Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 8955–8958, November 1989
Immunology

Human major histocompatibility complex contains a minimum of 19 genes between the complement cluster and HLA-B

THOMAS SPIES*, MAUREEN BRESNAHAN, AND JACK L. STROMINGER

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138

Contributed by Jack L. Strominger, August 15, 1989

ABSTRACT **A 600-kilobase (kb) DNA segment from the human major histocompatibility complex (MHC) class III region was isolated by extension of a previous 435-kb chromosome walk. The contiguous series of cloned overlapping cosmids contains the entire 555-kb interval between C2 in the complement gene cluster and HLA-B. This region is known to encode the tumor necrosis factors (TNFs) α and β, B144, and the major heat shock protein HSP70. Moreover, a cluster of genes, BAT1–BAT5 (HLA-B-associated transcripts) has been localized in the vicinity of the genes for TNFα and TNFβ. An additional four genes were identified by isolation of corresponding cDNA clones with cosmid DNA probes. These genes for BAT6–BAT9 were mapped near the gene for C2 within a 120-kb region that includes a HSP70 gene pair. These results, together with complementary data from a similar recent study, indicated the presence of a minimum of 19 genes within the C2–HLA-B interval of the MHC class III region. Although the functional properties of most of these genes are yet unknown, they may be involved in some aspects of immunity. This idea is supported by the genetic mapping of the hemopoietic histocompatibility locus-1 (*Hh-1*) in recombinant mice between TNFα and H-2S, which is homologous to the complement gene cluster in humans.**

The human major histocompatibility complex (MHC) is defined by the highly polymorphic class I (HLA-A, -B, and -C) and class II (DR, DQ, and DP) immune response genes. These encode cell-surface receptors for peptides derived from antigen by intracellular proteolysis, which mediate the clonal activation of cytotoxic and helper T lymphocytes, respectively (1–4). Various MHC class I and class II alleles are associated with susceptibility to a large number of diseases (5). Many of these are caused by cellular and/or humoral autoimmune responses and some are linked to certain class II DR and DQ alleles. Evidence has mounted that single polymorphic amino acid residues within the putative peptide binding cleft of class II molecules correlate with disease susceptibility (6). However, in most cases the role of the MHC in predisposition to disease is not understood. Within the MHC, allelic combinations of the distant class I and class II loci display linkage disequilibrium. Thus, the MHC may contain a number of presently unknown disease-related genes.

The MHC comprises a chromosomal segment of 3500 kilobases (kb). A molecular linkage map of the entire MHC has been established by pulsed-field gel electrophoresis (7, 8). The class I and class II gene families are located within distinct 1600- and 900-kb regions at the telomeric and centromeric end of the MHC, respectively, and flank the 1000-kb central MHC class III region. This interval contains within 100 kb a cluster of closely linked genes for several complement factors (C4A, C4B, Bf, and C2) and steroid 21-hydroxylase (21-OHA and 21-OHB) (9, 10). On the centromeric side, the gene for 21-OHB is 350 kb distant from the nearest class II locus, DR. Presently, no genes have been localized within this region. On the telomeric side, the gene for C2 is separated by 600 kb from the proximal class I locus, HLA-B. This interval includes the genes for the tumor necrosis factors (TNFs) α and β and the major heat shock protein HSP70 (7, 8, 11).

To identify genes within the MHC class III region, a 435-kb genomic segment centromeric to HLA-B has recently been isolated by chromosome walking with overlapping cosmids (12). In addition to HLA-B, the cloned region includes the genes for TNFα, TNFβ, and B144. Moreover, five previously unidentified genes, BAT1–BAT5 (HLA-B-associated transcripts), have been discovered. As an extension of these previous data, the present study provides a molecular analysis of the entire 600-kb C2–HLA-B interval. An additional four genes, BAT6–BAT9, were detected by isolation of corresponding cDNA clones. Together with a pair of genes encoding HSP70, the genes for BAT6–BAT9 were mapped within 120 kb of DNA near the gene for C2. Thus, the MHC class III region contains a surprisingly large number of genes with yet unknown functional properties. Like most MHC genes, the genes for BAT1–BAT9 may also be involved in some aspects of immunity. This hypothesis is supported by genetic mapping of the hemopoietic histocompatibility locus-1 (*Hh-1*) in H-2 recombinant mice between TNFα and H-2S, which is equivalent to the complement gene cluster in humans (13, †).

MATERIALS AND METHODS

Cosmid Cloning. Cosmid libraries were prepared from partial *Sau*3AI digests of total genomic DNA from the B-lymphoblastoid cell line Mann. The 35- to 45-kb fraction was obtained by centrifugation in sucrose gradients (10–40%) and was ligated with the vector pTCF, which had been linearized with *Bam*HI and treated with calf intestinal phosphatase (15). After *in vitro* packaging (Gigapack; Stratagene) and transduction into *Escherichia coli* 490A, $\approx 6 \times 10^5$ clones were plated onto nitrocellulose filters and the cosmid library was processed as described (16). DNA probes were labeled with [^{32}P]dCTP by random hexamer priming (17). Hybridizations were in 6× NaCl/Cit (1× NaCl/Cit is 0.15 M NaCl/0.015 M sodium citrate) containing 5× Denhardt's solution (1× Denhardt's solution is 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone), 10% (wt/vol) dextran sulfate, 0.2% $NaDodSO_4$, and sonicated salmon sperm DNA (100 μg/ml) for 16 hr at 65°C. High-stringency washing of filters was at 65°C in 0.1× NaCl/

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

Abbreviations: MHC, major histocompatibility complex; 21-OH, 21-hydroxylase; TNF, tumor necrosis factor; BAT, HLA-B-associated transcript; wp, walking probe(s).

*Present address: Division of Tumor Virology, Dana–Farber Cancer Institute, Harvard Medical School, Boston, MA 02115.

†Simmons, W. S., Bennett, M., Rembecki, R. M., Kumar, V., David, C. S. & Yuan, D., Seventh International Congress on Immunology, July 1989, Berlin, p. 99 (abstr.).

Cit/0.1% $NaDodSO_4$. Filters were exposed to Kodak XAR-5 film for 6–12 hr. Isolated cosmid clones were analyzed by restriction mapping and DNA blot hybridization. The cDNA probe for C2 was a 750-base-pair (bp) *Pvu* II fragment from pC274 (9). A 2.3-kb *Bam*HI/*Hin*dIII human cDNA fragment was used as a probe for HSP70 (18).

Chromosome Walking. Single-copy walking probes (wps) were obtained by a preannealing procedure that prevents hybridization of repeat sequences (12). Purified DNA fragments from distal cosmid insert locations were digested with *Sau*3AI, labeled, heat denatured in the presence of human placenta DNA (10 mg/ml) (Sigma type XIII; sonicated to yield an average size of 300–800 bp) in 0.1 ml of 0.9 M NaCl/0.09 M sodium citrate, incubated 1 min on ice, and preannealed 10 min at 65°C. The chromosome walk was an extension of a previously reported 435-kb stretch of overlapping cosmids (12). The cosmids S6B and S8B were cloned with wpN from K11C (see Fig. 1). From S8B, wpM was derived to isolate P17A. The short overlap of these two cosmids was confirmed by DNA blot hybridization of diagnostic restriction digests. The subsequent series of cosmids (S4A, S21A, T21A, and T27A) was identified by using wpL, wpK, wpJ, and wpH from P17A, S4A, S21A, and T21A, respectively. A probe from T27A (wpG) hybridized to the cosmids S10B and S22A, which together with S9B had been isolated by using the cDNA probe for C2. Inspection of restriction maps showed definitive overlaps shared among T27A, S22A, and S10B (see Fig. 1).

cDNA Cloning. A cDNA library of $\approx 1.5 \times 10^5$ recombinants from the human T-cell line HPB-ALL (kindly provided by B. Seed, Massachusetts General Hospital, Boston) (19) was screened with a series of cosmid probes by the preannealing procedure described above for wps. Isolated clones were matched and characterized by DNA and RNA blot hybridization, and the corresponding genes were mapped within the cosmids.

DNA and RNA Blot Hybridization. Transfer from gels onto nitrocellulose filters and blot hybridization of cosmid and total genomic DNA restriction digests and of RNA samples were according to standard protocols (20). RNA was prepared from the cell lines HPB-ALL (T cell), Raji (B cell), U937 (monocyte), and HeLa by the guanidinium thiocyanate method and cesium chloride step gradient centrifugation (21). Poly(A)$^+$ RNAs were selected by oligo(dT)-cellulose affinity chromatography and 2-μg samples were subjected to electrophoresis in 1% agarose/2.2 M formaldehyde gels (20).

RESULTS AND DISCUSSION

To characterize the genetic complexity of the entire 600-kb C2–HLA-B interval of the MHC class III region, a previous chromosome walk comprising a 435-kb DNA segment centromeric to HLA-B (12) was extended toward the gene for C2. A continuous stretch of overlapping cosmids was isolated by using a series of wps, the first of which originated from the cosmid K11C, which was nearest to the gene for C2 (Fig. 1). In addition, a cosmid library was screened with a C2 cDNA probe. The cosmids S9B, S22A, and S10B contained the closely linked genes for C2 and Bf. The position and relative orientation of the two genes were determined by cosmid DNA blot hybridization and by matching the restriction map with published data (9, 22). In the cosmids S22A and S10B, a region upstream from the gene for C2 overlapped the cosmid T27A, which had been cloned with wpH in the last step of the chromosome walk. This finding was confirmed by blot hybridization of diagnostic restriction digests of the cosmids S22A, S10B, and T27A with wpG from T27A used as a probe (Fig. 1). Thus, the genes for C2 and HLA-B were linked within a contiguous 600-kb cluster of overlapping cosmids representing a substantial portion of the MHC class

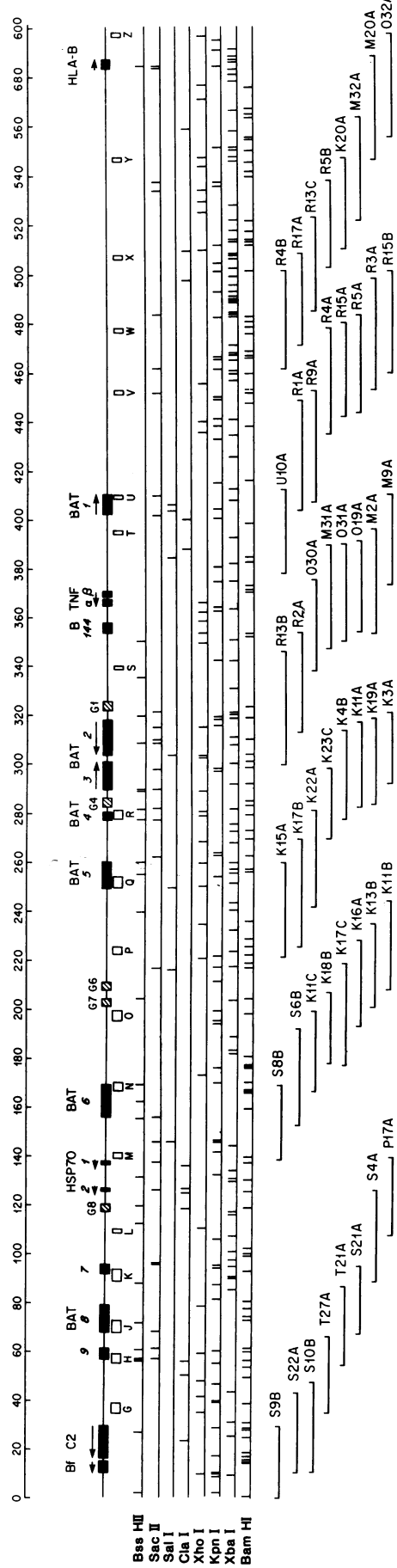

FIG. 1. Molecular structure of the 600-kb MHC interval between the complement gene cluster and HLA-B. This genomic segment is defined by a series of overlapping cosmids shown at the bottom. These clones were mapped by using the restriction enzymes *Bss*HII, *Sac* II, *Sal* I, *Cla* I, *Xho* I, *Kpn* I, *Xba* I, and *Bam*HI. The top line gives the scale (in kb). Solid boxes refer to genes. Hatched boxes are locations of the recently described transcripts G1, G4, G6, G7, and G8 (22). Arrows show the direction of transcription of genes. Open boxes indicate locations of wp. See text for further explanations.

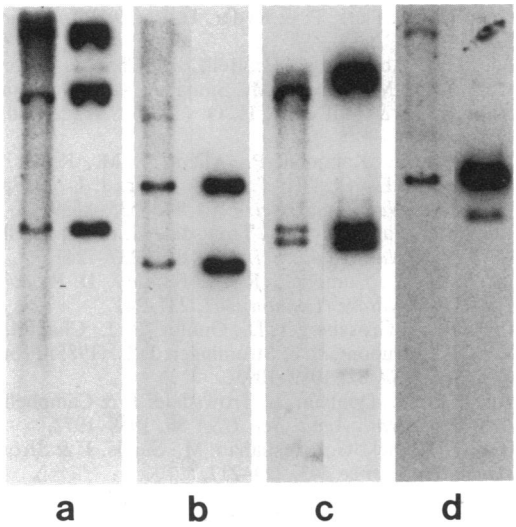

FIG. 2. Localization of the genes for BAT6–BAT9 within cloned cosmids and detection in total genomic DNA by blot hybridization. In parallel experiments, total Mann cell DNA (left lane of each pair) and cosmid clones (right lane of each pair) were compared. (*a*) The BAT6-4 cDNA probe hybridized to the cosmid S8B and identified corresponding fragments in total genomic DNA. Similarly, bands displayed on the total genomic DNA blots were matched by fragments in the cosmids S4A (*b*), S21A (*c*), and T27A (*d*) when the cDNA clones BAT7-11, BAT8-6, and BAT9-10, respectively, were used as probes. Restriction digests were with *Bam*HI in *a*, *c*, and *d*, and with *Xba* I in *b*.

III region. The distance between the genes for C2 and HLA-B was 555 kb. The genes for C2 and TNFα were separated by 335 kb. These results showed that previous linkage data derived from pulsed-field gel electrophoresis were 10–20% overestimates of the actual physical distances (7, 8).

In addition to the genes for HLA-B, C2, and Bf, the cloned MHC segment includes the genes for TNFα and TNFβ, B144, and BAT1–BAT5 (12). Moreover, two closely linked genes encoding a major heat shock protein, HSP70-1 and HSP70-2, have recently been localized 92 kb telomeric to the C2 gene (11). Both genes were identified in a corresponding position within the cosmid P17A by blot hybridization using a HSP70 cDNA probe (Fig. 1).

The previous discovery of the genes for BAT1–BAT5 near the genes for TNFα and TNFβ prompted a similar search for genes in the vicinity of the genes for C2 and HSP70. A large number of *Bss*HII and *Sac* II restriction sites occurring within 140 kb of DNA upstream from the gene for C2 indicated the presence of multiple islands of CpG-rich sequences (Fig. 1). CpG islands are known to be frequently associated with expressed genes (23, 24). Moreover, this correlation was exemplified by the genes for BAT2–BAT5, which lie within a region characterized by a high density of *Bss*HII and *Sac* II sites. In contrast, no gene has presently been found within the 175-kb BAT1–HLA-B interval, which is relatively depleted of CpG-rich sequences (Fig. 1). Altogether, these observations suggested strongly that additional genes were encoded near the genes for C2 and HSP70.

To expedite the identification of genes, a T-cell HPB-ALL cDNA library was screened with a series of total cosmid inserts from the C2–HSP70 region. This technical approach was chosen on the basis that most CpG island-associated genes, including the genes for BAT1–BAT5, are ubiquitously expressed (12, 23). Thus, genes encoded within the C2–HSP70 region were expected to be transcribed in T cells. The HPB-ALL cDNA library was screened successively with the labeled inserts from the cosmids S8B, S4A, S21A, and T27A (Fig. 1). To prevent hybridization of probes to repeat sequences, which are present in a small fraction of cDNAs, a preannealing procedure was used. Isolated clones were analyzed by restriction mapping and assigned to their homologous genomic locations by hybridization with cosmid DNA blots. Altogether, four genes, BAT6–BAT9, were identified. From each of the four sets of cDNA clones obtained, the clones BAT6-4 (4 kb), BAT7-11 (1.7 kb), BAT8-6 (3.2 kb), and BAT9-10 (1.6 kb) were the longest. These clones were hybridized with blots of total genomic DNA next to individual cosmid clones. The results demonstrated that all of the genes for BAT6–BAT9 were unique single-copy sequences (Fig. 2). The BAT6 gene was encoded within 14 kb of DNA, 17 kb upstream from the gene for HSP70-1 (Fig. 1). The location of the 6-kb BAT7 gene was 28 kb downstream from HSP70-2; 10.5 kb distant from BAT7, the gene for BAT8 was mapped within 13 kb of DNA. The 9-kb BAT9 gene was closely linked to BAT8 and was 28 kb upstream from the gene for C2 (Fig. 1).

All of the genes for BAT6–BAT9 were transcribed in a panel of cell lines including HPB-ALL (T cell), Raji (B cell), U937 (monocyte), and HeLa (Fig. 3). Thus, most of the genes within the MHC class III region between C2 and HLA-B are ubiquitously expressed. The BAT6, BAT8, and BAT9 mRNAs were $\approx$4, $\approx$3.2, and $\approx$3.0 kb long, respectively, and were expressed at moderately high levels. In contrast, the BAT7 (2.5 kb) transcript was very scarce.

Altogether, a minimum of 14 genes was encoded within the 555-kb C2–HLA-B MHC segment. In a similar recent study, a total of 10 transcripts, G1–G10 has been mapped within the C2–TNFα interval by probing RNA blots with cosmid DNA fragments (22). By comparison of cosmid map locations and lengths of mRNAs, the genes for BAT2, BAT3, BAT4, BAT7, and BAT9 corresponded to the transcripts G2, G3, G5, G9, and G10, respectively. However, the transcripts G1, G4, G6, G7, and G8 were not accounted for. By inference from published results, G1 was tightly linked to the 5′ end of the BAT2 gene. G4 was located between the genes for BAT3

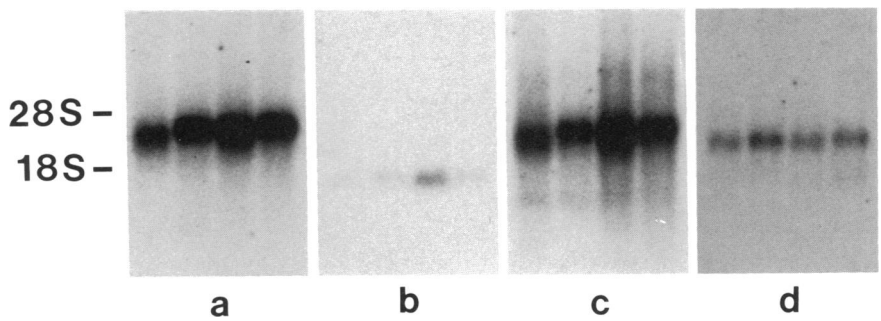

FIG. 3. RNA blot analysis of expression of the genes for BAT6–BAT9. (*a–d*) Lanes 1–4, poly(A)$^+$ RNA samples (2 μg each) from HPB-ALL, Raji, U937, and HeLa, respectively. (*a–d*) The BAT6-4, BAT7-11, BAT8-6, and BAT9-10 cDNA probes hybridized to single transcripts of 4, 2.5, 3.2, and 3.0 kb, respectively.

and BAT4. G6 and G7 were ≈50 kb and ≈35 kb telomeric to BAT6, respectively. G8 was 5 kb centromeric to HSP70-2 (Fig. 1). Thus, the available complementary data indicated the presence of at least 19 genes between the genes for C2 and HLA-B. This surprisingly large number suggests that the DR–21-OHB interval of the MHC class III region may also be replete with currently unidentified genes. Moreover, the present study was biased toward detection of genes that were ubiquitously expressed. Thus, the cloned 600-kb region may contain several more genes, which are expressed in a tissue-specific manner. This possibility is highlighted by the transcripts G1 and G7, which are restricted to monocytes (G1 and G7) and T cells (G1) (22).

Several lines of evidence support the hypothesis that some of the genes for BAT1–BAT9 may be related to immunity. Nucleotide sequence data suggest that several BAT gene products, such as HSP70, may belong to a cell stress response system (unpublished data). Moreover, a genetic locus controlling the expression of an alloantigen recognized by natural killer (NK) cells, *Hh-1*, has been mapped in recombinant mouse strains between TNFα and H-2S (13, †). Regarding the conserved overall organization of genes in the MHCs of mouse (H-2) and human, a human *Hh-1* homolog is presumably encoded between the genes for TNFα and C2. Thus, this putative genetic locus may be involved in the regulation of target cell lysis by alloreactive human $CD3^-$ and $CD16^+$ NK cells (14).

We thank B. Seed for providing the cDNA library. This work was supported by a grant from the German Rheumatism Research Center (Berlin) to T.S. and by National Institutes of Health Grant DK-30241 to J.L.S.

1. Kappes, D. & Strominger, J. L. (1988) *Annu. Rev. Biochem.* **57,** 991–1028.
2. Bjorkman, P. J., Saper. M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987) *Nature (London)* **329,** 512–518.
3. Buus, S., Sette, A. & Grey, H. M. (1987) *Immunol. Rev.* **98,** 115–141.
4. Townsend, A. R. M., Rothbard, J., Gotch, F. M., Bahadur, G., Wraith, D. & McMichael, A. J. (1986) *Cell* **44,** 959–968.
5. Svejgaard, A., Platz, P. & Ryder, L. P. (1983) *Immunol. Rev.* **70,** 193–218.
6. Todd, J. A., Acha-Orbea, H., Bell, J. I., Chao, N., Fronek, Z., Jacob, C. O., McDermott, M., Sinha, A. A., Timmerman, L., Steinman, L. & McDevitt, H. O. (1988) *Science* **240,** 1003–1009.
7. Carroll, M. C., Katzman, P., Alicot, E. M., Koller, B. H., Geraghty, D. E., Orr, H. T., Strominger, J. L. & Spies, T. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8535–8539.
8. Dunham, I., Sargent, C. A., Trowsdale, J. & Campbell, R. D. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7237–7241.
9. Carroll, M. C., Campbell, R. D., Bentley, D. R. & Porter, R. R. (1984) *Nature (London)* **307,** 237–241.
10. White, P. C., Grossberger, D., Onufer, B. J., Chaplin, D. D., New, M. I., Dupont, B. & Strominger, J. L. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1089–1093.
11. Sargent, C. A., Dunham, I., Trowsdale, J. & Campbell, R. D. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 1968–1972.
12. Spies, T., Blanck, G., Bresnahan, M., Sands, J. & Strominger, J. L. (1989) *Science* **243,** 214–217.
13. Rembecki, R. M., Kumar, V., David, C. S. & Bennett, M. (1988) *J. Immunol.* **141,** 2253–2260.
14. Ciccone, E., Viale, O., Pende, D., Malnati, M., Biassoni, R., Melioli, G., Moretta, A., Long, E. O. & Moretta, L. (1988) *J. Exp. Med.* **168,** 2403–2408.
15. Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10,** 6715–6732.
16. Steinmetz, M., Stephen, D., Dastoorniko, G. R., Gibb, E. & Romaniuk, R. (1986) in *Immunological Methods*, eds. Lefkovits, I. & Pernis, B. (Academic, New York), Vol. 3, pp. 1–19.
17. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
18. Wu, B., Hunt, C. & Morimoto, R. I. (1985) *Mol. Cell. Biol.* **5,** 330–341.
19. Aruffo, A. & Seed, B. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8573–8577.
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
21. Chirgwin, J., Przybyla, A., MacDonald, R. & Rutter, W. (1979) *Biochemistry* **18,** 5295–5299.
22. Sargent, C. A., Dunham, I. & Campbell, R. D. (1989) *EMBO J.* **8,** 2305–2312.
23. Bird, A. P. (1986) *Nature (London)* **321,** 209–213.
24. Lindsay, S. & Bird, A. P. (1987) *Nature (London)* **327,** 336–338.