MASTER IN BIOLOGY

Development of a method to identify RNA-binding proteins:Study of the transcription and translation mechanisms among Schmallenberg virus

Frau, Lisa

Award date:
2016

Awarding institution:
University of Namur

Faculté des Sciences

Development of a method to identify RNA-binding proteins:
Study of the transcription and translation mechanisms among Schmallenberg virus

Mémoire présenté pour l'obtention
du grade académique de master 120 en biochimie et biologie moléculaire et cellulaire

Lisa FRAU

Janvier 2016

Université de Namur
FACULTE DES SCIENCES

Université de Namur
FACULTE DES SCIENCES
Secrétariat du Département de Biologie
Rue de Bruxelles 61 - 5000 NAMUR
Téléphone: + 32(0)81.72.44.18 - Téléfax: + 32(0)81.72.44.20
E-mail: joelle.jonet@unamur.be - http://www.unamur.be

Development of a method to identify RNA-binding proteins:

Study of the transcription and translation mechanisms among Schmallenberg virus

FRAU Lisa

Abstract

After genomic sequencing analysis, Schmallenberg virus was categorized as a member of *Bunyaviridae* family, the *Orthobunyavirus* genus and the *Simbu* serogroup. Its tripartite genome is composed of negative sense RNA segments (L, M and S), encoding 6 proteins in total. These genomic segment, consisting in a coding region flanked by 3' and 5' UTR, are transcribed into mRNAs, constituted of a 5' host-derived sequence and lacking a poly(A) tail at the 3' end. However, poly(A) tail is crucial for the mRNA stability, the transcription termination and the mRNA translation initiation processes.

In silico analyses recently revealed a conserved stem-loop and a putative transcription termination signal both localized in the 3' UTR of the SBV antigenomic S segment, indicating a potential functional relevance in viral replication. Our objective was to develop an RNA-affinity purification technique coupled with mass spectrometry analysis (RAP-MS) to identify the RNA-binding proteins interacting with these two elements.

In brief, the RNA bait, containing both elements of interest, was immobilized on a solid support, in a reversible manner. As the virus is described to replicate in the cytoplasm, cytosolic extracts were incubated with the immobilized RNA. After washing steps to eliminate non-specifically bound proteins, the ribonucleoproteins complexes were specifically eluted from the solid matrix and subjected to mass spectrometry.

Although some results suggest that the conditions settled for protein capture on the RNA are not optimal, several proteins described to play a role in viral transcription/ translation/ replication processes identified by RAP-MS. However, these results have to be taken with caution since they are based on only one experiment. Therefore, the RAP-MS experiment has to be repeated with independent replicates.

If successful, this technique would confirm the presumed role of the two conserved elements in SBV transcription/translation processes and would highlight the proteins contributing to those mechanisms. More generally, it should provide further insight on transcription termination and translation initiation processes in the absence of poly(A) tail.

Master thesis (120) in biochemistry and molecular and cellular biology

January 2016

Supervisor: P. Renard **Co-supervisor :** B. Muylkens

Université de Namur
FACULTE DES SCIENCES
Secrétariat du Département de Biologie
Rue de Bruxelles 61 - 5000 NAMUR
Téléphone: + 32(0)81.72.44.18 - Téléfax: + 32(0)81.72.44.20
E-mail: joelle.jonet@unamur.be - http://www.unamur.be

Développement d'une méthode permettant l'identification de proteins se liant à l'ARN: application à l'étude des mécanismes de transcription et traduction chez le virus de Schmallenberg

FRAU Lisa

Résumé

Suite à une analyse génomique, le virus de Schmallenberg (SBV) a été classifié en tant que membre de la famille *Bunyaviridae*, du genre *Orthobunyavirus* et du sérogroupe *Simbu*. Son génome se compose de trois segments (L, M et S) d'ARN négatif simple brin, encodant 6 protéines au total. Ces segments génomiques, qui consistent en une séquence codante flanquée par des régions non transcrites (3' et 5' UTR), sont transcrits en ARNm, constitué d'une sequence dérivée de l'ARNm de l'hôte en 5' mais dépourvu de queue poly (A) à leur extrémité 3'. Or celle-ci est cruciale pour la stabilité de l'ARNm et les processus de terminaison de la transcription et d'initiation de la traduction.

Des analyses *in silico* ont révélé la présence d'une tige-boucle et d'un signal de fin de transcription potentiel dans la région 3' UTR du segment S antigénomique de SBV. De par leur localisation et leur conservation dans le sérogroupe *Simbu*, ces deux éléments sont suspectés d'intervenir dans la réplication virale. Notre objectif était de développer une technique de purification par affinité couplée à une analyse par spectrométrie de masse (RAP-MS) afin d'identifier les protéines interagissant avec ces deux éléments.

En bref, l'ARN d'intérêt, contenant à la fois la tige-boucle et le signal potentiel de terminaison de transcription, a été immobilisé sur un support solide, d'une manière réversible. Etant donné la réplication exclusivement cytoplasmique du virus, les extraits cytosoliques ont été incubés avec l'ARN immobilisé. Après avoir éliminé les protéines liées non spécifiquement à l'ARN par de multiples lavages, les complexes ribonucleoprotéiques ont été élués de la matrice solide et soumis à l'analyse par spectrométrie de masse.

Bien que les conditions établies pour la capture des protéines sur l'ARN ne soient pas optimales, plusieurs protéines connues pour participer aux mécanismes de transcription/ traduction/ réplication virale ont été identifiées par RAP-MS. Cependant, ces résultats doivent être considérés avec précaution car ils ne se basent que sur une seule expérience et doivent être confirmés par réplicats indépendants.

Si cette méthode est couronnée de succès, elle confirmera le rôle présumé des deux éléments dans les mécanismes de transcription/traduction et mettra en évidence les protéines contribuant à ces processus. D'une manière plus générale, cette technique devrait fournir davantage d'indications sur les mécanismes de fin de transcription et d'initiation de la traduction en absence de queue poly(A).

Mémoire de master 120 en biochimie et biologie moléculaire et cellulaire

Janvier 2016

Promoteur: P. Renard **Co-promoteur :** B. Muylkens.

Remerciements

Au terme de 10 mois de recherche appliquée et de 5 années d'études qui touchent à leur fin, quelques remerciements aux personnes qui ont contribué de près ou de loin à la réalisation de ce mémoire et à la réussite de ces 5 années s'imposent.

Mes premières pensées vont aux Seniors de l'URVI et de l'URBC : merci pour m'avoir accueilli au sein de votre département durant l'élaboration de ce mémoire. Parmi eux, je voudrais particulièrement remercier Mr Muylkens et Mme Renard, pour m'avoir ouvert les portes de leur laboratoire, pour la confiance qu'ils m'ont accordée, la conception de ce projet de recherche mais également leur talent en tant que co-promoteur et promotrice. Merci à Mr Muylkens, pour son humanité et son optimisme résistant à toute épreuve. A Mme Renard, merci pour votre disponibilité, votre patience, votre soutien et vos nombreux encouragements. Travailler dans vos équipes fut un réel plaisir pour moi et cette expérience n'a fait que confirmer mon intérêt pour la recherche.

La réussite d'un mémoire dépend à la fois d'un cadre de travail optimal et d'une bonne ambiance au sein d'un groupe. Pour ceci et leurs remarques constructives, j'aimerais exprimer ma gratitude envers tous les membres de l'URVI et de l'URBC.
Merci à Damien pour le partage de ses connaissances, son enseignement et sa patience. Merci à Laetitia et Isabelle, pour leur sympathie, leur humour et pour avoir fait régner la bonne humeur au sein de l'URVI. Evidemment, je ne peux oublier Antoine, envers qui j'aimerais témoigner ma reconnaissance pour m'avoir appris la rigueur et la précision. Merci pour ta disponibilité, tes conseils avisés, ta générosité et ta bonne humeur communicative. Merci à Maud, pour le temps qu'elle m'accordé et ses précieuses explications, à Elodie pour son aide durant ma première IF et à Noëlle pour l'analyse des IF. Merci à Marc pour s'être attelé à l'analyse des protéines en spectrométrie de Masse.

Un merci tout particulier aux mémorantes de l'URVI et de l'URBC, pour m'avoir appris la solidarité et le travail d'équipe. Merci pour avoir partagé mes joies et mes échecs et pour cette complicité qui nous a unies durant ces 10 derniers mois. Sans vous, ce mémoire n'aurait pas été le même. Merci à tous mes collègues BBMC, et ex-biomeds que j'ai eu la chance de côtoyer durant ces dernières années pour leur amitié.

Les dernières lignes sont réservées à mes parents que j'aimerais remercier de tout cœur pour leur éducation, leur soutien aux cours de ces années et pour m'avoir apporté tout ce qu'un enfant peut espérer recevoir. Vous rendre fiers restera mon premier objectif. Enfin, merci à Jérôme pour sa présence, son réconfort, ses distractions et son soutien.

Lisa.

Abbreviations

ADAR: Adenosine Deaminase Acting on RNA
AINOV: Aino Virus
AKAV: Akabane Virus
APOBEC: Apolipoprotein B mRNA Editing Enzyme, Catalytic polypeptide-like
ARE: AU-Rich Element
AUF1: AU-rich element RNA-binding protein 1
BHK-21: Baby Hamster Kidney cells
BSA: Bovine Serum Albumin
BUNV: Bunyamwera virus
$CaCl_2$: Calcium Chloride
CBF2: C promoter Binding Factor 2
CDAR: Cytosin Deaminase Acting on RNA
cDNA: complementary DNA
CHAPS: 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CIP: Calf Intestinal Phosphatase
CNS: Central Nervous System
CTA: C-terminal Arm
CTD: C-terminal Domain
cRNA: complementary RNA
Da: dalton
DAP-MS: DNA-affinity purification followed by mass spectrometry
DC-SIGN: Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
DDX3: DEAD-box RNA helicase 3
DMEM: Dulbecco's Modified Eagle Medium
DNA: Deoxyribonucleic Acid
dNTP: Deoxynucleotide Triphosphate
DTT: Dithiothreitol
DOUV: Douglas Virus
E. Coli: Escherichia Coli
EBNA2: Epstein-Barr Nuclear Antigen 2
EBV : Epstein Barr Virus
EDTA : Ethylene Diamine Tetra Acetic acid
eEF-1a: Eukaryotic Elongation Factor-1α
eIFs: Eukaryotic Initiation Factors
eIF2S: Eukaryotic Translation Initiation Factor 2
ENS: Non specific elution
ER: Endoplasmic Reticulum
ES: Specific Elution
ESb: beads from Specific Elution
ESI: Electrospray Ionization
EV71: Enterovirus 71
FBS: Fetal Bovine Serum
FBXO18: F-box Only Protein 18
FDR: False Discovery Rate
Gc: C-terminal Glycoprotein
GCN1L1: translational activator GCN-1 like protein
GMEM: Glasgow's Minimal Essential Medium
Gn: N-terminal Glycoprotein
gRNA: genomic RNA
H_2PO_4: dihydrogen phosphate
HB: Hypotonic Buffer
HBV: Hepatitis B Virus

HCl: Hydrogen Chloride
HCV: Hepatitis C Virus
HEPES: (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV: Human Immunodeficiency Virus
hnRNP: heterogeneous nuclear ribonucleoprotein
hnRNP A/B: heterogeneous nuclear ribonucleoprotein A/B
hnRNP A3: heterogeneous nuclear ribonucleoprotein A3
hnRNP D: heterogeneous nuclear ribonucleoprotein D
hnRNP E: heterogeneous nuclear ribonucleoprotein E
hnRNP K: heterogeneous nuclear ribonucleoprotein K
hnRNP U: heterogeneous nuclear ribonucleoprotein U
IBDV: Infectious Bursal Disease Virus
IPTG: Isopropyl β-D-1-thiogalactopyranoside
IRES: Internal Ribosome Entry Site
JEV: Japanese Encephalitis Virus
KCl : Potassium Chloride
KH_2PO_4 : Monopotassium Phosphate
L segment: Large segment
LACV: La Crosse encephalitis Virus
LB medium: Lysogeny Broth medium
LC: Liquid Chromatography
LC-MS: Liquid Chromatography – Mass Specrometry
Lsm: Like sm
M: Molar
M segment: Medium segment
m^7G cap: 7-methylguanylate cap
MCS: Multiple Cloning Site
mg: milligram
miRNAs: microRNAs
mM: millimolar
MOI: Multiplicity Of Infection
mRNA: messenger RNA
ms : milliseconds
MS: Mass Spectrometry
MSMS: tandem Mass Spectrometry
MX1: Interferon-induced GTP-binding protein
N protein : Nucleocapsid protein
Na_2MoO_4 : Sodium Molybdate
NaCl : Sodium Chloride
NaF: Sodium Fluoride
NaOH: Sodium Hydroxyde
NEF protein: Negative Regulatory Factor protein
NEP/NS2: Nuclear Export Protein
ng: nanogram
NH_4HCO_3 : Ammonium bicarbonate
nM: nanomolar
NP-40: Nonidet P40
NS1: Non-Structural protein 1
NS1-BP: NS1-Binding protein
NSm: Non-Structural protein NSm
NSs: Non-Structural protein NSs
NTA : N-terminal Arm
NTD : N-terminal Domain
NTP: Nucleotide Triphosphate
NUP98: Nuclear Pore Complex Protein 98

nusG: transcription termination protein nusG
OAS1/2: 2'-5'-Oligoadenylate Synthetase 1
PABP: Poly(A)-Binding Protein
PAF: Paraformaldéhyde
PBS: Phosphate Buffered Saline
PBS-EDTA: Phosphate Buffered Saline – Ethylene Diamine Tetra Acetic acid
PBS-BSA: Phosphate Buffered Saline – Bovine Serum Albumine (PBS-BSA)
PCBP: Poly(rC) Binding Protein
PCR: Polymerase Chain Reaction
PEAV: Peaton Virus
pmoles : picomoles
Poly(A) tail: Polyadenylation tail
PPS: (3-[3-(1,1-bisalkyloxyethyl)pyridin-1-yl]propane-1-sulfonate) Silent Surfactant
PTB: Polypyrimidine Tract-Binding protein
PTR: Polymerase I and Transcript Release factor
PVDF: Polyvinylidene difluoride
Q-TOF: Quadrupole - Time of flight
RAP-MS: RNA-Affinity Purification followed by Mass Spectrometry analysis
RBP: RNA-Binding Proteins
RdRp: RNA-dependent RNA polymerase
RNA: Ribonucleic Acid
RNP complexes : Ribonucleoprotein complexes
RPA2: Replication Protein A
RPL31 : Ribosomal Protein L31
RPM : Revolution Per Minute
RTC: Reverse Transcription Complex
rRNA: ribosomal RNA
RSS: RNA Secondary Structure
S segment: Small segment
S-ENS: Supernatant - non specific elution
S-ES: Supernatant – specific elution
SABOV: Sabo Virus
SANV: Sango Virus
SATV: Sathuperi Virus
SBV: Schmallenberg Virus
SDS: Sodium Dodecyl Sulfate
SHAV: Shamonda Virus
SIMV: Simbu Virus
siRNAs: Small Interfering RNAs
SPL: Scheduled Precursor List
tRNA: transfer RNA
TRIS: trishydroxyméthylaminométhane/ 2-amino-2-hydroxyméthyl-1,3-propanediol
UBC: Ubiquitin C
UTR: Untranslated Region
VSV: Vesicular Stomatitis viruses
WNV: West Nile virus
WT: Wild-Type
X-gal: 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
YB-1: Nuclease-sensitive element-binding protein 1
ZNFX1: NFX1-type Zinc Finger-containing protein 1
°C: Celsius degree
µg: microgram
µM: micromolar
µL: microliter

Content

Introduction

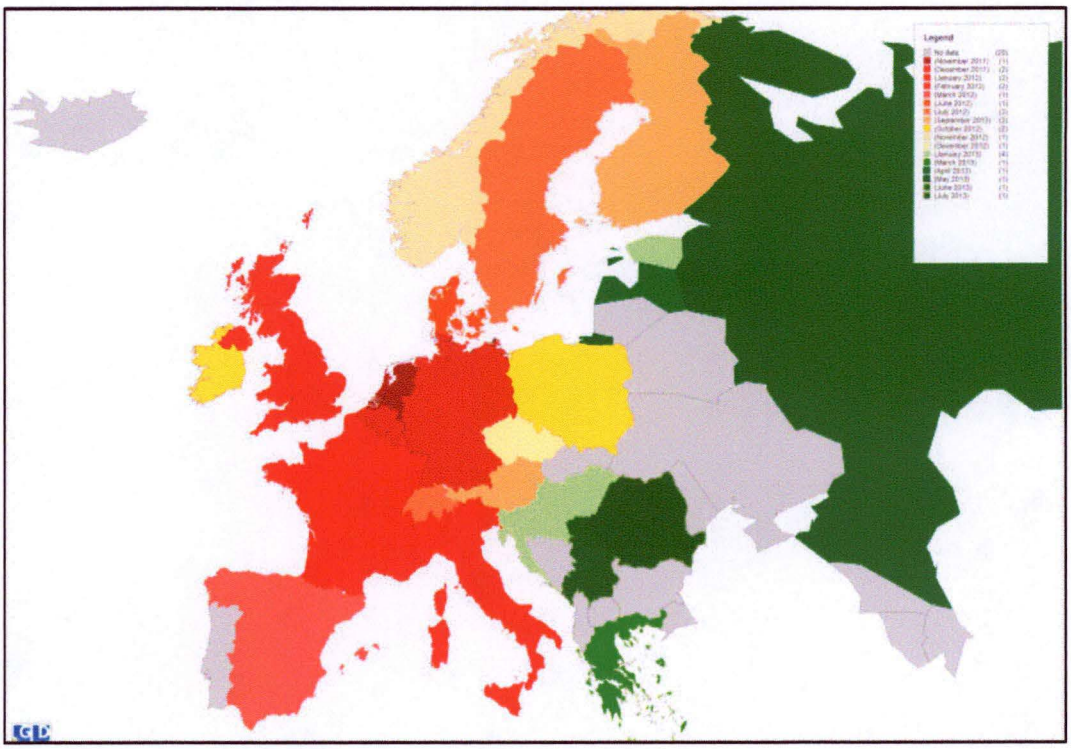

Figure 1 : Representation of distribution by date and country of SBV seropositive ruminants first detection by RT-qPCR or serological analyses (K. Lievaart-Peterson, 2015).

Legend: No data (20); November 2011 (1); December 2011 (2); January 2012 (2); February 2012 (2); March 2012 (1); June 2012 (1); July 2012 (2); September 2012 (2); October 2012 (2); November 2012 (1); December 2012 (1); January 2013 (4); March 2013 (1); April 2013 (1); May 2013 (1); June 2013 (1); July 2013 (1).

Introduction

1. Schmallenberg virus

1.1. Discovery

At the end of the summer and fall 2011, an unidentified syndrome characterized by a drop in milk production, hyperthermia and diarrhea among lactating cows emerged in north-west Germany and Netherlands. Later on, stillbirth, abortion and congenital malformations among the offspring of infected sheeps, goats, and cattles were reported (Beer et al., 2012), suggesting that this pathogen is able to cross the placental barrier (Hoffmann et al., 2012). Then, the virus spread across most of the European continent (figure 1). After further investigation, known pathogens were excluded and a new virus, named "Schmallenberg virus (SBV)" according to the town from where the first sample was isolated, was determined as the cause of this epizooty (Hoffmann et al., 2012).

A genomic analysis revealed that SBV belongs to the *Orthobunyavirus* genus from the *Bunyaviridae* family (Hoffmann et al., 2012) that includes 350 viruses divided into five categories: *Hantavirus, Nairovirus, Orthobunyavirus, Phlebovirus* and *Tospovirus*. Except *Tospoviruses* that are plant viruses, they all affect vertebrates. Up to now, more than 170 viruses split into 18 serogroups constitute the *Orthobunyavirus* genus (Doceul et al., 2013). SBV is part of one of them: the *Simbu* serogroup, as confirmed by genomic similarities shared with other *Simbuviruses* such as Aino, Akabane, Douglas, Sathuperi and Shamonda viruses (Hoffmann et al., 2012).

1.2. Origin

The emergence of a new virus strain may result from genetic (Domingo et al., 2010), environmental changes (Jones et al., 2008) or more presumably, a combination of both. Like any other negative-sense RNA viruses, SBV is highly susceptible to genetic modifications due to intrinsic properties. Indeed, the SBV RNA-dependent RNA polymerase (RdRp) lacks proof-reading function so the substitutions accumulate during the replication process. Besides, the segmented genome of SBV prompts to create genomic reassortment during co-infection. Genomic reassortment is a type of genetic recombination among RNA segmented viruses, defined as the mix of genetic material coming from 2 viruses infecting the same host cell. So, after co-infection with two segmented viruses, the reassortment process can not only generate viral particles identical to the parent viruses but it can also create progeny virions displaying genomic sequences derived from both of them.

Since its discovery in 2011, the historical and geographical origins of SBV have raised many questions in the scientific community. This deep interest led two different groups (Yanase et al. and Goller et al.) to distinct conclusions.

First, phylogenetic analyses of Yanase and coworkers suggested that SBV might be a reassortant virus resulting from the mix of genomic segments from 2 Simbuviruses, namely Shamonda (SHAV) and Sathuperi viruses (SATV), that have never been isolated in Europe (Doceul et al., 2013). They postulated that SBV would be composed of the small and the long segments from Shamonda virus while the medium segment likely comes from Sathuperi virus (Yanase et al., 2012).
On the contrary, results of full-genome examinations and serologic investigations led by *Goller et al* indicated that even if SBV was recently discovered, its origins seem to be more ancient that expected. Indeed, SBV could be an ancestor of Shamonda virus as it may have provided its small and long segments to SHAV during the reassortment process (Goller et al., 2012). Nevertheless, the SHAV medium segment provenance remains still unknown. Besides, this speculation also assumes that SBV must have existed in some parts of the world without causing any clinical signs detected or assigned to SBV (Doceul et al., 2013).

Altogether, despite extensive studies, the SBV origin still remains unclear. Both hypotheses proposed by *Yanase et al* and *Goller et al* seem plausible but further researches are required to determine which one has more credit.

1.3. Host range

Since the initial record of the Schmallenberg disease in cattle, sheeps and goats in 2011, this epizooty has spread to wild ruminants such as red deer and roe deer (Linden et al., 2012), chamois (Chiari et al., 2014), elk, bison (Larska et al., 2013a), buffalo, alpaca (Jack et al., 2012) and fallow deer (European Food Safety Authority, 2013). Besides, SBV seropositive dogs have also been reported in France (Sailleau et al., 2013) and Sweden (Wensman et al., 2013). Up to now, no evidence of SBV transmission in human has been described, indicating that the virus lacks zoonotic potential despite its ability to infect a wide range of ruminant species. Indeed, 301 sera analyses from persons working or living near farms and thus potentially exposed to SBV showed that no SBV-neutralizing antibodies and no presence of SBV viral genome sequences were detected (Reusken et al, 2012).

1.4. Clinical signs

SBV is an arbovirus, which means that it uses an arthropod vector, in this case a *Culicoid* (figure 3), to propagate (see point 1.5 *transmission*). The *Culicoid* bite triggers the host infection by SBV which can reach the striated muscles where the virus is able to replicate widely and to lead to a high viremia. Once in the bloodstream, the pathogen is able to spread throughout the entire body and infect any organ, including the central nervous system (CNS) after crossing the blood-brain barrier. Interestingly, the neuroinvasiveness appears to be age-related as young animals tend to be more sensitive than the older ones. So, the symptoms observed in the offspring born from SBV-infected females will be extremely different from the clinical signs described in adults (Elliott, 2014).

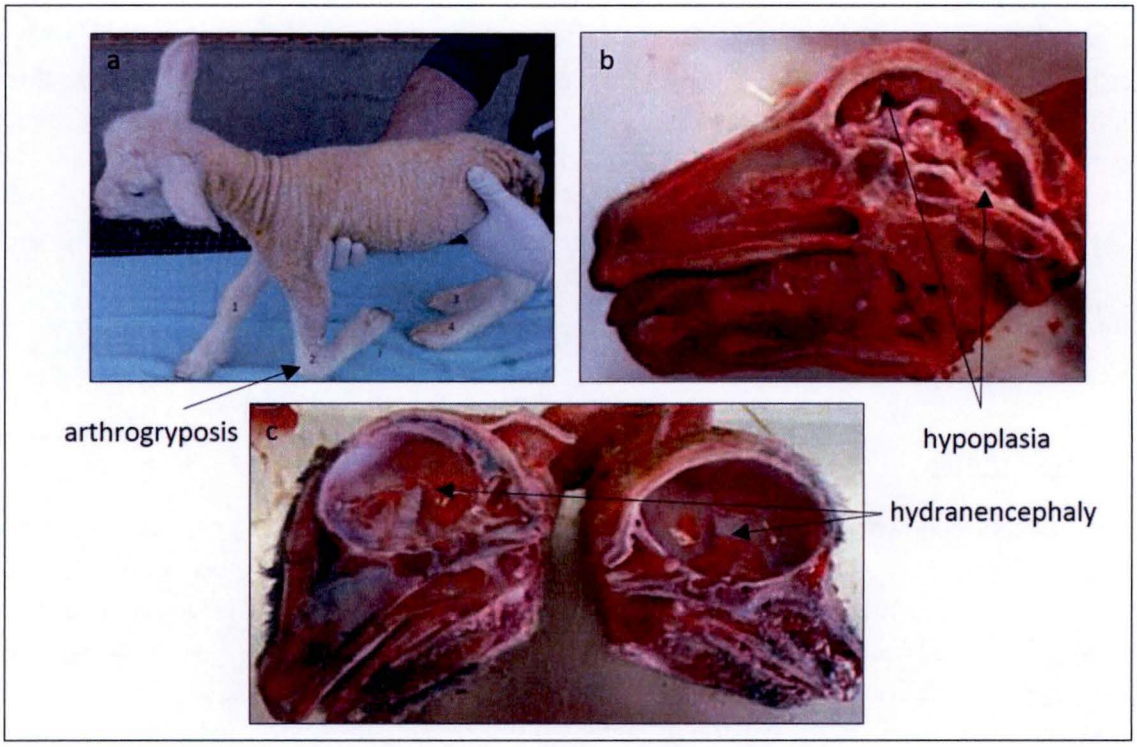

Figure 2 : Clinical signs caused by Schmallenberg virus in neonatal lambs. a) Newborn infected lamb showing arthrogryposis. **b)** Sagittal section of an infected lamb's head displaying cerebellum and cortical hypoplasia. **c)** Sagittal section of an infected lamb's head exhibiting hydronencephaly (pictures from Boseret et al., 2014 & Martinelle et al., 2014).

Lactating cows are ordinarily more affected than sheeps and goats by SBV infection, which is characterized by a 50% drop in milk production, loss of appetite, hyperthermia and diarrhea (Doceul et al., 2013 & Steukers et al., 2012). These symptoms usually disappear within a few days and the viremia period is quite short as it lasts 3 to 5 days in cattle. However, SBV RNA is still detectable in lymph nodes and in the spleen during up to 44 days (Wernike et al., 2013) although the persistence relevance in these organs remains unclear.

If SBV infects a serologically naïve pregnant female, the virus can disseminate in the placenta, reach the fetus and exert its teratogenic effect. More precisely, the progeny will suffer from severe congenital malformations consisting in neuro-musculo-skeletal disorders such as arthrogryposis, scoliosis, as well as abnormalities of the central nervous system like hydranencephaly, hypoplasia, amaurosis and even ataxia generally resulting in stillbirths and abortions (figure 2). Several cases of cryptorchidism have also been reported (Herder et al., 2012 & Garigliany et al., 2012).

These reports suggest that neurons can be considered as the major target for SBV replication, which could explain both severe lesions of the CNS and hypothalamo-hypophyso-glandular axes. This was confirmed by *in situ* hybridization and immunohistochemistry experiments led on brain sections extracted from infected newborn calves and lambs (Varela et al., 2013 & Hahn et al., 2012).

Furthermore, it was revealed that the severity of the injuries depends on the infection time point during the gestation period. If the contamination occurs when the fetal immune system is mature enough to control the infection, no symptoms is observed among the newborn. On the contrary, if the placenta development is not completed before the infection, the estimated risk to display any clinical sign reaches 28% (Garigliany et al., 2012).

To be clear, the time frame during which the offspring is sensitive to SBV infection (28-50 days for sheep, 60-180 for cow) corresponds to the development of placentomes (exchange unit between the mother and the fetus, 28^{th} day for sheep, 60^{th} day for cow) on one hand and the development of the blood brain barrier (50^{th} for sheep, 180 for cow) on the other hand. Before the 28^{th} day in sheep (or 60^{th} day for cows), no placentome is shaped and thus, even if SBV circulates in the mother bloodstream, it cannot be transmitted to the progeny as no mother-fetus exchanges are allowed. After of the blood-brain barrier formation is initiated, coinciding with the 50^{th}-60^{th} day in sheep (180^{th} for cows), the SBV access to the fetus CNS becomes limited. Once the barrier is functional (125^{th} day for sheep), SBV can no longer reach the brain of the offspring (Varela et al., 2013).

1.5. Transmission

No case of SBV horizontal transmission has ever been reported (Wernicke et al., 2013) but vertical transplacental transmission of the virus has been described in infected pregnant

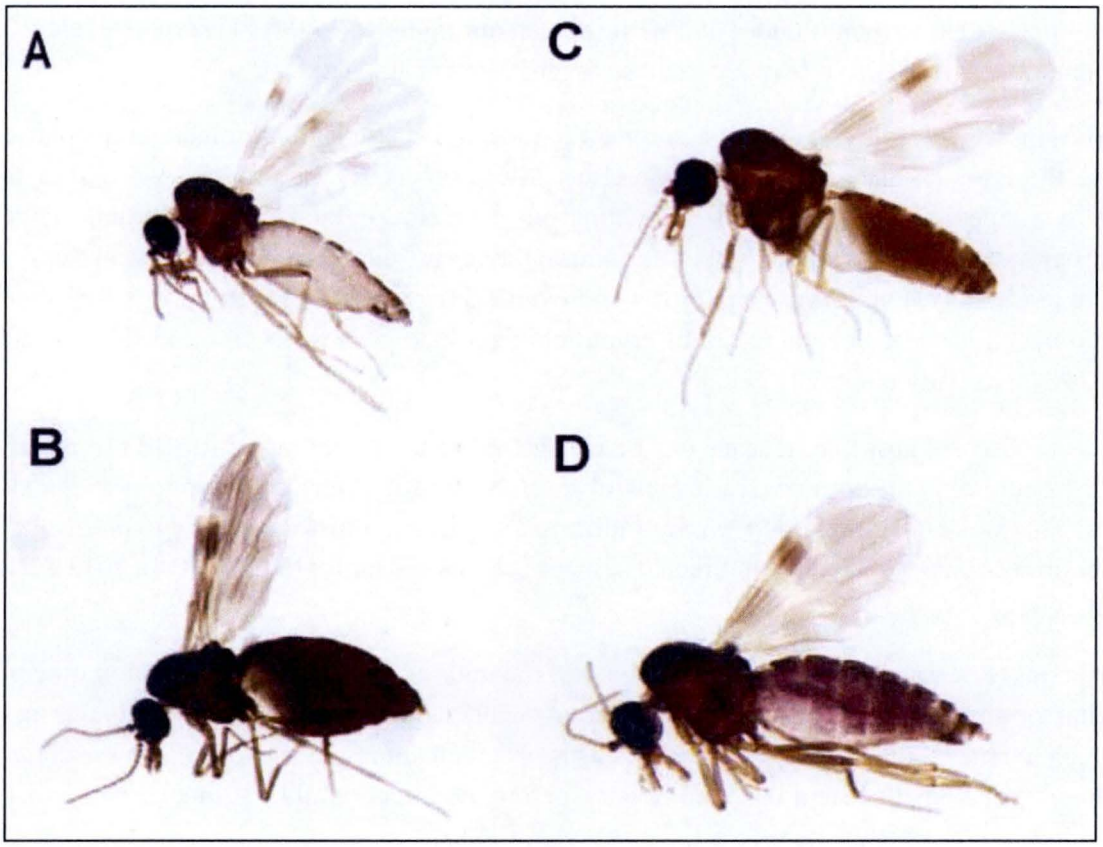

Figure 3 : Schmallenberg virus vector : *Culicoides*. *Culicoides obsoletus* parity forms: nulliparous (A); blood-fed (B); gravid (C); parous (D) (Larska et al., 2013b)

ruminant (Herder et al., 2012 & Garigliany et al., 2012). Indeed, potential hosts in contact with infected animals remained SBV-negative (Wernicke et al., 2013). Considering that SBV horizontal transmission (simple contact with infected animals) is very unlikely and that vertical transmission results in such severe lesions that they cause animal's death, an extended spreading of the virus as those observed in 2011 and 2012, likely involve a vector. This was confirmed by several groups who showed that small midges of the species *Culicoides* (figure 3), and more precisely *C. sonorensis (Veroseni et al., 2013) C. dewulfi*, *C. chiopterus*, *C. obsoletus* complex (de Regge et al., 2012) and C. *pulicaris* (Elbers et al., 2013), are SBV vectors. Moreover, it should be noticed that the virus transmission to mammalian hosts exclusively depends on female insects. Indeed, in contrast to males, females are hematophageous which means that they need to consume blood for nutritious or developmental purposes. In addition, even if mosquitoes have been proven not to be involved in SBV transmission (Wernicke et al., 2014), *Culicoides* may not be the sole SBV vectors as other blood-feeding arthropods such as bugs, ticks and black flies have not been tested so far.

The SBV transmission between the vector and the host can be summarized as following. First, an infected vector takes a blood meal on a vertebrate host and concomitantly injects pathogens. Secondly, while the virus is inside the host, it goes through an amplification process leading to high viremia. During this limited time of viremia (3-5 days), a significant quantity of pathogens is available to potential vectors. Thus, an uninfected arthropod female can become contaminated from feeding on this infected host by ingesting blood containing SBV viral particles. Once in the insect midgut, the pathogen replicates, disseminates through its body and establishes a persistent infection in its salivary glands. Next time this infected vector will bite another host, SBV will be transmitted and the cycle starts again.

Being disseminated by an arthropod makes SBV an arbovirus. Arboviruses are defined as a sub-type of viruses able to infect not only mammalian susceptible host but also blood-feeding arthropods. Thus, they maintain themselves by undergoing in life cycle in hosts, which carry the virus, and vectors, which spread the virus from one host to another (Mellor, 2000).

1.6. Viral genome and functions of the virus-encoded proteins

Like any other member of the *Orthobunyavirus* genus, SBV is a negative sense single-strand RNA virus with a tripartite genome (figure 4). The three segments (the short, the medium and the large segments) encode 6 proteins and share some organization similarities (figure 4). Indeed, each genomic segment consists of a coding region flanked by 3' and 5' untranslated regions (UTR) (Elliott, 2014). Within these UTR, complementary nucleotide motifs composed of the 10-20 terminal nucleotides (nt) allow the formation of a panhandle structure, considered as the replication and transcription promoter of the segments (Kohl et al., 2004) (figure 4). Besides, UTR sequences were also found to be used as packaging signal (Kohl et al., 2006) but little is known about this mechanism in *Orthobunyaviruses* compared to the influenza virus in which it has been extensively studied (Hutchinson et al., 2010). Interestingly, further analyses

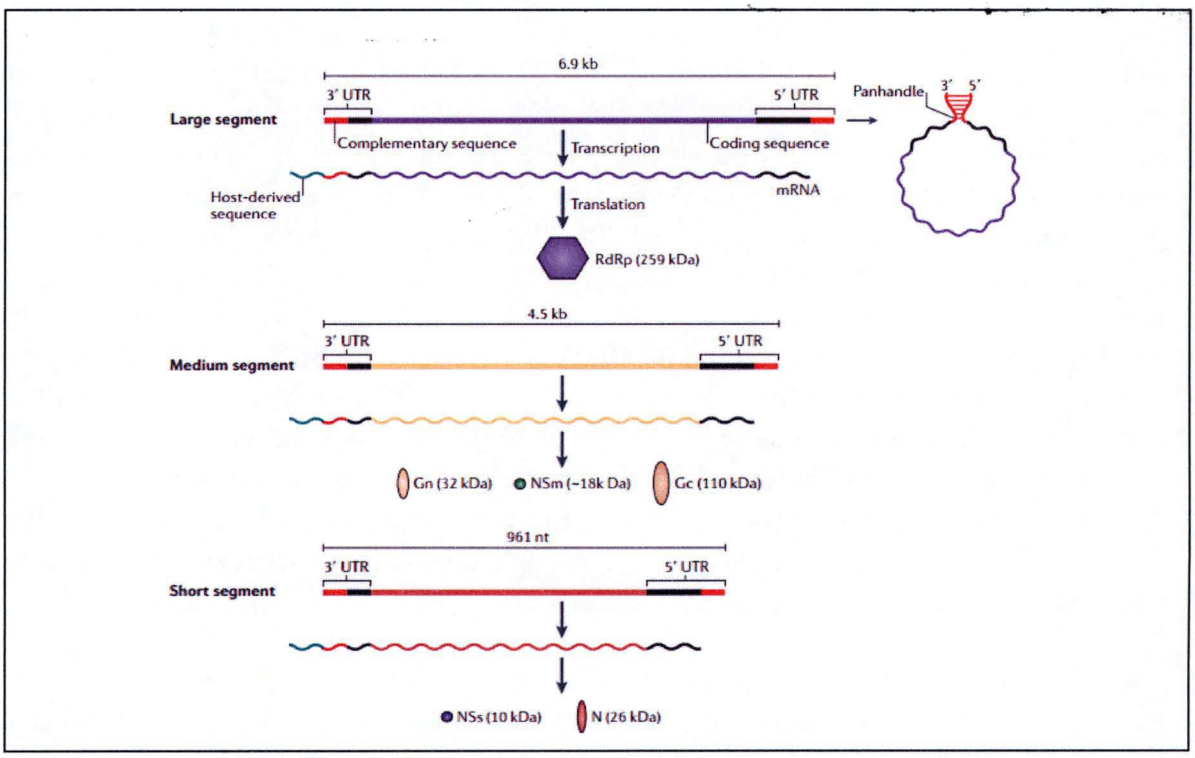

Figure 4 : Coding strategy of the orthobunyavirus genome. The three genomic segments are shown: the small, medium and large RNAs, with the total length above each segment (Bunyamwera virus used as model). These genomic RNAs are first transcribed into mRNAs, composed of a 5' host-derived sequence and truncated at the 3' end in comparison with the template. Both 3' and 5' terminal sequences of each genome segment are complementary (red). As a consequence, the base pairing between these 3' and 5' end regions induce the formation of a panhandle structure. The coding region of each segment, flanked by 3' and 5' UTR, is represented by distinct colored boxes. The L (large) segment encodes the RNA-dependent RNA polymerase (RdRp) whereas the M (medium) segment encodes a precursor polyprotein that is cleaved by host proteases to generate the viral glycoproteins Gn and Gc and the first non-structural protein (NS), NSm. Finally, the S (small) segment encodes the second non-structural protein NSs and the N protein (nucleocapsid) in an overlapping open reading frame but a different initiation codon is used for the translation. The estimated molecular weight of each encoded protein is also indicated (Elliott R.M., 2014).

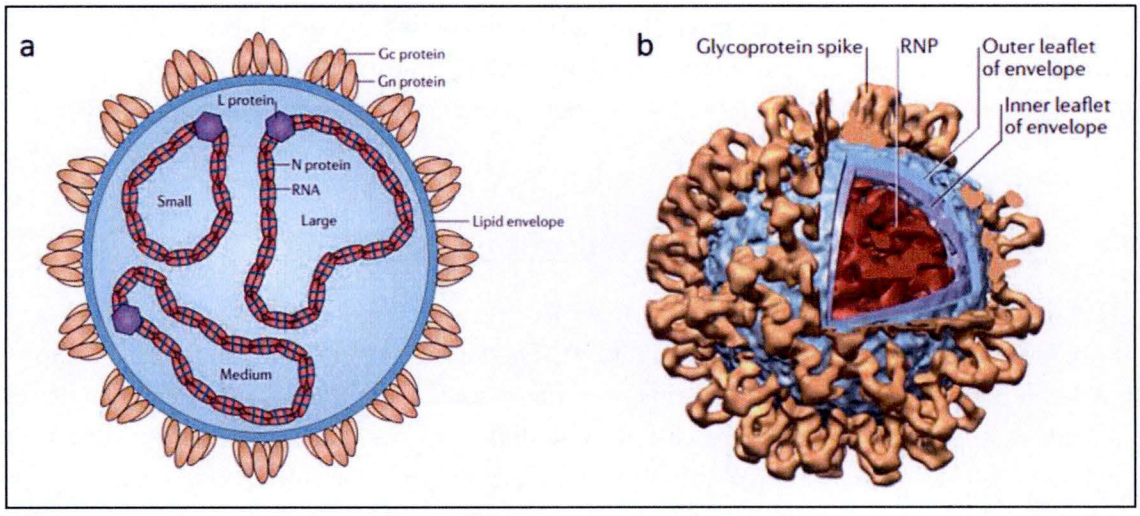

Figure 5 : Schematic representation of the Schmallenberg virus particle. a) SBV possesses a tripartite genome (small, medium and large segments) which is encapisdated by the Nucleocapsid protein (N protein) and associates with the RNA-dependent-RNA polymerase (RdRp) to constitute ribonucleoprotein complexes (RNP). These RNP are surrounded by a lipid envelope derived from the host cell Golgi complex, modified by incorporation of two glycoproteins, Gn and Gc. **b)** Illustration of tomographic reconstruction of Bunyamwera virus, apparented to SBV, displaying the ribonucleoprotein complexes, both outer and inner leaflet of the envelope and the global spiky appearance of the virion due to glycoproteins (Elliott, 2014).

have demonstrated a loss of cytopathogenicity and a reduced viral replication, both linked to deletions located within the UTRs (Mazel-Sanchez et al., 2012).

Since SBV only emerged 4 years ago, the accurate functions of the different genome segments are still poorly known. Therefore, data presented in the next paragraphs refer to other *Orthobunyaviruses* to summarize the assumed functions associated with each genome segment.

The large segment (L) encodes the RNA-dependent RNA polymerase (RdRp) (figure 4), also called L polymerase, involved in the viral replication and the production of three kinds of RNA: the genomic RNA (gRNA), the complementary RNA (cRNA) and the messenger RNA (mRNA) (Fields, 2007). It is essential to notice that the underlying mechanism of the molecular switch from transcription to replication remains unclear. Furthermore, the N-terminal domain of the RdRp might also display an endonuclease activity as it plays a key role in the cap-snatching process (detailed below) (Elliott, 2014).

The Medium segment (M) encodes a polyprotein precursor which is then cleaved by host proteases into the non-structural protein NSm and the two external glycoproteins Gn (N-terminal glycoprotein) and Gc (C-terminal glycoprotein) (figure 4) (Elliott, 2014).

Gn and Gc associate to constitute trimers of heterodimers in the endoplasmic reticulum (ER) and are transported to the Golgi complex thanks to a targeting signal mapped in the Gc transmembrane domain (Shi et al., 2004). These two envelope proteins, which confer a spiky appearance to the viral particle (figure 5), recognize yet unknown host cell membrane receptors to allow the interaction between the virus and the cell (Elliott, 2014). They are also thought to be the major target of antibodies produced by the immune response. However, the virus could be able to counteract the host immune detection thanks to a hyper-variable region contained into the M segment between NSm and Gc coding regions and ensuring a variation of the Gc protein sequence (Coupeau et al., 2013). Moreover, as they also interact with the Nucleoprotein and with each other, Gc and Gn glycoproteins seem to participate to the budding phenomenon and the virus assembly (Briese et al., 2013 ; Shi et al., 2007). Mutation of residues of the C-terminal domain of Gc resulting in impaired membrane fusion supports this hypothesis. It has also been postulated that Gn would play the role of a chaperone helping the appropriate traffic of Gc to the Golgi complex and preventing its sequestration in the ER (Shi et al., 2009).

Regarding NSm, it participates in virus assembly and morphogenesis since the protein is located in the Golgi complex (where the maturation occurs) independently of the other viral proteins. However, NSm mutant viruses showed distinct growth properties indicating that mutated NSm protein doesn't affect the virus viability but rather slows down the growth rate compared to the wild-type virus. (Shi et al., 2006).

Finally, the short segment (S) encodes the second non-structural protein NSs and the nucleocapsid protein (N) (figure 4) in an overlapping open reading frame but a different initiation codon is used for the translation (Elliott, 2014).

NSs, which is located in both cytosolic and nuclear compartments (Thomas et al., 2004), seems to take part to the virulence mechanism, to apoptosis as well as the replication cycle. Indeed, NSs is considered as the major virulence factor since silenced NSs expression strongly reduce virulence (Eifan et al., 2013). However, its participation to the apoptosis process is

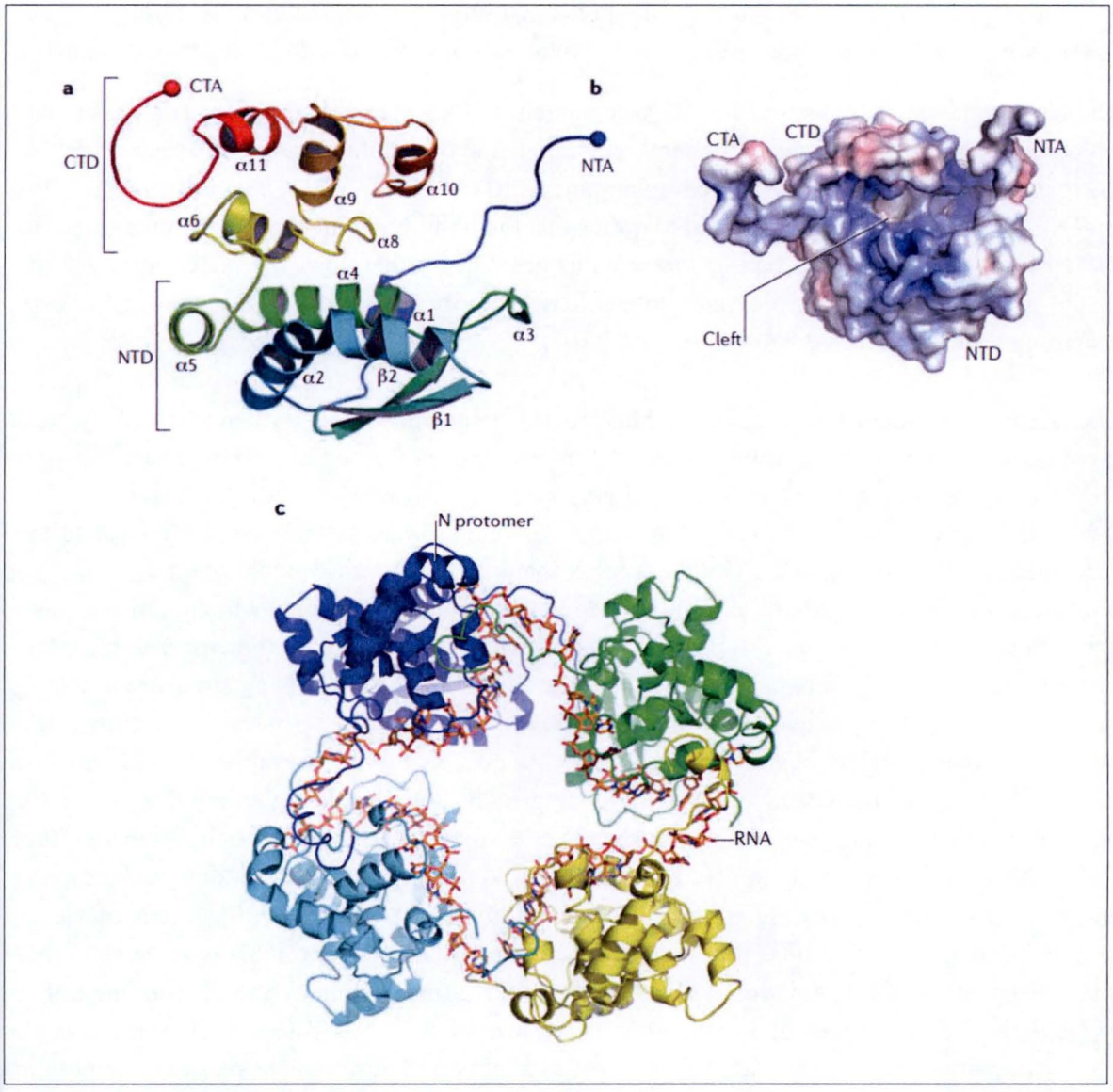

Figure 6 : Structure of SBV N protein. a) Crystal structure of SBV N protein protomer consisting of an N-terminal domain (NTD) in blue with an N-terminal arm (NTA) as well as a C-terminal domain (CTD) with a C-terminal arm (CTA) in red. **b)** Illustration of SBV N protomer displaying its electrostatic surface potential. Negatively charged residues are represented in red while positively charged residues are colored in blue. (Dong et al, 2013) **c)** Illustration of the SBV N protein-RNA complex. The tetrameric structure of the SBV N protein is formed thanks to close interactions between the NTA and CTA from each monomer (pictured in dark and light blue, green and yellow). This tetramer binds the RNA (represented in orange) (Elliott, 2014).

contested even within *Orthobunyavirus* genus: apoptosis is triggered in La Crosse encephalitis virus (LACV) (Blakqori et al., 2005) whereas it is delayed in Bunyamwera virus (BUNV) (Kohl et al., 2003). In terms of replication, the NSs involvement is not negligible since it induces transcriptional machinery inhibition by targeting RNA polymerase II subunit (RPBI) for degradation in vertebrate cells (Bridgen et al., 2001). It is also imperative to mention that even if NSs is responsible for shutting down the host immune antiviral response by interfering with type I interferon pathway (Doceul et al., 2013), some antiviral proteins such as viperin, Interferon-induced GTP-binding protein (MX1), 2'-5'-Oligoadenylate Synthetase 1 (OAS1/2) remain expressed during SBV infection (Blomström et al., 2015). Nevertheless, despite its numerous functions, NSs is not crucial for viral growth as NSs-lacking virus doesn't display any replication disparity. Besides, confirming the wide diversity of the *Orthobunyavirus* family, it has been reported that some members encode a severely truncated or even absent NSs-(Elliott, 2014).

The Nucleoprotein (N) (figure 6), which is the most abundant protein translated by the infected cells, is responsible for the encapsidation of the three segments (both gRNA and cRNA). More precisely, the N protein protomer is composed of an amino- and a carboxy-terminal domains both consisting in an N-terminal and C-terminal arms (figure 6a) and a cleft (figure 6b). Once the formation of a tetramer mediated by N- and C-terminal arms of each monomer is performed, a central groove containing positively charged residues is generated (figure 6c). This groove constitutes a binding site for the viral RNA so that the wrapped genome is protected from ribonucleases (Elliott, 2014) and the host innate immune system detection (ssRNA could be recognized by toll-like receptors triggering TNFα expression) (Dong et al., 2013(a et b) & Ariza et al., 2013). Once the gRNA encapsidation is performed, they assemble with the RdRp to constitute ribonucleoprotein complexes (RNP) surrounded by a lipid envelope originating from the host Golgi membrane. In addition, the N protein does not only interact with RNA but also with itself, the RdRp and the glycoproteins, although the molecular basis of these interactions must still be investigated (Elliott, 2014).

1.7. Replication cycle

As SBV replication cycle is still poorly described in the literature but predicted to be identical to other *Orthobunyaviruses*, the cycle detailed below is based on information collected on *Orthobunyaviruses*.

The replication cycle (figure 7) begins with the viral entry into the host cell, promoted by interactions between the Gn and Gc glycoproteins of the viral particle and the cell surface receptors. Some experiments conducted with La Crosse encephalitis virus (from *Orthobunyaviruses* family) have suggested that Gn would be the major protein involved in the attachment to invertebrate cells (Ludwig et al., 1991), whereas other groups studying California encephalitis virus (from *Orthobunyaviruses* family) have reported that Gc would function as the attachment protein for both vertebrate and invertebrate cells (Hacker et al., 1997). Although the host cellular receptors still remain unknown, studies have revealed that DC-SIGN (Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin), which is highly

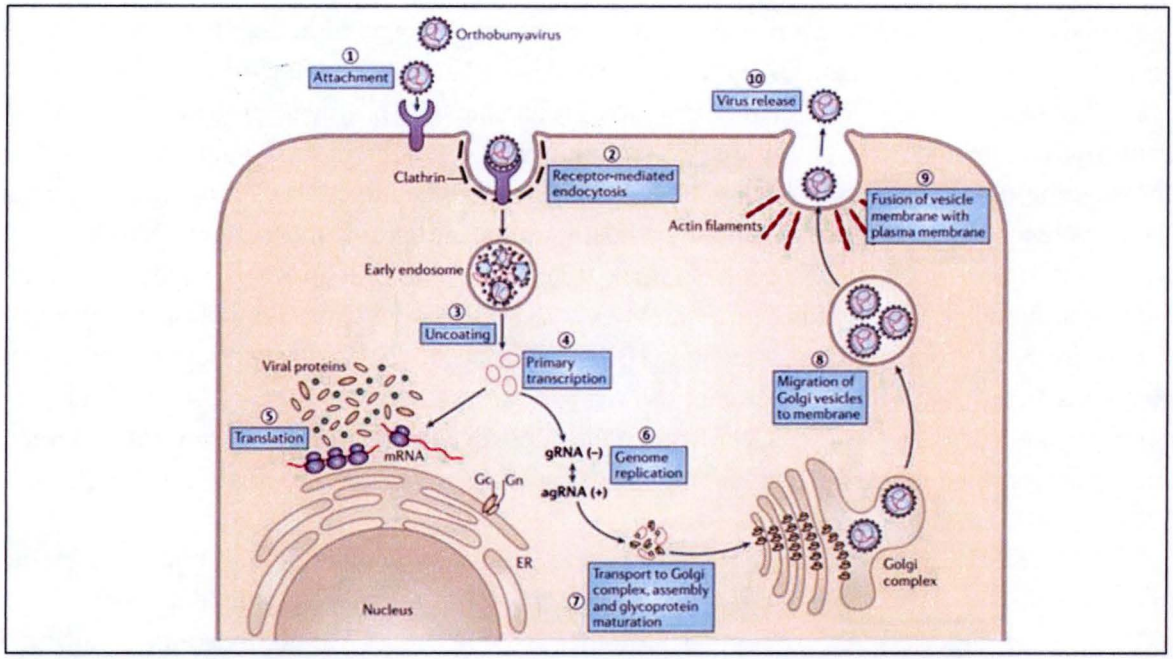

Figure 7 : The orthobunyavirus replication cycle. Replication begins with the recognition of host cell surface receptors by glycoproteins Gn and Gc (1) before the viral internalization via clathrin-mediated endocytosis (2). Acidification of endocytic vesicles results in the fusion between viral and endosomal membranes (3). Once in the host cell cytoplasm, the viral RNA-dependent RNA polymerase (RdRp) catalyses primary transcription of viral mRNAs (4) initiated by the cap-snatching process (not shown). These viral mRNAs are either translated into proteins (5) or used as templates to generate genomes copies (gRNA) (6). gRNA assemble with viral proteins and are transported to the Golgi complex whose membrane is modified by the incorporation of Gc and Gn. The viral particles bud from the Golgi complex (7) and are then transported to the cellular surface (8). After the fusion of both membranes (9), the new virions are released in the extracellular environment (10) (Elliott, 2014).

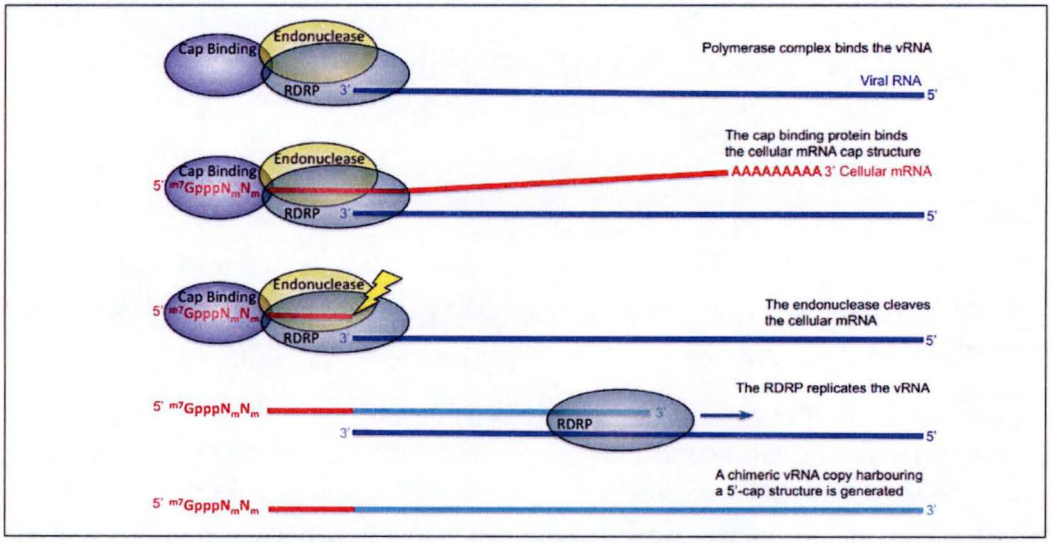

Figure 8 : RNA cap snatching process. The viral RNA-dependent RNA polymerase (RdRp) which recognizes viral RNA (dark blue), also binds the host cellular mRNA cap structure thanks to its cap binding activity (purple). Then, the RdRp cleaves the cellular mRNA downstream of the cap structure thanks to its endonuclease activity (yellow) and these short oligomers are used to prime the transcription process. This mechanism results in a chimeric vRNA (in red and light blue) harbouring a 5' cap structure derived from the host mRNA (picture from INTEC, 2013).

expressed on dermal dendritic cells, might be a candidate for the recognition phenomenon (Lozach et al., 2011). However, given the wide range of infected tissues, it is likely that other receptors are involved. The viral internalization continues with clathrin-mediated endocytosis (Hollidge et al., 2012) followed by the acidification of the vesicle. This pH decrease results in Gc conformational modifications promoting first the fusion between viral and endosomal membranes followed by the release of RNP in the host cell cytosol (Elliott, 2014).

Once in the cytoplasm, the primary transcription of viral mRNAs is catalyzed by the viral RNA-dependent RNA polymerase (RdRp), which executes the cap-snatching process (figure 8). Thanks to its endonuclease activity, the RdRp cuts the 5' ends of the host cell mature mRNAs about 12-18 nucleotides downstream the 5' 7-methylguanylate and the resulting short host-derived oligonucleotides are used to prime the viral mRNA synthesis (Reguera et al., 2010; Coupeau et al., 2013). This process has also been observed among influenza viruses, but for *Orthobunyaviruses* it takes place in the cytosol rather than in the nucleus as described for influenza viruses (Elliott, 2014).

Moreover, it has been discovered recently that none of the SBV mRNA is provided with a polyadenylation tail (poly(A) tail) (Coupeau et al., 2013). This intriguing feature raises questions since all the mRNA that have to be translated in a eukaryotic cells require a poly(A) tail, involved in mRNA stability and translation termination. This point, that provides the rationale of this research project, will be further developed in the section *1.8 SBV, a poly(A)-independent virus*, but we can already mention that the 3' end of the viral mRNAs is able to constitute a stem-loop structure suspected to be involved in transcription and/or translation termination process. Similarly, a putative transcription termination signal located 9–11 nucleotides downstream this hairpin was brought to light (figure 9) (Coupeau et al., 2013).

As a matter of fact, the transcription step will give rise to 3 kinds of RNA: the messenger RNA (mRNA), the genomic RNA (gRNA) and the complementary RNA (cRNA). On one hand, the three negative-sense genome segments (genomic RNAs) are used as template and are converted into positive-sense RNAs (complementary RNA, also named antigenomic RNA) necessary for genome replication. These nascent complementary RNAs are always associated with the N protein in the host cell to prevent hybridization with the template which would result in the formation of double stranded RNAs, possibly leading either to the aborted progression of the RdRp or to the detection by the host immune system. On the other hand, viral mRNAs are translated by host ribosomes to generate the six viral proteins described above (Elliott, 2014).

Following translation of the viral mRNAs, Gc and Gn glycoproteins associate to form trimers of heterodimers in the endoplasmic reticulum and are transported to the Golgi apparatus, using a targeting signal in the transmembrane domain in Gn (figure 7). Their glycosylation will be completed in this organelle. Meanwhile, new genome copies (gRNA) associate with N proteins and the RdRp to constitute RNPs which accumulate in the Golgi complex, and more precisely in the region where the membrane has been modified by the incorporation of Gn and Gc. Next, interactions between the C terminal domain of the glycoproteins and the RNPs appears to be required to promote budding into Golgi-derived vesicles (Shi et al., 2007). These vesicles containing new viral particles are then trafficked to the plasma membrane before being

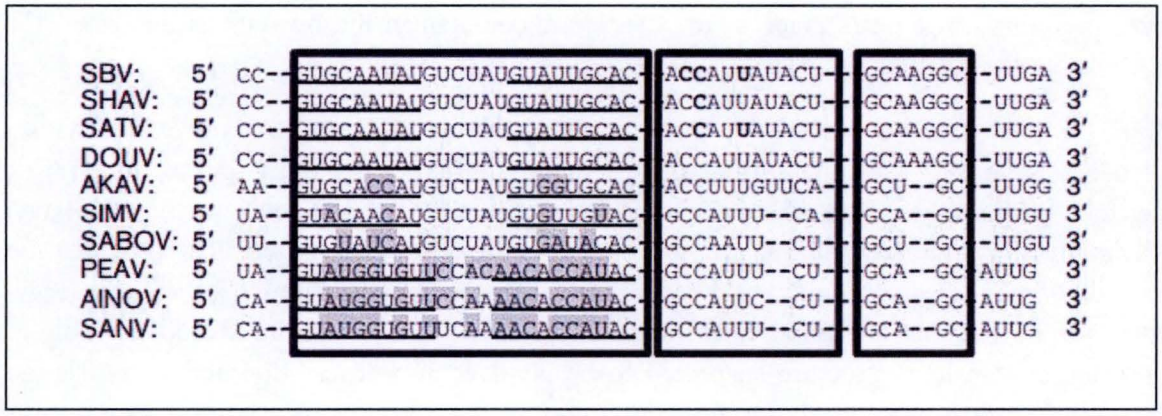

Figure 9 : Nucleotide alignment of conserved 3' UTR of several viruses belonging to the Simbu serogroup and comparison of the 3'UTR general organization. The representative selected viruses are: Schmallenberg virus (SBV), Shamonda virus (SHAV), Sathuperi virus (SATV), Douglas virus (DOUV), Akabane virus (AKAV), Simbu virus (SIMV), Sabo virus (SABOV), Peaton virus (PEAV), Aino virus (AINOV) and Sango virus (SANV). Underlined nucleotides correspond to the complementary nucleotides involved in the stem-loop structure whereas the highlighted nucleotides within this hairpin structure represent the modified nucleotides as compared to SBV. The boxes correspond to the conserved stem-loop, the 9–11 nucleotides downstream this hairpin and the identified putative termination signal, respectively (Coupeau et al., 2013).

subsequently released by exocytosis. Once in the extracellular environment, the nascent enveloped virions of 100nm are able to infect neighbouring cells and restart a new viral cycle (Elliott, 2014).

To date, distinct characteristics in *Orthobunyavirus* replication cycle have been observed, suggesting that replication mechanisms might differ in the mammalian host cell from the insect vector cell. First, during infection of invertebrate cells by bunyamwera virus (BUNV), the RdRp and the Nucleoprotein are confined in a large vacuole causing attenuated viral production (Lopez-Montero et al., 2011), a process which has not been reported in mammalian cells such as BHK-21 (Baby hamster kidney cells). Furthermore, BUNV assembly occurs in the whole Golgi complex of invertebrate cells resulting in a swelling and a disturbed secretory pathway in contrast to vertebrate cells wherein BUNV seems to accumulate only into peripheral Golgi stacks. Besides, the infectious particle release is quite contrasting depending on the infected arthropod or mammalian cell. Indeed, new BUNV virions are described to be released by budding in invertebrate cells and by extrusion in mammalian cells (Lopez-Montero et al., 2011). Finally, while a cytolytic viral replication cycle is observed in vertebrate cells, a persistent infection of arthropod cells is described in the literature (Elliott, 2014)

1.8. SBV, a poly(A) tail-independent virus

Despite progresses in terms of viral replication understanding, the transcription and translation mechanisms of Schmallenberg virus remain largely unclear.

Recently, it has been discovered that none of the SBV mRNA is provided with a polyadenylation tail (Coupeau et al., 2013). However, poly(A) tail is normally essential for the nuclear export, the RNA stability and the ribosome recruitment for the translation process (Guhaniyogi et al., 2001). Therefore, this lack of poly(A) tail among *Orthobunyaviruses* raises many questions and prompted researchers to look for alternative mechanism to control RNA stability and ribosome recruitment. A stem-loop structure spanning from nt 762–786 and localized in the 3'UTR of the antigenomic S segment (also named complementary S segment) of SBV (figure 9) was proposed as a putative functional substitute of the poly(A) tail (Coupeau et al., 2013). This hairpin structure is highly conserved within the *Simbu* serogroup and any mutation introduced in one strand of the stem-loop is counterbalanced by a complementary mutation in the other strand to maintain its stability. Remarkably, this hairpin structure is also present in the 3'UTR of the M and L segments, although less conserved among *Simbuviruses* (Coupeau et al., 2013). Similarly, a putative transcription termination signal (5'-GCAAGGC-3', nucleotides 798-804) located 9–11 nucleotides downstream this stem-loop was brought to light (figure 9). The conservation of such RNA elements and their localization suggest a potential functional relevance in viral replication. Indeed, while cap-snatching initiates the mRNA synthesis, the hairpin could play a key role in transcription termination as well as in translation initiation as it could partly replace the poly(A) tail. On the contrary, the termination signal, which is not present on viral mRNAs, would only be involved in transcription termination (Coupeau et al., 2013).

As a consequence, we hypothesize that these two elements may also constitute binding sites for RNA-binding proteins which could take part to transcription / translation mechanisms. To identify them, this project aims to set up an RNA-affinity purification followed by the identification of interacting proteins by mass spectrometry.

Determination of RNA-binding proteins (RBPs) contributing to transcription / translation processes will not be surprising considering that RNA viruses replication requires proteins from both viral and host provenance.

Several examples of RNA-binding proteins diverted from their genuine functions to the virus advantage have been already described in the literature (Li & Nagy, 2011; Lloyd, 2015). For instance, heterogeneous nuclear ribonucleoprotein K (hnRNP K) and NS1-Binding protein (NS1-BP) taking part to the splicing of RNA segments of influenza A viruses.

Influenza A virus genome consists in eight single-strand negative-sense RNA segments. Among them, two segments undergo to alternative splicing and produce numerous proteins: the NS segment encodes the non-structural protein (NS1) and the nuclear export protein (NEP/NS2) whereas the M segment encodes the matrix protein (M1) and an ion channel (M2). Although little is known about the molecular basis involved in splicing of influenza RNA segments, a complex of RNA-binding proteins composed of NS1-BP and hnRNP K, both host proteins, have been recently shown to be crucial players in this process. Indeed, low levels of NS1-BP and hnRNP K were correlated with reduced M2 levels as well as attenuated viral replication. Further investigations demonstrated that NS1-BP does not bind directly to M1 mRNA but this interaction seems to be dependent on hnRNP K (Tsai et al., 2013).

As most RNA viruses are described to replicate in the cytoplasm of infected cells, the RNA-dependent-RNA-polymerase should be able to discriminate the viral genome from other subtypes of host cellular RNAs. In positive-sense RNA viruses, several host RBPs have been identified to participate to the viral RNA recognition process while others have an impact on the replicase complex and thus positively or negatively regulate RNA synthesis. Viral RNAs stability and escape from host degradation pathways have also been shown to be affected by host RPBs. Although some of RBPs contribute to various events during infection, most of them appear to be responsible for specific functions. Besides, numerous RBPs such as Eukaryotic translation elongation factor 1A (eEF1A) and the Like Sm (Lsm) 1–7 complex seem to be conserved within positive-sense RNA viruses, suggesting potentially shared underlying mechanisms in host-virus interactions. By contrast, particular RBPs are also hijacked by viruses indicating that identical roles could be carried out by distinct RBPs (Li et al., 2013).

In addition to genomic amplification, recent studies revealed that cellular RBPs could also be implicated in numerous steps of infection process such as translation (Karakasiliotis et al., 2010).

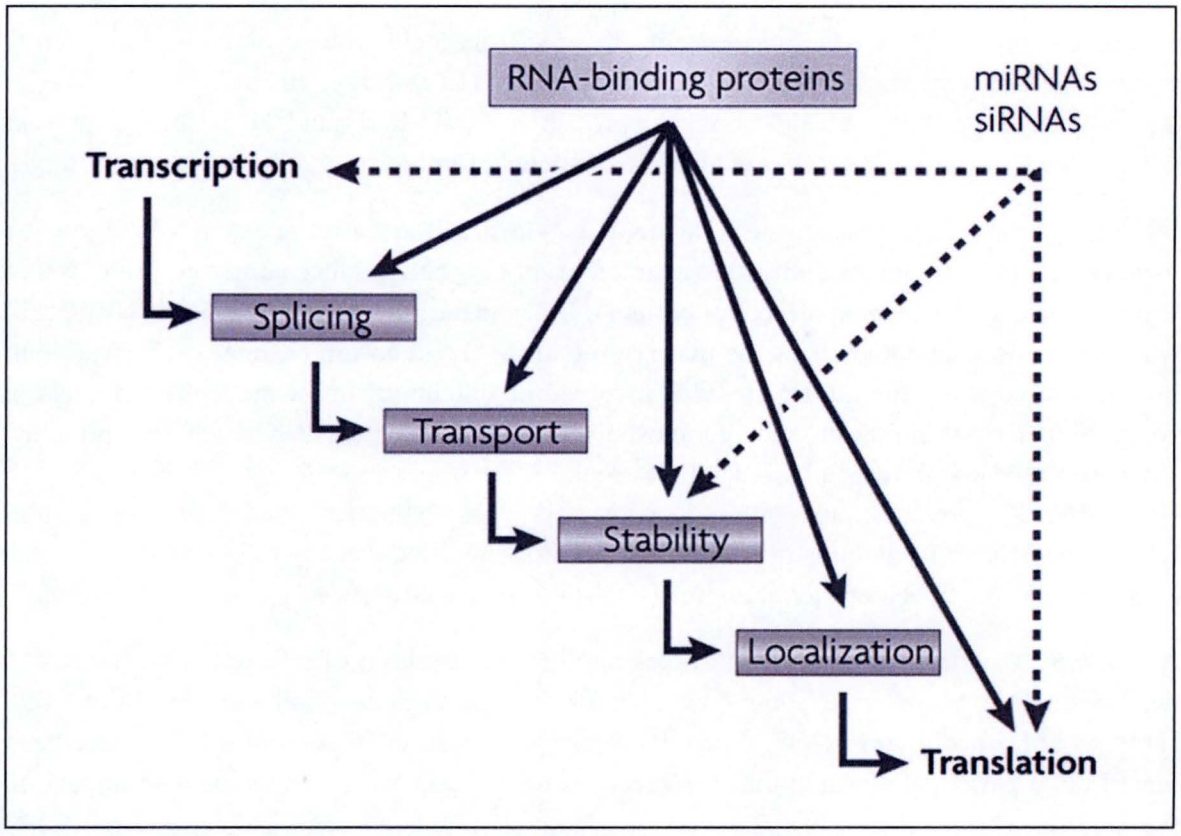

Figure 10 : Coordination and interconnection of gene regulation steps from transcription to translation. RNA-binding proteins and small non-coding RNAs (microRNAs (miRNAs) and small interfering RNAs (siRNAs)) supervise mRNAs regulation at multiple levels (Keene, 2007).

2. RNA-binding proteins

2.1. RNA-binding proteins in gene regulation

Life is dependent on carefully regulated gene expression ensured by an appropriate spatial and temporal control of RNA and protein amount. Regulation of gene expression includes a broad spectrum of mechanisms occurring in cells to turn the genetic information contained in a DNA sequence into a functional gene product. In other words, each step orchestrated in the RNA lifespan can be modified, whether transcriptional initiation; RNA processing, maturation, editing, stability and localization; mRNA translation or post-translational modification of a protein in order to raise, decrease or simply modulate RNA and protein quantity/activity (Glisovic et al., 2008) (figure 10).

All the steps involved in gene regulation but occurring before transcription and coordinated by DNA-binding proteins will not be discussed in the next paragraphs as the purpose of this work is focused on RNA-binding proteins.

Once the transcripts produced by DNA-dependent RNA polymerase, both modifications, distribution, stability and translation of the mRNA will be handled by RNA-binding proteins (RBP) and micro-RNA (miRNAs). In this way, the transcriptional rate and each step of the post-transcriptional regulation will be governed by RBP.

The splicing, consists in the pre-mRNA introns removal and the exons juxtaposition in order to produce a transcript ready for translation. Moreover, there exists a variant of this process named alternative splicing generating a wide range of proteins by choosing to include or exclude exons in the final mRNA. Consequently, by changing mRNAs exons combination, distinct proteins will be translated from alternatively spliced mRNAs. This mechanism is catalyzed by the spliceosome, a complex of RBP which binds to each of the sequences to be removed, loops it and finally cleaves it before joining the two extremities (Müller-McNicoll & Neugebauer, 2013).

Once genetic information packaging completed, the RNA is further processed by the addition of a 7-methylguanylate cap (m^7G cap) and a polyadenylation tail (poly(A) tail) to 5' and 3' mRNA extremities respectively, two modifications crucial for mRNA stability and translation, as detailed below. Moreover, the poly(A) tail is also involved in transcription termination as this process requires simultaneous interactions of the RNA polymerase with the transcript polyadenylation site and the terminator signal present in DNA sequence (Jackson et al., 2010). In addition, post-transcriptional regulation also implies RNA editing: nucleotide conversions occur by deamination of adenosine residues into inosine or of cytosine into uracil (Brennicke et al., 1999). These reactions, catalyzed by Adenosine Deaminase Acting on RNA (ADAR) enzymes or Cytosin Deaminase Acting on RNA (CDAR or "APOBEC") respectively (Brennicke et al., 1999), contribute to the extension of the encoded genes diversity.

After the RNA maturation step, numerous RBPs were demonstrated to associate with the transcripts in the nucleus and promote their export to the cytoplasm through close interactions

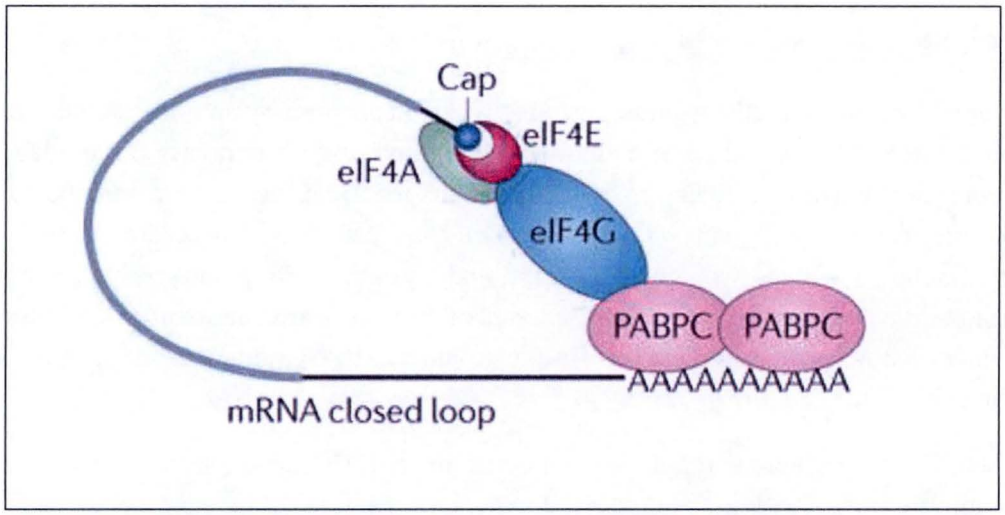

Figure 11 : Role of the cap structure and the poly(A) tail in translation. Both 5'-cap structure and poly(A) tail are described to play a crucial role in translation as they both recruit pre-intiation complexes (composed of 40S ribosomal subunit, the eIF2–GTP–MET–tRNA Met, eukaryotic translation initiation factor 3 (eIF3), eIF1, eIF1A and probably eIF5, not shown). The cap structure is bound by eIF4F comprising the cytoplasmic cap-binding protein eIF4E, the RNA helicase eIF4A and the scaffolding protein eIF4G while the poly(A) tail is recognized by the cytoplasmic poly(A)-binding protein PABPC. Through interactions between PAPBC and eIF4G, the two RNA extremities are brought closer and translation is promoted by an enhanced binding of eIF4E to the cap (Huntzinger & Izaurralde, 2011).

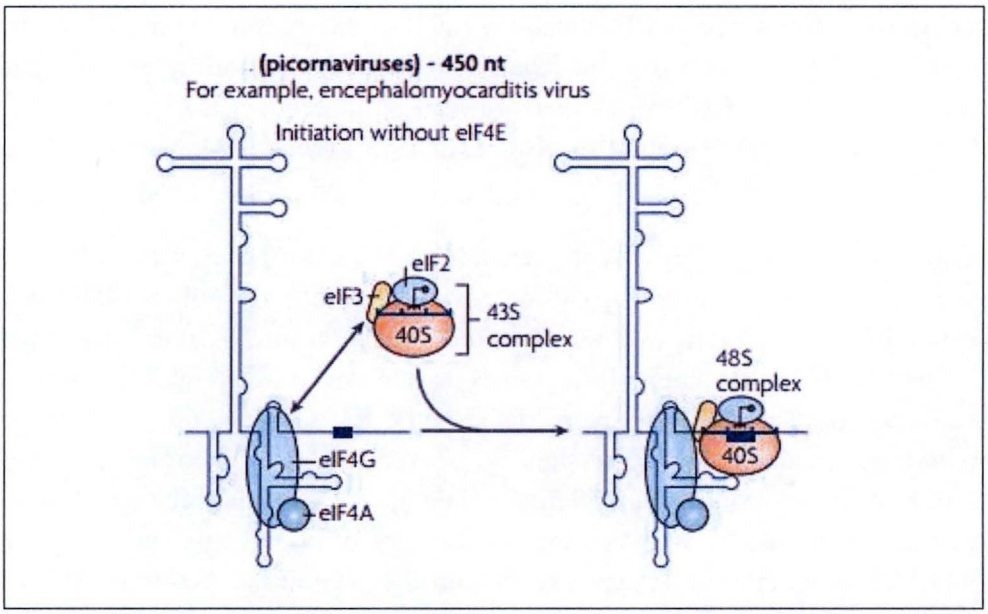

Figure 12 : Internal ribosome entry sites (IRES)-mediated translation initiation in picornaviruses. IRESs consist in highly structured RNA elements able to recruit the ribosome, in a cap-independent manner, to internal locations in mRNA. Viruses uses different mechanisms depending on interactions with eukaryotic initiation factors (eIFs) and/or 40S subunits (not shown). In picornaviruses, eIF4G bind the IRES; an interaction which is enhanced by eIF4A. Then 43S complexes are recruited without the eIF4E implication.

with proteins constituting the nuclear pore (Cullen et al., 2003; Müller-McNicoll & Neugebauer, 2013).

Once in the cytosol, mRNA stability can be jeopardized by 5' and 3' exonucleases degrading mRNA and thus decreasing its half-life. Protection of the transcript is ensured by the addition of the cap and the poly (A) tail (Glisovic et al., 2008). Besides, capping and polyadenylation processes also have a positive impact on translation: while the cap helps the ribosome binding to the RNA, the poly(A) tail is recognized by a poly(A)-binding protein (PABP) which mediates interactions between translation initiation factors (Jackson et al., 2010) (figure 11).

Let's mention that several of the above-mentioned RNA life steps are also controlled by microRNAs, as illustrated in figure 10. Indeed, these small non-coding RNA molecules are responsible for gene silencing by pairing to a complementary sequence of the targeted mRNA which results either in translational repression or in mRNA degradation. Although extremely powerful, the regulatory role of miRNAs will not be further discussed in this document. Let's just notice that RBPs also contribute to this phenomenon as they are required for the biogenesis and maturation of functional miRNAs (Valinezhad et al., 2014).

2.2. Control of RNA fate by RBP depends largely on conserved RNA structures

It has been known for years that the chemical intrinsic nature of RNA allows its folding into highly structured secondary and tertiary conformations, located in both coding and untranslated regions of RNA. As our knowledge of RNA biology and the listed processes in which RNP complexes operate is gradually increasing, it is becoming evident that conserved RNA structures - and their protein partners - are fundamental regulators of numerous cellular mechanisms, as illustrated below.

Transcription termination

In eukaryotes and bacteria, a role in transcription termination has been proposed for structured RNA. Indeed, in bacteria, this process implies a particular RNA folding located downstream the elongation complex, inducting its dissociation. A similar use of RNA secondary structure (RSS)-dependent termination has been demonstrated for eukaryotic polymerase III, involved in the production of many structural RNAs including ribosomal RNA, transfer RNA, spliceosomal RNA,... (Zenkin, 2014)

Translation

As just mentioned, in eukaryotic cells, translation initiation is generally governed by a cap-dependent mechanism, but some viral and cellular mRNAs escape this pathway and recruit the translational machinery thanks to internal ribosome entry sites (IRES). The ribosome subunits assemble around the targeted mRNA near the cap in the first case, whereas in the second case, it binds the IRES sequence which adopts a stable secondary structure localized mostly in the 5'UTR but also encountered in different positions of the transcript molecule (Jackson et al., 2010; Martínez-Salas et al., 2013) (figure 12).

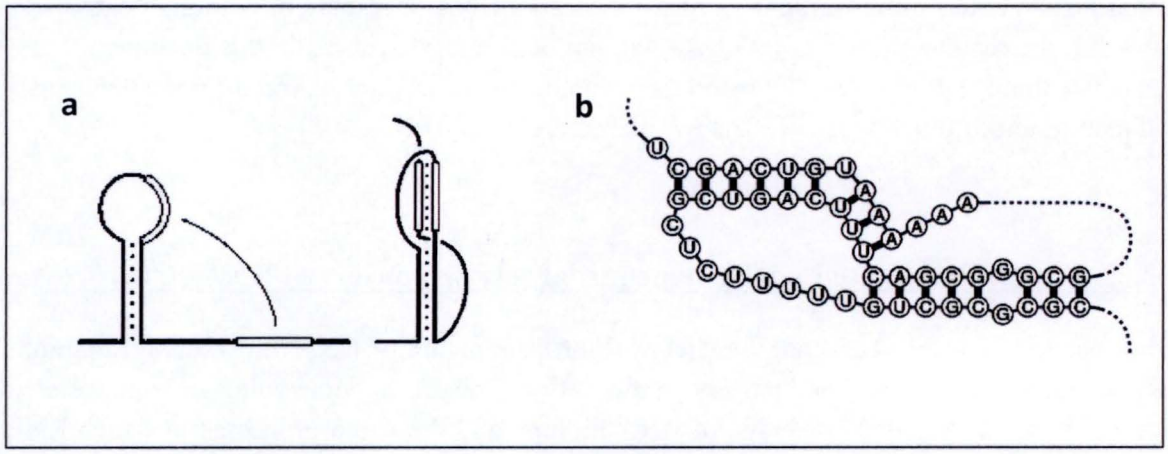

Figure 13 : Pseudoknots. a) Picture showing the mechanism of pseudoknots formation. This results in a structure with two helices wherein a strand from one of the two helices corresponds to a region localized between the segments corresponding to the two strands of the other helix. **b)** A model for human telomerase pseudoknot structure.

In cap-dependent translation, ribosome recruitment at the start codon can be modulated by mRNA secondary structures located between the cap and the initiation codon, in such a way that the ribosome progression on the mRNA is blocked, as observed for the gene of human ferritin (Hentze et al., 1987). All the eukaryotic mRNA are polyadenylated but the transcripts coding for histones. Those are characterized by a 3' end stem-loop structure playing an important role in translation (Marzluff et al., 2008).

Reading frame shifting

In particular cases, a shift of the open reading frame is required for protein expression. Indeed, the codons are encoded in two overlapping translational reading frames: the beginning of the gene is located in one phase and the end in another phase shifted by +1 or -1 nucleotide. Consequently, the ribosome must slip from a reading frame to another one in order to get a complete polypeptide (Staple & Butcher, 2005). This slippage is carried out at a specific site of the mRNA usually containing a secondary structure, called pseudoknot, which temporarily blocks the ribosome and promotes the change of the reading frame (Somogyi et al., 1993). This stable structure is formed thanks to a hairpin structure on one hand, and a pairing of nucleotides between a loop region and another RNA region located outside this hairpin on the other hand (figure 13a) (Staple & Butcher, 2005). Pseudoknots, which were first identified in the 3' end of turnip yellow mosaic virus genomic RNA, were also reported in bacteria, yeast but also human telomerase (figure 13b) where it helps the catalytic core constitution (Theimer et al., 2005).

Among various secondary structures of RNA, single hairpins are most often encountered. These latter, usually constituted of a terminal loop and a basal double-stranded RNA stem (often containing mismatches) thanks to intermolecular base pairing, differ from each other in many aspects. Indeed, structural variability, probably playing a role in RNA-protein interactions specificity, is due to the stem length, the number of mispairing, the loop size and the overall primary sequence of the hairpin as well as its position in the RNA molecule. Up to now, single stem loops have been reported to interact with many proteins and to be involved in mRNA localization in mammalian cells and in Drosophila oocytes and embryos, mRNAs protection from degradation (Svoboda & Di Cara, 2006), genomic RNA synthesis (Bhullar et al., 2014) and translation of mRNA of replication-dependent histones which do not possess poly(A) tail (Marzluff et al., 2008).

2.3. RNP conservation, assembly and complexity of RNA-protein interactions

RNP complexes are evolutionarily conserved

The few examples exposed above support that RNA–protein interactions are crucial for the regulation of gene expression at the post-transcriptional level. RBPs associate to RNA structural elements and function together to form ribonucleoprotein particles (RNPs) within the cell (Mitchell & Parker, 2014) to supervise all steps of RNA metabolism (Muller-McNicoll & Neugebauer, 2013).

Besides, RNP complexes relevance is further underscored by their conservation throughout evolution. Indeed, a recent study by Hogan and co-workers has highlighted a strong conservation of RBPs-RNA interactions involved in post-transcriptional gene regulation, that was already existing at the earliest stages of evolution (Hogan et al., 2015).

RBPs are generally sub-divided into main categories based on their evolutionary relationships, cell distribution and identified targets (reviewed in Gerstberger et al., 2014). Although some of them have been reported to display preference for multiple target interactions, the majority of RBPs are intended to associate with one RNA subtype. The larger group of RBPs consists in proteins binding messenger RNAs (mRNAs) (~700 proteins, 45% of human RBPs), followed by proteins committed to the cytosolic and mitochondrial ribosomes (~170 proteins) while RBPs interacting with transfer RNA (tRNA) (~150) and ribosomal RNA (rRNA) (~120) constitute smaller classes (Gerstberger et al, 2014).

Of approximately 110 RPBs families determined in humans, roughly half of them share at least 30% of homology with their homologs in yeast. Among them, cytosolic ribosomal proteins, displaying the highest degree of conservation (~57 % homology), are encountered. Nevertheless, such structural and functional similarities are not surprising considering that translational mechanisms are expected to be conserved between eukaryotes. By contrast, no homologs in yeast have been detected for most human mitochondrial ribosomal proteins suggesting early divergence in evolution (Gerstberger et al., 2014).

The RNP assembly is sequential

Studies on ribonucleoprotein particles within different species showed that RNPs usually contain one or more RNAs which are/is bound by at least one protein which broadly differ in size (Faoro, 2014). Besides, it was also reported that the same RBP was able to interact with mRNAs encoding related proteins or proteins with similar functions (McKee & Silver, 2007). RNPs assembly, occurring in the nuclear or cytosolic compartment through successive events, would be carried out in a sequential manner, meaning that each step is conditioned by the correct completion of the previous one. In this way, only competent RNP intermediates will undergo the next maturation stage, ensuring that final RNP will be fully functional (Oeffinger & Montpetit, 2015).

In eukaryotic cells, no less than 20 000 mRNAs are encountered (Ramsköld et al., 2009); a variety which is further increased by alternative splicing (Nilsen et al., 2010) and chemical modifications (Dominissini et al., 2012). In addition to mRNAs, various species of non-coding RNAs, (miRNAs), tRNAs and rRNA exist. The RNA length is highly variable, ranging from 22 nucleotides to approximately 10 000 nucleotides (Jankowsky & Harris, 2015). As soon as it arises from the RNA polymerase, any transcript is covered with multiple RBPs that interact with the RNA in a concomitant or successive manner, through interactions with other RBPs (Mitchell & Parker, 2014) or even via post-translational modifications (Jankowsky & Harris, 2015). In vertebrate cells, the number of RBPs was estimated to approximatively 2500 distinct proteins expressed either ubiquitously or in a tissue-specific manner (Keene, 2010).

Domain class	Subclass (superfamily): Family
Nucleotidyltransferases	PAPs: canonical PAPs and non-canonical PAPs
	Terminal uridylate transferases
	CCA-adding enzyme
	Guanylyltransferases
	RNA ligases
	2′–5′ PAPs
	RNA-dependent RNA polymerases
Ribonucleases	α/β[1]: RNase A, RNase H, 3′→5′ exonuclease, RNase II, RNase R, RNase E, RNase PH and metallolactamase
	α and β: RNase T2 and XRN1
	α: RNase III
	Decapping enzyme
RNA-modifying enzymes	tRNA synthetases: class I and class II
	Deaminases: ADAR, APOBEC, TADA and CDA
	Pseudouridine synthases
	Methyltransferases: RMT, SPOUT, radical SAM-dependent methyltransferase and FAD/NAD(p)
Helicases	Superfamily 1: UPF1-like
	Superfamily 2: SKI2-like, RIG-I-like, DEAD-box, DEAH, RHA, viral superfamily 2 and Cas3
	Superfamily 3
	Superfamily 4
	Superfamily 5
GTPase	EF-Tu, EF-G, BMS1, SNU114
RNA-binding domains	RRM, KH, S1, OB-fold, PUF, sRBD, zinc-fingers, PAZ, PIWI, LSM, KOW, MIF4G, NTF2, GAR, HEAT repeat, homeodomain and CSD

Table 1 : Classification of common protein domains that interact with RNA.

ADAR, adenosine deaminase that acts on RNA; APOBEC, apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like; BMS1, BMh-sensitive 1; Cas3, CRISPR-associated protein 3; CDA, cytidine deaminase; CSD, cold shock domain; dsRBD, double-stranded RNA-binding domain; EF-G, elongation factor G; EF-Tu, elongation factor thermo unstable; FAD/NAD(p), flavin adenine dinucleotide/nicotinamide adenine dinucleotide phosphate; GAR, glycine arginine rich; HEAT, Huntington, elongation factor 3, protein phosphatase 2A and TOR1; KH, K homology; KOW, Kyprides–Onzonis–Woese; LSM, like Sm; MIF4G, MA-3 and eIF4G; NTF2, nuclear transport factor 2; OB-fold, oligonucleotide/oligosaccharide-binding fold; PAPs, poly(A) polymerases; PAZ, PIWI, Argonaute and Zwille; PIWI, P-element induced wimpy testis; PUF, Pumilio and FBF; RHA, RNA helicase A; RIG-I, retinoic acid-inducible gene I; RMT, ribomethyltransferase; RRM, RNA-recognition motif; S1, similarity to ribosomal protein 1; SKI2, superkiller 2; SNU114, small nuclear ribonucleoprotein-associated 114; SPOUT, spoU and trmD RNA methylase; TADA, tight adherence protein A; UPF1, up-frameshift suppressor 1; XRN1, 5′–3′exoribonuclease. (Jankowsky & Harris, 2015)

RBP-RNA interactions are complex and diversified

Myriad of RBPs harbor numerous RNA binding domains (RBDs) such as the RNA-recognition motif, which is the most commonly detected RBP in mammalian proteins. However, some of RBDs are associated with functions not limited to simple RNA interaction such as RNA structural remodeling, chemical modification, hydrolysis, etc (table 1) (Jankowsky & Harris, 2015). Most of RNA binding sites are composed of 3 to 8 nucleotides and can vary widely without altering the binding ability of RBPs. Consequently, the number of possible RNA-protein interactions becomes extremely large. RBPs are generally categorized as "specific" or "nonspecific" according to their inherent affinity for RBDs. Specific proteins are described to interact preferentially with characterized RNA sequence, secondary structure or a combination of both. On the contrary, nonspecific proteins are defined by an association with RNA sites devoid of expected sequences or secondary structures (Jankowsky & Harris, 2015). Interestingly, it was also revealed that RNA-interacting proteins are not restricted to proteins containing at least one identified RBD: RBP can be found among proteins not predicted to bind RNA (Baltz et al., 2012).

Besides, interaction with proteins may force the RNA molecule to adopt an alternative conformation thereby assisting the recruitment of supplementary protein partners. Given the flexibility of RNA structure, RNA-protein associations become quite difficult to forecast (Gopinath, 2009).

In addition, changes in intra- and extracellular environment might result in RNP complexes remodeling. Indeed, RBPs binding capacity seems to be affected by changes in protein content of subcellular compartments and post-translational modifications (Kishore et al., 2010). Similarly, some modifications to RNA such as (peudouridine or N6-methyladenosine) have been shown to change in response to environmental signals, leading sometimes to a disrupted binding of a RBP due to RNA structure modification (Oeffinger & Monpetit, 2015; Carlile et al., 2014 & Liu et al., 2015).

Finally, it is essential to notice that each interaction between one protein and its targeted RNA site is ruled by RNA and proteins concentrations, by the affinity of the protein for RNA binding site, by the presence of competing proteins for this particular site or competing RNAs for interaction with the protein of interest (Jankowsky & Harris, 2015).

Considering the wide RNA diversity, the countless binding sites and the huge quantity of RPBs and RNA-associated proteins, in addition to their abundance variation as well as the multiple layers of their interaction regulation, the number of potential RNP complexes is likely to grow.

2.4. Viruses take advantage of RBP-RNA interactions

Several RBP-dependent mechanisms described above for eukaryotic cells have also been reported for viruses. This is the case for IRES, enabling the cap-independent translation (Jackson et al, 2010), for pseudoknots that allow the ribosomal slippage from a reading frame to another, which is needed for complete polypeptide expression (Staple & Butcher, 2005).

In addition, the RNA splicing also occurs in some viruses and, in the case of the Influenza A virus, it has been shown to involve both viral and host RBPs. Influenza A virus genome consists in eight single-strand negative-sense RNA segments. Among them, two segments undergo to alternative splicing and produce numerous proteins: the NS segment encodes the non-structural protein (NS1) and the nuclear export protein (NEP/NS2) whereas the M segment encodes the matrix protein (M1) and an ion channel (M2). Although little is known about the molecular basis involved in splicing of influenza RNA segments, a complex of RNA-binding proteins composed of NS1-binding protein (NS1-BP) and the host heterogeneous nuclear ribonucleoproteins K (hnRNP K), both host proteins, have been recently shown to be crucial players in this process. Indeed, low levels of NS1-BP and hnRNP K were correlated with reduced M2 levels as well as attenuated viral replication. Further investigations demonstrated that NS1-BP does not bind directly to M1 mRNA but this interaction seems to be dependent on hnRNP K (Tsai et al., 2013).

Beside the regular RNA processes occurring in non-infected cells, viral infection also involves specific RNA-related requirements, like for the packaging and viral replication of RNA viruses.

Regarding the packaging, the example of the Influenza virus, composed of eight single-stranded negative-sense, is instructive. As a random packaging of viral RNAs into budding particles is unlikely, a selective mechanism, packing one single copy of each segment, has been postulated. Gavazzi and coworkers highlighted the specific sequences required for packaging and underscored their folding into two hairpins facing each other and forming a kissing loop complex which could potentially promote interactions between viral RNA segments (Gavazzi et al., 2013).

As most RNA viruses are described to replicate in the cytoplasm of infected cells, the RNA-dependent-RNA-polymerase should be able to discriminate the viral genome from other subtypes of host cellular RNAs. In positive-sense RNA viruses, several host RBPs have been identified to participate to the viral RNA recognition process while others have an impact on the replicase complex and thus positively or negatively regulate RNA synthesis. Viral RNAs stability and escape from host degradation pathways have also been shown to be affected by host RPBs (Li & Nagy,2011 ; Lloyd, 2015).

Although some of RBPs contribute to various events during infection, most of them appear to be responsible for specific functions. Besides, numerous RBPs such as eeF1A and the Lsm 1–7 complex seem to be conserved within positive-sense RNA viruses, suggesting potentially shared underlying mechanisms in host-virus interactions. By contrast, particular RBPs are also hijacked by viruses indicating that identical roles could be carried out by distinct RBPs (Li et al, 2013).

Finally, viruses lacking 5'-cap or 3'-poly(A) tail are being forced to invent tricks to elude mRNA decay and promote translation. Polioviruses, from the *Picornaviridae* family, chose to adopt a 5' cloverleaf conformation, bound by poly(rC) binding protein (PCBP) (also called heterogeneous nuclear ribonuclear protein E; hnRNP E), subsequently stabilizing mRNAs through blocking the interaction with Xrn1 exonuclease. Replacement of this structural element, by 5' m7G cap impeded mRNAs degradation while a mutant PCBP protein, unable to

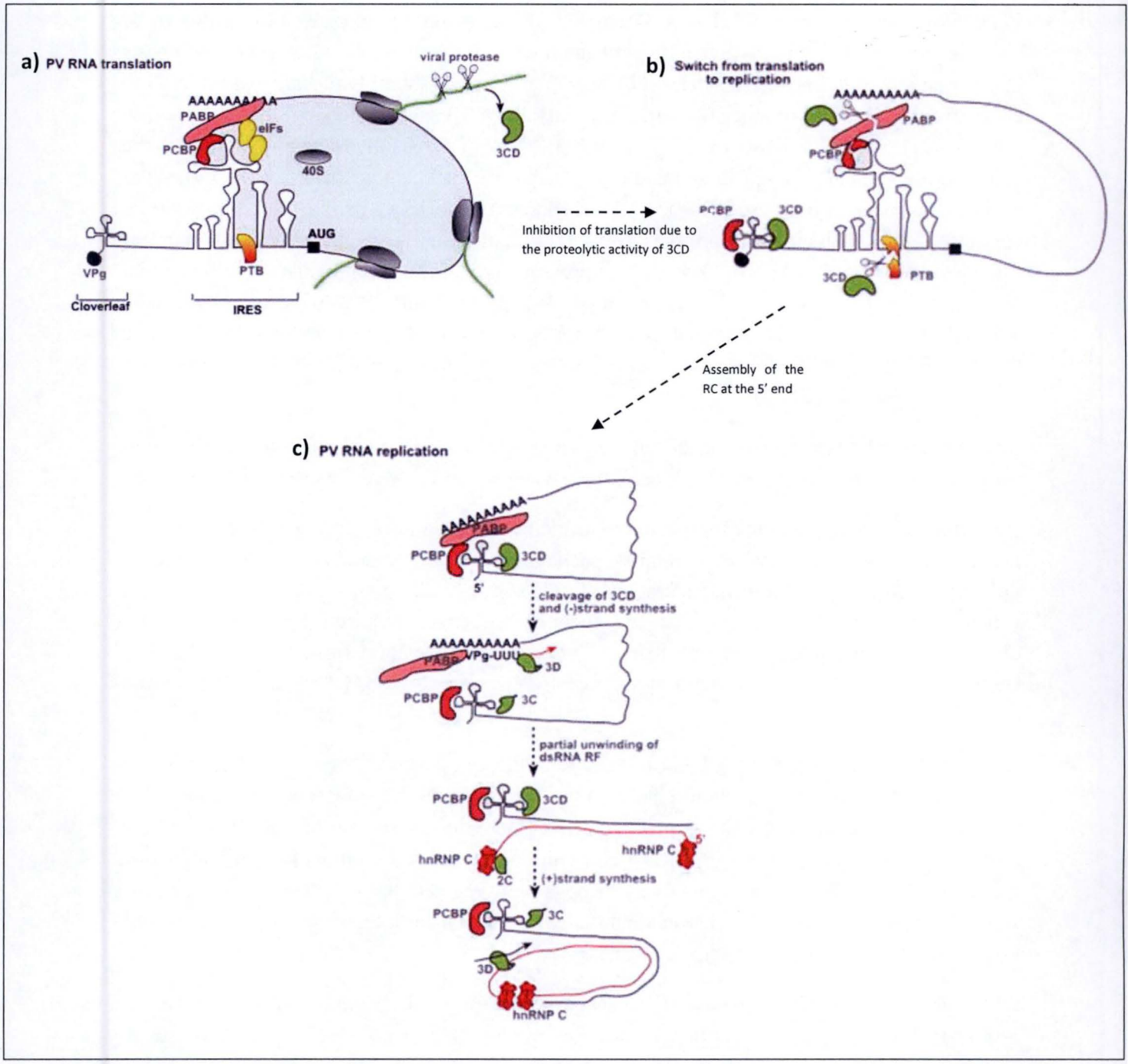

Figure 14 : A model of the roles of host RBPs in Polioviruses replication. a) Model picturing the role of host RBPs in polioviruses translation. Several host RBPs interact with stem-loop IV of internal ribosomal entry site (IRES) to initiate cap-independent translation. **b)** Model picturing the role of host RBPs in polioviruses in switch from translation to replication. PCBP2 is thought to be essential for switch from translation to replication due to its dual roles in translation and replication of polioviruses RNAs. PCBP act as a "molecular switch" whereas the viral 3CD is likely to be used as a "sensor" to measure protein translation. The involvement of PCBP is cleaved by viral protease 3C/3CD during viral infection. **c)** Model picturing the role of host RBPs in polioviruses RNA replication. Both 3CD protein and PCBP binds the 5′ cloverleaf structure and forms RNP with viral RNA to promote initiation of RNA synthesis. Thanks to an interaction of PCBP with PABP, which is bound to the poly(A)-sequence in the 3′UTR, the genome is circularized and the replication initiation complex formed at 5′-UTR, brought close to the 3′ end. hnRNP C was postulated to act as a chaperone and then, recruit viral 3CD replication protein to form an initiation complex for (+) RNA synthesis by bringing the 5′- and 3′-end sequences into close proximity (Li & Nagy, 2011).

interact with the cloverleaf structure, led to mRNA instability. In addition to its established role in mRNAs protection, the 5'-terminal cloverleaf structure, bound by polymerase precursor 3CD protein and PCBP, is also essential for initiation of RNA synthesis (figure 14b & 14c). Besides, PCBP also interacts with a second 5' UTR RNA structure and more precisely, with the stem-loop IV of internal ribosomal entry site, which is essential for translation of RNA genome (figure 14a) (Li & Nagy, 2011).

Likewise, a study conducted on Andes virus, a member of Hantavirus genus from the *Bunyaviridae* family, also lacking poly(A) tail, demonstrated that the 3' untranslated region of S Segment RNA, predicted to fold into a stem-loop structure, stimulates cap-dependent translation initiation (Vera-Otarola et al., 2010). Finally, poly-(A) tail-independent viruses might have developed a translation-enhancing alternative strategy is by the replacement of the poly(A) tail by structured 3'untranslated region, as shown for several viruses such as Bunyamwera virus, Japanese encephalitis virus and others (Blakqori et al, 2009; Bhullar et al., 2014).

However, it has to be mentioned that interaction of RNA structures with host proteins are not always beneficial for viral viability. Through examination of replication mechanisms of Japanese encephalitis virus (JEV), from the Flaviviridae family, also lacking of poly(A) tail at the 3' end of mRNA, Bhullar and coworkers revealed an innate immune host response. This mechanism involves an interaction of the polypyrimidine tract-binding protein (PTB) with the viral RNA which competitively inhibits the viral replication (Bhullar et al., 2014). More precisely, the PTB binds to the JEV 3' UTR RNA, predicted to form a hairpin, and prevents the association between the RNA and viral RNA-dependent RNA polymerase NS5 Protein, required for the efficient synthesis of genomic RNA (Bhullar et al., 2014).

2.5. Study of RNP complexes

Despite their fundamental roles, our understanding of RBP is quite restricted. So far, only functions of a few RBP have been defined thanks to the study of their tissue-specific expression profiles or through prediction analyses of RNA-binding domains and regulatory elements present in their sequence. Fortunately, a recent renewed interest in post-transcriptional regulation and the establishment of improved RNA-capture methods gave rise to a considerable evolution in the field of RBP biology as well as in understanding the RNP role in gene regulation.

In order to decipher the function of every RNP component, two complementary approaches have been used. The first one, called protein-centric methods, is focused on the determination of RNAs associated with a protein of interest while in the second one, named RNA-centric methods, the purpose is to identify numerous proteins interacting with a selected RNA used as a bait. In other words, RNA association with an individual RBP will be examined through protein-centric methods whereas characterization of multiple proteins interacting competitively, independently or cooperatively with a target RNA will be studied in RNA-centric approaches (Faoro, 2014). So, given the biological question developed in this thesis

Purification steps	Features	Advantages /Limitations
Bait sequence	Mostly several hundreds of nucleotides (to allow the formation of RNA secondary structures required for protein binding). Unfrequently: short RNA sequences (22 nt to 45 nt).	The longer the RNA sequence, the more complex is the protein mixture, making the identification of low abundant proteins more difficult.
Bait synthesis	In vitro synthesized (biotinylated or not), in vitro transcribed or in vivo transcribed.	RNAse sensitivity from the bait.
Solid support	Chromatographic columns or paramagnetic bead	Advantages of paramagnetic beads: easy to handle, large incubation/washes volumes and small elution volumes, the equilibrium is reached more easily with beads under agitation
Bait immobilization & elution possibilities	- Chemical modifications of RNA. - Hybrid nucleotidic probes, with biotinylated deoxyribonucleotides linked to the ribonucleotide bait. - Tags introduced in the RNA sequence: aptamers or polyA (for recovery with oligo(dT)-beads). - Antisense oligonucleotides bound to the chromatographic support are used to fish the RNA bait.	- Chemical modifications of nucleotides or incorporation of a foreign sequence in the RNA bait may modify the RNA structure and the RNP complex composition. - The recovery yield of the RNP complex might be low with aptamers (although not always defined) but could be improved with scaffolding technology). - The RNA fishing with antisense oligonucleotide does not require any RNA modification (chemical or sequence) but is restricted to RNA with accessible regions.
Protein extract	Cytosolic or nuclear extracts, depending on the biological question. Possible cross-linking with RNA.	More stringent washing conditions can be used when proteins and RNA are cross-linked, thereby limiting the presence of non-sequence specific interacting proteins.
Pre-clearing/ Blocking	Blocking of the immobilized bait with high concentrations of unspecific RNA (yeast tRNA), proteins (BSA), heparin, and/or salmon sperm DNA). Unspecific pre-clearing can be performed by (successive) chromatography steps; selection of the fraction retaining a RNA-binding activity.	- Blocking agents increase the noise during the MS-based identification step. - Pre-clearing is useful to decrease the complexity of protein mixture. - Successive chromatographies are time-consuming, require large amounts of material. - Unspecific pre-clearing represents a risk to lose the specific TFs. - Adsorbed blocking molecules might be eluted together with RNP complexes
NA–protein binding & washes conditions	Binding and washing buffer compositions could be adapted: - Ionic stringency - Detergent concentration - Unspecific competitors: RNAs, negatively charged molecules (heparin). - Sequence specific competitors	RNA competitors decrease the non-sequence specific binding, but they might also capture the proteins of interest. Therefore, the use of dedicated RNA sequences is advised.
Elution	- No elution: on bead digestion of RNP - Unspecific: salt, temperature, detergent, urea. - Specific: 1) displacement of aptamers with an excess of their ligand molecule (ex: biotin) 2) insertion of a cleavage sequence in the fishing device	- No elution: 17–26% of identified proteins are proteins adsorbed on the beads. - Unspecific elution: the proteins adsorbed on the solid support contaminate the eluate. - Specific elution: strongly advised to limit the number of adsorbed proteins identified, but this requires to modify the bait sequence

Table 2 : Major characteristics, advantages & limitations of RNA-affinity chromatographies. (Tacheny et al., 2013)

and as protein-centric techniques such as supershift or RNA-immunoprecipitation require a prior knowledge of potential protein partners, we will focus on solely RNA-centric approaches.

Several computational methods have been set up for mapping of RNA sequence interacting with proteins. Although such tools can provide information regarding the location of potential binding sites, sequence analysis is not sufficient to determine RBPs or predict all targeted RNA. Moreover, RNA folding, RBPs multidomains, RBDs accessibility and clustering contributing to the constitution of RNPs complexes are not taken into account (Faoro, 2014 & Tacheny et al., 2013). To cope with this problem, research over recent decades has turned to the use of high-throughput technologies such as Mass spectrometry (MS) analysis following a RNA-affinity purification (RAP-MS) (Butter et al, 2009).

Briefly, in RNA-affinity based RBP purification approaches, RNP complexes are assembled *in vivo* or *in vitro* and the RNA employed as a bait is immobilized on a solid support either covalently or not. Then, a series of washing steps are performed in order to eliminate non-specifically bound proteins which could possibly compromise the results. Finally, the RNP are eluted from the solid matrix and subjected to mass spectrometry (Tacheny et al., 2013). As MS-based proteomics enables the quantification and the study of complex proteins mixtures at high sensitivity, this powerful and wide spread device has clearly provided significant insight into mechanisms underlying binding phenomenon within RNP complexes. However, RAP-MS methods described in the literature display several drawbacks listed in table 2. The most important one probably comes from the fact that protein identification by mass spectrometry is made difficult when the peptide mixture contains a large number of different proteins (complex mix) spanning a large dynamic range. It is therefore essential to avoid as much as possible the presence contaminant/unspecific interacting proteins in the eluate, as those are usually largely abundant and mask the identification of low abundant proteins. In this master's thesis, we will take advantage of an efficient DNA-affinity method, due to a specific elution of DNA-proteins complexes, and try to suit it to RNA-affinity.

Objectives

Schmallenberg virus (SBV), which first emerged in Germany in 2011, rapidly spread across the European continent causing a large outbreak in ruminants on its path. Given the severe damages that it induced, researchers were prompted to determine its origin, transmission and pathogenicity. Genomic sequencing results revealed that Schmallenberg virus belongs to *Bunyaviridae* family, *Orthobunyavirus* genus and was part of the *Simbu* serogroup. Like any other member of this family, SBV is arthropod-borne virus and has a tripartite genome composed of negative sense RNA segments named large (L), medium (M) and small (S), according to their size, that encode 6 proteins in total. Each segment is constituted of a coding region flanked by 3' and 5' UTR involved in genome packaging as well as in initiation and termination of transcription process. These genomic RNAs are first transcribed into mRNAs, constituted of a 5' host-derived sequence and lacking a poly(A) tail at the 3' end. However, the poly(A) tail is usually crucial for the nuclear export, the mRNA stability, transcription termination and the mRNA translation initiation processes. So far, despite researches dedicated to the SBV replication, little is known about its poly(A)-independent transcription and translation mechanisms.

Nevertheless, *in silico* analyses recently reported a conserved hairpin and a putative transcription termination signal both localized in the 3' UTR of the antigenomic S segment, suggesting a potential functional relevance in viral replication. In order to investigate this hypothesis, our *first objective* is to develop an **R**NA-**a**ffinity **p**urification technique coupled with **m**ass **s**pectrometry analysis (RAP-MS) to identify the RNA-binding proteins interacting with either the stem-loop structure or the potential transcription termination signal.

This method will be adapted from a DNA-affinity purification followed by mass spectrometry (DAP-MS) technique that has been previously developed in the lab. It will involve the following steps.

First, an RNA bait suitable for RNA-affinity capture containing both the stem-loop and the putative transcription termination signal will be defined, amplified and cloned into a vector convenient for *in vitro* transcription.

In addition to wild-type (WT) RNA sequence, 5 mutated constructs will be generated in order to discriminate proteins binding specifically the sequence of interest from proteins interacting with both wild-type and mutated RNA, classified as unspecific and unlikely to participate to transcription/translation mechanisms. Among these mutants, two will explore the role of the transcription termination signal while 3 others will be conceived in order to investigate the hypothesis that the crucial determinant of the transcription and translation processes is the secondary structure rather than the primary sequence.

In parallel, we will have to select cytosolic or nuclear extracts, depending on the biological question, and to settle adequate binding conditions to favor interactions between the proteins and the RNA bait and consequently ensure the RNP complexes formation. Although the concept

seems quite simple, the real challenge lays in the fact that regulatory proteins are often present in low quantities compared to the highly abundant proteins binding the RNA in an unspecific manner.

Once WT and mutated RNA transcribed, the RNA molecules will be immobilized on a solid support, in a reversible manner. Therefore, we will design an antisense desthiobiotinylated deoxyribonucleotide probe and take advantage of the base pair complementarity of nucleic acids to immobilize the RNA bait on beads. It is important to design a strategy including a specific elution step of the RNP complexes, as a lot of abundant proteins are adsorbed on the chromatographic of support. An unspecific elution, by heat or other denaturing conditions, would contaminate the sample to be analysed by mass spectrometry. Therefore, we will evaluate the efficiency of the specific RNP complexes elution, that will be further processed for trypsin digestion, reverse-phase liquid chromatography separation, followed by a mass spectrometry analysis. To that aim, we will take advantage of reversibility of the desthiobiotin/streptavidin interactions and displace the RNP complexes by an excess of biotin, described to have a higher affinity for the streptavidin than the desthiobiotin does.

Finally, the *second objective* will consist in comparing the data obtained with the WT and the mutated RNA to identify the proteins most likely to be involved in SBV transcription termination or translation initiation mechanisms. Once the best candidates selected, their expression will be downregulated using siRNA and the potential influence on viral replication will be analyzed in infected cells. At the same time, the involvement of those proteins in the transcription and translation processes will be assessed.

Material & Methods

Purpose	Name	Primer sequence 5'→3'	T° of Melting (Tm)	Position (nt)
Reverse transcription	SFw	5'-AGTAGTGAACTCCACTATTAACTAC-3'	49.49 °C	1-25 of SBV S Segment
Reverse transcription	SRev	5'-AGTAGTGTTCTCCACTTATTAACTATC-3'	53.13°C	814-840 of SBV S Segment
PCR	PU	5'-TGTAAAACGACGGCCAGT-3'	54.3°C	2976-2993 of pGEM-T Easy (Empty)
PCR	RPU	5'-CAGGAAACAGCTATGACC-3'	47.38°C	176-194 of pGEM-T Easy (Empty)
PCR & Nested PCR	HFw	5'-CTAGGGGCCCGTATCAACATCTAAACCTCTTC-3'	72,91°C	1-22 of W-T PCR products
PCR & Nested PCR	HRev	5'-CTAGCCATGGAGTAGTGTTCTCCACTTATTAAC-3'	67,27°C	97-120 of W-T PCR products
Nested PCR	Fw5p	5'- ATTTCCGTAACCAATGTCTATGT-3'	55,47°C	36-58 of W-T PCR products
Nested PCR	Rev5p	5'- ACATTGGTTACGGAAATTGAAG-3'	56,58°C	31-52 of W-T PCR products
Nested PCR	Fw3p	5'- ATGTTGGTTACACCATTATACTGC-3'	56,07°C	55-78 of W-T PCR products
Nested PCR	Rev3p	5'-TAATGGTGTAACCAACATAGACAT-3'	57,04°C	49-72of W-T PCR products
Nested PCR	FwGc1	5'-CATTATACTTTAAGTTTTCTGTTAAGA-3'	52,79°C	69-95 of W-T PCR products
Nested PCR	RevGc1	5'-CAGAAAACTTAAAGTATAATGGTGT-3'	52,87°C	65-89 of W-T PCR products
Nested PCR	FwGc2	5'-TATACTCGAAGCGTTCTGTTAAGA-3'	56,64°C	72-95 of W-T PCR products
Nested PCR	RevGc2	5'-CAGAACGCTTCGAGTATAATGGT-3'	58,49°C	67-89 of W-T PCR products

Table 3 : Oligonucleotides used in Reverse Transcription (RT), Polymerase Chain Reaction (PCR) and nested-PCR.

Materials & methods

1. Reverse transcription and cDNA synthesis

To prepare the SBV S segment cDNA, we used RNA derived from the Na1 strain (SBV isolate from Namur) that had been previously extracted. The reverse transcription was carried out with the SuperScript III First-Strand Synthesis System (Invitrogen) and was initiated with specific primers for the cDNA synthesis of the S segment (SFw and SRev) (table 3). 1048.5ng RNA were mixed with 0.1µM of each primer and 0.2mM dNTP Mix. The sample was then incubated at 65°C for 5 minutes and cooled on ice for 1 minute before adding first-strand buffer 5x, 0.1M of dithiothreitol (DTT), 40U of RNase OUT and 200U of SuperScript III into the mix. The final solution was heated at 55°C for 60 minutes to allow the reverse transcription and then incubated at 70°C for 15 minutes to inactivate the enzymatic activity.

2. Polymerase Chain Reaction (PCR)

Several polymerase chain reaction (PCR) were executed, all performed with a PCR mix including 0.1µM of each primer, Green GoTaq buffer 1x (Promega), 0.2mM of dNTP mix, 1U (1U per final 50µL) of GoTaq DNA polymerase (Promega, 5U/µL) and Nuclease-Free water to reach the final volume.

The first one was executed to amplify the cDNA of the 120 last nucleotides of the S segment using the following pair of specific primers: FW 5'-CTAGGGGCCCGTATCAACATCTAAACCTCTTC-3' (HFw) and REV: 5'-CTAGCCATGGAGTAGTGTTCTCCACTTATTAAC-3' (HRev) (supplementary data, table 1). A total of 30 amplification cycles were performed: the first five cycles were composed of 30" at 94°C, 30"at 45°C, 30" at 72°C and the 25 others of 30"at 94°C, 30" at 58°C and 30'' at 72°c, and finally 8' at 72°C.

Other PCR were carried out for the screening of bacteria potentially containing the expected insert using the same PCR mix as previously described but composed of different primers : FW: 5'-TGTAAAACGACGGCCAGT-3' (PU) and Rev: 5'-CAGGAAACAGCTATGACC-3' (RPU). Thirty amplification cycles (30'' at 94°C, 30''at 55°C and 1' at 72°C) were performed.

3. Nested PCR to create muted sequences

Once the SBV cDNA synthesized, it was used as a template to create 5 mutants by nested PCR. To that end, specific primers containing the intended nucleotide substitutions were carefully designed (table...).
A first PCR was performed with a PCR mix including Green GoTaq buffer 1x (Promega), 0.2mM of dNTP mix, 1U of GoTaq DNA polymerase (Promega, 5U/µL), 0.1µM of each primer

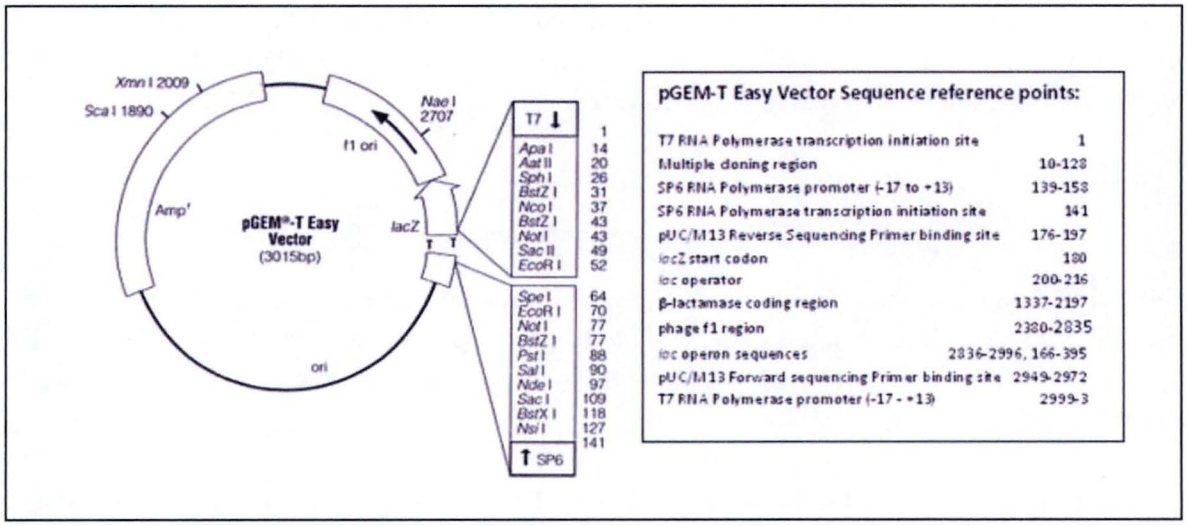

Figure 15 : pGEM-T Easy vector map. This vector was used in order to amplify the sequence of interest composed of the 120 last nucleotides of SBV S segment.

and Nuclease-Free water to reach a final volume of 50µL. In this case, the pair of primers were HFw and a reverse primer containing the intended mutations. A total of 35 amplification cycles were performed: the first five cycles were composed of 30" at 94°C, 30"at 48°C, 30" at 72°C and the 30 others of 30"at 94°C, 30" at 54°C and 30'' at 72°C, and finally 8' at 72°C.
Also, the same experiment was carried out with different primers: here, HRev and a forward primer composed of the expected substitutions were used.
Finally, a third PCR was performed using 1ng of each preceding purified PCR product. A total of 35 amplification cycles were performed: the first five cycles were composed of 30" at 94°C, 30"at 48°C, 30" at 72°C and the 30 others of 30"at 94°C, 30" at 54°C and 30'' at 72°C, and finally 8' at 72°C. Hfw and HRev primers were added after the first five cycles.

PCR products resulting from these three experiences were purified using the nucleospin Gel and PCR clean up kit (Macherey-Nagel) following the supplier's protocol (mentioned earlier).

4. Purification of PCR products

The nucleospin Gel and PCR clean up kit (Macherey-Nagel) were used to purify DNA fragments of PCR products from enzymatic reactions and agarose gels prior their introduction into pGEM-T easy plasmid. A controlled volume of binding buffer NT1 was mixed with the sample and incubated at 55°C to allow the agarose dissolution. Next, this solution was poured on a column composed of a silica membrane which binds the DNA fragments thanks to chaotropic salts contained in the NT1 buffer, followed by a 30" centrifugation at 11 000g. Two washing steps with 650µL of ethanol-containing buffer NT3 and two centrifugations at 11 000g for 30" and another at 11 000g for 1 minute were performed to remove any contaminants or residual ethanol from the sample. Once the silica membrane was dried, nuclease-free water was added into the column to elute the DNA. Then the column was incubated at room temperature for 3 minutes and centrifuged 1 minute at 11 000g to collect the purified DNA.

5. pGEM-T Easy vector purification

The pGEM-T Easy vector (figure 15) is a 3015bp high copy plasmid composed of a replication origin allowing the replication of the plasmid inside TG1 electro-competent *Escherichia Coli* (E. Coli). Furthermore, this vector also contains a gene encoding a β-lactamase providing an ampicillin resistance to transformed bacteria grown on agar medium added with ampicillin (200µg/mL). Hence, only transformed E. Coli will grow on this medium and will be selected. pGEM-T Easy vector also includes the Beta-galactosidase-encoding lacZ operon that is disrupted by the multiple cloning site (MCS) localized within this operon. Indeed, bacteria grown on medium supplemented with IPTG (40mg/L) and X-gal substrate (60mg/L) will constitute blue colonies if they have integrated the empty vector. On the contrary, they will be white if they contain the recombinant plasmid because the insertion of the sequence will discontinue the lacZ operon. Moreover, the vector was linearized by *EcoRV* and 3' terminal

thymidine was incorporated to both digested ends. This MCS is surrounded by a SP6 and a T7 RNA polymerase promoter. As a matter of fact, pGEM-T Easy plasmid was chosen due to the presence of this T7 RNA polymerase promoter required for the *in vitro* transcription.

After an overnight culture of TG1 competent bacteria, the pGEM-T Easy plasmid was extracted from bacteria and then purified thanks to the NucleoBond plasmid DNA purification kit (Macherey-Nagel) according to the manufacturer's instructions. The principle of this technique is based on a bacterial cells lysis followed by the loading of a column with this lysate. The anion-exchange resin contained into this column allows the binding of DNA which can then be collected by elution. First of all, the bacterial colony was picked from the plate and inoculated into a proper volume of LB medium and ampicillin (200µg/mL) to start an overnight bacterial culture. Thanks to a centrifugation at 6 000g for 15 minutes at 4°C, the bacteria were pelleted and then resuspended in Buffer S1 (50 mM Tris-HCl, 10 mM EDTA, 100 µg/mL RNase A, pH 8.0). Lysis buffer S2 (200 mM NaOH, SDS 1%) was incorporated into the sample and the whole solution was mixed gently and incubated at room temperature for 2-3 minutes in order to lyse the bacteria and denature chromosomal DNA as well as plasmid DNA. After the addition of neutralization buffer S3 (2.8 M KAc, pH 5.1) to induce the formation of a precipitate composed of chromosomal DNA and other cellular elements while the plasmid DNA remains in solution, the mixture was cooled on ice for 5 minutes. Meanwhile, the Nucleobond Column was equilibrated with buffer N2 (100mM Tris, 15% Ethanol, 900 mM KCl, 0.15% Triton X-100, adjusted to pH 6.3 with H_3PO_4). Once the column empty, this one was loaded with the lysate and then rinsed with buffer N3 (100 mM Tris, 15% Ethanol, 1.15M KCl, adjusted to pH 6.3 with H_3PO_4). This washing step is usually repeated a second time to eliminate impurities whereas plasmid DNA remains bound to the filter. Thereafter, thanks to the buffer N5 (100 mM Tris, 15% ethanol, 1M KCl, adjusted to pH 8.5 with H_3PO_4), the column was eluted and the DNA collected. Isopropanol was then added into the sample and the whole mixture underwent a 30 minutes centrifugation at 4°C at 15 000g to precipitate the DNA. The obtained pellet was then washed twice with ethanol 70%, followed by a 10 minutes centrifugation at room temperature at 15 000g. The ethanol 70% was removed and the pellet was dried. The purified plasmid DNA was finally resuspended in water.

6. Enzymatic cleavage of the insert and pGEM-T Easy vector

In order to clone the insert of interest into the final pGEM-T Easy vector, these two elements had to be digested by the same restriction enzymes to generate compatible overhangings. Therefore, *ApaI* and *NcoI* were chosen. Bovine Serum Albumin 10x (BSA) was diluted into restriction enzyme (*NcoI*/*ApaI*) buffer. 5.5µL of this mix was added with 1µL of restriction enzyme (*NcoI*/*ApaI*), 5µg of pGEM-T Easy plasmid and water to reach a final volume of 50µL. The sample was heated at 37°C (*NcoI*)/ 30°C (*ApaI*) for 1 hour. Finally, the solution was incubated at 80°C (for *NcoI*) or 65°C (for *ApaI*) to inactivate the enzymatic activity. Each enzymatic cleavage was followed by a phenol-chloroform-isoamylalcohol extraction step.

The pGEM-T Easy plasmid as well as the insert of interest have been digested by both restriction enzymes.

7. pGEM-T Easy dephosphorylation

46µg of doubly digested pGEM-T Easy plasmid, 15U of Calf Intestinal Phosphatase (CIP) (10U/µL NEB) and CIP buffer 10x were added into a nuclease-free microtube. This mix was heated at 37°C for 20 minutes, 56°C for 20 minutes and then cooled on ice for 3 minutes. Next, 10U of Calf Intestinal Phosphatase (CIP) (10U/µL, NEB) was added into this sample and the whole solution underwent the same incubation cycle as before. Finally, the phosphatase activity was inhibited by incubation of the mixture at 70°C for 20 minutes and the product was purified by phenol/chloroform extraction.

8. Phenol-chloroform-isoamylalcohol extraction

The principle of this technique is based on DNA extraction by phase separation. First, an equal volume of phenol-chloroform-isoamylalcohol is added into the tube containing the sample, thoroughly mixed and then centrifuged at 12 000g at 4°C for 5 minutes to separate the layers. The upper layer is carefully removed and placed into a new microtube and added with an equal volume of chloroforme-isomylalcohol (49-1). This step is followed by a centrifugation step at 12 000g at 4°C for 5 minutes. Once more, the supernatant is collected and placed into a new tube containing 0.9 volume of isopropanol added with 3 M sodium acetate. The whole sample was then incubated at -80°C for at least 1 hour to allow a complete precipitation of the DNA and centrifuged at 12 000g at 4°C for 15 minutes. Directly after the centrifugation, the supernatant is removed and the DNA precipitated into a pellet remains at the bottom of the tube (not always visible). Next, two washes with 75% ethanol each followed by a 10 minutes centrifugation at 10 000g are performed. Finally, the pellet is dried and resuspended into nuclease-free water.

9. Ligation of PCR products into a dephosphorylated pGEM-T Easy vector

To perform this ligation, a 3:1 ratio of purified amplicon and dephosphorylated and double digested pGEM-T Easy plasmid, T4 DNA ligase buffer 10x were mixed with 400 U of T4 DNA Ligase (NEB, 400 U/µL) and this sample was incubated at 12°C overnight.

10. Bacteria electroporation and culture

First, the ligation product was mixed with competent TG1 *E. coli* prior to inject this solution into a 2mm gap pre-cooled electroporation cuvette. Next, the cuvette was tapped on the bench to remove the bubbles present in the mixture and placed into the electroporator. Indeed, to allow the bacteria to integrate the plasmid potentially containing the insert of interest, *E. Coli* membrane had to be permeabilized thanks to an electroporation step performed at 2500V for 5ms. Immediately after electroporation, 150µL LB culture medium were added to ensure the recovery of the bacteria which were then cultured overnight on LB medium also containing X-

gal (60µg/mL), IPTG (40µg/mL) and ampicillin (200µg/mL). The next day, a PCR screen was performed to select colonies of bacteria which had incorporated the pGEM-T Easy plasmid with the insert of interest. To that end, the colonies were resuspended in a PCR mix previously described.

11. Preparation of template DNA and *in vitro* transcription

Linearized plasmid DNA and PCR products containing a T7 RNA polymerase promoter site can be used as templates to synthesize capped RNA thanks to the mMESSAGE mMACHINE®Kit. In this case, the pGEM-T Easy plasmid enclosing the insert of interest was digested with NcoI and then examined on gel to confirm the complete linearization of the plasmid.

Regarding the *in vitro* transcription reaction, the RNA polymerase Enzyme Mix and the thawed ribonucleotides (2X NTP/CAP) were stored on ice while the Reaction Buffer 10x was kept at room temperature to prevent a potential coprecipitation between the template DNA and the spermidine contained in the buffer if the reaction was executed on ice. Then, ribonucleotides (2X NTP/CAP), nuclease-free water, Reaction Buffer 10x, 1 µg of linearized plasmid and 2µL of RNA polymerase Enzyme Mix were sequentially added into a microtube before mixing thoroughly the solution. The sample was then incubated at 37°C for 6 hours to obtain a maximum amount of synthesized transcripts. 2U of TURBO DNase were added to eliminate the DNA template. Finally, 74µg of RNA were recovered by Spin column chromatography: RNeasy Micro Kit was used according the supplier's protocol. First, before pouring on a column composed of a silica membrane and centrifugation at 10 000rpm for 15 seconds, ethanol 60% were added to the RNA in order to promote selective binding. Once the flow-through discarded, 700 µL of Buffer RWT were added to the RNeasy MinElute spin and centrifuged at 10 000 rpm for 15 seconds. After the flow-through removal, 500µL of Buffer RPE were deposited into the column and centrifuged at 10 000 rpm for 15 seconds. Once again, the flow-through is eliminated, 500µL of ethanol 80% were incorporated into the column, and centrifuged 2 minutes at 10 000 rpm and then 5 minutes at 10 000 rpm to completely dry the membrane. Any contaminants washed away, total RNA is eluted in RNase-free water.

12. RNA immobilization tests

Strategy A : 15 pmoles of capture probe (5'- GATGAAGAGGTTTAGATGTTGATAC Desthiobiotin -3') were incubated in phosphate buffered saline $(PBS)_{50}$ (10 mM PO_4 pH 7.4, 50 mM NaCl) for 1 hour at room temperature on a rotary wheel with 0.5 mg of streptavidin-coupled magnetic beads (Dynabeads M-280 Streptavidin, Dynal) that had previously undergo four successive washes with PBS_{50}. In order to remove the unbound desthiobiotinylated oligonucleotide probe prior the addition of the RNA, three washes with PBS_{50} were executed. 20 pmoles of the RNA were then added and the whole mix was heated at 60°C for 5 minutes before undergoing a gradual cooling to reach the room temperature. Once more, 2 washes were

performed to eliminate the unbound RNA. In order to separate the DNA–RNA complexes from beads, an excess of biotin (5mM) was added for 2 hours at room temperature on a rotary wheel. The supernatant containing the RNA was collected.

Strategy B : 15 pmoles of desthiobiotinylated oligonucleotide capture probe (5'-GATGAAGAGGTTTAGATGTTGATAC Desthiobiotin -3') and 20 pmoles of RNA in phosphate buffered saline $(PBS)_{50}$ (10 mM PO_4 pH 7.4, 50 mM NaCl) were heated at 70°C for 5 minutes before undergoing a gradual cooling to reach the room temperature to allow the hybridization step between the two nucleic acids. The mixture was then incubated for 1 hour at room temperature on a rotary wheel with 0.5 mg of streptavidin-coupled magnetic beads (Dynabeads M-280 Streptavidin, Dynal) that had previously undergo four successive washes with PBS_{50}. In order to remove the unbound heterocomplex composed of the capture probe and the RNA, three washes with PBS_{50} were performed. Next, an excess of biotin (5 mM) was added for 2 hours at room temperature on a rotary wheel to separate the DNA–RNA complexes from beads. Finally, the supernatant containing heterocomplexes was collected.

In order to assess the efficiency of both strategies, supernatants coming from washes were analyzed thanks to 2100 Agilent bioanalyzer®. The Bioanalyzer is a device assessing quality control of RNA and DNA samples. The samples are separated by capillary electrophoresis in precast. The detection of the samples is made possible by the incorporation of a fluorescent dye, used to stain nucleic acids.

13. Cell culture

BHK-21 fibroblasts, an adherent cell line derived from baby hamster kidney, were required to ensure the *in vitro* multiplication of SBV. BHK-21 cells were cultured at 37°C and 5% CO_2 in Glasgow's Minimal Essential Medium (GMEM) supplemented with 10% fetal bovine serum (FBS), tryptose phosphate broth (2.95 g/L) and 1% penicillin (5000 U/mL) - streptomycin (5mg/mL) at 37°C and 5% CO_2. Once confluent, cells were rinsed with Phosphate Buffered Saline – Ethylene Diamine Tetra Acetic acid (PBS-EDTA) before addition of trypsin. Once the cells were detached, the trypsin activity was inhibited by the addition of complete GMEM. The cells are then resuspended and transferred into new flasks containing complete GMEM.

14. Infection of BHK-21 cells

BHK-21 cells at 70% confluency were infected with a controlled multiplicity of infection (MOI) of SBV (provided by the *Friedrich*-Loeffler-Institute). The viral suspension was incorporated into Dulbecco's Modified Eagle Medium (DMEM) supplemented with 1% FBS and 1% penicillin (5000 U/ml) - streptomycin (5000 µg/mL) and then administered to the BHK-21 cells. After 2 hours of incubation at 37°C with 5% CO_2, the infection medium was replaced by complete GMEM. BHK-21 cells were then incubated at 37°C with 5% CO_2 for 14 hours.

15. Immunofluorescence analysis

BHK-21 cells were seeded in 24 wells plates, on cover slips previously incubated with decomplemented FBS at 37°C for two days. Once cells reached 70% confluency, they were infected with SBV viral suspension as described above. 15 hours post-infection, GMEM was removed and cells were washed three times with PBS and then fixed using 500 µL per well of paraformaldéhyde 4% in PBS (PAF) for 10 minutes at room temperature. After overnight incubation at 4°C, PAF was removed, cells were rinsed three times with PBS and then permeabilized using PBS added with Triton 1% for 10 minutes at 37°C. Cells were washed three times with PBS added with 2% bovine serum albumine (PBS-BSA) and incubated with the rabbit primary antibody raised again the Nucleoprotein from SBV (Delphine Depierreux, master thesis 2014) diluted in PBS-BSA 2% at 1:100 for 2 hours at room temperature. After three washes of 10 minutes in PBS-BSA 2%, cells were incubated with the secondary antibody (Westburg, li926-32211 anti-rabbit IgG) and red phalloidin diluted in PBS-BSA 2% at 1:1000 and 1:100 respectively, for 1 hour at 37°C in the dark. Cells were rinsed once with PBS-BSA 2% and twice with PBS. Finally, BHK-21 cells were incubated with Hoechst diluted in PBS-BSA 2% at 1:500 and washed three times using PBS. Cover slips were deposited on mounting medium (Mowiol) on glass slides. Infected cells incubated without primary or secondary antibody were used as control. Few hours later, cells were observed using confocal microscopy.

16. Nuclear and cytosolic protein extraction

First, the uninfected and infected BHK-21 cells were washed once with cold PBS and once with PBS added with 5mM NaF and 1mM Na_2MoO_4. Once the PBS removed, Hypotonic Buffer (HB) 1X (20mM Hepes, 5mM NaF, 1mM Na_2MoO_4, 0.1mM EDTA) was added for 5 minutes to let the cells swell and weaken the plasma membrane. Then, the HB 1X was replaced by the lysis buffer, consisting of HB 1X containing 0.2% NP-40, to disrupt the plasma membrane without affecting the nucleus. After 5 minutes incubation on ice, the cells were scrapped and centrifuged for 30 seconds at 13000 rpm. The supernatant containing the cytosolic proteins as well as the nuclei-containing pellets were collected. For Western blot analyses, the pellet was resuspended into Dye Labelling Amino Acid (2M Thiourea, 7M Urea, CHAPS 4%, 30mM Tris) to lyse the nuclear membrane and harvest the proteins. The protein content of both fractions was assayed with the Pierce protein assay.

17. Western blot analysis

To assess the quality of nuclear and cytosolic proteins extracts, a western blot was performed. To that end, the proteins were resolved with a Mini-PROTEAN® TGX™ Gel 10%, BIORAD, #456-1033. A total amount of 10 µg of proteins from each sample were loaded per lane and SeeBlue Plus2 Pre-stained Protein Standard was used as molecular ladder. After the electrophoresis run at 100 volts for 1 hour, the proteins were actively transferred on a PVDF membrane (Immobilon, IPFL00010) using transfer buffer 1x (500 mM Glycine, 500 mM Bis-

Tris, 20.5 mM EDTA, 1 mM 4-chloro-1-butanol, 20 mL Methanol, 200 µL Antioxident agent) for 1 hour (at 64 mA). After the blocking step with Licor diluted 4x in 0.1 M phosphate buffered saline (PBS) (137 mM NaCl, 250 mM KH_2PO_4, 0.5 M $KHPO_4$) for 1 hour, the membrane was washed 3 times with PBS-TWEEN (0.1% TWEEN, PBS). This step was followed by the incubation of the membrane with the first rabbit primary antibody specific for NUP98 (Cell-Signalling 2598S, diluted 1:1000), a nuclear protein, overnight at 4°C and then the second mouse primary antibody raised against β-actin (Sigma A5441, diluted 1:30000) for 1 hour at room temperature. Next, the membrane was washed 3 times with PBS T for 5 minutes and twice with PBS for 2 minutes. The proteins were visualized thanks to the relevant secondary antibodies (Westburg, li926-68070 anti-mouse IgG and li926-32211 anti-rabbit IgG diluted 1:10000) conjugated with a different fluorochrome and the Odyssey® CLx Infrared Imaging System. After a rehydration step into methanol for 1 minute and PBS-Tween for 5minutes, the same procedure was repeated with a third rabbit primary antibody raised again the Nucleoprotein from SBV and an adapted secondary antibody.

18. RNA-affinity approach

30 pmoles of desthiobiotinylated oligonucleotide capture probe (5'-GATGAAGAGGTTTAGATGTTGATAC Desthiobiotin -3') were mixed with 40 pmoles of RNA in a final volume of 60 µL of phosphate buffered saline $(PBS)_{50}$ (10 mM H_2PO_4 pH 7.4, 50mM NaCl). This whole mixture was heated at 70°C for 5 minutes before undergoing a gradual cooling to reach the room temperature to allow the hybridization step between the two nucleic acids strands. The sample was then incubated for 1 hour at room temperature on a rotary wheel with 0.5 mg of streptavidin-coupled magnetic beads (Dynabeads M-280 Streptavidin, Dynal) that had previously undergo six successive washes with PBS_{50}.

To remove the unbound heterocomplex composed of the capture probe and the RNA, three washes with 400µL PBS_{50} were performed, followed by a last wash with 10Mm $MgCl_2$, 150mM NaCl, 20mM HEPES. Next, a volume equivalent to 1mg of cytosolic extracts was placed in the binding buffer (10 mM $MgCl_2$, 150 mM NaCl, 20 mM HEPES, 40 U of RNAse inhibitor (fermentas) and 50 µg of yeast tRNA (Invitrogen), final concentrations) on ice for 15 minutes before addition to magnetic beads for a 1 hour incubation on a rotary wheel at room temperature. The supernatant, considered as the flow-through of this chromatography, was collected for western blot analyses. Streptavidin-coupled magnetic beads were then washed once with 250 µL of binding buffer, three times with PBS_{50} + 0.1% Tween20 and twice with 50 mM NH_4HCO_3.

In order to separate the DNA–RNA complexes from beads, an excess of biotin (5mM) was added for 2 hours at room temperature on a rotary wheel and the supernatant containing heterocomplexes was collected. Finally, beads were rinsed with 10µL of 5mM biotin and both supernatants were pooled and prepared for digestion.

1 µL of PPS (3-[3-(1,1-bisalkyloxyethyl)pyridin-1-yl]propane-1-sulfonate) Silent Surfactant (Protein Discovery; 0.8 % final concentration)) was mixed with the supernatants before boiling

at 100°C for 5 minutes to denature the proteins. Then, DTT (Sigma; 5mM final concentration) was added into the microtube for a 30 minutes incubation at 50°C at 300 rpm. The reduced proteins were then alkylated for 30 minutes at room temperature in the dark with iodoacetamide solution at 15mM final concentration before being digested overnight at 37°C using trypsin (1 mg/ml in 50 mM NH_4HCO_3, 1 mM $CaCl_2$; Trypsin Gold, Mass Spectrometry Grade; Promega).

The next day, new streptavidin-coated magnetic beads underwent six consecutive washes: four with PBS_{50} followed by two with 50 mM NH_4HCO_3. Once the beads rinsed, they were mixed with the digestion product and incubated for 30 minutes at 21°C on a rotary wheel to remove the excess of free biotin.

The final step regards acidification of the sample by the addition of 1µL of 12 N HCl into the tube and incubation for 1 hour at 37°C at 300 rpm to hydrolyze the PPS. After centrifugation at 13 000 rpm for 5 minutes at 4°C, the supernatant was collected into a 75 mm x 500 mm C18 Dionex column (Acclaim PepMap 100 C18) and subjected to MS run.

19. Liquid chromatography coupled with mass spectrometry analysis and protein identification

Generated digestion products were analysed thanks to nano-liquid chromatography (LC) system Ultimate 3000 (Dionex) connected to a maXis Impact electrospray Ultra-High Resolution Q-TOF mass spectrometer (Bruker). Peptide separation occurs at 35°C through reverse-phase LC in the Ultimate 3000 LC system. First mobile phase was composed of 0.1% formic acid in water while the second one contained 0.1% formic acid in 80% acetonitrile. 15 µL of the sample were injected to the LC and the organic content of the mobile phase was linearly increased from 4% B to 40% during 280 minutes, from 40% B to 90% B for the next 10 minutes, maintained at 90% B for 10 minutes and then decreased at 4% B for the last 45 minutes. As the column is directly coupled with an electrospray ionization (ESI) Captive Spray (Bruker), the flow-through underwent an ionization step leading to the transformation of this liquid into electrically charged droplets. After a long reverse-phase chromatography separation, the peptide were sequenced by tandem mass spetcrometry.

Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by Data analysis version 4.2. All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.5.1) and X! Tandem (The GPM, thegpm.org; version CYCLONE (2010.12.01.1)). Mascot was set up to search the UniRef100 database (selected for Human or Rodentia or Bos or Schmallenberg (november 2015, 871143 entries)) Mascot and X! Tandem were searched with a fragment ion mass tolerance of 0,050 Da and a parent ion tolerance of 10,0 PPM. Gln->pyro-Glu of the N-terminus, oxidation of methionine and carbamidomethylation of cysteine were specified in Mascot as variable modifications. Scaffold (version Scaffold_4.4.8, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95,0 % probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 5,0 % probability to achieve an FDR less than 3,0 % and contained at least 2 identified peptides.

Results & Discussion

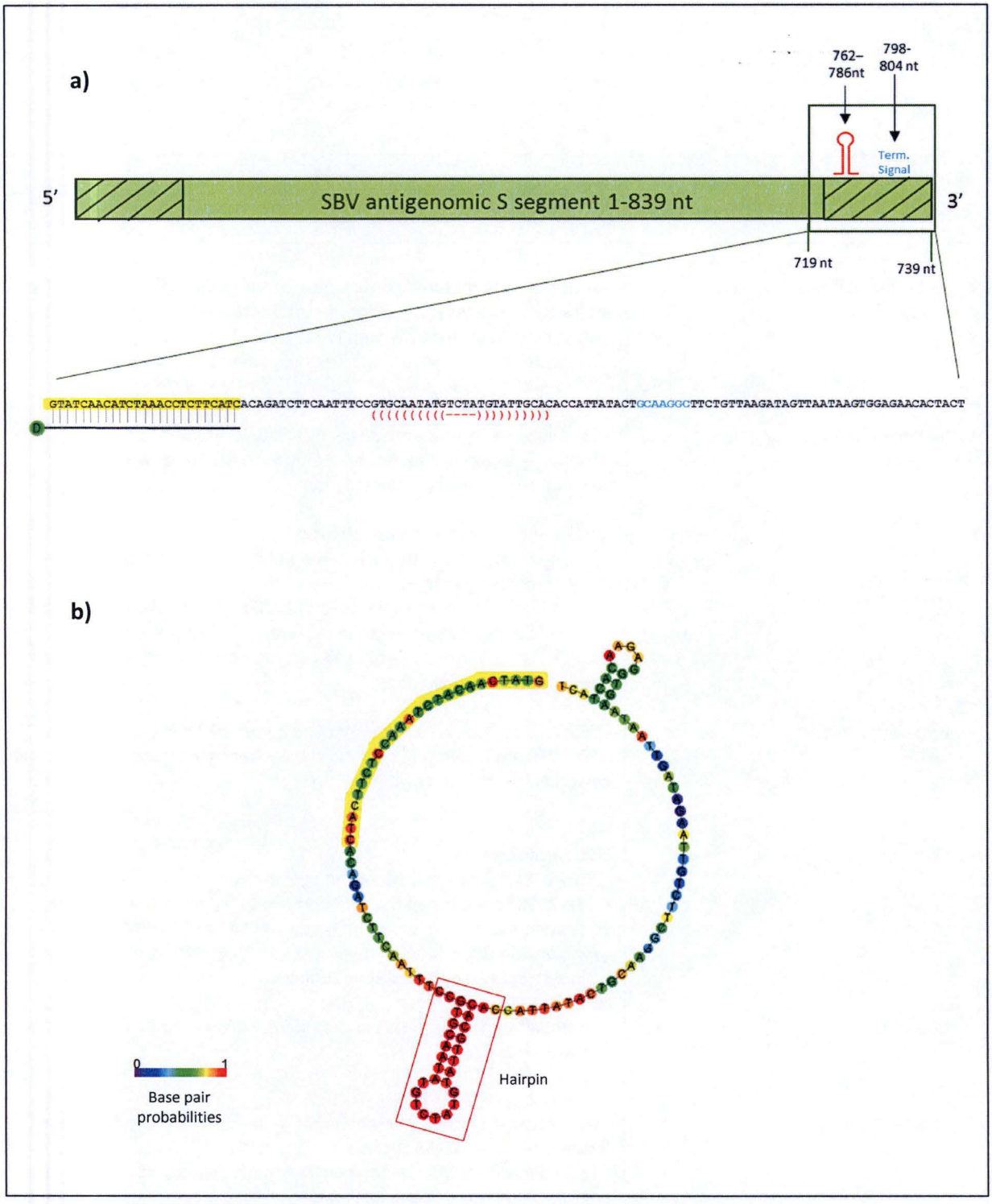

Figure 16 : Schematic representation and secondary structure prediction of the wild-type RNA sequence of interest. **a)** Representation of the wild-type RNA sequence of interest comprising the 120 last nucleotides of the SBV S segment. The hairpin is pictured in red whereas the putative transcription termination signal is represented in blue. The highlighted sequence in yellow is complementary to the antisense DNA capture probe. **b)** Secondary structure prediction of the WT RNA sequence at 37°C using the RNAfold software. The base pairing probabilities are shown below the sequence. The highlighted sequence in yellow is complementary to the antisense DNA capture probe.

Purification steps	Features
▪ Bait sequence selection and synthesis	- Wild-type and mutated constructs in order to distinguish specifically interacting proteins from unspecific interactants - Must be suitable for RNA-affinity capture - To synthesize by *in vitro* transcription, with the T7 polymerase.
▪ Bait immobilization on a solid support	- Solid support : paramagnetic beads will be preferred to chromatographic column because they authorize large washing volumes and a small elution volume - RNA Immobilization can be achieved by : ▪ Incorporation of a tag → modification of RNA sequence and conformation ▪ Chemical modifications → utilisation of modified nucleotides during the transcription process ▪ Complementary probe → no modification of RNA sequence and conformation
▪ Protein extract	- Cytosolic or nuclear extracts, depending on the biological question - The selected incubation conditions have been described by Butter and coworkers (Butter et al., 2013)
▪ Elution	Elution can be : - Unspecific: heating, denaturing agents - Specific: to separated the RNP complexes from the beads before protein analysis. An excess of biotin can be used to displace oligonucleotides functionalized with desthiobiotin and immobilized in a streptavidin-coated solid support
▪ Digestion	- Improved trypsin digestion in case some RBPs would be tightly bound to the RNA sequence
▪ LC-MSMS analysis	- Long reverse-phase chromatography separation to make the sample less complex for MSMS analysis - Two consecutive MS run and generation ofa scheduled precursor list containing the most sequenced ions

Table 4 : Steps of RNA-affinity purification coupled with mass spectrometry analysis and associated features.

Results & discussion

As a reminder, the 3' terminal ends of SBV mRNAs are not provided with a polyadenylation tail (poly(A) tail) and the transcription and translation processes are still poorly described in the literature. Fortunately, recent computer-assisted secondary structure prediction analysis revealed a conserved stem-loop structure comprising nucleotides 762–786 of the S segment and the existence of a putative termination signal (5'-GCAAGGC-3', nucleotides 798-804) both located in the 3' UTR (figure 16). The conservation of such RNA structures among the *Simbu* serogroup and their localization suggest a potential functional relevance in viral replication (Coupeau et al., 2013).

Our work aims to develop a method to identify the proteins interacting with both elements to decipher the mechanisms involved in the transcription and translation of the Schmallenberg virus.

1. Experimental design

The following paragraphs summarize the main steps of the RAP-MS strategy we aim to step up, and the critical elements considered to select the most suitable options (table 4).

First, the choice of the studied sequence is an essential step as it has to be suitable for immobilization on a solid support (see section *Selection of the RNA sequence to be immobilized*). Besides, both wild-type (WT) and mutated nucleotide constructs have to be generated in order to identify the proteins likely to participate to transcription/translation mechanisms.

Therefore, we have conceived several mutations with the intention to abolish the functions hypothetically associated with the stem-loop structure and the transcription termination signal. Let's mention that the same mutations are currently being functionally tested by another researcher of the lab, by reporter gene assay.The first two designed mutants (named Gc1 & Gc2) explore the role of the transcription termination signal (figure 17). As this signal could also form a short secondary structure, one mutation will be inserted in one mutant to break it whereas the second mutant will contain restored stem-loop structure.

The mutations designed to abolish the functions hypothetically associated with the stem-loop structure are based on literature analysis. A research conducted on Infectious bursal disease virus (IBDV; from the *Birnaviridae* family), reported that each inserted mutation in order to destabilize the secondary structure gave rise to a complementary mutation (or reversion) to maintain the stem-loop stability. No difference was noticed in the replication efficiency of these viruses compared to the WT. On the contrary, mutants with point mutations at the base of the hairpin or an elongated hairpin failed to yield viable viruses (Boot et al., 2004). Intriguingly, these findings are contradictory with a study conducted on Hepatitis C virus (HCV; from the *Flaviviridae* family) by Yi, Lemon and coworkers. Indeed, they revealed that substitutions at

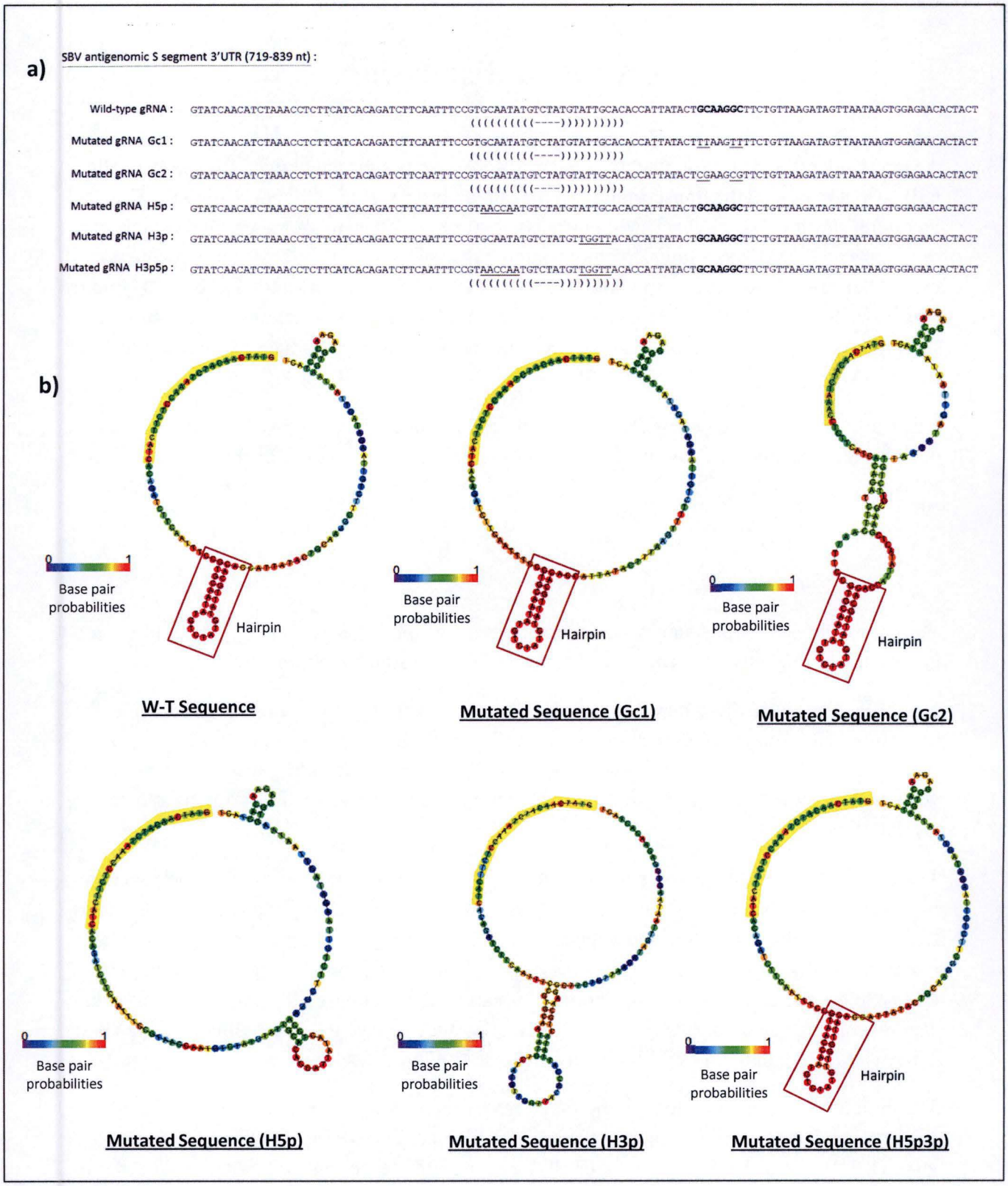

Figure 17 : Sequence and secondary structure predictions of both wild-type and mutated RNA sequences of interest. a) Alignment of wild-type or mutated sequences comprising the 120 terminal nucleotides from the SBV 3' UTR. The predicted conserved stem-loop structure is indicated below each sequence by brackets whereas nucleotides in bold represent the putative termination signal. For the mutated sequences, the underscored nucleotides correspond to the inserted mutations. **b)** secondary structure prediction of both wild-type and mutated RNA sequences of interest. The highlighted sequence in yellow is complementary to the antisense DNA capture probe. The base pairing probabilities are shown below each sequence.

some positions were well tolerated, as substitutions at the internal loop positions did not abolish replication of the RNA. By contrast, substitution at the base of the loop, near the ends of the duplex stem or a complete replacement of each of the positions within the loop by its Watson-Crick complement, did in both cases. Moreover, mutants characterized by a conserved stem-loop structure but modified primary sequence at the apical half of the duplex stem were lethal for replication whereas mutants with substitutions at the basal half maintained a low level of replication (Yi & Lemon, 2003). However, another study showed that restored stem-loop mutations resulted in drastic decline of translation in HCV (Luo et al., 2003).

Taken together, these results indicate that interactions between the bases located in the lower half of the loop might be critical for conformation of the loop and the recognition by the RNA-polymerase whereas the upper half of the loop does not seem to participate to RNA–RNA or RNA–protein interactions. Besides, these observations also suggest that even if RNA structure is required, primary nucleotide sequence of the duplex stem is also a major player for effective replication as secondary structure with comparable analogous stability but distinct specific sequence does not completely replace the initial hairpin.

To summarize, conservation of both the primary sequence and secondary structure of the stem-loop structure might to be fundamental for the viral viability and efficient replication. We will therefore test these two possibilities by comparing the proteins interacting with the WT sequence and with three different mutant sequences. The 3rd and the 4th mutants (named H5p & H3p) contain a mutation on one strand of the hairpin structure (either 3'p or 5'p) in order to break it (figure 17), whereas two other mutations were incorporated into the 5th mutant (H3p5p) to modify the primary sequence while preserving the stem-loop structure (figure 17). These last 3 mutants (H5p, H3p & H3p5p) have been conceived in order to test the hypothesis that the crucial determinant of the transcription and translation processes is the secondary structure rather than the primary sequence. As a reminder, this hypothesis relies on the observation that the introduction of a point mutation leads to the acquirement of a compensating mutation (Coupeau et al., 2013) (see point 1.8. *SBV, a poly(A) tail independent virus*).

Once WT and mutated RNAs obtained, the RNA bait has to be bound on a solid support. So far, several protocols to immobilize an RNA molecule have been reported (summarized in table 4), each one based on either chemical modifications, the incorporation of a tag into the RNA sequence or the capture by a complementary probe. Nevertheless, each method does bear limitations. Indeed, chemical modifications (ex: biotinylated nucleotides) or the introduction of a foreign sequence in the RNA might reshape the RNA conformation and subsequently modify both the RNP complex genesis and the putative protein partners (reviewed in Tacheny et al., 2013). On the other hand, using an antisense probe has the advantage to avoid any modification of RNA of interest, but one has to make sure that a region of the RNA is accessible for base pairing with the antisense probe.

Then, to ensure the RNP complexes formation, the immobilized bait has to be incubated with proteins in a binding buffer that is compatible with protein-RNA interactions. We have used the incubation conditions described by Butter and coworkers (Butter et al., 2009). Moreover,

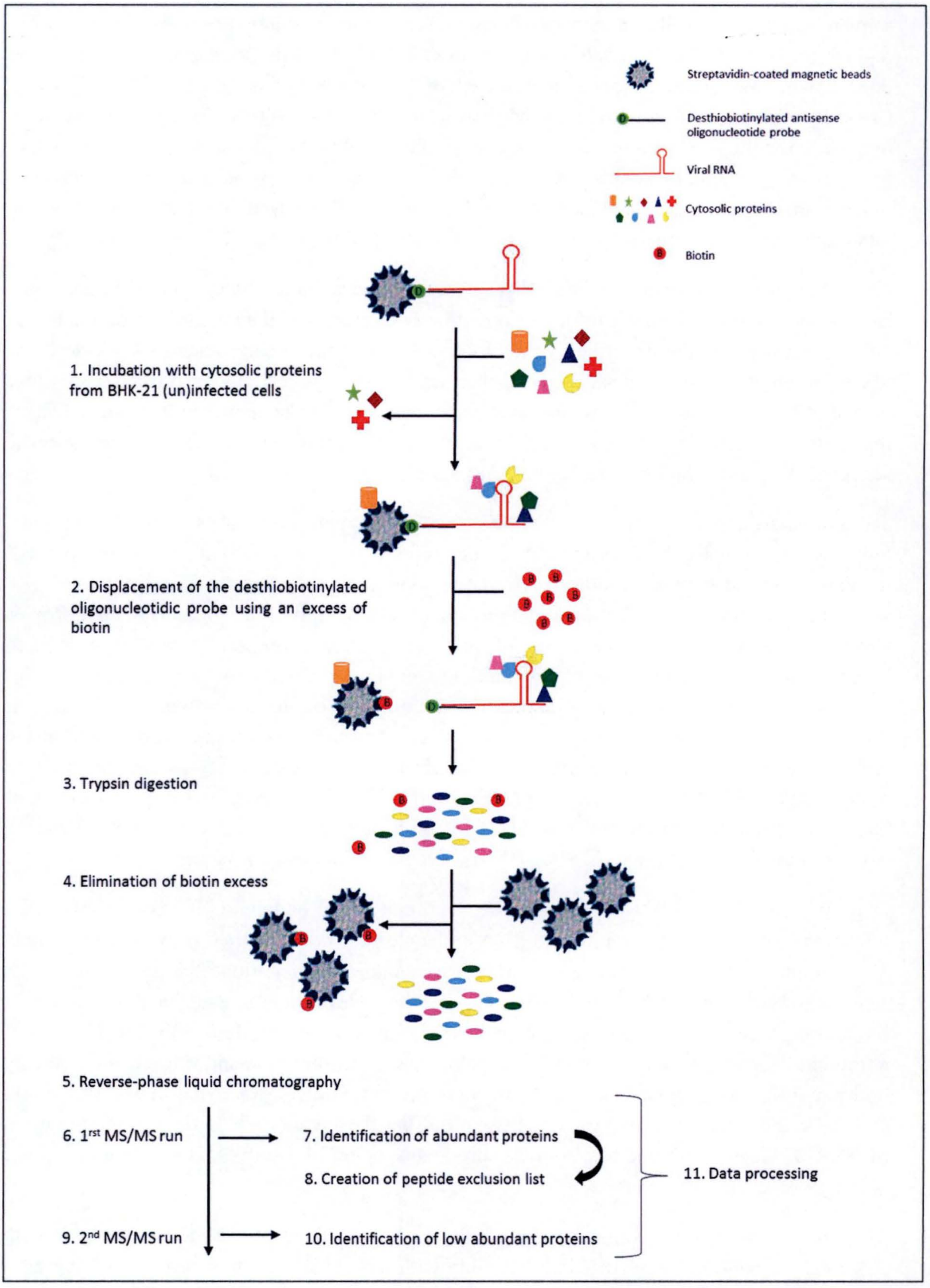

Figure 18 : Schematic representation of the method used to identify RNA-binding protein. (adapted from Tacheny et al., 2013)

the type of selected protein extracts (nuclear, cytosolic or total) is driven by the biological question.

Next, in most cases, the RNA-protein complexes are then eluted unspecifically (by heating, denaturing agents,...) before digestion and mass spectrometry analyses of proteins. The main drawbacks of such strategies is that the abundant proteins unspecifically adsorbed on the chromatographic support are denatured as well by such treatment and thus contaminate the eluate potentially containing the interacting partners of interest. As the latter might be less abundant, their identification by mass spectrometry (MS) in a complex mixture is compromised. Therefore, we have decided to set up a strategy compatible with a specific elution step (Figure 18) based on the reversibility of the desthiobiotin/streptavidin interaction. Indeed, as desthiobiotin has a lower affinity for streptavidin than biotin, an incubation with an excess of biotin will displace the desthiobiotinylated probe from the chromatographic support, streptavidin-coated magnetic beads, without denaturing the proteins (Hirsch et al., 2002). This approach has previously been used successfully in the lab to set up a DNA-affinity capture of transcriptional regulators, followed by mass spectrometry analysis (Tacheny et al., 2012). Therefore, we will generate an antisense desthiobiotinylated deoxyribonucleotide probe to take advantage of the base pair complementarity of nucleic acids to immobilize the RNA bait on a solid support. Indeed, the RNA will be captured by hybridization between an endogenous region of the RNA of interest and a desthiobiotinylated deoxyribonucleotide probe. In turn, the desthiobiotinylated deoxyribonucleotide probe will be immobilized on streptavidin-coated magnetic beads through streptavidin-desthiobiotin interactions. After incubation with the protein extracts, the RNA-protein complexes will be eluted by an excess of biotin (figure 18).

Proteins purified by RNA-affinity will then be subjected to improved trypsin digestion (Tacheny et al., 2012), before MS analysis (figure 18). MS identification was chosen over gel-dependent approaches as the latter were reported to suffer from poor detection of low abundant proteins (reviewed in Tacheny et al., 2013). Finally, as we expect a complex peptide mixture, two consecutive MS runs will be performed. The first one will allow the sequencing of the most abundant peptides present in the sample and, based on these results, a scheduled precursor list (SPL) containing the most sequenced ions will be defined (figure 18). In order to enable the MS to focus on the less abundant ions, any ions included in the SPL will be excluded from the second run analysis.

2. Selection of the RNA sequence to be immobilized

The 120 last nucleotides of SBV S segment containing the stem-loop structure and the putative transcription termination signal were selected as the sequence of interest (figure 16). This sequence includes the whole 3' UTR, and a small extension upstream the UTR that will be used to immobilize the RNA by hybridization with the oligonucleotide. Thanks to *in silico* analyses (RNAfold software), we have first verified the predicted secondary structure of the RNA and the accessibility for base pairing with an antisense probe.

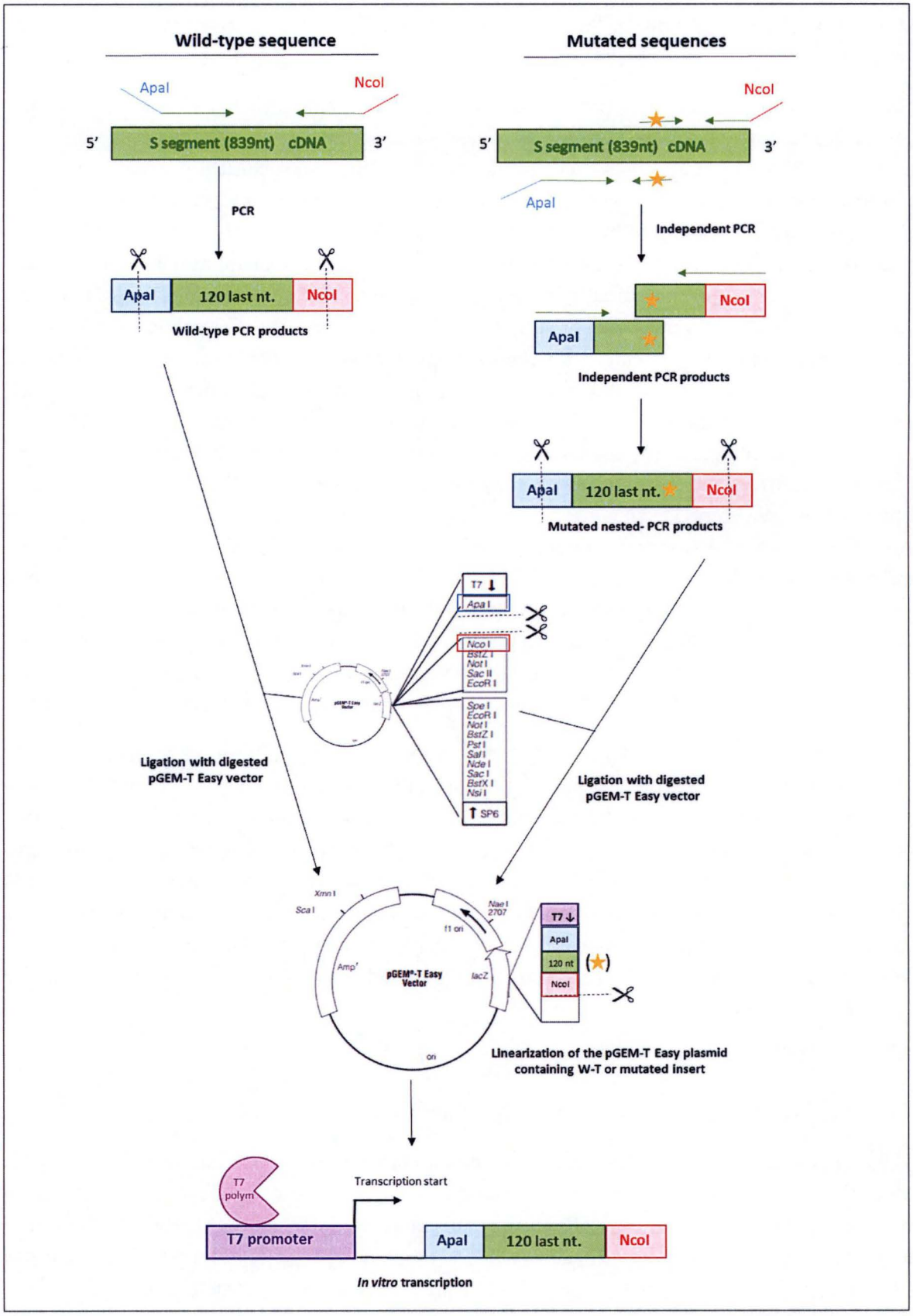

Figure 19 : Schematic representation of the different steps to synthesize the SBV S segment RNA fragment that has been used for RNA-affinity capture strategy.

3. Cloning

3.1. Construct of the insert of interest

First of all, the S Segment RNA derived from the SBV Na1 strain was retro-transcribed into cDNA. Then, thanks to carefully designed primers, a PCR was performed to amplify the 120 last nucleotides of this segment containing the stem-loop structure and the putative transcription termination signal. At the same time, *ApaI* and *NcoI* restriction sites, required for the future cloning in the final pGEM-T Easy plasmid, were introduced into the PCR products (figure 19). This RT-PCR amplification generated a fragment of 140bp, as expected (data not shown).

This insert was subsequently purified and digested by *ApaI* and *NcoI* restriction enzymes.

3.2. Preparation of the final pGEM-T Easy vector

Before the use of the final pGEM-T Easy vector required for the *in vitro* transcription of the sequence of interest, this plasmid needed to be prepared. First of all, after an overnight culture of TG1 competent bacteria transformed with the pGEM-T Easy vector, the plasmid was extracted and purified.

Secondly, the vector was restricted by *ApaI* and *NcoI* to generate overhangings compatible with the insert (figure 19). Consequently, 27 out of 3015 nucleotides of the pGEM-T Easy plasmid multiple cloning site were eliminated. Once the two enzymatic digestions and the purification completed, a CIP-mediated dephosphorylation was performed to prevent a potential vector recircularization. The proper double restriction of the plasmid was confirmed by electrophoresis (data not shown).

3.3. Ligation of the insert into the final pGEM-T Easy vector

The purified insert was cloned into the final pGEM-T Easy vector, previously digested with *ApaI* and *NcoI* enzymes (data not shown) (figure 19). After electroporation of this construct into TG1 competent bacteria, 12 white colonies grown on LB medium supplemented with X-gal, IPTG and ampicillin and transformed with this construct underwent a PCR screen using PU and RPU as primers (data not shown). Consequently, 207bp have to be added to the expected size of the insert (140bp). The migration on agarose gel revealed that the 11 tested colonies were transformed with the pGEM-T Easy vector containing the expected insert of interest (347bp). These results were further confirmed by sequencing (data not shown).

After amplification and purification of the plasmid, this construct was digested with *NcoI* restriction enzyme to linearize it (figure 19). Once digested, the plasmid was used as DNA template for *in vitro* transcription.

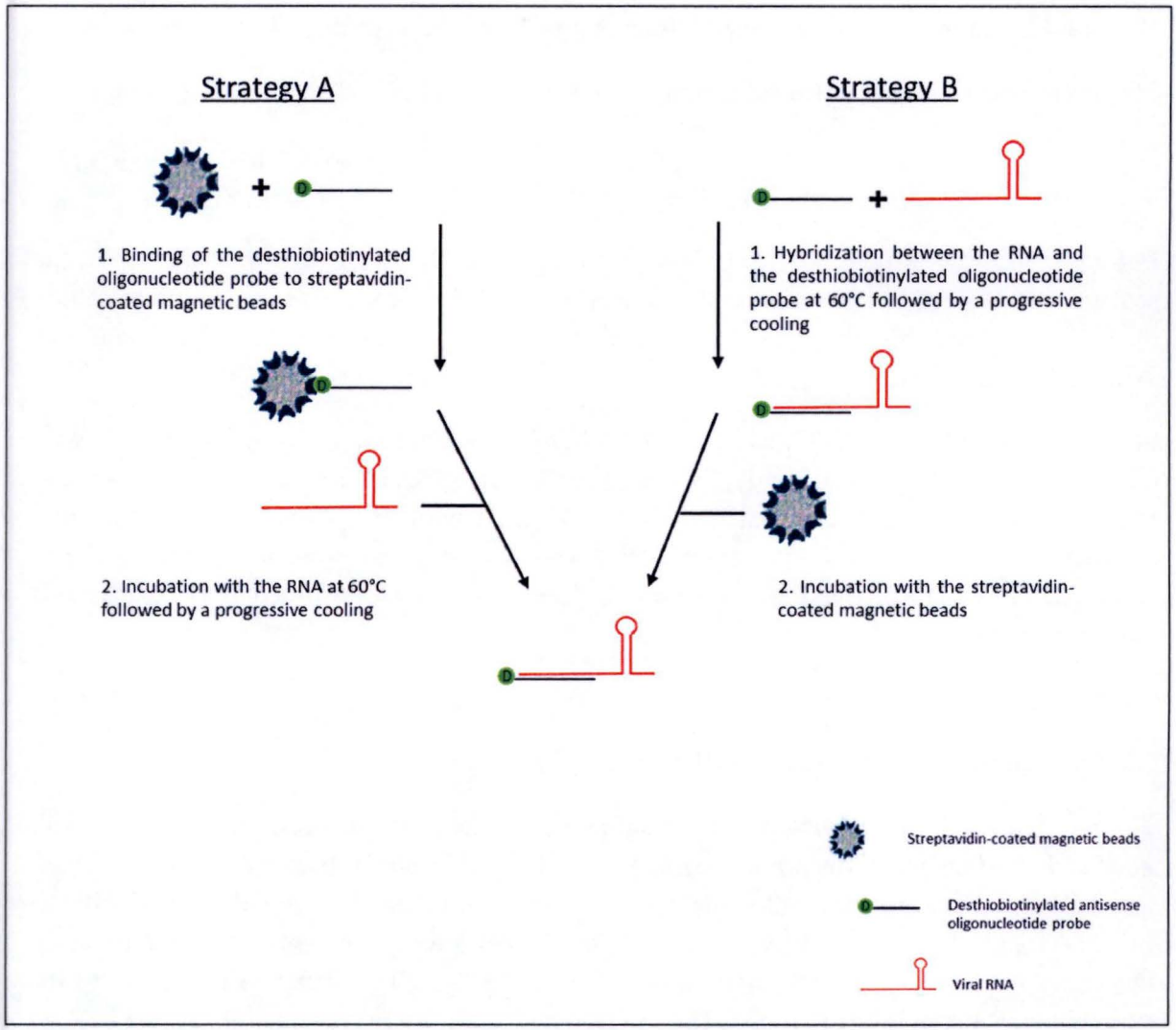

Figure 20 : Schematic representation of RNA immobilization strategies. The first strategy (strategy A) implies the previous binding of the desthiobiotinylated capture probe (black) to the solid support (grey) prior the addition of the RNA (red) for its immobilization. On the contrary, the second strategy (strategy B) consists of the hybridization between the capture probe and the RNA before the incorporation of the beads into the sample.

4. Generation of mutants

Five kinds of mutated RNA were generated by nested-PCR based on SBV S segment cDNA (figure 19) to have further insight into the influence of an altered hairpin/transcription termination signal on termination or translation mechanisms. Therefore, the mutations were inserted by PCR using specific primers containing the desired nucleotide substitutions. Once the mutated PCR products ready, they underwent the same experiments as the WT insert, namely digestion with *ApaI* and *NcoI* restriction enzymes and cloning into pGEM-T Easy restricted by the same enzymes (figure 19). Although the 5 mutants were successfully constructed, as confirmed by sequencing, only the H3p mutant was further *in vitro* due to time restriction.

5. *In vitro* transcription

pGEM-T Easy vector was selected as it contains a T7 polymerase promoter, required for the *in vitro* transcription (figure 19). Once the plasmid containing the WT or mutated inserts digested, the *in vitro* transcription reaction using the T7 polymerase was run. A total RNA (140b) amount of 75µg of WT RNA sequence and 62µg of H5p mutant RNA were collected.

6. RNA immobilization tests

The RNA of interest has to be immobilized on a solid support (streptavidin-coated beads) functionalized with a desthiobiotinylated deoxyribonucleotide probe complementary to a fragment of the RNA. An excess of biotin will then be used to separate the probe from the beads and collect the DNA/RNA heterocomplexes, eventually associated with RBP when the beads have been prior incubated with protein extracts (figure 18).

To immobilize the RNA bait, 2 strategies were considered (figure 20). The first one (strategy A) consists in binding first the desthiobiotinylated capture probe to the solid support followed by the addition of the RNA. On the contrary, in the second strategy (strategy B) the capture probe will be hybridized with the RNA before the addition of the beads into the sample.

Both methods are expected to present some advantages and drawbacks. To hybridize the RNA and DNA molecule, it is necessary to heat the sample in order to unfold the RNA secondary structures susceptible to hamper the base pairing between the two molecules. However, it means that in the first strategy, the heating will be performed in the presence of the streptavidin-coated beads, which may denature the streptavidin coating the magnetic beads and thus, hamper the interaction with the desthiobiotinylated probe. On the other hand, a major disadvantage is associated with the use of the second strategy. Indeed, the RNA from the DNA-RNA heterocomplex might fold in such a way that the desthiobiotin would be less accessible for streptavidin.

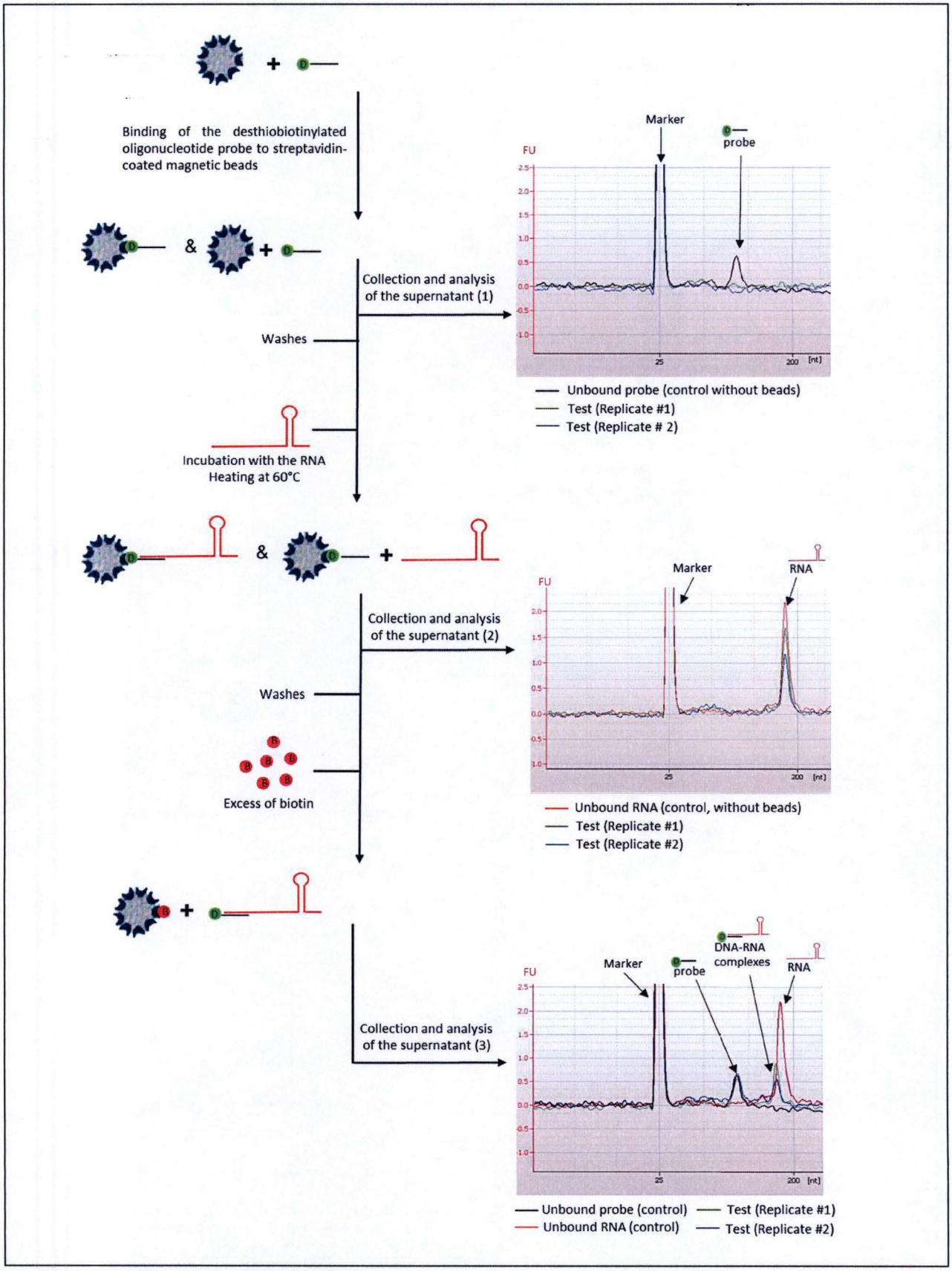

Figure 21 : Schematic representation and results of the RNA immobilization test using strategy A. The experiment was performed as described in the *Material and Methods* section with the following input quantities: 0.5mg of beads, 15pmol of the desthiobiotinylated probe, 20 pmol of the WT RNA.

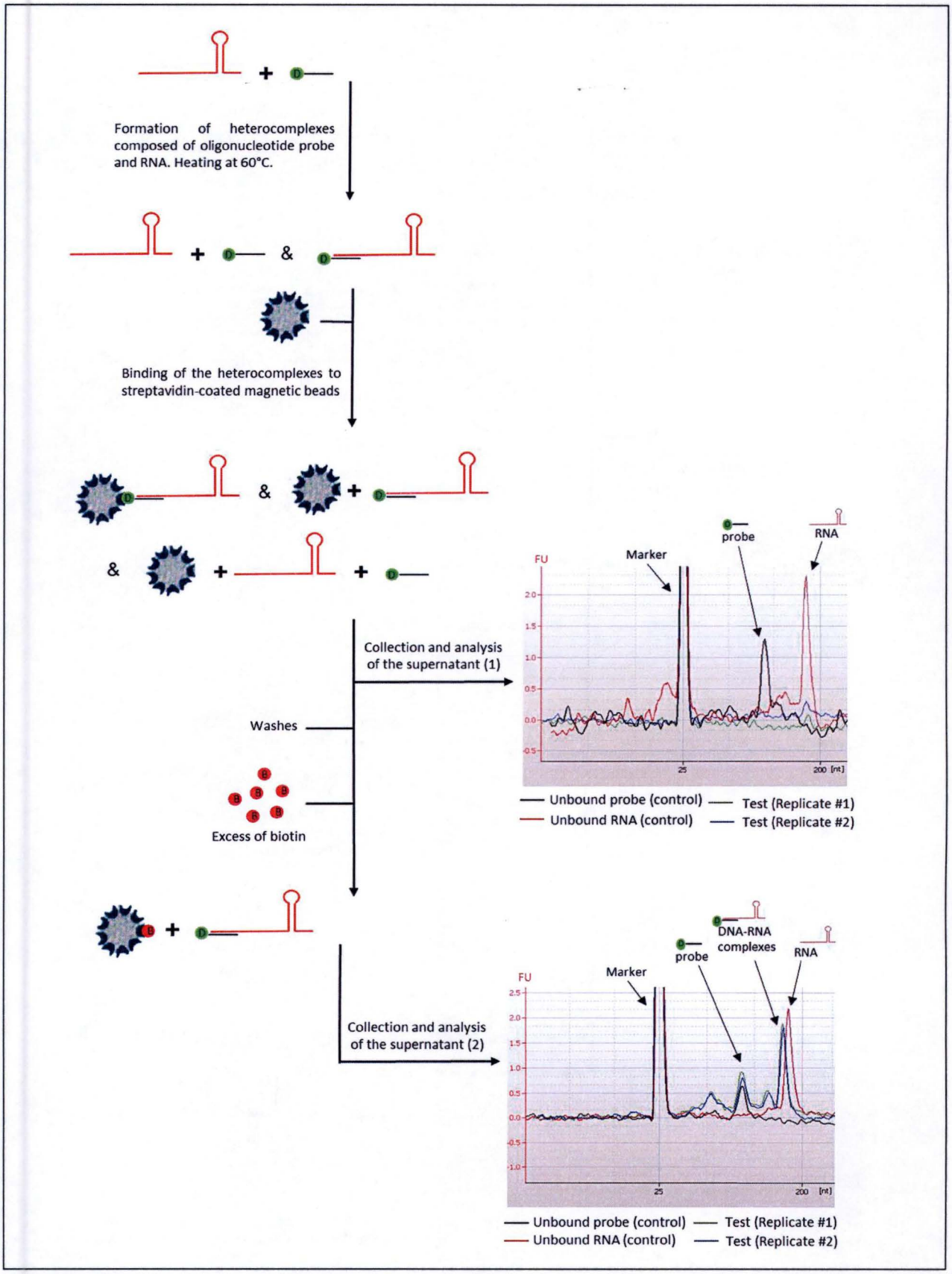

Figure 22 : Schematic representation and results of the RNA immobilization test using strategy B. The experiment was performed as described in the *Material and Methods* section with the following input quantities: 0.5mg of beads, 15pmol of the desthiobiotinylated probe, 20 pmol of the WT RNA.

6.1. Results from Strategy A

To know if RNA immobilization using the first strategy could be considered as a potential option, we assessed the binding functionality of the streptavidin after being heated up to 70°C. The results indicated that its capacity to bind a desthiobiotinylated probe was not altered by such treatment (data not shown), which is further confirmed by the literature (González et al., 1997).

As illustrated in figure 21, the desthiobiotinylated oligonucleotide probe was first incubated with streptavidin-coated beads before the collection of the supernatant (1), containing possible unbound probe. Then, the RNA was added and the whole mixture was heated at 60°C, to allow the hybridization between the nucleic strands, prior to undergo a gradual cooling. Once more, the supernatant (2) was conserved. Finally, thanks to a higher affinity of the streptavidin for the biotin than for the desthiobiotin (Hirsch, 2002), DNA–RNA complexes were specifically eluted from the beads (3). In order to evaluate the efficiency of the procedure, all the collected supernatants were subjected to 2100 Agilent bioanalyzer to obtain the electrophoresis profiles.

The results presented in figure 21 indicate that the desthiobiotinylated probe (green and blue curves) is not detectable in the supernatant (1) suggesting a 100% efficiency of the capture probe on the solid support. However, the large majority of the input RNA was recovered in the supernatant (2) (green and blue curves) leading to only a small amount of immobilized RNA collected after elution by the excess of biotin (3) (green and blue curves).

6.2. Results from Strategy B

On the contrary to the previous strategy, the desthiobiotinylated capture probe and the RNA bait were first heated at 60°C to allow the hybridization step before undergoing a progressive cooling (figure 22). This step was followed by the immobilization of the assembled DNA-RNA heterocomplexes on streptavidin-coated beads prior to collect the supernatant (1), potentially containing unbound complexes. Just as in the former strategy, an excess of biotin was used to separate DNA-RNA complexes from the beads and this elution was conserved (2) (figure 22).

Similarly, all the supernatants underwent 2100 Agilent analysis. The electrophoretic profiles showed that a fraction of DNA-RNA heterocomplexes was not bound to the beads as the supernatant (green and blue curves) (1) contains small amount of assembled complexes (figure 22). The excess of biotin allowed the recovery of the immobilized bait (supernatant 2) as well as two others peaks whose origin remains unkown (figure 22).

6.3. Selected method

The data shown above indicate that the strategy A is less efficient than strategy B to immobilize the RNA. In addition, while the oligonucleotide probe is efficiently captured on beads, the RNA

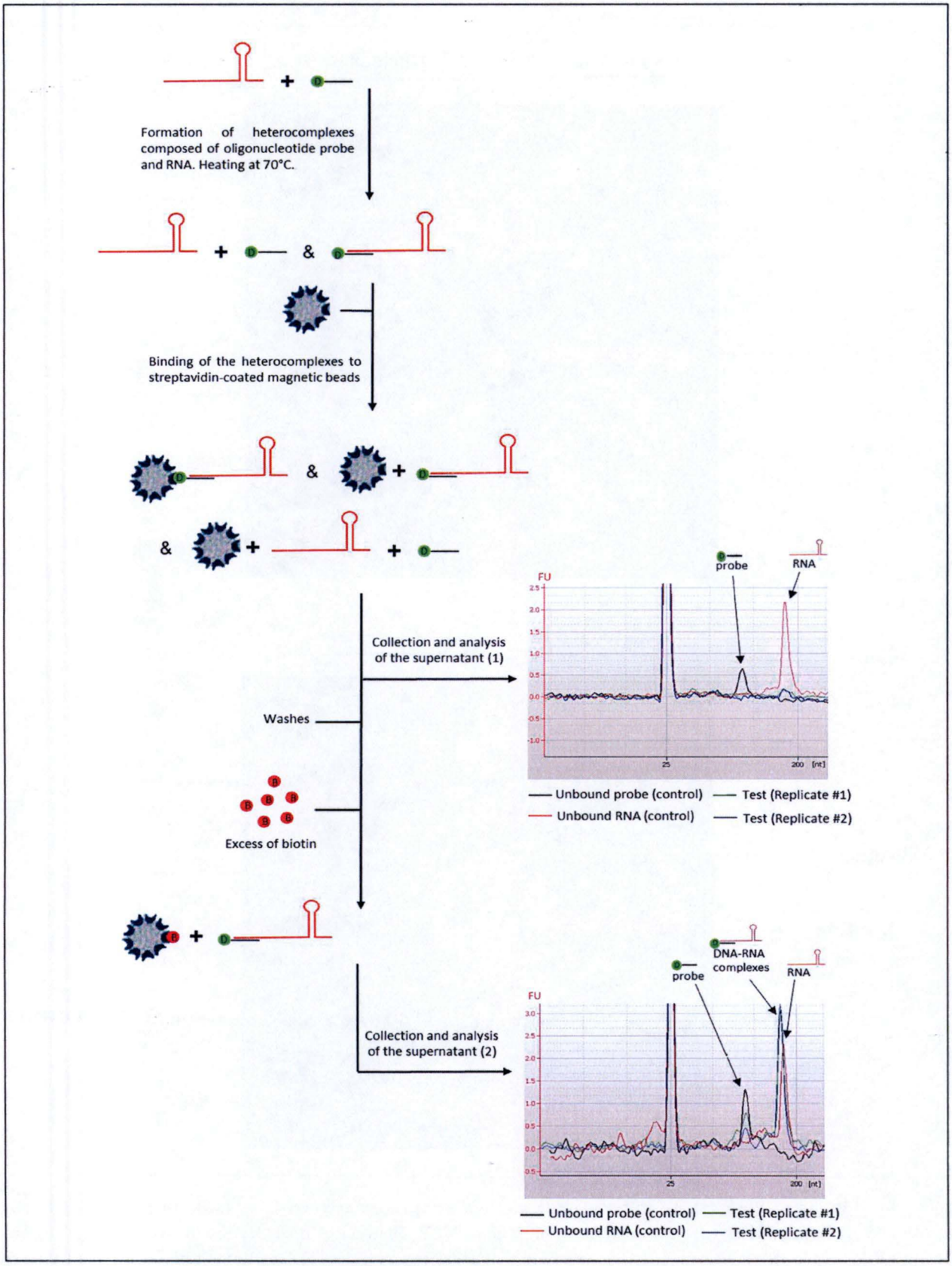

Figure 23 : Schematic representation and results of the RNA immobilization test using strategy B. The experiment was performed as described in the *Material and Methods* section with the following input quantities: 0.5mg of beads, 15pmol of the desthiobiotinylated probe, 20 pmol of the WT RNA. The difference with the experiment depicted in figure 22 is the temperature used for the hybridization between the oligonucleotide probe and the RNA.

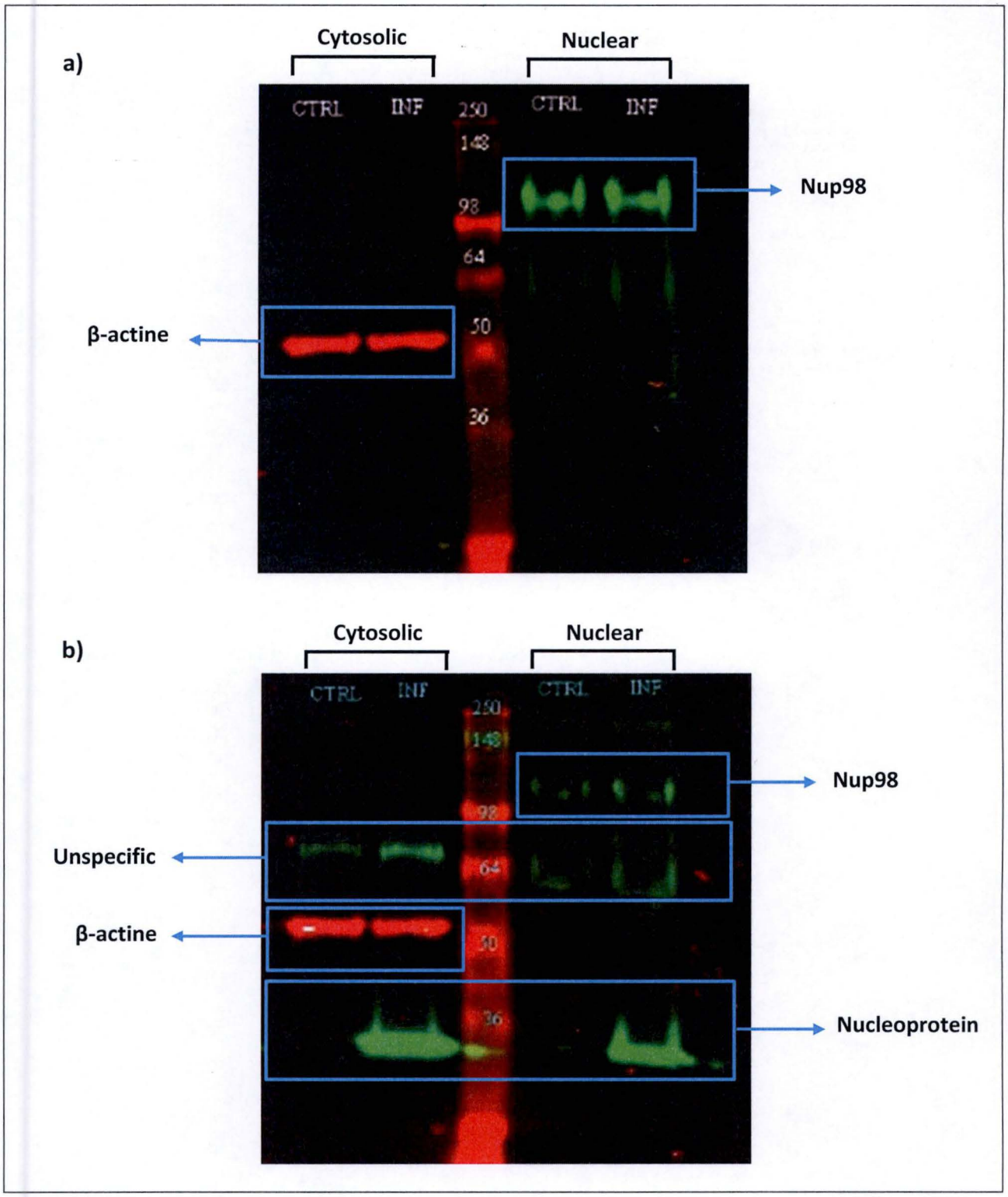

Figure 24 : Western Blot analysis of cytoplasmic and nuclear protein extracts from (un) infected BHK-21 cells. Nuclear and cytosolic proteins were extracted from uninfected (CTLR) BHK-21 cells and BHK-21 cells infected (INF) with SBV (MOI 1) for 14h. 10µg of each sample were loaded on a polyacrylamide gel. **a)** The Western Blot was performed with a rabbit primary antibody specific for NUP98 (Green signal), a second mouse primary antibody raised against β-actin (Red signal), followed by the corresponding secondary antibodies conjugated with a fluorochrome. **b)** The same membrane as in A was revealed with a rabbit IgG anti-Nucleoprotein and the corresponding secondary antibody (Green signal)

molecule cannot be fully immobilized in these experimental conditions. Several parameters can be modified in order to improve RNA-DNA hybridization.

First, the duplex can be stabilized by increasing salt concentration or decreasing the buffer pH as monovalent cations interact with phosphate groups on DNA/RNA strands in such a way that electrostatic repulsion is reduced. Besides, increasing the amount of RNA added to the oligonucleotide probe should favor the RNA/DNA hybrid formation. Also, rising the temperature allows the breakage of hydrogen bonds associated with each strand of nucleic acid and subsequently favor the hybridization step. The presence of organic solvent (ex: Formamid) is another example of factor affecting duplex formation as it diminishes the thermal stability so that hybridization can occur at lower temperature. However, as formamid covalently binds to proteins, the use of such component present a main drawback here as it could inactivate the streptavidin and/or the proteins contained in the protein extracts.

We thus pursued with strategy B, revised conditions used in this protocol, namely the temperature, the heating duration, the amount of RNA added to the same quantity of oligonucleotide probe, and the buffer composition, in order to optimize the RNA immobilization step.

From multiple trials carried out in order to improve the hybridization between the two nucleic acid strands, using the second strategy with higher RNA concentrations and a hybridization temperature reaching 70°C turned out to be the best option. Indeed, as illustrated in figure 23, the data showed that the entirety of the assembled DNA-RNA heterocomplexes was bound to the solid support as they were not detected in the supernatant (1) (green and blue curves). Finally, the displacement of the desthiobiotinylated probe by biotin allowed the recovery of most of the input RNA (figure 23). By integrating the areas under the curve, this amount of recovered RNA, as compared to the RNA input, was estimated at 75%.

Notably, every experiment was performed in duplicate.

7. Preparation of cytosolic extracts

Although hamsters are not natural hosts of SBV, baby hamster kidney-21 fibroblasts were chosen for the *in vitro* SBV replication because of their deficiency in type I interferon (Truant et al., 1977). Besides, BHK-21 is one of the only cells line able to replicate SBV. As the virus is described to replicate in the cytosol of host cell (Elliott, 2014), we have selected a protocol to prepare the cytosolic extracts that does not denature the proteins. In addition, eliminating the nuclear proteins will reduce the risk to contaminate the list of putative RNA-interacting proteins with non-relevant nuclear proteins.

In order to confirm that nuclear and cytosolic proteins were correctly extracted from BHK-21 (un)infected cells, a western blot was performed to reveal the presence of β-actin, NUP98 and SBV Nucleoprotein specific for the cytoplasmic, nuclear and infected extracts respectively. The Western Blot analysis revealed a 42 kDa and 98 kDa bands both in infected and uninfected cells, corresponding to the β-actin and NUP98 proteins respectively (figure 24a). As no signal

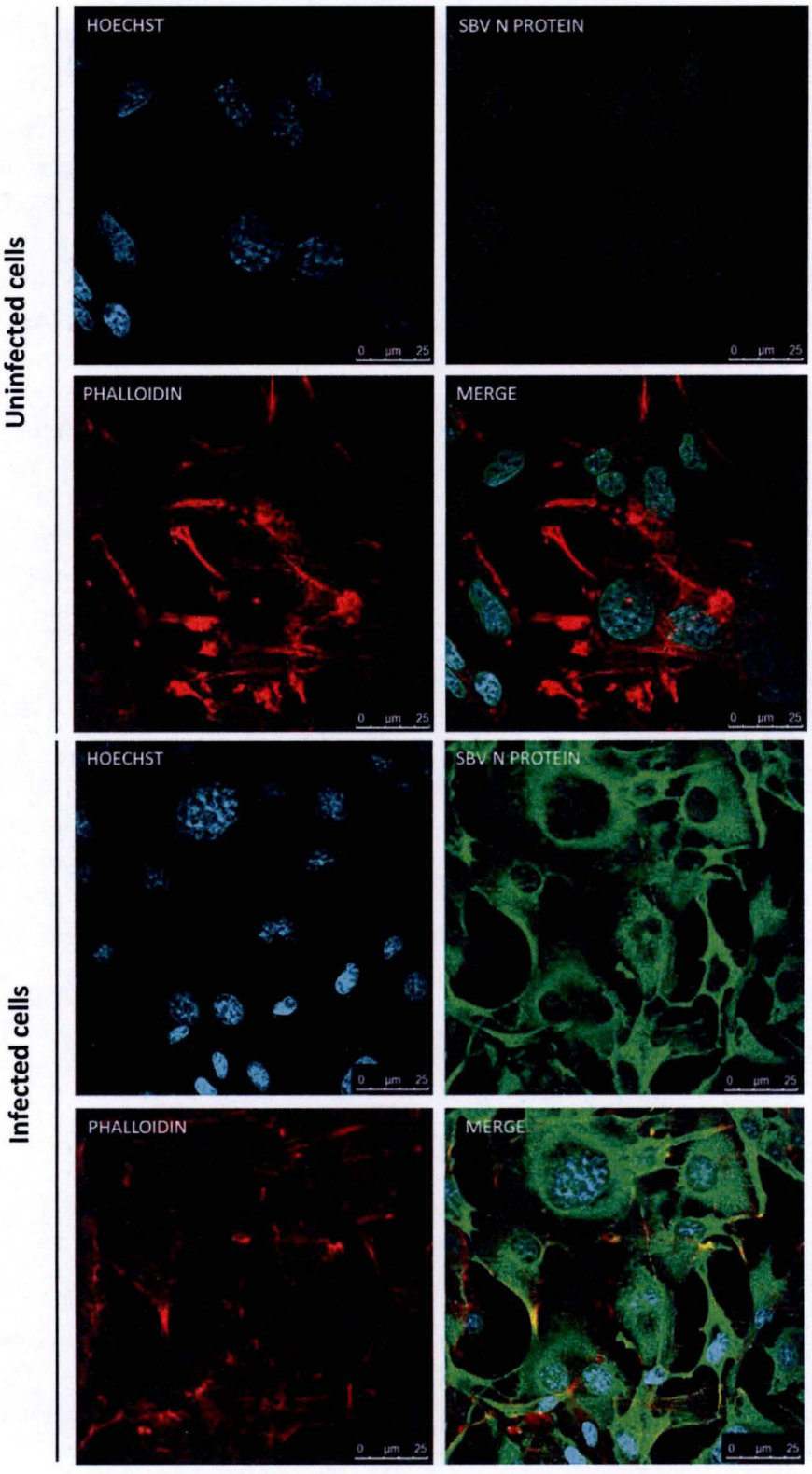

Figure 25 : Subcellular localization of the SBV nucleoprotein (N) in BHK-21 cells revealed by Immunofluorescence staining followed by confocal microscopy analysis. BHK-21 cells were infected or not with SBV at MOI 1.0 during 2 hours. Cells were fixed and permeabilized at 13h post-infection and then incubated with Hoechst (to stain DNA), phalloidin conjugated to a red fluorescent dye (to stain actin) and a rabbit antibody raised against the SBV Nucleoprotein at a dilution 1:100. Anti-N antibody was revealed with an anti-rabbit secondary antibody fluorescent in green.

for β-actin was detected in nuclear extracts and, conversely, no signal for NUP98 was observed in the cytosolic extracts, we considered them as uncontaminated. The membrane was then hybridized with an antibody raised against the SBV Nucleoprotein (figure 24b). The light signal at 66 kDa is an unspecific signal of the antibody interacting with a cytoplasmic protein of the host cells (Delphine Depierreux, master thesis 2014), while the signal observed at 25–30 kDa correlated with the presence of the SBV (predicted molecular weight: 26 181 Da).

However, while the Nucleoprotein was expected to localize in the cytosolic fractions, it was also detected in the nuclear extracts (figure 24b). This suggests that SBV (or at least the N protein) could locate both in the nucleus and in the cytoplasm, on the contrary to what has been described in the literature for *Orthobunyaviruses* (Elliott, 2014). Furthermore, this information could be supported by an experiment conducted on bunyamwera virus, the *Orthobunyavirus* reference. Indeed, Pennington and coworkers observed that enucleated BHK cells failed to produce bunyamwera virus progeny (Pennington & Pringle, 1978), although this study could simply underscore the need of nuclear protein(s) for viral replication. Another explanation involving a potential interaction between the N protein and the RdRp (Shi et al., 2006), which is a perinuclear membrane-associated protein (Kukkonen et al., 2004), is not excluded.

Anyway, the Western Blot result has cast doubt on SBV subcellular localization and thus on the subcellular protein extracts that must be selected for the RAP-MS experiment. Therefore, an analysis of the subcellular distribution of the SBV Nucleoprotein by immunofluorescence followed by confocal microscopy observation was performed.

8. Immunofluorescence analysis of the subcellular distribution of the nucleoprotein

To investigate the SBV localization, infected and uninfected BHK-21 cells were incubated with red phalloidin, Hoechst (blue) and primary antibody raised again the Nucleoprotein from SBV (green), revealing the cytoplasmic actin, the nucleus and the SBV N protein respectively (figure 25). The immunofluorescence results show a weak green signal, both in cytosol and nucleus of uninfected cells used as negative controls (figure 25), which can be attributed to a background noise due to partial specificity of anti-Nucleoprotein antibody which also recognizes host protein(s) (also noticed on the western blot analysis). On the contrary, infected cells display a stronger green intensity in the cytoplasmic compartment whereas the signal observed in the nucleus is comparable to the one detected in the uninfected cells (figure 25). Therefore, we can conclude that the N protein has a cytosolic distribution in infected cells and that the signal observed by western blot in the nuclear fraction was probably due the interaction between the N protein and the perinuclear membrane-associated RdRp (Kukkonen et al., 2004).

Once confident about SBV subcellular localization, the cytoplasmic protein extracts were prepared from uninfected and infected BHK-21 cells to perform the RNA-affinity purification of RBP.

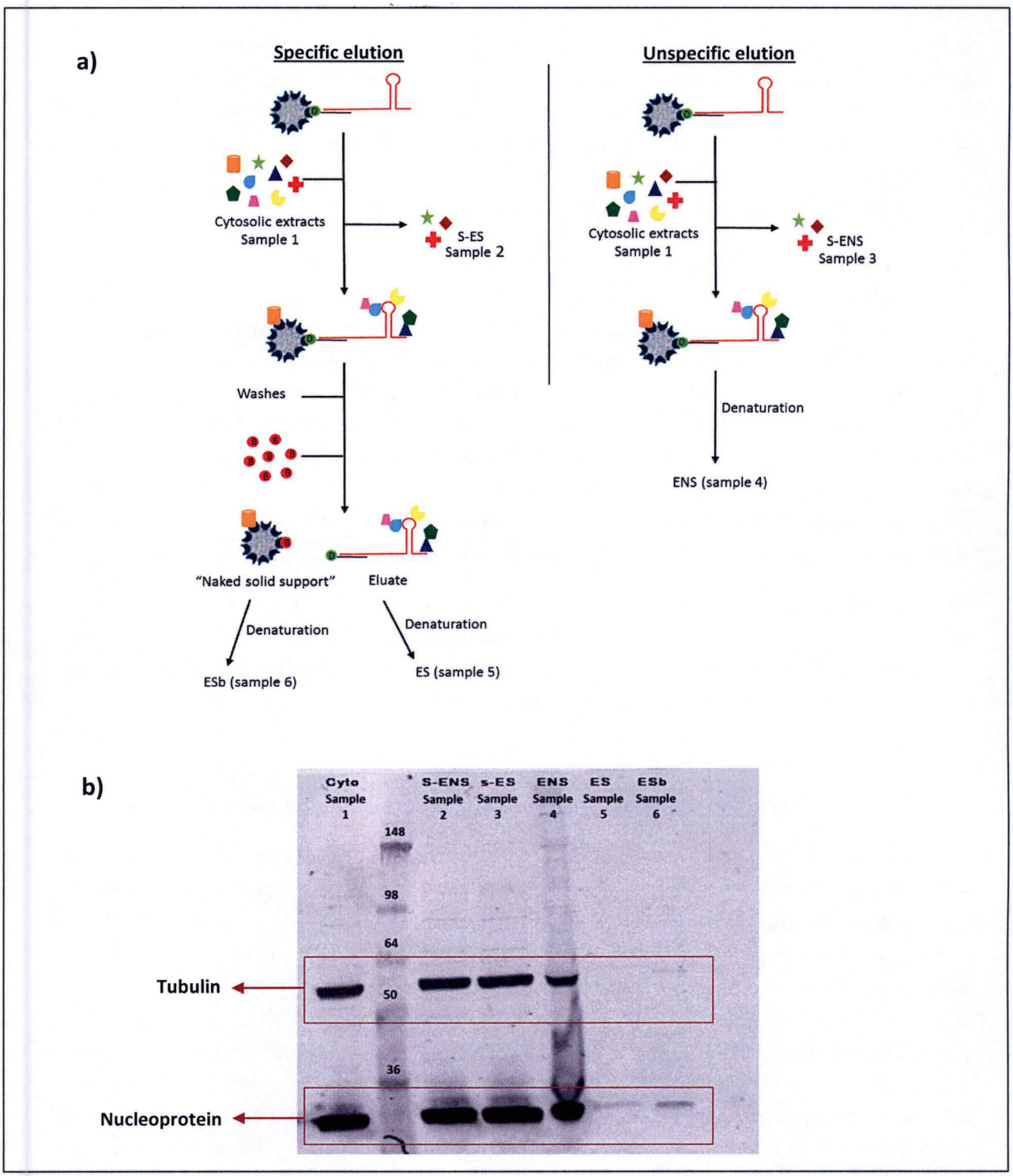

Figure 26 : Schematic representation of specific and unspecific elutions and western bloth analysis of the collected supernatants. **a)** Schematic representation of the experiment performed to evaluate the quality of specific and unspecific elution. 1mg of beads functionalized with the WT RNA sequence of interest were incubated with 1mg of cytosolic extracts (sample 1) prepared from BHK-21 cells infected for 13h with SBV (MOI 0.5). The surpernatants (= flow-through) of incubation were collected (samples 2 and 3). Then, the beads were washed before specific elution with biotin. The specific eluate (sample 5) was denatured by heat before WB analysis. The beads post-elution were also heated before WB analysis, in order to assess the presence of adsorbed proteins (sample 6). For the unspecific elution, the beads + nucleic acids + captured proteins were denatures before WB analysis (sample 4). **b)** Western blot analysis of the samples described in A. 20µg of proteins were prepared by SDS-PAGE for samples 1, 2 and 3. For samples 4, 5 and 6, all the collected material was loaded on the gel. The western bloth was performed with a mouse primary antibody specific for tubulin and a second rabbit IgG anti-Nucleoproten antibody, followed by the corresponding secondary antibodies conjugated with a fluorochrome.

9. Purification of RBP by RNA-affinity chromatography

Although the ultimate purpose is to identify, without any a priori, the RBPs interacting with the 3' UTR of SBV, it is important to validate technically, as much as possible, the different steps in the presence of the protein extracts, i.e. the protein-RNA incubation step, the washes, and the elution step (figure 18). Therefore, we have decided to follow the abundance of the nucleoprotein N along these steps, as this protein interacts with the SBV RNA, although this has not been specifically demonstrated for the SBV 3' UTR sequence studied in this project. However, the Bunyamwera virus (a SBV-related virus) N protein was shown to interact with both 5' UTR and 3' UTR (Mazel-Sanchez & Elliott, 2012). A Western Blot analysis has thus been conducted on the Nucleoprotein N to assess the binding conditions, the washes conditions, and the efficiency of the specific elution step.

9.1. Western blot analysis

The different experimental conditions and the obtained results are depicted in figure 26. The desthiobiotinylated capture probe mixed with the RNA bait was heated at 70°C for hybridization followed by a gradual cooling. Then, the constituted DNA-RNA heterocomplexes were immobilized on streptavidin-coated beads before the addition of cytosolic extracts, as described in the *material and methods* section and illustrated on figure 26a. Once the ribonucleoprotein complexes formed, the surpernatants (S-ENS/s-ES) containing unbound proteins were collected.

After several washes, an excess of biotin was used to recover the ribonucleoprotein complexes; the specific elution (ES) as well as the presumed naked beads (ESb) were conserved. In order to evaluate the role of the specific elution in RAP-MS, the same experiment was performed without the displacement of RNP complexes by the biotin (ENS). The beads linked to RNP complexes were submitted to denaturing conditions in order to collect all the proteins associated with the chromatographic support linked to the RNA. As a proof on concept, a western blot analysis was performed on S-ENS/s-ES, ES, ESb and ENS, revealing the presence of the N protein, expected to bind viral RNA.

Several important information can be extracted from the western blot presented in figure 26b.

- First, a weak band is detected in the specific eluate (sample 5: ES), indicating that the Nucleoprotein interacts with the 3'NTR.
- Second, while the Nucleoprotein was expected to be found only in ENS and ES (sample 4 and 5), it is mostly detected in the flow-through (samples 2 and 3: S-ENS/s-ES). Several hypotheses can be proposed to explain this observation: 1) an inappropriate RNA/protein ratio; 2) a large amount of N protein in infected BHK cytosolic extracts in comparison with the other proteins; 3) inappropriate conditions used for protein binding; and 4) a combination of both.

- Third, the Nucleoprotein is also present in ESb (sample 6), corresponding to the beads after the specific elution. This indicates an adsorption of the Nucleoprotein on the streptavidin-coated beads.
- Four, the tubulin, an abundant contaminant, is detected in S-ENS/s-ES, ENS and ESb (sample 2, 3, 4 and 6) but not in ES (sample 5). From these observations, it can be concluded that using unspecific elution does not allow the elimination of the tubulin, which will therefore contaminate the eluate potentially containing the interacting partners of interest.
- Five, the proteins detected in the non-specific elution (sample 4: ENS) should theoretically correspond to the sum of the proteins detected in the specific elution (sample 5: ES) and the proteins detected in the beads after the specific elution (sample 6: ESb). In other words, ENS should be equal to ES + ESb. However, the western blot analysis shows that it is not the case, as the signal intensity in ENS is much higher than the sum of ES and ESb conditions. This strongly suggests that washing conditions were too stringent.

9.2. Mass spectrometry analysis

Although the Western blot analysis of the RNA-affinity procedure discussed above indicates that the protocol still needs to be refined, we have nevertheless conducted a first RAP-MS experiment in parallel of the Western Blot analysis, with the samples derived from the specific elution.

The RNP specifically eluted from the magnetic beads were denatured and then submitted to trypsin digestion. The generated peptide mixture underwent a long reverse-phase chromatography separation followed by MSMS sequencing. Due to lack of time, no scheduled precursor list was built before a second run analysis (figure 18).

The results from MS analyses listed 138 protein clusters and 23 proteins (supplementary data, table 1).

Of note, the number of proteins detected by mass spectrometry (MS) was higher in uninfected cytosolic extracts (131 clusters) than in infected extracts (94 clusters). This observation could be due to chance, an insufficient quantity of competitor RNAs in the binding buffer or a blocking of RNA binding sites hypothetically due to the N protein.

Surprisingly, neither the Nucleoprotein nor the RNA-dependent-RNA polymerase from SBV were detected by MS sequencing, although the N protein was observed in western blot analysis of the specific elution sample (figure 24b). These conflicting information could be due to a refolding of the N protein post-denaturation, as reported for the ribonuclease A (Ribo et al., 2006), or to inefficient denaturation. Post-transcriptional modification of this protein is another hypothesis which cannot be excluded either, although none has been described so far. Other possible explanations involve a peptide mixture generated by trypsin digestion which is still too complex for MS sequencing, or a technical problem of the device itself. Regarding the RNA-

Protein classification	Detected proteins
▪ Ribosomal proteins	40S ribosomal protein S2 (RPS2) 40S ribosomal protein S3 (RPS3) 40S ribosomal protein S3A (RPS3A) 40S ribosomal protein S4 (RPS4) 40S ribosomal protein S6 (RPS6) 40S ribosomal protein S8-like (RPS8) 40S ribosomal protein S11 (RPS11) 40S ribosomal protein S12 (RPS12) 40S ribosomal protein S13 (RPS13) 40S ribosomal protein S14 (RPS14) 40S ribosomal protein S18 (RPS18) 40S ribosomal protein S19 (RPS19) 40S ribosomal protein S26-like (RPS26) 60S ribosomal protein L1 (RPL1) 60S ribosomal protein L4 (RPL4) 60S ribosomal protein L7a (RPL7A) 60S ribosomal protein L7-like (RPL7) 60S ribosomal protein L8 (RPL8) 60S ribosomal protein L12 (RPL12) 60S ribosomal protein L13 (RPL13) 60S ribosomal protein L13a (RPL3a) 60S ribosomal protein L14 (RPL14) 60S Ribosomal Protein L17 (RPL17) 60S ribosomal protein L18 (RPL18) 60S ribosomal protein L19 (RPL19) 60S ribosomal protein L26 isoform X1 (RPL36) 60S ribosomal protein L23 (RPL23) 60S ribosomal protein L27 (RPL27) 60S ribosomal protein L27a (RPL27a) 60S ribosomal protein L30-like (RPL30) 60S ribosomal protein L36a-like (RPL36a) 60S acidic ribosomal protein P0 isoform X2 (RPP2B) 60S acidic ribosomal protein P2 (RPLP2) tRNA-dihydrouridine(47) synthase [NAD(P)(+)] (DUS3L) tRNA-splicing endonuclease subunit Sen34 (SEN34)
▪ Translation-related proteins	Eukaryotic elongation factor 1-α (eEF1α) Eukaryotic elongation factor β (eEFβ) Eukaryotic elongation factor 2 (eEF2) Eukaryotic translation initiation factor 2 subunit 1 (eIF2S1)
▪ Host proteins hijacked by viruses	DEAD box polypeptide 3 Y-linked long isoform (DDX3Y) Heat shock protein 90 (Hsp90) Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/B1) Heterogeneous nuclear ribonucleoproteins A0 (hnRNP A0) Heterogeneous nuclear ribonucleoproteins A3 (hnRNP A3) Heterogeneous nuclear ribonucleoproteins D-like (hnRNP D) Heterogeneous nuclear ribonucleoprotein F (hnRNP F) Heterogeneous nuclear ribonucleoproteins U (hnRNP U) Nuclease-sensitive element-binding protein 1 (YBX-1) Nucleolysin TIAR (TIAR) Plasminogen activator inhibitor 1 RNA-binding protein (SERBP1)
▪ RNA-binding proteins	Neuroblast differentiation-associated protein protein (AHNAK) Cellular nucleic acid-binding protein (CNBP) Polymerase I and transcript release factor (PTRF) NFX1-type zinc finger-containing protein 1 (ZNFX1) RNA-binding protein Musashi homolog 2 (MSI2) Transcription termination protein nusG (nusG)
▪ DNA-binding proteins	DNA helicase B (HELB) DNA-topoisomérase 3α (TOP3A) F-box only protein 18 (FBXO18) Replication protein A 32 kDa subunit (RPA2) Replication protein A 70 kDa DNA-binding subunit (RPA1) Y-box binding protein 3 (YBX3)
▪ Technical contaminants	Actin (ACT) Acetyl-CoA carboxylase (ACAC) Bovine serum albumin (BSA) Mehylcrotonyl-CoA carboxylase (MCCA) Propionyl-CoA carboxylase (PCCA) Pyruvate carboxylase (PC) Tubulin (TUB)

Table 5 : Partial list of proteins identified by mass spectrometry analysis. Interesting proteins have been classified as ribosomal proteins, translation-related proteins or RNA-binding proteins. Technical contaminants and DNA-binding proteins have also been sequenced.

dependent RNA polymerase, Lodeiro and coworkers revealed that the replicase complex containing the RNA-dependent RNA polymerase recognizes a stem-loop structure located in UTR of the Dengue virus segments as promoter for RNA synthesis (Lodeiro et al., 2009). Therefore, we presumed a similar binding phenomenon for SBV RNA-polymerase but no detection by MS was observed. On the other hand, its absence in MS analysis results could also be due to the fact that the RNA polymerase is known to bind both 5' and 3' UTR which are complementary, forming a panhandle structure (figure 4). In others words, while the polymerase can recognize a double stranded RNA, described as the promoter for the transcription and replication of each RNA segment, it might be unable to bind a single stranded RNA (Elliott, 2014).

Besides, despite precautions, the number of technical contaminant is not negligible. For instance, a very large number of MSMS spectra of Bovine serum albumin, used in the preparation of streptavidin-coated magnetic beads, was observed (table 5 and supplementary data, table 1). Similarly, biotin-dependent enzymes including acetyl-CoA carboxylase, pyruvate carboxylase, Propionyl-CoA carboxylase and mehylcrotonoyl-CoA carboxylase which participate to amino acid or lipid metabolism, were identified (table 5 and supplementary data, table 1). We presume that these abundant metabolic proteins were unspecifically adsorbed on the beads, and then eluted in the presence of a high concentration of biotin.

Also, DNA-binding proteins such as replication protein A (RPA2), involved in DNA replication, repair and recombination or F-box only protein 18 (FBXO18), provided with DNA helicase activity were sequenced (table 5 and supplementary data, table 1). However, although the presence of DNA-binding proteins in the eluate could be interpreted as a contamination, it has to be considered with caution as dual DNA- and RNA-binding capacity have been reported for many proteins (reviewed in Hudson et Ortlund, 2014).

However, even if the RAP-MS protocol needs to be refined, and the experiment repeated with independent replicates with both the WT and mutated RNA sequences, the fact that numerous ribosomal proteins and translation-related factors were identified (table 5) is encouraging.

Dealing with a vast number of identified proteins, and without the data that will be generated in the near future with the mutant RNA, we focused on the proteins previously described in the literature to play a role in eukaryotic and viral transcription/translation/replication process (reviewed in Lloyd, 2015). Using this method, several potential regulator proteins including heterogeneous nuclear ribonucleoproteins (hnRNP A/B, hnRNP A3, hnRNP D, hnRNP U), poly(rC) binding protein 1 (PCBP1), Nucleolysin TIAR, DEAD-box RNA helicase 3(DDX3), Nuclease-sensitive element-binding protein 1 (YB-1) and elongation factor-1-alpha (eEF-1a) drew our attention.

PCPB1 / hnRNP E1

PCBP1, one of the four PCBP isoforms, is a family member of heterogeneous nuclear ribonucleoproteins (hnRNPs) which shuttle between the nucleus and the cytoplasm (Lloyd,

2015). Although the PCBP2 isoform has a predominant role since it is required for both cap-dependent translation initiation and RNA replication, the PCBP1 isoform also takes part into RNA metabolism (Luo et al., 2014).

Indeed, PCBP1, also known as heterogeneous nuclear ribonucleoprotein E1 (hnRNP E1), is able to bind the 3′ UTR of cellular mRNAs and subsequently stabilize mRNAs or control their translation. Besides, in polioviruses, interaction between PCBP1 and the 5′ UTR resulted in enhanced viral RNA replication (Luo et al., 2014).

More recently, Luo and colleagues reported that PCBP1 binds the 5′ UTR RNA of enterovirus 71 (EV71), predicted to form a secondary structure containing six stem-loops. However, solely stem-loop I and IV seems to be involved in viral RNA-protein interactions. Since stem-loop I has a cloverleaf conformation which was proved to be fundamental for poliovirus RNA synthesis and stem-loop IV was crucial for IRES-mediated translation of polioviruses RNAs, they investigated the biological function of PCBP1 in enterovirus 71 replication. Therefore, the subcellular localization of PCBP1 protein and viral RNA was first studied. Immunofluorescence assays revealed a distribution of PCBP1 in both the nucleus and cytoplasm of uninfected cells although PCBP1 was mainly located in the cytosol of infected cells, as well as colocalisation with viral RNA. Moreover, PCBP1 was demonstrated to facilitate EV71 replication through interaction with 5′ UTR, although the molecular basis of this regulation remains to be elucidated. Finally, Luo and coworkers also underscored a correlation between over-expression of PCBP1 and increased levels of EV71 viral protein expression on one hand, and knocked-down of PCBP1 and reduced viral titers on the other hand. These results, indicating a role for PCBP1 as positive regulator of viral protein expression and viral replication, are in agreement with former reports describing PCBP2, as an activator of viral RNA translation and RNA synthesis in polioviruses (Luo et al., 2014).

HnRNP D / AUF1 & TIAR

hnRNP D, also named AU-rich element RNA-binding protein 1 (AUF1) has been shown to affect transcriptional activation and mRNA translation.

Due to alternative splicing, the AUF1 family is composed of four isoforms: $p37^{AUF1}$, $p40^{AUF1}$, $p42^{AUF1}$, $p45^{AUF1}$, named according to their molecular weight. AUF1 proteins are known for the binding of adenylate/ uridylate-rich elements (AU-rich elements or AREs), commonly localized within the 3' UTR of mRNAs encoding cytokines, proto-oncogenes, inflammatory mediators and G protein-coupled receptors. Although considered for years as mRNA-destabilizing sequences, recent investigations have reported that ARE can play a role in mRNA stabilization or translational control (White et al., 2013).

The myc proto-oncogene was the first identified target of AUF1, which is reported to participate to mRNA decay. Indeed, a study revealed that the levels of myc translation were correlated with AUF1 abundance but inversely proportional to the expression of the ARE-binding protein TIAR, a translational suppressor (Liao et al., 2007). By contrast, in rhinovirus and poliovirus,

AUF1 binds the IRES at stem loop IV and negatively regulates viral translation. This negative regulation of translation also seems contrasted as mutations incorporated into the TIAR recognition motif in the hairpin, located in 3' UTR of *Flaviviruses,* hampered RNA replication but not translation. Besides, inhibited West Nile virus replication in TIAR knockout cells further supports a role for this factor in RNA synthesis (Lloyd, 2015).

In Epstein Barr virus, AUF1 is considered as a major component of C promoter binding factor 2 (CBF2), which allows the interaction of Epstein-Barr nuclear antigen 2 (EBNA2), a viral protein expressed in infected lymphocytes, with the latency C promoter. EBNA2 has been shown to stimulate the expression of this promoter directing the transcription of most of the latency genes (Fuentes-Pananá et al., 2000).

Notably, AUF1 binding affinity was reported to be highly sensitive to conserved RNA secondary (White et al., 2013).

hnRNP A/B et hnRNP A3

In the coronavirus mouse hepatitis virus, interactions between Polypyrimidine tract-binding protein (PTB) and hnRNPA1 allow a crosstalk between the viral UTRs. Indeed, the viral replication and transcription were decreased by mutations inserted in order to block the interaction between these two proteins. Intriguingly, the viral replication was not affected in cells lines which do not express hnRNP A1. This might be explained by the fact that viral RNA can also associate with hnRNP A2/B1, hnRNP A/B and hnRNP A3 replacing the initial role of hnRNPA1 (Lloyd, 2015).

Similarly, in hepatitis viruses, the downregulation of hnRNP A3 did not hamper viral replication (Real et al., 2013). However, the specific functions of hnRNP A3 are still largely unknown, as protein less studied than the other members of hnRNP A/B family.

hnRNP U

hnRNP U, an abundant nucleoplasmic phosphoprotein residing in eukaryotic cells, has been demonstrated to take part to the splicing of RNA segments of Vesicular Stomatitis viruses (VSV), a negative-strand RNA arbovirus from the *Rhabdoviridae* family. Although hnRNP U is predominantly found in the nucleus, immunofluorescence and confocal microscopic analyses have shown a colocalisation between this RBP and VSV RNA in the cytoplasm of infected cells. Moreover, the results also indicated that hnRNP U is packaged within the nascent virions. These results provide evidence that hnRNP U diffuses out in the cytosol following VSV infection and plays a fundamental role in VSV replication (Gupta et al., 1998).

Besides, the binding motif of hnRNP U has not been characterized yet. However, a secondary structure located in the 3'-terminal domain of genomic RNA was postulated to be recognized by hnRNP U as mutational analyses lead to reduced binding (Gupta et al., 1998).

DDX3

DDX3 is a member of the DEAD-box RNA helicase family, taking part in RNA transcription, splicing, and translation.

In hepatitis B virus (HBV) infected cells, DDX3 binds the polymerase and subsequently inhibits viral replication at the posttranscriptional level. Indeed, Ko and coworkers showed that DDX3 expression resulted in decreased levels of viral RNAs whereas DDX3 knockdown was correlated to their increase. However, the human immunodeficiency virus type 1 also hijacks DDX3 but in this case, it promotes viral replication and nuclear export and translation of unspliced mRNA. DDX3 also seems to participate to the replication of the West Nile virus (WNV) as DDX3 is sequestered towards viral replication sites, although the underlying molecular mechanism remains still unknown. In Japanese encephalitis virus (JEV) infected cells, DDX3 interacts with two components of the replicase complex and both 5' UTR and 3' UTR, supporting its role in viral replication. Besides, DDX3 also appears to be essential for RNA translation. Interestingly, after identification by the murine norovirus RNA affinity purification followed by a mass spectrometry analysis, DDX3 was shown to be an RNA-binding protein which was later on proved as viral RNA and protein synthesis enhancer. Despite its apparent crucial role in viral replication and RNA translation, most of the mechanisms by which DDX3 promotes these both events have yet to be deciphered (Ko et al., 2014; Valiente-Echeverría et al., 2015).

YB-1 & eEF1a

The stem-loop located in 3' UTR of *Flaviviruses* RNA was demonstrated to be bound by host factors such as YB-1 and the translation elongation factor eEF1a. Y box-binding protein 1 (YB-1), also interacting with DDX3 and numerous hnRNPs, was identified as a translation repressor as well as a potential regulator of RNA stability (Lloyd, 2015).

Contrary to what its name might suggest, the interaction of the eEF1a with this hairpin does not seem to have an effect on translation. However, eEF1a, encountered in both cytosolic and nuclear compartments, is thought to take part in RNA synthesis as immunofluorescence assays performed on cell infected with WNV revealed a colocalisation between eEF1a and replication complexes (Lloyd, 2015).

In human immunodeficiency virus (HIV), the eEF1a factor, which has been detected within the virions, appears to be involved in different stages of the viral cycle. First, eEF1a was identified as a component of the reverse transcription complex (RTC). Knockdown of eEF1a using siRNA resulting in a reduced reverse transcription and decreased levels of RTC further support this hypothesis. Secondly, a potential role in transcription, through stimulation of the RNA polymerase II, and in translation was brought to light. In addition to its function in the binding of aminoacyl-tRNAs to the ribosome, eEF1a was demonstrated to associate with both the viral protein Negative Regulatory Factor (NEF) and the tRNA. As Nef is known to interact with the 40S small ribosomal subunit, the 18S rRNA, and the tRNAs and was correlated with a

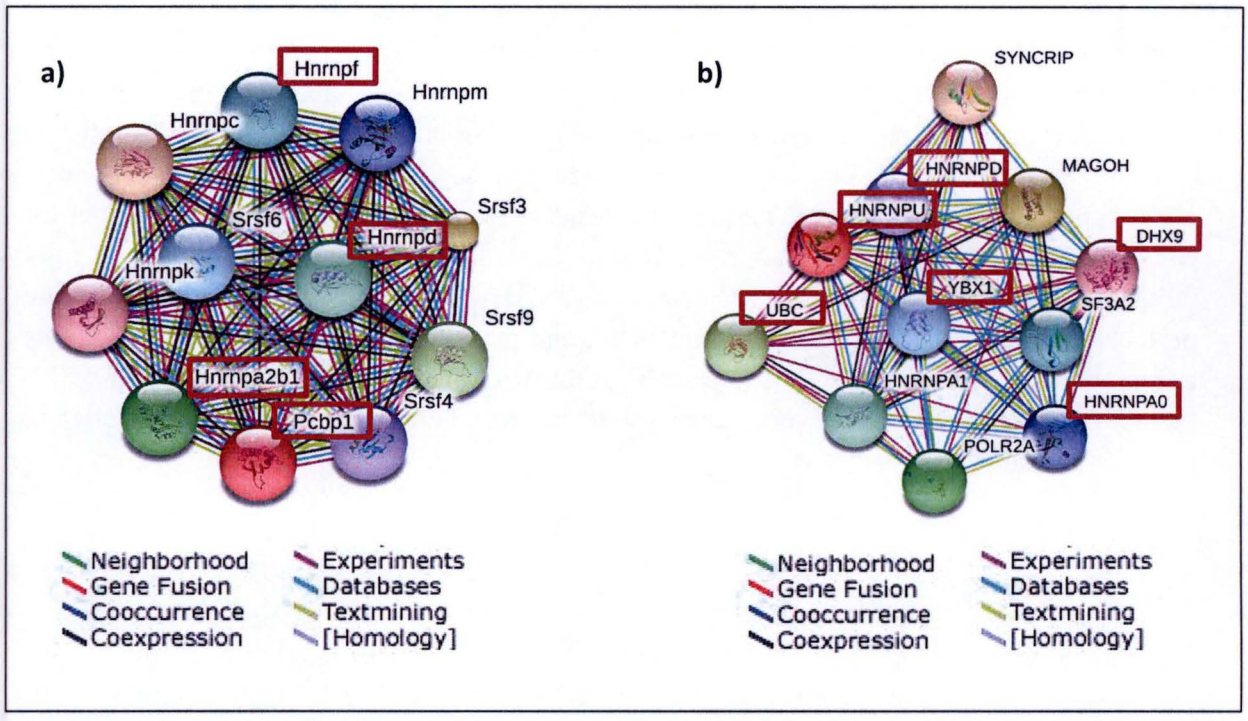

Figure 27 : predicted YBX1 and PCBP1 protein networks (STRING software). a) protein-protein interactions centralised on poly(rC) binding protein 1 (PCBP1), involving heterogeneous nuclear ribonucleoprotein C (hNnRNP C), serine/arginine-rich splicing factor 3 (Srsf3), serine/arginine-rich splicing factor 9 (Srsf9); heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNPA2/B1), heterogeneous nuclear ribonucleoprotein D (hnRNP D), heterogeneous nuclear ribonucleoprotein F (hnRNP F), serine/arginine-rich splicing factor 6 (Srsf6), heterogeneous nuclear ribonucleoprotein M (hnRNP M), serine/arginine-rich splicing factor 4 (Srsf4), heterogeneous nuclear ribonucleoprotein K (hnRNP K). Among these proteins, the factors detected by RAP-MS analysis are framed in red. **b)** protein-protein interactions centralised on Nuclease-sensitive element-binding protein (YB-1) , involving synaptotagmin binding (SYNCRIP) mago-nashi homolog (MAGOH) , ubiquitin C (UBC) polymerase (RNA) II (DNA directed) polypeptide A (POLR2A), heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1), splicing factor 3a, subunit 2 (SF3A2), heterogeneous nuclear ribonucleoprotein A0 (hnRNPA0), heterogeneous nuclear ribonucleoprotein D (hnRNPD), DEAH box polypeptide 9 (DHX9). Among these proteins, the factors detected by RAP-MS analysis are framed in red.

decreased translation, a translation regulation operated by eEF1a cannot be excluded (Abbas et al., 2015).

Besides, it is interesting to notice that the proteins mentioned above to take part in the translation / replication mechanisms of viruses may interact with each other. In order to investigate this hypothesis, the STRING v9.1 web software, a biological database of known and predicted protein-protein interactions was used. As illustrated in figure 27b, Y box binding protein 1 (YB-1) is likely to interact with hnRNP D, hnRNP A2/B1, hnRNP U, DHX9 (a member of DExD/H-box helicases, as DDX3), and Ubiquitin C, which was also detected. Also, poly(rC) binding protein 1 (PCBP1), hnRNP A2/B1, hnRNP D and hnRNP F seem to be part of the same protein network (figure 27a). Although totally hypothetic at this stage, one might speculate that at least some of those proteins are involved in the interactome re-built *in vitro* around the SBV 3' UTR.

As mentioned above, several actors of RNA translation have been identified in this preliminary RAP-MS experiment: 36 components of the 40S and 60S ribosomal subunits, translation initiation factor 2 (eIF2S1), translational activator GCN-1 like protein (GCN1L1), etc (table 5 and supplementary data, table 1). Similarly, proteins involved in transcription like polymerase I and transcript release factor (PTRF) and transcription termination protein nusG (nusG), taking part in transcription termination were also sequenced (table 5). Intriguingly, several proteins described to bind poly(A) sequences have also been recovered after RAP-MS analysis, although the RNA sequence used as a bait is deprived of such poly(A) tail. This is the case for NFX1-type zinc finger-containing protein 1 (ZNFX1) ahnak protein and Ribosomal Protein L31 (RPL31), a component of 60S ribosomal subunit (table 5 and supplementary data, table 1). However, the presence of such proteins in the RAP eluate might also be due to interactions with other RNA motifs or protein-protein interactions.

Conclusion & Perspectives

Conclusion and perspectives

Our main objective was to have further insight into the transcription and translation processes of the Schmallenberg virus by developing a procedure to capture and identify the RNA-binding proteins which interact with the stem-loop or the termination signal, both highly conserved among the *Simbu* serogroup. Given their conservation and their location in the in the 3' UTR segment of the virus, these two elements are suspected to play a key role in transcription/translation mechanism. To test this hypothesis, the method settled here was adapted from a DNA-affinity purification followed by mass spectrometry (DAP-MS), previously developed in the lab (Tacheny et al., 2012).

First, the SBV sequence composed of the 120 last nucleotides of the S segment and containing both stem-loop and transcription termination signal has been selected as the fragment of interest, amplified and cloned into a pGEM-T Easy plasmid containing a T7 promoter sequence.

In parallel, in addition to the wild-type RNA sequence, 5 mutated constructs have been conceived in order to have further insight into the influence of an altered hairpin/transcription termination signal on transcription termination or translation mechanisms. The generation of mutants aims to select the proteins likely to take part into transcription/translation mechanisms. Among these mutants, two have been conceived to explore the role of the transcription termination signal (Gc1 & Gc2). Two others (H3p & H5p) contain a mutation on one strand of the hairpin structure (either 3'p or 5'p) with the purpose of breaking it, whereas two mutations have been incorporated into a fifth mutant (H3p5p) to modify the primary sequence while preserving the stem-loop structure. These last 3 mutants have been designed in order to figure out whether the primary sequence or the secondary structure is the crucial determinant of the transcription and translation processes. Although the 5 mutants have been constructed, only the H5p has been used for RAP-MS experiment, due to time shortage.

Secondly, after *in vitro* transcription of both wild-type and the H5p mutated sequences by the T7 polymerase, the RNA bait was immobilized on a solid support. Therefore, we designed an antisense desthiobiotinylated deoxyribonucleotide probe complementary to an endogenous sequence of the RNA fragment of interest. More precisely, the RNA was hybridized with the probe and the subsequently assembled DNA-RNA heterocomplexes were captured on the streptavidin-coated magnetic beads through reversible streptavidin-desthiobiotin interactions. The amount of recovered RNA, as compared to the RNA input, was estimated at 75%.

Thirdly, once the cytoplasmic localization of SBV in infected BHK-21 cells confirmed by an immunofluorescence experiment, cytosolic proteins were extracted from infected and uninfected BHK-21 cells. The cytosolic proteins were then incubated with the immobilized bait to allow the RNP complexes formation. Thanks to a higher affinity of the streptavidin for the biotin than for the desthiobiotin, DNA–RNA complexes were then specifically eluted from the beads.

As validating as many technical steps as possible is essential, a western blot analysis was performed in order to assess the binding conditions, the washes conditions, and the efficiency of the specific elution step. Therefore, the abundance of the nucleoprotein N was followed along these steps as this protein interacts with the SBV RNA (Dong et al., 2013a et b). However, so far, this interaction has not been specifically demonstrated for the SBV 3 'UTR sequence studied in this project. Remarkably, comparison analysis between the Bunyamwera virus (BUNV), a SBV-related virus, and SBV N proteins revealed a highly similar structure (Ariza et al., 2013). Although a preferential interaction with 5'-terminal 32 nt was noticed, BUNV N protein was demonstrated to bind both 3' and 5' UTR. Especially, nt positions 20 to 33 of the 3′ and 5′ UTRs were shown to be crucial for genome segments packaging (Mazel-Sanchez & Elliott, 2012). Considering this information, following the abundance of the nucleoprotein N at different steps is likely a reliable approach to evaluate the conditions used for protein RNA binding, washes and specific elution. The western blot results highlighted that, as expected, using an unspecific elution contaminates the eluate with abundant proteins unspecifically adsorbed on the support (like tubulin). Obviously, such contaminant likely decreases the efficacy of the method in the identification of the true interacting proteins, that might be of low abundance. In addition, the Western Blot analysis of the Nucleoprotein also shows that the RNA-affinity protocol needs to be optimized on several aspects. Indeed, more promising results could be obtained by reducing the stringency of washes, increasing the RNA/protein ratio and by modifying the conditions used for protein binding.

Although the Western blot analysis of the RNA-affinity procedure indicates that the protocol still needs to be refined, a first RAP-MS experiment was conducted with the samples derived from the specific elution. Among the sequenced proteins, 36 ribosomal proteins, several translation-related proteins (i.e. elongation factor-1-alpha), poly(A) binding proteins (i.e. NFX1-type zinc finger-containing protein 1) and other host proteins described to be hijacked by other viruses (i.e. poly(rC) binding protein 1 or AU-rich element RNA-binding protein 1) and to play a role in viral transcription, translation, or replication processes were identified. The encouraging number of detected proteins likely to take part into these mechanisms further supports a role for the SBV studied sequence in transcription/translation processes. However, these results have to be taken with caution since they are based on only one preliminary experiment. In order to identify the proteins likely to be involved in transcription/translation mechanisms, the RAP-MS experiment has to be repeated with independent replicates with both the WT and mutated RNA sequences.

Considering that regulatory proteins are often present in low quantities compared to technical contaminants or to the highly abundant proteins binding the nucleic acid in an unspecific manner, a scheduled precursor list (SPL) containing the most sequenced ions could be defined before a second MS analysis. In this way, any peptide included in the SPL being excluded from the second run analysis, the MS will be more focused on the less abundant ions, and thus on proteins of interest. The significant benefit of this approach was demonstrated in the context of the study of DNA-binding proteins interacting with the HIV-1 5'LTR sequence (Tacheny et al., unpublished results). Although initially planned, no SPL was built in this project due to time restriction. However, a careful analysis of the results suggests that using a SPL could largely

improve the results, based on several arguments. First, to interpret the preliminary RAP-MS data, we have arbitrarily decided to consider the proteins containing at least 2 identified peptides. Nevertheless, if proteins identified on the basis of a single sequences peptide were considered, this would have nearly doubled the number of proteins to deal with (2465 proteins gathered in 607 clusters if one peptide is considered, as compared to 1433 gathered in 138 clusters if at least two peptides are required for consideration). This strongly suggests the necessity to use a SPL as many proteins are represented by low abundant peptides in a complex peptide mix.

Secondly, this SPL necessity is further supported by the fact that, the peptide mix has a large dynamic range: some technical contaminants like BSA peptides were sequenced 1041 times (16% of total spectra) indicating the relatively high abundance of this protein. While sequencing those irrelevant peptides, the mass spectrometer was less available to sequence the low abundant peptides of potential interest.

Finally, after the comparison of the data obtained with the WT and the mutated RNA, in the several independent experiments, some proteins likely to be involved in SBV transcription termination or translation initiation mechanisms could be identified. Once the best candidates selected, their expression could be downregulated using siRNA and the potential influence on viral replication could be analyzed in infected cells. At the same time, the involvement of those proteins in the transcription and translation processes will be assessed, using the luciferase reporter gene assays constructed in parallel. For instance, if PCBP1 is detected in the results of RAP-MS experiment repeated with independent replicates with the WT and not a mutated RNA sequence, it could be considered as a good candidate. Indeed, in addition to bind conserved RNA secondary structures, PCBP1 is described as a positive regulator of viral protein expression and viral replication in enterovirus 71 (Luo et al., 2014).

It seems also essential to mention that the functional role of the highly conserved stem-loop or putative transcription termination signal as substitutes of the poly(A) tail can be explored by reporter gene assay. For instance, the poly(A) tail of the luciferase gene, a reporter gene, could be replaced by the sequence corresponding to the SBV hairpin or transcription termination signal. If luminescence is observed, it means that either the SBV stem-loop or transcription termination signal is able to functionally substitute the poly(A) tail and conversely. Let's mention that both wild-type and mutated sequences described previously are currently being tested by another researcher of the lab. These experiments will not only reveal the most relevant mutations at the functional level but will also be useful to point the most interesting captured proteins that will be selected for functional analyses. In order to have further insight into viral protein-RNA interactions, this experiment should be performed on both infected and uninfected cells, in case a viral protein is required for a functional stem-loop.

While a cytolytic SBV replication cycle is observed in mammalian host cell, a persistent infection is visible in insect vector cell. To investigate the molecular basis of distinct characteristics in *Orthobunyaviruses* replication cycle, the RAP-MS experiment performed with proteins extracted from KC cells would be a judicious perspective.

In the future, it would also be interesting to study RNA–protein interactions *in vivo,* as RNA and protein content of RNPs seems to be reshaped in response to environmental signals. Besides, since RNA post-transcriptional modifications, reported to have an impact on RNA-binding activity, are not present in *in vitro* synthesized RNA molecules, some protein partners recruited through such RNA modifications, likely to take part into transcription/translation processes, are not captured in such *in vitro*-assembled RNP complexes. Interestingly, the development of RAP-MS performed on *in vitro*-assembled RNP complexes could also be useful for RAP-MS analysis of *in vivo*-assembled RNP complexes. Indeed, once the infected cells fixed by paraformaldehyde, the RNA sequence of interest with its associated RBP could be captured using a desthiobiotin oligonucleotide probe and then immobilized on streptavidin-coated magnetic beads through reversible streptavidin-desthiobiotin interactions, as described by Déjardin and Kingston (Déjardin & Kingston, 2009). The main difficulty of this technique would be to find a region of the RNA accessible for base pairing with the antisense probe. Indeed, RNA-binding proteins could mask the RNA sequence, hamper the hybridization between the RNA and the probe and subsequently compromise RNA immobilization.

If successful, the RAP-MS developed here, would confirm the presumed role of the conserved hairpin or the termination signal in SBV transcription/translation processes and would highlight the proteins contributing to those mechanisms. More generally, it should shed light not only on the biology of SBV, but also on fundamental biology processes such as translation initiation in the absence of poly(A) tail. As particularly suited for the investigation of RNA–protein interactions, this technique is expected to be applicable to the study of other RNA-binding proteins. Besides, numerous illnesses have been associated with improper regulation or inappropriate function of RNA-binding proteins, so deciphering RNA–protein interactions has also a major impact on understanding physiology and disease.

References

References

Abbas W., Kumar A. & Herbein G. (2015) The eEF1A Proteins: At the Crossroads of Oncogenesis, Apoptosis, and Viral Infections. *Frontiers in Oncology* 5, 75.

Ariza A., Tanner S.J., Walter C.T., Dent K.C., Shepherd D.A., Wu W., Matthews S.V., Hiscox J.A., Green T.J., Luo M., Elliott R.M., Fooks A.R., Ashcroft A.E., Stonehouse N.J., Ranson N.A., Barr J.N. & Edwards T.A. (2013) Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization. *Nucleic Acids Research* 41, 5912–5926.

Baltz A.G., Munschauer M., Schwanhäusser B., Vasile A., Murakawa Y., Schueler M., Youngs N., Penfold-Brown D., Drew K., Milek M., Wyler E., Bonneau R., Selbach M., Dieterich C., Landthaler M. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular Cell* 46, 674–690.

Banerjee A. K. (1980) 5'-terminal cap structure in eucaryotic messenger ribonucleic acids. *Microbiology Reviews* 44, 175–205.

Beer M., Conraths F.J. & van der Poel W.H. (2013) 'Schmallenberg virus'--a novel orthobunyavirus emerging in Europe. *Epidemiology and infection* 141, 1-8.

Bhullar D., Jalodia R., Kalia M. & Vrati S. (2014) Cytoplasmic translocation of polypyrimidine tract-binding protein and its binding to viral RNA during Japanese encephalitis virus infection inhibits virus replication. *PLoS One* 9, e114931.

Blakqori G. & Weber F. (2005) Efficient cDNA-based rescue of La Crosse bunyaviruses expressing or lacking the nonstructural protein NSs. *Journal of Virology* 79, 10420–10428.

Blomström A. L., Gu Q., Barry G., Wilkie G., Skelton J. K., Baird M., McFarlane M., Schnettler E., Elliott R.M., Palmarini M. & Kohl A. (2015) Transcriptome analysis reveals the host response to Schmallenberg virus in bovine cells and antagonistic effects of the NSs protein *BioMed Central Genomics* 16, 324.

Boot H. J. & Pritz-Verschuren S.B. E. (2004) Modifications of the 3'-UTR stem-loop of infectious bursal disease virus are allowed without influencing replication or virulence. *Nucleic Acids Research* 32, 211-222.

Boseret G., Claine F., Coupeau D., Muylkens B., Saegerman C., Chantraine F., Kirschvink N. (2014) Abstract Submission In Utero infection by Schmallenberg virus induces Arthrogryposis multiplexa congenital in ovine fetuses, COGI.

Brennicke A., Marchfelder A. & Binder S. (1999). RNA editing *FEMS Microbiology Review* 23, 297–316.

Briese T., Calisher C.H. & Higgs S. (2013) Viruses of the family Bunyaviridae: are all available isolates reassortants? *Virology* 446, 207-16.

Butter F., Scheibe M., Mörl M. & Mann M. (2009) Unbiased RNA-protein interaction screen by quantitative proteomics. *Proceedings of National Academy of Sciences of United States of America* 106, 10626-31.

Carlile T.M., Rojas-Duran M.F., Zinshteyn B., Shin H., Bartoli K.M. & Gilbert W.V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515, 143-6.

Chiari M., Sozzi E., Zanoni M., Alborali L.G., Lavazza A. & Cordioli P. (2014) Serosurvey for Schmallenberg virus in alpine wild ungulates. *Transboundary and emerging diseases* 61, 1-3.

Coupeau D., Claine F., Wigger L., Martin B, Kirschvink N. & Muylkens B. (2013) Characterization of messenger RNA termini in Schmallenberg virus and related Simbuviruses. *Journal of General Virology*, 94, 2399-405.

De Regge N., Deblauwe I., De Deken R., Vantieghem P., Madder M., Geysen D., Smeets F., Losson B., van den Berg T. & Cay A.B. (2012) Detection of Schmallenberg virus in different Culicoides spp. by real-time RT-PCR. *Transboundary & Emerging Diseases* 59, 471-5.

Depierreux D. (2014) Interaction du virus de Schmallenberg et des virus apparentés dans des cellules de mammifères et d'insectes vecteurs, *master thesis*.

Doceul V., Lara E., Sailleau C., Belbis G., Richardson J., Bréard E., Viarouge C., Dominguez M., Hendrikx P., Calavas D., Desprat A., Languille J., Comtet L., Pourquier P., Eléouët J-F., Delmas B., Marianneau P., Vitour D. & Zientara S. (2013) Epidemiology, molecular virology and diagnostics of Schmallenberg virus, an emerging orthobunyavirus in Europe. *Veterinary Research* 15, 44-31.

Domingo E. (2012) Mechanisms of viral emergence. *Veterinary research* 41, 38.

Dominissini D., Moshitch-Moshkovitz S., Schwartz S., Salmon-Divon M., Ungar L., Osenberg S., Cesarkas K., Jacob-Hirsch J., Amariglio N., Kupiec M., Sorek R. & Rechavi G. (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206.

Dong H., Li P., Elliott R. M. & Dong C. (2013a) Structure of Schmallenberg orthobunyavirus nucleoprotein suggests a novel mechanism of genome encapsidation. *Journal of Virology* 87, 5593–5601.

Dong H., Li P., Böttcher B., Elliott R.M. & Dong C. (2013b) Crystal structure of Schmallenberg orthobunyavirus nucleoprotein-RNA complex reveals a novel RNA sequestration mechanism. *RNA* 19, 1129-36.

Eifan S., Schnettler E., Dietrich I., Kohl A. & Blomstrom A. L. (2013) Non-structural proteins of arthropod-borne bunyaviruses: roles and functions. *Viruses* 5, 2447–2468.

Elbers A. R.W., Meiswinkel R., van Weezep E., M.M. Sloet van Oldruitenborgh-Oosterbaan & Kooi E.A. (2013) Schmallenberg Virus in Culicoides spp. Biting Midges, the Netherlands, 2011 *Emerging Infectious Disease* 19, 106–109.

European Food and Safety Authority Schmallenberg virus: analysis of the epidemiological data 2013 available from http://www.efsa.europa.eu/fr/supporting/pub/429e.

Faoro C. & Ataide S. F. (2014) Ribonomic approaches to study the RNA-binding proteome. *FEBS Letters* 588, 3649–64.

Fields Virology by Lippincott, Williams, Wilkins. (2007) Volume 2 5th Edition, 3091.

Fuentes-Pananá E.M., Peng R., Brewer G., Tan J. & Ling P.D. (2000) Regulation of the Epstein-Barr Virus C Promoter by AUF1 and the Cyclic AMP/Protein Kinase A Signaling Pathway. *Journal of Virology* 74, 8166–8175.

Garigliany M.M., Bayrou C., Kleijnen D., Cassart D., Jolly S., Linden A. & Desmecht D. (2012) Schmallenberg virus: a new Shamonda/Sathuperi-like virus on the rise in Europe. *Antiviral Research 95*, 82–7.

Gavazzi C., Yver M., Isel C., Smyth R.P., Rosa-Calatrava M., Lina B., Moulès V. & Marquet R. (2013) A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proceedings of the National Academy of Sciences* 110, 16604-9.

Gerstberger S., Hafner M., Ascano M. & Tuschl T. (2014) Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. *Advances in Experimental Medicine and Biology* 825, 1-55.

Glisovic T., Bachorik J.L., Yong J. & Dreyfuss G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters* 582, 1977-86.

Goller K.V., Höper D., Schirrmeier H. & Mettenleiter T. C. & Beer M. (2012) Schmallenberg Virus as Possible Ancestor of Shamonda Virus. *Emerging infectious diseases* 18, 1644-6.

González M., Bagatolli L.A., Echabe I., Arrondo J.L., Argaraña C.E., Cantor C.R. & Fidelio G.D. Interaction of biotin with streptavidin. (1998) Thermostability and conformational changes upon binding. *Journal of Biological Chemitry* 272, 11288-94.

Gopinath S.C.B. (2009) Mapping of RNA–protein interactions. *Analytica Chimica Acta* 636, 117–28.

Guhaniyogi J. & Brewer G. (2001) Regulation of mRNA stability in mammalian cells. *Gene* 265, 11-23.

Gupta A.K., Drazba J.A. & Banerjee A.K. (1998) Specific interaction of heterogeneous nuclear ribonucleoprotein particle U with the leader RNA sequence of vesicular stomatitis virus. *Journal of Virology* 72, 8532-40.

Hacker J. K. & Hardy J. L. (1997) Adsorptive endocytosis of California encephalitis virus into mosquito and mammalian cells: a role for G1. *Virology* 235, 40–47.

Hahn K., Habierski A., Herder V., Wohlsein P., Peters M., Hansmann F. & Baumgartner W. (2012) Schmallenberg virus in central nervous system of ruminants. *Emerging Infectious Diseases* 19, 154–155.

Hentze M.W., Caughman S.W., Rouault T.A., Barriocanal J.G., Dancis A., Harford J.B., Klausner R.D. (1987) Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science* 238, 1570-3.

Herder V., Wohlsein P., Peters M., Hansmann F. & Baumgartner W. (2012) Salient lesions in domestic ruminants infected with the emerging so-called Schmallenberg virus in Germany. *Veterinary Pathology* 49, 588–591.

Hirsch J.D., Eslamizar L., Filanoski B.J., Malekzadeh N., Haugland R.P., Beechem J.M. & Haugland R.P. (2002) Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Analytical Biochemistry* 308, 343–57.

Hoffmann B., Scheuch M., Höper D., Jungblut R., Holsteg M., Schirrmeier H., Eschbaumer M., Goller K. V., Wernike K., Fischer M., Breithaupt A., Mettenleiter T. C. & Beer M. (2012) Novel Orthobunyavirus in Cattle, Europe. *Emerging Infectious Diseases* 18, 469-72.

Hogan G.J., Brown P.O. & Herschlag D. (2015) Evolutionary Conservation and Diversification of Puf RNA Binding Proteins and Their mRNA Targets. *PLoS Biology* 13, :e1002307.

Hollidge B.S., Nedelsky N.B., Salzano M.V., Fraser J.W., González-Scarano F. & Soldan S.S. (2012) Orthobunyavirus entry into neurons and other mammalian cells occurs via clathrin-mediated endocytosis and requires trafficking into early endosomes. *Journal of Virology* 86, 7988–8001.

Hudson W.H. & Ortlund E.A. (2014) The structure, function and evolution of proteins that bind DNA and RNA. *Nature Review Molecular Cell Biology* 15, 749-60.

Huntzinger E. & Izaurralde E. (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Review Genetics* 12, 99-110.

Hutchinson E. C., von Kirchbach J. C., Gog J. R. & Digard P. (2010) Genome packaging in influenza A virus. *Journal of General Virology* 91, 313–328.

INTECH. (2013) Current Issues in Molecular Virology - Viral Genetics and Biotechnological Applications, Chapter 2: RNA 5′-end Maturation: A Crucial Step in the Replication of Viral Genomes.

Iroegbu C. U. & Pringle C. R. (1981) Genetic interactions among viruses of the Bunyamwera complex. *Journal of Virology* 37, 383–394.

Jack C., Anstaett O., Adams J., Noad R., Brownlie J. & Mertens P. (2012) Evidence of seroconversion to SBV in camelids. *The veterinary record* 170, 603.

Jackson R.J., Hellen C.U. & Pestova T.V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology* 11, 113-27.

Jankowsky E. & Harris M.E. (2015) Specificity and nonspecificity in RNA-protein interactions. *Nature Reviews Molecular Cell Biology* 16, 533-44.

Jones K.E., Patel N.G., Levy M.A., Storeygard A., Balk D., Gittleman J.L. & Daszak P. (2008) Global trends in emerging infectious disease. *Nature* 451, 990-3.

Keene J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nature Review Genetics* 8, 533-43.

Keene J.D. (2010) Minireview: global regulation and dynamics of ribonucleic acid. *Endocrinology* 151, 1391–7.

Kishore S., Luber S. & Zavolan M. (2010) Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Briefings in Functional Genomics* 9, 391-404.

Ko C., Lee S., Windisch M.P. & Ryu W.S. (2014) DDX3 DEAD-box RNA helicase is a host factor that restricts hepatitis B virus replication at the transcriptional level. *Journal of Virology* 88, 13689-98.

Kohl A., Clayton R.F., Weber F., Bridgen A., Randall R.E. & Elliott R.M. (2003) Bunyamwera virus nonstructural protein NSs counteracts interferon regulatory factor 3-mediated induction of early cell death. *Journal of Virology* 77, 7999–8008.

Kohl A., Dunn E. F., Lowen A. C. & Elliott R. M. (2004) Complementarity, sequence and structural elements within the 3′and 5′non-coding regions of the Bunyamwera orthobunyavirus S segment determine promoter strength. *Journal of General Virology* 85, 3269–78.

Kohl A., Lowen A. C., Leonard V. H. J. & Elliott R. M. (2006) Genetic elements regulating packaging of the Bunyamwera orthobunyavirus genome. *Journal of General Virology* 87, 177–187.

Kukkonen S. K. J., Vaheri A. & Plyusnin A. (2004) Tula hantavirus L protein is a 250 kDa perinuclear membrane-associated protein. *Journal of General Virology* 85, 1181–1189.

Larska M., Krzysiak M., Smreczak M., Polak M.P. & Zmudziński J.F. (2013a) First detection of Schmallenberg virus in elk (Alces alces) indicating infection of wildlife in Białowieża National Park in Poland. *Veterinary journal* 198, 279-81.

Larska M., Lechowski L., Grochowska M. & Żmudziński J.F. (2013b) Detection of the Schmallenberg virus in nulliparous Culicoides obsoletus/scoticus complex and C. punctatus--the possibility of transovarial virus transmission in the midge population and of a new vector. *Veterinary microbiology* 166, 467-73.

Li Z. & Nagy P.D. (2011) Diverse roles of host RNA-binding proteins in RNA virus replication. *RNA Biology* 8, 305-315.

Liao B., Hu Y. & Brewer G (2007) *Competitive binding of AUF1 and TIAR to MYC mRNA controls its translation.* Nature Structural Molecular Biology 14, 511-8.

Lievaart-Petersona K., Luttikholta S., Peperkampb K., Van den Broma R. & Vellema P. (2015) Schmallenberg disease in sheep or goats: Past, present and future. *Veterinary microbiology* pii: S0378-1135(15)00313-2.

Linden A., Desmecht D., Volpe R., Wirtgen M., Gregoire F., Pirson J., Paternostre J., Kleijnen D., Schirrmeier H., Beer M. & Garigliany M.M. (2012) Epizootic spread of Schmallenberg virus among wild cervids, Belgium, Fall 2011. *Emerging Infectious Diseases* 18, 2006-8.

Liu N., Dai Q., Zheng G., He C., Parisien M. & Pan T. (2015) N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* 518, 560-564.

Lloyd R.E. (2015) Nuclear proteins hijacked by mammalian cytoplasmic plus strand RNA viruses. *Virology* 479-480, 457-74.

Lodeiro M.F., Filomatori C.V. & Gamarnik A.V. (2009) Structural and functional studies of the promoter element for dengue virus RNA replication. *Journal of Virology* 83, 993-1008.

Lozach P.Y., Kühbacher A., Meier R., Mancini R., Bitto D., Bouloy M. & Helenius A. (2011) DC -SIGN as a receptor for phleboviruses. *Cell Host & Microbe* 10, 75–88.

Ludwig G. V., Israel B. A., Christensen B. M., Yuill T. M. & Schultz K. T. (1991) Role of La Crosse virus glycoproteins in attachment of virus to host cells. *Virology* 181, 564–571.

Luo G., Xin S. & Cai Z. (2003) Role of the 5'-proximal stem-loop structure of the 5' untranslated region in replication and translation of hepatitis C virus RNA. *Journal of Virology* 77, 3312-8.

Luo Z., Dong X., Li Y., Zhang Q., Kim C., Song Y., Kang L., Liu Y., Wu K. & Wu J. (2014) PolyC-binding protein 1 interacts with 5'-untranslated region of enterovirus 71 RNA in membrane-associated complex to facilitate viral replication. *PLoS One* 9, e87491.

Mazel-Sanchez, B. & Elliott, R. M. (2012) Attenuation of Bunyamwera orthobunyavirus replication by targeted mutagenesis of genomic UTRs and creation of viable viruses with minimal genome segments. *Journal of Virology* 86, 13672–13678.

Martinelle L., Dal Pozzo F., Kirschvink N., De La Grandière M.A., Thiry E. & Saegerman C. (2012) Le virus Schmallenberg ou l'émergence du premier Orthobunyavirus du sérogroupe Simbu en Europe. *Annales de Médecine Vétérinaire* 156, 07-24.

Martínez-Salas E., Lozano G., Fernandez-Chamorro J., Francisco-Velilla R., Galan A. & Diaz R (2013) RNA-binding proteins impacting on internal initiation of translation. *International Journal of Molecular Sciences* 4, 21705-26.

Marzluff W.F., Wagner E.J. & Duronio R.J. (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature Review Genetics* 9, 843-54.

McKee A.E. & Silver P.A. (2007) Systems perspectives on mRNA processing. *Cell Research* 17, 581–590.

Mellor P.S. (2000) Replication of arboviruses in insect vectors. Journal of comparative pathology 123, 231-47.bac

Mitchell S.F. & Parker R. (2014) Principles and properties of eukaryotic mRNPs. *Molecular Cell* 54, 547-58.

Muller-McNicoll M. & Neugebauer K.M. (2013) How cells get the message: Dynamic assembly and function of mRNA-protein complexes. *Nature Review Genetics* 14, 275–287.

Oeffinger M. & Montpetit B. (2015) Emerging properties of nuclear RNP biogenesis and export. *Current Opinion in Cell Biology* 34, 46-53.

Pennington T. H. & Pringle C. R. (1978) Negative strand viruses in enucleate cells. *Negative strand viruses and the host cell*, 457-464.

Phillips T. (2008) Regulation of Transcription and Gene Expression in Eukaryotes. *Nature Education* 1, 199.

Ramsköld D., Wang E. T., Burge C. B. & Sandberg R. (2009) An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Computational Biology* 5, e1000598.

Real C.I., Megger D.A., Sitek B., Jahn-Hofmann K., Ickenstein L.M., John M.J., Walker A., Timm J., Kuhlmann K., Eisenacher M., Meyer H.E., Gerken G., Broering R. & Schlaak J.F. (2013) Identification of proteins that mediate the pro-viral functions of the interferon stimulated gene 15 in hepatitis C virus replication. *Antiviral Research* 100, 654-61.

Reguera J., Weber F. & Cusack S. (2010) Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. *PLoS Pathogens* 6, e1001101.

Reusken C., van den Wijngaard C., van Beek P., Beer M., Bouwstra R., Godeke G.J., Isken L., van den Kerkhof H., van Pelt W., van der Poel W., Reimerink J., Schielen P., Schmidt-Chanasit J., Vellema P., de Vries A., Wouters I. & Koopmans M. (2012) Lack of Evidence for Zoonotic Transmission of Schmallenberg Virus. *Emerging Infectious Diseases* 18, 1746-54.

Ribó M., Font J., Benito A., Torrent J., Lange R. & Vilanova M. (2006) Pressure as a tool to study protein-unfolding/refolding processes: the case of ribonuclease A. *Biochimica Biophysica Acta* 1764, 461-9.

Sailleau C., Boogaerts C., Meyrueix A., Laloy E., Bréard E., Viarouge C., Desprat A., Vitour D., Doceul V., Boucher C., Zientara S., Nicolier A. & Grandjean D. (2013) Schmallenberg virus infection in dogs, France, 2012. *Emerging Infectious Disease* 19, 1896-8.

Shi X., Goli J., Clark G., Brauburger K. & Elliott R. M. (2009) Functional analysis of the Bunyamwera orthobunyavirus Gc glycoprotein. *Journal of General Virology* 90, 2483–2492.

Shi X., Kohl A., Leonard V.H., Li P., McLees A. & Elliott R.M. (2006) Requirement of the N-terminal regin of orthobunyavirus nonstructural protein NSm for virus assembly and morphogenesis. *Journal of Virology* 80, 8089-99.

Shi X., Kohl A., Li P. & Elliott R. M. (2007) Role of the cytoplasmic tail domains of Bunyamwera orthobunyavirus glycoproteins Gn and Gc in virus assembly and morphogenesis. *Journal of Virology 81*, 10151–10160.

Shi X., Lappin D. F. & Elliott R. M. (2004) Mapping the Golgi targeting and retention signal of Bunyamwera virus glycoproteins. *Journal of Virology* 78, 10793–10802.

Somogyi P., Jenner A.J., Brierley I. & Inglis S.C. (1993) Ribosomal pausing during translation of an RNA pseudoknot. *Molecular Cell Biology* 3, 6931-40.

Sonenberg N. & Hinnebusch A.G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731-45.

Staple D.W. & Butcher S.E. (2005) Pseudoknots: RNA Structures with Diverse Functions. *PloS Biology* 3, e213.

Steukers L., Bertels G., Cay A.B. & Nauwynck H.J. (2012) Schmallenberg virus: emergence of an Orthobunyavirus among ruminants in Western Europe. *Vlaams Diergeneeskundig Tijdschrift* 81, 119–127.

Svoboda P. & Di Cara A. (2006) Hairpin RNA: a secondary structure of primary importance. *Cellular & Molecular Life Sciences* 63, 901–918.

Tacheny A., Dieu M., Arnould T., Renard P. (2013) Mass spectrometry-based identification of proteins interacting with nucleic acids. *Journal of proteomics* 94, 89-109.

Tacheny A., Michel S., Dieu M., Payen L., Arnould T. & Renard P. (2012) Unbiased proteomic analysis of proteins interacting with the HIV-1 50 LTR sequence: role of the transcription factor Meis. *Nucleic Acids Research* 40, e168.

Theimer C.A. Blois C.A. & Feigon J. (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Molecular Cell* 17, 671-82.

Thomas D., Blakqori G., Wagner V., Banholzer M., Kessler N., Elliott R.M., Haller O. & Weber F. (2004) Inhibition of RNA polymerase II phosphorylation by a viral interferon antagonist. *Journal of Biological Chemistry* 279, 31471–31477.

Truant A.L. & Hallum J.V. (1977) A persistent infection of baby hamster kidney-21 cells with mumps virus and the role of temperature-sensitive variants. *Journal of Medical Virology* 1, 49-67.

Tsai P.L., Chiou N.T., Kuss S., García-Sastre A., Lynch K.W. & Fontoura B.M. (2013) Cellular RNA binding proteins NS1-BP and hnRNP K regulate influenza A virus RNA splicing. *PLoS Pathogens* 9, e1003460.

Valiente-Echeverría F., Hermoso M.A. & Soto-Rifo R. (2015) RNA helicase DDX3: at the crossroad of viral replication and antiviral immunity. *Reviews in Medical Virology* 25, 286-99.

Valinezhad O. A., Safaralizadeh R. & Kazemzadeh-Bavili M. (2014) Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation. *International journal of genomics*, 970607.

Varela M., Schnettler E., Caporale M., Murgia C., Barry G., McFarlane M., McGregor E., Piras I.M., Shaw A., Lamm C., Janowicz A., Beer M., Glass M., Herder V., Hahn K., Baumgärtner W., Kohl A. & Palmarini M. (2013) Schmallenberg virus pathogenesis, tropism and interaction with the innate immune system of the host. *PLoS Pathogens* 9, e1003133.

Vera-Otarola J., Soto-Rifo R., Ricci E.P., Ohlmann T., Darlix J.L. & López-Lastra M. (2010) The 3' untranslated region of the Andes hantavirus small mRNA functionally replaces the poly(A) tail and stimulates cap-dependent translation initiation from the viral mRNA. *Journal of Virology* 84, 10420-4.

Veronesi E., Henstock M., Gubbins S., Batten C., Manley R., Barber J., Hoffmann B., Beer M., Attoui H., Mertens P. P.C. & Carpenter S. (2013) Implicating Culicoides biting midges as vectors of Schmallenberg virus using semi-quantitative RT-PCR, *PLoS One* 8, e57747.

Wensman J.J., Blomqvist G., Hjort M. & Holst B.S. (2013) Presence of antibodies to Schmallenberg virus in a dog in Sweden. *Journal of Clinical Microbiology* 8, 2802-3.

Wernike K., Hoffmann B., Bréard E., Bøtner A., Ponsart C., Zientara S., Lohse L., Pozzi N., Viarouge C., Sarradin P., Leroux-Barc C., Riou M., Laloyb E., Breithaupt A. & Beer M. (2013) Schmallenberg virus experimental infection of sheep. *Veterinary Microbiology* 166, 461–466.

Wernike K., Jöst H., Becker N., Schmidt-Chanasit J. & Beer M. (2014) Lack of evidence for the presence of Schmallenberg virus in mosquitoes in Germany, 2011. *Parasites & Vectors* 7, 402.

White E.J., Brewer G. & Wilson G.M. (2013) Post-transcriptional control of gene expression by AUF1: mechanisms, physiological targets, and regulation. *Biochimica Biophysica Acta* 1829, 680-8.

Yanase T., Kato T., Aizawa M., Shuto Y., Shirafuji H., Yamakawa M. & Tsuda T. (2012) Genetic reassortment between Sathuperi and Shamonda viruses of the genus Orthobunyavirus in nature: implications for their genetic relationship to Schmallenberg virus. *Archives of virology* 157, 1611-6.

Yi M. & Lemon S.M. (2003) Structure-function analysis of the 3' stem-loop of hepatitis C virus genomic RNA and its role in viral RNA replication. *RNA* 9, 331-45.

Zenkin N. (2014) RNA secondary structure-dependent termination of transcription. *Cell Cycle* 13, 3-4.

Supplementary Data

#	Identified proteins	Accession number	Molecular Weight	I	NI
1	Cluster of Transcriptional activator protein Pur-beta n=1 Tax=Mus musculus RepID=PURB_MOUSE (O35295)	O35295 [30]	34 kDa	74	63
2	Cluster of ALB protein n=1 Tax=Bos taurus RepID=B0JYQ0_BOVIN (B0JYQ0)	B0JYQ0 [18]	69 kDa	41	35
3	Cluster of PREDICTED: replication protein A 70 kDa DNA-binding subunit n=1 Tax=Mesocricetus auratus RepID=UPI000359E136 (UPI000359E136)	UPI000359E136 [17]	68 kDa	24	20
4	Cluster of Replication protein A 32 kDa subunit n=1 Tax=Rattus norvegicus RepID=RFA2_RAT (Q63528)	Q63528 [20]	29 kDa	21	17
5	Cluster of PREDICTED: acetyl-CoA carboxylase 1 isoform X1 n=2 Tax=Mesocricetus auratus RepID=UPI0006603A3D (UPI0006603A3D)	UPI0006603A3D [43]	270 kDa	6	5
6	Cluster of PREDICTED: epiplakin n=1 Tax=Peromyscus maniculatus bairdii RepID=UPI00042ACD9F (UPI00042ACD9F)	UPI00042ACD9F [63]	323 kDa	25	19
7	Cluster of PREDICTED: single-stranded DNA-binding protein, mitochondrial isoform X1 n=2 Tax=Mesocricetus auratus RepID=UPI000359C7DE (UPI000359C7DE)	UPI000359C7DE [6]	17 kDa	22	17
8	Cluster of Transthyretin n=1 Tax=Bos taurus RepID=TTHY_BOVIN (O46375)	O46375	16 kDa	26	22
9	Cluster of PREDICTED: RNA-binding motif, single-stranded-interacting protein 2 isoform X3 n=1 Tax=Chinchilla lanigera RepID=UPI00069703FB (UPI00069703FB)	UPI00069703FB [22]	45 kDa	21	19
10	Cluster of Uncharacterized protein n=1 Tax=Ictidomys tridecemlineatus RepID=I3NCC6_ICTTR (I3NCC6)	I3NCC6 [41]	37 kDa	6	5
11	Cluster of PREDICTED: replication protein A 14 kDa subunit n=1 Tax=Mesocricetus auratus RepID=UPI000359CEC5 (UPI000359CEC5)	UPI000359CEC5 [6]	14 kDa	12	15
12	Cluster of Uncharacterized protein n=1 Tax=Bos taurus RepID=F1N076_BOVIN (F1N076)	F1N076 [33]	124 kDa	40	34
13	Cluster of PREDICTED: pyruvate carboxylase, mitochondrial n=1 Tax=Nannospalax galili RepID=UPI0004ED09C8 (UPI0004ED09C8)	UPI0004ED09C8 [14]	130 kDa	31	43
14	Cluster of PREDICTED: F-box only protein 18 n=1 Tax=Mesocricetus auratus RepID=UPI00065FD554 (UPI00065FD554)	UPI00065FD554 [20]	83 kDa	27	34
15	Cluster of Valyl-tRNA synthetase-like protein n=1 Tax=Cricetulus griseus RepID=A0A061IM03_CRIGR (A0A061IM03)	A0A061IM03 [43]	263 kDa	22	23
16	Cluster of PREDICTED: nestin n=1 Tax=Mesocricetus auratus RepID=UPI0006602EAB (UPI0006602EAB)	UPI0006602EAB [16]	221 kDa	24	33
17	Cluster of PREDICTED: LOW QUALITY PROTEIN: RNA-binding protein Musashi homolog 1 isoform X2 n=1 Tax=Cricetulus griseus RepID=UPI0004542AFA (UPI0004542AFA)	UPI0004542AFA [22]	63 kDa	12	14
18	Cluster of PREDICTED: LOW QUALITY PROTEIN: klotho-like n=1 Tax=Heterocephalus glaber RepID=UPI0006573018 (UPI0006573018)	UPI0006573018 [76]	113 kDa	23	26

19	Cluster of PREDICTED: E3 ubiquitin-protein ligase HERC2 n=1 Tax=Mesocricetus auratus RepID=UPI00066077F8 (UPI00066077F8)	UPI00066077F8 [28]	498 kDa	11	38
20	Cluster of PREDICTED: LOW QUALITY PROTEIN: tubulin beta-3 chain n=1 Tax=Chinchilla lanigera RepID=UPI00038E9530 (UPI00038E9530)	UPI00038E9530 [45]	95 kDa	22	24
21	Cluster of Peripheral-type benzodiazepine receptor-associated protein 1 n=1 Tax=Cricetulus griseus RepID=G3HJ76_CRIGR (G3HJ76)	G3HJ76 [2]	285 kDa	2	2
22	Cluster of Propionyl-CoA carboxylase alpha chain, mitochondrial n=1 Tax=Rattus norvegicus RepID=A0A0G2K401_RAT (A0A0G2K401)	A0A0G2K401 [19]	80 kDa	14	21
23	Cluster of PREDICTED: embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein n=1 Tax=Mesocricetus auratus RepID=UPI00035A0F89 (UPI00035A0F89)	UPI00035A0F89 [12]	40 kDa	15	15
24	Cluster of PREDICTED: NFX1-type zinc finger-containing protein 1 n=1 Tax=Mesocricetus auratus RepID=UPI00035A1FED (UPI00035A1FED)	UPI00035A1FED [16]	220 kDa	20	9
25	Cluster of PREDICTED: filamin-A isoform X3 n=1 Tax=Bos mutus RepID=UPI0003C0175F (UPI0003C0175F)	UPI0003C0175F [49]	275 kDa	2	13
26	Cluster of Pantetheinase n=1 Tax=Bos mutus RepID=L8J052_9CETA (L8J052)	L8J052 [5]	58 kDa	10	9
27	Cluster of Elongation factor 1-alpha 1 n=1 Tax=Heterocephalus glaber RepID=G5BNQ5_HETGA (G5BNQ5)	G5BNQ5 [22]	48 kDa	5	6
28	Cluster of PREDICTED: lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial n=1 Tax=Mesocricetus auratus RepID=UPI000359D279 (UPI000359D279)	UPI000359D279 [12]	53 kDa	11	12
29	Cluster of PREDICTED: anionic trypsin-2-like n=1 Tax=Cricetulus griseus RepID=UPI0004546481 (UPI0004546481)	UPI0004546481 [11]	30 kDa	9	8
30	Cluster of PREDICTED: propionyl-CoA carboxylase beta chain, mitochondrial n=1 Tax=Bos mutus RepID=UPI0003C03056 (UPI0003C03056)	UPI0003C03056 [15]	58 kDa	8	12
31	Cluster of Nucleolysin TIAR n=1 Tax=Cricetulus griseus RepID=A0A061I9X0_CRIGR (A0A061I9X0)	A0A061I9X0 [15]	39 kDa	10	9
32	Cluster of PREDICTED: peroxiredoxin-1 n=1 Tax=Mesocricetus auratus RepID=UPI000660F86A (UPI000660F86A)	UPI000660F86A [10]	26 kDa	10	11
33	Cluster of PREDICTED: microtubule-actin cross-linking factor 1 isoform X7 n=1 Tax=Homo sapiens RepID=UPI000387C899 (UPI000387C899)	UPI000387C899 [3]	870 kDa	0	0
34	Cluster of Putative uncharacterized protein n=1 Tax=Mus musculus RepID=Q3U6V3_MOUSE (Q3U6V3)	Q3U6V3 [3]	72 kDa	13	16
35	Cluster of PREDICTED: mitochondrial genome maintenance exonuclease 1 n=1 Tax=Mesocricetus auratus RepID=UPI000359A2AE (UPI000359A2AE)	UPI000359A2AE [4]	39 kDa	8	11
36	Cluster of PREDICTED: myosin-14 isoform X4 n=1 Tax=Homo sapiens RepID=UPI0005D02172 (UPI0005D02172)	UPI0005D02172 [44]	230 kDa	7	0

37	Cluster of Neuroblast differentiation-associated protein AHNAK n=4 Tax=Homo sapiens RepID=AHNK_HUMAN (Q09666)	Q09666 [32]	629 kDa	0	3
38	Cluster of PREDICTED: DNA topoisomerase 3-alpha n=1 Tax=Mesocricetus auratus RepID=UPI00065F74AB (UPI00065F74AB)	UPI00065F74AB [22]	22 kDa	5	20
39	Cluster of PREDICTED: heterogeneous nuclear ribonucleoprotein D0 n=1 Tax=Ictidomys tridecemlineatus RepID=UPI00067FC924 (UPI00067FC924)	UPI00067FC924 [5]	38 kDa	9	8
40	Cluster of Uncharacterized protein n=1 Tax=Bos taurus RepID=G3MY87_BOVIN (G3MY87)	G3MY87 [2]	207 kDa	11	9
41	Cluster of Carboxypeptidase Q n=3 Tax=Bos taurus RepID=CBPQ_BOVIN (Q17QK3)	Q17QK3 [6]	52 kDa	7	5
42	Cluster of Activated RNA polymerase II transcriptional coactivator p15 n=1 Tax=Rattus norvegicus RepID=TCP4_RAT (Q63396)	Q63396 [3]	14 kDa	8	6
43	Cluster of cDNA FLJ51903, highly similar to Stress-70 protein, mitochondrial n=1 Tax=Homo sapiens RepID=B7Z4T3_HUMAN (B7Z4T3)	B7Z4T3 [6]	70 kDa	5	9
44	Cluster of PREDICTED: kinesin-like protein KIF1B isoform X8 n=1 Tax=Mesocricetus auratus RepID=UPI00035973BC (UPI00035973BC)	UPI00035973BC [20]	130 kDa	0	12
45	Cluster of PREDICTED: cationic trypsin-3 n=1 Tax=Ictidomys tridecemlineatus RepID=UPI00067FC52A (UPI00067FC52A)	UPI00067FC52A [3]	14 kDa	5	5
46	Cluster of Vitamin D-binding protein n=1 Tax=Bos taurus RepID=F1N5M2_BOVIN (F1N5M2)	F1N5M2 [3]	53 kDa	7	8
47	Cluster of Polymerase I and transcript release factor n=1 Tax=Rattus norvegicus RepID=G3V8L9_RAT (G3V8L9)	G3V8L9 [10]	44 kDa	2	10
48	Cluster of PREDICTED: fascin n=1 Tax=Peromyscus maniculatus bairdii RepID=UPI00042AC6EC (UPI00042AC6EC)	UPI00042AC6EC [6]	54 kDa	10	4
49	Cluster of Protein Rcn1 n=1 Tax=Rattus norvegicus RepID=D3ZUB0_RAT (D3ZUB0)	D3ZUB0 [9]	38 kDa	5	8
50	MCG124046 n=1 Tax=Mus musculus RepID=Q9Z1R9_MOUSE	Q9Z1R9	26 kDa	2	2
51	Cluster of 60S ribosomal protein L6 n=1 Tax=Cricetulus griseus RepID=G3H278_CRIGR (G3H278)	G3H278 [11]	297 kDa	2	7
52	Cluster of Putative heat shock protein HSP 90-beta 4 n=1 Tax=Homo sapiens RepID=H90B4_HUMAN (Q58FF6)	Q58FF6 [25]	58 kDa	0	15
53	Cluster of Heterogeneous nuclear ribonucleoprotein A/B n=1 Tax=Cricetulus griseus RepID=G3H2E2_CRIGR (G3H2E2)	G3H2E2 [7]	24 kDa	7	6
54	Cluster of Hemoglobin subunit alpha n=1 Tax=Bos taurus RepID=HBA_BOVIN (P01966)	P01966	15 kDa	5	4
55	Cluster of PREDICTED: DNA helicase B n=1 Tax=Mesocricetus auratus RepID=UPI0006607B03 (UPI0006607B03)	UPI0006607B03 [3]	123 kDa	5	11
56	Cluster of PREDICTED: recQ-mediated genome instability protein 1 n=1 Tax=Microtus ochrogaster RepID=UPI00038BE1F4 (UPI00038BE1F4)	UPI00038BE1F4 [3]	69 kDa	2	4
57	Cluster of PREDICTED: cellular nucleic acid-binding protein isoform X3 n=1 Tax=Jaculus jaculus RepID=UPI000332F6CB (UPI000332F6CB)	UPI000332F6CB [7]	19 kDa	5	5

58	Cluster of tRNA-splicing endonuclease subunit Sen34 n=1 Tax=Fukomys damarensis RepID=A0A091DZH6_FUKDA (A0A091DZH6)	A0A091DZH6 [3]	57 kDa	5	7
59	Cluster of PREDICTED: LOW QUALITY PROTEIN: S-phase kinase-associated protein 1-like n=1 Tax=Mesocricetus auratus RepID=UPI00065FCE8F (UPI00065FCE8F)	UPI00065FCE8F [5]	19 kDa	4	5
60	Cluster of Transcription termination/antitermination protein NusG n=273 Tax=root RepID=NUSG_ECO57 (P0AFG1)	P0AFG1	21 kDa	7	6
61	Cluster of PREDICTED: collagen alpha-1(I) chain isoform X1 n=1 Tax=Homo sapiens RepID=UPI00051268FD (UPI00051268FD)	UPI00051268FD [2]	133 kDa	0	0
62	Cluster of PREDICTED: methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial isoform X1 n=2 Tax=Nannospalax galili RepID=UPI0004ED3205 (UPI0004ED3205)	UPI0004ED3205 [11]	79 kDa	5	4
63	Cluster of 40S ribosomal protein S4 n=1 Tax=Fukomys damarensis RepID=A0A091CW34_FUKDA (A0A091CW34)	A0A091CW34 [15]	27 kDa	3	6
64	PREDICTED: cationic trypsin-3 n=1 Tax=Dipodomys ordii RepID=UPI00065232FD	UPI00065232FD	26 kDa	5	3
65	Cluster of PREDICTED: poly(rC)-binding protein 1 n=1 Tax=Cricetulus griseus RepID=UPI0004542E98 (UPI0004542E98)	UPI0004542E98 [4]	38 kDa	5	5
66	Cluster of 40S ribosomal protein S12 n=1 Tax=Cavia porcellus RepID=H0UTX4_CAVPO (H0UTX4)	H0UTX4	15 kDa	4	4
67	Cluster of Hemoglobin subunit beta n=5 Tax=Bos RepID=HBB_BOVIN (P02070)	P02070 [2]	16 kDa	5	5
68	Cluster of PREDICTED: regulator of nonsense transcripts 1 isoform X2 n=1 Tax=Nannospalax galili RepID=UPI0004ED0B21 (UPI0004ED0B21)	UPI0004ED0B21 [4]	123 kDa	6	0
69	Cluster of Fumarylacetoacetase n=3 Tax=Homo sapiens RepID=FAAA_HUMAN (P16930)	P16930 [8]	46 kDa	2	4
70	Cluster of PREDICTED: beta-2-glycoprotein 1 n=1 Tax=Bos mutus RepID=UPI0003C02BC4 (UPI0003C02BC4)	UPI0003C02BC4 [2]	38 kDa	3	4
71	Alpha-1-acid glycoprotein n=1 Tax=Bos taurus RepID=A1AG_BOVIN	Q3SZR3 (+1)	23 kDa	4	0
72	Cluster of PREDICTED: 60S acidic ribosomal protein P2 isoform X1 n=1 Tax=Mesocricetus auratus RepID=UPI000359A69C (UPI000359A69C)	UPI000359A69C [7]	12 kDa	3	4
73	Cluster of Cullin-1 n=1 Tax=Homo sapiens RepID=A0A0C4DGX4_HUMAN (A0A0C4DGX4)	A0A0C4DGX4 [8]	87 kDa	0	8
74	Cluster of Uncharacterized protein n=1 Tax=Ictidomys tridecemlineatus RepID=I3MT23_ICTTR (I3MT23)	I3MT23 [9]	18 kDa	4	5
75	Cluster of Annexin n=1 Tax=Heterocephalus glaber RepID=G5AWC0_HETGA (G5AWC0)	G5AWC0 [7]	58 kDa	0	7
76	Cluster of PREDICTED: 60S acidic ribosomal protein P0 isoform X2 n=1 Tax=Cricetulus griseus RepID=UPI00022F38E4 (UPI00022F38E4)	UPI00022F38E4 [15]	35 kDa	2	4
77	Cluster of 60S ribosomal protein L27a n=3 Tax=Rodentia RepID=RL27A_MOUSE (P14115)	P14115 [9]	17 kDa	0	5

78	Cluster of PREDICTED: muscle-related coiled-coil protein n=1 Tax=Mesocricetus auratus RepID=UPI0003599463 (UPI0003599463)	UPI0003599463 [6]	40 kDa	0	7
79	Cluster of PREDICTED: 40S ribosomal protein S8-like, partial n=1 Tax=Mus musculus RepID=UPI0005ABA108 (UPI0005ABA108)	UPI0005ABA108 [5]	33 kDa	0	4
80	Legumain n=1 Tax=Bos mutus RepID=L8HX98_9CETA	L8HX98 (+1)	49 kDa	3	2
81	Cluster of Heterogeneous nuclear ribonucleoprotein F n=1 Tax=Fukomys damarensis RepID=A0A091DL21_FUKDA (A0A091DL21)	A0A091DL21 [10]	41 kDa	2	4
82	Cluster of 60S ribosomal protein L7a n=1 Tax=Cricetulus griseus RepID=G3ILW8_CRIGR (G3ILW8)	G3ILW8 [6]	27 kDa	0	4
83	Uncharacterized protein n=1 Tax=Ictidomys tridecemlineatus RepID=I3NFU5_ICTTR	I3NFU5	26 kDa	0	0
84	Cluster of Dipeptidyl peptidase 4 n=1 Tax=Bos taurus RepID=DPP4_BOVIN (P81425)	P81425 [3]	88 kDa	4	0
85	Cluster of PREDICTED: 60S ribosomal protein L4 n=1 Tax=Chinchilla lanigera RepID=UPI00038EEA1F (UPI00038EEA1F)	UPI00038EEA1F [7]	47 kDa	0	6
86	Cluster of Protein LOC100359960 n=1 Tax=Rattus norvegicus RepID=D3ZIY7_RAT (D3ZIY7)	D3ZIY7 [13]	30 kDa	4	5
87	Cluster of PREDICTED: 40S ribosomal protein S18 n=1 Tax=Mesocricetus auratus RepID=UPI0003595FC1 (UPI0003595FC1)	UPI0003595FC1 [4]	18 kDa	2	5
88	Cluster of PREDICTED: LOW QUALITY PROTEIN: interferon-induced very large GTPase 1-like n=1 Tax=Bos taurus RepID=UPI000572E9C6 (UPI000572E9C6)	UPI000572E9C6 [2]	280 kDa	0	0
89	Cluster of Putative uncharacterized protein n=1 Tax=Mus musculus RepID=Q3UK56_MOUSE (Q3UK56)	Q3UK56 [3]	27 kDa	4	5
90	Cluster of Calmodulin n=1 Tax=Heterocephalus glaber RepID=G5C0H6_HETGA (G5C0H6)	G5C0H6 [11]	10 kDa	4	3
91	Cluster of PREDICTED: 60S ribosomal protein L18 n=1 Tax=Mesocricetus auratus RepID=UPI0003594AE9 (UPI0003594AE9)	UPI0003594AE9 [5]	22 kDa	0	4
92	Protease do n=2 Tax=root RepID=H1BZ94_ECOLX	H1BZ94	51 kDa	5	4
93	Cluster of PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L12 isoform X2 n=1 Tax=Cricetulus griseus RepID=UPI0004543BFD (UPI0004543BFD)	UPI0004543BFD [4]	25 kDa	2	2
94	Cluster of Zinc finger protein 34 n=1 Tax=Fukomys damarensis RepID=A0A091DJB9_FUKDA (A0A091DJB9)	A0A091DJB9 [2]	86 kDa	0	3
95	Cluster of 60S ribosomal protein L13a n=1 Tax=Rattus norvegicus RepID=RL13A_RAT (P35427)	P35427 [3]	23 kDa	0	4
96	Cluster of DEAD box polypeptide 3 Y-linked long isoform n=3 Tax=Bos taurus RepID=D3IVZ2_BOVIN (D3IVZ2)	D3IVZ2 [15]	73 kDa	2	0
97	Cluster of PREDICTED: uncharacterized protein LOC101601625 n=1 Tax=Jaculus jaculus RepID=UPI0003330EB0 (UPI0003330EB0)	UPI0003330EB0 [2]	37 kDa	0	2
98	Cluster of PREDICTED: regulator of G-protein signaling protein-like isoform X1 n=1 Tax=Rattus norvegicus RepID=UPI0004E47DD4 (UPI0004E47DD4)	UPI0004E47DD4 [11]	141 kDa	3	2

99	Cluster of Polyadenylate-binding protein n=1 Tax=Heterocephalus glaber RepID=G5AP00_HETGA (G5AP00)	G5AP00 [6]	79 kDa	0	4
100	Cluster of Trypsin-1 n=1 Tax=Homo sapiens RepID=A0A087WW55_HUMAN (A0A087WW55)	A0A087WW55 [23]	27 kDa	3	0
101	Cluster of 60S ribosomal protein L7-like protein n=2 Tax=Cricetulus griseus RepID=A0A061IJF5_CRIGR (A0A061IJF5)	A0A061IJF5	29 kDa	0	0
102	Cluster of Heterogeneous nuclear ribonucleoprotein D-like n=1 Tax=Rattus norvegicus RepID=A0A0G2KAZ7_RAT (A0A0G2KAZ7)	A0A0G2KAZ7 [3]	46 kDa	5	2
103	Cluster of Plasminogen activator inhibitor 1 RNA-binding protein n=1 Tax=Cricetulus griseus RepID=A0A061HWU0_CRIGR (A0A061HWU0)	A0A061HWU0 [7]	38 kDa	0	2
104	Cluster of 60S ribosomal protein L23a n=1 Tax=Heterocephalus glaber RepID=G5AUJ4_HETGA (G5AUJ4)	G5AUJ4 [5]	18 kDa	0	3
105	Cluster of 40S ribosomal protein S14 n=1 Tax=Fukomys damarensis RepID=A0A091DLN1_FUKDA (A0A091DLN1)	A0A091DLN1	24 kDa	2	2
106	Cluster of Serpin A3-3 (Fragment) n=2 Tax=Bos mutus RepID=L8HMQ7_9CETA (L8HMQ7)	L8HMQ7 [5]	46 kDa	6	0
107	Cluster of Zinc finger and BTB domain-containing protein 11 n=1 Tax=Fukomys damarensis RepID=A0A091ELX9_FUKDA (A0A091ELX9)	A0A091ELX9 [2]	113 kDa	3	0
108	Uncharacterized protein KIAA1614 n=2 Tax=Homo sapiens RepID=K1614_HUMAN	Q5VZ46	127 kDa	0	0
109	Cluster of Reticulocalbin-2 n=1 Tax=Cricetulus griseus RepID=G3H0D2_CRIGR (G3H0D2)	G3H0D2 [5]	37 kDa	2	5
110	Cluster of Protein RGD1563705 n=1 Tax=Rattus norvegicus RepID=F1LTW8_RAT (F1LTW8)	F1LTW8 [2]	16 kDa	0	4
111	Cluster of Uncharacterized protein n=6 Tax=root RepID=O11993_BVDV (O11993)	O11993 [4]	559 kDa	0	2
112	Cluster of Protein LOC100911337 n=1 Tax=Rattus norvegicus RepID=M0R763_RAT (M0R763)	M0R763 [2]	14 kDa	0	2
113	Cluster of PREDICTED: methylcrotonoyl-CoA carboxylase beta chain, mitochondrial isoform X3 n=1 Tax=Fukomys damarensis RepID=UPI00053F5AA5 (UPI00053F5AA5)	UPI00053F5AA5 [5]	56 kDa	2	0
114	Cluster of Glutathione S-transferase P n=1 Tax=Mesocricetus auratus RepID=GSTP1_MESAU (Q60550)	Q60550	23 kDa	0	3
115	Cluster of PREDICTED: 60S ribosomal protein L19 n=1 Tax=Bos taurus RepID=UPI00005BF24C (UPI00005BF24C)	UPI00005BF24C	21 kDa	0	2
116	Cluster of PREDICTED: LOW QUALITY PROTEIN: 40S ribosomal protein SA n=1 Tax=Jaculus jaculus RepID=UPI00064CE1F9 (UPI00064CE1F9)	UPI00064CE1F9 [2]	31 kDa	0	4
117	Periplasmic dipeptide transport protein n=122 Tax=root RepID=DPPA_ECOLI	P23847	60 kDa	2	2
118	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC100763537 isoform X2 n=2 Tax=Cricetulus griseus RepID=UPI00045457F2	UPI00045457F2	63 kDa	0	2

119	Cluster of PREDICTED: LOW QUALITY PROTEIN: 40S ribosomal protein S3a-like n=1 Tax=Octodon degus RepID=UPI00062A75D7 (UPI00062A75D7)	UPI00062A75D7 [13]	30 kDa	0	4
120	Cluster of PREDICTED: LOW QUALITY PROTEIN: Y-box-binding protein 3 n=1 Tax=Cavia porcellus RepID=UPI000661F4CF (UPI000661F4CF)	UPI000661F4CF [5]	59 kDa	0	2
121	Cluster of Protein Bco2 n=1 Tax=Rattus norvegicus RepID=F1LMB4_RAT (F1LMB4)	F1LMB4	54 kDa	0	2
122	Cluster of Uncharacterized protein n=1 Tax=Bos mutus RepID=L8HQ10_9CETA (L8HQ10)	L8HQ10 [2]	66 kDa	0	2
123	Cluster of PREDICTED: mid1-interacting protein 1 n=1 Tax=Mesocricetus auratus RepID=UPI000359A091 (UPI000359A091)	UPI000359A091	20 kDa	0	3
124	Cluster of cDNA FLJ44920 fis, clone BRAMY3011501, highly similar to Heterogeneous nuclear ribonucleoprotein U n=1 Tax=Homo sapiens RepID=B3KX72_HUMAN (B3KX72)	B3KX72 [16]	83 kDa	0	2
125	PREDICTED: embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein n=1 Tax=Ictidomys tridecemlineatus RepID=UPI000681E2C8	UPI000681E2C8	55 kDa	4	3
126	Cluster of PREDICTED: elongation factor 1-beta n=1 Tax=Mesocricetus auratus RepID=UPI000359B718 (UPI000359B718)	UPI000359B718	25 kDa	0	2
127	Cluster of Uncharacterized protein n=1 Tax=Cavia porcellus RepID=H0UVA7_CAVPO (H0UVA7)	H0UVA7 [2]	69 kDa	0	3
128	Cluster of Putative interferon-induced protein like protein n=1 Tax=Cricetulus griseus RepID=A0A061I0D8_CRIGR (A0A061I0D8)	A0A061I0D8 [2]	42 kDa	0	2
129	Cluster of Alpha-actinin-2 n=1 Tax=Mus musculus RepID=D3YY95_MOUSE (D3YY95)	D3YY95 [4]	18 kDa	0	2
130	PREDICTED: non-lysosomal glucosylceramidase n=1 Tax=Dipodomys ordii RepID=UPI000650BE82	UPI000650BE82	104 kDa	0	2
131	Cluster of 40S ribosomal protein S19 n=1 Tax=Bos taurus RepID=E1BHA5_BOVIN (E1BHA5)	E1BHA5 [2]	16 kDa	0	2
132	Cluster of PREDICTED: 60S ribosomal protein L14 n=1 Tax=Mesocricetus auratus RepID=UPI000359FD7B (UPI000359FD7B)	UPI000359FD7B [3]	23 kDa	0	2
133	Cluster of Putative uncharacterized protein n=1 Tax=Mus musculus RepID=Q3TJZ1_MOUSE (Q3TJZ1)	Q3TJZ1 [11]	95 kDa	0	2
134	Cluster of Nuclear autoantigenic sperm protein n=1 Tax=Fukomys damarensis RepID=A0A091CRM7_FUKDA (A0A091CRM7)	A0A091CRM7	88 kDa	2	0
135	40S ribosomal protein S17 n=1 Tax=Bos taurus RepID=RS17_BOVIN	A5PK63 (+3)	16 kDa	0	2
136	Cluster of Heat shock 70 kDa protein 6 n=1 Tax=Bos mutus RepID=L8I314_9CETA (L8I314)	L8I314 [2]	71 kDa	0	2
137	Cluster of 40S ribosomal protein S6 n=1 Tax=Ictidomys tridecemlineatus RepID=I3N7K2_ICTTR (I3N7K2)	I3N7K2	28 kDa	0	2
138	Cluster of Ribosomal protein n=1 Tax=Rattus norvegicus RepID=D3Z9F6_RAT (D3Z9F6)	D3Z9F6 [3]	25 kDa	0	2
139	Protein disulfide-isomerase n=1 Tax=Homo sapiens RepID=B3KQT2_HUMAN	B3KQT2	55 kDa	0	2

140	WASH complex subunit FAM21 n=2 Tax=Mus musculus RepID=FAM21_MOUSE	Q6PGL7 (+3)	145 kDa	2	0
141	Cationic trypsin n=1 Tax=Bos taurus RepID=TRY1_BOVIN	P00760	26 kDa	0	2
142	Cluster of Kinesin-like protein n=1 Tax=Cricetulus griseus RepID=G3IAS8_CRIGR (G3IAS8)	G3IAS8 [24]	186 kDa	0	2
143	PREDICTED: 60S ribosomal protein L30-like n=1 Tax=Octodon degus RepID=UPI000333E1E1	UPI000333E1E1	13 kDa	0	2
144	Rpl31 protein n=1 Tax=Mus musculus RepID=Q58DZ1_MOUSE	Q58DZ1	8 kDa	0	2
145	Eukaryotic translation initiation factor 2 subunit 1 n=1 Tax=Cricetulus griseus RepID=G3H1M4_CRIGR	G3H1M4	36 kDa	0	0
146	Cluster of Similar to ribosomal protein L17 (Fragment) n=1 Tax=Bos taurus RepID=Q862J1_BOVIN (Q862J1)	Q862J1 [4]	10 kDa	0	2
147	PREDICTED: LOW QUALITY PROTEIN: vascular non-inflammatory molecule 3-like n=1 Tax=Chinchilla lanigera RepID=UPI00038EA961	UPI00038EA961	57 kDa	2	0
148	Reticulocalbin 2, EF-hand calcium binding domain n=2 Tax=Bos RepID=Q0VCQ9_BOVIN	Q0VCQ9	37 kDa	0	2
149	PREDICTED: LOW QUALITY PROTEIN: epiplakin n=1 Tax=Bos taurus RepID=UPI000383FE76	UPI000383FE76 (+1)	376 kDa	2	3
150	Cluster of 14-3-3 protein zeta/delta n=1 Tax=Fukomys damarensis RepID=A0A091DLH5_FUKDA (A0A091DLH5)	A0A091DLH5 [10]	25 kDa	0	2
151	Actin, cytoplasmic 1 n=1 Tax=Heterocephalus glaber RepID=G5B2F5_HETGA	G5B2F5	17 kDa	0	2
152	Fructose-bisphosphate aldolase n=2 Tax=Cricetulus griseus RepID=A0A061IB69_CRIGR	A0A061IB69 (+20)	45 kDa	0	0
153	Cluster of 40S ribosomal protein S26 n=1 Tax=Cricetulus griseus RepID=G3I583_CRIGR (G3I583)	G3I583 [3]	10 kDa	0	2
154	Cluster of Uncharacterized protein (Fragment) n=1 Tax=Bos mutus RepID=L8IBL5_9CETA (L8IBL5)	L8IBL5 [5]	13 kDa	0	2
155	Cluster of Serum albumin n=4 Tax=Mus RepID=ALBU_MOUSE (P07724)	P07724 [2]	69 kDa	5	4
156	Cluster of PREDICTED: heterogeneous nuclear ribonucleoprotein A0 n=1 Tax=Microtus ochrogaster RepID=UPI00038BCC78 (UPI00038BCC78)	UPI00038BCC78 [4]	30 kDa	2	0
157	Cluster of PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L13 n=1 Tax=Fukomys damarensis RepID=UPI00053FEA43 (UPI00053FEA43)	UPI00053FEA43 [5]	27 kDa	0	3

Table 1 :. Molecular weight, accession number, taxonomy and amount of peptides sequenced in Infected extracts (I) and uninfected extracts(NI) associated with proteins identified by RAP-MS analysis.