Accelerated Publications

HIV-1 Reverse Transcriptase: Crystallization and Analysis of Domain Structure by Limited Proteolysis

Denise M. Lowe,*,‡ Alastair Aitken,§ Christopher Bradley,‡ Graham K. Darby,‡ Brendan A. Larder,‡ Kenneth L. Powell,‡ Dorothy J. M. Purifoy,‡ Margaret Tisdale,‡ and David K. Stammers‡

Department of Molecular Sciences, Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, U.K., and National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.

Received August 25, 1988; Revised Manuscript Received September 30, 1988

ABSTRACT: Bacterially expressed recombinant HIV-1 reverse transcriptase is active as both a homodimer of M_r 66 000 subunits and a heterodimer of M_r 66 000 and 51 000 subunits. The heterodimer is formed by cleavage of a C-terminal fragment from one M_r 66 000 polypeptide, which occurs during purification and crystallization of reverse transcriptase. Thus, crystals obtained from purified M_r 66 000 polypeptide preparations consisted of an apparently equimolar mixture of M_r 66 000 and 51 000 polypeptides, which were apparently analogous to the M_r 66 000 and 51 000 polypeptides detected in HIV-infected cells and in virions. Limited proteolysis of the homodimer with α-chymotrypsin also resulted in cleavage to a stable M_r 66 000/51 000 mixture, and proteolysis with trypsin resulted in the transient formation of some M_r 51 000 polypeptide. These results are consistent with the reverse transcriptase molecule having a protease-sensitive linker region following a structured domain of M_r 51 000. Further digestion with trypsin resulted in cleavage of the M_r 51 000 polypeptide after residue 223, yielding peptides of apparent M_r 29 000 and 30 000. A minor peptide of M_r 40 000 was also produced by cleavage of the M_r 66 000 polypeptide after residue 223. About half the original M_r 66 000 polypeptides remained resistant to proteolysis and existed in complex with the above peptides in solution. During both chymotrypsin and trypsin digestion there was an increase in the reverse transcriptase activity caused by a doubling of V_{max} with little change in K_m for dTTP. The resistance of one subunit of the homodimer to protease digestion indicates that either these subunits are not symmetrically related in the homodimer or a conformational rearrangement occurs on proteolysis of one subunit, so as to protect the second subunit from a similar event.

Human immunodeficiency virus, the retrovirus which causes AIDS[1] (Barre-Sinoussi et al., 1983; Levy et al., 1984; Popovic et al., 1984), encodes a reverse transcriptase (EC 2.7.7.49), which is present in the virion (Varmus & Swanstrom, 1985), and converts the RNA genome into a double-stranded DNA copy. The region of the *pol* gene encoding RT has been cloned, and high levels of expression of a M_r 66 000 polypeptide (562 amino acids) have been achieved in *Escherichia coli* (Larder et al., 1987a,b). This protein exhibits reverse transcriptase activity and also low-level RNase H activity (Tisdale et al., 1988b). RT activity in virions is associated with two polypeptides, p66 and p51, which share a common N-terminus, suggesting that the smaller polypeptide is derived from the larger by a proteolytic cleavage (Di Marzo Veronese et al., 1986; Lightfoote et al., 1986). Similar processing has been seen in various prokaryotic and eukaryotic expression systems for HIV-1 RT (Hansen et al., 1987; Farmerie et al., 1987; Le Grice et al., 1987; Barr et al., 1987; Mous et al., 1988). During purification from our *E. coli* expression system some production of p51, associated with p66, occurs, although we are able to obtain homogeneous p66 RT (Tisdale et al., 1988a).

In order to gain a better understanding of these processing events, we have investigated the subunit composition of RT and the effect of limited proteolysis by α-chymotrypsin and trypsin on the molecular structure and catalytic activity of RT. This information will be useful in our studies on the crystallization of RT to determine its three-dimensional structure for use in the rational design of novel inhibitory molecules.

Experimental Procedures

Purification of Reverse Transcriptase. HIV-1 RT was expressed in *E. coli* (TG-1) from the recombinant plasmid pRT1 (Larder et al., 1987b). After a 1 M NaCl extraction procedure (Larder et al., 1987b), RT was purified by chromatography on a RTMAb8 immunoabsorbent column (Tisdale et al., 1988a).

Assay for Reverse Transcriptase Activity. Assays were performed at 37 °C in a buffer containing 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol, and 50 μg/mL poly(rA)–$(dT)_{18}$. The incorporation of [^{3}H]dTTP (2 Ci/mmol) was followed by absorption of the polymerized DNA onto DEAE-cellulose paper (Larder et al., 1987b). Kinetic data were analyzed by nonlinear least-squares and median estimate methods (Stammers et al., 1987).

Crystallization of RT. Purified p66 homodimer RT was set up to crystallize at 4 °C by using standard vapor diffusion techniques (McPherson, 1976) with 35% (w/v) ammonium sulfate as precipitant. The protein was dissolved in 50 mM HEPES, pH 6.8, at a concentration of 20 mg of protein/mL.

‡ Wellcome Research Laboratories.
§ National Institute for Medical Research.

[1] Abbreviations: AIDS, acquired immune deficiency syndrome; HIV-1, human immunodeficiency virus type 1; RT, reverse transcriptase; Tris, tris(hydroxymethyl)aminomethane; $NaDodSO_4$, sodium dodecyl sulfate; FPLC, fast protein liquid chromatography; p66, p51, p40, p30, and p29, proteins of apparent molecular mass 66 000, 51 000, 40 000, 30 000, and 29 000, respectively, as determined by $NaDodSO_4$–polyacrylamide gel electrophoresis; RT MAb, monoclonal antibody raised against reverse transcriptase.

0006-2960/88/0427-8884$01.50/0

Limited Proteolysis. RT was dissolved in 20 mM Tris-HCl (pH 8), 100 mM NaCl, and 0.5 mM dithiothreitol at about 1 mg of protein/mL and incubated with either 10 μg/mL α-chymotrypsin (Sigma, type I-S, from bovine pancreas) or 10 μg/mL trypsin (Sigma, type III, from bovine pancreas) either on ice or at 21 °C. Digestion with α-chymotrypsin or trypsin was stopped by diluting samples 100-fold into 10 mM phosphate buffer (pH 7), 0.1 mM EDTA, and 10% (v/v) glycerol containing 2 mM phenylmethanesulfonyl fluoride or 1 μg of trypsin inhibitor/mL (Sigma, type I-S, from soybean), respectively. Samples for analysis on $NaDodSO_4$–polyacrylamide gels were concentrated by precipitation with trichloroacetic acid by the method of Peterson (1977), which recovers protein quantitatively.

$NaDodSO_4$–Polyacrylamide Gel Electrophoresis. This was performed with a Tris–glycine buffer system based on that of Laemmli (1970). Gels were stained with Coomassie Brilliant Blue R250 and then scanned with a Joyce-Loebl gel scanner.

Western Blotting. This was performed by standard procedures (Banks et al., 1984), using the monoclonal antibodies RTMAb3, which reacts with an epitope mapped to between amino acids 233 and 315 from the N-terminus, and RTMAb8, which reacts with a conformational epitope, part of which maps between amino acids 531 and 539 at the C-terminus of RT (Tisdale et al., 1988a).

Immunoprecipitation. The T-cell line MT4 (10^7 cells/culture) was infected with HIV (strain HTLVIIIB) and after 48-h incubation at 37 °C (when a cytopathic effect was evident) labeled with [^{35}S]methionine (100 μCi/mL) for 16 h at 37 °C. Immunoprecipitations were carried out on pelleted cells as described by Banks et al. (1987), using 100 μL of lysis buffer/culture and immunoadsorption overnight at 4 °C followed by precipitation with *Staphylococcus aureus* protein A. Precipitates were washed extensively as described by Banks et al. (1987) and then analyzed by $NaDodSO_4$–polyacrylamide gel electrophoresis and fluorography.

FPLC Gel Filtration. A preparative Superose 12 FPLC column (1.6 × 50 cm) (Pharmacia) was equilibrated in 50 mM Tris-HCl (pH 7.5), 0.5 M NaCl, and 0.5 mM dithiothreitol. The sample (2 mL) was loaded at 4 °C and eluted isocratically at 0.5 mL/min, collecting 2-mL fractions. For analytical purposes a Superose 6 column (1 × 30 cm) was loaded with 0.1-mL samples and eluted at 0.5 mL/min, collecting 0.5-mL fractions.

Sequence Analysis of RT Polypeptides. The N-terminal sequence was obtained from peptides separated on a $NaDodSO_4$–polyacrylamide gel and electroblotted onto polyvinylidene difluoride membranes (Millipore), as described previously (Matsudaira, 1987). Sequence analysis was performed with Applied Biosystems 470A gas-phase and 477A pulsed liquid-phase peptide sequenators with an on-line 120A PTH-amino acid analyzer. C-Terminal sequencing was performed on peptide electroeluted from a phosphate-buffered $NaDodSO_4$–polyacrylamide gel (Weber et al., 1972), using carboxypeptidase Y and P digestion (Morrice et al., 1988). Amino acid analysis was performed on an Applied Biosystems 420A derivatizer–analyzer fitted with an on-line 130A PTC analyzer.

Results

Comparison of the Polypeptide Composition of Virus-Expressed and Crystalline RT. Purified recombinant RT has been used to obtain monoclonal antibodies to RT, one of which (RTMAb8) reacts on a Western blot with p66 but not with p51, which arises during the purification or is present in virus extracts (Tisdale et al., 1988a). This monoclonal antibody was used to immunoprecipitate RT from HIV-infected cells. When analyzed by $NaDodSO_4$–polyacrylamide gel electrophoresis, both p66 and p51 were detected in approximately equal amounts, indicating the presence in virus-infected cells of a p66/p51 complex (Figure 1).

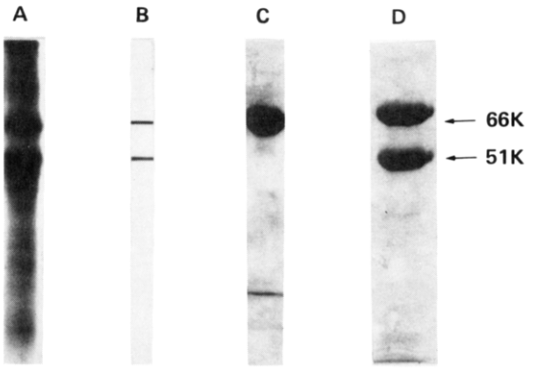

FIGURE 1: Analysis of polypeptide components of RT. Panel A shows an autoradiograph of HIV-infected cells immunoprecipitated with RTMAb8. Panel B shows the detection of RT-related polypeptides in a preparation of HIV particles (as supplied in the Du Pont Western blotting kit) using RTMAb3 [from Tisdale et al. (1988a)]. Panels C and D show stained gels of RT purified from recombinant *E. coli* and redissolved crystals of RT, respectively.

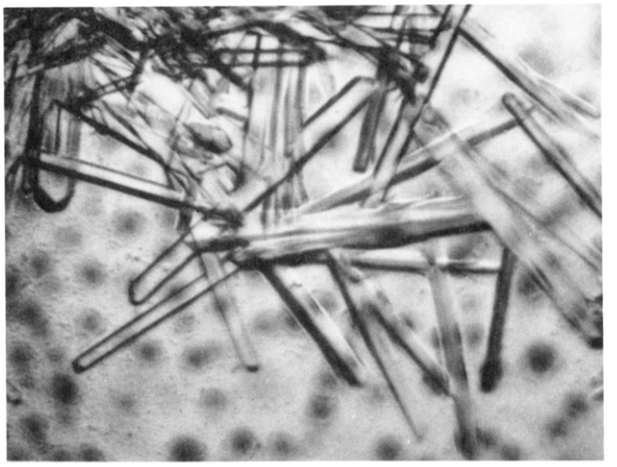

FIGURE 2: Crystals of HIV RT. The crystals shown in the micrograph were grown in ammonium sulfate at 4 °C.

Purified recombinant p66 RT was used for crystallization experiments. Crystals of RT grew as rods up to 0.4 mm in length (Figure 2). The redissolved crystals were analyzed by $NaDodSO_4$ gel electrophoresis and found to contain approximately equimolar amounts of p66 and p51, which were apparently identical with those detected in HIV-infected cells and in virions (Figure 1). The mother liquor also contained some p51, but not in equimolar amounts with the p66. The p51 from crystals did not react with RTMAb8. Sequence analysis indicated that cleavage to produce p51 had occurred at multiple sites approximately between residues 430 and 440 in the RT sequence (data not shown).

Determination of Native Relative Molecular Mass. Samples of homogeneous p66 RT and a preparation containing approximately equal amounts of p66 and p51 (judged by $NaDodSO_4$–polyacrylamide gel electrophoresis) were analyzed by gel filtration on a Superose 12 FPLC column. Both preparations eluted with a M_r between 110 000 and 130 000, consistent with p66 homodimers and p66/p51 heterodimers. No p66 or p51 monomers could be detected.

Limited Proteolysis by α-Chymotrypsin. A preparation of RT, consisting of p66 homodimers, was digested with α-chy-

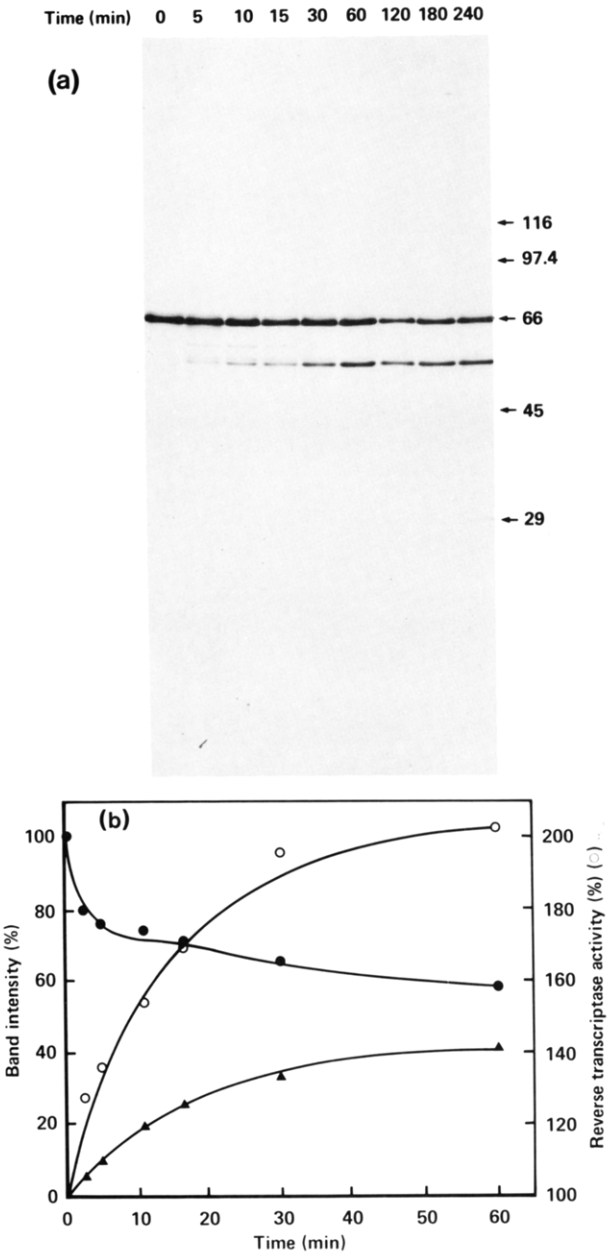

FIGURE 3: α-Chymotrypsin digestion of RT. Digestion was performed at 0 °C with 10 μg/mL α-chymotrypsin. Samples (8 μg of protein) were removed at the times indicated and analyzed on a 15% NaDodSO$_4$–polyacrylamide gel (Tris-glycine). (a) The stained gel is shown. (b) The gel shown in (a) was scanned to determine the band intensity corresponding to p66 (●) and p51 (▲). RT activity (○) was measured as described under Experimental Procedures.

motrypsin. During the first hour of exposure to α-chymotrypsin there was a transient production of some polypeptide of M_r about 60 000, but the major product was an apparently equimolar mixture of the original p66 and a newly formed polypeptide of M_r about 51 000 (Figure 3a). This pattern of peptides then remained stable even for many hours in the presence of α-chymotrypsin.

The relationship between this cleavage and the RT activity is shown in Figure 3b. An increase in the enzyme activity is seen which correlates with the cleavage to p51, reaching a maximum of about double the original activity of the p66 homodimer. A steady-state kinetic analysis of the undigested RT and a sample digested for 1 h with α-chymotrypsin revealed that the increase in activity is due to a doubling of the specific activity with very little change in K_m for dTTP (Table I).

Table I: Effect of Limited Proteolysis on the Kinetics of RT Activity[a]

enzyme	K_m for dTTP (μM)	sp act. [nmol min^{-1} (mg of protein)$^{-1}$]	x-fold increase in sp act.
p66 RT (undigested)	5 ± 1.0	700 ± 40	
p66/p51 RT (chymotrypsin digested)	6 ± 0.7	1610 ± 70	2.3
p66/p30/p29 RT (trypsin digested)	7 ± 0.9	1470 ± 80	2.1

[a] Assay conditions as described under Experimental Procedures.

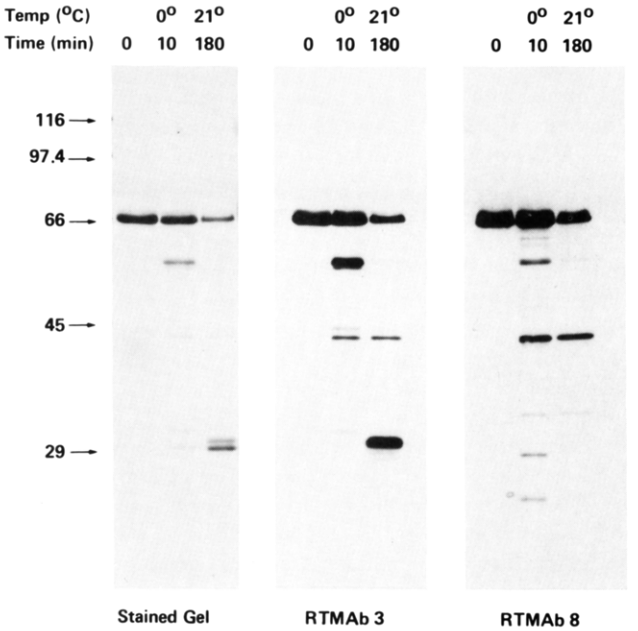

FIGURE 4: Western blot analysis of trypsin-digested RT with RTMAb3 and RTMAb8. The digests were performed with 10 μg/mL trypsin for the times and at the temperatures indicated, and samples (10 μg of protein) were analyzed on a 15% NaDodSO$_4$–polyacrylamide gel (Tris–glycine).

A Western blot of the p66/p51 digestion product was reacted with RTMAb8. This produced a reaction with p66 but none with p51. Sequence analysis of the C-terminus of the p51 indicated that cleavage had occurred at multiple sites.

Crystals of chymotrypsin-digested RT were grown under the same conditions used to grow the crystals shown in Figure 2. These crystals were rodlike but had a tendency to grow as multiple crystals.

Limited Proteolysis by Trypsin. Incubation of p66 RT with trypsin resulted in very rapid (<1 min) formation of a peptide or peptides of M_r about 51 000, which disappeared on continued digestion and was followed by appearance of initially a peptide of apparent M_r 29 000 and then another of apparent M_r 30 000 (Figure 4). In some digestions a peptide of M_r about 40 000 was also produced (Figure 4). In no case was more than half of the p66 protein cleaved. The mixture of p29/p30/p66 remained stable for many hours even at room temperature.

Analysis of the RT activity during trypsin digestion showed a rapid increase in activity, correlating with the formation of p51/p29/p30 and a decrease in the amount of p66. Kinetic analysis of the product after 4-h digestion showed that the specific activity had doubled with a small increase in the K_m for dTTP (Table I).

The mixture of p66, p40, p30, and p29 was passed down a Superose 6 FPLC column in an attempt to separate the peptides. However, a single peak of protein was eluted that

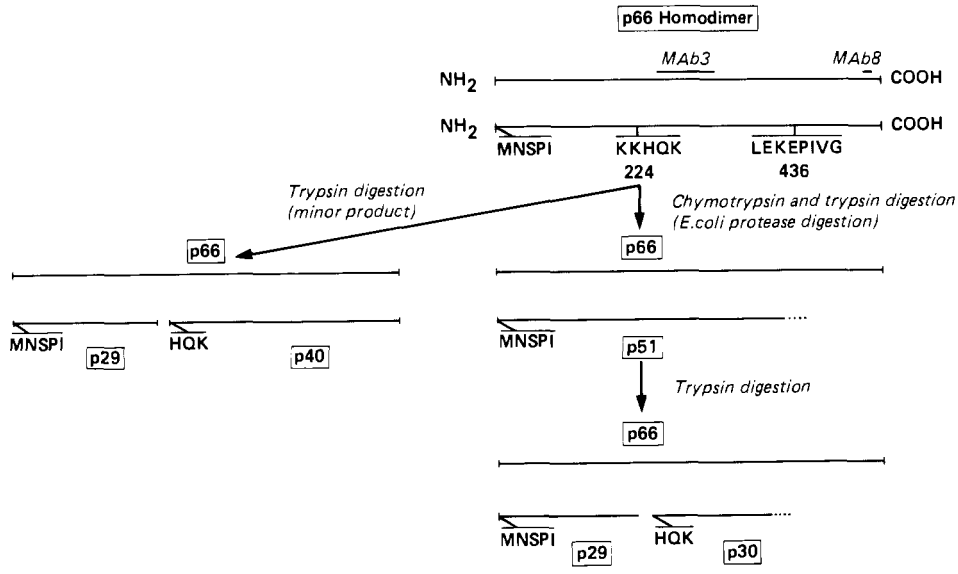

FIGURE 5: Proteolytic processing of RT.

had RT activity and showed the same polypeptide composition on $NaDodSO_4$–polyacrylamide gel electrophoresis as the loaded material.

Western blots were prepared of samples from limited trypsin digestion (10 min, 0 °C) and after prolonged trypsin digestion (3 h, 21 °C) and reacted with RTMAb3 and RTMAb8 (Figure 4). The p51 produced by limited digestion reacted intensely with RTMAb3 and also to a lesser extent with RTMAb8. Both RTMAb3 and RTMAb8 reacted with the p40 present under both digestion conditions. The p30 reacted with RTMAb3 whereas p29 did not. Neither p29 or p30 reacted with RTMAb8 (Figure 4).

N-Terminal sequence analysis of the peptides produced by trypsin digestion gave the same sequence, MNSPISP etc., for p66, p51, and p29 (Table II). The p30 and p40 peptides had the N-terminal sequence, HQKEPP etc. (Table II), which occurs at position 224 in the RT amino acid sequence, immediately following two lysines. These sequence data are in agreement with the known RT sequence (Wain-Hobson et al., 1985) except for the two additional amino acids present at the N-terminus of this recombinant protein (Larder et al., 1987b) and show that the N-terminal methionine has not been removed.

Table II: Automated Sequencer Analysis of Peptides from Trypsin-Digested RT[a]

cycle no.	1	2	3	4	5	6	7
sequence	M	N	S	P	I	S	P
yield (pmol)							
p66	4	3		3	2	2	2
p51	3	4	2	3	2	2	2
p29	4	5	4	5	4		
cycle no.	1	2	3	4	5	6	7
sequence	H	Q	K	E	P	P	M
yield (pmol)							
p30	5	5	1	3	2	3	1
p40	0.5	1		1	2	2	

[a] Details given under Experimental Procedures.

Discussion

Our bacterial expression system for HIV-1 RT codes for a recombinant polypeptide of M_r 66 000, which we have found to be active in solution as a homodimer. In agreement with other workers (Hansen et al., 1987; Farmerie et al., 1987; Mous et al., 1988), we see some formation of a M_r 51 000 RT-related polypeptide during our purification process, but most dramatically under conditions used for the crystallization of RT. The p66/p51 seen in crystalline RT appears identical with that seen in HIV-infected cells and virions (Figure 1) (Di Marzo Veronese et al., 1986; Chandra et al., 1986; Hansen et al., 1988; Tisdale et al., 1988a). This processing appears to involve several cleavages in the region approximately between amino acid residues 430 and 440 and only occurs on one subunit of the p66 homodimer, resulting in a p66/p51 heterodimer. The protease responsible for this cleavage is assumed to be of *E. coli* origin. Our expression vector, pRT1, unlike other constructs that have been described (Hansen et al., 1987; Farmerie et al., 1987; Mous et al., 1988), does not code for the HIV protease (Larder et al., 1987b). Hence, the cleavage seen in our recombinant RT is not a virus-specific processing event unless it was self-cleavage, a possibility that we cannot exclude. This is confirmed by our observations that limited proteolysis by both chymotrypsin and trypsin results in cleavage in the same region, producing an N-terminal p51 fragment. In the presence of chymotrypsin, this p66/p51 heterodimer is resistant to further proteolysis, whereas a second site in the p51, on the C-terminal side of amino acid residue 223 (lysine), is cleaved by trypsin (see Figure 5).

The time–course of trypsin cleavage indicates that while the major early product is p51, a small population is cleaved first at residue 223, resulting in the early formation of p29 and p40 (Figure 4). N-Terminal sequencing assigns the p29 to the N-terminus and p40 to the C-terminus, and this is confirmed by the reaction of p40 with both RTMAb's 3 and 8 (Figure 4). There is, however, an anomalous reaction of RTMAb8 with the p51 produced by trypsin digestion. Since RTMAb8 is thought to recognize a conformational epitope (Tisdale et al., 1988a), there may be some element in the trypsin-cleaved p51 that allows this weak reaction.

With continued trypsin digestion the p51 is cleaved virtually in half to yield p29 (223 amino acids) and p30 (212 amino acids), which remain complexed with p66. (p30 is actually the smaller peptide by molecular mass but runs in an anomalous position on $NaDodSO_4$–polyacrylamide gels, slightly above the p29.) During this time–course the amount of p40 remains approximately constant, suggesting that it has taken on a conformation that is now resistant to a second cleavage in the 430–440 region. This would account for the slightly denser p29 band compared to p30 detected on Coomassie Blue stained gels, even after prolonged trypsin digestion, rather than tending toward equimolar amounts (Figure 4).

Attempts to determine an exact cleavage site corresponding to the C-terminus of the p51 showed that neither the p51 recovered from crystalline preparations of RT nor that produced by chymotrypsin digestion had a unique C-terminus. This implies that there is a protease-susceptible region of the RT molecule around residues 430–440, which is presumably accessible in solution. There are possible cleavage sites for both trypsin and chymotrypsin in this sequence. Interestingly, a sequence (Lys-Glu-Pro-Ile) at residues 435–438 resembles the predicted HIV protease cleavage site (Pearl & Taylor, 1987). This could account for the presence of p66/p51 rather than p66 in HIV-infected cells and virions. The RT polypeptide from HIV, like that from other retroviruses (Moelling, 1976; Lai & Verma, 1978; Tanese & Goff, 1988), is known to have two distinct enzyme activities, namely, RT and RNase H. In the case of Moloney murine leukemia virus RT, RT and RNase H activities have been expressed separately and shown to reside in the N- and C-terminal domains, respectively (Tanese & Goff, 1988). Our previous work on mutagenesis of conserved regions in the N-terminal half of the HIV RT molecule has identified several residues that are essential for the reverse transcriptase activity of the protein (Larder et al., 1987a). We, and other workers, have shown evidence that the C-terminal domain encodes the RNase H activity (Tisdale et al., 1988b; Johnson et al., 1986; Hansen et al., 1988). The protease-sensitive region may simply represent a linker region between two domains of RT. So far, however, we have failed to identify a stable M_r 15 000 peptide corresponding to the C-terminal domain either during crystallization experiments or by limited proteolysis. Two small peptides (M_r <29 000) were detected by RTMAb8 after limited (0°, 10 min) digestion with trypsin, but these disappear on prolonged digestion (Figure 4). Similarly, with chymotrypsin digestion there appears to be at least one intermediate cleavage on the pathway to a stable p66/p51 (Figure 3a), suggesting that the M_r 15 000 C-terminal fragment is probably cleaved into two or more fragments making it difficult to detect. This pattern of limited proteolysis of HIV-1 RT contrasts with that seen for avian myeloblastosis virus RT where proteolysis resulted in substantial loss of polymerase activity and the RNase H activity could, subsequently, be recovered on a proteolytic fragment (Lai & Verma, 1978).

An interesting feature of the proteolysis of RT, observed both in crystals and after treatment with trypsin and chymotrypsin, is the apparent resistance of one subunit of the homodimer to proteolytic cleavage. This suggests either that conformational rearrangement of the enzyme occurs after proteolysis of one chain so as to protect the other chain from a similar cleavage or that the two chains of the homodimer are not symmetrically related in the native enzyme. The uncut p66 subunit remains bound to the proteolytic cleavage products (p51 or p30 and p29) as a stable complex in solution, whereas the C-terminal M_r 15 000 fragment is apparently lost. This would suggest that most of the important subunit interactions occur in the N-terminal M_r 51 000 part of the molecule.

We have shown that the p66/p51 and p66/p30/p29 complexes produced by chymotrypsin and trypsin digestion, respectively, have a higher V_{max} than the p66 homodimer by a factor of 2, with no significant change in the K_m for dTTP. Thus, it is possible that the loss of the C-terminal fragment from one subunit removes some structural constraint on the molecule, allowing it to function more efficiently as a reverse transcriptase. Loss of the C-terminal fragment from both subunits, although not seen experimentally, has been mimicked in a mutant, CTRT1, which codes for a polypeptide of M_r 51 000 (Tisdale et al., 1988a), which has much reduced RT activity compared to wild type. Other reports have failed to detect any RT activity associated with p51 from virions or infected cells (Starnes et al., 1988; Hansen et al., 1988). Thus, there appears to be some biochemical advantage in the heterodimer. It is still surprising that the second cut by trypsin, occurring in the center of the proposed N-terminal domain, has no detrimental effect on RT activity and might, therefore, be cutting in a surface loop not involved in the active site.

Data from the crystallographic study of the enzyme may eventually shed light on the interactions between the two subunits. The crystals of p66/p51 heterodimers (Figure 2) are insufficiently ordered to obtain a high-resolution structure, and work is continuing to obtain better crystals of this form and of the p66 homodimer form. A detailed comparison of the heterodimer and homodimer must await the resolution of these crystal structures.

ACKNOWLEDGMENTS

We thank V. Parmar, P. Ertl, A. Emmerson, C. Roberts, and C. Ross for excellent technical support, Dr. E. D'Souza and N. Oliver for help with virus and tissue cultures, F. Beg for the carboxypeptidase analysis, A. Harris for the peptide sequencing, and M. Parkinson and P. Ward for help in the preparation of the typescript.

Registry No. RT, 9068-38-6; dTTP, 365-08-2.

REFERENCES

Banks, L. M., Vaughan, P. J., Meredith, D., & Powell, K. L. (1984) *J. Gen. Virol. 65*, 1625–1630.

Banks, L., Spence, P., Androphy, E., Hubbert, N., Matlashewski, G., Murray, A., & Crawford, L. (1987) *J. Gen. Virol. 68*, 1351–1359.

Barr, P. J., Power, M. D., Lee-Ng, C. T., Gibson, H. L., & Luciw, P. A. (1987) *Bio/Technology 5*, 486–489.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axier-Blin, C., Vexinet-Brun, F., Rouzioux, C., Rozenbaum, W., & Montagnier, L. (1983) *Science (Washington, D.C.) 220*, 868–871.

Chandra, A., Gerber, T., Kaul, S., Wolf, C., Demirhan, I., & Chandra, P. (1986) *FEBS Lett. 200*, 327–332.

Di Marzo Veronese, F., Copeland, T. D., De Vico, A. L., Rahman, R., Oroszlan, S., Gallo, R. C., & Sarngadharan, M. G. (1986) *Science (Washington, D.C.) 231*, 1289–1291.

Farmerie, W. G., Loeb, D. D., Casavant, N. C., Hutchison, C. A., III, Edgell, M. H., & Swanstrom, R. (1987) *Science (Washington, D.C.) 236*, 305–308.

Hansen, J., Schulze, T., & Moelling, K. (1987) *J. Biol. Chem. 262*, 12393–12396.

Hansen, J., Schulze, T., Mellert, W., & Moelling, K. (1988) *EMBO J. 7*, 239–243.

Johnson, M. S., McClure, M. A., Feng, D.-F., Gray, J., & Doolittle, R. F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7648–7652.

Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.

Lai, M.-H. T., & Verma, I. M. (1978) *J. Virol. 25*, 652–663.

Larder, B. A., Purifoy, D. J. M., Powell, K. L., & Darby, G. (1987a) *Nature (London) 327*, 716–717.

Larder, B., Purifoy, D., Powell, K., & Darby, G. (1987b) *EMBO J. 6*, 3133–3137.

Le Grice, S. F. J., Beuck, V., & Mous, J. (1987) *Gene 55*, 95–103.

Levy, J. A., Hoffman, A. D., Kramer, S. M., Landis, J. A., Shimabukuro, J. M., & Oshiro, L. S. (1984) *Science (Washington, D.C.) 225*, 840–842.

Lightfoote, M. M., Coligan, J. E., Folks, T. M., Fauci, A. S., Martin, M. A., & Venkatesan, S. (1986) *J. Virol. 60*, 771–775.
Matsudaira, P. (1987) *J. Biol. Chem. 262*, 10035–10038.
McPherson, A., Jr. (1976) *Methods Biochem. Anal. 23*, 249–345.
Moelling, K. (1976) *J. Virol. 18*, 418–425.
Morrice, N., Geary, P., Cammack, R., Harris, A., Beg, F., & Aitken. A. (1988) *FEBS Lett. 231*, 336–340.
Mous, J., Heimer, E. P., & Le Grice, S. F. J. (1988) *J. Virol. 62*, 1433–1436.
Pearl, L. H., & Taylor, W. R. (1987) *Nature (London) 329*, 351–354.
Peterson, G. L. (1977) *Anal. Biochem. 83*, 346–356.
Popovic, M., Sarngadharan, M. G., Read, E., & Gallo, R. C. (1984) *Science (Washington, D.C.) 224*, 497–500.
Stammers, D. K., Dann, J. G., Harris, C. J., & Smith, D. R. (1987) *Arch. Biochem. Biophys. 258*, 413–420.
Starnes, M. C., Gao, W., Ting, R. Y. C., & Cheng, Y.-C. (1988) *J. Biol. Chem. 263*, 5132–5134.
Tanese, N., & Goff, S. P. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85I*, 1777–1781.
Tisdale, M., Ertl, P., Larder, B., Purifoy, D. J. M., Darby, G., & Powell, K. (1988a) *J. Virol. 62*, 3662–3667.
Tisdale, M., Larder, B., Lowe, D., Stammers, D., Purifoy, D., Ertl, P., Bradley, C., Kemp, S., Darby, G. K., & Powell, K. (1988b) *J. Antimicrob. Chemother.* (in press).
Varmus, H., & Swanstrom, R. (1985) in *RNA Tumor Viruses: Molecular Biology of Tumor Viruses* (Weiss, R., Teich, N., Varmus, H., & Coffin, J., Eds.) 2nd ed., pp 75–134, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., & Alizon, M. (1985) *Cell (Cambridge, Mass.) 40*, 9–17.
Weber, K., Pringle, J. R., & Osborn, M. (1972) *Methods Enzymol. 26*, 3–27.

Reduced Bovine Pancreatic Trypsin Inhibitor Has a Compact Structure†

Dan Amir‡ and Elisha Haas*,‡,§

Department of Life Science, Bar-Ilan University, Ramat Gan 52100, Israel, and Departments of Chemical Physics and Biophysics, Weizmann Institute of Science, Rehovot 76100, Israel

Received August 29, 1988; Revised Manuscript Received October 4, 1988

ABSTRACT: The conformation of reduced bovine pancreatic trypsin inhibitor (R-BPTI) under reducing conditions was monitored by measurements of nonradiative excitation energy-transfer efficiencies (E) between a donor probe attached to the N-terminal Arg^1 residue and an acceptor attached to one of the lysine residues (15, 26, 41, or 46) [Amir, D., & Haas, E. (1987) *Biochemistry 26*, 2162–2175]. High-excitation energy-transfer efficiencies that approach those found in the native state were obtained for the reduced labeled BPTI derivatives in 0.5 M guanidine hydrochloride (Gdn·HCl) and 4 mM DTT. Unlike the dependence expected for a random coil chain, E does not decrease as a function of the number of residues between the labeled sites. The efficiency of energy transfer between probes attached to residues 1 and 15 in the reduced state is *higher* than that found for the same pair of sites in the native state or reduced unfolded (in 6 M Gdn·HCl) state. This segment also shows high dynamic flexibility. These results indicate that the overall structure of reduced BPTI under folding (but still reducing) conditions shows a high population of conformers with interprobe distances similar to those of the native state. Reduced BPTI seems to be in a molten globule state characterized by a flexible, compact structure, which probably reorganizes into the native structure when the folding is allowed to proceed under oxidizing conditions.

Kinetic pathways of folding via intermediate states were proposed for the mechanisms by which proteins fold from unfolded states to their unique three-dimensional structures (Matheson & Scheraga, 1978; Creighton, 1978; Kim & Baldwin, 1982; Goldberg, 1985). Intermediate states, which restrict the conformational space available for the unfolded state, can direct the pathway of folding and accelerate the transition to the native state. Unraveling the structural characteristics of the initial states in the folding transition should help decipher the mechanism of folding and the specific amino acid sequence messages that code for the pathway.

Creighton (1978) has studied the pathway of folding of BPTI and showed that the first disulfide bonds are already formed in a nonrandom distribution; the 30–51 disulfide pair is dominant.

Ptitsyn and co-workers (Dolgikh et al., 1981) have studied the conformation of α-lactalbumin unfolded by acid and by temperature elevation and found evidence for a compact globule state with nativelike secondary structures and with slowly fluctuating tertiary structure. Ohgushi and Wada (1983) made similar observations and coined the term "molten globule state" to describe this conformation. Both groups suggested that the molten globule state is a common type of an early intermediate in the folding pathway of globular proteins.

In the present study we are interested in the structure of early folding intermediates of BPTI[1] prior to formation of

† Supported in part by the U.S.–Israel Binational Science Foundation (Grant 212/86) and by the National Institute of General Medical Sciences (Grant GM39372-1). This paper is part of a Ph.D. Thesis submitted by D.A. to the Senate of Bar-Ilan University.
* Address correspondence to this author at Bar-Ilan University.
‡ Bar-Ilan University.
§ Weizmann Institute of Science.

0006-2960/88/0427-8889$01.50/0