells. Nature *296*, 404–409.

Goodhardt, M., Cavelier, P., Akimenko, M. A., Lutfalla, G., Babinet, C., and Rougeon, F. (1987). Rearrangement and expression of rabbit immunoglobulin κ light chain gene in transgenic mice. Proc. Natl. Acad. Sci. USA *84*, 4229–4233.

Gubler, U., and Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. Gene *25*, 263–269.

Halligan, B. D., and Desiderio, S. V. (1987). Identification of a DNA binding protein that recognizes the nonamer recombinational signal sequence of immunoglobulin genes. Proc. Natl. Acad. Sci. USA *84*, 7019–7023.

Hendrickson, E. A., Schatz, D. G., and Weaver, D. T. (1988). The *scid* gene encodes a *trans*-acting factor that mediates the rejoining event of Ig gene rearrangement. Genes Dev. *2*, 817–829.

Hesse, J. E., Lieber, M. R., Mizuuchi, K., and Gellert, M. (1989). V(D)J recombination: a functional definition of the joining signals. Genes Dev. *3*, 1053–1061.

Hope, T. J., Aguilera, R. J., Minie, M. E., and Sakano, H. (1986). Endonucleolytic activity that cleaves immunoglobulin recombination sequences. Science *231*, 1141–1145.

Huber, H. E., Iida, S., Arber, W., and Bickle, T. A. (1985). Site-specific DNA inversion is enhanced by a DNA sequence element in *cis*. Proc. Natl. Acad. Sci. USA *82*, 3776–3780.

Johnson, R. C., and Simon, M. I. (1985). Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer. Cell *41*, 781–789.

Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M., Perry, H. M., and Gottesman, K. S. (1987). Sequences of Proteins of Immunological Interest, Fourth Edition (Bethesda, Maryland: National Institutes of Health).

Kahmann, R., Rudt, F., Koch, C. and Mertens, G. (1985). G inversion in bacteriophage Mu DNA is stimulated by a site within the invertase gene and a host factor. Cell *41*, 771–780.

Kataoka, T., Kondo, S., Nishi, M., Kodaira, M., and Honjo, T. (1984). Isolation and characterization of endonuclease J: a sequence-specific endonuclease cleaving immunoglobulin genes. Nucl. Acids Res. *12*, 5995–6010.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNA's. Nucl. Acids Res. *12*, 857–872.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribososmes. Cell *44*, 283–292.

Lachman, L. B., Hacker, M. P., Blyden, G. T., and Handschumacher, R. E. (1977). Preparation of lymphocyte-activity factor from continuous murine macrophage cell lines. Cell. Immunol. *34*, 416–419.

Landau, N. R., Schatz, D. G., Rosa, M., and Baltimore, D. (1987). Increased frequency of N-regional insertion in a murine pre-B-cell line infected with a terminal deoxynucleotidyl transferase retroviral expression vector. Mol. Cell. Biol. *7*, 3237–3243.

Lanford, R. E., White, R. G., Dunham, R. G., and Kanda, P. (1988). Effect of basic and nonbasic amino acid substitutions on transport induced by simian virus 40 T-antigen synthetic peptide nuclear transport signals. Mol. Cell. Biol. *8*, 2722–2729.

Lassar, A. B., Paterson, B. M., and Weintraub, H. (1986). Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. Cell *47*, 649–656.

Lieber, M. R., Hesse, J. E., Mizuuchi, K., and Gellert, M. (1987). Developmental stage specificity of the lymphoid V(D)J recombination activity. Genes Dev. *1*, 751–761.

Lieber, M. R., Hesse, J. E., Lewis, S., Bosma, G. C., Rosenberg, N., Mizuuchi, K., Bosma, M. J., and Gellert, M. (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.

Litman, G. W., Berger, L., Murphy, K., Litman, R., Hinds, K., and Erickson, B. W. (1985a). Immunoglobulin V_H gene structure and diversity in *Heterodontus*, a phylogenetically primitive shark. Proc. Natl. Acad. Sci. USA *82*, 2082–2086.

Litman, G. W., Murphy, K., Berger, L., Litman, R., Hinds, K., and Erickson, B. W. (1985b). Complete nucleotide sequences of three V_H genes in *Caiman*, a phylogenetically ancient reptile: evolutionary diversification in coding segments and variation in the structure and organization of recombination elements. Proc. Natl. Acad. Sci. USA *82*, 844–848.

Lowy, I., Pellicer, A., Jackson, J. F., Sim, G.-K., Silverstein, S., and Axel, R. (1980). Isolation of transforming DNA: cloning the hamster *aprt* gene. Cell *22*, 817–823.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Mitchell, P. J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371–378.

Nader, W. F., Burland, T., Cox, R. A., Dove, W. F., Monteiro, M. J., and Paul, E. C. A. (1986). Gene cloning and construction of genomic libraries in *Physarum*. In The Molecular Biology of *Physarum polycephalum*, W. F. Dove, J. Dee, S. Hatano, F. B. Haugli, and K. Wohlfarth-Bottermann, eds. (New York: Plenum Press), pp. 291–300.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA *85*, 2444–2448.

Perucho, M., Hanahan, D., Lipsich, L., and Wigler, M. (1980). Isolation of the chicken thymidine kinase gene by plasmid rescue. Nature *285*, 207–210.

Phillips, R. A., and Fulop, G. M. (1989). Pleiotropic effects of the *scid* mutation: effects on lymphoid differentiation and on repair of radiation damage. Curr. Topics Microbiol. Immunol. *152*, in press.

Pinney, D. F., Pearson-White, S. H., Konieczny, S. F., Latham, K. E., and Emerson, C. P., Jr. (1988). Myogenic lineage determination and differentiation: evidence for a regulatory gene pathway. Cell *53*, 781–793.

Raff, M. C., Megson, M., Owen, J. J. T., and Cooper, M. D. (1976). Early production of intracellular IgM by B lymphocyte precursors in mouse. Nature *259*, 224–226.

Raulet, D. H. (1989). The structure, function, and molecular genetics of the γ/δ T cell receptor. Annu. Rev. Immunol. *7*, 175–207.

Reynaud, C.-A., Anquez, V., Dahan, A., and Weill, J.-C. (1985). A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. Cell *40*, 283–291.

Risser, R., Kaehler, D., and Lamph, W. W. (1985). Different genes control the susceptibility of mice to Moloney or Abelson murine leukemia viruses. J. Virol. *55*, 547–543.

Sanger, R., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Sartoris, S., Cohen, E. B., and Lee, J. S. (1987). A rapid and improved method for generating cDNA libraries in plasmid and phage lambda vectors. Gene *56*, 301–307.

Schatz, D. G. (1989). Oligonucleotide hybridization in dried agarose gels. In Current Protocols in Molecular Biology, Vol. 1, F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, eds. (New York: Greene Publishing Associates and Wiley-Interscience), pp. 2.9.10–2.9.13.

Schatz, D. G., and Baltimore, D. (1988). Stable expression of immunoglobulin gene V(D)J recombinase activity by gene transfer into 3T3 fibroblasts. Cell *53*, 107–115.

Schlissel, M. S., and Baltimore, D. (1989). Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. Cell *58*, 1001–1007.

Schuler, W., Weiler, I. J., Schuler, A., Phillips, R. A., Rosenberg, N., Mak, T. W., Kearney, J. F., Perry, R. P., and Bosma, M. J. (1986). Rearrangement of antigen receptor genes is defective in mice with severe combined immune deficiency. Cell *46*, 963–972.

Schwager, J., Grossberger, D., and Du Pasquier, L. (1988). Organization and rearrangement of immunoglobulin M genes in the amphibian *Xenopus*. EMBO J. *7*, 2409–2415.

Seed, B. (1987). An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its receptor CD2. Nature *329*, 840–842.

Shaw, G., and Kamen, R. (1986). A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell *46*, 659–667.

Staudt, L. M., Clerc, R. G., Singh, H., LeBowitz, J. H., Sharp, P. A., and Baltimore, D. (1988). Cloning of a lymphoid-specific cDNA encoding a

protein binding the regulatory octamer DNA motif. Science *241*, 577–580.

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

Vogelstein, B., and Gillespie, D. (1979). Preparation and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA *76*, 615–619.

Weisberg, R. A., and Landy, A. (1983). Site-specific recombination in phage lambda. In Lambda II, R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 211–250.

Yancopoulos, G. D., and Alt, F. W. (1985). Developmentally controlled and tissue-specific expression of unrearranged V_H gene segments. Cell *40*, 271–281.

Yancopoulos, G., Blackwell, T. K., Suh, H., Hood, L., and Alt, F. W. (1986). Introduced T cell receptor variable region gene segments recombine in pre-B cells: evidence that B and T cells use a common recombinase. Cell *44*, 251–259.

GenBank Accession Number

The accession number for the human sequence reported in this paper is M29474, and for the mouse sequence is M29475.