Complete amino acid sequence of the nucleic acid-binding protein of bovine leukemia virus

Terry D. Copeland, M. Alice Morgan and Stephen Oroszlan

Biological Carcinogenesis Program, Basic Research Program-LBI, Frederick Cancer Research Facility, Frederick, MD 21701, USA

Received 28 March 1983

The complete amino acid sequence of the nucleic acid-binding protein p12 of bovine leukemia virus (BLV) has been determined. Peptides were generated by enzymatic digestion and formic acid cleavage, purified by reversed-phase liquid chromatography and subjected to automated Edman degradation. BLV p12 is a proline-rich linear polypeptide composed of 69 amino acids with M_r 7558. A comparison of the p12 structure to that of the avian and murine type C retroviral nucleic acid-binding proteins shows significant homology only in the putative binding domain. This conserved region is duplicated in BLV p12 as in the avian homolog.

Bovine leukemia virus *Retrovirus* *Nucleic acid binding protein* *Micro-sequencing* *Primary structure determination*

1. INTRODUCTION

Bovine leukemia virus (BLV) is an exogenous non-genetically transmitted retrovirus shown to be associated with enzootic bovine leukosis or lymphosarcoma [1,2]. BLV is immunologically distinct from other retroviruses and has not been readily classified into any of the well-defined groups [3]. Previously we found [4] that the major core protein p24 of BLV has amino acid sequence homology to type C virus major core proteins p30s especially to that of feline leukemia virus (FeLV), another naturally occurring infectious retrovirus. The extent of the structural homology of the major internal virion proteins suggested that BLV and FeLV may have originated from a common progenitor [4]. Substantial amino acid sequence data for the various structural proteins of a large number of retroviruses has been accumulated from studies conducted primarily in this laboratory [5]. To obtain further insight into the relationship between BLV and type C viruses of other species, extensive primary structure analyses of BLV *gag* gene products were undertaken. BLV p12, a low-M_r basic protein [6,7], has also been shown to bind to single-stranded nucleic acids [6]. Here, we report the complete amino acid sequence of BLV p12.

2. MATERIALS AND METHODS

2.1. *Virus*

BLV grown in fetal lamb kidney cells [8,9] was purified by conventional sucrose density gradient centrifugation and obtained from the Biological Products Laboratory of the NCI-Frederick Cancer Research Facility (Frederick MD).

2.2. *Protein purification*

BLV p12 was purified by neutral chloroform methanol extraction of aqueous virus suspensions [10] followed by reversed phase liquid chromatography (RPLC) as in [11].

2.3. *Protein carboxamidomethylation*

Prior to fragmentation, p12 was reduced with dithiothreitol and carboxamidomethylated (with iodoacetamide) as in [12]. To remove excess reagents and side products, the pH of the reaction

Published by Elsevier Science Publishers B.V.
00145793/83/$3.00

mixture was lowered to about 2 with aqueous trifluoroacetic acid (TFA) before loading it onto a Waters C_{18} RPLC column. The modified protein was readily obtained by elution with linear acetonitrile gradient (0–30%) containing 0.05% TFA. Protein eluted at about 26% acetonitrile.

2.4. *Digestion with Lys-C protease*

Thirty nmol of carboxamidomethyl p12 was dissolved in 0.5 ml 0.1 M sodium bicarbonate (pH 8.5) containing 20% acetonitrile [13]. Lys-C protease (Boehringer-Mannheim Biochemicals, Indianapolis IN) was added and allowed to react at an enzyme to substrate ratio of 1:75 under nitrogen for 24 h.

2.5. *Limited trypsin digestion*

The lysines in 25 nmol carboxamidomethyl p12 were blocked with methyl-acetimidate as in [12] and the modified protein was digested with TPCK-trypsin (enzyme: protein = 1:100) in 0.1 M sodium bicarbonate (pH 8.5) which was also 1 M guanidine hydrochloride. Reaction proceeded overnight at 37°C.

2.6. *Cleavage with formic acid*

Twenty nmol of carboxamidomethyl p12 was dissolved in 1.0 ml 70% formic acid [14], heated at 37°C overnight, diluted with water and lyophilized.

2.7. *Peptide purification*

After the allotted time for enzymatic digestions, the pH of the solutions was lowered to pH 2 with aqueous TFA, either with or without the prior addition of guanidine hydrochloride. Samples were loaded onto a Waters μBondapak C_{18} column. Peptides were eluted with a linear gradient of acetonitrile in water containing 0.05% trifluoroacetic acid. Peaks were collected manually and aliquots taken for amino acid analysis. Amino acid analyses were carried out on a Durrum Model 500 analyzer equipped with a PDP 8/A computing integrator.

2.8. *Sequence analysis of intact protein and peptides*

A Beckman sequenator updated to the level of a Model 890C and equipped with a cold trap was employed as in [15]. Peptides dissolved in aqueous acetonitrile as obtained from RPLC purification were added directly to the spinning cup, dried and subjected to semi-automated Edman degradation in the presence of Polybrene [16]. PTH amino acids were analyzed on a phenylalkyl support [17].

2.9. *Carboxyl-terminal analysis*

Four nmol of protein was digested with carboxypeptidase A as in [18]. Amino acids released at timed intervals were determined by amino acid analysis.

3. RESULTS AND DISCUSSION

3.1. *Sequence determination*

The complete amino acid sequence of BLV p12 together with the sequence strategy employed is illustrated in fig.1. Amino-terminal analysis of intact unmodified protein was described in [11] in a report on the immunology of p12. Peptides generated by Lys-C protease proved to be very useful. As illustrated in fig.1 these peptides were

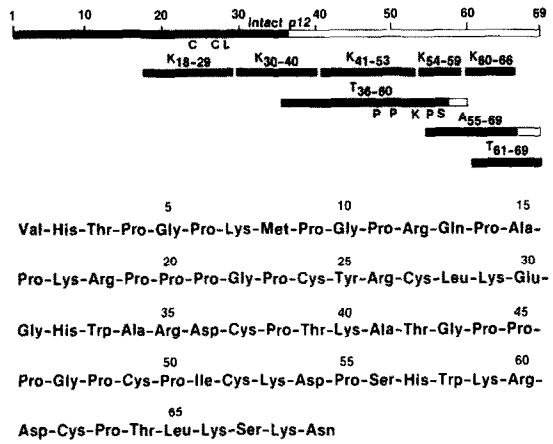

Fig.1. The complete amino acid sequence of BLV p12 and strategy for its determination. The structure shown was determined by semi-automated Edman degradation of intact protein and fragments derived by digestion with endoproteinase Lys-C (fragment designated with letter K) or trypsin (fragment designated with letter T) and by acid cleavage (fragment designated with letter A). The bars indicate the length (number of residues) of each peptide, while the shaded area indicates the extent to which sequence could be accurately determined. The letters under the bars indicate residues (single letter code, see legend to fig.2) which could not be unambiguously identified in the given peptide.

sequenced to the end. Peptides K_{1-7} and K_{8-17} were obtained by RPLC as a mixture and sequenced together. Peptides K_{67-68} and K_{69} were located by amino acid analysis in the salt breakthrough peak of the RPLC separation.

The NH_2-terminal sequence determination of peptides T_{36-60} and A_{55-69} provided the necessary overlap with Lys-C peptides in the COOH-terminal half of p12. Peptide T_{61-69} was completely sequenced and was found to be the COOH-terminal peptide of p12 as confirmed by the amino acid compositional data of intact protein and fragment, and also by COOH-terminal analysis of intact p12. From the results of carboxypeptidase A digestion the deduced sequence of p12 is -Lys–Asn-OH.

3.2. *Characterization of protein*

BLV p12 had the following composition (values calculated from the sequence are in parentheses): Asp 3.7(3), Asn(1), Thr 2.9(4), Ser 2.3(2), Glu 2.1(1), Gln(1), Pro 14.8(18), Gly 5.8(6), Ala 3.1(3), Val 1.3(1), Met 0.7(1), Ile 1.0(1), Leu 2.0(2), Tyr 0.8(1), His 2.2(3), Lys 7.4(8), Arg 4.8(5), Cys 5.5(6) and Trp(2). Tryptophan was not determined on the analyzer. The values for the amino acids are in good agreement with those derived from the complete sequence shown in fig.1.

BLV p12 is a linear polypeptide of 69 amino acids with a calculated M_r of 7558. An assessment of the number and nature of the charged amino acid side chains and end groups indicate that BLV p12 is a highly basic protein in agreement with the previously noted high isoelectric point [6]. It has 17 basic groups (3 histidine, 8 lysine, 5 arginine and one α-amino group) and only 5 carboxyl groups (3 aspartic acid, 1 glutamic acid and one α-carboxyl group). The number of proline residues is high: 18 out of 69 total residues. Thus proline is occupying 26% of the polypeptide chain. This high proline content is unusual among the retroviral nucleic acid-binding proteins analyzed. Twelve of the prolines appear in two internally homologous segments each comprising a total of 21 residues. The alignment of these internally duplicated regions is shown in fig.2. There are 14 amino acids out of 21 (66.7%) which are positionally identical (boxed). The most remarkable common feature of these repeated sequences is that they both contain the significant characteristic of the putative nucleic acid-binding domain of retroviral nucleic acid-binding proteins expressed as a set of 3 cysteines (circled) [12] periodically placed at positions n, $n + 3$ and $n + 13$. In addition to the 6 cysteines and most of the prolines, the other amino acids which appear positionally identical are glycine at residues 22 and 47, histidine 32 and 57, tryptophan 33 and 58, arginine 35 and 60, aspartic acid 36 and 61 and threonine 39 and 64. BLV p12 has been shown to bind to single-stranded DNA [6]. It remains to be determined whether both partially duplicated sequences are functionally important in nucleic acid-binding.

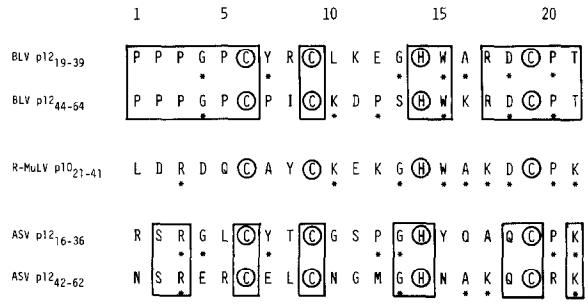

Fig.2. Internal duplication of sequence in BLV p12 and ASV p12. Alignment of fragments with the putative nucleic acid-binding domain of R-MuLV p10. Internally repeated amino acids are boxed. The cysteines and histidine conserved in all fragments are circled. Asterisks indicate amino acids which occur at identical alignment positions in at least 2 of 3 species. The single letter code is that of Dayhoff [20].

3.3. *Sequence comparison with other retroviral nucleic acid binding proteins*

In the nucleic acid-binding proteins of murine leukemia viruses (MuLV) the putative nucleic acid-binding domain occurs only once [12]. In the homologous structural protein p12 of avian sarcoma virus (ASV) this cysteine-containing region is duplicated as in BLV p12. The data indicate that ASV p12 and BLV p12 may have evolved by partial gene duplication. The alignments of the internally repeated segments of BLV p12 and ASV p12 [19] are shown together with the nucleic acid-binding domain of Rauscher (R)-MuLV p10 in fig.2. The cysteines and histidines (circled) are conserved in all fragments. Other amino acids which appear positionally identical in at least 2 of the 3 species are labeled with asterisks.

In general, the results of comparison of amino acid sequences of the NBPs of type C retroviruses suggest that the nucleic acid-binding regions are highly conserved.

ACKNOWLEDGEMENTS

We wish to thank Cathy V. Hixson and Young D. Kim for excellent technical assistance. Research was sponsored by the National Cancer Institute, DHHS, under contract no. N01-CO-23909 with Litton Bionetics. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

REFERENCES

[1] Olson, C. (1974) J. Am. Vet. Med. Assoc. 165, 630–632.
[2] Burny, A., Bruck, C., Chantrenne, H., Cleuter, Y., Dekegel, D., Ghysdael, J., Kettmann, R., Leclercq, M., Leunen, J., Mammerickx, M. and Portetelle, D. (1980) in: Bovine Leukemia Virus: Molecular Biology and Epidemiology (Klein, G. ed) pp.231–289, Raven Press, New York.
[3] Teich, N. (1982) in: Molecular Biology of Tumor Viruses: RNA Tumor Viruses (Weiss, R. et al. eds) pp.25–207, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
[4] Oroszlan, S., Copeland, T.D., Henderson, L.E., Stephenson, J.R. and Gilden, R.V. (1979) Proc. Natl. Acad. Sci. USA 76, 2996–3000.
[5] Oroszlan, S. and Gilden, R.V. (1980) in: Molecular Biology of RNA Tumor Viruses (Stephenson, J.R. ed) pp.299–344, Academic Press, New York.
[6] Long, C.W., Henderson, L.E. and Oroszlan, S. (1980) Virology 104, 491–496.
[7] Gupta, P. and Ferrer, J.F. (1980) J. Gen. Virol. 47, 311–322.
[8] Van Der Maaten, M.J., Miller, J.M. and Boothe, A.D. (1974) J. Natl. Cancer Inst. 52, 491–494.
[9] Gilden, R.V., Long, C.W., Hanson, M., Toni, R., Charman, H.P., Oroszlan, S., Miller, J.M. and Van Der Maaten, M.J. (1975) J. Gen. Virol. 29, 305–314.
[10] Olpin, J. and Oroszlan, S. (1980) Anal. Biochem. 103, 331–336.
[11] Morgan, M.A., Copeland, T.D. and Oroszlan, S. (1983) J. Virol. 46, 177–186.
[12] Henderson, L.E., Copeland, T.D., Sowder, R.C., Smythers, G.W. and Oroszlan, S. (1981) J. Biol. Chem. 256, 8400–8406.
[13] Henderson, L.E., Krutzsch, H.C. and Oroszlan, S. (1983) Proc. Natl. Acad. Sci. USA 80, 339–343.
[14] Versteegen, R.J., Copeland, T.D. and Oroszlan, S. (1982) J. Biol. Chem. 257, 3007–3013.
[15] Copeland, T.D., Grandgenett, D.P. and Oroszlan, S. (1980) J. Virol. 36, 115–119.
[16] Tarr, G.E., Beecher, J.F., Bell, M. and McKean, D.J. (1978) Anal. Biochem. 84, 622–627.
[17] Henderson, L.E., Copeland, T.D. and Oroszlan, S. (1980) Anal. Biochem. 102, 1–7.
[18] Oroszlan, S., Henderson, L.E., Stephenson, J.R., Copeland, T.D., Long, C.W., Ihle, J.N. and Gilden, R.V. (1978) Proc. Natl. Acad. Sci. USA 75, 1404–1408.
[19] Misono, K.S., Sharief, F.S. and Leis, J. (1980) Fed. Proc. 39, 1611.
[20] Dayhoff, M.O. (1976) in Atlas of Protein Sequence and Structure, vol.5, suppl.2, p.22, National Biomedical Research Foundation, Washington DC.