Complete amino acid sequence of human T-cell leukemia virus structural protein p15

Terry D. Copeland, Stephen Oroszlan, V.S. Kalyanaraman[+], M.G. Sarngadharan[+] and Robert C. Gallo*

Laboratory of Molecular Virology and Carcinogenesis, LBI-Basic Research Program, NCI-Frederick Cancer Research Facility, Frederick, MD 21701, [+]Department of Cell Biology, Litton Bionetics, Inc., Kensington, MD 20895 and **Laboratory of Tumor Cell Biology, National Cancer Institute, Bethesda, MD 20205, USA*

Received 7 July 1983; revised version received 5 September 1983

The complete amino acid sequence of human T-cell leukemia virus (HTLV) structural protein p15 has been determined. The intact protein and peptides generated by enzymatic digestion and acid cleavage were purified by reversed-phase liquid chromatography and subjected to semi-automated Edman degradation. HTLV p15 is a basic linear polypeptide composed of 85 amino acids with M_r 9458. The primary structure indicates that HTLV p15 is homologous to the nucleic acid binding proteins of other type-C retroviruses and especially related to bovine leukemia virus p12.

Human T-cell leukemia virus *Retrovirus* *Nucleic acid binding protein* *Micro-sequencing* *Primary structure determination*

1. INTRODUCTION

The isolation of human T-cell leukemia-lymphoma virus (HTLV) was first reported in [1]. Subsequently it has been shown that HTLV is an exogenous human retrovirus, etiologically associated with a subtype of adult T-cell leukemia [2], and a number of additional isolates of HTLV have been obtained [3,4]. Extensive biochemical and immunological studies provided evidence that HTLV is a unique retrovirus dinstinct from all previously known isolates [2,5]. Amino acid sequence analysis of the major internal protein p24 [6] confirmed these findings and revealed for the first time a structural relationship between HTLV p24 and protein homologs in other retroviruses, especially bovine leukemia virus (BLV). The initial NH_2-terminal analysis and the results of ongoing studies show an ~37% homology of the major core proteins indicating a distant but definitive evolutionary relationship between HTLV and BLV. Accumulated sequence data indicate that the basic nucleic acid binding proteins (NBP) of retroviruses are among the most highly conserved structural proteins encoded by the *gag* gene [7–9]. The amino acid sequence of BLV NBP (p12) has also been determined [10]. We here report the determination of the complete amino acid sequence of the low-M_r basic protein of HTLV, designated p15. HTLV p15 contains the putative nucleic acid binding domain involving a periodic placement of Cys residues [7–10]. This and other primary structure data indicate that HTLV p15 is a protein homolog of type C retrovirus NBPs.

The structure of HLTV-I_{CR} p15 was presented at the 1st Annual Meeting of the American Society for Virology, Ithaca, New York, August 2–5, 1982

2. MATERIALS AND METHODS

2.1. *Virus*

Several isolates of HTLV are available now and can be grouped based on their biological and immunological properties [3,4,11]. The strain designated HTLV-I_{CR}, grown in HUT-102 cells [1,3], was used here.

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

2.2. *Protein and peptide purification*

The details of the purification of HTLV p15 from sucrose gradient-banded virus are described in [12]. Briefly, sucrose-purified HTLV was disrupted and subjected to ion exchange chromatography on a phosphocellulose column. The peak eluting at 150–300 mM NaCl concentration containing mainly p15 and p24 was used for further purification by reversed-phase liquid chromatography (RPLC). The pH was lowered to pH 2 with aqueous trifluoroacetic acid (TFA). The sample was then applied to a Waters μBondapak C_{18} column which was developed by a specially designed acetonitrile gradient in 0.05% TFA for 110 min, as shown in fig.1. The same column was used to separate peptide mixtures. Mixtures were either directly applied to the column after the pH had been adjusted to ~2.0 with aqueous TFA or small amounts of 6 M guanidine hydrochloride were added to the sample before the pH was adjusted and then applied to the column. Peptides were eluted with a 0.05% TFA–acetonitrile (CH_3CH) gradient controlled by a Waters Model 660 solvent programmer and detected with a Waters 450 variable wavelength detector. Peak fractions were collected manually and aliquots taken for amino acid analysis.

2.3. *Amino acid analysis*

The composition of HTLV p15 and derived peptides were determined by amino acid analysis [13] on a Durrum 500 amino acid analyzer. Samples were hydrolyzed in vacuo with 6 N HCl containing 0.1% phenol at 110°C for 24 h.

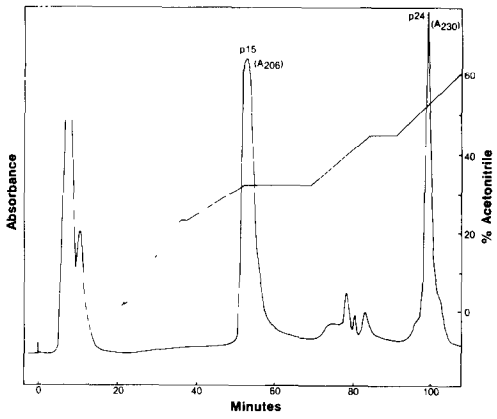

Fig. 1. Reversed-phase liquid chromatography (RPLC) elution profile of HTLV p15 and p24.

2.4. *Liquid-phase sequencing*

Semi-automated Edman degradation [14] was performed on a Beckman sequenator updated to the level of a Model 890C [15], in the presence of Polybrene [16]. A Waters μBondapak phenylalkyl column was used to identify and quantitate the phenylthiohydantoin (PTH) derivatives of amino acids [17]. Samples were loaded onto the column with a Waters WISP 910A automated injector.

2.5. *Carboxyl-terminal analysis*

HTLV p15 was digested with carboxypeptidase Y [18]. Amino acids released were determined by amino acid analysis.

2.6. *Reduction and S-carboxamidomethylation*

The procedure used for reductive alkylation of protein is based on the original method in [19]. HTLV p15 was reduced with dithiothreitol and *S*-carboxamidomethylated with iodoacetamide as in [7]. To remove excess reagents and side products, the pH of the reaction mixture was lowered to ~2 with aqueous trifluoroacetic acid and the mixture was loaded onto a C_{18} RPLC column and the modified protein was eluted with an acetonitrile gradient.

2.7. *Lys-C digestion*

S-carboxamidomethyl p15 was dissolved in 0.5 ml 0.1 M $NaHCO_3$ (pH 8.0) containing 20% acetonitrile [20]. Lys-C enzyme was added to give an enzyme/protein ratio of 1:100. The reaction proceeded for 24 h at room temperature. Digestion was stopped by addition of aqueous TFA.

2.8. *Limited trypsin digestion*

The lysines of *S*-carboxamidomethyl p15 were blocked with methylacetamide as in [7], the modified protein was dissolved in 0.1 M sodium bicarbonate (pH 8) with 1 M guanidine hydrochloride and digested with TPCK trypsin (enzyme/protein = 1:100) for 22 h.

2.9. *Formic acid cleavage*

S-carboxamidomethyl p15 was dissolved in 1.0 ml 70% formic acid [21], reacted at 37°C for 47 h, diluted with water and lyophilized.

3. RESULTS AND DISCUSSION

RPLC-purified HTLV p15 (fig.1) was homogenous by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) migrating with an apparent M_r of 15000, hence the designation p15.

The complete amino acid squence of HTLV p15 together with the sequencing strategy used is shown in fig.2. The primary structure was determined by semi-automated microsequence analysis utilizing NH_2-terminal Edman degradation of intact protein and fragments generated by various enzymatic and chemical cleavages (see legend to fig.2). Amino acid analyses were done on a Durrum 500 analyzer equipped with a PDP 8/A computing integrator. Values for the intact protein were based on the average of 3 samples hydrolyzed for 24, 48 and 72 h, respectively. HTLV p15 had the following composition (values calculated from the sequence are in parentheses): Asp 7.8 (6), Asn (2), Thr 4.1 (4), Ser 2.0 (2), Glu 8.9 (5), Gln (4), Pro 20.3 (20), Gly 4.9 (5), Ala 2.1 (2), Val 2.4 (3), Met 0 (0), Ile 2.8 (3), Leu 6.0 (6), Tyr 0 (0), Phe 0.9 (1), His 3.2 (3), Lys 5.9 (6), Arg 5.1 (5), Cys (6), Trp (2). Cysteine and tryptophan were not determined on the analyzer. The values for the other amino acids in excellent agreement with those derived from the complete sequence (fig.2). As with the intact protein, the compositions of the peptides (see fig.2) were also in good agreement with the number of residues found in the sequences. To determine the COOH-terminal sequence, 0.5 nmol of HTLV p15 was digested with carboxypeptidase Y [18] for 40 min. Subsequent analysis gave the following results for the released amino acids (nmol/nmol protein): Val 0.89, Glu 0.21, Gly 0.19. The deduced COOH-terminal sequence is then: -(Gly,Glu)-Val-OH.

HTLV p15 is a linear polypeptide consisting of 85 amino acid residues with a calculated M_r of 9458, a value which is substantially smaller than that estimated by SDS-PAGE. While this discrepancy between the actual molecular mass and electrophoretic mobility is not understood, it is a finding similar to previous observations made with low M_r highly polar retroviral proteins [7–10]. The structural data indicate that HTLV p15 is a basic

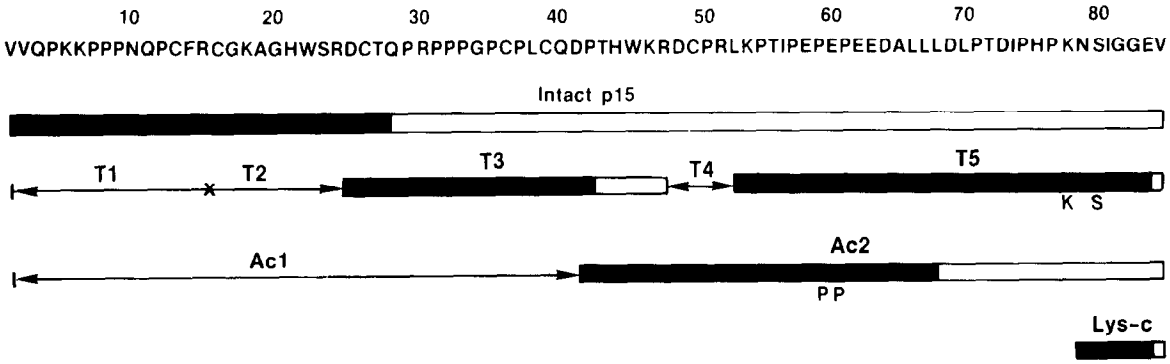

Fig. 2. The complete amino acid sequence of HTLV p15 and strategy for its determination. The structure shown was determined by semi-automated Edman degradation of intact protein and fragments derived from it by digestion with trypsin (fragments designated with letter T), by acid cleavage (fragments designated with letters Ac) and endoproteinase Lys-C (fragment designated Lys-c). The length (number of residues) of each peptide is indicated by horizontal bars or arrows. Five peptides (T1–T5) corresponding to the entire p15 molecule were purified from the limited (lysine residues were blocked) trypsin digest. Two peptides were obtained and purified after formic acid cleavage (Ac1 and Ac2). From the Lys-C digest of the unreduced protein only the carboxyl-terminal peptide was pursued. The intact protein and all peptides were analyzed for composition and sequenced. Those peptides required to prove the structure are shown with bars. The shaded area indicates the extent to which sequence could be accurately determined. The letters under the bars indicate residues (single letter code) which could not be unambiguously identified in the given peptide. The single letter code for amino acids is as follows: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lycine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

protein. It has 12 carboxyl groups (6 aspartic acid, 5 glutamic acid and 1 α-carboxyl group) and 15 basic groups (3 hystidine, 6 lysine, 5 arginine, and 1 α-amino group). It has no methionine or tyrosine. The most abundant amino acid is proline (23.5%). In general this high proline content is unusual for the NBPs of other retroviruses and is a property of HTLV shared only with BLV. In addition there are 6 cysteine residues in the structure of HTLV p15 as found for BLV p12 [10].

The alignment of the complete amino acid sequences of HTLV p15 and BLV p12 is shown in fig.3. The two proteins share the common NH_2-terminus valine (position number 1). In the first 69 positions of the alignment where numbering is based on the complete continuous sequence of BLV p12, there are 36 identities (52.5%), 20 differences and 3 gaps (deletions) involving a total of 13 positions (residues) shown by asterisks in the HTLV p15 sequence. The extent of nucleotide sequence identity between the genes coding for HTLV p15 and BLV p12 remains to be seen. Because of the degeneracy in the genetic code, it could be less than the 52.5% identity at the amino level. It should be noted however that in the 20 positions that are different, most of the substitutions could be the result of a single base change in the respective codon. In any event there are several identical amino acid sequence regions long enough to raise the possibility for the presence of identical nucleotide stretches detectable by nucleic acid hybridization. Nucleotide sequence homology has recently been found between molecularly cloned proviruses of HTLV and BLV using non-stringent molecular hybridization conditions (E. Gelman, R. Gallo, V. Manzari and F. Wong-Staal, in preparation).

```
                    5                   10                  15
BLV p12   Val-His-Thr-Pro-Gly-Pro-Lys-Met-Pro-Gly-Pro-Arg-Gln-Pro-Ala-
HTLV p15  Val-Val-Gln- * - * -Pro-Lys-Lys-Pro-Pro-Pro-Asn-Gln- * - * -

                   20                  25                  30
BLV p12   Pro-Lys-Arg-Pro-Pro-Pro-Gly-Pro-Cys-Tyr-Arg-Cys-Leu-Lys-Glu-
HTLV p15  * - * - * - * - * - * -Pro-Cys-Phe-Arg-Cys-Gly-Lys-Ala-

                   35                  40                  45
BLV p12   Gly-His-Trp-Ala-Arg-Asp-Cys-Pro-Thr-Lys-Ala-Thr-Gly-Pro-Pro-
HTLV p15  Gly-His-Trp-Ser-Arg-Asp-Cys-Thr-Gln- * - * -Pro-Arg-Pro-Pro-

                   50                  55                  60
BLV p12   Pro-Gly-Pro-Cys-Pro-Ile-Cys-Lys-Asp-Pro-Ser-His-Trp-Lys-Arg-
HTLV p15  Pro-Gly-Pro-Cys-Pro-Leu-Cys-Gln-Asp-Pro-Thr-His-Trp-Lys-Arg-

                   65                  70                  75
BLV p12   Asp-Cys-Pro-Thr-Leu-Lys-Ser-Lys-Asn
HTLV p15  Asp-Cys-Pro-Arg-Leu-Lys-Pro-Thr-Ile-Pro-Glu-Pro-Glu-Pro-Glu

                   80                  85                  90
HTLV p15  Glu-Asp-Ala-Leu-Leu-Leu-Asp-Leu-Pro-Thr-Asp-Ile-Pro-His-Pro

                   95
HTLV p15  Lys-Asn-Ser-Ile-Gly-Gly-Glu-Val
```

Fig. 3. Sequence alignment of BLV p12 and HTLV p15. Residues which are positionally identical are underlined. Asterisks idicate gaps introduced in the HTLV p15 sequence in order to maximize homology. Sequence of BLV p12 is taken from [10].

The polypeptide chain of HTLV p15 is 16 residues longer than that of BLV p12. The unique carboxyl-terminal sequence of HTLV beginning with position 70 of the alignment involves a total of 29 residues. This, together with the extent of amino acid sequence differences between HTLV p15 and BLV p12 in the comparable region (first 69 residues), emphasizes the distinctness of the two viruses and the uniqueness of HTLV.

Statistical analysis of the sequence data was carried out by the ALIGN program in [22]. This program is based on the mutation data matrix for scoring amino acid comparisons as published [22] and is designed to compute a numerical value for any given alignment of two sequences. When a break is introduced in order to obtain the best alignment, a penalty is assigned. Usually 300 random runs are performed to obtain the alignment score which is the number of standard deviations by which the maximum score for the real sequences exceeds the average maximum score for the random permutations. The alignment score calculated for BLV p12 and HTLV p15 was 7.5 (table 1), a highly significant value which indicates a definitive evolutionary relationship between HTLV and BLV. Alignments of the BLV p12 and HTLV p15 sequences with known sequences of NBPs of several type C viruses were also made. The alignment scores obtained are summarized in table 1. It is seen that HTLV p15 and BLV p12 are more related to each other than to NBPs of other retroviruses. An alignment score of 3.0 has been accepted as evidence implying common ancestry

Table 1

Alignment scores[a] for the basic nucleic acid binding proteins of various type C viruses[b]

	BLV_{p12}	$R\text{-}MuLV_{p10}$	$FeLV_{p10}$	ASV_{p12}
$HTLV_{p15}$	7.5	4.4	3.8	4.0
BLV_{p12}		4.1	2.4	2.8
$R\text{-}MuLV_{p10}$			13.0	4.6
$FeLV_{p10}$				3.7

[a]Number of standard deviations of real score above random score
[b]Sequences other than that of HTLV p15 were taken from previous publications: BLV p12 from [10]; R-MuLV p10 from [7]; FeLV p10 from [8] and ASV p12 from [24]

[22,23]. In this respect it is important to point out that the alignment of HTLV p15 with the NBPs of Rauscher-murine leukemia virus (R-MuLV), feline leukemia virus (FeLV) and avian sarcoma virus (ASV) resulted in scores of 4.4, 3.8 and 4.0, respectively (table 1). BLV p12 *vs* R-MuLV p10 scored a value of 4.1. The scores calculated for BLV p12 and FeLV p10 as well as for the BLV p12 and ASV p12 pair (2.4 and 2.8, respectively) are slightly below the value considered statistically significant (3.0 standard deviation above the mean) but FeLV p10 and ASV p12 show a significant relationship, not only to each other but to both R-MuLV p10 and HTLV p15. The combined results of these statistical analyses therefore permit us to group HTLV p15 and BLV p12 together with the NBPs of the other type C viruses (both mammalian and avian) in an evolutionarily related superfamily of proteins.

It has been previously shown that the putative nucleic acid binding domain including a set of 3 cysteines is duplicated in BLV p12 and ASV p12 [10]. Similar internal sequence homology also occurs in HTLV p15. The alignment of these homologous regions is shown in fig.4. It is seen that in these 19-residue long segments a total of 10 amino acids are positionally identical. These again include the 3 cysteines periodically placed at positions n (first cysteine in each set), $n+3$ and $n+13$.

Fig. 4. Internal duplication of sequence in HTLV p15. In the alignment of the fragments the repeated residues are boxed.

No internal sequence homology has been found in NBPs of the murine and feline type C viruses [7–10]. Therefore, on the basis of the occurrence of this partial sequence repetition due to gene duplication, HTLV p15, BLV p12 and ASV p12 may be classified as forming a distinct family within the superfamily of type C virus NBPs as proposed above.

The complete amino acid sequence of $HTLV\text{-}I_{CR}$ p15 will allow characterization of additional new members of the HTLV family [3,4], including the adult T cell leukemia virus (ATLV) recently isolated in Japan [25], which is already known to be highly related to $HTLV\text{-}I_{CR}$ [4,11] and therefore a member of the HTLV group. During the preparation of this manuscript, the complete nucleotide sequence of the proviral ATLV DNA was reported [26]. A comparison of the amino acid sequence of $HTLV\text{-}I_{CR}$ p15 with the deduced sequence of ATLV shows a single amino acid difference only. Residue Thr^{72} in $HTLV\text{-}I_{CR}$ p15 is exchanged for Ala in ATLV. This may reflect a strain difference between the two independently isolated HTLVs.

ACKNOWLEDGEMENTS

We wish to acknowledge the expert technical assistance of Young D. Kim. Research sponsored in part by the National Cancer Institute, DHHS, under contract no. N01-CO-23909 with Litton Bionetics, Inc. 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] Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D. and Gallo, R.C. (1980) Proc. Natl. Acad. Sci. USA 77, 7415–7419.

[2] Gallo, R.C. and Wong-Staal, F. (1982) Blood 60, 545–557.

[3] Kalyanaraman, V.S., Sarngadharan, M.G., Robert-Guroff, M., Miyoshi, I., Blayney, D., Golde, D. and Gallo, R.C. (1982) Science 218, 571–574.

[4] Popovic, M., Sarin, P.S., Robert-Guroff, M., Kalyanaraman, V.S. and Gallo, R.C. (1983) Science 219, 856–859.

[5] Kalyanaraman, V.S., Sarngadharan, M.G., Poiesz, B.J., Ruscetti, F.W. and Gallo, R.C. (1981) J. Virol. 38, 906–913.

[6] Oroszlan, S., Sarngadharan, M.G., Copeland, T.D., Kalyanaraman, V.S., Gilden, R.V. and Gallo, R.C. (1982) Proc. Natl. Acad. Sci. USA 79, 1291–1294.

[7] Henderson, L.E., Copeland, T.D., Sowder, R.C., Smythers, G.W. and Oroszlan, S. (1981) J. Biol. Chem. 256, 8400–8406.

[8] Copeland, T.D., Morgan, M.A. and Oroszlan, S. (1983) submitted.

[9] Morgan, M.A., Copeland, T.D. and Oroszlan, S. (1983) J. Virol. 46, 177–186.

[10] Copeland, T.D., Morgan, M.A. and Oroszlan, S. (1983) FEBS Lett. 156, 37–40.

[11] Popovic, M., Reitz, M.S., Sarngadharan, M.G., Robert-Guroff, M., Kalyanaraman, V.S., Nakao, Y., Miyoshi, I., Minowada, J., Yoshida, M., Ito, Y. and Gallo, R.C. (1982) Nature 30, 63–66.

[12] Kalyanaraman, V.S., Morar, M.J., Sarngadharan, M.G. and Gallo, R.C. (1983) Virology, in press.

[13] Spackman, D.H., Stein, W.H. and Moore, S. (1958) Anal. Chem. 30, 1190–1206.

[14] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.

[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] Martin, B., Svendsen, I. and Ottesen, M. (1977) Carlsberg Res. Commun. vol. 42, pp. 99–102.

[19] Crestfield, A.M., Moore, S. and Stein, W.H. (1963) J. Biol. Chem. 238, 622–627.

[20] Henderson, L.E., Krutzch, H.C. and Oroszlan, S. (1983) Proc. Natl. Acad. Sci. USA 80, 339–343.

[21] Versteegen, R.J., Copeland, T.D. and Oroszlan, S. (1982) J. Biol. Chem. 257, 3007–3013.

[22] Dayhoff, M.O. (1976) in: Atlas of Protein Sequence and Structure, vol.5, Suppl. 2, pp. 1–8, National Biomedical Research Foundation, Washington, D.C.

[23] Doolittle, R.F. (1981) Science 214, 149–159.

[24] Misono, K.S., Sharief, F.S. and Leis, J. (1980) Fed. Proc. 39, 1611.

[25] Yoshida, M., Miyoshi, I. and Hinuma, Y. (1982) Proc. Natl. Acad. Sci. USA 79, 2031–2035.

[26] Seiki, M., Hattori, S., Hirayama, Y. and Yoshida, M. (1983) Proc. Natl. Acad. Sci. USA 80, 3618–3622.