JOURNAL OF VIROLOGY, Feb. 1983, p. 885–888
0022-538X/83/020885-04$02.00/0

Vol. 45, No. 2

Amino-Terminal Amino Acid Sequence of p10, the Fifth Major *gag* Polypeptide of Avian Sarcoma and Leukemia Viruses

ERIC HUNTER,[1] J. CLAUDE BENNETT,[2] AJIT BHOWN,[2] ROBERT B. PEPINSKY,[3] AND VOLKER M. VOGT[3]*

Department of Microbiology[1] and Department of Medicine,[2] University of Alabama in Birmingham, Birmingham, Alabama 35294, and Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853[3]

Received 6 July 1982/Accepted 20 October 1982

We have identified p10 as a fifth *gag* protein of avian sarcoma and leukemia viruses. Amino-terminal protein sequencing of this polypeptide purified from the Prague C strain of Rous sarcoma virus and from avian myeloblastosis virus implies that it is encoded within a stretch of 64 amino acid residues between p19 and p27 on the *gag* precursor polypeptide. For p10 from the Prague C strain of Rous sarcoma virus the first 30 residues were found to be identical with the predicted amino acid sequence from the Prague C strain of Rous sarcoma virus DNA sequence, whereas for p10 from avian myeloblastosis virus the protein sequence for the same region showed two amino acid substitutions. Amino acid composition data indicate that there are no gross composition changes beyond the region sequenced. The amino terminus of p10 is located two amino acid residues past the carboxy terminus of p19, whereas its carboxy terminus probably is located immediately adjacent to the first amino acid residue of p27.

The internal structural proteins of retroviruses are encoded by nucleic acid sequences near the 5′ end of the viral genome, the *gag* gene. This gene is translated to yield a polyprotein precursor that is subsequently cleaved into the individual proteins of the virus (5). In the avian sarcoma and leukemia viruses (ASLV), four major proteins have been shown previously to derive from the *gag* precursor Pr76 (14). Gel filtration of ASLV proteins in 6 M guanidine hydrochloride yields, in addition to these four *gag* proteins and the two envelope glycoproteins, a protein peak that has been termed p10. This peak, which actually corresponds to a molecular weight of about 7,000 (14), is absent from chromatograms of virus metabolically labeled with radioactive lysine (6; Vogt, Eisenman, and Diggelmann, unpublished observations). At least two polypeptides appear to comprise the p10 peak, which has led to confusion in the literature (3, 8). In the accompanying paper (8), we describe the purification of a major, small protein from ASLV that we term p10 in accordance with the existing nomenclature. Several unusual properties of this protein explain why it had not been well characterized previously.

Comparison of the DNA sequence of Rous sarcoma virus (RSV; D. Schwartz, R. Tizard, and W. Gilbert, personal communication) and the partial sequences of the four major *gag* proteins shows that they are arrayed on Pr76 as follows: N-p19—p27-p12-p15-C (E. Hunter, A. Bhown, R. Eiseman, C. Pachel, D. Schwartz, and R. Tizard, personal communication). The DNA sequence between p19 and p27 encodes a stretch of 64 amino acid residues that had not previously been identified with any viral protein. The similarities in molecular weight and the fact that both the p10 peak in gel filtration and the polypeptide predicted to lie between p19 and p27 lack lysine suggested to us that p10 may be a fifth *gag* protein derived from this unassigned stretch of amino acids. In this communication we present protein sequence data that confirm this hypothesis.

The p10 polypeptide of the PrC-strain of RSV, from cells grown in culture, and of avian myeloblastosis virus (AMV), from leukemic chicken plasma, were purified chromatographically as described in the accompanying paper (8) and then submitted to automated Edman degradation. The results of these studies are summarized in Fig. 1 and 2. Figure 1A, line b, shows the resulting first 30 amino acid residues of p10 from the Prague C strain of RSV (PrC-RSV). Twenty-three of these residues are identical to the corresponding residues in the amino acid sequence (Fig. 1, line a) predicted from the nucleotide sequence of PrC-RSV (Schwartz et al., personal communication). The seven remaining residues were not identified. The amino terminal sequence for p10 from AMV (Fig. 1,

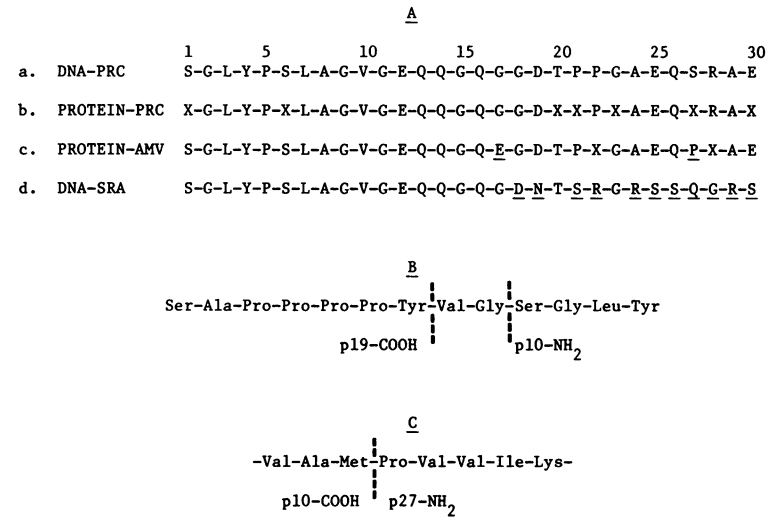

FIG. 1. Partial protein sequence of p10. p10 polypeptides were prepared from both AMV and from PrC-RSV by sequential column chromatography as described in the accompanying paper (8) followed by gel filtration chromatography (Bio-Rad P100) in 50% formic acid. Samples of 5 to 10 nmol of p10 were submitted to amino terminal amino acid sequencing by sequential Edman degradation on a Beckman 890C automated sequencer, with a modified dilute quadrol program (4) and polybrene (13) as described in detail elsewhere (1). The repetitive yields for the first 30 amino acids were greater than 94%. Phenyl thiohydantoin derivatives of amino acids were identified spectrophotometrically at 254 nm after high-pressure liquid chromatography. Phenyl thiohydantoin-threonine was detected simultaneously at 313 μm as dehydrothreonine. A modified methanol program (2) was used for these studies. Amino acids were confirmed by thin-layer chromatography by the method of Summers et al. (11). (A) Amino-terminal amino acid sequences of p10: (a) sequence predicted for PrC-RSV p10 from nucleic acid sequence (Schwartz et al., personal communication); (b) sequence derived for p10 of PrC-RSV by automated Edman degradation; (c) sequence derived for p10 of AMV by automated Edman degradation; (d) sequence predicted for SRA-RSV p10 from nucleotide sequence (12). Residues that differ from the sequence predicted from the PrC-RSV nucleic acid sequence are underlined. X indicates not determined. (B) Predicted arrangement of p19 and p10 on *gag* gene precursor from nucleotide sequence of PrC-RSV. C, Predicted arrangement of p10 and p27 on the *gag* gene precursor from nucleotide sequence.

line c) is very similar to the predicted p10 sequence from PrC-RSV, with only 2 of the 30 amino acids differing. These differences can be explained by single-base mutations. Two other residues were unidentified and thus could represent differences as well. The quantitative yields for six amino acids from AMV p10 are plotted in Fig. 2. For comparative purposes, we also include in Fig. 1A, line d, the p10 amino acid sequence predicted from the nucleotide sequence of the Schmidt-Ruppin A strain of RSV (12). As for AMV, the amino acid residues are conserved for the first 16 residues, after which some variation occurs. A major change in predicted amino acid sequence after residue 23 results from an apparent shift in the reading frame. Protein sequencing of p10 from the Schmidt-Ruppin A strain of RSV will be needed to verify the existence of this major alteration.

The similarity in the predicted amino acid composition of p10 for PrC-RSV with the observed composition of p10 from AMV and from RSV suggests that there is no major divergence between the two viruses in the amino acid sequence from residues 31 to 62. These amino acid compositions are shown in Table 1. Both proteins contain high levels of glutamic acid plus glutamine, proline, and glycine, and both lack lysine, isoleucine, phenylalanine, and cysteine. Two discrepancies between the compositions and the predicted sequence deserve comment. First, the yield of tyrosine after hydrolysis was variable, even though Edman degradation showed a normal yield of tyrosine at position 4 for the same preparation. In four separate hydrolyses, twice we observed values of 0.9 mol of tyrosine per mol of p10, and twice we observed values of 0.1 or less. Also, we have found p10 to be poorly labeled with ^{125}I in standard chloramine T-catalyzed iodination reactions. Thus it is possible that a fraction of the tyrosine residues is modified in some manner. Second, the yield of glycine in the compositional analysis of p10 from PrC-RSV was higher than predicted by the DNA sequence. It is likely that the excess glycine resulted from contamination of the polypeptide

TABLE 1. Comparative amino acid compositions[a]

Amino acid	Composition of p10 sequence from:			
	PrC-RSV DNA[b]	PrC-RSV	AMV[c]	AMV[c]
Aspartic acid	2	2.1	1.8	2.4
Threonine	3	3.0	2.5	2.7
Serine	4	2.4	2.6	2.7
Glutamic acid	10	9.8	10.5	11.0
Proline	8	9.9	8.0	9.0
Glycine	11	15.9	10.0	10.0
Alanine	9	8.8	9.7	10.3
Valine	4	3.4	3.8	3.8
Methionine[d]	1	0.3	0.3	0.4
Isoleucine	0	<0.05	0.1	0.1
Leucine	4	3.8	4.7	4.9
Tyrosine	1	<0.05	0.9	0.1
Phenylalanine	0	<0.05	<0.05	<0.05
Histidine	1	0.8	1.1	1.1
Lysine	0	0.2	0.1	0.1
Arginine	3	3.0	4.5	4.7

[a] For amino acid composition analyses, 10 nmol was hydrolyzed at 110°C with constant boiling HCl plus 0.01% ethane thiol in sealed evacuated ampules for 24 h and then analyzed for amino acid content on a Beckman model 118 CL analyzer. Values listed are moles of each amino acid per mole of p10.

[b] The numbers for the DNA sequence are derived from the stretch of 62 amino acids encoded by the *gag* gene that is amino terminal to p27.

[c] Independent analyses done on two different preparations of AMV p10.

[d] Methionine sulfoxide values were less than 0.02 in all determinations.

with free glycine at some stage in the purification, since greater than a molar amount of glycine was released in the first cycle of the Edman degradation (data not shown).

The amino-terminal serine of p10 is not contiguous with the carboxy-terminal tyrosine of p19; two amino acid residues, valine and glycine, separate these polypeptides, as inferred from the DNA sequence and carboxypeptidase analysis of p19 (Hunter et al., personal communication). This arrangement is shown schematically in Fig. 1B. Thus either two internal proteolytic cleavages are required to generate p19 and the amino terminus of p10, or else carboxypeptidase or aminopeptidase action removes the two amino acid residues that remain attached to p19 or to p10. The fact that the p10 species both from AMV and from RSV have the same N terminus, despite the very different sources of the virus preparations, suggests to us that the terminus is not the result of artifactual proteolysis in vitro.

If a single peptide bond scission acts to separate p10 from p27, then according to the DNA sequence, the C terminus of p10 should be methionine. This is the only methionine residue in the polypeptide. The predicted amino acid sequence for this region is shown in Fig. 1C, but in fact, purified p10 contains significantly less than one methionine. Amino acid compositions of p10 purified by chromatography (Table 1) or by organic extraction directly from virions (data not shown) showed 0.3 mol of methionine per mol of p10. Furthermore, we have found consistently that the relative yields of [^{35}S]methionine-labeled p10 are about fivefold lower than predicted by the sequence. We infer from these results that variable carboxypeptidase action, either in vivo or in vitro, removes at least one residue from the C terminus. However, the presence of some methionine means that at least a fraction of p10 consists of molecules with the entire series of 62 amino acid residues, consistent with the size predicted from gel filtration and from the amino acid composition.

In summary, these results confirm that the *gag* gene of ASLV codes for a 7-kilodalton protein designated p10, and that this protein is located between p19 and p27 on the polypeptide precursor Pr76. The earlier conclusion (14) that p10 is not a *gag* protein is thus in error. This conclusion was based on comparison of tryptic maps of [^{35}S]methionine-labeled p10 and Pr76. The fact that the single ^{35}S-labeled tryptic peptide of p10 ends at the C terminus of the protein, and thus does not have an exact counterpart in Pr76, in part may explain this discrepancy. More recent experiments involving pactamycin map-

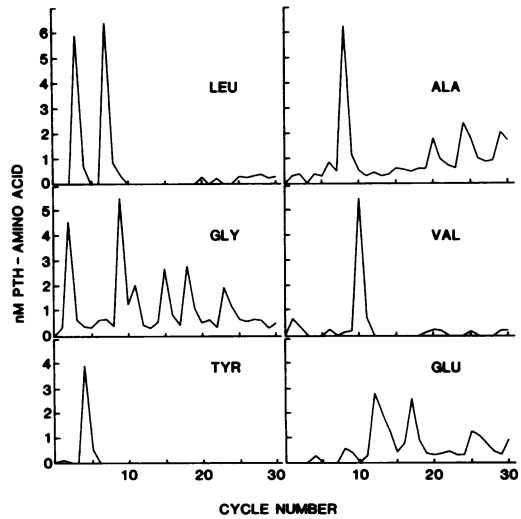

FIG. 2. Phenyl thiohydantoin-amino acid yields from AMV p10. AMV p10 polypeptide was purified and sequenced as described in the legend to Fig. 1. The yields (in nanomoles) of the phenyl thiohydantoin derivatives of six representative amino acids are plotted versus the cycle number of the samples from the automated sequencer.

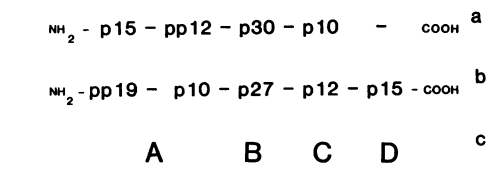

FIG. 3. Schematic representation of the *gag* gene functional regions. (a) Murine leukemia virus. (b) ASLV. (c) A, B, C, and D, Proposed functional regions of the *gag* gene.

ping of leucine-labeled polypeptides of Rous-associated virus 61 were interpreted to suggest that p10 derives from the aminoterminus of the *env* gene precursor (10). However, the data obtained in these experiments also are consistent with the localization of p10 between p19 and p27.

The addition of p10 to the set of avian *gag* proteins makes the organization of functional domains in the *gag* precursors from avain and murine retroviruses appear more closely similar. A schematic summary of these domains is shown in Fig. 3. Both classes of viruses code for about 65 kilodaltons of protein, represented by domains A through C, that is processed into four mature proteins. In addition, the *gag* gene of ASLV, but not that of murine viruses, encodes a domain D that contains the viral protease p15, which mediates *gag* precursor cleavage. The murine counterpart of domain D may be located elsewhere in the viral genome, perhaps in the *gag-pol* precursor where extensive coding regions remain to be characterized. Domain B, which contains the major structural protein of the virion (avian p27, murine p30) is believed to form the core shell. Domain C contains a small basic protein that forms a complex with the genomic RNA (avian p12, murine p10). Domain A gives rise to two proteins in both classes of viruses, but the functional correspondence of these proteins is not exact. The amino-terminal protein (avian p19, murine p15) is positioned close to the lipid bilayer, as inferred from the results of lipid-protein cross-linking (7). Two other properties associated with p19 in the avian viruses—the ability to bind specifically to the homologous viral RNA and the acquisition of radioactive label in [^{32}P]phosphate-labeled cells—are associated with p12 in the mammalian viruses (9). However, in another respect the proteins that lie between the N-terminal and core proteins do bear a resemblance: both avian p10 and murine p12 have a peculiar amino acid composition, with a high content of proline plus glycine (32% in both proteins). The functional significance of this property remains to be elucidated.

We thank Thomas Cornelius, Jane Ellen Smith, Karen Butler, and Edgar Hill for expert technical assistance and Curtis Fullmer for analysis of the amino acid compositions.

This work was supported by Public Health Service grants CA 20081, CA 29884, AM 03555, and AI 09153 from the National Institutes of Health. E.H. is the recipient of Public Health Service research career development award CA 00685 from the National Cancer Institute.

LITERATURE CITED

1. **Bhown, A. S., and A. F. H. S. Habeeb.** 1977. Structural studies on e-prototoxin of *Clostridium perfringens* type D. Localization of the site of tryptic scission necessary for activation to e-toxin. Biochem. Biophys. Res. Commun. **78**:889–896.
2. **Bhown, A. S., J. E. Drole, A. Weissenger, and J. C. Bennett.** 1978. Methanol solvent system for rapid analysis of phenyl-thiohydantoin amino acids by high pressure liquid chromatography. J. Chromatogr. **148**:532–535.
3. **Bolognesi, D. P., R. Luftig, and J. H. Shaper.** 1973. Localization of RNA tumor virus polypeptides. I. Isolation of further virus substructures. Virology **56**:549–564.
4. **Brauer, A. W., M. N. Hargolies, and E. Haber.** 1975. The application of 0.1 M quadrol to the microsequence of proteins and the sequence of tryptic peptides. Biochemistry **14**:3029–3035.
5. **Eisenman, R. N., and V. M. Vogt.** 1978. The biosynthesis of oncovirus proteins. Biochem. Biophys. Acta **473**:187–239.
6. **Fleissner, E.** 1971. Chromatographic separation and antigenic analysis of proteins of the oncornaviruses. I. Avian leukemia-sarcoma viruses. J. Virol. **8**:778–785.
7. **Pepinsky, R. B., and V. M. Vogt.** 1979. Identification of retrovirus matrix proteins by lipid-protein cross-linking. J. Mol. Biol. **131**:819–837.
8. **Pepinsky, R. B., and V. M. Vogt.** 1983. Purification and properties of a fifth major viral *gag* protein from avian sarcoma and leukemia viruses. J. Virol. **45**:648–658.
9. **Sen, A., C. J. Scherr, and G. J. Todaro.** 1977. Phosphorylation of murine type C viral p12 proteins regulates their extent of binding to the homologous viral RNA. Cell **10**:489–496.
10. **Shealy, D. J., A. G. Mosser, and R. R. Rueckert.** 1980. Novel p19-related protein Rous-associated virus type 61: implications for avian *gag* gene order. J. Virol. **34**:431–437.
11. **Summers, M. R., G. W. Smythers, and S. Oroszlan.** 1973. Thin layer chromatography of sub-nanomole amounts of phenyl thiohydantoin (PTH) amino acids on polyamide sheets. Anal. Biochem. **53**:624–628.
12. **Swanstrom, R., H. E. Varmus, and J. M. Bishop.** 1982. Nucleotide sequence of the 5′ noncoding region and part of the *gag* gene of Rous sarcoma virus. J. Virol. **41**:535–541.
13. **Tarr, G. E., J. F. Beecher, M. Bell, and D. J. McKean.** 1978. Polyquarternary amines prevent peptide loss from sequenators. Anal. Biochem. **84**:622–625.
14. **Vogt, V. M., R. Eisenman, and H. Diggelmann.** 1975. Generation of avian myeloblastosis virus structural proteins by proteolytic cleavage of a precursor polypeptide. J. Mol. Biol. **96**:471–493.