Domain Structure of Lassa Virus L Protein

Linda Brunotte, Michaela Lelke, Meike Hass, Katja Kleinsteuber, Beate Becker-Ziaja, and Stephan Günther

Department of Virology, Bernhard-Nocht-Institute for Tropical Medicine, 20359 Hamburg, Germany

Short title: Domain Structure of Lassa Virus L Protein

*Corresponding author: Stephan Günther
Bernhard-Nocht-Institute of Tropical Medicine
Bernhard-Nocht-Strasse 74
D-20359 Hamburg, Germany
Phone: +49 40 42818 421, Fax: +49 40 42818 378
e-mail: guenther@bni.uni-hamburg.de
The 200-kDa L protein of arenaviruses plays a central role in viral genome replication and transcription. This study aimed at providing evidence for the domain structure of L protein by combining bioinformatics with a stepwise mutagenesis approach using the Lassa virus minireplicon system. Potential inter-domain linkers were predicted using various algorithms. The prediction was challenged by insertion of flexible sequences into the predicted linkers. Insertion of 5 or 10 amino acid residues was tolerated at seven sites (S407, G446, G467, G774, G939, S1952, and V2074 in Lassa virus AV). At two of these sites, G467 and G939, L protein could be split into an N-terminal and a C-terminal part, which were able to trans-complement each other and re-constitute a functional complex upon co-expression. Co-immunoprecipitation studies revealed physical interaction between the N- and C-terminal domains, irrespective of whether L protein was split at G467 or G939. In confocal immunofluorescence microscopy, the N-terminal domains showed a dot-like, sometimes perinuclear, cytoplasmic distribution similar to full-length L protein, while the C-terminal domains were homogenously distributed in cytoplasm. The latter were re-distributed into the dot-like structures upon co-expression with the corresponding N-terminal domain. In conclusion, this study demonstrates two inter-domain linkers in Lassa virus L protein, at G467 and G939, suggesting that L protein is composed of at least three structural domains spanning residues 1 – 467, 467 – 939, and 939 – 2220. The first domain seems to mediate accumulation of L protein into cytoplasmic dot-like structures.
Lassa virus is a segmented negative-strand RNA virus of the family *Arenaviridae*. It belongs to the Old World complex of the arenaviruses, which also includes the prototype species of the family, lymphocytic choriomeningitis virus (LCMV). Lassa virus is endemic in the West African countries of Sierra Leone, Guinea, Liberia, and Nigeria. It is transmitted by rodents of the species *Mastomys natalensis* and causes Lassa fever in humans. The disease is characterized by bleeding, organ failure, and shock (15).

The arenavirus genome consists of two single-stranded RNA segments, each containing two genes in opposite directions, a coding strategy called ambisense (2). The S RNA segment encodes the nucleoprotein (NP) and the glycoprotein precursor. The L RNA encodes the small matrix protein Z (26, 34) and the 200-kDa L protein (31). The minimal viral trans-acting factors required for replication and transcription of the genome are NP and L protein (17, 20, 22). The central region of L protein from amino acid residue 1040 to 1540 harbors the RNA-dependent RNA polymerase (RdRp) (10, 13, 18, 23, 37). The enzyme mediates the synthesis of two RNA species, mRNA terminating in the intergenic region and non-capped genomic or antigenomic RNA representing a full-length copy of the genome (12, 25). The N terminus of L protein is involved in mRNA synthesis (21). It presumably contains a cap-snatching endonuclease generating primers for virus mRNA synthesis, similar as described for influenza virus PA protein (7, 16, 39). Two binding sites for Z protein have been mapped in L protein, one in the N terminus between residues 160 and 290 overlapping with the putative endonuclease and one in the RdRp (38). Like L proteins of other negative-strand RNA viruses (32, 33, 40), arenavirus L protein oligomerizes which may be important for function (28).

Because of its large size and multiple functions in viral genome replication and transcription, it is assumed that L protein is composed of several domains. Sequence comparison revealed the existence of four conserved regions (I, positions 1 – 300; II, 500 – 900; III, 1000 – 1600; IV, 1700 – 2000) (37), which could correspond to structural domains. This study aimed at providing experimental evidence for the domain structure of L protein. We combined inter-
domain linker prediction with a stepwise mutagenesis approach using the Lassa virus minireplicon system (17). The functional studies were complemented with co-immunoprecipitation and co-immunofluorescence experiments to demonstrate interaction between the L protein domains as well as their intracellular localization.

Material and Methods

Lassa virus replicon system. Plasmids pCITE-NP and pCITE-L expressing Lassa virus NP and L protein under the control of a T7 RNA polymerase promoter have been described previously (17). Minigenome (MG) plasmid pLAS-MG contains the T7 promoter followed by a single additional G residue, the 5' noncoding region, chloramphenicol acetyltransferase (CAT) gene, intergenic region, Renilla luciferase (Ren-Luc) gene in reverse orientation, and 3' noncoding region. For transfection, the MG, including T7 promoter, was amplified for 25 cycles with Phusion DNA polymerase (Finnzymes), 3 ng of linearized pLAS-MG as a template, and vector-specific primer pUC-fwd and primer LVS-3400-rev (CGCACAGTGGATCCTAGGCTATTGGA) to generate a functional 3' end. Amplified MG was purified by using PCR-purification kit (Macherey & Nagel) and quantified spectrophotometrically.

Mutagenesis of L gene and cloning. The functional cassette of pCITE-L (T7 RNA polymerase promoter, internal ribosome entry site, and L gene) was amplified by mutagenic PCR, and the resulting PCR products were used for transfection without prior cloning as described (18). PCR was performed with Phusion DNA polymerase (Finnzymes).

For insertion of a 5-amino acid linker (GDGSGX, where X represents the amino acid upstream of the linker, leading to duplication of this residue), two fragments were amplified by using 10 ng of linearized pCITE-L as a template and primer combinations pUC(–110)-fwd/Linker-5-rev and Linker-5-fwd/pUC(–123)-rev, respectively. Both Linker-5 primers
contained a 5' tail encoding the linker sequence. PCR products were gel-purified and fused together in a second PCR containing aliquots of both fragments as a template and primers pUC(–47)-fwd and pUC(–48)-rev.

To elongate the linker sequence from 5 to 10 amino acid residues (GDGSGGKADGX), purified PCR products obtained with primer combinations pUC(–110)-fwd/Linker-5-rev and Linker-5-fwd/pUC(–123)-rev were further amplified using primers pUC(–110)-fwd/Linker-10-rev and Linker-10-fwd/pUC(–123)-rev, respectively. Both Linker-10 primers contain a 5' tail encoding the linker sequence GKADG. PCR products were gel-purified and fused together in a second PCR containing aliquots of both fragments as a template and primers pUC(–47)-fwd and pUC(–48)-rev.

Constructs for expression of individual L protein domains were generated by PCR using 10 ng of pCITE-L as a template. N-terminal domains were amplified with primers pUC(–110)-fwd and L-N-rev adding an artificial stop codon to the domain. C-terminal domains were amplified with primers pUC(–123)rev and L-C-IRES-fwd adding IRES sequences with an artificial start codon to the domain. The T7 promoter and the complete IRES were amplified with primers pUC-fwd and L-C-IRES-rev and subsequently fused in a second PCR to the amplified C-terminal domain using primers pUC(–47)-fwd and pUC(–48)-rev.

For generation of hemagglutinin (HA)-tagged L gene PCR product, the pUC(–48)-rev primer in the fusion PCR was replaced by primer L-HA-rev adding a C-terminal HA tag to the L gene. Before transfection, PCR-based L gene constructs were purified by using a purification kit (Macherey & Nagel) and quantified spectrophotometrically. Selected N- and C-terminal domains (N1, N2, C1, and C2) were cloned into pCITE-2a (Novagen) and tagged with HA or 3xFLAG sequences. The presence of artificial mutations was ascertained by sequencing the final PCR product or plasmid.
Wild-type L gene and an inactive mutant with a mutation in the catalytic site of the RdRp were generated by PCR using primers binding to the L gene around amino acid position 1334. They served as positive and negative controls, respectively, for the transfection experiments.

**Cells and transfections.** BSR-T7/5 cells (4) stably expressing T7 RNA polymerase were grown in Glasgow’s minimal essential medium (GMEM, GIBCO) supplemented with 5% fetal calf serum (FCS). Every second passage, 1 mg Geneticin (GIBCO) per ml of medium was added to the cells. BHK-21 cells were grown in Dulbecco’s modified Eagle’s medium (DMEM, Gibco) supplemented with 10% FCS. All transfections were performed with 3 µl Lipofectamine 2000 (Invitrogen) per µg DNA in cell culture medium without supplements. Medium was replaced 4 h after transfection by fresh medium complemented with FCS.

**Replicon assay.** BSR-T7/5 cells were seeded at a density of 1 x 10^5 cells per well of a 24-well plate one day prior to transfection. Cells in a well were transfected with 250 ng of MG, 250-500 ng of L gene PCR product, 250 ng of pCITE-NP, and 10 ng of pCITE-FF-luc (expression construct for firefly luciferase) as a transfection control. The amount of transfected DNA was kept constant within an experiment by adding empty pCITE-2a vector. Cells were lysed in 100 µl passive lysis buffer (Promega) per well 24 h post transfection and 20 µl thereof were assayed for firefly luciferase and Ren-Luc activity using the Dual-Luciferase Reporter assay system (Promega) as described by the manufacturer. Ren-Luc levels were corrected with the firefly luciferase levels (resulting in standardized relative light units [sRLU]) to compensate for differences in transfection efficiency or cell density.

**Expression of L protein.** BSR-T7/5 cells were seeded at a density of 1 x 10^5 cells per well of a 24-well plate one day prior to transfection. They were inoculated with modified vaccinia virus Ankara expressing T7 RNA polymerase (MVA-T7) (35) at a multiplicity of infection (MOI) of 5 for 1 h before transfection. Cells in two wells of a 24-well plate were transfected with 1 µg of PCR product for expression of HA-tagged L protein mutants or domains and lysed in 100 µl passive lysis buffer (Promega) containing Complete Protease Inhibitor.
Cocktail (Roche) 24 h after transfection. Nuclei were pelleted by centrifugation, and the cytoplasmic lysate was mixed with 4x NuPAGE LDS Sample Buffer (Invitrogen) complemented with dithiothreitol (DTT). Proteins were separated by 4 – 12% Bis-Tris polyacrylamide gel electrophoresis (PAGE), transferred to nitrocellulose membrane (Schleicher & Schuell), and detected by immunoblotting (see below).

**Co-immunoprecipitation studies.** BHK-21 cells were seeded at a density of 3 x 10^5 cells per well of a 24-well plate one day before transfection and inoculated with MVA-T7 (35) at an MOI of 5 for 1 h before transfection. Cells in two wells of a 24-well plate were transfected with expression plasmids for L protein domains (1 µg HA construct and 1 µg 3xFLAG construct) and lysed 24 h post transfection in 200 µl lysis buffer (50 mM Tris-HCl [pH 7.5], 200 mM NaCl, 1 mM EDTA, 0.5% Triton-X100, 40 µl Complete Protease Inhibitor Cocktail/ml [Roche]). Nuclei were pelleted by centrifugation, and 150 µl supernatant were mixed with 300 µl 2x binding buffer (100 mM Tris-HCl [pH 7.5], 400 mM NaCl, 2 mM EDTA, 0.5% Triton-X100, 2 mg bovine serum albumin [BSA]/ml), 150 µl water, and 1 µl of anti-HA antibody solution (H6908, 0.5 – 0.7 mg/ml, Sigma-Aldrich) and incubated at 4°C overnight with gentle agitation. Protein G or nProtein A Sepharose 4 Fast Flow solution (30 µl as supplied by GE Healthcare) were added and the mix was incubated for 1 h at 4°C with gentle agitation. Sepharose-coupled antibody-protein complexes were precipitated by centrifugation and washed four times with ice-cold 1x binding buffer and once with ice-cold 1x TNE buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 1 mM EDTA). Precipitated proteins were mixed with 4x NuPAGE LDS Sample Buffer (Invitrogen) and DTT, separated by 4 – 12% Bis-Tris PAGE, transferred to nitrocellulose membrane (Schleicher & Schuell), and detected by immunoblotting (see below).

**Immunoblot analysis.** Nitrocellulose membranes (Schleicher & Schuell) were stained with Fast Green FCF (Roth), blocked in 1x Roti-Block (Roth) overnight at room temperature, and incubated with anti-HA (1:10,000, H6908, Sigma-Aldrich) or peroxidase-conjugated anti-
FLAG M2 antibody (1:10,000, A8592, Sigma-Aldrich) in phosphate-buffered saline (PBS)–0.2x Roti-Block for 1 h at room temperature. The anti-HA blot was additionally incubated with peroxidase-conjugated secondary antibody (Dianova) for 1 h at room temperature. After washing, protein bands were visualized by chemiluminescence using SuperSignal West Pico or Femto substrate (Pierce) and X-ray film (Kodak).

**Confocal immunofluorescence microscopy.** BSR-T7/5 cells were seeded at a density of 2 x 10^4 cells per well of an 8-well Nunc Lab-Tek II Chamber Slide (Nalge Nunc). Cells in a well were transfected with 50 – 150 ng per L gene expression plasmid, optional complemented with 25 ng pCITE-NP and 100 ng MG. The amount of transfected DNA was kept constant at 300 ng by adding empty pCITE-2a vector. 12 – 24 h post transfection, cells were fixed with 4% formaldehyde–4% saccharose in PBS for 20 min at room temperature. Cells were permeabilized with 0.02% Triton-X100 in PBS for 10 min, washed with PBS, blocked with 10% BSA in PBS for 30 min at 37°C, and incubated with mouse monoclonal anti-FLAG M2 (1:6, F1804, Sigma-Aldrich) and rabbit anti-HA (1:250, H6908, Sigma-Aldrich) in 10% BSA–PBS for one hour at 37°C. After washing, cells were incubated with 4',6-diamidino-2-phenylindole (DAPI, 1:10,000, Boehringer Ingelheim), anti-mouse-IgG coupled to rhodamine (1:25, Dianova), and anti-rabbit-IgG coupled to FITC (1:250, SIFIN, Berlin) in 10% BSA–PBS for 30 min at 37°C. Cells were washed, air-dried, and covered with ProLong Gold antifade reagent (Invitrogen). Confocal images were captured using a FluoView 1000 microscope (Olympus) and processed with CorelDraw X4 (Corel).

**Bioinformatics.** Inter-domain linkers in L protein were predicted using programs DomCut (http://www.bork.embl-heidelberg.de/~suyama/domcut/) (36) and Armadillo (http://www.armadillo.blueprint.org/) (8); the latter includes the DLI (8), REI (11), and GHL (14) algorithms. DomCut scores were normalized with the factor 20 to adjust them to the score range of the other algorithms. The output scores from all four predictions were combined by addition. Linker predictions were performed for L protein of the following
viruses: Lassa AV, CSF, NL, and Z148; Mopeia; Morogoro; lymphocytic choriomeningitis Armstrong and WE; Junin; Guanarito; Machupo; Pichinde; Pirital; Tacaribe; and Sabia. L protein secondary structure was predicted with SSpro (http://scratch.proteomics.ics.uci.edu/) (5) and Jpred3 (http://www.compbio.dundee.ac.uk/www-jpred/) (6) servers. Protein sequences and predicted secondary structures were aligned with MacVector software (MacVector).

Results

Prediction of inter-domain linkers. L protein sequences of 15 arenaviruses, including Lassa strains AV, NL, Z148, and CSF, were subjected to inter-domain linker prediction. The domain linker propensity index DLI (8), the residue entropy index REI (11), and the residue index GHL (14) implemented in the Armadillo program (8), as well as the DomCut index (36) were used for prediction. The indices are based on the distinct amino acid composition of inter-domain linkers in comparison to functional domains (DLI, GHL, and DomCut) or on the number of degrees of freedom on the angles $\phi$, $\varphi$, and $\chi$ for each amino acid residue (REI). In addition, the secondary structures of the 15 L proteins were predicted using SSpro and Jpred3 servers (5, 6). The lack of prediction of ordered secondary structure elements and a low degree of sequence conservation were considered further criteria for inter-domain linkers.

The average prediction for Lassa virus based on strains AV, CSF, NL, and Z148 is shown in Fig. 1A. Similar models were predicted for other arenaviruses (data not shown). Because further experimental studies were conducted with L protein of Lassa virus strain AV, the prediction for this strain is shown separately (Fig. 1A, dotted line). Altogether, 14 inter-domain linkers with a domain propensity score smaller than −2 were predicted in Lassa virus L protein. They were located in regions without predicted secondary structure elements or with ambiguous prediction. Eight linkers (central residues S407, G467, G774, G939, L1546, N1722, S1952, and V2074 in strain AV) were located in regions of high sequence variability.
One peak (C84) was excluded from further analysis because it was too close to the N terminus of the protein in a conserved region. Two peaks (V1113 and D1459) were located within the putative RdRp in rather conserved regions that had previously been shown to be less tolerant to mutation (18). Position 1113 is located in a predicted loop that overlaps with motif pre-A of the RdRp and supposedly forms the NTP tunnel. It was included in the mutagenesis study as a control to verify the experimental strategy aiming at identifying inter-domain linkers rather than loops within functional domain.

**Mutational challenge of predicted inter-domain linkers.** Previous experiments with Lassa virus L protein have shown that the introduction of Gly-Gly residues into loops of functional domains is often deleterious (18). We assumed that a true inter-domain linker could be stretched without affecting the function of the upstream and downstream domains. Therefore, a flexible sequence of five amino acid residues (GDGSG) was inserted into the proposed linkers to challenge the prediction (Fig. 1B, Ins-5-aa construct). Mutagenesis was performed using a previously described PCR strategy to facilitate rapid screening (18). Functionality of the mutant L proteins was analyzed in the Lassa virus minireplicon system (17).

Insertion of the artificial 5-amino acid linker at positions S407, G467, G774, G939, S1952, and V2074 did not interfere with L protein activity (Tab. 1). As expected, insertion into an intra-domain loop (V1113) was not compatible with L protein function. Since the linker prediction for position N1722 was quite strong (Fig. 1A), and to exclude positional effects of the insertion, the sequence was inserted at two additional sites within the predicted linker (A1712 and Y1726). However, no activity was observed (Tab. 1). Because residues S407 and G467 are located only 60 residues apart in a highly variable sequence connecting conserved regions I and II, we wondered whether the whole interconnecting sequence might act like a linker. Therefore, the GDGSG sequence was also inserted between S407 and G467 at position G446. L protein was active (Tab. 1) suggesting the existence of a long linker sequence spanning residues 407 to 467.
To verify expression of inactive mutants, their intracellular level was evaluated in immunoblot. All insertion mutants were expressed, although mutants Y583, G738, and L1546 showed a lower steady-state level than wild-type L protein (Fig. 1C). However, it is unlikely that this drop rendered the system completely inactive [reducing the amount of transfected L protein construct from 250 ng to 50 ng still allows for 25% Ren-Luc activity with a signal-to-noise (S/N) ratio of 60 (18)].

As a next step in the analysis, the insertion was extended from 5 to 10 amino acid residues (GDGSGGKADG) (Fig. 1B, Ins-10-aa construct). This experiment was performed with the seven mutants who retained activity following GDGSG insertion. Extending the linker at positions S407, G467, G774, G939, and V2074 did not reduce replicon activity, while a drop was seen at positions G446 and S1952 (Tab. 1). Expression of the G446 mutant was verified in immunoblot (Fig. 1C). In conclusion, insertion mutagenesis suggests that L protein contains five potential inter-domain linkers: a long linker region spanning S407 to G467, and shorter linkers around G774, G939, S1952, and V2074.

**Splitting of L protein into N- and C-terminal domains.** If the linkers connect individually folded functional domains, it might be possible to physically separate the domains while retaining functionality of L protein. To this end, the L gene was split at each of the residues S407, G446, G467, G774, G939, and V2074 into an N- and C-terminal part (Fig. 1B, Split constructs). An artificial stop codon was introduced at end of the N-terminal domain, while the C-terminal domain was attached to an artificial start codon. Corresponding N- and C-terminal domains reconstituting a full-length L gene were expressed from separate T7 promoter constructs together with the other components of the replicon system.

In the initial experiment, N- and C-terminal domains were generated by PCR mutagenesis and transfected without cloning. L protein remained functional when split at positions S407 (residues 1 – 409 plus 405 – 2220; note that the expression cassettes were constructed with a 4-residue overlap at the split site), G467 (residues 1 – 469 plus 465 – 2220), and G939...
(residues 1–941 plus 937–2220), respectively (Table 1). Separation at positions G446, G774, S1952, and V2074 led to loss of replicon activity. Some of these domains showed a low steady-state level in immunoblot analysis, e.g. the C terminus of G446 split and the N terminus of S1952 and V2074 split (Fig. 1C), suggesting reduced stability. To verify the positive results obtained with L protein split at G467 and G939, the corresponding domains were cloned into vector pCITE-2a and then tested (the constructs were called L-N1 + L-C1 and L-N2 + L-C2, respectively). Activity of the cloned domains was 218% and 33% for split at G467 and G939, respectively (Table 1). Transfection of each domain alone did not result in Ren-Luc activity above background. Similarly, expression of the 1–407 N-terminal domain together with the 467–2220 C-terminal domain, i.e. leaving out the 60 amino acid sequence between both domains, did not confer L protein activity. Thus, this sequence, despite high variability and flexibility, is indispensable for L protein function or domain interaction.

Altogether, these results indicate that L protein can be split at G467 or G939 into N- and C-terminal domains that most likely interact with each other and reconstitute a functional L protein by trans-complementation.

**Co-immunoprecipitation studies with N- and C-terminal domains.** To demonstrate physical interaction between N- and C-terminal domains, co-immunoprecipitation studies were performed. Expression constructs for L protein domains derived from split at residues G467 (L-N1 and L-C1) and G939 (L-N2 and L-C2) were subcloned and complemented with 3xFLAG and HA tags. The constructs were transfected either alone, or co-transfected with the corresponding domain to reconstitute full-length L protein (Fig. 2). Protein complexes were generally precipitated via the HA tag. The precipitate was analyzed in immunoblot using anti-HA and anti-FLAG antibodies. In replicon system, L protein reconstituted from tagged domains was still active, although at lower level compared to the untagged versions (L-N1-HA plus L-C1-FLAG, 130% Ren-Luc activity and S/N ratio of 118; L-N1-FLAG plus L-C1-HA, 35% Ren-Luc activity and S/N ratio of 55; L-N2-HA plus L-C2-FLAG, 6.2% Ren-Luc
activity and S/N ratio of 3.7; L-N2-FLAG plus L-C2-HA, 8.9% Ren-Luc activity and S/N ratio of 4.6).

All domains were expressed, irrespective of the type of tag (see the lysate lanes in Fig. 2). In the first experiment, the HA tag was attached to the N-terminal domains (L-N1-HA and L-N2-HA), while the FLAG tag was attached to the C-terminal domains (L-C1-FLAG and L-C2-FLAG). Upon co-expression, L-C1-FLAG co-precipitated with L-N1-HA (Fig. 2A, lane 4) and L-C2-FLAG co-precipitated with L-N2-HA (Fig. 2C, lane 4). The FLAG-tagged proteins were precipitated nearly quantitatively. Neither L-C1-FLAG nor L-C2-FLAG were precipitated in the absence of the corresponding HA-tagged domain (Fig. 2A and C, lane 2).

In the reciprocal experiment, the HA tag was attached to the C-terminal domains (L-C1-HA and L-C2-HA), while the FLAG tag was attached to the N-terminal domains (L-N1-FLAG and L-N2-FLAG). Consistent with the first experiment, L-N1-FLAG co-precipitated with L-C1-HA (Fig. 2B, lane 4) and L-N2-FLAG co-precipitated with L-C2-HA (Fig. 2D, lane 4). No background precipitation of FLAG-tagged protein was observed in the absence of the HA-tagged domains (Fig. 2B and D, lane 3).

These experiments provide evidence for physical interaction between N- and C-terminal domains, irrespective of whether L protein was split at G467 or G939. Neither NP nor MG are required as co-factors.

**Co-localization studies with N- and C-terminal domains.** To further substantiate the hypothesis of interaction between N and C terminus and to map the intracellular distribution of domains resulting from split at G467 (L-N1 and L-C1) and G939 (L-N2 and L-C2), co-localization studies were performed. The HA- and FLAG-tagged domains were detected using anti-HA and anti-FLAG, and FITC and rhodamine-conjugated secondary antibodies, respectively. Images were captured by confocal microscopy.

The intact full-length L protein showed a cytoplasmic distribution with some dot-like structures (Fig. 3A) irrespective of co-expression of MG or NP, which showed a granular
cytoplasmic distribution (Fig. 3B). In the first set of experiments, the HA tag was attached to the N-terminal domains (L-N1-HA and L-N2-HA), while the FLAG tag was attached to the C-terminal domains (L-C1-FLAG and L-C2-FLAG). Individually expressed, the N-terminal domains showed a dot-like distribution similar to full-length L protein (Fig. 3C and E, upper left image), while the C-terminal domains showed a homogenous cytoplasmic staining (Fig. 3C and E, upper right image). However, upon co-expression, both N- and C-terminal domains showed a dot-like distribution with nearly perfect co-localization (Fig. 3C and E, lower images). In the reciprocal experiment the tags were swapped; the HA tag was attached to the C-terminal domains (L-C1-HA and L-C2-HA), while the FLAG tag was attached to the N-terminal domains (L-N1-FLAG and L-N2-FLAG). In agreement with the first experiment, N- and C-terminal domains showed dot-like and homogenous distribution, respectively (Fig. 3D and F, upper images). Upon co-expression, both domains co-localized in dot-like structures (Fig. 3D and F, lower images).

In conclusion, L protein is distributed in cytoplasm and accumulates in dot-like structures. This distribution pattern is reflected by the N-terminal domains, irrespective of the split site, while the C-terminal domains are homogenously distributed. Co-expression of N- and C-terminal domains leads to re-distribution of the latter into the dot-like structures providing further evidence for interaction between both domains.

**Studies on L–L protein interactions.** It has been previously shown that L proteins can oligomerize (28). Having delineated domains within L protein, we wondered whether these domains play a role in L–L protein interaction. In a first set of experiments, the interaction between homologous domains was studied by co-immunoprecipitation. HA- and FLAG-tagged versions of each domain (L-N1, L-C1, L-N2, and L-C2) were co-expressed. Proteins were precipitated via the HA tag. All FLAG-tagged domains co-precipitated with their homologous HA-tagged version, indicating interaction between L-N1 and L-N1; L-N2 and L-N2; L-C1 and L-C1; and L-C2 and L-C2 (Fig. 4A). These experiments were complemented
by co-immunoprecipitation of HA-tagged full-length L protein with all FLAG-tagged domains. In agreement with the above experiments, L-N1, L-C1, L-N2, and L-C2 coprecipitated with full-length L protein (Fig. 4B). These data suggest that the individual domains can dimerize or oligomerize and that L protein contains at least two sites — one in the N terminus and one in the C terminus — mediating L–L protein interaction.

To further support the hypothesis of L–L protein interaction via N or C terminus, it was tested if individual domains can exert a dominant negative effect on wild-type L protein in replicon assay, as has been described for L protein mutants with mutations in the catalytic site of RdRp (18, 28). The dominant negative mutant L-D1334N served as a control; it reduced activity of wild-type L protein by >90% (Fig. 5A, lane 4). Co-expression of L-C1, L-C2, L-N1, or L-N2 with wild-type L protein did not significantly reduce replicon activity, irrespective of whether the D1334N mutation had been introduced into L-C1 or L-C2 (Fig. 5A). However, if the L-D1334N mutant was co-expressed with L protein split at G467 (L-N1 and L-C1) or G939 (L-N2 and L-C2), it exerted a dominant negative effect with reduction of replicon activity by about 87% (Fig. 5B). Consistent with the physical interaction studies, these functional data suggest that L protein reconstituted from N and C terminal domains is involved in L–L protein interactions.

Discussion

This study provides experimental evidence for the domain structure of Lassa virus L protein. Linkers of 5 or 10 amino acid residues were tolerated at seven sites (S407, G446, G467, G774, G939, S1952, and V2074) without loss of function, suggesting these sites represent linkers between domains or subdomains rather than intra-domain loops. At two of these sites, G467 and G939, the protein could be split into an N-terminal and a C-terminal domain, which were capable of reconstituting a functional L protein. Co-immunoprecipitation and co-
immunofluorescence experiments demonstrated physical interaction between the N- and C-terminal domains, irrespective of whether L protein was split at G467 or G939.

The 200-kDa L protein of arenaviruses has long been speculated to be composed of multiple domains. First evidence for a domain structure was provided by sequence comparison demonstrating the existence of four conserved regions interconnected by stretches of high sequence divergence (37). Our experimental data are largely consistent with this concept.

Insertion of 10 amino acids between conserved regions I and II (S407, G446, and G467), between regions II and III (G939), and between region IV and the C terminus (S1952 and V2074) was compatible with L protein function. However, none of the mutants with insertions between regions III and IV (L1546, A1712, N1722, and Y1726) was active suggesting that this part of the protein is structurally less flexible. Similar studies were performed with viruses of the Morbillivirus genus (rinderpest virus, measles virus, and canine distemper virus) (3, 9, 30), the Rhabdoviridae (vesicular stomatitis virus) (27), and the Orthobunyavirus genus (Bunyamwera virus) (29) supporting the view that the large L proteins of negative-strand RNA viruses function as multi-domain proteins. In these studies, short epitopes or green fluorescent protein (GFP) were inserted in putative linker or hinge regions and recombinant virus expressing epitope or GFP-tagged L protein was rescued. The recombinants were often attenuated and employed as a vaccine candidate or as a tool for studying L protein biology (29, 30). The present study provides a basis for generation of recombinant arenavirus expressing L protein internally tagged with epitopes or GFP, although it has to be taken into account that the findings observed in the context of the minireplicon assay may not easily be extended to the situation associated with an arenavirus infection. The identified inter-domain linker regions may still provide some functions required for steps in the virus life cycle that are not covered by the replicon system.

Definitive proof for the existence of two inter-domain linkers was obtained by splitting the L protein at G467 between conserved regions I and II, and at G939 between conserved regions
II and III. Expression of L protein as two polypeptides, which interact with each other and form a functional complex, closely resembles the situation with influenza virus. The functional homologues of L protein in influenza virus are three proteins — PB1, PB2, and PA — which assemble into a heterotrimeric polymerase complex (1). Given the evolutionary relatedness among the segmented negative-strand RNA viruses, it is conceivable that the L proteins of bunya- and arenaviruses are the result of joining functional subunits, which originally were expressed from separate genome segments. The fact that Lassa virus L protein can be split into subunits may be related to this hypothetical ancient feature.

The presence of two inter-domain linkers, at positions G467 and G939, suggests that L protein is composed of at least three structural domains: (i) from positions 1 – 467, (ii) from positions 467 – 939 and (iii) from positions 939 – 2220 (Fig. 6A). However, co-expression of all three domains from separate constructs did not result in measurable replicon activity (data not shown). Apparently, the central domain (residues 467 – 939) has to be covalently linked either to the N-terminal (residues 1 – 467) or to the C-terminal domain (residues 939 – 2220) to maintain activity. Putative enzymatic functions, which have been identified in L protein so far, can be allocated to two of the domains. A putative endonuclease has been mapped to the first 200 amino acids at the N terminus of L protein (21) and, hence, resides in the first domain. The putative RdRp has been mapped to positions 1040 to 1540 (37), which exactly corresponds to the N-terminal part of the third domain (Fig. 6A). Currently, there is no clue as to the functions residing in the second domain (residues 467 – 939) and the C-terminal 700 residues of the third domain. The two interaction sites with Z protein (38) map to first and third domain, overlapping with the putative endonuclease and the RdRp, respectively.

L proteins of various negative-strand RNA viruses, including paramyxoviruses (32, 33), LCMV (28), and Rift valley fever virus (40) as well as the heterotrimeric polymerase complex of influenza virus (19) form oligomers and this feature is assumed to be important for function. The interactions between (and within) L proteins of arenaviruses have not yet been...
resolved in detail. Our data suggest three different types of interactions: (i) interaction between N and C terminus of L protein (N–C), (ii) intermolecular L–L interaction through N terminus (N–N), and (iii) intermolecular L–L interaction through C terminus (C–C). The N–C interaction appears to be quite strong as it led to nearly quantitative co-immunoprecipitation of the interaction partner and induced considerable re-distribution of the C terminus in co-localization experiments. It offers possibilities for both intra- and intermolecular associations. First, the N–C interaction may stabilize the association of N and C terminus of the same L protein molecule (Fig. 6B), second, it may mediate dimerization of L protein molecules (a and b) through Na–Cb + Ca–Nb interaction (Fig. 6C), and third, it may lead to multimerization of L proteins (a, b, c, d ...) through Ca–Nb + Cb–Nc + Cc–Nd etc. interaction (Fig. 6D). Assuming that the N–C, N–N, and C–C interactions act via different interfaces and that one L molecule may interact with others through all three interfaces at the same time, L proteins could theoretically assemble into higher-order structures as has been reported for poliovirus (24). Formation of oligo- or multimeric structures at the sites of genome replication or transcription would also be consistent with the dominant negative effect exerted by arenavirus L proteins with a mutation in the active site of RdRp (18, 28). The observation that the L-D1334N mutant also exerts a dominant negative effect on the split L proteins suggests formation of complexes between full-length L protein, N terminal domain, and C terminal domain and thus supports the hypothesis of multiple interaction interfaces.

This study also provides insights into the cellular localization of L protein. While the full-length L protein and both L-N1 and L-N2 show a dot-like, sometimes perinuclear, cytoplasmic staining, both C-terminal fragments of the protein were homogenously distributed in the cytoplasm. This suggests that the accumulation of L protein in dot-like structures is mediated by the N-terminal domain (residues 1 – 467). Interestingly, the L protein of the phylogenetically related Bunyamwera virus was also found in punctate cytoplasmic structures in association with intracellular membranes (29). In which subcellular
compartment(s) Lassa virus L protein accumulates and which cellular proteins are involved, is a matter of future research.

Acknowledgments

This work was supported by a grant from the “Vereinigung der Freunde des Tropeninstituts Hamburg e.V.” to M. L., VIZIER integrated project grant LSHG-CT-2004-511960 of the European Union 6th Framework, and grant GU 883/1-1 from the German Research Foundation. The Bernhard-Nocht-Institute is supported by the Bundesministerium für Gesundheit and the Freie und Hansestadt Hamburg.
References


Table 1. Activity of L protein mutants in replicon system.

<table>
<thead>
<tr>
<th>Position of insertion or split</th>
<th>Insertion of 5 amino acid residues(^a)</th>
<th>Insertion of 10 amino acid residues(^a)</th>
<th>Split in N- and C-terminal domains(^a)</th>
<th>Split in N- and C-terminal domains (plasmids)(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% of wild-type(^c)</td>
<td>Signal-to-noise ratio(^d)</td>
<td>n exp</td>
<td>% of wild-type(^c)</td>
</tr>
<tr>
<td>S407</td>
<td>79.8</td>
<td>130.0</td>
<td>6</td>
<td>105.8</td>
</tr>
<tr>
<td>G446</td>
<td>41.8</td>
<td>144.0</td>
<td>5</td>
<td>10.7</td>
</tr>
<tr>
<td>G467</td>
<td>111.9</td>
<td>222.3</td>
<td>6</td>
<td>93.1</td>
</tr>
<tr>
<td>Y583</td>
<td>0.3</td>
<td>1.0</td>
<td>13</td>
<td>–</td>
</tr>
<tr>
<td>G738</td>
<td>3.4</td>
<td>6.8</td>
<td>13</td>
<td>–</td>
</tr>
<tr>
<td>G774</td>
<td>170.2</td>
<td>325.8</td>
<td>6</td>
<td>148.0</td>
</tr>
<tr>
<td>G939</td>
<td>80.4</td>
<td>148.8</td>
<td>7</td>
<td>138.2</td>
</tr>
<tr>
<td>V1113</td>
<td>0.4</td>
<td>0.9</td>
<td>12</td>
<td>–</td>
</tr>
<tr>
<td>L1546</td>
<td>2.0</td>
<td>7.8</td>
<td>12</td>
<td>–</td>
</tr>
<tr>
<td>A1712</td>
<td>1.4</td>
<td>1.0</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>N1722</td>
<td>1.0</td>
<td>1.9</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>Y1726</td>
<td>0.5</td>
<td>0.3</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>G1865</td>
<td>0.5</td>
<td>0.9</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>S1952</td>
<td>47.5</td>
<td>208.1</td>
<td>11</td>
<td>22.8</td>
</tr>
<tr>
<td>V2074</td>
<td>23.7</td>
<td>80.5</td>
<td>12</td>
<td>16.7</td>
</tr>
</tbody>
</table>

\(^a\) Expression cassettes for mutants or domains were generated by PCR and tested without cloning.

\(^b\) Domains were cloned into vector pCITE-2a before testing.

\(^c\) Renilla luciferase activity (sRLU) of mutant relative to wild-type activity (100%). Mean of n experiments as indicated in the table.

\(^d\) sRLU of mutant divided by sRLU of negative control (L protein containing D1334N mutation in the catalytic site of the RdRp). Mean of n experiments as indicated in the table.
Fig. 1: Bioinformatics and mutational analysis of inter-domain linkers in Lassa virus L protein. (A) Prediction of inter-domain linkers in Lassa Virus L protein by combination of DLI (8), REI (11), GHL (14), and DomCut (36) algorithms. The solid line represents the average prediction for Lassa virus based on strains NL, CSF, Z148, and AV. The dotted line shows the prediction for Lassa AV only. Negative scores indicate higher propensity for linker sequences. Conserved regions I to IV are shown above the graph; the RdRp is boxed (37). Sites selected for mutational analysis in replicon system are indicated by arrows. Linker predictions were challenged by insertion of flexible sequences of 5 or 10 amino acid residues (Ins-5-aa and Ins-10-aa) and by splitting the full-length L protein into N- and C-terminal domains (Split). Results from Table 1 are summarized as follows: ++, mutants with wild-type like activity; +, mutants with reduced activity; and −, defective mutants. n.d., not done. (B) Expression constructs for mutational challenge of predicted inter-domain linkers. All constructs have a T7 promoter (T7p) and an internal ribosomal entry site (IRES) to facilitate cap-independent translation. Inserted sequences are shown in grey boxes. Amino acid residues at the site of insertion (X) or split (X_{n-1}|X_n|X_{n+1}) were duplicated and the insertion or the split was placed between the duplicated residues. (C) Immunoblot analysis of HA-tagged L protein mutants inactive in replicon assay. BSR-T7/5 cells were infected with MVA-T7 and transfected with PCR products expressing L protein mutants containing a C-terminal HA tag. L protein in cytoplasmic lysate was separated by SDS-PAGE, blotted, and detected with anti-HA antibody. The size of insertion is indicated (Ins-5-aa or Ins-10-aa). For split mutants, either N terminus (Split-L-N) or C terminus (Split-L-C) was transfected; expression products are marked with arrows. The D1334N mutant contains a mutation in the catalytic site of the RdRp and served as negative control in the replicon experiments. Control cells were not infected and transfected (Neg. ctrl.) or infected with MVA-T7 but not transfected (MVA).
Fig. 2: Co-immunoprecipitation analysis of interaction between N- and C-terminal domains of L protein. BHK-21 cells were transfected with expression plasmids for HA- and FLAG-tagged domains shown on top of each panel. Cytoplasmic lysate was incubated with anti-HA antibody to precipitate protein complexes. Proteins in lysate (L) and immunoprecipitate (IP) were separated by SDS-PAGE, blotted, and detected with anti-FLAG antibody (upper blots in each panel) and anti-HA antibody (lower blots). (A) Interaction between L-N1-HA and L-C1-FLAG. (B) Interaction between L-N1-FLAG and L-C1-HA. (C) Interaction between L-N2-HA and L-C2-FLAG. (D) Interaction between L-N2-FLAG and L-C2-HA. The virtual size discrepancy between proteins in lysate and IP fraction is probably due to the different salt concentrations in both preparations.

Fig. 3: Confocal immunofluorescence analysis of intracellular localization of N- and C-terminal domains of L protein. BSR-T7/5 cells were transfected with expression plasmids for HA- and FLAG-tagged domains, or HA-tagged full-length L protein. N- and C-terminal domains were expressed separately (upper images in each panel) or co-expressed (lower images). Cells were fixed and proteins were visualized using anti-FLAG and anti-HA antibodies. (A) Analysis of full-length L protein. (B) Analysis of full-length L protein co-expressed with NP and minigenome (MG). (C) Analysis of L-N1-HA and L-C1-FLAG. (D) Analysis of L-N1-FLAG and L-C1-HA. (E) Analysis of L-N2-HA and L-C2-FLAG. (F) Analysis of L-N2-FLAG and L-C2-HA.

Fig. 4: Analysis of interaction between homologous domains and between full-length L protein and domains by co-immunoprecipitation. (A) BHK-21 cells were transfected with two expression plasmids for the same domain, one HA-tagged and one FLAG-tagged as shown on top. Cytoplasmic lysate was incubated with anti-HA antibody to precipitate protein complexes. Proteins in lysate and immunoprecipitate (IP) were separated by SDS-PAGE,
blotted, and detected with anti-FLAG antibody (upper blots) and anti-HA antibody (lower blots). The anti-FLAG lysate blot (top left) was exposed for a shorter time than the anti-FLAG IP blot (top right). Signals of the IgG heavy chain of the anti-HA antibody used for precipitation are marked with an asterisk. (B) BHK-21 cells were co-transfected with expression plasmids for HA-tagged full-length L protein and FLAG-tagged domains. IP was performed with anti-HA antibody, and lysate and precipitate were analyzed in immunoblot as described for panel A.

Fig. 5: Analysis of interaction between wild-type L protein, L protein domains, and dominant negative L protein mutants in replicon system. (A) Influence of individual L protein domains on activity of wild-type L protein. Expression constructs for N or C terminal domains (250 ng) were co-transfected at a ratio of 1:1 with the expression construct for wild-type L protein (250 ng) and the other replicon components (250 ng MG and 250 ng pCITE-NP). As controls, 250 ng of a translation-defective L gene construct lacking the internal ribosome entry site (L-del-IRES) or 250 ng construct for expression of a dominant negative L protein mutant with a mutation in the catalytic site of the RdRp (L-D1334N) were co-transfected with the replicon components. The expression construct for wild-type L protein was replaced by empty pCITE-2a vector in lanes 1 and 2. The D1334N mutation was also introduced into L-C1 and L-C2 (L-C1-D1334N and L-C2-D1334N). (B) Influence of dominant negative mutant L-D1334N on activity of L protein reconstituted from N and C terminal domains. 250 ng of L-del-IRES or 250 ng of L-D1334N were co-transfected with 250 ng MG, 250 ng pCITE-NP and, alternatively, (i) 250 ng wild-type L protein construct and 250 ng empty pCITE-2a, (ii) 250 ng L-N1 and 250 ng L-C1 constructs, or (iii) 250 ng L-N2 and 250 ng L-C2 constructs. Replicon activity obtained after co-transfection of L-del-IRES with L protein constructs (wild-type or reconstituted from domains) was set at 100%. Mean and SD of duplicate experiments or a representative experiment (*) are shown.
Fig. 6: Summary of experimental data and hypothetical interactions between N- and C-terminal domains. (A) The existence of two inter-domain linkers at positions 467 and 939 suggests that L protein is composed of at least three domains (black, grey, and white). The putative endonuclease (Endo) and the RdRp reside in the first and the third domain, respectively. (B – D) Possible modes of interaction due to presumed N–C binding interfaces in first and third domain (white circles with lines representing non-covalent bonds). (B) Intramolecular interaction. (C) Dimerization. (D) Multimerization.