1810–1814 Nucleic Acids Research, 1994, Vol. 22, No. 10

Analysis of regions of RAG-2 important for V(D)J recombination

Christina A.Cuomo and Marjorie A.Oettinger*
Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114 and
Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

Received March 3, 1994; Accepted April 7, 1994

ABSTRACT

The recombinase activating genes RAG-1 and RAG-2 operate together to activate V(D)J recombination, and thus play an essential role in the generation of immune system diversity. As a first step in understanding the function of the RAG-2 protein, we have tested a series of deletion and insertion mutations for their ability to induce V(D)J joining of a variety of model substrates. Mutants were assayed for their ability to induce deletional and inversional V(D)J joining, thereby testing their proficiency at forming both signal and coding joints, and, in some cases, for their ability to carry out recombination of both extrachromosomal and integrated recombination substrates. All these reactions were affected similarly by any one mutation. Although the RAG-2 protein shows extensive evolutionary conservation across its length, we found that the carboxy-terminal portion of RAG-2, including an acidic region, is dispensable for all forms of recombination tested. In contrast, all mutations we created in the N-terminal region severely decreased recombination. Thus, the core active region required for V(D)J recombination is confined to the first three-quarters of the RAG-2 protein.

INTRODUCTION

Developing B and T cells possess a unique recombinase activity that assembles antibody and T cell receptor genes from component gene segments. Recombination signal sequences, that flank all recombinationally competent gene segments, serve to target the recombinase activity to these sites (reviewed in 1). The recombination activating genes, RAG-1 and RAG-2, were identified as the only lymphoid-specific factors required to induce this V(D)J recombinase activity in fibroblasts (2,3). These two genes share no sequence similarity, and expression of both genes is required to induce recombination in non-lymphoid cells (3). In addition, mice with a disruption of either gene fail to carry out rearrangement of their T cell receptor or antibody gene segments, and no mature T or B cells develop (4,5).

While cell culture and transgenic mouse experiments demonstrate the critical dependence of V(D)J recombination on the expression of RAG-1 and RAG-2, the precise role of these proteins remains unclear. RAG-1 and RAG-2 might encode components of the recombinase itself, or they might be factors which regulate recombinase expression. While some potentially interesting sequence motifs have been noted in RAG-1, mutational analysis has demonstrated that those regions are not important for at least one RAG-1 function: the ability to induce deletional recombination resulting in signal joint formation in an extrachromosomal substrate (6–8). Indeed, RAG-1 deletion mutants lacking the first 383 amino acids were still almost fully active in this assay. In RAG-2, sequence analysis has not revealed any striking similarity to any other protein and the only region of note is a highly acidic stretch located near the C-terminus. To date, no biochemical function has been assigned to either RAG protein.

As an initial step toward understanding the role of RAG-2, we have carried out a mutational analysis of the protein to determine which regions are required for V(D)J joining.The function of RAG-2 can be assessed by measuring recombinase activity induced in fibroblasts. The complexity of the recombination event provides a number of different assays for RAG function. V(D)J joining results in the formation of two novel junctions, a signal joint and a coding joint, and reporter plasmids have been designed to look at the formation of one or both (9). Furthermore, depending on the orientation of the recombination signal sequences, the recombinase machinery is capable of carrying out either a deletional or inversional recombination event. Analysis of other site-specific recombination systems such as the DNA resolvases or invertases indicates that such events are often mechanistically distinct (reviewed in 10). Finally, recombination can be carried out on either integrated or extrachromosomal substrates.

Mutations made in the amino-terminus of the protein severely disrupted recombination of extrachromosomal substrates. However, unexpectedly, we have found that the C-terminal 25% of RAG-2 is not required for V(D)J recombination of any of the substrates tested here.

*To whom correspondence should be addressed at: Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA

MATERIALS AND METHODS

Construction of RAG-2 mutants

The starting molecule for all RAG-2 mutants, except the C-terminal deletion constructs, contained the entire RAG-2 coding sequence and one copy of a C-terminal influenza haemagglutinin epitope tag (derived from LG4.3 described in 11), in the CDM8 expression vector (12). Insertion mutants (R2CC19, R2CC20, R2CC21, R2CC8, R2CC6, R2CC7) were constructed by ligation of a linker into a *Hin*fI site. These mutations were sequenced to confirm that the oligonucleotide restored the degenerate *Hin*fI site on either side of a CGGCCG insertion. R2CC3, R2CC5, and R2CC4 were constructed by PCR site-directed mutagenesis with distinct oligonucleotides designed to insert or substitute specific residues (see Table 2). The starting molecule for the C-terminal deletion mutants was LG-1 (11) which contains the entire RAG-2 coding sequence with the 5′ end modified to begin with one copy of the influenza HA epitope tag (13). The deletions were constructed using unique upstream restriction enzyme sites and a common downstream *Eco*RV site. An oligonucleotide with translational stop codons in all three frames was introduced adjacent to the *Eco*RV site in all mutants except R2CC10 and R2CC11. The following additional amino acids were added at the end of RAG-2 sequence as a consequence of this mutagenesis: for R2CC10, none; for R2CC11, DRIPAARGIPLVLERPRL; for R2CC14, IEFLSS; for R2CC15, IEFLSS; for R2CC16, IEFLSS; for R2CC24, ESNS.

Extrachromosomal substrate assay

This assay was performed as previously described (3,9). HeLa cells were transiently transfected with 6.0 μg of M2CD7 (a RAG-1 expression plasmid), 4.8 μg of the appropriate RAG-2 expression construct, and 5 μg of a reporter plasmid [either pJH200 or pJH299 (9)] using a standard calcium phosphate procedure. Transfected cells were grown at 37°C, 5% CO_2 for 48 h and plasmid DNA was recovered by a modified alkaline lysis procedure. Recovered DNA was digested with *Dpn*I to remove any unreplicated plasmid stuck to the outside of cells which was not accessible for recombination. DNA was then electroporated into *E.coli* strain MC1061 and plated on either ampicillin or chloramphenicol/ampicillin plates. All substrates express ampicillin resistance constitutively while resistance to chloramphenicol is acquired as a result of V(D)J joining. $Chlor^r/amp^r$ colonies were counted after 15 h at 37°C and then transferred to a membrane and probed with the SJ2 oligonucleotide (3) to confirm proper V(D)J joining events. SJ2 hybridizes to sequence across an unmodified signal joint. Recombination frequency reflects the number of proper V(D)J joining events (SJ2 positive) divided by the number of total plasmids (number of Amp resistant colonies).

Integrated substrate assay

A fibroblast cell line (3TGR) (14), carrying the integrated substrate DGR, was transfected with RAG-1 DNA, pSV2-His linearized with *Eco*RI, and RAG-2 DNA as indicated. The

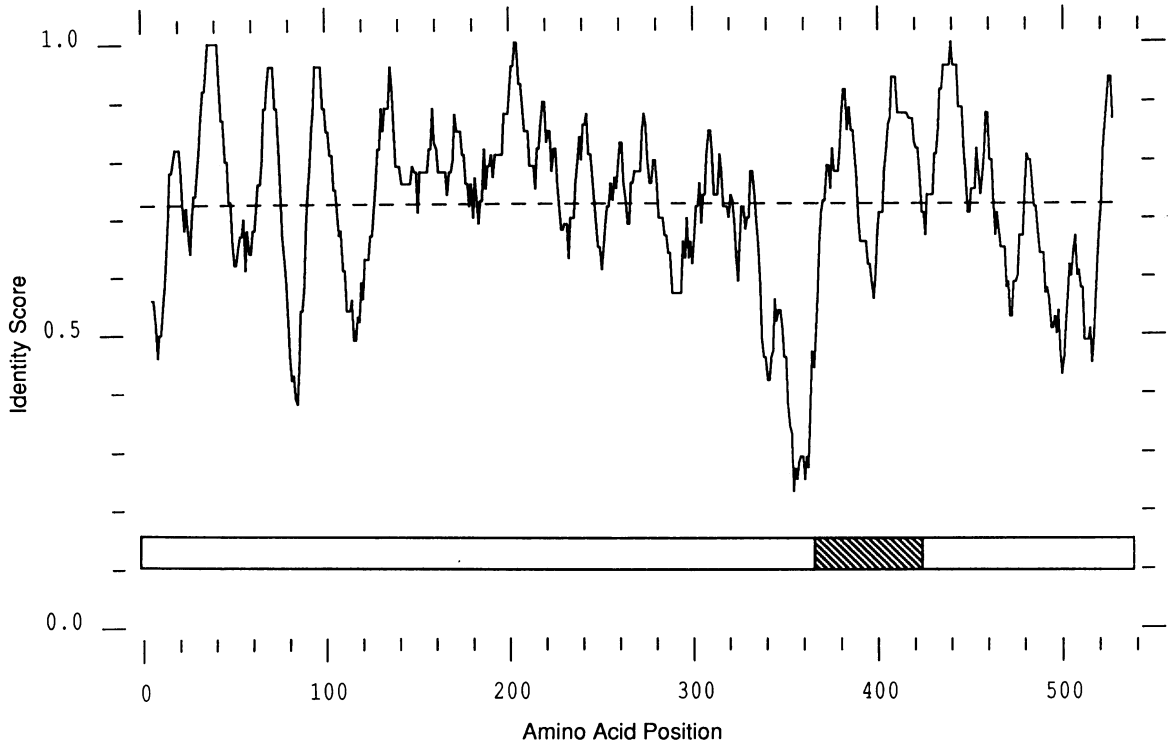

Figure 1. Identity of RAG-2 amino acid sequences from human, mouse, chicken, rabbit, and *Xenopus*. Sequences were compared using the GCG plotsimilarity program (19), in which all identical sequence comparisons are given a value of 1, and all other comparisons are given a value of 0. This graph represents these values averaged over a window of 10 amino acids and is a comparison of the following Genbank sequences: human (M94633), mouse (M33828), chicken (M58531), rabbit (M77667), and *Xenopus* (L19325). The bar at the bootom of the figure represents the RAG-2 sequence with the shaded box indicating the posititon of the acidic region (amino acid residues 352−410).

transfections were performed by standard calcium phosphate procedure, with a glycerol shock at 4–5 h. The cells were refed at 24 h. At 48 h the cells were split 1:2. At 72 h the cells were split into selective media. Selection was continued for 7–10 days, and medium was replaced every 3 days. Cells were stained with 2% mehylene blue and the number of colonies was counted.

RESULTS

The RAG-2 genes from five different species [human (15), mouse (3), rabbit (16), chicken (17) and *Xenopus* (18)] have been cloned and sequenced. We compared the amino acid conservation across these species using the Genetics Computer Group plot similarity program (19). The results of this comparison, diagrammed in Figure 1, show that these RAG-2 proteins share a high degree of sequence identity (an average of approximately 70%) across their entire length with no clear boundaries between more and less well-conserved regions. Thus, a series of deletion and insertion mutations was constructed across the length of RAG-2. The location of amino acid alterations are indicated in Figure 2. The ability of each mutant to support V(D)J joining activity in fibroblasts was measured in a standard transient transfection assay (see Materials and Methods). Two reporter constructs were used to allow separate measurements of signal joint formation upon deletional rearrangement (pJH200) or inversional recombination with retention of both signal and coding joints (pJH299) (9).

In order to insure that we would be able to detect both increases and decreases in recombinase activity, we examined the effect of changing the amount of transfected RAG-2 DNA on recombination frequency. With the amount of RAG-1 DNA held at 6.0 μg, recombination increased roughly in proportion to the amount of transfected RAG-2 DNA (Table 1), indicating that under these conditions, the assay was responsive to RAG-2 levels.

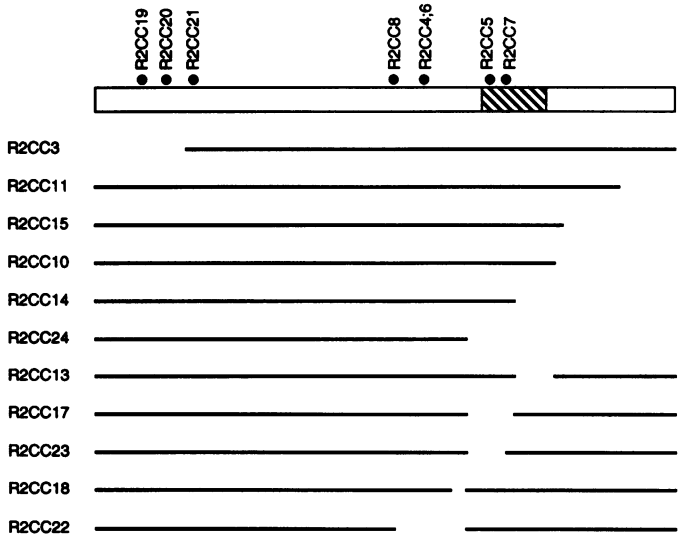

Figure 2. RAG-2 mutations. The open bar at the top represents the full-length RAG-2 sequence; the striped box within it represents the acidic region. The positions of all insertion mutations and R2CC5 (a substitution of three neutral amino acids for three conserved acidic residues) are shown as filled circles above the bar. The sequence remaining in each deletion mutant is drawn out linearly below the bar (see text and Table 2 for more detailed descriptions of each mutation).

Based on these results, we chose to use 6.0 μg of RAG-1 DNA and 4.8 μg of RAG-2 DNA for our mutational analysis.

The recombination frequencies measured for each mutant along with the number of plasmids screened are listed in Table 2. The results have been normalized to the wild-type recombination frequencies obtained in parallel transfections on the same day. While the absolute recombination frequencies varied between sets of transfections, the ratio of mutant to wild-type remained constant. As has been noted previously, the absolute recombination frequency measured for the inversional substrate pJH299 is typically 10-fold lower than for a deletional substrate that retains either a signal (pJH200) or coding (pJH290) joint.

Three insertional mutants were tested within the first 90 amino acids of RAG-2. Each of these mutations severely decreased recombination of both the deletional and inversional reporter constructs. Additionally, the mutant R2CC3, in which amino acids 2–81 are deleted, had no detectable activity in our assays. These results suggested that the N-terminus of RAG-2 is not tolerant of mutational alteration and that this region is necessary for RAG-2 function.

In the central region of the protein, five insertion mutations were tested. Four of these resulted in modest decreases in recombination frequency, while the fifth insertion, R2CC8 (inserting GRI at residue 257) reduced the frequency at least 1000-fold.

The effect of a number of internal deletions was also measured. R2CC18, 22, and 23 decreased recombination of both substrates below detectable levels. One additional deletion mutant, R2CC17, which is a slightly more extensive deletion in the same region of R2CC23, did retain very low but detectable levels of activity, indicating that sequences in this region are not absolutely required for function. Further, when the acidic region of RAG-2 was largely deleted (residues 382–409 in R2CC13), or when three conserved acidic amino acids outside the span of this deletion were changed to neutral residues (R2CC5), recombination was not significantly affected.

Because the acidic region lies close to the C-terminus, a series of deletions extending from this end of RAG-2 was generated. This region of RAG-2 is as well conserved as the N-terminal portions, thus our expectation was that such deletions would severely inhibit recombination activity. Surprisingly, a deletion mutant lacking amino acids 382–527 (R2CC14) was still fully functional in our assay. Indeed, recombination frequencies for this mutant consistently appeared elevated compared to wild-type, although it is unclear whether this small increase in recombinase activity reflects an increase in the amount of RAG-2 protein or

Table 1. The dependence of recombination frequency on amount of transfected RAG-2 expression plasmid

RAG-2 ug	RAG-1 ug	#Amp plasmids screened x100	normalized Recomb. Freq.
9.6	6.1	221	3.5
7.2	6.1	364	1.2
4.8	6.1	190	1.0
0.3	6.1	585	0.2

The recombination frequency for a deletional substrate (pJH200) in Hela cells was measured following transient transfection with the indicated amounts of expression plasmid. The total amount of DNA included in each transfection was held constant, with the addition of the appropriate amount of CDM8 vector DNA to compensate for changes in the amount of RAG-2 plasmid. All recombination frequencies represent the average of two or more duplicate transfections. The proper structure of the plasmid product was confirmed by hybridization with an oligonucleotide specific for the signal junction (see Materials and Methods).

an increase in its specific activity. As expected, two mutations that deleted less of the C-terminus, R2CC10 (482–527) and R2CC11 (419–527) were also functional. A deletion of residues further upstream, of amino acids 342–381 (R2CC17), drastically reduced recombinase activity, although recombinants were still generated. The most extensive deletion of amino acids 342–527 (R2CC24), which combines the deletions in R2CC14 and R2CC17, decreased recombination frequency below the level of detection.

The mechanistic constraints on carrying out recombination of a plasmid substrate are likely to be different from those for a substrate that is integrated into the chromosome. Thus, in an effort to find a function for the C-terminal one-fourth of RAG-2, we asked whether a mutant lacking this region could activate the recombination of an integrated substrate. A 3T3 cell line [3TGR (14)] containing two integrated copies of an inversional V(D)J recombination substrate (DGR) was co-transfected with a RAG-1 expression construct, pSV2-His as a selectable marker, and either vector alone or constructs expressing wild-type RAG-2 or a RAG-2 deletion mutant. V(D)J recombination of the reporter substrate results in expression of a GPT gene rendering the cells resistant to mycophenolic acid (MPA). The pSV2-His construct served as an internal control for transfection efficiency. At 72 h after transfection, cells were subject to either MPA or histidinol selection. The ratio of MPA^R colonies to His^R colonies is a measure of the level of recombinase activity transiently induced in the fibroblasts. The results of this experiment, summarized in Table 3, paralleled the results seen with the extrachromosomal substrates. The most extensive functional RAG-2 C-terminal deletion mutant, R2CC14, supported the recombination of the integrated substrate at nearly wild-type levels. A still larger deletion R2CC24, not functional in the plasmid assay, was not functional with the integrated substrate either. Thus, the C-terminal one-fourth of RAG-2 is also dispensable for recombination of integrated substrates.

Table 2. The recombination activity of RAG-2 mutants

	RAG-2 SEQ	Recomb. Freq. pJH200	#Amp plasmids screened x100	Recomb. Freq. pJH299	#Amp plasmids screened x100
R2RCD2		1.000		1.000	
R2CC3	del 2-81, ins AS	<0.001	1060	<0.001	2590
R2CC19	ins GRV 46	<0.001	4490	<0.004	12950
R2CC20	ins SAD 66	<0.001	13750	<0.004	17820
R2CC21	ins SAD 86	<0.001	20120	<0.004	19880
R2CC8	ins GRI 257	0.029	13680	0.027	2140
R2CC4	ins GG 302	0.409	1300	0.125	3188
R2CC6	ins TRP 304	0.464	201	0.239	255
R2CC7	ins SAD 375	0.846	148	0.368	132
R2CC5	EED 352 QQN	0.557	160	0.412	299
R2CC13	del 382-409, ins A	2.113	152	1.633	88
R2CC18	del 330-341	<0.001	12740	<0.004	23780
R2CC22	del257-342, ins GP	<0.001	3630	<0.005	1370
R2CC23	del 343-373	<0.001	12060	<0.004	16720
R2CC17	del 342-381	0.001	14600	<0.03	9460
R2CC24	del 342-527	<0.003	2050	<0.06	1720
R2CC14	del 383-527	6.787	2830	2.084	6730
R2CC10	del 419-527	1.122	355	0.664	1172
R2CC15	del 426-527	0.284	4980	0.259	10540
R2CC11	del 482-527	0.609	136	1.011	118

The effect of mutations in RAG-2 on recombination frequency of a deletional (pJH200) and inversional (pJH299) recombination substrate was measured in a transient transfection assay (see Materials and Methods). The total number of plasmid substrates screened is indicated as the number of ampicillin resistant colonies. Recombination frequencies are normalized to a value of 1 for wild-type RAG-2. Typical wild-type frequencies ranged from 1% to 6% in independent experiments. The amino acid alterations in each mutant are noted in one letter code in the second column. del = deletion, amino acid numbers listed are inclusive; ins = insertion, number indicates the position of the first inserted amino acid; for R2CC5 amino acids 352–354 were changed from EED to QQN.

DISCUSSION

The assembly of antigen receptor molecules by V(D)J recombination is a complex series of reactions subject to a variety of levels of regulation. By being able to induce at will the recombination of artificial substrates in non-lymphoid cells it is possible to explore the requirements of the basic V(D)J recombination reaction unencumbered by most regulatory modulation. By choosing the plasmid substrates tested here, we have been able to show that all aspects of the basic recombination reaction, the formation of signal joints by deletion and the formation of both signal and coding joints by inversional recombination, require the same regions of RAG-2. Mutations in the first 90 amino acids of RAG-2 severely inhibit all of these activities while the C-terminal 145 amino acids (at a minimum) are dispensable for these activities and for the recombination of integrated substrates (similar results have been observed by others; see accompanying paper by Sadofsky and Gellert). Furthermore, mutations that affect deletion affect inversion to roughly the same degree. Thus, the core region of RAG-2 required for the basic recombination reaction must be found complete in the first 382 amino acids of RAG-2.

In the absence of more detailed knowledge of the structure or function of RAG-2, it is necessary to rely on the pattern of mutational effects to identify essential and non-essential regions. The clustering of inactive mutants near the N-terminus makes it likely this region contains features essential for recombination.

In addition to defining a truncated core sequence required for RAG-2 activity, we have also shown that the large acidic stretch in RAG-2 is not required for its activity. Previous work has shown that deletion of the acidic region from 374–414 or the alteration of three conserved acidic residues to neutral does not

Table 3. The activity of RAG-2 mutants on an integrated recombination substrate

RAG-2	deletion	MPA^R Colonies	His^R Colonies	GPT/His ratio
R2RCD2	wild type	728	1178	0.62
none		0	478	0.00
R2CC24	342-527	0	1888	0.00
R2CC14	382-527	289	1568	0.18
R2CC13	382-409	880	3226	0.27

RAG-2 mutants are as described in Table 2. NIH 3T3 fibroblasts with an integrated recombination substrate (14) were co-transfected with the indicated RAG-2 expression constructs, wild-type RAG-1 expression plasmid and linearized pSV2-His as a co-selectable marker. The recombination frequency represents the ratio of the number of mycophenolic acid (MPA) resistant colonies (indicative of rearrangement) to the number of histidinol (His) resistant colonies (a reflection of transfection efficiency).

affect the ability of RAG-2 to induce the formation of signal joint deletions (6). The work presented here shows further that the surrounding region can also be removed with no effect on any of the recombination reactions measured.

Another site of interest that has been identified toward the C-terminus of RAG-2 is threonine 490 which had been shown to be subject to phosphorylation *in vivo* (20). Phosphorylation of this site appears to affect protein stability, indeed, substitution of an alanine at this site resulted in an almost 20-fold increase in protein half-life, but with no accompanying increase in recombinase activity (20). The removal of threonine 490 in our least extensive C-terminal deletion mutant (R2CC11) resulted in no appreciable increase in recombination frequency, in keeping with these previous results.

The dispensability of a large region of RAG-2 is reminiscent of results reported for the RAG-1 protein; a large N-terminal region of RAG-1 was found to be unnecessary for recombination of a deletional signal joint substrate (pJH200). However, in the RAG-1 gene there appears to be a natural boundary delimiting the essential segment. The conservation of the dispensable region of RAG-1 among the same five animal species is not nearly as high as in the required region (8). This is unlike the situation in the RAG-2 C-terminus, where there are blocks of highly conserved sequence throughout. The RAG-1 study left open the possibility that the N-terminus would be required for coding joint formation or recombination of integrated substrates. However, it is intriguing to speculate that both RAG proteins are composed of core regions required for recombination and other regions with as yet undefined functions. While these dispensable regions might have no function in lymphoid development, it is also possible that these regions will be involved in the regulation of recombinase activity or in the regulated recombination observed at endogenous loci.

ACKNOWLEDGEMENTS

The authors wish to thank M.J.Sadofsky and M.Gellert for communicating information prior to publication. We also wish to thank R.E.Kingston for critical reading of the manuscript. This work was supported by a grant from the NIH GM48026 (M.A.O.) and the PEW Scholar's program (M.A.O.)

REFERENCES

1. Gellert, M. (1992) *Annu. Rev. Genet.*, **22**, 425–446.
2. Schatz, D. G., Oettinger, M. A. and Baltimore, D. (1989) *Cell*, **59**, 1035–1048.
3. Oettinger, M. A., Schatz, D. G., Gorka, C.and Baltimore, D. (1990) *Science*, **248**, 1517–1523.
4. Shinkai, Y., Rathbun, G., Lam, K.-P., Oltz, E. M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., Stall, A. M.and Alt, F. W. (1992) *Cell*, **68**, 855–86.
5. Mombaerts, P., Iacomini, J., Johnson, R. S., Herrup, K., Tonegawa, S. and Papaioannou, V. E. (1992) *Cell*, **68**, 869–877.
6. Silver, D. P., Spanopolou, E., Mulligan, R. C.and Baltimore, D. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6100–6104.
7. Kallenbach, S., Brinkmann, T.and Rougeon, F. (1993) *Int. Immunol.*, **5**, 231–232.
8. Sadofsky, M. J., Hesse, J. E., McBlane, J. F.and Gellert, M. (1993) *Nucleic Acids Res.*, **21**, 5644–5650.
9. Hesse, J. E., Lieber, M. R., Gellert, M.and Mizuuchi, K. (1987) *Cell*, **49**, 775–783.
10. Craig, N. L. (1988) *Annu. Rev. Genet.*, **22**, 77–105.
11. Cuomo, C. A., Kirch, S. A., Gyuris, J., Brent, R.and Oettinger, M. A. (1994) *Proc. Natl. Acad. Sci.*, in press.
12. Seed, B. (1987) *Nature*, **329**, 840–842.
13. Colman, P. M., Laver, W. G., Varghese, J. N., Baker, A. T., Tulloch, P. A., Air, G. M.and Webster, R. G. (1987) *Nature*, **326**, 358–363.
14. Schatz, D. G.and Baltimore, D. (1988) *Cell*, **53**, 107–115.
15. Ichihara, Y., Hirai, M.and Kurosawa, Y. (1993) *Immunol. Lett.*, **33**, 277–284.
16. Fuschiotti, P., Harindranath, N., Mage, R. G., McCormack, W. T., Dhanarajan, P.and Roux, K. H. (1993) *Molec. Immunol.*, **30**, 1021–1032.
17. Carlson, L. M., Oettinger, M. A., Schatz, D. G., Masteller, E. L., Hurley, E. A., McCormack, W. T., Baltimore, D.and Thompson, C. B. (1991) *Cell*, **64**, 201–208.
18. Greenhalgh, P., Olesen, C.and Steiner, L. A. (1993) *J. Immunol.*, **151**, 3100–3110.
19. Genetics Computer Group (1991) GCG plotsimilarity program. Madison, Wisconsin.
20. Lin, W.-C.and Desiderio, S. (1993) *Science*, **260**, 953–959.