JOURNAL OF VIROLOGY, Apr. 1988, p. 1433–1436
0022-538X/88/041433-04$02.00/0

Vol. 62, No. 4

NOTES

Processing Protease and Reverse Transcriptase from Human Immunodeficiency Virus Type I Polyprotein in *Escherichia coli*

JAN MOUS,[1] EDGAR P. HEIMER,[2] AND STUART F. J. LE GRICE[1]*

Central Research Units, F. Hoffmann-La Roche Ltd., CH-4002 Basel, Switzerland,[1] and Peptide Research Department, Hoffmann-La Roche, Inc., Nutley, New Jersey 07110[2]

Received 24 September/Accepted 18 December 1987

Expression of the human immunodeficiency virus type I *pol* open reading frame in *Escherichia coli* led to several protease-mediated processing steps of the *pol* precursor polyprotein. Accumulation of two polypeptides with molecular sizes of 64 and 52 kilodaltons, with which reverse transcriptase activity is associated, was observed. The protease moiety of the precursor polyprotein accumulated as a 10-kilodalton species as a result of two specific cleavages. Furthermore, a single-amino-acid substitution in the putative active site of protease totally abolished processing of the precursor polyprotein.

In the absence of a suitable vaccine, chemotherapy offers an alternative approach by which the immunodeficiency caused by human immunodeficiency virus (HIV) infection (8) can be arrested. Protease, reverse transcriptase, and endonuclease, the products of the *pol* open reading frame (9), are critical enzymes against which it might be possible to develop an antiviral drug. These proteins are coordinately expressed as a precursor polyprotein from which, possibly through the *pol*-encoded protease, the individual components are released. Thus, drugs which inhibit protease may interrupt the life cycle of the virus. A better insight into protease action is however necessary. Since the *pol* open reading frame of HIV type I (HIV-I) has been expressed in *Escherichia coli* and *Bacillus subtilis* and leads to correctly processed reverse transcriptase polypeptides (3, 5, 10), bacterial expression systems offer a suitable means to analyze the sequential events of *pol* polyprotein cleavage.

The synthesis and cleavage of the HIV-I *pol* precursor polyprotein in our *E. coli* expression system is illustrated (Fig. 1). We (5) and other groups (3, 10) have demonstrated that two of the cleavage products are polypeptides with molecular sizes of 64 and 52 kilodaltons (kDa) with which reverse transcriptase activity is associated. Since protease is the first translation product of the *pol* open reading frame, we had expected our serum with HIV-I specific antibodies to detect this 18-kDa polypeptide. The inability to do so suggested either that protease was extremely labile in *E. coli* or that our serum had a low titer of the relevant antibodies. Amino acid sequence analysis (2, 6) suggested that the Phe-Pro bond between protease and reverse transcriptase is cleaved during processing. The Phe-Pro sequence which separates protease from reverse transcriptase also occurs within protease, suggesting that it might undergo a secondary internal cleavage after release from reverse transcriptase.

To determine whether the protease was internally cleaved, we repeated the experiment outlined in Fig. 1, except that samples were applied to the sodium dodecyl sulfate-polyacrylamide gel in triplicate. After electrophoresis and transfer to nitrocellulose (12), immunoreactive polypeptides were determined by using three separate antibody preparations. These were (i) a pool of sera with HIV-I-specific antibodies, (ii) antibodies to a synthetic peptide spanning amino acids 39 through 62 of protease (peptide I), and (iii) antibodies to a synthetic peptide spanning amino acids 86 through 108 of protease (peptide II). These peptide sequences lie on either side of the internal Phe-Pro bond of protease (Fig. 2). The results of this assay are shown (Fig. 3).

By using sera with HIV-I-specific antibodies (Fig. 3A), a 92-kDa polypeptide was detected 15 min after *pol* expression was induced. The size of this polypeptide is correctly

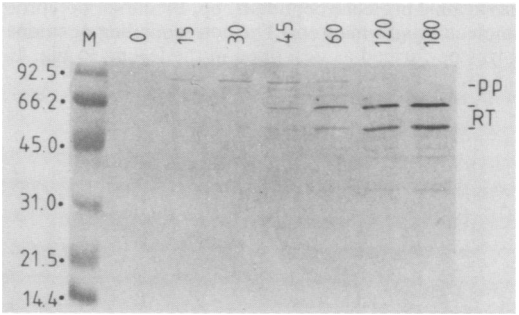

FIG. 1. Immunological analysis of processing of the HIV-I *pol* precursor polyprotein to 64- and 52-kDa polypeptides in *E. coli*. Our expression system (1) contains a *Bgl*II-*Xmn*I fragment from the *pol* region of the provirus (9). *E. coli* containing this recombinant clone pRTL10 was grown in antibiotic-supplemented L broth until mid-log phase (E_{600}, 0.7), after which induction of gene expression was initiated by the addition of isopropyl-β-D-thiogalactopyranoside to a final concentration of 400 μg/ml. At the times indicated (in minutes), cells from 1-ml cultures were suspended in 200 μl of sodium dodecyl sulfate-polyacrylamide gel sample buffer. After being heated at 90°C for 10 min, 10-μl samples were fractionated through 12.5% polyacrylamide gels, electroblotted onto nitrocellulose according to Towbin et al. (12), and then analyzed with serum containing HIV-I-specific antibodies (1). Lane M, Prestained protein molecular size markers (in kilodaltons). Note that only the amino terminus of endonuclease is expressed in the construction pRTL10. pp and RT, Precursor polyprotein and reverse transcriptase-like polypeptides, respectively.

* Corresponding author.

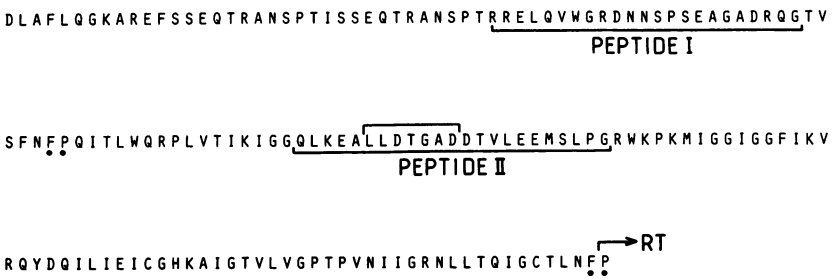

FIG. 2. Amino acid sequence of the HIV-I protease. The regions against which peptide antibodies were raised are underlined. The Phe-Pro bonds separating protease from reverse transcriptase (RT) and present within protease itself are indicated (●). Putative active site of protease has been overlined. Single-letter amino acid abbreviations are used.

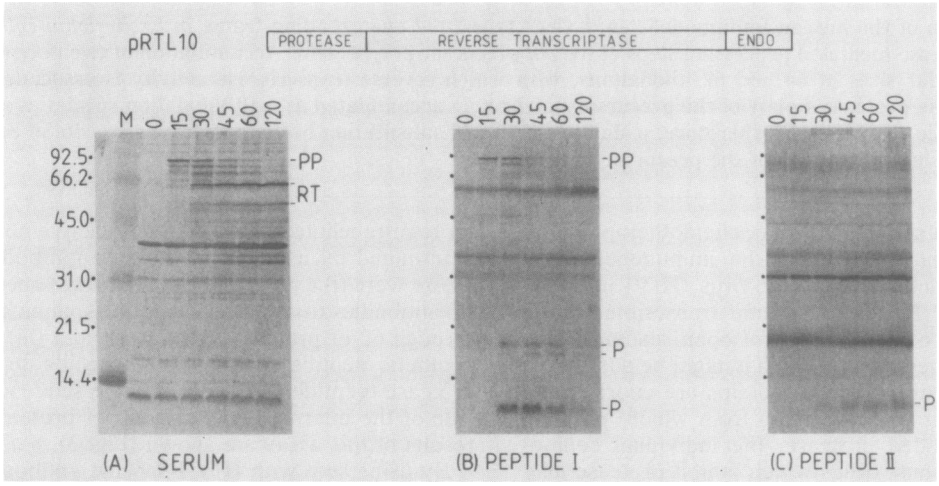

FIG. 3. Analysis of protease release from the *pol* precursor polyprotein in *E. coli* harboring recombinant plasmid pRTL10. The nature of the polyprotein synthesized from pRTL10 is indicated at the top. Postinduction times are indicated above each lane. (A) Immunoreactivity with sera containing HIV-I-specific antibodies. (B) Immunoreactivity with antibodies against protease peptide I. (C) Immunoreactivity with antibodies against protease peptide II. pp, Precursor polyprotein; RT, reverse transcriptase-like polypeptide; P, protease. Lane M, Prestained protein molecular size markers. The corresponding positions of marker proteins in panels B and C have been indicated (●). Immunological analysis was performed as described in the legend to Fig. 1.

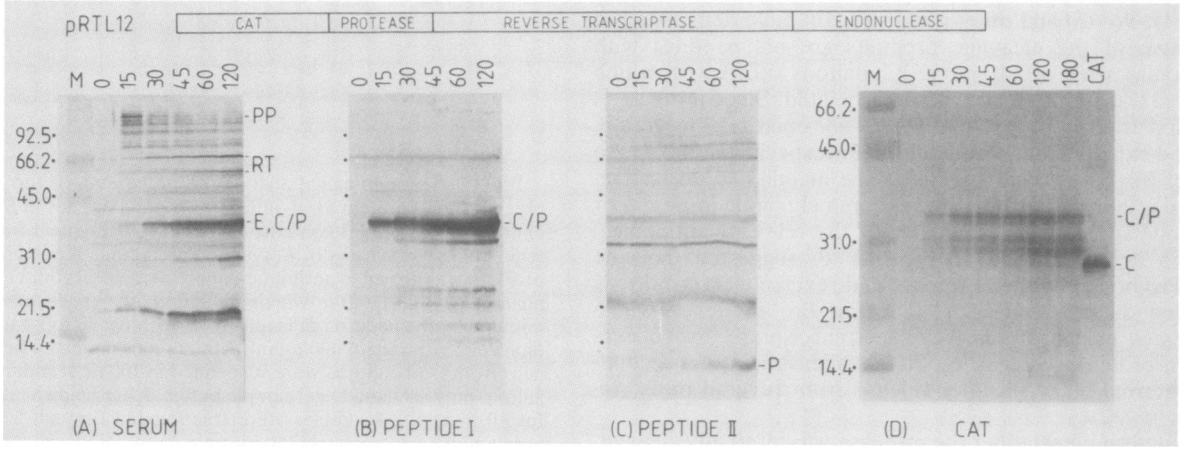

FIG. 4. Analysis of protease release from the *pol* precursor polyprotein in *E. coli* harboring recombinant clone pRTL12. The polyprotein synthesized from pRTL12 is indicated above the panels. (A) Immunoreactivity with sera containing HIV-I-specific antibodies. (B) Immunoreactivity with antibodies against protease peptide I. (C) Immunoreactivity with antibodies against protease peptide II. (D) Immunoreactivity with antibodies to CAT. Lanes M, prestained protein molecular size markers. pp, RT, and P, As in the legend to Fig. 3. C/P, CAT/protease fusion protein; E, endonuclease; C, CAT. Lane CAT in panel D is a sample of purified protein. Immunological analysis was as outlined in the legend to Fig. 1.

predicted by the *pol* fragment in the expression vector pRTL10, which contains protease (18 kDa), reverse transcriptase (64 kDa), and the amino-terminal portion of endonuclease (10 kDa). With time, the precursor polypeptide disappeared, and polypeptides of 64 and 52 kDa were observed. Antibodies against protease peptide I (Fig. 3B) detected two forms of the precursor polyprotein and additional polypeptides with sizes between 19 and 7 kDa. These smaller-sized polypeptides appeared after 30 min of induction and were relatively labile. Probing the same samples with antibodies to protease peptide II (Fig. 3C) revealed accumulation of a 10-kDa polypeptide. From the results obtained with antibodies to peptides I and II, it would indeed appear that full-length protease is released from the polyprotein but is subject almost simultaneously to a secondary internal cleavage to smaller polypeptides, one of which (the 10-kDa form) remains stable. Since peptides I and II lie on either side of the internal Phe-Pro bond, our results suggest that the internal cleavage occurs here. Internal cleavage of protease at this Phe-Pro bond predicts polypeptides of 7.8 and 10.8 kDa, strengthening the results illustrated in Fig. 3.

To further elucidate the secondary cleavage of protease, we cloned the entire *pol* gene as a carboxy-terminal fusion to *E. coli* chloramphenicol acetyltransferase (CAT) (7) with the intention of stabilizing the labile protease cleavage product as an carboxy-terminal fusion to CAT (Fig. 4). Probing fractionated cell proteins with sera with HIV-I-specific antibodies (Fig. 4A) revealed a large-size precursor polyprotein which would correspond to a fusion of CAT, protease, reverse transcriptase, and the entire endonuclease. With time, this precursor also disappeared and the 64- and 52-kDa reverse transcriptase-like polypeptides accumulated. In contrast to results from pRTL10, antibodies against protease

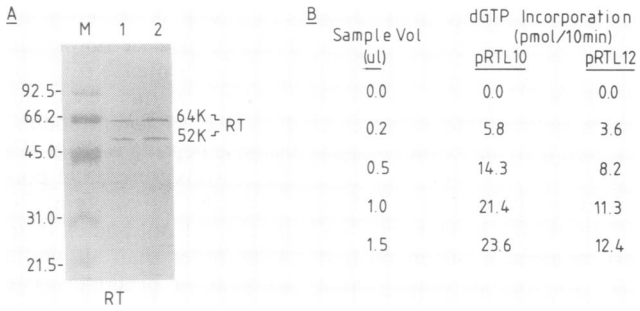

FIG. 5. Partial purification and analysis of reverse transcriptase from *E. coli* harboring the recombinant plasmids pRTL10 and pRTL12. (A) Immunological analysis of DEAE-Sephacel-purified enzymes with rabbit polyclonal antibodies against reverse transcriptase (RT). Lanes: M, Prestained protein molecular size markers; 1, RTL10 enzyme; 2, RTL12 enzyme. In both cases, reverse transcriptase was purified from 500-ml, 5-h induced cultures by DEAE-Sephacel ion exchange chromatography (5) and recovered as the non-DEAE-Sephacel-binding fraction. After pooling and concentration of these enzyme preparations (5), 5-μl aliquots were analyzed for immunological cross-reactivity. (B) Titration curves for RTL10 and RTL12 reverse transcriptase preparations from panel A. Assays (40 μl) contained 50 mM Tris (pH 8.0), 5 mM dithiothreitol, 120 mM KCl, 10 mM $MgCl_2$, 0.05% (vol/vol) Triton X-100, 5 μM dGTP, 10 μg of poly(rC) per ml, 5 μg of oligo(dG) per ml, 4 μCi [α-^{32}P]dGTP (800 Ci/mmol), and the amounts of DEAE-Sephacel-purified reverse transcriptase indicated. After a 10-min incubation at 37°C, a portion of the reaction mixture was spotted onto DEAE filters. Once dry, the filters were washed three times in 2× SSC (0.3 M NaCl, 0.03 M sodium citrate [pH 7.6]) and then dried, and cells were counted in 2 ml of scintillation fluid. Results are the averages of triplicate assays. Volume activities are reported rather than specific activities. As noted earlier (5), low protein concentration and the presence of Triton X-100 in the samples precluded accurate calculation of specific activities.

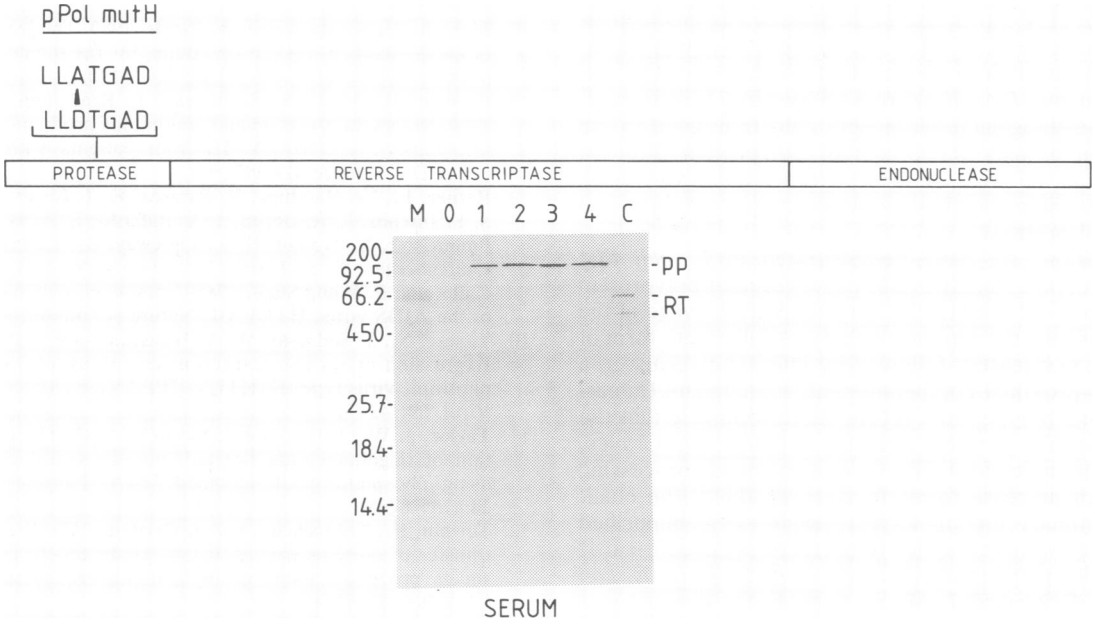

FIG. 6. Immunological analysis of the reverse transcriptase polyprotein produced in *E. coli* harboring the plasmid pPol mutH. (Top) Nature of mutation H which was introduced into the putative active site of protease. The one-letter amino acid notation was used to denote the active-site residues. The Asp-Ala mutation is depicted by an arrow. (Bottom) Expression of the *pol* polyprotein from strain pPol mutH after isopropyl-β-D-thiogalactopyranoside induction. Lanes: M, Prestained protein molecular size markers (note that certain of these differ from those in previous marker lanes); C, reverse transcriptase purified from *E. coli* pRTL10; 0, 1, 2, 3, and 4, times (in hours) after induction at which samples were prepared for immunological analysis. pp and RT, precursor polyprotein and reverse transcriptase, respectively. Illustrated here is an immunoblot which was probed with serum containing antibodies to HIV-I. Immunological analysis was performed as described in the legend to Fig. 1.

peptide I detected large amounts of a 34-kDa protein (Fig. 4B). This corresponds to the predicted size of full-length CAT plus the amino-proximal portion of protease released by internal cleavage. As further support of this result, no 7-kDa polypeptide was detected by the same antibodies. Probing with antibodies against protease peptide II (Fig. 4C) illustrated, however, that the stable carboxy-proximal portion of protease is still released from the polyprotein. Finally, probing lysates with polyclonal CAT antibodies (Fig. 4D) revealed a fusion protein 7 kDa larger than authentic CAT as well as some minor degradation products. The results shown in Fig. 4B and D are thus complementary, illustrating that the labile portion of protease has been stabilized as a carboxy-terminal fusion to CAT.

As proof that the 64- and 52-kDa polypeptides released from the polyproteins of pRTL10 and pRTL12 (Fig. 3 and 4) are indeed reverse transcriptase, a partial enzyme purification from both recombinant clones was undertaken (5), the results of which are presented (Fig. 5). Immunological analysis with rabbit polyclonal antiserum against HIV-I reverse transcriptase revealed both molecular size forms (Fig. 5A). Furthermore, reverse transcriptase activity can be detected in fractions which contain these polypeptides (Fig. 5B). The possibility that this activity results from an *E. coli* enzyme was eliminated by performing a similar purification procedure on the parental *E. coli* strain, in which no activity could be recovered (data not shown).

One possibility our results have as not yet excluded is that *E. coli* proteases and not the HIV-I protease moiety are responsible for the processing of the precursor polyprotein. Sequence alignment of HIV-I protease and other retroviral proteases reveals a highly conserved stretch of amino acids with similarity to sequences around the active site of amino acids Asp-Thr-Pro of aspartyl proteases (4). As a consequence, we have substituted the Asp residue at the putative active site of HIV-I protease with Ala via oligonucleotide-directed mutagenesis (11). The mutated protease was subsequently introduced into the context of the wild-type *pol* sequences in our *E. coli* expression system. In this case, isopropyl-β-D-thiogalactopyranoside induction leads to the accumulation of only the precursor polyprotein (Fig. 6), suggesting that the protease-reverse transcriptase as well as reverse transcriptase-endonuclease cleavages are mediated by the HIV-I protease.

In conclusion, we have demonstrated in *E. coli* that release of protease from the HIV-I *pol* precursor polyprotein is a two-step event initiated by autocatalytic cleavage of the precursor through the protease moiety. A full-length form of the protease is released, followed by internal cleavage to a labile amino-proximal polypeptide and a stable 10-kDa carboxy-proximal polypeptide; recent studies indicate that the latter polypeptide retains catalytic activity (M. Graves, personal communication). Whether an analogous processing leads to such protease forms in vivo remains to be elucidated. Preliminary immunological analysis of semipurified virus particles with antibodies to protease peptides I and II reveals similar polypeptides as those found here in *E. coli* (S. F. J. Le Grice and J. Mous, unpublished data). However, final proof requires N-terminal amino acid sequence analysis of these polypeptides. Experiments of this nature are presently underway.

We are grateful to M. Graves, Hoffmann-La Roche, Nutley, N.J., for communication of data before publication and for supplying peptide antibodies. We furthermore thank V. Beuck and R. Zehnle for excellent technical assistance as well as J. Martin and A. Kröhn, Hoffmann-La Roche, Welwyn Garden City, United Kingdom, for suggestions on mutation of the active site of protease.

LITERATURE CITED

1. **Certa, U., W. Bannwarth, D. Stüber, R. Gentz, M. Lanzer, S. F. J. Le Grice, F. Guillot, I. Wendler, G. Hunsmann, H. Bujard, and J. Mous.** 1986. Subregions of a conserved part of the HIV gp41 transmembrane protein are differentially recognised by antibodies of infected individuals. EMBO J. **5**:3051–3056.
2. **Di Marzo Veronese, F., T. D. Copeland, A. L. DeVico, R. Rahman, S. Orozlan, R. C. Gallo, and M. G. Sarngadharan.** 1986. Characterisation of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. Science **231**:1289–1291.
3. **Farmerie, W. G., D. H. Loeb, N. C. Casavant, C. A. Hutchison III, M. H. Edgell, and R. Swanstrom.** 1987. Expression and processing of the AIDS virus reverse transcriptase in *Escherichia coli*. Science **236**:305–308.
4. **Katoh, I., T. Yasunaga, Y. Ikawa, and Y. Yoshinaka.** 1987. Inhibition of retroviral protease activity by an aspartly proteinase inhibitor. Nature (London) **329**:654–656.
5. **Le Grice, S. F. J., V. Beuck, and J. Mous.** 1987. Expression of biologically active human T-cell lymphotropic virus type III reverse transcriptase in *Bacillus subtilis*. Gene **55**:95–103.
6. **Lightfoote, M. M., J. E. Coligan, T. M. Folks, A. S. Fauci, M. A. Martin, and S. Venkatesan.** 1986. Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. J. Virol. **60**:771–775.
7. **Marcoli, R., S. Iida, and T. A. Bickle.** 1980. The DNA sequence of an IS*1*-flanked transposon coding for resistance to chloramphenicol and fusidic acid. FEBS Lett. **110**:11–14.
8. **Popovic, M., M. G. Sarngadharan, E. Read, and R. C. Gallo.** 1984. Detection, isolation and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science **224**:497–500.
9. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **316**:277–284.
10. **Tanese, N., J. Sodorski, W. A. Haseltine, and S. P. Goff.** 1986. Expression of reverse transcriptase activity of human T-lymphotropic virus type III (HTLV-III/LAV) in *Escherichia coli*. J. Virol. **59**:743–745.
11. **Taylor, J. W., J. Ott, and F. Eckstein.** 1985. The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. Nucleic Acids Res. **24**:8765–8785.
12. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.