MASTER IN BIOCHEMISTRY AND MOLECULAR AND CELL BIOLOGY RESEARCH FOCUS

An Activator MicroRNA

Characterization in Marek's Disease Virus

Chasseur, Alexis

Award date:
2020

Awarding institution:
University of Namur

Faculté des Sciences

AN ACTIVATOR MICRORNA : CHARACTERIZATION IN MAREK'S DISEASE VIRUS

Mémoire présenté pour l'obtention

du grade académique de master 120 en biochimie et biologie moléculaire et cellulaire

Alexis CHASSEUR

Janvier 2020

« Il y a des gens qui peuvent être ailleurs quand ils veulent, ils n'ont pas besoin d'avoir un passeport »

Jacques Prévert

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

Un microARN activateur : caractérisation dans le virus de la maladie de Marek

CHASSEUR Alexis

Résumé

L'expression de la protéine ICP22 de l'herpesvirus GaHV-2 a été décrite comme étant induite par un mécanisme dépendant de la liaison d'un microARN. Les résultats préliminaires décrivent une augmentation de l'expression aussi bien au niveau transcriptionnel que traductionnel. Les mécanismes sous-tendant cette augmentation n'ont pas été élucidés.
L'objectif premier de ce mémoire de master était d'établir des constructions rapportrices de l'induction naturelle par le microARN dans le but de confirmer et de décortiquer les tenants et aboutissants de cette induction naturelle.
Les premiers résultats obtenus ont mis en lumière la présence d'une structure secondaire de l'ARN messager au niveau du site cible du microARN étudié. Cette structure en épingle à cheveux s'est révélée importante aussi bien au niveau transcriptionnel que traductionnel, au même titre que le microARN dont des effets à la fois directs et indirects ont été démontrés.
Bien que non-étudiés à l'heure actuelle, plusieurs modèles mécanistiques ont été proposés sur base des résultats et de la littérature. Ces mécanismes, non-exclusifs, invoquent la présence de la structure de l'ARN messager et du microARN.
Le virus devenant de plus en plus virulent au cours du temps, les vaccins approchant de l'obsolescence, les mécanismes proposés ouvrent la voie à de possibles nouveaux types de vaccins viraux atténués par l'expression d'éventuelles protéines délétères pour le cycle viral.

Mémoire de master 120 en biochimie et biologie moléculaire et cellulaire
Janvier 2020
Promoteur : Benoît Muylkens
Superviseur : Srdan Pejakovic

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

An activator microRNA : characterization in Marek's disease virus

CHASSEUR Alexis

Summary

Gene expression concerning the ICP22 protein, encoded by the *Gallid herpesvirus* type 2, was described as being induced by a mechanism depending on the binding of a microRNA to the messenger RNA. Preliminary results described an expression increase at both the transcriptional and translational levels. Mechanisms that underlie this increase were not elucidated yet.

The first aim of this master thesis was the establishment of constructs reporting the natural miRNA-mediated enhancement in order to confirm and to decipher the ins and outs of this natural enhancement.

The first obtained results shed light at a secondary structure belonging to the messenger RNA, localized at the microRNA binding site. This hairpin structure was revealed to have a prominent role at both the transcriptional and translational levels. The microRNA was also shown to play a substantial role which was both direct and indirect.

Although they were not studied yet, several mechanistic models were suggested based on the results and on literature. These non-exclusive mechanisms involve both the microRNA and the described hairpin.

As the virus becoming increasingly virulent over time, concomitant with the approaching obsolescence of the current vaccines, it is important to find new ways to protect the impacted animals. The suggested mechanisms open a gate to new alternative ways to engineer new vaccine strains, attenuated by the expression of proteins deleterious for the viral life cycle.

Master 120 in Biochemistry and Molecular and Cellular Biology

January 2020

Promoter : Benoît Muylkens

Supervisor : Srdan Pejakovic

Acknowledgement

Je remercie tout d'abord les membres du jury aussi bien pour en faire partie que pour leur accompagnement durant mes études. Madame Catherine Lambert De Rouvroit pour ses précieux conseils techniques lors des cours de bacheliers, Monsieur Thierry Arnould pour la gestion du master et les nombreux cours, Monsieur Marino Caruso qui a probablement 101 expériences en cours mais qui prend le temps de lire ce mémoire et Monsieur Charles Van Der Henst pour son aide depuis que je suis en master, aussi bien au plan des études que pour l'avenir.

Je remercie ensuite ma famille pour leur enthousiasme malgré la complexité de notre domaine. Ils ne comprennent probablement toujours pas ce que j'étudie mais ils restent présents. Je remercie plus particulièrement mon père et ma mère qui se sacrifient tous les jours pour que mes frères et sœurs et moi-même ayons une éducation plus que correcte.

Je remercie enfin mes amis et plus particulièrement Alexandre Braeckman, Cédric Vandenbulcke, Elias Aajja, François Warichet et Pauline Ponsard. Alexandre et Cédric m'ont permis d'énormément décompresser tout au long de mes études, je pense que je m'en serais bien mal sorti sans ces conversations de soirées qui n'avaient ni queue ni tête. Elias est probablement la meilleure et la plus importante rencontre que j'aie eue en « biomed », tu es vif d'esprit, drôle et atypique. Grâce à toi, les sciences étaient aussi drôles qu'époustouflantes. François Warichet est un sacré bonhomme géologue que j'ai rencontré en secondaire. Déjà à ce moment, on me disait que tu avais eu une bonne influence sur ma vie, tu as continué pendant le bachelier. On s'est un peu perdus de vue maintenant mais je ne serais pas là sans toi et je m'en souviendrai. Pauline, tu m'as hébergé, tu as affronté le master avec moi avec des bio qui faisaient peur au début, tu m'as fait rigoler, tu m'as permis de m'exprimer quand j'avais le cœur lourd pendant ce mémoire, merci pour tout, vraiment.

Je remercie mes collègues de bureau (dans l'ordre alphabétique pour ne pas faire de jaloux) Alexandra Decloux, Céline Istasse, Gabrielle Trozzi, Olivier Poncelet et également les autres qui sont arrivés ou partis en cours de route. Alexandra, tu es toujours notre maman, tu nous corriges quand on dit des bêtises, tu nous offres des mandarines,… Mais tu es aussi notre princesse un peu naïve qui nous fait rigoler… Tu seras toujours Guenièvre pour moi. Céline, tu as été trop cool dès le début, tu m'as permis d'avoir quelqu'un avec qui parler des circ quand je ne comprenais pas, tu ragotais avec Gabi et moi. Gabz', reste comme tu es. Comme je te l'ai déjà dit, tu es la personne la plus humaine du bureau, ça ramène de l'humain dans un monde de science ou tout peut être perçu comme froid. Olivier, on nous a un jour appelés les « petits rigolos » et je trouve ça beau car cela résume tout.

Je remercie Madison Warnier qui a été en première ligne de toutes mes crises de nerfs depuis le début de ce mémoire, qui m'a accompagné en vacances au moment où j'en avais le plus besoin car rien ne marchait, qui a supporté toutes mes histoires de labo et qui apportait des solutions bien pensées avec un avis extérieur, qui se pliait en quatre pour qu'on se voit au soir quand j'avais eu des grosses journées,… Tu en as fait beaucoup trop, je n'en méritais sans doute pas tant. Merci pour tout.

Je remercie enfin tous les membres de l'URVI. Laeti car c'était notre maman canard, Astrid pour ta gentillesse, la team APOBEC pour parler de science sans parler de poulet, Carole car elle partageait notre bureau sans se plaindre alors que nous n'étions pas des colocataires évidents à supporter, Chantal pour les biscuits sur la table et le café, Jean-Michel Vandeweerd pour ses petites blagues toujours bienvenues, Claire Diederich et Nathalie Kirschvink parce

qu'on est avant tout dans un labo vétérinaire et qu'il fallait quelqu'un pour nous le rappeler et Nicolas Burton pour ta conversation toujours aisée quelle que soit le sujet.

Damien, tu es probablement la plus grande raison de réussite des manips du labo, quelles qu'elles soient. Tu es ouvert, pédagogue, drôle et tu étais toujours le premier pour tout, que ce soit pour partager une bière ou pour partir dans des délires profonds, presque toujours scientifiques. Tu es passionné de biologie et tu transmets ça aisément, j'ai toujours aimé discuter du sujet avec toi, peu importe le sujet.

Srđan, I thank you a lot, you have your own paragraph because you deserve it. You were the first to welcome me in the lab, you taught me every experiment that I know now and your teaching was exactly what I expected. You were always at my back without being oppressive and when everything failed, you were there supporting me, until the end even if I lost my English every time I was a bit tired. When there is a problem in the lab, you are the first to try to solve it and when we need help, you are there too. At the beginning, you said me "stop thanking" but the truth is that I'm really thankful for everything you did for me.

Benoit, Boris Vian a dit un jour *« Ce qui m'intéresse, ce n'est pas le bonheur de tous les hommes mais celui de chacun »*. Je trouve que cette phrase te représente parfaitement. Si le labo tourne bien aujourd'hui, c'est grâce à toi, tu fais toujours attention à ce que tout le monde se porte bien individuellement, et c'est une immense qualité pour un patron. Je ne sais toujours pas comment tu fais pour assumer autant de responsabilités à la fois mais tu as sans doute un don pour ça. En ce qui me concerne, j'ai toujours essayé de te prendre le moins de temps possible mais tu m'en as donné bien plus que ça. Je te remercie finalement de m'avoir encouragé dans mes convictions, tu m'as encouragé à divaguer pendant un an à la recherche de chaque détail qui rendait le projet aussi attrayant. J'espère que c'est vraiment ça, la Science.

Merci pour tout, à tous.

Abbreviations

aa : amino acids

Ago : Argonaute

APC : Antigen-Presenting Cells

ARE : AU-Rich Element

CEF : Chicken Embryonic Fibroblasts

chTR : Chicken Telomerase RNA subunit

CIP : Calf Intestinal alkaline Phosphatase

circRNA : Circular RNA

CREB : cAMP Responsive Element Binding protein

CS : Chicken Serum

DGCR8 : DiGeorge syndrome Critical Region 8

DMEM : Dubelcco's Modified Eagle Medium

dORF : Downstream Open Reading Frame

Dpi : Days post-infection

dsRNA : Double-Stranded RNA

EBV : Epstein-Barr Virus

EF-Tu : Elongation Factor Tu

eIF : Eucaryotic Initiation Factor

eRF : Eucaryotic Release Factor

ESCDL-1 : Embryonic Stem Cell Derived Line 1

FBS : Fetal Bovine Serum

FSC : Forward Scatter

GaHV-2 : Gallid Herpesvirus 2

GFP : Green Fluorescent Protein

HCV : Hepatitis C Virus

HEPES : 4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid

HHV : Human Herpesvirus

HSV : Herpes Simplex Virus

ICP : Infected-Cell Protein

IE : Immediate Early

IgR : Intergenic Region

IRES : Internal Ribosome Entry Site

IR_L : Internal Repeat Long

IR_S : Internal Repeat Short

KSHV : Kaposi Sarcoma-associated Herpesvirus

lncRNA : Long Non-Coding RNA

LSU : Large ribosomal Subunit

MD : Marek's Disease

MDV : Marek's Disease Virus

Meq : MDV EcoRI Q fragment

MERE : Meq Responsive Element

miR-4 : MDV-miR-M4

miRE : miRNA Response Element

miRNA : Micro RNA

MN: *Macherey-Nagel* ©

mRNA : messenger RNA

NEAA : Non-Essential Amino Acids

NEB : *New-England Biolabs* ©

NLS : Nuclear Localization Signal

NGD : No-Go Decay

NMD : Nonsense-Mediated mRNA Decay

NSD : Non-Stop Decay

NXF1 : Nuclear Export Factor 1

ORF : Open Reading Frame

PABP : Poly-A-Binding Protein

PAN : Poly-A Nuclease

PAS : Polyadenylation Signal

PCR : Polymerase Chain Reaction

PS : Penicillin Streptomycin

RaQC : Ribosome-associated Quality Control

RISC : RNA-Induced Silencing Complex

RNAPII : RNA polymerase type II

RPMI : Roswell Park Memorial Institute

rRNA : Ribosomal RNA

RT : Room Temperature

saRNA : small activator RNA

sfRNA : Subgenomic Flavivirus RNA

SSC : Side Scatter

SSU : Small ribosomal Subunit

TERT : Telomerase Reverse Transcriptase

TMR : Telomeric Repeats

TNRC6 : Trinucleotide Repeat-Containing gene 6 protein

TR_L : Terminal Repeat Long

tRNA : Transfer RNA

TR_S : Terminal Repeat Short

TSN : Tudor-Staphylococcal/micrococcal-like Nuclease

TTP : Tristetraprolin

TumiD : Tudor-mediated miRNA Decay

U_L : Unique Long

uORF : Upstream Open Reading Frame

Upf : Up Factor

U_S : Unique Short

UTR : Untranslated Region

vIL-8 : viral Interleukin 8

vTR : viral Telomerase RNA subunit

Table of contents

1) Introduction 1
1.1) RNA as the basis of life 1
1.1.1) RNA secondary structures 1
1.1.2) Translational regulation 3
1.1.2.1) Canonical translation 3
1.1.2.2) Nonsense-mediated mRNA decay 5
1.1.2.3) Non-stop decay 5
1.1.2.4) No-Go Decay 7
1.2) MicroRNA's : Functions and regulations 9
1.3) Marek's Disease Virus 13
1.3.1) Background information 13
1.3.2) Viral cycle 13
1.3.3) Major viral genes 15
1.3.3.1) Meq 15
1.3.3.2) vTR 15
1.3.3.3) pp38 and pp24 17
1.3.3.4) vIL-8 17
1.3.3.5) LATs 17
1.3.3.6) miRNAs 19
1.3.3.7) MDV-miR-M4 19
1.4) ICP's 21
1.4.1) ICP4 21
1.4.2) ICP27 21
1.4.3) ICP0 23
1.4.4) ICP8 23
1.4.5) ICP22 25
1.5) Objectives 27
2) Material and methods 29
2.1) Viral and bacterial strains 29
2.2) Bioinformatic analysis 29
2.3) Site-directed mutagenesis 29
2.4) Digestion of the inserts 29
2.5) Vector preparation 31
2.6) DNA purifications 31

2.7) Cloning and screening 31
2.8) Mini/Midi Prep 33
2.9) Cell lines 33
2.9.1) ESCDL-1 cell line 33
2.9.2) MSB-1 cell line 35
2.9.3) DF-1 cell line 35
2.9.4) CEF 35
2.10) Transfection 35
2.11) Flow Cytometry 37
2.12) RNA extraction 37
2.13) DNase treatment 37
2.14) Reverse transcription 37
2.15) qPCR 37
2.16) Dactinomycin treatment 37
2.17) Protein extraction 39
2.18) Western blot 39
3) Results 41
3.1) Bioinformatic analysis of the ICP22-US10 bicistronic transcript reveals a strong structural hairpin located at the stop codon of the ICP22 ORF 41
3.2) Mutation of an internal poly-A signal within the bicistronic ICP22-US10 mRNA 43
3.3) Establishment of reporter constructs to decipher the mechanisms regulating the translation of the ICP22-US10 bicistronic transcript 45
3.4) Quantification of the transcript shows both direct and indirect miR4 effects on the amount of mRNA 47
3.5) MDV-miR-M4-3p might bind to the bicistronic transcript encoding for both ICP22 and US10 49
3.6) Protein expression is determined by several parameters : the sequence, the ORF position and the presence of miR4 51
3.7) mRNA half-life show a protective effect of miR4 on the bicistronic independent of the binding of miR4 to the miRE 55
4) Discussion 57
4.1) Suggested models to explain miRNA-dependent increase of the gene expression 59
4.2) miRNA dependent increase of the gene expression involves interaction with secondary RNA structure 63
4.3) Optimization process to increase significance of assays that were carried out during the master thesis 65
4.4) Conclusion and perspectives 69
5) References 71

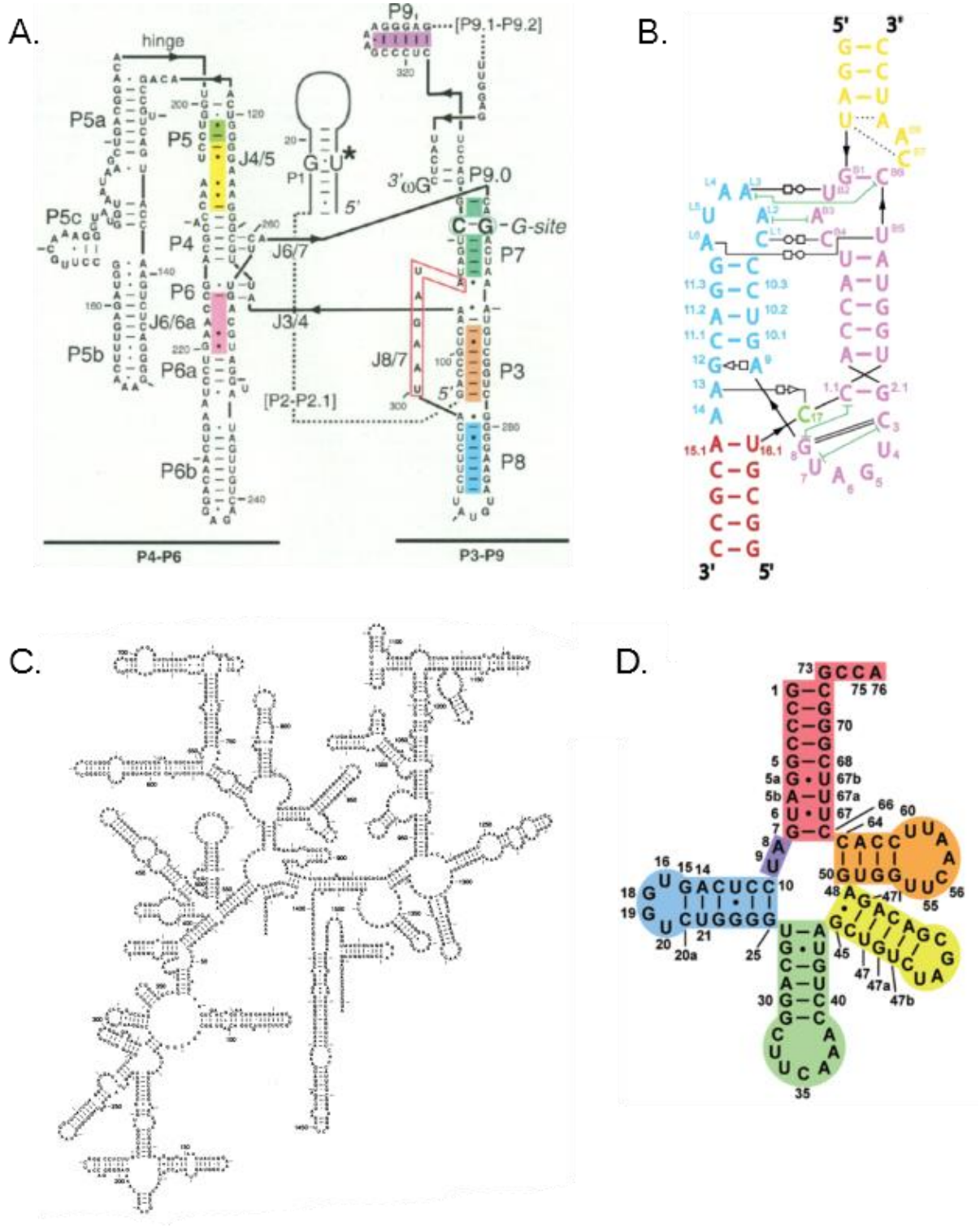


Figure 1 : **Examples RNA structures**. **A)** First ribozymes discovered in *Tetrahymena thermophila* in 1982. The paired helical elements are designated by “P" (from P1 to P9) and joining regions by "J". Some helical elements are not detailed but represented by dashed lines (Golden *et al*, 1998). **B)** Secondary structure of a hammerhead ribozyme and its boundaries with itself. The thick black lines represents the continuity of the backbone (Swadling *et al*, 2019). **C)** Structure of the 16S rRNA of an *E.coli* bacteria (Woese *et al*,1990). **D)** Secondary structure of the human selenocysteine-tRNA. The red color represents the acceptor arm, where the amino acid is bound. The orange color represents the T arm that binds to EF-Tu. The yellow color represents the extra arm, that allows selenocysteine to bind to the tRNA. The green color represents the anticodon arm. The blue color represents the D arm, that binds to EF-P (Itoh *et al*, 2013).

1) Introduction

1.1) RNA as the basis of life

The gene expression is still described as a linear pathway, known as central dogma. This dogma assumes that DNA is transcribed into a messenger RNA (mRNA) which is translated into a protein. This stereotypic view of the gene expression has successively been disrupted by discoveries such as ribozymes, retroviruses and post-transcriptional or post-translational modifications (Hartenian and Glaunsinger, 2019).

A major hypothesis about the origin of life suggests that the very first organic forms of life were RNA-based and protein-free. At the early time of life, RNA, structured into ribozymes, was able to self-replicate and to produce new RNA forms, such as aptamers. One major observation that supports this hypothesis is the ubiquity of the RNA in a lot of different shapes for basal functions of a cell. From small nucleolar RNAs to transfer RNAs and spliceosomes, RNA is always needed in order to produce proteins. However, proteins are not always necessary to synthetize RNA. And as a proof of concept, RNA viruses show that a DNA template is not needed to generate a simple life form (Mustafin and Khusnutdinova, 2019).

1.1.1) RNA secondary structures

RNA structures have many roles in a lot of pathways. Despite this fact, it is still difficult to understand and to know precisely the whole diversity of shapes that RNA can have. However, what is known today, thanks to bioinformatics, molecular biology and chemistry, allows making boundaries in the structure/function relationship that makes RNA the central component of the life (Swadling *et al.,* 2018; Weeks, 2010). The following paragraphs tend to show, in a non-exhaustive manner, the value of establishing secondary structures in RNA molecules. It is important to note that tertiary structures are common among RNAs but skipped in this paragraph.

3.9 billion years ago, in a world that was protein-free, emerged some molecules able to self-replicate. Ribozymes are probable candidates for being these molecules (Figure 1.A.). At this time, they synthetized themselves using minerals as chaperones to resist under extreme conditions, such as the extreme temperature variations. One of the first ribozyme was thus probably a RNA-dependent RNA polymerase (RNAP), beginning the RNA-only pathway that lead to life (Mustafin and Khusnutdinova, 2019; Swadling *et al.,* 2018).

Next to this, when RNA became abundant and diverse thanks to mutations caused by UV-light and poor fidelity of the first RNAP, hammerhead ribozymes (Figure 1.B.) may have appeared, increasing thereby the diversity in the RNA world. This ribozyme has a nucleolytic function either on itself or on other RNAs due to its conformation. It encompasses two loops which, when modified, gave its role to the molecule. More precisely, the mutations in the loops change the affinity of the hammerhead for itself or for other RNAs (De la Pena, 2003; Swadling *et al.,* 2018).

When amino acids have been present in the primitive surrounding environment (Miller, 1953), ribosomal RNA (rRNA) have emerged as highly structured RNA constitutive of the primitive ribosome, able to support the first protein synthesis. Even in the human cells, the ribosomal catalytic core is still only composed of RNAs, highly conserved in all the species (Swadling *et al.,* 2018). As an example, the 16S RNA of *Escherichia coli* is made of a large series of stem-loops (Figure 1.C) (Woese *et al.,* 1990).

To produce proteins in the proper way with the proper amino acids, adapter RNAs were generated to make a link between the RNA template and the amino acids. These adapters are the transfer RNAs (tRNA) and have one of the most known structure-function relationships

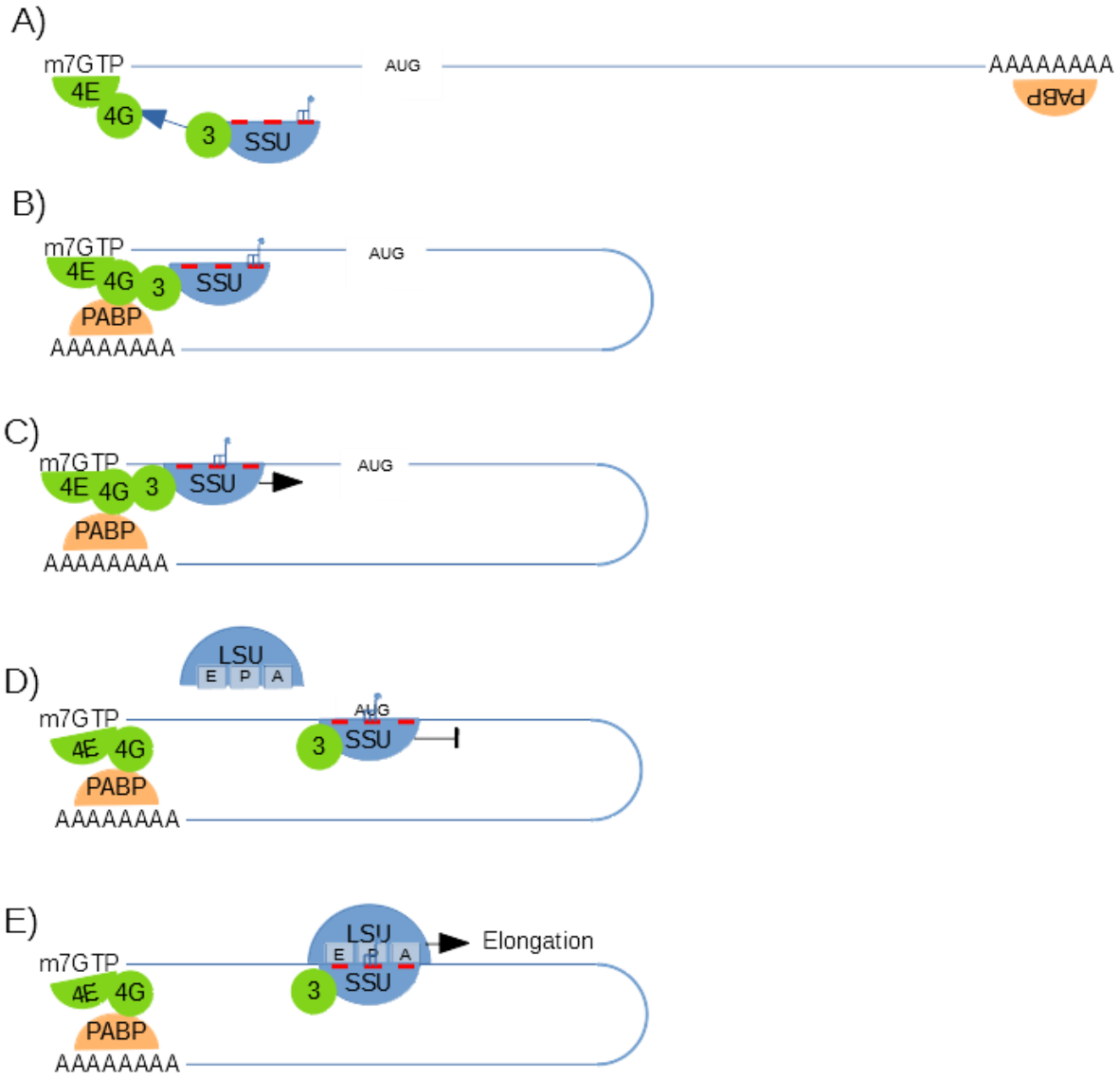


Figure 2 : **Summary of the initiation of the eukaryotic translation**. **A)** The mRNA displays both a 5'-m7-GTP cap and a 3'-poly A tail. The m7GTP-cap recruits eIF4E (4E) and the poly A tail recruits PABP. eIF4G (4G) is recruited on the m7GTP-cap by 4E. **B)** 4G recruits both eIF3 (3) bound to the SSU of the ribosome (with a tRNA on the A site) and PABP which makes the mRNA take a circular shape. **C)** The A site of the ribosomal SSU is left free by a methionylated tRNA that moves to the P site. This changes the ribosomal conformation, which allow it to go through the scanning of the mRNA. **D** and **E)** Thanks to the codon-anticodon interaction, the ribosome stalls at the start-codon. This stalling allows recruiting the LSU of the ribosome, which permits the synthesis of the protein (created from Weisser and Ban, 2019).

among RNAs. They are made of 3 (or 4 in some cases) hairpins and one arm of double stranded RNA (dsRNA). The double strand is the acceptor arm, and it makes the boundary with the amino acid at the 3' end. One of the hairpin is the anticodon and allows the recognition of the matching codon, according to which amino acid is bound to the first arm. The T and D loop are respectively bound with elongation factors (EF) Tu and P. Sometimes, an extra loop can exist and is used to help the ligation of the amino acid on the first arm (Figure 1.D.) (Itoh *et al.,* 2013; Katoh *et al.,* 2017; Ohtsuki and Watanabe, 2007).

Finally, over time, the RNA regulation became more and more complex and the functions diversified. RNA structures acquired the competency to regulate genic expression either in *cis* or in *trans,* but also to be used as a platform to recruit proteins (Swadling *et al.,* 2018). Viruses do not make an exception, they also use several RNA structures to complete their life cycle. In this paragraph, it was chosen to only focus on hairpins. For example, in Hepatitis C Virus (HCV), hairpins harbor major roles. First, the 5'end structures are used as riboswitches to monitor the replication rate. In this case, the structure can take two alternative conformations. This property is used by the virus to turn on (or not) its replication (Boerneke and Hermann, 2015). The second role of the structure is to be used as an internal ribosome entry site (IRES) to translate the viral proteins. The IRES are a common feature of the mRNA to launch translation in a cap-independent manner (Svoboda and Cara, 2006). Another role is involved during replication, where the RNA-dependent RNAP of HCV uses one strand of the hairpin as a primer to synthetize the other strand, in a mechanism named "copy-back replication" (Behrens *et al.,* 1996).

1.1.2) Translational regulation

Nowadays, one of the major role of the RNAs in the eukaryotic domain of life is the regulation of gene expression. Notably, the translation of a mRNA results usually in the expression of a protein. These proteins take a major place in the life cycle of a cell. Thus, the protein homeostasis is a critical scaffold that maintain the cell homeostasis (Hartenian and Glaunsinger, 2019).

During gene expression, several mechanisms ensure protein homeostasis. These mechanisms prevent the production of defective proteins that might have a deleterious dominant negative effect on the cell or even the organism. The detected mutated proteins are usually submitted to degradation, as well as the mutated transcript (Hildebrandt *et al.,* 2019; Simms *et al.,* 2017). Three major ways to decay aberrant mRNA are commonly described : Nonsense-mediated mRNA decay (NMD), Non-stop decay (NSD) and No-go decay (NGD) (Simms *et al.,* 2017).

1.1.2.1) Canonical translation

Translation initiation

The translation begins with the initiation step (Figure 2). The small ribosomal subunit (SSU), with a methionylated transfer RNA (tRNA), is recruited on the mRNA, at its 5'end, thanks to the 5' cap, recognized by initiation factors. The methionylated tRNA takes place first at the A site of the ribosomal subunit. During translation, the tRNA can take three places in the ribosome: A, P and E. The precise roles of these places can be summarized as follow: the tRNA arrives at the A site (A=aminoacyl), it elongates the nascent peptide at the P site (P=peptidyl) and it leaves the ribosome through the E site (E=exit). The binding of the mRNA to the SSU makes it take a circularized conformation, mediated by the interaction that gathers m_7GTP – eIF4E – eIF4G – PABP (eIF : eukaryotic initiation factor ; PABP : Poly-A binding protein). Then, the tRNA leaves the A site of the ribosome free and move to the P site, this configuration makes the SSU able to scan the mRNA until it reaches the start codon (AUG). There, it changes its

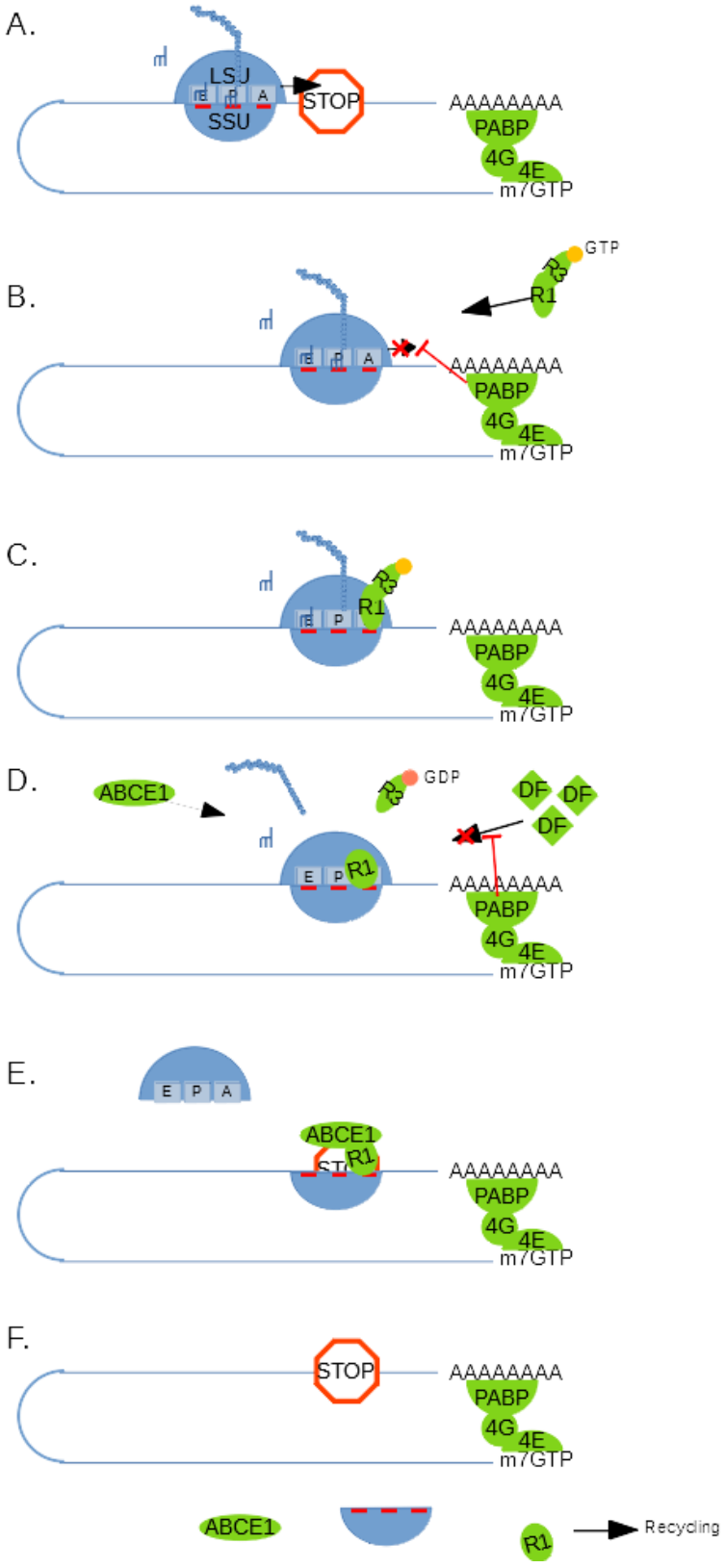


Figure 3: **Canonical termination of the translation. A** and **B)** At the stop codon, the ribosome pauses (mediated by its interaction with PABP) and recruits eRF1 (R1) bound to eRF3-GTP (R3-GTP). **C)** The hydrolysis of the GTP on eRF3 activates eRF1 which changes its conformation and **D)** releases the peptide. The proximity of PABP with the translation complex inhibits the binding of decay factors (DF), such as Upf-1. **E)** Next to the dissociation of eRF3, eRF1 recruits ABCE1 which allows recycling the translation machinery. **F)** eRF1 and ABCE1 are released from the mRNA when termination is completed.

conformation and the large ribosomal subunit (LSU) is able to bind to the SSU to form a proper ribosome, prone to translate the mRNA (Weisser and Ban, 2019).

Translation Termination

When reaching the stop codon (UAA, UAG or UGA) (Figure 3), the ribosome pauses and recruits eRF1 (eukaryotic release factor 1) to its A site. eRF 1 is bound to the eRF3 GTPase. The hydrolysis of a phosphate by eRF3 changes the conformation of eRF1 which activates the peptidyl-transferase activity of the ribosome and releases the peptide. The dissociation of eRF3 allows eRF1 recruiting ABCE1 which permits to detach the LSU and recycle the termination complex (Kervestin and Jacobson, 2012).

The GTPase activity of eRF3 is mediated by PABP, giving a role to PABP during the whole translation mechanism. Actually, the absence of PABP alters both initiation and termination, increasing the readthrough after nonsense codons (Kervestin and Jacobson, 2012).

Sometimes, in bicistronic transcripts, the SSU might remain bound to the mRNA and launch a new initiation event allowing the translation of the downstream ORF (dORF). This dORF is often less translated than the upstream one (uORF). The reinitiation is dependent on features such as the mRNA structures, the context surrounding both the uORF stop codon and the dORF start codon and the intergenic region (IgR) length (Wethmar, 2014).

1.1.2.2) Nonsense-mediated mRNA decay

When a premature termination codon (PTC) is scanned by the ribosome in an inappropriate context, the NMD pathway degrades the transcript. Inappropriate context can be a long 3' untranslated region (UTR) that does not permit the interaction between PABP and the ribosome-associated proteins, as well as the remaining presence of exon junction complexes that recruit Up factors or mRNA that are subject to frameshifting. In all these case, eRF3 does not leave the ribosome and Upf1 (Up factor 1) is recruited. The latter is phosphorylated right after being bound to Upf2 and 3. The active form of Upf1 has an ATPase and a helicase activities. It leads to the peptide degradation and to the recycling of the components of the ribosome (Celik *et al.*, 2015; Porrua and Libri, 2013, 2015).

1.1.2.3) Non-stop decay

Due to splicing or mutations, some mRNA are lacking a proper stop codon in the proper reading frame. This results in two types of transcripts: the nonstop transcripts and the stop codon-less transcripts. The nonstop transcripts encompass a poly-A tail since they come from a premature polyadenylation of the transcript. The second type do not encompass their poly-A tail, probably because of an endonucleolytic cleavage of the mRNA. Both of them are regulated by the NSD. This mechanism is poorly understood but when the mRNA carries its poly-A tail, it is assumed that the translation of the successive adenines, producing successive lysines, creates a strong positive charge around the exit tunnel of the ribosome, negatively charged. Thus, there is an interaction between the peptide and the ribosome, which stalls the ribosome on the poly-A. This stalling makes the ribosome empty on its A site for a long time. Thus, it recruits a complex made of Pelota and HBS1 which are parologs of eRF1 and eRF3 but they lack the sequence that recognize the stop codon and the sequence that allow the peptide release. It launches the release of the LSU but with the peptide still bound to it. This is recognized by a ribosome-associated quality control (RaQC) and the peptide is degraded. Concerning the mRNA, the recognition by the NSD complex triggers its degradation in a Ski7-mediated way (Simms *et al.*, 2017; Szádeczky-Kardoss *et al.*, 2018).

1.1.2.4) No-Go Decay

NGD occurs when the ribosome is stopped during elongation. This stop can be due to several features of the mRNA : premature termination codons, poly-basic amino acids stretches, rare codons, chemically modified nucleotides, or some secondary structures in the mRNA, such as hairpins. The stalling leads to ribosome collisions at the stalling site. These collisions trigger the phosphorylation of some ribosomal proteins, which recruits an RNA decay complex (Szádeczky-Kardoss *et al.,* 2018). This complex is made of HBS1 and Dom34 which trigger an endonucleolytic cleavage of the mRNA in the presence of ABCE1. This cleavage results in two mRNA pieces that are degraded by Xrn-1 (Ikeuchi and Toshifumi, 2016).

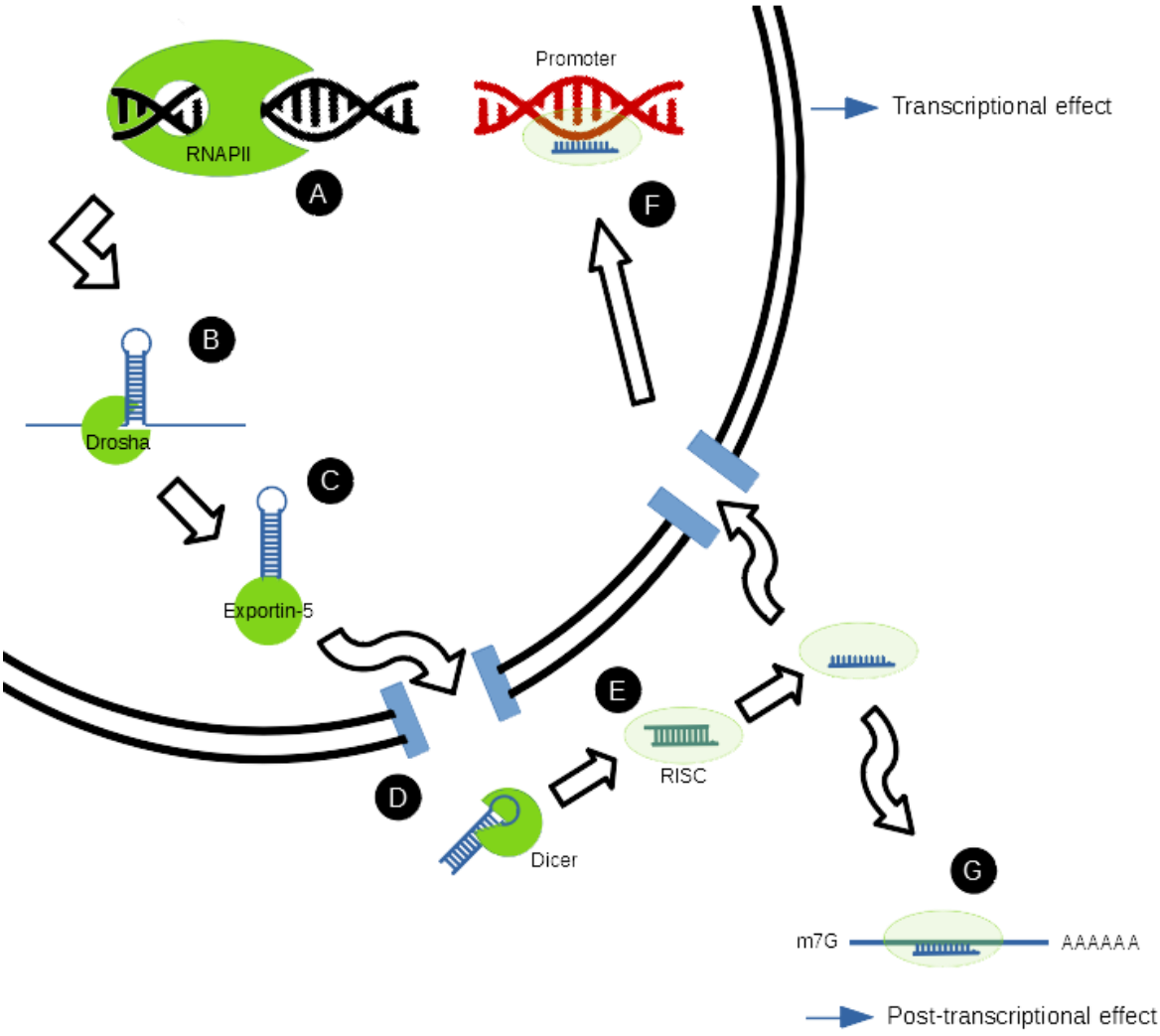


Figure 4 : **Animal miRNA biogenesis and function**. **A)** In the nucleus, miRNA are transcribed from DNA either from genic or intergenic regions by the RNAPII. **B)** A pri-miRNA transcript that contains a hairpin in its secondary structure is processed by Drosha to cleave the linear regions of the transcript. **C)** The hairpin is recruited by the exportin-5 which is used as a cargo to go out the nucleus. **D)** Outside the nucleus, the hairpin is processed by Dicer to cleave the loop of the hairpin. **E)** The double-stranded RNA is recruited by RISC, which select a strand following the stability of the 5' ends of the strands. The miRISC complex exert its function either on the DNA or on the RNA. **F)** Some miRNA, called saRNA, bind directly on the DNA at the promoters of some genes. They can alter the transcription positively or negatively by interacting with transcription factors, methylation or histones. **G)** Most of the known animal miRNA bind to mRNA. A perfect match (rare) cleaves the transcript, meanwhile an imperfect match alters the stability or the translation either positively or negatively.

1.2) MicroRNA's : Functions and regulations

MicroRNAs (miRNA) (Figure 4) are small non-coding RNAs transcribed by metazoans and plants discovered in 1998 (Fire *et al.,* 1998). They take part in RNA interference through a sequence-specific silencing of target genes. They can either destabilize mRNAs or alter their transcription and their translation either positively or negatively. They are mostly transcribed by RNAPII from either intronic or exonic region in the genome or even from intergenic regions. The transcripts (primiR) encompass a highly stable structure, a hairpin, that is processed by Drosha (a type III RNase) to build a pre-miR exported out of the nucleus. In the cytoplasm, Dicer (another type III RNase) cleaves the hairpin to make a double-strand RNA of an average size of 22 nucleotides. One of the strands is incorporated into the RNA-induced silencing complex (RISC) which contains an Argonaute protein (Ago) to become the miRISC complex. Usually, the selected strand is the one that has the less stably paired nucleotides at its 5' end. The non-selected strand is usually released and then degraded (Fischer, 2015; Krol *et al.,* 2010; Michlewski and Cáceres, 2019).

Drosha and Dicer

The reaction catalyzed by Drosha, in the nucleus, involves several proteins, including the double stranded RNA-binding protein DGCR8 (DiGeorge syndrome critical region 8). The reaction produces a ~70 nucleotides hairpin which is the pre-miRNA. Exportin-5 leads the pre-miRNA outside the nucleus to be processed by Dicer (Michlewski and Cáceres, 2019).

Dicer, with its cofactor proteins, processes the pre-miRNA to build the miRNA duplex incorporated into RISC with Ago. The cleavage processed by Dicer is often imprecise and can create several duplex variants that lead to different miRNAs at the end of the processing (Michlewski and Cáceres, 2019).

Seed Sequence

The most functional part of a miRNA is the seed sequence made of the nucleotides 2 to 8 at the 5'end of the miRNA mature sequence. The miRISC activity depends on the length of miRNA-mRNA hybridization. When the match between the mRNA and the miRNA is imperfect, there is a translational repression of the target or a recruitment of the mRNA turnover machinery. In the case of perfect match between the full miRNA sequence with the target mRNA, the cleavage of the mRNA is observed (Ameres and Zamore, 2013; Fischer, 2015; Krol *et al.,* 2010; Michlewski and Cáceres, 2019).

However, the seed is not the only interactional part of the miRNA. Three types of interactions have been described : 5' canonical, 5' dominant and 3' compensatory. In the first case, the seed sequence is paired with the mRNA and completed with several nucleotides paired in 3'. In the second case, only the seed is paired with the mRNA. In the third case, the seed sequence is paired badly but a 3' pairing allows the miRNA to keep its function (Brennecke *et al.,* 2005). A fourth case is also admitted, the so-called "G-bulge interaction" where the seed sequence matches almost perfectly with a sequence except on a G within the mRNA (Ameres and Zamore, 2013).

Mechanism of action

The most known mechanism have a negative impact on the gene expression. When a miRNA sequence does not fully match to its target, there is no cleavage of the target by Ago2. However, there is other ways of silencing a transcript. In animals, this is the case for most of the targeted transcripts. The miRISC complex either represses translation, or promotes deadenylation of the mRNA and leads it to decay (Ameres and Zamore, 2013).

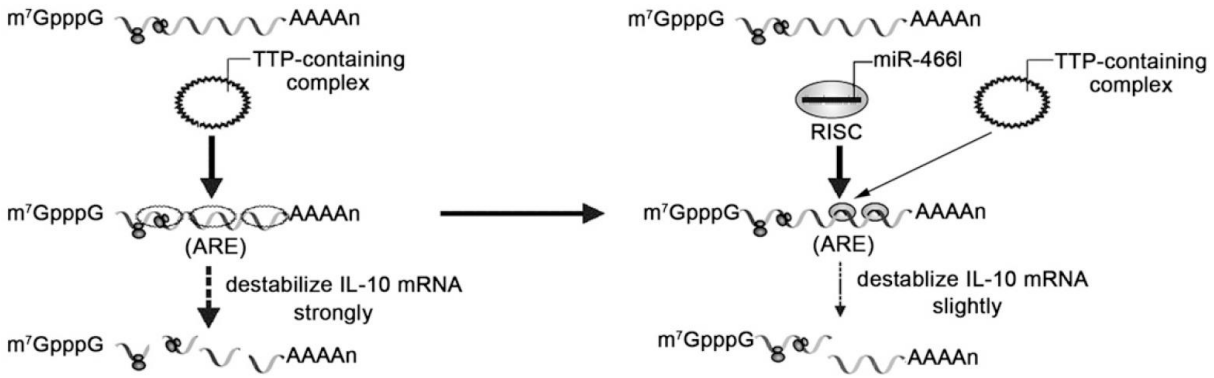


Figure 5 : **miRNA-mediated stabilization of a transcript by inhibition of the binding of a deletorious protein**. In the absence of miR466l, the mRNA encoding for IL-10 is targeted by TTP, which binds to ARE and destabilizes the transcript. In the presence of miR466l, the binding is inhibited. The miRNA-induced destabilization is weaker than the TTP-induced one (Ma *et al*, 2010).

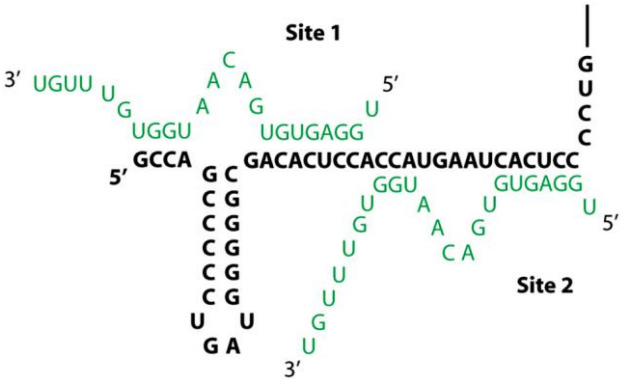


Figure 6 : **miRNA-induced stabilization of a secondary structure in a RNA**. The 5' end of the HCV RNA genome exhibits an hairpin that is stabilized by miR122, expressed in hepatocytes. The binding of miR122 is necessary at two sites to maintain the natural viral life cycle (Machlin *et al*, 2010).

The repression of the translation may occur at the initiation step. In this case the miRISC complex is assumed to interfere with the 5' cap and/or with initiation factors or even with the great ribosomal subunit, impairing its binding with the small one. Repression of the translation may also occur during elongation. In this case, the mechanism could involve a drop of the ribosome from the mRNA, a blocking of the ribosome or the proteolysis of the nascent protein. These three mechanisms are assumed to be specific to some genes and not widespread (Stroynowska-Czerwinska *et al.,* 2014).

The other widespread mechanism is the launching of the deadenylation of the target by the miRISC complex, leading the target to be degraded. This mechanism is more described than the previous ones (Wahle and Winkler, 2013) and relies on TNRC6 (Trinucleotide Repeat-Containing gene 6 protein), a protein of the miRISC complex. TNRC6 establishes first the interaction between miRISC and PABP and next, it recruits the PAN2-PAN3 (Poly-A Nucleases) complex that deadenylates the poly-A tail of the mRNA. TNRC6 is also able to recruit the CCR4-NOT complex through an interaction with the TOB protein to deadenylate the mRNA thanks to another pathway. Next to this deadenylation, the transcript is usually degraded into P bodies which are non-membranous organelles with a high concentration of the mRNA-turnover proteins (Stroynowska-Czerwinska *et al.,* 2014).

Degradation

Like the other types of RNAs, miRNA are submitted to a permanent turnover but the mechanism is still poorly understood. The Tudor-Staphylococcal/micrococcal-like Nuclease (TSN), which harbors an endonuclease activity, endorses the role of degrading miRNA in the Tudor-mediated miRNA decay (TumiD) pathway. To exert this effect, the miRNA is first detached from its target by Upf1 thanks to its helicase activity (Elbarbary *et al.,* 2017a, 2017b).

Upregulation of the gene expression mediated by miRNAs

Several examples of upregulation by miRNAs have been described. It can directly impact the gene expression at three different steps : during transcription, at the post-transcription level or during translation. For instance, a small activator RNA (saRNA) is recruited by an Ago protein and targets a promoter directly within the nucleus (Matsui *et al.,* 2013). The interaction between Ago, the miR, the promoter and the RNAPII regulates transcription either positively or negatively. saRNAs differ from miRNAs only by their function. The precise mechanism is still not known but seems to involve transcription factors, differential methylation of the promoter and remodeling of the chromatin structure (Huang, 2017; Setten *et al.,* 2018).

Another upregulation by a miR has been described by Ma *et al* (Ma *et al.,* 2010) (Figure 5). Tristetraprolin (TTP) is a human RNA-binding protein that recruits mRNA decay machinery when bound to an ARE (AU-rich element). This interaction leads to the decay of IL-10 mRNA in an *in vitro* culture of human macrophages. However, when miR466l binds to the transcript, it competitively antagonizes the binding of TTP and stabilizes the mRNA (Ma *et al.,* 2010).

In hepatitis C virus (HCV), a mechanism has been shown to upregulate viral abundance and translation of viral proteins. The virus, in order to stabilize its 5' terminal nucleotides, recruits two molecules of a liver-specific miRNA, miR-122. These two molecules form an oligomeric structural complex that stabilizes the genome of the virus by inhibiting the action of Xrn-1 on the viral RNA (Figure 6). The interaction between two miRNA and the 5' end of the viral genome also enhances the potency of an IRES, which increases the translation of the viral protein (Machlin *et al.,* 2011).

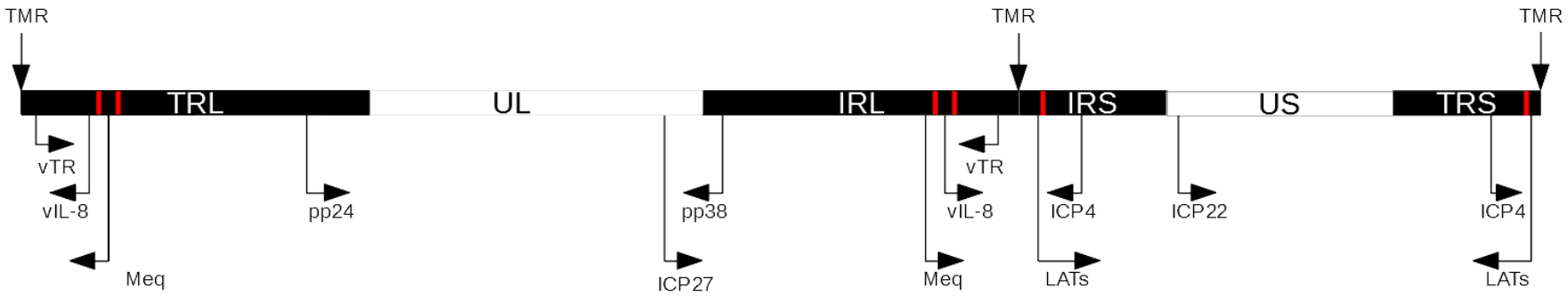


Figure 7 : **Representation of the genome of GaHV-2 with the position and orientation of the major viral genes.** The red lines show the regions coding for miRNAs.GaHV-2 has a peculiar genomic organization, the dsDNA genome encompasses two segments surrounded by inverted repeats. Major viral genes described in the text are represented by the black arrows. Telomeric repeats are also depicted since they are involved in the pathogenecity and in the viral life cycle.

1.3) Marek's Disease Virus

DNA viruses are a good way to study a lot of cellular mechanisms. They are especially a good model to study specific transcriptional mechanisms. They are proven to express miRNAs, mRNAs with alternative transcripts, and they have an easy-to-study genome, big enough to express a large series of transcripts and small enough to easily identify them. Gallid Herpesvirus 2 is a good way to enter herpesviruses biology since chicken is a cheap ethical model and since all the resources to study the virus are available to work both *in vitro* and *in vivo* (Boumart *et al.*, 2018; Muylkens *et al.*, 2010; Rasschaert and Laurent, 2015; Strassheim *et al.*, 2016).

1.3.1) Background information

Gallid Herpesvirus 2 (GaHV-2), also known as Marek's Disease Virus (MDV), is an alphaherpesvirus that is the causative agent of Marek's Disease (MD), first described by József Marek in 1907. The very first characterization of the symptoms of the disease described a polyneuritis syndrome with low mortality. In the 1960's, the post-war industrialization promoted intensive poultry industry. Concomitant with this period, the virus became increasingly virulent until an acute form of MD was discovered to induce visceral tumors, killing up to 30% of the infected fowls. This increasing death rate made GaHV-2 a burden for poultry industry. This lead to the development of attenuated vaccine strains (Figure 8) (Gennart and Muylkens, 2018).

Today, GaHV-2 is used as a relevant model for herpesvirus infections and virus-induced lymphomas. More precisely, the hyper virulent strain of the virus (RB1B), which is used in this master thesis, leads to immunosuppression, neurological paralysis and death in the susceptible chicken in a couple of weeks. This allows to have a fast, ethical and cheap insight into the disease (Boumart, 2018).

GaHV-2 has a particular genomic organization, the double stranded DNA genome is made of two segments each surrounded by inverted repeats (Figure 7). The first part is composed by TR_L (Terminal Repeat Long), U_L (Unique Long) and IR_L (Internal Repeat Long). The second part, which is shorter, is made of IR_S (Internal Repeat Short), U_S (Unique Short) and TR_S (Terminal Repeat Short). Based on this genome organization, GaHV-2 displays the same genomic arrangement as alphaherpesviruses infecting human such as Human Herpesvirus 1 and 2 (HHV-1 and HHV-2) (McPherson and Delany, 2016).

1.3.2) Viral cycle

The viral cycle is divided in four phases that are highly regulated on their gene expression pattern (Figure 9). Two productive phases are separated by the latency from which lymphocytes are transformed due to viral infection. This fourth phase takes place in the latently infected T lymphocytes, which then spread within the organism, forming the virus-induced lymphomas within the peripheral nerves and visceral organs (Davison and Nair, 2004).

The early productive phase starts with the infection of the chickens. The virions are found in the dust present in the environment of the naive chickens. Infected particles are inhaled by the animal and the Antigen-Presenting Cells (APC) become infected by the virions through the lung epithelium. 24 hours later, the virus is present in the primary lymphoid organs, invading B and T lymphocytes thanks to cell-to-cell contacts (Bertzbach *et al.*, 2018; McPherson and Delany, 2016).

Latency is established 7 days post-infection (dpi) mainly within CD4+ T lymphocytes. The genome of GaHV-2 is integrated within host telomeres which ensures the maintenance of the viral genome by the infected cell and it shelters the virus from the immune system while it is disseminating within the organism (Boodhoo *et al.*, 2016; Kheimar *et al.*, 2017). The integration

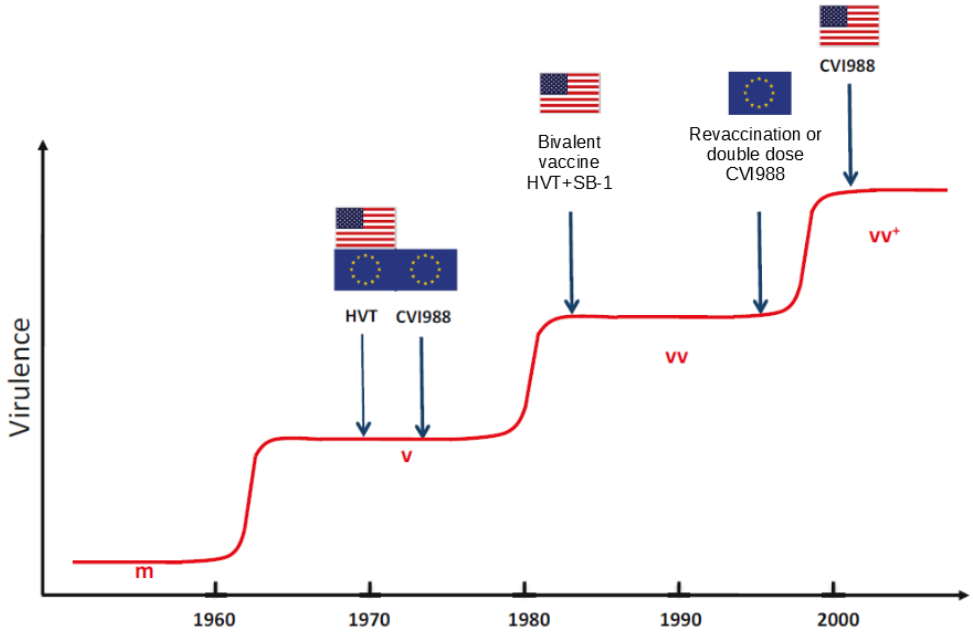


Figure 8 : Evolution of the pathogenecity of GaHV-2 over time, next to vaccination against the disease. Pathotypes are designated as follow : m = mild ; v = virulent ; vv = hypervirulent ; vv+ = hypervirulent +. RB1B, the strain used during this master thesis is a vv+ strain (Dambrine *et al*, 2016).

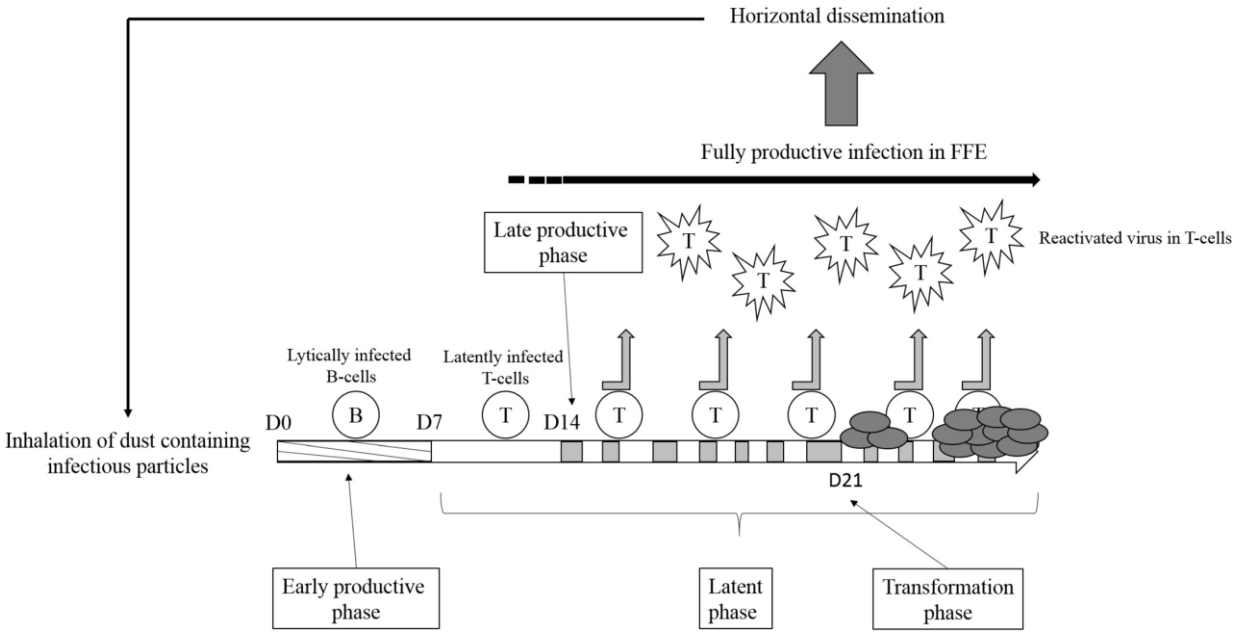


Figure 9 : **Viral life cycle of GaHV-2 during a natural infection**. Day 0 (D0) represent the moment when the chicken became infected by inhalation of dust containing viral particles. The days were counted in an artificial model of infection (natural way is slower). The grey arrows represent reactivation and the dark grey ellipses represent the production of lymphoma from transformed latently infected T cells (Gennart *et al*, 2015).

is needed for the virus to keep efficient transformation and reactivation rate (Kaufer *et al.,* 2011).

Around 14 dpi, the latently infected lymphocytes reactivate, leaving latency and starting the late productive phase. Virus then replicates within the whole organism at high rate, especially within the feather follicles. The shedding of feathers from the infected fowls releases virions in the surrounding environment. It is therefore associated with dust that can be inhaled by the naïve chickens that become infected (Boodhoo *et al.,* 2016; Gennart *et al.,* 2015).

1.3.3) Major viral genes

Viral genome encodes at least 103 identified proteins which are, for most of them, common to the other alphaherpesviruses. More than just being common, some genes are collinear between the alphaherpesviruses. The majority of these genes encode core proteins and are located in the unique regions of the genome. In the subsequent paragraphs, some genes considered as "major", following their involvement in the pathogenicity of the disease or in the viral life cycle, are depicted. These genes are embedded mostly in the repeated regions and are often specific to the pathotype and the gender of the virus (Davison, 2002).

1.3.3.1) Meq

Meq (MDV EcoRI Q fragment) is the major oncogene of the virus and is encoded from the gene *meq* located in the repeated regions surrounding the U_L segment of the genome. This protein is expressed during all the phases of the viral life cycle and was early discovered to have oncogenic properties since it shares properties with some of the cellular proto-oncogenes. Actually, its N-terminal domain is a b-ZIP (basic leucine zipper) similar to the one carried by the Jun/Fos proto-oncoproteins. Concerning its C-terminal domain, it encompasses a proline-rich region, similar with WT-1 which is a tumor suppressor, with a transactivation role (Gennart *et al.,* 2015; Kung *et al.,* 2001; Qian *et al.,* 1995).

Meq mainly acts as a transcription factor integrated to the cellular transcriptional network. Thanks to its b-ZIP domain, it can dimerize with either itself or another b-ZIP-containing protein, such as Jun for which it has a strong affinity. It also has affinity for some other cellular factors, such as cAMP response element binding protein (CREB), p53, pRB,… All these factors are either proto-oncogenes or tumor suppressor (Kung *et al.,* 2001; Qian *et al.,* 1995).

According to the factor bound to Meq, the complex acquires affinity for specific nucleotide sequences. Two major Meq Response Elements (MERE) exist within the genomes of both the virus and the host. MERE-1 is targeted by the Meq/Jun heterodimer meanwhile MERE-2 is targeted by the Meq/Meq homodimer (Kung *et al.,* 2001; Qian *et al.,* 1995).

The binding of the heterodimer Meq/Jun to a MERE-1 site transactivates genes that are related to cellular transformation and viral replication, such as Bcl-2 and ICP4 respectively. The MERE-2 sites are related to a decrease of gene expression. For example, the bidirectional promoter linked to pp38/pp24 and 1,8kb family of genes is repressed by the Meq/Meq homodimer. These genes are late lytic and thus, Meq prevents viral DNA replication and productive infection (Gennart *et al.,* 2015; Qian *et al.,* 1995).

1.3.3.2) vTR

The genome of GaHV-2 exhibits the same telomeric repeats (TMR) than the chicken ($TTAGGG_n$) at its ends. These TMR are dispensable for virus growth *in vitro* but not *in vivo*. Actually, the absence of TMR considerably reduces the incidence of the disease and its oncogenecity. Moreover, the deletant mutants are less able to integrate the host genome and to reactivate next to this (Kaufer *et al.,* 2011).

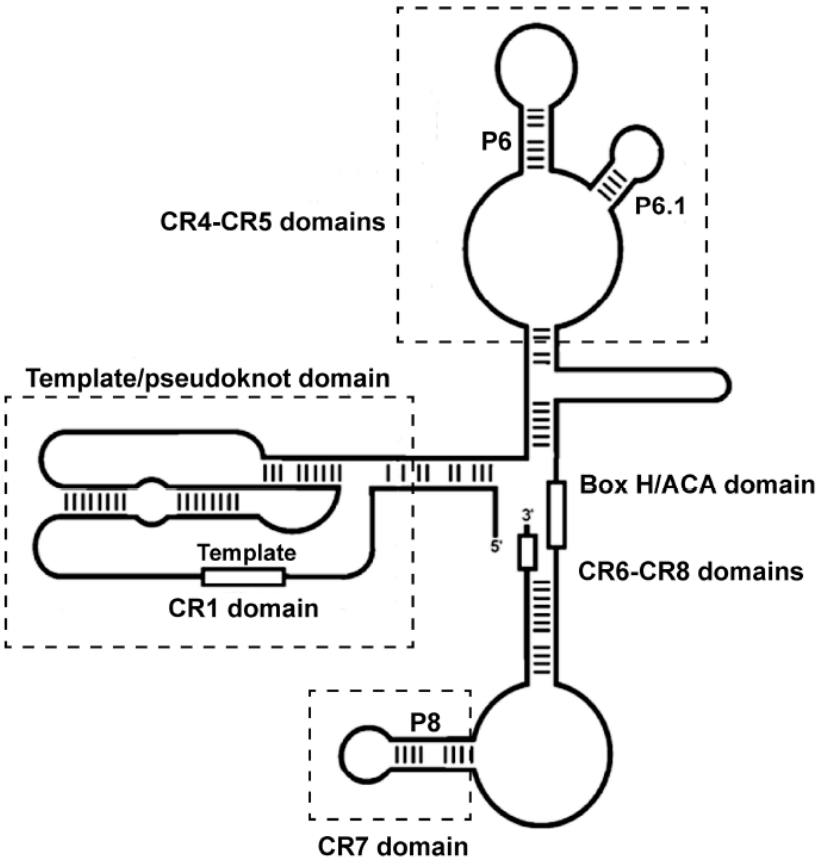


Figure 10 : **Secondary structure of the vTR of GaHV-2.** A highlight was put on the homologies between vTR and chTR represented as CR (Conserved Regions) domains (Kheimar *et al*, 2017).

Since the virus maintains itself within the telomeres, it is threatened by telomeres shortening. The virus evolved to discard this threat and therefore induces the cellular telomerase reverse transcriptase (TERT) by providing the RNA subunit that is essential for its activity. The virus genome encodes a viral telomerase RNA subunit (vTR) that shares 88% of its sequence with the chicken one (chTR). The conserved domain show a highly structured secondary conformation that display an even higher homology with chTR (Figure 10). vTR is present in two copies in both the TR_L and IR_L regions of the genome (Kheimar *et al.,* 2017).

As well as TMR, vTR is involved during transformation of GaHV-2-infected cells. Mutations in several parts of its sequence severely alter the transformation rate of the GaHV-2-infected cells (Trapp *et al.,* 2006). On the opposite, the overexpression of vTR leads to an increasing of the proliferation and the expression of adhesion factors (Kheimar *et al.,* 2017).

1.3.3.3) pp38 and pp24

A bidirectional promoter is located within the repeats surrounding the U_L segment of the viral genome near the U_L region. This bidirectional promoter controls the transcription of the 14kDa proteins associated transcripts and, in the opposite sense, of both pp24 and pp38 phosphoproteins. More precisely, pp24 gene overlaps the TR_L and U_L meanwhile pp38 overlaps IR_L and U_L (Boumart, 2018).

These two proteins, similar in their NH_2-terminal ends, assemble to form a heterodimer pp24/pp38 which transactivates its own promoter (Boumart, 2018). However the precise role of pp38 alone is still under debate. At first, pp38 was thought to be needed to maintain the transformation state of the lymphocytes (Xie *et al.,* 1996) but a recent study did not show any impact on the maintenance of the transformed state of the infected cells (Zhang *et al.,* 2019).

Nevertheless, pp38 was shown to interfere with apoptosis through an increase of mitochondrial succinate dehydrogenase. Since pp38 does not interact with the mitochondria, the precise mechanism is still under investigation (Piepenbrink *et al.,* 2009).

1.3.3.4) vIL-8

During the viral cycle, the virions do not locate within the bloodstream anytime. The only way used by the virus to spread is by cell-to-cell contacts. During this process, it is of first importance to establish close contacts between infected and non-infected cells. The role of the viral InterLeukin 8 (vIL-8) is to insure these contacts (Haertle *et al.,* 2017).

vIL-8 is a CXC chemokine homolog to the chicken CXCL13 present in two copies of the viral genome, in the long repeated regions. It acts as a chemoattractant for several cell types. Notably, B and T lymphocytes are reached by the virus in a vIL-8-dependent manner (Haertle *et al.,* 2017).

1.3.3.5) LATs

Latency Associated Transcripts are a series of long non-coding RNAs (lncRNA) transcribed from the IR_S and TR_S regions. Although the global function of these lncRNA is known to insure the latency of the virus in the infected cells, the precise mechanism still has to be described (Boumart, 2018; Rasschaert *et al.,* 2016).

The diversity in the transcripts and their splicing pattern could make their functions as much diverse. The transcription results in at least 22 transcripts encompassing from 3 to 15 exons. The first intron of the sequence encodes microRNAs that are known to inhibit both cellular and viral transcripts. This mechanism maintains latency within the infected cells and prevents detection by the immune system. The second specificity of these transcripts is their location which is antisense to ICP4 (see next point) in the genome of GaHV-2 but the LAT-encoding

gene is longer both upstream and downstream ICP4. Thus, the role of the LATs could be related to the function of ICP4 (Boumart, 2018; Rasschaert and Laurent, 2015; Rasschaert *et al.*, 2016).

1.3.3.6) miRNAs

Expression of viral miRNAs is a common feature among herpesviruses. 25 miRNAs from 13 premiRNAs have been shown to be expressed by GaHV-2 during the infection. They are gathered in three clusters. MDV-miR-M4 (miR-4) is encoded from the first cluster, also known as Meq-cluster because it is located upstream of the Meq oncogene, with 5 other miRNAs. The second cluster is named Mid-cluster and is located downstream of Meq. It encodes three pre-miRNAs. The third one, the LAT-cluster is located in the first intron of the LAT and encodes 5 miRNAs (Muylkens *et al.*, 2010; Zhuang *et al.*, 2017).

Viral miRNAs are widely studied for their involvement in the viral pathogenesis. Actually, they take part in the fine tuning of the gene expression, allowing the virus to keep a balance between immune evasion and spreading within the organism. Some of them regulate the transition to the late productive phase by targeting immediate early genes, other can also target the late genes and regulate the level of productive replication within the infected cells (Muylkens *et al.*, 2010).

1.3.3.7) MDV-miR-M4

miR-4 is known as a multi-functional miRNA. Both host and viral transcripts have been identified as targets of miR-4. One strand of the premiRNA (mdv-miR-M4-5p), which is the most expressed one, shares its seed sequence with gga-miR-155, its cellular ortholog. Thus, it downregulates the same set of cellular targets, having an effect on transcription factors, apoptosis signaling, cell differentiation, innate immunity and oncogenesis (Zhuang *et al.*, 2017). Viral mutants deleted for the miR-4 locus were produced and tested *in vitro* and *in vivo*. Despite its multi-functional aspect on cell cycle, deletion mutants did not show severe impairments on viral replication. However, these mutants lost their oncogenecity (Zhuang *et al.*, 2017). The effect of premiR-4 on viral transcripts have been characterized and showed interference with UL28 and UL32 expression (Muylkens *et al.*, 2010). These proteins are not characterized in GaHV-2 yet but they are essential factors needed for the cleavage and packaging processes in herpesviruses. A match was also found within the bicistronic transcript ICP22-US10. Precisely, the seed sequence of MDV-miR-M4-5p covers 7 nucleotides of the ICP22 open reading frame (ORF) including the stop codon. This sequence match was unexpectedly shown to increase the amount of ICP22-US10 bicistronic transcript and was shown to “turn on” the expression of the ICP22 protein (Boumart *et al.*, 2018).

1.4) ICP's

Infected-Cell Proteins (ICP's) are a group of immediate early (IE) proteins involved either directly or indirectly in transactivation of viral genes. In Alphaherpesviruses, 7 ICP's are described. However, only five of them are present in the GaHV-2 genome and four of them are characterized since ICP0 encoded protein has not been identified yet (Boumart, 2018; Previdelli *et al.,* 2019).

1.4.1) ICP4

ICP4 is an IE protein specific of the alphaherpesviruses. It is the major transactivator of the productive phase of the viral cycle. Within the alphaherpesviruses, the global identity rate for ICP4 is around 30% but some conserved domains share up to 55% of their genic identity. They are highly phosphorylated proteins which modify their localization and activity (Rasschaert and Laurent, 2015).

ICP4 of GaHV-2 is poorly expressed from the repeated regions surrounding U_S during latency. Its expression is extracellular at first and finally leads to reactivation (Rasschaert and Laurent, 2015). During the reactivation and the productive phase, the expression of ICP4 is increased. During reactivation, it localizes in the cytoplasm. During the fully-productive phase, ICP4 seems to localize shortly in the center of the nucleus and then, it relocalizes at the border, at the nuclear membrane (Boumart, 2018; Rasschaert and Laurent, 2015).

The most conserved domains are the 2nd and the 4th ones. The 2nd corresponds to the N-terminal domain of the protein, which allows targeting specific promoters and to multimerize. The 4th, with the 5th, corresponds to the C-terminal domain, which transactivates the genes by binding to promoter sequences and increases the recruitment of the RNA pol II (Rasschaert and Laurent, 2015).

Nevertheless, ICP4 has also a repressing function. The most characterized target for this repression is the negative feedback of ICP4 on its own promoter. In this case, ICP4 has a highly specific target ATCGT which make it a competitive inhibitor against the other transactivators (DeLuca, 2011; Rasschaert and Laurent, 2015).

The last role of ICP4, which has still to be fully described, is its involvement in the latency. ICP4 depletion triggers the apoptosis of GaHV-2-infected cells (Rasschaert and Laurent, 2015).

As the major transactivator protein encoded by the virus, ICP4 is tightly linked to several prominent viral genes. First, the LATs gene is antisense to ICP4. Their expression are negatively correlated. The LATs are assumed to have a negative effect on ICP4 to maintain latency. Next, a direct interaction of ICP4 with the pp38 promoter has been shown. There is also a link with Meq since there is a functional MERE-1 site within the ICP4 promoter. Finally, the protein is also regulated by the viral miRNAs, especially MDV-miR-M7 (Strassheim *et al.,* 2012).

1.4.2) ICP27

ICP27 is the only ICP that has been identified in every Herpesviridae. In GaHV-2, it is collinear with the gK protein (UL53) in the U_L part of the genome. This collinearity leads to two types of transcripts encoding the protein. The first one is transcribed from the gK promoter to build a bicistronic transcript encompassing both gK and ICP27. The second one, shorter, is made thanks to the ICP27 promoter, building a monocistronic transcript. These two promoters allow to monitor the expression pattern of the both genes. Actually, the gK promoter seems to be a hybrid promoter, which allows it to be active during the whole viral cycle since the ICP27 promoter is a canonical IE promoter only active during the lytic phase. However, gK is a L gene and ICP27 has a specific expression pattern (see further). Thus, according to the phase of the

viral cycle, the splicing of the pre-messenger RNA (pre-mRNA) regulates the expression pattern of the proteins (Strassheim *et al.,* 2016).

The splicing of the transcript encoded from the gK promoter can lead to the productive of either gK during the late productive phase or an alternative truncated version of the ICP27 protein during latency. How much the function of the protein is altered is not known and is hard to predict since the deleted part of the protein is the least conserved. Otherwise, when the expression of ICP27 relies on the more proximal promoter, the full sequence of ICP27 is expressed during the productive phase, like the other IE genes (Strassheim *et al.,* 2016).

The full-length ICP27 is made of several domains conserved or not among the herpesviruses. The most conserved domain is the homolog of the well-known UL69 superfamily. It is a folded domain which allows ICP27 to form complex with other proteins, including transcription factors or self-association (Tunnicliffe *et al.,* 2018). The two other conserved domains are the Nuclear Localization Signal (NLS) and the binding site for cellular factors like the Nuclear Export Factor (NXF1) (Amor *et al.,* 2011; Ote *et al.,* 2009; Strassheim *et al.,* 2016).

The functions of the full-length ICP27 are multiple and were precisely characterized, especially in HHV-1. Like the other ICP's of GaHV-2, most of its functions are related to the transcriptional and post-transcriptional regulation of the viral genes (Amor *et al.,* 2011; Strassheim *et al.,* 2016).

First, ICP27 is able to inhibit the pre-mRNA splicing by the spliceosome. Shortly, ICP27 binds to some splicing factors, for example SRp20 during HHV-1 infection, and stall the spliceosome factors at the splicing site in a "pre-spliceosomal" stage. Until now, it has been thought that this function could contribute to the shutoff of the cellular genes without affecting the viral genes that do not need any splicing for their expression. ICP27 is also interacting with SRPK, which is a kinase that phosphorylate the splicing factors, inhibiting its activity (Sandri-Goldin, 2008). During the infection by GaHV-2, only the first function described has been reported, inducing coalescence of the splicing factors in the nucleus and inhibiting the splicing.

Second, ICP27 has a role of transactivation by interacting with other transactivation proteins like ICP4. In HHV-1, ICP27 binds and stimulates RNA polymerase II (RNAPII). However, the binding of ICP27 to RNAPII could be due to a third unknown protein (Sandri-Goldin, 2008).

Third, ICP27 regulates the nuclear export of some proteins. Thanks to its NXF1 domain, it binds to export factors and, through a binding to the target protein, it takes the NXF1-dependent proteins out the nucleus (Ote *et al.,* 2009).

Finally, ICP27 also has a potential translational role. In several herpesviruses, ICP27 is able to bind to the translational machinery and monitor its activity. Notably, it has been shown to bind to the poly-A-binding protein (PABP) which is correlated with the inhibition of the nonsense-mediated mRNA decay (NMD) and with the initiation of the translation through interaction with collateral initiation factors (Fontaine-Rodriguez *et al.,* 2004).

1.4.3) ICP0

RLORF-1 is a putative ICP0 homolog in GaHV-2. This protein can switch on viral expression and trigger the reactivation from latent HHV-1 infection (Samrat *et al.,* 2017). However, in GaHV-2, a tagged protein was not revealed on a western blot and the phenotype of the infected cells was not affected. It seems that ICP0 is not expressed in GaHV-2 (Bertzbach *et al.,* 2019).

1.4.4) ICP8

Already characterized in HHV-1, ICP8 has been shown to be linked with the integration and the replication of the herpesviruses. By its ability to bind to ssDNA, ICP8 cooperates with UL9

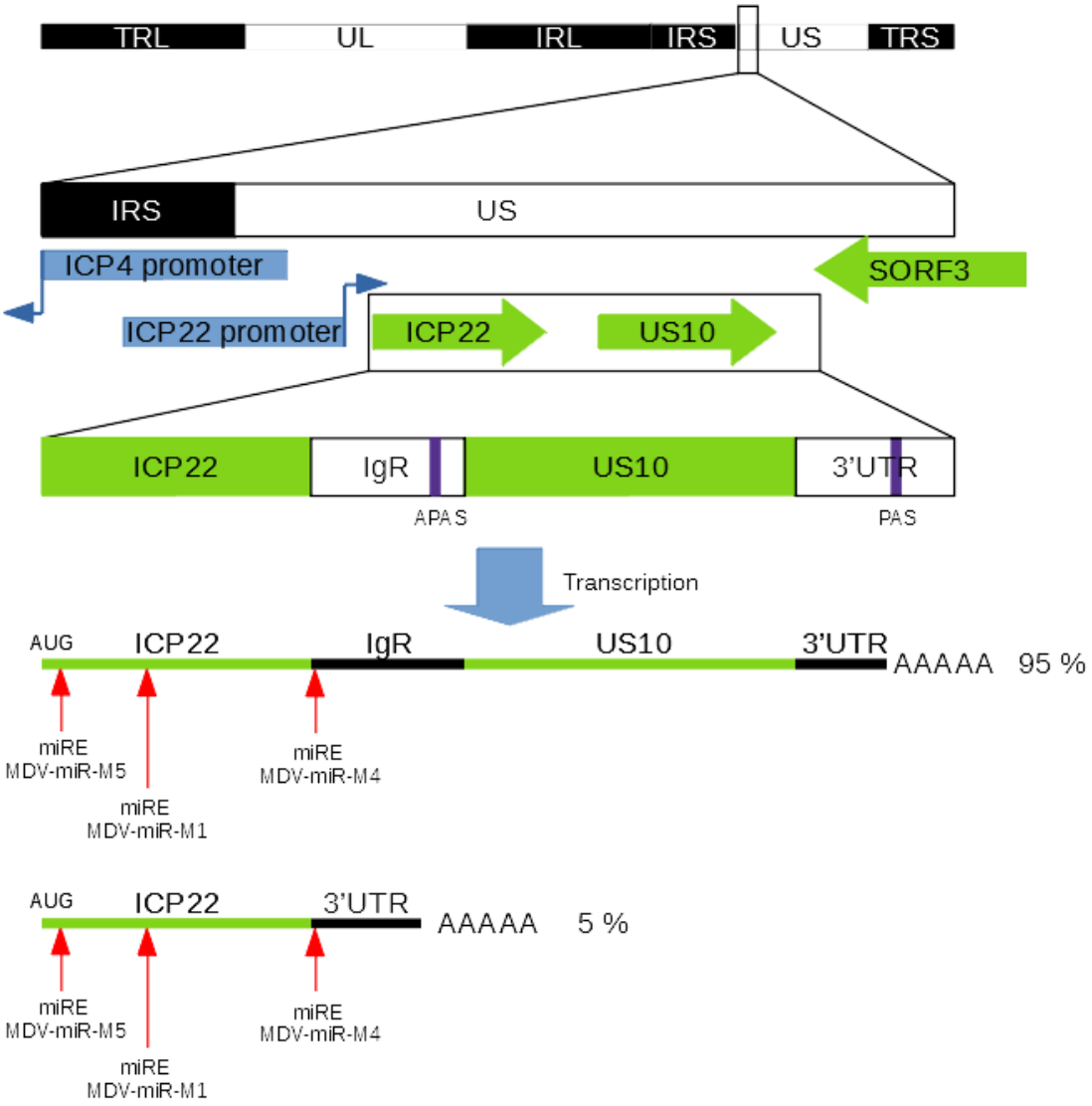


Figure 11 : **Localization of ICP22 on a genomic and transcript level**. ICP22 is transcribed from the U_S region of the genome. The transcription leads to two types of transcript : a monocistronic transcript encompassing only ICP22 (5% of the cases) polyadenylated thanks to an alternative poly-A signal (APAS) or a bicistronic transcript encompassing both ICP22 and US10 genes (95% of the cases) polyadenylated thanks to a canonical PAS. Three miRNA were shown to regulate the transcript encoding for ICP22: MDV-miR-M5; MDV-miR-M1 and MDV-miR-M4.

and UL12 to permit the recombination. During this mechanism, UL9 first opens the origin of replication. Next UL12 degrades one DNA strand of the virus. And then ICP8 binds to both the overhanging strand and the origin of replication. Self-interaction between the ICP8 proteins allow to get both these parts closer, increasing thereby the recombination mechanism (Weerasooriya *et al.,* 2019).

UL12 and ICP8, considered as a recombination complex, are also present in GaHV-2 and both are essential for the viral replication. Their roles seem to be the same as in an infection by HHV-1 but the precise mechanism still has to be clarified (Previdelli *et al.,* 2019).

1.4.5) ICP22

ICP22 is found in most of the alphaherpesviruses, it is only absent in iltoviruses (Chandra *et al.,* 2012). In the GaHV-2 genome. It is encoded from the US1 gene at the 5' end of the U_S region (Figure 11). Interestingly, like ICP27 and gK, ICP22 is expressed from a bicistronic transcript gathering ICP22 and US10. Actually, 95% of the transcripts are bicistronic, ending by a poly-A tail after US10 thanks to a canonical polyadenylation signal (PAS), AATAAA. A non-canonical signal, ATTAAA, is found between both ORF of ICP22 and US10 and leads to the polyadenylation of 5% of the transcripts (Boumart *et al.,* 2018).

Unlike the other ICP's expressed during a GaHV-2 infection, ICP22 has been characterized as an L gene expressed during the productive phase. During this phase, ICP22 is localized within the cytoplasm of the infected cells (Boumart *et al.,* 2018).

The gene encodes a protein made of 179 amino acids (aa) with a molecular weight of 29kDa when it is non-phosphorylated. This size includes a conserved Herpes_IE68 domain which confirm its homology with other IE proteins, especially the homology with the ICP22 encoded by other alphaherpesviruses (Boumart, 2018).

ICP22 has been shown to be important in maintaining the replication during progression of Marek's disease. Nevertheless, it does not affect either lymphoma formation *in vivo* or viral reactivation and its precise functions are still largely unknown (Parcells *et al.,* 1995). In Human Herpesvirus 1 and 3 (HHV-1 and HHV-3), it has a role in inhibiting the transcription by dephosphorylating the second Serin of the RNAP II. This inhibition is counterbalanced by VP16 which binds to the viral promoters and triggers the transcription of target genes. During an infection with HHV-1, interactions have been shown with pTEFβ which is involved in the dephosphorylation of the great subunit of the RNAP II (Zaborowska *et al.,* 2014).

In GaHV-2, a luciferase assay on several viral promoters showed the same effect in inhibiting the transcription. However, in this case, ICP22 remains within the cytoplasm and does not possess any nuclear localization signal (NLS). Thus, the precise mechanism of its activity remains to be elucidated (Boumart *et al.,* 2018). ICP22 could therefore be translocated by binding to a cellular or an unknown viral protein before or after phosphorylation.

The last role of ICP22, not shown in GaHV-2 but in HHV-1, is its link with the cellular life cycle. Actually, cells infected by HHV-1 produce less cyclines A and B and cdk1. It results in the blocking of the cell in the S phase of its life cycle which increases viral replication (Zaborowska *et al.,* 2014). The drop in viral replication in an ICP22-deletant-virus could be explained by this function of ICP22 but it still has to be described.

1.5) Objectives

Studying gene expression is essential in understanding the life cycle of GaHV-2, a virus that is considered as a major threat for the poultry industry. More precisely, regulating processes governing the gene expression is a prerequisite for depicting the life cycle of this virus that is lethal after a few weeks by inducing CD4+ T cell lymphoma. A hallmark of cancer is the decrease of the tumor suppressors concomitant with the increase of the proto-oncogenic factors. How GaHV-2 is able to tightly regulate its gene expression, for 1) invading the host ; 2) escaping the immune system; 3) producing lymphoma and 4) spread quickly in the organism before the death of the animal, is far from being fully described. Several genetic regulating mechanisms of GaHV-2 are already characterized, such as bidirectional promoters (pp38/pp24), bicistronic transcripts (ICP27-gK), alternative transcripts (LATs), and viral encoded miRNA produced from spliced transcripts.

GaHV-2 encoded miRNAs are potent negative regulators of both viral and cellular genes. Numerous studies have shown these noncoding RNAs involved in the switch between productive and latent phases of the infection by inhibiting the translation of viral genes supporting the early and late productive phases. Regarding cellular genes targeted by GaHV-2 miRNAs, several reports have shown that viral miRNAs are involved in the cellular transformation by downregulating a large set of genes controlling the cell proliferation. Thus, the GaHV-2 miRNAs are known as negative regulator of the gene expression fine tuning several processes underlying the progression of both the viral and the cellular cycles. No miRNA-induced activation was really deciphered up to now. Knowing the miRNA necessity during latency and the high diversity of the described roles attributed to miRNA, this fact could be surprising.

Recently, Boumart and collaborators (Boumart *et al.,* 2018) described a bicistronic transcript encompassing the ICP22 and US10 ORF. This peculiar transcript was shown to be targeted by three miRNAs. One of them, miR4, was unexpectedly shown to increase the amount of bicistronic transcript and was shown to "turn on" the expression of the ICP22 protein. It is still not known when miR4 binds to its target to enhance gene expression, neither how it happens.

The main goal of this project is to elucidate the whole pathway that triggers the upregulation of the expression mediated by a miRNA targeting a bicistronic transcript. This study aims at investigating the different functional parts of the whole transcript that might be involved in the upregulation triggered through miRNA-mRNA interactions. First the whole transcript will be analyzed thanks to bioinformatics. Primary or secondary RNA structures might explain the recruitment of proteins impairing the translation of the bicistronic transcript. This part of the project aimed at localizing these potential RNA structures. From this in silico analysis, the main working hypothesis will be addressed through the following questions : (1) Do specific RNA motifs have an impact on the translation of both ORF localized in the bicistronic transcript? (2) Is it possible to interfere with these motifs through miRNA-mRNA interaction?

Answering these questions might open a gate to a new RNA-mediated pathway of gene expression and think outside the box of the so-called "central dogma". To investigate these questions a reporter system will first be built by replacing both ORF of the bicistronic transcript with fluorescent reporter. These reporters will be used to monitor the expression of the genes both in the presence and in the absence of miRNA expression. Next to this, the half-life of the transcripts will be measured both in the presence and in the absence of miRNA expression.

Table 1 : List of primers used following the construction needed. The small letters represent a tail added to our primers to bind with the restriction enzymes to the constructs. The italic letters represent the restriction sites. The capital letters represent the sequence matching with the full sequence.

Constructions	**Bording primers Fwd/Rev**	**Overlap primers Rev/Fwd**
ICP22-miRE wt-pAmut-US10	acac*gctagc*ATGAGTCGTGATCGAGATCGAGC	GCCGTGTAAGATATATACATATTGGAGAGGTATGGCC
	acac*aagctt*CGAAGTAATCCTCATGCATACCCCA	TCCAATATGTATATATCTTACACGGCGCTCGTCTCCAG
GFP-miREwt-US10	acac*gctagc*ATGGTGAGCAAGGGCGAGGAGCT	GCAATTTACTGTCTACCGACAACTTGTACAGCTCGTCC
	acac*aagctt*CGAAGTAATCCTCATGCATACCCCA	CGAGCTGTACAAGTTGTCGGTAGACAGTAAATTGC
GFP-miREmut-US10	acac*gctagc*ATGGTGAGCAAGGGCGAGGAGCT	GGATAGTTAGATTATATCACTGTCTACCGACAACTTGTAC
	acac*aagctt*CGAAGTAATCCTCATGCATACCCCA	CGGTAGACAGTGATATAATCTAACTATCCAGACTTGAAGAG
GFP-UTR	acac*gctagc*ATGGTGAGCAAGGGCGAGGAGCT	GCGATTAGTTACTTGTACAGCTCGTCCATGCCGAG
	acac*aagctt*CGAAGTAATCCTCATGCATACCCCA	GGACGAGCTGTACAAGTAACTAATCGCACAATTATTAATAGG
ICP22-miREwt-GFP	acac*gctagc*ATGAGTCGTGATCGAGATCGAGC	GCTCACCATACCTCTCCAATATGTATATATCTTACACG
	acac*aagctt*CGAAGTAATCCTCATGCATACCCCA	CATATTGGAGAGGTATGGTGAGCAAGGGCGAGG
ICP22-miREmut-GFP	acac*gctagc*ATGAGTCGTGATCGAGATCGAGC	GGATAGTTAGATTATATCACTGTCTACCGACAAATCGTTC
	acac*aagctt*CGAAGTAATCCTCATGCATACCCCA	CGGTAGACAGTGATATAATCTAACTATCCAGACTTGAAGAG
GFP-hp-miREwt-US10	acac*gctagc*ATGGTGAGCAAGGGCGAGGAGCT	GAAATAGTTCCCGACTTGTACAGCTCGTCCATGCCG
	acac*aagctt*CGAAGTAATCCTCATGCATACCCCA	GCTGTACAAGTCGGGAACTATTTCTTCAGACTGCG
GFP-hp-miREmut-US10	acac*gctagc*ATGGTGAGCAAGGGCGAGGAGCT	GAAATAGTTCCCGACTTGTACAGCTCGTCCATGCCG
	acac*aagctt*CGAAGTAATCCTCATGCATACCCCA	GCTGTACAAGTCGGGAACTATTTCTTCAGACTGCG
qPCR-ICP22	CAACTGCCACATTCACATCCG	
	GCACATAACCGAGCGACATCG	
qPCR-GFP	CTACGGCAAGCTGACCCTGAAG	
	GTCGTGCTGCTTCATGTGGTCG	

2) Material and methods

2.1) Viral and bacterial strains

The very virulent strain of the GaHV-2 virus, RB1B (GenBank EF523390.1), is used for this research. This strain of the virus was isolated in the early 1980's, when it displayed resistance against immunity induced by SB-1 and HVT vaccine strains (Schat *et al.*, 1982).

The TG1 bacterial strain of *E.coli* was used to perform the amplification of the plasmids. These bacteria were made competent by a homemade treatment. A preculture was grown overnight in 10mL of LB medium at 37°C and put into 1L of 2YT medium. The main culture was grown until the optic density (600nm of wavelength) of the medium was at 0,45. The bacteria were put on ice during 2 hours before a first centrifugation (5000 rpm – 4°C – 15min). The supernatant was removed and the bacteria resuspended in 250mL of MilliQ water before a second centrifugation (5000rpm – 4°C – 15min). The supernatant was again removed and the bacteria resuspended in 200mL of MilliQ water before a third centrifugation (5000rpm – 4°C – 15min). This step was again repeated twice. The supernatant was finally removed and the bacteria resuspended in 20mL of sterile glycerol (10%) before being centrifugated a last time (5000rpm – 4°C – 15min). The supernatant was removed and the bacteria resuspended in 2mL of sterile glycerol (10%). The bacteria were then aliquoted 50µL aliquots and frozen at -80°C.

2.2) Bioinformatic analysis

ICP22-US10 bicistronic transcript was analyzed using web server of *RNA structure ProbKnot* from *Mathews Group*© [SOURCE]. This analysis produced a map of the nucleotides interactions in the full-length sequence of the ICP22-US10 bicistronic transcript.

Analysis of the structure of the sequence surrounding the stop codon of each analysed bicistronic transcript has been done using web server of *RNAfold* from University of Vienna (Institute for Theoretical Chemistry).

2.3) Site-directed mutagenesis

An overlapping PCR is made using at least two primers pairs [Table 1]. These primer pairs were designed with the *Geneious* software from *Biomatters*© and verified using web server of *NetPrimer* from *Premier Biosoft*© (http://www.premierbiosoft.com/NetPrimer/AnalyzePrimer.jsp).

The bording primer pairs contain restriction sites to allow the digestion/ligation. Following the organization of the plasmid, *NheI* was chosen for the 5'end and *HindIII* for the 3'end. The overlapping Polymerase Chain Reaction (PCR) was made using Q5 high fidelity enzyme in a 50µL mix following the manufacturer protocol (*New-England Biolabs* © (NEB)). It means 10µL of 5x buffer, 1µL of each primer (10µM), 1µL of dNTPs (total concentration of 20mM ; 5mM of each nucleotide), from 0.1 to 1ng of DNA depending on its origin (plasmid versus genomic) and 0.5µL of the enzyme (2U//µL). This enzyme can add 2000 nucleotides/minute, each reaction time was adapted following the length of the produced amplicons. Both the dehybridization and the hybridization steps were processed 30 seconds. To combine two amplicons in overlapping PCR, initial steps included 5 cycles without primers. The resulting overlapping fragment was amplified through 30 cycles with appropriate primers.

2.4) Digestion of the inserts

Double digestion were done on the inserts using *NheI* and *HindIII* restriction enzymes bought in NEB industry. This double digestion was done twice and the digested insert purified using *Macherey-Nagel* (MN) © kit and finally eluted in a 18µL final volume.

The first double digestion is done using a 50µL mix containing 5µL of 10x buffer, a maximum of insert DNA (1µg mx for plasmid DNA), 20U of each enzyme. For the second double digestion, 15U of each enzyme were added to the same mix described at the first step. The digestion requires, for each step, a digestion at 37°C during 1h and a denaturation at 8°C during 20 minutes.

2.5) Vector preparation

The same *Nhe*I/*Hind*III digestion as above was done on 20µg of plasmid *pcDNA* 3.1 (-) (*Invitrogen*©). This plasmid contains resistance gene against ampicillin and neomycin and a cytomegalovirus promoter for overexpression of the insert.

A separation of the restricted components is made on an electrophoresis TAE-agarose (1% agarose concentration) gel. The fragments were purified using MN kit and eluted in a final volume of 100µL. To avoid self-ligation, the plasmid was dephosphorylated twice using calf intestinal alkaline phosphatase (CIP) from NEB and purified using phenol/chloroform extraction. A first dephosphorylation is made in a 200µL mix containing 100µL of DNA (purified at the previous step), 20µL of 10x buffer and 100U of CIP. The second dephosphorylation was done by putting 80U of CIP in the same mix. Each step was realized by heating the mix 20 minutes at 37°C, 20 minutes at 56°C and 3 minutes at 4°C.

2.6) DNA purifications

Two ways to purify DNA were used. The first one used phenol/chloroform extraction, the second one used the MN kit.

Manufacturer protocol was used for MN purification. This includes a dissolution step for the gel in a NTI buffer (200µL of NTI for 100mg of gel), the obtained mix is put on a column from the kit for a centrifugation (11000xg-1 min) and a double wash with 675µL NT3, each wash is followed by another centrifugation step (1000xg-1min). A last centrifugation step is done to dry the column (11000xg-1 min). Finally the elution was done using 18µL of water at 37°C, incubated 1 minute on the column before a last centrifugation step (11000xg-1min).

For the phenol/chloroform extraction, the samples are diluted to obtain an initial volume of 300µL. 300µL of a phenol:chloroform:isoamylalcohol (24:25:1) mix is added before a centrifugation step (15min-13000xg-4°C). Supernatant is added to chloroform:isoamylalcohol (49:1) before another centrifugation step (15min-13000xg-4°C). Supernatant is added to a solution containing 600µL ethanol (100%) and 100µL sodium acetate (3M). One hour at -80°C is then required. A third centrifugation step (20min-14000xg-4°C) allows to recover the pellet. This pellet is resuspended in 1mL ethanol (75%) before a fourth centrifugation step (5min-7500xg-4°C). Supernatant is discarded and pellet is dried at 37°C during 5 minutes before being resuspended in 30µL of deionized water.

2.7) Cloning and screening

The first construct with a mutated cryptic PAS, was inserted into a pGEM-T easy vector (Promega©) by TA-cloning. This vector does not need any digestion step. Bacteria were transformed with this plasmid and grown in a LB-agar complemented with ampicillin (100µg/mL), Xgal (60µg/mL) and IPTG (40µg/mL) in order to perform a β-galactosidase screen. The white colonies were next screened by PCR (plasmid-specific primers) and then the clones presenting the proper insert were sent to sequencing (*Eurofins Genomics*©) and grown to make a MiniPrep.

The reporters are cloned into *pcDNA* 3.1 (-) previously prepared. The *pcDNA*-transformed bacteria are grown on LB-agar medium complemented with ampicillin (100µg/mL) in Petri dishes. Colonies are screened by directional PCR, forward primer being insert-specific and

reverse primer being plasmid-specific. The clones presenting the proper insert in the proper orientation are sent to sequencing (*Eurofins Genomics*©) and grown to make a MidiPrep.

Both ligations are made using the same mix adapted for T4 ligase from *Promega*©. 3µL of inserts (maximal concentration) are put with 100ng of plasmid, 5µL of 2x buffer and 1µL of enzyme.

The clones that grew overnight on the plates are screened by PCR using GoTaq enzyme (*Promega*©). Colonies were chosen randomly on the plates to be picked and suspended into 25µL of PCR mix prepared following the manufacturer protocol. It includes 5µL of 5x buffer, 0.5µL of dNTPs (20mM), 0.5µL of primers (10µM) and 0.125µL of enzyme (5U/µL).

2.8) Mini/Midi Prep

For the MiniPrep, the bacterial clones displaying the proper insert are grown overnight at 37°C in 5mL of LB-ampicillin (100µg/mL) to amplify the plasmid. This bacterial culture is used to make a MiniPrep with a *Qiagen*© kit following the manufacturer protocol. It begins by a centrifugation of the culture (6800xg-3min-room temperature (RT)). The supernatant is discarded and 250µL of resuspension buffer PI are added to resuspend the pellet. The lysis reaction is done using 250µL of P2 buffer for a maximum 5 minutes time. The lysis buffer is neutralized with 350µL of N3 buffer. A centrifugation (13000rpm-10min-RT) is done and the supernatant is put on the spin column furnished in the kit for another centrifugation step (13000rpm-1min-RT). The column is washed with 500µL of PB buffer, centrifuged twice to discard a maximum of the wash buffer. The elution was done using 50µL of deionized water with an incubation time of 1 minute followed by a last centrifugation step (13000rpm-1min-RT).

For the MIDI prep, the bacterial clones displaying the proper insert in the proper orientation are grown overnight at 37°C in 200mL of LB-ampicillin (100µg/mL) to amplify the plasmid. This bacterial culture is used to make a MIDI prep with a MN kit following the manufacturer protocol. It begins by a centrifugation of the culture (5000xg-15min-4°C). On ice, the supernatant is discarded and 8mL of the resuspension buffer [NOM] are added to resuspend the pellet. The lysis reaction is done by adding 8mL of lysis buffer and by mixing the tube gently. This step does not exceed a maximum 5 minutes time. The filter of the column and the column itself are drown by adding on them 12mL of equilibration buffer (EQU). The lysis reaction is neutralized by adding 8mL of neutralization buffer (NEU) and by mixing the tube gently. The supernatant of the reaction is put on the filter of the column and then 5mL of EQU buffer are added on it. The filter is then removed and a wash is done with 8mL of WASH buffer. Up from a 50mL falcon tube are put 5mL of elution buffer (ELU) on the column. The column is discarded and 3.5mL of isopropanol are added in the falcon tube. A centrifugation step (15000xg-30min-4°C) is used to precipitate the DNA. The supernatant is discarded and the pellet is washed with 2mL of ethanol (70%) and a centrifugation (7500xg-15min-4°C). The pellet is dried during 10 minutes before being resuspended in deionized water.

2.9) Cell lines

2.9.1) ESCDL-1 cell line

Embryonic Stem Cell Derived Line 1 (ESCDL-1) are a mesenchymal cell-derived lytic infection model developed by Vautherot *et al.*, 2017. These cells have some advantages compared to chicken embryonic fibroblasts (CEF) that were commonly used until now for GaHV-2 lytic infection. They are easy to grow, their phenotype is not altered after a long cell culture and a high number of passages.

They were cultured in a supplemented Dubelcco's Modified Eagle Medium (DMEM F12 1:1). The supplementation included 10% of Fetal Bovine Serum (FBS); 1% of Penicillin Streptomycin (PS), 1% of Non-Essential Amino Acids (NEAA) and 1% of sodium pyruvate. The cells were grown in an incubator at 37°C and 5% CO_2. The passages were made when cells were 85% confluent.

2.9.2) MSB-1 cell line

MSB-1 are a latently infected T CD4+ lymphocyte model isolated from a chicken spleen lymphoma. They were cultured in a supplemented Roswell Park Memorial Institute medium (RPMI). The supplementation includes 10% of FBS, 5% of Chicken Serum (CS), 1% of PS, 1% of NEAA, 1% of sodium pyruvate and 4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES) 25 mM. The cells are grown in an incubator at 41°C and 5% CO_2. The passages are made every 2 days to restore a cell density of 1/20.

2.9.3) DF-1 cell line

DF-1 are non-infected chicken fibroblasts self-immortalized derived from a chicken embryo. Cells were cultured in a DMEM supplemented with 10% FBS.

2.9.4) CEF

Chicken embryonic fibroblasts are extracted from 11-days-old chicken embryos. The extraction is made after the opening of the egg. The embryo was removed from the egg and put in the Petri dish containing PBS. The head of the embryo is cut right after the opening, as well as the wings and legs. Then, the organs are removed and only the epithelium is taken apart. The epithelium was trypsinized twice and the cells were gathered by two centrifugation (300xg-7min). The trash are removed as much as possible between the centrifugations. Next, the cells are specifically selected by a primary CEF culture medium. The primary CEF culture medium is a DMEM supplemented with 1% FBS, 1% CS and 1% PS. The primary CEFs are cultured over weekend and then a secondary culture medium is used. This secondary culture medium is a DMEM supplemented with 10% FBS and 1%CS.

2.10) Transfection

Two ways were used to transfect chicken cells. The first one, using nucleofection was performed using *Amaxa Nucleofection* kit (Lonza ©) on ESCDL-1 cell line. The cells were passed 24 hours before nucleofection and gelatin 2% was applied on 6-wells plates. Two hours before nucleofection, 5 x 10^5 cells/well were seeded on the gelatin and incubated in a DMEM medium containing 1% of FBS and 1% of CS, in order to make them grip the plate. 10^6 cells/condition were nucleofected in 100µL of *Amaxa Nucleofector Solution* specific for fibroblasts. The program was F-024, specific for fibroblasts. As fast as possible, 500µL of medium (1% FBS ; 1%CS) were added on the nucleofected cells. The cells were seeded on the plates (with the 5 x 10^5 cells seeded previously). The medium was changed 24 hours after the nucleofection and the cells were checked by fluorescent microscopy.

The second way of transfection was lipofection and was performed on DF-1 cell line. The cells were counted and seeded on 6-wells plates (1 000 000 cells/well) 16 hours before lipofection. The DNA was diluted in 400µL of Opti-MEM medium. 10µL of Lipofectamine 2000 were used/lipofection and diluted in 400µL of Opti-MEM medium. Both DNA and lipofectamine 2000 were mixed together during 20 minutes before nucleofection. During this time, the cells were washed 3 times with Opti-MEM. The mix (800µL/condition) was applied on the cells which were incubated 6 hours after lipofection. The medium was then replaced by a classical DMEM medium with 10% of FBS.

2.11) Flow Cytometry

One hour before analysis, the cells were detached from the plates using trypsin EDTA and resuspended into PBS. The cytometry was performed using a *FACSVerse* machine from BD Biosciences ©. The system allows to gather single cell data by passing individual cells in front of a laser. Detector around the cells measure the light intensity to create different data following different parameters : the side scatter (SSC) measure the granulosity of the cells, the forward scatter (FSC) measure their size and a GFP (Green Fluorescent Protein) filter measure their fluorescence. To detect these parameters, several signal amplificators are used. Signal amplification is monitored following the potential of the electrodes. In this case, the FSC amplificator was put at 100V, the SSC at 200V and the GFP at 320V.

2.12) RNA extraction

RNA extraction was performed using TRIzol extraction 24h post-transfection. At the beginning, the cells were either detached from the plate using trypsin-EDTA and pelleted by centrifugation (300xg – 7 min) or the TRI reagent was put directly on the cells. 1 mL of TRIzol was added for 5 million cells and incubated 5 minutes at RT. 200µL of chloroform were added and the tubes incubated during 10 min before a centrifugation (12000xg – 15min – 4°C). The aqueous phase, which stands up in the tube was extracted and put in 500µL of isopropanol and incubated during 10 min to precipitate the RNA. The tubes were centrifugated (12000xg – 10min – 4°C) and the supernatant removed. The pellets were washed with ethanol 70% before being centrifugated (7500xg – 10 min – 4°C). The pellet was dried and resuspended into 88µL of water to perform a DNase treatment.

2.13) DNase treatment

Since a lot of DNA was put into the cells through the transfection, it's important to carry out a DNase treatment to remove every DNA molecule that could interfere with the qPCR. It was performed using 10µL of DNase buffer and 2µL of DNase after having passed the samples at 65°C during 5 min. The DNase reaction spent 20 minutes and followed by a phenol/chloroform extraction (see 2.6).

2.14) Reverse transcription

Reverse transcription was done using SuperScript III kit from Thermofisher© following the manufacturer's protocol. It was chosen to reverse transcribe 2µg of RNA for each sample. The mix was thus containing 3µg of RNA, 1µL of oligodT (0,5µg/µL), 1µL of dNTPs (20mM) and ultrapure water up to 13µL. This mix was put at 65°C for 5 minutes and then put at 4°C for 1min. Then were added 4µL of buffer (5x), 1µL of DTT (0,1M), 1µL of RNase OUT and 1µL of SuperScript III enzyme (200U/µL). The reverse transcription was carried out at 50°C for 20 minutes and the enzyme was denatured at 70°C for 20 more minutes. The samples were used for qPCR without further treatment.

2.15) qPCR

qPCR were performed using Eco^{TM} Real-Time PCR System with the FastStart SYBR Green Master (Sigma Aldrich©) kit. 3 dilutions of the reverse transcribed products were done to carry out the qPCR (1/10;1/30;1/90). 2µL of each dilution was put into a mix containing 5µL of master mix (2x), 0,8µL of each primer (10µM) and 1,4µL of water. The analysis of the results was done using GAPDH to standardize the results.

2.16) Dactinomycin treatment

To measure the mRNA half-life, a treatment with actinomycin D (or Dactinomycin) (20µg/mL of medium) was applied to the transfected cells. This drug strongly binds to DNA and block the progression of the RNAPII, inhibiting the transcription. At each time-point (1h;4h;6h;12h;24h), the RNA was extracted using the TRIzol extraction and quantified using RT-qPCR.

2.17) Protein extraction

Protein extraction was carried out 48h post-transfection, using RIPA reagent. The cells were pelleted using trypsin-EDTA and centrifugated (300xg – 7min) and put into 50µL of RIPA mix. It includes 43,5µL of RIPA solution, 5µL of Triton (10x), 1µL of Complete (50x) and 0,5µL of PMSF (100x). This reaction was carried out on ice for 30 minutes. Next, the proteins were sonicated 3x1min and then centrifugated (14000 rpm – 15min – 4°C). The proteins were next dosed using a MultiSkan EX machine with Pierce reagent from the Thermo Scientific© BCA Protein Assay kit.

2.18) Western blot

25µg of proteins were put into lanes of a 10% acrylamide gel in the presence of SDS and migrated during 2 hours in order to perform an SDS PAGE separation. One lane was attributed to the scale dye. The running gel was made of 2,5 mL of acrylamide (37,5:1), 100µL of SDS (10%), 2,5 mL of Tris HCl (1,5M – pH 8,8), 4,75 mL of distilled water, 100µL of APS and 10µL of TEMED. The stacking gel was made of 300µL of acrylamide (same concentration), 30µL of SDS (10%), 760µL of Tris HCl (0,5M – pH 6,8), 1,875mL of distilled water, 30µL of APS and 3µL of TEMED. For the electrophoresis, the running buffer was made of 144,2 g of glycine, 30g of Tris, 10g of SDS ans up to 1L of distilled water. The transfer was done on a PVDF membrane at 4°C during 1h15. The transfer buffer was composed of 15,25g of Tris, 43,5g of glycine, 1L of methanol and up to 5L of distilled water. Next to the transfer, the membranes were blocked using TBST-BSA buffer. The TBST buffer is made of 500µL of Tween20 put into 500mL of TBS. 5% of BSA was put into this solution to have the proper blocking buffer. The blocking was carried out during 3h at RT. The primary antibody of HSP90 is a mouse monocolonal antibody that was provided by *abcam*©, it was diluted at 1:2500 for the experiment. The GFP antibody is a mouse monoclonal antibody that was provided by *invitrogen*©, it was diluted 1:2000 for the experiment. The ICP22 antibody is a rabbit monoclonal antibody that was provided by the Tours François Rabelais University, it was diluted 1:1000 for the experiment. The primary antibodies were incubated overnight at 4°C. Next to the incubation, the membranes were washed 3x10min at RT using a TBST buffer, prepared as previously described. The secondary antibody were diluted 1:2000 and put during 1h at RT. The secondary washing were done 3x15min using the same TBST buffer. The membranes were next revealed using the Pierce reagent and the ImageQuant LAS 4000 machine.

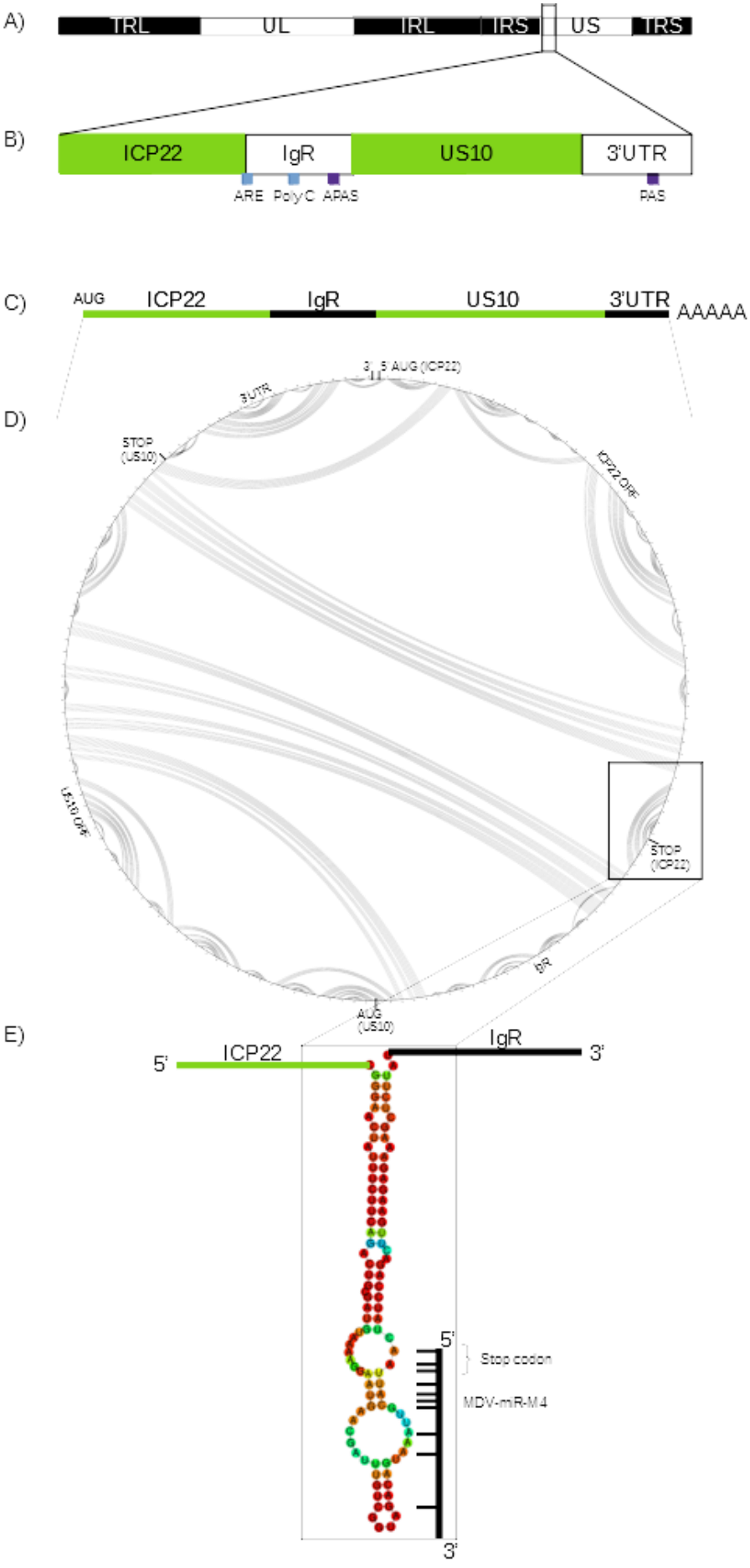


Figure 12 : **Bioinformatics analysis of the ICP22-US10 bicistronic transcript**. **A)** Structural organization of the GaHV-2 genome. It displays two unique regions, one short and one long. These two unique parts are surrounded by inverted repeats regions. **B)** Localization of the ICP22-US10 bicistronic transcript. The bioinformatics analysis revealed 2 relevant features in the primary structure of the transcript. First, a poly-C site and next an ARE overlapping the stop codon of the first ORF. **C)** mRNA produced in 95% of the case, induced by the canonical PAS. **D)** Prediction of ICP22-US10 internal interactions. The lines represent the probable interactions between base pairs. Both ICP22 and US10 ORFs are indicated. **E)** Bioinformatics analysis of the RNA folding of the region surrounding the stop codon of ICP22 ORF. The scale, from purple to red, indicates the increasing probability for an interaction to happen following the MFE. The target site of miR4 is shown as well as the position of the ICP22 stop codon.

3) Results

3.1) Bioinformatics analysis of the ICP22-US10 bicistronic transcript reveals a strong structural hairpin located at the stop codon of the ICP22 ORF

Previous data indicated that ICP22 is mainly expressed from a bicistronic transcript under the regulation of three miRNAs: MDV-miR-M1, MDV-miR-M5 and MDV-miR-M4 (miR4). In the case of miR4, the bicistronic transcript was shown to be increased by the miRNA-mRNA interaction at a site covering the stop codon of the first ORF of the transcript. To better understand the mechanism of expression of the 2 ORFs embedded within this major bicistronic transcript, we first analyzed the genetic structure of this peculiar transcript (Figure 12). Bioinformatics analysis was carried out to identify candidate motifs involved in any mechanism susceptible to trigger a decay pathway. A first structure, right after the stop codon was found to be a potential AU-rich element (ARE), this structure, made of 61 nucleotides, has a GC content of only 23%. Moreover, this region encompasses a stretch of 19 nucleotides including only A and U. A second interesting feature is a polyC stretch made of 8 consecutive C located in the intergenic region. Both of these features might recruit proteins that would either trigger the degradation of the mRNA or interact with Ago to disturb its endonucleolytic activity.

Bioinformatics analysis was completed to determine sequence self-complementarities over the full length mRNA. The resulting map of complementarities shows multiple interactions all over the sequence. We focused on the intergenic region between ICP22 and US10 ORFs. A highly structured hairpin was identified. It involved 70 nucleotides upstream the stop codon and 30 nucleotides downstream. This hairpin might hamper the progression of the ribosome through the mRNA. Moreover, this hairpin is likely to act as a platform recruiting proteins involved in the mRNA degradation. This structure surrounding the ICP22 stop codon could be the key for the upregulation involving MDV-miR-M4.

To support this hypothesis, a region showing weaker interaction and covering 15 nucleotides was identified around the stop codon. This stretch of nucleotides corresponds to the target site of miR-4. This makes the sequence reachable by the miRISC complex. Interactions with the miRNA may trigger a dehybridization process of the whole hairpin structure, which would leave the path free for the ribosome. In addition, the linearization of the mRNA following miRNA interaction might prevent protein interaction associated with the hairpin structure. From another point of view, it could also be interpreted that the miRNA could hamper the folding of this hairpin by binding to the mRNA before the closing of the structure.

To test the interaction between the activator miRNA and the highly structured bicistronic mRNA, reporter constructs were developed to decipher the mechanisms regulating the translation of the ICP22-US10 bicistronic transcript.

The initial step in generating the reporter constructs required the mutagenesis of the cryptic PAS to maintain the bicistronic nature of all constructs.

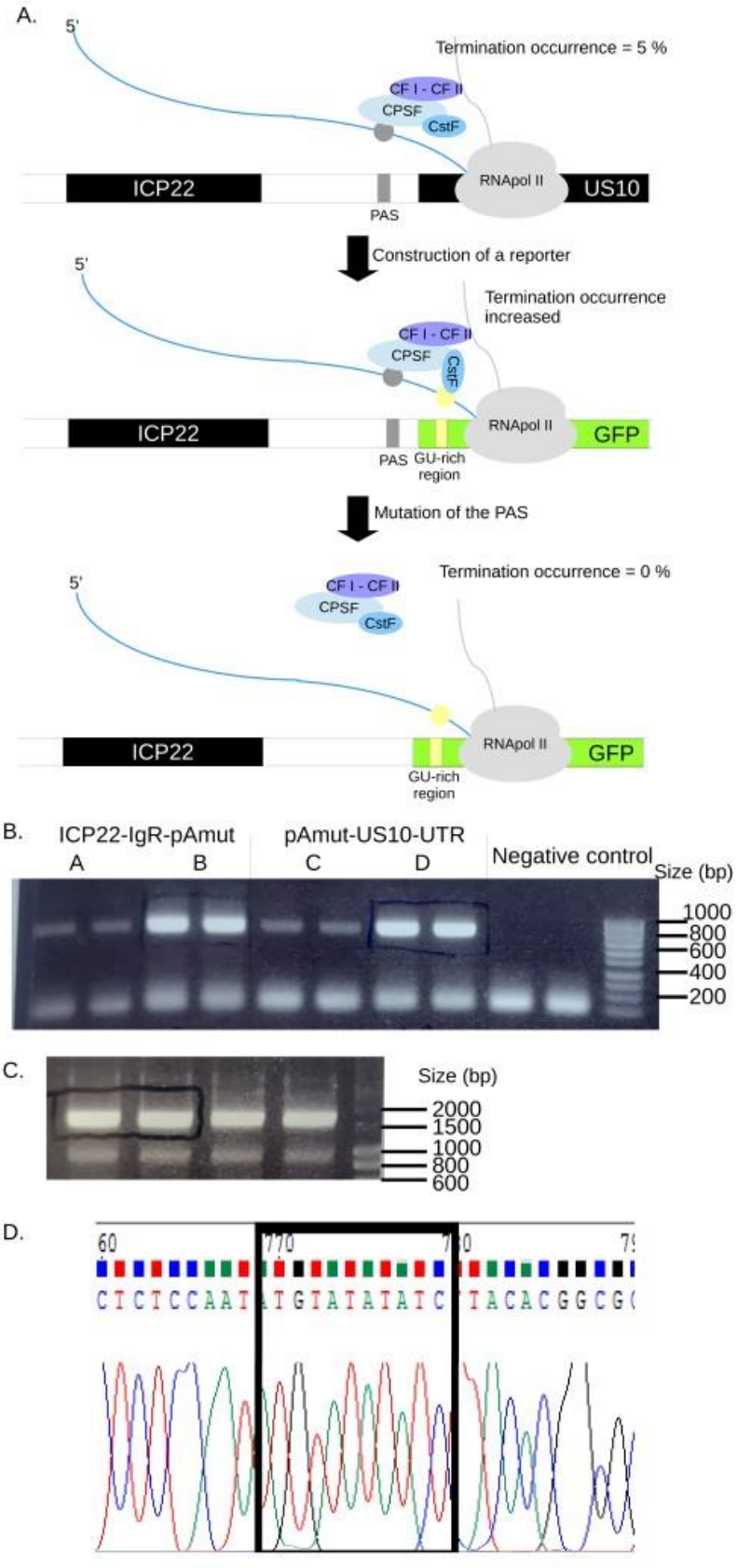


Figure 13 : **Mutation of the minor PAS of the bicistronic transcript**. A) Involvement of the alternative PAS with the bicistron transcription. When GFP is inserted downstream of the minor PAS, a GU-rich region is brought right after the signal which increases the termination at this site. Mutating the PAS standardize the transcription by only allowing bicistronic transcripts. B) Electrophoresis gel revealing the amplicons of ICP22-pAmut and pAmut-US10-UTR that are used for the overlap PCR. The surrounded bands were purified to perform the overlap PCR. The sizes of the initial amplicons are 830bp for the *NheI*-ICP22-pAmut fragment and 817bp for the pAmut-US10-UTR-*HindIII* fragment. Two concentrations of template DNA were tested: 0.01ng (lanes A and C) and 0.1ng (lanes B and D). C) Electrophoresis gel revealing the result of the overlapping PCR. All the lanes represent the same PCR product loaded in 4 pits. The size of the amplicons is 1623bp. D) Electrophoregram resulting from the Sanger sequencing at *Eurofins Genomics*© industries. The figure depicts the nucleotides bases ranging from the 760th position to the 790th. The mutated site is located between position 773 and 779.

3.2) Mutation of an internal poly-A signal within the bicistronic ICP22-US10 mRNA

Transcription termination occurs in a specific context. This context includes the PAS and a GU-rich region 20 nucleotides downstream the PAS (Garrido-Lecca *et al.,* 2016; Porrua and Libri, 2015).

Previous data obtained by 3' rapid amplification of cDNA ends (Boumart *et al.,* 2018) revealed the presence of two PAS in the sequence of ICP22, one minor and one major. The minor one leads to polyadenylation in 5% of the transcripts and is located upstream US10. However, this proportion may increase in the case of some of the constructs described in the next points, if a GU-rich region is brought downsteam the minor PAS. In order to standardize the reporter assays with the large variety of constructs, the cryptic PAS between ICP22 and US10 was removed by site-directed mutagenesis. The ATTAAA sequence was replaced by an ATATAT one. Two PCR fragments were generated to introduce the target mutation. The mutated construct was produced by overlapping PCR gathering mutated amplicons. After sequencing, the electrophoregram showed that mutagenesis of this cryptic PAS was successfully obtained. This initial step brings that reference material to generate reporter construct that will produce bicistronic mRNA in the most straightforward way.

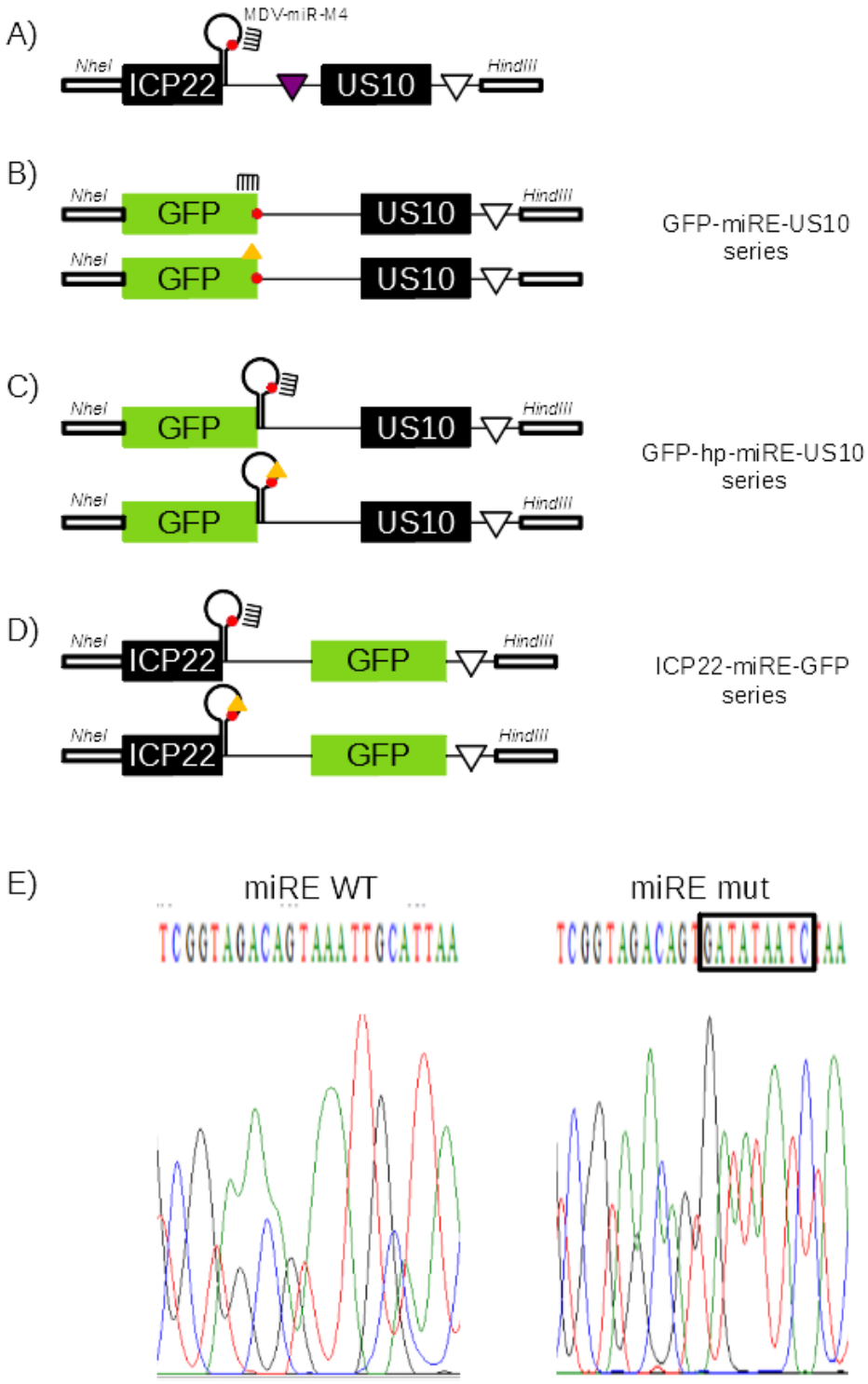


Figure 14 : **Design of the reporter constructs to decipher the mechanism involved in the regulation of ICP22-US10 bicistronic transcript.** A) Representation of the reporter constructs and the involved ORF. The white boxes represent the restriction sites, the black boxes represent the native genes, the green boxes represent the reporter genes, the lines represent the IgR, the yellow triangles represent the mutations, the white reversed triangles represent the canonical PAS, the indigo reversed triangle represent the APAS. B) Electrophoregram confirming the mutation of the miRE in the mutated constructs.

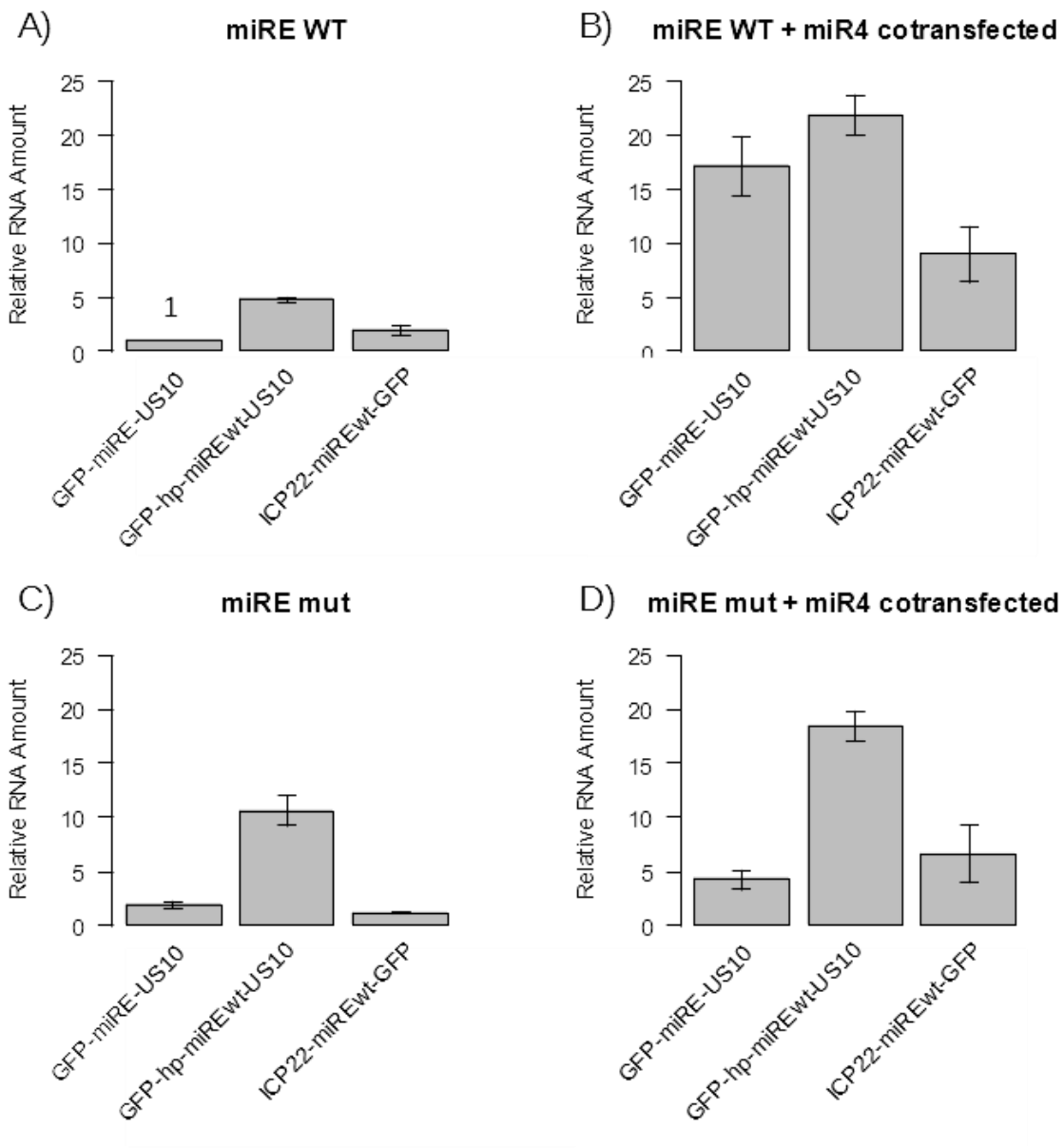
3.3) Establishment of reporter constructs to decipher the mechanisms regulating the translation of the ICP22-US10 bicistronic transcript

Three series of constructs were first established to test the hypothesis. The first constructs were built in order to replace ICP22 with a fluorescent marker, the GFP, that will be used as a reporter in further experiments. The GFP ORF was coupled with the 3'end of ICP22 ORF possessing the miR-4-miRE (miR4 miRNA Response Element) or with the mutated miRE. The PCR fragments amplifying the chimeric GFP-ICP22 3'end (745bp) and the US10 (1117bp) were obtained. Overlapping PCR was used to generate the resulting reporter bicistronic wild-type construct (1827bp). The reporter bicistronic mutant construct (devoid of the miRE) was generated following the same strategy (data not shown). This series of constructs is named hereafter GFP-miRE-US10.

The second series of constructs was produced to investigate the role of the hairpin that was found thanks to bioinformatics analysis. This series was established using the same strategy. However, the US10 fragment encompassed all the nucleotides involved in the structure, adding 45 nucleotides to the amplicon. This construct was then mutated on the miRE following the same strategy. This series is named hereafter GFP-hp-miRE-US10.

In the third series of constructs, we aimed to replace US10 by GFP. The GFP ORF was here coupled with the WT ICP22 ORF or with the ICP22 ORF mutated at the miR-4 binding site. In order to transfer the original major PAS of the bicistron into the reporter constructs, the strategy required to include 3 overlapping fragments, the first one amplifying either WT or mutated ICP22 ORF (845bp), the second one amplifying the GFP ORF (729bp) and the third one amplifying US10 3'UTR including the major PAS (165bp). The final construct was obtained by gathering these overlapping fragments (1702bp). This series is named hereafter ICP22-miRE-GFP.

The whole series of WT and mutated constructs generated above were produced with restriction sites adapted to molecular cloning downstream a CMV promoter. The integration into the same plasmid backbone (pcDNA 3.1 -) was used to overexpress the bicistronic reporter constructs in the presence of in the absence of miR-4 to test the interaction with the activator miR. The sequences of all the reporter constructs were confirmed by Sanger sequencing before transfection. Except the mutations introduced in the miRE to investigate its function, no sequence modifications were observed in the whole series of constructs.



g

Figure 15 : **Transcript quantification of the reporter constructs obtained from the ICP22-US10 bicistronic transcript**. Bar plot representations of the RNA relative quantifications next to lipofection of DF-1 cells with the different constructs. The cells were transfected with 500ng of pcDNA containing the constructs and with or without 5.000ng of pcDNA containing the miRNA. The miRNA response element miRE was mutated or not in the constructs. RNA was extracted 24h post-transfection. RNA abundance was normalized by GAPDH and the RNA abundance of the construct GFP-miREwt-US10 without miRNA cotransfection was set to 1. The data represent the mean of values obtained by a technical triplicate. The error bars represent the standard deviation. **A)** Relative quantification of the miREwt constructs, in the absence of miR4. **B)** Relative quantification of the miREwt constructs, in the presence of miR4. **C)** Relative quantification of the miREmut constructs, in the absence of miR4. **D)** Relative quantification of the miREmut constructs, in the presence of miR4.

3.4) Quantification of the transcript shows both direct and indirect miR4 effects on the amount of mRNA

A RNA quantification assay was designed to determine the importance of the sequence surrounding the stop codon of the ICP22 ORF. Four questions were addressed : 1) May a RNA secondary structure impact the gene expression at the transcriptional level? 2) Does the miRNA have an impact on the RNA abundance? And does the mutation of the miRE turn off this effect? 3) Does the hairpin change the fate of the transcript in the presence of the miRNA? 4) Does the position of the ORF impact the stability of the transcript?

Our data shows a five fold increase of the transcript abundance in the presence of the hairpin in the transcript (figure 15.A.) and a 2 fold increase of the transcript amount following the position of the ORF in the transcript. This result indicates that the presence of the hairpin in two series of the transcript alters positively the mRNA amount in the cell.

We next investigated the effect of miR4 expression. The quantified data showed that miR4 alters positively the amount of mRNA in each condition as revealed by the Figure 15.B.. More precisely, the cells cotransfected with both GFP-miRE-US10 and miR4 displayed a 17 fold increase compared to the non-cotransfected construct. The cells cotransfected with both GFP-hp-miRE-US10 and miR4 or both ICP22-miRE-GFP and miR4 were 5 fold increased compared to the non-cotransfected construct. These results show that miR4 increases the amount of transcript within the transfected cells and that the presence of the hairpin in the transcript sponges the boost triggered by the miRNA (17 fold increase vs 5 fold increase).

Strikingly, the mutation of the miRE showed differential effect on the mRNA amount (Figure 15.C. compared to Figure 15.A.). The GFP-miRE*-US10 construct and the GFP-hp-miRE*-US10 constructs showed a 2 fold increase following the mutation while the ICP22-miRE*-GFP construct was decreased by a factor 2 in the presence of the same mutation. This result suggests that both the miRE and the US10 sequence influenced the fate of the transcript. Surprisingly, adding miR4 to these mutated constructs (Figure 15.D. compared to Figure 15.C.) induced a 2 fold increase in transcript abundance for GFP-miRE-US10 and GFP-hp-miRE-US10 and a 3 fold increase for ICP22-miRE-US10. Taken together these data indicate that miR4 increases the abundance of the transcript even without acting directly on the first predicted miRE.

Following the latter result where an effect of miR4 was observed in the absence of a canonical miRE, it was decided to investigate additional properties of this part of the transcript. Through *in silico* analysis, additional miRNA binding sites were predicted towards to passenger strand produced from premiR4.

3.5) MDV-miR-M4-3p might bind to the bicistronic transcript encoding for both ICP22 and US10

To find any possible interaction between MDV-miR-M4 and the bicistronic transcript, IntaRNA web server was used. First, secondary interactions between miR4-5p and the transcript were looked for without success. Next to this, it was suggested that the small amount of miR4-3p produced from the premiR4 could bind the transcript. Since miR4 is produced thanks to a CMV promoter, it is possible that the premiRNA amount could be sufficient to produce mature miRISC complex incorporating miR4-3p which might interfere post-transcriptionally with the mRNA. Indeed, several miR4-3p binding sites were found in the mRNA at the level of US10 (Figure 16). Three sites were considered as candidates for binding miR4-3p in the bicistronic transcript. They are located at positions 464-486, 969-984 and 1100-1113. The possible interaction might explain in part why the mutation of the miRE does not turn off the effect of miR4 on the bicistronic transcript.

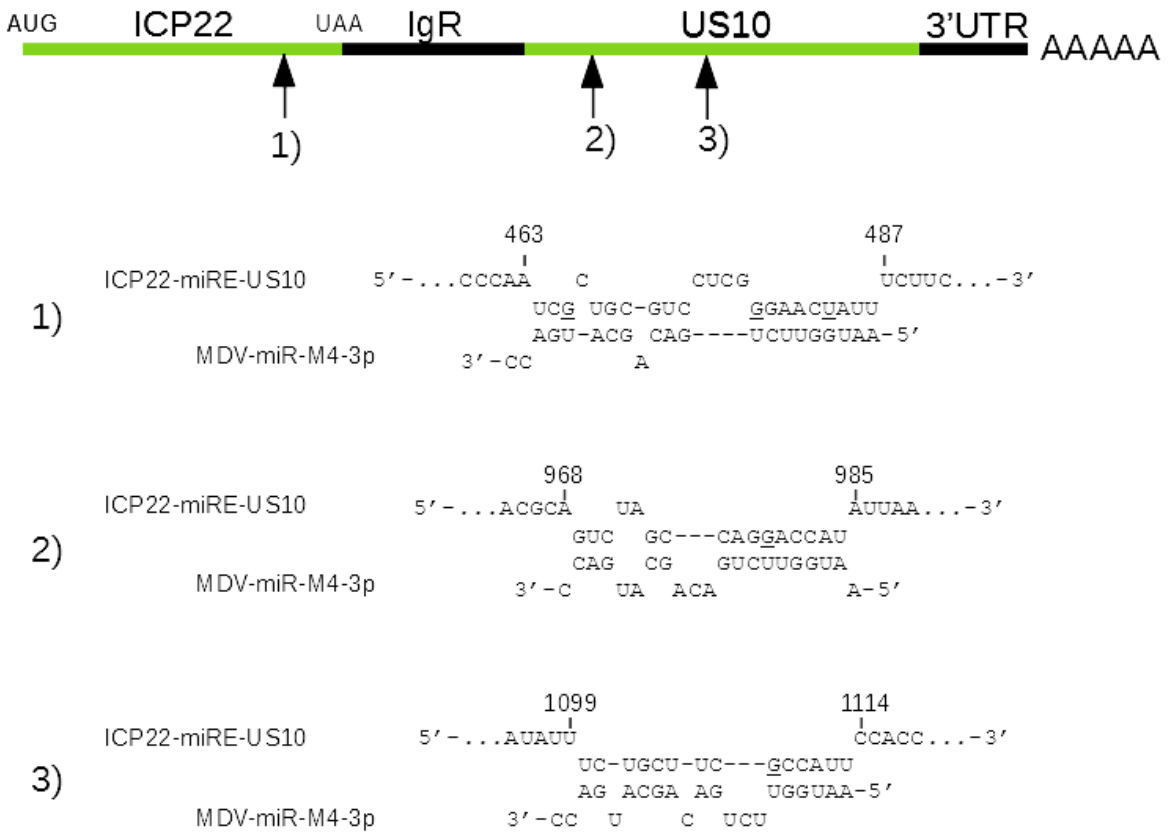


Figure 16: **Predicted miRE for MDV-miR-M4-3p**. The figure shows the positions of the nucleotides on the native bicistronic transcript. The underlined nucleotides are involved in a Wobble interaction with the matching nucleotide.

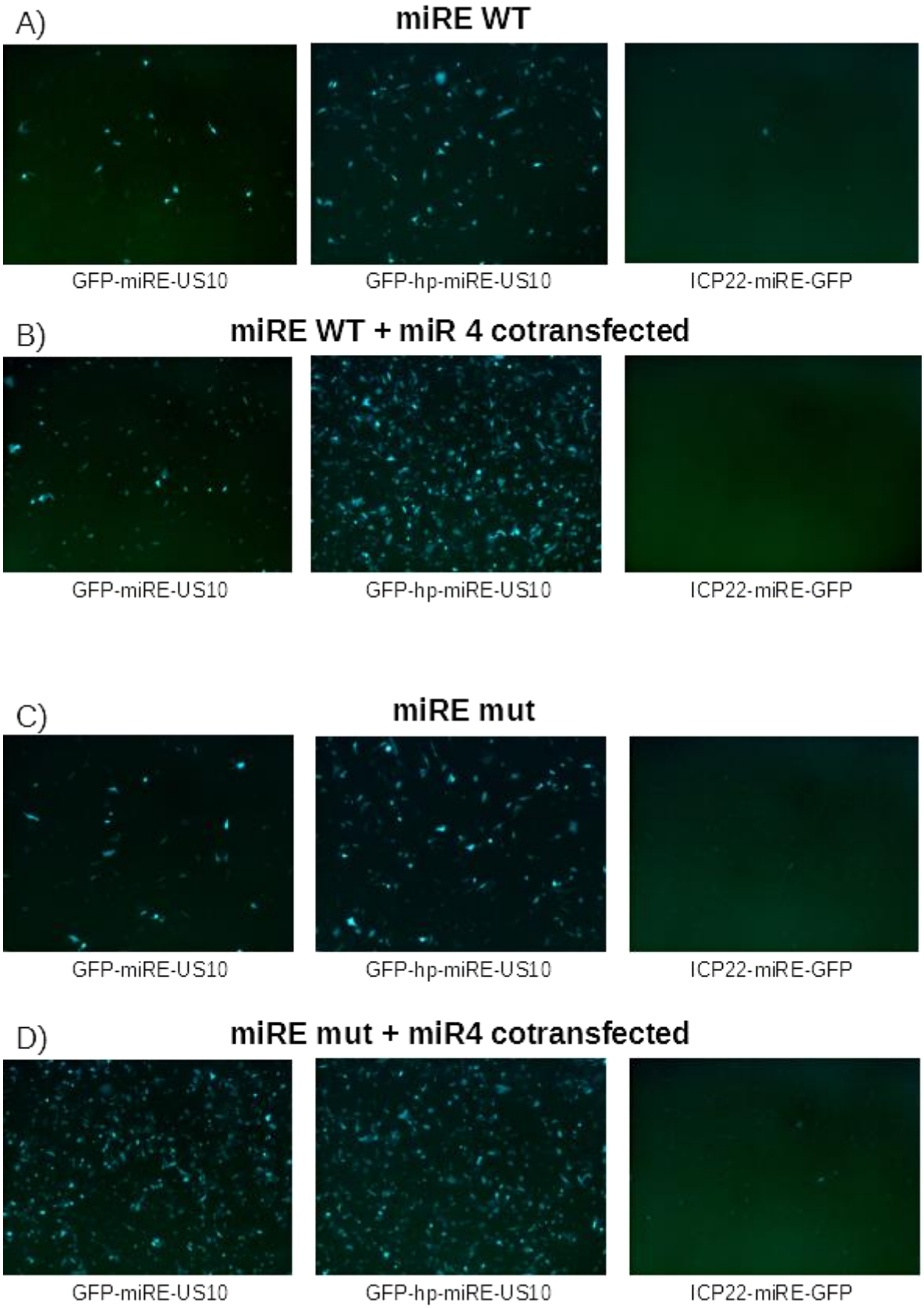


Figure 17: **Fluorescent microscopy of the cells transfected with the reporter constructs obtained from the ICP22-US10 bicistronic transcript**. At 24h post-transfection, constructs were cotransfected within the cells in the presence or in the absence of miR4 and with or without the mutation of the miRE. The figure displays the raw fluorescence of the cells following these (co)transfection.

3.6) Protein expression is determined by several parameters : the sequence, the ORF position and the presence of miR4

Fluorescent microscopy (Figure 17), western blot and flow cytometry (Figure 18) were used to analyze the protein expression from the different reporter constructs. In order to compare the data obtained in transcript quantification, protein expression was evaluated to address the same aspects: 1) the effect of the hairpin and of the position within the bicistronic transcript; 2) the influence of miR4 and 3) the impact of the mutation of the miRE.

Cells were transfected with the whole series of constructs in the presence or in the absence of miR4. Identical cell densities were compared in epifluorescence. To identify the factors influencing GFP expression. Rough comparison of the panels displayed in Figure 17 indicated that the position of the ORF within the transcript deeply impacted the protein expression when GFP replaces ICP22 or US10. Fluorescence was only visible when GFP is positioned in the upstream ORF (Figure 17.A.). The presence of the hairpin increased GFP expression (Figure 17.A.). Addition of miR4 influenced positively the GFP expression with the constructs where GFP replaced ICP22. This effect was only visible in the presence of the hairpin (Figure 17.B.).

Mutating miRE sequence did not reveal any effect in the absence of miR4 expression when Figure 17.C. was compared to Figure 17.A.. However, the mutation of the site revealed a striking increased fluorescence density in the construct where GFP replaced ICP22 devoid of the hairpin (Figure 17.D.).

Since epifluorescence observation is not a quantitative method, it was decided to assess the protein expression from the reporter constructs by using flow cytometry analysis and western blot.

According to protein expression of the cells transfected in the absence of miR4 with the miRE WT constructs (Figure 18.A.), the presence of the hairpin induced a 10 fold increase in the protein expression for the upstream ORF. However, when GFP replaced the downstream ORF, in the ICP22-miRE-GFP construct, fluorescence was slightly detected (20 times less than the GFP-miRE-US10 construct). These data show that both the position in the bicistronic transcript and the presence of an hairpin in the secondary structure were determinant for the GFP translation.

When miR4 was cotransfected with the miRE wt constructs (Figure 18.B.), the protein expression was increased for the tested conditions. More precisely, cells transfected with GFP-miRE-US10, GFP-hp-miRE-US10 and ICP22-miRE-GFP respectively increased the GFP expression by a factor 5, 4 and 6. What is worthy to note is that GFP is still barely expressed in the latter situation. This result demonstrate that the expression of miR4 has a positive effect on the protein expression, independently of both the hairpin and the position of the ORF in the bicistronic transcript.

The mutation of the miRE in the absence of miR4 increased the GFP expression in the whole series of constructs (Figure 18.C.). The two constructs including the hairpin (GFP-hp-miRE*-US10 and ICP22-miRE*-GFP) showed a 1,5 fold increase, meanwhile the construct without the hairpin (GFP-miRE*-US10) showed a 9 fold increase. Adding miR4 to these constructs (Figure 18.D.) increased the GFP expression by a factor 2 for the constructs possessing the hairpin (GFP-hp-miRE*-US10 and ICP22-miRE*-GFP) and by a factor 4 for the sequence devoid of the hairpin (GFP-miRE*-US10).

Altogether, these patterns of protein expression first show that the hairpin sponge a deleterious effect for the translation at the miRE, which is compensated by the mutation of this site. The

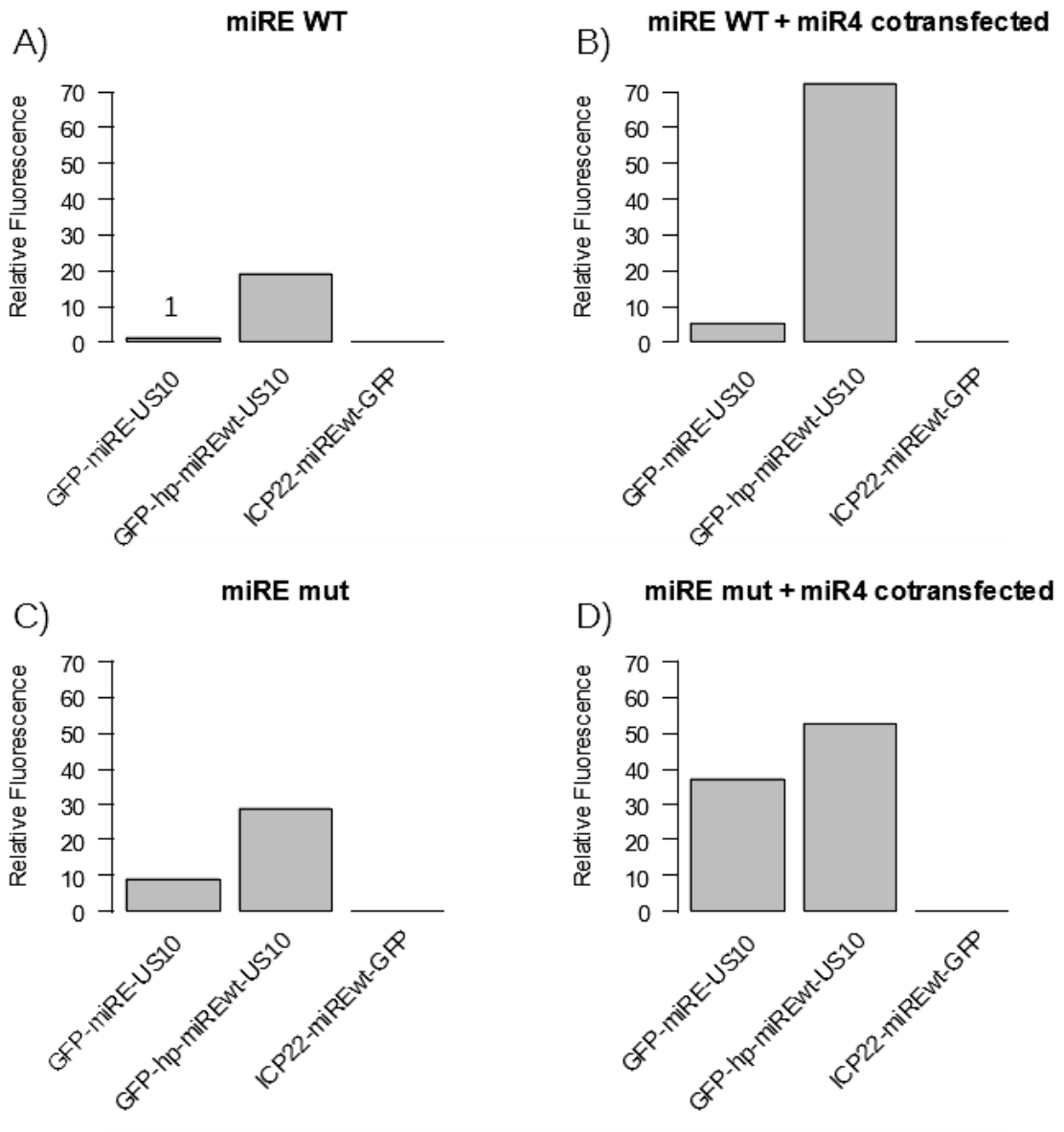


Figure 18: **Single cell analysis of the protein expression associated with the reporter constructs obtained from the ICP22-US10 bicistronic transcript**. Bar plot representations of the fluorescence relative quantifications following transfection of DF-1 cells with the different constructs. The cells were transfected with the constructs in the presence or in the absence of the miRNA. The miRNA response element (miRE) was mutated or not in the constructs. Cells were analysed by flow cytometry 24h post-transfection. Fluorescence of the construct GFP-miREwt-US10 without miRNA cotransfection was set to 1. **A)** Relative fluorescence of the cells next to transfection with miREwt constructs, in the absence of miR4. **B)** Relative fluorescence of the cells next to transfection with miREwt constructs, in the presence of miR4. **C)** Relative fluorescence of the cells next to transfection with miREmut constructs, in the absence of miR4. **D)** Relative fluorescence of the cells next to transfection with miREmut constructs, in the presence of miR4.

results also indicate that miR4 exert either an indirect enhancement of the mRNA translation or a direct enhancement at another unknown miRE.

In order to confirm the unexpected result of GFP expression induced by miR4 when the miRE was mutated, western blot analysis was carried out (Figure 19). This preliminary assay was also designed to monitor the slight expression of the GFP when it is positioned in the downstream ORF. The Figure 15 revealed similar expression pattern as those observed by flow cytometry.

Taken together, these results show again both direct and indirect effects of the miRNA on the mRNA translation. The presence of a hairpin dramatically impacts mRNA translation. Unexpectedly, the mutation of the miRE, in the presence of mIR4 (figure 18.B. compared to 18.D.), increased the expression of the GFP in the transcripts devoid of the hairpin (GFP-miRE-US10) meanwhile it decreased the expression in the sequences containing the hairpin (GFP-hp-miRE-US10 and ICP22-miRE-GFP). This last observation indicates that the binding between miR4 and the hairpin impact directly on the mRNA translation.

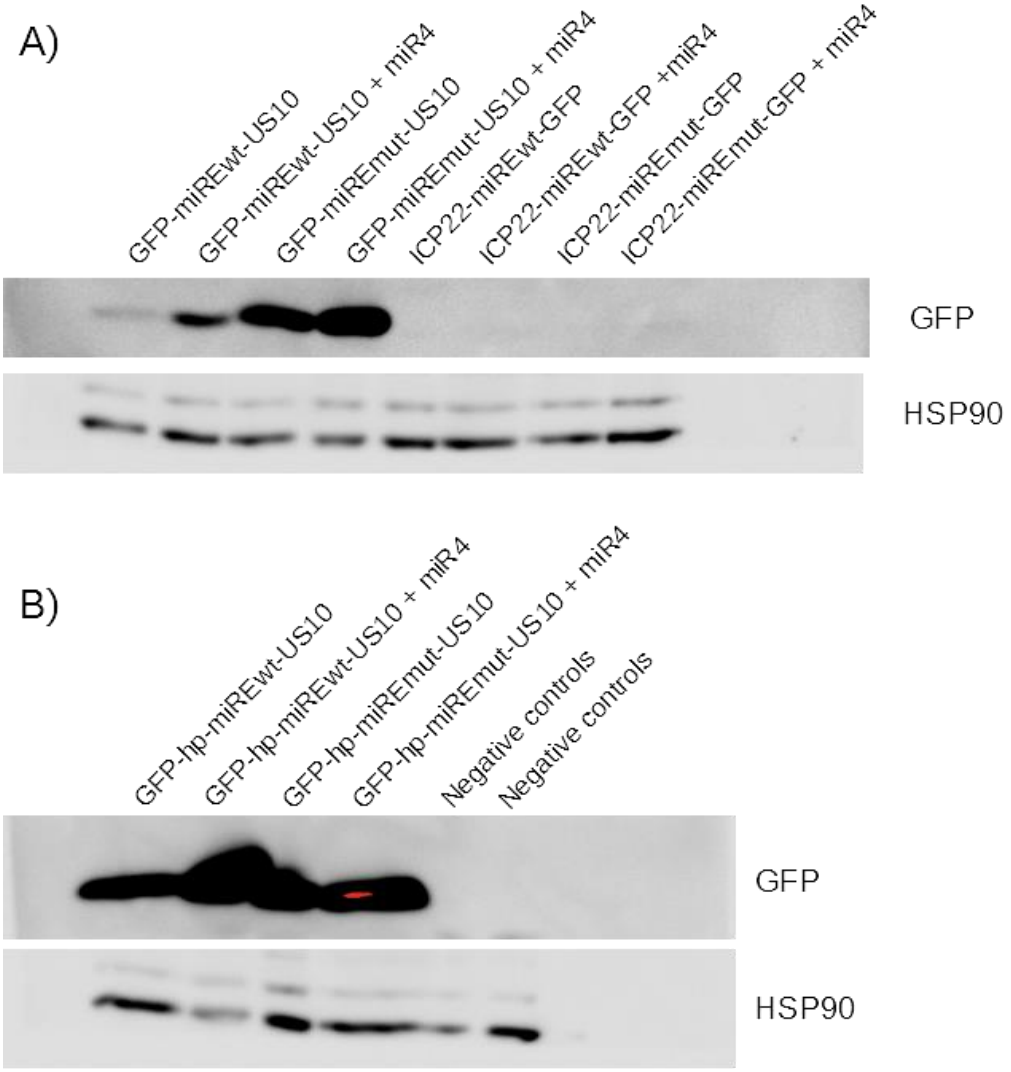


Figure 19: **Western blot analysis of the protein expression associated with the reporter constructs obtained from the ICP22-US10 bicistronic transcript.** A) Analysis of the protein abundance produced after transfection of the GFP-miRE-US10 and ICP22-miRE-GFP series of constructs. 25µg of protein were loaded on each lane. B) Analysis of the protein abundance produced after transfection of the GFP-hp-miRE-US10 series of constructs. 25µg of protein were loaded on each lane. Negative controls are a non-transfected cell line and a cell line transfected with a pcDNA 3.1 (-) encoding for miR4.

3.7) mRNA half-life show a protective effect of miR4 on the bicistronic independent of the binding of miR4 to the miRE

Since the mRNA amount was altered by the presence (or the absence) of miR4 in the cell. It was decided to investigate the effect of miR4 on the mRNA half-life of the bicistronic transcript. Investigating this aspect might show a role of a mRNA decay pathway on this specific mRNA. This decay would be inhibited in the presence of miR4 in the cell. It was decided to test this hypothesis on bicistronic transcripts that encompass the two sequences encoding the native viral proteins (ICP22 and US10). These transcripts were obtained by generating a new series of constructs with the WT or the mutated miRE. A similar cloning strategy was used as described above (see 3.2 and 3.3). The resulting constructs gathering ICP22 and US10 were transfected in the presence or in the absence of miR4. The transcription was inhibited 24h post-transfection and RNA extracted at 4h; 6h; 12h and 25h post-treatment.

The preliminary data show that adding miR4 to the WT and the mutated miRE constructs increased the half-lifes of the mRNA of a factor 1,6. It means that miR4 effectively inhibits a decay pathway involved in the fate of the mRNA but this inhibition is not only dependent on the binding of the miRNA with the predicted miRE.

However, these results must be reproduced and precised by carrying out multiple replicates, since the data seem not precise and since some samples failed to be quantified precisely. The clearly failed samples were removed from the data, having an influence on the results.

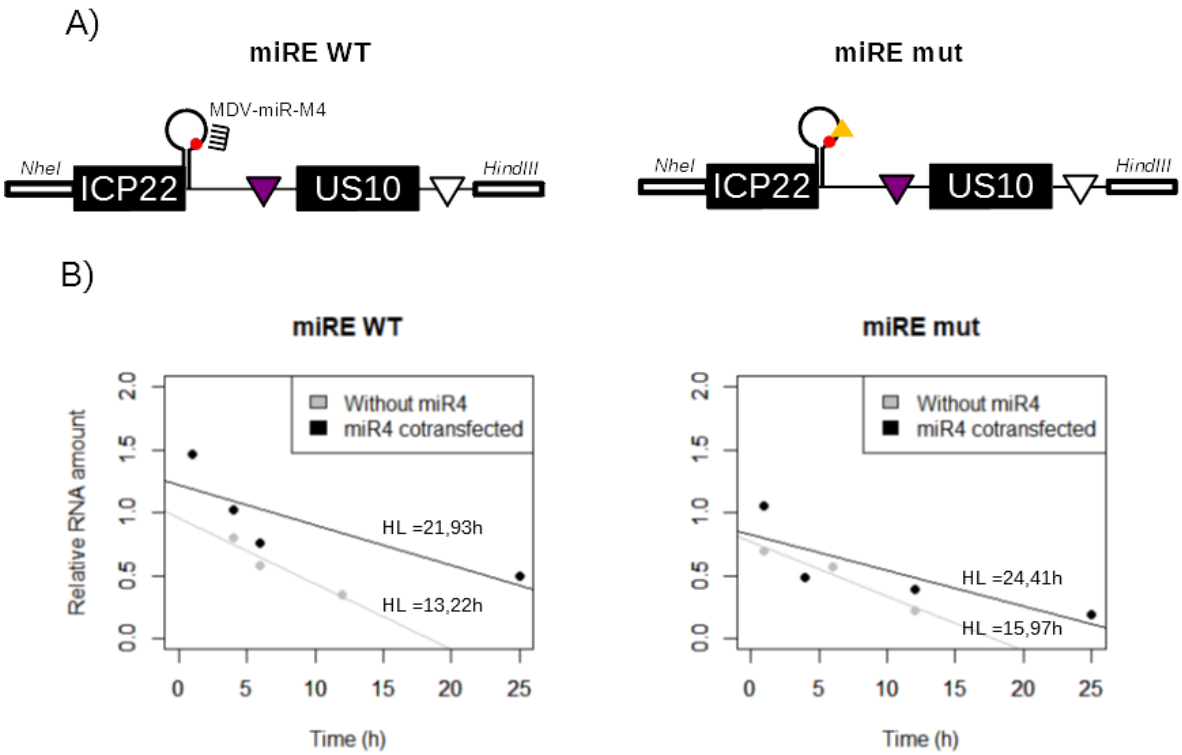


Figure 20: **miR4-mediated alteration of the mRNA half-life**. A) Newly synthetized constructs that were transcfected in the cells, in order to measure the half-life of the bicistronic transcript B) Relative RNA abundance over time after inhibition of the transcription by dactinomycin treatment. Both the constructs ICP22-US10 with or without a mutated miRE were transfected in the presence or in the absence of miR4. The dactinomycin treatment was carried out 24h post-transfection and the RNA was extracted at 1;4;6;12;25h post-treatment. The RNA amount was standardized using GAPDH (considered and confirmed as a stable transcript) and the ΔΔCt was calculated using the non-treated control at the first time-point. The R^2 of the linear regression are of 0,9351 for miRE WT without miR4; 0,6935 for miRE WT + miR4 cotransfected ; 0,9574 for miREmut without miR4 ; 0,6752 for miREmut + miR4 cotransfected.

4) Discussion

Understanding how gene expression is regulated is crucial since dysregulation of the gene expression is a hallmark of cancer. Indeed, every cancerous process displays a gene expression pattern with increased proto-oncogenes and decreased the tumor suppressor genes are decreased. Investigating such phenomenon in a model of virus-induced lymphomagenesis offers several advantages. In this context, we worked on the regulating processes of a specific bicistronic transcript encoding US10 and ICP22 proteins, the latter being known to play a role in tumorigenesis. ICP22 gene expression was shown to be positively impacted by a miRNA.

By using GFP reporters, it was aimed to decipher and to understand the whole pathway involved in the miRNA dependent increase of protein expression from a bicistronic transcript. Our data bring new findings that characterize the regulation of the expression of the viral genes encoded by the bicistronic ICP22-US10 transcript. The first insights were brought by Boumart and collaborators (Boumart *et al.*, 2018) when they proved the involvement of miR4 in the upregulation of the ICP22 translation following the binding with the stop codon of the ICP22 ORF. In their constructs, they only explored the regulation in a monocistronic model. This model revealed an essential role of miR4 in turning on ICP22 expression but did not allow to describe functions for each part of the bicistronic transcript. An original series of bicistronic reporters were set up to explore the mechanisms of expression through a diversified approach. By comparing both the transcript and the protein amounts in the different tested conditions, this approach permitted investigating the roles of the different portions of the transcript.

We revealed through transcript quantification the importance of the stop codon-associated hairpin, of the stop codon-associated miRE and of the concomitant expression of miR4. Regarding the hairpin, it was shown to increase the transcript amount. The presence of miR4 was confirmed to increase the mRNA amount within the cells while the mutation of the miRE had variable effect according to specific sequence arrangement. The position of the GFP ORF was shown to have an impact on the transcript abundance. This effect was associated with the presence or the absence of the US10 ORF. This observation is supported by the presence of two potential miRE for miR4-3p in the sequence of US10. It is likely that the binding of miR4-5p at the first miRE and the binding of miR4-3p to a secondary miRE triggers the formation of a complex able to inhibit a decay pathway.

Reporter constructs were designed to include or to exclude a highly structured hairpin within the bicistronic RNA. Data from RNA and protein quantifications pointed out the importance of the hairpin both in the abundance and the fate of the transcript. It is suggested that, more than just protecting directly the mRNA, miR4 targets directly mRNA decay factors, having thereby an indirect positive effect on the bicistronic transcript amount. This observation is confirmed by quantification of the mRNA over time after inhibiting the transcription with dactinomycin since miR4 induced an increase of the half-life of the mRNA even in the absence of the canonical miRE.

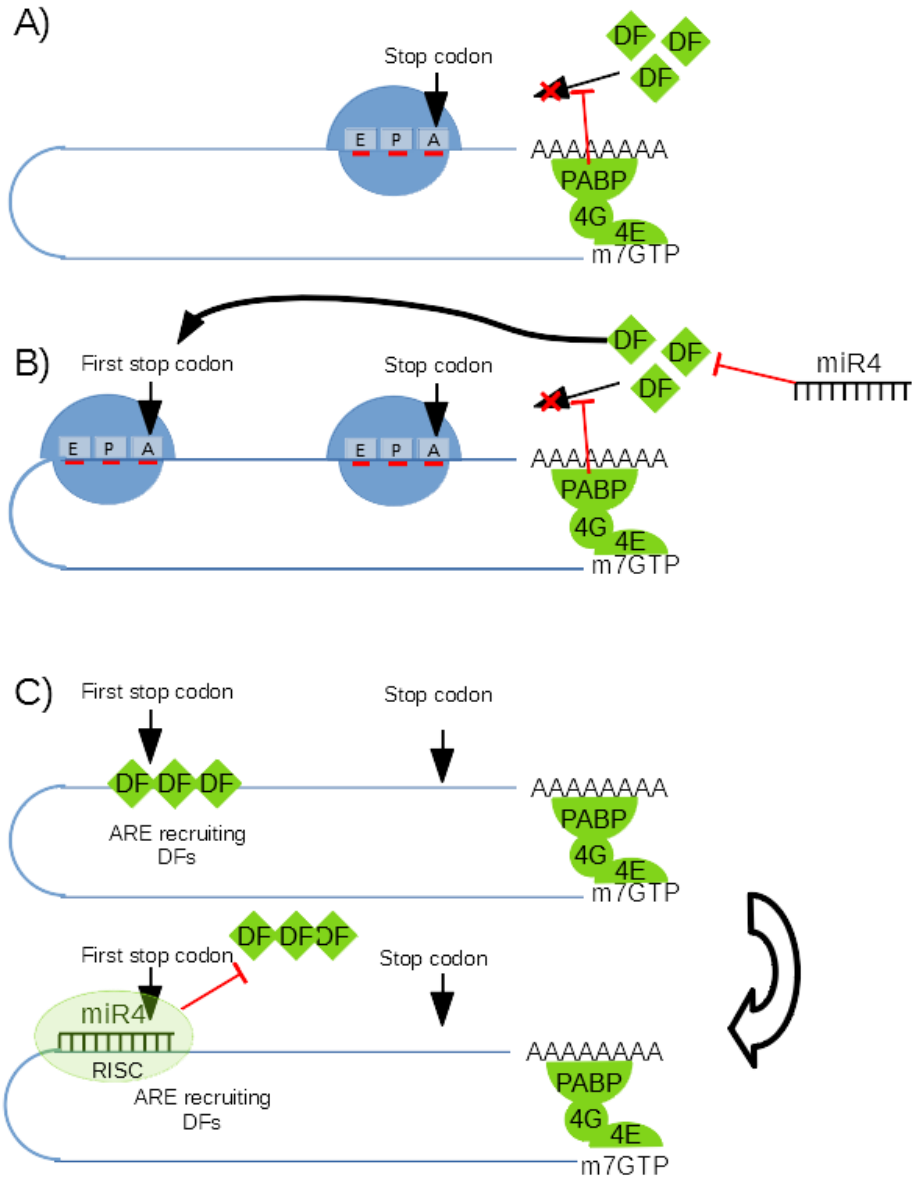


Figure 21 : **Suggested models of the miR4-induced upregulation of a bicistronic transcript.** DF = Decay Factors A) Canonical termination of the translation, as explained in Figure 3. B) Model of an indirect miR4-induced effect on the transcript. In this case, miR4 would inhibit some decay factors that impair the transcript amount. C) Model of a direct miR4-induced effect on the transcript. In this case, the binding of the miRISC complex would inhibit the recruitment of decay factors (such as Upf-1) at a site located far from the poly-A tail and PABP, the canonical inhibitor of Upf-1.

4.1) Suggested models to explain miRNA-dependent increase of the gene expression

Based on our data and literature, a series of non-exclusive models are suggested to explain the miR4 dependent enhancement of the gene expression. These models are developed and illustrated by previously published examples. The first model (Figure 21.B.) relies on the inhibition of decay protein by miR4. In a GaHV-2 infection, Tahiri-Alaoui and colleagues (Tahiri-Alaoui *et al.*, 2014) already demonstrated that kind of effect following indirect miRNA-mediated increase of the expression. In this example, the pp14 polyprotein was increased by the miRNAs produced from the LAT miRNA cluster. They showed how the miRNAs increased the recruitment of PABP to a transcript to increase the cap-independent translation. More precisely, the LAT-derived miRNAs were targeting the paip2-encoding transcript. Paip2 interferes negatively with PABP. By inhibiting this inhibitor, in a miRNA-mediated way, the total amount of available PABP increase, upregulating therefore the PABP-dependent translation. Based on this example, the model that is suggested is likely to happen since we proved an indirect effect of miR4 on the transcript stability. However, this indirect mechanism is not exclusive since we also proved another direct effect. Following this idea, several models were developed to provide with a comprehensive lecture of all the mechanisms that are likely to be involved in the miR4-dependent enhancement of the gene expression. Considering that each model might be associated with one or several alternative model, this offers a huge combinatory series of events that need to be investigated and confirmed one by one.

Stabilization of a transcript by a miRNA is already a known mechanism. An example is the TTP mechanism (Ma *et al.*, 2010), in which a miRNA protect a mRNA by binding to an ARE previously targeted by the TTP. This protein is unfortunately absent in birds (Lai *et al.*, 2013), including the chicken, and therefore cannot be the key of the miR4-mediated upregulation of the transcript. Nevertheless, this protein is an important anti-inflammatory factor, that could have a homolog in the chicken organism. Thus, this mechanism should not be discarded in the case of the miR4-mediated upregulation of the bicistronic transcript. To support this hypothesis, the bioinformatics analysis showed the presence of an ARE overlapping the stop codon of the ICP22 ORF. The second and third models are based on this ARE-identified pathway. Concerning the second model (Figure 21.C.), it assumes that the ARE is an instable part of the bicistronic transcript targeted by a protein homolog or not to TTP and that the binding of this protein is inhibited by the miRNA. During the viral infection, miR4 would synergize with ICP27 to exert the positive effect. Indeed, ICP27 has been shown to induce factors that stabilize ARE-containing transcript (Corcoran *et al.*, 2006) and was also shown to increase the expression of ICP22 (Boumart *et al.*, 2018).

The HCV illustration of miRNA-induced stabilization of an RNA molecule was described in the introduction of this master thesis (see 1.2). It was demonstrated that two miRNA strands, bound at two sites in the vicinity of a hairpin, stabilized the structure of the targeted RNA, insuring thereby its whole integrity. In the third illustrated model (Figure 22), we suggest a role of both miR4-5p and -3p on the stability of the transcript. This model is suggested to happen through three interactions : between miR4-5p and ICP22 ORF ; between miR4-3p and US10 ORF and between both miR4-5p and -3p. These three interactions might impact the secondary structure of the bicistronic transcript, leading to a supercoiled mRNA which would not be accessible for decay factors at the ARE level. It is important to note that this model does not exclude that another important structure might impact the conformational shift.

The second and third model share a common feature since they are likely to happen within the nucleus, at a early post-transcriptional level, impairing further binding of any decay factors on the mRNA.

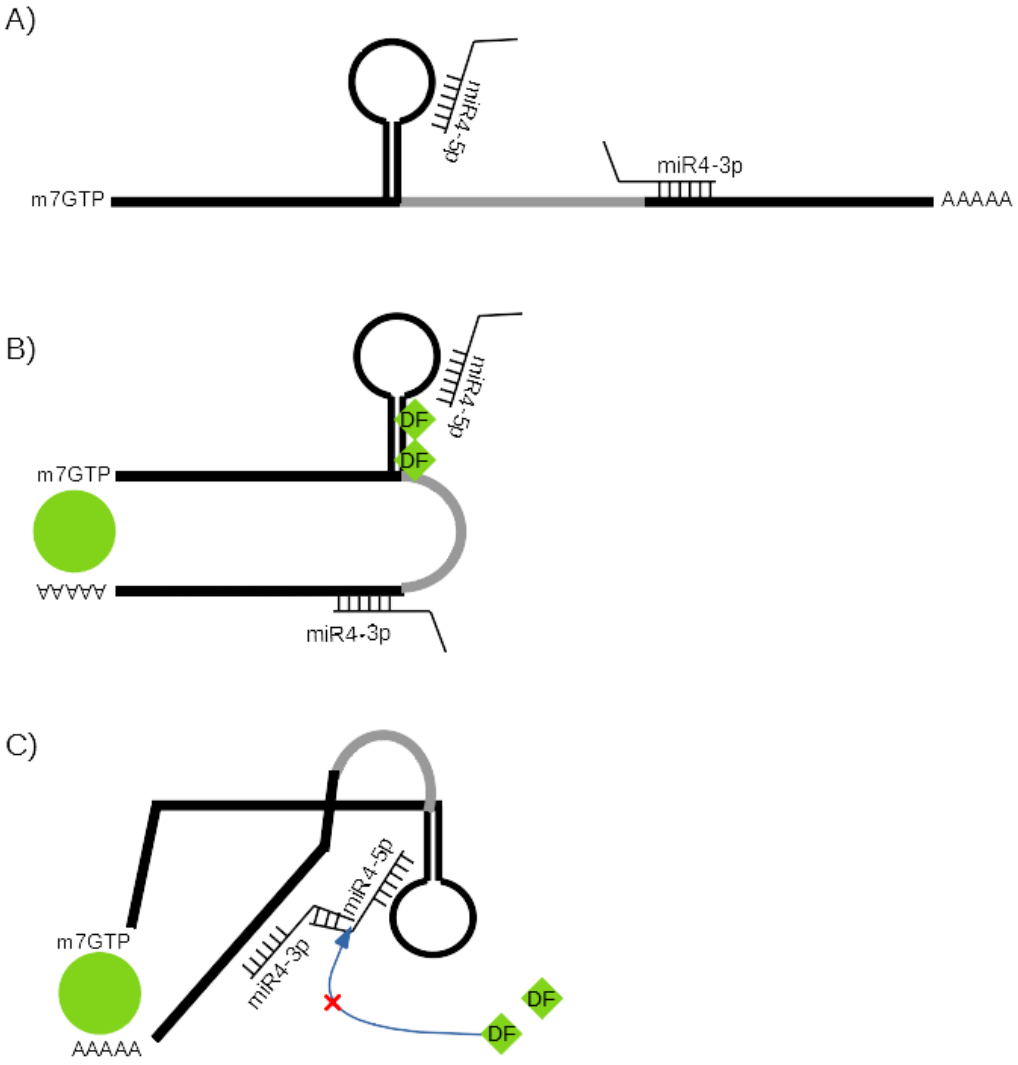


Figure 22 : **Suggested models of the miR4-induced upregulation of a bicistronic transcript (continuation).** DF = Decay Factors. ORFs are represented by black lines, intergenic sequence are represented in grey and proteins are represented in green. **A)** Example of a bicistronic transcript binding by a miRNA at two different sites and displaying a hairpin at the end of the first ORF. **B)** Canonical conformation of the same bicistronic transcript in the absence of miRNA-miRNA interaction. DFs are supposed to target the sequence nearby the miRNA target. **C)** Conformational change in the presence of an interaction between the two miRNA, protecting hampering the binding of the DFs to their target sequence.

In 2017, Wakiyama and collaborators (Wakiyama *et al.*, 2018) described an enhancement of the protein expression illustrated by using of a luciferase reporter system. This increase was mediated by an interaction between PABP and TNRC6, which take part in the RISC complex. The particularity of this example is that the expression boost happened only in the absence of a poly-A tail in the luciferase transcript and in an IRES-dependent translation. In the case of the ICP22-US10 bicistronic transcript, it was proven by a poly-T dependent 3'RACE that the transcript is polyadenylated by the US10 3'UTR. However, since PABP is able to interact with Ago, a model (Figure 23) is suggested to explain the miR4-induced enhancement of the gene expression. In this model, the binding of the miRNA with the first stop codon would affect the secondary structure of the bicistronic transcript by establishing an interaction between PABP and the stop codon. This interaction could either inhibit decay factors, such as Upf-1 which was proven to be inhibited by PABP, or increase the translation like in the described example.

The upregulation might also result from a transcriptional effect. saRNAs are a class of miRNAs that were shown to be functional within the nucleus after binding generally to a promoter to increase the transcription following the recruitment of transcription factors. The first known example of saRNA was described in 2008 (Place *et al.*, 2008). By targeting the e-cadherin gene with miRNA matching with 3'UTR, in order to knock-down the protein that they were studying, Place and collaborators have experienced one of those serendipity effects in research. Indeed, the candidate miRNA increased rather than decreased the e-cadherin expression. By investigating the process underlying the unexpected activation, the authors showed a strong match between the sequence of both the miRNA and the e-cadherin promoter. In the case of the ICP22-US10 bicistronic transcript, the miRE is located at the end of the ICP22 ORF. To our knowledge, a saRNA was never shown to exert its function at a distal site from the promoter. Moreover, the presence of a functional RNA hairpin involved in the upregulation does not provide with compelling evidence that the regulation occurs at the transcriptional level. Nevertheless, this hypothesis may be a fragment of the answer since the hairpin is not the only part involved in the regulation. Furthermore, it is not known if miR4 could target the CMV promoter that is used in the plasmid backbone that was used to clone the whole diversity of constructs. Preliminary data found one site susceptible of being targeted by miR4 in the CMV promoter but the predicted interaction is rather weak when compared with the given example.

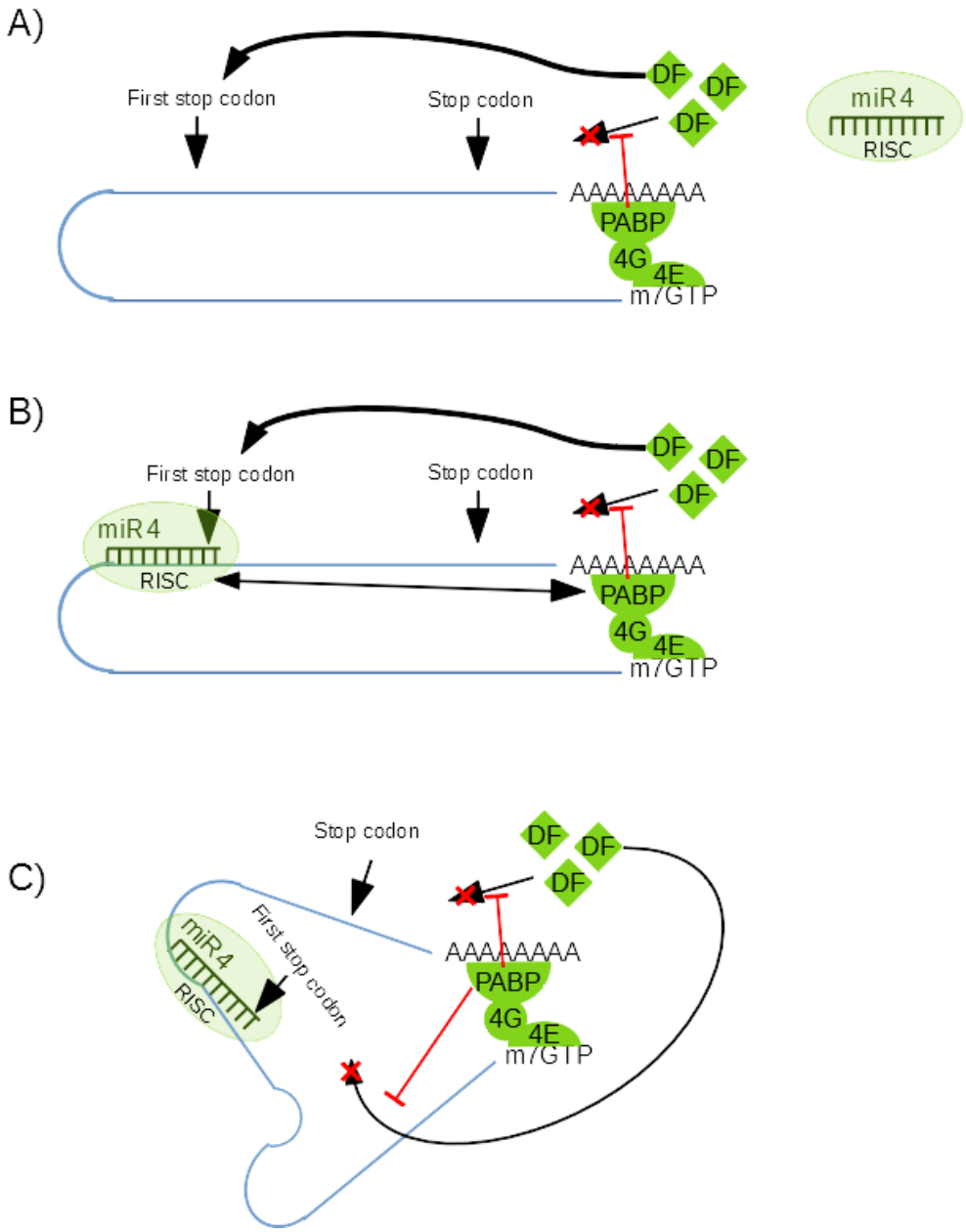


Figure 23 : **Suggested models of the miR4-induced upregulation of a bicistronic transcript (end). A)** Inhibition of the gene expression at a transcript level, induced at a premature termination codon by decay factors, such as Upf-1. **B)** In the presence of a miRNA at the first stop codon, the miRISC complex interacts with PABP through the TNRC6 protein. **C)** The interaction between RISC and PABP induce a conformational change that bring the premature stop codon in the neighborhood of PABP, which inhibits the decay factors.

4.2) miRNA dependent increase of the gene expression involves interaction with secondary RNA structure

In the initial description of the miR4-dependent ICP22 expression, another mechanism was suggested. The idea was that the ICP22-US10 bicistronic transcript was degraded through NMD in the absence of miR4. In this case, it is assumed that the stop codon of the ICP22 ORF is considered as a PTC by the translation machinery, recruiting thereby Upf-1. This idea is supported by several sequence features. First, in the case of a bicistronic transcript, the 3'UTR of the first ORF is pretty long, since it encompasses a second ORF. A recent study from May and collaborators (May *et al.,* 2018) described a mechanism used by a RNA virus to resist against NMD. Because of their polycistronic constitution, most of the RNA viruses display long 3'UTR associated to the ORFs localized near the 5' end of the unique viral transcript and are assumed to be targeted by NMD when infecting an eukaryotic cell. Most of their strategies to escape NMD are still to be described. In the case of Turnip Crinkle Virus (TCV), the 3'UTR has an unstructured sequence right after the stop codon. When an hairpin replaced this unstructured region, NMD was activated, impairing thereby the viral replication. This fact might explain why a bicistronic transcript like gK-ICP27 seems not to be targeted by NMD (Strassheim *et al.,* 2016). Indeed, bioinformatics analysis revealed no specific structures in the regions surrounding the stop codon of the gK ORF. In the case of the ICP22-US10 bicistronic transcript, the bioinformatics analysis revealed a highly structured hairpin. It was first assumed that this hairpin was unfolded by miR4 and the RISC, leading to a resistance against NMD but our data showed a miR4-mediated increase of the fluorescence even in the absence of the miRE. It is possible that the hairpin structure might be hampered by the mutation, but bioinformatics analysis (data not shown) did not reveal any change.

A sensitive question refers to the fact that miR4 expression, in the case of a binding to the bicistronic transcript, does not lead to Ago-RISC-mediated inhibition meanwhile the vast majority of the literature describe the inhibitory effects mediated by the miRNAs. It was recently described a role of the FAM120A protein which was associated to polynucleotide regions. This protein was identified in up to one third of all known miRNA transcripts.The presence of this protein (FAM120A) was demonstrated to decrease the effect of the miRNAs on their target. In the case of ICP22-US10, a poly-C site was found in the intergenic region of the transcript. FAM120A might bind with this poly-C stretch decreasing thereby the deleterious effect of RISC. Moreover, the involvement of an hairpin in the regulation of the bicistronic transcript suggest that this structure might act as a recruitment platform to catch proteins that would impair RISC activities. Another possibility is that this structure would be able to maintain decay factors at too long distance to exert their effect on the transcript. A described example of secondary structures that sponge decay proteins comes from RNA viruses belonging to the Flaviviridae family (Michalski *et al.,* 2019). More precisely, Zika virus and dengue virus both produce a subgenomic flavivirus RNA (sfRNA) that is strongly structured. It was shown that these specific RNA recruit a lot of proteins involved in decay mechanisms that are usually deleterious for the RNA genome of the virus. It is suggested that the hairpin observed in the bicistronic transcript embedding ICP22 and US10 might act the same way than the sfRNA by recruiting decay factors, inhibiting thereby their activities.

4.3) Optimization process to increase significance of assays that were carried out during the master thesis

The very first step of this project was to analyze widely the features of the mRNA, using a bioinformatics approach. At first glance, a highly structured hairpin was predicted *in silico*, this candidate structure becoming an important part of the global project. Nonetheless, RNA folding was not demonstrated through experimental approach neither *in vitro* nor *in vivo*. It cannot be excluded that the observed effects in the presence of the hairpin sequence depends more on the primary structure than the secondary one. There should be a new series of constructs to mutate the sequence silently, without impacting the hairpin, assessing thereby the role played either by the nucleotide sequence or by the RNA folding in the regulation of the gene expression.

As first wet lab experiments, an alternative PAS was mutated in the intergenic region of the bicistronic transcript. It was aimed to avoid background noise associated with monocistronic transcripts (5% of the total mRNA produced from the native ICP22-US10 transcriptional unit). However, even if the mutation was a success, it was never proven that it has really an effect on the polyadenylation process at this position. Further experiments should be carried out to confirm the absence of monocistronic transcript embedding only GFP. Nevertheless, qPCR results at least confirm the presence of bicistronic transcripts since the primers were designed on the GFP sequence. Since amplicons were detected when it was positioned in the downstream ORF, it indicates that the reporters mutated at the alternative PAS produce full length mRNAs.

Another way to increase the reliability of our results, based on an artificial model, is to change the selected reporter system with another one, such as luciferase in future constructs. At first, it was decided to use GFP. This reporter allows performing single cell analysis (Figure 18) and direct observation (Figure 17) by fluorescent microscopy. It was also convenient since the preliminary results (Boumart *et al.*, 2018) indicated a miR4-induced "on/off" effect. In addition the sequence length corresponding to GFP ORF is more representative of the two native genes encoded by the bicistronic transcript. However, quantification using GFP remains less precise than luciferase. Since it is now obvious that reporter luciferase constructs fit better for studying the miR4-mediated regulation, a new series of reporters would be established by using the renilla luciferase as reporter gene. It would allow having a more sensitive tool to measure the effect of both the hairpin and miR4 on the translation.

In the results and discussion, it was concluded that the position of the ORF in the sequence alter its expression since GFP was less expressed when placed at the second position of the transcript. However, when the transcript encompass both ICP22 and GFP, we predict a strong expression of ICP22. This protein was shown to impair the transcription with all the tested promoters (Boumart *et al.*, 2018). The precise mechanism still has to be described but is likely to involve the inhibition of the RNAPII. Therefore, the impairment of ICP22 on the transcription, including on the CMV promoter, would inhibit the expression of the whole transcript. Since GFP-hp-miRE-US10 and ICP22-miRE-GFP constructs are imbalanced in terms of ICP22 expression, this introduces an interpretation bias. Indeed, the lower amount of ICP22-miRE-GFP transcript compared to the GFP-hp-miRE-US10 transcript might be attributed to the presence of ICP22 only in the first situation. To investigate the roles of the hairpin and the miR4-miRE interaction without ICP22 interference, construct design would substitute both ORF with discriminative reporters.

Eventually, since the constructs were available 6 weeks before completing this manuscript, experiments were realized as single assays. Even if the reproducibility has not been tested yet, we observed dramatic effects induced by the presence of the hairpin and of miR4 expression in the different readouts (RT-qPCR, western blot, flow cytometry, epifluorescence, mRNA half-

life assessment). Data should be analyzed with appropriate statistical tests including biological replicates. At the present time, results are representative of technical replicates (qPCR).

4.4) Conclusion and perspectives

On the opposite of the linear view of the steps involved in the gene expression, reporter constructs designed from a viral bicistronic mRNA pointed out the importance of secondary structures as well as the positive influence of miRNA in the fate of a mRNA. Viruses are therefore adequate organisms to study molecular mechanisms that tightly regulate the gene expression. Three major findings emanated from this project: 1) a RNA hairpin increases the gene expression of an important viral protein; 2) the mRNA of the same gene is targeted by a miRNA, increasing its abundance and translation and 3) both direct and indirect effect of this miRNA were demonstrated. Altogether these data shed light at the different levels where gene expression might be regulated by RNA secondary structure or by miRNAs. To explore each of these levels, several non-exclusive models were suggested.

To decipher the whole pathway that was described in this master thesis, the first step would be the inhibition of some mentioned decay pathways. For example, the use of a competitive inhibitor of Upf-1 in the presence of the construct would increase the expression of the GFP in the absence of miR4 if this pathway is behind the miR4-mediated regulation.

The next step would be the use of the RaPID (RNA–Protein Interaction Detection) method (Ramanathan *et al.,* 2018) to understand the precise role of the hairpin in the involved pathway. This mechanism involves the generation of a new construct in which the hairpin would be surrounded by two structures linked to the BirA protein. This protein biotinylates all the surrounding protein which are therefore allowed to be purified using streptavidin-coupled beads. The analyze by MS/MS spectrometry would next allow identifying candidates that might interfere with the involved pathway.

To finalize these results, establishing recombinant viruses with the same features than the presented constructs would bring a confirmation of the existence of the pathway in the natural infection context.

This fact could be taken in account in the engineering of new vaccine strains. More precisely, it was never tried to overexpress a deleterious protein, to attenuate the virulence of a virus. We also suggest a new genetically engineered type of vaccine that would encompass a protein that inhibit latency, such as ICP4. In this case, the hairpin and miR4-miRE would be added to the ICP4 sequence to increase its expression during latency. This could inhibit the virus escape from the immune system and allow the chicken to develop antibodies against the disease before it turns out to induce lymphoma.

5) References

Ameres, S.L., and Zamore, P.D. (2013). Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol *14*, 475–488.

Amor, S., Strassheim, S., Dambrine, G., Remy, S., Rasschaert, D., and Laurent, S. (2011). ICP27 protein of Marek's disease virus interacts with SR proteins and inhibits the splicing of cellular telomerase chTERT and viral vIL8 transcripts. Journal of General Virology *92*, 1273–1278.

Behrens, S.E., Tomei, L., and De Francesco, R. (1996). Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. The EMBO Journal *15*, 12–22.

Bertzbach, L., Pfaff, F., Pauker, V., Kheimar, A., Höper, D., Härtle, S., Karger, A., and Kaufer, B. (2019). The Transcriptional Landscape of Marek's Disease Virus in Primary Chicken B Cells Reveals Novel Splice Variants and Genes. Viruses *11*, 264.

Bertzbach, L.D., Laparidou, M., Härtle, S., Etches, R.J., Kaspers, B., Schusser, B., and Kaufer, B.B. (2018). Unraveling the role of B cells in the pathogenesis of an oncogenic avian herpesvirus. Proceedings of the National Academy of Sciences *115*, 11603–11607.

Boerneke, M.A., and Hermann, T. (2015). Ligand-responsive RNA mechanical switches. RNA Biology *12*, 780–786.

Boodhoo, N., Gurung, A., Sharif, S., and Behboudi, S. (2016). Marek's disease in chickens: a review with focus on immunology. Vet Res *47*, 119.

Boumart, I. (2018). Régulation transcriptionnelle et post-transcriptionnelle du gène US1 codant pour la protéine ICP22 du virus de la maladie de Marek. Université de Tours.

Boumart, I., Figueroa, T., Dambrine, G., Muylkens, B., Pejakovic, S., Rasschaert, D., and Dupuy, C. (2018). GaHV-2 ICP22 protein is expressed from a bicistronic transcript regulated by three GaHV-2 microRNAs. Journal Of General Virology *99*, 1286–1300.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of MicroRNA–Target Recognition. PLoS Biol *3*, e85.

Celik, A., Kervestin, S., and Jacobson, A. (2015). NMD: At the crossroads between translation termination and ribosome recycling. Biochimie *114*, 2–9.

Chandra, Y.G., Lee, J., and Kong, B.-W. (2012). Genome sequence comparison of two United States live attenuated vaccines of infectious laryngotracheitis virus (ILTV). Virus Genes *44*, 470–474.

Corcoran, J.A., Hsu, W.-L., and Smiley, J.R. (2006). Herpes Simplex Virus ICP27 Is Required for Virus-Induced Stabilization of the ARE-Containing IEX-1 mRNA Encoded by the Human IER3 Gene. Journal of Virology *80*, 9720–9729.

Davison, A.J. (2002). Evolution of the herpesviruses. Veterinary Microbiology *86*, 69–88.

Davison, F., and Nair, V. (2004). Marek's disease: an evolving problem.

De la Pena, M. (2003). Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. The EMBO Journal *22*, 5561–5570.

DeLuca, N.A. (2011). Functions and Mechanism of Action of the Herpes Simplex Virus Type Regulatory Protein, ICP4. In Alphaherpesviruses : Molecular Virology, (UK), p.

Elbarbary, R.A., Miyoshi, K., Myers, J.R., Du, P., Ashton, J.M., Tian, B., and Maquat, L.E. (2017a). Tudor-SN–mediated endonucleolytic decay of human cell microRNAs promotes G $_1$ /S phase transition. Science *356*, 859–862.

Elbarbary, R.A., Miyoshi, K., Hedaya, O., Myers, J.R., and Maquat, L.E. (2017b). UPF1 helicase promotes TSN-mediated miRNA decay. Genes Dev. *31*, 1483–1493.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in. *391*, 6.

Fischer, S.E.J. (2015). RNA Interference and MicroRNA-Mediated Silencing: RNAi and MicroRNA-Mediated Silencing. In Current Protocols in Molecular Biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. (Hoboken, NJ, USA: John Wiley & Sons, Inc.), pp. 26.1.1-26.1.5.

Fontaine-Rodriguez, E.C., Taylor, T.J., Olesky, M., and Knipe, D.M. (2004). Proteomics of herpes simplex virus infected cell protein 27: association with translation initiation factors. Virology *330*, 487–492.

Garrido-Lecca, A., Saldi, T., and Blumenthal, T. (2016). Localization of RNAPII and 3′ end formation factor CstF subunits on *C. elegans* genes and operons. Transcription *7*, 96–110.

Gennart, I., and Muylkens, B. (2018). Study of the epigenetic silencing of a host microRNA in Marek's disease. Université de Namur.

Gennart, I., Coupeau, D., Pejaković, S., Laurent, S., Rasschaert, D., and Muylkens, B. (2015). Marek's disease: Genetic regulation of gallid herpesvirus 2 infection and latency. The Veterinary Journal *205*, 339–348.

Haertle, S., Alzuheir, I., Busalt, F., Waters, V., Kaiser, P., and Kaufer, B.B. (2017). Identification of the Receptor and Cellular Ortholog of the Marek's Disease Virus (MDV) CXC Chemokine. Front. Microbiol. *8*, 2543.

Hartenian, E., and Glaunsinger, B.A. (2019). Feedback to the central dogma: cytoplasmic mRNA decay and transcription are interdependent processes. Critical Reviews in Biochemistry and Molecular Biology 1–14.

Hildebrandt, A., Brüggemann, M., Rücklé, C., Boerner, S., Heidelberger, J.B., Busch, A., Hänel, H., Voigt, A., Möckel, M.M., Ebersberger, S., et al. (2019). The RNA-binding ubiquitin ligase MKRN1 functions in ribosome-associated quality control of poly(A) translation. Genome Biol *20*, 216.

Huang, V. (2017). Endogenous miRNAa: miRNA-Mediated Gene Upregulation. In RNA Activation, L.-C. Li, ed. (Singapore: Springer Singapore), pp. 65–79.

Ikeuchi, K., and Toshifumi, I. (2016). Ribosome-associated Asc1/RACK1 is required for endonucleolytic cleavage induced by stalled ribosome at the 3′ end of nonstop mRNA. Scientific Reports *6*.

Institute for Theoretical Chemistry RNAfold web server. University of VIenna.

Itoh, Y., Sekine, S. -i., Suetsugu, S., and Yokoyama, S. (2013). Tertiary structure of bacterial selenocysteine tRNA. Nucleic Acids Research *41*, 6729–6738.

Katoh, T., Iwane, Y., and Suga, H. (2017). Logical engineering of D-arm and T-stem of tRNA that enhances d-amino acid incorporation. Nucleic Acids Research *45*, 12601–12610.

Kaufer, B.B., Jarosinski, K.W., and Osterrieder, N. (2011). Herpesvirus telomeric repeats facilitate genomic integration into host telomeres and mobilization of viral DNA during reactivation. J Exp Med *208*, 605–615.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nature Reviews Molecular Cell Biology *13*, 700–712.

Kheimar, A., Previdelli, R., Wight, D., and Kaufer, B. (2017). Telomeres and Telomerase: Role in Marek's Disease Virus Pathogenesis, Integration and Tumorigenesis. Viruses *9*, 173.

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet *11*, 597–610.

Kung, H.-J., Xia, L., Brunovskis, P., Li, D., Liu, J.L., and Lee, L.F. (2001). Meq: An MDV-Specific bZIP Transactivator with Transforming Properties. In Marek's Disease, K. Hirai, ed. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 245–260.

Lai, W.S., Stumpo, D.J., Kennington, E.A., Burkholder, A.B., Ward, J.M., Fargo, D.L., and Blackshear, P.J. (2013). Life without TTP: apparent absence of an important anti-inflammatory protein in birds. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology *305*, R689–R700.

Ma, F., Liu, X., Li, D., Wang, P., Li, N., Lu, L., and Cao, X. (2010). MicroRNA-466l Upregulates IL-10 Expression in TLR-Triggered Macrophages by Antagonizing RNA-Binding Protein Tristetraprolin-Mediated IL-10 mRNA Degradation. J.I. *184*, 6053–6059.

Machlin, E.S., Sarnow, P., and Sagan, S.M. (2011). Masking the 5' terminal nucleotides of the hepatitis C virus genome by an unconventional microRNA-target RNA complex. Proceedings of the National Academy of Sciences *108*, 3193–3198.

Matsui, M., Chu, Y., Zhang, H., Gagnon, K.T., Shaikh, S., Kuchimanchi, S., Manoharan, M., Corey, D.R., and Janowski, B.A. (2013). Promoter RNA links transcriptional regulation of inflammatory pathway genes. Nucleic Acids Research *41*, 10086–10109.

May, J.P., Yuan, X., Sawicki, E., and Simon, A.E. (2018). RNA virus evasion of nonsense-mediated decay. PLoS Pathog *14*, e1007459.

McPherson, M.C., and Delany, M.E. (2016). Virus and host genomic, molecular, and cellular interactions during Marek's disease pathogenesis and oncogenesis. Poult. Sci. *95*, 412–429.

Michalski, D., Ontiveros, J.G., Russo, J., Charley, P.A., Anderson, J.R., Heck, A.M., Geiss, B.J., and Wilusz, J. (2019). Zika virus noncoding sfRNAs sequester multiple host-derived RNA-binding proteins and modulate mRNA decay and splicing during infection. J. Biol. Chem. *294*, 16282–16296.

Michlewski, G., and Cáceres, J.F. (2019). Post-transcriptional control of miRNA biogenesis. RNA *25*, 1–16.

Miller, S.L. (1953). A Production of Amino Acids under Possible Primitive Earth Conditions. Science *117*, 528–529.

Mustafin, R.N., and Khusnutdinova, E.K. (2019). The Role of Reverse Transcriptase in the Origin of Life. Biochemistry Moscow *84*, 870–883.

Muylkens, B., Coupeau, D., Dambrine, G., Trapp, S., and Rasschaert, D. (2010). Marek's disease virus microRNA designated Mdv1-pre-miR-M4 targets both cellular and viral genes. Arch Virol *155*, 1823–1837.

Ohtsuki, T., and Watanabe, Y. (2007). T-armless tRNAs and elongated elongation factor Tu. TBMB *59*, 68–75.

Ote, I., Lebrun, M., Vandevenne, P., Bontems, S., Medina-Palazon, C., Manet, E., Piette, J., and Sadzot-Delvaux, C. (2009). Varicella-Zoster Virus IE4 Protein Interacts with SR Proteins and Exports mRNAs through the TAP/NXF1 Pathway. PLoS ONE *4*, e7882.

Parcells, M.S., Anderson, A.S., and Morgan, R.W. (1995). Retention of Oncogenicity by a Marek's Disease Virus Mutant Lacking Six Unique Short Region Genes. J. VIROL. *69*, 11.

Piepenbrink, M.S., Li, X., O'Connell, P.H., and Schat, K.A. (2009). Marek's disease virus phosphorylated polypeptide pp38 alters transcription rates of mitochondrial electron transport and oxidative phosphorylation genes. Virus Genes *39*, 102–112.

Place, R.F., Li, L.-C., Pookot, D., Noonan, E.J., and Dahiya, R. (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. Proceedings of the National Academy of Sciences *105*, 1608–1613.

Porrua, O., and Libri, D. (2013). RNA quality control in the nucleus: The Angels' share of RNA. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1829*, 604–611.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat Rev Mol Cell Biol *16*, 190–202.

Previdelli, R., Bertzbach, L., Wight, D., Vychodil, T., You, Y., Arndt, S., and Kaufer, B. (2019). The Role of Marek's Disease Virus UL12 and UL29 in DNA Recombination and the Virus Lifecycle. Viruses *11*, 111.

Qian, Z., Brunovskis, P., Iii, F.R., Lee, L., and Kung, H.-J. (1995). Transactivation Activity of Meq, a Marek's Disease Herpesvirus bZIP Protein Persistently Expressed in Latently Infected Transformed T Cells. J. VIROL. *69*, 8.

Ramanathan, M., Majzoub, K., Rao, D.S., Neela, P.H., Zarnegar, B.J., Mondal, S., Roth, J.G., Gai, H., Kovalski, J.R., Siprashvili, Z., et al. (2018). RNA–protein interaction detection in living cells. Nat Methods *15*, 207–212.

Rasschaert, P., and Laurent, S. (2015). Régulation transccriptionnelle et post-transcriptionnelle des gènes LAT et ICP4 du virus de la maladie de Marek. Université de Tours.

Rasschaert, P., Figueroa, T., Dambrine, G., Rasschaert, D., and Laurent, S. (2016). Alternative splicing of a viral mirtron differentially affects the expression of other microRNAs from its cluster and of the host transcript. RNA Biology *13*, 1310–1322.

Samrat, S.K., Ha, B.L., Zheng, Y., and Gu, H. (2017). Characterization of Elements Regulating the Nuclear-to-Cytoplasmic Translocation of ICP0 in Late Herpes Simplex Virus 1 Infection. J Virol *92*, e01673-17.

Sandri-Goldin, R., M. (2008). The many roles of the regulatory protein ICP27 during herpes simplex virus infection. Front Biosci *Volume*, 5241.

Schat, K.A., Calnek, B.W., and Fabricant, J. (1982). Characterisation of two highly oncogenic strains of Marek's disease virus [1] [2]. Avian Pathology *11*, 593–605.

Setten, R.L., Lightfoot, H.L., Habib, N.A., and Rossi, J.J. (2018). Development of MTL-CEBPA: Small Activating RNA Drug for Hepatocellular Carcinoma. CPB *19*, 611–621.

Simms, C.L., Thomas, E.N., and Zaher, H.S. (2017). Ribosome-based quality control of mRNA and nascent peptides: mRNA and protein quality control on the ribosome. WIREs RNA *8*, e1366.

Strassheim, S., Stik, G., Rasschaert, D., and Laurent, S. (2012). mdv1-miR-M7-5p, located in the newly identified first intron of the latency-associated transcript of Marek's disease virus, targets the immediate-early genes ICP4 and ICP27. Journal of General Virology *93*, 1731–1742.

Strassheim, S., Gennart, I., Muylkens, B., André, M., Rasschaert, D., and Laurent, S. (2016). Oncogenic Marek's disease herpesvirus encodes an isoform of the conserved regulatory immediate early protein ICP27 generated by alternative promoter usage. Journal of General Virology *97*, 2399–2410.

Stroynowska-Czerwinska, A., Fiszer, A., and Krzyzosiak, W.J. (2014). The panorama of miRNA-mediated mechanisms in mammalian cells. Cell. Mol. Life Sci. *71*, 2253–2270.

Svoboda, P., and Cara, A.Di. (2006). Hairpin RNA: a secondary structure of primary importance. Cell. Mol. Life Sci. *63*, 901–908.

Swadling, J.B., Ishii, K., Tahara, T., and Kitao, A. (2018). Origins of biological function in DNA and RNA hairpin loop motifs from replica exchange molecular dynamics simulation. Phys. Chem. Chem. Phys. *20*, 2990–3001.

Szádeczky-Kardoss, I., Gál, L., Auber, A., Taller, J., and Silhavy, D. (2018). The No-go decay system degrades plant mRNAs that contain a long A-stretch in the coding region. Plant Science *275*, 19–27.

Tahiri-Alaoui, A., Zhao, Y., Sadigh, Y., Popplestone, J., Kgosana, L., Smith, L.P., and Nair, V. (2014). Poly(A) Binding Protein 1 Enhances Cap-Independent Translation Initiation of Neurovirulence Factor from Avian Herpesvirus. PLoS ONE *9*, e114466.

Trapp, S., Parcells, M.S., Kamil, J.P., Schumacher, D., Tischer, B.K., Kumar, P.M., Nair, V.K., and Osterrieder, N. (2006). A virus-encoded telomerase RNA promotes malignant T cell lymphomagenesis. J Exp Med *203*, 1307–1317.

Tunnicliffe, R.B., Collins, R.F., Ruiz Nivia, H.D., Sandri-Goldin, R.M., and Golovanov, A.P. (2018). The ICP27 Homology Domain of the Human Cytomegalovirus Protein UL69 Adopts a Dimer-of-Dimers Structure. MBio *9*, e01112-18, /mbio/9/3/mBio.01112-18.atom.

Vautherot, J.-F., Jean, C., Fragnet-Trapp, L., Rémy, S., Chabanne-Vautherot, D., Montillet, G., Fuet, A., Denesvre, C., and Pain, B. (2017). ESCDL-1, a new cell line derived from chicken embryonic stem cells, supports efficient replication of Mardiviruses. PLOS ONE *12*, e0175259.

Wahle, E., and Winkler, G.S. (2013). RNA decay machines: Deadenylation by the Ccr4–Not and Pan2–Pan3 complexes. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1829*, 561–570.

Wakiyama, M., Ogami, K., Iwaoka, R., Aoki, K., and Hoshino, S. (2018). MicroRNP-mediated translational activation of nonadenylated mRNAs in a mammalian cell-free system. Genes Cells *23*, 332–344.

Weeks, K.M. (2010). Advances in RNA structure analysis by chemical probing. Current Opinion in Structural Biology *20*, 295–304.

Weerasooriya, S., DiScipio, K.A., Darwish, A.S., Bai, P., and Weller, S.K. (2019). Herpes simplex virus 1 ICP8 mutant lacking annealing activity is deficient for viral DNA replication. Proc Natl Acad Sci USA *116*, 1033–1042.

Weisser, M., and Ban, N. (2019). Extensions, Extra Factors, and Extreme Complexity: Ribosomal Structures Provide Insights into Eukaryotic Translation. Cold Spring Harb Perspect Biol *11*, a032367.

Wethmar, K. (2014). The regulatory potential of upstream open reading frames in eukaryotic gene expression: uORF-mediated translational control. WIREs RNA *5*, 765–768.

Woese, C.R., Winker, S., and Gutell, R.R. (1990). Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". Proceedings of the National Academy of Sciences *87*, 8467–8471.

Xie, Q., Anderson, A.S., and Morgan, R.W. (1996). Marek's Disease Virus (MDV) ICP4, pp38, and meq Genes Are Involved in the Maintenance of Transformation of MDCC- MSB1 MDV-Transformed Lymphoblastoid Cells. J. VIROL. *70*, 7.

Zaborowska, J., Baumli, S., Laitem, C., O'Reilly, D., Thomas, P.H., O'Hare, P., and Murphy, S. (2014). Herpes Simplex Virus 1 (HSV-1) ICP22 Protein Directly Interacts with Cyclin-Dependent Kinase (CDK)9 to Inhibit RNA Polymerase II Transcription Elongation. PLoS ONE *9*, e107654.

Zhang, Y., Luo, J., Tang, N., Teng, M., Reddy, V.R.A.P., Moffat, K., Shen, Z., Nair, V., and Yao, Y. (2019). Targeted Editing of the pp38 Gene in Marek's Disease Virus-Transformed Cell Lines Using CRISPR/Cas9 System. Viruses *11*, 391.

Zhuang, G., Sun, A., Teng, M., and Luo, J. (2017). A Tiny RNA that Packs a Big Punch: The Critical Role of a Viral miR-155 Ortholog in Lymphomagenesis in Marek's Disease. Front. Microbiol. *8*, 1169.