Vol. 156, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 14, 1988 Pages 297-303

HIV-1 PROTEASE SPECIFICITY OF PEPTIDE CLEAVAGE IS SUFFICIENT FOR PROCESSING OF GAG AND POL POLYPROTEINS

Paul L. Darke†*, Ruth F. Nutt**, Stephen F. Brady**, Victor M. Garsky**, Terrence M. Ciccarone**, Chih-Tai Leu*, Patricia K. Lumma**, Roger M. Freidinger**, Daniel F. Veber** and Irving S. Sigal*

Departments of Molecular Biology* and Medicinal Chemistry**
Merck Sharp and Dohme Research Laboratories
West Point, Pennsylvania 19486

Received July 18, 1988

The mature proteins of retroviruses originate as a result of proteolytic cleavages of polyprotein precursors. Retroviruses encode proteases responsible for several of these processing events, making them potential antiviral drug targets. A 99-amino acid HIV-1 protease, produced by chemical synthesis or by expression in bacteria, is shown here to hydrolyze peptides corresponding to all of the known cleavage sites in the HIV-1 gag and pol polyproteins. It does not hydrolyze peptides corresponding to an env cleavage site or a distantly related retroviral gag cleavage site.

Retroviral proteins are initially translated in the form of large precursor fusions, gag, pol and env that are proteolytically processed to give the viral structural proteins and the virally encoded enzymes (1). Homologous virally encoded proteases have been partially characterized for several retroviruses (2-5) and similar proteases have been inferred from the DNA sequence of known retroviruses (6,7). Although the identities of the cleavages produced by these proteases have not been fully determined, in the case of murine leukemia virus a deletion of the protease gene resulted in unprocessed gag proteins and the loss of viral infectivity (8).

The genome organization for HIV-1 is shown in Fig. 1 with the genes encoding the 3 major viral precursor proteins outlined. The sites for in

† To whom correspondence should be addressed.

0006-291X/88 $1.50

vivo processing of the precursors to give the mature gag p17 (MA), p24 (CA), p7(NC) and p6, the reverse transcriptase, the integrase and the 2 envelope proteins gp41 and gp120 are known from amino acid sequencing of the polypeptides found in the HIV particles and are designated in Fig. 1 as I-IX (9-13). From the literature of retroviral cleavage site specificities (14), the mature HIV-1 protease would be expected to result from cleavages at positions V and VI in Fig. 1 (pol 69-167, numbering according to ref. 15). A protein having the expected N-terminus has been observed to be produced during *E. coli* expression of portions of the pol gene product (16). We have found that a polypeptide having the sequence pol 69-167 made by expression in *E. coli* cells (17, 18) or by chemical synthesis (19) has the ability to cleave HIV-1 gag p55 to give the gag p24 and p17 products. In this report we describe the ability of this protease to cleave peptides having sequences that correspond to the known *in vivo* cleavages of the virus. To ensure that the activities are intrinsic to the HIV-1 protease, we have used the chemically synthesized protease for all of the data reported, although the same specificity was observed for the microbially expressed enzyme. The 99 amino acid HIV-1 protease hydrolyzes peptides with a substrate specificity diverse enough to account for all of the known cleavage sites for the HIV-1 gag and pol precursors, but specific enough so as not to result in cleavage of the env precursor nor a gag precursor from the distantly related avian myeloblastosis virus (AMV).

MATERIALS AND METHODS

HIV Protease. The protease was synthesized by solid phase methods, purified and folded into the active form as described in ref. 19. HIV-1 protease was also purified from an *E. coli* (18) expression system. Both protease preparations were in 50 mM NaOAc, pH 5.5, 10% glycerol, 1 mM EDTA and 1 mM DTT prior to assay.

Peptide Hydrolysis assays. Synthetic peptides were purified by reverse phase HPLC and hydrolyzed by the protease (0.2 µg) at various concentrations in 50 mM Na acetate, pH 5.5. The reactions (40 µl) were initiated by addition of 5 µl protease and quenched after 30 min at 30°C with 160 µL 12% acetic acid. Peptides were separated on either a 15 cm or a 5.0 cm reverse phase C-18 column with linear gradients of water to acetonitrile (both 0.1% in trifluoroacetic acid), detected by absorbance at 220 nm and the peak areas integrated. In the case of peptides GAG 124-138, GAG 128-135, POL 59-72 and POL 162-174, synthetic product standards were independently

synthesized and shown to comigrate with the products observed. For peptides listed in Table II, large scale reactions were performed to greater than 33% hydrolysis and the resulting mixture subjected to gas phase sequencing to identify the cleavage sites (20). For each of the peptides SQNYPIV and SQNYPIVQ, there were 2 products of the enzyme reaction, one of which had the same retention time on HPLC and a UV absorption characteristic of tyrosine. In the case of SQNYPIV, the second product was shown to comigrate with synthetic PIV, thus the identity of the second product for SQNYPIVQ was assumed to be PIVQ.

Kinetic parameters were determined from assays performed in duplicate at 7 different concentrations of peptide. At least 2 but no more than 3 of the concentrations employed were above the Km values reported. The extent of loss of substrate was in no case greater than 7%. The extinction coefficient for a product on HPLC was either determined by comparison with a known quantity of synthetic product standard, or by preparing a reaction mixture in which more than half of the substrate was hydrolyzed. The quantity of the 2 products appearing in that mixture was calculated from the loss of substrate.

RESULTS AND DISCUSSION

Initial studies of peptide hydrolysis activity by the synthetic protease were conducted with a substrate peptide (gag 124-138, HSSQVSQNYPIVQNI) representing the cleavage site between p17 and p24 (I in Fig. 1). Following incubation of the peptide with the protease, HPLC analysis of the reaction mixture revealed two novel peaks which comigrated with the anticipated products, HSSQVSQNY and PIVQNI (Fig. 2). Gas phase sequencing (20) of the reaction mixture confirmed the generation of the predicted (10) PIVQNI sequence.

The minimal substrate recognition site for the HIV-1 protease was explored by synthesizing a series of consecutively smaller peptides having sequences that span cleavage site I (Table 1). The rates of cleavage reactions listed in Table 1 are relative to that observed for the 15 residue substrate, with all peptides tested at 1 mM. Only a 10% reduction in rate

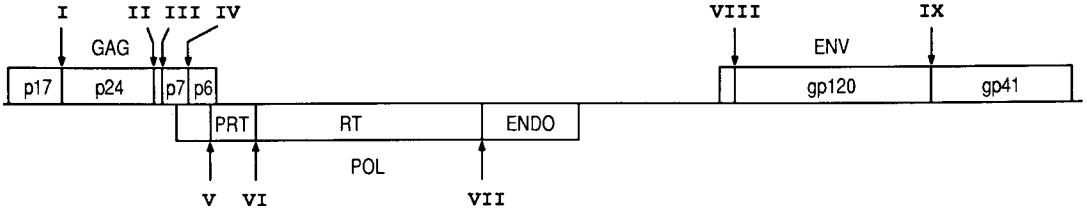

Figure 1. The major open reading frames of HIV-1. Numbered arrows indicate proteolytic processing sites within the polyprotein products.

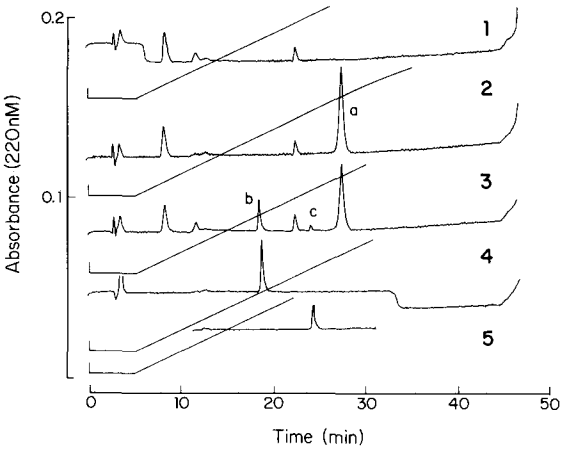

Figure 2. HPLC elution profile for the hydrolysis of a peptide by the synthetic HIV-1 protease. Tracing 1 represents a sample of protease in a reaction solution, in the absence of substrate. Tracing 2 represents GAG 124-138 (peak a) in the absence of enzyme. In tracing 3, the peak area of the substrate is decreased following treatment with the protease, and 2 new peaks (b and c) appear that have mobilities identical to those of the predicted products, GAG 124-132 (tracing 4) and 133-138 (tracing 5). Additional small peaks are due to components of the buffer in the protease preparation.

was observed when the 7 residue peptide corresponding to gag 129-135 (SQNYPIV) was used as a substrate. Removal of residues from either end of this minimal substrate resulted in the elimination of detectable activity. Peptides in which the serine of the 7 amino acid substrate was replaced with an acetyl group or the seryl-glutamine dipeptide with pyroglutamic acid were also not substrates. The Km for the hydrolysis of VSQNYPIV was determined to be 2.5 mM with a Vmax of 275 nmol/min/mg.

Having observed that the 99 amino acid HIV-1 protease had the ability to cleave a peptide of the sequence corresponding to the known cleavage site

Table 1. Length dependence of Hydrolysis Rates for Gag p17/p24 Peptide Substrates

Sequence	Relative Rate of Hydrolysis
HSSQVSQNY-PIVQNI	1.0
VSQNY-PIV	1.0
SQNY-PIVQ	0.95
SQNY-PIV	0.90
SQNY-PI	<0.05
QNY-PIV	<0.10
Acetyl-QNY-PIV	<0.10

Table 2. Hydrolysis of peptides corresponding to peptide cleavage sites

	Site	Sequence	Km (mM)	Vmax (nmol/min/mg)
PEPTIDES HYDROLYZED				
	HIV-1 sites[a]			
I	GAG 124-138	HSSQVSQNY-PIVQNI	N.D.[b]	>275
II	GAG 357-370	GHKARVL-AEAMSQV	2.3	100
III	GAG 370-383	VTNTATIM-MQRGNF	0.16	682
IV	GAG 440-453	SYKGRPGNF-LQSRP	13.9	382
V	POL 59-72	DRQGTVSFNF-PQIT	0.70	954
VI	POL 162-174	GCTLNF-PISPIET	N.D.	>120
VII	POL 721-734	AGIRKIL-FLDGIDK	6.1	145
	HIV-2 site			
	GAG 129-142	SEKGGNY-PVQHVGG	2.3	295
PEPTIDES NOT HYDROLYZED				
	HIV-1 site			
IX	ENV 511-524	RVVQREKR-AVGIGA	-	<2
	AMV site			
	p12/p15[c]	PAVS-LAMTMEHK	-	<2

[a] Numbering according to ref. 15.
[b] Km not determined due to low solubility.
[c] Sequence from ref. 22.

between gag p17 and gag p24, we examined its activity against a series of peptides listed in Table 2, where the peptides designated I-IX have sequences corresponding to several residues on either side of each of the cleavage sites I-IX shown in Fig. 1. Sequences were chosen to favor aqueous solubility while including at least 4 residues N-terminal and 3 residues C-terminal to the *in vivo* cleavage site. The protease was capable of cleaving all of the peptides corresponding to the 4 cleavages of gag (I-IV), the site at the N-terminus of the protease (V), the sites between the protease, reverse transcriptase and integrase (VI, VII), but not the site between env gp120 and gp41 (IX). Because of low solubility, we were unable to examine the peptide that corresponds to the cleavage of the signal peptide of the env precursor (VIII).

The specificity of the protease was further explored using peptides that had sequences for known *in vivo* cleavage sites of the closely related HIV-2 (21) and the more distantly related AMV virus (Table 2). The HIV-2 peptide was a substrate comparable to some of the HIV-1 substrates, while the AMV peptide, shown to be cleaved by the AMV protease (22), was not. The peptides where no cleavage was observed were tested at 2 mM. Increasing the

quantity of protease 10 fold (to 2 μg) or the time of reaction to 6 hr produced no loss of substrate or detectable products.

The peptide cleavages observed are efficient enough to suggest that all of these sites are processed by the HIV protease *in vivo*. Similarity among the retroviral polyprotein cleavage sites has been noted previously, with the deduction of a consensus sequence Ser(or Thr)-Xaa-Xaa-Tyr(or Phe)-Pro (7). We confirm that peptides (I and V) sharing this sequence are substrates, while noting that the dissimilar peptide III is also very effective. It is unusual that a single peptidase is able to cleave N-terminal to proline as well as N-terminal to primary amino acids. In this study, the utilization of enzyme derived by chemical synthesis avoids possible artifacts arising from contaminating activities and offers convincing evidence that the diverse cleavages are intrinsic to the HIV-1 protease sequence. Although a common feature of the cleavage site is the occurrence of hydrophobic amino acids on either side of the scissile bond, protease treatment of HIV-1 gag p55, which contains numerous hydrophobic residue pairs, has been previously shown to result in the highly specific cleavage at sites I and II (19).

On the basis of limited sequence similarity, the retroviral proteases have been compared to the family of aspartyl proteases that include pepsin and renin (14). The minimal length of 7 residues that we observe for an HIV protease substrate agrees closely with the minimal length for a renin substrate (23). In addition, a recent report on the peptide substrate specificity of avian sarcoma-leukosis virus protease (24) demonstrates that while a decamer is a substrate, a hexamer is not. Whereas renin has a high degree of sequence specificity for peptides as substrates, pepsin is much less stringent (25). The HIV protease activity characterized here is much less specific in its substrate sequence than renin, but a more detailed analysis is necessary before the structure-function determinants of an HIV-1 protease substrate will be known.

In conclusion, we have shown for the first time that the 99 amino acid HIV protease will function *in vitro* to cleave peptide substrates having the

sequences for 7 of the 9 known processing sites of HIV-1 polyproteins. The versatility of this protease reflects its importance in the life cycle of the virus and highlights its potential as an anti-viral target.

REFERENCES

1. Coffin, J. in Molecular Biology of Tumor Viruses (eds. Weiss, R. et al.) p. 261-268 (Cold Spring Harbor Laboratory, New York, 1982).
2. Sauer, R.T., Allen, D.W.,and Niall, H.D. (1981) Biochem. 20, 3784-3791.
3. Katoh, I., Yasunaga, T., Ikawa, Y., and Yoshinaka, Y. (1987) Nature 329, 654-656.
4. Yoshinaka, Y., Katoh, I., Copeland, T.D., Smythers, G.W., and Oroszlan, S. (1986) J. Virol. 57, 826-832.
5. Vogt, V.M., Wight, A., and Eisenman, R. (1979) Virology 98, 154-167.
6. Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D.W., Chen, I.S.Y., Miwa, M., and Sugimura, T. (1985) Proc. Nat. Acad. Sci. 82, 3101-3105.
7. Pearl, L.H., and Taylor, W.R. (1987) Nature 329, 351-354.
8. Crawford, S. and Goff, S.P. (1865) J. Virol. 53, 899-907.
9. Wain-Hobson, A., Sonigo, P., Danos, O., Cole, S., and Alizon, M. (1985) Cell 40, 9-17.
10. Sanchez-Pescador, T., Power, M.D., Barr, P.J., Steimer, K.S., Stempien, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolph, A., Levy, J.A., Dina, D., and Luciw, P. (1985) Science 227, 484-492.
11. Lightfoote, M.M., Coligan, J.E., Folks, T.M., Fauci, A.S., Martin, M.A., and Venkatesan, S. (1986) J. Virol. 60, 771-775.
12. Veronese, F.D., DeVico, A.L., Copeland, T.D., Oroszlan, S., Gallo, R.C., and Sarngadharan, M.G. (1985) Science 229, 1402-1405.
13. Veronese, F.D., Rahman, R., Copeland, T.D., Oroszlan, S., Gallo, R.C. and Sarngadharan, M.G. (1987) Aids Res. and Human Retroviruses 3, 253-264.
14. Pearl, L.H. and Taylor, W.R. (1987) Nature 328, 482.
15. 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 Jr., S.R., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. and Wong-Stall, F. (1985) Nature 313, 277-284.
16. DeBouck, C., Gorniak, J.G., Strickler, J.E., Meek, T.D., Metcalf, B.W., and Rosenberg, M. (1987) Proc. Nat. Acad. Sci. 84, 8903-8906.
17. Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A.F., Scolnick, E.M., and Sigal, I.S. (1988) Proc. Nat. Acad. Sci. USA 85, 4686-4690.
18. Darke, P.L. manuscript in preparation.
19. 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., and Veber, D.F. (1988) Proc. Nat. Acad. Sci. USA in press.
20. Hunkapillar, M.W. and Hood, L.E. (1978) Biochem. 17, 2124-2133.
21. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier and Alizon, M. (1987) Nature 326, 662-669.
22. Copeland, T.D. and Orozslan, S. in Peptides: Synthesis, Structure and Function (eds. Rich, D.H. and Gross, E.) 497-500 (Pierce, Rockford, IL 1981).
23. Skeggs, L.T., Lentz, K.E., Kahn, J.R. and Hochstrasser, H. (1968) J. Exp. Med. 128, 13-34.
24. Kotler, M., Katz, R.A., Dahno, W., Leis, J., and Skalka, A.M. (1988) Proc. Nat. Acad. Sci. 85, 4185-4189.
25. Powers, J.C., Harley, A.D. and Myers, D.V. (1977) Adv. in Exp. Med. and Biol. 95, 141-147.