Volume 16 Number 16 1988 Nucleic Acids Research

Use of RecA protein to enrich for homologous genes in a genomic library

Behnaz Taidi-Laskowski, Dolly Tyan[1], Saul M.Honigberg[2], Charles R.Radding[2] and F.Carl Grumet*

Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, [1]University of California, Los Angeles, CA and [2]Department of Human Genetics, Yale University School of Medicine, New Haven, CT, USA

Received March 25, 1988; Revised and Accepted 22 July, 1988

ABSTRACT

RecA protein-coated probe has been utilized to enrich genomic digests for desired genes in order to facilitate cloning from genomic libraries. Using a previously cloned HLA-B27 gene as the recA-coated enrichment probe, we obtained a mean 108x increase in the ratio of specific to nonspecific plaques in lambda libraries screened for B27 variant alleles of estimated 99% homology to the probe. Class I genes of lesser homology were less enriched: 6.7x for non-B27 genes of estimated >95% homology and 3.7x for other Class I genes of >80% homology. Loss of genomic DNA during the enrichment procedure can, however, restrict application of this technique whenever starting genomic DNA is very limited. Nevertheless, the impressive reduction in cloning effort and material makes recA enrichment a useful new tool for cloning homologous genes from genomic DNA.

INTRODUCTION

In generating a useful recombinant DNA library, enrichment of the starting digest of genomic DNA for genes of interest should allow one to screen a smaller population of clones with a corresponding reduction in time and effort. A new method for enriching for specific sequences in complex mixtures of DNA was devised by Honigberg et al (1), using recA protein from E. coli. A related method has been described by Rigas et al (2). RecA protein has the ability to promote homologous pairing between single stranded DNA (ssDNA) and duplex DNA (dsDNA). Nucleoprotein filaments, produced by the binding of recA protein to ssDNA, form networks with the dsDNA, concentrating the DNA and consequently speeding the search for homologous sequences. When homologous pairing occurs, joints are formed. Since such heteroduplex joints are more stable than the non-specific contacts of recA nucleoprotein filaments with heterologous duplex DNA, the latter can be selec-

tively washed out of the networks and discarded in a supernatant fraction.

Such enrichment of whole genome digests prior to cloning should be particularly useful in investigations of highly polymorphic families of closely related genes. One such example is the HLA system, in which certain alleles show strong linkage disequilibrium with particular diseases (e.g.,HLA-B27 and spondyloarthropathies), motivating study of selected variant and related HLA alleles (3). In addition to the extensive polymorphism at each of the classical, serologically defined HLA Class I (HLA-A, B, and C) and Class II (HLA-DQ, DR) loci, there exist a substantial number of "silent", closely related loci (e.g., estimated 20-40 additional Class I loci) further complicating cloning and characterizating of alleles of interest (4). During investigations directed toward cloning variant B27 alleles in diseased and normal populations, it seemed reasonable to evaluate the recA enrichment concept. The present report therefore describes the first successful application of the recA enrichment method to genomic cloning.

MATERIAL AND METHODS

Unless otherwise indicated, common techniques, reagents and protocols were as described by Maniatis, Fritsch & Sambrook (5). Enzymes were obtained from New England Biolabs, (Beverly, MA) and used according to manufacturers directions.

Preparation of Size Selected Genomic DNA.

Genomic DNA (200mcg) from two individuals, HS (HLA-A2, A25; B8, B27, Bw4, Bw6; Cw1, Cw7) and GDel (HLA-A2, A3; B18, B27, Bw4, Bw6; Cw1, Cw-) was digested with EcoRI and electrophoresed (30 volts, 20cm) overnight on a 0.6% agarose gel (SeaPlaque or SeaKem GTG agarose, FMC, Rockland, ME). Size selected (6.5-9.5kb) and ethidium bromide stained fragments were cut out and recovered either by gel melting or electroelution. The DNA was precipitattated with 3M sodium acetate and ethanol and stored at 4^{o}C for up to 23 months

Preparation of Homologous DNA Enrichment Probe.

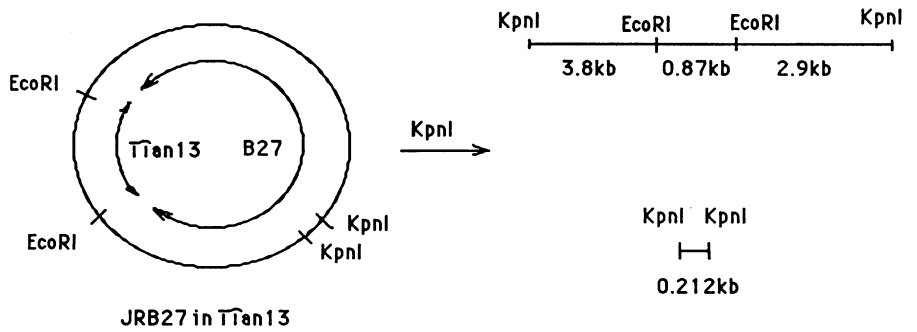
JRB27, a genomic HLA-B27 gene cloned in pian13 vector (6), was digested with KpnI. This enzyme cuts twice, within the B27 gene,

yielding a linear 7.6kb fragment with pianl3 in the middle and the B27 fragments on each end, and a minor 212bp fragment (Fig 1).

Enrichment with RecA Protein.

The enrichment was achieved in three stages using a modification of the procedure previously described (1). First, single-stranded DNA/recA protein filament was prepared for use as an enrichment probe. The single-stranded DNA was generated by denaturing linear duplex DNA from the JRB27 clone. Second, the enrichment probe was incubated with duplex DNA from a genomic digest; joint molecules were formed between the probe and the homologous duplex DNA molecules in the mixture, which were the targets of the probe. In the final stage, the probe was then precipitated, along with its paired target genomic fragments while the non-homologous DNA remained in the supernatant.

To prepare the enrichment probe, the linearized JRB27 plasmid clone (12mcg/ml) was denatured in a volume of approximately 70mcl in a 500 mcl microfuge tube. The tube was sealed with parafilm and heated at 94°C in a circulating water bath for four minutes The denatured DNA was immediately cooled in an ice water bath for one minute. A solution (4°C) was added to the DNA to raise the volume of the reaction mixture to 190mcl, and the concentrations to 20mM Tris-HCl (pH 7.5), 1.4mM DTT, 7mcg/ml BSA, 1.0mM $MgCl_2$, 1.2mM ATP,6mM phosphocreatine and 10U/ml creatine phosphokinase. To coat the probe, recA protein (7) was added to a concentration



Legend for Figure 1

KpnI digestion of JRB27 in pianl3. This linearized ds construct was then denatured, and as ssDNA, complexed with recA to form the enrichment probe.

of 0.23mg/ml and the mixture was incubated for ten minutes at 37°C. $MgCl_2$ was added (to 12mM),the reactions were next mixed on a Vortex mixer (Scientific Products, Evanston, IL) briefly and single-stranded DNA binding protein was immediately added to a concentration of 0.013mg/ml. After five minutes at 37°C, the duplex DNA from a genomic digest (1mcg) was added commonly in a volume of 10mcl. The total volume of the reaction mixture was now 200mcl.

After twenty minutes of incubation at 37°C, single stranded DNA from phage M13 (8) was added to the mixture at a concentration of 5mcg/ml. By making multiple non-specific contacts with the enrichment probe, this carrier DNA allows the selective precipitation of the probe molecules and the attached duplex DNA. After an incubation of four minutes, NaCl (to 75mM) was added to the reaction mixture to dissociate most of the non-specific contacts between the probe and non-homologous duplex DNA (1,9). The mixture was incubated for an additional four minutes at 37°C and then centrifuged in an Eppendorf microfuge (Brinkman Instruments, Inc., Westbury, NJ) for three minutes at room temperature. The supernatant was removed with an Eppendorf pipetman and discarded, leaving a few microliters with the pellet which was often invisible. Care was taken to avoid touching the bottom of the tube with the pipet tip.

The pellet was resuspended in 40mcl of a solution of 5mM ATP (4°C) by vortexing the tube at low speed for thirty seconds and then placing it in an ice water bath for thirty seconds and repeating these alternating steps for a total of three minutes. A solution was added to raise the concentration of the final mixture to 20mM Tris-HCl, 12mM $mgCl_2$, 6mM phosphocreatine, and 10U/ml creatine phosphokinase and the volume of the reaction mixture to 200mcl. The mixture was incubated at 37°C for four minutes, NaCl was added to 75mM and the incubation was continued for a further four minutes.

The mixture was again centrifuged and the supernatant discarded. The pellet was resuspended and incubated as described in the previous paragraph, and then precipitated once more. This final pellet was suspended in 300mcg/ml proteinase K (Boehringer - Mannheim, Indianapolis, IN) in a volume of about 5mcl and

incubated for fifteen minutes at 37°C and fifteen minutes at 50°C to digest the protein and disrupt joint molecules.

Control reactions.

Following the procedure just described, control reactions containing labeled duplex DNA were used to estimate the amount of homologous and non-homologous DNA that precipitates in each cycle. The probe in these reactions was prepared by denaturing unlabeled duplex DNA from phage M13, linearized at the HincII site. In the reaction to measure the recovery of homologous DNA, the duplex DNA was a mixture of unlabeled DNA from a PvuII digest from phage lambda and (^{3}H)-DNA from either phage M13 or phage G4 which is not homologous to the M13 DNA enrichment probe; the lambda DNA was in 100-fold excess. In the reaction to measure the level of non-homologous DNA precipitating, the duplex (^{3}H)-DNA was from phage G4 linearized at the BglI site. The final precipitate was resuspended and counted in a scintillation counter as was each supernatant.

Preparation of the Classical Unenriched Genomic Library.

One mcg of size selected genomic DNA was ligated with 5mcg lambda gtWES EcoRI arms (BRL, Gaithersburg, MD) overnight at 15°C. The ligated material (24mcl) was packaged at 4mcl per package mix (Gigapack Plus, Stratagene Cloning Systems, La Jolla, CA). Packaged phage was pooled and titrated in LE392 host cells on NZCYM agarose plates and the library was plated at approximately 5x10^3 plaque forming units (pfu)/plate. Plates were incubated at 37°C for 10-12 hours and stored at 4°C for 4-6 hours. Replicate plaque lifts were made onto nitrocellulose filters (BA85, 0.45mcm, 82.5mm, Schleicher and Schuell, Keene, NH), so that each filter was probed only once. Filters were probed as described (6) with nick translated pJY150 [the whole genomic HLA-B7 gene in pBR328 (4)] or purified EI7, (a fragment from the HLA-B7 gene, that functions as a HLA-B locus specific probe in Southern blotting (11)]. Under the high stringency wash conditions used to probe these plaque lifts, EI7 hybridizes primarily to HLA-B locus, but also shows some cross-hybridization to HLA-C locus genes, whereas pJY150 hybridizes to all HLA Class I genes. After washing, filters were then exposed to XAR-5 film (Kodak, Rochester, NY) for 3-16 hours. Standard plaque purifica-

tion and lysate preparation were performed and Class I plaques identified were plaque purified, cloned and run on Southern blots for further characterization, e.g., to identify specific B locus alleles (11).

Preparation of the RecA Protein Enriched Library.

The enriched genomic DNA was estimated from control reactions to range from 3.75-12.0ng and to represent about 1/2-3/4 of input DNA homologous to B27. A 5-8mcl solution containing 1-1.6mcg homologous enrichment probe DNA, 1mcg of single stranded, circular M13 carrier DNA, 10-15mM Tris-HCl (pH7.5), 5-6mM $MgCl_2$, 20-25mM NaCl, and 300mcg/ml proteinase K was extracted 1-2x with phenol/chloroform (1:1). This DNA was ligated to 5mcg of lambda gtWES EcoRI arms as described above for the unenriched DNA. Subsequent preparative steps (packaging, plating, plaque lifting and probing) were also performed as described above, again using the F^- host strain LE392 to avoid growth of M13.

Southern Blots.

These were performed by our modifications (5) of standard procedures. Under the stringent washing condition used for Southern blots, the pJY150 probe hybridizes to all HLA Class I genes while the EI7 probe is HLA-B locus specific. With this method, B27 and other B locus and C locus clones were readily identifiable by their characteristic restriction fragment length polymorphism (RFLP) pattern with EcoRI and TaqI and the EI7 probe.

RESULTS

In order to be certain that DNA of the enrichment probe JRB27 would not be cloned and confused with the variant B27 genes sought from the genomic libraries, target B27 genes were selected that could be distinguished unequivocally from probe by RFLP. When EI7 is used to probe whole genomic DNA Southern blots under high stringency conditions, it yields only one band per B locus allele. As seen in Table I, in each of the three individuals involved, the two B locus alleles are readily differentiated and the two target B27 genes could easily be distinguished from JRB27.

B27 enriched (using the JRB27 enrichment probe) and unenriched genomic libraries were prepared from HS and GDel as de-

Table I.
RFLP PATTERNS OF HLA-B GENES STUDIED

DNA donor**	HLA-B Locus alleles	HLA-B allele RFLP band* size (kb) with enzyme: EcoRI	EcoRI + TaqI	EcoRI + BstEII
JR***	B44	6.9	3.4	2.2
	B27	6.9	3.4	4.7
HS	B8	6.9	4.4	
	B27	6.9	2.5	
GDel	B18	6.9	3.4	
	B27	6.9	2.5	

* In all cases, definite assignment of each band to a specific HLA-B locus allele was made through family studies (9).
** DNA from each donor was digested with the enzymes noted and Southern blots were probed with EI7.
*** Previously, the B27 gene from JR had been cloned and that gene, designated JRB27, in pianl3, was used as the enrichment probe in this report.

scribed. Generally, the entire library was plated, however for the unenriched GDel library only 1.7% was plated, the remainder being committed to other experiments. HLA Class I pfu were plaque purified and cloned, and HLA gene assignment was made on the basis of RFLP patterns of Southern blots. Yields for each library are shown and compared in Tables II and III. For HS DNA, the relative frequency of B27 gene clones was increased by enrichment from 1.7×10^{-6} to 2.7×10^{-4}. Thus the specific/nonspecific pfu ratio was increased 158.8 fold for these B27 gene clones. Among the group of strongly homologous but non-B27, other EI7(+) pfu, the second B locus allele, B8, of HS was not found in the

Table II
CONSTRUCTION AND CHARACTERIZATION OF GENOMIC LIBRARIES

GENOMIC DNA								pfu		
								pJY150 (+)		
								EI7(+)		EI7(-)
Donor	storage (days)	Start* ng	Enriched	Estimated final* ng ligated in gtWES	% of library plated	Number of plates **	Total	B27	other ***	
HS	660	2500	+	3.75	100	18	8.1×10^4	22	5	9
		1000	-	1000	100	240	1.2×10^6	2	2	n.d.
GDel	14	4000	+	12.0	100	10	4.0×10^4	9	16	6
		1000	-	1000	1.7	50	2.5×10^5	1	15	10

* For unenriched libraries, all of the starting genomic (start) DNA was ligated as phage insert. For enriched libraries, the small percentage of the starting genomic DNA, greatly enriched for genes homologous to JRB27, left after the recA enrichment procedure was ligated as phage insert.
** All libraries plated at ~5×10^3 pfu/plate.
*** RFLP of these cloned plaques showed only 13% to be non-B locus (i.e., Cw1 or Cw7) genes.
n.d. = Not done.

Table III.
RELATIVE ENRICHMENT OF GENOMIC LIBRARIES

	GENOMIC DNA		INSERTION EFFICIENCY	RATIO OF SPECIFIC/NONSPECIFIC pfu		
	Donor	Enriched	Total pfu / μg insert DNA	B27 / total	Other* EI7(+) / total	pJY150(+), EI7(-) / total
A. pfu ratios	HS	+	2.2×10^{7}	2.7×10^{-4}	6.2×10^{-5}	1.1×10^{-4}
		-	1.2×10^{6}	1.7×10^{-6}	1.7×10^{-6}	n.d.**
	GDel	+	3.3×10^{6}	2.3×10^{-4}	4.0×10^{-4}	1.5×10^{-4}
		-	1.5×10^{7}	4.0×10^{-6}	6.0×10^{-5}	4.0×10^{-5}
B. Relative enrichment***	HS	+ vs -		158.8x	36.5x	-
	GDel	+ vs -		57.5x	6.7x	3.8x
	mean			~108x	~22x	~4x

* By RFLP, only 13% of these clones were non-B locus, i.e., the cross-hybridizing HLA-C locus.
** Not done.
*** Relative increase of the ratios of specific/nonspecific pfu for enriched vs. unenriched libraries; eg $(2.7 \times 10^{-4}/1.7 \times 10^{-6}) = 158.8$.

unenriched but occurred once in the enriched library; the remaining pfu in this group were B-like but not further identified, or were C locus genes. Overall, for other strongly homologous EI7(+) (but non-B27) clones, relative pfu frequency was enriched 36.5 fold.

Similarly, for GDel DNA, enrichment improved the B27 pfu relative frequency from 4.0×10^{-6} to 2.3×10^{-4}, an increase of 57.5 fold. Among the group of highly homologous, other (non-B27) EI7 (+) pfu, the second B locus allele, B18, of GDel was identified by RFLP and found at twice the frequency in the enriched compared to the unenriched library (1.0×10^{-4} vs 0.5×10^{-4}); the remainder of this group were primarily B-like but not further identified. Overall, all Class I pfu frequencies were moderately increased; 6.7 fold for other (non-B27) highly homologous, EI7(+) pfu, and 3.8 fold for lower homology, EI7(-) other Class I genes.

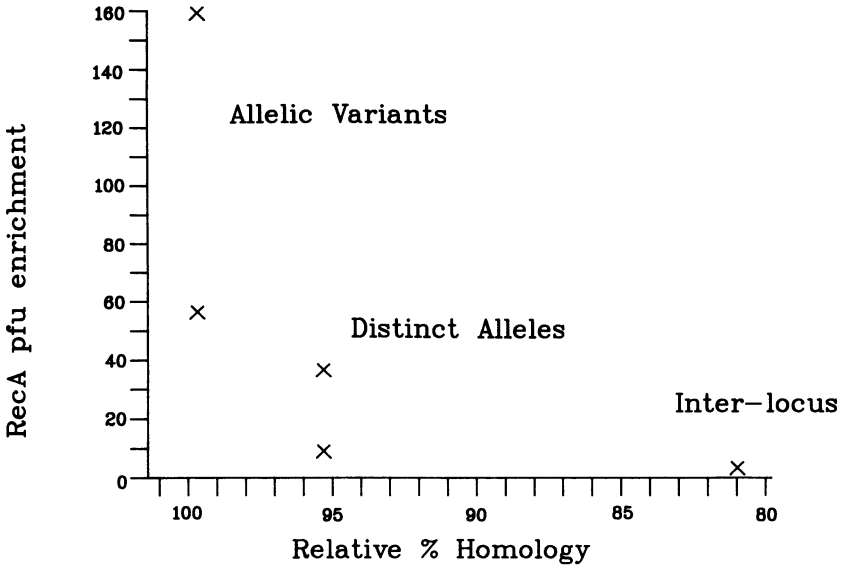
Although recA enrichment functions by large scale removal of non-homologous DNA, some homologous material is also lost in the process. The control reactions with labeled DNA (see Methods) project estimates of this loss, yielding results similar to those reported previously when the probe was also prepared from denatured DNA (1). In reactions where homologous DNA was labeled,

62% ± 18% of this DNA was recovered (all data on the control reactions are given as the average ± standard deviation, n=4). When heterologous DNA was radiolabeled, 0.27% ± 0.08% of this non-specific DNA was recovered. Enrichment was calculated as the recovery of specific DNA divided by the recovery of non-specific DNA; this enrichment was 230 ± 29 for the control reactions. The several-fold lower enrichment detected for the experiments with genomic DNA, relative to control reactions with labeled DNA, may be due to some non-homology (e.g., in 5' and 3' flanking regions) between target and probe in the former case. Alternatively, the lower enrichment in the experiments with genomic DNA may result from a variation between the enriched and unenriched samples in the efficiency of any of the steps involved in these experiments but not in the control experiments. For example, if selection for packaging only phage genomes that contain inserts is weaker for the enriched sample, the effect will be to lower the apparent enrichment relative to the control experiments. Additionally, in control reactions the probe is made from single stranded M13, whereas the probe used for the genomic digests was made from denatured dsDNA which has complementary strands that could affect enrichment efficiency. The data in tables II and III also permit a rough measure of the unavoidable loss of homologous genomic DNA actually experienced. We estimate that if each library had identical insertion, packaging and total cloning efficiency per mcg of insert DNA, then the recA enrichement procedure removed greater than 99% of the "nonspecific" DNA, but only 1/5 of the specific, highest homology DNA. These assessments of specific and nonspecific genomic DNA recovery in actual experiments are reasonably close to and thus consistent with the yields predicted by the control experiments for homologous and for nonhomologous DNA. In a similar manner, calculations can be made for the less homologous, non-B27, EI7(+) and the E17 (-) Class I genes as well; e.g., for GDe1 DNA the former is 2% and the latter 1% of DNA retained.

DISCUSSION

In the present study, recA enrichment has been utilized for the first time to facilitate cloning from genomic libraries. Specifi-

cally, from the large, polymorphic family of HLA Class I genes, the variant B27 alleles sought were significantly enriched among all pfu's in the genomic libraries prepared from two different individuals. In the most practical terms, the (mean) 108 fold increase in pfu frequency of the targeted clones means that once the first member of a gene family is available to use for recA enrichment, closely related genes may readily be cloned from enriched libraries only 1% of the size of a nonenriched library; i.e., enrichment will reduce the total pfu, plates, lifts, labelled probe and related work by approximately 99%. Although not as impressive a saving in effort and material can be effected for less homologous members of the family of the target gene, several fold improved efficiency is still obtained. To fully quantitate the enrichment for genes of lesser homology, detailed sequence data for the HLA genes cloned here would be needed; nevertheless some general correlations between degree of homology and pfu enrichment may be approximated based on the following estimations from this study and from other reports characterizing HLA genes and antigens (12,13,14). First, for very closely related genes (e. g., serologically identical or nearly identical variants of an HLA allele, e.g., B27 variants) the DNA sequence homology will be of the order of 99%. Second, within a Class I locus, sequence homology between readily distinguishable alleles will be of the order of 95%. Third, with the exception of occasional small stretches of high similarity, DNA sequence homology between alleles of different HLA Class I loci can be expected to be of the order of 80%. Combining these points with the observed pfu enrichment (Table III), the general trend for correlation between enrichment and homology can be shown by Figure 2. These data agree qualitatively with the correlation expected from previous observations showing that the formation of stable joint molecules by recA protein requires a high degree of homology (15). Enrichment is clearly a function of sequence similarity, however some critical parameters of similarity remain to be defined. For example, how important is the length of homologous regions; will any interspersed regions of non-homology negate the effect of adjacent streches of homology; how much will different temperatures and salt concentrations affect en-



Legend for Figure 2
General correlation between recA pfu enrichment of genomic libraries for target genes of different degrees of homology to enrichment probes. This correlation is based upon estimations described in the text, considering, for this discussion, that the group of high homology non-B27, EI7 (+) pfu (87% of which were B locus or B-like genes) represent intra-locus allelic degrees of homology. From Table III, x = relative enrichment values for different specific pfu in each library.

richment of less homologous genes; and what are the optional sizes and configurations (e.g., is a short internal non-homology sequence desirable) for the enrichment probe? Clarification of these questions should provide even more efficient recA enrichment, however even without those answers the current enrichment procedure offers a significant degree of enrichment when applied to genomic libraries.

Although recA improves the efficiency of screening genomic libraries where ample starting DNA is available, it will probably be of restricted utility in screening libraries with very limited starting DNA. This constraint arises from the loss of starting DNA inherent in the recA enrichment procedure itself. In the process of removing almost all of the nonspecific DNA from the library by recA enrichment, recovery of specific DNA is incom-

plete, ranging from about 20% of the most homologous, to about 1% of the least homologous of the Class I genes. Nevertheless, if the amount of starting DNA is not a severely limiting factor, the impressive and significant reduction in cloning effort and material afforded by recA enrichment is clearly worth the price of the loss of some of the starting DNA. For example, cloning the desired B27 gene from either of the experimental enriched libraries of this study could have been done with only 1-2 plates each, whereas the unenriched libraries would have required a minimum of 50-150 plates, with commensurate increases in plaque lifts and radioisotope for probing.

In conclusion, the recA enrichment procedure is rapid, inexpensive and allows ligation and packaging of enriched DNA directly, without further purification. This procedure has been shown to be a powerful new tool capable of greatly reducing the work and expense of cloning homologous genes from a library of genomic DNA.

ACKNOWLEDGEMENTS

Supported in part by NIH Grants HL35094, HL33811 and P01-CA16038. The authors wish to express their gratitude to Dr. Debra Hiraki for encouragement and stimulating discussions, to Mr. Edward Chan for technical assistance and to Ms. Judy Campbell and Ms. Miriam Guzman for their preparation of the Manuscript.

*To whom correspondence should be addressed

REFERENCES

1. Honigberg, S.M., Rao, B.J. and Radding, C.M. (1986) PNAS 83, 9586-9590.
2. Rigas, B., Welcher, A.A., Ward, D.C., and Weissman, S.M. (1986) Proc. Nat. Acad. Sci 83, 9591-9595
3. Turek P., Engleman, E.G. and Grumet, F.C. (1985) Immunol. Rev. 86, 71-92.
4. Srivastava R., Duceman B.W., Biro P.A., Sood A.K. and Weissman S.M. (1985) Immunol. Rev. 84, 93-122.
5. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: a laboratory mannual Cold Spring Harbor Laboratory, Cold Spring, Harbor New York.
6. Ness, D.B. and Grumet, F.C. (1987) Human Immunology 18, 65-73.

7. Shibata, T., Cunningham, R.P. and Radding, C.M. (1981) J. Biol. Chem. 256, 7557-7564.
8. DasGupta, C., Shibata, T., Cunningham, R.P. and Radding, C.M. (1980) Cell 22, 437-446.
9. Tsang, S.S., Chow, S.A. and Radding, C.M. (1985) Biochemistry 24, 3226-3232.
10. Muniyappa, K., Shaner, S.L., Tsang, S.S. and Radding, CM. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 2757-2761.
11. Grumet, F.C., Fish, L., Mossazadeh, J., Ness, D. and Duceman, B.W. (1984) Mol. Biol. Med., 1,501-9.
12. Koller, B.H. and Orr, H.T. (1985) J. Immunol. 134, 2727-2733.
13. Sodoyer, R., Nguyen, C., Strachan, T., Santoni, M.J., Damotte, M., Trucy, J. and Jordan, B.R. (1985) Ann, Inst. Pasteur/Immunol. 136C, 71-84.
14. Weiss, E.H., Kuon, W., Dorner, C., Lang M. and Riethmuller, G. (1985) Immunobiol. 170(5) 367-380.
15. DasGupta, C. and Radding, C.M. (1982) Nature 295, 71-73.