STRUCTURAL ANALYSIS OF HUMAN RESPIRATORY SYNCYTIAL VIRUS P PROTEIN: IDENTIFICATION OF INTRINSICALLY DISORDERED DOMAINS

Fernando M. Simabuco[1], John M. Asara[2], Manuel C. Guerrero[2], Towia A. Libermann[2], Luiz F. Zerbini[2], Armando M. Ventura[1*]

[1]Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, Brasil.; [2]Beth Israel Deaconess Medical Center, Harvard Institutes of Medicine, Boston, MA.

ABSTRACT

Human Respiratory Syncytial Virus P protein plus the viral RNA, N and L viral proteins, constitute the viral replication complex. In this report we describe that HRSV P protein has putative intrinsically disordered domains predicted by *in silico* methods. These two domains, located at the amino and caboxi terminus, were identified by mass spectrometry analysis of peptides obtained from degradation fragments observed in purified P protein expressed in bacteria. The degradation is not occurring at the central oligomerization domain, since we also demonstrate that the purified fragments are able to oligomerize, similarly to the protein expressed in cells infected by HRSV. Disordered domains can play a role in protein interaction, and the present data contribute to the comprehension of HRSV P protein interactions in the viral replication complex.

Key words: Human respiratory syncytial virus; P protein; intrinsically disordered domains; oligomerization

INTRODUCTION

The Human Respiratory Syncytial Virus (HRSV) is an enveloped non-segmented negative sense RNA virus that belongs to the *Paramyxoviridae* family, *Pneumovirinae* subfamily (3). HRSV is considered the most important pathogen leading to respiratory disease in infants and neonates worldwide, which might present serious symptoms, like pneumonia and bronchiolitis (6). HRSV genome has about 15,000 nucleotides and contains 10 genes encoding 11 proteins. The P protein or Phosphoprotein has 241 amino acids and is phosphorilated in serines located at positions 116, 117, 119, 232 and 237 (12,15).

The P protein interacts with two other viral proteins: the N protein, which interacts with the viral RNA (vRNA) and forms the nucleocapsid; and the L protein, which is the major subunit of the vRNA polymerase (3). One of P protein functions is to allow specificity of the N protein to the vRNA encapsidation, and another is to confer stability of the L protein in the ribonucleo-complex (3). It is also known that P protein is able to oligomerize in tetramers and the oligomerization domain is located in the central region (9).

Intrinsically disordered proteins lack stable secondary and tertiary structure under physiological conditions and in the absence of a biding ligand (5). Previous studies by Karlin et al. (7) demonstrated that N and P proteins of *Paramyxovirinae*

*Corresponding Author. Mailing address: Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Avenida Professor Lineu Prestes, 1374, São Paulo-SP, Brasil 05508-900.; Tel.: 55 11 3091-7276 Fax: 55 11 3091-7354.; E-mail: amventur@icb.usp.br

subfamily members have intrinsically disordered domains that might be related with protein functions. A model was proposed for the Measles virus replication process based on the flexibility of disordered domains (1), and similar studies have not yet been made for the *Pneumovirinae* subfamily.

We describe here *in silico* predictions of intrinsically disordered regions for the HRSV P protein. The protein was expressed in bacteria and purified as previously described (13) and a degradation pattern of the recombinant protein was observed during the purification steps, however, the protein fragments were still able to form oligomers. Furthermore, we show by mass spectrometry analysis that the degraded protein regions are part of the predicted intrinsically disordered regions and the degradation is not occurring at the oligomerization domain, previously characterized by Llorente *et al.* (9,10).

MATERIALS AND METHODS

Prediction of disordered domains

The approaches used here for predicting intrinsically disordered regions are similar to the ones described by Karlin et al. (7). First, the HRSV (strain A2) P protein sequence (AAC14897) was submitted to Predictor Of Naturally Disordered Regions (PONDR) server (http://www.pondr.com/) using VL3 method. In a second approach, P protein sequence was submitted to Hydrophobic Cluster Analysis (HCA) using the DRAWHCA software (http://www.lmcp.jussieu.fr/~soyer/www-hca/hca-form.html). The prediction of secondary structure by PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html) was also used to show the presence or lack of secondary structure.

Protein expression, purification and analysis

P protein was expressed in bacteria fused with Maltose Binding Protein (MBP) and purified using amylose resin column. After digestion with enterokinase to cleave the fused proteins, we observed a degradation pattern for P protein as previously reported (13). Despite the degradation, anion exchange chromatography (AEC) was performed using an AKTÄ Fast Protein Liquid Chromatography System (Amersham) to separate MBP from P protein fragments. Digested MBP-P protein was applied onto a HiTrap DEAE Fast Flow column (Amersham). The elution was performed using a NaCl gradient combining Buffer A (100 mM TrisHCl, pH 8.2, 25 mM NaCl and 1mM EDTA) and Buffer B (Buffer A plus 1 M NaCl). Fractions were collected and analyzed in SDS-PAGE.

Western Blot analysis under non-denaturing conditions was performed after PAGE without SDS. The proteins were transferred to Hybond-ECL Nitrocellulose membrane (Amersham), which was blocked and incubated with anti-P monoclonal antibody C771 (11). The detection of the primary antibody was performed using an anti-mouse peroxidase-conjugated antibody (KPL). Native P protein was extracted, under non-denaturing conditions, from HEp-2 cells infected with HRSV at a multiplicity of infection (MOI) of 10.

Mass spectrometry analysis

For the mass spectrometry analysis of the degraded P protein fragments, the following protocol was performed. After SDS-PAGE, the Coomassie blue stained gel bands were excised, washed with acetonitrile 50% and cut. Cysteine residues were reduced with 10 mM dithiothreitol and alkylated with 10 mM iodoacetamide. Gel slices were then washed and digested with 250 ng of modified trypsin (Promega) in 50 mM ammonium bicarbonate (pH 8.3) overnight at 37°C. Peptides were extracted from the gel pieces with 20 mM ammonium bicarbonate, followed by 40% acetonitrile and 2% formic acid. An aliquot of peptide mixture was applied onto a nanoflow reversed-phase liquid chromatography tandem mass spectrometry system at a flow rate of about 300 nL/min. Data dependent MS/MS spectra were collected using a Thermo Scientific LTQ 2D linear ion trap mass spectrometer. MS/MS spectra were searched versus a database containing the HRSV P protein sequence using the SequestÒ algorithm.

RESULTS

As we can see in Fig. 1A, only a small segment of the sequence was predicted by PONDR as ordered, indicating that

P protein has unstructured properties in both N and C-terminals. The HCA results are shown in Fig. 1B. The first feature that is noteworthy is the high density of amino acids that promote disorder, such as proline, glycine and serine (4, 5), between amino acids 1 to 100 (marked with a black square). Furthermore, the same region, as well as the region comprising amino acids 200 to 241, has small quantity of hydrophobic clusters, indicating lack of structure, since hydrophobic clusters tend to form secondary structure elements (5). On the other hand, the region comprising amino acids 101 to 200 has several hydrophobic clusters, indicating a structured region or even a domain. This region has already been reported and experimentally tested to be rich in alpha-helices and involved in the P protein oligomerization (9, 10).

After expression in bacteria fused with MBP and digestion with enterokinase, P protein was separated from MBP by AEC. All degraded P protein fragments, presenting molecular weights smaller than the one expected for P protein (27.2 kDa), were eluted together in a single peak (Fig. 2A fractions 7-8; Fig. 2B lanes 4 and 5). A Western Blot analysis under non-denaturating conditions of the degraded P protein fragments was performed. As seen in Fig. 2C, lane 1, even though the recombinant P protein was degraded, it was still able to form oligomers that reached, approximately, the same size of tetramers formed by native P protein (Fig. 2C, lane 2). The slightly lower molecular weight observed in P protein fragments oligomers can be attributed to the degradation and the fact that native P protein, expressed in infected eukaryotic HEp-2 cells, is phosphorilated.
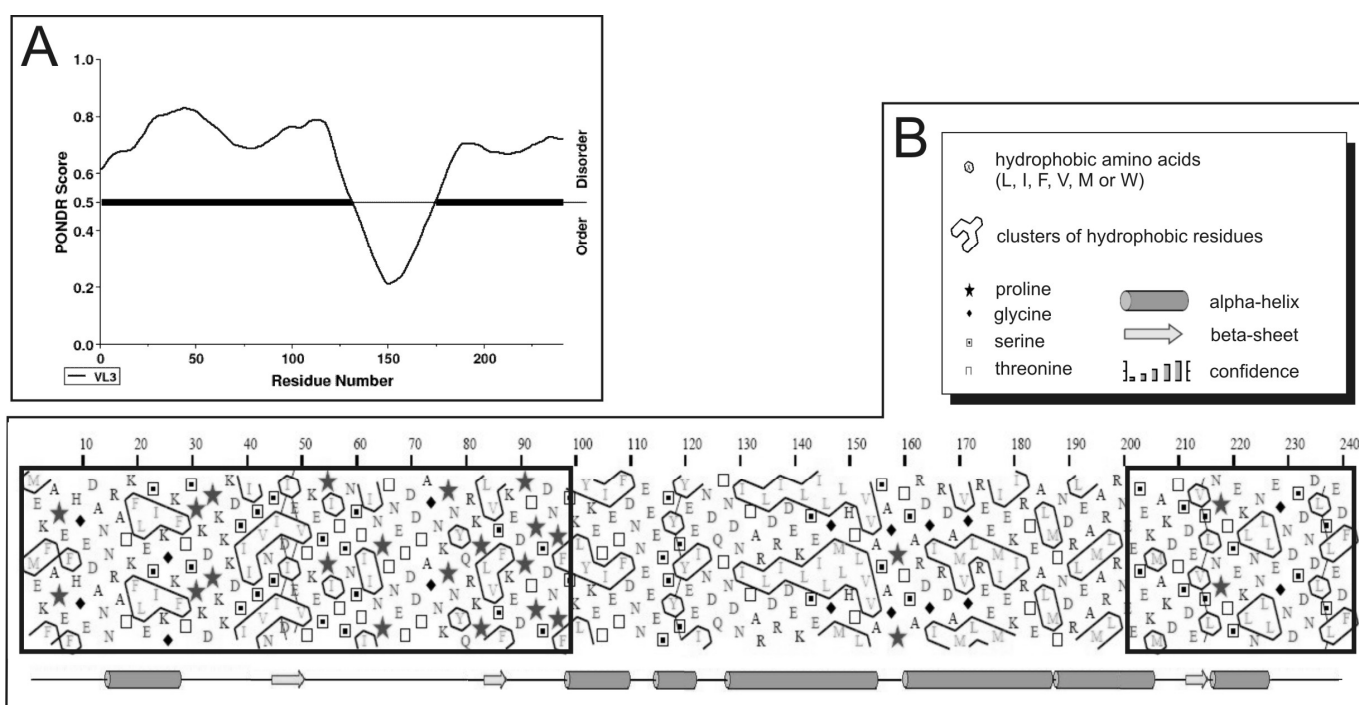

Figure 1. Prediction of disordered domains for HRSV P protein using different methods. (A) HRSV strain A2 P protein sequence was submitted to PONDR server using VL3 algorithm. The regions marked with dark lines were predicted to be disordered (B) Hydrophobic Cluster Analysis was performed using DRAWHCA software. Hydrophobic amino acids were grouped in clusters and special amino acids like proline, glycine, serine and threonine are distinguished from others. Prediction of secondary structure elements, such as alpha-helices and beta-sheets, was made by PSIPRED and is shown below HCA. The black squares delimitate the putative intrinsically disordered domains.

After SDS-PAGE of the degraded P protein, a higher and a lower molecular weight band (bands A and B in Fig. 2B, lanes 4 and 5) were selected, excised from the gel and digested with trypsin. Peptides were extracted and submitted to mass spectrometry analysis and the obtained data aligned with P protein sequence. The results are presented in Fig. 3 and reveal that P protein degradation is occurring in N-terminal region, indicated by lack of peptide matches in that region for both bands (Fig. 3, panels A and B). For the smaller band (Fig. 3, panel B), the degradation is more extensive in the N-terminal and a C-terminal degradation can also be seen. The C-terminal portion was still present in the bigger band (Fig. 3, panel A), detected by a non interrupted match of peptides. In the same way, the region between amino acids 50 to 100 is present in both bands and was also detected by non interrupted peptide matches. These results indicate that N and C-terminal regions contain flexible structures, more exposed to the trypsin digestion, that overlap with the same predicted as disordered by PONDR (Figure 1A). They also present few hydrophobic clusters predicted by HCA, indicating lack of secondary structure (6), also evidenced by PSIPRED secondary structure prediction (amino acids 1-100 and 201-241, Figure 1B).

On the other hand, the region between amino acids 101 and 200 has a drop in the peptide matches, with some gaps, indicating a region that is well structured and more resistant to the trypsin digestion. This region overlaps the ordered region predicted by PONDR (Figure 1A). It also contains high density of hydrophobic clusters predicted by HCA, indicating presence of secondary structure, also evidenced by PSIPRED secondary structure prediction (amino acids 101-200, Figure 1B).
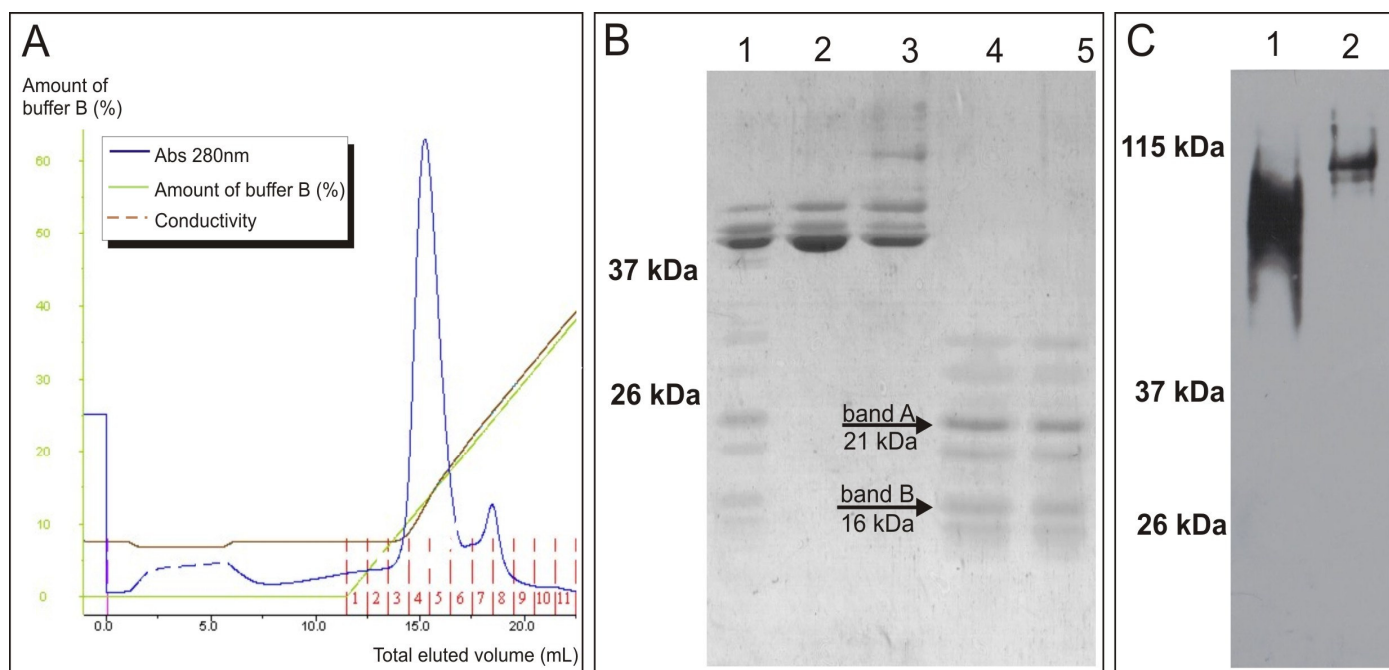

Figure 2. Oligomerization of P protein purified by anion exchange chromatography (AEC). (A) AEC protein elution profile. Absorbance at 280 nm is presented by a dark line, conductivity is presented by a dark dashed line, and concentration of buffer B is presented by a grey line. Elution fractions are delimitated and marked with numbers in the graphic. (B) SDS-PAGE of AEC elution fractions (Coomassie Blue staining). 1: MBP-P protein digested with enterokinase that was initially applied into the column. 2: fraction 4. 3: fraction 5. 4: fraction 7. 5: fraction 8. (C) Western Blot analysis under non-denaturing conditions. 1: P protein obtained from MBP-P enterokinase digestion and purified by AEC (fraction 7, from panel A). 2: Protein extract from HEp-2 cells infected with HRSV strain A2.
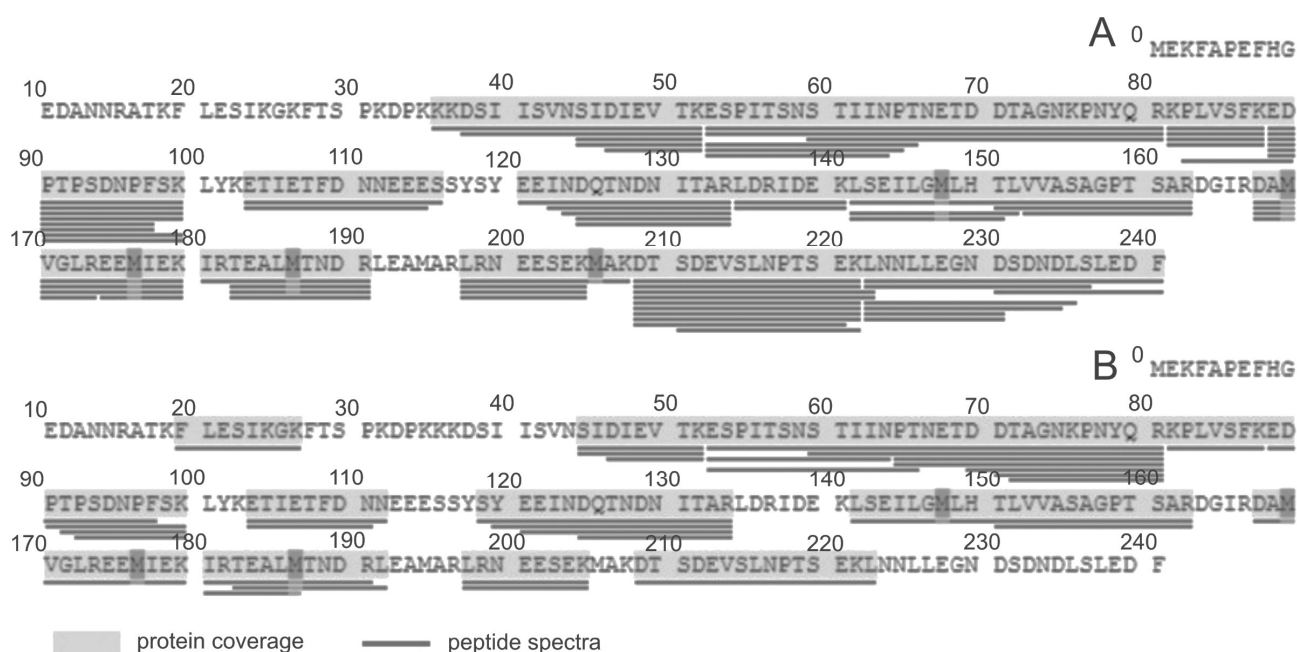
Figure 3. Mass spectrometry analysis of P protein degradation. The bands A and B (fig. 2B) were cut from poliacrilamide gel and digested with trypsin. Peptides were extracted and analyzed in a mass spectrometer and peptide sequences found were aligned with HRSV P protein. The peptides coverage is shown in P protein sequence by a grey marker, and their match frequencies are shown by dark lines under P sequence. (A) Mass spectrometry analysis of band A. (B) Mass spectrometry analysis of band B.

DISCUSSION

Previous data characterized *in vitro* an oligomerization domain in P protein central region and indicated that P oligomers could also occur in HRSV infected cells (2, 9, 10). The *in silico* data presented here indicate that the P protein degradation, previously reported by our group (13), is not occurring in the region of the oligomerization domain (Figure 1). It is well established that intrinsically disordered proteins are more sensitive to proteases (5) and this hyper sensibility is an experimental support to the identification of intrinsically disordered domains (7). The lack of a rigid structure might have exposed the sequences near the N and C-terminals of P protein, expressed in bacteria in fusion with MBP, to residual protease activity after digestion with enterokinase to cleave the fused proteins, causing the observed degradation (Figure 2). The regions near the N and C-terminals also have an increased sensitivity to trypsin digestion in the mass spectrometry

analysis, since they generate more peptide matches (Figure 3), indicating that they are part of unstructured domains.

There are several examples of proteins that have intrinsically disordered domains with defined functions (4, 14). Most of these proteins make interactions with other proteins, as the intrinsically disordered domains allow interactions with high specificity but low affinity (5). In a detailed molecular model proposed for the Measles virus replication complex, the N and P proteins intrinsically disordered regions flexibility are used to explain the interactions needed for the replication process to occur (1). A study with Bovine RSV, closely related to HRSV virus, shows that the regions between amino acids 161-180 and 221-241 interact with N protein (8). For HRSV P protein the disordered C-terminal structure might also be important for the interaction with N protein in the viral replication context, were it could folds upon binding to N protein, undergoing a disorder-to-order transition. The detailed characterization of the role of these intrinsically disordered

domains, however, demands more studies.

ACKNOWLEDGEMENTS

REFERENCES

1. Bourhis, J.M.; Canard, B.; Longhi, S. (2006). Structural disorder within the replicative complex of measles virus: functional implications. *Virology*. 344, 94-110.

2. Castagné, N.; Barbier, A.; Bernard, J.; Rezaei, H.; Huet, J.C.; Henry, C.; Da Costa, B.; Eléouët, J.F. (2004). Biochemical characterization of the respiratory syncytial virus P-P and P-N protein complexes and localization of the P protein oligomerization domain. *J. Gen. Virol.* 85, 1643-1653.

3. Collins, P.L.; Chanock, R.M.; Murphy, B.R. (2001). Respiratory syncytial virus. *In*: Fields, B.N., Howley, P.M., Griffin, D.E., Lamb, R.A., Martin, M.A., Roizman, B., Straus, S.E., Knipe, D.M. (eds). *Fields Virology*. Lippincott Williams & Wilkins, Philadelphia, USA, p. 1443-1485.

4. Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; Ausio, J.; Nissen, M.S.; Reeves, R.; Kang, C.; Kissinger, C.R.; Bailey, R.W.; Griswold, M.D.; Chiu, W..; Garner, E.C.; Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26-59.

5. Ferron, F.; Longhi, S.; Canard, B.; Karlin, D. (2006). A practical overview of protein disorder prediction methods. *Proteins*. 65, 1-14.

6. Holberg, C.J.; Wright, A.L.; Martinez, F.D.; Ray, C.G.; Taussig, L.M.; Lebowitz, M.D. (1991). Risk factors for respiratory syncytial virus-associated lower respiratory illnesses in the first year of life. *Am. J. Epidemiol.* 133, 1135–1151.

7. Karlin, D.; Ferron, F.; Canard, B.; Longhi, S. (2003). Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* 84, 3239-3252.

8. Khattar, S.K.; Yunus, A.S.; Samal, S.K. (2001). Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N-P and P-L interactions using a minigenome system. *J. Gen. Virol.* 82, 775-779.

9. Llorente, M.T.; García-Barreno, B.; Calero, M.; Camafeita, E.; López, J.A.; Longhi, S.; Férron, F.; Varela, P.F.; Melero, J.A. (2006). Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. *J. Gen. Virol.* 87, 159-169.

10. Llorente, M.T.; Taylor, I.A.; López-Viñas, E.; Gomez-Puertas, P.; Calder, L.J.; García-Barreno, B.; Melero, J.A. (2008). Structural properties of the human respiratory syncytial virus P protein: evidence for an elongated homotetrameric molecule that is the smallest orthologue within the family of paramyxovirus polymerase cofactors. *Proteins*. 72, 946-958.

11. Orvell, C.; Norrby, E.; Mufson, M.A. (1987). Preparation and characterization of monoclonal antibodies directed against five structural components of human respiratory syncytial virus subgroup B. *J. Gen. Virol.* 68, 3125–3135.

12. Sánchez-Seco, M.P.; Navarro, J.; Martínez, R.; Villanueva, N. (1995). C-terminal phosphorylation of human respiratory syncytial virus P protein occurs mainly at serine residue 232. *J. Gen. Virol.* 76, 425-430.

13. Simabuco, F.M.; Carromeu, C.; Farinha-Arcieri, L.E.; Tamura, R.E; Ventura, A.M. (2007). Production of polyclonal antibodies against the human respiratory syncytial virus nucleoprotein and phosphoprotein expressed in *Escherichia coli*. *Protein Expr. Purif.* 53, 209-215.

14. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*. 41, 415-427.

15. Villanueva, N.; Hardy, R.; Asenjo, A.; Yu, Q.; Wertz, G. (2000). The bulk of the phosphorylation of human respiratory syncytial virus phosphoprotein is not essential but modulates viral RNA transcription and replication. *J. Gen. Virol.* 81, 129–133.