(1984); E. Vilmer *et al.*, *Lancet* **1984-I**, 753 (1984); S. Harada, Y. Koyanagi, N. Yamamoto, *Science* **229**, 563 (1985).
4. L. Ratner *et al.*, *Nature (London)* **313**, 277 (1985); S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole, M. Alizon, *Cell* **40**, 9 (1985); R. Sanchez-Pescador *et al.*, *Science* **227**, 484 (1985); M. A. Muesing *et al.*, *Nature (London)* **313**, 480 (1985).
5. W. G. Robey *et al.*, *Science* **228**, 593 (1985); F. diMarzo Veronese *et al.*, *ibid.* **229**, 1402 (1985); L. Kitchen *et al.*, *Nature (London)* **312**, 367 (1984); J. Schüpbach *et al.*, *Science* **224**, 503 (1984); J. S. Allan *et al.*, *ibid.* **228**, 1091 (1985).
6. W. C. Goh, C. Rosen, J. Sodroski, D. Ho, W. Haseltine, in preparation.
7. J. Sodroski *et al.*, *Science* **227**, 171 (1985); S. K. Arya, C. Guo, S. F. Josephs, F. Wong-Staal, *ibid.* **229**, 69 (1985); J. Sodroski, R. Patarca, C. Rosen, F. Wong-Staal, W. Haseltine, *ibid.*, p. 74; J. Sodroski, C. Rosen, W. Haseltine, unpublished observations.
8. C. A. Rosen, J. G. Sodroski, W. A. Haseltine, *Cell* **41**, 813 (1985).
9. A. Dayton, J. Sodroski, C. Rosen, W. C. Goh, W. Haseltine, *ibid.*, in press.
10. J. S. Allan *et al.*, *Science* **230**, 810 (1985).
11. A. B. Rabson *et al.*, *ibid.* **229**, 1388 (1985).
12. G. M. Shaw *et al.*, *ibid.* **226**, 1165 (1984); A. G. Fisher, E. Collalti, L. Ratner, R. C. Gallo, F. Wong-Staal, *Nature (London)* **316**, 262 (1985).
13. C. A. Rosen, J. G. Sodroski, K. Campbell, W. A. Haseltine, *J. Virol.* **57**, 379 (1986).
14. S. Z. Salahuddin *et al.*, *Virology* **129**, 51 (1983).
15. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
16. D. P. Grandgenett, A. C. Vora, R. D. Schiff, *Virology* **89**, 119 (1978); D. P. Grandgenett, R. J. Kraus, P. J. Hippenmeyer, *ibid.* **130**, 257 (1983); D. Schwartzberg *et al.*, *Cell* **37**, 1043 (1984); L. A. Donehower and H. E. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6461 (1984); P. Hippenmeyer and D. Grandgenett, *Virology* **137**, 358 (1984); A. T. Panganiban and H. M. Temin, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7885 (1984).
17. T. H. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3856 (1984).
18. C. Queen and D. Baltimore, *Cell* **33**, 729 (1983).
19. A. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1977).
20. M. Essex *et al.*, *Science* **220**, 859 (1983); M. Essex *et al.*, *ibid.* **221**, 1061 (1983).
21. H. M. Rho, B. J. Poiesz, F. W. Ruscetti, R. C. Gallo, *Virology* **112**, 355 (1981).
22. We thank F. Wong-Staal, B. Hahn, G. Shaw, and R. Gallo for providing the pHXBc2 plasmid; D. Bolognesi, M. Hirsch, D. Ho, and J. Groopman for patient sera; Z. Salahuddin for providing the C8166 cell line; A. Dayton, M. F. McLane, J. Allen, T.-H. Lee, M. Cloyd, and D. Celander for discussions; K. Campbell and R. Zaya for technical assistance; and C. Owens and T. Parks for editorial assistance. This work was supported by grants to W.H. from the American Cancer Society and the State of Massachusetts. J.S. is a Special Fellow of the Leukemia Society of America and a recipient of NIH grant CA40658. C.R. is supported by NIH grant CA07580.

12 December 1985; accepted 27 January 1986

Identification of HTLV-III/LAV *sor* Gene Product and Detection of Antibodies in Human Sera

NANCY C. KAN, GENOVEFFA FRANCHINI, FLOSSIE WONG-STAAL, GARRETT C. DUBOIS, W. GERARD ROBEY, JAMES A. LAUTENBERGER, TAKIS S. PAPAS*

The nucleotide sequence of the genome of HTLV-III, the infectious agent etiologically associated with the acquired immune deficiency syndrome, predicts a small open reading frame, termed *sor*, located between the *pol* and *env* genes. A DNA segment containing 82 percent of the *sor* region was inserted into a prokaryotic expression vector, pJL6, to determine whether *sor* encodes a viral protein and to gain some insight into its possible function. The bacterially synthesized *sor* protein reacted with sera from individuals infected with HTLV-III, indicating that *sor* is expressed as a protein product or products that are immunogenic in vivo. Antibodies to the purified, bacterially synthesized *sor* protein were found to react specifically with the same protein and also with a protein of molecular weight 23,000 (23K) in HTLV-III–infected H9 cell extracts. The 23K protein comigrated with a protein immunoprecipitated by the serum of a hemophiliac patient with antibodies to HTLV-III, suggesting that this protein is probably the *sor* gene product.

HUMAN T-LYMPHOTROPIC VIRUS type III (HTLV-III/LAV) has been implicated as the causative agent of the acquired immune deficiency syndrome (AIDS) in humans (*1–5*). The published nucleotide sequences of three isolates of HTLV-III and related viruses predict the characteristic retroviral *gag*, *pol*, and *env* genes as well as two small open reading frames, *sor* and 3′-*orf* (*6–8*). The two small open reading frames in the HTLV-III genome are unusual in that they have no equivalents similar in size and location in the genomes of other exogenous human retroviruses such as HTLV-I and HTLV-II (*9, 10*). The *sor* region is located between the *pol* and *env* genes and can encode a protein of 192 amino acids with an estimated molecular weight of 22,500 (22.5K). Its initiator methionine overlaps with the end of *pol* and its terminator precedes the beginning of *env*. The 440 nucleotides between *sor* and *env* contain additional open reading frames that have the potential to encode small novel proteins. One of these has been shown to constitute the second exon of the *trans*-activating gene (tat_{III}) (*11, 12*). The putative initiator methionine of the 3′-*orf* is located 4 base pairs downstream from the stop codon of the *env* protein. Its open reading frame extends into the U3 element of the 3′ long terminal repeat (LTR). The native gene product of 3′-*orf* in HTLV-III–infected cells has been identified as a 27K protein and shown to be immunogenic in vivo (*13, 14*). In this report, we show that the HTLV-III *sor* gene product is a 23K protein that is also immunogenic in vivo.

HTLV-III is an exogenous retrovirus that infects mainly the $T4^+$ human T cells. Unlike most other nondefective retroviruses, infection with HTLV-III generally leads to cell death in vitro and to depletion of T cells with OKT4 phenotype in patients with AIDS (*2, 5, 15*). The small open reading frames may contribute to this cytopathic effect by encoding viral proteins that either directly or indirectly alter the cell metabolism. DNA probes derived from the *sor* region hybridize to RNA species in HTLV-III–infected cells (*12, 16*). However, whether the *sor*-related messenger RNA's (mRNA's) are translated into distinct protein species remains to be studied, and the functions of the putative *sor* gene products are unknown.

In one approach to these problems we have expressed the *sor* region in *Escherichia coli*. The Alu I–Alu I fragment of the BH10 clone (*17*) of HTLV-III corresponding to the segment from nucleotides 4724–5201 [numbered according to Ratner *et al.* (*6*)] was isolated from an Eco RI subclone. The 5′ Alu I site is located 34 amino acids downstream from the putative initiator methionine, while the 3′ Alu I site coincides with the termination codon of *sor*. Therefore, only 35 amino acids may have been deleted with respect to the native protein. The Alu I fragment was joined to the expression vector pJL6 (*18, 19*) at the Hind III site (Fig. 1). Thirteen amino acids at the NH_2-

N. C. Kan, J. A. Lautenberger, T. S. Papas, Laboratory of Molecular Oncology, National Cancer Institute, National Institutes of Health, Frederick, MD 21701-1013.
G. Franchini and F. Wong-Staal, Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892.
G. C. DuBois, Program Resources, Inc., Frederick Cancer Research Facility, Frederick, MD 21701-1013.
W. G. Robey, Virus Control Unit, National Cancer Institute, National Institutes of Health, Frederick, MD 21701-1013.

*To whom requests for reprints should be addressed.

terminus, including the initiator methionine for the *sor*-containing protein, are provided by the phage λ *c*II gene. The structure of plasmid psor2 was verified by the chain terminator DNA sequencing method (*20*) with the use of double-stranded psor2 DNA as described by Zagursky *et al.* (*21*). The DNA sequence of about 100 nucleotides clearly demonstrates that psor2 contains an open reading frame starting at the *c*II gene in the vector that extends in frame into the *sor* open reading frame. The bacterially synthesized *sor* protein, with an estimated molecular weight of 20K, thus contains 17 amino acids derived from the vector at its amino terminus. The remaining 157 amino acids are derived from the HTLV-III *sor* region.

The *c*II-*sor* hybrid gene in psor2 is under the transcriptional control of the λ *p*L promoter located immediately upstream from the *c*II gene. psor2 was placed in *E. coli* MZ1 that contains a temperature-sensitive λ repressor (*18*). When total *E. coli* proteins were metabolically labeled with either [^{35}S]methionine or [^{35}S]cysteine, a new protein of 20K was produced in temperature-induced MZ1 cells containing psor2 (Fig. 2A, lane 2). This protein was observed as two species that were resolved on sodium dodecyl sulfate–polyacrylamide gels (SDS-PAGE). It was not observed in uninduced cells, or in induced cells containing the vector alone, or psor1 with an out-of-frame *sor* region (Fig. 2A, lanes 1 and 3). Coomassie blue staining showed that the *sor* protein constituted approximately 10 percent of total *E. coli* proteins (Fig. 2B). This *sor* protein was purified by successive extractions with salts and detergents as described by Krippl *et al.* (*22*) and represented about 80 percent of the protein at this stage of purification (Fig. 2C, lane 1). It was further purified by high-performance liquid chromatography (HPLC) either on a cation-exchange column or on a reverse-phase C_{18} column as described by Samuel *et al.* (*23*). The cation-exchange column resolved the two major species of *sor* protein and revealed a possible third species of intermediate mobility on SDS-PAGE (Fig. 2C, lanes 2 to 4). The two prominent bands seen in Fig. 2C, lane 1, coeluted from the C_{18} column, and the NH_2-terminal sequence of both species was determined by sequential Edman degradation with an Applied Biosystems protein sequencer model 470A. Although the sample was composed of two bands seen on SDS-PAGE, only a single NH_2-terminal sequence could be detected. This sequence is composed of the first 13 amino acids of the NH_2-terminus from the λ *c*II gene. Residues 14–17 are derived from the ligation procedure and residues 18–29 are from the NH_2-terminus of the Alu I–Alu I fragment of BH10 clone. This result indicates that the two species of *sor* protein are not due to internal initiation of the cloned fragment and could be the result of proteolytic cleavage at the COOH-terminus or other post-translational modification of the protein.

Table 1. Frequency of human natural antibodies against bacterially synthesized *sor* protein.

Clinical status	Number tested	Number positive	
		Total	Percentage
Healthy donors	16	4	25
Healthy at risk	29*	12	41
ARC	22*	7	32
AIDS	28*	15	54

*All these sera were also positive for antibodies to other viral antigens (p24 or gp41 or both).

The partially purified *sor* protein shown in Fig. 2C, lane 1, was used in a Western blot assay to test its immunoreactivity with human sera. Serum samples from 79 patients seropositive for other HTLV-III antigens (p24 or gp41 or both) (*24, 25*) and from 16 normal blood donors with no detectable antibodies to the HTLV-III *gag* and *env* proteins were obtained on a random basis. The 79 seropositive patients included 29 individuals "at risk," such as male homosexuals, intravenous drug users, and relatives of AIDS patients; 22 patients with AIDS-related complex (ARC) (*4*); and 28 patients with AIDS. Representative Western blots are shown in Fig. 2D. We detected antibodies that reacted with the *sor* protein in some people of all the groups that were analyzed, including the healthy donors (Table 1). The healthy donors presumably have not been exposed to HTLV-III (all of them were seronegative for other viral proteins), suggesting that sera from these people may recognize a normal cellular protein or another foreign protein that shares antigenic determinants with the *sor* protein. Therefore, the percentage of patients whose sera reacted with the purified HTLV-III *sor* protein in vitro may not reflect the real frequency with which antibodies to *sor* protein were produced in the infected people. Our results indicate that the *sor* protein is immunogenic

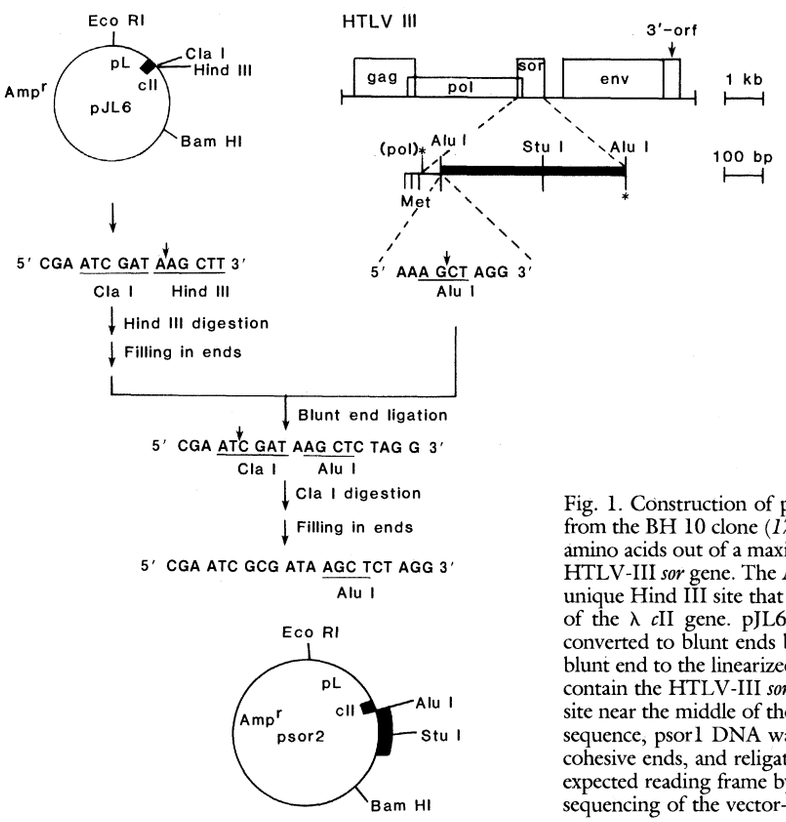
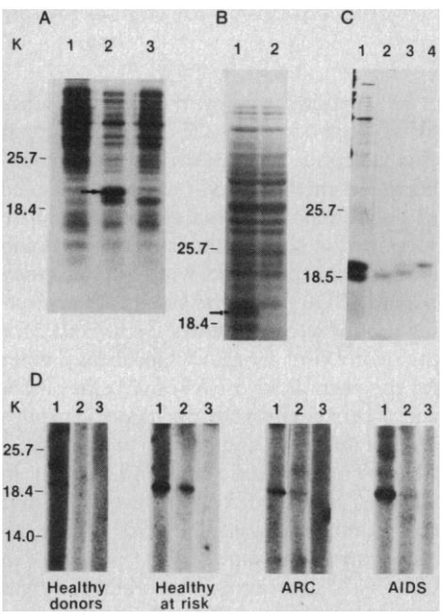


Fig. 1. Construction of plasmid psor2. The Alu I–Alu I fragment of HTLV-III is derived from the BH 10 clone (*17*) corresponding to positions 4724 to 5201 (*6*). It can encode 157 amino acids out of a maximum of 192 determined by *sor* and thus covers at least 82% of the HTLV-III *sor* gene. The Alu I fragment was joined to the expression vector pJL6 (*18*) at the unique Hind III site that is located 42 base pairs downstream from the initiator methionine of the λ *c*II gene. pJL6 DNA was cleaved with Hind III and the cohesive ends were converted to blunt ends by *E. coli* polymerase. The Alu I fragment was then ligated by the blunt end to the linearized pJL6 DNA by T4 DNA ligase. The plasmid psor1 was shown to contain the HTLV-III *sor* sequence cloned in the correct orientation by observing the Stu I site near the middle of the Alu I fragment. To adjust the frame of translation of the inserted sequence, psor1 DNA was cleaved with Cla I, treated with *E. coli* polymerase to fill in the cohesive ends, and religated by T4 ligase. A plasmid, termed psor2, was shown to have the expected reading frame by observing a new Nru I site replacing the Cla I site, and by direct sequencing of the vector-insert junction. Asterisks indicate termination codons.

in vivo and that its prevalence is higher in infected patients than in normal people.

The *sor* protein purified by reversed-phase HPLC (*23*) was used to elicit antibodies. The Western blot assay shown in Fig. 3A demonstrated that the rabbit antiserum reacted strongly with both species of the *sor* protein purified by the cation-exchange HPLC column. The rabbit antiserum was then used to identify the native *sor* protein in HTLV-III–infected H9 cells (*2*). Cells were metabolically labeled and the cell extract was immunoprecipitated with the rabbit antiserum. Only the immune serum recognized a 23K protein in the HTLV-III–infected cells (Fig. 3B, lanes 1 and 2). The serum of a hemophiliac individual without symptoms of AIDS, but seropositive for HTLV-III envelope antigens, also reacted with a protein from the same cell extract which comigrated with the 23K protein. A protein of the same size in HTLV-III–infected cells was also detected with the sera from HTLV-III–seropositive people and was identified as the native *sor* gene product by partial amino acid sequencing (*26*).

Open reading frames revealed by nucleotide sequence analysis are potential but not necessarily functional genes. One way to verify the coding potential of an open reading frame is to insert it, in frame, into a prokaryotic expression vector that can translate almost any foreign gene. Furthermore, antibodies to the bacterially synthesized foreign protein can facilitate the identification of the native product or products in the virus-infected cells. For example, the 3′-*orf* gene, which was first identified in the HTLV-III nucleotide sequence as an open reading frame, has been expressed in bacteria (*14*). In this report we have identified the *sor* gene, another HTLV-III gene first predicted from its nucleotide sequence by expressing it in *E. coli* through the use of a phage λ promoter. We found no specific correlation between the detection of antibodies against *sor* in patients with AIDS or ARC and the different stages of the disease. The low percentage of seropositivity against the *sor* antigen in the infected people may reflect either a low immunogenicity of the protein or an infrequent exposure of the *sor* antigen to the immune system.



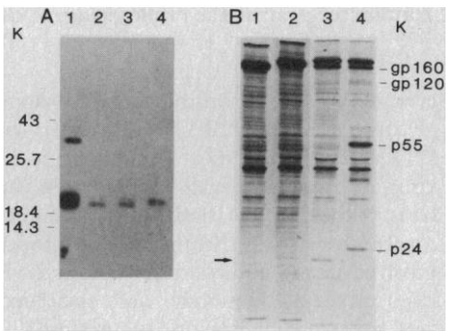


Fig. 2 (left). Analysis of the bacterially synthesized *sor* protein and its immunoreactivity with human sera. (A) Plasmid psor2 was grown in DC646 cells and used later to transform *E. coli* MZ1 that contains a temperature-sensitive λ repressor (*c*I857). MZ1 cells containing vector pJL6 (lane 1), plasmid psor2 (lane 2), or plasmid psor1 (lane 3), were grown in supplemented M56 media (*27*) at 32°C to an optical density at 590 nm of about 0.4. A portion (50 μl) was metabolically labeled with [^{35}S]methionine at 42°C for 30 seconds, and total *E. coli* proteins were visualized on 10% SDS-PAGE by autoradiography as described (*18*). The size of the new protein in lane 2 is consistent with the predicted molecular weight of the HTLV-III Alu I fragment (157 amino acids) plus the NH_2-terminus from the vector (17 amino acids). The results obtained with [^{35}S]cysteine were identical to the ones shown above. (B) *Escherichia coli* MZ1 cells containing either the vector pJL6 (lane 1), or the plasmid psor2 (lane 2), were induced at 42°C for 1 hour, and total *E. coli* proteins were analyzed on 10% SDS-PAGE stained with Coomassie blue as described (*23*). (C) The *c*II-*sor* hybrid protein was extracted from a 2-liter culture of MZ1 cells containing psor2 as described by Krippl *et al.* (*22*) and Samuel *et al.* (*23*). The protein in the potassium thiocyanate pellet fraction (1 mg) was solubilized in 50 m*M* tris, *p*H 7.5, 2 m*M* DTT, and 7*M* urea, and a portion was analyzed on 12.5% SDS-PAGE stained with Coomassie blue (lane 1). Approximately 400 μg of the protein was applied to a cation-exchange Waters Protein-Pak SP-5PW HPLC column equilibrated with 20 m*M* sodium phosphate, *p*H 7.0 with 7*M* urea. The two species of the *sor* protein were eluted in different fractions with a linear 25-minute gradient of 0 to 80% of 0.75*M* NaCl in the equilibration buffer. Portions of the peak fractions were analyzed on 12.5% SDS-PAGE stained with Coomassie blue (lanes 2 to 4). (D) The partially purified *sor* protein fraction shown in (C), lane 1 was subjected to electrophoresis on 15% SDS-PAGE, transferred to a nitrocellulose filter, and reacted with human sera (1:100 dilution) by the Western blot assay as described (*28*). The antigen-antibody complexes were detected by using ^{125}I-labeled protein A. Fig. 3 (right). Characterization of rabbit antiserum to HTLV-III *sor* protein. A rabbit was immunized with 100 μg of the bacterial *sor* protein purified by reverse-phase high-pressure liquid chromatography (*23*). The initial immunization was followed by two 100-μg booster doses 10 and 20 days later. (A) Western blot analysis (*28*). The fractions containing the *sor* protein were subjected to electrophoresis on 10% SDS-PAGE and immobilized on a nitrocellulose filter (*28*). Immune complexes were detected with ^{125}I-labeled protein A. Lanes 1 to 4 correspond to those shown in Fig. 2C, lanes 1 to 4. (B) Immunoprecipitation analysis. HTLV-III–infected H9 cells were labeled with equal activities of [^{35}S]methionine and [^{35}S]cysteine and lysed as described (*29, 30*). The cell extract was immunoprecipitated with: lane 1, rabbit preimmune serum; lane 2, the rabbit antiserum as described above; lane 3, serum from a hemophiliac patient; and lane 4, serum from another hemophiliac patient, followed by electrophoresis on 10% SDS-PAGE. The presence of gp160 in lane 3 and gp160, gp120, p55, and p24 in lane 4 indicate that the patients have been exposed to the virus.

REFERENCES AND NOTES

1. F. Barré-Sinoussi *et al.*, *Science* **220**, 868 (1983).
2. M. Popovic, M. G. Sarngadharan, E. Read, R. C. Gallo, *ibid.* **224**, 497 (1984).
3. M. S. Gottlieb *et al.*, *N. Engl. J. Med.* **305**, 1425 (1981).
4. J. J. Goedert and W. A. Blattner, *AIDS, Etiology, Diagnosis, Treatment and Prevention*, V. T. DeVita, S. Hellman, S. A. Rosenberg, Eds. (Lippincott, Philadelphia, 1985), pp. 1–30.
5. F. Wong-Staal and R. C. Gallo, *Nature* (*London*) **317**, 395 (1985).
6. L. Ratner *et al.*, *ibid.* **313**, 277 (1985).
7. S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole, M. Alizon, *Cell* **40**, 9 (1985).
8. R. Sanchez-Pescador *et al.*, *Science* **227**, 484 (1985).
9. M. Seiki, S. Hattori, Y. Hirayama, M. Yoshida, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3618 (1983).
10. K. Shimotohno *et al.*, *ibid.* **82**, 3101 (1985).
11. J. Sodroski *et al.*, *Science* **227**, 171 (1985).
12. S. K. Arya, C. Guo, S. F. Josephs, F. Wong-Staal, *ibid.* **229**, 69 (1985).
13. J. S. Allan *et al.*, *ibid.* **230**, 810 (1985).
14. G. Franchini *et al.*, in preparation.
15. D. Klatzmann *et al.*, *Science* **225**, 59 (1984).
16. A. B. Rabson *et al.*, *ibid.* **229**, 1388 (1985).
17. B. H. Hahn *et al.*, *Nature* (*London*) **312**, 166 (1984).
18. J. A. Lautenberger, D. Court, T. S. Papas, *Gene* **23**, 75 (1983).
19. J. A. Lautenberger, L. Ulsh, T. Y. Shih, T. S. Papas, *Science* **221**, 858 (1983).
20. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
21. R. J. Zagursky, K. Baumeister, N. Lomax, M. L. Berman, *Gene Anal. Tech.* **2**, 89 (1985).
22. B. Krippl, B. Ferguson, M. Rosenberg, H. Westphal, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6988 (1984).
23. K. P. Samuel, C. S. Flordellis, G. DuBois, T. S. Papas, *Gene Anal. Tech.* **2**, 60 (1985).
24. M. G. Sarngadharan, M. Popovic, L. Bruch, J. Schüpbach, R. C. Gallo, *Science* **224**, 506 (1984).
25. F. di Marzo Veronese *et al.*, *ibid.* **229**, 1402 (1985).
26. T. H. Lee *et al.*, *ibid.* **231**, 1546 (1986).
27. C. R. Merril, M. E. Gottesman, S. L. Adhya, *J. Bacteriol.* **147**, 875 (1981).
28. K. P. Samuel *et al.*, *Science* **226**, 1094 (1984).
29. W. G. Robey, W. J. Kuenzel, G. F. Vande Woude, P. J. Fischinger, *Cancer Res.* **42**, 2523 (1982).
30. W. G. Robey *et al.*, *Science* **228**, 593 (1985).

31. We thank T. Pry for helping us to purify the *sor* protein and T. Matthews for providing serum from a hemophiliac that was used to precipitate the *sor* protein. N.C.K. thanks P. Lapis for help and support during this project.

8 November 1985; accepted 28 January 1986