

# The N-Terminal Domain of the Arenavirus L Protein Is an RNA Endonuclease Essential in mRNA Transcription

Benjamin Morin<sup>1</sup>, Bruno Coutard<sup>1</sup>, Michaela Lelke<sup>2</sup>, François Ferron<sup>1</sup>, Romy Kerber<sup>2</sup>, Saïd Jamal<sup>1</sup>, Antoine Frangeul<sup>1</sup>, Cécile Baronti<sup>3</sup>, Rémi Charrel<sup>3</sup>, Xavier de Lamballerie<sup>3</sup>, Clemens Vonrhein<sup>4</sup>, Julien Lescar<sup>1,5</sup>, Gérard Bricogne<sup>4</sup>, Stephan Günther<sup>2</sup>, Bruno Canard<sup>1\*</sup>

**1** Architecture et Fonction des Macromolécules Biologiques, CNRS and Universités d'Aix-Marseille I et II, UMR 6098, ESIL Case 925, Marseille, France, **2** Department of Virology, Bernhard-Nocht-Institute for Tropical Medicine, Hamburg, Germany, **3** Unité des Virus Emergents UMR190, Université de la Méditerranée & Institut de Recherche pour le Développement, Marseille, France, **4** Global Phasing Ltd., Cambridge, United Kingdom, **5** School of Biological Sciences, Nanyang Technological University, Singapore, Singapore

## Abstract

*Arenaviridae* synthesize viral mRNAs using short capped primers presumably acquired from cellular transcripts by a 'cap-snatching' mechanism. Here, we report the crystal structure and functional characterization of the N-terminal 196 residues (NL1) of the L protein from the prototypic arenavirus: lymphocytic choriomeningitis virus. The NL1 domain is able to bind and cleave RNA. The 2.13 Å resolution crystal structure of NL1 reveals a type II endonuclease  $\alpha/\beta$  architecture similar to the N-terminal end of the influenza virus PA protein. Superimposition of both structures, mutagenesis and reverse genetics studies reveal a unique spatial arrangement of key active site residues related to the PD...(D/E)XK type II endonuclease signature sequence. We show that this endonuclease domain is conserved and active across the virus families *Arenaviridae*, *Bunyaviridae* and *Orthomyxoviridae* and propose that the arenavirus NL1 domain is the *Arenaviridae* cap-snatching endonuclease.

**Citation:** Morin B, Coutard B, Lelke M, Ferron F, Kerber R, et al. (2010) The N-Terminal Domain of the Arenavirus L Protein Is an RNA Endonuclease Essential in mRNA Transcription. *PLoS Pathog* 6(9): e1001038. doi:10.1371/journal.ppat.1001038

**Editor:** Michael J. Buchmeier, University of California Irvine, United States of America

**Received:** April 7, 2010; **Accepted:** July 12, 2010; **Published:** September 16, 2010

**Copyright:** © 2010 Morin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Toscana virus and H5N1 influenza endonuclease domains were produced as part of the European Virus Archive (EVA) project (European FP7 Capacities Project no 228292, <http://www.european-virus-archive.com/>). This work was supported in part by the VIZIER integrated project (LSHG-CT-2004-511960) of the European Union 6th Framework Programme (FP6), the Direction Générale de l'Armement (DGA 09cà402), the Fondation pour la Recherche Médicale (FRM-Aide aux équipes), the German Research Foundation grant GU 883/1-1, and a grant from the "Vereinigung der Freunde des Tropeninstituts Hamburg e.V." to ML. JL was supported by grant 08/1/22/19/589 from BMRC and an ATIP from CNRS. We gratefully acknowledge the ID14-4 beamline staff and equipment at the European Synchrotron Radiation facility, Grenoble, France. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [bruno.canard@afmb.univ-mrs.fr](mailto:bruno.canard@afmb.univ-mrs.fr)

## Introduction

The *Arenaviridae* family includes 22 viral species into a single genus Arenavirus, with new species awaiting classification [1,2]. They cause chronic and asymptomatic infections in rodents, and occasional transmission to man may result in life-threatening meningitis and/or hemorrhagic fever. *Lymphocytic choriomeningitis virus* (LCMV) is the prototypic species and first arenavirus isolated in 1933. Because its natural host is the common house mouse (*Mus musculus*), LCMV is the only known arenavirus presumably exhibiting a worldwide distribution. LCMV is a human pathogen of significant clinical relevance, causing central nervous system disease, congenital malformation, choriomeningitis, and systemic and highly fatal infection in immuno-compromised, organ transplant recipient patients [3,4,5,6]. Humans are generally infected through the respiratory tract after exposure to aerosols, or by direct contact with infectious material.

Arenaviruses are enveloped viruses with a bisegmented negative single-strand RNA genome. Each RNA segment, called large (L; ~7.2 kb) and short (S; ~3.5 kb), contains two open reading frames in mutually opposite orientations and use an ambisense coding strategy to direct the synthesis of two polypeptides [7]. Between the

two open reading frames of each segment resides a non-coding intergenic region (IGR), composed of a sequence predicted to form a stable hairpin structure [8]. The S RNA encodes the viral nucleoprotein (NP; ~63 kDa) and glycoprotein precursor (GPC; ~75 kDa), whereas the L RNA encodes a small RING finger protein (Z; ~11 kDa) and a large protein (L; ~250 kDa) which is the viral RNA-dependent RNA polymerase (RdRp). The two RNA genomes are encapsidated by the NP, which is the most abundant protein in virions and infected cells, and act as templates for two fundamentally different processes, RNA replication and transcription. During RNA replication, the L protein first binds to the 3'-end of RNA templates and reads them from end to end to direct the synthesis of encapsidated full-length anti-genomes. During transcription, the RdRp stops RNA synthesis at a pause site located near the IGR [7]. The newly synthesized mRNA molecules have a non-polyadenylated 3'-end with a heterogeneous sequence mapped within the predicted hairpin in the IGR [9]. Furthermore, non template-directed sequences have been identified at the 5'-end of the subgenomic mRNA [10]. These sequences are variable in length [9,10,11] and terminate with a 5'-cap structure, which suggests the presence of a cap-snatching mechanism for arenaviruses. In this process, originally described for influenza viruses [12,13] and

## Author Summary

The *Arenaviridae* virus family includes several life-threatening human pathogens that cause meningitis or hemorrhagic fever. These RNA viruses replicate and transcribe their genome using an RNA synthesis machinery for which no structural data currently exist. They synthesize viral mRNAs using short capped primers presumably acquired from cellular transcripts by a ‘cap-snatching’ mechanism thought to involve the large L protein, which carries RNA-dependent RNA polymerase signature sequences. Here, we report the crystal structure and functional characterization of an isolated N-terminal domain of the L protein (NL1) from the prototypic arenavirus: lymphocytic choriomeningitis virus. The NL1 domain is able to bind and cleave RNA. The 2.13 Å resolution crystal structure of NL1 reveals a type II endonuclease  $\alpha/\beta$  architecture similar to the N-terminal end of the influenza virus PA protein. Superimposition of both structures and mutagenesis studies reveal a unique spatial arrangement of key active site residues related to the PD... (D/E)XK type II endonuclease signature sequence. Reverse genetic studies show that mutation of active site residues selectively abolish transcription, not replication. We show that this endonuclease domain is conserved and active across the virus families: *Arenaviridae*, *Bunyaviridae* and *Orthomyxoviridae* and propose that the arenavirus NL1 domain is the *Arenaviridae* cap-snatching endonuclease.

bunyaviruses [14], the viral RdRp binds cellular mRNAs caps and ‘steals’ them using an endonuclease activity, located in the influenza PA subunit [15,16], and presumably in L protein of bunyaviruses. These short capped RNAs are then used as primers for mRNA synthesis. The arenavirus L protein is an essential element in genome replication and transcription [17]. It is the largest viral protein composed of approximately 2200 amino-acid (aa) residues, and sequence analysis using homologous proteins led to the prediction of several conserved domains [18,19]. A biological function can be inferred for the L3 domain containing conserved and typical RdRp signature sequence motifs [19,20]. For Tacaribe virus, both domains L1 and L3 interact with the Z protein [21]. By analogy with influenza and bunyaviruses, the L protein may also carry activities and domains responsible for a cap-snatching mechanism that would account for the sequence diversity found at the 5′-end of RNA transcripts. The expression and purification of such a large viral polymerase is problematic and has not been documented.

We report here the first crystal structure of an *Arenaviridae* L protein domain at 2.13 Å resolution, that of the N-terminus domain of the LCMV L protein. We show that this domain is able to bind nucleotides, with a preference for UTP, and RNA. Structural comparison with the N-terminal part of the influenza virus PA protein characterizes unambiguously the domain as an endonuclease. Sequence and secondary structure analysis of L proteins from various *Bunyaviridae* family members predict that their N-terminal end carries a similar endonuclease activity, that we demonstrate for Toscana virus (TOSV) (genus *Phlebovirus*, family *Bunyaviridae*). Activity assays and mutagenesis show that the arenavirus endonuclease exhibits sequence-specificity with a preference for uracil-containing substrates. Lastly, reverse genetics studies correlate expression of endonuclease activity with the selective production of mRNA, making the N-terminus domain of the L protein a likely candidate to be involved in the cap-snatching mechanism of arenaviruses.

## Results

### Delineation of an Arenavirus L Protein Domain and its Crystal Structure

Based on aa sequence conservation across arenaviruses and on the presence of a potential nucleotide-binding site, we designed cDNA constructs encoding aa residues 1 to ~250 for the N-terminal end of four arenavirus (Pirital virus (PIRV), Lassa fever virus (LASV), Parana virus (PARV), and LCMV) L proteins. All four domains were expressed as soluble recombinant proteins. We observed a self-limited proteolysis of the Parana arenavirus N-terminus L domain which prompted us to refine boundaries into a shorter 196 aa form, hereafter named “NL1”, fully included in the previously predicted arenavirus L1 domain (1–250 aa) [19]. The construct was expressed in *E.coli* and purified, but yielded crystals diffracting to 8 Å. However, the homologous 196 residues domain of LCMV yielded well-diffracting crystals. The atomic structure of NL1 was first determined by the SAD technique with seleno-methionylated crystals that diffracted to 3.4 Å. The structure was refined using a native data set at 2.13 Å resolution (Table 1). Two NL1 molecules are present within the asymmetric unit. Residues 1–191 are visible for one molecule whilst only 1–175 could be modelled for the other NL1 molecule owing to high mobility of the C-terminal end of helix  $\alpha 7$ .

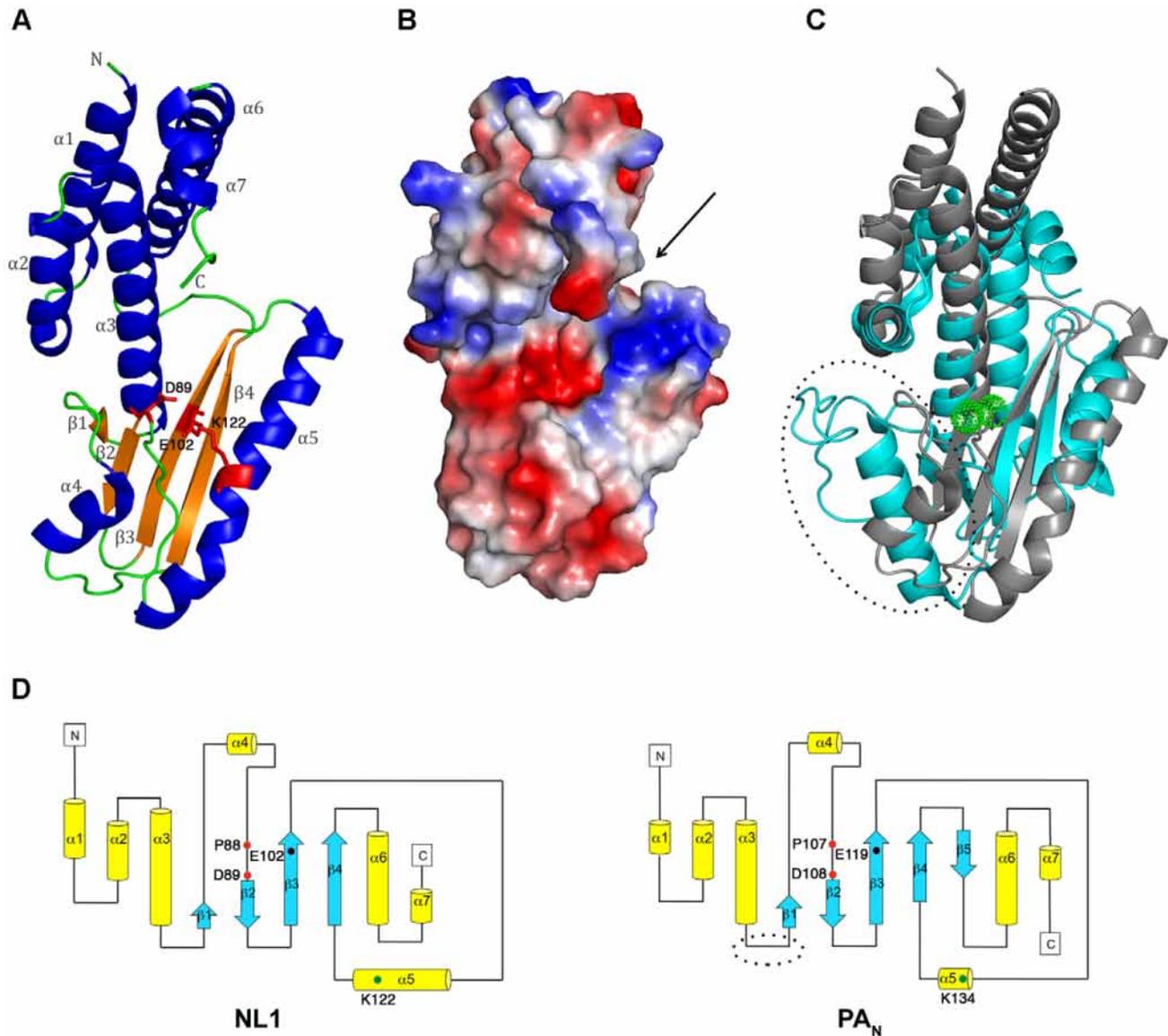
### The LCMV NL1 Domain Exhibits a Type II Endonuclease Fold

The LCMV NL1 monomer structure has approximate dimensions of 59 Å × 37 Å × 27 Å. It features four mixed  $\beta$ -strands forming a twisted plane surrounded by seven  $\alpha$ -helices (Figure 1A). The two anti-parallel strands  $\beta 1$  and  $\beta 2$  are connected by helix  $\alpha 4$ , whereas the two parallel strands  $\beta 3$  and  $\beta 4$  are

**Table 1.** Data Collection and Refinement Statistics.

Data collection and refinement statistics	LCMV NL1
Space group	C222 <sub>1</sub>
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	145.0, 159.3, 52.6
Resolution (Å)	107.25–2.13 (2.19–2.13)*
Measured/unique reflections	231,821/33,999
<i>R</i> <sub>merge</sub>	3.9 (45.2)
<  <i>I</i> / $\sigma$  >	27.1 (1.9)
Completeness (%)	97.9 (97.8)
Redundancy	6.8 (2.9)
Resolution (Å)	72.50–2.13
No. reflections	33,931
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	19.86/22.21
No. atoms	3,297
Protein	3,011
Water	253
<i>B</i> -factors	
Protein	57.6
Water	68.2
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.14

\*Values in parentheses are for the highest-resolution shell.  
doi:10.1371/journal.ppat.1001038.t001



**Figure 1. NL1 structure and comparisons with the influenza  $PA_N$  structure.** **A**, Cartoon-representation of the NL1 structure. Secondary-structure elements are labelled and colored as follows:  $\alpha$ -helices, blue,  $\beta$ -strands, orange, and loops, green. Side-chains for residues within the NL1 endonuclease active site are shown as red sticks and labelled. **B**, Electrostatic surface representation showing NL1 in the same orientation as in panel **A**. The arrow indicates the putative RNA binding groove and the active site crevice. Negative charges are in red and positive charges in blue and neutral in white. **C**, Superimposition (view in the same orientation as in **A**) of the structures of NL1 (grey) and  $PA_N$  (PDB code:2W69, cyan) highlighting their shared structural core as well as variations in the form of an extra loop only present in the  $PA_N$  structure (circled). The two  $Mn^{2+}$  ions in the  $PA_N$  structure active site are depicted as green spheres. **D**, Topology diagrams of the NL1 (left) and  $PA_N$  (right) structures.  $\alpha$ -helices are represented as yellow tubes and  $\beta$ -strands are blue arrows. The extra-loop of  $PA_N$  protein is circled as in panel **C**. Key residues from the endonuclease active site (PD, E/D, and K), are schematically depicted by colored dots and labelled, highlighting the fact that they project from conserved structural elements between the influenza  $PA_N$  protein and the arenavirus NL1 domain.  
doi:10.1371/journal.ppat.1001038.g001

connected by the long helix  $\alpha 5$ . These two helices run parallel to the central  $\beta$ -sheet and are disposed at the same side of the latter. On the opposite side of the  $\beta$ -sheet, helix  $\alpha 3$  is surrounded at its extremity by two N-terminal ( $\alpha 1$  and  $\alpha 2$ ) and C-terminal helices ( $\alpha 6$  and  $\alpha 7$ ). A search for similar protein folds using the DALI server [22] returned the PA N-terminal domain structure that was recently identified as a type II endonuclease domain [15,16]. The structural match with published molecular structures of the influenza PA N-terminal domains ( $PA_N$ ) returns a Z-score of 5.7 and an r.s.m.d. of 3.9 Å for 121 superposed aa (PDB code 3EBJ) and Z-score 5.2, r.s.m.d. 4 Å for 122 aa (PDB code 2W69). As was

the case for  $PA_N$ , other type II endonuclease proteins are also recovered: the Tt1808 hypothetical protein from *Thermus Thermophilus* HB88 (PDB code 1WDJ, Z-score 3.8, r.s.m.d. 3.4 Å for 81 aa), and the restriction endonuclease *SdaI* (PDB code 2IXS, Z-score 3.6, r.s.m.d. 6.3 Å for 104 aa).

The  $\beta$ -sheet forms a negatively charged cavity creating a binding site for divalent cations, whilst above that cavity, the C-terminal end of helix  $\alpha 5$  forms a positively charged patch and a concave surface that is likely to accommodate the RNA substrate (Figure 1B, arrow). The PA protein constitutes one subunit that associates with PB1 and PB2 to form the heterotrimeric influenza

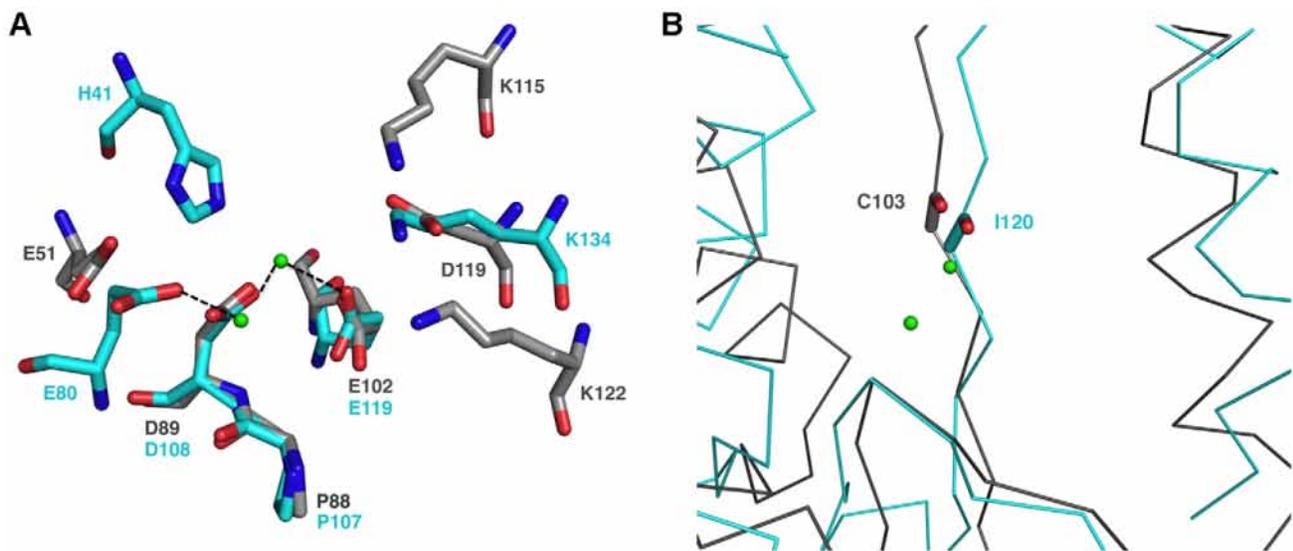
virus polymerase. Its N-terminal domain PA<sub>N</sub> hosts the RNA cap-snatching endonuclease activity [15,16]. Both NL1 and PA<sub>N</sub> share a similar core structure. Except for the absence of a fifth  $\beta$ -strand in NL1, all other secondary structure elements are conserved (Figure 1C) and the overall topology of these two structures is very similar (Figure 1D), albeit with interesting differences in the vicinity of the PA<sub>N</sub> active site (discussed below). At the aa sequence level, NL1 shares the conserved active site sequence motif characteristic of type II endonucleases: PD...(D/E)XK. In NL1, the corresponding residues are P88, D89...E102, and either K115 or K122 (Figure S1A, B). The identity of the distal lysine is not certain since it is found at different positions in the primary sequence, as is the case for influenza virus. The influenza PA<sub>N</sub> domain was crystallized either in the presence of magnesium or manganese ions in the active site which comprises five conserved catalytic residues: H41, E80, D108, E119 and K134. A structural superimposition of the arenavirus NL1 and influenza PA<sub>N</sub> active sites shows that the side-chains of three evolutionary-conserved residues within arenaviruses (P88, D89 and E102) closely superimpose with P107, D108 and E119 of the influenza virus PA<sub>N</sub> protein, pointing to a common function for these residues (Figure 2A and Figure S1B). Upon superimposition with PA<sub>N</sub>, one Mn<sup>2+</sup> ion needed for the enzymatic reaction coordinated by D108 in the PA<sub>N</sub> active site, falls at right distances to be coordinated by the carboxylate side-chains of D89 and E102. NL1 was crystallized without metal ions and a water molecule is found close to the position that should be occupied by the divalent metal. Interestingly, no close structural match is found neither for H41 nor K134 of the influenza virus PA<sub>N</sub>. This points to differences between the two active sites since His41 was proposed to play a catalytic role in the influenza PA<sub>N</sub>. However, we note that another possible contributor could be NL1 C103 main-chain carbonyl as it superimposes quite well with PA<sub>N</sub> I120 main-chain carbonyl (Figure 2B). The triad made of K115, D119, and K122 in NL1 is spatially equivalent to K134 in PA<sub>N</sub>. In summary, despite no aa sequence homology, the active site structures of the influenza PA<sub>N</sub>

and LCMV NL1 domains are clearly related but not identical (Figure 1C, 2), strongly suggesting that these two domains exhibit closely related enzymatic activities (see below).

### The NL1 Endonuclease Fold is Conserved Amongst *Bunyaviridae*

In addition to *Arenaviridae* and *Orthomyxoviridae*, *Bunyaviridae* is the other family of virus to possess a segmented negative-strand RNA genome. It contains four genera of animal viruses (*Orthobunyavirus*, *Phlebovirus*, *Nairovirus*, *Hantavirus*) and one genus of plant virus (*Tospovirus*) [23]. Although the genomic organisation differs between these three virus families, *Bunyaviridae* are also thought to use a cap-snatching mechanism to prime mRNA synthesis [24]. Arenaviruses, and *Bunyaviridae* share a conserved RdRp motif within their large L protein, as well as a conserved N-terminal domain [18]. Amino-acid sequence alignments, assisted by secondary structure prediction, of the N-terminal part of LCMV and *Bunyaviridae* L protein reveal that the latter also possesses the conserved active site motifs characteristic of type II endonucleases (Figure S2A). However, we could identify the catalytic motif within the L protein N-terminal end for only four out of the five bunyavirus genera: *Orthobunyavirus*, *Phlebovirus*, *Hantavirus* and *Tospovirus*. The L protein of *Nairovirus* is much larger (~4000 aa) than the L protein of other members of the *Bunyaviridae* family (~2200 aa). The putative endonuclease catalytic motif was located after aa ~700, the N-terminal of *Nairovirus* L protein being assigned as a so-called OTU-like domain [25].

Secondary structure predictions were used to draw the topology diagram of the NL1-like domain for each genera (Figure S2B). As expected from the sequence alignment, each genus seems to share a  $\beta$ -sheet with a variable number of  $\beta$ -strands. Furthermore, the PD catalytic motifs are in each case located in a loop before a  $\beta$ -strand, as expected. The PUMV, HLCV and RVFV NL1-like domains are more closely related to LCMV NL1 than are the TOMV and CCGV. The TOMV NL1-like domain contains 6  $\beta$ -strands and shares the PD motif just upstream the first  $\beta$ -strand,



**Figure 2. The endonuclease active site.** **A**, Structure-based superimposition of the endonuclease active site from the influenza PA<sub>N</sub> protein and the arenavirus NL1 domain. Putative active site residues of NL1 are shown as grey sticks and the active site of PA<sub>N</sub> (PDB code 2W69) in cyan. The two Mn<sup>2+</sup> ions present in the PA<sub>N</sub> structure (but not in the present NL1 domain crystal structure) are shown as light green spheres with their closest ligand indicated by a dashed line. **B**, C-alpha trace ribbon-representation of the superimposition of the endonuclease active site from the influenza PA<sub>N</sub> (cyan) protein and the arenavirus NL1 domain (grey). The carbonyl main-chain of PA<sub>N</sub> I120 and NL1 C103 are shown in sticks. The metal ions are shown as light green spheres.

doi:10.1371/journal.ppat.1001038.g002

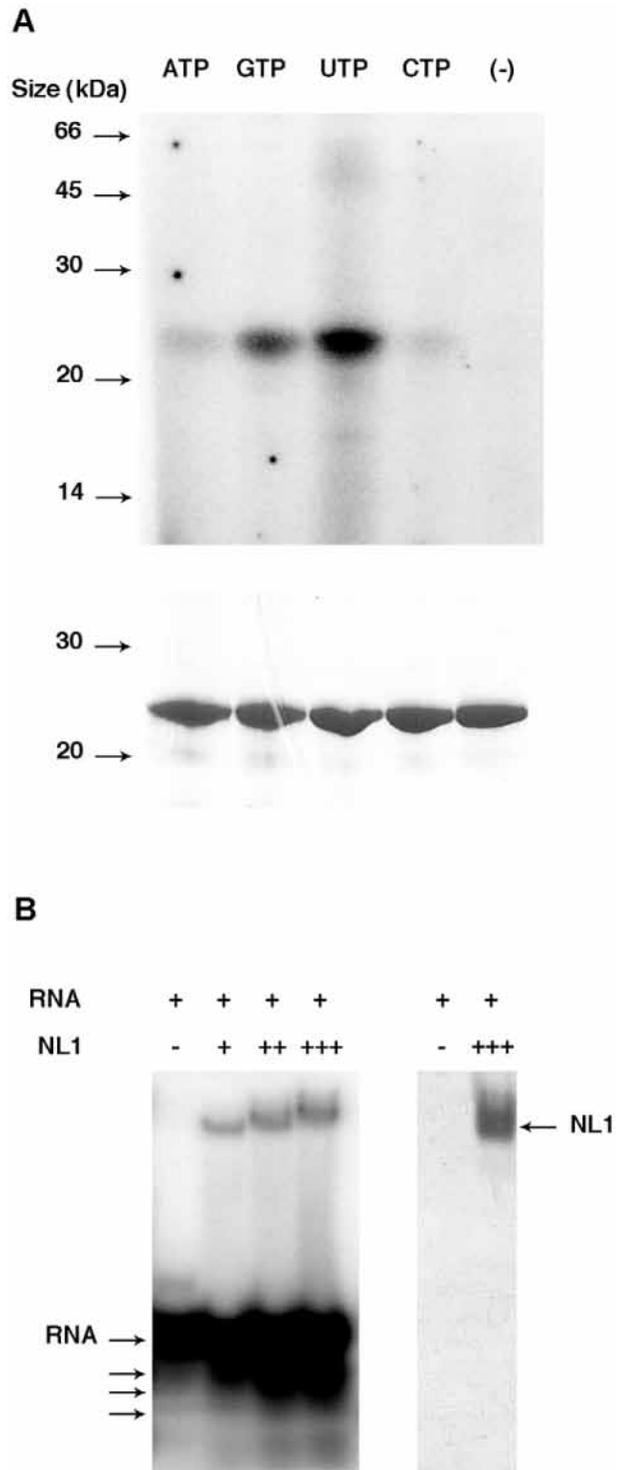
whereas it is just upstream the second  $\beta$ -strand in the case of NL1 and  $PA_N$ . Finally, the structural organization of the putative CCGV endonuclease domain seems to diverge even further from the others. Indeed, whereas the conserved lysine is shared by the same helix for all the domains, that of *Nairovirus* may be located at the end of the  $\beta_4$  strand (Figure S2B). Thus we conclude that the endonuclease motif is conserved across four animal virus genera *Orthobunyavirus*, *Phlebovirus*, *Nairovirus* and *Hantavirus*.

### NL1 is a $Mn^{2+}$ -Dependent RNA Endonuclease

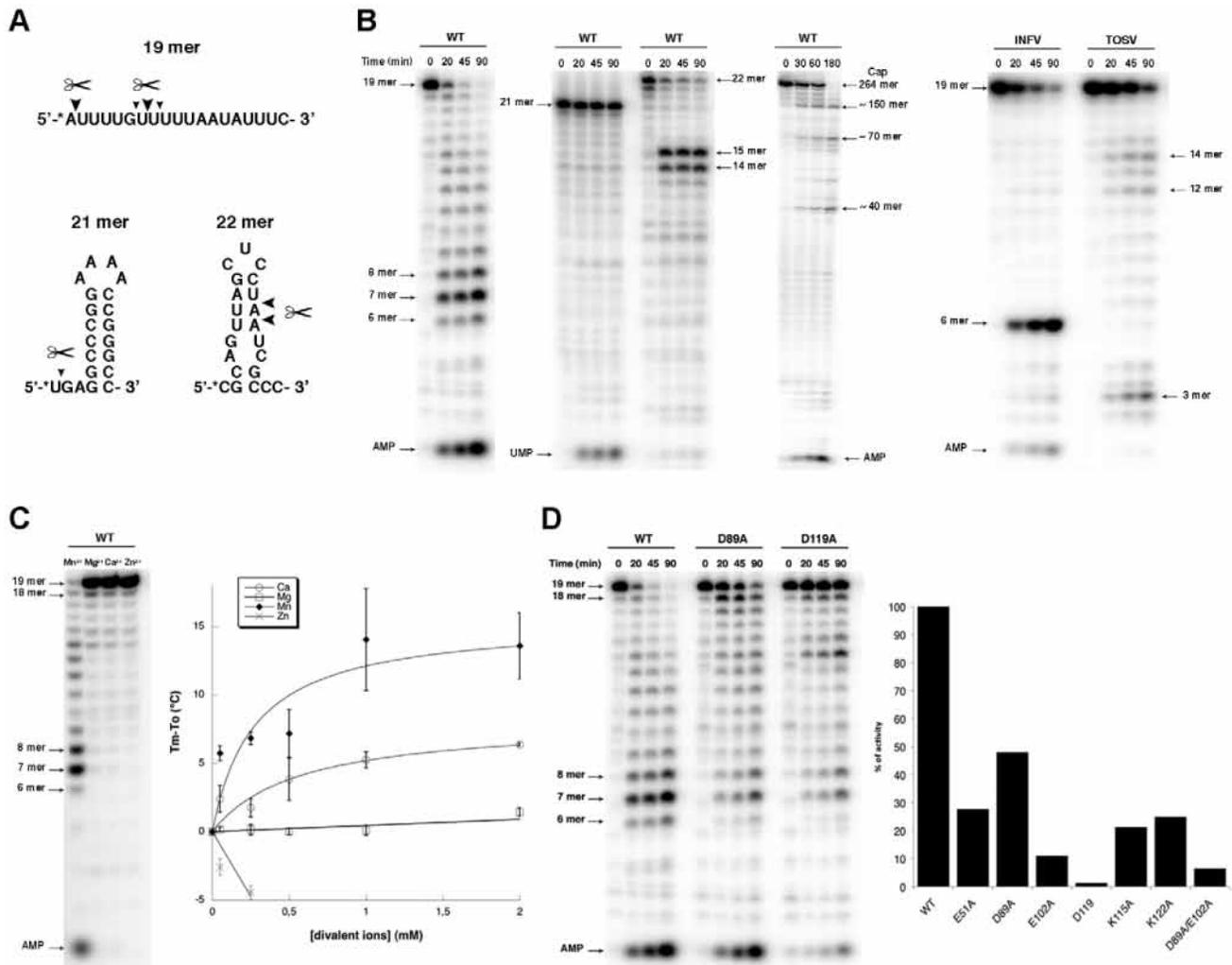
Recent crystal structures of complexes of  $PA_N$  with three different nucleoside monophosphates show that  $PA_N$  binds nucleotides [26]. The ability of NL1 to bind nucleotides was investigated using UV-crosslink experiments. We observe that NL1 binds NTPs, preferably UTP and GTP, whereas ATP and CTP show a weaker association (Figure 3A). The  $PA_N$  structures were determined in complex with ATP, CTP and UTP but not GTP [26] whereas NL1 bind GTP in a stronger fashion than ATP or CTP. The crystal structure relatedness to the endonuclease fold would suggest that the NL1 domain is able to bind RNA rather than nucleotides. We tested RNA binding by NL1, and found that indeed, NL1 binds RNA (Figure 3B). The band shift assay is also suggestive that the RNA substrate is cleaved under the assay conditions, as judged by degradation products at the bottom of the gel under the labeled RNA oligo (Figure 3B). Therefore, we surmise that nucleotide binding properties observed here reflect the ability of NL1 to bind RNA with some sequence specificity in the cap-snatching pathway (see below).

Several synthetic RNA oligonucleotides were used to characterize the endonuclease activity (Figure 4). NL1 is able to cleave ssRNA having no stable secondary structure at specific sites indicating a preference for the presence of uracil (Figure 4A, B), and adenosine to a lesser extent. Likewise, a moderately stable RNA hairpin containing uracil ( $\Delta G = -3.4$  kcal/mole) is cleaved down to a 14/15-mer product whereas a stable ( $\Delta G = -14.7$  kcal/mole) RNA hairpin devoid of uracil remains unattacked even in its single stranded regions (Figure 4A, B). PolyU RNA is cleaved randomly down to a 8-mer product with a better efficiency than polyA, whereas polyC is not a substrate for NL1 (not shown). A 5'-terminal nucleoside uracil or adenosine 5'-monophosphate is also cleaved and the 5'-monophosphate RNA end apparently competes for internal cleavage. A 5'-capped RNA of 264 nucleotides in length also acts as a substrate. It is cleaved at several specific positions indicated by the sequential appearance of band products over time (Figure 4B). This indicates that the cap structure does not seem to be a direct RNA binding determinant. A *Phlebovirus* (Toscana) virus endonuclease domain was prepared according to bio-informatic predictions described above. Its endonuclease activity was compared to both that of arenavirus NL1 and the influenza H5N1 endonuclease [16]. The enzymes were equally active using short RNA substrates, although it is apparent that sequence-specific cleavage is different for each enzyme: the influenza enzymes prefers cleavage at puric sites, Toscana virus and LCMV enzymes prefer adenosine- and uracil-containing sites (Figure 4B). NL1 is ~90-fold more active in the presence of  $Mn^{2+}$  than  $Mg^{2+}$ , and shows background activity with  $Ca^{2+}$  and  $Zn^{2+}$  (Figure 4C and not shown). The  $Mn^{2+}$  ion has also a significant stabilizing effect as judged by thermostability studies, whereas  $Zn^{2+}$  has a deleterious effect.

Mutagenesis analysis of most residues identified as part of the active site (Figure 2A) impaired the endonuclease activity. The most drastic effect was observed for D119, but residual activity was scored for E51, D89, and less for E102 (Figure 4D). As these three residues might coordinate metal ions as proposed above, defective



**Figure 3. Nucleotide and RNA binding assays of LCMV NL1 domain.** **A**, Cross-linking assay. 7  $\mu$ g of purified protein were incubated in the absence (-) or presence of each indicated radiolabelled NTP. The mixture was then UV-irradiated and loaded onto a denaturing polyacrylamide gel. The latter was analyzed by autoradiography (top) and Coomassie blue staining (bottom). **B**, Band shift assay. Radiolabelled RNA was incubated with increasing quantities (1.4  $\mu$ g (+), 4.2  $\mu$ g (++) and 7  $\mu$ g (+++)) of NL1 protein. Reaction mixture was then analyzed by PAGE, and the gel was visualized by autoradiography (left) and Coomassie blue staining (right). Apparent degradation products are indicated by arrows under the RNA input arrow. doi:10.1371/journal.ppat.1001038.g003



**Figure 4. Endonuclease activity of NL1.** **A**, Nucleotide sequence of the radiolabelled RNA used for the activity assays. The \* indicates the radiolabelled nucleotide. Big and small triangles indicates the primary and secondary cleavage site by wild-type (WT) NL1, respectively. **B**, Kinetics of endonuclease activity of WT NL1 on different substrates (left), and for Influenza and Toscana virus (INFV, TOSV) endonuclease domain (right). Activity assays were performed as described in Materials and Methods, using 3.3  $\mu$ M of RNA and 1  $\mu$ M of protein. Reactions were quenched by the addition of EDTA/formamide, and analyzed using 20% polyacrylamide/7M urea gel. Substrate and degradation product sizes are indicated. **C**, Divalent cations effect on the NL1 activity. The reaction was allowed to proceed during 45 min as described in Materials and Methods. The divalent cation assay (left) was run during 45 min without intermediate points. Titration of divalent ions on NL1 by thermal shift assay (right). T<sub>m</sub> is the melting temperature of NL1 with the divalent ions ; T<sub>0</sub> is the melting temperature of the protein alone. **D**, Mutational analysis of NL1 domain on the endonuclease activity. Kinetics were performed as described above, with WT, D89A and D119A mutants (Left). Graph showing the % of endonuclease activity determined using Fujilmager normalized quantitation for the different mutants (right). doi:10.1371/journal.ppat.1001038.g004

metal-binding due to a point mutation might be compensated by the presence of the remaining two adjacent acidic residues. A double mutant D89A/E102A shows further reduced but not abolished activity. Mutations K115A and K122A generated strongly altered activity, but the similar level of residual activity does not allow the identification of which lysine is predominant in catalysis.

### The Endonuclease Activity is Essential for RNA Transcription, not Replication

The effect of 33 mutations in L1 on virus RNA and protein expression was studied in a cell-based mini-replicon system. The LCMV L protein mediates the synthesis of two RNA species: first, capped mRNA terminating within the intergenic region, and second, antigenomic RNA being a full-length copy of the genomic

RNA template [9,27]. This dual role in RNA synthesis is recapitulated in the mini-replicon system. It contains all trans-acting factors (L protein and NP) required for transcription and replication of a genome analogue containing Renilla luciferase as a reporter gene (mini-genome). Reporter gene expression was measured in luciferase assay (Table 2), while RNA synthesis was measured in Northern blot (Figure 5), in which luciferase mRNA and antigenome can easily be distinguished due to their size difference. Wild-type (WT) L protein led to expression of high levels of Renilla luciferase (2–3 log units signal-to-noise ratio) as well as Renilla luciferase mRNA and antigenome in a ratio of about 1:1. Expression of mutant L protein was verified by immunoblotting (Figure S3).

The phenotype of mutants E41A, E41Q, K44A, S54A, C60A, T108S, F116A, D142N, and W155A is similar to that of wild-type.

**Table 2.** Functional Analysis of L Protein Mutants in LCMV Mini-Replicon System.

Mutant <sup>1</sup>	Renilla luciferase activity (sRLU)		RNA expression level (Northern blot signal)	
	% of wild-type <sup>2</sup>	Signal-to-noise ratio <sup>3</sup>	Antigenome level, % of wild-type <sup>4</sup>	mRNA-to-antigenome ratio, relative to wild-type <sup>5</sup>
E41A	44.3	243.2	21.9	2.01
E41Q	42.4	310.7	25.5	1.95
K44A	30.2	215.8	23.6	0.93
<b>E51A</b>	<b>1.3</b>	<b>8.5</b>	<b>28.5</b>	<b>0.39</b>
<b>E51Q</b>	<b>0.5</b>	<b>3.6</b>	<b>32.9</b>	<b>0.26</b>
S54A	23.5	175.6	28.3	0.80
C60A	47.9	357.7	62.4	1.13
<b>D89A</b>	<b>0.5</b>	<b>3.5</b>	<b>107.6</b>	<b>0.20</b>
<b>D89N</b>	<b>0.5</b>	<b>3.8</b>	<b>138.5</b>	<b>0.19</b>
<b>E102A</b>	<b>0.9</b>	<b>3.5</b>	<b>119.3</b>	<b>0.22</b>
<b>E102N</b>	<b>3.3</b>	<b>13.6</b>	<b>78.2</b>	<b>0.26</b>
F104A	0.6	3.8	—	—
R106A	4.4	32.5	—	—
T108S	29.7	212.7	48.0	0.89
F112A	0.3	1.8	—	—
K115A	0.9	6.8	—	—
F116A	27.4	209.4	39.1	1.79
<b>D119A</b>	<b>2.0</b>	<b>12.5</b>	<b>47.5</b>	<b>0.06</b>
<b>D119N</b>	<b>8.0</b>	<b>43.0</b>	<b>46.4</b>	<b>0.07</b>
<b>K122A</b>	<b>1.6</b>	<b>9.3</b>	<b>55.3</b>	<b>0.04</b>
<b>D129A</b>	<b>1.3</b>	<b>6.5</b>	<b>56.5</b>	<b>0.04</b>
<b>D129N</b>	<b>1.3</b>	<b>7.4</b>	<b>54.1</b>	<b>0.05</b>
D142A	2.8	17.6	—	—
D142N	42.8	270.6	66.1	0.93
R144A	2.3	10.2	—	—
W155A	34.7	203.3	91.0	0.91
R161A	3.4	17.6	—	—
E179A	40.0	239.4	109.9	0.46
E179Q	25.4	174.6	105.0	0.44
E182A	32.6	201.3	154.9	0.40
E182Q	29.3	184.9	103.1	0.42
Y183A	16.9	98.8	82.3	0.48
R185A	0.3	1.4	—	—

<sup>1</sup>Mutants with selective defect in mRNA synthesis are shown in boldface.

<sup>2</sup>Standardized relative light unit (sRLU) value (wild-type = 100%). Mean of  $\geq 2$  independent transfection experiments.

<sup>3</sup>sRLU value of mutant divided by sRLU value of negative control mutant containing a mutation in the catalytic site of the RNA-dependent RNA polymerase. Mean of  $\geq 2$  independent transfection experiments.

<sup>4</sup>Antigenome signals in Northern blots were quantified via intensity profiles (wild-type = 100%).

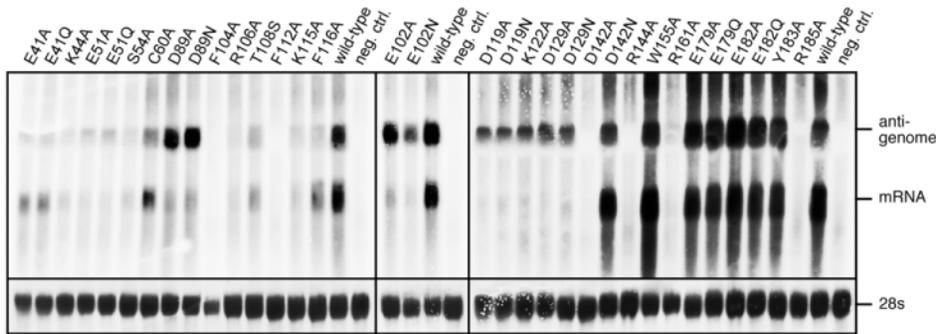
<sup>5</sup>RNA signals in Northern blots were quantified and the mRNA-to-antigenome signal ratio was calculated. The wild-type ratio was set at 1 for each experiment (i.e. the signal ratio of a mutant was normalized with the wild-type ratio) to render independent blots comparable.

doi:10.1371/journal.ppat.1001038.t002

Mutants E179A, E179Q, E182A, E182Q, and Y183A also express luciferase and RNA at high level, but the steady-state level of mRNA relative to that of antigenome is reduced by about 50%. Mutants F104A, R106A, F112A, K115A, D142A, R144A, R161A, R185A neither express Renilla luciferase nor any RNA species, indicating that global functions of L protein are affected.

The most interesting phenotype is observed with mutants D89A, D89N, E102A, E102N, D119A, D119N, K122A, D129A, and D129N. They synthesize antigenome close or equal to wild-type level, but are defective in mRNA and, thus, reporter gene

expression (Figure 5 and Table 2, shown in boldface). A similar phenotype is seen with mutants E51A and E51Q, though associated with reduced antigenome level. These data indicate that residues E51, D89, E102, D119, K122, and D129A are essential for viral mRNA synthesis, but not required for expression of uncapped RNA species. With the exception of the D129 residue located at the surface of the protein remote from the endonuclease active site, it is remarkable that these transcription-null mutants form the catalytic site (Figure 2) and match precisely those of the PD...(D/E)XK endonuclease type II signature sequence.



**Figure 5. Mutational analysis of the L protein in the context of the LCMV replicon system.** Synthesis of antigenomic RNA and Renilla luciferase mRNA was analyzed by Northern blotting. Negative control cells (neg. ctrl.) expressed mini-genome, NP, and an L protein mutant with a mutation in the catalytic site of the RNA-dependent RNA polymerase. The methylene blue-stained 28S rRNA is shown below the blots as a marker for gel loading and RNA transfer. Each panel represents an independent experiment with separate controls. Careful examination of the blots revealed residual signals at the mRNA position for some mutants negative in Renilla luciferase assay. Thus, these signals do not correspond to functional mRNA, but may be prematurely terminated antigenome.  
doi:10.1371/journal.ppat.1001038.g005

**Discussion**

The structural and functional results presented here show that the LCMV NL1 domain is an RNA endonuclease. The uncoupling of RNA replication from transcription and selective disappearance of mRNA when NL1 active site residues are mutated strongly suggests that this activity is involved in cap-snatching.

The identification of the arenavirus endonuclease is in line with the recent discovery of the PA<sub>N</sub> endonuclease domain of influenza virus. Whereas the active site of influenza virus features a cluster of three acidic residues, the active site of arenavirus contains four acidic residues (E51, D89, E102 and D119), as well as two important lysine residues K115 and K122 neighboring D119 (Figure 2A). The NL1 active site resembles but is clearly distinct from that of influenza PA<sub>N</sub>. Indeed, there is no histidine in the catalytic center, and the arenavirus NL1 nuclease has some specific features both upstream and downstream of the PD signature sequence. We define the arenavirus endonuclease motif as E-X<sub>98</sub>-P-D-X<sub>(11,13)</sub>-E-X<sub>12</sub>-K-X<sub>3</sub>-D-X<sub>2</sub>-K. The most obvious difference with the only known related RNA endonuclease, that of influenza virus PA<sub>N</sub>, is a divergence upstream the PD motif in structural elements carrying the E51 residue (Figure 1C), and the presence of a triad K...D...K at the distal side of the latter signature sequence (Figure 2A). Contrary to PA<sub>N</sub> which shares a conserved and essential histidine involved in the binding of both the metal ion and a nucleotide onto helix α3 [15,16,26], NL1 does not possess this conserved histidine residue. Instead, NL1 has a glutamic acid residue E51, which might reflect a different nucleobase specificity as detected in our nuclease assays (Figure 4). Likewise, residues downstream the PD motifs are distinct from the consensus sequence, and differently organized into a triad including two lysines. The presence of water molecules and previous structural models for influenza PA<sub>N</sub> allows to propose putative positions of metal ions, coordinated by D89 and E102.

The first step in the general mechanism for phosphodiester hydrolysis is the preparation of the attacking nucleophile by deprotonation, usually involving a general base deprotonating a water molecule. Lysine is often considered as this general base candidate in endonucleases but is not strictly conserved [28,29]. Here, there are no indications against D119 being this general base. Alternately, it could well be either lysine K115 or K122. Both are oriented towards the active site, and they could well have their pKa lowered by D119 in order to initiate the reaction. Reverse genetic studies provide evidence for K122, not K115. Indeed, mRNA

production is selectively abolished and clearly uncoupled from RNA synthesis in the case of K122A mutant, while the K115A mutant was completely defective preventing interpretation of its role in the endonuclease catalytic site. Although it is not known if uncapped mRNAs are synthesized and degraded for the transcription-null mutants, the most plausible scenario is that primer shortage prevents significant capped mRNA synthesis. Overall, the replicon data presented here closely match those obtained on the closely related Lassa arenavirus using a similar replicon system [30]. Arenaviruses may thus use two clearly independent and distinct RNA synthesis priming mechanisms: one is dependent on an active endonuclease carried by the N-terminus of the L protein, and the other might be linked to the observation that an extra G residue is found at the 5'-end of arenavirus genomes and antigenomes. The latter G bases would thus reflect a yet-uncharacterized priming mechanism unrelated to the U/A cleavage sequence preference of NL1.

NL1 also binds nucleotides, but the NTP binding site should differ from that of PA<sub>N</sub>. Indeed, the influenza PA<sub>N</sub> histidine 41 is involved in binding the nucleobase of the presumed incoming RNA substrate. The NL1 endonuclease does not share the same sequence specificity, and E51 is positioned at a spatially equivalent position.

The cap structure does not seem to be a direct RNA binding determinant (Figure 4B), as endonucleolytic cleavage is not directed to cleavage sites preferentially in the vicinity of the cap. We thus infer that an independent cap-binding site way exist elsewhere in viral proteins to bind and select cellular mRNAs, a possibility reminiscent of influenza for which PA carries the endonuclease activity and PB2 the cap binding site [15,16,31].

Structure and sequence alignment studies show that the N-terminal endonuclease domain of the L protein is also conserved in the *Bunyaviridae* family, although the *Nairovirus* endonuclease domain is not located into the N-terminal end of the protein. These findings were confirmed by the endonuclease activity of the N-terminal end of the L protein of TOSV (Figure 4B). Thus, we provide evidence that all three segmented negative single-strand RNA virus species share an endonuclease domain probably involved in the cap-snatching process during the viral life cycle. These data raise the question of a possible common ancestor for these viruses. Indeed, these three virus families use a cap-snatching mechanism involving binding and cleavage of cellular mRNA caps subsequently used by a large primer-dependent RNA-dependent RNA polymerase. It seems more plausible that the L gene has evolved by divergence over time, rather than by multiple

acquisitions of several activities converging into a common structure, at least in the case of the endonuclease. Furthermore, our study raises the interesting possibility that other activities involved in RNA replication/transcription might be discovered by comparative analysis of *Orthomyxoviridae* PB1, PB2, PA and *Arenaviridae/Bunyaviridae* L proteins.

To our knowledge, a single crystal structure of a functional arenavirus protein is currently available, that of the Machupo virus glycoprotein GP1 in complex with its human receptor, TIR1 [32]. Our results provide an arenavirus L domain structure, with a role consistent with the hypothesis of a cap-snatching mechanism suggested for arenaviruses [9,10]. The strategy used here to produce individually active domains might be useful to further characterize the *Arenaviridae/Bunyaviridae* large L protein which had so far resisted all biochemical characterization attempts.

The influenza, *Arenaviridae* and *Bunyaviridae* endonucleases are so far the only three examples of RNA endonucleases similar to type II DNA restriction endonucleases. The presence of such an endonuclease suggests that it could serve as a fruitful target for antiviral strategies against these two families, since such kind of inhibitors have been reported in the case of the influenza virus [33,34,35].

## Materials and Methods

### Cloning, Expression and Purification of LCMV NL1 Domain

The LCMV NL1 cDNA (Armstrong strain, aa 1 to 196) was cloned into pDest14 with a N-terminus hexa-histidine tag and expressed in *E. coli* Rosetta (DE3) pLysS (Novagen), at 17°C in 2YT medium overnight after induction with 500 µM IPTG. Cell pellets from harvested cultures were resuspended in 50 mM Tris buffer, pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.1% Triton, 5% Glycerol. Lysozyme (0.25 mg/ml), PMSF (1 mM), DNase I (2 µg/ml), and EDTA free protease cocktail (Roche) were added before sonication. IMAC chromatography of clarified lysates was performed on a 5 ml His prep column (Akta Xpress FPLC system, GE Healthcare) eluted with imidazole. Size exclusion chromatography was performed on preparative Superdex 200 column (GE Healthcare) pre-equilibrated in 10 mM Imidazole, pH 8.0, 50 mM NaCl, 2 mM DTT. Protein was concentrated (28 mg/ml) using a centrifugal concentrator. For enzymatic studies, WT and mutants were expressed in the *E. coli* BL21 star strain (Invitrogen) and further purified on HiTrap Q sepharose 1 ml column (GE Healthcare) to remove *E. coli* RNase contaminants. Proteins eluted in a linear gradient from 50 mM to 1 M NaCl in 10 mM Hepes buffer, pH 7.5, 2 mM DTT. A synthetic gene of the H5N1 PA<sub>N</sub> endonuclease was designed as described [16]. The Toscana virus (strain France AR\_2005, aa 2 to 233) cDNA was obtained from infected cell cultures. Both ORFs were cloned as a N-terminal Thioredoxin-Hexahistidine fusion in pETG20A. The tag was cleaved using TEV protease before a final gel filtration.

### Crystallization

Crystals grew in LiSO<sub>4</sub> 250 mM, citrate 50 mM, isopropanol 5.5%, using the hanging drop vapor diffusion method in Linbro plates by mixing 1 µl of protein solution with 1 µl of reservoir solution. Crystals were cryoprotected by dipping in a solution containing 65% of crystallization buffer and 35% of a buffer made of size exclusion chromatography buffer/glycerol (50/50). Crystals were cryo-cooled in liquid N<sub>2</sub>. The crystals belong to space group C222<sub>1</sub> and have two molecules per asymmetric unit. Despite repeated attempts, crystal soaked in the above buffer supplemented with various concentrations of MnCl<sub>2</sub> yielded crystals diffracting to >4 Å.

### Data Collection and Structure Determination

Diffraction intensities were recorded on the ID14-4 beamline at the European Synchrotron Radiation facility, Grenoble, France. Data were processed and integrated with MOSFLM [36]. Scaling and merging of the intensities was performed with SCALA and programs from the Collaborative Computational Project, No. 4 (CCP4) suite [37]. The structure was determined using SAD data from one selenomethionylated protein crystal diffracting to 3.4 Å resolution with SHARP/autoSHARP, followed by density modification with SOLOMON and DM. An initial model was built using BUCCANEER and completed in COOT, followed by refinement using BUSTER (see Text S1). Details of structure determination are given as supplemental material. Data from a native crystals diffracting to a 2.13-Å resolution were collected on an ADSC QUANTUM 315r at a wavelength of 0.9835 Å. The structure was refined with BUSTER and COOT using this data set (Table 1) [38]. The atomic coordinates have been deposited at the PDB (3JSB).

### Sequence Retrieval

A PHI-BLAST search using the sequence corresponding to the L1 domain and the signature of the *Arenaviridae* endonuclease motif *i.e.* P-D<sub>x(11,13)</sub>-E<sub>x(12)</sub>-K<sub>x(3)</sub>-D<sub>x(2)</sub>-K; was performed against non-redundant databases [39]. After 3 iterations, Batai and Kairi viruses both belonging to *Orthomyxoviridae*, appears in the section with an E-Value below threshold. A fourth iteration including these two sequences allows retrieving the entire family of orthomyxoviruses, with E-value comprised between 3e<sup>-18</sup> and 2e<sup>-4</sup>.

A standard CDD search from the sequence of Tensaw virus allows retrieving all the L of the *Bunyaviridae* family hitting the pfam 04196 [40].

### Sequence Comparison

A multiple sequence alignment of the N-terminal end of the L protein from LCMV, HLCV, BUNV, HANV, PUMV, RVFV, TOSV, TOMV, WTMV, CCGV, DUGV, was first performed with the T-coffee algorithm (<http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee.cgi/index.cgi>). Using the secondary structure prediction of the endonuclease domain of L proteins, the putative conserved active site residues were identified and placed correctly in the alignment.

### UV-Crosslink Experiments

7 µg of purified protein were incubated for 15 min at 25°C, with 0.5 µl of the various α-<sup>32</sup>P NTP (0.4 µCi/µl) in 10 µl of reaction buffer containing 10 mM Imidazole, pH 8.0, 50 mM NaCl, 2 mM DTT. The reaction mixtures were then exposed to UV light (254 nm) for 6 min at 5 mm distance. The crosslinked species were separated in a 15% polyacrylamide denaturing gel and visualized by autoradiography using photo-stimulated plates and a Fujilmager (Fuji).

### RNA Binding Experiments

The RNA 5'-AUUUUGUUUUUAAUUAUUUC-3' (Ambion) was [<sup>32</sup>P] 5'-end labeled, and 0.4 µM of radiolabelled RNA was incubated 20 min at 25°C without and with 1.4 µg, 4.2 µg and 7 µg of protein in 10 µl of 10 mM Imidazole, pH 8.0, 50 mM NaCl, 2 mM DTT. Reaction mixtures was analyzed by PAGE and visualized by autoradiography.

### Ion Binding Assays

Titration curves with CaCl<sub>2</sub>, MnCl<sub>2</sub>, MgCl<sub>2</sub> and ZnCl<sub>2</sub> were performed at 1 mg/ml protein in gel filtration buffer using thermal shift assay. Technical details can be found in [41].

## Endonuclease Assays

Endonuclease activity was assayed using 4 different heteromeric RNA substrates: an unstructured 19 mer as described above, a 21 mer stable hairpin (5'-UGAGGCCCGGAAACCGGGGCC-3' (Ambion),  $\Delta G = -14.7$  Kcal/mole), a 22 mer moderately stable hairpin (5'-CGCAGUUAGCUCCUAAUCGCC-3' (Ambion),  $\Delta G = -3.4$  Kcal/mole), and a long 264 mer RNA corresponding to the SARS-CoV 5'-genome sequence. The latter was radiolabelled with a cap structure at its 5'-end using the ScriptCap m7G Capping System (Epicentre *Biotechnologies*) with [ $\alpha$ - $^{32}$ P]GTP. Endonuclease assays were carried out using 3.3  $\mu$ M of radiolabeled RNA in a buffer containing 40 mM Tris-base, pH 7.5, 100 mM NaCl, 10 mM  $\beta$ -Mercaptoethanol and 2 mM MnCl<sub>2</sub>. Reactions were initiated by the addition of 1  $\mu$ M of protein and incubated at 37°C, and stopped by the addition of EDTA/formamide. Reactions products were analyzed using denaturing polyacrylamide gel electrophoresis (20% polyacrylamide, 7 M urea in TTE buffer (89 mM Tris, 28 mM taurine, 0.5 mM EDTA) and analyzed by autoradiography.

## Mutagenesis and Reverse Genetics Assays Using a LCMV Mini-Replicon System

The LCMV replicon system is based on strain Armstrong clone 13 and has been established in analogy to the Lassa virus replicon described previously [42]. BSR T7/5 cells constitutively expressing T7 RNA polymerase [43] were transiently transfected with T7 promoter-driven expression constructs for L protein, nucleoprotein (NP), mini-genome (MG) containing Renilla luciferase reporter gene, and firefly luciferase as a transfection control. L protein mutants were generated as described [44]. One day after transfection, total RNA was prepared for Northern blotting and cell lysate was assayed for firefly and Renilla luciferase activity. Renilla luciferase levels were normalised with firefly luciferase levels resulting in standardized relative light units (sRLU). Northern blot was performed using an antisense  $^{32}$ P-labeled riboprobe targeting the Renilla luciferase gene. Autoradiography was quantified on a PhosphorImager (Amersham Biosciences). To verify protein expression, hemagglutinin (HA)-tagged L protein was expressed in BSR T7/5 cells inoculated with modified vaccinia virus Ankara expressing T7 RNA polymerase (MVA-T7) [45] and detected in immunoblot using anti-HA antibody.

## Supporting Information

**Figure S1** Sequence Alignment of Viral Endonuclease Domains. **A**, Structure-based sequence alignment of NL1 with the two influenza PAN (3EBJ, 2W69), Tt1808 (1WDJ) and SdaI (2IXS), showing the structurally-conserved endonuclease motif (Highlighted in red). **B**, Sequence alignment of arenavirus NL1: Lymphocytic choriomeningitis virus (LCMV), Dandenong virus (DANV), Mopeia virus (MOPV), Morogoro virus (MORV), Mobala virus (MOBV), Ippy virus (IPPV), Lassa virus (LASV), Lujo Virus (LUJV), Parana virus (PARV), Pichinde virus (PICV), Allpahuayo virus (ALLV), Chapare virus (CHAV), Tamiami virus (TAMV), Whitewater Arroyo virus (WWAV), Bear Canyon virus (BCNV), Flexal virus (FLEV), Pirital virus (PIRV), Amapari virus (AMAV), Guanarito virus (GTOV), Cupixi virus (CPXV), Machupo virus

## References

- Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, et al. (2009) Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog* 5: e1000455.
- Charrel RN, de Lamballerie X, Emonet S (2008) Phylogeny of the genus Arenavirus. *Curr Opin Microbiol* 11: 362–368.
- Barton LL, Budd SC, Morfitt WS, Peters CJ, Ksiazek TG, et al. (1993) Congenital lymphocytic choriomeningitis virus infection in twins. *Pediatr Infect Dis J* 12: 942–946.
- Barton LL, Hyndman NJ (2000) Lymphocytic choriomeningitis virus: reemerging central nervous system pathogen. *Pediatrics* 105: E35.

(MACV), Junin virus (JUNV), Tacaribe virus (TCRV), Sabia virus (SABV), Oliveros virus (OLVV), Latino virus (LATV). Residues in a solid red background are strictly conserved. The blue point indicates the key active site residues.

Found at: doi:10.1371/journal.ppat.1001038.s001 (4.29 MB TIF)

**Figure S2** Conservation of NL1 Between Arenaviruses and Bunyaviruses. **A**, Sequence alignment showing endonuclease motif (highlighted in red) of the NL1 domain from LCMV and the five genera of the *Bunyaviridae* families. Each genera is represented by two viruses: *Orthobunyavirus*: Human La Cross Virus (HLCV) and Bunyamwera virus (BUNV), *Hantavirus*: Hantaan virus (HANV) and Puumala virus (PUMV), *Phlebovirus*: Rift valley fever virus (RVFV) and Toscana virus (TOSV), *Tospovirus*: Tomato virus (TOMV) and Watermelon silver mottle virus (WTMV), *Nairovirus*: Crimean-Congo hemorrhagic fever virus (CCGV) and Dugbe virus (DUGV). Position of the start of the motif is labelled in blue for each virus. **B**, Based on secondary structure predictions topology diagrams were drawn for the NL1 domains from PUMV, RVFV, TOMV, HANV, CCGV. Colours are the same as in Figure 1D. N-terminal and C-terminal position are labelled with the aa position. The two red dots, the black dot and the green dot indicate the PD, E/D, and K residues, respectively, from the key active site. Principal conserved secondary structures are labelled as for LCMV NL1 domain.

Found at: doi:10.1371/journal.ppat.1001038.s002 (8.97 MB TIF)

**Figure S3** Verification of L protein expression by immunoblotting. L protein used for analysis in replicon assay was tagged with HA tag, expressed under T7 promoter control in cells inoculated with modified vaccinia virus Ankara expressing T7 RNA polymerase (MVA-T7), and detected in immunoblot using anti-HA antibody. MVA, cells inoculated with MVA-T7 but not transfected; neg. ctrl., L mutant containing a mutation in the catalytic site of the RdRp.

Found at: doi:10.1371/journal.ppat.1001038.s003 (1.13 MB TIF)

**Text S1** Supplementary Methods.

Found at: doi:10.1371/journal.ppat.1001038.s004 (0.08 MB DOC)

## Acknowledgments

We thank Joëlle Boretto, Violaine Lantez, Karen Dalle for excellent technical assistance in preparing cDNAs and proteins used throughout, Lukas Flatz and Daniel Pinschewer for providing L and NP gene plasmids to facilitate construction of the LCMV mini-replicon system, Boris Klempa for helpful discussion and support on Bunyavirus work, Rolf Hilgenfeld and Hugues Tolou for sharing information and brainstorming at the initial phase of the project.

## Author Contributions

Conceived and designed the experiments: B. Morin, R. Charrel, S. Günther, B. Canard. Performed the experiments: B. Morin, B. Coutard, M. Lelke, F. Ferron, R. Kerber, S. Jamal, A. Frangeul, C. Baronti, R. Charrel, X. de Lamballerie, J. Lescar. Analyzed the data: B. Morin, B. Coutard, M. Lelke, F. Ferron, R. Kerber, C. Vonnrhein, G. Bricogne, S. Günther, B. Canard. Contributed reagents/materials/analysis tools: G. Bricogne, B. Canard. Wrote the paper: B. Morin, J. Lescar, B. Canard. Coordinated the study: B. Canard.

5. Fischer SA, Graham MB, Kuehnert MJ, Kotton CN, Srinivasan A, et al. (2006) Transmission of lymphocytic choriomeningitis virus by organ transplantation. *N Engl J Med* 354: 2235–2249.
6. Palacios G, Druce J, Du L, Tran T, Birch C, et al. (2008) A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med* 358: 991–998.
7. Meyer BJ, de la Torre JC, Southern PJ (2002) Arenaviruses: genomic RNAs, transcription, and replication. *Curr Top Microbiol Immunol* 262: 139–157.
8. Salvato MS, Shimomaye EM (1989) The completed sequence of lymphocytic choriomeningitis virus reveals a unique RNA structure and a gene for a zinc finger protein. *Virology* 173: 1–10.
9. Meyer BJ, Southern PJ (1993) Concurrent sequence analysis of 5' and 3' RNA termini by intramolecular circularization reveals 5' nontemplated bases and 3' terminal heterogeneity for lymphocytic choriomeningitis virus mRNAs. *J Virol* 67: 2621–2627.
10. Raju R, Raju L, Hacker D, Garcin D, Compans R, et al. (1990) Nontemplated bases at the 5' ends of Tacaribe virus mRNAs. *Virology* 174: 53–59.
11. Polyak SJ, Zheng S, Harnish DG (1995) 5' termini of Pichinde arenavirus S RNAs and mRNAs contain nontemplated nucleotides. *J Virol* 69: 3211–3215.
12. Plotch SJ, Bouloy M, Krug RM (1979) Transfer of 5'-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro. *Proc Natl Acad Sci U S A* 76: 1618–1622.
13. Plotch SJ, Bouloy M, Ulmanen I, Krug RM (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* 23: 847–858.
14. Bishop DH, Gay ME, Matsuoka Y (1983) Nonviral heterogeneous sequences are present at the 5' ends of one species of snowshoe hare bunyavirus S complementary RNA. *Nucleic Acids Res* 11: 6409–6418.
15. Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, et al. (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* 458: 914–918.
16. Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, et al. (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* 458: 909–913.
17. Lopez N, Jacamo R, Franze-Fernandez MT (2001) Transcription and RNA replication of tacaribe virus genome and antigenome analogs require N and L proteins: Z protein is an inhibitor of these processes. *J Virol* 75: 12241–12251.
18. Muller R, Poch O, Delarue M, Bishop DH, Bouloy M (1994) Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J Gen Virol* 75(Pt 6): 1345–1352.
19. Vieth S, Torda AE, Asper M, Schmitz H, Gunther S (2004) Sequence analysis of L RNA of Lassa virus. *Virology* 318: 153–168.
20. Lukashevich IS, Djavani M, Shapiro K, Sanchez A, Ravkov E, et al. (1997) The Lassa fever virus L gene: nucleotide sequence, comparison, and precipitation of a predicted 250 kDa protein with monospecific antiserum. *J Gen Virol* 78(Pt 3): 547–551.
21. Wilda M, Lopez N, Casabona JC, Franze-Fernandez MT (2008) Mapping of the tacaribe arenavirus Z-protein binding sites on the L protein identified both amino acids within the putative polymerase domain and a region at the N terminus of L that are critically involved in binding. *J Virol* 82: 11454–11460.
22. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DALI-Lite v.3. *Bioinformatics* 24: 2780–2781.
23. Nichol STBJ, Elliott RM (2005) Virus Taxonomy, VIIIth Report of the ICTV. In: Fauquet CM, Mayo AM, Maniloff J et al eds London: Elsevier Academic Press. pp 695–716.
24. Gro MC, Di Bonito P, Accardi L, Giorgi C (1992) Analysis of 3' and 5' ends of N and NSs messenger RNAs of Toscana Phlebovirus. *Virology* 191: 435–438.
25. Frias-Staheli N, Giannakopoulos NV, Kikkert M, Taylor SL, Bridgen A, et al. (2007) Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. *Cell Host Microbe* 2: 404–416.
26. Zhao C, Lou Z, Guo Y, Ma M, Chen Y, et al. (2009) Nucleoside monophosphate complex structures of the endonuclease domain from the influenza virus polymerase PA subunit reveal the substrate binding site inside the catalytic center. *J Virol* 83: 9024–9030.
27. Garcin D, Kolakofsky D (1990) A novel mechanism for the initiation of Tacaribe arenavirus genome replication. *J Virol* 64: 6196–6203.
28. Newman M, Strzelecka T, Dorner LF, Schildkraut I, Aggarwal AK (1994) Structure of restriction endonuclease BamHI and its relationship to EcoRI. *Nature* 368: 660–664.
29. Pingoud A, Fuxreiter M, Pingoud V, Wende W (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci* 62: 685–707.
30. Lelke M, Brunotte L, Busch C, Gunther S (2010) An N-terminal region of Lassa virus L protein plays a critical role in transcription but not replication of the virus genome. *J Virol* 84: 1934–1944.
31. Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, et al. (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* 15: 500–506.
32. Abraham J, Corbett KD, Farzan M, Choe H, Harrison SC. Structural basis for receptor recognition by New World hemorrhagic fever arenaviruses. *Nat Struct Mol Biol*.
33. De Clercq E, Neyts J (2007) Avian influenza A (H5N1) infection: targets and strategies for chemotherapeutic intervention. *Trends Pharmacol Sci* 28: 280–285.
34. Hsieh HP, Hsu JT (2007) Strategies of development of antiviral agents directed against influenza virus replication. *Curr Pharm Des* 13: 3531–3542.
35. Parkes KE, Ermert P, Fassler J, Ives J, Martin JA, et al. (2003) Use of a pharmacophore model to discover a new class of influenza endonuclease inhibitors. *J Med Chem* 46: 1153–1164.
36. Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* 55: 1690–1695.
37. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50: 760–763.
38. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126–2132.
39. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
40. Finn RD, Mistry J, Tate J, Coghill P, Heger A, et al. The Pfam protein families database. *Nucleic Acids Res* 38: D211–222.
41. Malet H, Coutard B, Jamal S, Dutartre H, Papageorgiou N, et al. (2009) The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. *J Virol* 83: 6534–6545.
42. Hass M, Gölnitz U, Müller S, Becker-Ziaja B, Günther S (2004) Replicon system for Lassa virus. *J Virol* 78: 13793–13803.
43. Buchholz UJ, Finke S, Conzelmann KK (1999) Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. *J Virol* 73: 251–259.
44. Hass M, Lelke M, Busch C, Becker-Ziaja B, Günther S (2008) Mutational evidence for a structural model of the Lassa virus RNA polymerase domain and identification of two residues, Gly1394 and Asp1395, that are critical for transcription but not replication of the genome. *J Virol* 82: 10207–10217.
45. Sutter G, Ohlmann M, Erfle V (1995) Non-replicating vaccinia vector efficiently expresses bacteriophage T7 RNA polymerase. *FEBS Lett* 371: 9–12.
46. Kabsch W Xds. *Acta Crystallogr D Biol Crystallogr* 66: 125–132.
47. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62: 72–82.
48. Vornrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364: 215–230.
49. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58: 1772–1779.
50. Bricogne G, Vornrhein C, Flensburg C, Schiltz M, Paciorek W (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* 59: 2023–2030.
51. Cowtan K (1994) An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31: 34–38.
52. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62: 1002–1011.
53. Abrahams JP, Leslie AG (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* 52: 30–42.
54. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501.
55. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, et al. (2010) BUSTER version 2.X. Global Phasing Ltd. Cambridge, United Kingdom. .
56. Leslie AG (1992) MOSFLM - Recent changes and future developments. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 35: 18–19.
57. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240–255.
58. Smart O, Brandl M, Flensburg C, Keller P, Paciorek W, et al. (2008) Refinement with Local Structure Similarity Restraints (LSSR) Enables Exploitation of Information from Related Structures and Facilitates use of NCS. Annual Meeting of the American Crystallographic Association.