Fine mapping and characterization of the L polymerase-binding domain of the respiratory syncytial virus phosphoprotein

Julien Sourimant1, 3, Marie-Anne Rameix-Welti1, 3, 4, Anne-Laure Gaillard1,*, Didier Chevret2, Marie Galloux1, Elyanne Gault3, 4, and Jean-François Eléouët1#

1 Unité de Virologie et Immunologie Moléculaires (UR892), INRA, Jouy-en-Josas, 78352, France
2 INRA, UMR1319 Micalis, Plateforme d’Analyse Protéomique de Paris Sud-Ouest, 78352 Jouy-en-Josas, France
3 INSERM U1173, UFR Simone Veil, Versailles-Saint-Quentin University, 78180 Saint-Quentin en Yvelines, France
4 AP-HP, Hôpital Ambroise Paré, Laboratoire de Microbiologie, Boulogne-Billancourt, 92100, France

* These authors contributed equally to this work
# Corresponding author:

jean-francois.eleouet@jouy.inra.fr
Abstract

The minimum requirement for an active RNA-dependent RNA polymerase of respiratory syncytial virus is a complex made of two viral proteins, the polymerase large protein L and the phosphoprotein P. Here, we have investigated the domain on P responsible for this critical P-L interaction. Using recombinant proteins and serial deletions, a L binding site was mapped in the C-terminal region of P, just upstream from the N-RNA binding site. The role of this molecular recognition element of about 30 amino acid residues in L-P interaction and RNA polymerase activity was evaluated in cellula using an RSV minigenome system and site-directed mutagenesis. The results highlighted the critical role of hydrophobic residues located in this region.

Importance. Respiratory syncytial virus (RSV) is the leading cause of lower respiratory tract illness in infants. Since no vaccine or good antivirals are available against RSV, it is essential to better understand how the viral machinery functions in order to develop new antiviral strategies. Like all negative strand RNA viruses, RSV codes for its own machinery to replicate and transcribe its genome. The core of this machinery is composed of two proteins, the phosphoprotein P and the large protein L. Here, using recombinant proteins, we have mapped and characterized the P domain responsible for this L-P interaction and formation of an active L-P complex. These findings extend our understanding of the mechanism of action of RSV RNA polymerase and allow us to define a new target for the development of drugs against RSV.

Introduction

Human respiratory syncytial virus (RSV) is the leading cause of acute respiratory infections in infants worldwide and is the primary cause of infant hospitalization for respiratory infections (1). Moreover, RSV is increasingly recognized as a significant cause of disease in the elderly population and can often be fatal for patients with compromised immune systems (2). Like its human counterpart, the bovine RSV (BRSV) also constitutes a major cause of respiratory disease in calves, resulting in substantial economic losses to the cattle industry worldwide (3). Despite the substantial health and economic burden caused by RSV illness, there is currently no human vaccine or antiviral drug available. The only significant preventive treatment available is prophylaxis with palivizumab (Synagis), a humanized monoclonal antibody that has provided about 50% protection to high-risk children.
Therefore, there is an urgent need for discovering compounds capable of blocking RSV infection. Protein-protein interactions are potential targets for antiviral chemotherapy (4). The viral RNA-dependent RNA polymerase (RdRp) complex represents an attractive target for drug discovery, because the different components have no cellular ortholog, and are highly conserved between RSV strains. The mechanism of action of this complex involves highly specific and regulated protein-protein and RNA-protein interactions that we need to understand to facilitate drug design approaches.

RSV belongs to the Pneumovirus genus of the Paramyxoviridae family, order Mononegavirales. The RSV genome is a single strand, negative sense RNA of about 15 kb in length that contains 10 transcriptional units encoding 11 proteins. The RSV genomic RNA is packaged by the viral nucleoprotein (N) at all times, forming a N-RNA complex called nucleocapsid. The viral RdRp uses this ribonucleoprotein complex as a template for mRNA transcription and genomic or antigenomic RNA replication (5). The RdRp is composed of two multifunctional proteins, the large protein (L) harboring the capping, methyltransferase, and polyadenylation activities (5) and the tetrameric phosphoprotein (P). The RSV RdRp in transcription mode also associates the viral protein M2-1, which acts as an anti-terminator / elongation factor (6), (7), (8). All the components of the RdRp machinery can be found concentrated into cytoplasmic inclusion bodies (IBs), which are, by analogy with Rhabdoviridae (9), (10), considered as viral factories where viral RNA synthesis take place. Complete atomic structures of N and M2-1 are now available (11), (12), (13). There is no atomic structure resolved, even partially, for P or for L. However, the crystal structure of human metapneumovirus (HMPV) P oligomerization domain, formed by a tetrameric coiled-coil, has recently been resolved (14). Since RSV and HMPV P proteins are highly similar, with 78% of identical residues between RSV and HMPV P proteins, the RSV P oligomerization domain should include at least residues 130-152, where a coiled-coil domain is also predicted (15).

The P protein is the main L cofactor and is essential for the formation of an active polymerase complex, allowing the L protein to gain access to the nucleocapsid where the viral genome is sequestered, P interacting with both the L protein and the N protein simultaneously (16). The P protein was shown to present multiple sites of phosphorylation, at threonine residues 46 and 108, serine residues 30, 39, 45, 54, 116, 117, 119, 156, 161, 232 and 237, and potentially also at Ser 86, 94 and 99 (17), (18), (19), (20), (21), (22), (23), (24), (25), (26). However, the major phosphorylation sites of P are dispensable for RSV replication in vitro (24), and the exact role of phosphorylation for P activity is...
still debated. The RSV P protein forms highly stable tetramers and can be divided into three domains, an N-terminal domain (P_{NTD}, residues 1-120), a central oligomerization domain (P_{OD}, residues 120-160), and a C-terminal domain (P_{CTD}, residues 161-241) (22) (15), (27), (28). P_{NTD} and P_{CTD} are predicted as disordered regions (29), although some short putative α-helices have been predicted between residues 14-25 and 220-228 (27). Although it is now well established that the last 9 C-terminal residues of P_{CTD} are critical for binding to N-RNA complexes (30) (31), a second region encompassing residues 161-180 could also be involved in N-binding (32). However, the L-binding domain of P is still debated. In this work we have investigated the P-L interactions using recombinant proteins. The RSV L protein was expressed using a baculovirus vector. This recombinant L protein was able to bind in vitro to recombinant P purified from bacteria. Serial deletions of P showed that a main L binding domain extend from residue 212 to 239 in the C-terminal region of P. Using site directed mutagenesis, we identified critical hydrophobic residues for P-L interaction, which were also shown to be critical for the function of the polymerase, using a RSV minigenome system. Altogether, our results reveal that the domain (212-239) of P constitutes a molecular recognition element (MoRE).

**Materials and Methods**

**Plasmid constructs for bacterial expression.**

Plasmid pGEX-P was described previously (15). The sequences of P with C-terminal deletions were obtained by introducing stop codons at the appropriate site in the coding sequence of pGEX-P using the QuickChange II Site-Directed Mutagenesis kit (Agilent Technologies, Les Ulis, France). N-terminal deletions of the P sequence were obtained by PCR amplification using Phusion High-Fidelity DNA Polymerase (ThermoScientific) and cloned into pGEX-4T-3 (GE Healthcare Life Sciences) at BamHI-XhoI sites. Internal deletions of the P sequence were obtained by PCR amplification using 5’ phosphorylated internal primer pair as described by Byrappa et al, 1995 (33).

**Expression and purification of recombinant proteins from *E. coli*.** Escherichia coli BL21 (DE3) pLysS chemically competent cells (Novagen, Madison, WI) transformed with pGEX-4T-3 derived constructs were grown at 37 °C for 8 h in 100 ml of Luria-Bertani (LB) medium containing ampicillin (100 µg/ml). The same volume of fresh LB medium was then added, and protein expression was
induced by the addition of 80 µg/ml isopropyl-β-D-thiogalactoside (IPTG) to the medium. The bacteria were incubated for 15 h at 28 °C and then harvested by centrifugation. Bacterial pellets were resuspended in lysis buffer (50 mM Tris-HCl [pH 7.8], 60 mM NaCl, 1 mM EDTA, 2 mM DTT, 0.2 % Triton X-100, 1 mg/ml chicken egg lysozyme) supplemented with a complete protease inhibitor cocktail (Roche, Mannheim, Germany). Lysates were incubated for 1 h on ice, treated for 15 min with 1 U/ml Benzonase nuclease (Novagen, Madison, WI), and centrifuged at 4 °C for 30 min at 10,000 g. Glutathione-Sepharose 4B beads (GE Healthcare, Uppsala, Sweden) were added to clarified supernatants and incubated at 4 °C for 4 h. The beads were then washed two times in lysis buffer, once in phosphate-buffered saline (PBS) and then stored at 4 °C in an equal volume of PBS.

Plasmid constructs for mammalian expression.

Plasmids designed for the expression in BSRT7/5 cells of RSV proteins N, P, M2-1, and L, called pN, pP, pM2-1, and pL were described previously (34) (31). They all contain a T7 transcription promoter, an encephalomyocarditis virus ribosome internal entry site (IRES) to enhance gene expression, and a T7 transcription terminator. The pL-HA and pL-EGFP plasmids were previously described (34). The pM/Luc subgenomic replicon, which encodes the firefly luciferase (Luc) gene under the control of the M-SH gene start sequence, was derived from the pM/SH subgenomic replicon (8) and was described previously (31). Point mutations were introduced into pP by site-directed mutagenesis as described above. For expression of P fragments in BSRT7/5 cells, P fragments were PCR amplified and subcloned into pP in place of the P ORF at KpnI-BamHI sites between the IRES and the T7 terminator. All sequences were from human RSV strain Long, ATCC VR-26 (Genbank accession n° AY911262.1). All constructs were verified by sequencing.

L protein constructs and expression in insect cells. A codon-optimized sequence of the RSV L protein ORF (Long strain) was synthesized (Proteogenix, Schiltigheim, France) and cloned into pFastBac Dual vector (Invitrogen) under the control of the polyhedrin promoter at BamHI-SalI sites. The synthetic L ORF was designed with a HA epitope tag (YPYDVPDYASLGGP) inserted between residues 1738-1739 of the original sequence, and a hexa-histidine tag was added at the C-terminus of the L ORF. For P+L co-expression, a codon-optimized sequence of the RSV P protein ORF (Long Strain) was also synthesized (Proteogenix, Schiltigheim, France) and cloned into the L-ORF.
containing pFastBac Dual vector under the control of the p10 promoter at KpnI-Smal sites. Recombinant baculoviruses were recovered using the Bac-To-Bac Baculovirus Expression System (Invitrogen). A baculovirus expressing no foreign protein (Bac) was used as a control. Sf9 cells from Spodoptera frugiperda were infected at a multiplicity of infection (MOI) of 1 for 72 h, then pelleted and stored at -80 °C.

**Pulldown assays.** Sf9 cells were infected with baculoviruses expressing L (Bac-L) or LNTD (Bac-LNTD) at a MOI of 1 for 72 h, then harvested, resuspended in ice cold lysis buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 10% Glycerol, 0.5% Triton X-100) supplemented with a complete protease inhibitor cocktail (Roche, Mannheim, Germany). Cell lysis was performed with Dounce homogenizer and the lysate was incubated on ice for 1h. Lysates were clarified 30 min at 10,000 g at 4 °C. Beads coated with purified recombinant GST-P proteins (wild-type or mutants) were incubated for 30 min at room temperature with Sf9 lysates. The beads were extensively washed with lysis buffer, then boiled in an equal volume of Laemmli sample buffer, and analyzed by SDS-PAGE and silver staining or immunoblotting.

**Immunoblotting.** Samples were boiled in Laemmli buffer and proteins were resolved by SDS-PAGE and transferred for 2 h at 100 V at 4 °C onto nitrocellulose membranes. The membranes were incubated in blocking solution (PBS-0.1 % [vol/vol] Tween 20 supplemented with 5 % [wt/vol] non-fat milk) for 1 h. Blots were rinsed with PBS-0.1 % [vol/vol] Tween 20 and incubated overnight at 4 °C with primary antibodies in blocking solution. The membranes were rinsed as described above and incubated for 30 min with the appropriate HRP-conjugated secondary antibodies diluted in blocking solution. The membranes were rinsed as described above, and immunodetection was performed using Clarity Western ECL Substrate (Biorad). Western blots were either exposed on a X-ray film or imaged with a GeneGnome machine (Syngene).

**Antibodies.** The following primary antibodies were used for immunofluorescence and/or immunoblotting: a mouse monoclonal anti-P protein antibody (clone 21/2P, kindly provided by José Melero, Madrid), rabbit anti-P and anti-N antisera described previously (15), a mouse monoclonal anti-β-actin antibody (mAbcam 8224), and a rat monoclonal anti-HA antibody conjugated with horseradish
peroxidase (Roche). Secondary antibodies directed against rabbit IgG coupled to Alexafluor-594 (Invitrogen) were used for immunofluorescence. Secondary antibodies directed against mouse and rabbit IgG coupled to horseradish peroxidase (HRP) (PARIS, Compiègne, France) were used for immunoblotting experiments.

Cell culture and transfections. BHK-21 cells (clone BSRT7/5) constitutively expressing a cytoplasmic form of the T7 RNA polymerase (35) were grown in Dulbecco modified essential medium (Lonza, Cologne, Germany) supplemented with 10% fetal calf serum (FCS), 2 mM glutamine, and antibiotics. Cells were transfected using Lipofectamine 2000 (Invitrogen, Cergy-Pontoise, France) as described by the manufacturer. Spodoptera frugiperda (Sf9) cells were cultured at 27 °C in Grace’s insect medium (Gibco) supplemented with 10% foetal calf serum (FCS). Sf9 transfection for baculoviruses recovery was done using lipofectin (Invitrogen).

Minigenome transcription assay. BSRT7/5 cells at 90% confluence in 24-well dishes were transfected with a plasmid mixture containing 0.25 µg of pM/Luc, 0.25 µg of pN, 0.25 µg of pP, 0.125 µg of pL, and 0.06 µg of pM2-1, as well as 0.06 µg of p-β-Gal plasmid (Promega) to normalize transfection efficiencies. The dicistronic subgenomic replicon pM/Luc contains the authentic M-SH gene junction and the Luc reporter gene downstream of the gene start sequence present in this gene junction (36) (37) (34). The assay was performed three times, and each independent transfection was done in triplicate. Cells were harvested at 24 h post-transfection and then lysed in luciferase lysis buffer (30 mM Tris pH 7.9, 10 mM MgCl₂, 1 mM DTT, 1 % Triton X-100, and 15 % glycerol). Luciferase activities were determined for each cell lysate with an Infinite200 Pro microplate reader (Tecan, Männedorf, Switzerland) and normalized to β-Gal expression levels.

Fluorescence microscopy. Immunofluorescence microscopy was performed with cells grown on coverslips and previously transfected with pN, pP and pL-EGFP (34). At 24 h post-transfection, cells were fixed with 4 % paraformaldehyde (PFA) for 25 min, made permeable and blocked for 30 min with PBS containing 0.1 % Triton X-100, 3 % bovine serum albumin (BSA). Cells were then successively incubated for 1 h at room temperature with primary and secondary antibody mixtures diluted in PBS containing 0.3% BSA. For nucleus labeling, cells were incubated with Hoechst 33342 (Invitrogen)
during incubation with secondary antibodies. Coverslips were mounted with Prolong gold antifade reagent (Invitrogen). Cells were observed with a Nikon TE200 microscope equipped with a CoolSNAP ES2 (Photometrics) camera, and images were processed with Meta-Vue (Molecular Devices) software.

Coimmunoprecipitation assay. BSR7/5 cells were infected with modified vaccinia virus strain Ankara expressing T7 RNA polymerase (MVA-T7) (38, 39) at a MOI of 3, 45 min before transfection. Medium was replaced with fresh DMEM and the cells were cotransfected with pP (wild-type or mutant plasmids), and pL-HA with Lipofectamin 2000. After 16h, medium was replaced by PBS containing 2.5 mM EDTA for 5 min at room temperature. Cells were then resuspended and pelleted in microcentrifuge tubes for 5 min at 2000 g. Cells were lysed for 1 h at 4 °C in ice-cold lysis buffer (50 mM Tris HCl pH 7.4, 2 mM EDTA, 1 mM DTT, 150 mM NaCl, 0.5 % Triton X-100, 10 % glycerol) with a complete protease inhibitor cocktail (Roche), and coimmunoprecipitation experiments were performed on cytosolic extracts. Cell lysates were incubated for 1 h at 4°C with an anti-HA monoclonal antibody or Anti-P polyclonal antibody coupled to protein G coated magnetic beads (Dynabeads, invitrogen). The beads were then washed 2 times with lysis buffer and once with PBS, and proteins were eluted in Laemmli buffer at 95°C for 5 min and then subjected to SDS-PAGE and immunoblotting as described above.

Results

Expression of recombinant HA-tagged L protein
Based on previous studies it is known that RSV L protein can be expressed in insect cells using recombinant baculovirus and codon-optimized L sequence (40). However, expression of L was conditioned by the presence of P. We previously determined that a L variant with insertion of the HA epitope tag between residues 1738-1739 in the variable region 2 (VR2) (aa 1718-1764) is functional and can be detected by immunofluorescence, using anti-HA antibodies (34). Based on these results and data, a codon-optimized version of the L open reading frame containing the HA epitope tag inserted between residues 1738-1739 was used to engineer a recombinant baculovirus expression vector (Bac-L). A P/L dual baculovirus expression vector (Bac-L/P) was also engineered for co-
expression of L-HA and P. Lysates from Bac-L-, Bac-L/P- or control Bac-infected cells were analyzed by Western blot using anti-HA and anti-P antibodies. As shown in Fig. 1, P was only observed in Bac-L/P infected cells, as expected. Two major bands were observed for L with lysates of Bac-L-infected cells, one migrating with an apparent mass of ~250 kDa, as expected for full-length L, the other one corresponding to a truncated form (~200 kDa) of L (designated L*). The same bands were observed with lysates of Bac-L/P-infected cells but the expression levels were higher for L in the presence of P, as expected (Fig.1). Analysis of these two bands from a representative gel by excision, trypsin digestion and mass spectrometry, revealed that the two bands contained L polypeptides, and that the smaller L* fragment was truncated at the C-terminus end, after residue 1805 of the HA-tagged L, corresponding to residue 1789 of the original L sequence. The relative abundance of this shorter form of L increased during storage of purified complexes at 4°C. This observation confirmed the high instability of L, and that the presence of P was not sufficient to prevent L cleavage in this system.

In conclusion, although the level of expression of L is lower when it is expressed alone in this system, co-expression of L and P would result in the formation of a P/L complex, which is obviously not suitable for mapping of L-P binding sites. We thus used Bac-L infected cell lysates for further studies.

**The C-terminal region of P contains an L binding site**

We first tried to determine whether the recombinant L produced in baculovirus infected insect cells could bind to P in vitro by using GST pulldown assays in the presence of either full-length P or the truncated forms P[1-126], P[1-161], P[120-160], P[127-241], or P[161-241]. These P fragments were selected according to the molecular organization of RSV P (see Introduction and Fig. 2A). A P construct deleted of the tetramerization domain (P[Δ122-160] or PΔOD) was also included in this first test. Full-length and truncated versions of the P protein fused to GST were expressed in bacteria, purified on glutathione-Sepharose beads and incubated with lysates of Bac-L-infected cells. After extensive wash, complexes were resolved by SDS-PAGE. The presence of L was analyzed by silver staining and Western blot, while GST-P fragments were revealed by Coomassie blue staining. As shown in Fig. 2B, L was pulled-down by GST-P, GST-P[127-241] and GST-P[161-241]. No binding of L was observed when using either PNTD (GST-P[1-126]), POD (GST-P[120-160]) or both associated domains (GST-P[1-160]). Surprisingly, we observed that deletion of the P oligomerization domain
(P[Δ122-160]) significantly increased the efficiency of L binding to P, in a reproducible way. In contrast to what was observed with other constructs, the band corresponding to L was clearly visible after silver staining when GST-P[Δ122-160] was used for these binding assays. Bands corresponding to proteins of molecular weight higher than that of L were also present in all samples. Altogether, these results revealed that recombinant GST-P fusion proteins expressed in bacteria are capable of interacting with the recombinant L present in the Bac-L infected cell lysates, and that an L binding domain is located in PCTD.

**Fine mapping of a L molecular recognition element in P**

Since we obtained stronger L-P binding with P[Δ122-160], we used this oligomerization domain-deleted construct instead of full-length P for further GST pulldown assays and fine mapping of L binding domain(s). This construct was then named Psod. A total of 8 overlapping deletions of 20 or 30 amino acids each were generated throughout the 202 amino acids of Psod protein. Complexes were resolved by SDS-PAGE, GST-P constructs were visualized by Coomassie blue staining, and the presence of L was determined by Western blot. Fig. 3A shows that, among the 8 constructs tested, two of them, GST-Psod[Δ197-226] and GST-Psod[Δ212-241], lost their capacity to bind L. These data confirmed that a L-binding domain is present at the C-terminus of P.

Based on these results and in order to map this L-binding domain with more accuracy, a second series of 12 internal overlapping deletions of 10 amino acids in the C-terminal half of Psod between residues 177-241 were engineered. All the deletions in the C-terminal region extending form amino acids 216 to 241 abrogated P-L interaction (Fig. 3B). These results revealed that the region of P encompassing amino acid residues 216 to 241 is necessary for L binding at least in vitro.

It has been previously determined that the last 9 residues of P are necessary and sufficient for

in vitro binding to ring-shaped N-RNA complexes, the last Phe residue being particularly critical (31), (41). Results obtained with P deletions suggested that L and N-RNA binding domains could overlap.

We thus introduced a series of STOP codons between residues 232-241 of the original P sequence in the GST-Psod construct, and tested the interaction with L by GST pulldown. As shown in Fig. 3C, efficient L binding, similar to what observed with full-length GST-Psod, was recovered with GST-Psod constructs lacking only the two last residues of P (Asp-Phe). These data showed that the domain...
extending from residue Leu$^{216}$ to Leu$^{239}$ of P, thus partially overlapping the N-RNA binding domain, is necessary for efficient binding to L.

**Effect of targeted P gene mutations on RSV polymerase activity**

Formation of a P-L complex is necessary for RNA polymerase activity. We then used a functional test to identify the residues critical for RdRp activity within this 216-239 region. This functional assay consist in using an intracellular plasmid-based mini-replicon system, which has been described previously to study RSV RNA synthesis (31), (34), (12), (41), (13). In this system, the subgenomic replicon pMI/Luc is cotransfected in BSRT7/5 cells expressing T7 RNA polymerase, together with plasmids pCMV-β-gal, pN, pL, pM2-1, and pP (WT or mutants), resulting in the replication and transcription of the minigenome and expression of reporter luciferase. Hence, the production of Luc protein is dependent on these processes. Luciferase activities were determined and normalized based on β-galactosidase expression. In a first attempt to identify conserved residues among Pneumovirus, the C-terminal regions of P proteins of human RSV, BRSV, ovine (ORSV) RSV strains, pneumonia virus of mice (PVM), and canine pneumovirus (CPV) were aligned. Whereas the sequences of PCTD can be aligned until residue 230, two different alignments could be obtained for the most C-terminal 10 residues (Fig. 4A). Compared with all RSV strains, PVM and CPV have an insertion of 7 residues in this region corresponding to RSV P residues 230-241. Thus, from this sequence comparison two stretches of residues can be distinguished, the first one corresponds to RSV P residues 216-230, the second one to RSV P residues 231-241, these two domains being separated by a linker of variable length among pneumoviruses.

Since the residues 233-241 of RSV P constitute an N-RNA binding domain (31), mutagenesis of this region could affect not only L-P but also P-N-RNA binding and therefore RNA synthesis. Based on these data, we restricted site-directed mutagenesis to amino acids located in the 212-232 region. It is worth noting that sequence alignments revealed the presence of two residues highly conserved among pneumoviruses, L216 and L223, and the conservation of two others hydrophobic residues at positions 226 and 227 for RSV P (Fig. 4A). Residues were then substituted by Ala and the effect on RNA polymerase activity was analyzed using the minigenome system. As shown in Fig. 4B, single substitutions L216A, L223A, and L227A reduced RNA polymerase activity to background levels, while L226A had no significant effect. Double (L216A/L223A, L216A/L227A, and L223A/L227A) or triple
(L216A/L223A/L227A) mutants were also generated. In these cases, RdRp activity was no longer detectable. Among the 7 other Ala substitutions that were tested, three of them, K222A, N230A, and D231A, induced a loss of 50% of the polymerase activity, whereas P218A, E221A, N224A and E228A mutations had no visible effect.

Ser and Thr residues representing potential targets for phosphorylation, which could modulate P function and viral RNA synthesis, we substituted T219 and S220 by either Ala (phospho-ablatant) or Asp (phospho-mimetic). S232, the main phosphorylation site of RSV P (20), (21), was also targeted for mutagenesis. Surprisingly, T219A and S220A mutants presented a slight increase of polymerase activity, whereas substitution of S220, and to a lesser extent T219, by Asp had an inhibitory effect on viral polymerase. Thus, although it is not known whether these two residues can be phosphorylated during the natural viral cycle, the presence of a negative charge in this region is deleterious for RSV polymerase activity. In contrast, concerning the main P phosphorylation site S232, no effect on RNA polymerase was observed with S232A and S232D mutants, confirming previous reports (23), (24).

Since the periodicity (i, i+3 and i, i+4) of critical residues identified suggests that this domain could fold into an α-helix (see Fig.7), the N224 residue was substituted by Pro, a known alpha-helical breaker residue. This mutation was shown to reduce RNA polymerase activity to background levels (Fig. 4B), thus supporting the hypothesis of a structured domain of interaction. Finally, as assessed by Western blot analysis, all of the P mutants used in this study were expressed in similar amounts in BSRT7/5 cells (Fig. 4C).

Critical residues of P for interaction with L

To validate that the amino acid residues identified here as critical for RdRp activity were involved in direct P-L interaction, the P mutants generated above, and more specifically L216A, L223A, L227A, single, double or triple mutants, were tested for their ability to bind L protein in mammalian cells. We also studied the interaction of L with PΔ[211-236] as negative control of interaction, and mutants of P N224A or N224P, and S220D. BSRT7/5 cells were cotransfected with pP (WT and mutants) and pL-HA and infected with MVA-T7 to enhance L expression. Before or after immunoprecipitation with P-specific antiserum, proteins were separated and visualized by Western blot (Fig. 5A). The ratio L/P was determined in all cases (Fig. 5B). Although L was still co-immunoprecipitated with P single mutants L216A, L223A, and L227A, L binding was strongly affected.
by P double (L216A/L223A, L223A/L227A, and L216A/L227A) or triple (L216A/L223A/L227A) mutations, by S220D and N224P substitutions, or by deletion of the (211-236) region of P (Fig. 5). On the contrary, N224A mutation had no visible effect. When co-expressed with P mutants for which L-binding was reduced, the L expression levels in cell lysates before immunoprecipitation were also affected, reaching a minimum of ~40% with L216A+L223A P mutant, compared to WT P (Fig.5B). In conclusion, these results point to hydrophobic amino acid residues as critical for P-L interaction, and indicate that L is protected from degradation when stable P-L complexes are formed. Moreover, the loss of interaction of L with the N224P mutant of P suggests that the L molecular recognition element (MoRE) of P could fold into an α-helix when binding to L. These results were also correlated with minigenome assay results.

Mechanisms of recruitment of RSV L polymerase to cytoplasmic inclusion bodies (IBs), where viral proteins involved in RNA synthesis are concentrated and RNA synthesis is believed to take place, are not well understood. Co-expression of RSV P and N proteins is sufficient to induce the formation of IBs similar to those observed in RSV infected cells (42). To investigate whether P mutations affecting P-L interaction could also affect the recruitment of L to IBs, BSRT7/5 cells were cotransfected with pL-EGFP, pN and pP (WT and mutants). IBs were revealed by anti-N antiserum and the presence of L-EGFP in these structures was analyzed by epifluorescence. As shown in Fig. 6, L-EGFP clearly co-localized with IBs when co-expressed with WT P and N proteins. When expressed in the presence of the P single mutant L216A, the L-EGFP fluorescence in IBs was attenuated. In contrast, L-EGFP was not detected in IBs when using the double L216A/L223A mutant of P. Similar results were obtained in the presence of S220D and N224P mutants. It is noteworthy that IBs revealed by anti-N antibodies were observed whatever the P mutant considered (Fig.6 and data not shown). As formation of IBs, depends on N-P interactions (42), this revealed that these P mutants were not affected in their capacity to interact with N.

In conclusion, these results revealed that P residues identified as critical for polymerase activity and located in the 216-232 region are involved in the interaction with the L protein as well as recruitment of L in IBs.

Discussion
For two members of the Mononegavirales order, i.e. vesicular stomatitis virus (VSV), a rhabdovirus, and RSV, a paramyxovirus, it has been determined that the L polymerase is capable of RNA synthesis in vitro, using naked RNA as a substrate, L initiating RNA synthesis de novo at the 3' terminus of the Leader (Le) RNA (43) (40). For VSV it was also demonstrated that P serves as a processivity factor for L, even with naked RNA, and that full processivity of the P–L hetero-complex additionally requires the template-associated N protein (43). In living cells, the genomic RNA of RSV is not naked but wrapped by the N protein forming a N-RNA complex in which RNA is sequestered and hidden. Reading of the RNA sequence by the polymerase needs i/ a specific targeting to this template, which is mediated by the interaction between the nucleocapsid and the C-terminus of P (31), and ii/ the local opening of the N-RNA complex, since RNA is sequestered into a basic groove located between the two α-helical lobes of N (11). It is now well established that, in infected cells, a L-P complex is the minimum functional RdRp (44), P mediating interactions between L and the N-RNA template. Although how N-RNA disassembles before RNA transcription and replication remains unknown, we can speculate that P, within the L-P complex, participates directly or indirectly to this mechanism. Thus, recently the N-terminal domain of mumps virus P has been shown to bind nucleocapsid and appeared to induce uncoiling of the helical nucleocapsid (45). Understanding how P and L interact and function is critical not only to understand mechanisms of viral replication, but also to design compounds capable of interfering with these protein-protein interactions in order to block RSV replication. For RSV, these investigations have been mainly hampered by difficulties to express, visualize or purify the L protein. Recently, a recombinant functional RSV L protein was obtained by using a codon-optimized L gene expressed with a baculovirus vector (40). However, attempts to express and purify L without P were unsuccessful, indicating that the RSV P protein might be necessary to stabilize L. Since our aim was to map the L binding domain of P, we focused our efforts on the expression of L alone. Using a similar strategy, a codon-optimized L gene was designed and expressed with a baculovirus expression vector. Insertion of a HA epitope tag in the variable region 2 of L allowed us to analyze the expression of L in Bac-L infected insect cells by Western blot. Two major bands were observed, one corresponding to full-length L, the other one to a shorter form, migrating with an apparent MW of ~200 kDa and including residues 1-1789 of L. This shorter band was also observed when L was co-expressed with P. These results are in agreement with those published by Noton et al. (40) and...
indicate that the C-terminal region of L located downstream of the variable region 2 is more susceptible to degradation than the rest of the protein, even in the presence of P.

The expression level of L in baculovirus infected insect cells was sufficient to detect it by Western blot, or by SDS-PAGE and silver staining after in vitro binding to GST-P. Since L present in the cell lysates was soluble at low concentration and capable of binding specifically to GST-P, this result indicated that the L protein was correctly folded. We estimate the L concentration to approximately 1 to 10 ng per ml of cell lysate. However, our attempts to concentrate L were unsuccessful, resulting in protein aggregation (data not shown). Furthermore, when L was purified on beads charged with Ni\(^{2+}\), it was not possible to recover it as a soluble form after elution with imidazole.

Our results are in agreement with the data obtained for HPIV3 for which it was shown that in the absence of P, L tends to aggregate during concentration step (46). Thus it is likely that RSV P binds to or modifies domains of L which are responsible for L aggregation at higher concentrations.

We thus used lysates from insect cells expressing this recombinant L protein to map L-binding domain(s) of P using truncated forms of P. We first observed that, in our in vitro system using GST-P fusion proteins purified from E. coli, L-P binding was facilitated when the P oligomerization domain was removed (PΔ[122-160]). In addition, the isolated P oligomerization domain was not able to bind to L. Finally, an L binding domain was mapped to the region encompassing residues 216-239 of P. However, because of instability of the P protein fragments when expressed in mammalian cells (data not shown), it was not possible to validate these results in mammalian cells.

Validation of the P 216-239 region as an L MoRE, and identification of residues critical for the interaction with L was further carried out by conjugating functional assays and co-immunoprecipitation experiments from mammalian cells, using P point mutants. Based on sequence alignment of the C-termini of different paramyxoviruses, we focused on residues 216 to 232 of P. Among the 15 residues substituted and tested using a minigenome assay, 8 were shown to be critical for the polymerase activity. More specifically, mutations of the three hydrophobic residues L216, L223, and L227 by Ala induced a reduction of 80-90% of the polymerase activity. It is also noteworthy that substitution of residues T219 and S220 by Asp, which mimic phosphorylation, partially or totally impairs the polymerase activity, respectively. These data suggest that, although these residues have not previously been identified as P phosphorylation sites, phosphorylation of these residues could be implicated in the regulation of the P-L interaction. In order to study the role of these residues of P in
the interaction with L, we thus performed immunoprecipitation assay of L in the presence of these mutants. Due to low L expression levels in BSRT7/5 cells transfected with T7 driven L expression vector, we used the attenuated recombinant vaccinia virus MVA-T7, which overexpresses the T7 RNA polymerase with reduced cytopathic effects in mammalian cells (38) (39), to increase L expression levels. Using full-length P mutated on Leu residues identified as critical for polymerase activity, we showed that co-immunoprecipitation of L was strongly reduced with these P mutants.

A short α-helix has been predicted in the 220-228 region of P (27). Our results are in agreement with this prediction, since substitution of L216, L223 and L227 by Ala, which are clustered on the same side of the theoretical helix (Fig.7), abolished polymerase activity. Furthermore, substitution of Asn224 by Pro, which is known to break α-helices, abrogated both RdRp activity and L-P interaction. Thus it is likely that this region exists as, or folds into an α-helix when binding to its L partner, and that L-P interaction mainly involves hydrophobic interactions. When the three P mutants, L216A+L223A, S220D, and N224P were substituted to WT P for co-expression with L in mammalian cells, L expression levels were reduced to 40%, 75%, and 56%, respectively, and a strong reduction of P-L interactions was observed by immunoprecipitation, indicating that this putative α-helix is critical for L-P interaction and hence L stabilization. These results were correlated with the absence of L-EGFP in cytoplasmic IBs when co-expressed with the same P mutants. Although stability of L-EGFP was probably also reduced by 25-60% in the presence of the three P mutants, these results suggest that the 216-239 L-binding region of P is also critical for the recruitment of L in cytoplasmic IBs.

Attempts to map the L-binding domain of RSV P have been done previously with human RSV (HRSV) and bovine RSV (BRSV). For BRSV, Khattar et al. mapped the L-binding site to a region of P encompassing amino acids 121–160 (32), which was further identified as the HRSV P oligomerization domain (P_{OD}) (22) (15) (27). On the other hand, Asenjo et al. observed that for HRSV, deletion of the P_{OD} reduced by ~50% but did not abolish its capacity to interact with L, whereas deletion of amino acids 203-241 had a more pronounced negative effect on P-L interaction (47). In both studies, the authors used transfection in HEp-2 cells and infection with a T7 expressing vaccinia virus. Since no anti-L antibodies were available, visualizing the L protein after co-immunoprecipitation with P was done by $^{35}$S-methionine radioactive labeling. Despite this, co-immunoprecipitation of L with P was poor and difficult to quantify. Since P or L sequence conservations are very high between BRSV and HRSV, it is unlikely that BRSV and HRSV use different P-L interaction domains. Our results are more in agreement with this prediction, since substitution of L216, L223 and L227
agreement with those obtained by Asenjo et al., and it is more likely that destabilization of P oligomerization in vivo strongly lowers P—L avidity, which should be compensated by the use of dimeric GST in vitro and high protein concentrations in the GST pulldown assays. Our results also indicate that the presence of GST at the N-terminus of full-length P generates higher order oligomers, which are less adapted for GST pulldowns and protein-protein interaction studies.

Comparison with other Mononegavirales. Due to technical difficulties, few reports describe the mapping and characterization of the L binding domain of P. The L binding domain within the P protein of Sendai virus has been mapped to aa 412-479 (48), in the C-terminal region of P and just downstream the P tetramerization domain extending from amino acid 320 to 429 (49). For rabies virus (RV) and VSV, the L binding region has been mapped to the N-terminal region of P (50) (51) (52). Interestingly, in RV this L binding site recovers the N binding domain which recognizes specifically to the RNA-free form of N called N°. Our results show that for RSV, L- and nucleocapsid-binding domains overlap 7 residues (aa 233-239). It is thus possible that L and N compete for binding to P. Given the respective extents of L and nucleocapsid binding domains of P, which are of 27 and 9 amino acid residues, respectively, L could have a higher affinity for P compared to N. Since L is produced in very small amounts compared to N in RSV-infected cells, this could also promote P-L interactions. However, since RSV P is a tetramer (15), (27), (53), it is also possible that P binds to L and to NC simultaneously, using only one or two protomer(s) for interaction with NC and the other(s) protomer(s) for L binding. However, to address this issue, higher amounts and protein concentration for L are required for in vitro competition assays.

Acknowledgements
We thank David Hughes and Patrick England for critical reading of the manuscript, Damien Vitour (ANSES, Maisons-Alfort) for anti-N polyclonal rabbit serum, José Melero (Madrid) for providing monoclonal antibodies, Bruno Da Costa (INRA, Jouy-en-Josas) for providing Bac baculovirus, and Jean-François Vautherot (INRA, Tours-Nouzilly) for providing MVA-T7.

References


44. Grosfeld H, Hill MG, Collins PL. 1995. RNA replication by respiratory syncytial virus (RSV) is directed by the N, P, and L proteins; transcription also occurs under these conditions but requires RSV superinfection for efficient synthesis of full-length mRNA. Journal of virology 69:5677-5686.


Figure 1. Expression of recombinant RSV HA-tagged L protein in insect cells. Recombinant baculoviruses expressing L-HA alone (Bac-L) co-expressing L-HA and P proteins (Bac-L/P), or expressing no foreign proteins (Bac) were used to infect Sf9 cells. At 72 h pi, cells were lysed and the presence of L-HA and P in the soluble fraction was analyzed by Western blot using anti-HA and anti-P antibodies. L* corresponds to the truncated form of L (residues 1-1789 of the original L sequence). Lysate from mock-infected Sf9 cells (Mock) was also included as a negative control. Beta-actin was used as an internal standard. Western blots were imaged with a GeneGnome machine (Syngene).

Figure 2. A major L binding domain is present in the C-terminal region of RSV P. (A) Schematic diagram of the GST-P fusion protein and of P deletion mutants used for P-L interaction studies. The names of the mutant GST-P proteins indicate the first and last remaining amino acids or deleted amino acids (GST-P(Δ122-160)) and the remaining sequences are shown as solid bars. The N-terminal (P_{NTD}), P oligomerization (P_{OD}) and P C-terminal (P_{CTD}) domains are indicated by arrows at the top of the diagram. (B) P and truncated forms of P fused to GST and purified from bacteria on glutathione-Sepharose beads were incubated with lysates from Sf9 cells infected with the recombinant baculovirus expressing L-HA (Bac-L). After extensive washing, the presence of L-HA was analyzed by SDS-PAGE and silver staining (B, upper panel) or Western blot (B, middle panel) using anti-HA antibodies. The black star on the right of the silver stained gel indicates contaminant proteins of higher molecular weight than L. GST-P proteins were visualized by SDS-PAGE and Coomassie blue staining (lower panel).

Figure 3. Mapping the region of P involved in L binding by serial deletions. (A, B) Internal overlapping deletions of 20 to 30 (A) or 10 (B) amino acids were made throughout GST-P_{OD}. (C) Stop codons were introduced between residues 232 and 241 of the original P sequence. The diagrams of the P deletion mutants used in these assays are depicted on the left of the corresponding gels. Numbers indicate the first and the last residues of the deleted region of P. The presence of L-HA after GST pulldown was analyzed by Western blot using anti-HA antibodies and exposed on X-ray films. GST-P_{OD}-derived complexes were visualized by SDS-PAGE and Coomassie blue staining.
Figure 4. Effect of P point mutations on RSV polymerase activity. (A) Sequence alignments of the C-terminal region of P of the pneumoviruses human RSV (HRSV), bovine RSV (BRSV), ovine RSV (ORSV), pneumonia virus of mice (PVM) and canine pneumovirus (CPV) (accession codes AAX23990.1, NP_048051.1, Q83956.1, Q5MKM7.1, and AHF88957.1 respectively) by Clustal Omega and prepared with ESPript 3. Two possible alignments between the most C-terminal residues of the different P proteins are shown. (B) Analysis of RSV polymerase activity with WT and P substitution mutants. BSRT7/5 cells were transfected with RSV pP, pN, pL, and pM2-1 plasmids and an RSV specific minigenome containing the firefly luciferase reporter gene, together with p-β-Gal constitutively expressing β-galactosidase. Luciferase activity, measured 24 h after transfection, was normalized by β-galactosidase activity, and the luciferase activity gained with WT P set to 100%. The mean value and confidence intervals (error bars) result from 3 separate experiments performed in triplicate. A negative control was run without P. (C) Expression of the different mutants was controlled in cell lysates by Western blotting using polyclonal anti-P rabbit serum after 24 h of transfection and compared to α-tubulin used as an internal standard.

Figure 5. Point mutations affecting the P-L interaction in mammalian cells. BSRT7/5 cells infected with MVA-T7 were transfected with pP (WT and mutants) and pL-HA plasmids. P proteins were immunoprecipitated from cell lysates 16 h after transfection using a rabbit P anti-serum, and the presence of L-HA in the samples was analyzed by Western blot using anti-HA antibodies. Upper panel, cell lysates; lower panel, immunoprecipitated products. (B) Western blots were imaged with a GeneGnome machine (Syngene) and the L/P ratio was calculated before (black bars) or after (grey bars) immunoprecipitation. The L/P ratio is expressed as a percentage of the value obtained for WT P.

Figure 6. Effect of point mutations targeting the L-binding domain of P on the recruitment of L to cytoplasmic inclusion bodies. BSRT7/5 cells were transfected with pP, pN, and pL-EGFP plasmids. Cells were fixed 24 h after transfection and the presence of L-EGFP in cytoplasmic inclusion bodies was observed by epifluorescence. The RSV N protein was revealed by immunofluorescence using rabbit polyclonal anti-N (1:500) and Alexa Fluor 594 goat anti-rabbit (1:2000) antibodies. Nuclei were stained with Hoechst 33342. Scale bars, 10 µm.
Figure 7. Helical wheel projection of the L-binding domain of P encompassing residues 216-234.

(A) Summary of the effects of amino acid substitutions in the 216-239 region of P on RdRp activity, L-P interaction as assessed by immunoprecipitation (CoIP), and on the presence of L in IBs. The amino acid sequence of the 210-241 region of P is indicated at the top. The nucleocapsid binding domain (NCBD) and L binding domain (LBD) are indicated, and the predicted α-helix is represented. In the first column, (-), (+), and (+) indicate RdRp activities of ≤ 5%, ≤ 50%, and ≥ 50%, respectively. (B) Assuming the MoRE of P adopts a helical conformation upon binding of L, we made a helical wheel representation of this region using the Helquest program. This projection shows three leucines (L216, L223, and L227) clustered on one side of the P molecular recognition helix, which most likely constitute the direct interaction interface with L. Hydrophobic residues are in yellow, negatively charged in red, and positively charged in blue. Red N and C letters indicate the N- and C-terminus, respectively. Amino acids for which RdRp activity was reduced to < 50% or < 5% are indicated by one or two stars, respectively. The residue N224, for which substitution by Pro abrogated the polymerase activity, is indicated by the sign (#).
Fig. 2
Fig. 3
Fig. 4
### Table A

<table>
<thead>
<tr>
<th>F mutant</th>
<th>RdRp activity</th>
<th>L CoIP</th>
<th>L in IBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>L216A</td>
<td>-</td>
<td>*</td>
<td>+</td>
</tr>
<tr>
<td>P16A</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>T219A</td>
<td>±</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>T219D</td>
<td>±</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>S220A</td>
<td>±</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>S220D</td>
<td>±</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>E221A</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>K222A</td>
<td>±</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>L232A</td>
<td>-</td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td>N224A</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>N224P</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L226A</td>
<td>±</td>
<td>+</td>
<td>nd</td>
</tr>
<tr>
<td>L227A</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>N230A</td>
<td>±</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>D231A</td>
<td>±</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>S232A</td>
<td>+</td>
<td>+</td>
<td>nd</td>
</tr>
<tr>
<td>S232D</td>
<td>+</td>
<td>+</td>
<td>nd</td>
</tr>
<tr>
<td>L233A</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L234A</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L235A</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### Diagram B

![Diagram](http://jvi.asm.org/)  

**Fig. 7**