The Lassa fever virus L gene: nucleotide sequence, comparison, and precipitation of a predicted 250 kDa protein with monospecific antiserum

Igor S. Lukashevich,1† Mahmoud Djavani,1 Keli Shapiro,1 Anthony Sanchez,2 Eugene Ravkov,2 Stuart T. Nichol2 and Maria S. Salvato1

1 Department of Pathology and Lab Medicine, University of Wisconsin Medical School, 1300 University Avenue, Madison, WI 53706, USA
2 Special Pathogen Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA

Lassa fever virus (LAS), the most dangerous human pathogen among the Arenaviridae, has two single-stranded RNA genomic segments termed large (L) and small (S) (Lukashevich et al., 1984). The LAS S RNA (3.4 kb) encodes the nucleoprotein (NP) and the envelope glycoprotein (GP) in an ambisense coding arrangement (Clegg & Oram, 1985; Auperin et al., 1986). Arenavirus L RNA segments have been described for lymphocytic choriomeningitis (LCM), Tacaribe (TAC) and Pichinde (PIC) viruses (Salvato & Shimomaye, 1989; Iapalucci et al., 1989; D. G. Harnish, S. Zheng & S. Polyak, unpublished results). The L segment (7.2 kb) encodes the viral RNA-dependent RNA polymerase (RdRp or L protein) on the genome-complementary strand and an 11 kDa ring-finger protein (Z) on the genomic strand. Both of these proteins are involved in regulation of transcription and replication processes and one or both may be implicated in the pathogenic potential of the virus (Salvato, 1993).

We report here the sequence of the LAS L gene. LAS L RNA was obtained by immunoprecipitating nucleocapsids from infected cells, extracting them with phenol–chloroform, and isolating the L RNA segment on linear 15–30% sucrose/0.5% SDS gradients as described (Lukashevich et al., 1984, 1988). The majority of the LAS L gene sequence was obtained from four partially overlapping clones (details available from the authors on request). Initially, approximately 3000 cDNA clones were obtained using purified LAS RNA as template; however, only three clones hybridized to a LAS L RNA probe. Sequences at the 5’ end were identified by dideoxy sequencing of uncloned nested PCR products derived from the 5’–3’ ligated junction. Briefly, gradient purified LAS L RNA was treated with tobacco acid pyrophosphatase to remove cap structures and circularized with T4 RNA ligase (Epicenter Technologies), followed by RT–PCR using nested primers flanking the junction region. The use of gradient purified LAS L RNA was critical for RT–PCR amplification across the ligated 5’–3’ area, as only the purified RNA template gave L-specific PCR products with expected sizes.

Similar to the LAS S RNA and genomic RNA segments of other arenaviruses (Auperin et al., 1982), the 3’ end of the LAS L RNA contains 19 nucleotides complementary to the 5’ end. Computer-assisted analysis revealed the presence of one major (L) open reading frame (ORF) on the genome-complementary strand (GenBank accession no. U63094), and a second ORF corresponding to an 11 kDa ring-finger protein on the 5’ end.

The complete nucleotide sequence of Lassa virus L gene has been deposited in GenBank under accession no. U63094.

The large (L) RNA segment of Lassa fever virus (LAS) encodes a putative RNA-dependent RNA polymerase (RdRp or L protein). Similar to other arenaviruses, the LAS L protein is encoded on the genome-complementary strand and is predicted to be 2218 amino acids in length (253 kDa). It has an unusually large non-coding region adjacent to its translation start site. The LAS L protein contains six motifs of conserved amino acids that have been found among arenavirus L proteins and core RdRp of other segmented negative-stranded (SNS) viruses (Arena-, Bunya- and Orthomyxoviridae). Phylogenetic analyses of the RdRp of 20 SNS viruses reveals that arenavirus L proteins represent a distinct cluster divided into LAS–lymphocytic choriomeningitis and Tacaribe–Pichinde virus lineages. Monospecific serum against a synthetic peptide corresponding to the most conserved central domain precipitates a 250 kDa product from LAS and lymphocytic choriomeningitis virus-infected cells.
The LAS L ORF is preceded by a 157 nucleotide non-coding region, five times longer than that for LCM, TAC and PIC L RNA (Salvato et al., 1989; Iapalucci et al., 1989; D. G. Harnish, S. Zheng & S. Polyak, unpublished data). The L ORF initiates at nucleotide 158 within AAC AUGGAG and ends with an amber codon at nucleotide 6812. It encodes 2218 amino acid residues with a calculated molecular mass of 253 kDa, and is rich in A + U (60%).

The LAS L protein contains 12.8% acidic (D, E) and 13.9% basic (K, R, H) amino acids. Assuming that K and R = +1, H is +0.5, and D and E are −1 at neutral pH, the charges of acidic and basic residues in LAS L protein are equivalent (± = 0). The estimated isoelectric point of the protein is 6.1.

The LAS L protein contains nine clusters of basic amino acids with a positive net charge of +4 or more (positions 246–251, 331–343, 632–642, 776–783, 860–868, 1255–1265, 1492–1505, 1772–1780 and 2180–2202); some or all of these clusters may be essential for interaction with template RNA. The carboxy-terminal sequence of the protein is highly acidic, since 7 of 14 carboxy-terminal residues are aspartic or glutamic acid. A similar carboxy-terminal acidic domain is found in the L proteins of LCM, tomato spotted wilt and Sin Nombre viruses (Salvato et al., 1989; de Haan et al., 1991; Chizhikov et al., 1995).

Aromatic (F, W, Y) and hydrophobic (I, L, M, F, W, Y, V) residues represent 9.4% and 32.6% of the total amino acids, respectively. In good agreement with the composition of other L proteins, the LAS L protein exhibits a high content of leucine + isoleucine (17.8%). A hydrophilic profile (not shown) suggests that the L protein has neither strongly hydrophobic nor hydrophilic character.

Sequence similarity between LAS, LCM, TAC and PIC arenaviruses is more extensive at the amino acid level than at the nucleotide level. LAS L protein has 67.2/48.0% amino acid similarity/identity with LCM L protein; TAC and PIC L proteins have less similarity/identity with LAS L protein: 59.1/37.6% and 58.7/38.6%, respectively. As seen in Fig. 1, stretches of conserved residues are found in the amino-terminal and central portions of arenavirus and bunyavirus L proteins that are not found among other RdRp (Muller et al., 1994). The most conserved sequence has been located in a central domain of the arenavirus L proteins (amino acid position 1000–1400). Residues 1121–1152 of the LAS L protein are identical to residues 1113–1144 in the LCM L protein (Salvato et al., 1989) and were used to produce anti-L rabbit serum.

The centre of the LAS L protein contains at least six strictly conserved amino acid motifs that have been found among arenavirus L proteins and core polymerases of segmented negative-stranded (SNS) viral RdRp (Poch et al., 1990; Muller et al., 1994). The LAS RdRp motifs were aligned with LCM sequences as shown in Fig. 1(b). The motifs represent stretches of highly conserved sequences that are linked by less conserved variable regions. It has been proposed that motifs A and C are involved in the RdRp catalytic activity and two strictly conserved aspartate residues located close to each other are important for interacting with metal ions. The strictly conserved lysine residue of motif D also seems to be important for catalytic activity because of its proximity to the aspartate of motif A (Ishihama & Barbier, 1994; Muller et al., 1994). Motif B is common to all RdRp and contains a strictly conserved Gln residue, generally thought to allow mobility of the peptide backbone, and probably essential for RNA binding. The conserved sequence EF/YXS (E motif) located downstream
sequences of LAS L protein are not conserved, leading to the suggestion that these sequences may interact with cell-encoded trans-acting transcription factors (Chizhikov et al., 1995). Taken together, the characteristics of the LAS L protein support the possibility that it is the viral RdRp.

Phylogenetic analysis of viruses is most valuable when it is based on sequences of viral polymerase. Amino acid replacement mutations in the RdRp genes are much less frequent than those in structural genes and the degree of variability of RdRp is lowest among all viral proteins, even though the frequency of silent mutations is equivalent amongst viral genes (Komase et al., 1995). The RdRp molecular structure is closely correlated with virus replication and may therefore be a more constant phylogenetic characteristic than envelope protein genes for example (Ishihama & Barbier, 1994).

Phylogenetic relationships between SNS viruses on the basis of their core RdRp sequences were analysed by weighted maximum parsimony (PAUP 3.1.1). Highly conserved regions with flanking sequences were aligned as described by Muller et al. (1994) and Marriott & Nuttall (1996) and analysed using PAUP. As seen in Fig. 2(a), the RdRp of SNS viruses represent a phylogenetic tree divided into three major branches: (i) arena-, phlebo-, nairo- and orthomyxoviruses [including Dhori (DHO) virus, an arbovirus]; (ii) bunya- and tospoviruses; and (iii) hantaviruses. A similar phylogenetic distribution of the Bunyaviridae RdRp sequences was noted recently using different methods of phylogenetic analysis and/or including in the data set entire L sequences (Chizhikov et al., 1995; Roberts et al., 1995; Elliott, 1996). Clustering together bunyaviruses (Bunyamwera and La Crosse viruses) and tospoviruses (tomato spotted wilt virus, a plant pathogen transmitted by thrips) reflects a phylogenetic relationship and suggests the existence of a common progenitor. Tenuivirus RdRp is more closely related to the phleboviruses than to the tospoviruses (Fig. 2(a)), suggesting that tospo- and tenuiviruses spread independently between animal and plant hosts (Elliott, 1996).

All arenavirus L proteins cluster into one branch divided into two lineages, LAS–LCM and TAC–PIC, in agreement with serological studies and NP and GP sequence data (Clegg, 1993; Bowen et al., 1996; Gonzalez et al., 1996). The clustering of influenza virus RdRp core sequences (including RdRp of DHO virus) together with phlebo-, nairo- and arenaviruses, as seen in Fig. 2(a), has also been noted recently by Marriott & Nuttall (1996). It may reflect unknown phylogenetic relationships between the core RdRp sequences of these SNS viruses, or indicate some limitations of using RdRp as a phylogenetic marker (Zanotto et al., 1996). When the phylogenetic data set was expanded to include the aligned amino-terminal motifs of Arena- and Bunyaviridae in the parsimony analysis, the shortest tree (Fig. 2b) was quite similar to that noted above (Fig. 2a) and also divided into three branches: arena–phleboviruses, bunya–tospoviruses and hantaviruses.

Antisera to a peptide from the conserved central domain of LAS L protein, residues 1121–1153, was used for radioimmune

Fig. 2. Genetic comparison of SNS viruses based on the analysis of RdRp sequences. Phylogenetic analysis of amino acid differences between RdRp was carried out by maximum parsimony (PAUP 3.1.1) using a heuristic search option and a weight matrix based on the minimum number of nucleotide substitutions required to convert one amino acid to another. Bootstrap confidence limits were calculated from 500 replicates and are indicated below the lines. Only branches present in > 50% of the trees are shown. Horizontal branch lengths are indicated above the lines and are proportional to the number of substitution steps. (a) Phylogenetic tree built on the basis of the RdRp core sequences of SNS viruses. (b) Phylogenetic analysis of arena- and bunyavirus RdRp sequences. Amino-terminal conserved motifs (Muller et al., 1994) and core sequences were included in a 678 residue data set. All RdRp sequences were taken from the latest release of the GenBank and EMBL databases: LCM (J04331); TAC (J04340); PIC (D. G. Harnish, S. Zheng & S. Polyak, unpublished data); BUN, Bunyamwera (X14383); LAC, La Crosse (U12396); RVF, rift valley fever (X56464); TSW, tomato spotted wilt (D10066); HTN, Hantaan 76–118 (X55901); SEO, Seoul 80–39F (X56492); PUU, 1993; Bowen et al., 1996; Gonzalez et al., 1996). Clustering together bunyaviruses (Bunyamwera and La Crosse viruses) and tospoviruses (tomato spotted wilt virus, a plant pathogen transmitted by thrips) reflects a phylogenetic relationship and suggests the existence of a common progenitor. Tenuivirus RdRp is more closely related to the phleboviruses than to the tospoviruses (Fig. 2a), suggesting that tospo- and tenuiviruses spread independently between animal and plant hosts (Elliott, 1996).

Fig. 2. Genetic comparison of SNS viruses based on the analysis of RdRp sequences. Phylogenetic analysis of amino acid differences between RdRp was carried out by maximum parsimony (PAUP 3.1.1) using a heuristic search option and a weight matrix based on the minimum number of nucleotide substitutions required to convert one amino acid to another. Bootstrap confidence limits were calculated from 500 replicates and are indicated below the lines. Only branches present in > 50% of the trees are shown. Horizontal branch lengths are indicated above the lines and are proportional to the number of substitution steps. (a) Phylogenetic tree built on the basis of the RdRp core sequences of SNS viruses. (b) Phylogenetic analysis of arena- and bunyavirus RdRp sequences. Amino-terminal conserved motifs (Muller et al., 1994) and core sequences were included in a 678 residue data set. All RdRp sequences were taken from the latest release of the GenBank and EMBL databases: LCM (J04331); TAC (J04340); PIC (D. G. Harnish, S. Zheng & S. Polyak, unpublished data); BUN, Bunyamwera (X14383); LAC, La Crosse (U12396); RVF, rift valley fever (X56464); TSW, tomato spotted wilt (D10066); HTN, Hantaan 76–118 (X55901); SEO, Seoul 80–39F (X56492); PUU, 1993; Bowen et al., 1996; Gonzalez et al., 1996). Clustering together bunyaviruses (Bunyamwera and La Crosse viruses) and tospoviruses (tomato spotted wilt virus, a plant pathogen transmitted by thrips) reflects a phylogenetic relationship and suggests the existence of a common progenitor. Tenuivirus RdRp is more closely related to the phleboviruses than to the tospoviruses (Fig. 2a), suggesting that tospo- and tenuiviruses spread independently between animal and plant hosts (Elliott, 1996).

All arenavirus L proteins cluster into one branch divided into two lineages, LAS–LCM and TAC–PIC, in agreement with serological studies and NP and GP sequence data (Clegg, 1993; Bowen et al., 1996; Gonzalez et al., 1996). The clustering of influenza virus RdRp core sequences (including RdRp of DHO virus) together with phlebo-, nairo- and arenaviruses, as seen in Fig. 2(a), has also been noted recently by Marriott & Nuttall (1996). It may reflect unknown phylogenetic relationships between the core RdRp sequences of these SNS viruses, or indicate some limitations of using RdRp as a phylogenetic marker (Zanotto et al., 1996). When the phylogenetic data set was expanded to include the aligned amino-terminal motifs of Arena- and Bunyaviridae in the parsimony analysis, the shortest tree (Fig. 2b) was quite similar to that noted above (Fig. 2a) and also divided into three branches: arena–phleboviruses, bunya–tospoviruses and hantaviruses.

Antisera to a peptide from the conserved central domain of LAS L protein, residues 1121–1153, was used for radioimmune

of the lysine residue in motif D has been found only in SNS viruses and it is thought to be involved in the ‘capsnatching’ mechanism of transcription that is described only for SNS viruses (Kolakofsky & Hacker, 1991). Also, a conserved ATP-binding motif [GXGXGXK] has been located at position 932–953 within the LAS L protein. Similar to other defined L proteins from SNS and non-SNS viruses, carboxy-terminal
precipitation of a 250 kDa protein from infected cells (Fig. 3). Vero or BHK-21 cells were infected with 5 p.f.u. LAS or LCM per cell, labelled 48 h later with 50 µCi/ml [35S]methionine for 2 h, and lysed in 1% NP40, 50 mM Tris–HCl, pH 8, 150 mM NaCl, 10% glycerol, 1 mM CaCl₂, 0.5 mM MgCl₂ (or 0.1 mM EDTA), 1 mM PMSF, 0.1% aprotinin and 0.1% iodoacetamide. Proteins were precipitated by incubation with LAS-specific sera followed by suspension with 10% formalin-fixed Staphylococcus aureus (Sigma) or Protein A–Sepharose-CL4B (Pharmacia). Precipitates were washed in the above buffer without detergent and analysed on 7–5% acrylamide gels (Fig. 3). Initial attempts to precipitate L protein from infected cells were unsuccessful and yielded only low molecular mass products (Fig. 3a, lanes 1 and 2). When the lysis buffers were changed to include magnesium (instead of EDTA), a high molecular mass protein appeared on polyacrylamide gels (Fig. 3a, lane 3, and 3b, lanes 2–5). We do not know whether this is due to stabilization of RNA–L protein complexes or to inhibition of proteases. Published attempts to immunoprecipitate large RdRp often describe the co-precipitation of small proteins thought to be nucleocapsid proteins; however, the small proteins seen in Fig. 3(a) are likely to be polymerase fragments since they are replaced by a high molecular mass protein when precipitation conditions are changed. The 250 kDa protein precipitated by monospecific serum from LAS-infected cells co-migrates with the largest LAS gene product precipitated by serum from LAS-infected monkeys. Thus, the protein sequence predicted from the LAS nucleotide sequence has been confirmed as a product of infected cells.

This work was supported by NIH Grant AI32107 (M.S.) and an International Scientific Foundation Grant, MWF000 (I.L.). We thank Dr A. Vladyko (BRIEM, Belarus) for LAS-specific monkey serum. We thank Dorothy Cheung for immunoprecipitation of infected cell extracts in Fig. 3(a), Doug Horejsh for extensive help in phylogenetic analysis, and Dr C. David Pauza for critical reading of the manuscript. Our ability to complete the LAS L RNA sequence benefitted from unpublished technical
information from the laboratories of Peter J. Southern, Christopher Clegg, Delsworth Harnish and David Auperin.


References


Received 16 September 1996; Accepted 11 November 1996