Vol. 159, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 28, 1989 Pages 87-94

CHEMICAL SYNTHESIS AND EXPRESSION OF THE HIV-1 PROTEASE GENE IN E.COLI

John M. Louis*, Ewald M. Wondrak+, Terry D. Copeland+, C. A. Dale Smith*, Peter T. Mora* and Stephen Oroszlan+

* Division of Cancer Biology and Diagnosis, National Cancer Institute, N.I.H., Bethesda, Maryland 20892

+ Laboratory of Molecular Virology and Carcinogenesis, BRI-Basic Research Program, NCI-Frederick Cancer Research Facility, Frederick, MD 21701

Received December 12, 1988

SUMMARY: The 297bp HIV-1 protease gene was constructed from five discrete synthetic fragments and expressed in E.coli. A soluble protein product of 11.5 Kd was detected by immunoblotting using protease specific antisera. A quantitative assay system, utilizing a synthetic nonapeptide spanning the cleavage site between p17-p24 in the gag polyprotein, was used to measure the specific protease activity in crude extracts. The protease hydrolyzed tyrosyl-proline bonds with an approximate specific activity of 43 pmoles/min/ μg of total protein. The chemical synthesis of the protease gene and it's expression provides a feasible method for rapid mutant analysis, important for structure-function studies and rational design of potential inhibitors.

The genome size mRNA of human immunodeficiency virus (HIV), the causative agent of the Acquired Immunodeficiency Syndrome (AIDS), is translated into two polyproteins: Pr55, the product of the gag gene, and Pr160, the product of the gag-pol gene (1-5). To produce infectious progeny virions from immature virus particles, cleavage of the polyproteins into the smaller mature structural proteins and into the replication enzymes (reverse transcriptase and integrase) is necessary. This cleavage is accomplished by the viral protease. It has been demonstrated in other retroviral systems that deletion mutants lacking protease sequences are non-infectious (6,7). In the HIV-1 it was found that an active aspartyl-type protease is necessary for the processing of the gag-pol polyprotein and for the viral replication, as studied by mutational analysis (8-10).

The precise coding region of several retroviral proteases (11-13) and that of the HIV-1 protease gene have been determined (14). Previous work of others was based on cloning the protease gene with flanking sequences of various lengths of HIV-1 DNA, and by expressing such constructs in E. coli autoprocessing of the virus protease was observed (8-10,15-19). In this

0006-291X/89 $1.50

communication we report the chemical synthesis of the protease gene and the conditions for expression in E. coli. We present evidence that the product of the synthetic gene is a functional protease having the cleavage site specificity of the natural enzyme, using a quantitative assay system especially suitable for measuring specific proteolytic activity in crude cellular extracts.

MATERIALS AND METHODS

PLASMID, BACTERIAL STRAINS AND CHEMICALS: Plasmid PKK233-2, a procaryotic expression vector (Pharmacia), was used to transform a lac-q host, E.coli JM105 or RB791 (20,21). All chemicals utilized in the synthesis of oligonucleotides were from Applied Biosystems Inc. T4 polynucleotide kinase, DNA ligase, and Klenow fragment of E.coli DNA polymerase I were obtained from New England Biolabs. Restriction endonucleases, phenylmethylsulfonylfluoride (PMSF) and isopropylbeta-D-thiogalactopyranoside (IPTG) were from Boehringer Mannheim, Bethesda Research Laboratories and Promega respectively.

CLONING THE 297bp PROTEASE GENE: DNA fragments shown in Fig 1, were synthesized using an ABI DNA synthesizer (model 380B). The full-length synthetic fragments were purified by standard techniques (22). The synthetic protease gene constructed as described in fig 1 was ligated to the plasmid PKK233-2 and used for transformation of E. coli, JM105 (21). Recombinant clones (PR-C, see below) were screened by colony hybridization using a 62 bp fragment (fragment 3 in figure 1) labelled by kinasing (23) and sequenced by the dideoxy method (24).

ANTIBODIES TO THE HIV-1 PROTEASE: Polyclonal antibodies were raised in rabbits against 1) a totally synthetic protease protein (14) of HIV-1 and 2) a tridecapeptide corresponding to the C terminus of the protease, as described (25).

ANALYSIS OF THE EXPRESSED PROTEINS: Recombinant clones were grown to log phase, induced, and lysed by sonication in various buffers (see legend to Figure 2). Total cell extracts were analyzed by $NaDodSO_4$/PAGE and subjected to immunoblot analysis (26,27).

ASSAY FOR THE ACTIVITY OF THE EXPRESSED PROTEASE: The nonapeptide, Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln-amide, used as the substrate, was synthesized according to methods previously published (14,28). After incubation (37°C) of the substrate with appropriate aliquots of cell extracts, (see legend to Fig. 3), the reactions were terminated by adding 100 μl of 8M Guanidine HCl and 5 μl of 20% trifluoroacetic acid (TFA). The substrate and the cleavage products were separated by reversed-phase high pressure liquid chromatography (RP-HPLC) on a μBondapak C_{18} column (Waters Associates). After injection of the sample, the column was eluted with 0.05% trifluoroacetic acid (TFA) in double distilled water. After the buffer components and unbound material were eluted, the column was subjected to a linear gradient of 0 to 30% acetonitrile in water containing 0.05% TFA to separate the substrate and the cleaved products in a time period of 35 minutes monitored at 206 nm. The majority of the bacterial components were eluted subsequently with a 30% to 60% acetonitrile gradient. Peak fractions were collected, lyophilized and analyzed for amino-acid composition using a Pico-Tag amino acid analyzer (Waters Associates).

RESULTS AND DISCUSSION

The 297bp nucleotide sequence of the protease gene of HIV-1, in the pol open reading frame (clone HTLV-III$_{BH10}$; 4,14,29-32), and its complement were synthesized as five individual fragments of approximately 60 bases, as shown in Figure 1.

Three clones (PR-C, PR-H, and PR-J) bearing the correct coding sequence of 297bp in the vector PKK233-2 were analyzed for the optimal induction of the gene. Figure 2 shows a typical Western blot analysis of the gene product. Clone PR-C when induced for various periods of time with .4mM IPTG expressed a single, unfused protein of 11.5kd (figure 2A). Expression was maximal after 30 min of induction and decreased to about 25% at 60 min. After 120 min there was no detectable protease protein in the soluble extract, suggesting either its highly labile nature or conversion into insoluble forms of the expressed product. Growth characteristics of clone PR-C was similar to the control cells bearing the plasmid PKK233-2.

The results of the induction for 30 min with varying concentrations of inducer are shown in Figure 2B. Induction with IPTG in the range of 1mM to 4mM resulted in maximum amount of expression. Similar data were obtained using other identical clones.

fragment 1

5' CCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGGGCAActaa

fragment 2

AGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAGTTTGCCAGGAagat

fragment 3

GGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAgata

fragment 4

CTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTcaac

fragment 5

ATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAAATTTT 3'

FIGURE 1: Strategy for the synthesis of the HIV-1 protease gene. The 3' overhangs (shown in lower case) were provided for the fragments to selectively ligate the appropriate fragments to form the correct coding sequence. Translational initiation codon ATG and termination codon TAA were provided at the appropriate ends of the protease gene. Oligonucleotides were provided at the 5' and 3' end of the gene, having cohesive ends compatible to the restriction enzyme sites NcoI and Hind3 respectively for cloning into the vector PKK233-2. Appropriate complementary fragments (fragments 2,3 and 4) were mixed in equimolar concentrations and annealed in a buffer of 50 mM Tris HCl pH 7.5, 10mM $MgCl_2$ and 100mM NaCl and kinased (10x kinase buffer contained 0.7M Tris HCl pH 7.6, 0.1M $MgCl_2$, 0.05 M DTT and 4mM ATP). Fragment 5 and complementary fragment to fragment 1 were kinased prior to annealing. All annealed fragments were mixed in equimolar concentration for ligation (10x ligase buffer contained 0.5M Tris HCl pH 7.8, 0.1M $MgCl_2$, 0.2M DTT, 500 mg/ml BSA and 4mM ATP) overnight at 15° C (24). This resulted in the construction of the protease gene with unphosphorylated 5' termini. Underlined is the sequence for the putative active site in fragment 2 and another nucleotide sequence in fragment 5 that is also highly conserved in retroviral proteases (14).

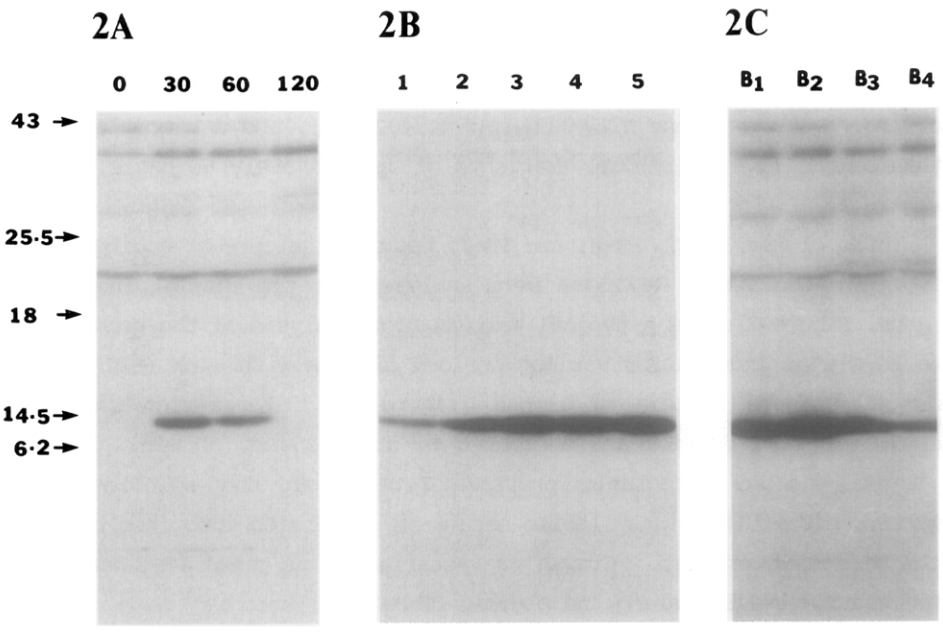

FIGURE 2: Expression of the synthetic protease gene in E.coli. Clone PR-C was grown in 5-50 ml Luria broth to an optical density of 0.4 A600nm and induced by IPTG. The cells were sonicated in B_1 buffer and analyzed by immunoblotting (26,27) using a mixture of the two protease specific rabbit polyclonal antibodies described in Materials and Methods. Figure 2A shows the induction of the gene with 0.4mM IPTG at various periods of time in minutes. Figure 2B shows the induction for 30 min with increasing concentrations of inducer IPTG. 1 to 5 represent mM concentration of IPTG at 0.28, 0.56, 1.12, 2.24, and 4.48 respectively. Figure 2C shows the analysis after 30 min induction with 1mM IPTG and lysing the cells in various buffers. B_1 denotes lysis of cells in 50mM Tris-HCl at pH 7.0, 150mM NaCl, 1mM EDTA, 1mM PMSF, 1mM DTT and 0.5% NP-40. B_2: same as B_1, but without NaCl and EDTA. B_3: 50mM potassium phosphate at pH 6.0, 1mM PMSF and 1mM DTT. B_4: same as B_3 with a pH of 6.5. Positions of protein molecular weight markers are indicated on the left in kilodaltons.

Cells when solubilized by sonication in 50mM Tris-HCl at pH 7.0, 1mM dithiothreitol (DTT), 1mM PMSF and 0.5% nonidet P-40 (NP-40) released the maximum amount of the protease in soluble form (Figure 2C). By comparing aliquots of soluble extract and insoluble pellet by Western blot analysis, similar to those in Figure 2, we estimated that 50% to 70% of the total protease product was solubilized by this buffer. In all experiments there was a single specific band of 11.5 kd (see Figure 2) corresponding to the protein previously characterized as the intact protease protein isolated from purified virus (14,25).

To assess the activity of the cloned HIV-1 protease, a synthetic nonapeptide Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln-amide, corresponding to the HIV-1 p17-p24 cleavage site of the gag precursor (cleavage between the Tyr and Pro residues) (25) was used as a substrate. The substrate in reaction buffer was mixed with aliquots of various cell extracts (see legend to Figure

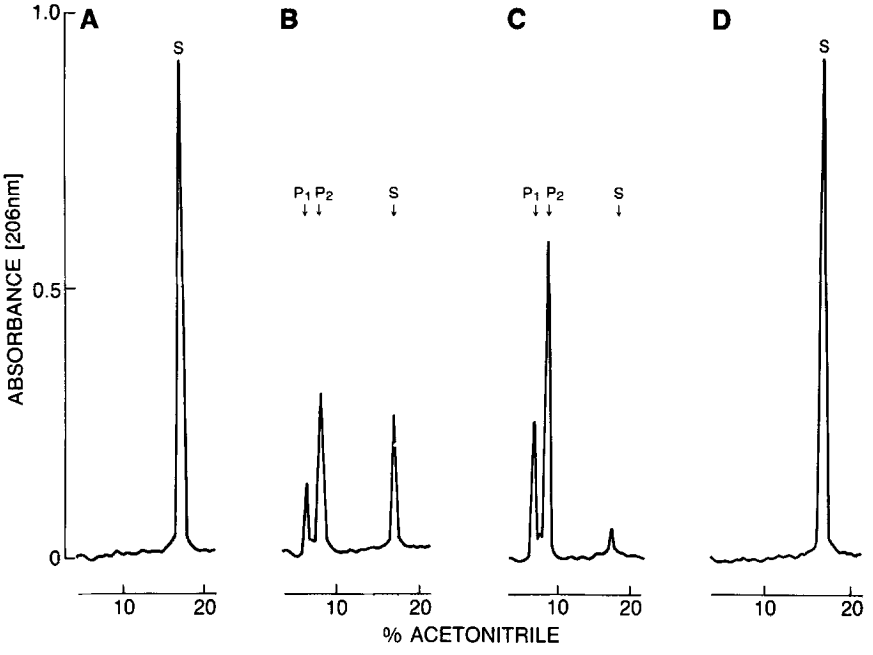

FIGURE 3: Activity of the expressed protease using the nonapeptide as substrate and lysates of induced (ABC) and uninduced (D) cultures (5-50 ml) of clone PR-C. Protease assays in a total volume of 75 µl were carried out with 7.5 µl of cell lysate containing 22.5 µg protein and 25 µg of substrate (in 30 µl of double distilled water) added to 37.5 µl (2X) reaction buffer providing a final concentration of 0.25 M potassium phosphate, pH 7.0, 0.5% (v/v) NP-40, 5% (v/v) glycerol, 5 mM DTT and 2 M NaCl. The reaction mixture was incubated at 37°C and aliquots of 25 µl each were taken at 0 hr (A), 1 hr (B), 3 hr (C) and analyzed by RP-HPLC. D shows the profile after incubation of substrate with uninduced extract after 6 hr. S denotes the substrate and P1 and P2 cleavage products 1 and 2 respectively. No other major peaks other than the products and the substrate were observed eluting between 5-30% acetonitrile. The absorbance peak of product 1, the tetrapeptide Pro-Ile-Val-Gln-NH_2 having a terminal amide, is substantially smaller than that of the pentapeptide which has a free COOH-terminal tyrosine.

3) and incubated at 37° C. Equal aliquots of incubation mixture were taken at various time points and analyzed by RP-HPLC. The substrate in the 0 hour sample eluted as a single peak as shown in Figure 3A. After incubation for 1 hour, two newly appearing peaks, products labelled 1 and 2, can be seen, correlating with a significant decrease of the substrate peak. Subsequent amino acid analysis of the recovered peaks demonstrated that product 1 and product 2 were the expected cleavage products, the tetrapeptide Pro-Ile-Val-Gln-amide and the pentapeptide Val-Ser-Gln-Asn-Tyr , proving a Tyr-Pro bond cleavage which is identical to the determined natural cleavage site. The products were recovered in approximately equimolar amounts. In addition the products were identified by independent synthesis and chromatography. Extended incubation for 3 hours showed that more than 95% of the substrate had been cleaved (fig 3C) indicating progression of the hydrolysis of the

Tyr-Pro bond. Product 1 and product 2 were again recovered in equimolar amounts, as determined by amino acid analysis.

Bacterial protease activities are often difficult to distinguish in crude extracts from the viral protease, in assays using the viral polyproteins. As seen in figure 3D, no cleavage products were detected in incubation mixtures using the extracts from uninduced cells of clone PR-C and of control cells bearing plasmid PKK233-2 (not shown), even after 6 hours of incubation. The sensitive and specific quantitative assay using a synthetic nonapeptide described above is suitable for measuring the proteolytic activity of the viral protease in crude extracts. Synthetic peptide substrates have been used before to assay avian (28,33) and HIV protease activities (14,34-36) and were shown to represent the processing of the polyproteins. We found that the enzyme in very low quantities (7.5 μg total protein of the bacterial lysate) was capable of cleaving 80% of the nonapeptide substrate within 60 minutes and to the extent of 95% in 3 hours (Fig. 3).

Figure 4 shows the amount of substrate cleaved as a function of time using crude bacterial lysate. From the slope of the curve an initial rate of reaction of 43 pmoles substrate cleaved per min/μg total protein was estimated. From experiments at varying substrate concentrations an apparent Km of 0.6 mM was calculated.

When scaled-up to one litre, clone PR-C showed the same levels of protease induced as the 5-10 ml cultures, as determined by Western blotting. Crude lysates (7.2 μg protein/μl) were also tested for viral protease activity using the nonapeptide substrate. After a 3 hr incubation, all the substrate in the reaction mixture was cleaved, giving rise to the correct

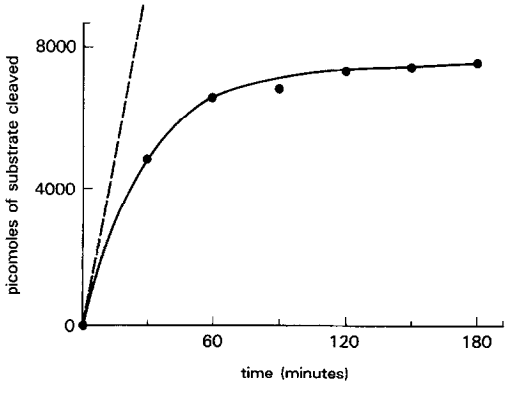

FIGURE 4: Time course of substrate cleavage. Protease assays were carried out under the conditions described in legend to Fig. 3 in a total volume of 200 μl. 25 μl aliquots were taken out of the reaction mixture at 0, 30, 60, 90, 120, 150 and 180 min of incubation and analyzed as described in Materials and Methods. From the initial slope a specific activity of 43 picomoles/min/μg total protein was estimated.

cleavage products. The activity was therefore comparable to the activity obtained using lysates made from 5 to 50 ml cultures.

In conclusion, we demonstrate that the polypeptide product of the synthetic gene corresponds to the 11.5 kd protease protein in the virions (25) and that this 99 amino acid residue protein is an active protease as shown by the specific cleavage of the synthetic nonapeptide used as substrate. Since the synthesis of this gene was in five discrete fragments, any mutations such as substitution, deletion, insertion, etc. in any region of the gene can be easily introduced and their effects can be rapidly tested. The detailed biochemical and structural characterization of HIV protease will facilitate the design and development of specific inhibitors.

ACKNOWLEDGMENTS: We thank Dr. Alan Rabson for encouragement and support, C. Hixson for amino acid analysis, S. Briker for excellent technical assistance, and Cheri Rhoderick for typing the manuscript. Research sponsored in part by National Cancer Institute,. DHHS under contract number NO1-CO-74101 with Bionetics Research Inc. EMW was supported by a Deutsche Forschungsgemeinschaft fellowship: WO 375/1-1.

REFERENCES

1. Dickson, C., Eisenman, R., Fan, H., Hunter, E. & Teich, N. (1984) in RNA Tumor Viruses, Second Edition, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), 513 -648.
2. Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. & Montaganier, L. (1983) Science 220, 868-870.
3. Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P. & Markham, P. (1984) Science 224, 500-502.
4. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R. Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., & Wong-Staal, F. (1985) Nature 313, 277-284.
5. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. (1985) Cell 40, 9-17.
6. Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Okada, T. & Oroszlan, S. (1985) Virology 145, 280-292.
7. Crawford, S. & Goff, S. P. (1985) J. Virol. 53, 899-907.
8. Hansen, J., Billich, S., Schulze, T., Sukrow, S. & Moelling, K. (1988) EMBO J. 7, 1785-1791.
9. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, A. F., Scolnick, E. M., & Sigal, I. S. (1988) Proc. Natl. Acad. Sci. USA 85, 4686-4690.
10. Seelmeier, S., Schmidt, H., Turk, V., von der Helm, K. (1988) Proc. Natl. Acad. Sci. USA 85, 6612-6616.
11. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. (1985) J. Virol. 55, 870-873.
12. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. (1985) Proc. Natl. Acad. Sci. USA 82, 1618-1622.
13. Yoshinaka, Y., Katoh, I., Copeland, T. D., Smythers, G. W. & Oroszlan, S. (1986) J. Virol. 57, 826-832.
14. Copeland, T. D. & Oroszlan, S. (1988) Gene Anal. Tech. 5, 109-115.

15. Kramer, R. A., Schaber, M. D., Skalka, A. M., Ganguly, K., Wong-Staal, F. & Reddy, E. P. (1986) Science 231, 1580-1584.
16. Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W., & Rosenberg, M. (1987) Proc. Natl. Acad. Sci. USA 84, 8903-8906.
17. Mous, J., Heimer, E. P., & Le Grice, S. F. J. (1988) J. Virol. 62, 1433-1436.
18. Graves, M. C., Johann Lim, J., Heimer, E. P. & Kramer, R. A. (1988) Proc. Natl. Acad. Sci. USA 85, 2449-2453.
19. Farmerie, W. G., Loeb, D. D., Casavant, C., Hutchinson, C. A., III, Edgell, M. H. & Swanstrom, R. (1987) Science 236, 305-308.
20. Yanisch-Perron, C., Vierira, J., & Messing, J. (1985) Gene 33, 103-119.
21. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.
22. Maxam, A. M. & Gilbert, W. (1980) in Meth. Enzymol. 65. Nucleic Acids Part I, eds. Grossman, L. & Moldare, K. (Academic Press Inc.) 499-559.
23. Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) Molecular Cloning : A Laboratory Manual. (Cold Spring Harbor Laboratory, Cold Spring Harbor NY) 366-380.
24. Berger, S.L., & Kimmel, A.R. (1987) in Guide to Molecular Cloning Techniques. Meth. Enzymol. 152. (Academic press Inc. NY) 94-562.
25. Henderson, L. E., Copeland, T. D., Sowder, R. C., Schultz, A. M. & Oroszlan, S. (1988) in Human Retrovirusus, Cancer and AIDS: Approaches to Prevention and Therapy. ed. D. Bolognesi (Alan R. Liss) 135-147.
26. Towbin, H., Staehelin, T., & Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350-4354.
27. Burnette, W. N. (1981) Anal. Biochem. 112, 195-203.
28. Copeland, T. D. & Oroszlan, S. (1981) in Peptides: Synthesis - Structure - Function (Proceedings of the Seventh American Peptide Symposium) eds. Rich, D. H. & Gross, E. (Pierce Chemical Company), 497-500.
29. Veronese, F. D. M., Copeland, T. D., DeVico, A. L., Rahman, R., Oroszlan, S., Gallo, R. C. & Sarngadharan, M. G. (1986) Science 231, 1289-1291.
30. Lightfoote, M. M., Coligan, J. E., Folks, T. M., Fauci, A. S., Martin, M. A. & Venkatesan, S. (1986) J. Virol. 60, 771-775.
31. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J. & Varmus, H. E. (1988) Nature 331, 280-283.
32. Lillehoj, E. P., Salazar, F. H. R., Mervis, R. J., Raum, M. G., Chan, H. W., Ahmad, N., & Venkatesan, S. (1988) J. Virol. 62, 3053-3058.
33. Kotler, M., Katz, R. A., Danho, W., Leis, J. & Skalka, A. M. (1988) Proc. Natl. Acad. Sci. USA 85, 4185-4189.
34. Schneider, J. & Kent, S. B. H. (1988) Cell 54, 363-368.
35. Nutt, R. F., Brady, S. F., Darke, P. L., Ciccarone, T. M., Colton, C. D., Nutt, E. M., Rodkey, J. A., Bennett, C. D., Waxman, L. H., Sigal, I. S., Anderson, P. S., & Weber, D. F. (1988) Proc. Natl. Acad. Sci. USA 85, 7129-7133.
36. Darke, P. L., Nutt, R. F., Brady, S. F., Garsky, V. M., Ciccarone, M. T., Leu, C-T., Lumma, P. K., Freidinger, R. M., Veber, D. F. and Sigal, I. S. (1988) Biochem. Biophys. Res. Commun. 156, 297-303.