JOURNAL OF VIROLOGY, Oct. 1980, p. 115-119
0022-538X/80/10-0115/05$02.00/0

Vol. 36, No. 1

Amino Acid Sequence Analysis of Reverse Transcriptase Subunits from Avian Myeloblastosis Virus

TERRY D. COPELAND,[1] DUANE P. GRANDGENETT,[2] AND STEPHEN OROSZLAN[1]*

Biological Carcinogenesis Program, Frederick Cancer Research Center, Frederick, Maryland 21701,[1] and Saint Louis University Medical Center, Institute for Molecular Virology, St. Louis, Missouri 63110[2]

The NH_2-terminal amino acid sequences of the α and β chains of avian myeloblastosis $\alpha\beta$ DNA polymerase were determined by using microsequence analysis in the subnanomole range and were found to be identical up to 17 residues. The common sequence was as follows: Thr-Val-Ala-Leu-His-Leu-Ala-Ile-Pro-Leu-Lys-Trp-Lys-Pro-Asn-His-Thr-. This result provides convincing chemical evidence that the α chain is derived from the NH_2-terminal region of the β chain by proteolytic cleavage, whereas the amino acid composition for these α and β subunits and p32 DNA endonuclease suggests that the latter is derived from the carboxyl-terminal region of the β chain.

The RNA-dependent DNA polymerase, reverse transcriptase, of avian type C retroviruses has been shown to possess multiple enzymatic activities, among which the transcription of RNA to DNA and the activities of RNase H and DNA endonuclease have been studied most extensively. The capability of this DNA polymerase to catalyze PP_i exchange and pyrophosphorolysis has also been reported (24, 26). Virion-bound avian retrovirus DNA polymerase appears in multiple molecular forms. The enzymatic, biochemical, and immunological properties of the various structural forms have been summarized in a recent review (5). The major molecular species is a heterodimer of two structurally related polypeptide chains, designated α and β, with molecular weights of 63,000 and 92,000, respectively. This species possesses all of the above-described enzymatic activities. Other enzymatically active (polymerase and RNase H) forms have been identified to be the α subunit and a dimer of the β chain (β_2). Tryptic fingerprint analyses provided convincing evidence that the α subunit may be derived by proteolytic cleavage from the β subunit (6, 22). Subsequently, another polymerase-specific polypeptide, termed p32, possessing endonuclease activity, was identified and shown to be structurally related to the β, but not to the α, subunit (10, 23).

In this communication, we report the results of the chemical analyses: the amino acid compositions and sequences of the above-specified *pol* gene-encoded polypeptides of avian myeloblastosis virus (AMV). The compositional data indicated that the β chain consists of the α chain and p32. The results of sequence analyses provided convincing evidence that the α and β chains of AMV DNA polymerase have identical NH_2 termini. A discussion of *gag-pol* gene overlap is presented.

MATERIALS AND METHODS

Virus. AMV, BAI strain A, was obtained from plasma of leukemic chickens. Concentrated plasma virus was kindly provided by Joseph Beard of Life Sciences Inc., St. Petersburg, Fla., through the Virus Cancer Program of the National Cancer Institute.

Purification of $\alpha\beta$ DNA polymerase and its subunits. The $\alpha\beta$ DNA polymerase was purified from disrupted virions by heparin-Sepharose 4B chromatography as previously described (7). To dissociate the subunits, the purified enzyme was denatured by heating at 100°C for 2 min in a solution of 1% sodium dodecyl sulfate (SDS), 1% 2-mercaptoethanol, 0.01 M sodium phosphate (pH 7.1), and 10% glycerol. Subsequently, the α and β subunits were separated by SDS-polyacrylamide gel electrophoresis on cylindrical gels made of 6% acrylamide (9, 16). The gels were soaked in electrophoresis buffer for 2 weeks before use. To facilitate localization of the proteins in the gel, ^{125}I-labeled $\alpha\beta$ DNA polymerase was added to the cold enzyme before denaturation and electrophoresis. After the run, gels were sliced into 2-mm pieces and counted in a gamma counter. Appropriate gel slices were pooled and extracted with 0.1 M sodium phosphate buffer (pH 7.1) containing 0.1% SDS, 0.02% sodium azide, and 0.5 mM phenylmethylsulfonyl fluoride. All samples containing the isolated subunits were dialyzed against six changes (2 liters each) of 0.005% ultrapure SDS (Bio-Rad Laboratories) at 4°C for 60 h. The dialyzed samples (total volume, 2 to 3 ml) were then lyophilized, suspended in 0.4 ml of water, and analyzed. The purity of the α and β chains is shown in Fig. 1.

Purification of p32. The p32 endonuclease was purified from detergent-lysed virions by phosphocellulose and polyuridylic acid-Sepharose 4B chromatographies as previously described (10). For chemical analysis, p32 was further purified by SDS-polyacryl-

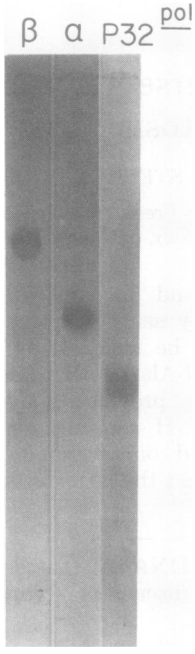

FIG. 1. *SDS-polyacrylamide gel electrophoresis pattern of purified α and β subunits of AMV DNA polymerase and p32. A portion of each sample for sequencing or analysis of amino acid composition was subjected to SDS-polyacrylamide gel electrophoresis as described in the text. Electrophoresis was from top to bottom. The migration of these proteins as shown above does not reflect their true molecular weights.*

amide gel electrophoresis as described above. The purity of p32 is also shown in Fig. 1.

Amino acid analysis. Protein samples were hydrolyzed in vacuo at 110°C for 24 h with 6 M HCl containing 0.1% phenol and analyzed on a Durrum 500 amino acid analyzer. The raw data were treated by a computer program (1) as previously described (13).

Microsequence analysis. Semiautomated microsequence analyses utilizing Edman degradation (4) were performed with a Beckman sequencer model 890B which has been updated to a model 890C and equipped with a cold-trap accessory. In separate runs, 700 pmol of the AMV reverse transcriptase α chain and 560 pmol of the β chain were subjected to stepwise degradation in the presence of Polybrene (15, 17, 27) to prevent protein washout. First, Polybrene (1.5 mg) was introduced into the sequencer cup with a synthetic dipeptide (15), either glycylglycine or tyrosylglutamic acid (120 nmol) in water, and, after drying, the mixture was subjected to six cycles of the Beckman protein program no. 121178. To degrade the proteins, the same dilute Quadrol program was used with modifications as described previously (18). The protein to be sequenced was double coupled in the first cycle. Single coupling and single cleavage were used thereafter. Sequencer fractions containing the thiazolinone derivatives of amino acids were converted to their phenylthiohydantoin derivatives by 1 N HCl as previously described (19), frozen, and lyophilized. To identify phenylthiohydantoin derivatives of amino acids, either samples of or the entire residue from each cycle was analyzed by high-performance liquid chromatography on a phenylalkyl support as previously described (12).

Carboxypeptidase digestion. Digestions with carboxypeptidases A, B, and Y (all purchased from Pierce Chemical Co., Rockford, Ill.) were performed by conventional methods as previously described (11, 20).

RESULTS

Amino acid composition. The amino acid compositions of AMV DNA polymerase α and β chains and of p32 are shown in Table 1. The compositions of the α and β chains were derived from two independent analyses of different batches of both subunits. Only a single analysis was performed on p32. Cysteine and tryptophan were not determined. The molecular weights calculated from the compositional data by a computer program were found to be ~58,000 for the α chain, ~92,000 for the β chain, and ~32,600 for p32. These values were in good agreement with previous estimates based on migration relative to standard proteins in SDS-polyacrylamide gel electrophoresis. In general, the sum of the number of each amino acid found in the α chain and p32 agreed reasonably well (equal or slightly smaller) with the residue number calculated for the β chain, the larger subunit. Ex-

TABLE 1. *Amino acid composition of AMV reverse transcriptase α and β subunits and endonuclease p32*

Amino acid[a]	No. of residues per protein		
	p32[b]	α Chain[c]	β Chain[c]
Lysine	17	22	43
Histidine	8	13	23
Arginine	22	32	56
Aspartic acid	29	43	65
Threonine	21	32	60
Serine	27	42	60
Glutamic acid	38	64	100
Proline	16	43	64
Glycine	42	45	69
Alanine	25	57	94
Valine	14	30	50
Methionine	4	7	9
Isoleucine	10	15	27
Leucine	20	68	94
Tyrosine	6	8	11
Phenylalanine	8	19	26

[a] Cysteine and tryptophan were not determined.
[b] Results of a single analysis after 24 h of digestion. Molecular weight of p32 is 32,636.
[c] Average values of two independent analyses of different samples. Hydrolysis was carried out for 24 h. Molecular weights of the α and β chains are 57,898 and 91,387, respectively.

ceptions were aspartic acid, serine, and glycine. The sum of these residues in the α chain and p32 was greater than the number found in the β chain. This apparent discrepancy may be attributed to background contamination derived from polyacrylamide gels or dialysis tubing or both (2). Thus, the compositional data corroborate previous findings which suggest that both the α chain and p32 may be derived from the β chain by proteolytic cleavage (6, 22, 23).

The hydrophobicity of the proteins was estimated from the compositional data by dividing the amino acids into three classes: polar, intermediate, and nonpolar (3). Table 2 shows that p32 is much less hydrophobic than either of the two subunits of $\alpha\beta$ DNA polymerase.

NH_2-terminal sequences. Highly purified α and β chains were available in minute amounts. To determine the NH_2-terminal amino acid sequence, approximately 700 pmol of the α subunit was degraded in a single run. The yields of various amino acids at each sequencer cycle are shown in Fig. 2. Unambiguous identification of residues up to position 17 was possible. The NH_2-terminal sequence (see Fig. 4) of the α chain is based on the quantitative data presented in Fig. 2. For the sequence analysis of the β chain, 560 pmol of purified protein was used. The results of the stepwise Edman degradation for this subunit are shown in Fig. 3, and the sequence is shown in Fig. 4. As shown, the NH_2-terminal amino acid sequences of the two subunits were identical up to position 17 and probably higher.

Carboxyl-terminal amino acids. Digestion of the α chain with carboxypeptidase A for different time intervals (1 and 20 min), followed by quantitative determination of the free amino acids released (0.64 nmol of tyrosine and 0.82 nmol of phenylalanine per nmol of protein after 20 min of digestion), provided results allowing

TABLE 2. *Polarities of AMV pol gene-encoded polypeptides*

Polypeptide	% of total amino acids[a]		
	Polar[b]	Intermediate[c]	Nonpolar[d]
α Chain	29.8	25.9	44.3
β Chain	31.0	26.2	42.8
p32	34.5	33.9	31.6

[a] Calculated from data of Table 1 by the method of Capaldi and Vanderkooi (3). Cysteine and tryptophan were not determined.

[b] Includes lysine, arginine, aspartic acid, and glutamic acid.

[c] Includes histidine, serine, threonine, glycine, and tyrosine.

[d] Includes alanine, valine, leucine, isoleucine, methionine, proline, and phenylalanine.

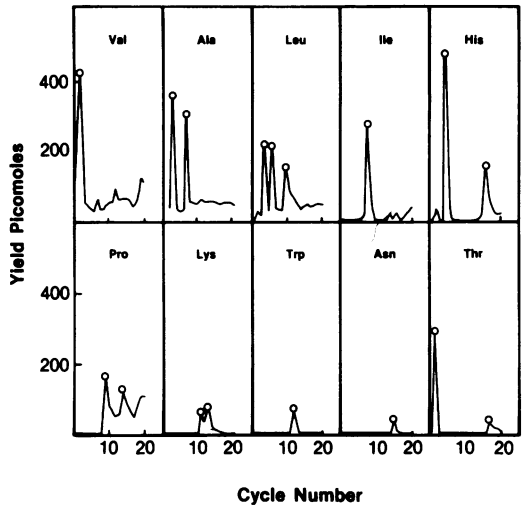

FIG. 2. *Yields of amino acids obtained by degrading 700 pmol of the α chain. The data plotted were derived from analysis by high-performance liquid chromatography. The yields were normalized to 100% of the sample. Assignments* (O) *are as indicated.*

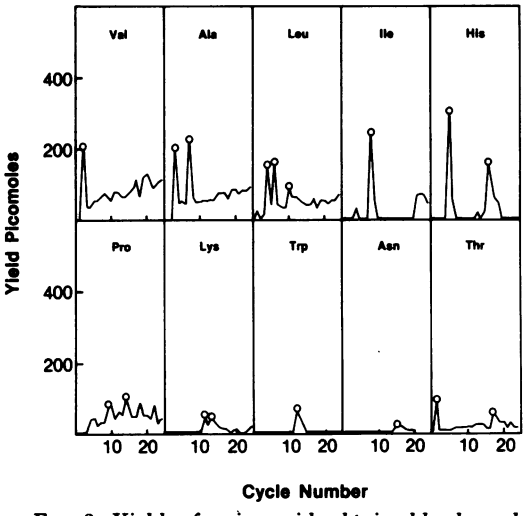

FIG. 3. *Yields of amino acids obtained by degrading 560 pmol of the β chain. The data plotted were derived from analysis by high-performance liquid chromatography. The yields were normalized to 100% of the sample. Assignments* (O) *are as indicated.*

us to determine the carboxyl-terminal sequence to be -Tyr-Phe-OH. Digestion of the β chain or p32 for 20 min or longer with the same enzyme or carboxypeptidase B, or with the mixture, even after complete unfolding of the molecules in the presence of SDS and urea, resulted in no release of free amino acids. Similarly, digestion with carboxypeptidase Y did not liberate amino acids in measurable quantities from the β chain or p32. The reason for this effect is unknown.

```
                1                   5                  10                  15
α chain  Thr-Val-Ala-Leu-His-Leu-Ala-Ile-Pro-Leu-Lys-Trp-Lys-Pro-Asn-

β chain  Thr-Val-Ala-Leu-His-Leu-Ala-Ile-Pro-Leu-Lys-Trp-Lys-Pro-Asn-


               16                  20
α chain  His-Thr- X -Val- X -Ile- X -Gln-

β chain  His-Thr-Pro-Val- X -Ile- X -Gln-
```

FIG. 4. *NH_2-terminal amino acid sequences of AMV reverse transcriptase α and β chains. Sequence assignments beyond residue 17 are tentative.*

DISCUSSION

The results of chemical analyses presented in this communication clearly show the NH_2-terminal sequence identity of the α and β chains of AMV DNA polymerase. This finding and the amino acid compositional data suggest that the α polymerase and p32 endonuclease are derived from the NH_2- and the carboxyl-terminal regions of the β polypeptide, respectively, by proteolytic cleavage.

Previously, pactamycin mapping of the α and β subunits (25) and tryptic peptide analysis of various *pol* gene products (21) suggested that the α and β chains share the same NH_2 terminus.

Although our data suggest that p32 endonuclease is derived from the β polypeptide by proteolytic cleavage in virions, the primary *pol* precursor to p32 is as yet undefined. Antiserum directed against p32 (23) immunoprecipitates $Pr130^{pol}$, a polypeptide containing *pol* peptides and p15 antigenic determinants, from Rous sarcoma virus-infected cells (4a). Whether or not $Pr130^{pol}$ is a primary transcript for p32 is unknown.

The nearly complete nucleotide sequence of the *gag-pol* region of Rous sarcoma virus has been recently determined (D. Schwartz, R. Tizard, and W. Gilbert, personal communication). The comparison of the nucleotide sequence of Rous sarcoma virus and the NH_2-terminal amino acid sequences of the α and β chains of AMV DNA polymerase revealed a perfect match and, thus, established the proper reading frame as well as the precise location of the pertinent coding portion of the genomic RNA. These data also suggest that the NH_2-terminal (5′) region of the *pol* gene of genomic RNA of the helper virus of the AMV complex and that of Rous sarcoma virus are highly conserved. The complete amino acid sequence of AMV p15 (R. Sauer, D. Allen, and H. Niall, personal communication) and of AMV pp12 (T. S. Misono, F. S. Sharief, and J. Leis, Fed. Proc., p. 1611, 1980) have been determined. Comparison of AMV p15 or pp12 amino acid sequences with those derived from the α and β chain indicates that there is no gene overlap within these specific regions of p15 and pp12 with that of the β chain. A similar conclusion has also been reached by Rettenmier et al. (21) by tryptic peptide analysis of various *gag* and *pol* gene products in contrast to a previous suggestion of gene overlap of *gag-pol* (14).

Both the α and β chains were found to be much more hydrophobic than p32. Since ionic detergents and dimethyl sulfoxide or 1,4-dioxane were shown to be effective agents which dissociate the α and β subunits (8), it is assumed that noncovalent bonds, such as hydrophobic interactions strengthened by short-range polar interactions (i.e., hydrogen bonding), are responsible for their firm binding in αβ DNA polymerase of AMV.

ACKNOWLEDGMENTS

This work was supported by contract NO1-CO-75380 with the National Cancer Institute, Bethesda, Md., Public Health Service research grant CA-16312 from the National Cancer Institute, and Specialized Cancer Center CORE grant CA-25843. D.P.G. is a recipient of an American Cancer Society Faculty Research Award.

We thank Gary W. Smythers for amino acid analysis and end group determinations.

LITERATURE CITED

1. **Boyer, S. H., A. N. Noyes, M. L. Boyer, and K. Marr.** 1973. Hemoglobin α chains in apes: primary structures and the presumptive nature of back mutation in a normally silent gene. J. Biol. Chem. **248**:992-1003.
2. **Brown, W. E., and G. C. Howard.** 1980. Amino acid composition of proteins eluted from polyacrylamide gels: background considerations. Anal. Biochem. **101**: 294-298.
3. **Capaldi, R. A., and G. Vanderkooi.** 1972. The low polarity of many membrane proteins. Proc. Natl. Acad. Sci. U.S.A. **69**:930-932.
4. **Edman, P., and G. Begg.** 1967. A protein sequenator. Eur. J. Biochem. **1**:80-91.

4a. **Eisenman, R. N., W. S. Mason, and M. Linial.** 1980. Synthesis and processing of polymerase proteins of wild-type and mutant avian retroviruses. J. Virol. **36**: 62-78.

5. **Gerard, G. F., and D. P. Grandgenett.** 1980. Retrovirus reverse transcriptase, p. 345-394. *In* J. Stephenson (ed.),

Molecular biology of RNA tumor viruses. Academic Press, Inc., New York.

6. **Gibson, W., and I. M. Verma.** 1974. Studies on the reverse transcriptase of RNA tumor viruses. Structural relatedness of two subunits of avian RNA tumor viruses. Proc. Natl. Acad. Sci. U.S.A. **71**:4991–4994.
7. **Golomb, M., and D. P. Grandgenett.** 1979. Endonuclease activity of purified RNA-directed DNA polymerase from avian myeloblastosis virus. J. Biol. Chem. **254**: 1606–1613.
8. **Grandgenett, D. P.** 1976. Dissociation of $\alpha\beta$ DNA polymerase of avian myeloblastosis virus by dimethyl sulfoxide. J. Virol. **17**:950–961.
9. **Grandgenett, D. P., G. F. Gerard, and M. Green.** 1973. A single subunit from avian myeloblastosis virus with both RNA-directed DNA polymerase and ribonuclease H activity. Proc. Natl. Acad. Sci. U.S.A. **70**:230–234.
10. **Grandgenett, D. P., A. C. Vora, and R. D. Schiff.** 1978. A 32,000-dalton nucleic acid-binding protein from avian retroviruses cores possesses DNA endonuclease activity. Virology **89**:119–132.
11. **Hayashi, R., S. Moore, and W. H. Stein.** 1973. Carboxypeptidase from yeast: large scale preparation and the application to COOH terminal analysis of peptides and proteins. J. Biol. Chem. **218**:2296–2302.
12. **Henderson, L. E., T. D. Copeland, and S. Oroszlan.** 1980. Separation of amino acid phenylthiohydantoins by high-performance liquid chromatography on phenylalkyl support. Anal. Biochem. **102**:1–7.
13. **Henderson, L. E., T. D. Copeland, G. W. Smythers, H. Marquardt, and S. Oroszlan.** 1978. Amino-terminal amino acid sequence and carboxyl-terminal analysis of Rauscher murine leukemia virus glycoproteins. Virology **85**:319–322.
14. **Hizi, A., M. A. McCrae, and W. K. Joklik.** 1978. Studies on the amino acid sequence content of proteins specified by the *gag* and *pol* genes of avian sarcoma virus B77. Virology **89**:272–284.
15. **Hunkapiller, M. W., and L. E. Hood.** 1978. Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier (polybrene), and high pressure liquid chromatography. Biochemistry **17**:2124–2133.
16. **Kacian, D. L., K. F. Watson, and S. Spiegelman.** 1971. Purification of the DNA polymerase of avian myeloblastosis virus. Biochim. Biophys. Acta **246**:365–383.
17. **Klapper, D. A., C. E. Wilde, and J. C. Capra.** 1978. Automated amino acid sequence of small peptides utilizing polybrene. Anal. Biochem. **85**:126–131.
18. **Oroszlan, S., T. D. Copeland, L. E. Henderson, J. R. Stephenson, and R. V. Gilden.** 1979. Amino-terminal sequence of bovine leukemia virus major internal proteins: homology with mammalian type C virus p30 structural proteins. Proc. Natl. Acad. Sci. U.S.A. **76**: 2996–3000.
19. **Oroszlan, S., T. D. Copeland, M. R. Summers, G. Smythers, and R. V. Gilden.** 1975. Amino acid sequence homology of mammalian type C RNA virus major proteins. J. Biol. Chem. **250**:6232–6239.
20. **Oroszlan, S., L. E. Henderson, J. R. Stephenson, T. D. Copeland, C. W. Long, J. N. Ihle, and R. V. Gilden.** 1978. Amino- and carboxy-terminal amino acid sequences of proteins coded by *gag* gene of murine leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **75**:1404–1408.
21. **Rettenmier, C. W., R. E. Karess, S. M. Anderson, and H. Hanafusa.** 1979. Tryptic peptide analysis of avian oncovirus *gag* and *pol* gene products. J. Virol. **32**:102–113.
22. **Rho, H. M., D. P. Grandgenett, and M. Green.** 1975. Sequence relatedness between the subunits of avian myeloblastosis virus reverse transcriptase. J. Biol. Chem. **250**:5278–5280.
23. **Schiff, R. D., and D. P. Grandgenett.** 1978. Virus-coded origin of a 32,000-dalton protein from avian retrovirus cores: structural relatedness of p32 and the β polypeptide of the avian retrovirus DNA polymerase. J. Virol. **28**:279–291.
24. **Seal, G., and L. A. Loeb.** 1976. On the fidelity of DNA replication: enzyme activities associated with DNA polymerases from RNA tumor viruses. J. Biol. Chem. **251**: 975–981.
25. **Shealy, D., and R. R. Rueckert.** 1978. Proteins of Rous-associated virus 61, an avian retrovirus: common precursor for glycoproteins gp85 and gp35 and use of pactamycin to map translational order of proteins in the *gag*, *pol*, and *env* genes. J. Virol. **26**:380–388.
26. **Srivastava, A., and M. J. Modak.** 1980. Enzymatic activities associated with avian and murine retroviral DNA polymerases. J. Biol. Chem. **255**:2000–2004.
27. **Tarr, G. E., J. F. Beecher, M. Bell, and D. J. McKean.** 1978. Polyquaternary amines prevent peptide loss from sequenators. Anal. Biochem. **84**:622–627.