



# **In vitro translation of virally-encoded replication polyproteins to recapitulate polyprotein maturation processes**

Johann Habersetzer, Mohamed Debbah, Marie-Laure Fogeron, Anja Böckmann, Stéphane Bressanelli, Sonia Fieulaine

## **► To cite this version:**

Johann Habersetzer, Mohamed Debbah, Marie-Laure Fogeron, Anja Böckmann, Stéphane Bressanelli, et al.. In vitro translation of virally-encoded replication polyproteins to recapitulate polyprotein maturation processes. Protein Expression and Purification, 2020, 10.1016/j.pep.2020.105694 . hal-03009803

**HAL Id: hal-03009803**

**<https://hal.science/hal-03009803>**

Submitted on 3 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

***In vitro* translation of virally-encoded replication polyproteins to recapitulate  
polyprotein maturation processes**

Johann Habersetzer<sup>a</sup>, Mohamed Debbah<sup>a</sup>, Marie-Laure Fogeron<sup>b</sup>, Anja Böckmann<sup>b</sup>,  
Stéphane Bressanelli<sup>a</sup>, Sonia Fieulaine<sup>a,\*</sup>

<sup>a</sup> Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC),  
91198, Gif-sur-Yvette, France

<sup>b</sup> Molecular Microbiology and Structural Biochemistry, Labex Ecofect, UMR 5086  
CNRS/Université de Lyon, 7 passage du Vercors, 69367, Lyon, France

**Keywords:** polyprotein, virus, maturation process, wheat-germ cell-free expression system

**Abbreviations:**

(+)RNA viruses: single-stranded, positive-sense RNA viruses

CFS: cell-free sample

huNoV: human norovirus

ORF: open reading frame

P: pellet

RdRp: RNA dependent RNA polymerase

S: supernatant

S-B: supernatant enriched with Strep-Tactin<sup>®</sup> beads

TYMV: *Turnip yellow mosaic virus*

VRC: viral replication complex

WGE: wheat-germ extract

**\* Corresponding author.**

E-mail address: sonia.fieulaine@i2bc.paris-saclay.fr (S. Fieulaine).

## **ABSTRACT**

Single-stranded, positive-sense RNA viruses encode essential replication polyproteins which are composed of several domains. They are usually subjected to finely regulated proteolytic maturation processes to generate cleavage intermediates and end-products. Both polyproteins and maturation products play multiple key roles that ultimately allow synthesis of viral genome progeny. Despite the importance of these proteins in the course of viral replication, their structural properties, including the conformational changes regulating their numerous functions, are poorly described at the structural level. This lack of information is mainly due to the extreme difficulty to express large, membrane-bound, multi-domain proteins with criteria suitable for structural biology methods. To tackle this challenge, we have used a wheat-germ cell-free expression system. We firstly establish that this approach allows to synthesize viral polyproteins encoded by two unrelated positive-sense RNA viruses, a human norovirus and a plant tymovirus. Then, we demonstrate that these polyproteins are fully functional and are spontaneously auto-cleaved by their active protease domain, giving rise to natural maturation products. Moreover, we show that introduction of point mutations in polyproteins allows to inhibit the proteolytic maturation process of each virus. This allowed us to express and partially purify the uncleaved full-length norovirus polyprotein and the tymoviral RNA-dependent RNA polymerase. Thus, this study provides a powerful tool to obtain soluble viral polyproteins and their maturation products in order to conduct challenging structural biology projects and therefore solve unanswered questions.

## 1. Introduction

Single-stranded, positive-sense RNA ((+)RNA) viruses constitute the largest group of eukaryotic viruses [1]. Using a largely similar replication cycle, they infect numerous hosts that are very different including humans, animals and plants. One hallmark of (+)RNA viruses is the synthesis of replication polyproteins made up of numerous domains, constituting a precursor for the production of viral non-structural proteins [2]. Some of these domains display enzymatic activities required for viral RNA synthesis, while others, containing hydrophobic sequences, are involved in rearrangements of infected cells membranes and in the formation of the membrane-bounded viral replication complex where viral RNA synthesis takes place [3], [4].

Viral polyproteins are usually further processed in one or several steps by dedicated proteases, giving rise to one or several cleavage intermediate(s) and then end-products [2], [5], [6]. Viral protein domains often display distinct functions depending on their context, *i.e.* embedded in precursor, intermediate(s) or as end-products. The set of functions played by each viral protein domain and protein is viewed as an ingenious strategy evolved by (+)RNA viruses to cope with their very small genome coding for only few proteins [6]–[8]. In addition, viral polyproteins are highly reconfigurable entities, subjected to important conformational changes that modify the relative orientation of their domains with respect to each other [9]–[12]. The tunable spatial arrangements of each domain thus allows the on/off switch of numerous viral functions, a phenomenon that has been particularly described for the viral protease and RNA polymerase domains [10], [13], [14]. Finally, polyprotein processing leads to temporal and spatial regulation of protein activity, depending on polyprotein cleavage state [15] and regulating specific steps of the viral replication cycle [16].

However, despite the crucial importance of viral polyprotein maturation process and the evidence that numerous steps of viral replication, including the synthesis of new viral genomes, strongly depend on the ability of viral replication polyproteins to be processed, very little is known about the associated structural changes allowing viral protein precursors, intermediate products and end-products to play different functions. Only few structures of precursors or intermediates have been solved [15]. This is mainly due to the difficulties to express these large proteins that are composed of several domains that can assume different conformations. However, it has recently been shown that a wheat-germ cell-free expression system [17] allows for the production of notoriously difficult-to-express membrane-bound viral proteins with criteria suitable for structural biology projects, *i.e.* solubility, purity and quantity [18]–[23]. We have applied this *in vitro* translation system to the expression of

replication polyproteins encoded by two unrelated (+)RNA viruses (a human norovirus and a plant tymovirus), and thus show that this approach is perfectly suitable for the production of pure, soluble and functional viral polyproteins. We also show that viral polyproteins can be engineered to produce unprocessed protein species and have thus successfully purified a human norovirus uncleaved polyprotein of 190 kDa. By this approach, we also have succeeded to produce the tymoviral RNA-dependent RNA polymerase (RdRp), a protein related to the alphavirus-like virus polymerases that are extremely difficult to produce outside the context of their polyproteins and for which no structure is currently available [24].

## 2. Materials and methods

### 2.1. Plasmid construction

cDNA encoding viral polyproteins were inserted into the expression vector pEU-E01-MCS (CellFree Sciences, Japan) that contains a translational enhancer and a SP6 promoter driving expression of the encoded proteins [25].

Nucleotide sequences coding for ORF1 from two huNoV GII.4 strains including a Sydney 2012 variant (Syd) (GenBank accession no JX459908) and a Den Haag 2006b variant (DH) (GenBank accession no. AB447456) were used in this work. Full length ORF1s were amplified by PCR using pBMH\_NoroSydney and pBMH\_Saga\_DenHaag plasmids [3] as templates respectively, then digested by SpeI and NotI and cloned into similarly digested pEU-E01-MCS to give pEU\_Sydney and pEU\_DenHaag plasmids. To prevent maturation of norovirus polyproteins by the viral protease NS6, the catalytic Cys of each construct was mutated in Ala by Quick-change mutagenesis (Stratagene) to produce pEU\_Sydney\_C1147A and pEU\_DenHaag\_C1147A. Primers used for cloning and mutagenesis are given in Table I.

**Table I. Primers used for cloning and mutagenesis of norovirus plasmids.**

SpeI and NotI restrictions site are underlined, modified nucleotides are shown in lower case.

Sequence (from 5' to 3')	Orientation	Purpose
GC <u>ACTAGT</u> ATGTGGAGCCACCCGAGTTCGAAAAAAGATGGCGTCTAACGACGCT	sense	Design of pEU_Sydney
GCGCGGCCGCTTATTTTCGAACTGCGGGTGGCTCCACTCGACGCCATCTTCATTCA	antisense	
GC <u>ACTAGT</u> ATGTGGAGCCACCCGAGTTCGAAAAAAGATGGCGTCTAACGACGC	sense	Design of pEU_DenHaag
GCGCGGCCGCTTATTTTCGAACTGCGGGTGGCTCCATTCGACGCCATCTTCATTCA	antisense	
GCACAACACCAGGCGACgcGGCTGCCCCCTACATT	sense	Mutagenesis of pEU_Sydney
GCAATGTAGGGGACGCCggcGTCGCCTGGTGTGTGT	antisense	
GCACCACACCAGGTGATgccGGCTGTCCCTACATC	sense	Mutagenesis of pEU_DenHaag
GCGATGTAGGGACAGCCggcATCACCTGGTGTGGT	antisense	

The ORF coding for 206K polyprotein of wild-type TYMV was cloned into the pEU-E01-MCS vector by RD-Biotech company (<https://www.rd-biotech.com/>), between SpeI and BamHI restriction sites, using E17-stop69K plasmid [26] as template. C783S and K874R875L876L877G878S879Q880/AAAAAAA mutations were then introduced into the resulting pEU\_206K plasmid by sub-cloning BstEII/XhoI digested fragment of pΩ-140K-C783S and pΩ-206K-7B plasmids [27] already containing the defined mutations. Resulting plasmids were called pEU\_206K-CS and pEU\_206K-A/S2.

Plasmids were amplified through transformation of DH5α (Invitrogen) or NEB Turbo (New England Biolabs) competent cells and isolated using NucleoBond Xtra Maxi kit (Macherey-Nagel). Plasmids were then purified by a phenol/chloroform extraction. All designed plasmids, listed in Table II, were sequenced by Sanger sequencing (GATC Biotech, Constance, Germany). For each construct, Strep-tags II (WSHPQFEK) were fused at both the N- and the C-termini of the polyproteins. Strep-tags are generally preferable to His-tags in the wheat-germ cell-free expression system (unpublished data and references [23], [25], see also Discussion) and were used for Western-blotting and for affinity purification steps.

**Table II. Expression plasmids for cell-free expression.**

Name	Description	Tags
<b>Noroviruses</b>		
pEU_Sydney	full-length Norovirus Sydney ORF1, residues 2-1699 wt	Strep-tag in N-ter and C-ter
pEU_Sydney_C1147A	full-length Norovirus Sydney ORF1, residues 2-1699 mutation C1147A	Strep-tag in N-ter and C-ter
pEU_DenHaag	full-length Norovirus Den Haag ORF1, residues 2-1699 wt	Strep-tag in N-ter and C-ter
pEU_DenHaag_C1147A	full-length Norovirus Den Haag ORF1, residues 2-1699 mutation C1147A	Strep-tag in N-ter and C-ter
<b>Tymovirus</b>		
pEU_206K	full-length TYMV ORF1, residues 2-1844 wt	Strep-tag in N-ter and C-ter
pEU_206K-CS	full-length TYMV ORF1, residues 2-1844 mutation C783S	Strep-tag in N-ter and C-ter
pEU_206K-A/S2	full-length TYMV ORF1, residues 2-1844 mutation K874R875L876L877G878S879Q880/AAAAAAA	Strep-tag in N-ter and C-ter

## 2.2. Wheat-germ cell-free expression protocol

### 2.2.1. *In vitro* transcription

*In vitro* transcription was performed as described in Takai *et al.* [25]. Plasmids were resuspended at a final concentration of 100 ng/μL in transcription buffer (CellFree Sciences)

supplemented with 2.5 mM of a rNTPs mix (Promega), 1 U/ $\mu$ L of RNasin (Promega) and 1 U/ $\mu$ L of SP6 polymerase (Promega). The whole mixture was incubated at 37°C during 7 hours and transcripts were used for *in vitro* translation.

### **2.2.2. *In vitro* translation**

Wheat-germ extracts were prepared according to Fogeron *et al.* [28] using non-treated Miradoux durum wheat seeds (Sud-Céréales, France).

*In vitro* translation was done using the bilayer method in 96-well plates as described by Takai *et al.* [25]. The bottom layer is composed of 10  $\mu$ L of synthesized RNA from *in vitro* transcription and 10  $\mu$ L of a mix containing homemade wheat-germ extract, 0.3 mM of a mix containing the twenty natural unlabeled amino acids (Cambridge Isotopes Laboratories) and 40  $\mu$ g/mL of creatine kinase (Roche). The upper layer is composed of 200  $\mu$ L of feeding buffer SUB-AMIX (CellFree Sciences) containing 30 mM Hepes-KOH, pH 7.6, 100 mM potassium acetate, 2.7 mM magnesium acetate, 16 mM creatine phosphate, 0.4 mM spermidine, 1.2 mM ATP, 0.25 mM GTP, 4 mM DTT, supplemented by 0.3 mM of unlabeled amino acids mix (Cambridge Isotopes Laboratories). When translation was performed in the presence of detergents (Anatrace), those were added in both layers. Translation reactions were done at 18°C (huNoV) or 22°C (TYMV) during 16 hours. All experiments were performed three times.

### **2.2.3. Sample analysis**

Samples were prepared according to the method described by Fogeron *et al.* [19], [28]. Following *in vitro* translation, samples were incubated with home-made benzonase on a rotating wheel during 1 hour at 4°C (huNoV) or room temperature (TYMV) and then centrifuged during 30 min at 4°C at 20,000 g. The pellets were resuspended in equivalent volume of supernatant with wash buffer (IBA Lifesciences) composed of 100 mM Tris-HCl, 150 mM NaCl and 1 mM EDTA, pH 8. Supernatants were then incubated with Strep-Tactin<sup>®</sup> XT magnetic beads (IBA) during 30 min at 4°C (huNoV) or 1 hour at room temperature (TYMV). Beads were washed 3 times with 10 column volumes (CV) of wash buffer and then resuspended with Blue loading buffer (50 mM Tris-HCl pH 6.8, 2% SDS (w/v), 10% glycerol (v/v), 4%  $\beta$ -mercaptoethanol, 0.02% bromophenol blue (w/v)).

For analysis, samples were submitted to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using 10% acrylamide Tris-Glycine gels [29] and proteins were electro-transferred on nitro-cellulose membrane (GE Healthcare Life Sciences). Membranes

were incubated with Strep-MAB classics primary antibodies (IBA) and then with anti-mouse IgG HRP conjugate secondary antibodies (Promega). Western-blot was revealed using the enhanced chemiluminescence method using ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences) on a CCD camera (Gbox, Syngene).

### 2.3. Large scale expression and purification

In order to purify polyproteins, translation reactions were up-scaled in 6-well plates [25]. The bottom layer was composed of 250  $\mu$ L of RNA obtained by *in vitro* transcription as described above and 250  $\mu$ L of home-made wheat-germ extract supplemented with amino acids and creatine kinase. The upper layer was composed of 5.5 mL of feeding buffer containing SUB-AMIX and amino acids as described above. Detergent was added in each layer as described in Fogeron *et al.* [19].

huNoV 190K polyprotein with mutation C1147A (see **Table II**) was produced in 18 wells at 18°C, in the presence of 0.1% (w/v) MNG-3. Following this step, samples were prepared as described before (part 2.3.3) and 108 mL of supernatant were incubated with 3.5 mL of Strep-tag affinity resin (Strep-Tactin<sup>®</sup> Superflow high capacity, IBA) on a rotary wheel during 3 hours at 4°C. The resin was then washed 2 times with 1 CV of wash buffer supplemented with 0.1% MNG-3 and 3 times with wash buffer supplemented with 0.02% MNG-3. Then elution was performed with 6 CV of elution buffer composed of 100 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, 0.02% MNG-3 and 2.5 mM desthiobiotin, pH 8. The protein concentration was estimated from the calculated extinction coefficient at 280 nm ( $231,030 \text{ M}^{-1} \text{ cm}^{-1}$ ). Purification process was monitored by SDS-PAGE and Western-blot with anti-Strep-tag antibodies (IBA) and then with anti-mouse IgG HRP conjugate secondary antibodies (Promega).

TYMV 206K polyprotein with mutation A/S2 (see **Table II**) was produced in 6 wells at 22°C in presence of 0.1% Brij-58. Samples were then prepared as previously described (see part 2.3.3) and 36 mL of supernatant were incubated with 4 mL of Strep-tag affinity resin (Strep-Tactin<sup>®</sup> Superflow high capacity, IBA) on a rotary wheel during 21 hours at 4°C. The resin was then washed 5 times with 1 CV of wash buffer supplemented with 0.1% Brij-58. Twenty mL of elution buffer containing 5 mM desthiobiotin were added to the resin and the mixture was incubated 24 hours at 4°C under agitation. The resin was then packed into empty PD-10 column (GE Healthcare) and protein was eluted in 11 fractions. The resin was finally washed 4 times with 2 CV of wash buffer containing 0.1% Brij-58. Purification process was monitored by Western-blot with anti-Strep-tag antibodies (IBA) and anti-mouse IgG HRP



conjugate secondary antibodies (Promega), or anti-66K primary antibodies [30] and then with anti-rabbit IgG HRP conjugate secondary antibodies (Promega).

Larger scale expression and purification of huNoV and TYMV polyproteins were realized at least twice.

### 3. Results

#### 3.1. Successful expression and maturation of norovirus polyprotein

Noroviruses are the most common causative agents of acute viral gastroenteritis worldwide. They belong to *Caliciviridae*, a family of the picornavirus-like superfamily [1]. Noroviruses are divided into seven genogroups (from GI to GVII) which are further subdivided into several genotypes. Genogroups I, II and IV primarily infect humans, with genogroup II genotype 4 (GII.4) causing the majority of infections [31].

The human noroviruses (huNoV) genome is composed of three open reading frames (ORFs). ORF2 and ORF3 encode structural capsid proteins, while ORF1 encodes a polyprotein (around 190 kDa) which contains several non-structural domains involved in viral replication (NS1-2, NS3, NS4, NS5, NS6, NS7) [32], [33] (**Fig.1A**). Functions of C-terminal domains were characterized: NS5 is covalently linked to the viral genome, NS6 is a cysteine protease involved in the polyprotein maturation process and NS7 is the viral RNA-dependent RNA polymerase. On the opposite, N-terminal domains encompassing NS1-2, NS3 and NS4 remain less characterized. Mouse and human proteins are known to localize to cellular membranes [3], [32], [34] and contribute to viral membrane alterations of infected cells whose NS4 would be the main organizer of the norovirus replication complex [3]. While it is known that there are six end-products resulting from the norovirus polyprotein processing [35]–[41] (**Fig.1A**), the exact number of cleavage intermediates as well as their potential functions are still poorly documented. Small scale *in vitro* translation assays in a rabbit reticulocyte lysate, transformation of human hepatoma cell line Huh7 with plasmid encoding proteins from a human NoV, as well as infection of BV-2 murine cells with a murine norovirus revealed the presence of different cleavage intermediates including NS1-2-NS3, NS4-NS5-NS6-NS7, NS6-NS7, NS4-NS5, NS4-NS5-NS6 and NS5-NS6 [3], [16], [35]. Among these cleavage intermediates, those including the membrane-interacting NS1-2, NS3 and NS4 would play an essential role in the localization of NS6 and NS7 to the membrane replication complex [16]. We tested whether wheat-germ cell-free expression system would be an appropriate tool to obtain huNoV polyproteins in milligram amount, sufficient for cryo-electron microscopy or X-ray crystallography. In this work, we have used the wild-type ORF1 polyprotein encoded

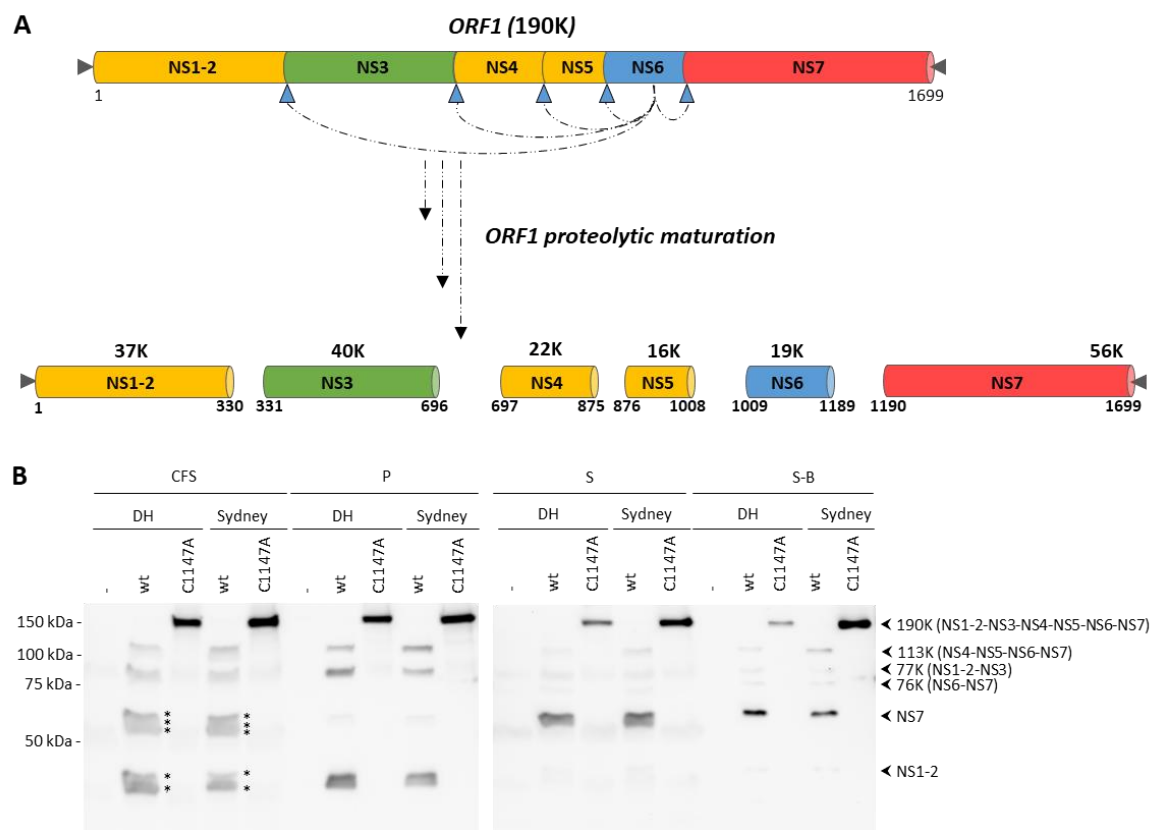
by two variants (Den Haag 2006b and Sydney 2012) belonging to the most clinically relevant huNoV GII.4. The nature of cleavage products resulting from polyprotein processing was observed after gel migration and Western-blot (**Fig.1B**). As Strep-tags were fused to both polyprotein extremities (**Fig.1A**), only cleavage products containing NS1-2 or NS7 domains are detectable. In the wild-type (wt) context, several bands which are not present in the negative control are detected in the translation reaction mixture also called cell-free sample (CFS) for the two variants (DH and Sydney) (**Fig.1B**, “CFS” samples, lanes wt). Using the estimated molecular weights and available data regarding *in vitro* processing of norovirus polyproteins [3], [35], we attributed these bands to the two end-products NS7 (56 kDa) and NS1-2 (37 kDa) and the two cleavage intermediates NS4-NS5-NS6-NS7 (113 kDa) and NS1-2-NS3 (77 kDa). NS7 and NS1-2 proteins migrate as a triplet and a doublet, respectively. Although the nature of these different bands is uncharacterized, they could reflect degradation or post-translational modifications including phosphorylation [42]–[45]. When the catalytic Cys1147 of huNoV protease NS6 is mutated to Ala, the four bands/multiplets are absent and only one strong band around 200 kDa is detected for both genotypes in the total cell-free sample (**Fig.1B**, “CFS” samples, lanes C1147A). This band represents the uncleaved polyprotein resulting from the inactivation of NS6 enzymatic activity. The absence of this band in a wild-type context proves the efficiency of the maturation process during cell-free expression as only cleavage products are detected. The cell-free system is therefore suitable to produce the human norovirus polyprotein and/or several of its processing cleavage forms belonging to both DH and Sydney genotypes. Moreover, it represents a very good tool to selectively produce uncleaved polyproteins or specific processed forms.

Wishing to examine the solubility profile of the uncleaved 190K polyprotein and cleavage products, CFS samples were centrifuged and the repartition of these products in both soluble (S) and insoluble (P) fractions were analyzed. NS4-NS5-NS6-NS7, NS1-2-NS3 and the doublet of NS1-2 are mainly detected in the insoluble fraction, (**Fig.1B**, “P” and “S” samples, lanes wt). In contrast, NS7 is mainly recovered in the soluble fraction (**Fig.1B**, “P” and “S” samples, lanes wt). However, only the two higher bands of NS7 triplets are detected. The lower is likely partitioned between insoluble and soluble fractions and therefore below the detection limit of the Western-blot. The uncleaved 190K polyprotein is found both in the insoluble and the soluble fractions (**Fig.1B**, “P” and “S” samples, lanes C1147A). Note that the level of soluble uncleaved polyprotein is higher with the Sydney 2012 variant.

We went then further and tested the accessibility of the Strep-tags fused to the proteins. Supernatant samples were incubated with Strep-tag affinity beads and the amount of

bound proteins was evaluated by Western-blot. If uncleaved 190K polypeptide and NS7 are enriched in higher amounts than the other forms (NS4-NS5-NS6-NS7, NS1-2-NS3, NS6-NS7) (**Fig.1B**, “S-B” samples), all proteins interact with beads showing that Strep-tag could be used to purify proteins by affinity chromatography. Of note, only one band is detected for NS7. The second band could correspond to a protein conformation with a buried Strep-tag not able to bind the magnetic beads. Moreover, a very faint band is detected for the putative intermediate NS6-NS7 protein.

Taken together, these results show that the cell-free expression system can be used to produce and purify polyproteins, uncleaved and final proteins belonging to two variants (Den Haag 2006b and Sydney 2012) of huNoV genogroup II genotype 4 (GII.4).



**Figure 1: Proteolytic processing of human norovirus replication polyprotein in a wheat-germ cell-free expression system.**

**A.** Schematic view of the human norovirus (huNoV) polyprotein containing NS1-2, NS3, NS4, NS5, NS6 and NS7 domains. Blue triangles illustrate cleavage sites recognized by the protease NS6, giving rise to six end-products at the end of the proteolytic maturation process. The calculated molecular mass (shown in bold) of each protein is given in kilo Daltons. Strep-tags, represented by dark triangles, have been fused to both extremities of the polyprotein.

**B.** Analysis of huNoV polyproteins processing using wheat-germ cell-free expression system. We expressed the wt ORF1 of two GII.4 strains of huNoV, variants Den Haag 2006b (DH) and Sydney 2012 (Sydney). Protein samples were analyzed by SDS-PAGE followed by Western-blot using anti Strep-tag antibodies.

-: negative control performed in absence of transcript. wt: wild-type polyprotein where Pro domain is active and cleaves the polyprotein. C1147A: mutation of the catalytic Cys of Pro inhibiting polyprotein maturation. CFS is the total reaction mixture, P and S are respectively the insoluble and soluble proteins obtained after centrifugation during 30 min at 20,000 g of the CFS. S-B corresponds to the supernatant enriched through incubation with Strep-Tactin magnetic beads. 20  $\mu$ L of samples (CFS, P and S) and 10  $\mu$ L of S-B samples were loaded on the gel. The black arrow heads indicate the different proteins derived from the 190K proteolytic maturation process. Molecular weights markers are indicated on the left.

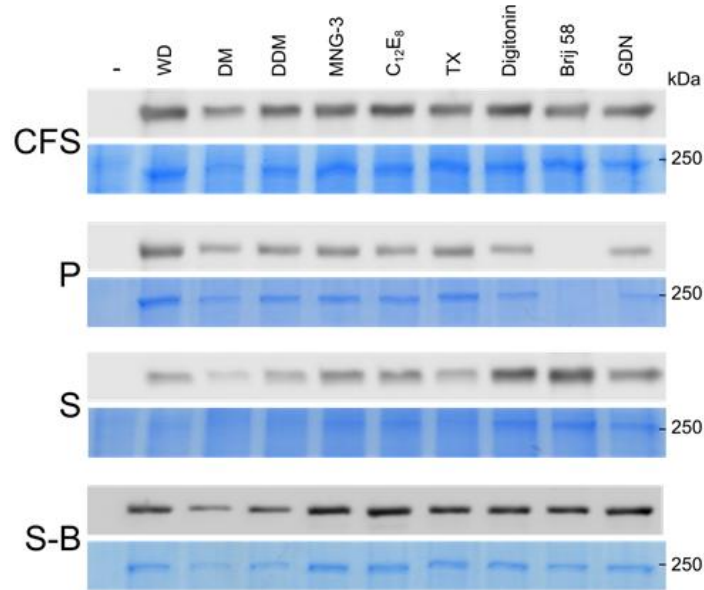
### **3.2. Improvement of the solubility of huNoV uncleavable mutant polyprotein by detergents**

As the C1147A mutant of 190K polyprotein expressed by the Sydney 2012 variant of huNoV GII.4 was expressed and enriched with the better efficiency, we were particularly interested in it. In order to increase the amount of soluble protein, we took advantage of the possibility to add detergents in the wheat-germ cell-free translation reaction mixture to produce detergent-solubilized forms of various proteins that are usually insoluble in cell-based systems [19]. We therefore conducted a detergent screening to identify which detergents could significantly increase the solubility of huNoV Sydney 2012 C1147A 190K polyprotein.

Uncleavable 190K polyprotein was produced in the absence or presence of different detergents (DM, n-decyl- $\beta$ -D-maltoside; DDM, n-dodecyl- $\beta$ -D-maltoside; MNG-3, lauryl maltose neopentyl glycol; C12E8, dodecyl octaethylene glycol ether; TX, Triton X-100; Digitonin; Brij-58; GDN, glyco-diosgenin) at a concentration of 0.1% (w/v). These detergents were chosen for their non-ionic properties and their low critical micelle concentrations (CMC) as previously described [19]. In the total cell-free sample (**Fig.2**, “CFS” samples), the amount of protein translated in presence of DM or GDN is slightly lower than in the absence of detergent (WD), showing that these detergents affect, in a different manner, the translation efficiency. For the other detergents (DDM, MNG-3, C12E8, TX and Digitonin), the amount of protein is equivalent in presence of 0.1% (w/v) and in the absence of detergent (WD).

The solubilization efficiency for each detergent was evaluated by comparing the amount of protein found in the insoluble (P) and soluble (S) fractions. In S fractions (**Fig.2**, “S” samples) and by comparison with the absence of detergent, while the amount of soluble uncleavable 190K polyprotein is lower in presence of DM, DDM and TX, the presence of other detergents (MNG-3, C12E8, Digitonin, Brij-58 and GDN) lead to higher amounts of soluble uncleavable 190K polyprotein with Digitonin and Brij-58 giving best results. All detergents allow the protein to bind to the Strep-tag affinity beads (**Fig.2**, “S-B” samples), the amount of bound protein being higher with MNG-3 and C12E8 as seen in the stained

acrylamide gel. These two detergents therefore appear to be the most promising to both solubilize and purify the huNoV uncleavable 190K polyprotein.



**Figure 2: Production of huNoV 190K polyprotein C1147A mutant by wheat-germ cell-free expression system in presence of several detergents.**

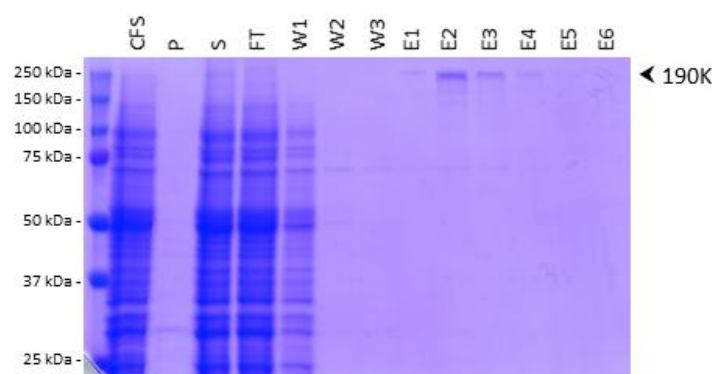
The huNoV 190K polyprotein from GII.4 Sydney strain and containing the catalytic mutation C1147A was expressed by wheat-germ cell-free expression system in absence (WD, without detergent) or in presence of eight different detergents: DM, n-decyl- $\beta$ -D-maltoside; DDM, n-dodecyl- $\beta$ -D-maltoside; MNG-3, lauryl maltose neopentyl glycol; C<sub>12</sub>E<sub>8</sub>, dodecyl octaethylene glycol ether; TX, Triton X-100; Digitonin; Brij58; GDN, glyco-diosgenin.

Protein samples were analyzed by Western-blot using anti Strep-tag antibodies (upper panel) or SDS-PAGE and Coomassie Blue staining (lower panel). For labels, see Fig.1. 20  $\mu$ L of each sample were loaded on the gel.

### 3.3. Purification of human norovirus replication mutant polyprotein

We then selected MNG-3 to solubilize and purify the C1147A mutant of the 190K polyprotein from the Sydney 2012 variant, and we scaled up by a factor of 250 the expression step. The uncleavable polyprotein was translated in medium-scale reactions using the wheat-germ cell-free expression system in the presence of 0.1% MNG-3. Following centrifugation, the supernatant of the total cell-free reaction mixture was purified by Strep-tag affinity chromatography, with buffers containing 0.02% MNG-3 (**Fig.3**). No polyprotein is detectable after SDS-PAGE and Blue Coomassie staining in flow-through (FT) and washes (W1-3). In contrast, one band corresponding to the uncleavable polyprotein is clearly seen in elution fractions, indicating that this step allowed for the efficient purification of the protein. Using this protocol, the amount of purified polyprotein in the whole elution fractions was estimated by UV-Vis spectrophotometry and gives a yield of around 0.25 mg of purified polyprotein per

mL of wheat-germ extract. This quantity is sufficient to set up conditions in order to conduct cryo-EM experiments while the search for crystallization conditions requires further increase in the amount of produced protein.



**Figure 3: Purification of huNoV 190K polyprotein C1147A mutant by affinity chromatography.** HuNoV 190K polyprotein mutant C1147A from GII.4 Sydney strain was produced and purified in presence of MNG-3. Protein samples were analyzed by SDS-PAGE and Coomassie Blue staining. CFS, total translation mixture; S, supernatant; P, pellet; FT, flow-through; W, wash fractions; E, elution fractions. 20  $\mu$ L of each sample were loaded on the gel.

### 3.4. Successful expression and maturation of TYMV polyprotein

*Turnip yellow mosaic virus* (TYMV) is a plant (+)RNA virus belonging to the genus Tymovirus. Organization and processing of TYMV polyprotein are strikingly similar to those of a number of other viruses from alphavirus-like superfamily, including human pathogens belonging to *Togaviridae* (e.g., *Rubella virus* (RUBV), *Chikungunya virus* (CHIKV)) or *Hepeviridae* (e.g., *Hepatitis E virus* (HEV)) [1], [27], [46]. It is therefore recognized as an excellent lower complexity study model for (+)RNA viruses of the alphavirus-like superfamily [27], [47].

TYMV genome has three ORFs, including ORF1 that codes for the replication polyprotein, the only viral protein necessary for replication [48]. The TYMV 206K polyprotein is constituted of 1844 amino acids residues with a molecular weight of  $\sim$  206kDa. From the N- to the C-terminus, TYMV 206K contains four distinct functional domains (*i.e.*, MT, PRO, HEL and POL) (**Fig.4A**), corresponding to methyltransferase, protease, NTPase/helicase and RdRp activities [49]. 206K is cleaved by the TYMV PRO domain [50], first between HEL and POL domains [51], [52], generating a 140K intermediate ( $\sim$ 140 kDa) and a 66K protein containing the RdRp domain (**Fig.4A**). 140K is subsequently cleaved between PRO and HEL domains to yield a 98K protein encompassing the MT and PRO domains and a 42K protein containing HEL domain [27], [53] (**Fig.4A**). While TYMV 206K is translated and subjected to the first proteolytic event in the cytoplasm of infected cells, the

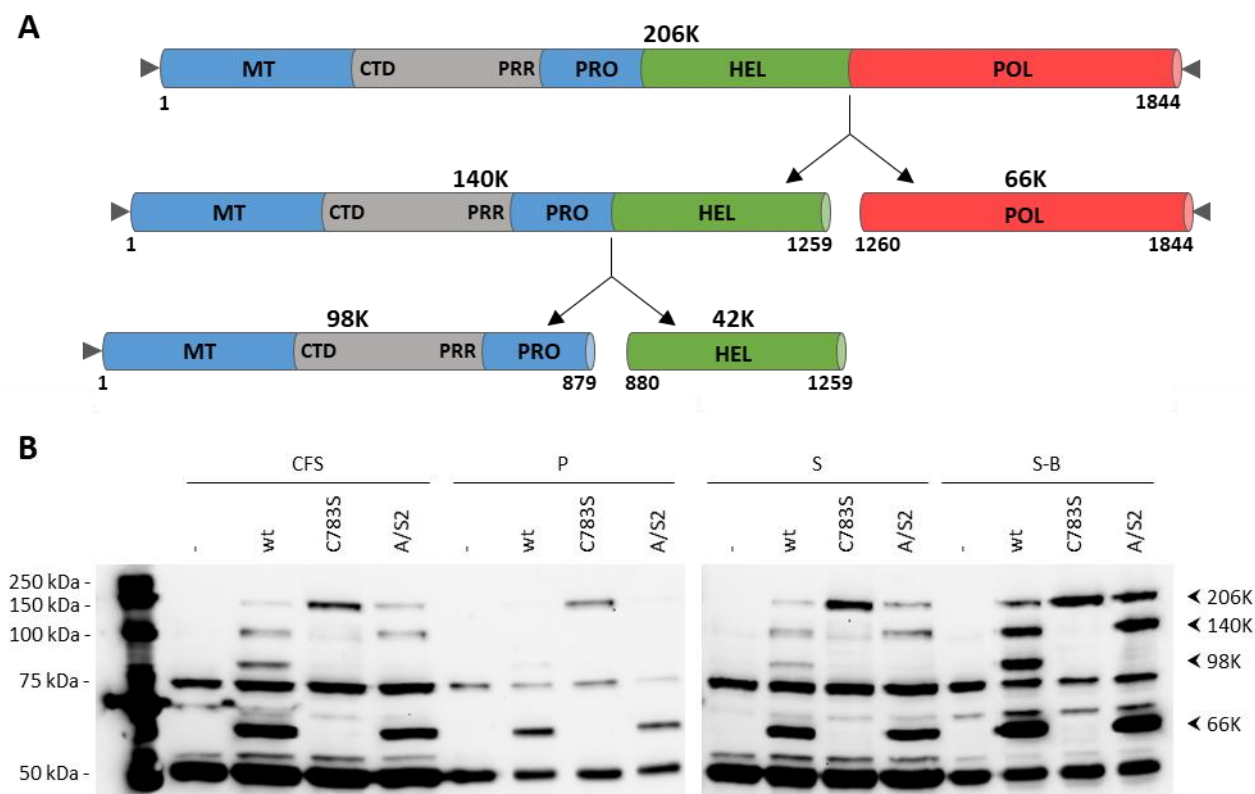
140K intermediate is targeted towards outer chloroplast membranes [26], inducing the formation of membrane invaginations that host the VRC. Through specific interactions, the 140K protein allows the recruitment and the targeting of the viral RdRp to the VRCs where replication of viral genome occurs [53].

In this work, in order to acquire structural data of TYMV 206K polyprotein and its maturation products, we produced the TYMV 206K polyprotein and its maturation products using the wheat-germ cell-free expression system. This approach had previously been used to characterize the TYMV polyprotein processing [50], [54] but there, the translation step was performed using RNA purified from virions as template. Here, we cloned the TYMV ORF1 sequence in the vector specifically designed for the cell-free expression system [25]. With the TYMV ORF1 transcript obtained by *in vitro* transcription, we then tested the expression of the wt full-length TYMV ORF1 in small-scale reactions. Since we tagged the 206K polyprotein with Strep-tags at both extremities (**Fig.4A**), we could detect translated proteins present in the cell-free reaction by Western-blotting using anti-Strep-tag antibodies. We first analyzed the cell-free sample (CFS). Compared to the negative control where no transcript is added, four bands are seen. Given the apparent molecular weights and what has been previously reported in *in vivo* and *in vitro* experiments including wheat-germ system [27], [50], [54], [55], these bands are attributed to the 206K precursor polyprotein, the 140K intermediate and the 66K and 98K end-products (**Fig.4B**, “CFS” samples). Note that, as the 42K protein is not flagged with any Strep-tag, it cannot be detected by Western-blot. This experiment thus demonstrates the functional integrity of the TYMV 206K when expressed by wheat-germ cell-free expression. This result is also confirmed by two mutated variants of the protein. Replacement of the catalytic Cys783 of the PRO domain by a Ser completely inhibits the proteolytic maturation of the polyprotein leading to the accumulation of an uncleavable form of the 206K precursor polyprotein (**Fig.4B**, “CFS” samples, lane C783S). Mutation of the PRO↓HEL cleavage site prevents the second cleavage step [27], leading to the disappearance of the band corresponding to the 98K end-product (**Fig.4B**, “CFS” samples, lane A/S2).

To our knowledge, the solubility of the resulting proteins has not been reported. We thus separated the soluble and insoluble fractions of the CFS by centrifugation. 206K and 66K are found partitioned between insoluble and soluble proteins while 140K and 98K appear mainly in the soluble fraction (**Fig.4B**, “P” and “S” samples).

Finally, we determined whether fused Strep-tags to 206K polyprotein could be used to purify proteins. We therefore incubated the CFS supernatant with Strep-tag affinity magnetic

beads. All proteins are efficiently enriched by this approach (**Fig.4B**, “S-B” samples), meaning that the partial purification of proteins by Strep-tag affinity chromatography is feasible. However, the amount of proteins is very low and the bands are not detected on SDS-PAGE after Coomassie blue staining.



**Figure 4: Expression and maturation of TYMV 206K replication polyprotein by wheat-germ cell-free expression system.**

**A.** Schematic representation of the TYMV 206K polyprotein, containing domains corresponding to methyltransferase (MT), protease (PRO), helicase (HEL) and RNA-polymerase (POL) activities. A proline-rich disordered region (PRR) is found upstream the PRO domain. The chloroplast targeting domain (CTD) is located downstream the MT domain. Strep-tags, represented by dark triangles, have been fused to both extremities of the polyprotein.

**B.** Analysis of TYMV 206K polyprotein expression using wheat-germ cell free expression system. Three variants of the protein were studied: the native full-length form (wt), a catalytic mutant where the catalytic cysteine of the PRO domain were replaced by a Ser (C783S) and a variant where the PRO↓HEL cleavage site (KRLGSG<sup>880</sup>) was replaced by seven Ala (mutant A/S2 as described in ref [27]). No transcript was added in the translation mixture corresponding to the negative control (-). 20 µL of each sample were loaded on the gel. For labels, see Fig.1.

### 3.5. Improvement of TYMV polymerase solubility by detergents

Since they replicate the RNA genome, RdRp encoded by (+)RNA viruses are essential factors of viral propagation. In this respect they are prime targets for development of new antiviral drugs. Knowledge of polymerases three-dimensional structure is obviously

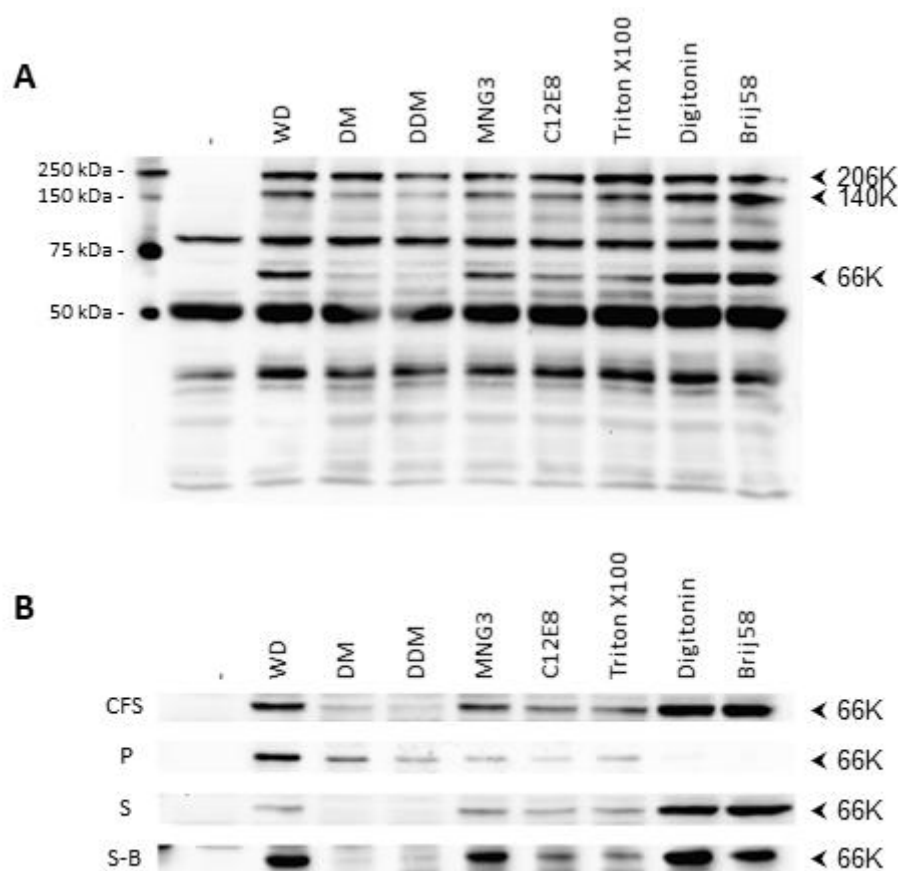


determinant to conduct drug-design projects. RdRp of (+)RNA viruses have been classified in three superfamilies [1], [56]. Structures of RdRp from (+)RNA viruses belonging to picornavirus-like superfamily (including *Caliciviridae* and *Picornaviridae*) and flavivirus-like superfamily (including *Flaviviridae*) have been solved [57] but there is a striking lack of structural information for the third alphavirus-like superfamily, to which the TYMV RdRp belongs [24], [57]. Actually, numerous attempts aiming to express a recombinant alphavirus-like RdRp in heterologous systems systematically failed because of the insolubility of the produced protein (unpublished results from our group and others). In some cases, the polymerase can be produced in high amounts and purified in denaturing conditions, as it has been done for TYMV 66K for example [58]. However, after attempts at refolding by removing chaotropic agents, 66K displays no enzymatic activity (unpublished data), which suggests a non-native structure. It seems therefore essential to produce 66K, and probably other alphavirus-like RdRp, in its native context to respect its native folding. Since we have shown above that TYMV 66K RdRp is partially found in the soluble fraction when expressed by wheat-germ cell-free expression system starting from the translation of the full-length polyprotein (**Fig.4B**), we then have attempted to increase the amount of soluble 66K by conducting a detergent screening as described above for huNoV polyprotein. This experiment was carried out with the variant A/S2 of TYMV 206K [27], which allows the translation of a 206K polyprotein that is cleaved to give the 66K end-product and an uncleavable form of 140K intermediate (**Fig.5A**, lane WD).

The behavior of 66K in the absence of detergent is slightly different to the observation made in the first experiment. Indeed, the ratio between insoluble and soluble protein is modified and 66K is now mainly found insoluble (**Fig.5B**, lane WD). WGEs contain residual lipids that contribute to slightly solubilize proteins (references [19], [20], [22] and **Fig.1B** (“S” samples)). However, the nature and the amount of these lipids might vary from one WGE batch to another and the use of two different WGE batches between the two series of experiments could therefore explain the difference of 66K solubility in absence of detergent.

Proteolytic maturation of TYMV polyprotein is strongly inhibited in the presence of 0.1% DM or DDM, leading to faint bands corresponding to 140K and 66K proteins (**Fig.5A**, lanes DM and DDM). Consequently, only a small amount of 66K is released through polyprotein maturation (**Fig.5B**, “CFS” samples, lanes DM and DDM) and the protein is essentially found in the insoluble fraction (**Fig.5B**, “P” samples, lanes DM and DDM). Presence of 0.1% MNG-3, C12E8 or TX in the translation reaction mixture does not drastically modify expression and maturation of 206K polyprotein (**Fig.5A**), nor the solubility

of 66K (**Fig.5B**, “S” samples). In contrast, the solubility of 66K is dramatically improved when polyprotein is translated in the presence of 0.1% Digitonin or Brij-58, which therefore appear as the best detergents to produce soluble 66K without inhibiting cell-free reaction nor interfering with Strep-tag affinity step (**Fig.5B**, “S” and “S-B” samples). We then tried to optimize the concentration of the two latter detergents by analyzing the solubility of 66K in a decreasing detergent concentration range. This search for the optimal concentration showed that 0.1% Digitonin or Brij-58 is the minimal concentration that can be used to produce the protein under a soluble form.



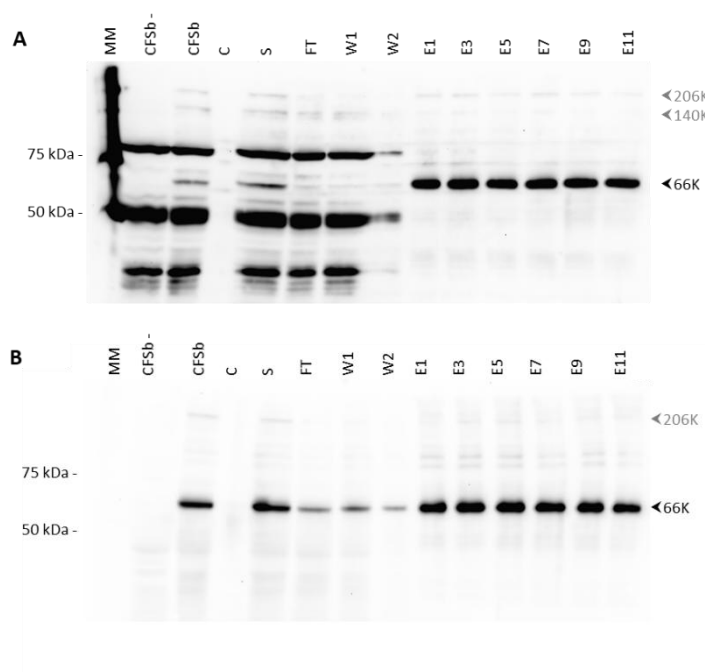
**Figure 5: Production of TYMV 66K RNA polymerase using wheat-germ cell-free expression system in presence of several detergents.**

**A.** The A/S2 mutant of the TYMV 206K polyprotein was expressed using wheat-germ cell-free expression system in absence (WD, without detergent) or in presence of eight different detergents. The Western-blot for cell-free samples (CFS) using anti Strep-tag antibodies is shown. 206K, 140K and 66K proteins are indicated with arrow heads.

**B.** 66K RNA polymerase obtained after spontaneous proteolytic processing of the polyprotein was detected in each sample (CFS as in A, P, S and S-B) by Western-blot using anti-Strep-tag antibodies. 20 µL of each sample were loaded on the gel. For conditions and labels, see Fig. 1 and 2.

### 3.6. Partial purification of TYMV RNA-dependent RNA-polymerase

In the next step, we scaled up by a factor of 150 the production and the purification of TYMV 66K. For that purpose, the polyprotein carrying the mutation A/S2 was translated in medium-scale reactions using wheat-germ cell-free expression system in the presence of 0.1% Brij-58. The 206K polyprotein was translated and partially processed, leading to the production of soluble 66K end-product (**Fig.6**, lanes CFS, P and S). The purification procedure of 66K has to be adapted to deal with the low affinity of the protein to the Strep-tag affinity resin. Indeed, in order to maximize the amount of Strep-tagged 66K bound to the resin, the binding step was extended to 21h. This unusually long time was necessary to ensure that almost no protein was lost in the flow-through or in washing fractions (**Fig.6**, lanes FT and W1-2). Again, elution step had to be optimized and elution of bound 66K was realized by mixing the resin and the elution buffer during 24h under agitation. The protein was thus completely eluted from the resin (**Fig.6**, lanes E1-11). However, whereas TYMV 66K is easily detected by Western-blot using anti-Strep-tag or anti-66K antibodies (**Fig.6**), the band corresponding to the protein is not visible after Coomassie blue staining of a polyacrylamide gel, indicating that the amount of purified protein is low, below 100 ng. Thus, this protocol is not suitable to obtain the *ca* milligram amount required for structural characterization. Nevertheless, we show here that the wheat-germ cell-free expression system allows to produce the soluble TYMV 66K polymerase. We therefore establish a proof-of-concept that this expression system is a very promising tool to produce other alphavirus-like RdRp and therefore obtain for the first time structural information for these polymerases.



**Figure 6: Purification of TYMV 66K RNA-polymerase by affinity chromatography.**

The A/S2 mutant of the TYMV 206K polyprotein was expressed using wheat-germ cell-free expression system in presence of 0.1% Brij58. The 66K RNA polymerase produced through polyprotein maturation was then partially purified by Strep-tag affinity chromatography. Each fraction collected during the purification was analyzed by Western-blot using anti-Strep-tag antibodies (A) or anti-66K antibodies (B): CFSb -: total translation mixture without transcript, treated with benzonase; CFSb, total translation mixture treated with benzonase (CFSb); P, pellet; S, supernatant; FT, flow-through; W, wash; E, elution; W', final wash after elution. 20  $\mu$ L (A) or 10  $\mu$ L (B) of each sample were loaded on the gel.

**4. Discussion**

Polyproteins encoded by single stranded positive- sense RNA ((+)RNA) viruses constitute the cornerstone of viral replication. Indeed, they harbor all the components that are necessary for viral replication, especially the RdRp that replicates viral RNA genome, and they are directly involved in the formation of viral replication complexes (RCs). Viral polyproteins are usually short-lived proteins. They are processed by a dedicated and finely regulated proteolytic pathway to generate intermediate proteolytic products and finally end-products. The maturation process is extremely important in the course of viral replication. Indeed, it controls enzymatic activities of each domain encompassed in the polyprotein. For example, the synthesis of RNA progeny by RdRp is highly regulated during the infection cycle, from viral entry and RNA genome release to encapsidation and egress from cells. The viral RdRp uses first the positive-sense single stranded RNA molecule as a template to synthesize the complementary minus-strand RNA. This negative-sense RNA molecule is then used by RdRp to synthesize new plus-strand RNA molecules that are later encapsidated in new viral particles. It has been shown in several independent reports that the successive RdRp strand-preferences are triggered for several viruses by the successive proteolytic cleavages (reviewed in [24]). Nevertheless, the correlation between viral polyprotein cleavage and the switch from negative- to positive-sense RNA synthesis remains poorly understood. In addition, the polyprotein maturation process controls the cellular localization of viral proteins. Indeed, viral proteins contain structural elements that allow their targeting and association to intracellular membranes, and that trigger membrane rearrangements [3], [59], [60]. These rearrangements may depend on cleavage intermediates as well as end-products, and the proteolytic steps giving rise to these intermediates are a pre-requisite for the formation of viral RCs [60], [61]. Since RNA is replicated inside the viral RCs, the polyprotein maturation process eventually tightly regulates both the temporal and the spatial aspects of viral RNA genome replication.

Viral polyproteins and their maturation products are clearly key factors of viral replication. Moreover, as cleavage products' functions are often distinct from those of the precursor polyproteins, it remains necessary to characterize all these proteins by biochemical and structural approaches. The structures of recombinant domains such as RNA polymerases and proteases have been solved but those of polyproteins or large cleavage intermediates are still poorly documented [15]. Indeed, due to their multi-domain architecture, their flexibility and their ability to interact with membranes, viral polyproteins are highly difficult to produce in usual systems and their synthesis usually requires specific expression systems such as mammalian cells, insect cells or plants. Of note, Kobayashi *et al.* have recently developed a very attractive and promising baculovirus-silkworm pupae system that allowed them to obtain 1.5 mg of well-folded and active *Tomato mosaic virus* (ToMV) 130K replication protein [62]. However, biological pathways of living cells or organisms used as expression platforms can be strongly affected by the overexpression of heterologous proteins, especially by foreign membrane proteins that interact strongly with host intracellular membranes. We were therefore interested in acellular expression systems, and we chose the wheat-germ cell-free expression system that presents many advantages that are discussed below. Of interest, in addition to being particularly adapted for large eukaryotic proteins [44], this system has proven its efficiency for notoriously difficult-to-express proteins, especially membrane proteins [18], [20]–[23]. Moreover it is highly modular and can be easily supplemented by many agents such as detergents [19], [28], labeled amino acids [21], [53], or even liposomes to reconstitute a membranous environment [44]. In addition, in order to facilitate the purification process of the proteins translated by this system, affinity tags can be easily added during the cloning step in the dedicated pEU vector [25]. Of note, the popular His-tags that are efficient in other expression systems are not satisfactory with the wheat-germ cell-free expression system. This is due to binding of many WGE proteins to nickel resins that requires a strong optimization of the purification procedure when using a His-tag (unpublished data and reference [23]) or pre-treat the extract with a nickel resin [25]. It is therefore preferable to use Strep-tags or FLAG-tags as described by Novikova *et al.* [23]. Strep-tags are also convenient for detection of the protein of interest in the initial tests, although several proteins in the extract also react with anti-Strep-tag antibodies. This sometimes also leads to artifacts when a contaminating band may be detected only in some experiments or lanes and be mistaken for a possible cleavage product. It is necessary to check reproducibility and always helpful to cross-check with different antibodies, *e.g.* against different tags or with a specific antibody as soon as it becomes available. Finally, some post-translational modifications such

as phosphorylation have been reported for proteins translated in the WGE cell-free expression system [45], [63], [64]. However some other modifications such as glycosylation do not occur in this system [44], which can affect the conformation and/or the function of the protein of interest.

The aim of our study was to determine whether the wheat-germ cell-free expression system could be suitable to produce milligram amounts of soluble full-length viral polyproteins in order to consider structural biology projects. In this work, we have studied polyproteins naturally encoded by (+)RNA viruses. The genome of these viruses being composed of positive-sense single-stranded RNA, it behaves as a cellular mRNA-like and is directly translated by the infected host translation machinery without any RNA synthesis or splicing step. As a consequence, the *in vitro* transcription step can be skipped and the *in vitro* translation step can be realized using RNA viral genome purified from infected host. However, note that laboratory biosafety principles have to be applied to handle genomes of human pathogenic viruses, which is not the case with plant viruses such as TYMV. Moreover, this approach has its limitations if lethal mutations are introduced in the viral genome because it is not replicated and its purification from infected host thus becomes impossible. In this case, the cDNA corresponding to the wild-type or mutated viral genome can be inserted in the appropriate vector as described [25]. Since (+)RNA viruses infect eukaryotic hosts, there is no codon usage bias with the wheat-germ translation machinery, a point that considerably simplifies the cloning strategy. In this study, we have cloned into the pEU-E01-MCS vector the cDNA corresponding to ORF1 of norovirus genogroup II genotype 4 (GII.4) and 206K-ORF of TYMV. We then proceeded to *in vitro* transcription step followed by *in vitro* translation by the wheat-germ cell-free expression system, leading to the expression of the two polyproteins. Of note, the molecular weight of these viral polyproteins is around 200 kDa, a result that confirms that one strength of the wheat-germ cell-free expression system compared to other expression systems is the possibility to produce large proteins [44]. Importantly we could observe the self-cleavage of the two wild-type polyproteins through their active protease domains, indicating that they are fully functional. Moreover, as the cleavage products that are observed correspond perfectly to those obtained in *in vivo* experiments using infected cells, we conclude that the wheat-germ cell-free expression system provides a very suitable method to produce full-length viral polyproteins and to recapitulate their proteolytic maturation process. Importantly, mutations can be readily introduced in coding sequences of viral polyproteins. For example, replacement of catalytic Cys of the protease domain leads to complete inhibition of the maturation process in both cases shown

here, and the translated polyprotein remains in an uncleavable form. This creates tremendous opportunities to obtain high resolution data of large viral polyproteins and document their conformation before their processing. This is particularly relevant to polyproteins such as TYMV's, whose first cleavage triggers a cascade of molecular events starting the targeting to and remodeling of specific membranes. The role(s) of the first cleavage(s) of norovirus polyprotein are more complex and less well established, but at any rate optimization of the protocol to produce and purify the uncleavable polyprotein encoded by HuNoV is currently under way.

Another advantage of the viral polyprotein self-processing is the production of cleavage intermediates and end-products with their authentic N-terminal extremity. Indeed, the cloning step of these proteins requires to add an initiator methionine (iMet), and sometimes several linker residues, to the N-ter extremity of the recombinant viral domain. Depending on the resulting N-terminal sequence and the cellular translation system, the iMet is excised or not, and the nature of subsequent N-terminal co- and post-translational modifications varies [65]. This cascade of events leads to a protein with an N-ter extremity that can be very different to the one obtained in natural conditions during viral infection. These unnatural protein modifications can alter the function of viral protein. In addition, in some cases, overexpressed recombinant proteins present heterogeneous N-ter extremities, which is not compatible with structural studies. One solution to this issue is to add an N-ter fusion protein followed by a proteolytic cleavage site in order to generate the true N-ter extremity of the recombinant protein after cleavage. However, the fusion protein can mask the ligand binding site or the active site of the recombinant protein. Moreover, the cleavage site is not always accessible to the protease and the cleavage is therefore not possible. In contrast, the self-cleavage of viral polyproteins when expressed by cell-free wheat-germ expression system surely produces cleavage intermediates and end-products with their authentic ligand accessible surface and N-terminal extremity.

Third, viral polyproteins' maturation products may be misfolded when expressed in recombinant form because they are translated in an unnatural context. In contrast, the translation of a full-length viral polyprotein, from its natural N- to C-terminal extremities, favors the proper folding of all its domains. Consequently, the self-cleavage of viral polyproteins in the translation mix allows the release of well-folded maturation products. With this proof-of-concept approach, we have successfully produced the soluble RdRp encoded by TYMV, a striking progress that presents a first step towards structural work of an RdRp encoded by viruses from the alphavirus-like superfamily. Furthermore, since they

interact with intracellular membranes, viral polyproteins contain numerous hydrophobic surfaces and tend to be insoluble when expressed in heterologous systems. In wheat-germ cell-free expression system, the solubility of translated proteins is easily managed by the use of detergents. Such an approach has proved its efficiency with notoriously difficult-to-express transmembrane proteins [19]–[21], [53]. In this study, we show that the use of detergents is also convenient for viral proteins that are not inserted to membranes but associated to membranes, again such as TYMV RdRp. Our experience (this study and previous works cited above) indicates that there is no universal detergent that would be suitable for all proteins. This requires an empirical trial-and-error process to identify the correct detergent for each new protein (*i.e.*, a detergent that provides a compromise between the solubility of the protein and its affinity for the chromatography affinity resin if a tag is fused), something a cell-free system is particularly suited to (see above).

In summary, this study provides a proof-of-concept for the efficient expression of soluble and fully functional virally-encoded replication polyproteins and the demonstration that mutations can be easily introduced to produce uncleavable proteins or particular maturation products. Even though expression yield and protein purity need to be optimized to progress towards biochemical and structural characterization, polyprotein in the wheat-germ cell-free expression system represents a significant step forward in understanding the functional roles of the proteolytic maturation process.

### **Conflicts of interest**

The authors declare that they have no conflict of interest.

### **Acknowledgments**

We thank V. Lohmann and I. Jupin for the gift of plasmids encoding noroviruses and TYMV polyproteins, respectively, from which plasmids used in this work were constructed. Anti-66K antibodies were a kind gift of I. Jupin. Benzonase was a generous gift from J.M. Betton. We thank R. Ghouil for contributing to initial stages of this project. SB's and AB's groups are funded by the Agence Nationale de la Recherche sur le Sida et les hépatites virales (France REcherche Nord et sud Sida-hiv Hépatites) and JH was funded by an ANRS postdoctoral fellowship during part of this work.

### **References**



- [1] E. V. Koonin, V. V. Dolja, and M. Krupovic, "Origins and evolution of viruses of eukaryotes: The ultimate modularity," *Virology*, vol. 479–480, pp. 2–25, May 2015, doi: 10.1016/j.virol.2015.02.039.
- [2] C. U. Hellen, H. G. Kräusslich, and E. Wimmer, "Proteolytic processing of polyproteins in the replication of RNA viruses," *Biochemistry*, vol. 28, no. 26, pp. 9881–9890, Dec. 1989.
- [3] S. Y. Doerflinger *et al.*, "Membrane alterations induced by nonstructural proteins of human norovirus," *PLoS Pathog.*, vol. 13, no. 10, p. e1006705, Oct. 2017, doi: 10.1371/journal.ppat.1006705.
- [4] D. Paul, V. Madan, and R. Bartenschlager, "Hepatitis C virus RNA replication and assembly: living on the fat of the land," *Cell Host Microbe*, vol. 16, no. 5, pp. 569–579, Nov. 2014, doi: 10.1016/j.chom.2014.10.008.
- [5] H. G. Kräusslich and E. Wimmer, "Viral proteinases," *Annu. Rev. Biochem.*, vol. 57, pp. 701–754, 1988, doi: 10.1146/annurev.bi.57.070188.003413.
- [6] V. E. Spall, M. Shanks, and G. P. Lomonosoff, "Polyprotein Processing as a Strategy for Gene Expression in RNA Viruses," *Seminars in Virology*, vol. 8, no. 1, pp. 15–23, Feb. 1997, doi: 10.1006/smv.1997.0102.
- [7] M. R. Herod *et al.*, "Genetic economy in picornaviruses: Foot-and-mouth disease virus replication exploits alternative precursor cleavage pathways," *PLoS Pathog.*, vol. 13, no. 10, p. e1006666, Oct. 2017, doi: 10.1371/journal.ppat.1006666.
- [8] A. Spear, S. A. Ogram, B. J. Morasco, L. E. Smerage, and J. B. Flanagan, "Viral precursor protein P3 and its processed products perform discrete and essential functions in the poliovirus RNA replication complex," *Virology*, vol. 485, pp. 492–501, Nov. 2015, doi: 10.1016/j.virol.2015.07.018.
- [9] C. Bussetta and K. H. Choi, "Dengue virus nonstructural protein 5 adopts multiple conformations in solution," *Biochemistry*, vol. 51, no. 30, pp. 5921–5931, Jul. 2012, doi: 10.1021/bi300406n.
- [10] I. M. Moustafa, D. W. Gohara, A. Uchida, N. Yennawar, and C. E. Cameron, "Conformational Ensemble of the Poliovirus 3CD Precursor Observed by MD Simulations and Confirmed by SAXS: A Strategy to Expand the Viral Proteome?," *Viruses*, vol. 7, no. 11, pp. 5962–5986, Nov. 2015, doi: 10.3390/v7112919.
- [11] W. G. Saw, A. Pan, M. S. Subramanian Manimekalai, A. Grüber, and G. Grüber, "Structure and flexibility of non-structural proteins 3 and -5 of Dengue- and Zika viruses in solution," *Prog. Biophys. Mol. Biol.*, Aug. 2018, doi: 10.1016/j.pbiomolbio.2018.08.008.
- [12] Y. Zhao *et al.*, "Flexibility of NS5 Methyltransferase-Polymerase Linker Region Is Essential for Dengue Virus Replication," *J. Virol.*, vol. 89, no. 20, pp. 10717–10721, Oct. 2015, doi: 10.1128/JVI.01239-15.
- [13] L. L. Marcotte *et al.*, "Crystal structure of poliovirus 3CD protein: virally encoded protease and precursor to the RNA-dependent RNA polymerase," *J. Virol.*, vol. 81, no. 7, pp. 3583–3596, Apr. 2007, doi: 10.1128/JVI.02306-06.
- [14] G. Shin, S. A. Yost, M. T. Miller, E. J. Elrod, A. Grakoui, and J. Marcotrigiano, "Structural and functional insights into alphavirus polyprotein processing and pathogenesis," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 41, pp. 16534–16539, Oct. 2012, doi: 10.1073/pnas.1210418109.
- [15] S. A. Yost and J. Marcotrigiano, "Viral precursor polyproteins: keys of regulation from replication to maturation," *Curr Opin Virol*, vol. 3, no. 2, pp. 137–142, Apr. 2013, doi: 10.1016/j.coviro.2013.03.009.
- [16] E. Emmott *et al.*, "Polyprotein processing and intermolecular interactions within the viral replication complex spatially and temporally control norovirus protease activity," *J Biol Chem*, vol. 294, no. 11, pp. 4259–4271, Mar. 2019, doi: 10.1074/jbc.RA118.006780.
- [17] T. Sawasaki, T. Ogasawara, R. Morishita, and Y. Endo, "A cell-free protein synthesis system for high-throughput proteomics," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, no. 23, pp. 14652–14657, Nov. 2002, doi: 10.1073/pnas.232580399.

- [18] M.-L. Fogeron *et al.*, "Functional expression, purification, characterization, and membrane reconstitution of non-structural protein 2 from hepatitis C virus," *Protein Expr. Purif.*, vol. 116, pp. 1–6, Dec. 2015, doi: 10.1016/j.pep.2015.08.027.
- [19] M.-L. Fogeron *et al.*, "Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins," *Protein Expr. Purif.*, vol. 105, pp. 39–46, Jan. 2015, doi: 10.1016/j.pep.2014.10.003.
- [20] M.-L. Fogeron *et al.*, "Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus," *J. Biomol. NMR*, vol. 65, no. 2, pp. 87–98, Jun. 2016, doi: 10.1007/s10858-016-0040-2.
- [21] V. Jirasko *et al.*, "Proton-Detected Solid-State NMR of the Cell-Free Synthesized  $\alpha$ -Helical Transmembrane Protein NS4B from Hepatitis C Virus," *Chembiochem*, Dec. 2019, doi: 10.1002/cbic.201900765.
- [22] G. David *et al.*, "Structural Studies of Self-Assembled Subviral Particles: Combining Cell-Free Expression with 110 kHz MAS NMR Spectroscopy," *Angewandte Chemie International Edition*, vol. 57, no. 17, pp. 4787–4791, 2018, doi: 10.1002/anie.201712091.
- [23] I. V. Novikova *et al.*, "Protein structural biology using cell-free platform from wheat germ," *Advanced Structural and Chemical Imaging*, vol. 4, no. 1, p. 13, Nov. 2018, doi: 10.1186/s40679-018-0062-9.
- [24] M. K. Pietilä, K. Hellström, and T. Ahola, "Alphavirus polymerase and RNA replication," *Virus Res.*, Jan. 2017, doi: 10.1016/j.virusres.2017.01.007.
- [25] K. Takai, T. Sawasaki, and Y. Endo, "Practical cell-free protein synthesis system using purified wheat embryos," *Nat Protoc*, vol. 5, no. 2, pp. 227–238, Feb. 2010, doi: 10.1038/nprot.2009.207.
- [26] D. Prod'homme, A. Jakubiec, V. Tournier, G. Drugeon, and I. Jupin, "Targeting of the turnip yellow mosaic virus 66K replication protein to the chloroplast envelope is mediated by the 140K protein," *J. Virol.*, vol. 77, no. 17, pp. 9124–9135, Sep. 2003.
- [27] A. Jakubiec, G. Drugeon, L. Camborde, and I. Jupin, "Proteolytic processing of turnip yellow mosaic virus replication proteins and functional impact on infectivity," *J. Virol.*, vol. 81, no. 20, pp. 11402–11412, Oct. 2007, doi: 10.1128/JVI.01428-07.
- [28] M.-L. Fogeron, A. Badillo, F. Penin, and A. Böckmann, "Wheat Germ Cell-Free Overexpression for the Production of Membrane Proteins," *Methods Mol. Biol.*, vol. 1635, pp. 91–108, 2017, doi: 10.1007/978-1-4939-7151-0\_5.
- [29] U. K. Laemmli, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature*, vol. 227, no. 5259, pp. 680–685, Aug. 1970, doi: 10.1038/227680a0.
- [30] D. Prod'homme, S. Le Panse, G. Drugeon, and I. Jupin, "Detection and subcellular localization of the turnip yellow mosaic virus 66K replication protein in infected cells," *Virology*, vol. 281, no. 1, pp. 88–101, Mar. 2001, doi: 10.1006/viro.2000.0769.
- [31] T. N. Hoa Tran, E. Trainor, T. Nakagomi, N. A. Cunliffe, and O. Nakagomi, "Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: global distribution of genogroups, genotypes and GII.4 variants," *J. Clin. Virol.*, vol. 56, no. 3, pp. 185–193, Mar. 2013, doi: 10.1016/j.jcv.2012.11.011.
- [32] J. L. Hyde and J. M. Mackenzie, "Subcellular localization of the MNV-1 ORF1 proteins and their potential roles in the formation of the MNV-1 replication complex," *Virology*, vol. 406, no. 1, pp. 138–148, Oct. 2010, doi: 10.1016/j.virol.2010.06.047.
- [33] J. L. Hyde, S. V. Sosnovtsev, K. Y. Green, C. Wobus, H. W. Virgin, and J. M. Mackenzie, "Mouse Norovirus Replication Is Associated with Virus-Induced Vesicle Clusters Originating from Membranes Derived from the Secretory Pathway," *J Virol*, vol. 83, no. 19, pp. 9709–9719, Oct. 2009, doi: 10.1128/JVI.00600-09.
- [34] B. T. Cotton *et al.*, "The Norovirus NS3 Protein Is a Dynamic Lipid- and Microtubule-Associated Protein Involved in Viral RNA Replication," *J. Virol.*, vol. 91, no. 3, Feb. 2017, doi: 10.1128/JVI.02138-16.

- [35] G. Belliot, S. V. Sosnovtsev, T. Mitra, C. Hammer, M. Garfield, and K. Y. Green, "In vitro proteolytic processing of the MD145 norovirus ORF1 nonstructural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells," *J. Virol.*, vol. 77, no. 20, pp. 10957–10974, Oct. 2003, doi: 10.1128/jvi.77.20.10957-10974.2003.
- [36] S. J. Blakeney, A. Cahill, and P. A. Reilly, "Processing of Norwalk virus nonstructural proteins by a 3C-like cysteine proteinase," *Virology*, vol. 308, no. 2, pp. 216–224, Apr. 2003.
- [37] M. E. Hardy, T. J. Crone, J. E. Brower, and K. Ettayebi, "Substrate specificity of the Norwalk virus 3C-like proteinase," *Virus Res.*, vol. 89, no. 1, pp. 29–39, Oct. 2002.
- [38] B. Liu, I. N. Clarke, and P. R. Lambden, "Polyprotein processing in Southampton virus: identification of 3C-like protease cleavage sites by in vitro mutagenesis," *J. Virol.*, vol. 70, no. 4, pp. 2605–2610, Apr. 1996.
- [39] B. L. Liu, G. J. Viljoen, I. N. Clarke, and P. R. Lambden, "Identification of further proteolytic cleavage sites in the Southampton calicivirus polyprotein by expression of the viral protease in *E. coli*," *J. Gen. Virol.*, vol. 80 ( Pt 2), pp. 291–296, Feb. 1999, doi: 10.1099/0022-1317-80-2-291.
- [40] E. L. Seah, J. A. Marshall, and P. J. Wright, "Open reading frame 1 of the Norwalk-like virus Camberwell: completion of sequence and expression in mammalian cells," *J. Virol.*, vol. 73, no. 12, pp. 10531–10535, Dec. 1999.
- [41] S. V. Sosnovtsev *et al.*, "Cleavage map and proteolytic processing of the murine norovirus nonstructural polyprotein in infected cells," *J. Virol.*, vol. 80, no. 16, pp. 7816–7831, Aug. 2006, doi: 10.1128/JVI.00532-06.
- [42] Z. Lateef, G. Gimenez, E. S. Baker, and V. K. Ward, "Transcriptomic analysis of human norovirus NS1-2 protein highlights a multifunctional role in murine monocytes," *BMC Genomics*, vol. 18, Jan. 2017, doi: 10.1186/s12864-016-3417-4.
- [43] E. S. Baker, S. R. Luckner, K. L. Krause, P. R. Lambden, I. N. Clarke, and V. K. Ward, "Inherent Structural Disorder and Dimerisation of Murine Norovirus NS1-2 Protein," *PLOS ONE*, vol. 7, no. 2, p. e30534, Feb. 2012, doi: 10.1371/journal.pone.0030534.
- [44] M. Harbers, "Wheat germ systems for cell-free protein expression," *FEBS Lett.*, vol. 588, no. 17, pp. 2762–2773, Aug. 2014, doi: 10.1016/j.febslet.2014.05.061.
- [45] G. David *et al.*, "Phosphorylation and Alternative Translation on Wheat Germ Cell-Free Protein Synthesis of the DHBV Large Envelope Protein," *Front Mol Biosci*, vol. 6, p. 138, 2019, doi: 10.3389/fmolb.2019.00138.
- [46] E. V. Koonin and V. V. Dolja, "Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences," *Crit. Rev. Biochem. Mol. Biol.*, vol. 28, no. 5, pp. 375–430, 1993, doi: 10.3109/10409239309078440.
- [47] M. W. van der Heijden and J. F. Bol, "Composition of alphavirus-like replication complexes: involvement of virus and host encoded proteins," *Arch. Virol.*, vol. 147, no. 5, pp. 875–898, May 2002, doi: 10.1007/s00705-001-0773-3.
- [48] J. J. Weiland and T. W. Dreher, "Infectious TYMV RNA from cloned cDNA: effects in vitro and in vivo of point substitutions in the initiation codons of two extensively overlapping ORFs," *Nucleic Acids Res.*, vol. 17, no. 12, pp. 4675–4687, Jun. 1989, doi: 10.1093/nar/17.12.4675.
- [49] T. W. Dreher, "Turnip yellow mosaic virus: transfer RNA mimicry, chloroplasts and a C-rich genome," *Mol. Plant Pathol.*, vol. 5, no. 5, pp. 367–375, Sep. 2004, doi: 10.1111/j.1364-3703.2004.00236.x.
- [50] M. D. Morch, G. Dugeon, P. Szafranski, and A. L. Haenni, "Proteolytic origin of the 150-kilodalton protein encoded by turnip yellow mosaic virus genomic RNA," *J. Virol.*, vol. 63, no. 12, pp. 5153–5158, Dec. 1989.
- [51] K. L. Bransom, S. E. Wallace, and T. W. Dreher, "Identification of the cleavage site recognized by the turnip yellow mosaic virus protease," *Virology*, vol. 217, no. 1, pp. 404–406, Mar. 1996, doi: 10.1006/viro.1996.0131.
- [52] G. Kadaré, M. Rozanov, and A. L. Haenni, "Expression of the turnip yellow mosaic virus proteinase in *Escherichia coli* and determination of the cleavage site within the 206 kDa

- protein," *J. Gen. Virol.*, vol. 76 ( Pt 11), pp. 2853–2857, Nov. 1995, doi: 10.1099/0022-1317-76-11-2853.
- [53] A. Jakubiec *et al.*, "Assembly of turnip yellow mosaic virus replication complexes: interaction between the proteinase and polymerase domains of the replication proteins," *J. Virol.*, vol. 78, no. 15, pp. 7945–7957, Aug. 2004, doi: 10.1128/JVI.78.15.7945-7957.2004.
  - [54] W. Zagorski, M. D. Morch, and A. L. Haenni, "Comparison of three different cell-free systems for turnip yellow mosaic virus RNA translation," *Biochimie*, vol. 65, no. 2, pp. 127–133, Feb. 1983.
  - [55] D. Prod'homme, S. Le Panse, G. Dugeon, and I. Jupin, "Detection and subcellular localization of the turnip yellow mosaic virus 66K replication protein in infected cells," *Virology*, vol. 281, no. 1, pp. 88–101, Mar. 2001, doi: 10.1006/viro.2000.0769.
  - [56] E. V. Koonin, "The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses," *J. Gen. Virol.*, vol. 72 ( Pt 9), pp. 2197–2206, Sep. 1991, doi: 10.1099/0022-1317-72-9-2197.
  - [57] S. Venkataraman, B. V. L. S. Prasad, and R. Selvarajan, "RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution," *Viruses*, vol. 10, no. 2, Feb. 2018, doi: 10.3390/v10020076.
  - [58] F. Héricourt, S. Blanc, V. Redeker, and I. Jupin, "Evidence for phosphorylation and ubiquitinylation of the turnip yellow mosaic virus RNA-dependent RNA polymerase domain expressed in a baculovirus-insect cell system," *Biochem. J.*, vol. 349, no. Pt 2, pp. 417–425, Jul. 2000.
  - [59] L. Moriceau, L. Jomat, S. Bressanelli, C. Alcaide-Loridan, and I. Jupin, "Identification and Molecular Characterization of the Chloroplast Targeting Domain of Turnip yellow mosaic virus Replication Proteins," *Front Plant Sci*, vol. 8, p. 2138, 2017, doi: 10.3389/fpls.2017.02138.
  - [60] D. Oudshoorn *et al.*, "Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with Coronaviral RNA Replication," *MBio*, vol. 8, no. 6, 21 2017, doi: 10.1128/mBio.01658-17.
  - [61] A. Lulla, V. Lulla, and A. Merits, "Macromolecular assembly-driven processing of the 2/3 cleavage site in the alphavirus replicase polyprotein," *J. Virol.*, vol. 86, no. 1, pp. 553–565, Jan. 2012, doi: 10.1128/JVI.05195-11.
  - [62] C. Kobayashi *et al.*, "Purification and functional characterization of tomato mosaic virus 130K protein expressed in silkworm pupae using a baculovirus vector," *Protein Expr. Purif.*, vol. 154, pp. 85–90, Feb. 2019, doi: 10.1016/j.pep.2018.10.001.
  - [63] S. Nakamura, "Possible role of phosphorylation in the function of chicken MyoD1," *J. Biol. Chem.*, vol. 268, no. 16, pp. 11670–11677, Jun. 1993.
  - [64] J. O. Langland, L. A. Langland, K. S. Browning, and D. A. Roth, "Phosphorylation of plant eukaryotic initiation factor-2 by the plant-encoded double-stranded RNA-dependent protein kinase, pPKR, and inhibition of protein synthesis in vitro," *J. Biol. Chem.*, vol. 271, no. 8, pp. 4539–4544, Feb. 1996, doi: 10.1074/jbc.271.8.4539.
  - [65] C. Giglione, S. Fieulaine, and T. Meinel, "N-terminal protein modifications: Bringing back into play the ribosome," *Biochimie*, vol. 114, pp. 134–146, Jul. 2015, doi: 10.1016/j.biochi.2014.11.008.