Research Article

RAG-1 and RAG-2, Adjacent Genes That Synergistically Activate V(D)J Recombination

MARJORIE A. OETTINGER, DAVID G. SCHATZ, CAROLYN GORKA, DAVID BALTIMORE

The vast repertoire of immunoglobulins and T cell receptors is generated, in part, by V(D)J recombination, a series of genomic rearrangements that occur specifically in developing lymphocytes. The recombination activating gene, RAG-1, which is a gene expressed exclusively in maturing lymphoid cells, was previously isolated. RAG-1 inefficiently induced V(D)J recombinase activity when transfected into fibroblasts, but cotransfection with an adjacent gene, RAG-2, has resulted in at least a 1000-fold increase in the frequency of recombination. The 2.1-kilobase RAG-2 complementary DNA encodes a putative protein of 527 amino acids whose sequence is unrelated to that of RAG-1. Like RAG-1, RAG-2 is conserved between species that carry out V(D)J recombination, and its expression pattern correlates precisely with that of V(D)J recombinase activity. In addition to being located just 8 kilobases apart, these convergently transcribed genes are unusual in that most, if not all, of their coding and 3′ untranslated sequences are contained in single exons. RAG-1 and RAG-2 might activate the expression of the V(D)J recombinase but, more likely, they directly participate in the recombination reaction.

V(D)J RECOMBINATION IS THE COMBINATORIAL PROCESS by which developing lymphocytes begin to generate their enormous range of binding specificities from a limited amount of genetic information. Variable (V), joining (J), and sometimes diversity (D) gene segments at seven different loci (α, β, γ, and δ for T cell receptors, μ, κ, and λ for the immunoglobulin genes of B cells) are joined by this site-specific recombination reaction (*1, 2*). The assembly process is tightly regulated, occurring in a preferred temporal order (D_H joins to J_H before V_H joins to D_HJ_H; μ segments rearrange before κ) (*3–7*) and in a lineage-specific manner (for example, T cell receptor loci are never fully rearranged in B cells) (*1, 2*). Rearrangements are mediated by recombination signal sequences (RSS's) that flank all recombinationally competent V, D, and J gene segments. RSS's, necessary and sufficient to direct recombination, consist of a dyad-symmetric heptamer, an AT-rich nonamer, and an intervening spacer region of either 12 or 23 bp (*8, 9*). These signals are conserved among the different loci and species that carry out V(D)J recombination (*10–13*) and are functionally interchangeable (*14–17*), suggesting that the joining reaction is catalyzed by a single, evolutionarily conserved, V(D)J recombinase.

The recombination machinery used by developing lymphocytes is not yet understood. A number of events must occur during the joining reaction; these include the recognition of RSS's, endonucleolytic cleavage at or near the signal border, base trimming and addition (the joints of coding sequences are imprecise), and ligation of the cleaved ends. Some of these activities may be carried out by proteins found in many cell types, but others are likely to be specific to recombinationally active lymphocytes. It is this latter class of pre-B cell– and pre-T cell–specific factors that we refer to as the V(D)J recombinase.

We previously reported that transfection of genomic DNA into NIH 3T3 cells, a fibroblastoid line that lacks V(D)J joining activity, could stably induce the expression of the V(D)J recombinase (*18*). Detection of recombinase activity after genomic transfection relied on the use of an integrated recombination substrate (DGR), which on rearrangement mediated by recombination signal sequences, conferred resistance to the drug mycophenolic acid (MPA). The frequency with which we obtained fibroblasts that had carried out a correct rearrangement event suggested that a single genetic locus was responsible for the induction of recombinase activity. Tagging the genomic DNA with an oligonucleotide allowed us to follow this locus through several rounds of transfection and to identify a tagged DNA fragment that cosegregated with recombinase activity (*19*). This fragment served as the starting point for a chromosomal walk that ended with the identification of the recombination activating gene (RAG-1), a lymphoid-specific gene whose expression correlates precisely with recombinase activity.

Transfection of fibroblasts with cloned DNA capable of expressing RAG-1 demonstrated that this gene could induce V(D)J recombination (*19*). However, we were puzzled by the fact that four independent RAG-1 cDNA's and an 18-kb genomic clone induced recombination at the same frequency as total genomic DNA (*20*). We had expected that a purified gene would raise the transfection efficiency about 100- to 1000-fold over that seen with the genomic DNA.

The inefficiency of the cloned genomic DNA, phage 12C.2, could be explained by its lack of a promoter and the absence of the nucleotides encoding the NH_2-terminal 36 amino acids (*19*), a region well conserved between humans and mice. There was, however, no obvious explanation for the inefficiency of the four independent cDNA's. Because all four cDNA's behaved equivalently in our assay, their poor function was unlikely the result of mutation.

In an attempt to explain the inefficiency with which cloned RAG-1 sequences induced V(D)J recombination, we compared the structures of the cDNA and genomic clones. Surprisingly, this structural

The authors are at the Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142.

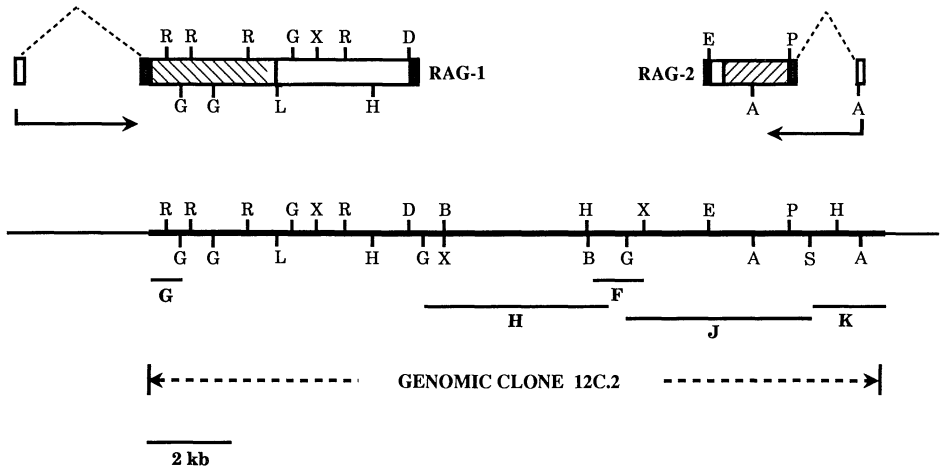

Fig. 1. Schematic diagram of the RAG locus. A partial restriction map of the RAG-1 (M6) and RAG-2 (MR2-1) cDNA clones and the genomic clone (12C.2) is indicated along with the relative positions of RAG-1 and RAG-2 sequences with respect to each other and to genomic DNA. Arrows indicate the direction of transcription. The RAG-1 cDNA is approximately 6.6 kb [corrected from the previous report (*19*)]. The RAG-2 cDNA is 2.1 kb excluding the poly(A) tail, and was obtained from a library prepared from the mouse pre-B cell line 22D6 (*19*). The genomic clone, 12C.2, is approximately 18 kb and represents the genomic configuration at the RAG locus. The locations of probes F, G, H, J, and K are indicated. Probes F and G, used in the chromosomal walk to RAG-1, have been described (*19*). Probes H, J, and K are 5.1-kb Bgl II, 4.5-kb Bgl II–Sal I, and 2.0-kb Sal I fragments, respectively, and are derived from phage 12C.2. Hatched areas indicate coding sequences. Shaded areas at the 5′ and 3′ ends of the RAG-1 and RAG-2 main exons indicate regions that are outside of the diagnostic restriction enzyme sites and may not be contained in those exons. The 5′ untranslated region may be encoded by more than one exon and is shown as a single exon for convenience only. The first 700 nucleotides of 12C.2, probe G, are colinear with the RAG-1 cDNA (*19*). All restriction fragment lengths from nucleotide 201 of the RAG-1 cDNA to the indicated Dra II site in RAG-1 and from the first Pst I site to the Eco RV site in RAG-2 were shown to be indistinguishable from those in the genomic clone by standard techniques. Hybridization with appropriate fragments of cDNA demonstrated that all sequences upstream from the Dra II site of the RAG-1 cDNA clone (110 bp upstream from the end of the clone) were confined to the first 6.6 kb of the genomic clone (to the left of the Bgl II site). A subset of the following restriction enzyme sites are indicated: A, Rsal; B, Bam HI; D, Dra II; E, Eco RV; G, Bgl II; H, Hind III; L, Cla I; P, Pst I; R, Eco RI; S, Sal I; X, Xba I.

mapping (Fig. 1) indicated that a single exon encoded most, if not all, of the RAG-1 structural gene and 3′ untranslated region and that the RAG-1 sequences were confined to the leftmost 6.6 kb of the genomic clone. Thus, genomic clone 12C.2 contained 12 kb of sequence whose function, if any, was unknown.

Identification of RAG-2. The unusual genomic structure of RAG-1, in conjunction with the genetic evidence regarding the inefficiency with which the RAG-1 cDNA and genomic clones induced recombination, prompted us to consider the possibility that the activation and stable expression of recombinase activity was dependent on two closely linked genes, RAG-1 and RAG-2, both largely contained in genomic clone 12C.2. By this model, the cloned RAG-1 cDNA would be expected to activate recombination poorly because it does not provide RAG-2 function. Similarly, the genomic clone, containing only one intact gene, RAG-2, would induce recombination inefficiently. Finally, the tight linkage of RAG-1 and RAG-2 would have resulted in the two genes behaving as a single genetic locus when total genomic DNA was originally transfected (*18*).

We tested the two-gene hypothesis by asking if cotransfection of the 12C.2 genomic clone and the RAG-1 cDNA cloned in an expression construct would increase the frequency with which transfected fibroblasts carried out rearrangement. Therefore, 3TGR fibroblasts, NIH 3T3 cells containing two copies of the DGR recombination substrate, were transfected either with the RAG-1 cDNA alone, the genomic clone alone, or the genomic and cDNA clones together. Two selection protocols were followed. In one case, cells were tested for rearrangement 4 days after transfection without prior selection for the uptake of DNA. Strikingly, six independent cotransfections of the genomic and RAG-1 cDNA clones each yielded 20 to 25 MPA-resistant colonies, whereas no resistant cells were observed after transfection with either the genomic or cDNA clones individually, or with total genomic DNA. In the alternative protocol, cells transfected with a co-selectable marker were selected for DNA uptake for 6 to 9 days before being tested with MPA. Under these conditions, approximately 500 colonies were obtained after the fibroblasts were transfected with the combination of cDNA and genomic clones, while transfection with either clone alone yielded no colonies (*21*). By comparison, transfection with genomic DNA gave from 5 to 30 colonies. The enormous increase in the number of fibroblasts that carried out rearrangements after transfection of the cDNA and genomic clones together strongly supported the hypothesis that the RAG-1 cDNA could be complemented by a second gene in genomic clone 12C.2. However, an alternative explanation was that recombination between the two transfected DNA's had generated a fully functional RAG-1 locus. Isolation of the putative second gene, as described below, ruled out this latter possibility.

We screened for a second gene in the genomic clone by hybridization of probes spanning the remaining sequences of 12C.2 (probes H, J, and K in Fig. 1) to filters containing poly(A)$^+$ RNA from recombinase positive (rec$^+$) and recombinase negative (rec$^-$) cell lines. While probes H and K failed to give detectable hybridization to any of the RNA samples, probe J hybridized to a predominant 2.0- to 2.2-kb mRNA present in a rec$^+$ pre-B cell (38B9) and the rec$^+$ genomic transfectant TRX-1 but not in a rec$^-$ mature B cell (WEHI-231) or in the recipient fibroblast, 3TGR (representative results of hybridization to poly(A)$^+$ RNA with cloned sequences corresponding to the 2.2-kb mRNA are shown in Fig. 4). These results indicated the presence of a second gene, RAG-2, in the genomic clone. Furthermore, the mRNA size and expression pattern indicated that it was likely to be distinct from RBP-2, the only other previously cloned gene suggested as a component of the V(D)J recombination machinery (*22*).

Probe J was used to screen a mouse pre-B cell cDNA library. We obtained 40 positives from this size-selected library out of 10^6 recombinants. The insert sizes of ten phage were determined and one clone with a 2.2-kb insert (referred to below as the RAG-2 cDNA clone) was chosen for further study.

Sequence of RAG-2. We determined the nucleotide sequence of the RAG-2 cDNA clone (Fig. 2). The sequence, 2062 bp followed by a poly(A) tail, contains a single long open reading frame capable of encoding a polypeptide of 527 amino acids with a molecular size of 58 kD. The putative ATG initiator codon is the first found in the sequence and lies downstream of an in-frame stop codon. The open reading frame is flanked on the 5′ and 3′ sides by untranslated

regions of 156 and 262 nucleotides, respectively.

No similarity was found between the deduced amino acid sequence of RAG-2 and that of RAG-1 or of any sequence contained in the National Biomedical Research Foundation protein database. Comparison of nucleotide sequences of the two RAG cDNA's did not reveal any shared sequences nor did a comparison of RAG-2 with the GenBank database. RAG-2 contains a lengthy acidic region; of 60 amino acids, 25 are acidic (Fig. 2, underlined region).

Synergy of RAG-1 and RAG-2. To demonstrate the biological activity of the RAG-2 cDNA clone, we used an assay for recombinase activity in which the extrachromosomal recombination substrate pJH200 (*17, 23*) was transiently transfected into mammalian cells, recovered after 48 hours, and then introduced into bacteria. This substrate contains the bacterial ampicillin and chloramphenicol resistance genes but confers resistance to chloramphenicol (Cam) only after V(D)J rearrangement or other nonspecific recombination events. Cam-resistant colonies, representing correct V(D)J rearrangement, were identified by hybridization with an oligonucleotide probe. This probe is complementary to the back-to-back heptamer sequence formed by fusion of the two RSS's after deletional V(D)J recombination of pJH200 (Fig. 3). Thus, recombination frequency is measured as the number of colonies resistant to both Amp (ampicillin) and Cam that hybridize to the oligonucleotide divided by the total number of Amp-resistant colonies.

When pJH200 was introduced into our standard recipient fibroblast, 3TGR, either alone or with either one of the RAG cDNA's cloned in a mammalian expression vector, no Cam^r colonies that hybridized to the oligonucleotide were observed. In contrast, hundreds of rearranged plasmids were recovered from fibroblasts transfected with both RAG-1 and RAG-2. This represented a recombination frequency of approximately 1.0 percent and at least a

```
   1  GCCGAGTTTA ATTCCTGGCT TGGCCGAAAG GATTCAGAGA GGGATAAGCA GCCCCTCTGG CCTTCAGTGC CAAAATAAGA AAGAGTATTT CACATCCACA AGCAGGAAGT   110
 111  ACACTTCATA CCTCTCTAAG ATAAAAGACC TATTCACAAT CAAAA

 156  ATG TCC CTG CAG ATG GTA ACA GTG GGT CAT AAC ATA GCC TTA ATT CAA CCA GGC TTC TCA CTT ATG AAT TTT GAT GGC CAA GTT TTC TTC   245
   1   M   S   L   Q   M   V   T   V   G   H   N   I   A   L   I   Q   P   G   F   S   L   M   N   F   D   G   Q   V   F   F    30

 246  TTT GGC CAG AAA GGC TGG CCT AAG AGA TCC TGT CCT ACT GGA GTC TTT CAT TTT GAT ATA AAA CAA AAT CAT CTC AAA CTG AAG CCT GCA   335
  31   F   G   Q   K   G   W   P   K   R   S   C   P   T   G   V   F   H   F   D   I   K   Q   N   H   L   K   L   K   P   A    60

 336  ATC TTC TCT AAA GAT TCC TGC TAC CTC CCA CCT CTT CGT TAT CCA GCT ACT TGC TCA TAC AAA GGC AGC ATA GAC TCT GAC AAG CAT CAA   425
  61   I   F   S   K   D   S   C   Y   L   P   P   L   R   Y   P   A   T   C   S   Y   K   G   S   I   D   S   D   K   H   Q    90

 426  TAT ATC ATT CAC GGA GGG AAA ACA CCA AAC AAT GAG CTT TCC GAT AAG ATT TAT ATC ATG TCT GTC GCT TGC AAG AAT AAC AAA AAA GTT   515
  91   Y   I   I   H   G   G   K   T   P   N   N   E   L   S   D   K   I   Y   I   M   S   V   A   C   K   N   N   K   K   V   120

 516  ACT TTC CGT TGC ACA GAG AAA GAC TTA GTA GGA GAT GTC CCT GAA CCC AGA TAC GGC CAT TCC ATT GAC GTG GTG TAT AGT CGA GGG AAA   605
 121   T   F   R   C   T   E   K   D   L   V   G   D   V   P   E   P   R   Y   G   H   S   I   D   V   V   Y   S   R   G   K   150

 606  AGC ATG GGT GTT CTC TTT GGA GGA CGT TCA TAC ATG CCT TCT ACC CAG AGA ACC ACA GAA AAA TGG AAT AGT GTA GCT GAC TGC CTA CCC   695
 151   S   M   G   V   L   F   G   G   R   S   Y   M   P   S   T   Q   R   T   T   E   K   W   N   S   V   A   D   C   L   P   180

 696  CAT GTT TTC TTG ATA GAT TTT GAA TTT GGG TGT GCT ACA TCA TAT ATT CTC CCA GAA CTT CAG GAT GGG CTG TCT TTT CAT GTT TCT ATT   785
 181   H   V   F   L   I   D   F   E   F   G   C   A   T   S   Y   I   L   P   E   L   Q   D   G   L   S   F   H   V   S   I   210

 786  GCC AGA AAC GAT ACC GTT TAT ATT TTG GGA GGA CAC TCA CTT GCC AGT AAT ATA CGC CCT GCT AAC TTG TAT AGA ATA AGA GTG GAC CTT   875
 211   A   R   N   D   T   V   Y   I   L   G   G   H   S   L   A   S   N   I   R   P   A   N   L   Y   R   I   R   V   D   L   240

 876  CCC CTG GGT ACC CCA GCA GTG AAT TGC ACA GTC TTG CCA GGA GGA ATC TCT GTC TCC AGT GCA ATC CTC ACT CAA ACA AAC AAT GAT GAA   965
 241   P   L   G   T   P   A   V   N   C   T   V   L   P   G   G   I   S   V   S   S   A   I   L   T   Q   T   N   N   D   E   270

 966  TTT GTT ATT GTG GGT GGT TAT CAG CTG GAA AAT CAG AAA AGG ATG GTC TGC AGC CTT GTC TCT CTA GGG GAC AAC ACG ATT GAA ATC AGT  1055
 271   F   V   I   V   G   G   Y   Q   L   E   N   Q   K   R   M   V   C   S   L   V   S   L   G   D   N   T   I   E   I   S   300

1056  GAG ATG GAG ACT CCT GAC TGG ACC TCA GAT ATT AAG CAT AGC AAA ATA TGG TTT GGA AGC AAC ATG GGA AAC GGG ACT ATT TTC CTT GGC  1145
 301   E   M   E   T   P   D   W   T   S   D   I   K   H   S   K   I   W   F   G   S   N   M   G   N   G   T   I   F   L   G   330

1146  ATA CCA GGA GAC AAT AAG CAG GCT ATG TCA GAA GCA TTC TAT TTC TAT ACT TTG AGA TGC TCT GAA GAG GAT TTG AGT GAA GAT CAG AAA  1235
 331   I   P   G   D   N   K   Q   A   M   S   E   A   F   Y   F   Y   T   L   R   C   S   E   E   D   L   S   E   D   Q   K   360

1236  ATT GTC TCC AAC AGT CAG ACA TCA ACA GAA GAT CCT GGG GAC TCC ACT CCC TTT GAA GAC TCA GAG GAA TTT TGT TTC AGT GCT GAA GCA  1325
 361   I   V   S   N   S   Q   T   S   T   E   D   P   G   D   S   T   P   F   E   D   S   E   E   F   C   F   S   A   E   A   390

1326  ACC AGT TTT GAT GGT GAC GAT GAA TTT GAC ACC TAC AAT GAA GAT GAT GAA GAT GAC GAG TCT GTA ACC GGC TAC TGG ATA ACA TGT TGC  1415
 391   T   S   F   D   G   D   D   E   F   D   T   Y   N   E   D   D   E   D   D   E   S   V   T   G   Y   W   I   T   C   C   420

1416  CCT ACT TGT GAT GTT GAC ATC AAT ACC TGG GTT CCG TTC TAT TCA ACG GAG CTC AAT AAA CCC GCC ATG ATC TAT TGT TCT CAT GGG GAT  1505
 421   P   T   C   D   V   D   I   N   T   W   V   P   F   Y   S   T   E   L   N   K   P   A   M   I   Y   C   S   H   G   D   450

1506  GGG CAC TGG GTA CAT GCC CAG TCG ATG GAT TTG GAA GAA CGC ACA CTC ATC CAC TTG TCA GAA GGA AGC AAC AAG TAT TAT TGC AAT GAA  1595
 451   G   H   W   V   H   A   Q   S   M   D   L   E   E   R   T   L   I   H   L   S   E   G   S   N   K   Y   Y   C   N   E   480

1596  CAT GTA CAG ATA GCA AGA GCA TTG CAA ACT CCC AAA AGA AAC CCC CCC TTA CAA AAA CCT CCA ATG AAA TCC CTC CAC AAA AAA GGC TCT  1685
 481   H   V   Q   I   A   R   A   L   Q   T   P   K   R   N   P   P   L   Q   K   P   P   M   K   S   L   H   K   K   G   S   510

1686  GGG AAA GTC TTG ACT CCT GCC AAG AAA TCC TTC CTT AGA AGA CTG TTT GAT TAA   1739
 511   G   K   V   L   T   P   A   K   K   S   F   L   R   R   L   F   D   *     527

1740  TTTAGCAAAA GCCCCTCAGA CTCAGGTATA TTGCTCTCTG AATCTACTTT CAATCATAAA CATTATTTTG ATTTTTGTTA CTGAAATCTC TATGTTATGT TTTAGTTATG  1849
1850  TGAATTAAGT GCTGTTGTGA TTTATTGTTA AGTATAACTA TTCTAATGTG TGTTTTTTAA CATCTTATCC AGGAATGTCT TAAATGAGAA ATGTTATACA GTTTTCCATT  1959
1960  AAGGATATCA GTGATAAAGT ATAGAACTCT TACATTATTT TGTAACAATC TACATATTGA ATAGTAACTA AATACCAATA AATAAACTAA TGCACAAAAA GTT          2062
```

Fig. 2. Nucleotide sequence of the mouse RAG-2 cDNA and predicted amino acid sequence of the RAG-2 protein. The nucleotide sequence of the entire RAG-2 cDNA clone MR2-1 is shown including 155 and 440 nucleotides of 5′ and 3′ untranslated sequences, respectively. The deduced amino acid sequence from nucleotides 156 to 1584 is also indicated, with the acidic region underlined.

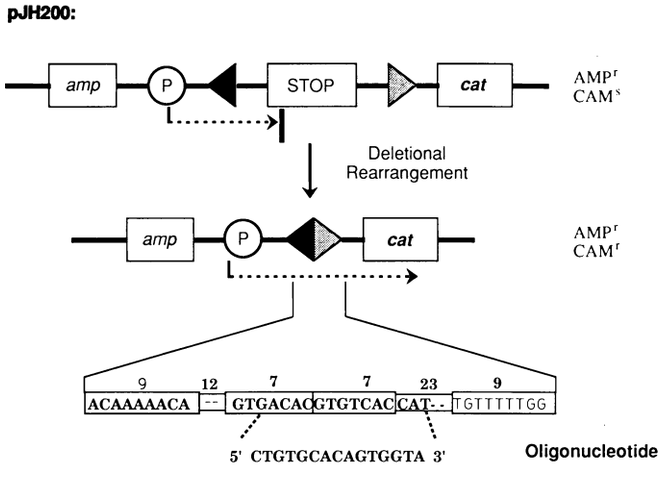

Fig. 3. The extrachromosomal recombination substrate pJH200 (*17*). Two recombination signal sequences (RSS's, indicated as triangles) flank a prokaryotic transcriptional terminator (labeled "stop"). The RSS's are flanked on one side by a bacterial promoter (P) and on the other, by the *cat* gene. Before V(D)J recombination, *cat* expression is prevented by the terminator. After rearrangement, the terminator is deleted and the gene is transcribed. Thus, *Escherichia coli* containing a rearranged plasmid will be Cam^r. The sequence complementary to the indicated oligonucleotide probe is formed only after correct V(D)J rearrangement.

1000-fold increase in recombinase activity over that induced by either RAG cDNA alone (Table 1). Sequencing confirmed that the plasmid recovered from colonies that hybridized to the oligonucleotide had indeed undergone normal V(D)J recombination. We thus concluded that the RAG-2 cDNA clone is biologically active and that RAG-1 and RAG-2 act synergistically to activate V(D)J recombination.

Because RAG-1 is highly conserved between mouse and human, we asked if a human RAG-1 cDNA clone could synergistically activate recombination together with the murine RAG-2 clone. The level of recombinase activity after transfection of NIH 3T3 fibroblasts with human RAG-1 and mouse RAG-2 expression constructs was approximately the same as that induced by the combination of murine RAG's 1 and 2 in the same cell line (Table 1). We had previously shown that transfection of human genomic DNA into mouse fibroblasts activates V(D)J recombination with the same efficiency as murine genomic DNA (*18*), while human RAG-1 cDNA clones induce rearrangement as inefficiently as their murine counterparts (*19*). Taken together, these results suggest that there must be a human RAG-2 homolog which is closely linked to human RAG-1.

While we were unable to detect recombinase activity in fibroblasts transiently transfected with either RAG-1 or RAG-2 alone, we had shown that RAG-1 could induce recombination on its own, albeit at a low frequency. We wanted to know if RAG-2 could also activate V(D)J recombination at a detectable frequency. To detect such rare events, we transfected fibroblasts (3TGR) containing DGR with the RAG-2 cDNA cloned in an expression vector, selected sequentially for uptake of DNA and rearrangement, and obtained a single MPA-resistant clone. Because this assay has no measurable background (*19*), this clone was likely to represent an authentic V(D)J recombination event. Indeed, we confirmed that the recombination substrate in this clone had undergone a normal V(D)J recombination event, with Southern (DNA) blot analysis to distinguish between the rearranged and unrearranged configuration of DGR. This finding suggested that the rearrangements seen after transfection with the 12C.2 genomic clone—a clone containing all of the RAG-2 coding sequences, but only a portion of RAG-1 (as shown below)—were probably a result of RAG-2 expression from 12C.2 and not a consequence of RAG-1 expression, as we had initially believed.

We next investigated whether expression of the RAG-1 and RAG-2 cDNA's was sufficient for the continued expression of recombinase activity. We measured the recombination frequency in cells stably expressing one or both RAG mRNA's by transiently transfecting these cells with the pJH200 recombination substrate. Stable expression of RAG-1 and RAG-2 mRNA was confirmed by Northern blot hybridization. Ongoing rearrangement was observed in fibroblasts stably transfected with both cDNA's, in transfectants containing the combination of the RAG-1 cDNA and the genomic clone, and in previously derived genomic transfectants (*19*) containing the entire RAG genomic locus (Table 2). The recombination frequencies varied from transfectant to transfectant, but were within the range normally seen in developing lymphocytes (*17*). In contrast, no recombinase activity was detected in cells expressing only one RAG mRNA (Table 2). Of particular interest was the absence of detectable recombinase activity in cell line MRH-1, a pool of fibroblasts transfected with just the RAG-1 cDNA clone, in which a normal rearrangement of the integrated recombination substrate had occurred. The absence of activity in these cells suggested that the rare fibroblasts that carry out a rearrangement after transfection

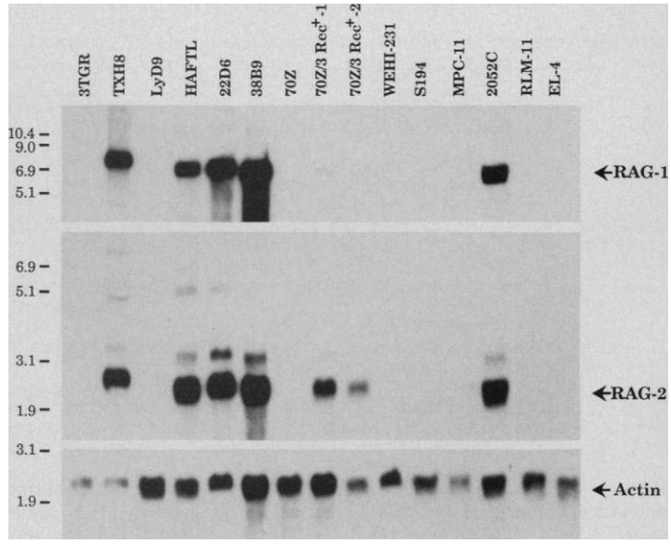

Fig. 4. RAG-2, like RAG-1, is expressed only in pre-B and pre-T cells and transfected fibroblasts (Northern blot analysis). A filter containing 16 μg of once-selected poly(A)$^+$ RNA or 8 μg of twice-selected poly(A)$^+$ RNA (22D6) was prepared by standard techniques. The cells surveyed include 3TGR fibroblasts; the genomic transfectant TXH8; the pro-B cell line LyD9 (*24*), which shows the germline configuration for immunoglobulin heavy and light chain genes; pre-B cells: HAFTL, 22D6 and 38B9, 70Z/3 cells, and 70Z/3 cells previously selected for V(D)J recombinase activity; mature B cell lines: WEHI-231, S194, and MPC-11; a pre-T line, 2052C, and the mature T cell lines RLM11 and EL-4. The filter was sequentially hybridized with probes specific for the entire RAG-2 cDNA [probe MR2-T, from the first nucleotide of the RAG-2 cDNA to the end of the poly(A) tail], the structural gene of RAG-1 (probe MR1-C, nucleotides 1 to 3380 of the RAG-1 cDNA) and γ-actin (*32*). Probe MR2-T hybridized to an approximately 2.1-kb mRNA indicated as RAG-2, along with several other bands of lesser intensity. Probe MR1-C hybridized to a 6.6-kb band indicated as RAG-1. Since the blot was stripped between probings and hybridization conditions and exposure times are not identical, the relative abundance of RAG-1 and RAG-2 mRNA's cannot be inferred from these data. Exposure times: RAG-2, 19 hours; RAG-1, 5 days; and actin, 10 hours. Northern transfer and hybridizations were carried out as described (*19*). The cell lines have been described previously (*3, 18, 33, 34*).

Table 1. Transient expression of recombinase activity. Transient expression of recombinase activity was assayed by a modification of the recombination assay described previously (*17, 23*). Approximately 2×10^6 fibroblasts of the cell line indicated were transfected with 10 μg of the recombination substrate pJH200 (Fig. 3) and 6 μg of RAG-1 or 4.8 μg of RAG-2 cDNA expression constructs. The mouse RAG-1 expression construct has been described (*19*). The human RAG-1 and mouse RAG-2 constructs are the H36 (*19*) and the MR2-1 cDNA inserts, respectively, which were excised with Not I and subcloned into the mammalian expression vector CDM8 with BstXI adaptors (*35*). The numbers of transformants (Amp^r or $oligo^+Amp^rCam^r$) has been corrected for a plating dilution factor. Because the CDM8 expression vector does not carry an Amp^r marker, the number of Amp^r transformants represents only the number of pJH200 plasmids recovered. The percentage of recombination, *R*, is calculated as the percentage of Amp^rCam^r colonies that hybridize to the oligonucleotide ($oligo^+$), divided by the total number of Amp^r colonies. DNA's were introduced into fibroblasts by calcium phosphate transfection (*19*). Plasmid was recovered 48 hours later by a rapid alkaline lysis procedure. The recovered plasmid DNA was digested with Dpn I, unless otherwise indicated, to eliminate plasmids that had not replicated and therefore may not have entered the cell and been available for recombination. The plasmid DNA was then introduced into *E. coli* strain MC1061 by electroshock (*36*), and then selected for Amp^r or Amp^r and Cam^r combined. Doubly resistant (Amp^rCam^r) bacteria were transferred to nitrocellulose membranes for colony–filter hybridization with a ^{32}P-labeled oligonucleotide probe. The sequence of the oligonucleotide, which corresponds to a portion of the sequence of the fused signal joints (Fig. 3) after recombination is 5′-CTGTGCACAGTGGTA-3′. Hybridizations were carried out at room temperature in 6× SSC (standard saline citrate) and 10× Denhardt's and washed in 6× SSC at 45°C for 2 hours.

Cell line	DNA	*N**	Amp^r	$Oligo^+$ Cam^rAmp^r	*R*
3TGR	0	2	184,000	0	0
3TGR	RAG-1	3	60,400	0	0
3TGR	RAG-2	3	50,000	0	0
3TGR	RAG-1 + RAG-2	2	70,600	490	0.7
NIH 3T3	RAG-1 + RAG-2	3	193,600	2,166	1.1
NIH 3T3	Human RAG-1 + mouse RAG-2	1	73,200	372	0.5

*Total number of independent transfections.

with a single cDNA do not arise from the selection of fibroblasts that have stably activated expression of the other RAG mRNA. Rather, these clones may reflect rearrangement induced by the transient transcription of one RAG mRNA in a transfected fibroblast expressing the other RAG cDNA, or they may indicate that both RAG-1 and RAG-2 can induce rearrangement by themselves, but with low efficiency.

Concordant expression of RAG-1 and RAG-2. In our initial characterization of RAG-1 we found that its pattern of expression as determined by Northern blot analysis correlated precisely with the known pattern of expression of V(D)J recombinase activity (*19*). Using the RAG-2 cDNA as a probe, we found an identical expression profile: RAG-2 mRNA was detectable only in pre-B and pre-T cell lines but not in cells earlier in the B lineage [LyD9 (*24*)] or in mature B and T cells (Fig. 4). RAG-2 mRNA was also detected in RNA from adult thymus (data not shown). The tight correlation between RAG-2 expression and recombinase activity is further supported by analysis of the late pre-B cell line 70Z/3. We reported previously that although the majority of 70Z/3 cells do not have detectable recombinase activity, rare rec^+ cells could be selected following infection with a recombination substrate (*18*). RAG-1 mRNA is detectable in these subclones, though not in the bulk culture (*19*). The same holds true for RAG-2 and indeed, the relative increase in expression of RAG-2 in rec^+ 70Z/3 cells is greater than it is for RAG-1 (Fig. 4). Other cell lines and tissues were surveyed with a sensitive PCR assay for RAG-2 expression. The RAG-2 signal was detectable in RNA from fetal liver and the pre-T line 2017 (*25*). In contrast, no RAG-2 was detected in RNA from a fibroblastoid line, LTK^-, from a macrophage line, P388D1 (*26*), or from an erythroid line, MEL. In addition, we have found no cell line or tissue that expresses only one of the two RAG mRNA's.

The RAG locus. The unusual genomic structure of RAG-1 and the close proximity of the two recombination activating genes made the genomic organization of RAG-2 of particular interest. Comparison of the restriction map of the cloned RAG-2 cDNA with genomic sequences indicated that the RAG-2 structural gene is contained entirely within the genomic clone (Fig. 1), as was predicted by the functional studies. The orientation of the two genes is such that they are convergently transcribed. Fragment lengths between the Pst I site four nucleotides 3′ to the ATG and the Eco RV site 97 nucleotides 5′ to the poly(A) tail in the cDNA are indistinguishable from those observed with genomic sequences, suggesting that all sequences between these two sites are present on a single exon. At least one intron must be present in the region 5′ to the Pst I site, because the distances between Rsa I sites (see Fig. 1) are not conserved between the cDNA and genomic clones. Thus the two genes at the RAG locus lie approximately 8 kb apart, in inverted orientation, but with similar exon-intron configurations.

The requirement for both RAG-1 and RAG-2 to efficiently activate V(D)J recombination and the known conservation of RAG-1 sequences between species carrying out V(D)J recombination led us to test whether RAG-2 was also conserved. Using the entire RAG-2 cDNA as a probe, we detected one or more bands in DNA from hamster, cow, dog, rabbit, opossum, and turtle (Fig. 5). Hybridization to human DNA and chicken DNA was also detected.

Molecular implications of two recombination activating genes. We previously suggested that RAG-1 either directly encodes a lymphoid-specific component of the recombination machinery or serves as a regulator of a pathway leading to the expression of the V(D)J recombinase (*19*). We favored the first hypothesis because (i) the RAG-1 expression pattern correlated precisely with recombinase activity, (ii) RAG-1 appeared to activate V(D)J recombination in transfected fibroblasts without inducing any other lymphoid-specif-

Table 2. Stable expression of recombinase activity. The level of recombinase activity in cell lines stably transfected with the indicated RAG cDNA's, the genomic clone (12C.2), or total genomic DNA was measured. The various cell lines were transiently transfected with pJH200 and the recombination frequencies were determined as outlined in Table 1. Cell line L-4, a single clone resulting from selection for V(D)J recombination after transfection with total genomic DNA has been described (*19*). Cell line MR1-H is a population of cells transfected with the RAG-1 cDNA that have rearranged their endogenous recombination substrate, DGR. The cell line indicated as 51P represents three independent populations of cells expressing RAG-2; each population was transiently transfected once and the results were combined. The remaining cell lines are all pools of clones derived from independent transfections with the DNA indicated. Cell line 51P was grown only in the presence of histidinol to select for the uptake of DNA. All other cell lines were grown in MPA (in addition to histidinol) to select for the uptake of DNA and rearrangement. Drug selections were carried out as previously described (*18*).

Cell line	DNA	*N*	Amp^r	$Oligo^+$ Cam^rAmp^r	*R*
MR1-H	RAG-1	2	228,000*	0	0
51P	RAG-2	3	390,000	0	0
31Ap	RAG-1 + 12C.2	1	260,000	2,200	0.85
34Ap	RAG-1 + 12C.2	1	68,000	310	0.45
34Bp	RAG-1 + 12C.2	1	90,000	240	0.26
52Cp	RAG-1 + RAG-2	1	46,000	4	0.008
52CHp	RAG-1 + RAG-2	1	30,000	126	0.42
54Ap	RAG-1 + RAG-2	1	240,000	78	0.03
L-4	Total genomic	2	220,000*	112	0.05

*Not Dpn I–digested.

ic properties, (iii) it was conserved through evolution in other species carrying out recombination, and (iv) RAG-1 shared no significant sequence similarity with any known transcription factor. These arguments are strengthened by the finding of a second gene with similar properties.

Models for the mechanism by which the RAG locus induces V(D)J recombination (*19*) must be modified to reflect the existence of RAG-2. The two genes RAG-1 and RAG-2 could encode functionally distinct subunits of a single lymphoid-specific V(D)J recombinase complex, or they could encode two separate factors carrying out different steps of the recombination reaction. Other site-specific recombination systems have requirements for reaction-specific factors that provide catalytic and specificity determinants (*27*). In addition, nonspecific factors are usually recruited to complete the reaction. Although the reaction-specific activities are often encoded by a single gene, recombination systems requiring two specific genes are also known—as with the λ Int and Xis proteins (*28*) and the bacteriophage Mu proteins, Mu A and Mu B (*29*). Thus, the requirement for two specific factors, RAG-1 and RAG-2, for efficient V(D)J recombination would not be unprecedented.

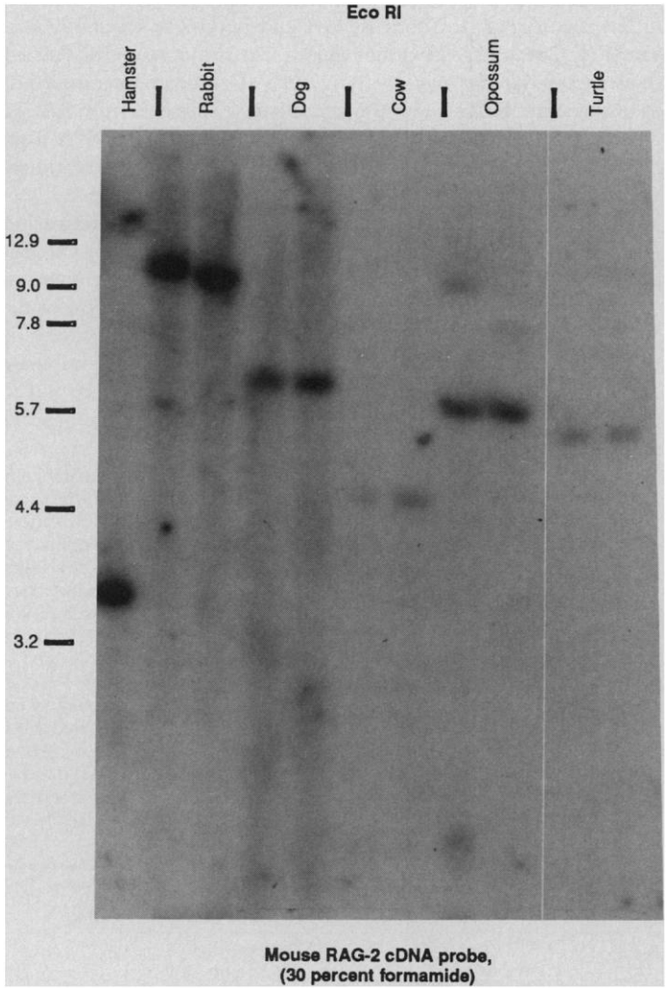

Fig. 5. The RAG-2 gene is conserved through evolution. DNA (8 μg) from the indicated species was digested with Eco RI, subjected to electrophoresis through a 0.8 percent agarose gel, and transferred to a Zetabind membrane. The two lanes for each species (except hamster) are from male and female animals, respectively. Hybridization with the RAG-2 cDNA probe MR2-T (see legend to Fig. 4) was carried out in 30 percent formamide at 42°C. Final washing conditions were 2× SSC, 0.1 percent SDS, 55°C. Marker sizes are shown in kilobases; exposure, 4 days.

If the RAG proteins play a regulatory rather than a catalytic role, they could do so by forming a heteromeric transcription factor—as with Fos/Jun (*30*) or HAP2/HAP3 (*31*) complexes where both proteins are important for activity. Alternatively, each RAG product could be an independent transcription factor that acts on distinct sets of genes in the recombinase pathway or synergistically activates transcription of recombinase genes. More indirect or complex regulatory mechanisms cannot be excluded.

Evolutionary considerations. The structure of the RAG locus is unusual for the genome of a higher eukaryote. First, for each gene, most (if not all) of the structural sequences and 3′ untranslated regions are encoded by a single exon, separated by at least one intron from the 5′ untranslated region. While there are certainly examples of mammalian genes devoid of introns, they are rare. More striking is the compact organization of RAG-1 and RAG-2 on the chromosome; such organization is more commonly seen in genomes of prokaryotic and viral origin. In mammalian cells, adjacent genes with related or identical function usually reflect the occurrence of gene duplication events (for example, the major histocompatibility complex, immunoglobulin, and globin genes). The absence of sequence similarity between RAG-1 and RAG-2 suggests that this is not the case here.

RAG-1 and RAG-2 are adjacent in the mouse genome and, by inference, in the human genome as well. Because both of these genes are required for recombinase expression and show comparable evolutionary conservation, it is reasonable to assume that they may also be linked in more distantly related organisms that carry out V(D)J recombination. One possibility for the origin of the RAG locus might be that it initially evolved as part of a viral or fungal recombination system. At some point, the virus or fungal genome could have been introduced into vertebrates or perhaps protochordates where the recombination system evolved to play its current role in V(D)J recombination.

REFERENCES AND NOTES

1. T. K. Blackwell and F. W. Alt, in *Molecular Immunology*, B. D. Hames and D. M. Glover, Eds. (IRL Press, Washington, DC, 1988), pp. 1–60.
2. M. M. Davis, in *ibid.*, pp. 61–79.
3. F. W. Alt *et al.*, *EMBO J.* **3**, 1209 (1984).
4. M. G. Reth, P. Amirati, S. Jackson, F. W. Alt, *Nature* **317**, 353 (1985).
5. D. H. Raulet, R. D. Garman, H. Saito, S. Tonegawa, *ibid.* **314**, 103 (1985).
6. L. E. Samelson *et al.*, *ibid.* **315**, 765 (1985).
7. H. R. Snodgrass, Z. Dembic, M. Steinmetz, H. von Boehmer, *ibid.*, p. 232.
8. S. Tonegawa, *ibid.* **302**, 575 (1983).
9. J. E. Hesee, M. R. Lieber, K. Mizuuchi, M. Gellert, *Genes Dev.* **3**, 1053 (1989).
10. G. W. Litman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2082 (1985).
11. G. W. Litman *et al.*, *ibid.*, p. 844.
12. C.-A. Reynaud, V. Anquez, A. Dahan, J.-C. Weill, *Cell* **40**, 283 (1985).
13. J. Schwager, D. Grossberger, L. Du Pasquier, *EMBO J.* **7**, 2409 (1988).
14. G. D. Yancopoulos *et al.*, *Cell* **44**, 251 (1986).
15. D. Bucchini *et al.*, *Nature* **326**, 409 (1987).
16. M. Goodhardt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4229 (1987).
17. M. R. Lieber, J. E. Hesse, K. Mizuuchi, M. Gellert, *Genes Dev.* **1**, 751 (1987).
18. D. G. Schatz and D. Baltimore, *Cell* **53**, 107 (1988).
19. D. G. Schatz, M. A. Oettinger, D. Baltimore, *ibid.* **59**, 1035 (1989).
20. The genomic DNA referred to here comes from the TRX-1 fibroblast. This fibroblast is itself a transfectant that stably expresses recombinase activity. The DNA from this cell line is tenfold more efficient at inducing recombination upon transfection into fibroblasts than is DNA from other nontransfected sources.
21. The absence of MPA-resistant cells after transfection with the individual cloned genes contrasts with previous results (*19*) and was probably due to a much decreased number of fibroblasts that took up DNA in this experiment and to the shortened time that cells were grown before MPA selection.
22. N. Matsunami *et al.*, *Nature* **342**, 934 (1989).
23. J. E. Hesse, M. R. Lieber, M. Gellert, K. Mizuuchi, *Cell* **49**, 775 (1987).
24. R. Palacios, H. Karasuyama, A. Rolink, *EMBO J.* **6**, 3687 (1987).
25. R. Risser, D. Kaehler, W. W. Lamph, *J. Virol.* **55**, 547 (1985).
26. L. B. Lachman *et al.*, *Cell. Immunol.* **34**, 416 (1977).
27. N. L. Craig, *Annu. Rev. Genet.* **22**, 77 (1988).
28. A. Landy, *Annu. Rev. Biochem.* **58**, 913 (1989).
29. K. Mizuuchi and R. Craigie, *Annu. Rev. Genet.* **20**, 385 (1986).
30. T. Curran and B. R. Franza, *Cell* **55**, 395 (1988).
31. S. Hahn and L. Guarente, *Science* **240**, 317 (1988).
32. T. Enoch, K. Zinn, T. Maniatis, *Mol. Cell. Biol.* **6**, 801 (1986).

33. F. Alt, N. Rosenberg, L. S. E. Thomas, D. Baltimore, *Cell* **27**, 381 (1981).
34. E. A. Kabat *et al.*, *Sequences of Proteins of Immunological Interest* (NIH, Bethesda, MD, 1987).
35. B. Seed, *Nature* **329**, 840 (1987).
36. W. J. Dower, J. F. Miller, C. W. Ragsdale, *Nucleic Acids Res.* **16**, 6127 (1988).
37. We thank R. Mosher, J. Bogan, and D. Page for the DNA's used in the Southern blot, K. Struhl for many helpful discussions, and K. Struhl, D. Black, and S. Smale for critical reading of the manuscript. Supported by NIH grant GM39458 and by fellowships from the Whitaker Health Sciences Fund (M.A.O.), and the Life and Health Insurance Medical Research Fund (O.G.S.). The GenBank accession number for the mouse RAG-2 sequence reported in this article is M33828.

30 April 1990; accepted 23 May 1990

Radar Images of Asteroid 1989 PB

STEVEN J. OSTRO, JOHN F. CHANDLER, ALICE A. HINE, KEITH D. ROSEMA, IRWIN I. SHAPIRO, DONALD K. YEOMANS

Radar observations of the near-Earth asteroid 1989 PB, made shortly after its optical discovery, yield a sequence of delay-Doppler images that reveal it to consist of two distinct lobes that appear to be in contact. It seems likely that the two lobes once were separate and that they collided to produce the current "contact-binary" configuration.

THIS ARTICLE PRESENTS RADAR OBSERVATIONS THAT PROvide two-dimensional images of a kilometer-sized asteroid and, quite unexpectedly, reveal it to have an extremely bifurcated, "double-lobed" shape. The existence of such an object has important implications for theories of the origin and evolution of asteroids and meteorite parent bodies as well as for our understanding of the cratering record of the terrestrial planets and the moon.

The imaged object, 1989 PB, is one of some 140 known near-Earth asteroids (NEAs). The NEA population is thought to include extinct cometary nuclei as well as fragments of mainbelt asteroids, and many NEAs might share (or be) the parent bodies of some meteorites (*1*). The known NEAs have sizes that range from ~50 m to ~50 km; the entire population probably contains $\sim 10^3$ objects with dimensions ≥ 1 km. Typical NEA lifetimes against collision with an inner planet or gravitational ejection from the solar system are only 10^7 to 10^8 years, but the lunar cratering record indicates that the flux of impacting projectiles has not changed very dramatically during the past 3 billion years, implying a balance between average rates of NEA depletion and replenishment (*2*). NEAs are among the most accessible objects for spacecraft rendezvous missions, but they are very difficult to study with ground-based optical telescopes and their physical characteristics are poorly known (*3*).

Radar observations can resolve NEAs spatially if the echoes are strong enough. On 9 August 1989, Helin (*4*) discovered a rapidly moving object on photographic plates taken at Palomar Observatory. Two days later, orbital calculations showed that the asteroid, designated 1989 PB by the Minor Planet Center (*5*), was approaching Earth and would pass through the Arecibo Observatory's declination window during 19–22 August at distances likely to provide echoes much stronger than in previous asteroid radar investigations (*6*). The asteroid would make its closest approach to Earth (0.027 AU, or 11 lunar distances) on 25.2 August UTC, while the Voyager 2 spacecraft was making its closest approach to Neptune. The Goldstone 70-m antenna was occupied with Voyager communications throughout the Neptune encounter and it was unavailable for use as a radar telescope until 30 August. We observed 1989 PB at Arecibo on 19–22 August and at Goldstone on 30 August (Table 1), and here we report our principal results.

Orbit refinement. Radar observations place stringent demands on the accuracy of predictions of echo time delay and Doppler frequency as well as pointing coordinates (*7*). However, errors in prediction ephemerides for freshly discovered NEAs grow rapidly, since orbits must be estimated from optical astrometric data that span short arcs. Uncertainties in a 19 August ephemeris based on optical observations obtained through 11 August (including "prediscovery" measurements found on photographic plates dating back to 1 August) would have compromised the radar observations and may even have prevented detection of echoes altogether. Optical measurements of the position of 1989 PB during 12–18 August were hampered by the nearby, nearly full moon. Fortunately, R. McNaught made critical astrometric measurements during 12–15 August at Siding Spring Observatory, Australia. In addition, two observers in Great Britain (J. D. Shanklin at Cambridge and B. Manning at Stakenbridge) made useful measurements during the 17 August total lunar eclipse.

On 19 August, using an ephemeris based on optical data from 1 to

Table 1. Geocentric distance, right ascension (RA), and declination (DEC) for radar observations of 1989 PB at epochs of the asteroid's Arecibo transit on 19–22 August and the midpoint of Goldstone observations on 30 August.

Date (1989)	Time (UTC) (hh:mm)	Distance (AU)	RA (h)	DEC (deg)	Observatory	Wavelength (cm)
19 Aug	06:33	0.060	23.93	7.0	Arecibo	13
20 Aug	06:34	0.052	0.01	11.3	Arecibo	13
21 Aug	06:36	0.044	0.13	17.2	Arecibo	13
22 Aug	06:43	0.037	0.30	25.4	Arecibo	13
30 Aug	19:35	0.057	10.78	45.0	Goldstone	3.5

S. J. Ostro, K. D. Rosema, and D. K. Yeomans are at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109. J. F. Chandler and I. I. Shapiro are at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138. A. A. Hine is at the National Astronomy and Ionosphere Center, Box 995, Arecibo, PR 00613.