Vol. 168, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 16, 1990 Pages 30-36

IDENTIFICATION OF A HUMAN IMMUNODEFICIENCY VIRUS-1 PROTEASE CLEAVAGE SITE WITHIN THE 66,000 DALTON SUBUNIT OF REVERSE TRANSCRIPTASE

Mary C. Graves[1*], May C. Meidel[2], Yu-Ching E. Pan[2], Michael Manneberg[3], Hans-Werner Lahm[3], and Fiona Grüninger-Leitch[4]

Departments of [1] Molecular Genetics and [2] Protein Biochemistry
Hoffmann-La Roche Inc., Roche Research Center, Nutley, NJ 07110

[3] ZFE-Bio and [4] ZFE-MB, Departments of Central Research Units
F. Hoffmann-La Roche & Co., Ltd., CH-4002 Basel, Switzerland

Received February 20, 1990

SUMMARY: The human immunodeficiency virus-1 reverse transcriptase is a heterodimer of related 51 and 66 kDa subunits. The smaller subunit arises by viral protease-catalyzed cleavage of the carboxy-terminal domain of the 66 kDa species. Comparison of the amino acid composition analyses of the isolated 51 kDa and 66 kDa subunits indicates that the carboxyl terminus of 51 kDa is Phe440. This site was confirmed in vitro using purified recombinant protease and a peptide spanning the postulated cleavage area. The sequence surrounding this site does not show significant homology to other protease cleavage sites in the viral gag and pol precursors; thus, this new information may contribute to our understanding of the sequence specificity of the viral protease.

The role of retroviral PR in virus maturation and its processing of the viral gag and pol precursor proteins is becoming well understood (1, 2). In human immunodeficiency virus-1 (HIV-1), the PR domain within the 156 kDa gag-pol precursor is released to yield an 11 kDa mature form (3), which is active as a homodimer. All HIV-1 PR cleavage sites within the gag and pol proteins have been identified with the exception of a site within RT (4).

HIV-1 RT can be isolated from virus or from recombinant expression systems (when co-expressed with PR) as a 1:1 heterodimer of related 66 and 51 kDa proteins (3, 5-8). The N-terminal sequence of the two subunits are identical (5, 6) indicating that

* To whom correspondence should be sent.

ABBREVIATIONS: HIV, human immunodeficiency virus; RT, reverse transcriptase; PR, protease.

0006-291X/90 $1.50

PR-directed cleavage occurs in the C-terminal domain. This region has a demonstrated susceptibility to many proteases in addition to the viral PR, such as trypsin, and chymotrypsin, and those from E. coli, (9). The heterodimer RT displays RNA- and DNA-dependent DNA polymerase as well as RNase H activities (5, 6, 9, 10). The 15 kDa fragment produced by PR action on 66kDa RT also has been postulated to have RNase H activity (11). However, determing the HIV-PR processing site by N-terminal analysis of the 15 kDa peptide has not been possible; since it is unstable in E. coli (MCG observations) quantites sufficient for analysis have not been obtained.

We sought to determine the sequence of the internal RT processing site which generates the C-terminus of 51 kDa RT for two reasons. First, knowledge of a new cleavage site will provide additional insight on the substrate specificity of HIV PR, which continues to be an attractive target for drug therapy. Second, knowledge of the 51 kDa C-terminus and the N-terminus of the putative RNase H domain will aid in molecular cloning and biochemical characterization of the authentic species.

MATERIALS AND METHODS

Preparation and Amino Acid Analysis of 51 and 66 kDa RT. The HIV RT gene was cloned and co-expressed with the HIV PR gene in E. coli, under which conditions the RT accumulates as a heterodimer of 66 and 51 kDa subunits. The purification and characterization of the enzyme have been described elsewhere (12). Briefly, the 66 kDa RT gene is cloned such that several additional residues are included in the N- and C-termini of RT. None of the additions significantly affect activity. The N-terminus of both 51 and 66 kDa subunits contain an additional Met-Arg-Ser-Gln-Leu- prior to the Pro of the authentic N-terminus. The C-terminus of only the 66 kDa RT contains an additional Arg-Ser-$(His)_6$ affinity tail, which facilitates purification by nickel-chelate chromatography (12). The purified heterodimer was subjected to electrophoresis on a 7.5 % SDS-PAGE, which resolves the 66 and 51 kDa subunits. The proteins were electrophoretically transferred to a polyvinylidene difluoride (PVDF) membrane (13) and stained with Ponceau S to allow precise excision of the two proteins. Membrane strips were cut into fine pieces and transferred to a hydrolysis tube containing 2 ml of 6 N HCl; 1 mg of phenol was added and the protein hydrolyzed for 24 hr at 110°C under argon. Duplicate analyses of both proteins were performed.

Preparation of HIV PR. The molecular cloning and expression of the mature 11 kDa form of PR in E. coli (3) , and the purification of the PR (14) have been previously described. Briefly, the PR present in the soluble fraction of E. coli lysates was subjected to sulphopropyl-silica strong cation exchange (Toyopearl) and gel filtration (Sephadex G-50) chromatography. The enzyme concentration was 0.3 mg/ml and estimated to be 50% pure. Confirmatory experiments were performed with pure PR obtained by further gel filtration chromatography.

Synthesis and Analyses of Peptides. The hexadecapeptide Glu-Lys-Glu-Pro-Ile-Val-Gly-Ala-Glu-Thr-Phe-Tyr-Val-Asp-Lys-Arg, which represents RT residues 430-445,

was synthesized (15) and purified. Determination of the amino acid composition of isolated peptide products following PR treatment was by standard methods (16).

Peptide Assay. The 10 µl reaction contained 7 nmol hexadecapeptide and 11 nmol of PR in 300 mM sodium phosphate, pH 6.0, 1 mM DTT, and 10% glycerol, and was incubated at 37°C. At the indicated time points, the reaction was terminated by the addition of trifluroacetic acid to 1.5%, and the entire reaction was analyzed by HPLC (Waters 15 cm Novapak C18 column). The buffer was 0.1 % trifluoroacetic acid in water with a 10-40% acetonitrile gradient at a flow rate of 1 ml/min. Absorbance at 215 nm was monitored. Fractions containing peptide products were quantitated and analyzed for amino acid content as above.

RESULTS AND DISCUSSION

Amino Acid Analysis of 51 and 66 kDa RT. When portions of the HIV pol gene are initially expressed in E. coli in the absence of a functional viral PR, the major species observed is the precursor form (3). However, co-expression of the viral PR results in the rapid accumulation of the characteristic 1:1 ratio of 51: 66 kDa RT heterodimer. In order to identify this viral PR cleavage site, we have determined the amino acid composition of the two RT subunits. Heterodimer RT was produced in E. coli co-expressing viral PR and purified (12). The two RT subunits were separated and the amino acid composition of each species was determined. The composition of 66 kDa RT, where the exact protein sequence is known (17), was in excellent agreement with that predicted (not shown). The composition of the 51 kDa RT is shown in Table 1. The data are consistent with the PR cleavage being between Phe440-Tyr441 (numbering from RT N-terminus). Several possible sites in this region were analyzed against the observed data, but none demonstrated the excellent agreement as seen with the site at Phe440. A second line of evidence corroborates this as the processing site. Subtraction of the 51 kDa residues from those in 66 kDa RT yields the data in Table 1. Once again, agreement is evident between observed data and that predicted from a Phe440-Tyr441 cleavage site.

Cleavage of Peptide Substrate. We sought to confirm this sequence as the PR cleavage site by testing a peptide substrate in vitro since PR accurately processes peptides that contain sequences of known cleavage sites (reviewed in 18). A hexadecapeptide (Glu430-Arg445, Fig. 1) encompassing the general region thought to contain the cleavage site, was incubated with purified PR and the resulting products were analyzed. Unexpectedly, a total of five products (Fig. 1) were apparent after prolonged digestion, although two (1 and 5) were the major products

Table 1

Amino Acid Composition of RT Subunits

	51 RT		51/66 RT Difference	
Amino Acid	Predicted[1]	Observed[2]	Predicted	Observed
Lys	49	49.2	12	12.4
His	7	7.3	7	7.4
Arg	14	13.3	5	5.4
Trp	18	N.D.	1	N.D.
Asx	32	32	12	12
Thr	32	30.4	9	9.2
Ser	15	12.3	6	6.3
Glx	66	66.4	18	18.9
Pro	35	34.2	3	3.7
Gly	25	25.3	9	10.0
Ala	15	15.7	11	11.4
Cys	2	N.D.	0	-
Val	28	27.8	11	11.0
Met	7	6.4	0	-
Ile	32	30.3	10	10.7
Leu	38	37.1	13	11.9
Tyr	17	16.8	5	5.2
Phe	13	12.9	0	-

N.D. Not determined.

[1] The values represent the residues predicted when the carboxyl terminus of 51 kDa RT is Phe440. The predicted residues include those from the modification of the RT amino terminus described in Methods.

[2] The values were normalized to the Asx residues. The average of two independent determinations is given.

after brief digestion or when PR concentration was low. The results demonstrate that two sites are recognized within this peptide sequence, although with different efficiencies. The kinetics of the appearance and, in some cases, disappearance of

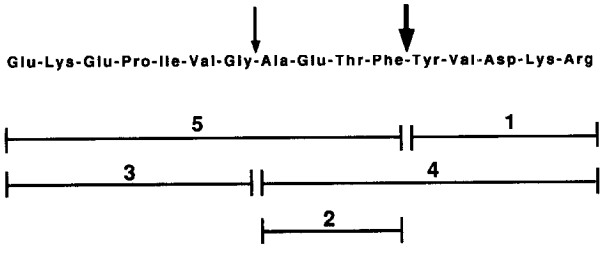

Figure 1. Sequence of peptide spanning the putative C-terminus of HIV-1 51 kDa RT. The peptide represents residues Glu430 through Arg445 of RT. Large arrow, predominant processing site; small arrow, minor processing site. Numbered bars below signify products and their order of elution by reverse phase HPLC.

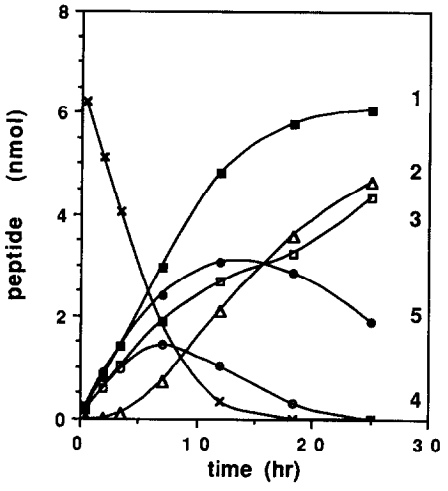

Figure 2. Kinetics of appearance and disappearance of peptide products after HIV-1 PR treatment. Substrate, **X**; fragment 1, ■; 2, △; 3, □; 4, ○; 5, ●.

products during the course of an extensive digestion is shown in figure 2; note that products 4 and 5 eventually give rise to product 2. We assign the major site of cleavage as occurring between Phe-Tyr: products 1 and 5 are the first to appear, and product 1 is the only one achieving 100 % yield during the time of the experiment. Furthermore, product 5 (which contains the minor site) is relatively long-lived, and product 4 (which contains the major site), has a shorter lifetime. Several control experiments (not shown) confirm that the peptide cleavages are caused by HIV-1 PR. The same results were obtained with pure PR, and no hydrolysis was observed with an extract of E. coli lacking HIV-1 PR. Appearance of all the products is inhibited by pepstatin A, which inhibits HIV-1 and other aspartyl proteinases.

From this work, we conclude that the C-terminus of 51 kDa RT is Phe440. Most likely, after this site is hydrolyzed in the RT molecule, no further processing occurs. If the remaining 15 kDa peptide is a discrete RNase H species, we would predict Tyr441 is the N-terminus (see 19). This cleavage site differs from that predicted by others on the basis of the residual cleavage observed in recombinant systems lacking the viral PR (9, 20, 21). By analogy to mammalian and E. coli polymerases, the internal cleavage site on the 66 kDa RT is located in the "tether region", a protease-sensitive hinge which joins the polymerase domain to the RNase H domain (22). Lowe and workers (9) demonstrated the general protease susceptibility of this region and predicted that E. coli proteases as well as trypsin and chymotrypsin cleaved at several locations within residues 430-440. Several groups employed this information

to predict PR cleavage between Glu432-Pro433 since many viral sites contain a Pro residue at the P1' subsite. However, peptides encompassing the sequences Leu429-Val435 (20) or the larger peptide Tyr427-Glu438 (21) , which contains the secondary site observed here, were not hydrolyzed by HIV-1 PR. We conclude the lack of activity is due to the fact that the peptides did not include the major cleavage site described here. Thus, while there is a sensitive hinge region in 66 kDa RT that is attacked by many proteases, the exact cleavage site differs between heterologous proteases and the viral PR. The information presented here should aid in the accurate design of recombinant 51 kDa RT and 15 kDa proteins.

The sequence of the newly identified cleavage site is also of particular interest since it does not share significant sequence homology to the other PR cleavage sites within gag and pol (4). The exception is the Phe which is common in position P1; also, Phe (but not Tyr) is found in the P1' position in two other cases. In those cleavage sites with Phe in the P1 subsite, Asn is frequently found in P2 and Ser (or Thr) in P4. However, in the site described here, we observe Thr and Ala, respectively, indicating that this site is very different from those previously observed. The enzyme's ability to accommodate a substrate with two aromatic residues in positions P1 and P1' may be surprising, especially considering that Phe in subsite P1 is usually followed by Pro in the other viral sites. However, the enzyme is a perfectly symmetrical dimer (23, 24) and, therefore, two identical hydrophobic pockets exist in the native protein that are capable of binding these larger residues. The site recognized with lower efficiency (Gly//Ala-Glu) demonstrates some agreement with other viral sites with a Gln or Glu residue in the P2' position. However the small Gly residue in the important P1 position may prevent significant recognition/cleavage by the viral PR. Clearly, this enzyme accomodates many types of sequences within the binding cleft, but still demonstrates tight stringency for only ten sites in the viral gag and pol precursors. Thus, while some requirements for PR substrate recognition can be defined, much remains to be discovered about the mechanism of substrate flexibilty inherent in this enzyme.

ACKNOWLEDGMENTS

We wish to thank W. Danho for the synthesis of the peptide, Stuart Le Grice for the RT plasmid pRT/BCG, and Robert Crowl and Iain Sim for critical reading of the manuscript.

REFERENCES

1. Skalka, A.M. (1989) Cell 56, 911-913.
2. Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A.F., Skolnick, E.M., and Sigal, I.S. (1988) Proc. Natl. Acad. Sci. (USA) 85, 4686-4690.
3. Graves, M.C., Lim, J.J., Heimer, E.P., and Kramer, R.A. (1988) Proc. Natl. Acad. Sci. (USA) 85, 2449-2453.
4. Henderson, L.E., Benveniste, R.E., Sowder, R., Copeland, T.D., Schultz, A.M., and Oroszlan, S. (1988) J. Virol. 62, 2587-2595.
5. Di Marzo Veronese, F., Copeland, T.D., DeVico, A.L., Rahman, R., Oroszlan, S., Gallo, R.C., and Sarngadharan, M.G. (1986) Science 231, 1289-1291.
6. Lightfoot, M.M., Coligan, J.E., Folks, T.M., Fauci, A.S., Martin, M.A., and Venkatesan, S. (1986) J. Virol. 60, 771-775.
7. Tanese, M., Sodroski, J., Haseltine, W.A., and Goff, S.P. (1986) J. Virol. 59, 743-745.
8. Farmerie, W.G., Loeb, D.D., Casavant, N.C., Hutchison, C.A., III, Edgell, M.H., and Swanstrom, R. (1987) Science 236, 305-308.
9. Lowe, D.M., Aitken, A., Bradley, C. Darby, G.K., Larder, B.A., Powell, K.L., Purifoy, D.J.M., Tisdale, M., and Stammers, D.K. (1988) Biochemistry 27, 8884-8889.
10. Hansen, J., Schulze, T., and Moelling, K. (1987) J. Biol. Chem. 262, 12393-12396.
11. Hansen, J., Schulze, T., Mellert, W., and Moelling, K. (1988) EMBO J. 7, 239-243.
12. Le Grice, S.F.J. and Grüninger-Leitch, F. (1990) Eur. J. Biochem. in press.
13. Matsudaira, P. (1987) J. Biol. Chem. 262, 10035-10038.
14. Graves, M.C., Lim, J.J., Zicopoulos, M.A., Stoller, T.J., Miedel, M.C., Pan, Y.-C.E., Danho, W., and Nalin, C.M. (1989) in Proteases of Retroviruses (V. Kostka, ed.) pp.83-92, Walter de Gruyter, Berlin.
15. Kotler, M., Katz, R.A., Danho, W., Leis, J., and Skalka, A.M. (1988) Proc. Natl. Acad. Sci. (USA) 85, 4185-4189.
16. Pan, Y.-C.E. and Stein, S. (1986) In: Methods of Protein Microcharacterization (J.E. Shively, ed.) Humana Press, Clifton, NJ. pp 105-119.
17. 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., and Wong-Staal, F. (1985) Nature 313, 277-284.
18. Kay, J. and Dunn, B.M. (1990) Biochim. Biophys. Acta (in press).
19. Schatz, O., Cromme, F.V., Grüninger-Leitch, F., and Le Grice, S.F.J. (1989) FEBS Lett. 257, 311-314.
20. Billich, S., Knoop, M.-T., Hansen, J., Strop, P., Sedlacek, J., Mertz, R., and Moelling, K. (1988) J. Biol. Chem. 263, 17905-17908.
21. Mizrahi, V., Lazarus, G.M., Miles, L.M., Meyers, C.A., and Debouck, C. (1989) Arch. Biochem. Biophys. 273, 347-358.
22. Johnson, M.S., McClure, M.A., Feng, D.-F., Gray J., and Doolittle, R.F. (1986) Proc. Natl. Acad. Sci. (USA) 83, 7648-7652.
23. Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W., Sigal, I.S., Darke, P.L., and Springer, J.P. (1989) Nature 337, 615-620.
24. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B., Baldwin, E., Weber, I., Selk, L.M., Clawson, L., Schneider, J., Kent, S.B.H. (1989) Science 245, 616-621.