Université de Sherbrooke

Structural definition of substrate recognition by model RNA capping enzymes and the identification of a novel class of viral RNA capping enzymes

By

Issur Moheshwarnath

Département de biochimie

Thesis submitted to the *Faculté de médecine et des sciences de la santé*

In partial fulfillment of the requirements for the obtention of the degree

Philosophiae doctor (Ph.D.) in Biochemistry

March 2011

Your file Votre référence
ISBN: 978-0-494-83281-3

Our file Notre référence
ISBN: 978-0-494-83281-3

Members of the evaluating commitee

Dr. Martin Bisaillon

Dr. Éric Marsault

Dr. Léa Brakier-Gingras

Dr. Xavier Roucou

Résumé (en français)

La structure coiffe est une unité protectrice située à l'extrémité 5' des ARNm eucaryotes. Cette structure est essentielle pour le transport et la traduction des ARNm. La coiffe est synthétisée co-transcriptionellement par trois activités enzymatiques consécutives : (1) Une ARN triphosphatase (RTase) qui hydrolyse le phosphate terminal de l'extrémité 5' de l'ARN; (2) Une ARN guanylyltransférase (GTase) transfère un groupement GMP à cette extrémité diphosphorylée et; (3) une ARN (guanine-N7) méthyltransférase (MTase) qui ajoute un groupement méthyle à la position N7 de la coiffe.

La cristallographie a permis d'élucider la structure de plusieurs enzymes impliquées dans la synthèse de la structure coiffe. En dépit du fait que ces structures ont permis de mieux comprendre plusieurs aspects mécanistiques du fonctionnement de ces enzymes, certains points nébuleux persistent, notamment, sur les interactions enzymes-substrats. Parmi les premières études cristallographiques des enzymes de la synthèse de la coiffe, la structure de la protéine Cet1, la RTase de *Saccharomyces cerevisiae*, fut élucidée en complexe avec une molécule de sulfate. La difficulté à élucider la structure du complexe enzyme-ligands est probablement liée à l'instabilité thermodynamique de ce complexe. Afin d'avoir une meilleure compréhension de l'activité RTase de *S. cerevisiae* une analyse en profondeur du site actif de cette enzyme en complexe avec un substrat approprié reste à être établie.

Les nombreuses études cristallographiques sur la GTase du virus *Paramecium bursaria chlorella virus*-1 (PBCV-1) ont permis d'élucider le chemin réactionnel macroscopique de cette famille d'enzyme. Cette protéine a été cristallisée en présence de plusieurs ligands et sous différentes conformations; ce qui a permis de visualiser certaines étapes de la réaction de la synthèse de la coiffe. Malgré tout, en dépit du fait que le substrat naturel, le GTP, a été co-cristallisé avec la protéine, d'autres études permettant la compréhension de la spécificité de la GTase envers la guanine doivent être réalisées. De plus, un mécanisme moléculaire de la catalyse de la réaction GTase manque toujours à nos connaissances.

L'importance de la structure coiffe pour le processus de traduction est incontestable. La relation entre la coiffe et la machinerie traductionnelle a été étudiée indirectement par l'étude des protéines qui lient la coiffe. La plupart des études qui investiguaient directement la structure coiffe étaient restreintes à évaluer l'inhibition de la traduction par des analogues de la coiffe. Des études sur les effets des analogues de la coiffe en 5' n'ont débuté que depuis peu, et de plus, la fonction principale de la méthylation en N7 de la coiffe n'a pas encore été adressée.

Cette thèse vise donc à fournir un aperçu de la dynamique structurale des interactions enzyme-ligand de la RTase de *S. cerevisiae* et de la GTase de PBCV-1. Nous montrons que des analogues de purines peuvent être un outil utile pour l'étude de plusieurs processus cellulaires, tels que la traduction. Au cours de mes études, nous avons découvert une nouvelle classe de GTase virale dans la famille des *flavivirus*, fournissant ainsi un aperçu plus succinct du complexe de réplication de ce virus.

Key words: Analogues, ARNm, flavivirus, Protéines virales, Structure coiffe, Traduction,

SUMMARY:

The RNA cap structure is a fundamental feature of most known eukaryotic mRNAs and some viral RNAs. It is important for the stability, transport and translation of mRNAs. It is co-transcriptionally synthesized via the action of 3 consecutive enzymatic reactions: (1) A RNA triphosphatase which cleaves off the 5' terminal phosphate of nascent RNAs; (2) A RNA guanylyltransferase which transfers a GMP moiety onto the acceptor RNA; (3) A RNA (guanine-N7) methyltransferase which methylates the cap guanine at the N7 position.

Through the end of the 1990's until now, the crystal structures of several capping enzymes have been solved. However, these structures, although very insightful in themselves, failed to provide any instructive information on several key issues regarding enzyme-substrate interactions. For instance, one of the first breakthrough crystallographic studies in RNA capping chemistry led to the elucidation of the yeast RNA triphosphatase structure (the Cet1 protein). However, in the crystal structure, the Cet1 protein was bound to a sulphate molecule, which was hypothesised to be mimicking the product of the RNA triphosphatase reaction- a phosphate molecule. The inability to capture the RNA triphosphatase in complex with its ligands is probably on account of the inherent thermodynamic instability of this protein when bound to RNA or a nucleotide. A structural definition of the active site of the yeast RNA triphosphatase in complex with an appropriate substrate is still lacking.

In addition, the elucidation of the structure of the RNA guanylyltransferase of the *Paramecium bursaria chlorella virus*-1 (PBCV-1) in several different conformations has been a landmark study which greatly contributed towards the understanding of the catalytic pathway of this model enzyme. On the other hand, despite the presence of the natural substrate-GTP, within the active site of the enzyme, the rationale behind the GTP specificity of RNA guanylyltransferase remains poorly understood. Moreover, a molecular mechanism for the RNA guanylyltransferase reaction is still missing. Finally, the importance of the RNA cap for the process of eukaryotic translation is undisputable. However, the relationship between the RNA cap and translation has been mostly studied indirectly through proteins which bind to the cap structure. Most studies pertaining directly to the impact of the binding of the RNA cap structure have been restricted to investigating

the inhibitory potential of various cap analogues on the translation process. Studies on the effects of modified RNA caps at the 5' ends of RNAs have only started in the last few years, and more importantly, the necessity of the N7-methyl group on RNA cap analogues had not been addressed.

This thesis therefore aims to provide a structural insight into the structural dynamics of enzyme-ligand(s) interactions of the model *S. cerevisiae*'s RNA triphosphatase and the PBCV-1 RNA guanylyltransferase. In addition, we show that purine analogues can be a useful tool for the study of several cellular processes, such as RNA translation. In the process we have uncovered a novel class of RNA capping enzyme in the *flavivirus* genus of the *Flaviviridae* family of RNA viruses, thus providing a more succinct insight into the flaviviral replication complex.

I dedicate this thesis

to the memory of my father,

Mr. Hemraze Issur

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A103R: The RNA guanylyltransferase of *Paramecium bursaria chlorella virus-1*

ATP: Adenosine triphosphate

CBP 20/80: Cap Binding Proteins 20 and 80

Cet1: The Capping enzyme triphosphatase of *Saccharomyces cerevisiae*

CPSF: Cleavage and polyadenylation specificity factor

Dcp1/Dcp2: Decapping enzymes 1 and 2

eIF4E/eIF4G: Eukaryotic initiation factors 4 E and G

GTP: Guanosine triphosphate

IRES: Internal Ribosome Entry Site

Lys: Lysine

mRNA: Messenger RNA

NS5: The non-structural protein 5

NTP: Nucleoside triphosphate

PAP: Poly(A) Polymerase

PABP: Poly(A) Binding Protein

PBCV-1: *Paramecium bursaria chlorella virus-1*

Pol II: RNA Polymerase II

Poly(A): Poly adenosine

RNA: Ribonucleic acid

SWI/SNF: SWItch/Sucrose NonFermentable

TAFs: TBP Associated Factors

TBP: TATA Binding protein

TTM: Triphosphate tunnel metalloenzyme

INTRODUCTION

1. The lifecycle of eukaryotic mRNAs

The life cycle of mRNAs starts with their transcription and ends ultimately in their demise by degradation (Gu and Lima 2005). Eukaryotic mRNAs differ from their bacterial counterparts in several ways namely by undergoing several co-transcriptional modifications such as the addition of a RNA cap at their 5' end, the synthesis of a poly(A) tail at the 3' end following cleavage, and the splicing of introns to form mature functional mRNAs (Varani 1997). These modifications are crucial to establish the fate of the mRNA. Appropriately modified mRNAs will be eligible for transport to the cytoplasm and translation by ribosomes (Hamm and Mattaj 1990; Shatkin and Manley 2000). On the other hand, a pre-mRNA possessing a defect in maturation (a splicing defect for instance) is rapidly targeted for degradation (Isken and Maquat 2007).

An mRNA, from its synthesis to its degradation, is bound by several proteins and protein complexes throughout its life cycle. For instance, following synthesis of the 5' end by RNA Polymerase II and capping by the RNA capping machinery, the RNA cap structure is bound by the CBP20/80 complex in the nucleus. This interaction is important for the splicing of mRNAs as well as their transport from the nucleus to the cytoplasm (Izaurralde, Lewis et al. 1994; Visa, Izaurralde et al. 1996; Lewis and Izaurralde 1997; Newbury 2006). The CBP20/80 complex is exchanged for the eIF4E protein of the translation initiation complex in the cytoplasm. The binding of the RNA cap structure to eIF4E is fundamental for cap-dependent translation to occur. The 3' poly(A) tail also plays a vital role in the life cycle of mRNAs. It is bound by several proteins, the most notable one being PABP, which is vital to improve translation efficiencies of mRNAs (Kuhn and Wahle 2004).

The 5' and 3' modifications brought about on an mRNA also contribute to the stability of mRNAs. The 5' RNA cap structure blocks the 5' end in an unusual 5'-5' linkage, thus effectively preventing 5' exonucleases from degrading the mRNA. For a regulated degradation of mRNAs, decapping enzymes cleave off the cap structure to aid in RNA degradation (Newbury 2006). Therefore, due to the critical roles that RNA modifications play in the life cycle of mRNAs, most enzymes directly involved in the synthesis of these modifications have been shown to be essential. Extensive genetic studies in the model

organism budding yeast have shown that all components of the RNA capping machinery are crucial for survival (Shatkin and Manley 2000). In addition, in several viruses, abrogating any step of RNA cap synthesis leads to defective viral replication (Bisaillon and Lemay 1997). This underlines the fundamental importance of the RNA cap structure in the life cycle of eukaryotic mRNAs.

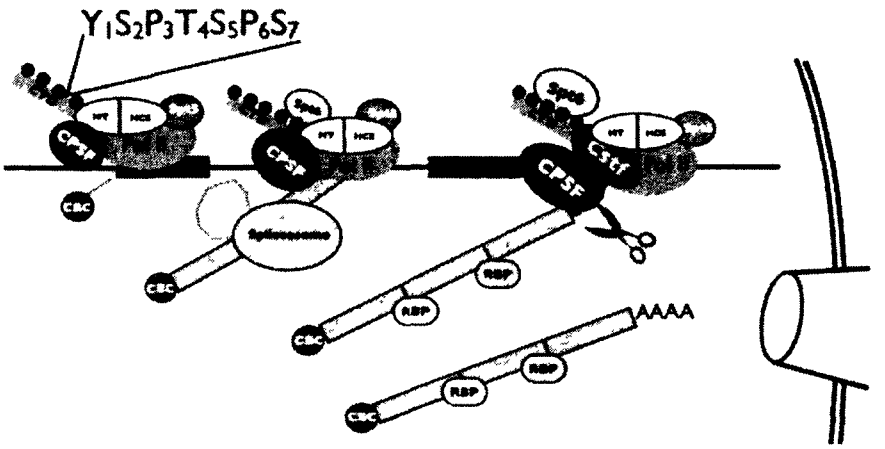
1.1 Synthesis and co-transcriptional modifications

Eukaryotic mRNAs are exclusively transcribed by the DNA-dependent RNA Polymerase II (Pol II). The transcription process is guided by several tightly regulated steps. Initially in the first steps of transcription, a pre-initiation complex consisting of the polymerase and transcription factors (such as TBP, TAFs and SWI/SNF units) binds to the promoter region of the gene of interest, and RNA synthesis begins. Initiation is followed by promoter clearance, when initial contacts between the promoter and the polymerase are broken. The polymerase complex then enters the elongation phase of transcription, culminating finally into transcription termination when the polymerase dissociates from the DNA template and the RNA transcript is released (Howe 2002). During the tightly regulated process of mRNA synthesis, several modifications are also brought about co-transcriptionally on the mRNA (Fig. 1) (Moore and Proudfoot 2009).

The first modification to occur is the addition of the 5' RNA cap structure. *In vivo*, cap addition occurs early during transcription, when about 25-30 nucleotides have been polymerized and the 5' end is protruding from the RNA binding pocket of the RNA Pol II. The enzymes involved in RNA cap synthesis are specifically recruited to the Pol II pre-initiation complex. The phosphorylation level of the CTD of the PolII holoenzyme mediates this specific recruitment at the beginning of the transcription process. The evolutionary conserved proteins, CBP20 and CBP80, co-transcriptionally bind as a heterodimer to the RNA cap structure shortly after its synthesis. The RNA cap structure retains its association to this cap-binding complex (CBC) throughout its transcription, co-transcriptional processing and nucleocytoplasmic export (Moore and Proudfoot 2009).

During transcription elongation, intron removal occurs. This step is catalyzed by the assembly of a macromolecular ribonucleoprotein complex, the spliceosome. The splicing

reaction has been extensively characterized in several organisms and will only be alluded to in this thesis (Martinez-Contreras, Cloutier et al. 2007). Prior to transcription termination, another key modification to occur on the pre-mRNA is the addition of a poly(A) tail. This is a multi-step mechanism which is initiated by the recognition of the Poly(A) cleavage site by processing factors. Following endonucleolytic cleavage of the pre-mRNA, a chain of adenosine nucleotides is added to the transcript by a Poly(A) polymerase (PAP), to which Poly(A) Binding Protein (PABP) binds co-transcriptionally. At the end of these steps, a mature mRNA is released from the transcription bubble for transport to the cytoplasm (Shatkin and Manley 2000; Moore and Proudfoot 2009). This is summarized in Figure 1.



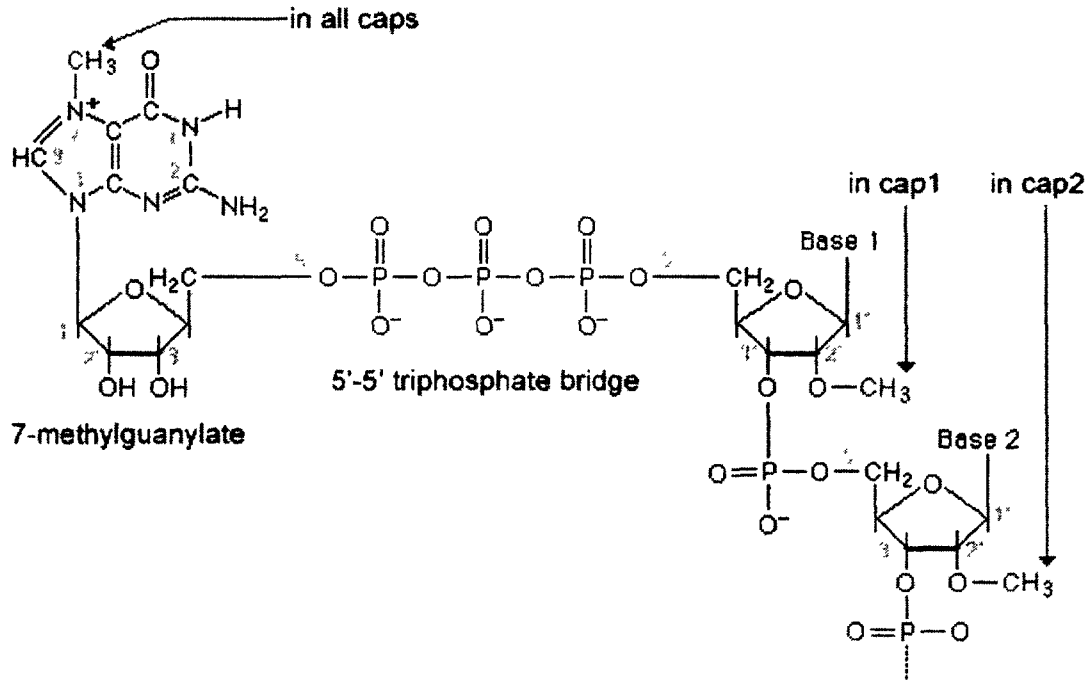
Introduction, Figure 1: Global overview of mRNA synthesis and co-transcriptional modifications brought on the pre-mRNA in human cells *(adapted from Perales and Bentley. Mol Cell, 2009)*

mRNAs are transcribed by RNA Polymerase II, which possess a hepta-peptide repeat in its CTD. The level of phosphorylation of the CTD dictates transcription progress. The RNA is bound by several proteins or protein complexes during its synthesis. The CBP 20/80 complex binds to the 5' cap following its synthesis by HCE and IIT (human capping enzyme and human methyltransferase). The spliceosome binds to excise introns, while the Cleavage and polyadenylation specificity factor (CPSF) binds to the pre-mRNA for cleavage and polyadenylation. Following these modifications, mRNAs can be exported for translation to the cytoplasm. Scissors depict RNA cleavage at the poly(A) site.

2. The RNA cap structure

2.1. Structure and function

The RNA cap structure consists of a N7-methyl guanosine residue added at the 5' ends of eukaryotic mRNAs via a 5'-5' triphosphate bridge (Fig. 2). The capped guanosine residue is methylated at the N7 position in most mRNAs and forms a functional RNA cap structure known as a cap 0. In some organisms and viruses, the ribose of the first nucleotide of the mRNA may be methylated at the 2' position. This leads to the formation of a cap 1 structure. Additional methylations on the 2' position of subsequent nucleotides in the RNA lead to the formation of cap 2, 3 etc structures (Bisaillon and Lemay 1997). The 5' RNA cap has several essential functions in eukaryotic cells. In addition to being essential for cap-dependent translation for protein synthesis, it plays several key roles in the life cycle and life span of mRNAs.



Introduction, Figure 2: The structure of the mRNA cap

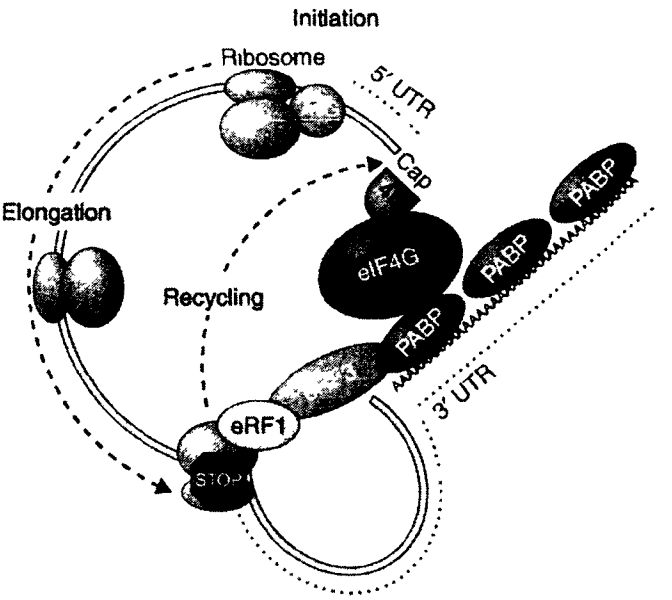
In green is the N7-methyl guanosine residue linked via a 5'-5' triphosphate bridge (in grey) onto the 5' end of the RNA (in blue). In red are the methylation positions on mRNAs.

Stability of mRNAs: The unconventional 5'-5' bond in the RNA cap structure protects the 5' ends of mRNAs against most exonucleases since they only hydrolyze 3'-5' bonds. During the regulated degradation of mRNAs in cells, decapping enzymes (the Dcp1/Dcp2 complex in baker's yeast) degrades the 5'-5' linkage prior to the 5'to 3' degradation process (Newbury 2006). In cells, the cap structure is strongly bound by the Cap Binding Complex (CBC 20/80 in metazoans and CBC 1/2 in yeast) in the nucleus and eIF4E/eIF4G in the cytoplasm, which limits its access by the decapping machinery, thus effectively protecting the mRNA against 5' to 3' degradation (Moore and Proudfoot 2009).

Nucleo-cytoplasmic transport of mRNAs: Following transcription of mRNAs within the nucleus, their transport to the cytoplasm for translation is ensured by the Cap Binding Complex. The Nuclear Cap Binding Complex binds exclusively to the RNA cap structure and is then recognized by the Nuclear Pore Complex, for transport into the cytoplasm, where, following the first pioneer round of translation, it is replaced by the eIF4E/eIF4G complex (Maquat 2004).

Translation efficiency of mRNA: Translation efficiency is significantly improved by the presence of the 5' cap. The 5' cap plays a key role in recruiting protein co-factors, and ultimately the small ribosomal subunit to the 5' ends of mRNAs. In addition, the 5' cap, when bound by the cytoplasmic cap binding complexes (eIF4E/eIF4G), lead to the circularization of the mRNA molecule by interactions with poly(A) binding proteins bound at the 3'ends of mRNAs (Fig.3). Such interactions enable an efficient recycling of ribosomes, thus improving translational efficiency of mRNAs (Livingstone, Atas et al. 2010).

Splicing of the 5' proximal intron: The RNA cap structure has also been shown to be important to promote intron excision during the splicing of the 5' proximal intron (Lewis, Izaurralde et al. 1996; Lewis and Izaurralde 1997). The precise mechanism by which this is achieved is still under study in model organisms (Raczynska, Simpson et al. 2010).



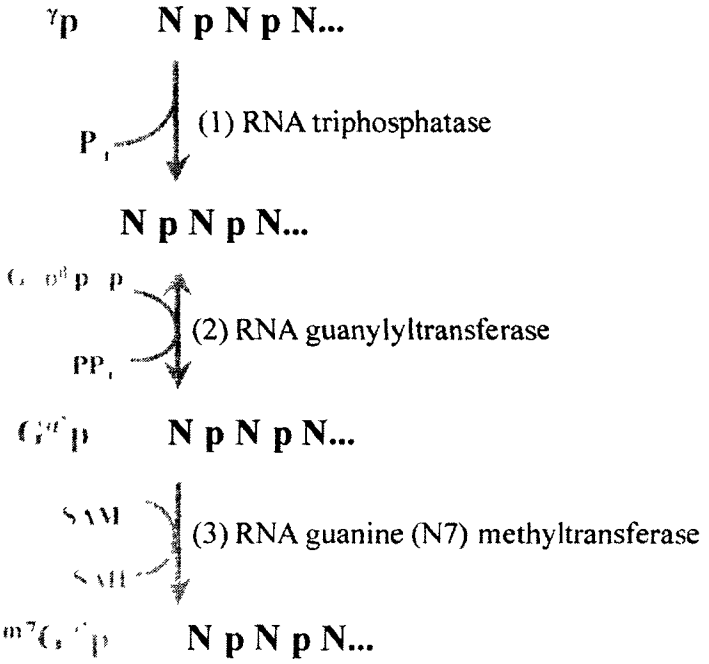
Introduction, Figure 3: Importance of the RNA cap and Poly(A) tail *(adapted from Sonenberg and Dever, Curr. Opin. Struct. Biol., 2003)*

The RNA cap structure is bound by eIF4E a component of the eIF4F complex during translation initiation. eIF4G, another component of the initiation complex binds to PABP, the poly(A) binding protein. Thus a mature mRNA is effectively circularised, which ensures rapid recycling of ribosomes for efficient translation.

2.2 Synthesis of the RNA cap structure

RNA cap synthesis is a co-transcriptional event requiring 3 essential enzymatic reactions (Fig. 4). During RNA transcription, the 5' end of the nascent RNA is triphosphorylated. A RNA triphosphatase is first required to cleave off the terminal phosphate (γ) of this RNA to yield a diphosphorylated end, onto which a RNA guanylyltransferase will transfer a GMP moiety. Finally a RNA (guanine-N7) methyltransferase transfers a methyl group onto the N7 position of the cap guanine nucleotide. These 3 consecutive reactions lead to the formation of a functional cap structure called cap 0 (Fig. 4) (Shuman 1995). As mentioned previously, in some organisms and viruses, additional methylation events may occur on the ribose moiety of the first nucleotide(s) of the RNA molecule, thereby leading to the formation of cap 1, 2 etc structures. RNAs harbouring such cap structures (Cap 0, 1, 2 etc), are translation competent. The importance of these additional methylations is still under investigation. However, in some viruses, a recent report suggests that 2'O methylation of viral mRNAs serve to subvert the host antiviral response (Daffis, Szretter et al. 2010).

Mutant viruses of the *flavivirus*, *poxvirus* and *coronavirus* genera, which are deficient in their 2'O- methyltransferase activity, are more sensitive to the antiviral action of *Interferon-1*. It is speculated that 2'O methylation has evolved in some eukaryotic genomes, in order to differentiate self from non-self mRNAs (Zust, Cervantes-Barragan et al. 2010). In addition to these types of RNA cap structures, there also exists RNAs harbouring a trimethylguanosine (TMG) cap structure. This cap differs from the typical cap 0 structure of mRNAs with the addition of two methyl groups at the exocyclic N2 position. A subset of RNA Pol II transcribed cellular RNAs, including small nuclear RNAs (snRNA), small nucleolar RNA (snoRNA), and the telomerase RNA are thus capped (Reddy, Singh et al. 1992). These RNAs are transported within the cell for various functions and are not aimed for translation.

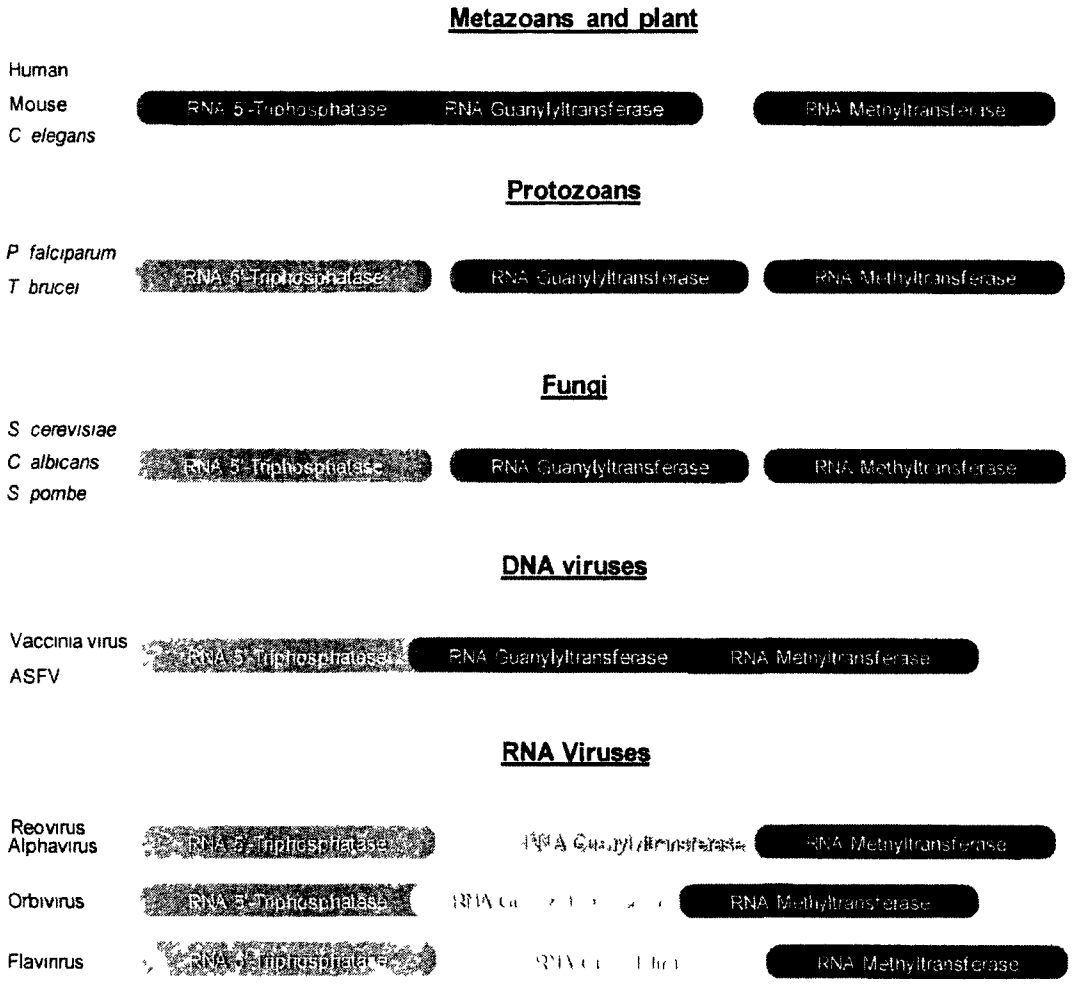


Introduction, Figure 4: RNA Cap Synthesis

(1) Cleavage of the terminal (γ) phosphate by a RNA triphosphatase followed by release of inorganic phosphate (P_i). (2) Transfer of a GMP moiety onto the diphosphorylated end by a RNA guanylyltransferase with the concomitant release of pyrophosphate (PP_i) (3) Addition of a methyl group at the N^7 position of the cap guanine by a RNA (guanine-N7) methyltransferase with S-Adenosyl Methionine (SAM) as co-factor S-Adenosyl-homocysteine (SAH) is released as a by product of the reaction.

2.3 Functional Organization of RNA capping enzymes

From viruses to metazoans, the RNA cap structure is identical. However, the physical organization and to some extent the mechanism of the enzymes responsible for its synthesis differ significantly across the various taxa (Ghosh and Lima 2010). While in all known unicellular eukaryotes, each RNA modifying activity for RNA cap synthesis lies on a separate polypeptide, in higher order eukaryotes like plants and metazoans, the RNA (guanine-N7) methyltransferase is segregated from the RNA triphosphatase and RNA guanylyltransferase which are fused together in a single polypeptide (Fig. 5). On the other hand, with regards to viruses, no common theme distinct to any viral group emerges. However, the genomic organization as well as the mechanism of some of the proteins involved in RNA capping (more precisely the RNA triphosphatase and the RNA guanylyltransferase) diverges widely from those in unicellular eukaryotes or from those in plants and metazoans (Fig. 5).



Introduction, Figure 5: Schematic depiction of the organization of capping enzymes across various taxa *(adapted from Issur et al, Wiley's RNA, 2010)*

Each box represents an enzymatic activity. A space between boxes infers enzymatic activities harboured by distinct polypeptides while fused boxes represent enzymatic activities residing on the same protein. The colour code represents the structural and/or mechanistic divergence between members of the same family of enzymes

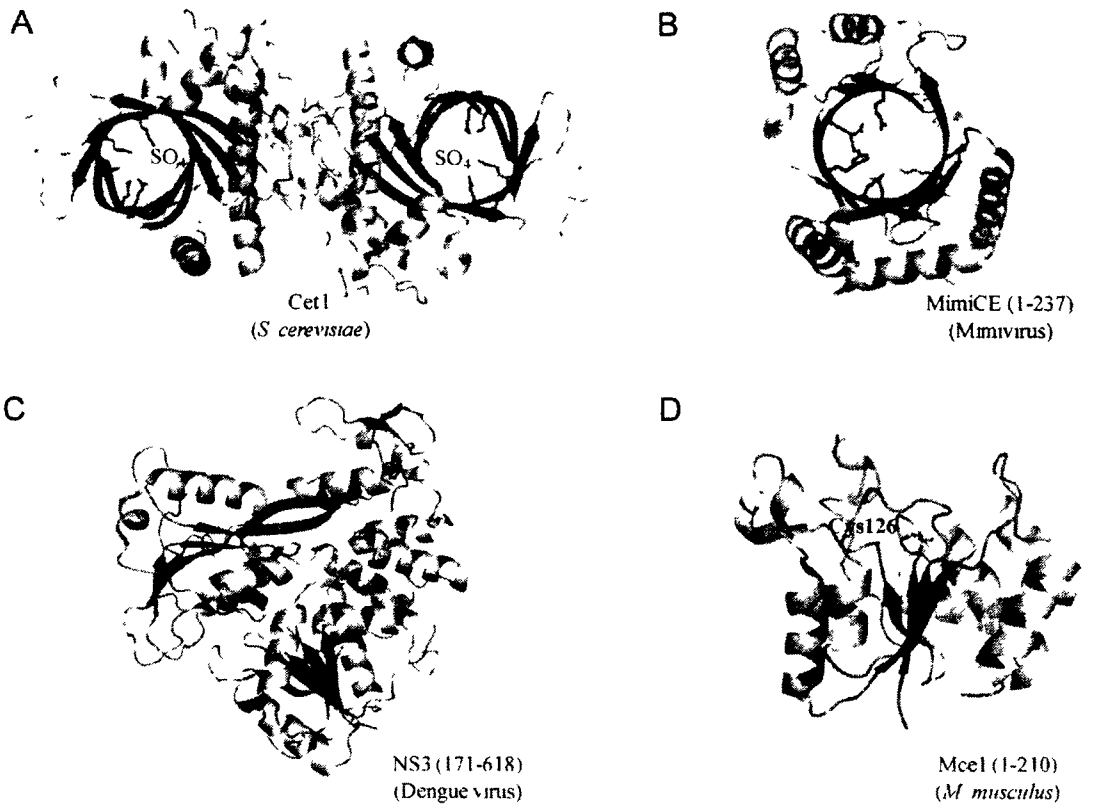
2.3.1 RNA triphosphatase

The RNA triphosphatase catalyzes the first step of RNA cap synthesis. Across the various eukaryotic lineages, RNA triphosphatases differ significantly with respect to structure and catalytic mechanism and can be grouped into 2 distinct families: (1) metal-dependent RNA triphosphatases of lower eukaryotes such as fungi and protozoans (2) metal-independent RNA triphosphatases of nematodes, metazoans and plants (Fig. 6 and 7).

2.3.1.1. Metal-dependent RNA triphosphatases

The metal-dependent RNA triphosphatase has been identified in several eukaryotic lineages ranging from DNA viruses to fungi and protozoans. These RNA triphosphatases belong to the triphosphate tunnel metalloenzyme (TTM) family of phosphohydrolases and in contrast to metal-independent RNA triphosphatases, can hydrolyze NTPs (Ghosh and Lima 2010). The initial crystallization of the *S. cerevisiae*'s RNA triphosphatase, the Cet1 protein, revealed its homodimeric nature, whereby two equivalent active sites are present around a topologically closed eight-stranded anti-parallel β barrel (Fig. 6A) (Lima, Wang et al. 1999). Biochemical and mutational evidences also suggest that this *then-novel* fold also encompasses the RNA triphosphatases of other fungi (*Schizosaccharomyces pombe*, *Candida albicans*), protozoan parasites (*Plasmodium falciparum, Trypanosoma brucei, Encephalitozoon cuniculi, Giardia lambia*) and some DNA viruses (*Chlorella* virus, poxvirus, baculovirus) (Shuman 2002). The recent elucidation of the Mimivirus RNA triphosphatase domain crystal structure has provided the first structural evidence for the inclusion of viral RNA triphosphatases in the TTM clade (Fig. 6B) (Benarroch, Smith et al. 2008). Most interestingly, the TTM fold is more widely distributed across the various taxa than initially expected. Its finding within the archael and bacterial domains of life suggests a deeper evolutionary origin. The family of metal-dependent RNA triphosphatases is exemplified by the model *S. cerevisiae*'s RNA triphosphatase, the Cet1 protein. Its active site consists of several basic residues presumably important for coordinating the triphosphate moiety of pppRNA, as well as some basic residues coordinating metal ions which directly interact with the pppRNA triphosphate moiety (Fig. 7A). Phosphohydrolysis by RNA triphosphatases of the TTM clade is purported to occur in a one-step in-line catalytic mechanism. The metal ions are suggested to be important for activating a water

molecule for nucleophilic attack on the γ-phosphate as well as to stabilize the subsequent phosphorane transition state intermediate formed (Bisaillon and Shuman 2001; Bisaillon and Bougie 2003). A crucial glutamate residue (Glu 433 in Cet1) is suggested to act as a general base by abstracting a proton from a water molecule coordinated by the aforementioned glutamate residue and potentially a metal ion, thus effectively rendering the coordinated water molecule a general acid for nucleophilic attack on the γ-phosphate. This mechanism is still under investigation.



Introduction, Figure 6: Structures of yeast, viral and metazoan RNA triphosphatases (All structures were generated using PYMOL)

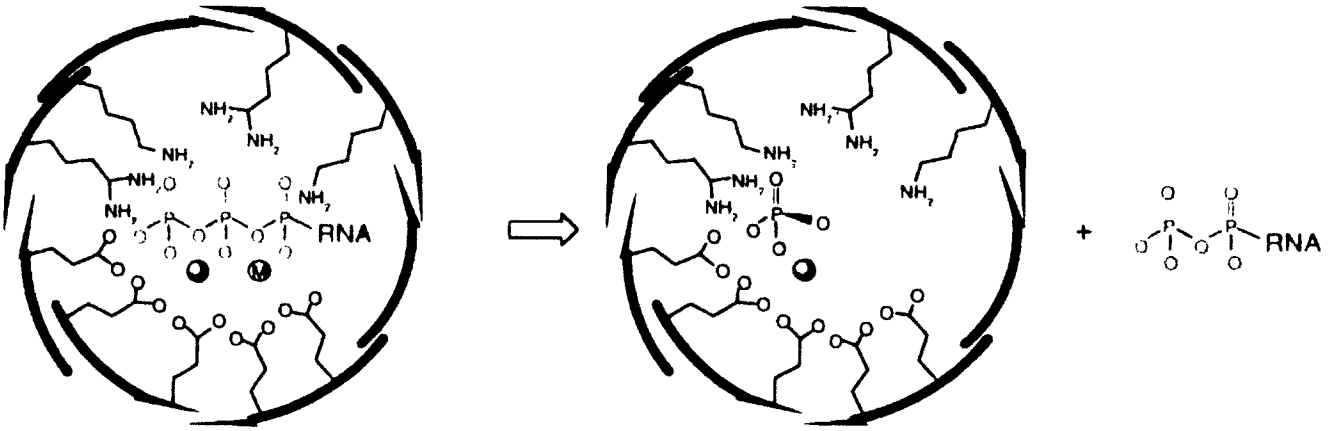
β-sheets are coloured in red, alpha helices in orange and random coils in gray. (A) The homodimeric RNA triphosphatase Cet1 (PDB 1D8H) from S. cerevisiae with a view looking into the triphosphate tunnel with a co-crystallized sulphate shown in sticks. (B) The monomeric RNA triphosphatase domain of mimivirus (PDB 3BGY). (C)The NS3 protein from Dengue virus corresponding to its RNA triphosphatase domain is shown (PDB 2BHR). (D) The murine RNA triphosphatase domain (PDB 1I96). The active site cysteine (Cys126) is denoted in stick representation and labelled.

2.3.1.2. Metal ion- independent RNA triphosphatase

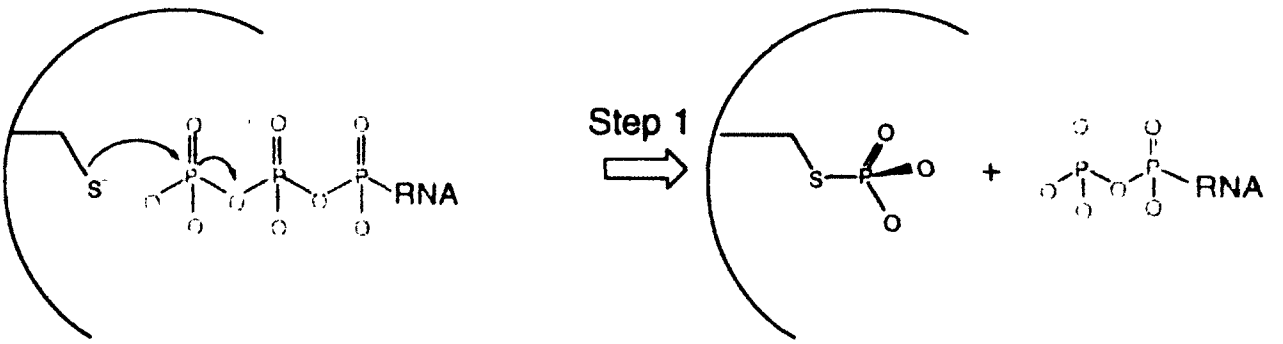
In higher eukaryotes the metal-independent RNA triphosphatase belonging to the cysteine-phosphatase enzyme superfamily, prevails. The elucidation of the crystal structure of the mouse RNA triphosphatase provided the first structural evidence that the mammalian RNA triphosphatase (Mce1), which is defined by a central five-stranded parallel β-sheet flanked with several α-helices, is entirely unrelated to the TTM-clade of RNA triphosphatases (Fig. 6D) (Fabrega, Shen et al. 2003). Phosphohydrolysis by these metal-independent RNA triphosphatases occurs in two steps not unlike the protein phosphatases of the same family. In the first step of the reaction, the conserved cysteine (Cys 126 in Mce1) attacks the 5' γ-phosphate of the RNA transcript to form a covalent protein-cysteinyl-*S*-phosphate intermediate with the concomitant release of the 5' ppRNA product. In the second step, the covalent phosphoprotein is hydrolyzed to liberate inorganic phosphate (Fig. 7B) (Ghosh and Lima 2010).

With the ultimate intent to find novel antiviral and anti-microbial targets, this thesis is more concerned with RNA triphosphatases of the TTM clade, more specifically, the model Cet1 protein. The relative instability induced to the Cet1 protein upon ligand binding is probably the reason behind the lack of structural data of these enzymes bound to their substrates (Bisaillon and Bougie 2003). TTM RNA triphosphatases have only been co-crystalized with magnesium ions and sulphate or acetate bound within the active site. The bound sulphate/acetate molecule is purported to act as the γ-phosphate of the substrate RNA. In this thesis, through computational and biochemical methods we provide the first insight into how the tunnel active site of the TTM RNA triphosphatases binds to their substrates.

(A) Metal-dependent RNA triphosphatase (Cet1 tunnel)



(B) Metal-independent RNA triphosphatase



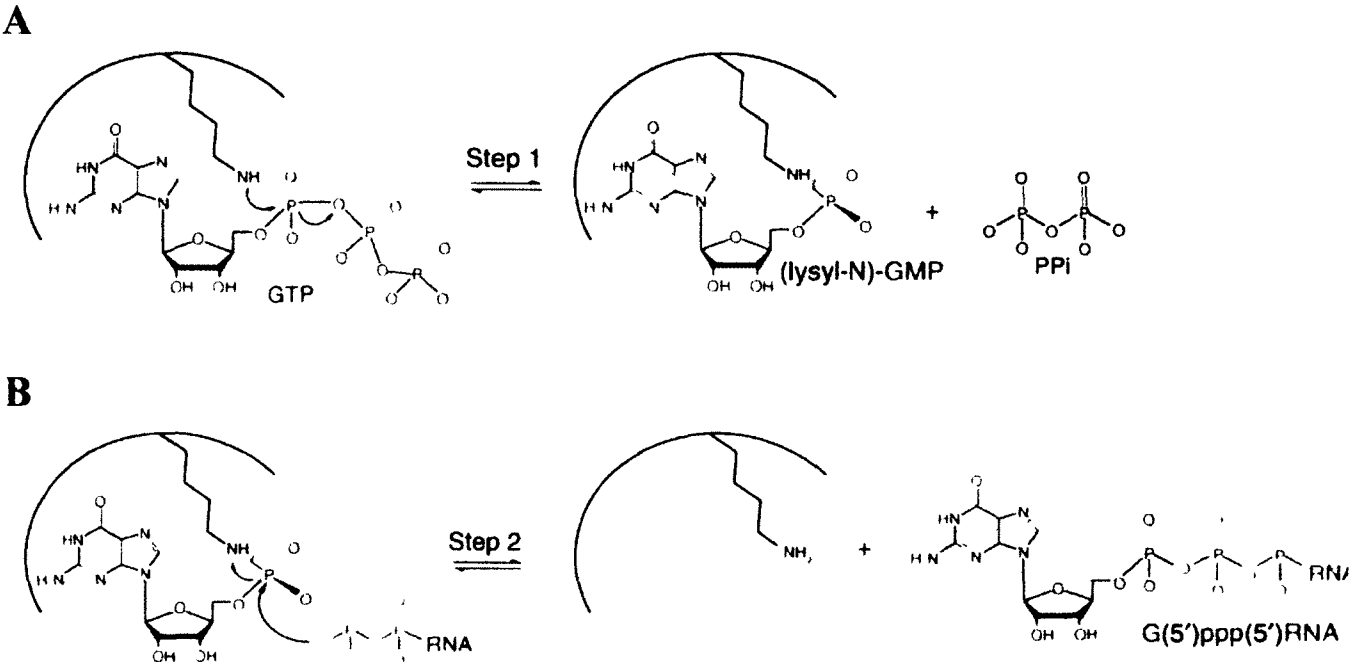
Introduction, Figure 7: Schematic depiction of the RNA triphosphatase mechanism *(adapted from Ghosh and Lima, Wiley Wire's RNA, 2010)*

The metal-dependent (top) and the metal-independent (bottom) RNA triphosphatase activities catalyze the removal of the γ phosphate from pppRNA. The tunnel topology of the metal dependent RNA triphosphatase Cet1 is shown with the side chains that coordinate pppRNA and 2 divalent metal ions, one derived from biochemical studies (circle with the letter M) and one derived from structural studies (solid sphere in blue).

Metal independent RNA triphosphatases operate in 2 steps. In the first step (shown above) a covalent protein-cysteinyl-S-phosphate intermediate is formed with the γ-phosphate from pppRNA following its hydrolysis. The covalent intermediate is then hydrolyzed to liberate inorganic phosphate in solution.

2.3.2. RNA guanylyltransferase

The RNA guanylyltransferase carries out the second step of RNA cap synthesis. This enzyme catalyzes the formation of a chemical bond between the 5' end of a GMP moiety and the 5' diphosphorylated end of mRNA, thus effectively forming an exonuclease-resistant 5'-5' linkage. With only a few exceptions in some viral families, the RNA guanylyltransferase reaction is very conserved across all eukaryotic lineages (Fig. 5) (Doherty and Suh 2000; Shuman and Lima 2004).



Introduction, Figure 8: Schematic depiction of the RNA guanylyltransferase mechanism *(adapted from Ghosh and Lima, Wiley Wire's RNA, 2010)*

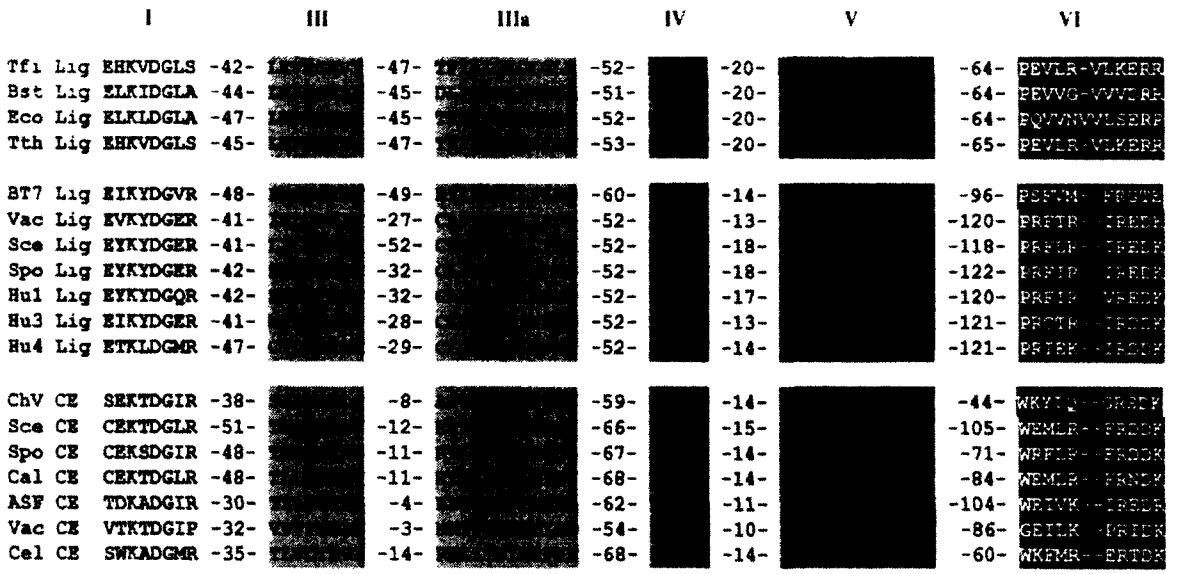
(A) RNA guanylyltransferase catalyzes capping in a two-step reaction. In step 1, the enzyme binds GTP and magnesium to catalyze transfer of GMP to the active site lysine to form a covalent enzyme-(lysyl-N)-GMP intermediate.(B) In step 2, the enzyme binds the ppRNA (colored red) to catalyze transfer of the GMP to form GpppRNA.

2.3.2.1.Mechanism of RNA guanylyltransferase

Two distinct catalytic events, in a ping-pong type reaction, define the RNA guanylyltransferase mechanism (Fig.8). In the first step, attack of the α-phosphorus of GTP by the capping enzyme results in the release of pyrophosphate and formation of a covalent enzyme-(*lysyl-N*)-GMP intermediate (Fig.8A). The second step entails the transfer of this covalently bound GMP onto an acceptor RNA (Fig.8B). The RNA substrate in the second step of the reaction is specifically a diphosphorylated RNA formed from the hydrolysis of the terminal 5' phosphate by an RNA triphosphatase. Both steps of this reaction require the presence of a divalent metal ion and have been shown to be reversible (Souliere, Perreault et al. 2008). The catalytic lysine, the ε amino group of which forms the covalent phosphoamidate intermediate in the first step, is part of a KxDG motif (Motif I), one of six co-linear conserved sequences (I-VI) defining the active site of this family of enzymes (Fig.9).

2.3.2.2. Nucleotidyltransferase superfamily

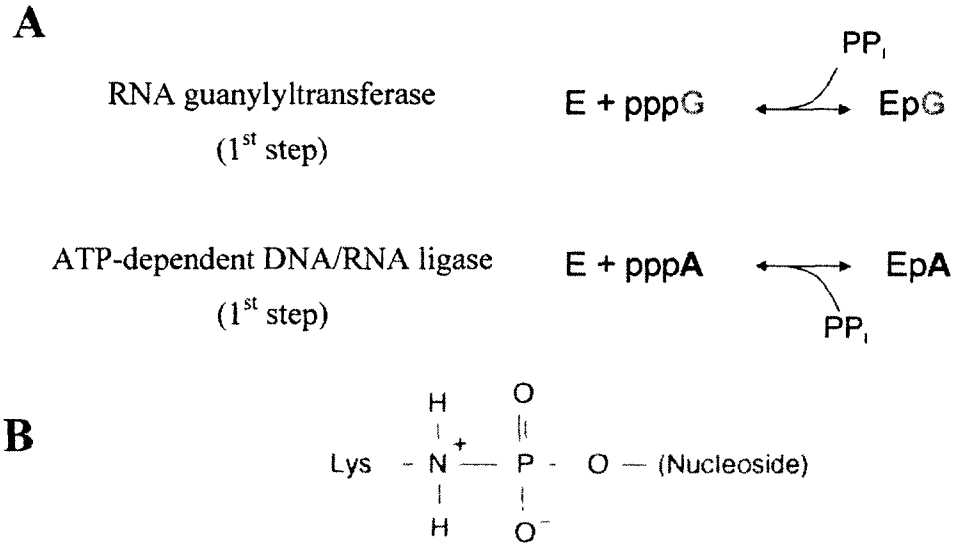
Interestingly, the mechanism as well as the conserved motifs (I-VI) of capping enzymes are shared by ATP-dependent RNA and DNA ligases, which along with the former belong to the broader nucleotidyltransferase superfamily (Fig. 9) (Doherty and Suh 2000). In fact, crystal structures of the *Chlorella* virus RNA guanylyltransferase and of the T7 DNA ligase bound to GTP and ATP respectively have revealed a common tertiary fold in which the conserved motifs are assembled together at the enzyme's active site (Fig. 10). Both enzymes can be split into 2 distinct domains: a larger N-terminal domain encompassing motifs I, III, IIIa, IV and V which form the nucleotide binding pocket, and a smaller C-terminal OB fold domain, which includes motif VI. Indeed, the enzyme-AMP intermediate, analogous to the phosphoamidate intermediate during cap synthesis, formed when RNA/DNA ligases react with ATP, clearly highlights the mechanistic similarity shared by the members of this superfamily (Fig. 11).



Introduction, Figure 9: Sequence alignments of RNA guanylyltransferases (abbreviated CE) and ATP-dependent ligases (abbreviated Lig) *(extracted from Doherty and Shu, Nucleic Acids Res, 2000)*

Motifs I, II, III, IIIa, IV, V and VI are the six conserved elements in NAD^+- and ATP-dependent DNA/RNA ligases and RNA guanylyltransferases. The upper part of the alignment consists of the NAD^+dependent ligases encoded by T. filiformis (Tfi), B. stearothermophilus (Bst), E. coli (Eco) and T. thermophilus (Tth). Below these are aligned sequences for the ATP-dependent DNA ligases (Lig) of bacteriophage T7 (BT7), vaccinia virus (Vac), Saccharomyces cerevisiae (Sce), Schizosaccharomyces pombe (Spo) and human ligases I (Hu1), 3 (Hu3) and 4 (Hu4). The last block of the alignment includes the RNA guanylyltransferases encoded by Chlorella virus PBCV-1 (ChV), S. cerevisiae, S. pombe, Candida albicans (Cal), African swine fever virus (ASF), vaccinia virus and Caenorhabditis elegans (Cel). The catalytic lysine residue is highlighted in red for each protein.

Introduction, Figure 10: Structural similarity of DNA ligases and RNA guanylyltransferases *(adapted from Doherty and Shu, Nucleic Acids Res., 2000)*
The location of the conserved sequences as defined in figure 9 are indicated, with the same color coding for the T7 DNA ligase and the PBCV-1 RNA guanylyltransferase.



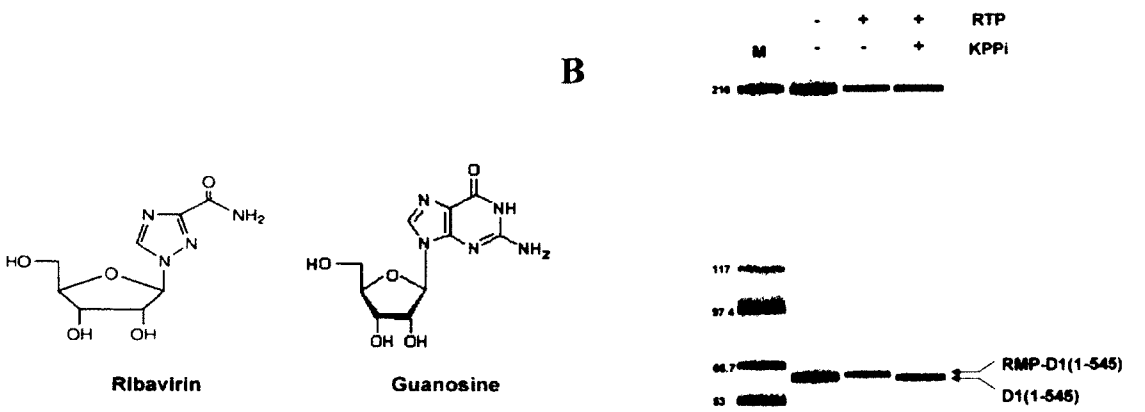
Introduction, Figure 11: Mechanistic similarity between RNA guanylyltransferases and ATP-dependent DNA/RNA ligases.

(A) The first step of the nucleotidyltransferase reaction. RNA guanylyltransferases form a covalent bond with GMP, while ATP-dependent DNA/RNA ligases form a covalent bond with AMP (B) In both cases, the covalent bond is a phosphoamidate bond formed between a nucleoside and a lysine residue.

2.3.2.3. Nucleotide specificity of RNA guanylyltransferases

Despite the availability of more structural and biochemical data on several members of the nucleotidyltransferase superfamily, an accurate understanding of the substrate specificity of RNA capping enzyme still remains elusive. As yet, from a biological perspective, studies pertaining to this particular issue are scarce. It is worth mentioning that through domain swapping between the T7 ligase and the PBCV-1 RNA guanylyltransferase, an ATP capping enzyme has been previously generated (Doherty 1999). However, modulation of the substrate specificity of RNA guanylyltransferases through the rational design of point mutations remains to be achieved. While mutational analyses have yielded substantial data on protein-substrate interactions, gauging the importance of each with regards to their importance in substrate recognition is tedious and limited in its scope.

In 2004, the report that, *in vitro*, ribavirin triphosphate (RTP) could be used as a cap donor by the vaccinia virus RNA capping enzyme revived interest in the possibility of modulating the substrate specificity of RNA guanylyltransferases (Fig.12) (Bougie and Bisaillon 2004). This was the first report that an RNA capping enzyme could potentially use an alternative substrate to GTP to generate a modified RNA cap structure at the 5' end of an mRNA. In the wake of this study, we decided to investigate into the propensity of a model RNA guanylyltransferase to use artificial substrates as cap donors. Synthetic cap analogues have been extensively studied with regards to RNA stability and translation. However, few studies have aimed to probe into the capping machinery itself with a view to understand the underlying interactions leading to substrate discrimination. In this thesis we present a thorough study of the structural requirements of a ligand to be a cap donor by the model PBCV-1 RNA guanylyltransferase. We use purine analogues bearing various modifications at different positions and by gauging the effect of each modification, we traced the essential recognition motifs on the guanosine residue which determine its inherent ability to act as a cap donor.



Introduction, Figure 12: A viral RNA guanylyltransferase can form a covalent intermediate with Ribavirin monophosphate *(adapted from Bougie et Bisaillon, JBC, 2004)*

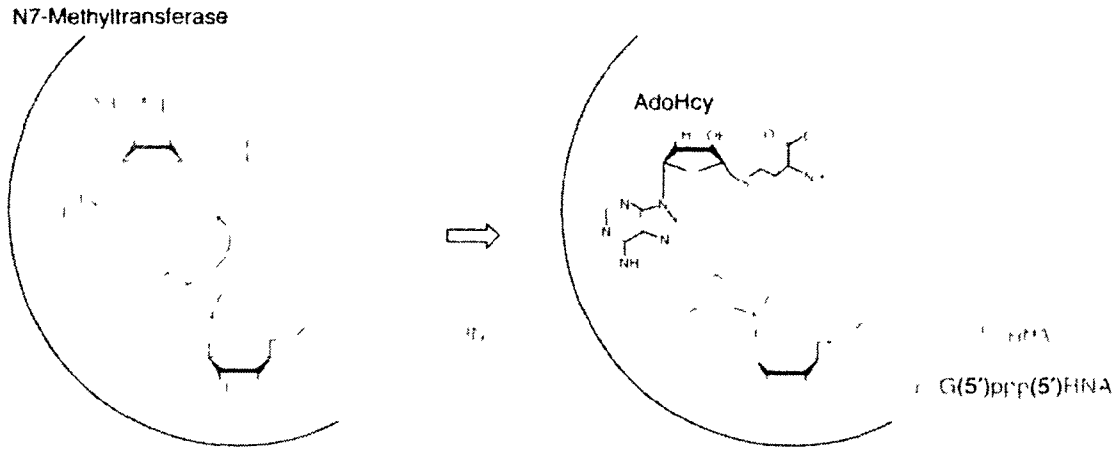
Left: Structural difference and similarity between Ribavirin and Guanosine.
Right: Visualisation of the enzyme-ribavirin monophosphate (D1-RMP) complex formation by capillary electrophoresis. Ribavirin triphosphate (RTP) and potassium pyrophosphate (KPPi) were added where indicated. The location of the covalent complex with Ribavirin monophosphate is indicated to the right.

2.3.3. RNA (guanine-N7) methyltransferase

The last essential step during RNA capping chemistry is carried out by the RNA (guanine-N7) methyltransferase, which catalyzes the transfer of a methyl group from S-Adenosyl Methionine (SAM) onto the N7 position of the capped guanine residue (Fig. 13). The methylation of the N7 residue is crucial for the formation of a functional RNA cap structure which can be recognised by cap binding proteins for nuclear export and for the formation of the translation initiation complex.

In our study we used the model RNA (guanine-N7) methyltransferase from *S. cerevisiae*. The MTase of *S. cerevisiae* bears a strong similarity with that of the protozoan parasite *Encephalitozoon cuniculi*; the structure of which had previously been determined in complex with SAH and therefore has been extensively characterized (Fabrega, Hausmann et al. 2004). The absence of stacking interactions within the RNA MTase crystal structure as compared to other cap binding proteins necessarily infers a need for a precise

orchestration of amino acids in the active site to ensure optimal ligand positioning for nucleophilic attack onto the methyl donor (SAM). A model of the MTase mechanism is shown in figure 13.



Introduction, Figure 13: Schematic depiction of the RNA (guanine-N7) methyltransferase mechanism *(adapted from Ghosh and Lima, Wiley Wire's RNA, 2010)*

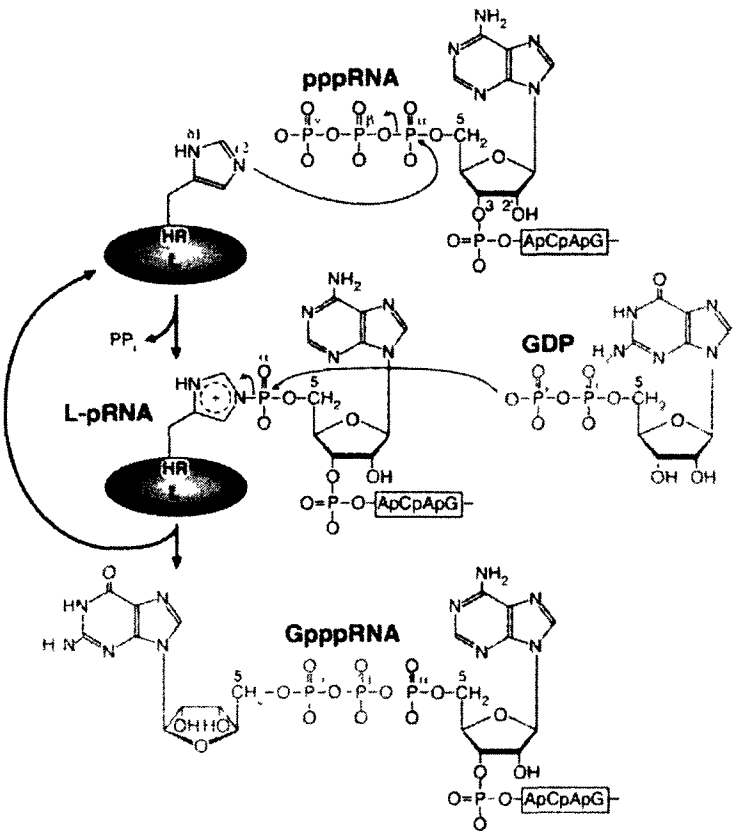
Methyltransferase activities: RNA (guanine-N7) methyltransferase binds S-adenosylmethionine (AdoMet) (colored green) and GpppRNA (colored as above) to catalyze transfer of the methyl group (colored green) to the N7 position of the guanine.

2.4. Unconventional RNA capping apparatus

The RNA cap structure is exclusively found in eukaryotes. The aforementioned RNA capping mechanism is the general RNA capping pathway that is found in all known organisms from fungi to mammals. However, in several viruses alternative pathways for cap formation have been described. These pathways not only differ in terms of the enzymes required, but also in the mechanism as well as in the order of the reactions involved in RNA cap formation. A few examples are discussed below with a particular emphasis on the RNA capping apparatus of the flaviviruses.

Vesicular stomatitis virus (VSV) exemplifies the *Rhabdoviridae* family of the *Monegavirales* order of viruses. VSV is a single stranded RNA virus of negative sense. The

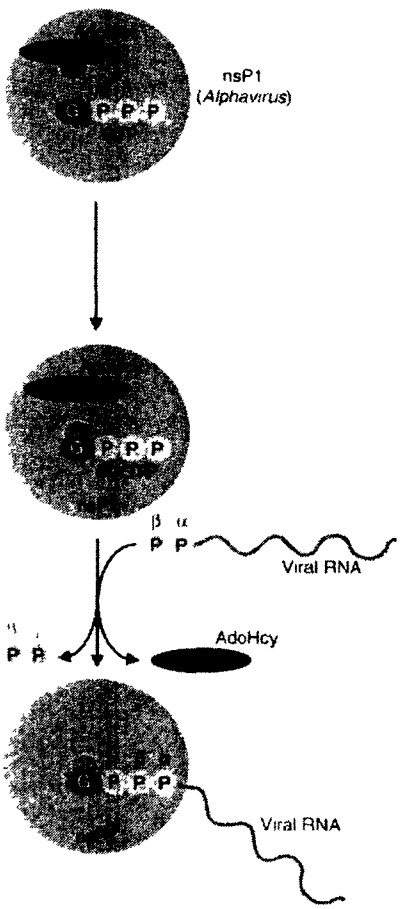
L protein of VSV has a RNA-dependent RNA polymerase activity which ensures viral replication. In addition to this, the L protein also possesses RNA guanylyltransferase activity, without possessing any of the conserved motifs of the nucleotidyltransferase superfamily (Fig. 9). In fact, the L protein uses the 5' triphosphate end of an RNA to form a covalent enzyme-RNA adduct through a phosphoamide bond between the the 5' α-phosphate and a histidine residue on the protein. It has been demonstrated that the L protein also possesses a GTPase activity *in vitro*, which leads to the formation of GDP – the nucleophile in the second step of the reaction to generate the 5' capped RNA (Fig.14) (Ogino, Yadav et al.; Ogino and Banerjee 2007). The L protein also possesses RNA (guanine-N7) methyltransferase activities.



Introduction, Figure 14: Schematic depiction of the proposed model of the unconventional RNA capping with GDP by VSV *(adapted from Ogino, et al, PNAS, 2010)*

The L polymerase forms a histidine-mediated covalent bond with the α-phosphate at the 5'end of the RNA. The covalently bound monophosphate RNA is then transferred onto a diphosphorylated Guanosine residue (GDP) to form the RNA cap structure

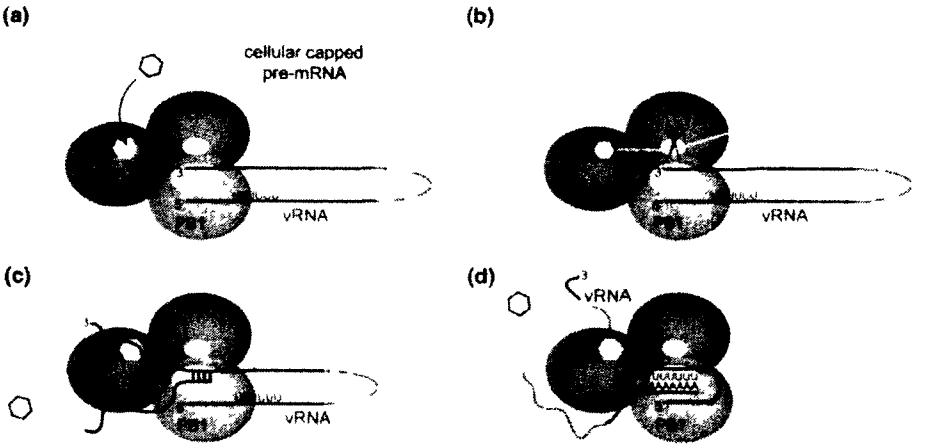
In alphaviruses the nsp1 protein is responsible for capping of nascent transcripts. The nsp1 protein lacks the conserved motifs of the nucleotidyltransferase superfamily. The nsp1 protein can also add a methyl group at the N7 position of the guanine cap. by using S-adenosyl methionine (SAM) as substrate. From biochemical analysis of the nsp1 protein. it has been discovered that very much unlike cellular GTases nsp1 forms a covalent complex not with GMP but instead with an ^{m7}GMP molecule. Nsp1employs a unique capping mechanism whereby the addition of the methyl group at the N7 position occurs prior to the the guanylation of the nascent RNA molecule (Fig.15) (Bisaillon and Lemay 1997).



Introduction, Figure 15: Schematic depiction of the proposed model of the unconventional RNA capping with ^{m7}GTP by *alphaviruses* *(adapted from Ghosh and Lima, Wiley Wire's RNA. 2010)*

The nsp1 protein, by its (guanine-N7) methyltransferase activity, transfers a methyl group at the N7 position of a GTP molecule. The N7-methyl GTP is then hydrolyzed and transferred onto an acceptor RNA.

It should however be mentioned that instead of using the host RNA capping machinery, or encoding for a viral RNA capping apparatus, some viruses like those of the genus *bunyaviridae* snatch the cellular RNA cap structure from cellular mRNAs and use these snatched capped short RNAs to synthesize viral mRNAs (Ghosh and Lima 2010). The mechanism used by the *influenza* virus is shown in figure 16.



Introduction, Figure 16: Schematic depiction of the model of the cap-snatching mechanism by the *influenza* virus *(adapted from Ruigrok RW, et al. Current opinion in structural biology, 2010)*

(a) Host mRNAs bind to the cap binding domain on the PB2 subunit. (b) The bound cellular mRNAs are cleaved by the endonuclease activity of the PA subunit. (c) The 5' capped cleaved cellular RNA fragment is used to elongate the viral mRNA by the nucleotidyltransferase activity of the PB1 subunit and by using the viral RNA as template. (d) The chimeric viral RNA is polyadenylated by a polymerase stuttering mechanism at the oligo-U sequence near the 5' end of the viral RNA.

2.5. The other unconventional RNA guanylyltransferases and the case of the *flavivirus* genus

Unconventional RNA guanylyltransferases have been observed in several families of RNA viruses. In several human pathogenic viruses, the identity of the RNA capping apparatus is still unknown. For instance, the RNA (guanine-N7) methyltransferase of members of the *coronavirus* genus, of which the deadly SARS virus forms part of, has only been recently identified on the nsp1 protein of the virus (Chen, Cai et al. 2009). The identity of the RNA guanylyltransferase is still under investigation. In the *flavivirus* genus, of the *Flaviviridae* family, the identity of the complete RNA capping apparatus has only recently been confirmed.

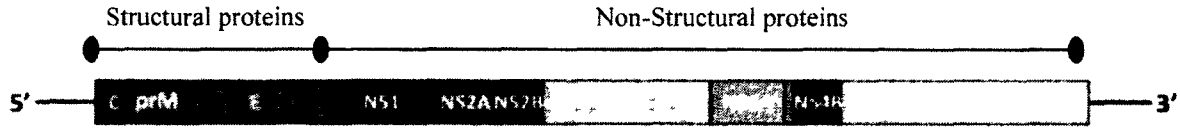
The *flavivirus* genus belongs to the *Flaviviridae* family of viruses along with the *Hepacivirus* and *Pestivirus* genera. These genera differ in numerous ways. One major divergence is the presence of an IRES at the 5' ends of the RNA genomes of the *Hepacivirus* and *Pestivirus* genera whereas members of the the *flavivirus* genus harbour a RNA cap 1 structure. All members of the flavivirus genus possess an RNA cap 1 structure at the 5' end of their RNA genomes. The organization of the RNA genome and the nature of the 5' cap are shown in figure 17 (Bollati, Alvarez et al. 2010).

The *flavivirus* genus comprises several medically important pathogens, the most notorious one being the Dengue Fever virus. This family also includes the Yellow Fever virus, the West Nile virus and the Japanese Encephalitis virus, amongst others (Ecker, Sampath et al. 2005). Flaviviruses are arthropod-born viruses, possessing a complex life cycle involving 2 distinct hosts, mosquito and human (Bollati, Alvarez et al. 2010). The replication cycle of flaviviruses is shown in figure 18.

The viral E protein mediates the attachment of the virions to the host cell surface; and the virions penetrate cells by receptor mediated endocytosis. In the low pH of the endosome, fusion of the viral and host membranes occurs which leads to the release of the nucleocapsid and the viral RNA into the cytoplasm of the cell. The translation of the viral RNA generates a polyprotein which is co-translationally and post-translationally processed by host encoded proteases (eg furin) and the viral encoded protease (NS2B/NS3) to form 3

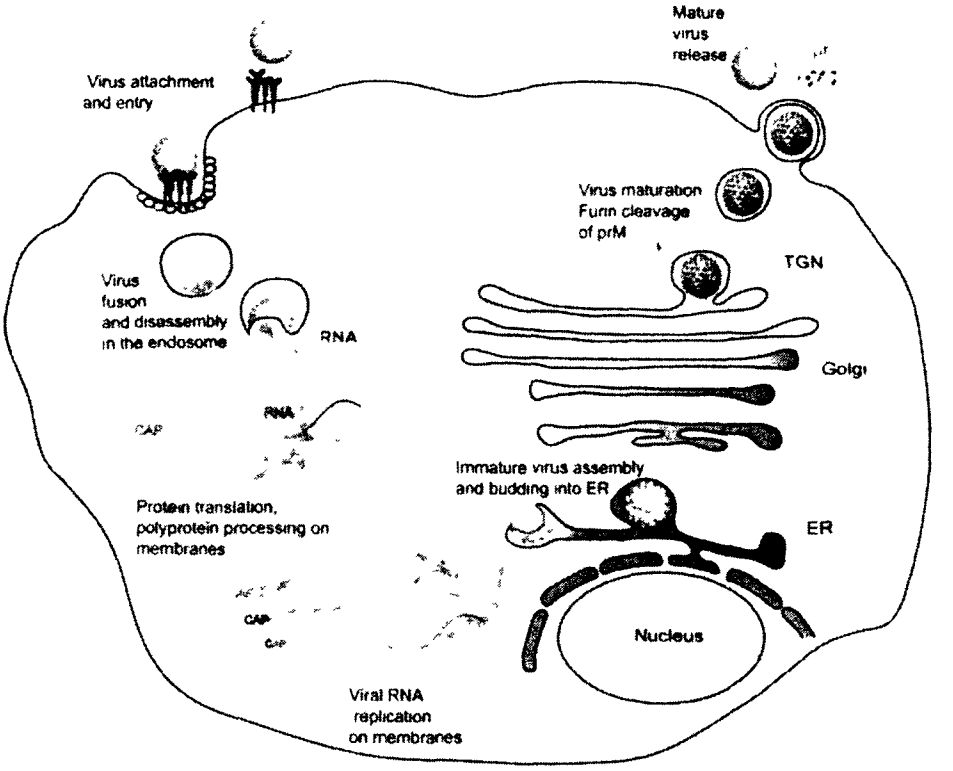
structural and 7 non-structural proteins (Fig.17). The replication of the viral RNA occurs within the membranes of the rough endoplasmic reticulum; assembly of virus particles takes place within the lumen of the ER, and virus maturation prior to release occurs through the Golgi apparatus (Sampath and Padmanabhan 2009).

Among the viral proteins translated, the NS3 (~70 kDa) and NS5 (~105 kDa) proteins are the most characterized, possessing enzymatic activities essential for viral replication. The NS3 protein possesses (1) serine protease activity with the co-factor NS2B, which is required for the polyprotein processing; (2) helicase/NTPase activity required for unwinding the viral RNA; and (3) RNA triphosphatase activity which is essential for the synthesis of the viral RNA cap structure (Lescar, Luo et al. 2008). The NS5 protein is the largest flaviviral protein. It is a multi-functional protein which harbours in its C-terminal domain the RNA-dependent RNA polymerase activity of the virus and in its C-terminal domain a (guanine-N7) and a (2'O) methyltransferase activities, which are required for the synthesis of the RNA cap 1 structure (Dong, Zhang et al. 2008).



Introduction, Figure 17: Schematic linear depiction and organization of the RNA genome of flavivirus *(inspired from Bollati, et al. Antiviral Research, 2010)*

The flaviviral RNA genome possesses a RNA cap 1 structure at its 5' end This RNA genome encodes for a polyprotein which is organized into 2 distinct parts: structural and non-structural proteins as indicated.



Introduction, Figure 18: Schematic depiction of the replication cycle of viruses of the *flavivirus* genus *(taken from Sampath and Padmanabhan Molecular targets for flavivirus drug discovery, 2009)*

Virion enters host cells by receptor-mediated endocytosis. Following liberation of the RNA from virions, viral RNA replication occurs in the rough endoplasmic reticulum (RER). Virion assembly occurs within the lumen of the Endoplasmic Reticulum (ER) while virus maturation occurs via the Golgi apparatus.

As mentioned previously the flavivirus genome possesses a cap 1 structure at its 5' end. Of the 3 essential activities required for RNA cap synthesis, only the RNA triphosphatase (on the NS3 protein) and the RNA methyltransferase (on the NS5) protein have been identified. The RNA guanylyltransferase had escaped identification. The report of a co-crystal structure of GTP with the N-terminal domain of the NS5 protein of the Wesselsbron virus (a member of the *flavivirus* genus) led us to speculate that the NS5 protein also possessed RNA guanylyltransferase activity (Bollati, Milani et al. 2009). In this thesis, we present the first biochemical proof that the NS5 proteins of flaviviruses do possess RNA guanylyltransferase activity.

3. Research Objectives

Enzymes involved in the synthesis of the RNA cap structure are essential for survival. In all eukaryotic organisms, hindering the RNA capping process severely hampers cell growth or leads ultimately to cellular death. In viruses encoding RNA capping apparatuses, inhibiting any of the components of this apparatus severely undermines viral replication. The study of RNA capping enzymes is important because due to their essentiality they can prove to be potential anti-microbial targets. In addition, with regards to viral RNA capping enzymes, they are valuable tools for the mechanistic study of RNA capping enzymes in general on account of their small sizes, which englobe all the essential catalytic motifs for activity. Therefore, it is of fundamental importance to study and make a thorough characterization of RNA capping enzymes, with a view to evaluate their similarity with related enzymes in human and thus, gauge their potency as anti microbial targets.

The yeast RNA triphosphatase is the model enzyme for the study of most fungal, protozoan, and trypanosomal related enzymes, including that from the deadly malaria parasite, *Plasmodium falciparum*. The crystal structure of this enzyme has been solved in complex with a sulphate molecule, purported to mimic the γ-phosphate of the 5' end of an mRNA. Ligand (RNA or NTP) binding thermodynamically destabilizes the protein relative to its free unbound form. This is probably why the structure of the enzyme-ligand complex cannot be elucidated by crystallography for now. The yeast RNA triphosphatase, which belongs to the TTM family of metalo-enzymes, can prove to be an efficient drug target, mainly on account of the fact that the metazoan RNA triphosphatase belongs to the TTM-unrelated metal-independent cysteine phosphatase family. A more acute understanding of the interactions of the enzyme with RNA, could potentially prove to be useful for the design of anti-microbial agents. Therefore, our aim is to probe into the molecular determinants for ligand binding by the model enzyme of the TTM family of RTases, the *S. cerevisiae*'s RTase.

RNA guanylyltransferases are conserved across various viral lineages as well as most of the known eukaryotic taxa. These enzymes belong to the nucleotidyltransferase superfamily, which also includes the ATP dependent or NAD+ dependent DNA/RNA ligases. Despite sharing the same conserved motifs in the same order, as well as displaying very similar

tertiary structure, ligases and RNA guanylyltransferases possess very different nucleotide specificity. Structures of both RNA guanylyltransferases and ligases have been elucidated bound to GTP and ATP respectively. However, even though the molecular contacts between the ligand and the enzyme are known, modulating the substrate specificity of these enzymes has proven be unfeasible. A precise understanding of their substrate specificity could potentially pave the way for the design of novel proteins with novel activities, which could be of use for drug development. In this study our aim is to find the essential recognition elements of GTP which render it able to act as RNA cap donor. The PBCV-1 RNA guanylyltransferase was the model used since it has been co-crystallised several times in complex with GTP. In addition, the PBCV-1 RNA guanylyltransferase, at 330 amino acids, is the smallest known enzyme of this family. It comprises all the essential motifs required for RNA capping chemistry in its short sequence, thus making it an ideal tool to probe into the active site of RNA capping enzymes.

Finally, in the course of my PhD, I investigated into the identity of RNA capping apparatus of the *flavivirus* genus, which includes human pathogenic viruses like the Dengue Fever virus, Yellow Fever virus and the West Nile virus. Up until very recently their RNA capping apparatus had only been partially identified. The RNA triphosphatase activity was shown to reside on the NS3 protein, which also possesses RNA helicase and a serine protease activity. The NS5 protein, which harbours the RNA dependent RNA polymerase activity of the virus, was shown to possess RNA (guanine-N7) and RNA (2'O) methyltransferase activities. The identity of the RNA guanylyltransferase was still unknown. The elucidation of the NS5 protein of several flaviviruses in complex with either GTP or Ribavirin triphosphate, led to the speculation that the NS5 protein could potentially also possess the RNA guanylyltransferase activity of the virus. In this study, we went out to formally identify the NS5 protein as possessing RNA guanylyltransferase activity. The importance of this study lies mainly in the fact that a novel family of RNA guanylyltransferase, unrelated to the metazoan RNA guanylyltransferase, has been identified. Therefore, this activity could be a potent antiviral target for the rational design of anti-flaviviral drugs.

General objectives

The main aim of my research is to probe into the substrate interactions of model enzymes involved in the synthesis of the RNA cap structure, in order to better understand their substrate specificity.

Specific objectives

My research has been focused on essentially 3 different enzymes from different organisms. This has been grouped accordingly in 3 chapters each with the following specific objectives:

(1) Structural characterization of the yeast RNA triphosphatase bound to a nucleotide by computational modelling and validation of the computational model by mutational analysis of the protein and through the use of nucleotide analogues.

(2) Understanding the nucleotide specificity of the model PBCV-1 RNA guanylyltransferase by the use of nucleotide analogues and probing into the consequences of modified RNA cap structures on the process of RNA translation *in cellulo*.

(3) Investigating into the identity and characterization of the RNA guanylyltransferase of the West Nile virus.

RESULTS

Chapter I – Understanding protein-ligand interactions of the model yeast RNA triphosphatase

1.1. ARTICLE:

Nucleotide analogs and molecular modeling studies reveal key interactions involved in substrate recognition by the yeast RNA triphosphatase

Moheshwarnath Issur, Simon Despins, Isabelle Bougie and Martin Bisaillon

Article published in *NUCLEIC ACIDS RESEARCH*, 2009, Vol. 37, No. 11: 3714-3722

CONTRIBUTIONS

I performed 50% of the experiments, analysed all the results and participated in the preparation of the manuscript. SD expressed and purified the Cet1 protein, and helped for the phosphohydrolase assays with BioMol Green. IB generated the expression vector of the Cet1 protein. MB had the original idea, helped in the generation of the docking models, provided the funding and wrote the manuscript.

SUMMARY ***(en français)***

Notre étude de la protéine Cet1, révèle en détail les résidus spécifiques requis pour les interactions d'un ARN triphosphatase avec un ligand. Cette information n'était pas disponible auparavant car la structure cristalline de la protéine en complexe avec un nucléotide ou un oligonucleotide n'a pas encore été résolue. Grâce à des algorithmes d'arrimage moléculaire (<<*molecular docking*>>), nous avons produit un modèle de la protéine Cet1 liée à un nucléotide GTP. Selon ce modèle, la molécule entière de GTP et non pas seulement la queue triphosphate pénètre à l'intérieur de la structure en tunnel de la protéine Cet1. Ce modèle prédit également que, en plus des résidus Arg393, Lys456 et Arg458, qui avaient été observés en coordination avec une molécule de sulfate dans le site actif dans la structure cristalline de la protéine Cet1, plusieurs autres acides aminés (Glu305, Glu433 et Arg458 pour ne citer que ces trois) pourraient être contraignants pour la liaison de la queue triphosphate. La plupart des acides aminés qu'on a identifiés ont été précédemment démontrés par mutagenèse comme étant importants pour la catalyse. Ces acides aminés ont été postulés pour être importants pour la liaison des phosphates du substrat ou de l'ion métallique par des contacts spécifiques avec des molécules d'eau. Bien que des implications non liées à la coordination des phosphates par ces acides aminés ne puissent être exclues, notre modèle suggère que les chaînes latérales de ces acides aminés sont directement impliquées dans la coordination de la partie triphosphate du substrat, afin de permettre son alignement optimal dans le site actif de la protéine. Dans le but de confirmer expérimentalement la validité de notre modèle, des analogues de nucléotides ont été utilisés comme sondes afin de caractériser les déterminants moléculaires de l'interaction de l'ARN triphosphatase de *S. cerevisiae* avec un nucléotide. Tous les analogues de nucleotides analysés peuvent inhiber, quoiqu'à des degrés différents, la reaction d'ARN triphosphatase de la protéine Cet1, mettant ainsi en évidence la flexibilité structurelle du site actif de cette enzyme. Durant cette étude on a remarqué que plusieurs analogues de nucléotides peuvent se lier fortement à l'enzyme, sans pour autant être hydrolysés par l'activité ARN triphosphatase de la protéine. Ces molécules pourront potentiellement être utilisées comme point de départ pour la conception et le développement d'agents anti-fongiques ayant comme cible l'ARN triphosphatase.

SUMMARY

Our study of the *S. cerevisiae*'s RNA triphosphatase, the Cet1 protein, reveals for the first time the specific residues required for the interactions of an RNA triphosphatase with a ligand. Through extensive computational docking procedures, we have produced a model of the Cet1 protein bound to a GTP nucleotide. This computerised model structure predicts that the whole GTP molecule and not only the triphosphate tail is located within the tunnel structure of the yeast RNA triphosphatase. This model also predicts that in addition to Arg393, Lys456 and Arg458 residues, which were previously observed in the crystal structure of *S. cerevisiae*'s RNA triphosphatase to coordinate the bound sulphate, several other amino acids (Glu305, Glu433 and Arg458 to mention only these three) could be binding to the phosphates. Most of the identified amino acids were previously demonstrated to be important for catalysis by mutational analysis. They were postulated to be involved in water-mediated contacts with the phosphates or the divalent metal ion. Although the implication of these amino acids in making interactions unrelated to the binding of the phosphates cannot be excluded, our model suggests that their side chains are directly involved in coordinating the triphosphate moiety of the nucleotide substrate, for the optimal alignment of the substrate for the nucleotide triphosphatase activity of tunnel shaped RNA triphosphatase enzymes. Finally, in order to experimentally confirm the validity of our model, we used nucleotide analogues to probe into the molecular determinants of the interactions of the yeast RNA triphosphatase with a nucleotide. Of the 17 nucleotide analogues tested all could inhibit the RNA triphosphatase reaction, albeit to different extents, thus highlighting the structural flexibility of this enzyme's active site. More interestingly, several analogues could strongly bind to the enzyme, but were not efficiently hydrolyzed. These molecules could be strong starting points for the design of nucleotide based inhibitors of the RNA triphosphatase of pathogens.

ABSTRACT

RNA triphosphatases (RTPases) are involved in the addition of the distinctive cap structure found at the 5' ends of eukaryotic mRNAs. Fungi, protozoa, and some DNA viruses possess an RTPase that belongs to the triphosphate tunnel metalloenzyme family of enzymes that can also hydrolyze nucleoside triphosphates. Previous crystallization studies revealed that the phosphohydrolase catalytic core is located in a hydrophilic tunnel composed of antiparallel β-strands. However, all past efforts to obtain structural information on the interaction between RTPases and their substrates were unsuccessful. In the present study, we used computational molecular docking to model the binding of a nucleotide substrate into the yeast RTPase active site. In order to confirm the docking model, and to gain additional insights into the molecular determinants involved in substrate recognition, we also evaluated both the phosphohydrolysis and the inhibitory potential of an important number of nucleotide analogues. Our study highlights the importance of specific amino acids for the binding of the sugar, base, and triphosphate moieties of the nucleotide substrate, and reveals both the structural flexibility and complexity of the active site. These data illustrate the functional features required for the interaction of an RTPase with a ligand, and pave the way to the use of nucleotide analogues as potential inhibitors of pathogenic RTPases.

INTRODUCTION

Eukaryotic mRNAs harbor a distinctive m7GpppN cap structure at their 5' ends (1). The structure is added shortly after the initiation of transcription by a series of three sequential enzymatic reactions (2-4). The first step involves the hydrolysis of the 5' triphosphate end of the nascent mRNA by an RNA triphosphatase to form a diphosphate extremity. The addition of GMP to the diphosphate end is then mediated by an RNA guanylyltransferase, or capping enzyme. Finally, the GpppN cap is methylated by an RNA (guanine-N7) methyltransferase. Since its discovery three decades ago, numerous studies have demonstrated the importance of the cap structure for the stability, transport, and translation of mRNAs (reviewed in refs. 2 and 5).

An important number of enzymes involved in the synthesis of the cap structure have been found in different eukaryotic organisms ranging from fungi, protozoans, viruses, plants, and metazoans (6). Numerous structural and functional studies have also contributed to elucidate the basic features of these enzymes (reviewed in ref. 5). Interestingly, significant structural and mechanistic differences are found in the RNA triphosphatase (RTPase) component of the capping machinery. Metazoan and plant RTPases belong to the cysteine phosphatase family which also includes numerous protein tyrosine phosphatases (7, 8). However, structural and biochemical studies have shown that despite sharing an HCxxxxxR(S/T) motif, a phosphoenzyme intermediate and a core α/β-fold with other cysteine phosphatases, the precise mechanism of phosphoanhydride cleavage by these RTPases differs from the one used by protein phosphatases to hydrolyze phosphomonoesters (7, 8). The most important difference is the absence of a carboxylate general acid catalyst in metazoan and plant RTPases (8). Finally, the RTPases of this family are divalent cation-independent and are not able to hydrolyze NTPs.

Fungi, protozoa, and some DNA viruses possess an RTPase that belongs to the triphosphate tunnel family of metal-dependent phosphohydrolases that can also hydrolyze NTPs (9-14). These enzymes harbor two glutamate-containing motifs that are essential for catalysis and that coordinate the essential metal cation (9). The initial crystallization of the *S. cerevisiae* RTPase revealed a novel fold in which the catalytic core is located in a hydrophilic tunnel composed of eight antiparallel β-strands (15). Interestingly, this particular fold appears to be more widely distributed in the various taxa than initially expected, being found in archael and bacterial homologs, thus suggesting a deep evolutionary origin (16). The analysis of the crystal structure of the yeast RTPase revealed the presence of a single sulfate ion which is coordinated by the side chains of three essential amino acids (Arg393, Lys456, and Arg458). It was suggested that the side chain interactions with this sulfate ion reflect the contacts made by the protein with the γ-phosphate of the RNA or NTP

substrates (15). Numerous mutational studies have also contributed to the identification of a dozen additional residues that are essential for the enzymatic activity through their interactions with the divalent metal ion, or through their water-mediated contacts with either the metal ion or the sulfate ion (9, 17-19). More recently, analysis of the crystal structure of the RTPase component of mimivirus, a giant virus of amoeba, also revealed a minimized tunnel fold and an active site strikingly similar to the yeast enzyme (20). However, all past efforts to obtain structural information on the interaction between RTPases and their substrates were unsuccessful. This is perhaps not surprising since thermodynamic studies have shown that the binding of the RNA or nucleotide substrates to RTPases results in a destabilization of the enzymes (21, 22). Available crystals of the RTPases of the triphosphate tunnel metalloenzyme (TTM) family do not provide any information on the contacts between the enzyme and the triphosphate, sugar, or base moieties of the phosphohydrolase substrate.

In the present study, we used computational molecular docking to model the binding of a nucleotide substrate into the yeast RTPase active site. In order to confirm the generated model, and to gain additional insights into the molecular determinants involved in substrate recognition and catalysis, we also evaluated the phosphohydrolysis of an important number of nucleotide analogues. Our study highlights the importance of specific residues for the binding of the sugar, base, and triphosphate moieties of the nucleotide substrate, and reveals both the structural flexibility and complexity of the active site.

MATERIALS AND METHODS

Molecular docking

Docking calculations were carried out using the DockingServer software and the Dreiding force field was used for energy minimization of GTP using built-in Chemaxon tools in DockingServer (23). PM6 semi-empirical charges calculated by MOPAC2007 were added to the ligand atoms. Non-polar hydrogen atoms were merged, and rotatable bonds were defined (24). Docking calculations were carried out using the coordinated of the *S. cerevisiae* RNA triphosphatase (Protein Data Bank 1d8h). Essential hydrogen atoms, Kollman united atom type charges, and solvation parameters were added with the aid of AutoDock tools (25). Affinity (grid) maps of 20×20×20 Å grid points and 0.375 Å spacing were generated using the Autogrid program (25). AutoDock parameter set- and distance-dependent dielectric functions were used in the calculation of the van der Waals and the electrostatic terms, respectively. Docking simulations were performed using the Lamarckian genetic algorithm (LGA) and the Solis & Wets local search method (26). Initial position, orientation, and torsions of the ligand molecules were set randomly. Each docking experiment was derived from 2 different runs that were set to terminate after a maximum of 2,500,000 energy evaluations. The population size was set to 150. During the search, a translational step of 0.2 Å, and quaternion and torsion steps of 5 were applied.

Cet1 Expression and Purification

The Cet1 protein was expressed and purified as described before (21).

Competition assay

GTPase reactions were performed in reaction mixtures (20 μL) containing 50 mM Tris-HCl, pH 7.0, 5 mM DTT, 2 mM $MnCl_2$, 1 μM of the *S. cerevisiae* RTPase, and 20 μM [γ-^{32}P]GTP. The reactions were incubated for 15 min at 30 °C. Initially, the reactions were carried out both in the absence or presence of 100 μM of nucleotide analogues (or tripolyphosphate). The reactions were quenched by the addition of 5 μL of 5M formic acid. Aliquots of the reactions were spotted on polyethyleneimine-cellulose thin-layer

chromatography plates. The plates were developed in a solution of 1M formic acid and 0.5 M LiCl and the released inorganic phosphate was quantitated by scanning the plates with a PhosphorImager (Molecular Dynamics). The average of at least two single independent experiments is presented.

The IC_{50} values were evaluated by performing the standard GTPase assay in the presence of increasing concentrations of nucleotide analogues (or tripolyphosphate) ranging from 0 to 200 μM. The K_i values were determined by performing GTP assays with GTP concentrations ranging from 0 to 100 μM in the presence of 0, 5, 10, or 20 μM of analogues (or tripolyphosphate). The average of at least two single independent experiments is presented.

Phosphohydrolase assay

The reaction mixtures (20 μL) containing 50 mM Tris-HCl, pH 7.0, 5 mM DTT, 2 mM $MnCl_2$, 1 μM of the *S. cerevisiae* RTPase, and various concentrations of substrates (GTP, nucleotide analogues, or tripolyphosphate) were incubated for 15 min at 30 °C. The reactions were quenched by the addition of 400 μL of malachite green reagent (BIOMOL Research Laboratories). The released inorganic phosphate was measured by monitoring the A_{620}. The values were extrapolated to a standard curve for phosphate. Background levels of contaminating phosphate were subtracted in all cases.

RESULTS AND DISCUSSION

Molecular docking

Although important structural information is available from the currently available crystal structures of the members of the TTM family, the substrates (RNA or NTP) are conspicuously absent from all these structures (15, 20). We set out to initially use the power of molecular docking to provide information on the interaction between the yeast RTPase and a nucleotide substrate. Guanosine triphosphate (GTP) was selected as the substrate since this purine nucleotide is frequently encountered as the initiating nucleotide in eukaryotic mRNAs (27). Extensive computational docking and structure optimizations were then used to generate a model of the enzyme-GTP complex. More than 2,000,000 energy evaluations were performed in order to provide an accurate description of the enzyme-substrate interactions. The model underwent 150 rounds of steepest descent energy minimization, and did not contain energetically unfavourable bonds, angles, or torsions.

The molecular docking model predicts that the GTP substrate is located in the middle of the tunnel structure of the yeast RTPase (Fig. 1). The space-filling analysis suggests that the tunnel can accommodate the ribose, base, and phosphate moieties of the GTP molecule, thereby implying that the entire nucleotide, and not only the phosphates, is entering the tunnel (Fig. 1E). The molecular docking model provides instructive findings on the interaction between specific residues and the nucleotide substrate. For instance, multiple side chains (Summarized in Table 1) are contacting the α, β, and γ phosphates of GTP. In addition to the Arg393, Lys456, and Arg458 residues that were previously observed in the crystal structure of the *S. cerevisiae* RTPase (15), six other amino acids appear to be involved in the coordination of the phosphates (Fig. 2). Some of these residues, such as Arg393, Lys409, Lys456, and Glu494 are solely contacting one phosphate, while others (Glu305, Glu433, Arg458, Asp471, and Glu496) are contacting two adjacent phosphates (Fig. 2A-D). Interestingly, all of these phosphate-contacting residues were previously shown to be essential for catalysis through mutational studies thereby highlighting their importance in catalysis (9, 17-19). Based on the crystal structure of the enzyme (15), some of these essential residues were previously proposed to be important for catalysis through their ability to make water-mediated contacts with the phosphate or the essential divalent metal ion (Glu433 and Glu494). Similarly, other residues such as Lys409 were thought to be indirectly involved in catalysis via their interactions with other essential side chains (15, 17). Although the role of these amino acids in making important interactions unrelated to the binding of phosphate cannot be excluded, the current model clearly reveals that the side chains of these amino acids are involved in the coordination of the triphosphate

moiety of the nucleotide substrate. The previously observed ability of the TTM family of RTPases to hydrolyze tripolyphosphate (28-30), which completely lacks the sugar and base components of the nucleotide, is probably a reflection of the excessive number of interactions between the active site residues and the α-, β-, and γ-phosphates.

Analysis of the docking model between the enzyme and GTP suggests that only four amino acids are responsible for the coordination of the hydroxyl groups of the ribose. These are Glu307 and Lys409, which are contacting the 3'OH group, and Arg393 and Arg454, which are coordinating the 2'OH group of the sugar (Fig. 2A and 2D; Table 1). Previous mutational studies have shown that each of these four residues is essential for catalysis (9, 17-19). The molecular docking model also highlights key amino acids involved in the binding of the guanine base. These amino acids are either contacting the exocyclic 2-amino (Ser429, and Asn431) or the 6-oxo (Lys427) groups of the pyrimidine ring of guanine (Fig. 2A and 2D). Previous structure-function analyses of the amino acids that are contacting the guanine base in the docking model indicate that these residues are not essential for catalysis (17, 19). Moreover, pi-pi and cation-pi stacking interactions are also occurring between His411 (non-essential) and the pyrimidine ring of guanine, while Pro341 is engaged in hydrophobic interactions with the imidazole ring of guanine (Fig. 2A and 2C). The non-essentiality of the amino acids contacting the guanine base is not surprising since the enzyme can efficiently hydrolyze both purine and pyrimidine nucleotides. Both the coordination geometry and the nature of the amino acids contacting the purine/pyrimindine rings are likely modified according to the precise nature of the substrate.

Nucleotide analogues to probe the active site

In order to experimentally confirm the docking model, and to gain additional insights on the molecular determinants involved in the formation of the enzyme-substrate complex, we have used nucleotide analogues to monitor their effects on the reaction chemistry. The nucleotide analogues displayed various modifications on both the ribose and the guanine base of GTP (Fig. 3). We initially monitored the ability of 17 analogues of GTP to inhibit the activity of the yeast RTPase by evaluating both the IC_{50} and K_i values for each molecule (Table 2). The informative finding is that every analogue tested had the ability to inhibit the RTPase reaction, albeit to different extents, thus highlighting the high structural flexibility of the active site (Fig. 4A). All the nucleotide analogues used in the current study were competitive inhibitors of the RTPase reaction indicating that they bind to the active site of the enzyme. A typical example using the analogue A-12 is shown in figure 4B-E. In order to gain additional information on the functional flexibility of the catalytic center, we determined the ability of the various nucleotide analogues to be hydrolyzed by the enzyme. The

yeast RNA triphosphatase catalyzed the phosphohydrolysis of all nucleotide analogues tested in the current study with specific activity ranging from 7 to 138 % of GTP (Table 2). Interestingly, some of the analogues bound strongly to the enzyme active site (as evidenced from the low IC_{50} and K_i values) but were not efficiently hydrolyzed. These analogues (A_4, A_8, A_{11}, A_{18}, A_{20}, and A_{24}) had IC_{50} values ranging from 2 to 16 μM with low specific phosphohydrolase activities varying from 7 to 18% of the GTP substrate (Table 2).

Our study indicates that the active center of the yeast RTPase is highly flexible and can accommodate nucleotide substrates displaying a number of unusual chemical modifications. For instance, analogues harboring modifications on the hydroxyl moieties of the ribose (2'OH and 3'OH) were hydrolyzed with a high level of efficiency by the enzyme. The addition of a methyl group to either the 2' or 3' hydroxyls (A-13 and A-17) had no significant effect on the phosphohydrolase activity (Table 2). Most strikingly, A-12 which lacks both ribose hydroxyls (2',3'-dideoxy-GTP) was hydrolyzed very efficiently by the yeast RTPase. This was unexpected since amino acids contacting the ribose hydroxyls (Glu307, Arg393, Lys409, Arg454) were previously shown to be essential for catalysis through mutational studies (9, 17-19). However, we observe that two of these residues (Arg393 and Lys409) are also contacting the α-phosphate of the bound GTP, while Glu307 coordinates the essential divalent cation. Moreover, Arg454 forms a salt bridge with Glu492, an interaction which is important for the stabilization of the tunnel architecture (15, 17). Therefore, we conclude that the importance of these amino acids is not directly related to their ability to bind to the hydroxyls of the ribose, but rather lies in their other functions namely through the coordination of the α-phosphate and metal ion, or in the stabilization of the tunnel structure.

Some of the analogues used in the current study contain chemical modifications on the guanine base of the GTP molecule (Fig. 3). Although the *S. cerevisiae* RTPase is active on both purine and pyrimidine nucleotides (ATP, CTP and UTP were used as substrates with specific activities of 103, 103, and 109 % of the GTP substrate, respectively), analysis of the GTP analogues helped to illuminate the flexibility and complexity of the RTPase active site. For instance, the addition of a chemical group to the C8 position of guanine had a negative effect on the phosphohydrolysis activity. Analogues harboring such modification (A-10 and A-24) were used inefficiently by the enzyme. Although it can be argued that the addition of a bromo- or iodo- group at this position can potentially alter the electronic properties of the guanine ring, analysis of the docking model indicates that steric hindrance is the likely explanation for the limited hydrolysis of these substrates. Analysis of the enzyme-GTP model indicates that the space is occupied by Lys474

and Glu476 of the β10 strand that comprises part of the walls of the tunnel. The presence of halogen elements with large atomic radius at the C8 position of the guanine base reveals the conformational importance of amino acids not directly involved in catalysis. The importance of these amino acids in forming an optimal nucleotide binding pocket could not be inferred from previous mutational analyses. Similarly, the addition of a methyl group at the N1 position of guanine (A-4, N1-methyl-GTP) decreases the catalytic activity of the enzyme by 5-fold (Table 2). In our docking model, Lys427 occupies the space that is filled by the additional methyl group thereby hindering the optimal positioning of the substrate and the concomitant hydrolysis of the analogue (Fig. 2A, 2C, and 2D).

The 6-oxo group of the guanine ring also appears important for substrate recognition and hydrolysis. The phosphohydrolysis activity of all the analogues which harbored chemical modifications at this position (A-5, A-6, A-9, A-19) was lower as compared to the hydrolysis of GTP (Table 2). We initially suspected that the inability of these analogues to form hydrogen bonds through this position was the likely explanation to explain their low catalytic usage. However, Lys427 is the only proximal amino acid which can potentially coordinate the 6-oxo group of guanine through hydrogen-bonding, and previous studies have shown that the substitution of this residue by alanine has no effect on the catalytic activity. Steric hindrance might again be a factor since analogues with larger substituents (A-5, O6-methyl-GTP; A-9, 6-thio-GTP; and A-19, 6-thio-methyl-GTP) were less hydrolyzed relative to A-6 (6-chloro-GTP) which harbored a smaller substituent at the same position on the guanine ring. Closer examination of the enzyme-GTP docking model indicates that three amino acids are in the vicinity of the 6-oxo group of guanine (Fig. 5A). These are His338 and Thr340 of the β2 strand and Thr489 of the β11 strand, which are just too far removed to hydrogen-bond with the 6-oxo group but might interfere with the larger substituents displayed by the analogues. The importance of these amino acids was therefore investigated by generating three distinct enzymatic mutants. His338, Thr340, and Thr489 were individually substituted for alanine, and the mutant polypeptides were expressed and purified in parallel with the wild-type enzyme (Fig. 5B). The GTP substrate was hydrolyzed efficiently by the mutants polypeptides as evidenced from the specific activities of 98, 69, and 79 % of wild-type activity for the H338A, T340A, and T489A mutants, respectively (data not shown). The informative finding is that the replacement of these lateral side chains by alanine resulted in an increase in the phosphohydrolysis of analogues harboring larger substituents at the 6-oxo position (A-5, A-9, and A-19). For instance, the relative hydrolysis of 6-thio-methyl-GTP (A-19) increased from 0.38 for the wild-type enzyme to 0.60, 0.85, and 0.81 for the H338A, T340A, and T489A mutants, respectively (Fig. 5C). A similar increase in phosphohydrolysis was also observed for both the A-5

(O6-methyl-GTP) and A-19 (6-thio-methyl-GTP) analogues when the His338, Thr340, and Thr489 lateral side chains were individually replaced by alanine (Fig. 5C). However, these substitutions had no positive effect on the phosphohydrolysis of A-6 (6-chloro-GTP) which harbored a smaller substituent at the same position on the guanine ring. Overall, the mutagenesis data confirmed that steric hindrance prevents the phosphohydrolysis of analogues harboring larger substituents at the 6-oxo position.

One of the most striking features of the GTP-enzyme complex formation is the importance of the 2-amino group of guanine. In that regard, A-11, A-18, and A-20 are particularly interesting since they only differ with A-6, A-9, and A-19 by the lack of the 2-amino moiety (Fig. 3). The absence of the amino group drastically reduces the hydrolysis of the analogues (compare A-6 with A-11, A-9 with A-18, and A-19 with A-20) highlighting its importance for efficient catalysis (Table 2). The importance of this functional group is also underscored by the relative inability of the enzyme to hydrolyze A-8 which only differs with GTP by harboring a 6-oxo group at this position instead of the amino group (Table 2). The most likely explanation for the importance of the 2-amino group is that it can act as hydrogen-bond donor with specific amino acids. Analysis of our docking model with bound GTP reveals that two amino acids (Ser429 and Asn431) are potentially contacting the 2-amino group of guanine through hydrogen-bonding. We hypothesize that Ser429 is likely the key amino acid responsible for coordination of the 2-amino group. Previous conservative substitution analyses have revealed the importance of the hydroxyl moiety of the lateral side chain of Ser429 for the activity of the protein (18). Moreover, earlier *in vivo* studies have indicated that the substitution of this residue by alanine elicits a cold-sensitive phenotype, thereby highlighting the importance of this residue in the proper folding of the enzyme (18). As for Asn431, previous mutational analyses have shown that it is not critical for hydrolysis (17), and its distance from the 2-amino group suggests that it makes a rather weak hydrogen bond with guanine.

The RTPase component of the capping machinery is an attractive target for the future development of inhibitors targeting pathogenic fungi, viruses, and protozoans since these RTPases are structurally and mechanistically different from their human counterpart. Some of the analogues identified in the current study appear as good starting blocks for the design of more specific inhibitors. These analogues bound very tightly to the yeast RTPase active site but were not efficiently hydrolyzed. In that respect both A-20 and A-11, which are structurally related with a substitution of the 6-oxo group and the lack of the 2-amino group of guanine, possess high inhibitory potential (Table 2). Although tripolyphosphate, which was previously shown to inhibit other RTPases of the TTM family (31), displays even higher inhibitory potential *in vitro*, it is

unlikely that such a molecule could eventually display any specificity towards pathogenic RTPase *in vivo*. Because our study revealed the intrinsic flexibility of the RTPase active site, we believe that additional chemical modifications of A-20 and A-11 could lead to molecules displaying a very robust and specific inhibition of pathogenic RTPases.

CONCLUSION

The current study contributes to highlight both the complexity and flexibility of the TTM family active site. It is clear that the RTPase reaction requires a very precise alignment between the active site residues, the substrate, and the metal ion cofactor. Two observations support this conclusion. First, the hydrolysis of RNA triphosphate ends is activated by magnesium, but not by manganese or cobalt, whereas the NTPase activity is supported by manganese and cobalt, but not magnesium. Second, unnatural substrates such as nucleotide analogues or tripolyphosphate can still be hydrolyzed, albeit very inefficiently, even if the conformation of the active site is less than optimal, lacking several important contacts with the substrate. In summary, our structural docking model, coupled with the use of an important number of nucleotide analogues highlights the importance of specific residues for the binding of the nucleotide substrate, and reveals both the structural flexibility and complexity of the active site. Our study illustrates the important structural and functional features for the interaction of an RTPase with a ligand, and opens the way to the use of nucleotide analogues as potential inhibitors of pathogens.

REFERENCES

1. Shatkin,A.J. (1976) Capping of eucaryotic mRNAs. *Cell* **9**, 645-653.

2. Furuichi,Y. and Shatkin,A.J. (2000) Viral and cellular mRNA capping: past and prospects. *Adv. Virus Res.* **55**, 135-184.

3. Shuman,S. and Schwer,B. (1995) RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. *Mol. Microbiol.* **17**, 405-410.

4. Bisaillon,M. and Lemay,G. (1997) Viral and cellular enzymes involved in synthesis of mRNA cap structure. *Virology* **236**, 1-7.

5. Shuman,S. (2001) Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog. Nucleic. Acids Res. Mol. Biol.* **66**, 1-40.

6. Shuman,S. (2002) What messenger RNA capping tells us about eukaryotic evolution. *Nat. Rev. Mol. Cell Biol.* **3**, 619-625.

7. Tagaki,T., Moore,C.R., Diehn,F. and Buratowski,S. (1997) An RNA 5'-triphosphatase related to the protein tyrosine phosphatases. *Cell* **89**, 867-873.

8. Changela,A., Ho,C.K., Martins,A., Shuman,S. and Mondragon,A. (2001) Structure and mechanism of the RNA triphosphatase component of mammalian mRNA capping enzyme. *EMBO J.* **20**, 2575-2586.

9. Ho,C.K., Pei,Y. and Shuman,S. (1998) Yeast and viral RNA 5' triphosphatases comprise a new nucleoside triphosphatase family. *J. Biol. Chem.* **273**, 34151-34156.

10. Ho,C.K. and Shuman,S. (2001) A yeast-like mRNA capping apparatus in *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3050-3055.

11. Hausmann,S., Altura,M.A., Witmer, M., Singer,S.M., Elmendorf,H.G. and Shuman, S. (2005) Yeast-like mRNA capping apparatus in *Giardia lamblia*. *J. Biol. Chem.* **280**, 12077-12086.

12. Hausmann,S., Vivares,C.P. and Shuman,S. (2002) Characterization of the mRNA capping apparatus of the microsporidian parasite *Encephalitozoon cuniculi*. *J. Biol. Chem.* **277**, 96-103.

13. Myette,J. and Niles,E.G. (1996) Characterization of the vaccinia virus RNA 5'-triphosphatase and nucleotide triphosphate phosphohydrolase activities. Demonstrate that both activities are carried out at the same active site. *J. Biol. Chem.* **271**, 11945-11952.

14. Martins,A. and Shuman,S. (2001) Mutational analysis of baculovirus capping enzyme Lef4 delineates an autonomous triphosphatase domain and structural determinants of divalent cation specificity. *J. Biol. Chem.* **276**, 45522-45529.

15. Lima,C.D., Wang,L.K. and Shuman,S. (1999) Structure and mechanism of yeast RNA triphosphatase: an essential component of the mRNA capping apparatus. *Cell* **99**, 533-543.

16. Gong,C., Smith,P. and Shuman,S. (2006) Structure-function analysis of Plasmodium RNA triphosphatase and description of a triphosphate tunnel metalloenzyme superfamily that includes Cet1-like RNA triphosphatases and CYTH proteins. *RNA* **12**, 1468-1474.

17. Bisaillon,M. and Shuman,S. (2001) Structure-function analysis of the active site tunnel of yeast RNA triphosphatase. *J. Biol. Chem.* **276**, 17261-17266.

18. Bisaillon,M. and Shuman,S. (2001) Functional groups required for the stability of yeast RNA triphosphatase *in vitro* and *in vivo*. *J. Biol. Chem.* **276**, 30514-30520.

19. Pei,Y., Ho,C.K., Schwer,B. and Shuman,S. (1999) Mutational analyses of yeast RNA triphosphatases highlight a common mechanism of metal-dependent NTP hydrolysis and a means of targeting enzymes to pre-mRNAs *in vivo* by fusion to the guanylyltransferase component of the capping apparatus. *J. Biol. Chem.* **274**, 28865-28874.

20. Benarroch,D., Smith,P. and Shuman,S. (2008) Characterization of a trifunctional mimivirus mRNA capping enzyme and crystal structure of the RNA triphosphatase domain. *Structure* **16**, 501-512.

21. Bisaillon,M. and Bougie,I. (2003) Investigating the role of metal ions in the catalytic mechanism of the yeast RNA triphosphatase. *J. Biol. Chem.* **278**, 33963-33971.

22. Bougie,I. and Bisaillon,M. (2006) Inhibition of a metal-dependent viral RNA triphosphatase by decavanadate. *Biochem. J.* **398**, 557-567.

23. Mayo,S.L., Olafson,B.D. and Goddard,W.A. (1990) DREIDING: a generic force field for molecular simulations *J. Phys. Chem.* **94**, 8897-8909.

24. Stewart,J.P.P. (2007) Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Modeling* **13**, 1173-1213.

25. Morris,G.M., Goodsell,D.S., Halliday,R.S., Huey,R., Hart,W.E., Belew,R.K. and Olson,A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639-1662.

26. Solis,F.J. and Wets,R.J.B. (1981) Minimization by Random Search Techniques. *Math. Operat. Res.* **6**, 19-30.

27. Frith,M.C., Valen,E., Krogh,A., Hayashizaki,Y., Carninci,P. and Sandelin,A. (2008) A code for transcription initiation in mammalian genomes. *Genome Res.* **18**, 1-12.

28. Pei,Y., Schwer,B., Hausmann,S. and Shuman,S. (2001) Characterization of *Schizosaccharomyces pombe* RNA triphosphatase. *Nucleic Acids Res.* **29**, 387-396.

29. Keppetipola,N., Jain,R. and Shuman,S. (2007) Novel triphosphate phosphohydrolase activity of *Clostridium thermocellum* TTM, a member of the triphosphate tunnel metalloenzyme superfamily. *J. Biol. Chem.* **282**, 11941-11949.

30. Jain,R. and Shuman,S. (2008) Polyphosphatase activity of CthTTM, a bacterial triphosphate tunnel metalloenzyme. *J. Biol. Chem.* **283**, 31047-31057.

31. Gong,C. and Shuman,S. (2002) *Chlorella* virus RNA triphosphatase. Mutational analysis and mechanism of inhibition by tripolyphosphate. *J. Biol. Chem.* **277**, 15317-15324.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Canadian Institutes for Health Research (CIHR). The RNA group is supported by a grant from the CIHR and by the Université de Sherbrooke. M.B. is a Chercheur Boursier Junior 2 from the Fonds de recherché en santé du Québec and a member of the Infectious diseases group of the Centre de Recherche Clinique Étienne-Lebel.

FUNDING

This work was funded by a grant from the Canadian Institutes for Health Research. MB is a Chercheur Boursier Junior 2 from the Fonds de recherche en santé du Québec and a member of the Infectious diseases group of the Centre de Recherche Clinique Étienne-Lebel.

FIGURE LEGENDS

Figure 1. Molecular docking model for the binding of GTP to the yeast RNA triphosphatase. (**A**) Ribbon diagram looking into the tunnel exit of the enzyme with bound GTP. Numerous amino acids are interacting with the nucleotide. (**B**) A view looking into the triphosphate tunnel entrance. (**C**) Side view of the enzyme following a 90° rotation to the left depicting both the entrance (*left*) and exit (*right*) of the tunnel. (**D**) A three-quarter exploded view of the tunnel with bound GTP. (**E**) Space-filling surface view looking into the tunnel entrance of the enzyme with bound GTP.

Figure 2. Active site of the yeast RNA triphosphatase. The GTP molecule is coordinated by an elaborate network of interactions. The side chains of an important number of amino acids are contacting the phosphates, ribose, and guanine base of the substrate. Four different views are depicted in order to observe the various interactions.

Figure 3. Nucleotide analogues used in the current study. The GTP analogues harbored various modifications on both the ribose and the guanine base of GTP.

Figure 4. Inhibition of the phosphohydrolase activity by nucleotide analogues. (**A**) The standard GTPase activity was performed in the presence of 20 uM [γ-^{32}P]GTP. The various nucleotide analogues were added at a single concentration of 100 uM and the inhibition of the GTPase activity was evaluated by monitoring the release of radiolabeled inorganic phosphate which was separated from the GTP substrate by thin-layer chromatography. (**B**) Dose-response inhibition of the GTPase activity by A_{12} (2',3'-dideoxy-GTP). (**C** and **D**) Competitive inhibition of the GTPase reaction by A_{12}. The GTPase activity was evaluated in the absence (■) or presence of 5 (□) or 10 μM (O) of A_{12}. (**E**) Dixon plot of the inhibition performed at each fixed concentration of GTP substrate, 5 (■), 10 (▲), or 50 mM (□).

Figure 5. Steric hindrance caused by analogues harboring large substituents at the 6-oxo position of the guanine ring of GTP. (**A**) The His338, Thr340, and Thr489 residues of the enzyme are in the vicinity of the 6-oxo group of the GTP substrate. (**B**) Aliquots (2 μg) of the purified wild-type (WT, lane 1), H338 A (lane 2), Thr340A (lane 3), and Thr489A (lane 4) mutant proteins were analyzed by electrophoresis through a 12.5% polyacrylamide gel containing 0.1% SDS, and visualized with Coomassie Blue Dye. The positions and sizes (in kDa) of the size markers are indicated on the left. (**C**) Normalized

phosphohydrolase activities of the wild-type (WT), H338A, T340A, and T489A mutants. The phosphohydrolase specific activities were calculated from the slopes of the titration curves and normalized to the specific activity for the hydrolysis of GTP by the wild-type enzyme. GTP, O6-methyl-GTP (A-5), 6-chloro-GTP (A-6), 6-thio-GTP (A-9), and 6-thio-methyl-GTP (A-19) were used as substrates.

Table 1. Key interactions between the active site residues of the yeast RNA triphosphatase and GTP as predicted by the molecular docking model.

Amino acid	Interaction	Distance (Å)
Lys427	O6 of guanine	3.9
Ser429	NH_2 of guanine	2.6
Asn431	NH_2 of guanine	3.9
Glu307	3'OH of ribose	4.4
Arg393	2'OH of ribose	2.6
Lys409	3'OH of ribose	3.5
Arg454	2'OH of ribose	3.9
Glu305	β-PO_4 and γ-PO_4	3.2 and 3.7
Arg393	α-PO_4	3.4
Lys409	α-PO_4	3.2
Glu433	β-PO_4 and γ-PO_4	3.1 and 3.6
Lys456	α-PO_4	3.6
Arg458	α-PO_4 and β-PO_4	3.2 and 2.9
Asp471	α-PO_4 and β-PO_4	4.1 and 4.4
Glu494	α-PO_4	4.3
Glu496	β-PO4 and γ-PO4	3.3 and 3.9

Table 2. Inhibition of the GTPase activity by nucleotide analogues.

Molecule	IC_{50} (μM)	K_i (μM)	Phosphohydrolase specific activity (% of GTP) [a]	Inhibitory potential [b]
GTP	87	16	100	1.0
dGTP	76	14	99	1.1
A_3	83	15	88	1.2
A_4	8.1	1.5	18	60
A_5	150	27	17	3.4
A_6	53	9.6	70	2.4
A_8	10	1.8	10	86
A_9	16	2.9	41	13
A_{10}	93	17	20	4.7
A_{11}	2.0	0.4	16	268
A_{12}	6.1	1.1	138	10
A_{13}	32	5.9	109	2.5
A_{15}	49	8.8	111	1.6
A_{17}	59	11	116	1.3
A_{18}	8.1	1.5	12	89
A_{19}	61	11	38	3.8
A_{20}	2.0	0.4	7.2	597
A_{22}	113	21	81	0.9
A_{24}	8.1	1.5	7.2	149
Tripoly-PO_4	0.4	0.1	3.6	5965

[a] The phosphohydrolase specific activities were calculated from the slopes of the titration curves and normalized to the specific activity for the hydrolysis of GTP.

[b] The inhibitory potential is defined by the following equation:
$(K_{i\ GTP}/K_{i\ Molecule}) \times (Spec.\ activity_{\ GTP}/Spec.\ activity_{\ Molecule})$

Figure 1. Issur *et al.* 2009

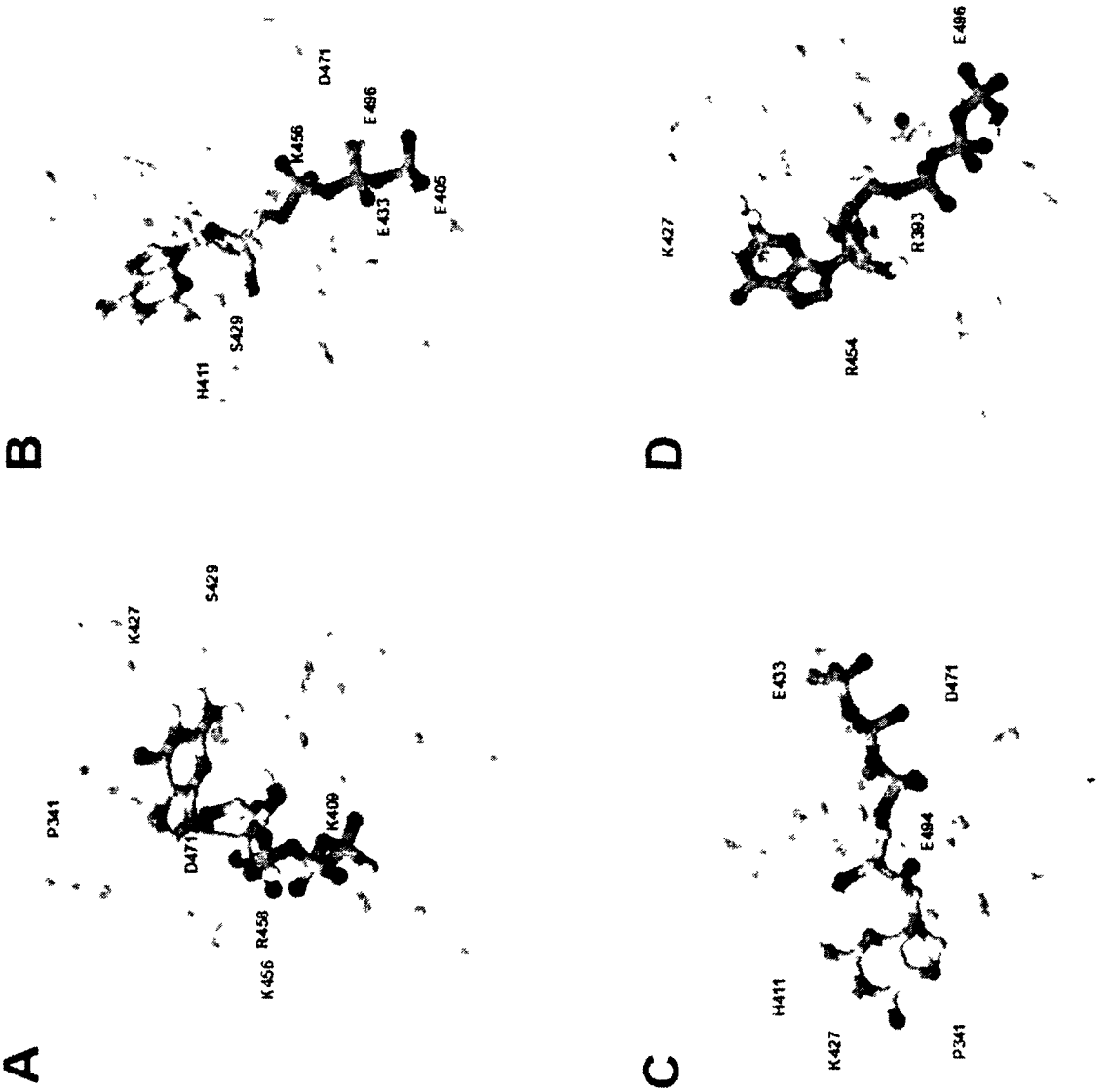


Figure 2. Issur *et al*. 2009

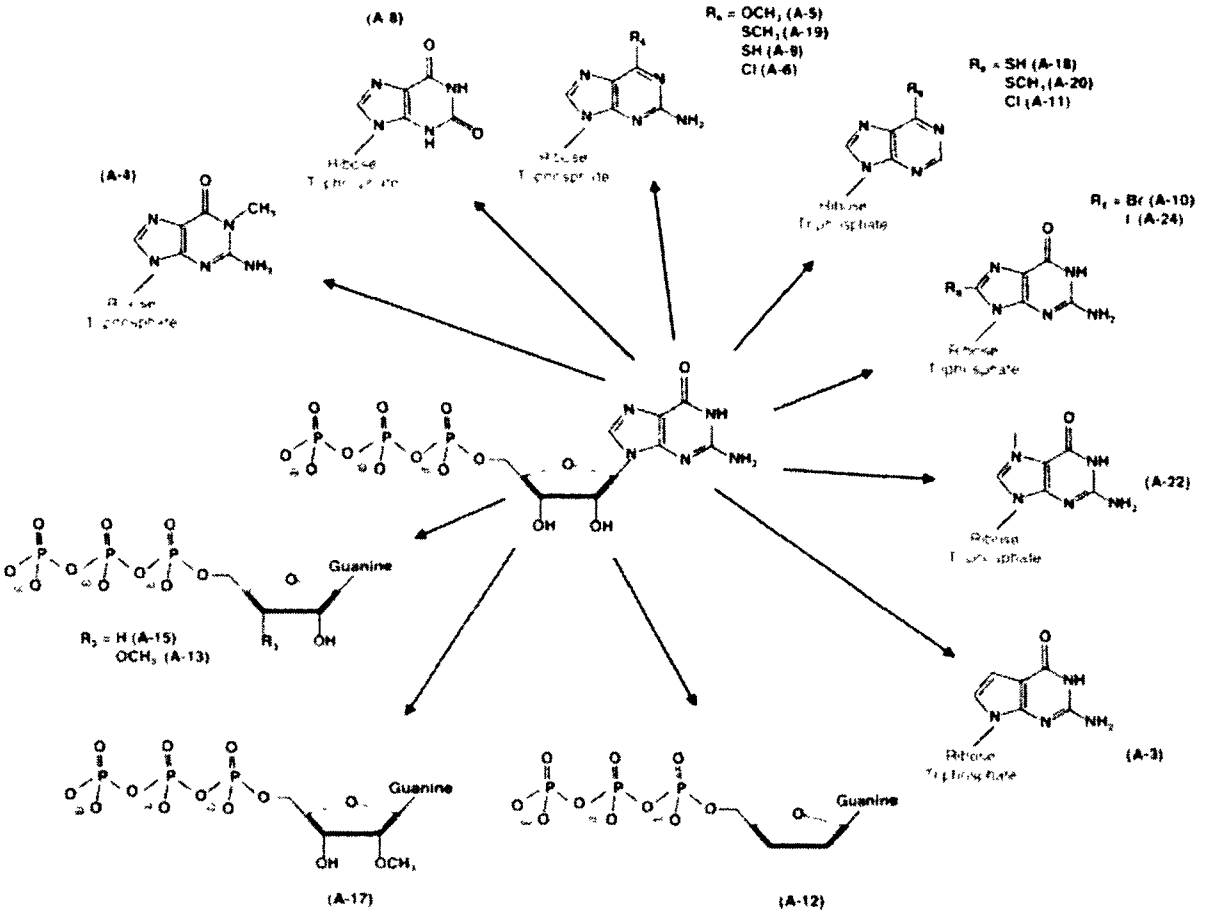


Figure 3. Issur *et al.* 2009

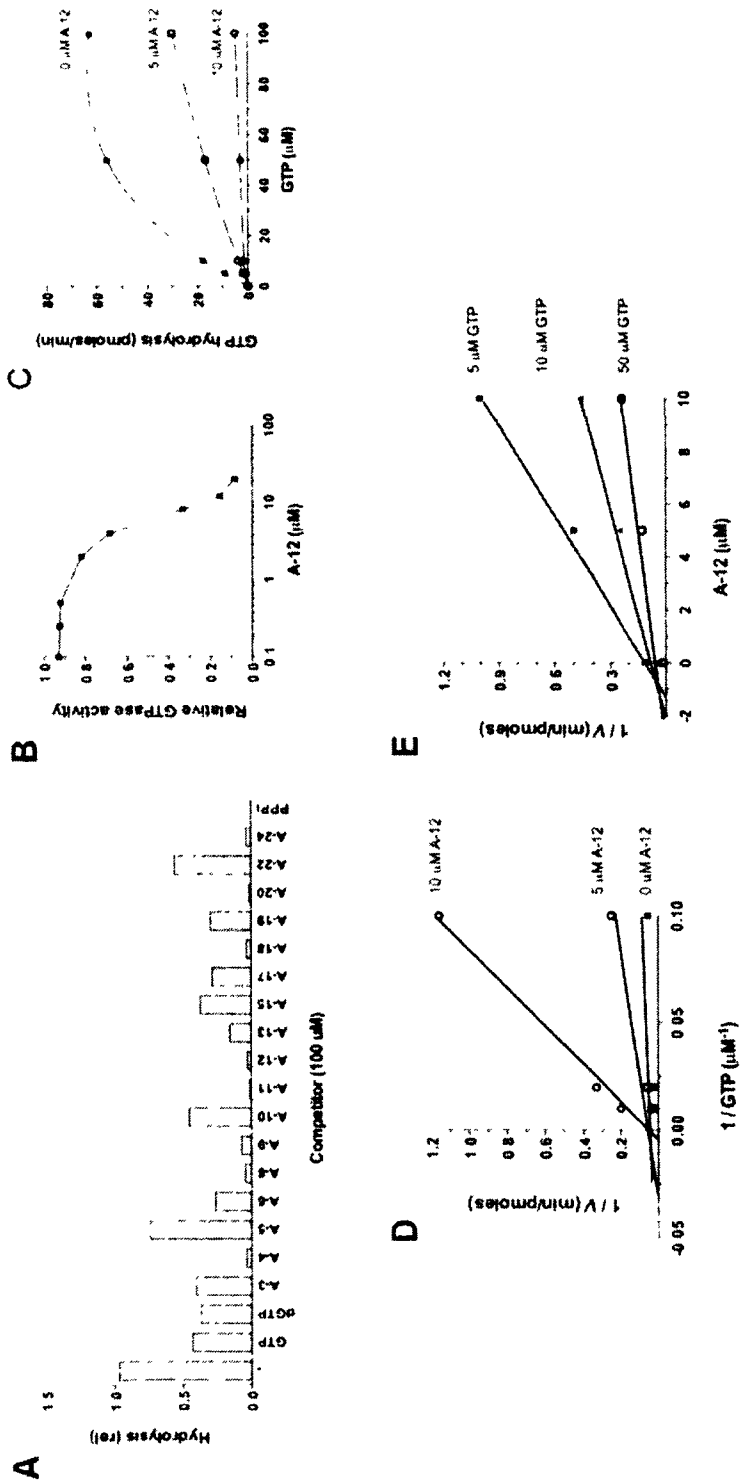


Figure 4. Issur *et al.* 2009

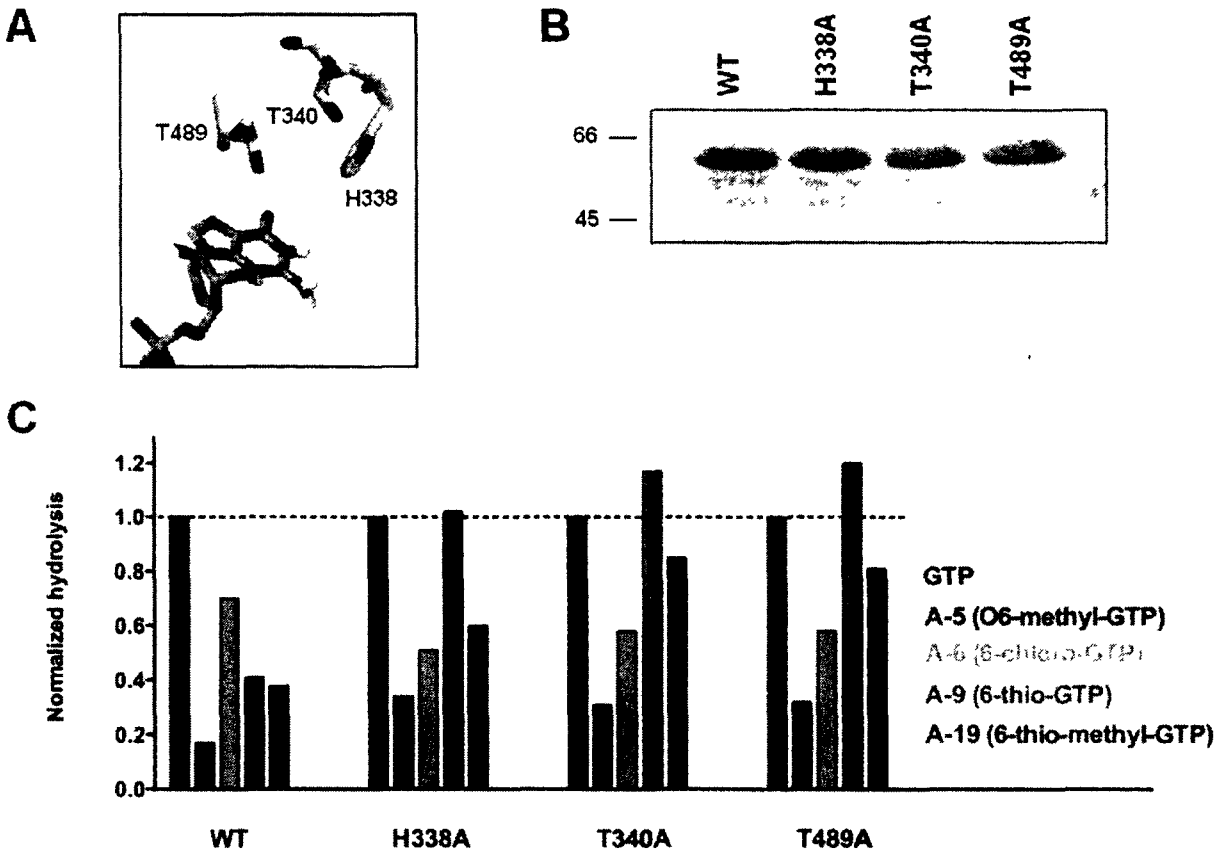


Figure 5. Issur *et al.* 2009

Chapter II –Probing into the GTP specificity of a model RNA guanylyltransferase

2.2. ARTICLE:

Biosynthesis of novel RNA cap analogues reveals new insights into the translation process

Moheshwarnath Issur, Simon Despins, Isabelle Bougie and Martin Bisaillon

Article submitted in Nucleic Acids Research, March 2011,

CONTRIBUTIONS

I performed most of the experiments, analysed all the results and prepared the whole manuscript. SD contributed to figure 3B of the manuscript by expressing and purifying the eIF4E protein and performing RNA binding assays. IB generated the expression vector of the PBCV-1 RNA guanylyltransferase and RNA triphosphatase. MB provided the funding and the necessary guidance for the completion of the project.

SUMMARY ***(en français)***

En utilisant la même méthodologie que dans l'étude précédente, nous avons effectué une étude approfondie du site actif de l'ARN guanylyltransférase modèle de PBCV-1, la protéine A103R. Bien que, cette enzyme modèle a été co-cristallisée avec le GTP et le GMP dans son site actif, le rationnel de la spécificité de cette protéine envers le GTP reste un sujet de débat. En effet, même si plusieurs interactions entre l'enzyme et le groupement guanine de GTP ont été mises en évidence, l'incapacité de moduler la spécificité de l'enzyme par le biais des mutations ponctuelles indique qu'il y a des lacunes dans notre compréhension du fonctionnement de ces enzymes. Dans cette étude, grâce à l'utilisation des analogues de purines, notre intention est de fournir une image plus détaillée des diverses interactions qui entrent en jeu lorsque l'enzyme se lie au GTP, catalyse la formation d'un intermédiaire covalent avec le GMP ainsi que le transfert de ce GMP à un ARN accepteur. Ainsi nous avons pu démontrer que la Lys188, en établissant des ponts hydrogène avec le O6 et le N7 participe à médier la spécificité de l'enzyme envers le GTP. La Lys188 est un résidu conservé chez toutes les ARN guanylyltransférases conventionelles et est absent chez les ADN/ARN ligases dépendantes de l'ATP. La conversion de ce résidu en alanine abolit toute activité de l'enzyme quoique la liaison au GTP ne soit pas affectée. Nous avons aussi démontré que le 3'OH de la molécule de GMP est fondamental pour la coordination de l'ARN accepteur pour permettre à la deuxième étape de la réaction d'ARN guanylyltransférase d'avoir lieu. En plus de plusieurs autres grandes conclusions que cette étude a établies, nous avons également été en mesure de générer plusieurs molécules d'ARN possédant des structures coiffes artificielles. Notre objectif ultime est d'évaluer l'impact de ces structures d'ARN modifiées sur divers processus cellulaires. Jusqu'à présent, nous nous sommes concentrés sur la capacité de ces structures coiffes modifiées à soutenir la traduction des ARN. L'affinité de ces nouveaux analogues de la coiffe avec la protéine eIF4E, un des facteurs clés impliqués dans la reconnaissance par la machinerie traductionnelle a aussi été évalué. Nos résultats démontrent qu'en présence de certaines modifications compensatoires sur la structure coiffe, l'absence du groupement méthyle (essentiel pour la reconnaissance de la coiffe naturelle par eIF4E), situé à la position N7, n'a pas de répercussion négative sur les propriétés traductionnelles d'un ARN.

SUMMARY

Using the same methodology as in the previous study, we have performed a thorough investigation into the active site of the model PBCV-1 RNA guanylyltransferase, the A103R protein. Although, this model enzyme has been co-crystallised with both GTP and GMP in its active site, the rationale behind its specificity remains a matter of debate. Indeed, even though several interactions of the enzyme with the guanine moiety of GTP has been pinpointed to contribute to the guanine specificity of RNA guanylyltransferases, the inability to modulate the specificity of the enzyme through point mutations is indicative of only a partial understanding of the enzyme's interactions with its substrates. Therefore, through the use of several purine analogues our intent is to provide a broader picture of the various interactions coming into play when the enzyme binds to GTP, catalyzes the formation of a covalent intermediate with GMP as well as transfer this bound molecule onto an acceptor RNA. Thus we have been able to show that Lys188, through hydrogen bond type of interactions with both the O6 and partially with the N7 position of guanine, mediates the GTP specificity of the enzyme. Lys188 is a conserved residue among all conventional RNA guanylyltransferases known and is absent in all known ATP-dependent RNA/DNA ligases. Alanine conversion of this residue abolishes all activities of this enzyme. We have also demonstrated that the 3'OH of guanosine, while relatively unimportant for the first step of the reaction, is fundamental for the coordination of the acceptor RNA molecule to enable the second step of the RNA guanylyltransferase reaction. In addition to several other major conclusions that this study led to, we were also ultimately able to generate, from a biological perspective, several RNA molecules possessing unnatural RNA cap structures. Our future aim is to evaluate the impact of these modified RNA cap structures on several key cellular events. Up until now, we have concentrated ourselves on the ability of these modified RNA cap structures to support the translatability of mRNAs. We also monitored the binding affinity of these novel cap structures to the eIF4E protein. The novel cap structures revealed that in the light of compensatory modifications on either the base or the ribose moiety of the capped residue, the presence of the N7-methyl group in the classical N7-methyl guanosine cap is not indispensible. Overall, this study highlights the importance of several features of the RNA cap structure itself in ensuring the various functions the RNA cap structure is subjected to.

Abstract

The incorporation of synthetic nucleotide analogues in RNAs has been of considerable use for the study of various aspects of RNA metabolism, most notably the steps of pre-mRNA splicing, mRNA turnover and the translation process. Of particular interest have been RNAs capped with synthetic analogues. These are traditionally generated by chemically synthesising cap dinucleotide analogues, which are incorporated at the 5' ends of RNAs during *in vitro* transcription. In this study we describe a simple and fast method to generate RNAs capped with modified nucleotides by using a purified RNA capping enzyme and purine analogues. Thus, we identified several key structural determinants at each step of the RNA capping reaction and generated RNAs harbouring several novel 5' cap analogues. Moreover, we monitored the binding affinity of these novel cap structures to the eIF4E protein and evaluated their translational properties *in cellulo*. The novel cap structures revealed that in the light of compensatory modifications on either the base or the ribose moiety of the capped residue, the presence of the essential N7-methyl group in the classical N7-methyl guanosine cap is not indispensible. Several modified RNA cap structures lacking this N7-methyl group conserved binding to eIF4E and were translation competent when present at the 5' ends of RNAs *in cellulo*.

INTRODUCTION

The addition of a 5' cap structure to RNA transcripts synthesized by RNA polymerase II is a fundamental eukaryotic process (Shuman 1995). The N7-methyl guanosine cap in mRNAs is essential for their stability, maturation, transport and translation. RNA cap synthesis occurs in 3 consecutive steps. An RNA triphosphatase (RTase) first cleaves off the 5' terminal phosphate of the RNA molecule, to form a diphosphorylated end. An RNA guanylyltransferase (GTase) transfers a GMP moiety onto the diphosphorylated end, followed by an RNA (guanine-N7) methyltransferase (MTase) which adds a methyl group at the N7 position to form the classical RNA cap structure (Shuman 2001). On account of the essentiality of all the proteins involved in RNA cap synthesis, the study of the various roles of the cap structure is often restricted to the analysis of cap binding proteins. The advent of RNA cap analogues has provided a new tool to directly investigate RNA cap metabolism and function, either as competitive inhibitors or as alternative structures at the 5' ends of RNAs (Darzynkiewicz, Stepinski et al. 1989; Carberry, Darzynkiewicz et al. 1990). So far, RNA cap analogues have been solely generated through chemical synthesis. However, although a wide range of RNA cap analogues have been chemically synthesized, the structural variety of possible RNA cap analogues is restricted by the method. The requirement of protected nucleosides as starting materials, the poor solubility of RNA cap analogues in organic solvents as well as their susceptibility to hydrolysis in both acidic and basic conditions during the removal of the protecting groups, make the production of RNA cap analogues a very challenging task (Strenkowska, Kowalska et al. 2010). In addition, the incorporation of these chemically synthesized cap analogues at the 5' ends of RNAs is heavily dependent on the conditions of *in vitro* transcription. This further limits the variability of possible cap structures at the 5' ends of RNAs.

In an attempt to seek for an alternative way to generate novel RNA cap analogues, we decided to probe into the guanine specificity of RNA guanylyltransferases. GTases first hydrolyze GTP to form a covalent enzyme-*(lysyl-N)*-GMP intermediate prior to transferring the GMP moiety onto an acceptor RNA (Fig. 1A). GTases share the same mechanistic profile as well as six co-linear conserved motifs, including the catalytic KxDG motif, with ATP-dependent ligases (Supplementary figure 1) (Shuman and Lima 2004). Crystal structures of the *Paramecium bursaria Chlorella virus-1* (PBCV-1) GTase and T7 DNA ligase bound to GTP and ATP respectively have revealed a common tertiary fold in which these conserved motifs are brought together at the active site of the enzyme (Hakansson, Doherty et al. 1997). However, in spite of numerous structural and biochemical data on these enzymes, modulation of the substrate specificity of GTases or ligases through the rational design of point mutations remains to be achieved. The previous finding that the

antiviral nucleotide analogue, ribavirin triphosphate, could be used as substrate by a viral GTase led us to interrogate whether the active sites of GTases were flexible enough to accommodate other nucleotide analogues and catalyze an RNA capping reaction to form potential novel RNA cap structures (Bougie and Bisaillon 2004). While synthetic nucleotide analogues have been extensively studied with regards to RNA stability and translation, few studies have aimed to probe into the capping machinery itself, in order to understand the underlying interactions involved in GTP binding for RNA cap synthesis (Grudzien, Stepinski et al. 2004; Stepinski, Zuberek et al. 2005; Grudzien-Nogalska, Jemielity et al. 2007; Grudzien-Nogalska, Stepinski et al. 2007; Kowalska, Lewdorowicz et al. 2008). In order to understand the modulation of substrate specificity in RNA capping enzymes, this report aims to identify the essential interactions at each step of the RNA capping pathway. Synthetic purine analogues were therefore evaluated for their efficiency to substitute GTP for the two steps of the RNA capping reaction by the PBCV-1 GTase. The fact that the PBCV-1 GTase has been isolated in several different conformations rendered it the ideal tool to delve further into the RNA capping mechanism (Doherty, Hakansson et al. 1997). Several novel cap structures bearing unusual substitutions on the base moiety were thus generated *in vitro*. While assessing the impact of these modifications on the translation machinery *in cellulo*, key features for translation-competent RNA cap structures were unravelled.

METHODS:

Expression and purification of the relevant proteins

Recombinant His-tagged PBCV-1 GTase, PBCV-1 RTase, *S. cerevisiae*'s MTase, human MTase and murine eIF4E proteins were expressed in bacteria as described before (Mao, Schwer et al. 1996; Saha, Schwer et al. 1999; Bougie and Bisaillon 2006; Slepenkov, Darzynkiewicz et al. 2006; Souliere, Perreault et al. 2008).

Preparation and purification of RNA substrates

RNA substrates were synthesized with the MAXIscript kit (Ambion) using T7 RNA polymerase. For the evaluation of RNA capping efficiencies, a 50 nt long RNA was synthesized using a modified version of the T7 promoter that will allow the initiation of the transcription with a single guanosine residue (5'-TAATACGACTCACTATA↓G A_{49}-3', where ↓ indicates the initiation start site). *In cellulo* translation assays, were done with the *firefly* luciferase RNA bearing a 60 nt long poly(A) tail at its 3' end, which was synthesized from an *HpaI* digested *plucA*$_{60}$ plasmid (a generous gift from Dr. Rhoads, Louisiana State University). Following transcription, the RNA molecules were purified on denaturing 8M UREA-PAGE and visualized by ultraviolet shadowing. The corresponding band was excised and then eluted from the gel by an overnight incubation in 0.1% SDS and 0.5 M ammonium acetate. The RNA was then precipitated with ethanol and quantified by spectrophotometry at 260 nm.

Nucleotide Analogues

All nucleotide analogues used in this study were purchased from Jena Biosciences (Germany) and Tri-Link Biotechnologies (USA).

First step RNA guanylyltransferase reaction

The first step of the GTase reaction was carried out by incubating the purified PBCV-1 GTase (10 μM) with the appropriate substrates (GTP or nucleotide analogues) in a buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, 5 mM $MgCl_2$ and 1.5 μg/ml of inorganic yeast pyrophosphatase (Roche) for 1hr at 30^0C.

Inhibition assay and determination of the IC_{50}

Inhibition of the first step of the GTase reaction was evaluated by carrying out the standard first step RNA guanylyltransferase reaction in the presence of 0.2 pmol of [α-^{32}P] GTP and 2 mM of either unlabelled GTP or each unlabelled nucleotide analogue separately (Fig. 2A). The IC_{50} of nucleotide analogues was determined by carrying out the standard first step reaction in the presence of 0.2 pmol of [α-^{32}P] GTP and increasing concentrations of up to 2 mM of each nucleotide analogue. The reactions were stopped by the addition of EDTA to 10 mM and SDS to 1% and analyzed by electrophoresis through a 12.5% polyacrylamide gel containing 0.1% SDS. The radiolabelled proteins were visualized by autoradiography of the gel. Radiolabelled covalent complex formation was quantified by scanning the gel with a PhosphorImager (Amersham Biosciences).

Evaluation of the formation of the covalent intermediate

The standard first step reaction was carried out in the presence of each nucleotide analogue (2 mM) in the presence or absence of potassium pyrophosphate (1 mM). The reactions were resolved on SDS-PAGE followed by Coomassie blue staining.

***In vitro* RNA capping reaction**

RNA capping was performed by incubating the purified PBCV-1 GTase (10 µM) and PBCV-1 RTase (5 µM), either GTP (2 mM) or a nucleotide analogue (2 mM) with the appropriate RNA in a buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, 5 mM $MgCl_2$ and RnaseOut (1 unit) (Invitrogen).

Transfer of nucleotide analogues to RNA

The transfer of GTP or of a nucleotide analogue onto RNA was assayed by performing the *in vitro* RNA capping reaction in the presence of an internally labelled 50 nt long RNA substrate (10 pmol). The RNA was extracted with phenol/chloroform and recovered by ethanol precipitation before being analyzed on a denaturing UREA-PAGE. Autoradiography of the gel with a PhosphorImager enabled discrimination between a cap and uncapped RNA. Alternatively, a 50 nt long RNA molecule with a 5' α-labelled guanosine residue (5'-pp^{32}p RNA-3') was used. To evaluate capping efficiency, the GTase reaction mixture was heated to 95^{0}C for 3 minutes before being adjusted to 50 mM sodium acetate (pH 5.2) and digested with nuclease P1 (5µg) and alkaline phosphatase (1 unit) at 37^{0}C for 1 hr. The products were then analyzed by thin layer chromatography (TLC) on a polyethyleneimine-cellulose plate developed with 0.5 M LiCl and 1 M formic acid, following which, the extent of cap formation was measured by scanning the TLC with a PhosphorImager. Cap

formation is inferred from the presence of unhydrolyzable 5'-5' triphosphate bridge ($Npp^{32}pG$) on the TLCs.

Cell culture

Human embryonic kidney cells (HEK293) were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum.

Transfection

The day before transfection 600,000 cells/well were distributed in a 6-well plate. $LucA_{60}$ RNA (5 μg) was transfected using the Qiagen Transmessenger kit according to the manufacturer's protocol for 3 hr. Each RNA species was transfected twice per experiment. For each RNA species transfected, the cells in one of the wells were washed with PBS and harvested, while in the other, the cells were incubated at 37°C in pre-warmed complete medium for 6 hr, following which, the cells were washed and harvested.

Determination of translation efficiency

The harvested cells were split into 2 tubes. Cells in one of the tubes were lysed in Luciferase Cell Culture Lysis Reagent (Promega) and luciferase activity was measured according to the manufacturer's protocol (Promega). Total protein concentration of cell extracts was determined by the Bio-Rad dye binding method, using bovine serum albumin as the standard. Relative luciferase units (RLU) reads were rationalized onto the total protein concentration in the extracts and the capping efficiency of each nucleotide analogue tested. The translation efficiency was determined by normalising the data relative to the natural N7-methyl guanosine cap $lucA_{60}$ RNA.

Determination of RNA stability

Harvested cells were lysed in Qiazol and RNA was extracted according to the manufacturer's protocol (Qiagen). Real-time quantitative PCR analysis using total RNA extracts was performed as described previously (Bachand, Lackner et al. 2006). Briefly, 1 μg of total RNA was treated with Promega DNase RQ1 and reverse transcribed using Qiagen Omniscript RT. cDNAs were diluted 20-fold and analyzed on an Eppendorf Realplex PCR instrument using PerfeCTa™ SYBR Green Supermix kit (Quanta Biosciences). $LucA_{60}$ RNA levels were quantified relative to the GAPDH RNA using the $\Delta\Delta^{Ct}$ method as previously described (Bachand, Lackner et al. 2006). RNA stability was determined by comparing relative $lucA_{60}$ RNA levels 6 hr post-transfection relative to 0 hr

post-transfection. The data presented has been normalized relative to the stability of the N7-methyl guanosine capped $lucA_{60}$ RNA.

In vitro RNA binding assay to eIF4E

The binding of RNAs harbouring modified RNA cap structures to the eIF4E protein was evaluated by fluorescence spectroscopy as described previously (Benzaghou, Bougie et al. 2006). Briefly, excitation was performed at a wavelength of 290 nm fluorescence using a Hitachi F-2500 fluorescence spectrophotometer. Background emission was eliminated by subtracting the signal from the buffer containing the RNA substrate. The extent to which the RNA binds to purified eIF4E protein was determined by monitoring the fluorescence emission of a fixed concentration of proteins with and without the appropriate RNA (2 μM). All data has been rationalized relative to the capping efficiency of each nucleotide analogue (Table 1) and is presented relative to the fluorescence quenching observed when eIF4E binds the natural N7-methyl guanosine cap.

RESULTS

Formation of a covalent intermediate with nucleotide analogues

In order to probe into the structural flexibility of the active site of an RNA guanylyltransferase, we decided to evaluate the relative propensity of the PBCV-1 GTase to accommodate modified substrates through the use of nucleotide analogues. The nucleotide analogues harboured various modifications on both the ribose and the guanine base of GTP (Fig. 1C).

We initially monitored the ability of a library of 22 synthetic nucleotide analogues to inhibit the PBCV-1 GTase. In order to identify nucleotide analogues which could efficiently substitute GTP in the reaction under steady state conditions, the candidate molecules (2 mM) were incubated in the presence of [α-^{32}P]GTP and the PBCV-1 capping enzyme (10 μM). The presence of non-labelled GTP (2 mM) led to a maximum loss of the radiolabelled EpG signal while ATP at the same concentration had no apparent effect. Each synthetic analogue inhibited the intermediate formation to a different extent (Fig. 2A). The IC_{50} of the nucleotide analogues were then determined. Increasing concentrations of non-labelled GTP or nucleotide analogue were added to the standard GTase reaction. An example, using ITP (A_2) as an inhibitor is shown in figure 2B. IC_{50} values ranged from 100 μM (high inhibition e.g. GTP) to 2.0 mM or more (low inhibition e.g. 2'dGTP). The fact that some nucleotide analogues led to inhibition of the PBCV-1 GTase activity at high concentrations (> 2.0 mM) raised the possibility that they might be acting as chelator for the essential divalent metal ion co-factor, and therefore these analogues were not included in later analysis.

In order to monitor the formation of the covalent intermediate, the PBCV-1 GTase was incubated in the presence of magnesium ions (5 mM) and the nucleotide analogues (2 mM). The reaction products were then analyzed by SDS-PAGE and visualised by Coomassie Blue staining (Fig. 1D). In each reaction with a nucleotide analogue, a slower migrating species relative to the unbound protein was observed. We deduced that this band corresponds to the enzyme covalently bound to a nucleotide analogue (E-NMP). Pyrophosphate, the reaction product, promoted the release of NTP through reversal of the reaction, thus indicating that the use of different nucleotide analogues did not alter the reaction reversibility (Fig. 1A and 2D). The informative finding is that all identified inhibitors of the first step of the GTase reaction could form the covalent intermediate.

Nucleotide analogues as RNA cap donors

Having confirmed that the PBCV-1 GTase can efficiently form covalent intermediates with several nucleotide analogues, we next questioned whether these intermediates had conserved the ability to be transferred onto a 5'-diphosphate RNA. The PBCV-1 GTase was therefore incubated with the appropriate nucleotide analogue, a 5'-α terminally labelled RNA and the PBCV-1 RTase. The reaction products were digested by nuclease P1 and alkaline phosphatase and resolved by thin layer chromatography. The resolved chromatogram revealed the presence of a digestion resistant species corresponding to GpppG when GTP was added to the RNA capping reaction mixture, thus confirming the transfer of GMP onto an acceptor RNA (Fig. 2F). The covalent intermediates formed from various purine triphosphate analogues showed varying degrees of efficiency in their ability to act as cap donors (Table 1). However, some of the analogues tested, such as A_5, A_8, A_9, and A_{10}, were clearly not transferable onto RNA. We conclude that the formation of the covalent E-NMP intermediate does not necessarily imply the completion of the second step of the GTase reaction. A second assay was performed to demonstrate the transfer of nucleotide analogues onto RNA. A purified ^{32}P-internally labelled RNA was incubated with the PBCV-1 GTase, the PBCV-1 RTase, magnesium ions and each nucleotide analogue separately. The reaction products were analyzed by UREA-PAGE. The addition of GTP or a nucleotide triphosphate (A_2) analogue led to the formation of a slower migrating species corresponding to the capped RNA (Fig. 2E). ATP, on the other hand had no effect. This clearly confirmed that some nucleotide analogues can act as cap donors and be transferred onto RNA, while some others cannot.

Translation and stability of differentially capped RNAs *in cellulo*

Several studies have demonstrated that the N7-methyl guanosine cap structure is essential for the process of cap dependent translation, since the association of a functional translation initiation complex requires the prior interaction of the RNA cap structure with the eukaryotic initiation factor 4E (eIF4E), a phylogenetically conserved subunit of the heterotrimeric eIF4F initiation complex (Sonenberg 2008). Having generated several novel 5' cap structures bearing new modifications on both the base and the ribose moieties, we decided to investigate the structural link between translation and the nature of the cap structure by performing *in cellulo* translation assays with these modified RNAs. It is worth mentioning that we first evaluated the capacity of each modified cap to accommodate a methyl group at the N7 position, by using the human and the yeast MTases. RNAs blocked with A_2, A_6, A_7, A_{13}, A_{14} and A_{22} could be methylated by an MTase to different extents (Supplementary table 2). However, all attempts to methylate an RNA blocked with A_1 were unsuccessful.

Firefly luciferase RNA (5μg) harbouring modified caps at their 5' ends were transfected into HEK293 cells. An outline of the experimental procedure is shown in figure 3A. In order to ensure reproducibility, each transfection procedure was performed in parallel with a positive (^{m7}G cap) and a negative (uncap) control RNA. Luc A_{60} RNA levels relative to the GAPDH RNA level (normalising control) were evaluated just after transfection and 6 hr post-transfection by qRT-PCR using specific primers which amplify sequences near the 5' end. The results were normalized onto the relative amount of lucA_{60} RNA possessing a natural cap (Fig. 3B).

In addition, for each transfection assay, the luciferase activity of total cell extracts was also determined and corrected for the capping efficiency of each nucleotide analogue and translation efficiency was determined relative to the positive control for each assay (Fig. 3B). Interestingly, the translation efficiencies of the RNAs investigated did not reflect their relative efficiency as methyl acceptors. RNA capped with 3'O-Me GTP (A_{22}), despite its poor propensity to be a N7-methyl acceptor, displayed higher translation efficiency as compared to RNAs capped with A_2, A_{13} or A_{14}. Moreover, our results suggest that a methyl group at the O6 position render the presence of a methyl group at the N7 position futile; the O6-Me guanosine cap (A_7), being able by itself to sustain RNA translation. Our results are in agreement with previous studies which showed that RNAs possessing an N7-(3'O) dimethyl guanosine cap structure ($^{m7}A_{22}$) are translationally more active than RNAs possessing the natural cap (Kuhn, Diken et al.; Kalek, Jemielity et al. 2005; Kore and Shanmugasundaram 2008). However, most intriguingly, the absence of the N7-methyl group (A_{22} cap) does not impact translation efficiency negatively. The translation efficiency of A_{22} and $^{m7}A_{22}$ remain similar (within experimental error) despite the absence of the methyl group at the N7 position and consequently of the positive charge on the cap structure. We therefore concluded that for the process of translation *in cellulo*, the presence or absence of the essential N7-methyl modification of the natural N7-methyl guanosine cap becomes secondary in the light of other modifications on either the base or even the ribose moiety of the 5' RNA cap structure.

In vitro binding to eIF4E

In an effort to further analyze the translation profile obtained, we set out to determine the binding affinity of each novel RNA cap structure obtained to the eIF4E protein. The eIF4E protein harbours eight conserved tryptophan residues within its cap binding slot (Marcotrigiano, Gingras et al. 1997; Matsuo, Li et al. 1997). Therefore, binding affinity was evaluated by monitoring the quenching of the intrinsic fluorescence of the protein when incubated with a 30 nt long RNA possessing a natural or modified cap structure (Fig. 3B). Our results echo previous studies in that for cap dependent

translation to occur, binding to eIF4E is a fundamental requirement. For instance, the affinity of eIF4E to an A_{22} capped RNA relative to a naturally capped RNA was more than 2 fold higher, in spite of the lack of the N7-methyl group on this cap analogue. This relates directly with the higher translation profile obtained for the A_{22} capped $lucA_{60}$ RNA. Overall, these results indicate that cap dependent translation can be sustained in the absence of the N7 modification of the RNA cap structure, provided that alternative modifications enable proper binding to the eIF4E protein.

DISCUSSION

RNA cap analogues are important biological tools for the study of RNA metabolism, and could prove to be potent novel therapeutic agents. In an attempt to look for an easier and more effective way than chemical synthesis to generate novel cap structures at the 5' ends of RNAs, we report that RNA cap analogues can be formed by enzymatic synthesis using nucleotide analogues. The differential ability of nucleotide analogues to be used by the PBCV-1 GTase was used as a measure of the flexibility of the active site, and we were able to show that the active site can accommodate several unnatural substrates, and can efficiently cap an RNA molecule with a modified substrate. Thus, we were able to gain a better understanding of the interactions involved in substrate binding and catalysis in GTases and in the translation process.

C2 and O6 modified analogues

Our analogues indicate that the exocyclic amino group of GTP is not crucial for substrate discrimination by GTases. ITP (A_2) efficiently substitutes GTP in both steps of the GTase reaction (Table 1). This discloses a relative lack of importance of the exocyclic C2 amino group for substrate binding by the PBCV-1 GTase, and in accordance, 2-amino-ATP (A_4), which harbours an exocyclic amino group equivalent to GTP, could not inhibit the first step of the GTase reaction.

It has previously been speculated from crystallographic data that GTP specificity of capping enzymes is mediated by the interactions of the O6 of GTP via a hydrogen bond with a conserved lysine residue (Lys188 in the PBCV-1 GTase) (Fig. 1B) (Hakansson, Doherty et al. 1997; Fabrega, Shen et al. 2003). A sequence alignment between RNA capping enzymes tentatively places this conserved lysine residue in a motif (denoted IIIc), which ostensibly seems to be absent among DNA ligases (Supplementary figure 1). Interestingly, while alanine mutation of this lysine residue abolishes GTase activity, binding of GTP remains unaffected (Supplementary figure 3). Therefore, it is presumed that other amino acids are contributing to GTP binding. Nucleotide analogues harbouring modifications at the O6 position were used to investigate this issue. With the exception of 6-thio ITP (A_9) (IC_{50}=1.7 mM) and 6-Me-thio ITP (A_{10}) (IC_{50}=1.2 mM), O6 modified analogues that were tested did not differ significantly relative to GTP ((IC_{50}=0.1 mM) in their ability to inhibit the first step of the reaction. At most, the fact that 6-Me thio GTP (A_8) (IC_{50}=0.43 mM) is a poorer substrate than 6-thio GTP (A_6) (IC_{50}=0.22 mM) or O6-Me GTP (A_7) (IC_{50}=0.15 mM) may be indicative of steric hindrance for substrate binding. The similar IC_{50} of A_5, A_6 and A_7 indicate that the extent of Lys188 interactions with nucleotide analogues bearing C6 modifiers is permeable to

steric constraints and variations in the electrostatic potential. However, in all cases, the transfer onto an acceptor RNA was heavily compromised during the second step of the GTase reaction. While O6-Me GTP (A_7) was transferred ~50% as efficiently as GTP onto an RNA, A_5 and A_6 (8% and 16 % respectively relative to GTP) were very poor cap donors. Our results indicate that along with Lys188 and O6 hydrogen bonding, it is the very architecture of this region of the active site which may be mediating substrate binding and contributing to the reaction progress. In addition, in the light of the fact that only O6-Me GMP (A_7) retains appreciable transferability onto an RNA, we inferred that only oxygen at the C6 position could preserve the subtleties of interactions with Lys188, thereby allowing accurate substrate alignment for the intermediate complex formation while also retaining the transferability of the bound nucleotide analogue onto an acceptor RNA.

N7 and C8 modified analogues

Additional proof of the involvement of other factors in determining cap donor specificity was obtained through N7 (A_{15} and A_{16}) and corroborated with C8 (A_{12-14}) modified purine analogues. Neither 7-deaza GTP (A_{15}) nor N7-methyl GTP (A_{16}) could inhibit the first step of the reaction (Table 1). This is suggestive of an important role of the N7 position. The poor inhibition by A_{16} may also be partly due to its positively charged imidazole ring which may be impeding stacking interactions with Phe146 in the GTP binding site (Fig. 1B). C8 modified analogues displayed a differential utilization by the PBCV-1 GTase based on the nature of the substituent. In fact, while the presence of a hydroxyl group at the C8 position (A_{12}) abolished inhibition (IC_{50} > 2.0 mM), halide substituted C8 analogues (A_{13-14}) were less detrimental in that an appreciable IC_{50} was obtainable for the case of 8-bromo GTP (A_{13}) (IC_{50}=1.5 mM) and 8-iodo GTP (A_{14}) (IC_{50}=0.42 mM) (Table 1). The decreasing tendency of C8 substituted GTP analogues (8-I GTP > 8-Br GTP >> 8-Oxo GTP) to inhibit the first step of the GTase reaction was suggestive of an influence of the relative electronegativity of each substitution (χ of I < Br < O) on the guanine base. A direct consequence of the presence of electronegative substitutions at the C8 position is a net decrease in the electron density by inductive effect at the N7 position. Therefore, taken together, N7 and C8 modifications (A_{12-16}) indicate that both N7 and O6 likely mediate interactions responsible for cap donor specificity. In contrast to GTP, the lone pair of electrons at the N7 position in ATP is less available for interactions due to the basic chemical properties of the base (Kmpf, Kapinos et al. 2002). The requirement for interactions with both O6 and N7 could also be explaining why O6 modifiers (A_{6-11}) led to such a wide difference in their effectiveness between the first and the second steps of the reaction.

2' modified GTP analogues

Previous studies have shown that 2'dGTP (A_{18}) cannot be hydrolyzed by the PBCV-1 GTase (Ho, Van Etten et al. 1996). This is suggestive of an important role of the 2'OH group. This might be due to the following reasons: (i) the 3`-*endo* conformation, which is preferred by ribonucleotides but not deoxyribonucleotides may be important; (ii) the inductive effect of the 2'OH group may be important; (iii) the 2'OH group coordinates a critical metal ion or (iv) binding of the 2'OH may be a pre-requisite for proper alignment of the triphosphate moiety.

With the exception of the *poxvirus* capping enzyme, exemplified by the vaccinia virus D1 protein, most known capping enzymes are unable to hydrolyze 2'dGTP (Venkatesan and Moss 1980; Venkatesan and Moss 1982). Since the D1 protein shares the same conserved motifs as the PBCV-1 GTase, we think that the 2'OH must have a preponderant role in ligand binding rather than catalysis. Therefore, the hypothesis that the 2'OH may be participating in the coordination of a critical metal ion is unlikely. The sugar moiety of 2'F-2'dGTP (A_{20}) has a net preference for the 3'-*endo* conformation and the 2'F substituent has a stronger inductive effect than the hydroxyl group in GTP (Roitzsch, Fedorova et al. 2010). The fact that the 2'F substituent negatively affects the first step of the GTase reaction by drastically increasing the IC_{50} (1.4 mM for A_{20} compared to 0.10 mM for GTP) is a clear indication that neither the preference for the 3'-*endo* conformation nor the inductive effect explains the preference of the PBCV-1 GTase for GTP over 2'dGTP. Thus far, our results indicate that the 2'OH is involved in ligand binding/positioning, without necessarily being involved in catalysis. Previous structure-function analysis in budding yeast and mouse capping enzymes have shown the 2' OH group to be interacting with an essential Motif III glutamate – Glu131 in the PBCV-1 GTase, with which it forms a hydrogen bond (Fig. 1B and Supplementary figure 1) (Shuman, Liu et al. 1994). The pH dependency of the inhibition of the first step of the GTase reaction by 2'O-Me GTP (A_{19}) further substantiated this conclusion. A_{19} does not inhibit the GTase reaction at pH 7.5 (Table 1). However, decreasing the pH reduced the ease of formation of the radiolabelled covalent EpG intermediate, thereby indicating that at lower pH, A_{19} can effectively inhibit the reaction (Supplementary figure 2). Decreasing pH infers protonation of basic residues (like Glu131) which renders possible the otherwise unfavourable interaction between the glutamate residue and the 2' oxygen of 2'O-Me GTP at a pH of 7.5. Likewise, 2'F-2'dGTP also shows a pH dependency for the first step of the reaction, albeit to a lesser extent that 2'O-Me GTP. We conclude that the essentiality of the 2'OH group in RNA capping resides mainly in proper ligand positioning.

3' modified GTP analogues

The role of the 3' OH group was evaluated through the different levels of inhibition and activity displayed by 3'dGTP (A_{21}) and 3'O-Me GTP (A_{22}). The absence of the 3'OH group (A_{21}) did not affect the first step of the reaction, while its substitution by 3' O-Me (A_{22}) led to an increase in the IC_{50}; a change that we attributed to steric hindrance due to the methyl group. However this trend was not conserved in the second step of the reaction. RNA capping with 3'dGTP was markedly inferior relative to capping with 3' O-Me GTP (10% and ~40% respectively relative to GTP). It has previously been speculated that the role of the 3'OH was to coordinate the acceptor RNA molecule. The structure of the *Candida albicans* GTase has been solved in an open conformation receptive for RNA binding, with a phosphate ligand, speculatively attributed to be mimicking the terminal phosphate of RNA, near the 3' hydroxyl group of the bound GMP (Fabrega, Shen et al. 2003). Crystal structures of other nucleotidyltransferases in complex with their ligands also seemed to suggest that the covalently bound ligand contributes to the formation of a novel active site, where the ligand itself plays a fundamental functional role in the coordination of the incoming RNA or DNA substrate (Nair, Nandakumar et al. 2007). Our results add weight to this hypothesis. While the absence of a hydroxyl group at the 3' position is detrimental to RNA coordination, the presence of an O-Me group at the same position can clearly rescue capping efficiency, albeit to a limited extent only. Therefore, we believe that the role of the 3'OH lies mainly in the coordination of the diphosphate RNA ligand.

***In cellulo* properties of artificial cap structures**

The subset of RNAs generated with modified cap structures at their 5' ends were evaluated for their translational properties in HEK293 cells. Regarding the relative stability of RNAs capped with various analogues *in cellulo*, our results were consistent with the fact that the presence of a blocking residue at the 5' end of an RNA was sufficient to protect it from rapid degradation as compared to an uncapped RNA (Bougie and Bisaillon 2004). Previous studies addressing the low ability of ^{m7}IMP and ^{m7}IDP to inhibit translation have demonstrated their poor binding affinity for the mammalian eIF4E, and in agreement, a very low translation efficiency was observed for a N7-methyl inosine ($^{m7}A_2$) capped RNA(Adams, Morgan et al. 1978; Ueda, Maruyama et al. 1991). In addition, our results concerning an N7-(3'O)-dimethyl guanosine ($^{m7}A_{22}$) capped RNA are consistent with previous observations that an RNA capped with $^{m7}A_{22}$ are translationally more active than RNAs possessing the natural cap (Fig. 3B) (Jemielity, Fowler et al. 2003; Kalek, Jemielity et al. 2005; Kore and Shanmugasundaram 2008).

On the other hand, our observations with regards to the translational properties of RNAs capped with A_7 and A_{22} were very unexpected. A_7 and A_{22} capped RNA could support translation *in cellulo* in spite of the absence of the N7-methyl modification. To the best of our knowledge, this is the first report of an N7-methyl deficient capped RNA which can support cap-dependent translation *in cellulo*. Previous reports have evaluated the importance of the N7-methyl addition solely in the context of the natural N7-methyl guanosine cap. Compensation for the absence of the N7-methyl group by alternative modifications on the cap structure had not been looked into before. Our i*n vitro* binding studies indicate that the relative binding affinity of eIF4E to the modified cap structures was enough to explain the observed translation efficiency. While a GpppG capped RNA is not bound by eIF4E, the presence of a methyl group at the O6 (A_7) or 3'-O (A_{22}) clearly restores eIF4E binding, and thus ensures translation. The major implication of this result is that cap binding by eIF4E does not necessarily require a positively charged capped nucleotide, as is the case for the natural RNA cap structure. Crystallographic data on eIF4E indicates that the cap binding slot is essentially divided into a positively charged region which binds the phosphate bridge, and a hydrophobic region where the charged N7-methyl guanosine cap is stacked between two conserved tryptophan residues through cation-π-π interactions (Marcotrigiano, Gingras et al. 1997). The N7-methyl group is involved in Van der Waal's interactions only, and has been shown to be substitutable with various alkyl groups without being deleterious to eIF4E binding (Brown, McNae et al. 2007). It was therefore surmised that the main purpose of the N7-methyl group is to confer a positive charge to the cap. Since, most previous studies used the premise that cap binding to the eIF4E protein requires a positively charged RNA cap analogue, the necessity of the N7-methyl group in the light of other modifications on the RNA cap structure had not been addressed before. In view of the fact that neither A_7 nor A_{22} can conserve the cation-π-π interactions of the natural cap with eIF4E, we conclude that other factors are necessarily rendering A_7 and A_{22} interactions with eIF4E favourable.

Several contradictory observations have been reported with regards to the ionic state of the cap structure on eIF4E binding. Previous studies on the pH dependency of eIF4E binding to cap analogues have suggested that the cap preferably binds in its enolate tautomer (in which N1 is deprotonated) rather than in the keto form (Rhoads, Hellmann et al. 1983; Matsuo, Li et al. 1997). NMR studies of cap analogue binding to eIF4E showed that the imino proton (at N1 position) was absent, thus further indicating that binding occurs in the enolate tautomeric form of the ligand (Matsuo, Li et al. 1997). On the other hand, in the crystal structure of the yeast eIF4E bound to ^{m7}GDP, the keto tautomer is present in the cap binding site (Marcotrigiano, Gingras et al. 1997).

Other *in vitro* studies have pointed out that tightest binding to eIF4E is achieved by the keto tautomeric form of the N7-methyl guanosine cap (Niedzwiecka, Marcotrigiano et al. 2002). The deprotonated imino group of A_7 is reminiscent of the enolate tautomer of guanosine. Since we detect translation of both A_7 and $^{m7}A_7$ capped RNA, this infers that for the occurrence of eIF4E binding *in cellulo*, there is no need for the N1 proton of the cap to be engaged in a hydrogen bond within the cap binding site. Therefore, from our current results we suggest that cap binding by eIF4E preferentially occurs in the enolate form of the ligand. The protonation equilibrium of the enolate and keto tautomers probably guides eIF4E binding to the natural cap. Therefore, we surmised that the fact that the imino group of A_7 is permanently deprotonated, greatly favours eIF4E binding, which may also explain the higher translation efficiency observed despite the absence of the N7-methyl modification.

The ribose moiety of the capped guanosine residue has very seldom been described to be involved in cap binding by eIF4E. In fact, the stacking interactions of the base, and interactions between the phosphate chain and positively charged amino acids are the main energetic contributions for cap binding to eIF4E (Niedzwiecka, Stepinski et al. 2002; Niedzwiecka, Stepinski et al. 2003). It is postulated that the phosphate bridge of the RNA cap acts as an anchor to enable the capped N7-guanosine to interact within the cap binding slot. The observation that RNA blocked with 3'O-Me guanosine (A_{22}) was efficiently translated was very unexpected. Two hypotheses may explain these results: either the conformation of the A_{22} cap renders the anchoring of the phosphate chain more energetically favourable thus effectively decreasing the energetic requirement of the stacking interactions of the base for tight binding of the cap analogue with eIF4E; or conformational changes upon cap binding may be stabilizing the eIF4E-cap analogue complex. This is presently being investigated. Another interesting finding of this study is that binding to eIF4E does not necessarily imply formation of a translation competent initiation complex. RNA blocked with either of $^{m7}A_{13}$ or $^{m7}A_{14}$ is poorly translated, while retaining an appreciable relative binding affinity to eIF4E (0.4 and 0.8 fold relative to a natural cap respectively). Therefore, we hypothesize that conformational changes on cap binding by eIF4E rather than the cap binding in itself may be playing a crucial role in the formation of a translation competent initiation complex. Overall, our results indicate that cap-dependent translation can occur in the absence of the N7-methyl group on the cap structure provided that alternative modifications enable appropriate binding to eIF4E.

The importance of the N7-methyl group of the cap structure has been highlighted in several studies. In the light of our results which indicate that some N7-methyl deficient RNA cap analogues are biologically active, it raises the question of the functional role of this methyl group. At which step

of RNA metabolism is the N7-methyl group most important? The use of more novel RNA cap analogues for the study of other aspects of RNA metabolism would certainly prove to be crucial to address this question. Moreover, our results have major implications for the design and synthesis of potential therapeutic agents targeting the eIF4E protein. eIF4E is over-expressed in various cancers, and is under investigation as a potential drug target (Ruggero, Montanaro et al. 2004; Graff, Konicek et al. 2007; Graff, Konicek et al. 2008). The finding that some N7-methyl deficient RNA cap analogues are able to bind eIF4E has a direct impact on the various steps involved in the chemical synthesis of cap analogues and also raises the possibility of using RNA capping enzymes to generate them.

FUNDING

Canadian Institutes for Health Research (CIHR) grant. CIHR and Université de Sherbrooke grant to RNA group. M.B. is a Chercheur Boursier Junior 2 from the Fonds de recherche en santé du Québec and a member of the Infectious diseases group of the Centre de Recherche Clinique Étienne-Lebel. Doctoral training award from the Fonds de Recherche en Santé du Québec (to M.I.)

ACKNOWLEDGEMENTS

We would like to thank Dr. Robert Rhoads (Louisiana State University) for the generous gift of the *plucA*$_{60}$ plasmid. We also thank Dr Xavier Roucou (Université de Sherbrooke) for having provided the necessary infrastructure for the manipulation of human cell lines. Our thanks also go to Antannas Staskevicius for technical advices on cell culture experiments and Dominique Lesvesque for technical advices on RNA manipulation.

FIGURE LEGENDS

Figure 1: The RNA capping mechanism and the nucleotide analogues tested (A) The two-step RNA guanylyltransferase reaction. (B) The GTP binding site of the PBCV-1 RNA guanylyltransferase (PDB 1CKN). Residues shown are those interacting with the base and the sugar moiety. (C) Nucleotide analogues used in this study.

Figure 2: Biosynthesis of novel RNA cap structures (A) The PBCV-1 GTase was incubated with [α-^{32}P]GTP in the absence or presence of unlabelled GTP (2 mM) (lanes 1 and 2) or purine analogues (2 mM) (lanes 3-12). An autoradiogram of the SDS-PAGE gel is shown. The location of the EpG complex is indicated on the right. (B) Competitive inhibition of the EpG complex formation by ITP (A_2). Increasing concentrations of A_2 (0.0625, 0.125, 0.25, 0.5, 1.0, 2.0 mM) were added to the standard GTase reaction containing [α-^{32}P]GTP. An autoradiogram of the SDS-PAGE gel is shown. The location of the EpG complex is indicated on the *left*. (C) Dose-response inhibition of the PBCV-1 GTase by unlabelled GTP and ITP (A_2). (D) Formation of the enzyme covalent intermediate. The PBCV-1 GTase was incubated with GTP (lanes 2 and 3) or nucleotide analogues (lanes 4 to7) in the presence of either yeast pyrophosphatase or potassium pyrophosphate (5 mM). A Coomassie blue stain of the SDS-PAGE gel is shown. (E) Unlabelled GTP or nucleotide analogue (2 mM) was incubated with a ^{32}P-radiolabeled RNA substrate of 50 nt in the presence of the PBCV-1 GTase and RTase. The RNA samples were extracted with phenol/chloroform, recovered by ethanol precipitation, and analyzed on a 20% UREA-PAGE. An autoradiogram of the gel is shown. The position of the unblocked RNA of 50 nt is indicated. (F) The RNA capping reaction was carried out with a 50 nt long 5' terminally labelled RNA on the α phosphate. The reaction mixtures were heated for 5 min at 70^{0}C, adjusted to 50 mM NaOAc, pH 5.2, and subjected to digestion by nuclease P1 for 60 min at 37^{0}C and then adjusted to 50 mM Tris-HCl, pH 8, and digested by an alkaline phosphatase. The reaction products were analyzed by thin layer chromatography on a PEI-cellulose plate developed with 0.5 M LiCl and 1M formic acid. An autoradiogram of the plate is shown. The positions of the chromatographic origin (*ori*), GpppG, and of inorganic phosphate (P_i) are indicated. P_i in lane 1 (negative control) migrates higher because this reaction was done in the absence of any RNA capping proteins. (*) indicates the position of the radiolabelled phosphate in the capped dinucleotide.

Figure 3: *In cellulo* and *in vitro* properties of the novel cap analogues (A) Schematic representation of the experimental procedure for the determination of the translation efficiency of differentially capped lucA_{60} RNA in HEK293 cells. (B) Comparative analysis of the relative

translation of differentially capped $lucA_{60}$ RNAs with respect to the relative stability of each RNA in HEK293 cells, and with respect to their relative binding affinity to the eIF4E protein. Binding to eIF4E was determined by fluorescence spectroscopy with a 30 nt long RNA molecule. The relative RNA stability was evaluated by quantifying the amount of $lucA_{60}$ RNA relative to the GAPDH RNA by qRT-PCR 0 hr and 6 hr post-transfection. The relative translation efficiency was experimentally determined by quantifying *firefly* luciferase activity relative to the amount of total protein 6 hr post-transfection. Experimental data was adjusted relative to the capping efficiency (as determined in Table 1) of each analogue, and rationalized onto the ^{m7}G cap. The error associated with each data set is less than ± 0.1. (*) indicates more than 1.5 fold difference relative to the translation efficiency of a naturally capped RNA. (†) indicates more than 2 fold difference relative to the binding observed for the natural ^{m7}G capped RNA.

REFERENCE

1. Shuman, S., *Capping enzyme in eukaryotic mRNA synthesis.* Prog Nucleic Acid Res Mol Biol, 1995. **50**: p. 101-29.

2. Shuman, S., *Structure, mechanism, and evolution of the mRNA capping apparatus.* Prog Nucleic Acid Res Mol Biol, 2001. **66**: p. 1-40.

3. Darzynkiewicz, E., et al., *Inhibition of eukaryotic translation by nucleoside 5'-monophosphate analogues of mRNA 5'-cap: changes in N7 substituent affect analogue activity.* Biochemistry, 1989. **28**(11): p. 4771-8.

4. Carberry, S.E., et al., *A spectroscopic study of the binding of N-7-substituted cap analogues to human protein synthesis initiation factor 4E.* Biochemistry, 1990. **29**(13): p. 3337-41.

5. Strenkowska, M., et al., *Towards mRNA with superior translational activity: synthesis and properties of ARCA tetraphosphates with single phosphorothioate modifications.* New J Chem. **34**(5): p. 993-1007.

6. Shuman, S. and C.D. Lima, *The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases.* Curr Opin Struct Biol, 2004. **14**(6): p. 757-64.

7. Hakansson, K., et al., *X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes.* Cell, 1997. **89**(4): p. 545-53.

8. Bougie, I. and M. Bisaillon, *The broad spectrum antiviral nucleoside ribavirin as a substrate for a viral RNA capping enzyme.* J Biol Chem, 2004. **279**(21): p. 22124-30.

9. Kowalska, J., et al., *Synthesis and characterization of mRNA cap analogs containing phosphorothioate substitutions that bind tightly to eIF4E and are resistant to the decapping pyrophosphatase DcpS.* RNA, 2008. **14**(6): p. 1119-31.

10. Grudzien-Nogalska, E., et al., *Synthesis of anti-reverse cap analogs (ARCAs) and their applications in mRNA translation and stability.* Methods Enzymol, 2007. **431**: p. 203-27.

11. Grudzien-Nogalska, E., et al., *Phosphorothioate cap analogs stabilize mRNA and increase translational efficiency in mammalian cells.* RNA, 2007. **13**(10): p. 1745-55.

12. Stepinski, J., et al., *Novel dinucleoside 5',5'-triphosphate cap analogues. Synthesis and affinity for murine translation factor eIF4E.* Nucleosides Nucleotides Nucleic Acids, 2005. **24**(5-7): p. 629-33.

13. Grudzien, E., et al., *Novel cap analogs for in vitro synthesis of mRNAs with high translational efficiency.* RNA, 2004. **10**(9): p. 1479-87.

14. Doherty, A.J., et al., *Crystallization of the RNA guanylyltransferase of Chlorella virus PBCV-1.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 4): p. 482-4.

15. Bougie, I. and M. Bisaillon, *Inhibition of a metal-dependent viral RNA triphosphatase by decavanadate.* Biochem J, 2006. **398**(3): p. 557-67.

16. Souliere, M.F., J.P. Perreault, and M. Bisaillon, *Kinetic and thermodynamic characterization of the RNA guanylyltransferase reaction.* Biochemistry, 2008. **47**(12): p. 3863-74.

17. Mao, X., B. Schwer, and S. Shuman, *Mutational analysis of the Saccharomyces cerevisiae ABD1 gene: cap methyltransferase activity is essential for cell growth.* Mol Cell Biol, 1996. **16**(2): p. 475-80.

18. Slepenkov, S.V., E. Darzynkiewicz, and R.E. Rhoads, *Stopped-flow kinetic analysis of eIF4E and phosphorylated eIF4E binding to cap analogs and capped oligoribonucleotides: evidence for a one-step binding mechanism.* J Biol Chem, 2006. **281**(21): p. 14927-38.

19. Saha, N., B. Schwer, and S. Shuman, *Characterization of human, Schizosaccharomyces pombe, and Candida albicans mRNA cap methyltransferases and complete replacement of the yeast capping apparatus by mammalian enzymes.* J Biol Chem, 1999. **274**(23): p. 16553-62.

20. Bachand, F., et al., *Autoregulation of ribosome biosynthesis by a translational response in fission yeast.* Mol Cell Biol, 2006. **26**(5): p. 1731-42.

21. Benzaghou, I., et al., *Energetics of RNA binding by the West Nile virus RNA triphosphatase.* FEBS Lett, 2006. **580**(3): p. 867-77.

22. Sonenberg, N., *eIF4E, the mRNA cap-binding protein: from basic discovery to translational research.* Biochem Cell Biol, 2008. **86**(2): p. 178-83.

23. Kuhn, A.N., et al., *Phosphorothioate cap analogs increase stability and translational efficiency of RNA vaccines in immature dendritic cells and induce superior immune responses in vivo.* Gene Ther. **17**(8): p. 961-71.

24. Kalek, M., et al., *Synthesis and biochemical properties of novel mRNA 5' cap analogs resistant to enzymatic hydrolysis.* Nucleosides Nucleotides Nucleic Acids, 2005. **24**(5-7): p. 615-21.

25. Kore, A.R. and M. Shanmugasundaram, *Synthesis and biological evaluation of trimethyl-substituted cap analogs.* Bioorg Med Chem Lett, 2008. **18**(3): p. 880-4.

26. Matsuo, H., et al., *Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein.* Nat Struct Biol, 1997. **4**(9): p. 717-24.

27. Marcotrigiano, J., et al., *Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP.* Cell, 1997. **89**(6): p. 951-61.

28. Fabrega, C., et al., *Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II.* Mol Cell, 2003. **11**(6): p. 1549-61.

29. Kmpf, G., et al., *Comparison of the acid–base properties of purine derivatives in aqueous solution. Determination of intrinsic proton affinities of various basic sites.* Journal of the Chemical Society, Perkin Transactions 2, 2002: p. 1320-1327.

30. Ho, C.K., J.L. Van Etten, and S. Shuman, *Expression and characterization of an RNA capping enzyme encoded by Chlorella virus PBCV-1.* J Virol, 1996. **70**(10): p. 6658-64.

31. Venkatesan, S. and B. Moss, *Donor and acceptor specificities of HeLa cell mRNA guanylyltransferase.* J Biol Chem, 1980. **255**(7): p. 2835-42.

32. Venkatesan, S. and B. Moss, *Eukaryotic mRNA capping enzyme-guanylate covalent intermediate.* Proc Natl Acad Sci U S A, 1982. **79**(2): p. 340-4.

33. Roitzsch, M., O. Fedorova, and A.M. Pyle, *The 2'-OH group at the group II intron terminus acts as a proton shuttle.* Nat Chem Biol. **6**(3): p. 218-224.

34. Shuman, S., Y. Liu, and B. Schwer, *Covalent catalysis in nucleotidyl transfer reactions: essential motifs in Saccharomyces cerevisiae RNA capping enzyme are conserved in Schizosaccharomyces pombe and viral capping enzymes and among polynucleotide ligases.* Proc Natl Acad Sci U S A, 1994. **91**(25): p. 12046-50.

35. Nair, P.A., et al., *Structural basis for nick recognition by a minimal pluripotent DNA ligase.* Nat Struct Mol Biol, 2007. **14**(8): p. 770-8.

36. Adams. B.I.., et al., *The effect of "cap" analogs on reovirus mRNA binding to wheat germ ribosomes. Evidence for enhancement of ribosomal binding via a preferred cap conformation.* J Biol Chem, 1978. **253**(8): p. 2589-95.

37. Ueda, H., et al., *Expression of a synthetic gene for human cap binding protein (human IF-4E) in Escherichia coli and fluorescence studies on interaction with mRNA cap structure analogues.* J Biochem, 1991. **109**(6): p. 882-9.

38. Jemielity, J., et al., *Novel "anti-reverse" cap analogs with superior translational properties.* RNA, 2003. **9**(9): p. 1108-22.

39. Brown, C.J., et al., *Crystallographic and mass spectrometric characterisation of eIF4E with N7-alkylated cap derivatives.* J Mol Biol, 2007. **372**(1): p. 7-15.

40. Rhoads, R.E., et al., *Translational recognition of messenger ribonucleic acid caps as a function of pH.* Biochemistry, 1983. **22**(26): p. 6084-8.

41. Niedzwiecka, A., et al., *Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins.* J Mol Biol, 2002. **319**(3): p. 615-35.

42. Niedzwiecka, A., et al., *Thermodynamics of 7-methylguanosine cation stacking with tryptophan upon mRNA 5' cap binding to translation factor eIF4E.* Nucleosides Nucleotides Nucleic Acids, 2003. **22**(5-8): p. 1557-61.

43. Niedzwiecka, A., et al., *Positive heat capacity change upon specific binding of translation initiation factor eIF4E to mRNA 5' cap.* Biochemistry, 2002. **41**(40): p. 12140-8.

44. Graff, J.R., et al., *Therapeutic suppression of translation initiation factor eIF4E expression reduces tumor growth without toxicity.* J Clin Invest, 2007. **117**(9): p. 2638-48.

45. Ruggero, D., et al., *The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis.* Nat Med, 2004. **10**(5): p. 484-6.

46. Graff, J.R., et al., *Targeting the eukaryotic translation initiation factor 4E for cancer therapy.* Cancer Res, 2008. **68**(3): p. 631-4.

Table 1: RNA capping reaction with nucleotide analogues

Analogues		IC_{50} for E-GMP formation (mM) [a]	Formation of a covalent complex[b]	Relative efficiency of RNA capping[c]
		1st Step	*1st Step*	*1st and 2nd Steps*
	GTP	0.10	+	1.0
	ATP	>2.0	-	-
A_1	N1-Me GTP	0.08	+	1.1
A_2	ITP	0.34	+	1.3
A_3	XTP	>2.0	-	-
A_4	2-Amino ATP	>2.0	-	-
A_5	2-Amino-6 Cl purine RTP[d]	0.16	+	0.08
A_6	6-thio GTP	0.22	+	0.16
A_7	O6-Me GTP	0.15	+	0.48
A_8	6-Me thio GTP	0.43	+	0.07
A_9	6- thio ITP	1.7	+	-
A_{10}	6-Me thio ITP	1.2	+	-
A_{11}	6-Cl purine RTP[d]	>2.0	-	-
A_{12}	8-Oxo GTP	>2.0	-	-
A_{13}	8-Bromo GTP	1.5	+	0.40
A_{14}	8-Iodo GTP	0.42	+	1.0
A_{15}	N7-Me GTP	>2.0	-	-
A_{16}	7-deaza GTP	>2.0	-	-
A_{17}	ddGTP	>2.0	-	-
A_{18}	2'dGTP	>2.0	-	-
A_{19}	2'O-Me GTP	>2.0	-	-
A_{20}	2' F-2' dGTP	1.4	+	0.16
A_{21}	3'dGTP	0.15	+	0.15
A_{22}	3' O-Me GTP	0.45	+	0.38

[a] IC_{50} were determined from the dose response inhibition of the GTase activity as indicated in figures 2B and 2C for A_3.
[b] The formation of the covalent complex was determined as indicated in figure 2D for each nucleotide analogue.
[c] The efficiency of RNA capping was calculated by quantifying the nuclease P1 and alkaline phosphatase digestion resistar products formed for GTP and each nucleotide analogue as indicated in figure 2F and normalizing it onto the value obtaine with GTP.
[d] RTP stands for Ribose triphosphate

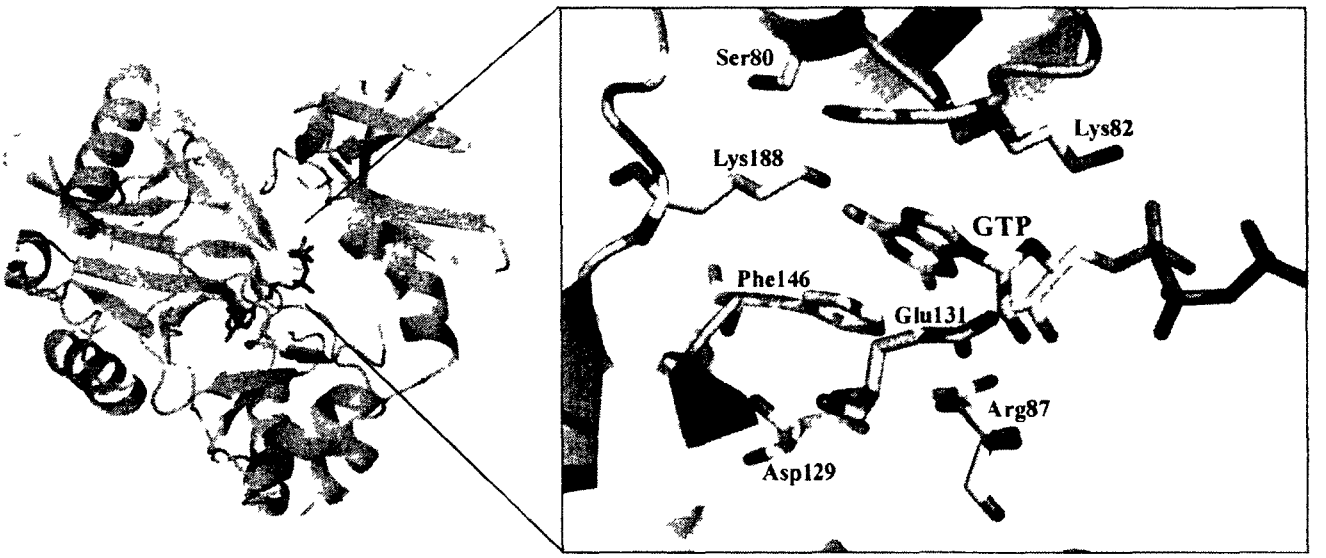
Figure 1

A

$$E + p_\gamma p_\beta p_\alpha G \underset{}{\overset{Mg^{2+}}{\rightleftharpoons}} Ep_\alpha G + PP_i \quad \text{(Step 1)}$$

$$Ep_\alpha G + ppRNA \underset{}{\overset{Mg^{2+}}{\rightleftharpoons}} Gp_\alpha ppRNA + E \quad \text{(Step 2)}$$

B



C

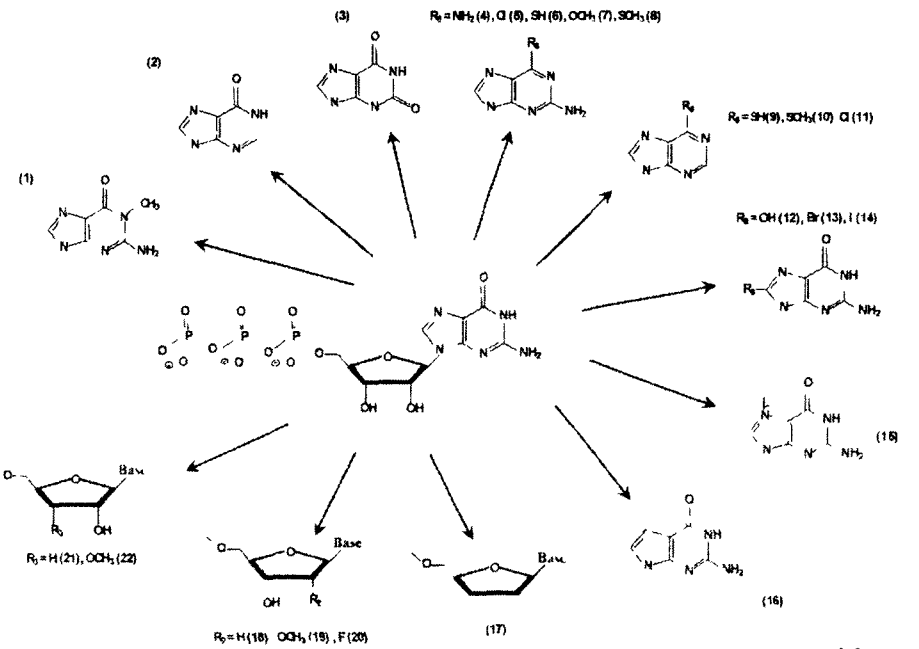


Figure 2

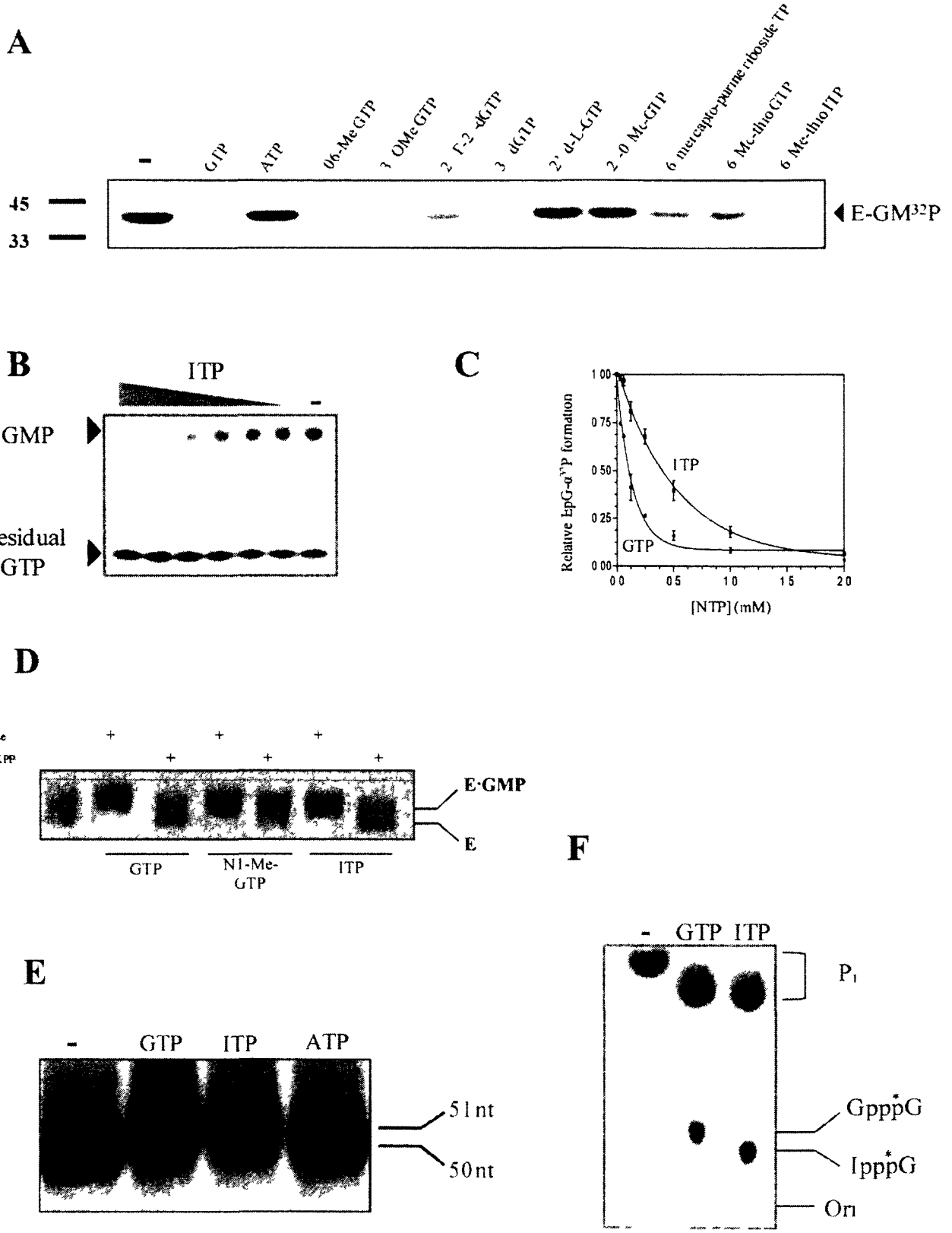
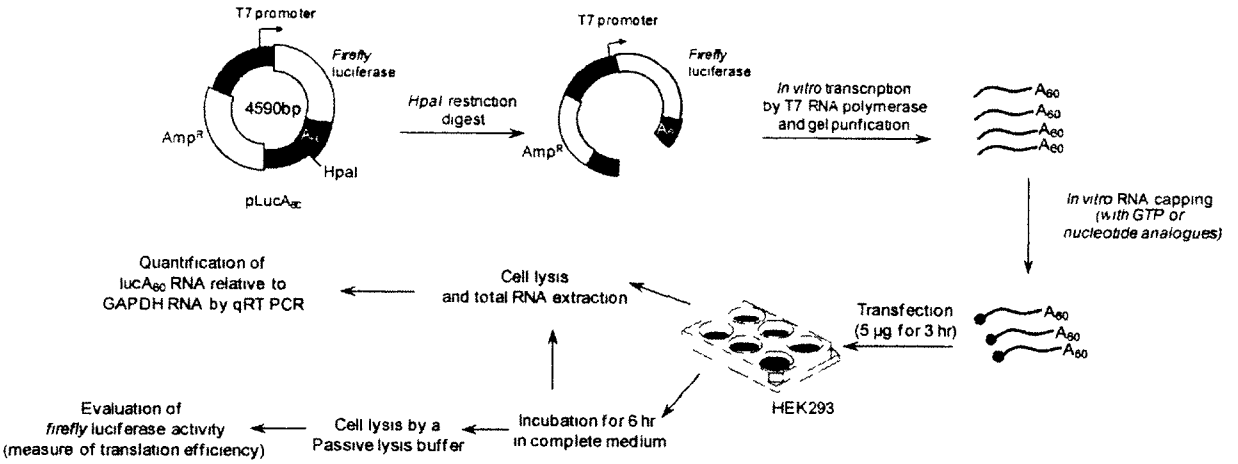
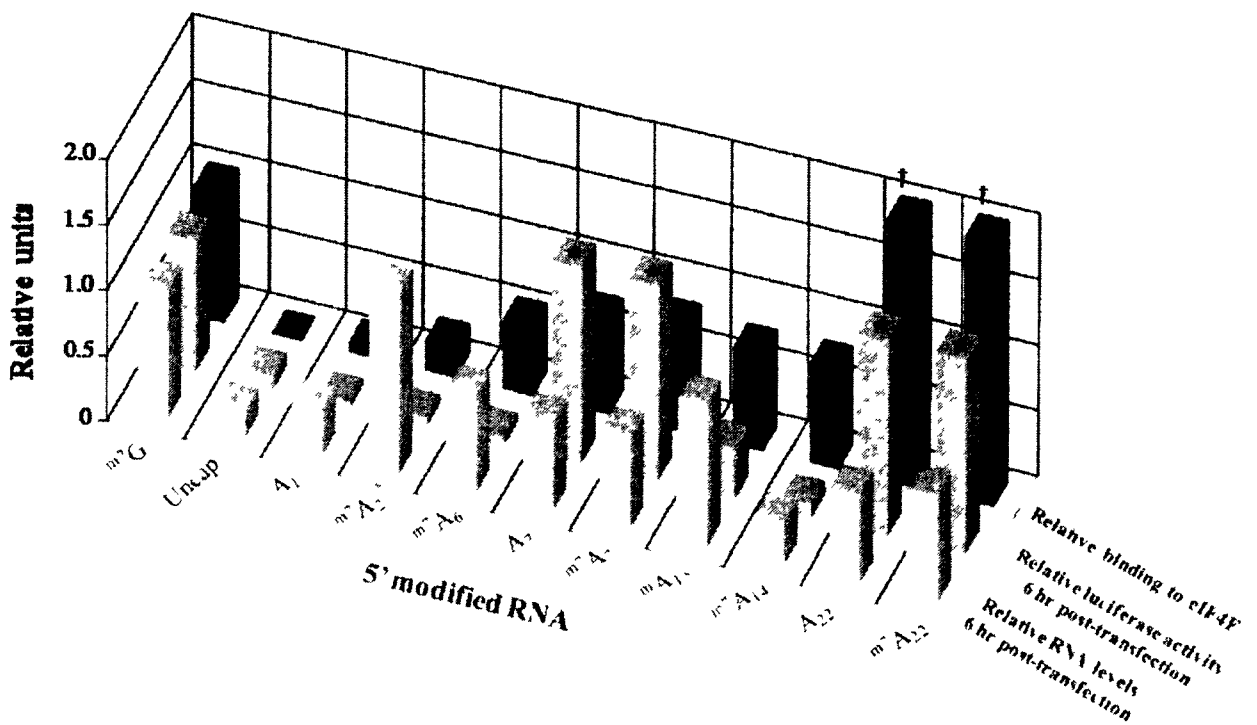

A
-
GTP
ATP
06-Me GTP
3' OMe-GTP
2' F-2' -dGTP
3' dGTP
2', d-L-GTP
2'-0 Me-GTP
6 mercapto-purine riboside TP
6 Me-thio GTP
6 Me-thio ITP
45
33
E-GM32P
B
ITP
E•GMP
Residual GTP
C
Relative EpG-α32P formation
1 00
0 75
0 50
0 25
0 00
0 0
0 5
1 0
1 5
2 0
[NTP] (mM)
ITP
GTP
D
Pyrophosphatase
KPP
E·GMP
E
GTP
N1-Me-GTP
ITP
E
-
GTP
ITP
ATP
51 nt
50 nt
F
- GTP ITP
Pi
GpppG
IpppG
Ori

Figure 3

A


T7 promoter
Firefly luciferase
4590bp
Amp^R
HpaI
pLucA60
HpaI restriction digest
In vitro transcription by T7 RNA polymerase and gel purification
In vitro RNA capping (with GTP or nucleotide analogues)
Transfection (5 µg for 3 hr)
HEK293
Incubation for 6 hr in complete medium
Cell lysis by a Passive lysis buffer
Cell lysis and total RNA extraction
Quantification of lucA60 RNA relative to GAPDH RNA by qRT PCR
Evaluation of firefly luciferase activity (measure of translation efficiency)

B


Relative units
2.0
1.5
1.0
0.5
0
Unc-ap
5' modified RNA
Relative binding to eIF4F
Relative luciferase activity 6 hr post-transfection
Relative RNA levels 6 hr post-transfection

Supplementary Table 1

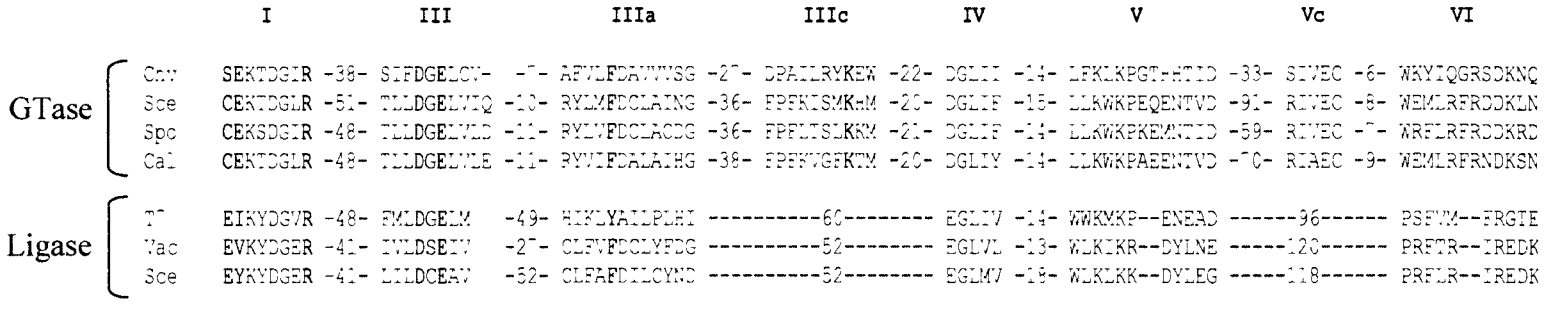
Supplementary Table 1: RNA (guanine-N7) methyltransferase activities with RNA capped with nucleotide analogues

	RNA capped with[a]	Efficiency of N7-methylation[b]
-	**GTP**	+++
(A_1)	N1-Me GTP	-
(A_2)	ITP	+++
(A_6)	6-thio GTP	+
(A_7)	O6-Me GTP	+++
(A_{13})	8-Bromo GTP	++
(A_{14})	8-Iodo GTP	+++
(A_{22})	3' O-Me GTP	+

[a] Nucleotides analogues presenting more than 30% RNA capping efficiency were selected to undergo N7-methylation by the *S. cerevisiae*'s RNA (guanine-N7) methyltransferase.

[b] The efficiency of N7-methylation was determined by comparing the release of nuclease P1 and alkaline phosphatase digestion resistant products (^{m7}NpppG (where N is a nucleotide analogue) relative to NpppG), following incubation with the S. *cerevisiae*'s MTase and S-Adenosyl-methionine.

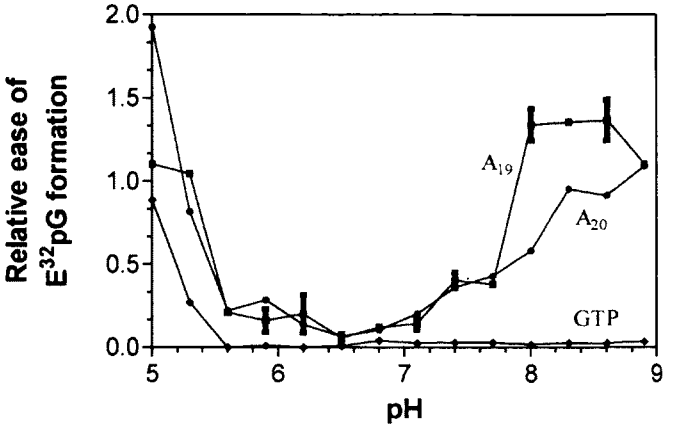
Supplementary Figure 1



Supplementary Figure 1: Structural conservation in GTases and ligases. The amino acid sequences of RNA guanylyltransferases from Paramecium bursaria Chlorella virus-1 (Chv), *S. cerevisiae* (Sce), *S. pombe* (Spo) and *C. albicans* (Cal) are aligned with ligases from the T7 phage (T7), Vaccinia virus (Vac) and Sce (*S. cerevisiae*)

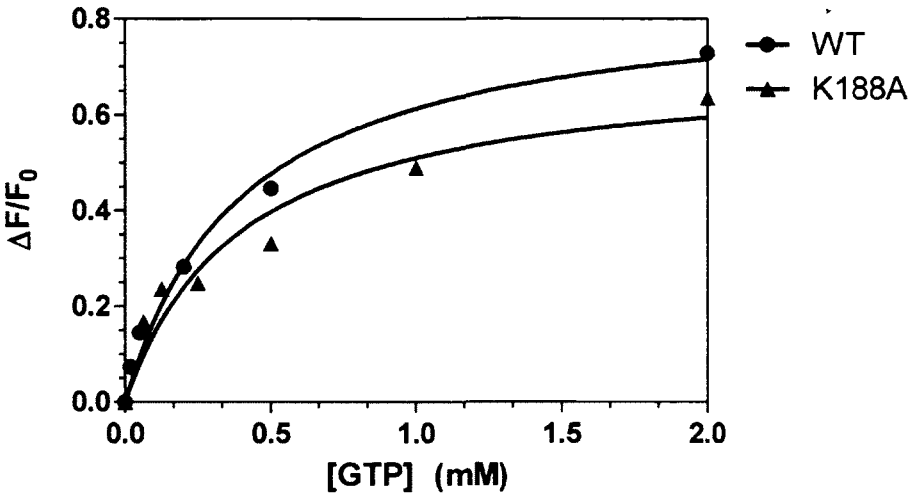
Supplementary Figure 2

A



Supplementary Figure 2: pH dependency of the inhibition by 2' modified nucleotide analogues (A) pH dependency for the inhibition by A_3. The PBCV-1 GTase was incubated with [α-^{32}P]GTP in the presence of either GTP (0.5 mM) or A_{19} or A_{20} (0.5 mM) in a standard GTase buffer ranging from a pH of 5 to 9. The reactions were resolved by SDS-PAGE and analyzed by a Phosphorimager. The formation of the radiolabelled E-GMP complex was quantified and its relative ease of formation is plotted as a function of the pH. High ease of formation implies low inhibition by either GTP or the nucleotide analogue, while low ease of formation implies high inhibition by the unlabelled nucleotides.

Supplementary Figure 3



Supplementary figure 3: Binding of GTP to the wild-type and K188A mutant of the PBCV-1 GTase. Increasing amounts of GTP were added to a 2 μM solution of the enzyme in binding buffer (50 mM Tris/HCl, pH 8.0, and 50 mM KOAc) and the emission spectrum was scanned from 310 to 440 nm, following excitation of tryptophan residues at 290 nm.

Chapter III – Identification of a novel class of viral RNA guanylyltransferase

3.1. ARTICLE:

The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure

Moheshwarnath Issur*, Brian J. Geiss*, Isabelle Bougie, Frederic Picard-Jean, Simon Despins, Joannie Mayette, Sarah E. Hobdey and Martin Bisaillon

Article published in *RNA*, Dec 2009, 15: 2340 - 2350

CONTRIBUTIONS

I had the original idea and made the actual demonstration that NS5 possess RNA guanylyltransferase activity and characterized this activity. BJG of the Colorado State University provided purified versions of the NS5 protein from the various flaviviruses analyzed in this study. IB helped in the characterization of the RNA guanylyltransferase activity. FPJ, SD and JM also helped in the characterization of the activity although their work was not included in the final paper and provided helpful comments thoughout the project. SEH is a student in the lab of BJG who helped in the expression and purification of all purified versions of the NS5 proteins used in this study. MB provided the necessary funding and was instrumental to the establishment of the collaboration with the lab of BJG.

SUMMARY ***(en français)***

L'extrémité 5' du génome des flavivirus abrite une structure coiffe d'ARN dite de type 1. Cette structure est générée par l'activité ARN triphosphatase de la protéine virale NS3, l'ARN (guanine-N7) méthyltransférase et l'ARN (2'O) méthyltransférase de la protéine virale NS5, et une ARN guanylyltransférase qui n'avait pas encore été identifiée. En raison du rôle crucial de la structure coiffe pour le métabolisme et la stabilité des ARNm, l'identification complète du complexe de synthèse de la structure coiffe des flavivirus est apparue comme un défi important. Nous avons étudié la capacité de la protéine NS5 du virus du Nil occidental à catalyser la réaction ARN guanylyltransférase. La protéine NS5 est la plus grosse protéine non-structurale chez les flavivirus, hébergeant, en plus de ses activités ARN méthyltransférases, l'activité ARN polymérase ARN-dépendante qui est cruciale pour la réplication et la transcription du génome viral. La protéine NS5 ne possède aucun des motifs conservés chez les ARN guanylyltransférases conventionelles (Fig. 10). Dans cette étude, nous avons montré que le GTP peut être utilisé comme substrat par l'enzyme pour former un lien covalent enzyme-GMP, caractéristique du mécanisme d'ARN guanylyltransférase. Nos données indiquent que le groupement GMP peut être transféré à la fin diphosphate d'un transcrit d'ARN pour former la structure coiffe d'ARN virale. Cette étape est dépendante de la nature du premier nucléotide en 5' de l'ARN utilisé. Ceci pourrait expliquer pourquoi cette activité de la protéine NS5 a échappé à toute identification jusqu'à présent. Nous avons aussi d'autres résultats expérimentaux qui démontrent que la protéine NS5 d'autres virus du genre *flavivirus* possèderait l'activité ARN guanylyltransférase, notamment les protéines NS5 des virus de la fièvre jaune, de la dengue et de l'encéphalite japonaise. Finalement, nous démontrons que la protéine NS3 des flavivirus stimule l'activité ARN guanylyltransférase de la protéine NS5, et que ces deux enzymes sont suffisantes pour catalyser la formation *de novo* d'une structure coiffe d'ARN. Grâce à cette démonstration biochimique qui établit clairement que la protéine NS5 possède une activité ARN guanylyltransférase, nous étions donc en mesure de proposer un modèle pour le complexe de réplication flavivirus. Comme l'activité ARN guanylyltransférase est indispensable, cette étude a des implications majeures pour l'élaboration de médicaments antiviraux contre les flavivirus pathogènes comme les virus de la fièvre dengue, de la fièvre jaune, de l'encéphalite japonaise et du Nil Occidental.

SUMMARY

The 5'-end of the flavivirus genome harbors an RNA cap 1 structure which is generated by the virus-encoded RNA triphosphatase found on the NS3 protein, RNA (guanine-N7) methyltransferase and nucleoside 2'-O methyltransferase both found on the NS5 protein, and an as yet unidentified RNA guanylyltransferase. Because of the crucial role of the RNA cap structure for RNA metabolism and stability, the complete identification of the molecular species involved in the capping of flavivirus RNA appeared as an important challenge. In the present study, we investigated the ability of the West Nile virus NS5 protein to catalyze the elusive RNA guanylyltransferase reaction. The NS5 protein is the largest non-structural protein of flaviviruses, harboring, in addition to its RNA methyltransferase activities, the RNA dependent RNA polymerase activity required for the replication and transcription of the viral genome. Moreover, it does not possess any of the classically conserved motifs of conventional RNA guanylyltransferases (Fig. 10). In this study I demonstrated that GTP can be used as a substrate by the enzyme to form a covalent GMP-enzyme intermediate, characteristic of the mechanism of known RNA guanylyltransferases. Furthermore, our data indicate that the GMP moiety can be transferred to the diphosphate end of an RNA transcript to form the classical RNA cap structure; a step which is highly dependent on the nature of the initiating nucleotide of the RNA used. We also provide experimental evidence that the NS5 protein from other flaviviruses harbour the active site of the RNA guanylyltransferase. Finally, we also demonstrate that the flavivirus NS3 protein stimulates the RNA guanylyltransferase activity of the NS5 protein, and that both enzymes are sufficient to catalyze the de novo formation of a methylated RNA cap 1 structure. Our study provides the first biochemical evidences that flaviviruses encode an RNA capping enzyme and we were, thus, able to propose a model for the flaviviral replication complex. Since RNA guanylyltransferase activity is known to be essential, this study has major implications for the elaboration of antiviral drugs against major pathogenic flaviviruses like Dengue Fever virus, Yellow Fever Virus, Japanese Encephalitis virus and West Nile virus.

ABSTRACT

The 5'-end of the flavivirus genome harbors a methylated $^{m7}GpppA_{2'OMe}$ cap structure, which is generated by the virus-encoded RNA triphosphatase, RNA (guanine-N7) methyltransferase, nucleoside 2'-*O* methyltransferase, and RNA guanylyltransferase. The presence of the flavivirus guanylyltransferase activity in NS5 has been suggested by several groups but has not been empirically proven. Here we provide evidence that the N-terminus of the flavivirus NS5 protein is a true RNA guanylyltransferase. We demonstrate that GTP can be used as a substrate by the enzyme to form a covalent GMP-enzyme intermediate via a phosphoamide bond. Mutational studies also confirm the importance of a specific lysine residue in the GTP binding site for the enzymatic activity. We show that the GMP moiety can be transferred to the diphosphate end of an RNA transcript harboring an adenosine as the initiating residue. We also demonstrate that the flavivirus RNA triphosphatase (NS3 protein) stimulates the RNA guanylyltransferase activity of the NS5 protein. Finally, we show that both enzymes are sufficient and necessary to catalyze the *de novo* formation of a methylated RNA cap structure *in vitro* using a triphosphorylated RNA transcript. Our study provides biochemical evidence that flaviviruses encode a complete RNA capping machinery.

INTRODUCTION

The 5'-end of eukaryotic mRNAs and many viral mRNAs harbors a ^{m7}GpppN cap structure that plays a crucial role in the translation and stability of mRNAs (Shuman 2001). Synthesis of the cap structure involves three distinct enzymatic activities (Furuichi and Shatkin 1976). In the first step of the reaction, the RNA 5'-triphosphate end of the nascent RNA is hydrolyzed by an RNA triphosphatase (RTase) to form a diphosphate end. An RNA guanylyltransferase (GTase) then catalyzes a two-step reaction in which it utilizes GTP as a substrate to form a covalent GMP-enzyme intermediate. The GMP moiety is then transferred to the diphosphate end of the RNA transcript in the second step of the reaction to form the GpppN structure. Lastly, the guanosine base is methylated by an RNA (guanine-N7) methyltransferase (N7MTase) to form the typical ^{m7}GpppN cap structure. The ribose of the nucleotides adjacent to the cap structure can also be 2'-*O* methylated to various extents by a nucleoside 2'-*O* methyltransferase (2'OMTase) (Shatkin 1976).

The *Flavivirus* genus includes more than 70 human pathogens such as West Nile, yellow fever, and Dengue viruses (Heinz et al. 2000). The flavivirus genome is a single-stranded RNA of positive polarity that possesses a methylated cap 1 structure at its 5' end (^{m7}GpppA$_{2'OMe}$-RNA) (Brinton 2002). The genome encodes for a single open reading frame of 11 kb in length that is translated into a large polyprotein precursor which is processed into three structural (capsid C, membrane M, and envelope E) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) by both viral and cellular proteases (Brinton 2002; Nowak et al. 1989; Wengler et al. 1991). The NS5 protein is the largest non-structural protein of flaviviruses. It harbors the RNA-dependent RNA polymerase (RdRp) activity which is required both for replication and transcription of the viral genome (Choi et al. 2004). The N-terminal domain of NS5 also contains N7MTase and 2'OMTase activities, both of which are involved in the synthesis of the RNA cap 1 structure (Geiss et al. 2009; Egloff et al. 2002). The RTase activity is mediated by the C-terminal domain of NS3 (Benarroch et al. 2004a) while the precise nature of the GTase remains uncertain. Recent cross-linking and fluorescence polarization assays have shown that GTP can bind to the MTase domain of NS5 with high affinity, and a thorough biochemical understanding of how the methyltransferase domain binds GTP is available

(Geiss et al. 2009; Bollati et al. 2009). Crystal structures of the flavivirus MTase complexed with GTP show how the MTase can bind to GTP and suggest a possible site for GMP-enzyme formation (Geiss et al. 2009; Egloff et al. 2002; Benarroch et al. 2004a; Bollati et al. 2009; Zhou et al. 2007; Assenberg et al. 2007; Egloff et al. 2007). Moreover, the MTase domain of the NS5 protein of Wesselsbron virus was recently demonstrated to covalently bind GMP upon incubation with GTP, hinting to a possible role in RNA capping chemistry (Bollati et al. 2009). However, despite the structural and biochemical evidence suggesting that the NS5 MTase domain can bind GTP and form a GMP-enzyme complex, there has been no experimental demonstration of GTase activity i.e. there is no evidence that the GMP moiety can be transferred to an acceptor RNA.

In the present study we investigated if the flavivirus NS5 protein is a true RNA GTase which can catalyze the formation of the viral RNA cap structure. We demonstrate that GTP can be used as a substrate by the flavivirus NS5 protein to form a covalent GMP-enzyme intermediate, and that the presence of NS3 is stimulatory to the formation of the GMP-enzyme intermediate. Furthermore, our data indicate that GMP can be transferred from NS5 to the diphosphate end of an acceptor RNA transcript. Finally, we show that NS3 and NS5 are sufficient for the formation of the complete RNA cap structure *in vitro*. These data demonstrate that the flavivirus NS5 protein is a true RNA GTase.

RESULTS

Formation of an enzyme-GMP covalent intermediate

The first step of the RNA GTase reaction entails the nucleophilic attack of the α-phosphate of GTP by the enzyme and the subsequent formation of a covalent enzyme-GMP intermediate (Shuman and Hurwitz 1981; Shuman 1982; Venkatesan and Moss 1982). The ability of the purified full-length NS5 ($NS5^{FL}$) protein (Fig. 1A) from West Nile virus (WNV) to form a covalent GMP-enzyme intermediate was detected by label transfer from [α-^{32}P]GTP to the enzyme in the presence of magnesium ions. A single SDS-stable GMP-enzyme complex was detected following SDS-PAGE (Fig. 1B). No labelling was observed upon incubation with [γ-^{32}P]GTP (data not shown). Since the presence of pyrophosphate has been shown to strongly inhibit the forward RNA guanylyltransferase reaction (Ho et al. 1996), we therefore decided to investigate the effect of pyrophosphate on the reaction catalyzed by the WNV $NS5^{FL}$ protein. Our results demonstrate that the addition of pyrophosphate inhibited the formation of the covalent enzyme-GMP intermediate. We therefore conclude that the WNV $NS5^{FL}$ protein is active in the formation of a protein-GMP covalent complex, and that pyrophosphate is a potent inhibitor of the reaction.

The formation of the covalent $NS5^{FL}$-GMP complex was quantified as described under "Materials and Methods." As can be seen in figure 1C, the yield of enzyme-GMP formation in the presence of 3 μM (150 pmoles) of $NS5^{FL}$ increased as a function of GTP concentration. The titration assay indicated that half-saturation was reached at a concentration of 1 mM GTP. At this GTP concentration, we estimated that at least 30% of the $NS5^{FL}$ molecules were labelled with GMP *in vitro* (50 pmoles of enzyme-GMP complex). As observed for other GTases (Ho et al. 1998; Cong and Shuman 1995), this substoichiometric GMP labeling is likely due to the fact that a significant fraction of the enzyme is already guanylylated and remains so during purification of the enzyme from bacteria. Moreover, GTase activity was only observed in the presence of magnesium ions. Manganese, cobalt, calcium, copper, nickel, and zinc could not support the formation of the $NS5^{FL}$-GMP complex (Fig. 1D).

In order to determine the nature of the enzyme-GMP linkage, the radiolabelled $NS5^{FL}$-GMP complex was isolated by gel filtration, and submitted to chemical treatment. In the classical GTase reaction, the nucleophilic attack on the α-phosphate of GTP by the enzyme results in the formation of a covalent intermediate in which GMP is typically covalently linked via a phosphoamide bond to the attacking lysine residue of the enzyme (Shuman and Hurwitz 1981; Pena et al. 1993; Cong and Shuman 1993). The reaction products resulting from chemical treatments were analyzed by thin layer chromatography (Fig. 1E). Our data indicate that the $NS5^{FL}$-GMP intermediate was resistant to NaOH treatment, but treatments with HCl or NH_2OH resulted in the release of GMP, which is indicative of a phosphoamide linkage.

The NS5 N-terminal methyltransferase domain forms a covalent linkage with GMP

The N-terminal MTase domain of NS5 has been shown to bind to GTP (Geiss et al. 2009; Bollati et al. 2009; Egloff et al. 2002), and recent evidence suggests that an enzyme-GMP complex can be formed within this domain of the Wesselsbron virus NS5 protein (Bollati et al. 2009). We wanted to validate to results from this paper and see if the enzyme-GMP complex can be formed by other flavivirus MTase domains. Incubation of the WNV NS5 MTase domain (aa 1-268) in the GTase reaction resulted in the transfer of a GMP moiety to the protein, indicating that the GMP-enzyme bond is formed within the first 268 amino acids of the NS5 protein (Fig. 2A). Note that the extent of $NS5^{1-268}$-GMP formation (55 pmoles / 150 pmoles of enzyme) was comparable to the one observed for the full-length protein. In order to evaluate whether this activity was common to other flaviviruses, we next tested the ability of other flavivirus NS5 proteins to form a protein-GMP covalent complex. Truncated versions of the NS5 protein from Dengue virus (DV) and Yellow Fever Virus (YFV) were incubated with radiolabelled GTP, and the formation of the enzyme-GMP covalent complex was evaluated by monitoring the transfer of radiolabelled GMP from [α-^{32}P]GTP to the enzyme. Again, the formation of SDS-stable GMP-enzyme complexes was detected following SDS-PAGE for both enzymes (Fig. 2A). We hereby conclude that the formation of a GMP-enzyme complex is an intrinsic property of the N-terminal portion of the flavivirus NS5 protein.

A recent paper by Bollati *et al* (Bollati et al. 2009) suggested that the GMP-enzyme bond is formed at Lys-29. However, the mass spectroscopy data reported in that paper appears to be in error, as the fragment reported as the GMP-modified peptide was the same size (2225.12 Mhz) as a commonly observed methylated trypsin fragment (Ding et al. 2003). Moreover, we were not able to detect a guanylylated K29 peptide by mass spectroscopy when the enzyme was incubated with GTP (data not shown). In order to shed light into the specific residues involved in the formation of the enzyme-GMP complex, a mutational analysis was performed on the DV NS5 protein. The Lys29 residue, which interacts with the a-phosphate of GTP (Geiss et al. 2009; Egloff et al. 2002) was therefore substituted for alanine, and the mutant polypeptide was expressed and purified in parallel with the wild-type enzyme (Fig. 2B). The effect of this single alanine mutation on the formation of the covalent enzyme-GMP complex was then investigated. Alanine substitution of Lys29 resulted in a significant decrease in the ability of the enzyme to covalently bind GMP (Fig. 2C). The extent of $NS5^{K29A}$-GMP formation at 1 mM GTP (8 pmoles of enzyme-GMP / 3 μM of enzyme) was about 15% of the wild-type enzyme. We conclude that the residue is important for catalysis. However, the fact that a residual activity is still present raises speculations on its precise role in the reaction chemistry (see Discussion).

NS5 is a true RNA guanylyltransferase

RNA GTases catalyze a two-step reaction in which they initially form a covalent enzyme-GMP intermediate which is transferred to an acceptor RNA molecule in the second part of the reaction. The flavivirus NS5 protein clearly has the ability to form a covalent GMP-enzyme intermediate, but can it transfer GMP to an RNA substrate containing a 5'-diphosphate end? We initially attempted to monitor the transfer of GMP to an acceptor RNA harboring a 5'-diphosphate end. This substrate was initially synthesized by *in vitro* transcription before being subjected to phosphohydrolysis in order to specifically remove the g-phosphate. This RNA transcript harbors a guanosine as the initiating residue and is typically used in biochemical assays to monitor the transfer of the GMP moiety to an acceptor RNA. The ability of the enzyme to transfer GMP to this RNA molecule was tested by incubating the WNV $NS5^{FL}$ protein with [α^{32}P]GTP, the RNA substrate (81 nucleotides)

containing a diphosphate 5'-end, and magnesium ions. The products of the reaction were extracted with phenol/chloroform to remove the radiolabelled protein, and the RNA recovered by ethanol precipitation. However, following electrophoresis on a denaturing polyacrylamide gel (Fig. 3), we were unable to detect an efficient transfer of the GMP moiety from the flavivirus $NS5^{FL}$-GMP complex to this acceptor RNA (<0.05 pmoles of GMP transferred / 7 pmoles of $NS5^{FL}$). Since the initiating residue of the flavivirus mRNAs is always an adenosine, we surmise that the nature of this residue might influence the RNA guanylyltransferase reaction. We therefore synthesized an RNA substrate harboring a 5'-diphosphate end with an adenosine as the initiating nucleotide (see Materials and Methods). The transfer of radiolabelled GMP to the RNA substrate was confirmed by electrophoresis on a denaturing polyacrylamide gel, indicating that the $NS5^{FL}$ protein has the ability to transfer the GMP moiety to an acceptor RNA substrate harboring an adenosine as the initiating residue (Fig. 3). We evaluated that 5 pmoles of GMP were transferred to the RNA substrate in the presence of 7 pmoles of protein. We conclude that the NS5 protein is a true RNA GTase that catalyzes both steps of the typical RNA capping reaction.

The NS3 protein stimulates the RNA guanylyltransferase activity

Allosteric regulation frequently plays a crucial role in enzymatic catalysis. We therefore investigated the ability of various effectors to modulate the GTase activity of the WNV NS5 protein. The addition of S-adenosyl-L-methionine (AdoMet), which serves as the methyl donor during the MTase reaction, did not affect the formation of the enzyme-GMP intermediate (data not shown). Similarly, the addition of other nucleotides did not result in the stimulation of the activity (data not shown). The NS5 flexible linker region between the MTase and RdRp domains has been shown to be a site of interaction between the NS3 and NS5 proteins (Yon et al. 2005). As NS3 encodes the RTase enzyme, it seemed likely that the formation of the RNA cap structure is dependent on the interaction between these two proteins. To determine if NS3 can alter the GTase activity, we performed the GTase assay with the full-length WNV NS5 protein at a concentration 2 μM in the presence of the WNV $NS3^{168-618}$ protein (10 μM). We observed that the addition of the purified WNV NS3 protein, which has no GTase activity of its own, significantly increased the formation of the $NS5^{FL}$-GMP complex (Fig. 4A). The activity increased from 15 pmoles of

$NS5^{FL}$-GMP complex / 100 pmoles of $NS5^{FL}$ to 45 pmoles of $NS5^{FL}$-GMP / 100 pmoles of $NS5^{FL}$ when the assay was performed in the presence of $NS3^{168-618}$. Moreover, we observed that the extent of $NS5^{FL}$-GMP complex formation was proportional to the concentration of input NS3 in the 4-20 μM and plateaued thereafter (Fig. 4B). A maximal stimulation of 6-fold was reached at a ~6:1 molar ratio of NS3 to NS5.

Synthesis of the RNA cap structure by the NS3 and NS5 proteins

We finally tested the ability of the flavivirus NS3 and NS5 proteins to catalyze the *de novo* formation of a methylated RNA cap structure. An 81 nucleotide RNA substrate harboring a 5'-triphosphate end with an adenosine as the initiating nucleotide was incubated with the WNV NS3 protein (RTase). The resulting RNA product was then incubated with the WNV $NS5^{FL}$ protein and [α-^{32}P]GTP in the presence or absence of *S*-adenosyl-methionine. The products of the reaction were extracted with phenol/chloroform to remove the radiolabelled protein, and the RNA products were recovered by ethanol precipitation. The transfer of the radiolabelled GMP to the RNA substrate was confirmed by electrophoresis on a denaturing polyacrylamide gel, thereby indicating that the NS5 protein has the ability to transfer a GMP moiety to the 5'-diphosphate end of an acceptor RNA substrate that was generated through the RNA 5'-triphosphatase activity of NS3 (Fig. 5A). Note that the transfer of the radiolabelled GMP moiety to the RNA was not observed in control reactions performed in the absence of either NS3 or $NS5^{FL}$ (Fig. 5A). Aliquots of the RNA samples were then digested with nuclease P1 and alkaline phosphatase, and analyzed by polyethyleneimine-cellulose thin layer chromatography (Fig. 5B). The reaction performed in the presence of *S*-adenosyl-methionine displayed a faster migrating species corresponding to the RNA cap ^{m7}GpppA structure, as judged from the unlabelled markers, visualized under ultraviolet light. These data indicate that NS3 and NS5 are the sole flaviviral enzymes involved in genomic RNA cap formation, and that we have observed the formation of a flavivirus RNA cap structure *in vitro*.

DISCUSSION

Our study provides biochemical evidences that the flavivirus NS5 protein is a true RNA GTase that has the ability to form a covalent protein-GMP intermediate and subsequently transfer the GMP moiety to a 5'-diphosphate RNA. There have been several pieces of evidence used to suggest that NS5 MTase domain contains the flaviviral GTase activity, including structural and biochemical analysis of MTase GTP binding, limited data showing covalent formation of a GMP-enzyme complex, and structural data with cap analogs to propose a mechanism for RNA capping (Geiss et al. 2009; Egloff et al. 2002; Egloff et al. 2007). These reports, while important to understand the MTase activity of the enzyme, did not demonstrate that the MTase domain catalyzes a true GTase reaction. The data that we present in the current study fills this gap by demonstrating that NS5 can catalyze both steps of the RNA GTase reaction. Moreover, the enzyme can synthesize a complete methylated cap structure in the presence of NS3 using a nascent triphosphorylated RNA transcript harboring a 5' adenine base. Therefore, we conclude that the N-terminus of NS5, in combination with the NS3 RTase, comprises the complete flavivirus capping enzyme.

The identification of the flaviviral RNA GTase highlights the large diversity in the physical organization of the RNA capping machineries. For instance, the RTase and GTase activities of metazoans and plants are catalyzed by a single protein which contains both active sites while both the N7MTase and 2'OMTase activities reside on separate single-domain enzymes (Shuman 2000). In yeast, the RTase, GTase, and N7MTase are all catalyzed by separate proteins (Mao et al. 1995; Shibagaki et al. 1992). Interestingly, an important variety of molecular organizations has been observed for viruses. For instance, the poxviruses harbor a single protein that contains the active site for the RTase, GTase, and N7MTase while a different protein is responsible for the 2'OMTase activity (Higman et al. 1994; Schwer et al. 2006). In alphaviruses, both the GTase and N7MTase activities are catalyzed by a single enzyme while the RTase active site is located on a different peptide (Ahola and Kääriäinen 1995; Vasiljeva et al. 2000). Mammalian reoviruses harbor a single multi-functional protein containing the active sites of the GTase, N7MTase, and 2'OMTase while the RTase reaction is performed by a distinct protein (Bisaillon and Lemay 1997a;

Luongo et al. 2000). This particular RNA capping machinery is reminiscent of what is observed in flaviviruses where the NS3 protein contains the active site of the RTase while the NS5 protein is responsible for the N7MTase, 2'OMTase, and GTase activities. However, a distinction between the two viral families is the additional presence of the RNA polymerase active site on the multi-functional enzyme (NS5) (Oh et al. 1999). The multiple essential enzymatic activities catalyzed by the flavivirus NS5 protein clearly make it an attractive target for future drug development.

The GTases of metazoans, plants, and DNA viruses are members of the covalent nucleotidyl transferase superfamily, which also includes DNA and RNA ligases (Cong and Shuman 1993). These GTases harbor a consensus KxDG motif in which the ε-amino group of the active-site lysine is covalently bound to GMP through a phosphoamide bond (Cong and Shuman 1993). However, the GTases of several RNA viruses lack this precise consensus sequence found in DNA viruses and cellular GTases. For instance, this consensus signature is not found in the GTase of alphaviruses and several genera of the *Reoviridae* family (Ahola and Kääriäinen 1995; Supyani et al. 2007; Cleveland et al. 1986; Hsiao et al. 2002; Liu et al. 1992; Le Blois et al. 1992; Qiu and Luongo 2003; Mohd Jaafar et al. 2005; Suzuki et al. 1996). Similarly, the KxDG consensus sequence is conspicuously absent form the NS5 protein of flaviviruses. Henceforth we speculate that the residue involved in the formation of the covalent enzyme-GMP intermediate is either a lysine located outside the classically consensus sequence or mediated via a completely different amino acid residue since formation of the enzyme–GMP intermediate can potentially occur with the imidamino group of a histidine residue and the guanido-amino group of an arginine residue (Shuman and Hurwitz 1981). Mutagenesis studies of the DV NS5 protein by us and others have highlighted the importance of Lys29 in the binding of GTP (Geiss et al. 2009; Bollati et al. 2009). However, a residual GTase activity is still present upon substitution of the lysine residue by alanine which raises questions about the precise role of the lysine residue in the catalytic activity. Interestingly, the residue next to the Lys29 residue is also a lysine (Lys-30) in the DV NS5 protein. It is possible that altering the structure of the protein upon replacement of the Lys29 residue by alanine may allow the Lys30 residue to come closer than it normally would and form a weak bond with the GMP

moiety. Surprisingly, the Lys29 position is not strictly occupied by a lysine residue in all flaviviruses; in some members of the family such as West Nile virus, Japanese encephalitis, and St. Louis encephalitis virus, an arginine is found at the position normally occupied by the lysine (Bollati et al. 2009). This is somewhat surprising since all known GTases identified so far are thought to bind the GMP moiety through a lysine residue (Shuman and Hurwitz 1981; Pena et al. 1993; Cong and Shuman 1993). The precise biochemical consequences of this conservative substitution remain to be investigated.

The crystal structures of different members of the nucleotidyl transferase superfamily have been determined and have provided insightful information on the GTase reaction chemistry (Fabrega et al. 2003; Doherty et al. 1997; Hakansson et al. 1997; Lee et al. 2000; Odell et al. 2000). Members of the family are characterized by a common tertiary fold that consists of an N-terminus, which encompasses the nucleotide-binding pocket, and a C-terminal oligonucleotide binding-fold domain (Odell et al. 2000; Hakansson and Wigley 1998; Benarroch et al. 2004b; Malet et al. 2007; Yap et al. 2007). Examination of the *Chlorella* virus GTase crystals have revealed that a large conformational change occurs on GTP binding, shifting the structure from an open to a closed state (Hakansson et al. 1997; Hakansson and Wigley 1998). On the basis of these crystallographic studies, a model has been suggested, in which the conformational change encountered on GTP binding would promote metal ion binding and guanylylation (Hakansson et al. 1997; Hakansson and Wigley 1998). In contrast, recent crystallographic data suggests the presence of important structural and mechanistic differences between the flavivirus NS5 protein and the members of the nucleotidyl transferase superfamily (Hakansson and Wigley 1998; Benarroch et al. 2004b; Malet et al. 2007; Yap et al. 2007; Ago et al. 1999; Bressanelli et al. 1999; Lesburg et al. 1999). The GTase-MTase domain of the flavivirus NS5 protein is characterized by three subdomains with a core region that resembles the topology observed in the catalytic domain of all other AdoMet-dependent MTases (Geiss et al. 2009; Bollati et al. 2009; Lesburg et al. 1999). The GTP-binding site located on the N-terminal appendage of the classical MTase core region shows a previously unreported fold and a structurally novel way of promoting the specific binding of GTP by which the specific contacts to the guanine base are exclusively provided by main-chain atoms to the 2-amino group and the

N1 nitrogen via a water bridge (Geiss et al. 2009). This is in contrast to the nucleotidyl transferase superfamily GTases in which the specificity for GTP is achieved through interactions with the 6-oxo and 2-amino group of guanine (Fabrega et al. 2003; Doherty et al. 1997; Hakansson et al. 1997; Lee et al. 2000; Odell et al. 2000). Finally, binding of the GTP molecule to the flavivirus GTase active site is not characterized by large conformational changes in the architecture of the enzyme (Geiss et al. 2009).

Our study demonstrates that the GTase activity of NS5 is stimulated by the presence of NS3. Future biochemical studies will undoubtedly reveal key insights on this catalytic regulation. Interestingly, a similar stimulation of the GTase (Ceg1) upon binding to the RTase (Cet1) is also observed in *Saccharomyces cerevisiae* where the Cet1-Ceg1 interaction increases the extent of formation of the Ceg1-GMP complex (Ho et al. 1999; Hausmann et al. 2001; Ho et al. 1998). This interaction is crucial for the stabilization of the yeast GTase which is inherently thermolabile (Hausmann et al. 2001). Moreover, the interaction also enhances the affinity of the enzyme for the GTP substrate (Ho et al. 1998). Such an allosteric stimulation of an enzyme involved in RNA capping chemistry is also encountered in the vaccinia virus capping machinery where the active site of the N7MTase is located within the C-terminal portion of the virus-encoded D1 protein (Schnierle et al. 1992; Mao and Shuman 1994). The D1 protein has a weak intrinsic MTase activity which is stimulated allosterically by the binding of the viral D12 protein (Mao and Shuman 1994). The basal level of MTase activity can be stimulated by as much as 100-fold by the addition of the catalytically inert D12 protein (Mao and Shuman 1994). NS3 is known to interact with the flexible linker region on NS5, positioning NS3 between the NS5 RdRP and MTase-GTase domains (Yon et al. 2005). With the current demonstration that NS3 can stimulate the NS5 GTase activity and that NS3 and NS5 are sufficient for RNA cap formation, a model of the entire RNA capping process can now be proposed (Fig. 6). In this model, the nascent positive strand RNA immediately interacts with the NS3 RTase following polymerization by the NS5 C-terminal RdRp, leading to the removal of the γ-phosphate from the RNA. The modified RNA would then be in position to interact with the NS5 MTase-GTase domain. Association of NS3 with NS5 increases the GTase activity of the MTase-GTase domain, and results in the transfer of GMP to the modified RNA,

resulting in the synthesis of a capped RNA transcript. Subsequently, the modified RNA would be in excellent position for N7-methylation to generate the RNA cap structure. Future studies will seek to rigorously test this model for RNA cap formation.

Because of the crucial role of the RNA cap structure for mRNA metabolism and stability, potent inhibitors of pathogen GTases are an attractive area for drug development. For instance, an elegant study recently demonstrated the ability of non-nucleoside inhibitors to inhibit the mRNA guanylylation of the human respiratory syncytial virus transcripts (Liuzzi et al. 2005). These GTase inhibitors were shown to exhibit antiviral activities both *in vitro* and in a mouse model of infection. The broad-spectrum antiviral ribavirin triphosphate, a nucleoside analog, has also been shown to directly interact with the guanylyltransferase of vaccinia virus thereby inhibiting the guanylylation of RNA transcripts (Bougie and Bisaillon 2004). However, because of the high overall similarity between viral and cellular GTases, it was initially suggested that mechanism-based inhibitors might not be specific and show adverse side effects on human cells (Bougie and Bisaillon 2004). Nonetheless, mounting evidence suggests that subtle structural and mechanistic differences exist between the RNA viruses and cellular GTases. These differences will undoubtedly be exploited in the near future for the rational design of potent antivirals against these medically important viruses.

MATERIALS AND METHODS

Expression and purification of the flavivirus NS5 protein

Expression and purification of the NS5 and NS3 protein from West Nile virus, Dengue virus, and Yellow fever virus was performed as described before (Geiss et al. 2009; Bougie and Bisaillon 2009; Benzaghou et al. 2006). Briefly, recombinant full-length WNV NS5 protein (strain NY1999, aa 1-905), truncated WNV NS3 protein (strain NY99, aa 168–618), and the NS5 MTase domains from DEN (strain 16681, aa 1–267), YF (strain 17D, aa 1–268), and WNV (strain NY1999, aa 1–268) were expressed from inducible T7 expression plasmids that contain a carboxy-terminal 6-histidine tag. The plasmids were transformed into *Escherichia coli* BL21(DE3) and cultures were induced with 400 μM IPTG overnight at 22 °C. The bacterial pellets were collected, sonicated, and the histidine-tagged proteins were purified from clarified lysates by affinity chromatography based on the affinity of the tag for immobilized nickel. The eluted proteins were further purified over a HiTrap-S cation-exchange column (Amersham). The isolated proteins were > 99% pure as estimated from SDS-PAGE and Coomasie blue staining.

Assay for enzyme-GMP complex formation

The assay was performed by incubating the enzyme (3 μM) with 1 mM [α-^{32}P]GTP in a buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, and 5 mM $MgCl_2$ for 60 min at 37 °C. The reactions were stopped by the addition of EDTA to 10 mM and SDS to 1%. The reactions were analyzed by electrophoresis through a 12.5% polyacrylamide gel containing 0.1% SDS. The radiolabelled proteins were visualized by autoradiography of the gel. The extent of covalent complex formation was quantitated by scanning the gel with a PhosphorImager (Amersham Biosciences). Alternatively, the NS5-GMP labelled complex was isolated by gel filtration and treated with either 0.5 M NaOH or 0.5 M HCL at 70 °C for 10 min or with 3.8 M hydroxylamine (NH_2OH) in 0.2 M sodium acetate (pH 4.8) at 37 °C for 20 min. The sample products were analyzed by thin layer chromatography on a polyethyleneimine-cellulose plate and developed with 0.75 M KH_2PO_4.

Preparation of the 5'-diphosphate RNA substrate

An RNA substrate of 81 nt was synthesized using a pair of complementary

oligonucleotides including a modified T7 RNA promoter (TAATACGACTCACTATAAGC) which allowed the synthesis of an RNA transcript (pppAGCCATG...) that harbored an adenosine as the initiating nucleotide. A standard T7 RNA promoter was also used as a control to generate an RNA harboring a guanosine as the initiating nucleotide. The sequences of these RNAs are not derived from WNV and contain a random sequence having no bearing to the sequence or structure of the 5' terminus of the WNV positive stranded RNA. In both cases, the RNA substrates of 81 nucleotides were synthesized with the MAXIscript kit (Ambion) using T7 RNA polymerase. The corresponding bands were excised and then eluted from the gel by an overnight incubation in 0.1% SDS and 0.5 M ammonium acetate. The 5'-triphosphate RNAs were then precipitated with ethanol and quantified by spectrophotometry. An RNA triphosphatase reaction was performed in a buffer containing 50 mM Tris/HCl (pH 7.0), 5 mM DTT, 1 mM $MgCl_2$, 1 pmol of RNA, and 1 μg of purified *Chlorella* virus RTase. The reaction mixture was incubated at 30 °C for 30 min, and stopped by the addition of formic acid (1.25 M). The RNA substrate was purified on a denaturing 20% polyacrylamide gel and visualized by ultraviolet shadowing. The corresponding band was excised and then eluted from the gel by an overnight incubation in 0.1% SDS and 0.5 M ammonium acetate. The 5'-diphosphate RNA was then precipitated with ethanol and quantified by spectrophotometry.

Transfer of GMP to RNA

The transfer of GMP to the 5'-diphosphate RNA was assayed by monitoring the transfer of [^{32}P]GMP to the RNA substrate (1 μg) in a buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 5 mM DTT, 2 μM of purified $NS5^{FL}$ protein, and 1 mM [α-^{32}P]GTP. The reaction was incubated at 37 °C for 60 min the RNA was extracted with phenol/chloroform and recovered by ethanol precipitation. The RNA was analyzed on a denaturing 10% polyacrylamide gel. The gel was then scanned with a PhosphorImager (Amersham Biosciences).

Synthesis of the RNA cap structure by the WNV NS3 and NS5 proteins

The RNA capping reaction was performed in a buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, 1 μg of triphosphate-terminated RNA (81 nucleotides) and 2 μM of

purified WNV $NS3^{168-618}$ protein. The reaction was incubated at 37 °C for 60 min. The RNA was extracted with phenol/chloroform, recovered by ethanol precipitation, adjusted to 5 mM $MgCl_2$, and incubated with 3 μM of purified WNV $NS5^{FL}$ protein and 1 mM [α-^{32}P]GTP. The sample was incubated at 37 °C for 60 min. The RNA sample was extracted with phenol/chloroform, recovered by ethanol precipitation, and analyzed on a denaturing 10% polyacrylamide gel.

Alternatively, aliquots of the RNA capping reactions were adjusted to 50 mM NaOAc, pH 5.2, and digested with nuclease P1 (5 μg) for 60 min at 37 °C. The reactions were then adjusted to 50 mM Tris-HCl, pH 8.0, and digested with alkaline phosphatase (1 unit) for 30 min at 37 °C. The reaction products were analyzed by thin layer chromatography on a polyethyleneimine-cellulose plate developed with 0.4 M ammonium sulfate.

ACKNOWLEDGMENTS

We thank Dominique Lévesque for technical assistance. This work was funded by grants from the Canadian Institutes for Health Research and the Natural Sciences and Engineering Research Council of Canada to MB, and a grant from the Rocky Mountain Regional Center for Excellence (U54 AI-065357) to BJG. MB is a Chercheur Boursier Junior 2 from the Fonds de recherche en santé du Québec and a member of the Infectious diseases group of the Centre de Recherche Clinique Étienne-Lebel.

REFERENCES

Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K. and Miyano, M. 1999. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure* **7:** 1417-1426.

Ahola, T. and Kääriäinen, L. 1995. Reaction in alphavirus mRNA capping: Formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl. Acad. Sci. USA* **92:** 507-511.

Assenberg, R., Ren, J., Verma, A., Walter, T.S., Alderton, D., Hurrelbrink, R.J., Fuller, S.D., Bressanelli, S., Owens, R.J., Stuart, D.I. and Grimes, J.M. 2007. Crystal structure of the Murray Valley encephalitis virus NS5 methyltransferase domain in complex with cap analogues. *J Gen Virol.* **88:** 2228-2236.

Benarroch, D., Selisko, B., Locatelli, G.A., Maga, G., Romette, J.L. and Canard, B. 2004a. The RNA helicase, nucleotide 5'-triphosphatase, and RNA 5'-triphosphatase activities of Dengue virus protein NS3 are Mg^{2+}-dependent and require a functional Walker B motif in the helicase catalytic core. *Virology* **328:** 208-218.

Benarroch, D., Egolff, M.P., Mulard, L., Guerreiri, C., Romette, J.L. and Canard, B. 2004b. A structural basis for the inhibition of the NS5 Dengue virus mRNA 2'-*O*-methyltransferase domain by ribavirin 5'-triphosphate. *J. Biol. Chem.* **279:** 35638-35643.

Benzaghou, I., Bougie, I., Picard-Jean, F. and Bisaillon, M. 2006. Energetics of RNA binding by the West Nile virus RNA triphosphatase. *FEBS Lett.* **580:** 867-877.

Bisaillon, M. and Lemay, G. 1997a. Characterization of the reovirus □1 protein RNA 5'-triphosphatase activity. *J. Biol. Chem.* **272:** 29954-29957.

Bisaillon, M. and Lemay, G. 1997b. Viral and cellular enzymes involved in synthesis of mRNA cap structure. *Virology* **236:** 1-7.

Bollati, M., Milani, M., Mastrangelo, E., Ricagno, S., Tedeschi, G., Nonnis, S., Decroly, E., Selisko, B., Lamballerie, X., Coutard, B., Canard, B. and Bolognesi, M. 2009. Recognition of the RNA cap in Wesselsbron virus NS5 methyltransferase domain: implications for RNA capping mechanisms in Flavivirus. *J. Mol. Biol.* **385:** 140-152.

Bougie, I. and Bisaillon, M. 2004. The broad spectrum antiviral nucleoside ribavirin as a substrate for a viral RNA capping enzyme. *J. Biol. Chem.* **279:** 22124-22130.

Bougie, I. and Bisaillon, M. 2009 Metal ion-binding studies highlight important differences between flaviviral RNA polymerases. *Biochim. Biophys. Acta* **1794:** 50-60.

Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R.L., Mathieu, M., De Francesco, R. and Rey, F.A. 1999. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **96:** 13034-13039.

Brinton, M.A. 2002. The molecular biology of West Nile Virus, a new invader of the western hemisphere. *Annu. Rev. Microbiol.* **56:** 371-402.

Choi, K.H., Groarke, J.M., Young, D.C., Kuhn, R.J., Smith, J.L., Pevear, D.C. and Rossmann, M.G. 2004. The structure of the RNA-dependent RNA polymerase from bovine viral diarrhea virus establishes the role of GTP in *de novo* initiation. *Proc. Natl. Acad. Sci. USA* **101:** 4425-4430.

Cleveland, D.R., Zarbl, H. and Millward, S. 1986. Reovirus guanylyltransferase is L2 gene product lambda 2. *J. Virol.* **60:** 307–311.

Cong, P. and Shuman, S. 1993. Covalent catalysis in nucleotidyl transfer: A KTDG motif essential for enzyme-GMP complex formation by mRNA capping enzyme is conserved at the active sites of RNA and DNA ligases. *J. Biol. Chem.* **268:** 7256-7260.

Cong, P. and Shuman, S. 1995. Mutational analysis of mRNA capping enzyme identifies amino acids involved in GTP binding, enzyme-guanylate formation, and GMP transfer to RNA. *Mol. Cell. Biol.* **15:** 6222-6231.

Ding, Q., Xiao, L., Xiong, S., Jia, Y., Que, H., Guo, Y. and Liu, S. 2003. Unmatched masses in peptide mass fingerprints caused by cross-contamination: an updated statistical result. *Proteomics* **3:** 1313-1317.

Doherty, A.J., Hakansson, K., Ho, C.K., Shuman, S. and Wigley, D.B. 1997. Crystallization of the RNA guanylyltransferase of *Chlorella* virus PBCV-1. *Acta Crystallogr. D. Biol. Crystallogr.* **53:** 482-484.

Egloff, M.P., Benarroch, D., Selisko, B., Romette, J.L. and Canard, B. 2002. An RNA cap (nucleoside-2'-*O*)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.* **21:** 2757-2768.

Egloff, M.P., Decroly, E., Malet, H., Selisko, B., Benarroch, D., Ferron, F. and Canard, B. 2007. Structural and functional analysis of methylation and 5'-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. *J Mol Biol.* **372:** 723-36.

Fabrega, C., Shen, V., Shuman, S. and Lima, C.D. 2003. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **11:** 1549-1561.

Furuichi, Y. and Shatkin, A.J. 2000. Viral and cellular mRNA capping: Past and prospects. *Adv. Virus Res.* **55:** 135-184.

Geiss, B.J., Thompson, A.A., Andrews, A.J., Sons, R.L., Gari, H.H., Keenan, S.M. and Peersen, O.B. 2009. Analysis of flavivirus NS5 methyltransferase cap binding. *J. Mol. Biol.* **385:** 1643-1654.

Hakansson, K., Doherty, A.J., Shuman, S., and Wigley, D.B. 1997. X-ray crystallography reveals a large conformational change during guanylyl transfer by mRNA capping enzymes. *Cell* **89:** 545-553.

Hakansson, K. and Wigley, D.B. 1998. Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA-capping. *Proc. Natl. Acad. Sci. USA* **95:** 1505-1510.

Hausmann, S., Ho, C.K., Schwer, B. and Shuman, S. 2001. An essential function of *Saccharomyces cerevisiae* RNA triphosphatase Cet1 is to stabilize RNA guanylyltransferase Ceg1 against thermal inactivation. *J. Biol. Chem.* **276:** 36116-36124.

Heinz, F.X., Purcell, M.S., Gould, E.A., Howard, C.R., Houghton, M., Moormann, R.J.M., Rice, C.M. and Thiel, H.J. 2000. In *Virus Taxonomy* (Academic press, San Diego), pp 860-878.

Higman, M.A., Christen, L.A. and Niles, E.G. 1994. The mRNA (guanine-7-) methyltransferase domain of the vaccinia virus mRNA capping enzyme: Expression in *Escherichia coli* and structural and kinetic comparison to the intact capping enzyme. *J. Biol. Chem.* **269:** 14974-14981.

Ho, C.K., Van Etten, J.L. and Shuman, S. 1996. Expression and characterization of an RNA capping enzyme encoded by *Chlorella* virus PBCV-1. *J. Virol.* **70:** 6658-6664.

Ho, C.K., Schwer, B. and Shuman, S. 1998. Genetic, physical, and functional interactions between the triphosphatase and guanylyltransferase components of the yeast mRNA capping apparatus. *Mol. Cell. Biol.* **18:** 5189-5198.

Ho, C.K., Sriskanda, V., McCraken, S., Bentley, D., Schwer, B. and Shuman, S. 1998. The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxy-temrinal domain of RNA polymerase II. *J. Biol. Chem.* **273:** 9577-9585.

Ho, C.K., Lehman, K. and Shuman, S. 1999. An essential surface motif (WAQKW) of yeast RNA triphosphatase mediates formation of the mRNA capping enzyme complex with RNA guanylyltransferase. *Nucleic Acids Res.* **27:** 4671-4678.

Hsiao, J., Martinez-Costas, J., Benavente, J. and Vakharia, V.N. 2002. Cloning, expression, and characterization of avian reovirus guanylyltransferase. *Virology* **296:** 288–299.

Le Blois, H., French, T., Mertens, P.P.C., Burroughs, J.N. and Roy, P. 1992. The expressed VP4 protein of bluetongue virus binds GTP and is the candidate guanylyl transferase of the virus. *Virology* **189:** 757–761.

Lee, J.Y., Chang, C., Song, H.K., Moon, J., Yang, J., Kim, H.K., Kwon, S.T. and Suh, S.W. 2000. Crystal structure of NAD(+)-dependent DNA ligase: Modular architecture and functional implications. *EMBO J.* **19:** 1119-1129.

Lesburg, C.A., Cable, M.B., Ferrari, E., Hong, Z., Mannarino, A.F. and Weber, P.C. 1999. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat. Struct. Biol.* **6:** 937-943.

Liu, M., Mattion, N.M. and Estes, M.K. 1992. Rotavirus VP3 expressed in insect cells possesses guanylyltransferase activity. *Virology* **188:** 77–84.

Liuzzi, M., Mason, S.W., Cartier, M., Lawetz, C., McCollum, R.S., Dansereau, N., Bolger, G., Lapeyre, N., Gaudette, Y. et al. 2005. Inhibitors of respiratory syncytial virus replication target cotranscriptional mRNA guanylylation by viral RNA-dependent RNA polymerase. *J. Virol.* **79:** 13105-13115.

Luongo, C.L., Reinisch, K.M., Harrison, S.C. and Nibert, M.L. 2000. Identification of the guanylyltransferase region and active site in reovirus mRNA capping protein □2. *J. Biol. Chem.* **275:** 2804-2810.

Malet, H., Egloff, M.P., Selisko, B., Butcher, R.E., Wright, P.J., Roberts, M., Gruez, A., Sulzenbacher, G., Vonrhein, C., Bricogne, G., Mackenzie, J.M., Khromykh, A., Davidson, A.D. and Canard, B. 2007. Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. *J. Biol. Chem.* **282:** 10678-10689.

Mao, X. and Shuman, S. 1994. Intrinsic RNA (guanine-7) methyltransferase activity of the vaccinia virus capping enzyme D1 subunit is stimulated by the D12 subunit: Identification of amino acid residues in the D1 protein required for subunit association and methyl group transfer. *J. Biol. Chem.* **269:** 24472-24479.

Mao, X., Schwer, B. and Shuman, S. 1995. Yeast mRNA cap methyltransferase is a 50-kilodalton Nowak, T., Farber, P.M., Wengler, G. and Wengler, G. 1989. Analyses of the terminal sequences of West Nile virus structural proteins and of the in vitro translation of these proteins allow the proposal of a complete scheme of the proteolytic cleavages involved in their synthesis. *Virology* **169:** 365-376.

Mohd Jaafar, F., Attoui, H., Mertens, P.P.C., de Micco, P. and de Lamballerie, X. 2005. Structural organization of an encephalitic human isolate of Banna virus (genus Seadornavirus, family Reoviridae). J. Gen. Virol. **86:** 1147–1157.

Odell. M., Sriskanda. V., Shuman, S. and Nikolov, D.B. 2000. Crystal structure of eukaryotic DNA ligase illuminates the mechanism of nick-sensing and strand-joining. *Mol. Cell* **6:** 1183-1193.

Oh, J.W., Ito, T. and Lai, M.M.C. 1999. A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. *J. Virol.* **73:** 7694-7702.

Pena, L., Yanez, J., Revilla, Y., Vinuela, E. and Salas, M.L. 1993. African swine fever virus guanylyltransferase. *Virology* **193:** 319-328.

Qiu, T. and Luongo, C.L. 2003. Identification of two histidines necessary for reovirus mRNA guanylyltransferase activity. *Virology* **316:** 313–324.

Schnierle, B.S., Gershon, P.D. and Moss, B. 1992. Cap-specific mRNA (nucleoside-*O*2'-) methyltransferase and poly(A) polymerase stimulatory activities of vaccinia virus are mediated by a single protein. *Proc. Natl. Acad. Sci. USA* **89:** 2897-2901.

Schwer, B., Hausmann, S., Schneider, S. and Shuman, S. 2006. Poxvirus cap methyltransferase: Bypass of the requirement for the stimulatory subunit by mutations in the catalytic subunit and evidence for intersubunit allostery. *J. Biol. Chem.* **281:** 18953-18960.

Shatkin, A.J. 1976. Capping of eukaryotic mRNAs. *Cell* **9:** 645-653.

Shibagaki, Y., Itoh, N., Yamada, Y., Hagata, S. and Mizumoto, K. 1992. mRNA capping enzyme: Isolation and characterization of the gene encoding mRNA guanylyltransferase subunit from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **267:** 9521-9528.

Shuman, S. and Hurwitz, J. 1981. Mechanism of mRNA capping by vaccinia virus guanylyltransferase: Characterization of an enzyme-guanylate intermediate. *Proc. Natl. Acad. Sci. USA* **78:** 187-191.

Shuman, S. 1982. RNA capping by HeLa cell RNA guanylyltransferase. Characterization of a covalent protein-guanylate intermediate. *J. Biol. Chem.* **257:** 7237-7245.

Shuman, S. 2000. Structure, mechanism, and evolution of the mRNA-capping apparatus. *Prog. Nucleic Acids Res. Mol. Biol.* **66:** 1-40.

Supyani, S., Hillman, B.I. and Suzuki, N. 2007. Baculovirus expression of the 11 mycoreovirus-1 genome segments and identification of the guanylyltransferase-encoding segment. *J. Gen. Virol.* **88:** 342-350.

Suzuki, N., Kusano, T., Matsuura, Y. and Omura, T. 1996. Novel NTP binding property of rice dwarf phytoreovirus minor core protein P5. *Virology* **219:** 471–474.

Vasiljeva, I., Merits, A., Auvinen, P. and Kääriäinen, L. 2000. Identification of a novel function of the Alphavirus capping apparatus. RNA 5'-triphosphatase activity of Nsp2. *J. Biol. Chem.* **275:** 17281-17287.

Venkatesan, S. and Moss, B. 1982. Eukaryotic mRNA capping enzyme-guanylate covalent intermediate. *Proc. Natl. Acad. Sci. USA* **79: 340**-344.

Wengler, G., Czaya, G., Farber, P.M. and Hegemann, J.H. 1991. *In vitro* synthesis of West Nile virus proteins indicates that the amino-terminal segment of the NS3 protein contains the active centre of the protease which cleaves the viral polyprotein after multiple basic amino acids. *J. Gen. Virol.* **72:** 851-858.

Yap, T.L., Xu, T., Chen, Y.L., Malet, H., Egloff, M.P., Canard, B., Vasudevan, S.G. and Lescar, J. 2007. Crystal structure of the Dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-Angstrom resolution. *J Virol.* **81:** 4753-4765.

Yon, C., Teramoto, T., Mueller, N., Phelan, J., Ganesh, V.K., Murthy, K.H.M. and Padmanabhan, R. 2005. Modulation of the nucleoside triphosphate/RNA helicase and 5'-RNA triphosphatase activities of Dengue virus type 2 nonstructural protein 3 (NS3) by interaction with NS5, the RNA-dependent RNA polymerase. *J. Biol. Chem.* **280:** 27412-27419.

Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., Guo, Y., Bernard, K.A., Shi, P.Y. and Li, H. 2007. Structure and function of flavivirus NS5 methyltransferase. *J Virol.* **1:** 3891-903.

FIGURE LEGENDS

Figure 1. Formation of a NS5-GMP covalent intermediate. (**A**) An aliquot (1 μg) of the purified full-length NS5 ($NS5^{FL}$) protein from WNV was analyzed by electrophoresis through a 12.5% polyacrylamide gel containing 0.1% SDS and visualized by staining with Coomassie Blue dye. The positions and sizes (in kDa) of the molecular weight markers (L) are indicated on the left. (**B**) Formation of an enzyme-GMP covalent intermediate. The $NS5^{FL}$ enzyme (3 μM) was incubated for 60 min at 37 °C with 1 mM [α-^{32}P]GTP in a buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, and 5 mM $MgCl_2$. Potassium pyrophosphate (PPi) (1 mM) was also added to the reaction where indicated. A control reaction was also performed with 1 μg of the RNA GTase from *Paramecium bursaria Chlorella* virus-1 (A103R). The reactions were stopped by the addition of EDTA to 10 mM and SDS to 1%, and analyzed by electrophoresis through a 12.5% polyacrylamide gel containing 0.1% SDS. An autoradiogram of the gel is shown. The positions and sizes (in kDa) of the molecular weight markers are indicated on the left. (**C**) The extent of $NS5^{FL}$-GMP complex formation is plotted as a function of GTP concentration. A double-reciprocal plot of the data is shown in the inset. (**D**) Divalent cation specificity. Formation of the $NS5^{FL}$-GMP complex was performed in the presence of 5 mM divalent cation as specified. Mg, Mn, Co, Ca, and Zn were added as chloride salts; Cu and Ni were added as sulfates. (**E**) The $NS5^{FL}$-GMP labelled complex (EpG) was isolated by gel filtration and treated with either 0.5 M NaOH or 0.5 M HCL at 70 °C for 10 min or with 3.8 M hydroxylamine (NH_2OH) in 0.2 M sodium acetate (pH 4.8) at 37 °C for 20 min. The sample products were analyzed by thin layer chromatography on a polyethyleneimine-cellulose plate and developed with 0.75 M KH_2PO_4. An autoradiogram of the plate is shown. The positions of the chromatographic origin (ori), the nucleotidyl-enzyme complex, and unlabelled GMP marker, visualized under ultraviolet light, are indicated.

Figure 2. The NS5 N-terminal methyltransferase domain is active in the formation of a protein-GMP covalent complex. (**A**) The N-terminal MTase domain of the NS5 proteins (3 μM) of Dengue virus (DEN; strain 16681, aa 1-267), Yellow fever virus (YFV; strain 17D, aa 1-268) and West Nile virus (WNV; strain NY99, aa 1-268) were incubated for 60 min at 37 °C with 1 mM [α-^{32}P]GTP in a buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM

DTT, and 5 mM $MgCl_2$. The reactions were stopped by the addition of EDTA to 10 mM and SDS to 1%, and analyzed by electrophoresis through a 12.5% polyacrylamide gel containing 0.1% SDS. An autoradiogram of the gel is shown. The positions and sizes (in kDa) of the molecular weight markers are indicated on the left. (**B**) Mutational analysis of the N-terminal domain of the Dengue virus NS5 protein. Lys29 was substituted for alanine in the DEN (aa 1-267) construct, and the mutant polypeptide was expressed and purified in parallel with the corresponding wild-type enzyme. Aliquots (3 μM) of the purified fractions were analyzed by electrophoresis through a 12.5% polyacrylamide gel containing 0.1% SDS and visualized by staining with Coomassie Blue dye. The positions and sizes (in kDa) of the molecular weight markers are indicated on the left. (**C**) Effect of the K29A mutation on the formation of the covalent enzyme-GMP complex formation. The reactions were performed with 3 μM of the mutant and wild-type proteins, and analyzed as described in panel A. The positions and sizes (in kDa) of the molecular weight markers are indicated on the left of the autoradiogram.

Figure 3. Transfer of GMP to an acceptor RNA. 5'-diphosphate acceptor RNAs of 81 nucleotides (1 μg) initiating either with a guanosine (*left*) or adenosine (*right*) were incubated with 1 mM [α-^{32}P]GTP in a reaction mixture containing 50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 5 mM DTT, and 0.3 μM of purified WNV $NS5^{FL}$ protein. A control reaction was also performed with 1 μg of the RNA GTase from *Paramecium bursaria Chlorella* virus-1 (A103R). The RNA samples were extracted with phenol/chloroform, recovered by ethanol precipitation, and analyzed on a denaturing 10% polyacrylamide gel. An autoradiogram of the polyacrylamide gel is shown.

Figure 4. Stimulation of the GTase activity by the NS3 protein. (**A**) The WNV $NS5^{FL}$ protein (2 μM) was incubated in the absence or presence of purified WNV $NS3^{168-618}$ protein (10 μM) for 60 min at 37 °C with 1 mM [α-^{32}P]GTP in a buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, and 5 mM $MgCl_2$. As a control, the purified $NS3^{168-618}$ protein was also assayed in the absence of $NS5^{FL}$. The reactions were stopped by the addition of EDTA to 10 mM and SDS to 1%, and analyzed by electrophoresis through a 12.5% polyacrylamide gel containing 0.1% SDS. An autoradiogram of the gel is shown.

The positions and sizes (in kDa) of the molecular weight markers are indicated on the left. (**B**) Increasing concentrations of $NS3^{168-618}$ were used in the GTase assay, and the formation of the $NS5^{FL}$-GMP complex was monitored through electrophoresis and autoradiography. The intensities of the $NS5^{FL}$-GMP complexes were normalized to the signal obtained in the absence of $NS3^{168-618}$ protein.

Figure 5. Synthesis of the RNA cap 1 structure by the NS3 and NS5 proteins. (**A**) The RNA capping reaction was initiated by performing an RTase assay in a buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM DTT, 20 pmol of triphosphate-terminated RNA, and 2 μM of purified WNV $NS3^{168-618}$ protein. The reaction was incubated at 37 °C for 60 min. The RNA was extracted with phenol/chloroform, recovered by ethanol precipitation, adjusted to 5 mM $MgCl_2$, and the GTase reaction was performed in the presence of 3 μM of purified WNV $NS5^{FL}$ protein and 1 mM [α-^{32}P]GTP. The sample was incubated at 37 °C for 60 min. The RNA sample was extracted with phenol/chloroform, recovered by ethanol precipitation, and analyzed on a denaturing 10% polyacrylamide gel (lane 3). An autoradiogram of the polyacrylamide gel is shown. A control reaction was also performed with the RTase (A449R) and GTase (A103R) of *Paramecium bursaria Chlorella* virus (PBCV-1) (lane 4). Control reactions were also performed in the presence of WNV $NS3^{168-618}$ (lane 1) or $NS5^{FL}$ proteins (lane 2). The position of the radiolabelled phosphate, originating from the [α-^{32}P]GTP substrate, is denoted by an asterisk. (**B**) Aliquots of the RNA capping reactions were adjusted to 50 mM NaOAc, pH 5.2, and digested with nuclease P1 (5 μg) for 60 min at 37 °C. The reactions were then adjusted to 50 mM Tris-HCl, pH 8.0, and digested with alkaline phosphatase (1 unit) for 30 min at 37 °C. The reaction products were analyzed by thin layer chromatography on a polyethyleneimine-cellulose plate developed with 0.5 M LiCl/ 1 M formic acid. An autoradiogram of the plate is shown. Lane 1, Control reaction performed in the absence of protein; Lane 2, Control reaction performed with the *Chlorella* virus GTase (A103R); Lane 3, Reaction performed with the WNV $NS5^{FL}$ protein; Lane 4, The RNA from the standard GTase reaction (performed with A103R) was extracted with phenol/chloroform and recovered by ethanol precipitation. The purified RNA was then subjected to a standard MTase reaction with the *S. cerevisiae* MTase (Abd1) in the presence of 1 mM *S*-adenosyl-methionine; Lane 5, The

RNA from the Gtase reaction performed with the WNV $NS5^{FL}$ protein was extracted with phenol/chloroform and recovered by ethanol precipitation. The purified RNA was then subjected to a methylation reaction with the $NS5^{FL}$ in the presence of 1 mM *S*-adenosyl-methionine; The positions of the chromatographic origin (ori), GpppA, and ^{m7}GpppA are indicated.

Figure 6. RNA cap synthesis by the flavivirus NS5 and NS3 proteins. The flavivirus NS5 protein harbors the catalytic center of the RNA-dependent RNA polymerase (RdRp) and initiates the synthesis of the triphosphorylated RNA transcript. The C-terminal region of NS3 is responsible for the helicase activity which is necessary to allow progression of the polymerase along the template strand. The nascent RNA then interacts with the NS3 RTase leading to the removal of the γ-phosphate from the RNA. The modified RNA would then be in position to interact with the NS5 MTase-GTase domain. The interaction between NS3 and NS5 increases the GTase activity of the NS5 protein, generating a covalent intermediate (EpG*) in which the GMP moiety is linked to the Lys-29 residue of NS5. Subsequently, the NS5 protein can perform the transfer of GMP to the acceptor RNA resulting in the synthesis of a capped RNA transcript. The capped RNA would then be methylated to generate the RNA cap structure.

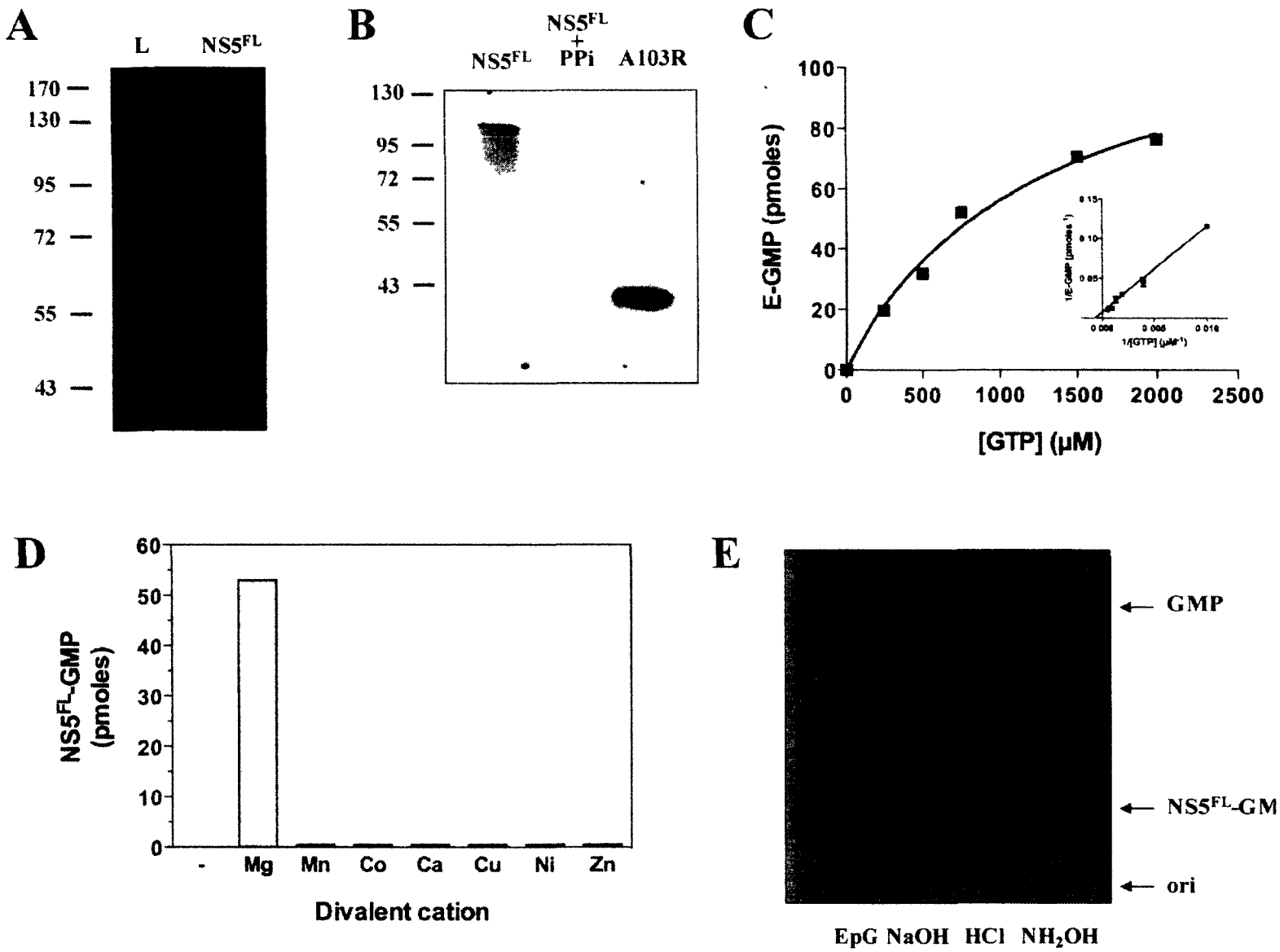


Figure 1, Issur, Geiss et al., 2009

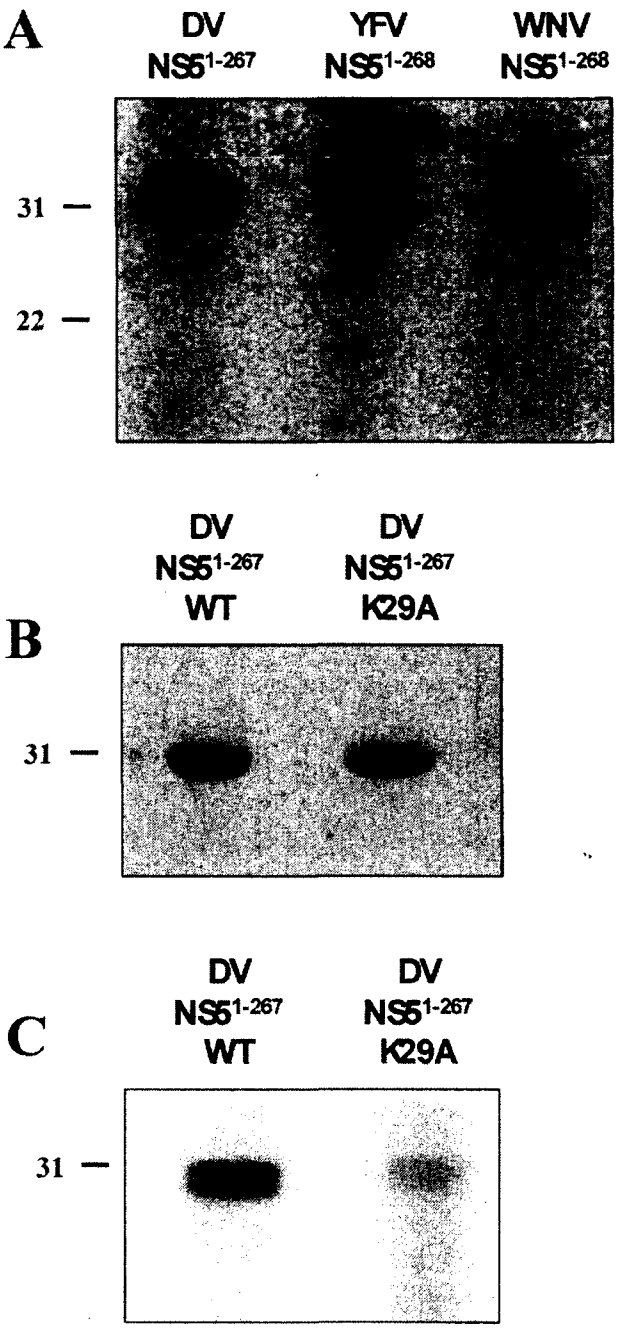


Figure 2, Issur, Geiss et al., 2009

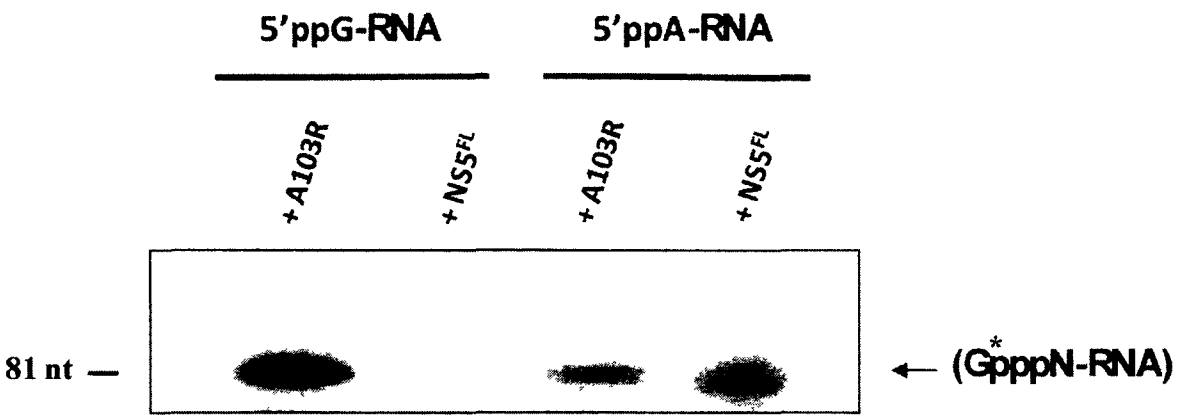


Figure 3, Issur, Geiss et al., 2009

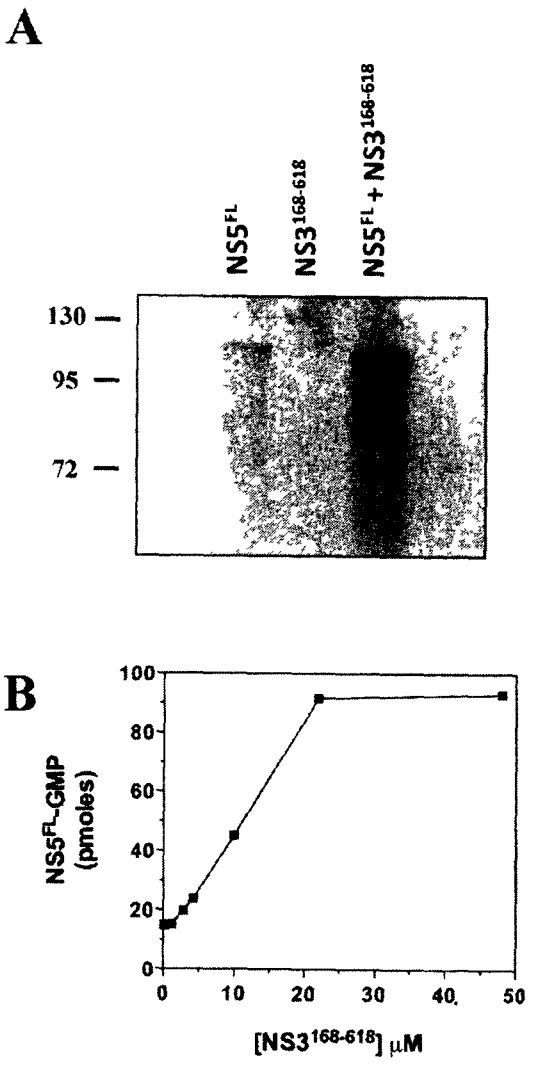


Figure 4, Issur, Geiss et al., 2009

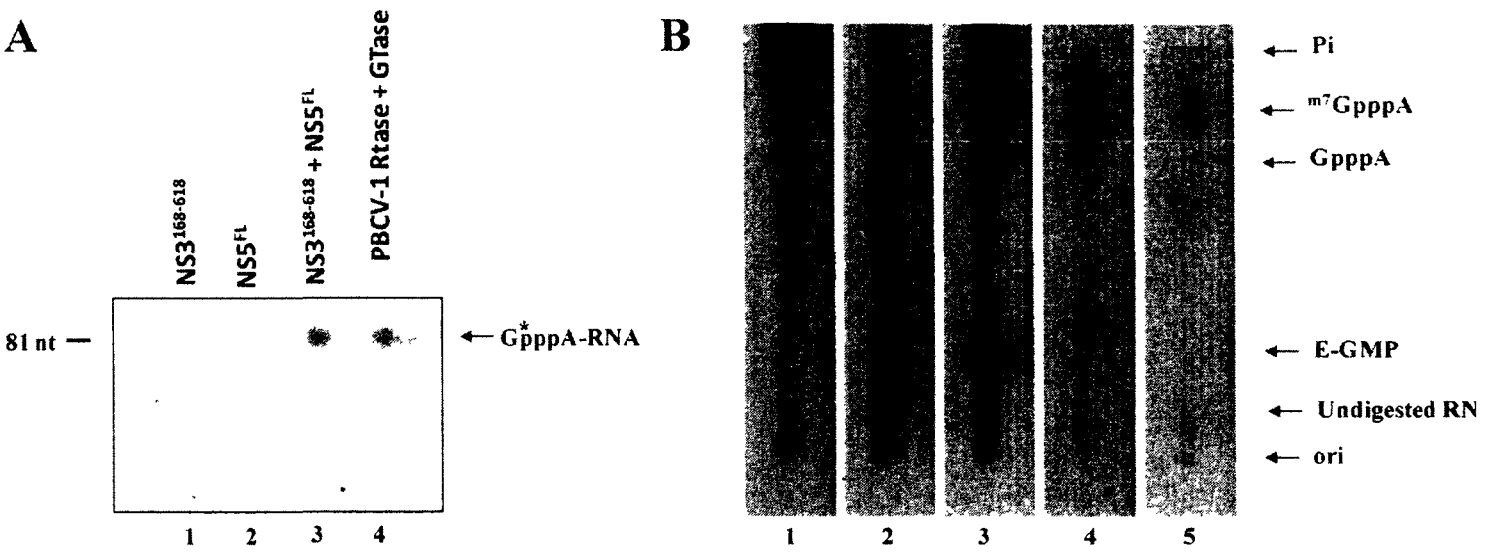


Figure 5, Issur, Geiss et al., 2009

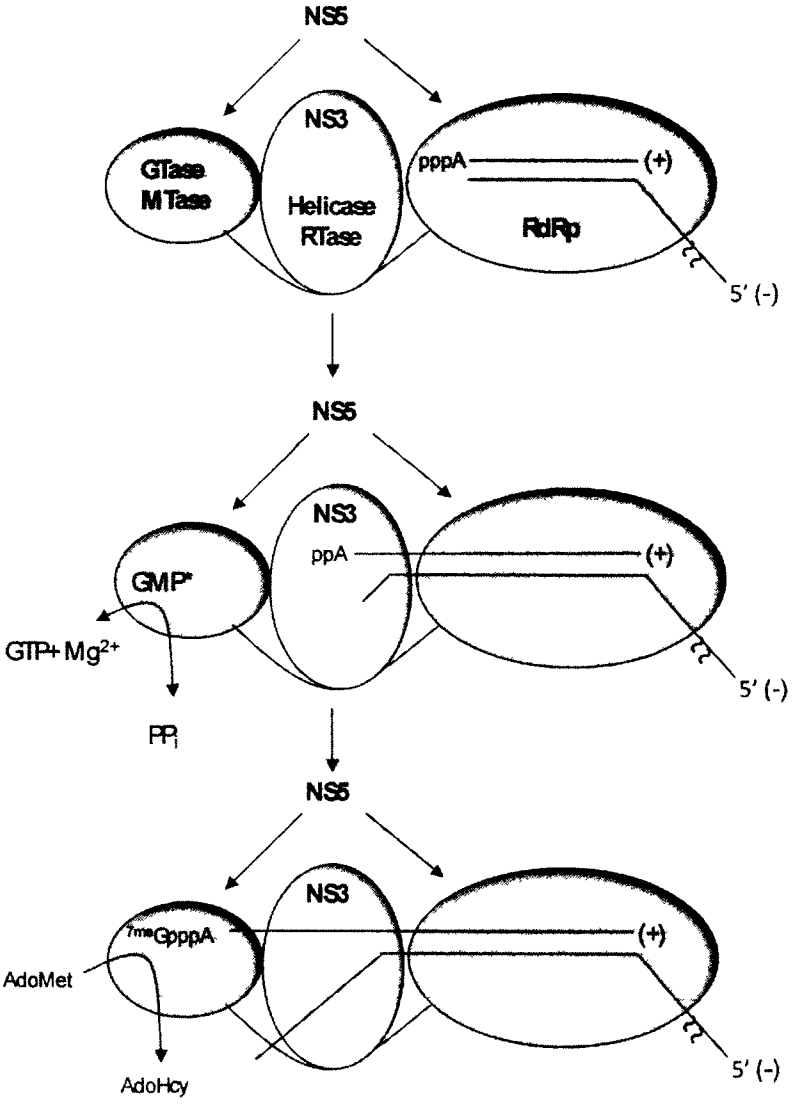


Figure 6, Issur, Geiss et al., 2009

DISCUSSION

Developing inhibitors against fungal and viral RNA triphosphatases

Fungal and viral RNA triphosphatases are one of the most potent anti-microbial targets, because of the high degree of divergence between these enzymes and the human RTase. As mentioned in the introduction, fungal and viral RTases differ from their mammalian counterparts in terms of their mechanism of action, and also in terms of their tertiary structure (Introduction, figures 5 and 6).

In addition to showing that the tunnel architecture of the microbial TTM RTase is far more flexible than initially expected, our studies pinpoint several interesting candidates as potential inhibitors of the TTM RTase. Some of these candidate molecules are N1-Me GTP, XTP, 6-Chloro purine riboside triphosphate, 6-thio purine riboside triphosphate and 8-Iodo GTP. All of these 5 molecules are characterized by a good binding affinity to the *S. cerevisiae*'s RTase (Cet1), while being poorly hydrolyzed. Their K_i range from 0.4 to 1.8 μM. By comparison, the K_i of these inhibitors have also been evaluated against a viral RTase, the West Nile virus RTase (NS3). Table 1 provides a brief summary of the major differences between the inhibition of Cet1 and NS3. The informative finding here is that while the K_i for Cet1 is in the low micromolar range, the K_i for NS3 is much higher. This implies that these molecules are better inhibitors of the fungal RTase than the viral one. In order to complete this analysis on the potency of these purine analogues as starting points for the rational design of drugs, the K_i of these inhibitors for the mammalian RTase need to be evaluated. However, its value is expected to be much higher than the K_i obtained for the NS3 protein, because the metal ion-independent mammalian RTase, in contrast to the fungal RTase of the TTM family, cannot hydrolyze nucleotide triphosphates (NTPs) (Ghosh and Lima 2010). The substrate of the mammalian RTase is exclusively 5' triphosphorylated RNA molecules.

Discussion, Table 1: Comparison of the K_i of 4 potent inhibitors for the RTase of *S. cerevisiae* (Cet1) against the K_i of the West Nile virus RTase (NS3)

Inhibitor	K_i (mM)	
	Cet1[1]	NS3[2]
N1-Me GTP	0.0015	2.6
XTP	0.0018	>2.6
6-Cl purine RTP[3]	0.0004	NA
6-thio RTP[3]	0.0015	0.8
8-Iodo GTP	0.0015	>2.6

1: Extracted from Issur M, et al, NAR, 2009 (Issur, Despins et al. 2009)
2: Extracted from Despins S, et al, NAR 2010 (Despins, Issur et al. 2010)
3: RTP stands for Riboside triphosphate

The RTases of fungi, trypanosomes and of several viruses are the founding members of the TTM family of phosphohydrolases. The Cet1-like TTM fold is conserved in the mimivirus RTase, as well as in several proteins of bacterial and archael origin (Jain and Shuman 2008). With the exception of eukaryal RTases of fungi and trypanosomes, this TTM fold has not been discovered yet in any other higher order eukaryotic phosphohydrolases. Therefore, I believe that the design of drugs specific to the TTM RTase is an achievable goal. We plan to perform *in vivo* studies in a Δ*xrn1 S. cerevisiae* strain with several of the inhibitors. Xrn1 is a 5'-3' exonuclease which degrades uncapped RNA. If the RNA capping machinery is indeed the major target, the Δ*xrn1* strain will enable us to visualize the formation of cap deficient RNAs in the presence of these inhibitors. It should be noted that in order to improve the potency of the inhibitors, the addition of two or more modifications on the purine base is being considered. For instance, the effects of the presence of an N1-methyl group in addition to a 6-Chloro modification on the K_i for the fungal RTase are supplementary information which would be important, for the generation of more potent and specific compounds. The impact of these results does not apply solely to the fungal

RTase. The implications are much broader, since the TTM RTase is also present in several very pathogenic organisms, including the malaria parasite, *Plasmodium falciparum* (Ho and Shuman 2001). Any major advances in the development of inhibitors against the model TTM RTase of *S. cerevisiae* are also applicable to the malaria parasite RTase.

The TTM-RNA triphosphatase mechanism

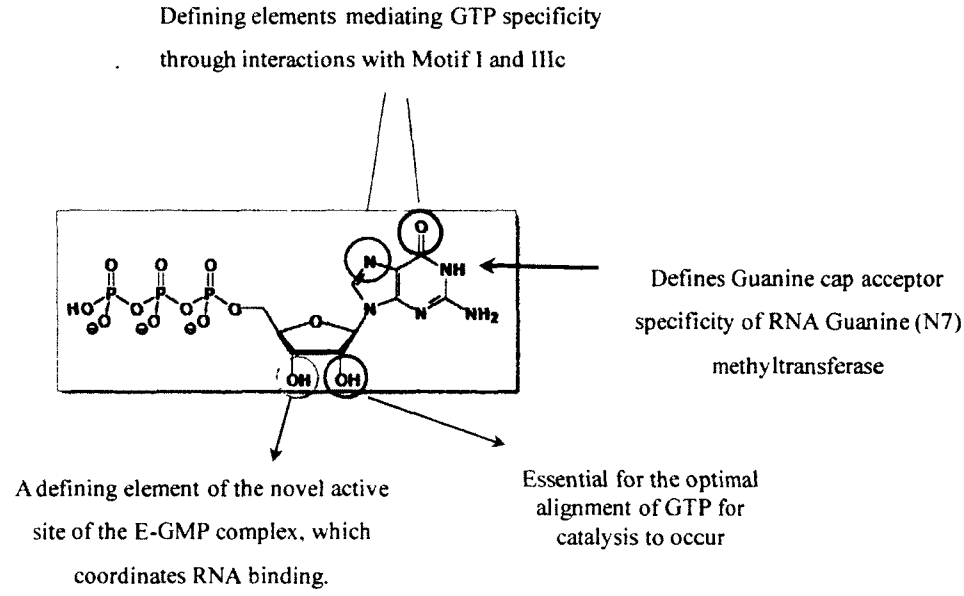
In all crystal structures of RTases of the TTM clade, the active sites of the proteins are devoid of substrates. The reaction products and a metal ion have been found within the tunnel architecture of the protein, while all attempts to co-crystallize a nucleotide or RNA substrate with the protein have failed. This is not surprising since our lab has previously shown that the protein is thermodynamically unstable when bound to a nucleotide or RNA substrate (Bisaillon and Bougie 2003). The use of molecular docking has provided the first insight into how a ligand is bound within the tunnel architecture of the enzyme.

From this model, several other aspects of the RTase reaction can now be investigated. From our model which indicates all the essential residues contacting the triphosphate tail, and by re-analyzing previous mutational data, we can now correctly infer the catalytic mechanism for phosphohydrolysis. Based on our model, *Lanthier et al*, (unpublished data) are presently investigating the residue(s) involved in the nucleophilic attack on the phosphate tail. They have thus been able to conclusively point out that a water molecule coordinated by a conserved glutamate residue (Glu433 in Cet1) is responsible for the nucleophilic attack on the phosphate tail. In our model, this glutamate residue is coordinating both the β and γ phosphates of the nucleotide. Several previous studies have shown that the TTM RTase uses a two-metal ion catalytic mechanism, without however any definite proof of two divalent metal ions binding to the protein. One metal ion has been found in the crystal structure which is not involved in catalysis but in the stabilization of the reaction products. We think that the unidentified metal ion is the catalytic metal ion which may be coordinated by both the glutamate residue and the attacking water molecule. This is currently being investigated in the lab. In addition our data show that not only the triphosphate tail, but the whole nucleotide enters the tunnel of the protein. It would be of great fundamental interest to investigate into how product release occurs. It would be very

interesting to know whether the "entrance" and the "exit" of the tunnel have any functional significance. And more importantly, what happens when the true ligand, RNA, is present?

The cap donor specificity of the RNA guanylyltransferase reaction

In our study with regards to the use of nucleotide analogues by the PBCV-1 GTase, we have demonstrated that the active site of the conventional GTase can accommodate ligands possessing several different types of modifications. In addition, the PBCV-1 GTase can transfer several of these modified analogues onto RNAs to form novel RNA cap structures at their 5' ends. Some of these cap analogues could be methylated at the N7 position by either the yeast or human RNA (guanine-N7) methyltransferase. In addition to generating several differentially capped RNAs, this procedure allowed us to gain a better insight into the guanine specificity of both the GTase and MTase reactions. This is summarized in figure 1.



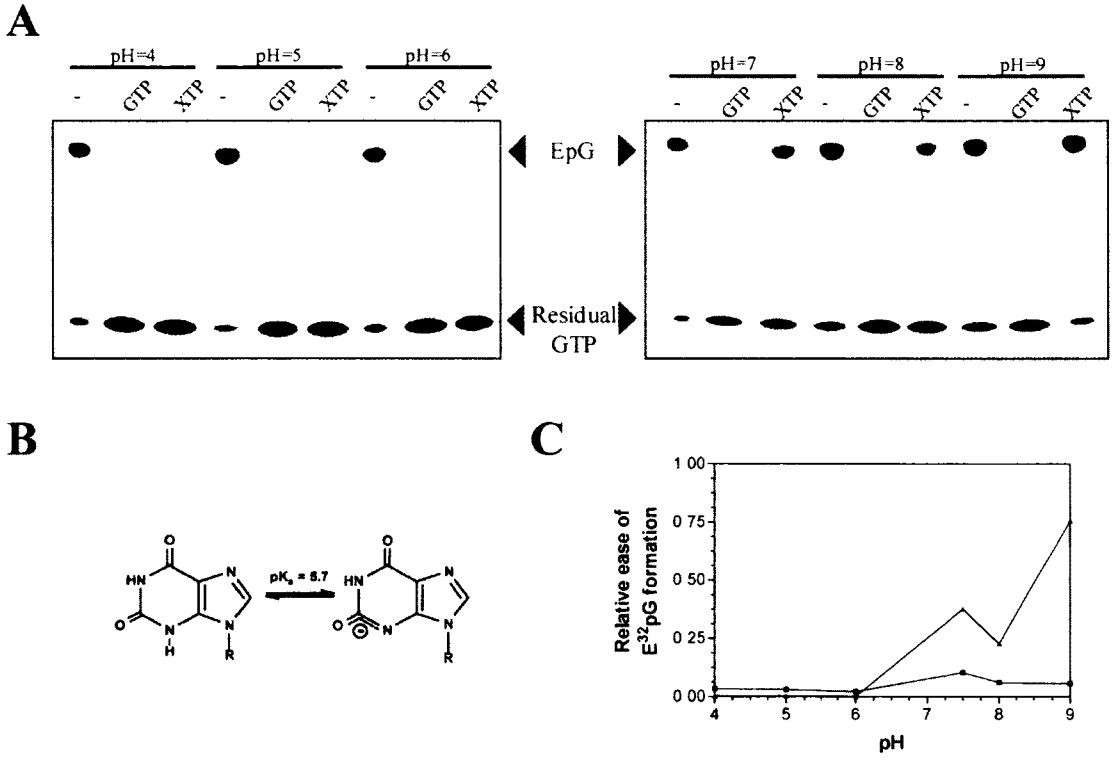
Discussion, figure 1: A brief summary of the major conclusions of Chapter II on the guanine specificity of the PBCV-1 GTase and the *S. cerevisiae* MTase.

Encircled in red are positions which are most important for the first step of the GTase reaction. In green is the 3' position whose importance resides mainly in the second step of the reaction. The red arrow pointing towards the N1 position indicates the position which was most detrimental to N7-methylation by the S. cerevisiae's MTase.

Having deciphered the major contributions of the base and ribose moieties for ligand binding, the next question we are asking ourselves is how to broaden the range of nucleotide analogues which can be used by the PBCV-1 GTase. This can be achieved by generating enzymes harbouring different mutations within the GTP binding. However, enzymes of the nucleotidyltransferase superfamily, of which the PBCV-1 GTase is a member, are known to be very sensitive to any mutation within the GTP binding site (Wang, Deng et al. 1997). For instance, alanine mutation of Lys188, which interacts with O6 and N7 of GTP in the PBCV-1 GTase, leads to complete loss of activity of the enzyme, even if it does not contact the phosphate tail of the GTP molecule. Therefore, we propose to look into the effects of mutations of conserved non-essential residues around the known essential amino acids interacting with GTP, on the GTP specificity of the capping enzyme. For instance, Lys188, which interacts with O6 of guanine, is embedded between two Motif I residues (Ser80 and Glu81) located just besides the catalytic lysine residue (Lys82). We intend to look into importance of these two conserved, non-essential, Motif I residues on capping chemistry and on the GTP specificity of the reaction. Our ultimate goal is to generate mutant enzymes which could carry out the GTase reaction with purine analogues reminiscent of ATP, in order to create a wider range of RNA cap structures.

We have also shown that the use of several analogues is dependent on the pH of the reaction. The case of 2' modified GTP analogues is presented in Chapter II. We also have evidence that the use of Xanthosine triphosphate, a C2 modified GTP analogue can also be influenced by pH (Fig. 3). XTP harbours an exocyclic keto group at the C2 position, instead of the exocyclic amino group in GTP. The poor inhibition of the first step GTase reaction by XTP was intriguing since our data clearly showed that the C2 amino group was unimportant for RNA capping. ITP which lacks this exocyclic amino group is as efficient as GTP for RNA capping. It has been reported that the keto group in free xanthine is in an ionised enol form at pH greater than 5.7 (Kulikowska, Kierdaszuk et al. 2004). We believe that this ionised enol form predominates in XTP at the reaction's conditions, thus rendering it inert as substrate for RNA capping. To test this hypothesis, competition assays with radiolabelled [α-^{32}P]GTP were recently performed at various pH in the presence of XTP or GTP (Fig. 3). Decreasing the pH of the reaction rescued inhibition by XTP, while at higher pH (pH > 6) XTP could no longer inhibit the reaction. In the light of these data, we think

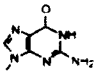
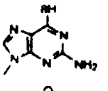
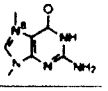
that the poor inhibition by XTP is due to the predominance of the ionised enol form of the molecule. The negative charge on the molecule may be affecting ligand binding. Taken together, these results indicate that the range of possible RNA cap analogues that can be generated by RNA capping enzymes can be widened by varying the pH of the reactions.



Discussion, figure 2: pH dependency of the inhibition by XTP

(A) The PBCV-1 GTase was incubated with [α-^{32}P]GTP and magnesium ions in various buffer solutions with pH ranging from 4 to 9. For each pH, a reaction with no inhibitor (-), one with GTP (2.0 mM), and one with XTP (2.0 mM) were carried out. At a given pH, if GTP or the nucleotide analogue was inhibiting, a loss of the EpG signal was observed relative to the reaction with no inhibitor (-). (B) Equilibrium between the keto- and ionised enol form of Xanthosine (C) Quantification of the ease of formation of the radiolabelled EpG complex in the presence of unlabelled GTP (■) and unlabelled XTP (▲). High ease of formation implies poor inhibition; and vice-versa.

With the same intent to generate more RNA cap analogues, unconventional RNA guanylyltransferases can also be used. *Geiss et al*, have recently shown that the NS5 protein, the newly identified RNA guanylyltransferase from the *flavivirus* genus can bind several unnatural substrates in its N-terminal domain (Table 2) (Geiss, Thompson et al. 2009). However, no report on whether NS5 can form a covalent intermediate with any of the GTP analogues tested nor whether any of these substrates can be transferred on an acceptor RNA has been made yet. The use of unconventional RNA capping enzymes can also be broadened to generate methylated nucleotide analogues. Some viral GTases are known to use N7-methyl GTP as their substrate instead of GTP (Bisaillon and Lemay 1997). Generating N7-methylated nucleotide analogues to be used as substrates by unconventional GTases could also be an interesting alternative. This will reduce the number of steps required to produce methylated (at the N7 position) cap RNAs.

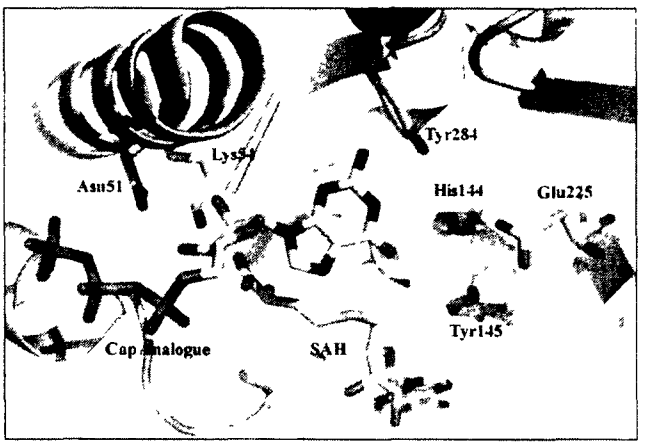
GTP analog	K_i	Standard error	Fold affinity reduction	Purine ring structure
GTP	40 nM	10 nM	0	
GDP	100 nM	30 nM	3	
GMP	950 nM	70 nM	29	
Guanosine	260 μM	8 μM	6400	
dGTP	610 nM	70 nM	15	
ddGTP	770 nM	70 nM	19	
ATP	350 μM	39 μM	8750	
ITP	120 nM	30 nM	3	
XTP	480 nM	110 nM	12	
6-Methylthio GTP	230 nM	90 nM	6	
m^{7}GpppG	100 nM	20 nM	2.6	
m^{7}GpppA	170 nM	20 nM	4.1	

Discussion, Table 2: Binding of purine nucleotides and dinucleotides to the N-terminal domain of the Dengue virus NS5 protein *(Extracted from Geiss et al, JMB, 2009)*

Binding studies with the NS5 protein and GTP-bodipy in the presence of various nucleotide analogues was performed. The $NS5^{Den}$ (1-267) protein was incubated with GTP-Bodipy γ-phosphate-labelled fluorescent analogue (10 nM) and an unlabelled nucleotide analogue. Fluorescent polarization assays were made to determine K_i.

The methyl acceptor specificities of the RNA (guanine-N7) methyltransferase reaction

The final step in the formation of the cap structure is the addition of a methyl group at the N7 position by an RNA (guanine-N7) methyltransferase. We tested the ability of each of the cap analogs to be methylated at the N7 position by *S. cerevisiae*'s MTase. We used the available crystal structure of the *E. cuniculi*'s MTase to analyze our results.



Discussion, figure 3: The RNA cap binding site of *E. cuniculi* RNA (guanine-N7) methyltransferase

Magnification of the cap binding site of the E. cuniculi RNA (guanine-N7) methyltransferase (PDB 1RI5). Residues shown are those interacting with GTP, which is purported to be mimicking a cap analogue.

Briefly, all but one of the modified cap structures could accommodate a methyl group at the N7 position. An RNA capped with N1-Me GTP could not be methylated at the N7 position. By analyzing the crystal structure of the MTase of *E. cuniculi* bound to a cap analogue and S-adenosyl homocysteine (SAH), we deduced that hydrogen bonding between the N1 hydrogen and an essential glutamate residue (Glu225 in figure 3) within the active site may be crucial for cap binding and optimal alignment of the capped residue for nucleophilic attack on the SAM co-factor. We also inferred that this interaction may mediate guanine specificity, since the N1 position of ATP is deprotonated and exposes a lone pair of electrons to the solvent. Finally, the reason why RNA capped with 3'O-Me GTP was a relatively poor methyl acceptor is likely due to steric hindrance, since in the active site of the *E. cuniculi* MTase, several residues are seen to be interacting with the ribose moiety of the cap guanosine (Supplementary table 1 in Chapter II and Fig.3 of the discussion).

Impact of the novel cap analogues on the process of translation

In an attempt to investigate the biological impact of these novel RNA cap analogues, we synthesized *firefly* luciferase RNAs harbouring these modified caps at their 5' ends and a 60 nt long Poly(A) tail at their 3' ends. These were transfected into HEK293 cells and evaluated for their translation efficiency. We were very surprised to detect that some cap analogues could efficiently support translation in spite of the absence of the N7-methyl group. These were the O6-Me guanosine and the 3'O-Me guanosine capped luciferase RNA (Chapter II). As mentioned previously, all previous studies on the effect of RNA cap analogues on the process of translation have been conducted either with GpppG, N7-methyl or N7-modified GpppG cap analogues. The impact of alternative modifications on the essentiality of the N7-methyl group has been overlooked.

In order to better understand the *in cellulo* effects, the binding of each cap analogue to the murine eIF4E was evaluated *in vitro*. The binding affinity of both 3'O-Me guanosine and O6-Me guanosine cap analogues were higher than the GpppG cap dinucleotide and their binding to eIF4E was sufficient to explain the higher translation profile observed (Chapter II). Again, with the same intent to better understand the binding of N7-methyl deficient cap analogues to eIF4E, I am presently doing some *in silico* binding experiments. Since the eIF4E protein has been co-crystallised with ^{m7}GDP, I have performed docking experiments with 3'O-Me GDP and O6-Me GDP to the eIF4E protein (PDB 1L8B). Our i*n silico* results are consistent with good binding affinity to eIF4E (Table 3). The estimated free energy of binding for the docking of ^{m7}GDP (positive control), 3'O-Me GDP and O6-Me GDP are very similar (Table 3).

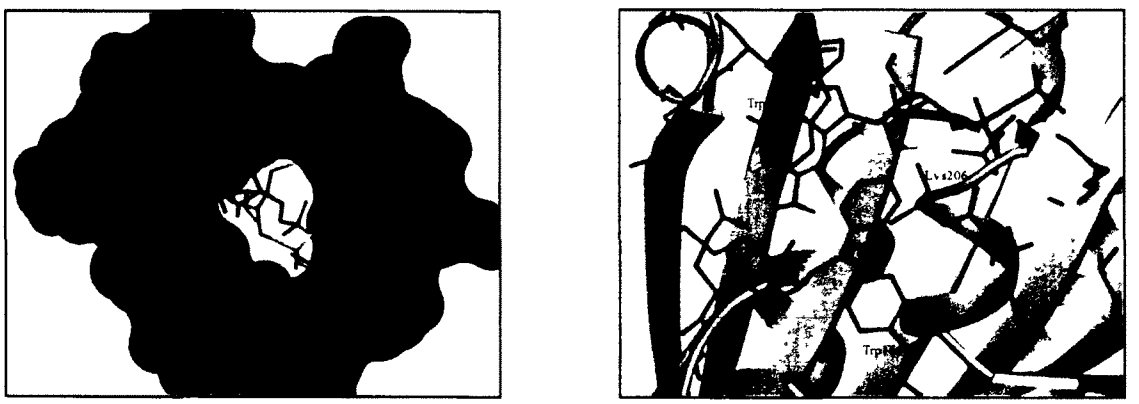
Discussion, Table 3: *In silico* parameters for the best selected docking result*

Docked ligand	Estimated Free Energy of Binding	Estimated Inhibition Constant, K_i	vdW+Hbond+desolv Energy	Electrostatic Energy	Total Intermolecu Energy
^{m7}GDP	-8 96 kcal/mol	270 76 nM	-6 40 kcal/mol	-4 82 kcal/mol	-11 23 kcal/mol
3'O-Me GDP	-8 40 kcal/mol	693 11 nM	-5 77 kcal/mol	-5 22 kcal/mol	-10 99 kcal/mol
O6-Me GDP	-10 41 kcal/mol	23 34 nM	-6 86 kcal/mol	-5 03 kcal/mol	-11 89 kcal/mol

* Docking was performed on the murine eIf4E (PDB 1L8B) crystal structure, from which the bound ^{m7}GDP ligand was removed. ^{m7}GDP was docked backed into the structure as a positive control.

The *in silico* positioning of 3'O-Me GDP is very similarly to the *in crystallo* position of the ^{m7}GDP ligand (Fig.4A). The major difference between the proposed model and the crystallographic structure is an 180^0 flip of the guanine base in the docking model relative to the crystal structure. The ribose moiety has rotated around the glycosidic bond to be significantly differently positioned relative to the ribose of ^{m7}GDP. A direct consequence of this rotation is the possibility of hydrophobic interactions between the 3' O-Me group and the Trp102 (Fig.4B). From these *in silico* data, we hypothesize that the 3'O-methyl group is enhancing interactions of the cap residue with eIF4E, thus increasing binding affinity and subsequently translation.

A **B**

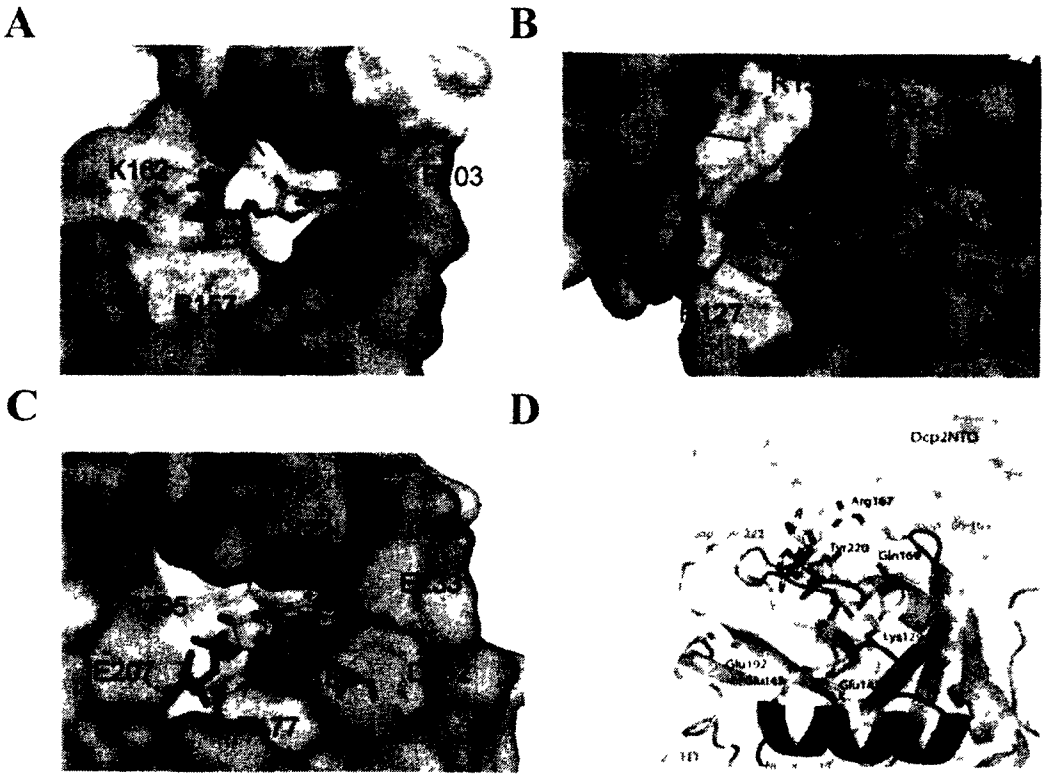


Discussion, figure 4: *In silico* docking of 3'O-Me GDP to eIF4E (PDB 1L8B)

(A) Comparision of the in crystallo ^{m7}GDP with the in silico 3'O-Me GDP, within the cap binding site of the murine eIF4E protein. In blue is the ^{m7}GDP found in the crystal structure, and in yellow is the 3'O-Me GDP of the docking model. (B) Interactions of the 3'O-Me GDP with murine eIF4E as predicted by the docking model.

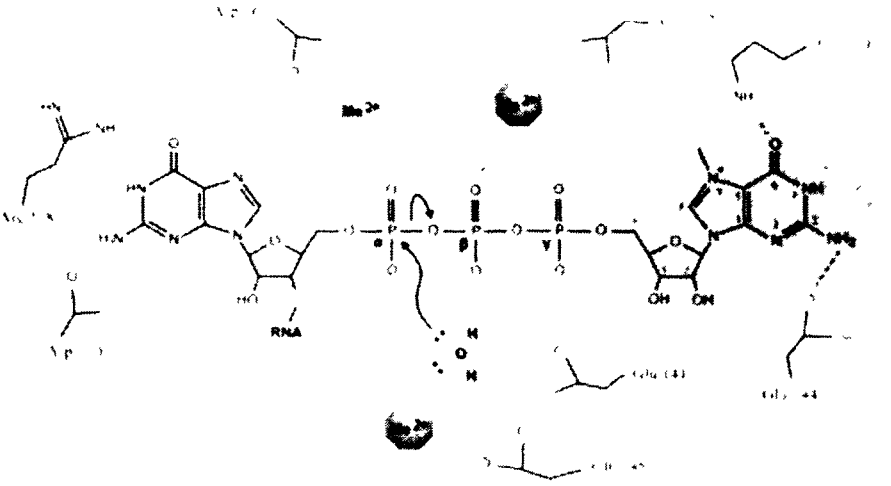
Novel RNA cap analogues to study RNA metabolism

The fact that some N7- methyl deficient capped RNA can sustain cap-dependent translation *in cellulo* has important implications for other aspects of RNA metabolism. The RNA cap structure interacts with several different proteins during its life cycle, starting from CBP20 after its synthesis to decapping enzymes (Dcp2 and DcpS) for its degradation. How do N7-deficient cap analogues interact with cap binding proteins which have different modes of recognizing the cap structure? For instance, decapping enzymes do not recognize the cap structure by π-π stacking interactions like most cap binding proteins (Fig. 5). Instead, decapping enzymes bind the capped guanosine residue by a network of interactions involving hydrogen bonds, polar and non-polar interactions (Fig. 5D). A previous report by *Soulière et al*, showed that 3'O-Me GTP could inhibit decapping activity *in vitro* with similar kinetic constants to GTP, while O6-Me GTP was a poorer inhibitor in comparison (Souliere, Perreault et al. 2010). From the kinetic experiments in this report, we deduced that 3'O-Me GTP could compete for the cap binding site, while O6-Me GTP could not. In addition, neither the proposed model in this report nor previous crystallographic data identified any interactions with the N7-methyl residue (Fig. 6). It would be very interesting to look into the decapping profile of N7-methyl and N7-methyl deficient RNA cap analogues, in order to understand the impact of the novel cap analogues identified in Chapter II, on RNA metabolism. Finally, since both CBP20 and VP39 (a 2'O MTase) recognize the cap structure essentially by π-π stacking interactions not unlike eIF4E, can they also bind to the N7-methyl deficient 3'O-Me guanosine and O6-Me guanosine cap structures? These are important questions that need to be addressed in the near future, since RNA cap analogues are potential candidates as therapeutic agents for the treatment of certain cancers.



Discussion, figure 5: Recognition of the RNA cap structure by different RNA cap binding proteins *(extracted from Fetcher and Brownlee, J Gen Virol, 2005 and She et al, NSMB, 2006)*

(A) Structure of the murine eIF4E protein bound to ^{m7}GDP (PDB 1L8B). (B) Structure of the human CBP20 bound with the cap analogue ^{m7}GpppG (PDB 1H2T) (C) Structure of VP39 from Vaccinia virus bound to a hexameric capped RNA (PDB 1AV6)) (D) Structure of SpDcp1-SpDcp2 bound to a nucleotide (PDB 2A6T)



Discussion, figure 6: Proposed model for cap recognition by a decapping enzyme *(extracted from Soulière et al, Nucleic Acids Res, 2010)*

A proposed model for the recognition of the RNA cap structure by the D10 decapping enzyme from Vaccinia virus. This model was elaborated from inhibition studies of the decapping activity with nucleotide analogues.

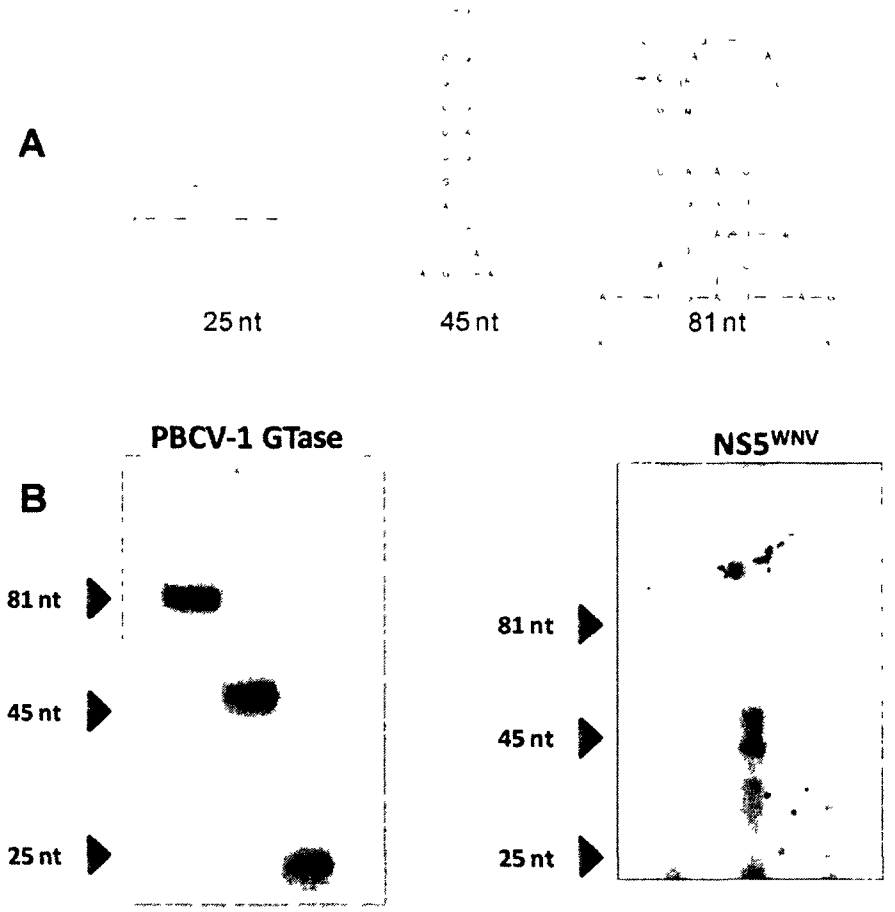
Novel RNA capping enzyme

During my PhD, I made the first biochemical demonstration that the NS5 protein of viruses of the *flavivirus* genus from the Flaviviridae family harboured the RNA guanylyltransferase activity. Up until that point, only the RNA triphosphatase and the RNA (guanine-N7) and RNA (2'O) methyltransferases had been identified. These are harboured on the NS3 protein and on the N-terminal domain of the NS5 protein, respectively. The elusive RNA guanylyltransferase activity had escaped identification. There have been several previous indications that the NS5 protein could potentially harbour this activity. For instance, the N-terminal domain of the NS5 protein from the *Murray Valley Encephalitis* virus has been crystallized with GTP. However, in this structure the GTP was assumed to be a snapshot of the protein in a pre-N7-methylation state whereby the GTP is mimicking the cap guanosine residue. More recently, the N-terminal domain of the NS5 protein from the *Wesselsbron* virus was co-crystallized with GTP, and the mass spectroscopic data in this report seemed

to indicate that this protein fragment could form a covalent bond with GMP (Bollati, Milani et al. 2009). As pointed out in Chapter III though, the mass spectroscopic data in this report by *Bollati et al*, appears to be erroneous (Fig. 5). The fragment reported as a GMP modified peptide of NS5 corresponds to a commonly observed methylated trypsin fragment (Ding, Xiao et al. 2003). In addition, no proof that the NS5 protein could indeed cap an RNA, had yet been reported. In my research, I showed that the NS5 protein is in fact an RNA guanylyltransferase, possessing a specificity towards the initiating nucleotide of the RNA to be capped.

Characterization of the RNA guanylyltransferase activity of NS5^{WNV}

Several aspects of the RNA guanylyltransferase reaction by the NS5^{WNV} protein are presently being investigated in the lab. One of the issues which interests me most is the specificity of the protein towards an RNA initiating with adenine. I therefore decided to look whether the RNA structure itself may potentially be influencing the detected RNA guanylyltransferase activity. Since RNA capping is a co-transcriptional event, and also since only the positive strand of the virus genome is capped, I decided to look into the various conformations that a newly transcribed RNA, corresponding to the positive strand of the virus, may adopt during its synthesis by the bio-informatics tool, MFOLD. I identified several RNA conformations, of which the three major ones are indicated in figure 7A. These RNA molecules were synthesized *in vitro*, di-phosphorylated by the PBCV-1 RNA triphosphatase, and subjected to RNA capping by either the PBCV-1 RNA guanylyltransferase or the NS5^{WNV} protein (Fig.7B). Most interestingly, while the PBCV-1 RNA guanylyltransferase did not show any sequence specificity, the NS5^{WNV} protein had a marked preference for the 45 nucleotide long RNA sequence. We hypothesise that this RNA conformation or sequence is optimal for RNA capping by NS5^{WNV} protein potentially due to conformational changes it may be causing to the protein. However, this remains to be confirmed. It is however not surprising that a particular conformation/sequence seems to favour RNA capping. This particular sequence is absent from the 5' end of the negative strand of the virus, which might explain why it is not capped during the virus replication cycle.



Discussion, figure 7: RNA structure and RNA capping by $NS5^{WNV}$

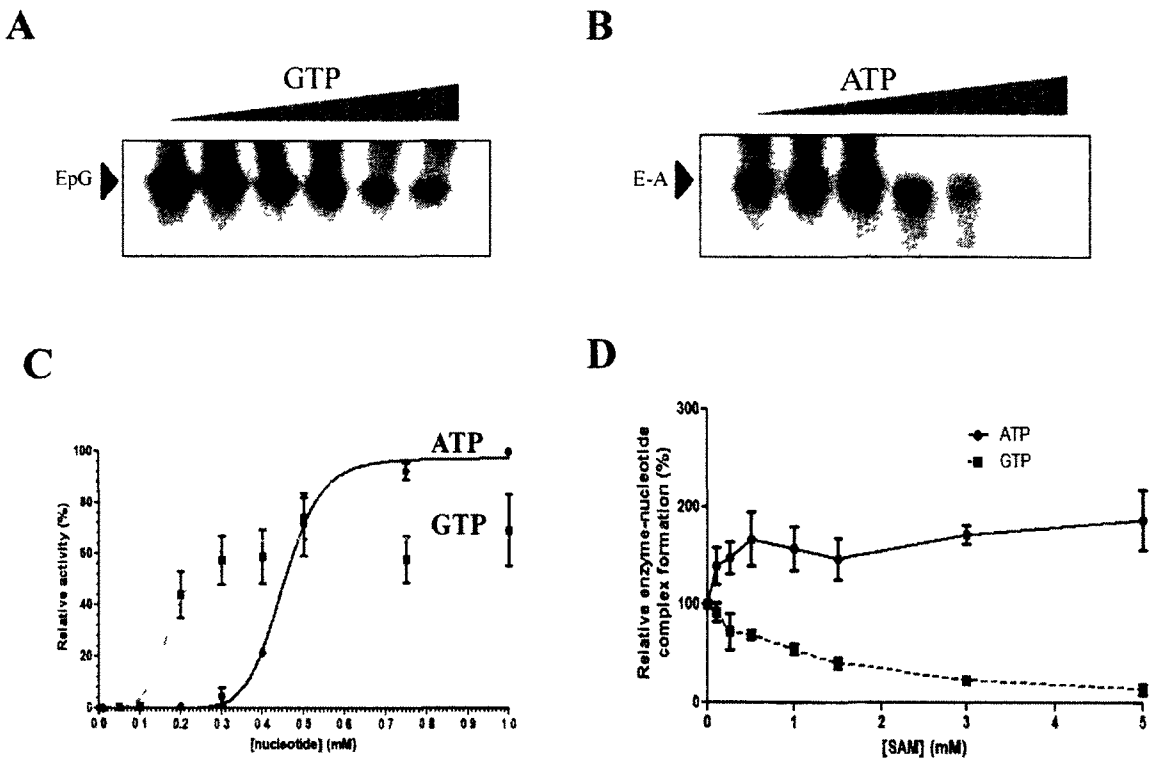
(A) MFOLD predictions for the structure of the first 80 nucleotides of the positive strand of the West Nile Virus genome. The structure has been generated for the first 25, 45 and 81 nt. (B) RNA guanyltransferase reaction by the $NS5^{WNV}$ and PBCV-1 GTase in the presence of the 25 nt, 45 nt and 81 nt diphosphorylated fragments and [α-^{32}P]GTP. The RNAs were extracted by phenol/chloroform and precipitated by ethanol and resolved by UREA-PAGE. An autoradiogram of the gel is shown. To the left is the reaction by the PBCV-1 GTase which does not shown any sequence specificity, while to the right is the $NS5^{WNV}$ which prefers the 45 nt long RNA fragment.

Most curiously, while conventional GTases are very specific to GTP, we discovered in the lab that the $NS5^{WNV}$ protein can also form a covalent complex with ATP (Fig. 8A and 8B). The nature and the importance of this complex are still undetermined. This activity is currently under investigation. Our preliminary characterization revealed that at low

nucleotide concentrations, GTP is hydrolyzed while ATP hydrolysis starts at higher nucleotide concentrations (Fig. 8C). This result is consistent with cellular concentrations of GTP which are usually much lower than that of ATP. Therefore, we deduced that at physiological conditions, the $NS5^{WNV}$ protein can likely form a covalent complex with both GTP and ATP. In order to further characterise this activity, we performed inhibition assays with *S-Adenosyl methionine* (SAM) (Fig. 8D). The results were very suggestive of two different locations for guanylylation and adenylation on the $NS5^{WNV}$ protein. If adenylation and guanylation occurred at the same position, the effect of the inhibitor would have been similar for both reactions. In contrast we observed that while SAM did not negatively impact adenylation, the guanylylation reaction was greatly compromised. Considering that guanylylation occurs in very close proximity to the (guanine-N7) and (2'O) methylation active sites, the negative impact of SAM binding to guanylylation is not wholly unexpected. The fact that SAM does not have the same effect on adenylation suggests that another active site may be involved. Our hypothesis is that the C-terminal domain of the $NS5^{WNV}$ protein, which also harbours RNA-dependent RNA polymerase activity, may be involved in the adenylation activity. Previous reports on the *bunyavirus* RNA polymerase activity have shown that guanylylation of the RdRp is used as a priming mechanism for the polymerase reaction. This may be a plausible explanation for the adenylation of the $NS5^{WNV}$ protein. If this is the case, adenylation would be closely associated to the polymerase active site. Mutations within the polymerase active site which abrogates polymerase activity, more precisely within the essential metal ion binding site would most certainly help to provide an answer to this hypothesis. If adenylation is associated with polymerase activity, we should observe a sharp decrease in adenylated NS5. However, other hypotheses could also explain the adenylation of NS5. For instance, the NS5 protein may also possess ligation activity. Ligases, similar to RNA guanylyltransferases, form an enzyme-nucleotide complex (refer to Introduction, figure 10).

I believe that a better understanding of both the adenylation and guanylylation properties of the NS5 protein would shed some light on the organization of the replication complex of the virus. Questions on how the NS5 protein juggles between RNA polymerase activity, RNA guanylyltransferase activity and RNA (guanine-N7) and (2'O) methyltransferase activities remain to be answered. In addition, prior to RNA capping, the newly synthesized

RNA molecule needs to interact with the NS3 protein for RNA triphosphatase activity. In chapter III, we proposed a simple model whereby the NS3 protein interacts at the linker region between the N- and C- terminal domains of NS5 to ensure this activity. However, according to new theories, NS5 may be a dimer or a multimer. How do all the components of the capping machinery coordinate with the polymerase complex is a key question to be answered in the future.



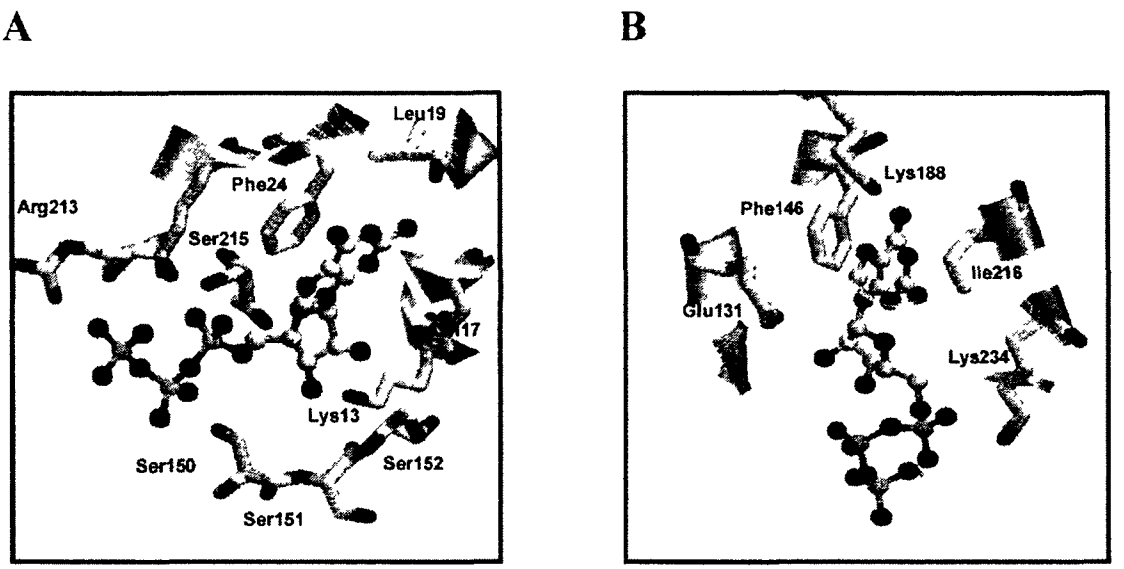
Discussion, figure 8: Guanylylation and adenylation by $NS5^{WNV}$

(A) $NS5^{WNV}$ *has been incubated with [α-*^{32}*P]GTP in the appropriate buffer containing magnesium ions and increasing concentrations of unlabelled GTP up to 1 mM. The formation of the E-GMP covalent complex was monitored by resolving the reactions on an SDS-PAGE An autoradiogram of the PAGE is shown. (B) The* $NS5^{WNV}$ *has been incubated with [α-*^{32}*P]ATP in the same buffer as above and increasing concentrations of unlabelled ATP up to 2 mM The formation of an Enzyme-adenylate covalent complex was observed by resolving the reactions on an SDS-PAGE An autoradiogram of the PAGE is shown. (C) Quantification of the relative guanylylation and adenylation of the* $NS5^{WNV}$ *protein (D) Inhibition of the Enzyme-guanylate (■) or –adenylate (●) complex by S-Adenosyl-methionine (SAM)*

Importance as a drug target

According to the World Health Organization, viruses of the *flavivirus* genus are major causes of mortality and morbidity worldwide (Gould and Solomon 2008). The various genotypes of the Dengue virus alone lead to thousands of deaths per year and almost 50 million people are infected each year, particularly in developing nations (Hotez 2008). Epidemic outbreaks of Dengue virus around the world have increased since 2000 ((WHO) 2009). Other flaviviruses, most notably the Yellow Fever virus and the Japanese Encephalitis virus, cause thousands of deaths each year and are major causes of morbidity in poor countries (Gould and Solomon 2008). These viruses are transmitted principally by the *Aedes aegypti* mosquito, which is known to reproduce in the tropical and some sub-tropical regions of the world (Pastorino, Nougairede et al. 2010). However, due to climate change, the reproductive ground of this mosquito is increasing and consequently, the transmission of flaviviruses may become an issue for developed nations in the near future. The development of antiviral drugs against flaviviruses has become pressing.

The demonstration that the NS5 proteins of flaviviruses possess RNA guanylyltransferase activity has key implications for the design of anti-flaviviral drugs. The NS5 protein is an unconventional RNA guanylyltransferase, because it does not possess any of the conserved motifs of classical GTases, which also includes the human GTase. In addition, GTP binding seems to be mediated by different amino acids in the NS5 protein (Fig. 9). Therefore, the use of nucleotide analogues could potentially prove to be of antiviral interest. As shown in chapter II, the interactions of the classical GTase with its substrates are governed by several key parts of the guanine base and ribose moiety. *Geiss et al*, have drawn a partial picture of the interactions of GTP with the NS5 protein (Discussion, Table 2) (Geiss, Thompson et al. 2009). However, a more detailed analysis of the NS5 interactions with its substrates is required. Comparative analysis of the substrate specificities of the GTP binding sites of NS5 and classical GTase will enable the design of inhibitors specific to the essential GTase activity of the NS5 protein.



Discussion, figure 9: Comparison of the GTP binding site of the N-terminal domain of the NS5 protein of Murray Valley Encephalitis virus with the PBCV-1 GTase

Residues contacting the GTP molecule within 4.0 Å are shown. (A) The GTP binding site of the N-terminal domain of the NS5 protein (PDB 3EVD) (B) The GTP binding site of the PBCV-1 GTase (PDB 1CKM)

CONCLUSIONS

This thesis highlights the coming together of different aspects of research in biological sciences in order to find a solution to the lack of structural data on protein-substrate interactions with regards to several RNA capping enzymes. From the pressing need to find potent anti-microbial and antiviral agents, such integrative approaches offer a fast alternative for the finding of novel drug targets. My research draws a structural link between basic chemical biology, fundamental biochemistry and bio-informatics. Bio-informatics predictions have been successfully applied for the better understanding of protein-substrates interactions. The use of chemical biology to probe into the active site of the yeast RNA triphosphatase has demonstrated the relative plasticity of the tunnel architechture of the TTM fold, and revealed key determinants of the metal-dependent RNA triphosphatase for the rational design of anti-microbial drugs. In addition, through the use of chemical biology, a subset of RNAs harbouring differentially modified caps has been generated which guided some pioneering works on the recognition of an uncharged cap structure by the mammalian eIF4E protein, for cap-dependent translation. Our results greatly enhance the knowledge of cap recognition by several cap binding proteins, and raise the question on the exact role of the N7-methyl group of the cap structure.

In addition, through this research, we laid the foundation for the elaboration of a model of the flaviviral replication complex by identifying one of the core activities essential to the translation of the viral RNA. Our identification of the flaviviral RNA guanylyltransferase paves the way for more research in order to understand the organization of the replication complex, and ultimately for the design of flaviviral specific antiviral drug. At present, on account of the high degree of multi-tasking of the NS5 protein, delimiting the RNA guanylyltransferase site from the RNA (guanine-N7) and (2'O) methyltransferase sites is a very challenging task. In the near future, with the advent of novel bio-informatics tools which account for the dynamism of the active site residues as well as the flexibility of the ligand, such questions would be accurately addressed. Only by evolving at the crossroads of biochemistry, chemical biology, biophysics, cellular biology and bio-informatics can new tools be developed to address some key questions in biological research. This thesis highlights some of the benefits from this approach.

ACKNOWLEDGEMENTS

I sincerely thank my research supervisor, Dr. Martin Bisaillon for having given me this great opportunity to embark on graduate studies in his lab. I thank him for having given me free reign to pursue my ideas, and also to have been of great advice and support throughout my graduate studies. I particularly thank him for having allowed me to attend the 24th Annual Symposium of the Protein Society in 2010; a meeting which was immensely instructive for my future plans.

I thank all members of the lab of Dr. Martin Bisaillon, particularly Isabelle Bougie for her technical support, and Frédéric Picard-Jean for his constructive criticisms on my work. Several past members of the lab have been of considerable help during my formative years. I will make a special mention of Amélie Parent, who has been patient enough to show me the ropes at my beginnings in the lab.

My thanks also go to Dr. Xavier Roucou, for having allowed me to share in the expertise of his lab with regards to cellular culture. I thank him for having provided the necessary infrastructure for all the cell culture experiments I carried out. I also thank all the members of his lab for their dynamism, friendliness, help and support. I particularly thank Antanas Staskevicius for his advices, during the initial stages of my cell culture experiments.

I also extend my thanks to all members of the lab of Dr. Jean-Pierre Perreault for their willingness to share their materials and expertise particularly with regards to RNA manipulation. My thanks go particularly to Dominique Lesvesque for providing helpful comments on several of my experiments.

My thanks also go to the Riboclub, for great talks and great conversations with members. I thank all my colleagues in the department of biochemistry, particularly to the friends I have made. I thank all of them for their support.

My thanks also go to Dr Brian J Geiss of the Colorado State University, for having allowed me to make a brief internship in his lab and for being a great collaborator.

Finally, on a personal level I thank Jenny for her faith in my abilities and her support throughout my graduate studies.

REFERENCES

Adams, B. L., M. Morgan, et al. (1978). "The effect of "cap" analogs on reovirus mRNA binding to wheat germ ribosomes. Evidence for enhancement of ribosomal binding via a preferred cap conformation." J Biol Chem **253**(8): 2589-95.

Bachand, F., D. H. Lackner, et al. (2006). "Autoregulation of ribosome biosynthesis by a translational response in fission yeast." Mol Cell Biol **26**(5): 1731-42.

Benarroch, D., P. Smith, et al. (2008). "Characterization of a trifunctional mimivirus mRNA capping enzyme and crystal structure of the RNA triphosphatase domain." Structure **16**(4): 501-12.

Benzaghou, I., I. Bougie, et al. (2006). "Energetics of RNA binding by the West Nile virus RNA triphosphatase." FEBS Lett **580**(3): 867-77.

Bisaillon, M. and I. Bougie (2003). "Investigating the role of metal ions in the catalytic mechanism of the yeast RNA triphosphatase." J Biol Chem **278**(36): 33963-71.

Bisaillon, M. and G. Lemay (1997). "Viral and cellular enzymes involved in synthesis of mRNA cap structure." Virology **236**(1): 1-7.

Bisaillon, M. and S. Shuman (2001). "Structure-function analysis of the active site tunnel of yeast RNA triphosphatase." J Biol Chem **276**(20): 17261-6.

Bollati, M., K. Alvarez, et al. (2010). "Structure and functionality in flavivirus NS-proteins: perspectives for drug design." Antiviral Res **87**(2): 125-48.

Bollati, M., M. Milani, et al. (2009). "Recognition of RNA cap in the Wesselsbron virus NS5 methyltransferase domain: implications for RNA-capping mechanisms in Flavivirus." J Mol Biol **385**(1): 140-52.

Bougie, I. and M. Bisaillon (2004). "The broad spectrum antiviral nucleoside ribavirin as a substrate for a viral RNA capping enzyme." J Biol Chem **279**(21): 22124-30.

Bougie, I. and M. Bisaillon (2006). "Inhibition of a metal-dependent viral RNA triphosphatase by decavanadate." Biochem J **398**(3): 557-67.

Brown, C. J., I. McNae, et al. (2007). "Crystallographic and mass spectrometric characterisation of eIF4E with N7-alkylated cap derivatives." J Mol Biol **372**(1): 7-15.

Carberry, S. E., E. Darzynkiewicz, et al. (1990). "A spectroscopic study of the binding of N-7-substituted cap analogues to human protein synthesis initiation factor 4E." Biochemistry **29**(13): 3337-41.

Chen, Y., H. Cai, et al. (2009). "Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase." Proc Natl Acad Sci U S A **106**(9): 3484-9.

Daffis, S., K. J. Szretter, et al. (2010). "2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members." Nature **468**(7322): 452-6.

Darzynkiewicz, E., J. Stepinski, et al. (1989). "Inhibition of eukaryotic translation by nucleoside 5'-monophosphate analogues of mRNA 5'-cap: changes in N7 substituent affect analogue activity." Biochemistry **28**(11): 4771-8.

Despins, S., M. Issur, et al. "Deciphering the molecular basis for nucleotide selection by the West Nile virus RNA helicase." Nucleic Acids Res **38**(16): 5493-506.

Ding, Q., L. Xiao, et al. (2003). "Unmatched masses in peptide mass fingerprints caused by cross-contamination: an updated statistical result." Proteomics **3**(7): 1313-7.

Doherty, A. J. (1999). "Conversion of a DNA ligase into an RNA capping enzyme." Nucleic Acids Res **27**(16): 3253-8.

Doherty, A. J., K. Hakansson, et al. (1997). "Crystallization of the RNA guanylyltransferase of Chlorella virus PBCV-1." Acta Crystallogr D Biol Crystallogr **53**(Pt 4): 482-4.

Doherty, A. J. and S. W. Suh (2000). "Structural and mechanistic conservation in DNA ligases." Nucleic Acids Res **28**(21): 4051-8.

Dong, H., B. Zhang, et al. (2008). "Flavivirus methyltransferase: a novel antiviral target." Antiviral Res **80**(1): 1-10.

Ecker, D. J., R. Sampath, et al. (2005). "The Microbial Rosetta Stone Database: a compilation of global and emerging infectious microorganisms and bioterrorist threat agents." BMC Microbiol **5**: 19.

Fabrega, C., S. Hausmann, et al. (2004). "Structure and mechanism of mRNA cap (guanine-N7) methyltransferase." Mol Cell **13**(1): 77-89.

Fabrega, C., V. Shen, et al. (2003). "Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II." Mol Cell **11**(6): 1549-61.

Geiss, B. J., A. A. Thompson, et al. (2009). "Analysis of flavivirus NS5 methyltransferase cap binding." J Mol Biol **385**(5): 1643-54.

Ghosh, A. and C. D. Lima (2010). "Enzymology of RNA cap synthesis." Wiley Interdisciplinary Reviews : RNA **1**(1): 152-172.

Gould, E. A. and T. Solomon (2008). "Pathogenic flaviviruses." Lancet **371**(9611): 500-9.

Graff, J. R., B. W. Konicek, et al. (2008). "Targeting the eukaryotic translation initiation factor 4E for cancer therapy." Cancer Res **68**(3): 631-4.

Graff, J. R., B. W. Konicek, et al. (2007). "Therapeutic suppression of translation initiation factor eIF4E expression reduces tumor growth without toxicity." J Clin Invest **117**(9): 2638-48.

Grudzien, E., J. Stepinski, et al. (2004). "Novel cap analogs for in vitro synthesis of mRNAs with high translational efficiency." RNA **10**(9): 1479-87.

Grudzien-Nogalska, E., J. Jemielity, et al. (2007). "Phosphorothioate cap analogs stabilize mRNA and increase translational efficiency in mammalian cells." RNA **13**(10): 1745-55.

Grudzien-Nogalska, E., J. Stepinski, et al. (2007). "Synthesis of anti-reverse cap analogs (ARCAs) and their applications in mRNA translation and stability." Methods Enzymol **431**: 203-27.

Gu, M. and C. D. Lima (2005). "Processing the message: structural insights into capping and decapping mRNA." Curr Opin Struct Biol **15**(1): 99-106.

Hakansson, K., A. J. Doherty, et al. (1997). "X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes." Cell **89**(4): 545-53.

Hamm, J. and I. W. Mattaj (1990). "Monomethylated cap structures facilitate RNA export from the nucleus." Cell **63**(1): 109-18.

Ho, C. K. and S. Shuman (2001). "A yeast-like mRNA capping apparatus in Plasmodium falciparum." Proc Natl Acad Sci U S A **98**(6): 3050-5.

Ho, C. K., J. L. Van Etten, et al. (1996). "Expression and characterization of an RNA capping enzyme encoded by Chlorella virus PBCV-1." J Virol **70**(10): 6658-64.

Hotez, P. (2008). Forgotten people, forgotten diseases: the neglected tropical diseases and their impact on global health and development. Washington, DC, American Society for Microbiology Press.

Howe, K. J. (2002). "RNA polymerase II conducts a symphony of pre-mRNA processing activities." Biochim Biophys Acta **1577**(2): 308-24.

Isken, O. and L. E. Maquat (2007). "Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function." Genes Dev **21**(15): 1833-56.

Issur, M., S. Despins, et al. (2009). "Nucleotide analogs and molecular modeling studies reveal key interactions involved in substrate recognition by the yeast RNA triphosphatase." Nucleic Acids Res **37**(11): 3714-22.

Izaurralde, E., J. Lewis, et al. (1994). "A nuclear cap binding protein complex involved in pre-mRNA splicing." Cell **78**(4): 657-68.

Jain, R. and S. Shuman (2008). "Polyphosphatase activity of CthTTM, a bacterial triphosphate tunnel metalloenzyme." J Biol Chem **283**(45): 31047-57.

Jemielity, J., T. Fowler, et al. (2003). "Novel "anti-reverse" cap analogs with superior translational properties." RNA **9**(9): 1108-22.

Kalek, M., J. Jemielity, et al. (2005). "Synthesis and biochemical properties of novel mRNA 5' cap analogs resistant to enzymatic hydrolysis." Nucleosides Nucleotides Nucleic Acids **24**(5-7): 615-21.

Kmpf, G., E. Kapinos, et al. (2002). "Comparison of the acid–base properties of purine derivatives in aqueous solution. Determination of intrinsic proton affinities of various basic sites." Journal of the Chemical Society, Perkin Transactions 2: 1320-1327.

Kore, A. R. and M. Shanmugasundaram (2008). "Synthesis and biological evaluation of trimethyl-substituted cap analogs." Bioorg Med Chem Lett **18**(3): 880-4.

Kowalska, J., M. Lewdorowicz, et al. (2008). "Synthesis and characterization of mRNA cap analogs containing phosphorothioate substitutions that bind tightly to eIF4E and are resistant to the decapping pyrophosphatase DcpS." RNA **14**(6): 1119-31.

Kuhn, A. N., M. Diken, et al. "Phosphorothioate cap analogs increase stability and translational efficiency of RNA vaccines in immature dendritic cells and induce superior immune responses in vivo." Gene Ther **17**(8): 961-71.

Kuhn, U. and E. Wahle (2004). "Structure and function of poly(A) binding proteins." Biochim Biophys Acta **1678**(2-3): 67-84.

Kulikowska, E., B. Kierdaszuk, et al. (2004). "Xanthine, xanthosine and its nucleotides: solution structures of neutral and ionic forms, and relevance to substrate properties in various enzyme systems and metabolic pathways." Acta Biochim Pol **51**(2): 493-531.

Lescar, J., D. Luo, et al. (2008). "Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target." Antiviral Res **80**(2): 94-101.

Lewis, J. D. and E. Izaurralde (1997). "The role of the cap structure in RNA processing and nuclear export." Eur J Biochem **247**(2): 461-9.

Lewis, J. D., E. Izaurralde, et al. (1996). "A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site." Genes Dev **10**(13): 1683-98.

Lima. C. D., L. K. Wang, et al. (1999). "Structure and mechanism of yeast RNA triphosphatase: an essential component of the mRNA capping apparatus." Cell **99**(5): 533-43.

Livingstone, M., E. Atas, et al. (2010). "Mechanisms governing the control of mRNA translation." Phys Biol **7**(2): 021001.

Mao, X., B. Schwer, et al. (1996). "Mutational analysis of the Saccharomyces cerevisiae ABD1 gene: cap methyltransferase activity is essential for cell growth." Mol Cell Biol **16**(2): 475-80.

Maquat, L. E. (2004). "Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics." Nat Rev Mol Cell Biol **5**(2): 89-99.

Marcotrigiano, J., A. C. Gingras, et al. (1997). "Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP." Cell **89**(6): 951-61.

Martinez-Contreras, R., P. Cloutier, et al. (2007). "hnRNP proteins and splicing control." Adv Exp Med Biol **623**: 123-47.

Matsuo, H., H. Li, et al. (1997). "Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein." Nat Struct Biol **4**(9): 717-24.

Moore, M. J. and N. J. Proudfoot (2009). "Pre-mRNA processing reaches back to transcription and ahead to translation." Cell **136**(4): 688-700.

Nair, P. A., J. Nandakumar, et al. (2007). "Structural basis for nick recognition by a minimal pluripotent DNA ligase." Nat Struct Mol Biol **14**(8): 770-8.

Newbury, S. F. (2006). "Control of mRNA stability in eukaryotes." Biochem Soc Trans **34**(Pt 1): 30-4.

Niedzwiecka, A., J. Marcotrigiano, et al. (2002). "Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins." J Mol Biol **319**(3): 615-35.

Niedzwiecka, A., J. Stepinski, et al. (2003). "Thermodynamics of 7-methylguanosine cation stacking with tryptophan upon mRNA 5' cap binding to translation factor eIF4E." Nucleosides Nucleotides Nucleic Acids **22**(5-8): 1557-61.

Niedzwiecka, A., J. Stepinski, et al. (2002). "Positive heat capacity change upon specific binding of translation initiation factor eIF4E to mRNA 5' cap." Biochemistry **41**(40): 12140-8.

Ogino, T. and A. K. Banerjee (2007). "Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus." Mol Cell **25**(1): 85-97.

Ogino, T., S. P. Yadav, et al. "Histidine-mediated RNA transfer to GDP for unique mRNA capping by vesicular stomatitis virus RNA polymerase." Proc Natl Acad Sci U S A **107**(8): 3463-8.

Pastorino, B., A. Nougairede, et al. (2010). "Role of host cell factors in flavivirus infection: Implications for pathogenesis and development of antiviral drugs." Antiviral Res **87**(3): 281-94.

Raczynska, K. D., C. G. Simpson, et al. (2010). "Involvement of the nuclear cap-binding protein complex in alternative splicing in Arabidopsis thaliana." Nucleic Acids Res **38**(1): 265-78.

Reddy, R., R. Singh, et al. (1992). "Methylated cap structures in eukaryotic RNAs: structure, synthesis and functions." Pharmacol Ther **54**(3): 249-67.

Rhoads, R. E., G. M. Hellmann, et al. (1983). "Translational recognition of messenger ribonucleic acid caps as a function of pH." Biochemistry **22**(26): 6084-8.

Roitzsch, M., O. Fedorova, et al. "The 2'-OH group at the group II intron terminus acts as a proton shuttle." Nat Chem Biol **6**(3): 218-224.

Ruggero, D., L. Montanaro, et al. (2004). "The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis." Nat Med **10**(5): 484-6.

Saha, N., B. Schwer, et al. (1999). "Characterization of human, Schizosaccharomyces pombe, and Candida albicans mRNA cap methyltransferases and complete replacement of the yeast capping apparatus by mammalian enzymes." J Biol Chem **274**(23): 16553-62.

Sampath, A. and R. Padmanabhan (2009). "Molecular targets for flavivirus drug discovery." Antiviral Res **81**(1): 6-15.

Shatkin, A. J. and J. L. Manley (2000). "The ends of the affair: capping and polyadenylation." Nat Struct Biol **7**(10): 838-42.

Shuman, S. (1995). "Capping enzyme in eukaryotic mRNA synthesis." Prog Nucleic Acid Res Mol Biol **50**: 101-29.

Shuman, S. (2001). "Structure, mechanism, and evolution of the mRNA capping apparatus." Prog Nucleic Acid Res Mol Biol **66**: 1-40.

Shuman, S. (2002). "What messenger RNA capping tells us about eukaryotic evolution." Nat Rev Mol Cell Biol **3**(8): 619-25.

Shuman, S. and C. D. Lima (2004). "The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases." Curr Opin Struct Biol **14**(6): 757-64.

Shuman, S., Y. Liu, et al. (1994). "Covalent catalysis in nucleotidyl transfer reactions: essential motifs in Saccharomyces cerevisiae RNA capping enzyme are conserved in Schizosaccharomyces pombe and viral capping enzymes and among polynucleotide ligases." Proc Natl Acad Sci U S A **91**(25): 12046-50.

Slepenkov, S. V., E. Darzynkiewicz, et al. (2006). "Stopped-flow kinetic analysis of eIF4E and phosphorylated eIF4E binding to cap analogs and capped oligoribonucleotides: evidence for a one-step binding mechanism." J Biol Chem **281**(21): 14927-38.

Sonenberg, N. (2008). "eIF4E, the mRNA cap-binding protein: from basic discovery to translational research." Biochem Cell Biol **86**(2): 178-83.

Souliere, M. F., J. P. Perreault, et al. (2008). "Kinetic and thermodynamic characterization of the RNA guanylyltransferase reaction." Biochemistry **47**(12): 3863-74.

Souliere, M. F., J. P. Perreault, et al. (2010). "Insights into the molecular determinants involved in cap recognition by the vaccinia virus D10 decapping enzyme." Nucleic Acids Res **38**(21): 7599-610.

Stepinski, J., J. Zuberek, et al. (2005). "Novel dinucleoside 5',5'-triphosphate cap analogues. Synthesis and affinity for murine translation factor eIF4E." Nucleosides Nucleotides Nucleic Acids **24**(5-7): 629-33.

Strenkowska, M., J. Kowalska, et al. (2010). "Towards mRNA with superior translational activity: synthesis and properties of ARCA tetraphosphates with single phosphorothioate modifications." New J Chem **34**(5): 993-1007.

Ueda, H., H. Maruyama, et al. (1991). "Expression of a synthetic gene for human cap binding protein (human IF-4E) in Escherichia coli and fluorescence studies on interaction with mRNA cap structure analogues." J Biochem **109**(6): 882-9.

Varani, G. (1997). "A cap for all occasions." Structure **5**(7): 855-8.

Venkatesan, S. and B. Moss (1980). "Donor and acceptor specificities of HeLa cell mRNA guanylyltransferase." J Biol Chem **255**(7): 2835-42.

Venkatesan, S. and B. Moss (1982). "Eukaryotic mRNA capping enzyme-guanylate covalent intermediate." Proc Natl Acad Sci U S A **79**(2): 340-4.

Visa, N., E. Izaurralde, et al. (1996). "A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export." J Cell Biol **133**(1): 5-14.

Wang, S. P., L. Deng, et al. (1997). "Phylogeny of mRNA capping enzymes." Proc Natl Acad Sci U S A **94**(18): 9573-8.

(WHO), W. H. O. (2009). DENGUE GUIDELINES FOR DIAGNOSIS, TREATMENT, PREVENTION AND CONTROL. Geneva, Switzerland, WHO.

Zust, R., L. Cervantes-Barragan, et al. (2010). "Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5." Nat Immunol **12**(2): 137-43.

ANNEX

1. Co-authors approbation letters