

# Etude structurale de protéines virales impliquées dans la réplication et régulation du cycle de multiplication d'un virus de plante

Charlotte Robin

▶ **To cite this version:**

Charlotte Robin. Etude structurale de protéines virales impliquées dans la réplication et régulation du cycle de multiplication d'un virus de plante. Sciences agricoles. Université Paris Sud - Paris XI, 2011. Français. NNT : 2011PA114853 . tel-00661702

**HAL Id: tel-00661702**

**https://theses.hal.science/tel-00661702**

Submitted on 20 Jan 2012

**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.

# UNIVERSITÉ PARIS-SUD 11

**ECOLE DOCTORALE** :

INNOVATION THÉRAPEUTIQUE : DU FONDAMENTAL A L'APPLIQUÉ
PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES

**DISCIPLINE :** STRUCTURE, FONCTION ET INGÉNIERIE DES PROTÉINES

SÉRIE DOCTORAT N° 1148

**THÈSE DE DOCTORAT**
soutenue le 5 Décembre 2011

par

**Charlotte ROBIN**

# Etude structurale de protéines virales impliquées dans la réplication et régulation du cycle de multiplication d'un virus de plante

| | | |
|---|---|---|
| **Directeur de thèse :** | Stéphane Bressanelli | CR1 (VMS, UPR CNRS 3296, Gif/Yvette) |
| **Composition du jury :** | | |
| *Président du jury :* | Jean-Michel Rossignol | Professeur (LGBC, Université de Versailles / Saint Quentin en Yvelines, Versailles) |
| *Rapporteurs :* | Patrice GOUET | Professeur (IBCP, UMR 5086 CNRS et Université de Lyon, Lyon) |
| | Rob RUIGROK | Professeur (EMBL, UMR 5233 CNRS et Université Joseph Fourier, Grenoble) |
| *Examinateur :* | Daniel Picot | Docteur (IBCP, UMR 7099 CNRS et Paris Diderot, Paris |

# Résumé

Le virus de la mosaïque jaune du navet (TYMV) est un excellent modèle pour la réplication des virus à ARN simple brin de polarité positive. Ce petit virus de plante code l'ensemble des protéines nécessaires à sa réplication sous forme d'un précurseur polyprotéine (206K). Du N- au C-terminus, celui-ci porte une activité méthyltransférase, protéase à cystéine (PRO), hélicase et ARN-polymérase ARN dépendante (POL). Comme tous les virus à ARN(+) connus, son complexe de réplication est étroitement lié à des membranes cellulaires. Dans le cas du TYMV, c'est l'enveloppe des chloroplastes qui est impliquée. Un des acteurs clés de la réplication est la polymérase qui permet la synthèse de nouveaux génomes. La régulation de son activité implique dans un premier temps son clivage de la 206K par PRO. Ainsi libérée, elle est recrutée aux chloroplastes par une interaction directe avec le domaine PRO du produit de clivage 140K, afin de former le complexe de réplication. Un second clivage par PRO contenu dans 140K, à la jonction PRO-hélicase, permet de poursuivre le cycle de réplication. Récemment, il a également été montré que l'activité POL était régulée par la voie ubiquitine-protéasome durant le cycle viral. Ubiquitinilée par la cellule hôte, elle est adressée au protéasome où elle sera dégradée. Cependant, PRO, grâce à sa seconde fonction ubiquitine hydrolase, est capable de la protéger de cette dégradation. Afin de caractériser d'un point de vue structural ce mécanisme de régulation de la réplication, nous avons cristallisé, à l'aide d'un contaminant, PRO et avons résolu sa structure. L'empilement cristallin est tel que le C-terminus d'un domaine PRO est inséré dans la crevasse catalytique du domaine PRO suivant, nous fournissant ainsi des informations structurales sur son activité endopeptidase. Dans un second temps, afin d'avoir des informations sur sa seconde activité, nous avons réalisé un complexe stable entre PRO et une molécule d'ubiquitine afin de le cristalliser et résoudre sa structure. Enfin, nous avons initié l'étude cristallographique de la polymérase.

**Mots clés** : Tymovirus, réplication, polymérase, protéase, ubiquitination, cristallographie

# Abstract

Turnip Yellow Mosaic Virus is an excellent model for positive-stranded RNA virus replication. It's a small plant virus whose replication machinery is encoded in the viral genome as a single polyprotein (206K). From N- to C-terminus, this 206K harbors a methyltransferase, a cysteine proteinase (PRO), a helicase and an RNA-dependent RNA polymerase (POL). As in all RNA(+) viruses known, the replication complex is bound to cell membranes. For TYMV, it ,is the chloroplast envelope that is implicated. A key component in the replication process is the polymerase, that allows the synthesis of new genomes. The regulation of its activity involves initially its cleavage by PRO from the 206K. Once liberated, the polymerase is recruited to the chloroplasts through a direct interaction with the PRO domain, contained in 140K, in order to form the replication complex. A second cleavage of 140K by PRO at the PRO-helicase junction allows to continue the replication cycle. Recently, it has also been shown that the POL activity was regulated by the ubiquitin-proteasome system during the viral multiplication cycle. Ubiquitinilated by the host cell, POL is addressed to the proteasome where it is degraded. However, PRO, due to its second function as an ubiquitin hydrolase, is able to protect POL from its degradation. In order to characterize this mechanism of replication regulation with a structural point of view, we crystallized, assisted by contaminant, PRO and resolved its structure. In the crystal packing, the C-terminal of a PRO domain is inserted into the catalytic cleft of the next PRO domain, thus providing us structural informations on its endopeptidase activity. In a second step, in order to obtain information about its second activity, we made a stable complex between PRO and a molecule of ubiquitin in order to cristallise it. Finally, we initiated the crystallographic study of POL.

**Keywords**: Tymovirus, replication, polymerase, proteinase, ubiquitination, crystallography

# Remerciements

Je tiens tout d'abord à remercier le Pr Patrice Gouet et le Pr Rob Ruigrok de m'avoir fait l'honneur d'être les rapporteurs de mon manuscrit de thèse. Je remercie également le Pr Hermann Van Thilbeurgh et le Dr Daniel Picot d'avoir accepté de juger ce travail.

J'exprime plus particulièrement ma gratitude à Stéphane Bressanelli, mon directeur de thèse, pour m'avoir accueillie au sein de son équipe depuis mon Master pour ce projet que j'ai beaucoup apprécié. On parle généralement de confiance, de soutien, de patience, de liberté accordée, de compétences, d'aide… lorsque l'on remercie son directeur de thèse. Je pense que toutes ces qualités définissent en fait sa fonction. Je tiens donc à le remercier tout simplement d'avoir été mon directeur de thèse et surtout de l'avoir été avec enthousiasme, optimisme et humour.

Je remercie également Isabelle Jupin, qui est à l'origine de ce projet et avec qui j'ai pris plaisir à collaborer depuis le début. Je la remercie plus spécifiquement pour sa persévérance et ses nombreux conseils en biologie cellulaire et moléculaire. Je remercie également Mélanie qui a aussi rendu cette collaboration et mes expériences de clonage très agréables.

Après 4 ans au laboratoire de Virologie Moléculaire et Structurale, je tiens bien sûr à remercier Yves Gaudin, le directeur de cet institut pour avoir accepté ma venue dans l'équipe de biocristallographie. Suite à ces années passées au sein de ce laboratoire, des tas d'histoires scientifiques ou non se sont accumulées faisant de ma thèse une expérience inoubliable. Pour cela, je remercie tous les membres de la VMS. Pour les moments d'amitié, nos discussions, nos déjeuners et nos covoiturages, je remercie vivement Déborah. Pour leur bonne humeur et leur humour je remercie Célia, Cynthia et Maya. Pour les longs moments un peu « geek » passés avec elle, je remercie Virginie. Et enfin, un remerciement spécial pour Lionel, mon voisin de bureau et envahisseur de paillasse qui a travaillé avec moi sur le projet « navet ». Je le remercie pour tous les moments très drôles qu'il m'a fait partager, pour ses nombreuses formations et ses débats scientifiques, pour son soutien et pour sa gentillesse.

Il est maintenant tant de remercier ceux sans qui ma vie de thésarde ne serait pas aussi équilibrée et épanouie. Tout d'abord mes amis d'escalade : les PSUCiens qui me permettent de passer d'excellents week-end et de me dépasser presque chaque soir. Parmi eux, je voudrais mettre en valeur la gentillesse et l'humour de Pietro, José et Stéphane et adresser un grand merci à Julia qui est toujours là pour partager les moments, joyeux comme douloureux et me faire rire à chaque fois.

Ensuite, je voudrais remercier mes amis. Ceux qui partagent ma vie depuis plusieurs années et avec qui je compte bien en passer encore de nombreuses : Halim, Samira, Zélia, Jérémy, les Trichets et Romain. Merci aussi à un autre Romain pour me faire aimer toujours plus la cristallographie, les congrès, le synchrotron et la lourdeur.

Finalement, il y a ma vie de thésarde mais il y a ma vie de façon plus générale. Je voudrais remercier du fond du cœur ma famille pour tout ce qu'elle m'apporte chaque jour. Je remercie ma grand-mère pour sa gentillesse et ses encouragements, mon oncle pour ses messages et son aide à des moments inespérés et Augustin pour ses « blagues sur blagues ».

Un gigantesque merci à Emilie, ma sœur, pour son humour inconditionnel, pour les folies qu'elle m'a fait vivre, pour son affection, son rein hypothétique et toutes les autres qualités qui font que je l'adore.

J'adresse un merci tout aussi énorme à mes parents débordants d'amour, d'humour, de générosité et de gentillesse à mon égard. S'ils sont fiers de moi comme ils me le disent souvent, je le suis au moins autant d'eux.

Mon dernier merci revient à Warody pour son soutien essentiel à l'aboutissement de ma thèse et pour son aide précieuse lors de la rédaction de ce manuscrit. Je le remercie de m'accompagner chaque jour dans tout, de ses bétises, de la vie que l'on partage et de tout ce qu'il fait pour moi. Merci de ce que tu es et qui fait que je t'aime à ce point !

A Richard Kahn,

Je voudrais ajouter quelques lignes supplémentaires pour Richard. Initialement rangé dans les remerciements dédiés aux amis avec lesquels je compte passer encore de nombreux moments, le voici tristement placé ici. Je tiens malgré tout à le remercier pour nos longues discussions sur les sujets de la vie, mes vacances grenobloises, ses talents de pâtissier, ses « cours » de français, ses récits scientifiques ou non, ses astuces FIP et surtout pour son amitié !

# Table des matières


RESUME ..... 3
ABSTRACT ..... 5
REMERCIEMENTS ..... 7
TABLE DES MATIERES ..... 10
INDEX DES FIGURES ..... 13
INDEX DES TABLES ..... 14
ABREVIATIONS ..... 15
CODE UNE ET TROIS LETTRES DES ACIDES AMINES ..... 16

**INTRODUCTION ..... 17**

LES VIRUS ..... 20

*Généralités ..... 20*
*Classification ..... 21*
*Les virus à ARN (+) ..... 22*

LE VIRUS DE LA MOSAÏQUE JAUNE DU NAVET : MODELE D'ETUDE ..... 24

*Caractéristiques du TYMV ..... 24*
*Infection et transmission ..... 25*
*Mécanisme de réplication ..... 26*
*Organisation du génome de TYMV ..... 26*
*La polyprotéine nécessaire à la réplication ..... 28*
*L'ARN polymérase dépendante de l'ARN ..... 30*
*L'hélicase ..... 32*
*Le dernier produit de clivage : la méthyltransférase et la protéase ..... 33*

L'UBIQUITINATION ET LA DEUBIQUITINATION ..... 36

*L'ubiquitination ..... 36*
*Le système de dégradation ubiquitine/protéasome ..... 37*
*La déubiquitination ..... 38*
*L'ubiquitination et la déubiquitination chez les virus ..... 39*
*La protéase du TYMV et les OTU protéases ..... 40*

LA BIO-CRISTALLOGRAPHIE ..... 43

*La cristallogenèse ..... 43*
*Principe ..... 44*
*Technique ..... 45*
*Préparation des cristaux ..... 45*
*Diffraction des rayons X ..... 46*
*Espace réciproque, espace direct et transformée de Fourier ..... 47*
*Le problème des phases ..... 48*
*Le remplacement moléculaire ..... 50*
*Le remplacement isomorphe ..... 51*
*Utilisation de la diffusion anomale ..... 52*
*Reconstruction de la structure ..... 52*

**PARTIE I : CRISTALLISATION ASSISTEE PAR UN CONTAMINANT (ARTICLE 1) ..... 56**

RESUME ..... 58

ARTICLE 1 : IN PRAISE OF IMPURITY: 30S RIBOSOMAL S15 PROTEIN - ASSISTED CRYSTALLIZATION OF TURNIP YELLOW MOSAIC VIRUS PROTEINASE ..... 60

*Synopsis* ..... *61*
*Abstract* ..... *61*
*Introduction* ..... *62*
*Results and discussion* ..... *62*
*Cloning, expression and purification* ..... *62*
*Crystallization* ..... *64*
*Data collection and MIRAS phasing from hexagonal crystals in a single drop* ..... *65*
*Native data* ..... *65*
*Derivative data* ..... *65*
*MIRAS phasing and density modification with autoSHARP* ..... *66*
*Model rebuilding and reassessment of the asu content* ..... *67*
*Conclusion* ..... *69*
*Aknowledgments* ..... *72*
*References* ..... *72*

COMPLEMENT AUX RESULTATS DE L'ARTICLE 1 : PHASAGE ALTERNATIF ..... 71

*Purification partielle de TYMVpro* ..... *72*
*Remplacement moléculaire à partir de la structure résolue du contaminant* ..... *74*
*Conclusion* ..... *75*

**PARTIE II ETUDE STRUCTURALE DE LA PROTEASE DU TYMV (ARTICLE 2) ..... 76**

RESUME ..... 78

ARTICLE 2 : CRYSTAL STRUCTURE OF A VIRAL MULTIFUNCTIONAL PROTEINASE/UBIQUITIN HYDROLASE IN POLYPROTEIN PROCESSING MODE ..... 82

*Abstract* ..... *83*
*Introduction* ..... *84*
*Experimental procedures* ..... *86*
*Protein expression, purification and crystallyzation* ..... *86*
*MIRAS phasing, model building and refinement* ..... *87*
*Structure analysis* ..... *87*
*Results* ..... *89*
*Overall protein architecture* ..... *89*
*A view of the N-terminal product complex in viral polyprotein processing* ..... *92*
*Structural basis of PRO proteinase sequence specificity* ..... *94*
*Peptidase-substrate recagnition patches* ..... *94*
*PRO has an ubiquitin hydrolase (DUB) fold distantly related to the Ovarian Tumor (OTU) domain class of DUB enzymes* ..... *95*
*Comparison with available OTU DUB substrate complexes* ..... *95*
*Discussion* ..... *97*
*A substrate recognition lobe unique to Tymoviridae* ..... *99*
*Footnotes* ..... *99*
*References* ..... *100*

COMPLEMENTS AUX RESULTATS DE L'ARTICLE : VUE STRUCTURALE DE LA FONCTION DUB ..... 100
*Obtention d'un complexe stable entre TYMVpro et une ubiquitine suicide* ..... *100*
*Matériel et méthode* ..... *104*
*Expression/Production/purification de TYMVpro* ..... *104*
*Cristallisation de TYMVpro* ..... *106*
*Collecte des données et phasage par MIRAS* ..... *106*
*Expression et production de l'HAUb-activée* ..... *106*
*Purification d'Ubi-MESNa* ..... *106*
*Synthèse et purification d'HAUb-Br2, d'HAUb-Br3 ou d'HAUb-CL* ..... *106*
*Mise en contact pour la formation du complexe* ..... *106*
POURSUITE DU PROJET ..... 111

**PARTIE III : OBTENTION D'UN FRAGMENT SOLUBLE DE LA RDRP DU TYMV ..... 112**
INTRODUCTION ..... 114
RESULTATS ..... 117
*Expression du domaine TYMVpol* ..... *117*
*Production et lyse bactérienne* ..... *118*
*Repliement de la 66K* ..... *120*
*Dichroïsme circulaire et tests d'activité* ..... *121*
*Protéolyse ménagée et identification du domaine polymérase stable* ..... *123*
*Expression d'un fragment soluble* ..... *124*
MATERIEL ET METHODES ..... 125
*Essais sur la production insoluble* ..... *125*
*Expression de la 66K* ..... *104*
*Solubilisation de la 66K* ..... *106*
*Purification de la 66K* ..... *106*
*Protéolyse ménagée* ..... *106*
*Production soluble de 66K : pESPRIT* ..... *126*
*Préparation de constructions utilisées par la plateforme* ..... *104*
DISCUSSION ..... 129

**PERSPECTIVES ET DISCUSSION ..... 131**
DECOUVERTES RECENTES ..... 134
ETUDE STRUCTURALE DE LA FONCTION DUB ..... 134
RESOLUTION DE LA STRUCTURE DE LA 66K ..... 135
ROLE DE L'UBIQUITINATION REVERSIBLE DU TYMV ..... 136
CARACTERISATION DE CE MECANISME DE REGULATION AU NIVEAU CELLULAIRE ..... 136

**BIBLIOGRAPHIE ..... 139**

# Index des Figures

Figure 1 : Diversité virale illustrée à l'aide de trois exemples ..... 21
Figure 2 : Importance des virus à ARN positifs ..... 23
Figure 3 : Capside icosaédrique du TYMV résolue par microspie électronique ..... 24
Figure 4 : Cycle de multiplication du TYMV ..... 27
Figure 5 : Organisation du génome viral du TYMV ..... 28
Figure 6 : Régulation par phosphorylation réversible de la RdRp du TYMV ..... 31
Figure 7 : Intéraction RdRp/protéase lors de la formation du complexe de réplication ..... 33
Figure 8 : Mécanisme catalytique des protéases à cystéine ..... 34
Figure 9 : Mécanisme de régulation de la RdRp conduisant à sa dégradation ..... 35
Figure 10 : Mécanisme d'ubiquitination des protéines substrat ..... 37
Figure 11 : Différentes possibilité d'ubiquitination des protéines cibles ..... 37
Figure 12 : Différentes cibles reconnues par les déubiquitinases et déubiquitination ..... 39
Figure 13 : Représentation des méthodes utilisant le phénomène de diffusion de vapeur ..... 45
Figure 14 : Représentation de la loi de Bragg ..... 47
Figure 15 : Principe de la diffraction des rayons X en bio-cristallographie ..... 48
Figure 16 : Obtention du signal de l'atome lourd ..... 51
Figure 17 : Analyse de la composition des cristaux hexagonaux sur gel Tris –tricine ..... 73
Figure 18 : Nomenclature des sites P ..... 79
Figure 19 : Simulation d'accession d'une molécule d'ubiquitine dans l'empilement cristallin ..... 101
Figure 20 : Synthèse des dérivés HAUb ..... 102
Figure 21 : Carte plasmidique de la construction utilisée pour l'expression du domaine TYMVpro ..... 104
Figure 22 : Analyse des étapes de purification du domaine TYMVpro ..... 105
Figure 23 : Spectre de masse du produit de synthèse de l'HAUb-CL ..... 108
Figure 24: Analyse sur gel du complexe HAUb-VME / TYMVpro ..... 109
Figure 25 : Analyse de la solubilisation de la 66K sur gel de polyacrylamide ..... 120
Figure 26 : Analyse sur gel de polyacrylamide 10% des fractions issus de la purification de TYMVpol ..... 122
Figure 27 : spectre de dichroïsme circulaire de la 66K solubilisée et purifiée ..... 122
Figure 28 : Analyse par spectrométrie de masse des produits de protéolyse ménagée ..... 123
Figure 29 : Carte du plasmide pESPRIT002 ..... 127
Figure 30 : Carte du plasmide pACYC pour l'expression du domaine TYMVpro ..... 128
Figure 31 : Prédiction de désordre à partir de la séquence de la 66K ..... 130

# Index des Tables

Table 1 : Principales DUB rencontrées dans le monde viral ..... 42

Table 2 : Différents tests d'expression réalisés ..... 118

Table 3 : Conditions de production testées ..... 119

Table 4 : Tampons de lyse utilisés ..... 119

Table 5 : Méthodes de lyse utilisées ..... 119

# Abréviations

| | |
|---|---|
| **Å** | Angstroem |
| **ADN** | Acide DésoxyriboNucléique |
| **ARN** | Acide RiboNucléique |
| **ATP** | Adénosine TriPhosphate |
| **CCP4** | Collaborative Computational Project, Number 4 |
| **CD** | Circular Dichroism |
| **DO** | Densité Optique |
| **DTT** | 1,4-DiThioThreitol |
| **DUB** | Deubiquitinating Enzyme |
| ***E. coli*** | *Escherichia coli* |
| **EDTA** | Ethylene Diamine Tetra Acetic Acid |
| **EMBL** | European Molecular Biology Laboratory |
| **ESPRIT** | Expression of Soluble Proteins by Random Incremental Truncation |
| **ESRF** | European Synchrotron Radiation Facility |
| **GTP** | Guanosine TriPhosphate |
| **HEPES** | Acide 4-(2-hydroxyéthyl)-piperazine-1-éthanesulfonique |
| **ICTV** | International Committee on Taxonomy of Viruses |
| **kDa** | Kilodalton |
| **M** | Mole par litre |
| **MAD** | Multi Anomalous Diffusion |
| **mg** | Milligramme |
| **MIR** | Multi Isomorphous Replacement |
| **MIRAS** | Multi Isomorphous Replacement with Anomalous Scattering |
| **mL** | Millilitre |
| **mm** | Millimètre |
| **mM** | Millimole par litre |
| **ORF** | Open Reading frame |
| **$P^3$** | Protein Production Plateform |
| **PDB** | Protein data bank |
| **PHENIX** | Python-based hierarchical environment for integrated crystallography |
| **RdRp** | RNA-dependent RNA polymerase |
| **SAD** | Single Anomalous Diffusion |
| **SDS-PAGE** | Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis |
| **SIR** | Single Isomorphous Replacement |
| **SIRAS** | Single Isomorphous Replacement with Anomalous Scattering |
| **SOLEIL** | Source Optimisée de Lumière d'Energie Intermédiaire du LURE |
| **TRIS** | Tris(hydroxymethyl)amino méthane |
| **TYMV** | Turnip Yellow Mosaïc Virus |
| **VMS** | Virologie Moléculaire et Structurale |
| **XDS** | X-ray Detector Software |
| **µL** | Microlitre |

## *Code une et trois lettres des acides aminés*

| | | |
|---|---|---|
| G | Gly | Glycine |
| A | Ala | Alanine |
| V | Val | Valine |
| L | Leu | Leucine |
| I | Ile | Isoleucine |
| S | Ser | Sérine |
| T | Thr | Thréonine |
| C | Cys | Cystéine |
| M | Met | Méthionine |
| D | Asp | Acide aspartique |
| N | Asn | Asparagine |
| E | Glu | Acide glutamique |
| Q | Gln | Glutamine |
| K | Lys | Lysine |
| R | Arg | Arginine |
| H | His | Histidine |
| F | Phe | Phénylalanine |
| Y | Tyr | Tyrosine |
| W | Trp | Tryptophane |
| P | Pro | Proline |

# INTRODUCTION

# Les virus

## Généralités

Les virus sont des entités biologiques extrêmement simples composées d'une capside protéique renfermant le génome viral. Ces deux éléments constituent la base minimale et forment la nucléocapside. Certains virus peuvent aussi posséder une capside plus complexe constituée de plusieurs couches, une enveloppe lipidique, un tégument reliant l'enveloppe et la capside, des protéines de spicule, des enzymes et d'autres éléments plus divers. L'ensemble de tous ces composants forme ce qu'on appelle le virion (Figure 1).

La première définition d'un virion a été donnée en 1953 par André Lwoff, qui l'a complétée par la suite (Lwoff 1971) pour aboutir aux caractéristiques suivantes :

- Le virion renferme un seul type d'acide nucléique. L'ADN et l'ARN ne coexistent pas dans la particule virale, contrairement aux cas des êtres vivants. L'acide nucléique viral porte l'intégralité de l'information génétique du virus et constitue ce qu'on appelle le génome viral.
- Le virion se reproduit à partir de son acide nucléique, simplement par réplication de son génome viral.
- Le virus est un parasite intracellulaire strict. Il ne peut se reproduire qu'après son entrée dans une cellule hôte, vivante. Etant dépourvu d'organite et de machinerie cellulaire lui permettant d'assurer sa propre réplication, il doit alors les détourner de la cellule qu'il parasite. Des systèmes biologiques vivants sont donc nécessaires pour reproduire un virus en laboratoire.
- Le virion a dans la plupart des cas une structure présentant des éléments de symétrie particuliers.
- Il ne croît pas, ni ne se scinde, et se multiplie grâce à la synthèse et à l'assemblage de ses différents constituants.

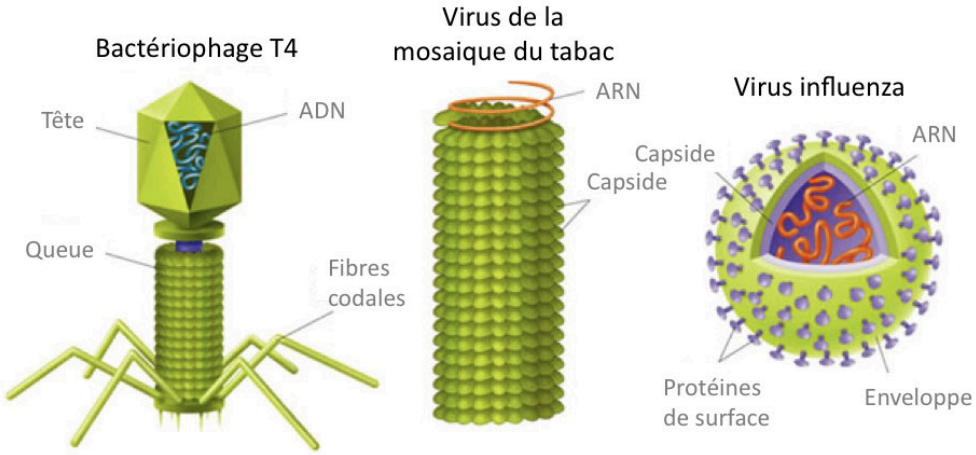



**Figure 1 : Diversité virale illustrée à l'aide de trois exemples**
Le bactériophage T4 possède une capside allongée complexe renfermant son génome, ainsi une queue à la base de laquelle est attachée de longues fibres. Il fait partie des virus à symétrie complexe. Le virus de la mosaïque du tabac est un virus hélicoïdal composé simplement d'une nuclécapside cylindrique. Le virus influenza est un virus enveloppé et décoré de glycoprotéines.

## Classification

Il existe de nombreuses façons de classer les virus. Certaines s'appuient sur l'hôte infecté, d'autres sur le mode de réplication, ou même sur le type de génome et de structure du virus. On peut également les distinguer en fonction des maladies qu'ils engendrent et des symptômes qu'ils provoquent par exemple.
L'ICTV (International Committee on Taxonomy of Viruses) a mis en place une base de données (http://www.ictvdb.org) se voulant universelle, et au sein de laquelle les virus sont distingués selon différents taxons. Du plus au moins élevé, on trouve l'ordre (noté par le suffixe -virales), la famille (-viridae), la sous-famille (-virinae) et le genre (-virus). Parfois, au sein d'un genre, on trouve également une catégorie supplémentaire : l'espèce.
En 2009, la base de donnée comptait six ordres : les Caudovirales englobant 3 familles, les Herpesvirales (3 familles), les Mononegavirales (4 familles), les Nidovirales (3 familles), les Picornavirales (5 familles), les Tymovirales (4 familles) et les virus non assignés à un ordre regroupés en 65 autres familles.

Les critères principaux utilisés pour l'assignation d'un virus à une catégorie sont ici :

- Son type de génome et son organisation. On distingue alors les virus à ADN des virus à ARN et ceux dont les génomes sont simple ou double brin. Dans le cas des virus « simples brins », on les subdivise également en fonction de la polarité de leur génome : négative ou positive.
- La structure du virion. Il existe en effet des symétries de capside particulières. Trois types sont alors dissociés : les capsides icosaédriques, hélicoïdales ou complexes (Figure 1).
- La stratégie de réplication. De façon simplifiée, on différencie les processus cytoplasmiques des processus nucléaires.

## Les virus à ARN (+)

Les virus à ARN de polarité positive constituent la plus grande classe de virus connus et englobent de nombreux pathogènes infectant aussi bien les animaux que les humains, les bactéries, les champignons et même les plantes (Figure 2). Plus d'un tiers des virus possèdent un génome pouvant servir directement de matrice. Cela signifie qu'il agit tel un ARN messager et est traduit directement pour la synthèse de nouveaux génomes.

Au sein de cette classe, on trouve de nombreuses super-familles (Van Regenmortel 2000) assez divergentes. Cependant, des caractéristiques communes à tous ces virus existent et plus particulièrement lorsque l'on s'intéresse à leur mécanisme de réplication. Tous utilisent en effet leur ARN génomique comme matrice pour la traduction et la réplication. Tous subissent également un recrutement et une localisation de leurs complexes de réplication au niveau de membranes intracellulaires et produisent un grand excès d'ARN de polarité positive par rapport aux ARN de polarité négative.

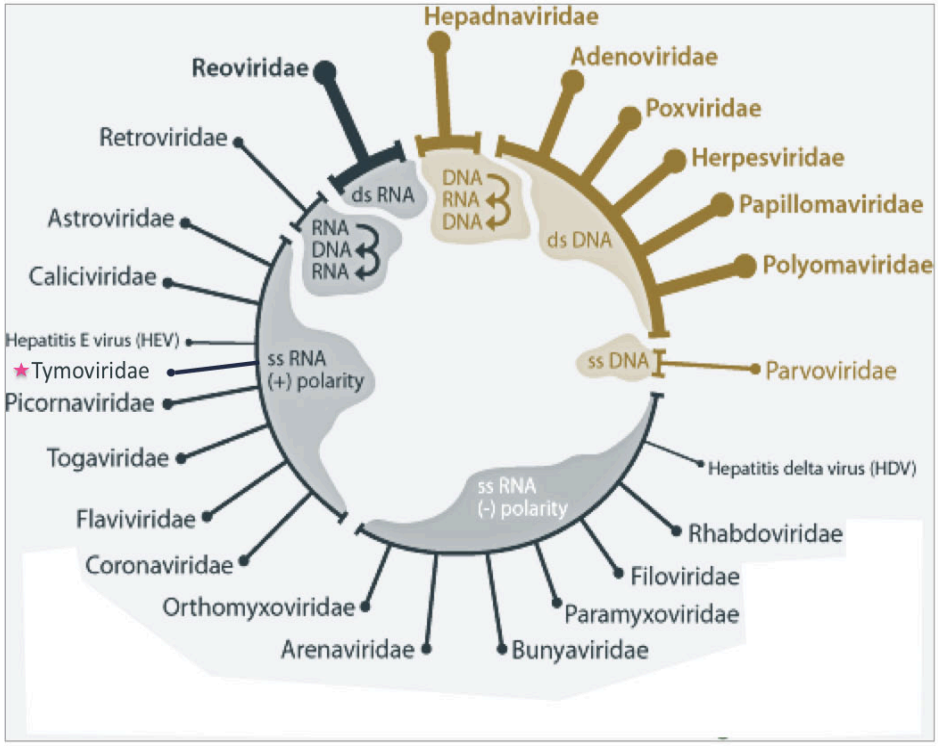



**Figure 2 : Importance des virus à ARN positifs**
En bleu sont représentés les virus à ARN et en Marron les virus à ADN. On constate ici que les virus à ARN constituent environ 2/3 des virus connus et les virus à ARN simple brin comprenant les tymoviridae sont largement représentés au sein de cette classe.

# Le virus de la mosaïque jaune du navet : modèle d'étude

Afin d'étudier plus en détail la classe des virus à ARN(+), leur mécanisme de réplication et les interactions hôte/pathogène impliquées au cours du cycle viral, nous avons choisi de nous intéresser au virus de la mosaïque jaune du navet (abrégé TYMV pour son acronyme anglais Turnip Yellow Mosaic Virus). Ce choix s'explique en partie par le fait que l'on dispose de nombreuses informations sur ce petit virus de plante. Sa réplication est particulièrement bien caractérisée au niveau moléculaire et cellulaire et de nombreux outils sont disponibles pour compléter nos investigations et en faire un modèle d'étude des virus à ARN simple brin de polarité positive.

## Caractéristiques du TYMV

Le TYMV est un tymovirus appartenant à la famille des Tymoviridae et à l'ordre des Tymovirales. C'est un virus sphérique, de 30 nm de diamètre environ, non enveloppé et dont la capside simple couche possède une symétrie icosaédrique de triangulation T=3 (Figure 3). Comme la quasi-totalité des virus de plante, il ne possède pas d'enveloppe et est dit « nu ».

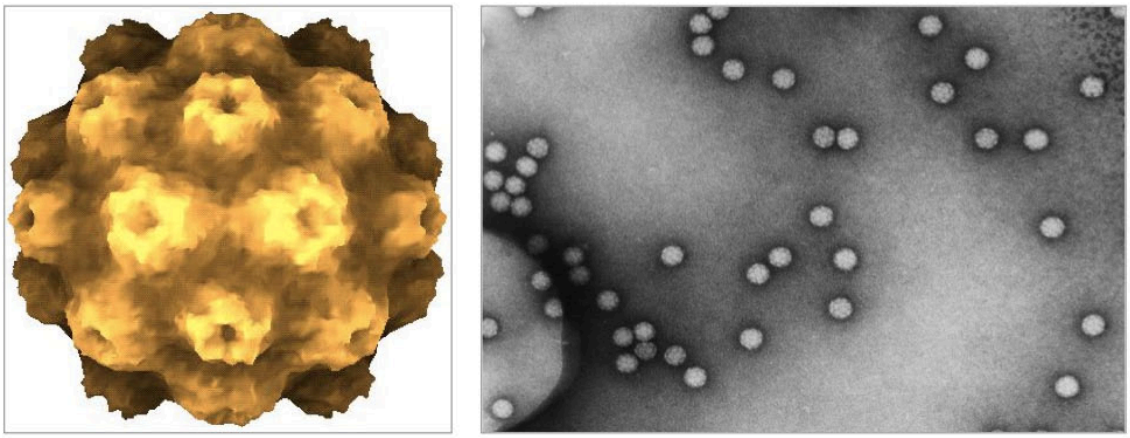


**Figure 3 : Capside icosaédrique du TYMV résolue par microspie électronique**
A gauche la reconstruction et à droite les particules virales vues par microscopie (Canady 1996).

## Infection et transmission

TYMV affecte le système photosynthétique des plantes dycotylédones parmis lesquelles on trouve les brassicacées et la plante modèle *Arabidopsis thaliana*. Le fait que TYMV soit capable d'infecter *Arabidopsis thaliana* (Camborde 2007) accentue notre choix pour cette étude fondamentale. Puisque cette petite plante modèle est très largement utilisée dans la recherche scientifique, de nombreuses informations sont disponibles dans une base de données (http://www.arabidopsis.org) et son génome est presque intégralement séquencé. Sa culture en laboratoire est aussi parfaitement maîtrisée ainsi que toutes les régulations qui régissent son développement. L'infection par le TYMV induite par inoculation mécanique est, de façon intéressante, mise au point par nos collaborateurs et permet de réaliser des études *in vivo* (Camborde 2007).

Une fois infectée par TYMV, les plantes présentent une mosaïque jaune au niveau des feuilles. Ces taches sont dues à la décoloration des zones dont le système photosynthétique est touché. La transmission naturelle des virus de plante se fait de façon particulière. En effet, les cellules végétales ont une paroi épaisse qui ne permet pas l'entrée du virus par endocytose ou fusion. Dans le cas de notre modèle, l'entrée se fait par effraction grâce à un insecte vecteur qui va injecter le virus directement dans la cellule végétale. Une autre originalité des plantes est la présence de pores permettant de relier les cellules végétales entres elles et appelés plasmodesmes. Grâce à ces connections, le virus n'est pas contraint de réaliser un cycle de multiplication « complet » pour propager son infection à une nouvelle cellule voisine. Par ce biais, il n'a en effet pas à former de nouveaux virions, ni à sortir de la cellule où il s'est multiplié pour ensuite entrer dans une cellule voisine. Il peut directement transmettre son acide nucléique néo-synthétisé par réplication, via ces pores, avec l'aide d'une protéine virale de mouvement (Figure 4).

## Mécanisme de réplication

Une fois libre dans la cellule, après l'entrée et la décapsidation, le génome des virus à ARN de polarité positive va être utilisé immédiatement comme matrice pour la traduction et la réplication. L'ARN génomique qui est directement messager, va être traduit par la machinerie de l'hôte pour fournir les protéines virales nécessaires au cycle de multiplication (Figure 4). Parmi ces protéines, certaines sont essentielles pour la réplication et d'autres pour la synthèse de nouveaux virions. Les protéines impliquées dans la réplication vont être les premières à entrer en action et vont s'associer entre elles et avec des protéines de la cellule hôte pour former le complexe de réplication. Ce complexe est toujours retrouvé au niveau de structures membranaires particulières lors de la réplication des virus à ARN(+). Pour le TYMV, ces structures ont été identifiées et l'on sait que les membranes externes des chloroplastes sont directement impliquées dans la formation de vésicules périphériques dans lesquelles a lieu la synthèse d'ARN (Prod'homme 2001 et 2003). Pour que cette localisation membranaire ait lieu, ces virus codent dans leur génome, un gène codant pour une protéine non structurale ayant de l'affinité pour les membranes. Celle-ci, grâce à des interactions protéine/protéine, va pouvoir recruter les autres sous-unités du complexe de réplication, directement ou non.

Une fois formé, ce complexe va permettre la synthèse d'un brin d'ARN complémentaire de polarité négative à partir du génome. Via ce brin d'ARN(-) néo-formé, le complexe de réplication réalise ensuite la synthèse de nouveaux ARN(+) génomiques, en large excès. Ces nouveaux génomes sont a leur tour traduits, envoyés vers des cellules voisines, ou alors encapsidés pour infecter de nouvelles cellules. Le brin d'ARN de polarité négative peut également servir de matrice lors de la transcription pour permettre l'expression des gènes codés via des ARN subgénomiques.

## Organisation du génome de TYMV

Le génome de TYMV est constitué d'un ARN monocaténaire simple brin de 6,3 kb (Morch 1988) (Figure 5). L'extrémité 5' est coiffée et l'extrémité 3' se replie à la manière d'un ARNt pouvant lier une valine. Comme pour de nombreux virus à ARN(+), plusieurs stratégies d'optimisation du génome sont retrouvées afin de limiter la taille de celui-ci.

On distingue 3 phases ouvertes de lecture (ORF) le long de cet ARN, dont la dernière est traduite depuis un ARN subgénomique. La première ORF (ORF1) code pour une protéine non-structurale de 69 kDa (69K) responsable du mouvement de cellule à cellule lors de l'infection virale (Bozarth 1992). La deuxième phase ouverte de lecture (ORF2) permet la synthèse d'une polyprotéine, strictement nécessaire à la réplication, de 206 kDa (206K) (Weiland 1989 et 1993).

Cette deuxième ORF chevauche la première et débute seulement 7 nucléotides après le début de l'ORF1. Enfin, un ARN subgénomique, englobant l'extrémité 3' donne, lorsqu'il est traduit au cours de l'infection virale, la protéine de capside (CP) de 20 kDa.

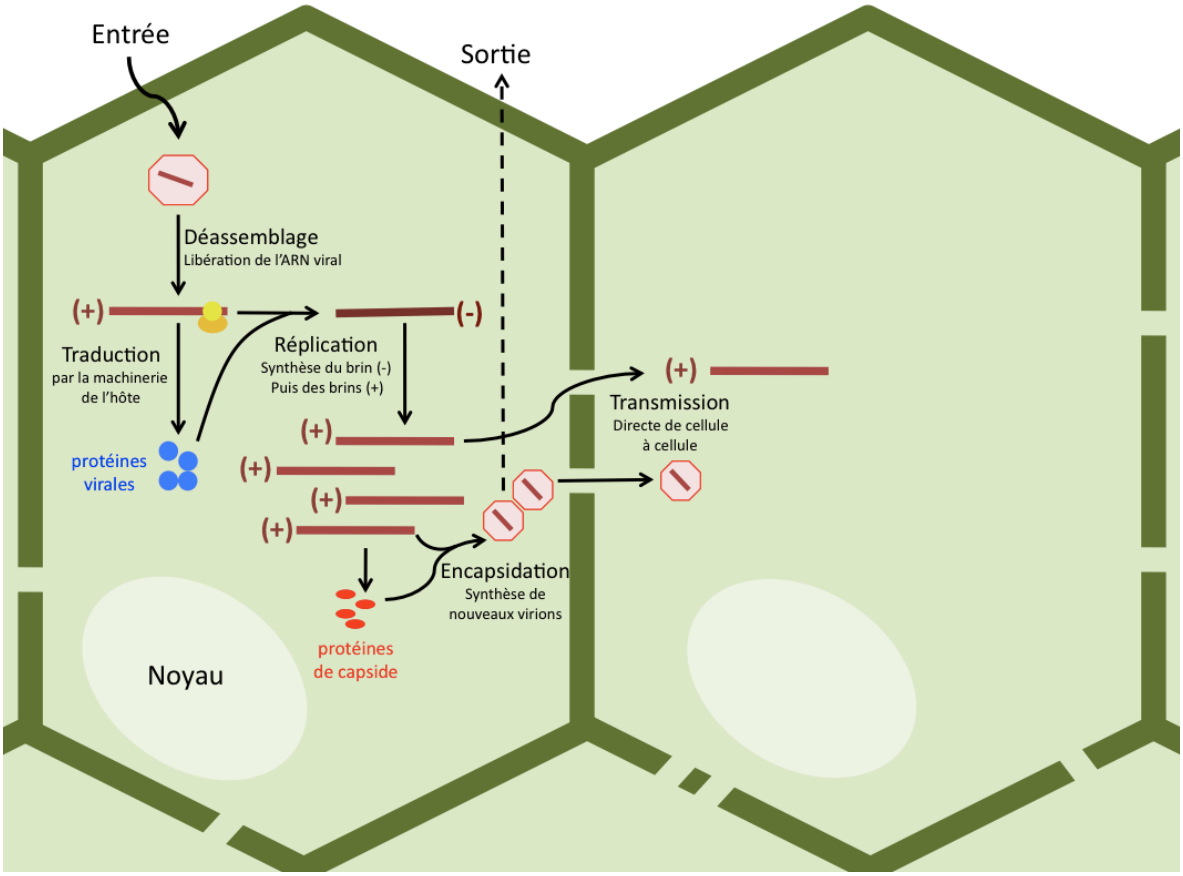



**Figure 4 : Cycle de multiplication du TYMV**

Après son entrée dans la cellule par effraction, le virus se désassemble pour libérer son génome (en rouge foncé) dans le cytoplasme de la cellule infectée. Celui-ci va être directement traduit par la machinerie de la cellule hôte (en jaune) pour fournir les protéines virales nécessaires à la réplication (en bleu). Celles-ci vont alors s'assembler au niveau de structures membranaires pour former le complexe de réplication et la synthèse de brins de polarité négative est alors réalisée (en marron). Celui-ci va ensuite servir de matrice à la synthèse de brins (+). Ces brins (+) vont être à leur tour traduits et conduire à la synthèse des protéines de capside (en rouge). Certains vont être encapsidés pour former de nouveaux virions qui vont soit ressortir de la cellule pour infecter des cellules plus éloignées, soit passer directement par les plasmodesmes pour infecter les cellules voisines. Enfin, certains brins (+) vont passer directement dans la cellule voisine via ces plasmodesmes et un nouveau cycle de réplication pourra directement avoir lieu.

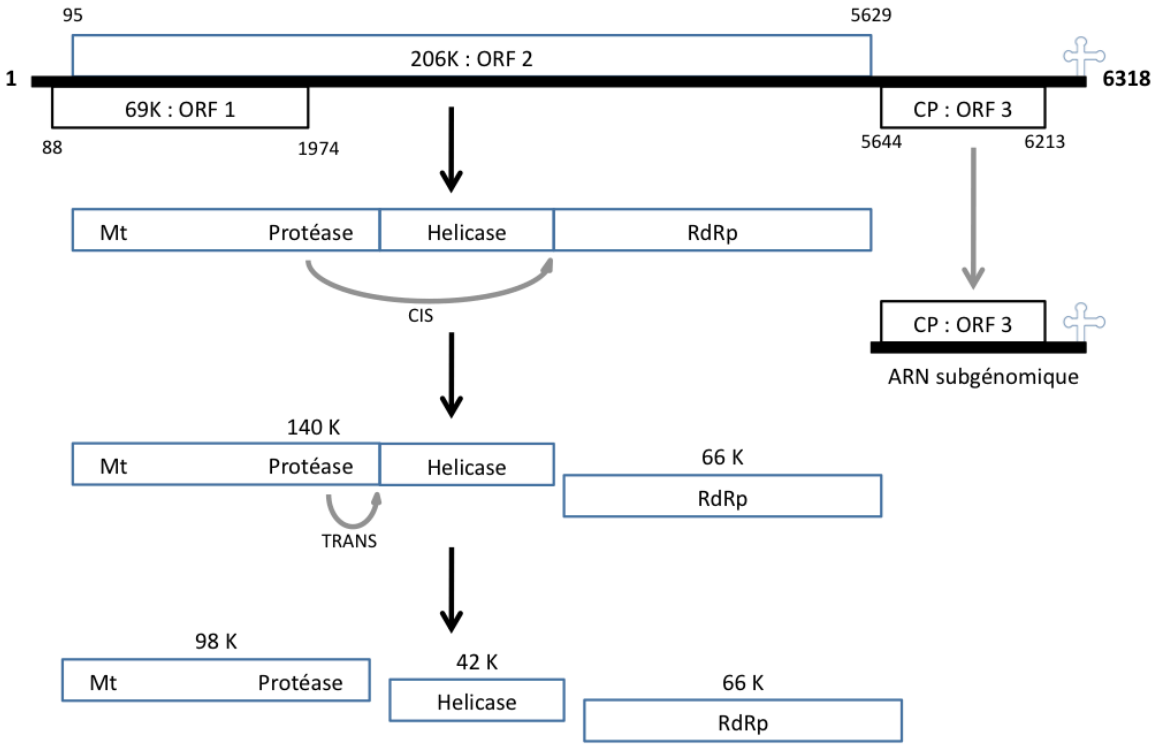



**Figure 5 : Organisation du génome viral du TYMV**
Le génome est constitué de 6318 bases. La séquence nucléotidique allant de 88 à 1974 correspond à la première phase ouverte de lecture et code pour la protéine de mouvement. Une deuxième phase ouverte de lecture est comprise entre les bases 95 et 5629 et contient les gènes codant pour les protéines nécessaires à la réplication. Finalement, une dernière ORF est portée par un ARN sub-génomique en 3' du génome suivi d'une structure de type ARNt. L'ORF 2 va donc être traduite en un précurseur polypeptidique composé, du N-terminal au C-terminal, d'un domaine méthyltransférase, d'un domaine protéase, hélicase et enfin ARN-polymérase ARN dépendante. Afin de maturer les différentes sous untiés de la réplicase, le domaine protéase va cliver, en CIS, la RdRp. Ainsi deux produits de clivage sont libérés, un de 140 kDa et un de 66 kDa contenant donc la RdRp. L'hélicase est ensuite à son tour maturée par un second clivage réalisé par la protéase toujours, mais probablement en trans. Cette fois, un fragment de 98 kDa et un de 42 kDa sont obtenus.

## La polyprotéine nécessaire à la réplication

En raison de la petite taille de leur génome, les virus à ARN(+) ont généralement recours à des précurseurs polypeptidiques. Dans le cas de TYMV, toutes les protéines impliquées dans la réplication sont exprimées au sein d'un seul précurseur : la polyprotéine 206K. De l'extrémité N-terminale à l'extrémité C-terminale, on trouve un domaine de type méthyltransférase(MT), un domaine portant une activité cystéine protéase (PRO), puis une NTPase-hélicase (HEL) et finalement l'ARN polymérase dépendante de l'ARN (RdRp). Les domaines MT et PRO sont séparés par un domaine riche en prolines d'environ 200 acides aminés probablement non structuré.

Grâce à la protéase PRO, la maturation protéolytique de cette réplicase est réalisée. En effet, celle-ci va effectuer plusieurs clivages pour libérer les différentes sous unités (Figure 5).
Tout d'abord, une première coupure, en cis, entre les acides aminés $Ala^{1259}$ et $Thr^{1260}$ permet de libérer la RdRp de 66 kDa et un polypetide de 140 kDa portant les domaines MT, PRO et HEL.
Un second clivage va ensuite libérer le domaine hélicase. Pour cela, un domaine PRO va agir, en trans, au niveau C-terminal d'un autre domaine PRO (Bransom 1991, 1995 et 1996, Jakubiec 2007). Deux nouveaux produits de maturation sont ainsi formés : l'un de 42 kDa portant le domaine HEL et l'autre de 98 kDa constitué de la méthyltransférase, du domaine riche en proline et de la protéase. Le processus de maturation n'est sans doute pas encore connu dans sa totalité, mais les sites de coupures mis en évidence jusqu'à présent sont parfaitement documentés, et l'on sait que ce second clivage a lieu entre les acides aminés $Ser^{879}$ et $Gln^{880}$.
Le premier clivage est nécessaire à l'assemblage du complexe de réplication et à la synthèse du brin(-). Le second clivage quant à lui influe sur la synthèse de brin(+) et sur l'ARN subgénomique (Jakubiec 2007).

*L'ARN polymérase dépendante de l'ARN*

La protéine non-structurale clé de la réplication est l'ARN polymérase ARN dépendante. En effet, c'est elle qui permet la synthèse de nouveaux ARN à partir d'une matrice ARN.
Elle est généralement identifiée par alignement de séquences grâce à des motifs conservés dans l'enchaînement primaire des RdRp virales (Koonin 1991, Poch 1989).
Ces motifs sont notés de différentes manières mais la nomenclature la plus utilisée les distingue par des lettres allant de A à F. La majeure partie des motifs conservés sont situés dans un domaine appelé la paume et dans lequel on trouve le site catalytique et une séquence GDD localisée dans le motif C strictement conservée chez toutes les RdRp. D'autres résidus nécessaires au fonctionnement de l'enzyme sont également retrouvés dans ces motifs tels que des résidus impliqués dans la chélation d'ions divalents, dans le déterminisme des nucléotides à incorporer, dans la fidélité de la polymérisation ou encore dans l'agencement du site actif. Le motif F quant à lui est localisé dans un autre domaine appelé doigts et contient des résidus chargés essentiels à l'incorporation des nucléotides lors de la synthèse d'un nouveau brin d'ARN.

Dans le cas du TYMV, aucune donnée structurale sur la RdRp n'est pour le moment connue. En effet, lors de notre recherche d'homologue nous n'avons pas pu déceler une RdRp proche dont la structure aurait déjà été résolue. La résolution de sa structure est donc essentielle si l'on veut obtenir des informations structurales sur celle-ci. En revanche, sur le plan moléculaire, cette RdRp est très bien caractérisée et de récentes découvertes concernant son mode de régulation ont été mises en évidences.
La polymérase du TYMV fait environ 66 kDa et elle sera notée de façon simplifiée 66K. Dans sa partie N-terminale, on trouve une séquence PEST. Les séquences PEST sont riches en proline (P), acide glutamique (E), sérine (S) et thréonine (T). Elles sont généralement non structurées et jouent le rôle de signal pour la dégradation des protéines (Rechsteiner 1996). En aval de cette séquence, se trouve le domaine contenant l'activité polymérase et incluant les motifs conservés des RdRp virales.
Au cours du cycle viral, l'activité de la polymérase est hautement régulée afin d'accomplir convenablement les différents étapes de la réplication. Cette enzyme intervient tout d'abord pour la synthèse du brin(-) et réalise ensuite la synthèse du brin(+) en large excès. Bien souvent, une modification post-traductionnelle est un moyen efficace et rapide de réguler l'activité des enzymes et, parmi ces modifications, la phosphorylation a été mise en évidence chez plusieurs RdRp virales. Dans le cas du virus de la dengue et du virus de la mosaïque du concombre par exemple, la phosphorylation de leur polymérase affecte la formation du complexe de réplication en modulant les interactions inter-protéines (Kapoor 1995, Kim 2002).

Chez la 66K, son mode de régulation par phosphorylation est connu de façon bien plus avancée. En effet, trois résidus ont été précisément identifiés par nos collaborateurs comme étant phosphorylés au cours du cycle viral. Parmi ces résidus, on trouve la $Ser^{326}$ localisée dans le motif catalytique A de la paume dont la phosphorylation/déphosphorylation altère le ratio entre les brins(-) et les brins(+) au cours de la réplication (Jakubiec 2006). Les deux autres résidus, la $Thr^{64}$ et de la $Ser^{80}$, sont localisés dans la partie N-terminale de la séquence PEST et leur phosphorylation constitue un signal pour l'ubiquitination de la polymérase, entraînant alors son adressage au protéasome où elle sera dégradée (Jakubiec 2006, Camborde 2010) (Figure 6).

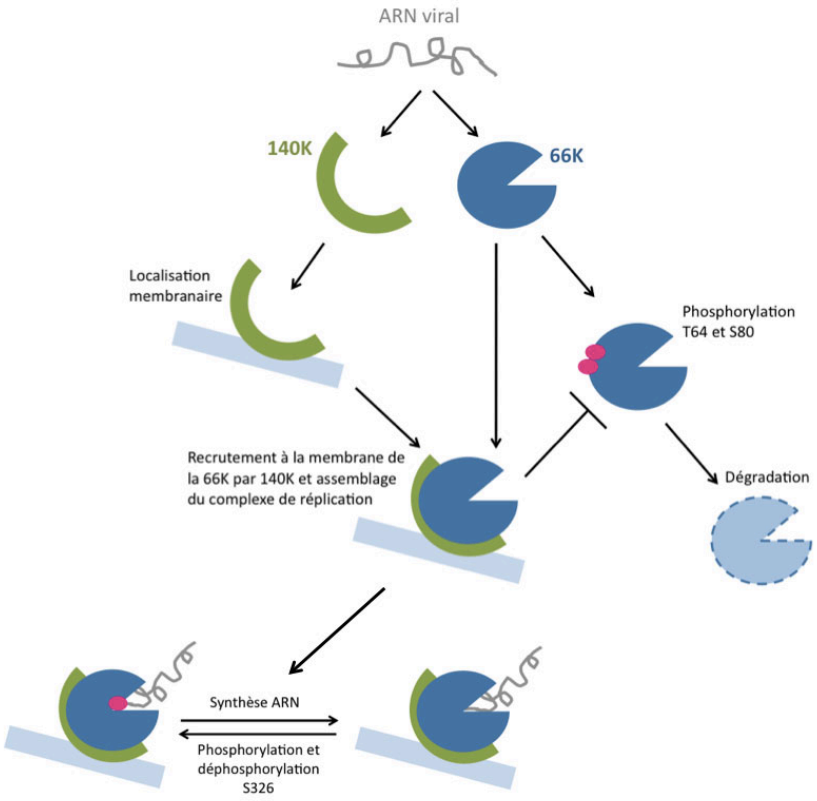



**Figure 6 : Régulation par phosphorylation réversible de la RdRp du TYMV**
A droite est représentée la phosphorylation (en rose) associée à la thréonine 64 et la sérine 80 et conduisant à la dégradation de la RdRp (en bleu foncé). Dans la seconde partie est représenté le mécanisme de protection lié à la protéase (en vert) qui peut inhiber cette dégradation par recrutement de la RdRp au niveau des structures membranaires chloroplastiques (en bleu clair) lors de la formation du complexe de réplication. Dans la partie inférieure du schéma, la phosphorylation réversible de la sérine 326, régulant la synthèse d'ARN est finalement représentée.

*L'hélicase*

La deuxième protéine à être libérée lors de la maturation de la polyprotéine non-structurale 206K est la NTPase/hélicase. Cette enzyme n'est pas toujours présente chez les virus à ARN de polarité positive mais on la retrouve presque systématiquement chez ceux dont le génome dépasse 6 kb.
Les hélicases sont des enzymes capables de dissocier les brins d'acides nucléiques complémentaires. Pour cela, elles rompent les liaisons hydrogènes qui les relient. Elles ont généralement une activité NTPase ATP dépendante associée car, pour dérouler les duplex d'ARN ou d'ADN, l'hydrolyse de nucléosides 5'-triphosphates (NTP) est nécessaire.

Ces enzymes sont classées en 3 familles selon les motifs conservés qu'elles abritent, qui sont notés I, Ia, II, III, IV, V, et VI (Gorbalenya 1989). Lorsqu'une protéine possède au moins 3 de ces motifs, il est alors évident de la classer dans le genre des protéines de type hélicase.
Dans le cas de la NTPase/hélicase de TYMV, les 7 segments conservés sont présents et elle a donc pu être identifiée par homologie de séquence sans difficulté (Kadaré 1996). L'activité ATPase, GTPase et la capacité de liaison à l'ATP et à l'ARN ont par la suite été confirmée en laboratoire.
Au cours de la réplication, les hélicases peuvent jouer plusieurs rôles. Elles permettent notamment de relarguer les ARN néoformés de la matrice afin de laisser la place à une nouvelle synthèse concomitante et permettent de défaire les structures double brins de la matrice et parfois de dérouler l'extrémité 3' pour la réplication.

*Le dernier produit de clivage : la méthyltransférase et la protéase*

Finalement la maturation de la 206K conduit à la formation d'un dernier produit de clivage de 140 kDa contenant plusieurs domaines fonctionnels. Une des fonctions de ce polypeptide est la localisation membranaire du complexe de réplication. En effet, il possède un domaine ayant de l'affinité pour les membranes capable de recruter ensuite la 66K via une interaction directe entre le domaine protéase du produit de clivage précédent et la partie N-terminale de la 66K (Figure 7).
On trouve dans ce produit de clivage, en plus du domaine protéase, le domaine riche en proline et le domaine méthyltransférase. Ce dernier domaine a lui aussi été identifié par homologie de séquence après comparaison avec d'autres polypeptides précurseurs de virus à ARN(+).
En revanche, son activité n'a pas été démontrée par la suite en laboratoire. Peu d'informations sont connues pour cette enzyme. On sait qu'elle joue un rôle dans la méthylation de la coiffe et, que chez le TYMV, l'extrémité 5' concernée est coiffée en $^{m7}$GpppG lors de la synthèse ARN. A l'inverse, la protéase située en C-terminal a été plus largement étudiée.

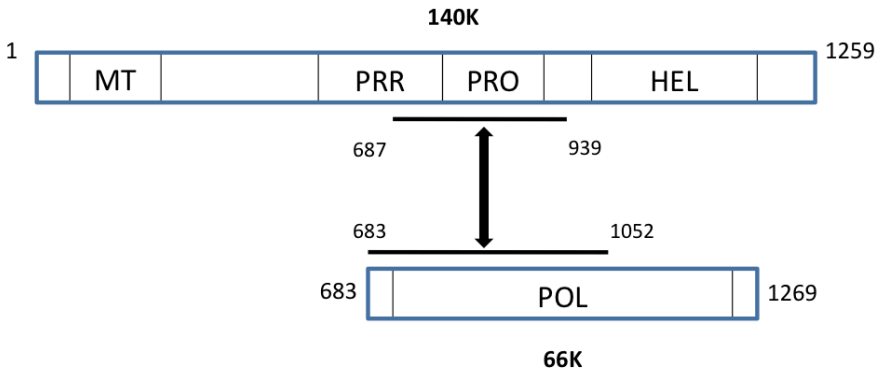



**Figure 7 : Intéraction RdRp/protéase lors de la formation du complexe de réplication**
Lors du recrutement à la membrane de la polymérase pour former le complexe de réplication, la région d'interaction a été localisée dans le segment des résidus 687 à 939 au niveau du produit de clivage 140K. Cette région se lie à la partie N-terminale de la 66K englobant la séquence PEST et donc la T64 et la S80.

La protéase est l'enzyme permettant la maturation de toutes les protéines citées précédemment impliquées dans le complexe de réplication. Les protéases sont des enzymes qui catalysent l'hydrolyse des liaisons peptidiques. On les classe en fonction de leur site actif et des acides aminés qui composent celui-ci. Ainsi, on trouve les protéases à sérine, les protéases à cystéine, les protéases à thréonine, les métallo-protéases et les protéases aspartiques.

Les protéases à cystéines sont les plus nombreuses et interviennent dans divers processus cellulaires aussi bien chez les bactéries que les virus ou encore les eucaryotes. Au sein de cette classe, on trouve plusieurs sous-catégories selon qu'on les identifie par clan ou par famille. Les clans se basent sur l'évolution ce qui entraîne qu'au sein d'un même clan, des protéases n'ayant que très peu de similarités au niveau de leur séquence peuvent être retrouvées. En revanche, les familles s'appuient sur une évolution identifiée au niveau de la séquence.

La famille des protéases de type papaïne est majoritairement représentée chez tous les règnes cités précédemment. Ces protéases ont un site catalytique caractéristique composé au minimum d'une diade qui comprend une cystéine et une histidine. Parfois, une asparagine s'ajoute à ces deux résidus pour former une triade catalytique. Les résidus $Cys^{783}$ et $His^{869}$ constituent la diade catalytique de la protéase du TYMV et ont permis de l'assigner à la famille des protéases à cystéine lors de la recherche par homologie de séquence.

Lors de la réaction d'hydrolyse de la chaine peptidique, l'histidine va tout d'abord activer la cystéine catalytique. Cette activation se traduit par la déprotonation de son groupement thiol. Une fois activée,

elle va pouvoir réaliser une attaque nucléophile sur le carbone de la liaison peptidique à couper de la protéine cible. Afin de rammener l'histidine catalytique à son état d'origine, une déprotonation va ensuite avoir lieu permettant dans le même temps la coupure du subtsrat. Un produit de clivage est ainsi libéré tandis qu'un autre forme un intermédiaire de type thioester avec la protéase. Une molécule d'eau va finalement permettre l'hydrolyse de cet intermédiaire et le retour à l'état protonée de la cystéine catalytique. Ainsi, le relargage du deuxième produit de clivage est achevé (Figure 8).

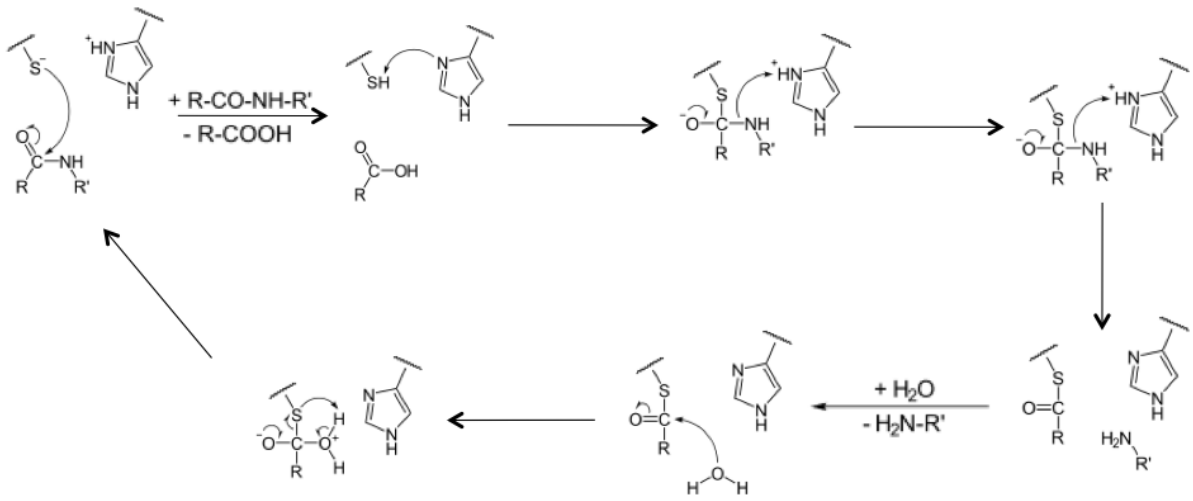



**Figure 8 : Mécanisme catalytique des protéases à cystéine**

Nos collaborateurs ont récemment mis en évidence un second rôle pour cette enzyme. En effet, une activité ubiquitine hydrolase ou déubiquitinase (DUB) lui a été attribuée (Chenon 2011). Au cours de cette étude, il a été montré que cette seconde fonction permettait une régulation forte de la réplication. En effet, la 66K, ubiquitinilée en vue de sa dégradation, peut, grâce à cette DUB, être stabilisée. PRO est capable de cliver les chaînes de polyubiquitine liées à la 66K et permet alors de déjouer son adressage au protéasome. De plus, lors du recrutement de la 66K, la protéase inhibe la phosphorylation, probablement par masquage des sites concernés lors de l'interaction directe entre les deux protéines (Figure 9).

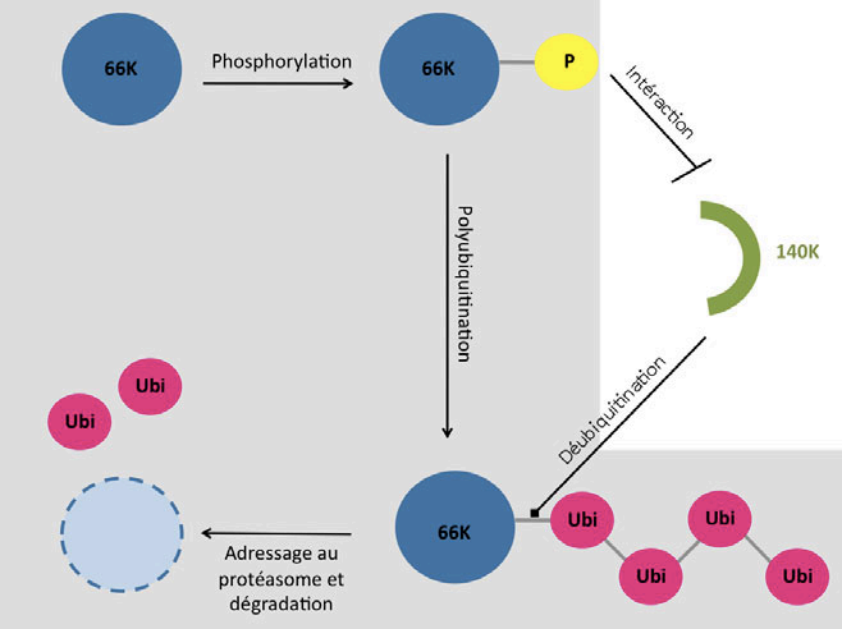



**Figure 9 : Mécanisme de régulation de la RdRp conduisant à sa dégradation**
La polymérase du TYMV (en bleu foncé) est phosphorylée (représentation en jaune), au cours du cycle viral en vue de sa régulation. La phosphorylation de la thréonine 64 et de la sérine 80 constitue un signal pour la polyubiquitination (représentée par une chaine d'ubiquitines roses), entraînant alors son adressage au protéasome où elle sera dégradée. En parallèle, son interaction avec la protéase (en vert) peut masquer ces sites de phosphorylation et inhiber ce mécanisme dégradatif. Si la phosphorylation et la polyubiquitination ont déjà eu lieu, la protéase peut alors cliver l'étiquette polyubiquitine et contre-carrer l'adressage au protéasome.

L'ubiquitination et la déubiquitination sont des moyens très complexes de réguler le cycle viral. L'hôte comme le virus peut y avoir recours pour réaliser divers processus biologiques ou se défendre et ce, de multiples façons. Au vu des derniers éléments mis en évidence par nos collaborateurs sur les interconnexions régulatrices entre la polymérase et la protéase du TYMV, nous nous sommes intéressé de plus près à ces deux enzymes et aux phénomènes de phosphorylation et d'ubiquitination qui leur sont liés.

# L'ubiquitination et la déubiquitination

## L'ubiquitination

L'ubiquitine est une petite protéine composée de 76 acides aminés et faisant 8,5 kDa. Grâce à différentes enzymes, elle est capable de lier les résidus lysine de protéines cibles servant alors d'étiquette reconnue lors de différents processus biologiques (Pickart 2004). Pour cela, l'extrémité C-terminale de l'ubiquitine est tout d'abord activée par E1 une « ubiquitin activating enzyme » qui s'y associe via une liaison thioester ATP dépendante au niveau de sa cystéine catalytique. Une deuxième enzyme E2 appelée « ubiquitin conjugating enzyme » accueille ensuite cet intermédiaire et s'y attache à son tour par une liaison thioester au niveau de sa cystéine catalytique. Finalement, une dernière enzyme E3 va permettre l'attachement de l'ubiquitine à la protéine cible. Cette « ubiquitin protein ligase enzyme » catalyse la formation d'une liaison isopeptidique entre une lysine de la protéine à étiqueter et le dernier acide aminé de l'ubiquitine : la glycine 76. L'ensemble de ces étapes constitue l'ubiquitination (Passemore 2004) (Figure 10).
L'ubiquitination varie selon le devenir des protéines ubiquitinilées. On en distingue 3 types généraux : la mono-ubiquitination lorsqu'une seule ubiquitine est attachée sur la protéine cible, la poly-ubiquitination si c'est une chaîne d'ubiquitines qui s'associe sur une lysine et la multi-ubiquitination si la protéine porte des ubiquitines sur plusieurs de ses résidus lysine (Figure 11). Un jeu de combinaison entre les différentes variétés d'enzymes E2 et E3 permet de moduler le type d'ubiquitination. Ainsi, selon le nombre d'ubiquitines liées et le mode de fixation sur la cible, les protéines ne vont pas avoir la même destinée (Clague 2010).

Parmi les fonctions de l'ubiquitination, on trouve la relocalisation des protéines dans la cellule, leur activation ou leur dégradation, la réparation de l'ADN, le contrôle de la transcription de gènes, la transduction du signal, le transport membranaire des protéines, l'autophagie ou encore la réponse immunitaire (Haglund 2005, Welchman 2005, Azzo 2005). La première fonction attribuée à l'ubiquitination et la plus étudiée jusqu'à présent a été l'envoi vers le protéasome des protéines à dégrader.

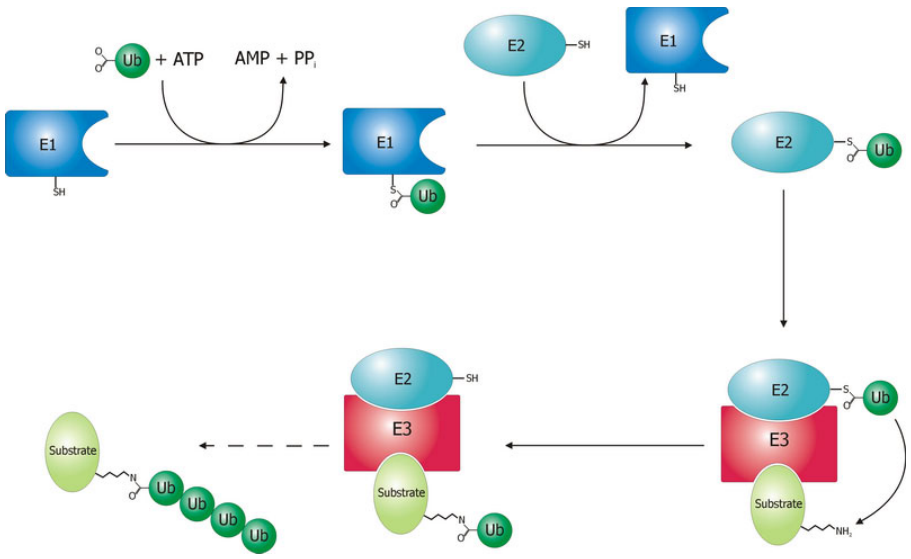



**Figure 10 : Mécanisme d'ubiquitination des protéines substrat**
L'extrémité C-terminale de l'ubiquitine (en vert) à est activée par E1 (en bleu foncé) grâce à une liaison thioester ATP dépendante au niveau de sa cystéine catalytique. Ensuite, E2 (en bleu clair) remplace l'enzyme E1 de façon similaire, mais l'énergie provient cette fois de l'hydrolyse de la liason E1/substrat. Finalement, E3 (en rouge) permet de relier l'intermédiaire E2/ubiquitine et la protéine cible (en vert clair). La catalyse de la formation d'une liaison isopeptidique entre une lysine de la protéine à étiquetter et le dernier acide aminé de l'ubiquitine : la glycine 76 est alors réalisé. Cette ubiquitination peut être simple ou multiple.

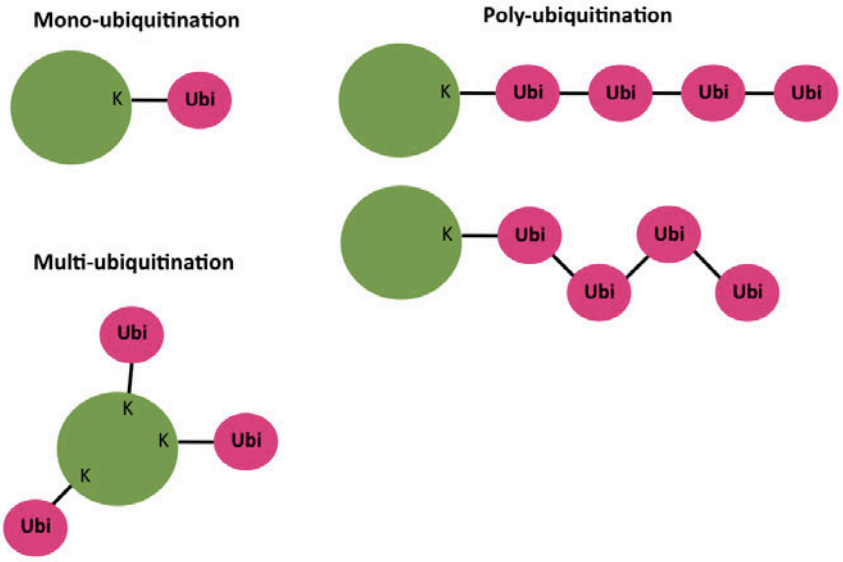



**Figure 11 : Différentes possibilité d'ubiquitination des protéines cibles.**
En vert est schématisé la cible protéique à étiquetter et en rose les molécules d'ubiquitine se liant aux lysines K.

## Le système de dégradation ubiquitine/protéasome

La dégradation par le protéasome des protéines ubiquitinilées a une influence dans de nombreux mécanismes et en particulier lors du cycle cellulaire, de l'apoptose, de la réponse immunitaire, de l'expression de gènes, de la signalisation ou encore de la réparation de l'ADN. Cette dégradation sert généralement à l'élimination de protéines mal repliées ou encore à la régulation quantitative, au cours de ces processus cellulaires, d'une protéine clé.

Pour qu'une protéine soit adressée vers le protéasome, il faut qu'une polyubiquitination d'au moins quatre ubiquitines ait lieu (Lodish 2008). Une fois cette étiquette mise en place, la protéine à dégrader est envoyée vers un complexe multi-enzymatique : le protéasome. Ce complexe riche en protéase va reconnaître la protéine et va la cliver en plusieurs peptides de petite taille grâce à un mécanisme ATP dépendant.

## La déubiquitination

L'ubiquitination est un processus réversible. Il est en effet possible d'hydrolyser la liaison covalente liant l'ubiquitine à sa protéine cible. Cela permet notamment de recycler l'ubiquitine libre dans la cellule et de réguler les nombreux processus biologiques cités précédemment (Reyes-Turcu 2009).

Tout comme l'ubiquitination, sa réaction inverse est un moyen de régulation puissant et certains virus ont d'ailleurs su en tirer avantage. Les enzymes assurant cette fonction sont appelées ubiquitine hydrolases ou déubiquitinases (DUB). Ce sont donc elles qui permettent l'hydrolyse des précurseurs d'ubiquitine et des étiquettes d'ubiquitine après l'ubiquitination (Figure 12).

Certaines endopeptidases impliquées dans la maturation des précurseurs polyprotéiques sont également capables de lier d'autres substrats de type ubiquitine et avoir ainsi une activité déubiquitinase supplémentaire. Cela a notamment été montré pour certaines protéases virales comme celle du virus de l'hépatite E (Karpe 2011), du virus de la fièvre aphteuse (Wang 2011) ou encore du virus du syndrome respiratoire aigu sévère (Barretto 2006). D'autres virus codent également pour des déubiquitinases strictes comme le virus de la Fièvre hémorragique de Crimée-Congo (Capodagli 2011) ou encore celui du virus de l'Herpes (Kattenhorn 2005).

L'activité ubiquitine hydrolase de ces protéases virales a été identifiée notamment grâce à des outils mis au point pour la détection spécifique de l'activité des DUBs à cystéine basés sur des substrats suicides

(Borodovsky 2002, Hemelaar 2004, Love 2007). Cette technique de détection se sert d'ubiquitines modifiées auxquelles un groupe fonctionnel est greffé, en faisant donc un substrat suicide. Il suffit alors de mettre en présence la protéase à cystéine dont on suspecte une activité DUB et une ubiquitine greffée du bon groupe pour avoir une réponse immédiate et rapide. Cette technique sera plus amplement présentée dans les compléments de résultats de la Partie II. Grâce en partie à ces outils, la protéase de TYMV est venue s'ajouter à la liste des protéases virales multifonctions.

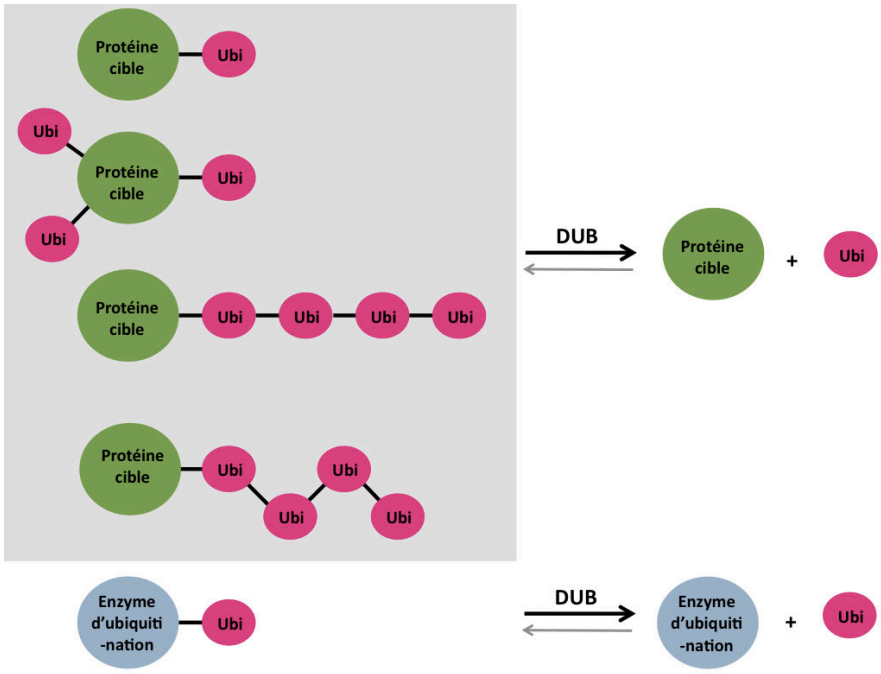



**Figure 12 : Différentes cibles reconnues par les déubiquitinases et déubiquitination**
En vert est schématisée la cible protéique étiquetter par les molécules d'ubiquitine en rose. Les enzymes liant l'ubiquitine lors de l'ubiquitination sont quant à elles représentées en bleu.

## L'ubiquitination et la déubiquitination chez les virus

Plusieurs études récentes montrent que les virus ont su utiliser l'ubiquitination de la cellule hôte afin d'en tirer avantage pour leurs propres processus biologiques ainsi que pour se défendre des actions antivirales. Une liste non exhaustive de DUBs rencontrées dans le monde viral et de leurs rôles divers lors de la réplication est présentée Table 1.

Dans le cas du TYMV, notre virus modèle, on sait que la protéase est capable de stabiliser la polymérase en déjouant son adressage au protéasome. Cette déubiquitinase / endopeptidase est actuellement la

seule protéase virale dont la double fonction a été parfaitement caractérisée, aussi bien in-vitro qu'in-vivo et dont on connaisse le substrat naturel (Camborde 2010, Chenon 2011).

## La protéase du TYMV et les OTU protéases

La protéase de TYMV n'a pu être assignée à un clan dans la base de donnée MEROPS (http://http://merops.sanger.ac.uk). Cette base de données est spécialisée dans la classification des protéases. S'appuyant sur les similitudes existantes au niveau de la séquence, elle attribue tout d'abord chaque protéase enregistrée à une famille. Si aucune homologie n'est assignable avec certitude par comparaison de séquences, une nouvelle famille est établie. C'est le cas pour la PRO/DUB des tymovirus, seul membre de la famille C21. Puis, si une homologie plus lointaine peut être établie par comparaison des structures de membres de deux familles, ces familles sont regroupées dans des clans.
Chez le TYMV, des résidus caractéristiques autour des résidus constituant la diade catalytique ont donné à penser qu'elle pourrait être homologue aux déubiquitinases de type OTU appartenant au clan CA et arborant également ces résidus conservés (Makarova 2000). Mais le motif propre à ces déubiquitinases n'est pas retrouvé dans son ensemble et il n'est donc pas possible d'affirmer cette homologie. Ainsi, la famille C21 est restée sans assignation de clan.
Les déubiquitinases OTU appartiennent à la superfamille des protéines à domaine OTU. Cette superfamille tire son nom et son origine de la découverte du gène otu (ovarian tumor). Ce gène a était mis en évidence chez la drosophile, où il est nécessaire à la formation des ovaires. Au sein de celui-ci, une zone code pour une région d'environ 130 acides aminés, très conservée chez plusieurs espèces eucaryotes, bactériennes et virales. Cette région porte l'activité DUB (Glenn 2001) et est très semblable aux protéases à cystéine de type papaïne. Le mécanisme catalytique est ainsi le même que celui présenté Figure 8.

Les OTU ubiquitines hydrolases interviennent notamment dans la protection des cellules de l'apoptose, l'inhibition de l'activation des cellules lymphoïdes par régulation négative de la voie NFκB en contrôlant l'état d'ubiquitination des protéines impliquées dans la transduction du signal (Evans

2004) ou encore dans l'action contre la réponse inflammatoire et anti-virale de la cellule hôte dans le cas d'une infection virale (James 2011). Leur étude et la résolution de leur structure intéressent donc une large communauté scientifique.

Ainsi, l'étude de la protéase du TYMV, notre virus modèle, a été entreprise au cours de cette thèse ainsi que celle de la polymérase. Les fonctions de ces deux enzymes sont étroitement liées à différents stades du cycle viral et leurs études structurales simultanées nous a grandement intéressé afin d'identifier notamment ces interconnections. Leurs rôles dans la réplication étant essentiels et ce virus étant modèle, elles ont été largement étudiées d'un point de vue cellulaire lors de précédentes études. De solides bases sont donc plantées pour nous permettre de comprendre comment se déroulent précisément leurs fonctions au cours du cycle viral, d'un point de vue structural. Pour cette étude, nous avons eu recours alors à une méthode de résolution de structure tridimensionnelle appellée « biocristallographie aux rayons X ».

Table 1 : Principales DUB rencontrées dans le monde viral

| DUBs<br>- virus - | Processus biologiques impliqués ou suspectés | Références |
|---|---|---|
| PLpro<br>- SARS -<br>- Fièvre aphteuse - | Stabilisation de protéines virales en déjouant l'adressage au protéasome<br>Régulation négative de la réponse interféron | Harcourt et al. 2004<br>Lindner et al. 2005 |
| UL36<br>- HSV1 - | Stabilisation de protéines virales | Kattenhorn et al. 2005 |
| UL36<br>- MDV - | Pathogenèse<br>Formation des cellules lymphoïdes T | Jarosinski et al. 2007 |
| Prv-pUL36<br>- PRV - | Pathogenèse<br>Neuro-invasion<br>Sortie du virion | Bottcher et al. 2008 |
| BPLF1<br>- EBV - | Stabilisation de la β-cathenin dans les cellules infectées en latence<br>Augmentation de l'activité de la ribonucléotide réductase | Whitehurst et al. 2009 |
| HBx<br>- HBV - | Régulation négative de la réponse interféron | Jiang et al. 2010 |
| USP11<br>- Influenza - | Déubiquitination de la protéine NP<br>Régulation de la réplication des ARN | Liao et al. 2010 |
| $PLP1^{pro}$<br>- TGEV - | Stabilisation de protéines virales en déjouant l'adressage au protéasome<br>Régulation négative de la réponse interféron | Snijder et al. 2010 |
| nsp2<br>- PRRSV - | Inhibition de la voie NFκB | Wojdyla et al. 2010 |
| PRO<br>- TYMV - | Stabilisation de l'ARN-polymérase ARN-dépendante en inhibant son adressage au protéasome | Chenon et al. (2011) |

# La bio-cristallographie

La fonction d'une protéine est généralement reliée à sa structure tridimensionnelle et la connaissance de sa structure permet souvent de mieux comprendre le mécanisme fonctionnel associé à cette protéine. Grâce à la bio-cristallographie, une méthode biophysique, on peut accéder à la structure tridimensionnelle des macromolécules biologiques à une résolution atomique. Cette méthode repose sur la diffraction des rayons X par un cristal de protéine pure. Cela implique en premier lieu l'obtention d'une préparation pure de la protéine d'intérêt, qui va ensuite être cristallisée. Les cristaux recherchés doivent avoir une taille et des propriétés de diffraction convenables. Après leur passage dans un faisceau de rayons X, les données collectées vont alors êtres analysées afin de reconstruire la structure et plus précisément déterminer les coordonnées spatiales des atomes constituant la protéine. Pour y parvenir il faudra toutefois résoudre ce qu'on appelle le problème des phases.

## La cristallogenèse

*Principe*

Un cristal est un objet périodique composé d'éléments répétés de façon ordonnée dans tout son volume. Dans le cas qui nous intéresse ici, l'élément constituant la base du cristal ou l'empilement cristallin est la protéine d'intérêt. La cristallogenèse consiste donc à trouver les conditions qui vont permettre à la protéine étudiée de précipiter de façon ordonnée lors de son passage de l'état soluble à l'état solide. C'est l'étape la plus incertaine et souvent la plus limitante. En effet, il n'existe pas de règle stricte pour accéder directement à ces conditions. La croissance cristalline va dépendre des propriétés intrinsèques de chaque protéine et il faut donc procéder par tâtonnement jusqu'à l'obtention des premiers signes cristallins. Ces signes seront ensuite affinés afin d'aboutir à une condition où la protéine va former un réseau tridimensionnel suffisamment important pour l'obtention de cristaux de plus grand volume et utilisables pour l'expérience de diffraction des rayons X.

Pour que ce réseau se forme, il va falloir diminuer les interactions entre le solvant et la protéine au profit d'interactions protéine-protéine non covalentes. Les paramètres favorisants ces interactions et sur lesquels on peut jouer sont nombreux. Tout d'abord, on peut moduler les paramètres directs de la protéine comme sa concentration et sa pureté. On peut également faire varier les paramètres de l'expérience comme la température, le pH, la présence d'agents précipitants ou d'additifs. Pour réaliser cette mise au point, plusieurs méthodes ont été développées.

*Technique*

La technique la plus employée et utilisée ici pour la cristallisation de la protéase du TYMV est la méthode de diffusion de vapeur (Mc Pherson 1995). Elle repose sur l'équilibration entre une goutte constituée de la solution pure de la protéine d'intérêt et d'une solution de précipitation et un réservoir contenant un grand volume de la solution de précipitation. Le dispositif est fermé hermétiquement, de sorte à produire une diffusion de vapeur entre la goutte et le réservoir. Un déplacement d'équilibre se produit ainsi et plus généralement par évaporation d'eau de la goutte vers le réservoir puisque ce dernier possède une concentration en agent précipitant plus importante. Ce phénomène entraîne finalement la diminution du volume de la goutte, augmentant alors la concentration de la protéine et de l'agent précipitant. Au cours de cette concentration, on peut, dans les cas favorables, atteindre une condition donnant naissance à des cristaux. L'équilibre peut également être déplacé par évaporation de composés volatiles dans la goutte. En pratique, on peu suspendre la goutte sur une lamelle retournée au dessus du réservoir dans le cas de la méthode de la goutte suspendue ou alors la déposer dans un puits sur la paroi du réservoir dans le cas de la goutte assise (Figure 13).

Au laboratoire, deux techniques sont mises en œuvre classiquement pour réaliser ce phénomène de diffusion de vapeur. Dans un premier temps, des kits, constitués de 96 puits contenant différentes solutions de précipitations, sont testés au moyen d'un robot afin de couvrir rapidement un large spectre de solutions. La méthode de la goutte assise est utilisée dans ce cas. Puis, une fois les premiers signes de cristallisation observés, une phase d'optimisation est réalisée en plus grand volume, de façon manuelle, via la technique de la goutte suspendue. Cette fois, des boîtes de 24 puits sont utilisées.

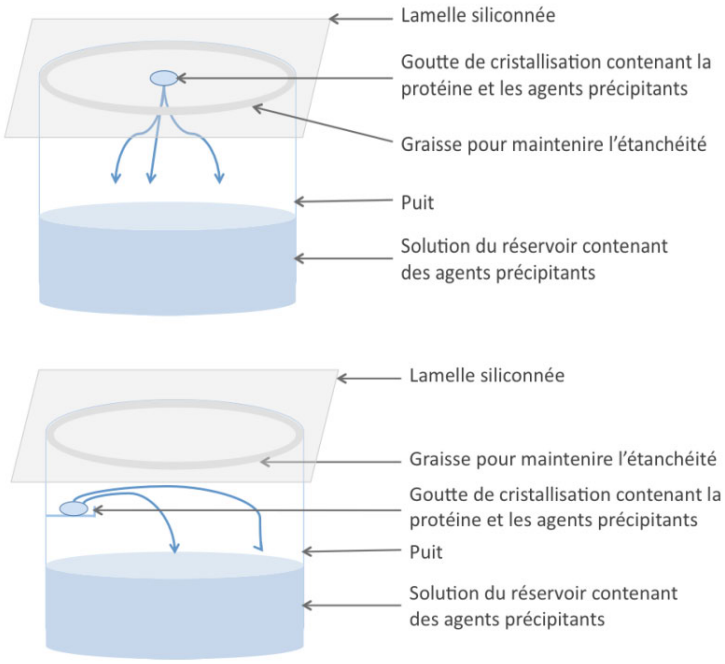



**Figure 13 : Représentation des méthodes utilisant le phénomène de diffusion de vapeur**
Au dessus, la méthode de la goutte suspendue sur une lamelle retournée afin qu'elle se retrouve en vis à vis du réservoir. En dessous, la méthode de la goutte assise où la goutte est déposée dans un puit sur la paroi du réservoir.

Une fois les cristaux d'apparence et de taille convenable obtenus, l'expérience de diffraction des rayons X peut commencer.

## Préparation des cristaux

Avant toute analyse par diffraction, il va falloir préparer les cristaux en fonction de la source de rayons X. Pour cette étude, seule une source synchrotron a été employée et, dans ce cas, les cristaux ont été placés sur un support spécifique permettant de les positionner correctement devant le faisceau de rayons X mais aussi de les transporter et de les préparer à l'expérience.
La préparation des cristaux est effectivement nécessaire car ce sont des objets fragiles, qui subissent des dommages lors du passage des rayons X. Ces dommages sont dûs principalement aux radiations ionisantes et aux réactions radicalaires engendrées par ces rayons (Burmeister 2000, Ravelli 2000).

Afin de limiter la dégradation causée, on réalise alors l'expérience de diffraction à une température de 100K en projetant continuellement un flux d'azote sur le cristal. Toutefois, cette étape peut provoquer la formation de cristaux parasites de glace qui peuvent gêner la diffraction voire détruire le cristal. Pour limiter la formation de glace, on congèle rapidement les cristaux dans l'azote liquide afin de vitrifier le solvant présent autour du et dans le cristal. Ainsi piégé dans un état amorphe, le solvant ne viendra pas ou peu gêner la diffraction (Mayer 1985, Dubochet 1988). Cette étape de vitrification peut être optimisée grâce à l'utilisation d'agents cryo-protectants ralentissant la formation de glace tels que le glycérol.
Une fois les cristaux disposés sur leur support et congelés convenablement, on peut les placer au travers du faisceau de rayons X afin de collecter les données issues de la diffraction.

## Diffraction des rayons X

La diffusion des rayons X engendrée par une seule protéine aboutit à un signal trop faible pour pouvoir être détecté et analysé. La solution consiste donc à faire diffuser simultanément un grand nombre de protéines afin d'additionner leurs signaux et rendre le total détectable et exploitable. Pour que l'addition soit possible, les protéines doivent diffuser de façon similaire, c'est à dire être ordonnée dans l'espace. Le cristal est ainsi utilisé.
Nous avons vu précédemment qu'un cristal est formé par la répétition d'un motif dans les trois dimensions de l'espace. Le plus petit élément retrouvé par translation tout le long du réseau cristallin est appelé maille élémentaire. Ces répétitions de la maille définissent des plans parallèles caractérisés par les indices de Miller h, k et l. Ces plans réticulaires sont situés à des distances définies $d_{hkl}$ qui dépendent de la disposition et de la taille des atomes dans le réseau cristallin.
Le phénomène de diffraction se produit lorsque les rayons X incidents, monochromatiques et parallèles, vont frapper une même famille de plans réticulaires et être émis en phase. Leurs diffusions vont dans ce cas, s'additionner et conduire à une onde émise constructive. Ce sont ces ondes en phase que l'on mesure et qui conduisent à une tache de diffraction chaque fois qu'elles rencontrent le détecteur. L'ensemble des taches collectées constitue finalement le cliché de diffraction qui sera interprété par la suite.

La diffraction se produit lorsque la loi de Bragg est respectée : $\mathbf{2d_{hkl}\ \sin\theta = n\lambda}$ avec $\lambda$ la longueur d'onde du faisceau incident, $\theta$ l'angle du faisceau incident et n un nombre entier (Figure 14).

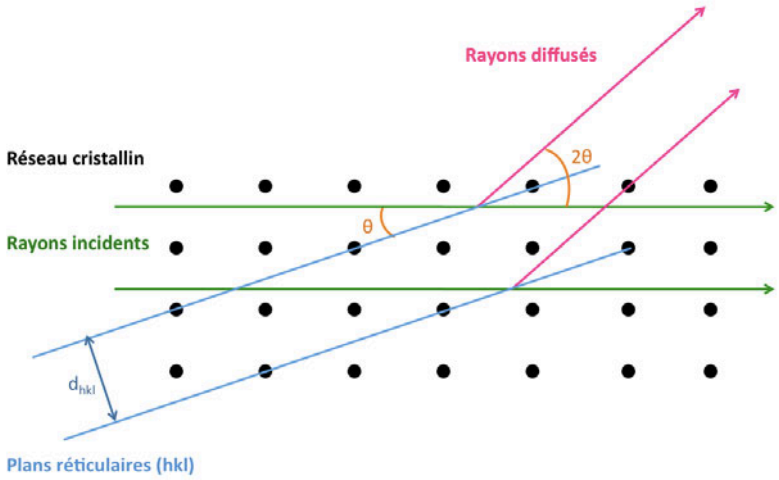



Figure 14 : Représentation de la loi de Bragg
Réflexion des rayons X par une même famille de plans réticulaires espacés d'une distance $d_{hkl}$.

## Espace réciproque, espace direct et transformée de Fourier

Il existe une relation exacte entre l'ensemble des ondes diffractées dans l'espace réciproque décrites ci-dessus et la densité électronique du cristal dans l'espace direct : c'est la transformée de Fourier. Ainsi, la fonction de densité électronique peut s'écrire :

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl}\, e^{-2\pi i\,(hx+ky+lz)}$$

Avec (x, y et z) les coordonnées dans l'espace direct, (h, k et l) les indices de Miller, V le volume de la maille du cristal et $F_{hkl}$ le facteur de structure caractérisant les ondes diffractées.

Reconstituer la densité électronique du cristal équivaut alors à déterminer l'ensemble des ondes diffractées qui, comme toutes ondes, sont définies par une amplitude $|F_{hkl}|$ et une phase $\phi_{hkl}$. Ces deux paramètres sont reliés au facteur de structure $F_{hkl}$ grâce à la relation :

$$\mathbf{F_{hkl} = |F_{hkl}|\, e^{i\phi hkl}}$$

Pour l'amplitude $|F_{hkl}|$, on tire l'information directement du cliché de diffraction. En effet, le détecteur de rayons X mesure les intensités des taches et donc des ondes diffractées $I_{hkl}$ (Figure 15). Or $I_{hkl}$ est directement proportionnelle au carré de $|F_{hkl}|$ via la relation :

$$\mathbf{I_{hkl} = k|F_{hkl}|^2}$$

En revanche, l'information relative à $\phi_{hkl}$ est perdue au cours de l'expérience de diffraction et il va falloir la retrouver. C'est ce qu'on appelle résoudre le problème des phases.

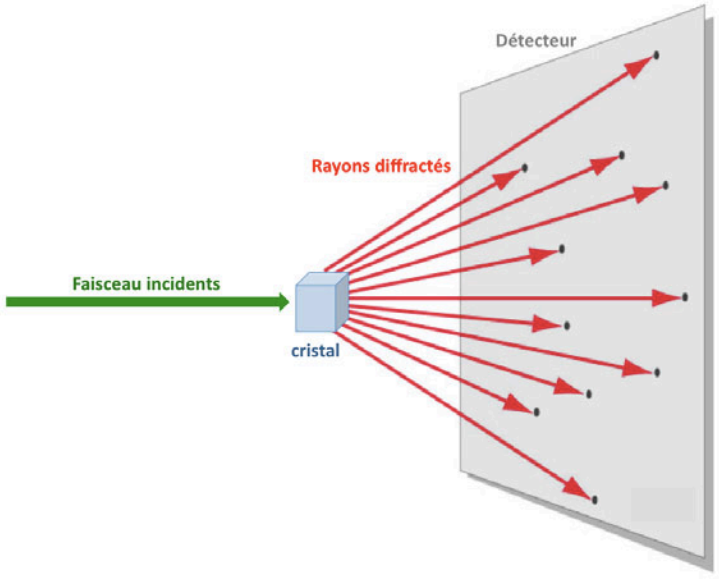



**Figure 15 : Principe de la diffraction des rayons X en bio-cristallographie**
Un faisceau incident monochromatique et parallèle (en vert) est envoyé sur le cristal (en bleu). A la rencontre des électrons des atomes constituant la molécule cristallisée, les rayons X vont être déviés dans toutes les directions de l'espace. On place alors un détecteur (en gris) derrière le cristal afin de collecter les taches (en noir) correspondantes aux rayons diffusés en phase, soit les rayons diffractés (en rouge).

## Le problème des phases

Pour trouver l'information manquante relative à la phase, plusieurs stratégies existent. Les plus utilisées en biologie sont : le phasage par remplacement moléculaire, par remplacement isomorphe ou par utilisation de la diffusion anomale.

*Le remplacement moléculaire* (Rossmann 1962 et 1972, Machin 1985, Navaza 1994)

Le remplacement moléculaire n'est possible que si l'on possède un modèle homologue à la protéine cristallisée. Une telle homologie est certaine si le modèle a au moins 30 % d'identité de séquence avec la protéine cristallisée.

Si l'on dispose d'un tel modèle, on va pouvoir se servir de sa structure comme première approximation pour celle de la protéine étudiée. De cette façon, on va positionner le modèle dans la maille et on va calculer les phases et les amplitudes correspondantes à un tel positionnement. On va ensuite renouveler cette opération, par un jeu de rotations et de translations, afin de se rapprocher le plus précisément possible des informations mesurées lors de l'expérience de diffraction. Lorsque l'accord entre les données calculées et les données mesurées est suffisamment bon, on considère que le positionnement du modèle dans la maille reflète assez précisément ce que l'on a en réalité et, on peut se servir des phases calculées et des intensités mesurées pour retrouver le facteur de structure. Celui-ci est donc seulement une approximation et la transformée de Fourier inverse permettant de remonter à la densité électronique sera donc elle aussi approximée. Cette approximation conduit cependant à une densité électronique suffisamment proche de la densité recherchée. A partir de celle-ci, on va donc pouvoir y reconstruire la structure de notre protéine et une série d'affinements permettra de corriger les erreurs faites lors de l'obtention de la densité électronique.

Bien que le nombre de structures résolues augmente chaque jour, on ne trouve pas toujours de modèle adéquat pour résoudre le problème des phases via cette technique. Parfois, aucun homologue de structure connue ne peut être détecté ou que les homologues trouvés peuvent s'avérer trop éloignés pour le remplacement moléculaire ou pour l'affinement à partir de la solution de remplacement moléculaire.

De plus, avant que des modèles ne soient disponibles, des techniques plus anciennes permettaient de résoudre le problème des phases. Ce sont sur celles-ci que l'on s'est orienté pour le cas de la protéase du TYMV puisque la recherche d'homologue à partir de sa structure primaire n'a pas abouti. Ces méthodes alternatives sont basées sur les propriétés de diffraction des atomes lourds et l'on trouve ainsi les méthodes de remplacement isomorphe ou celles utilisant la diffusion anomale pour arriver à résoudre le problème des phases.

*Le remplacement isomorphe*

La résolution du problème des phases par remplacement isomorphe nécessite l'incorporation d'atomes lourds. Ceux-ci sont soit ajoutés directement dans la protéine lors de son expression soit après, dans le cristal formé. Dans le premier cas, on remplace les méthionines de la protéine d'intérêt par des sélénométhionines. Ainsi, on connaît précisément le nombre et la localisation dans la séquence de ces atomes lourds. En revanche, dans le deuxième cas, les atomes lourds sont ajoutés dans une solution où l'on va tremper les cristaux. Comme les cristaux sont constitués pour moitié de solvant, les atomes lourds vont pouvoir s'infiltrer dans le réseau protéique et s'y fixer lorsqu'ils vont rencontrer un site approprié. Cette fois, leur fixation n'est pas garantie ni caractérisée. Une fois que l'on possédera de tels cristaux appelés dérivés, on va réaliser une expérience de diffraction et comparer alors les données expérimentales natives et dérivées.

Si les cristaux sont isomorphes, c'est à dire s'ils ont le même groupe d'espace et des paramètres de maille similaires, on peut alors relier les facteurs de structure correspondants aux différents intervenants comme suit : $\mathbf{F_{PH}} = \mathbf{F_P} + \mathbf{F_H}$ avec $F_{PH}$ le facteur de structure du dérivé, $\mathbf{F_P}$ le facteur de structure du cristal natif et $\mathbf{F_H}$ pour les atomes lourds uniquement. Afin de retrouver le facteur $F_P$ qui nous intéresse, on va tout d'abord déterminer expérimentalement les modules $|F_{PH}|$ et $|F_P|$ et déduire $|F_H|$ en observant les différences entre les clichés de diffraction (Figure 16).

Ensuite, grâce à la méthode de Patterson (Patterson 1934, Harker 1936) ou par le biais de méthodes directes, on va pouvoir déterminer la position des atomes lourds et déterminer la sous-structure qu'ils forment. Celle-ci va alors nous permettre de calculer la phase relative à $F_H$ qui sera alors déterminé en module et en phase. Finalement, on pourra remonter à $F_P$ et tracer ensuite la densité électronique correspondante à la protéine.

Toutefois, si l'on utilise la méthode SIR (Single Isomorphous Replacement), soit un seul dérivé lourd,

on aboutit à deux solutions possibles. En effet, on ne peut discerner la solution correspondante à la structure réelle de celle correspondant à son miroir. Afin de lever l'ambiguïté, on a alors recours à la méthode MIR (Multiple Isomorphous Replacement). Cela signifie qu'il faut au moins deux dérivés lourds différents et donc au minimum 3 cristaux isomorphes : le natif et les deux dérivés.

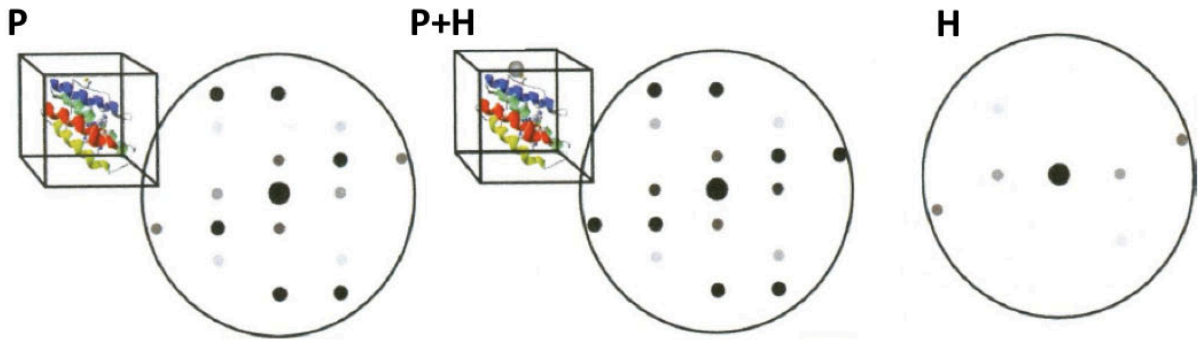



**Figure 16 : Obtention du signal de l'atome lourd**
A gauche est représenté la protéine native cristallisée dans l'unité asymétrique et le cliché de diffraction correspondant. Au centre, on trouve la même représentation mais pour la protéine contenant un atome lourd. Finalement, à gauche, on a la différence des intensités entre les clichés de diffractions représentés par la protéine + atome lourd et la protéine seule (Rhodes 2006).

Il existe finalement d'autres astuces pour retrouver les phases tout en s'affranchissant de la nécessité d'isomorphisme des cristaux. En effet, il est possible d'utiliser la diffusion anomale pour le phasage.

*Utilisation de la diffusion anomale* (Hendrickson 1991)

La diffusion des rayons X par les électrons libres d'un atome du cristal, entraîne un déphasage des rayons diffusés de 180° par rapport aux rayons incidents. Le facteur de diffusion f, pour un électron, est considéré réel. Ce facteur, pour un atome donné et à une longueur d'onde donnée, est une grandeur complexe, mais sa partie imaginaire est généralement négligeable. Cependant, la partie imaginaire du facteur de diffusion n'est plus négligeable dans certains cas, par exemple dans le cas où la longueur d'onde des rayons incidents est proche du seuil d'absorption d'un atome lourd j présent dans le cristal, c'est à dire lorsque la fréquence de l'onde incidente est proche d'une fréquence de transition électronique d'un atome j.

En effet, les rayons diffusés par ces atomes subissent un déphasage par rapport aux ondes de référence

diffusées par un électron à l'origine. Le facteur de diffusion s'écrit alors :

$f_\lambda = f_0 + f_\lambda' + i f_\lambda''$

avec $f_0$ le facteur de diffusion à l'origine, f' la composante réelle de f et f'' sa composante imaginaire, pour une longueur d'onde $\lambda$ donnée.

Le déphasage engendré brise alors la loi de Friedel, qui stipule que pour les paires de réflexions d'indice (hkl)=(+) et (-h-k-l)=(-), $|F_{hkl}| = |F_{-h-k-l}|$. En présence de diffusion anomale par un atome lourd H, les modules des facteurs de structure totaux $F_{PH}(+)$ et $F_{PH}(-)$ ne sont plus égaux :

$$\mathbf{F}_{PH}(+) = \mathbf{F}_P(+) + \mathbf{F}_H(+) + \mathbf{iF''}_H(+) \quad \textbf{et} \quad \mathbf{F}_{PH}(-) = \mathbf{F}_P(-) + \mathbf{F}_H(-) - \mathbf{iF''}_H(-)$$

Avec un raisonnement similaire à la méthode par remplacement isomorphe, il est alors possible de retrouver $F_P$ et donc de retrouver les phases.

On s'affranchit, avec l'utilisation de la diffusion anomale, de l'isomorphisme puisque l'on a uniquement besoin d'un cristal contenant un atome lourd. Les méthodes utilisant uniquement la diffusion anomale sont le SAD (Single anomalous scattering) si l'on collecte les données de diffraction à une seule longueur d'onde, et le MAD (Multiple Anomalous Scattering) si l'on fait varier la longueur d'onde au cours de l'expérience de diffraction.

Comme pour le SIR, après phasage avec la méthode SAD, on obtient deux solutions. En revanche, dans le cas du SAD, elles ne sont pas équivalentes et l'on peut généralement discerner la plus probable. Le MAD permet cependant d'obtenir des phases de plus haute qualité et n'aboutit qu'à la bonne solution. En revanche cette dernière méthode n'est réalisable que sur une source de lumière synchrotron où l'on peut faire varier la longueur d'onde du faisceau incident.

Certaines stratégies combinent également le remplacement isomorphe à l'utilisation de la diffraction anomale. On trouve par exemple le SIRAS (Single Isomorphous Replacment Anomalous Scattering) qui permet de lever l'ambiguïté résiduelle de la méthode SIR. Et, on trouve également le MIRAS (Multiple Isomorphous Replacement Anomalous Scattering) qui permet de compléter la méthode MIR. C'est d'ailleurs cette méthode qui nous a permis de résoudre le problème des phases dans le cas de l'étude structurale du domaine TYMVpro. Pour chacune des méthodes de phasage, on aboutit donc à la densité électronique. Mais celle-ci doit être améliorée et il faut reconstruire la structure protéique dans cette densité.

### Reconstruction de la structure

Les phases initiales contiennent généralement des biais dues aux erreurs et au caractère incomplet de la structure initiale dans le cas du remplacement moléculaire, aux défauts d'isomorphisme entre cristal natif et cristal(aux) dérivé(s) dans le cas des méthodes MIR/MIRAS/SIR/SIRAS, aux dégâts d'irradiation et aux erreurs de mesure de façon générale. On doit, avant toute reconstruction structurale, les améliorer afin d'obtenir une densité électronique correspondante interprétable. Plusieurs méthodes existent pour cette étape et l'on peut d'ailleurs en combiner plusieurs pour aboutir à un meilleur résultat. Parmis ces méthodes, les plus communes sont les méthodes d'optimisation basée sur l'aplatissement de solvant (Wang 1985), la moyennation des molécules reliées par une symétrie non cristallographique ou sur l'histogramme de densité (Zhang 1990). Une fois ces méthodes appliquées, on aboutit alors à une carte de densité améliorée que l'on va pouvoir interpréter afin d'y reconstruire la structure tridimensionnelle de la protéine cristallisée.

La première reconstruction va aboutir à un modèle structural approximatif qu'il va falloir ensuite affiner afin d'améliorer l'accord entre les facteurs de structure observés ($F_O$) et ceux calculés ($F_C$) à partir de ce modèle. Pour ce faire, on alterne reconstruction manuelle par le biais de programme de visualisation et affinement numérique des paramètres du modèle reconstruit tel que la position des atomes et leur agitation thermique.

Pour nous aider lors de la reconstruction manuelle d'un modèle, on utilise des cartes de densité calculées à partir du modèle précédent telles que les cartes de densité électronique $2mF_O\text{-}DF_C$ et $mF_O\text{-}DF_C$. On note respectivement ces cartes de manière simplifiée $2F_O\text{-}F_C$ et Fo-Fc. La figure de mérite m est une estimation de la précision de la phase et, avec le coefficient D, elle permet de prendre en compte les erreurs apportées par le modèle incomplet et par l'inexactitude de la position des atomes reconstruits.

Ainsi, la carte $2F_O\text{-}F_C$ correspond à la carte de densité du nouveau modèle à résoudre et prend en compte les erreurs déterminées à partir du modèle construit au cycle précédent d'affinement. La deuxième carte correspond quant à elle à la carte de densité résiduelle. Elle permet de mettre en évidence les erreurs du modèle précédent permettant ainsi de les corriger lors de la construction du modèle suivant. Dans cette carte, les pics positifs vont traduire des zones où il manque des électrons et les pics négatifs vont au contraire nous informer d'une zone anormalement excédante en électrons.

On va donc poursuivre cette procédure d'affinement jusqu'à ce que l'on soit satisfait d'un modèle qui deviendra alors la représentation finale de la structure tridimensionnelle cristallisée. Pour nous

permettre d'évaluer les modèles reconstruits et valider le modèle final, il existe des outils et des paramètres que l'on peut observer. Tout d'abord, le facteur R calculé par la relation suivante :

$$R = \frac{\sum_{hkl} \| Fo | - | Fc \|}{\sum_{hkl} \| Fo \|}$$

Il permet de rendre compte de l'accord entre les facteurs de structure observés et les facteurs de structure calculés. Ce facteur va diminuer nécessairement au cours de l'affinement puisque l'on cherche précisément à minimiser les différences entre les données calculées et observées au risque d'ailleurs de suraffiner le modèle. Afin de nous permettre de vérifier qu'on ne se trouve pas dans ce cas là, on regarde également le facteur $R_{free}$. Ce second facteur est calculé à partir de réflexions aléatoires représentant entre 5 à 10 % de l'ensemble. De plus, ces réflexions n'étant pas prises en compte lors de l'affinement, le $R_{free}$ sera un indicateur qualité moins biaisé que le R. C'est finalement l'écart entre ces deux facteurs qui nous permettra de nous rendre compte de la justesse de notre affinement. En effet, plus le rapport entre le nombre d'observations et le nombre de paramètres à affiner est faible, plus cet écart est important. Cela signifie que, dans le cas d'un suraffinement du modèle, le nombre d'observations ne sera plus suffisant et le $R_{free}$ va tendre à augmenter tandis que le facteur R diminuera.

Finalement, les paramètres structuraux du modèle sont vérifiés et comparés aux valeurs caractéristiques des structures protéiques avant toute étude fine. Ainsi, on passe en revue des pramètres tels que la stéréochimie du modèle, les distances interatomiques, les angles des liaisons, les contacts atomiques…
Toutes ces étapes nécessaires à l'obtention de la structure tridimensionnelle d'une protéine sont maintenant assistées par des programmes informatiques. Dans les cas les plus favorables, une fois les clichés de diffraction obtenus, on peut résoudre la structure en seulement quelques heures. Il existe cependant des cas plus complexes et d'autres surprenants où l'on peut passer de nombreuses semaines à essayer de retrouver la carte de densité. Le cas original de la protéase du virus de la mosaïque jaune du navet en est le parfait exemple et vous est présenté dans la première Partie à suivre.

# PARTIE I

# Cristallisation assistée par un contaminant (article 1)

# Résumé

Au cours de la résolution de la structure du domaine TYMVpro, nous nous sommes rendu compte qu'un contaminant était aussi présent dans l'unité asymétrique. Après l'étude de l'empilement cristallin, nous avons pu comprendre comment celui-ci avait assisté la croissance de nos cristaux de protéase et débloquer nos longues et infructueuses tentatives de cristallogenèse. L'histoire originale de cette contamination positive est décrite dans l'article intitulé "In praise of impurity: 30S ribosomal S15 protein - assisted crystallization of Turnip Yellow Mosaic Virus proteinase".

Lors des essais de cristallisation du domaine TYMVpro, nous n'obtenions que des aiguilles ne présentant pas de diffraction. Après de nombreuses tentatives d'optimisation, des petits hexagones sont apparus, après la formation des oursins composés des aiguilles habituelles. Ces hexagones d'environ 50 x 50 x 40 $\mu m^3$ diffractaient jusqu'à 2Å de résolution malgré leur petite taille. Un premier jeu de données natif a été collecté, mais la structure n'a pu être résolue directement par remplacement moléculaire faute d'homologue de structure connue disponible. De plus, TYMVpro ne possède aucune méthionine dans sa séquence nous permettant d'avoir recours à un dérivé sélénométhionine. Nous avons donc réalisé l'incorporation d'atomes lourds dans les cristaux pour résoudre le problème des phases. Trois dérivés ont ainsi été obtenus. Le premier, isomorphe au cristal natif, avec de l'acétate de mercure, présentait un fort signal anomal mais à basse résolution uniquement. Les seconds dérivés NaI et CsCl en revanche avaient un signal plus étendu mais n'étaient pas isomorphes par rapport au cristal natif. Les traitements SAD et SIRAS des dérivés séparés n'ont pas abouti contrairement au MIRAS réalisé avec l'ensemble des jeux de données collectés.

Compte tenu du groupe d'espace et des informations concernant le domaine TYMVpro, nous pensions trouver au minimum 2 molécules dans l'unité asymétrique et plus probablement 3, chacune de ces solutions correspondant respectivement à un pourcentage en solvant de 60 et 40% environ. Etonnament, seulement une molécule a pu être construite et il restait alors beaucoup de densité électronique non assignée.

La qualité de la carte était cependant suffisante pour identifier par endroits les résidus non assignés et la séquence du contaminant S15 de la sous unité 30S du ribosome de *E. coli* a pu être identifiée. Après analyse sur gel plus résolutif et spectrométrie de masse, cette contamination a été confirmée.

L'analyse ultérieure de l'empilement cristallin nous a permis de comprendre comment seule une fraction contaminée a pu donner naissance a ces cristaux et nous permettre d'obtenir la structure de TYMVpro.

L'empilement cristallin est tout d'abord composé d'hélices de protéases où une molécule insère son extrémité C-terminale dans le site catalytique de la molécule suivante et ce, de façon infinie. Cet assemblage donne alors naissance aux aiguilles obtenues tout d'abord lors des essais de cristallisation et ne présentant pas de diffraction.

Ce n'est que lors d'essais avec une fraction plus contaminée que des hexagones sont apparus. L'empilement cristallin nous montre finalement que la protéine S15 a en fait ponté les hélices de TYMVpro permettant alors d'étendre le réseau dans une dimension plus large. Ce phénomène a pu avoir lieu notamment grâce à la forte concentration initiale en protéase. Ainsi, la protéase s'est auto-assemblée en hélice diminuant ainsi la concentration en TYMVpro libre dans la goutte. A un instant t, la protéase et le contaminant se sont alors retrouvés dans des quantités favorables à leur co-cristallisation.

A posteriori, nous avons constaté que la structure de S15 était déjà résolue (à une résolution inférieure à la nôtre). Après des tests de phasage par remplacement moléculaire à partir de celle-ci, nous avons pu reconstruire sans problème la structure de TYMVpro.

# ARTICLE 1 (soumis à Acta crystallographica section F le 28/10/11)

# In praise of impurity: 30S ribosomal S15 protein - assisted crystallization of Turnip Yellow Mosaic Virus proteinase

**Charlotte Robin[a], Lionel Beaurepaire[a], Mélanie Chenon[b], Isabelle Jupin[b] and Stéphane Bressanelli[a]***

[a]Virologie Moléculaire et Structurale, CNRS UPR 3296, INRA USC, Centre de Recherche de Gif, Avenue de la Terrasse, Gif-sur-Yvette, 91190, France

[b]Virologie Moléculaire, CNRS - Univ Paris Diderot, Sorbonne Paris Cité, Institut Jacques Monod, UMR 7592, 15 rue Hélène Brion, 75205 Paris Cedex 13, France

Correspondence email: stephane.bressanelli@vms.cnrs-gif.fr

**Keywords: Tymoviridae; Proteinase; 30S ribosomal S15 protein**

### Synopsis

Diffraction-quality crystals of the Turnip Yellow Mosaic Virus proteinase/ubiquitin hydrolase could only be obtained with a protein preparation heavily contaminated with *E. coli* 30S ribosomal S15 protein (Robin et al, submitted). Crystal packing reveals the basis of this observation.

### Abstract

Turnip Yellow Mosaic Virus is an excellent model for eukaryotic, positive-stranded RNA virus replication. Correct processing of the replication polyprotein is dependent on the virally-encoded cystein proteinase (PRO) domain. Crystalline needles obtained with highly pure preparations of the recombinant 17.6 kDa PRO did not diffract. In contrast, small hexagonal prisms obtained together with the needles in the same conditions but with a poorly purified preparation diffracted to 2 Å resolution and allowed structure determination by MIRAS (Robin et al, submitted). It turns out that the hexagonal crystals contain stoechiometric amounts of PRO and of *E. coli* 30S ribosomal S15, a 10.1 kDa protein commonly co-purified by immobilised metal affinity chromatography. The solvent content is nearly 70 %, with S15 bridging parallel, infinite helices of PRO across large solvent channels. With hindsight, this spurious interaction not only yielded diffraction-quality crystals but also would have allowed structure determination by molecular replacement, using S15 as a search model and subsequent automatic rebuilding of the asymmetric unit.

## 1. **Introduction**

Turnip yellow mosaic virus (TYMV) is a small positive-stranded RNA virus whose replication machinery is encoded in the viral genome as a single polyprotein (206K). From N- to C-terminus, 206K harbours a methyltransferase, a cysteine proteinase (PRO), a helicase (HEL or 42K) and an RNA-dependent RNA polymerase (POL or 66K). As with all known positive-stranded RNA viruses, the replication complex is bound to a cellular membrane, the chloroplast envelope for TYMV. Precise temporal regulation of 66K activity is achieved by a series of steps: First, 66K is cleaved from 206K by PRO. The two products (66K and 140K) remain associated through an interaction that has been mapped to the PRO region and 140K is targeted to the chloroplast, forming the nascent replication complex (Jakubiec et al, 2004). A second cleavage by PRO in 140K occurs at the PRO-HEL junction, triggering the regulation of synthesis of the positive rather than the negative strand (Jakubiec et al, 2007). Recently, we reported that the ubiquitin-proteasome system regulates availability of 66K during infection of plant cells (Camborde et al, 2010) and that PRO is a functional ubiquitin hydrolase *in vitro* and *in vivo* (Chenon et al, in press). Here we describe the crystallization and structure determination of the recombinant PRO, the structure of which is described elsewhere (Robin et al, submitted).

## 2. Results and discussion

### 2.1. Cloning, expression and purification

The coding sequence of the PRO domain (residues 728-879 of 206K) was inserted into a modified pGEX plasmid with an in-frame N-terminal 6His-tag. The resulting protein construct's expected molecular weight is 17.6 kDa. This plasmid was transformed into *Escherichia coli* BL21 rosetta (DE3). An overnight culture was used to inoculate 1-L of LB media containing 50 µg/L carbenicillin and 25 µg/L chloramphenicol. This culture was grown at 37°C to optical density (OD600) 0.6. Expression was induced by the addition of 0.5 mM isopropyl-b-D-thiogalactopyranoside (IPTG) and the cells were grown for 4 hours at 30°C. The cell pellet was harvested, frozen and stored at -20°C.

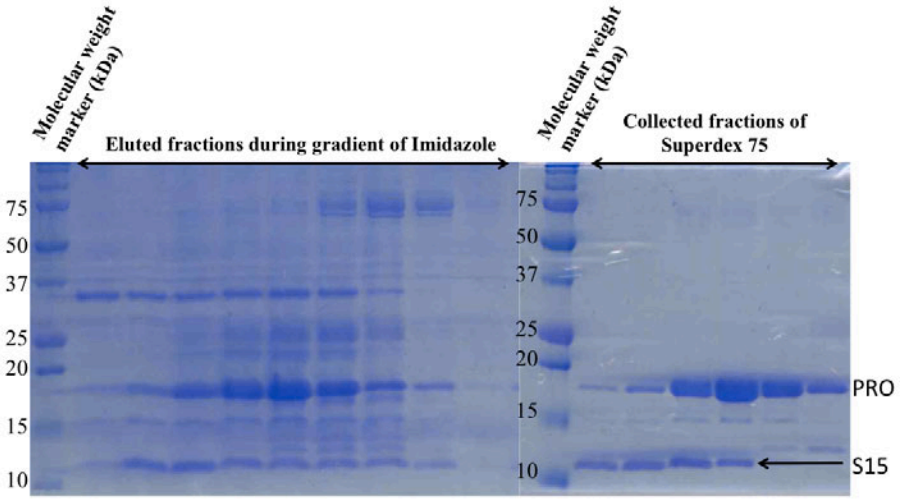
Approximately 5 g of cell pellet was resuspended in 20 mL of lysis buffer (100 mM Tris-HCl pH 7.5, 350 mM NaCl, 25 mM imidazole, 1 mM DTT, 0.5 % Triton X-100, 2 mg/ml Lysozyme, and 1 U/ml Benzonase) and incubated for 60 min on ice. Lysis was completed by 5 freeze/thaw cycles (70 K/303 K).

The disrupted cell lysate was centrifuged at 8000 x*g* during 30 min and the supernatant was loaded onto a 1 mL Ni2+-NTA agarose column (Qiagen) preequilibrated with buffer A (100 mM Tris-HCl pH 7.5, 350 mM NaCl, 25 mM imidazole, 1 mM DTT). The column was washed with 50 mL of buffer A, followed by 10 mL of washing buffer A² (100 mM Tris-HCl, pH 6.0, 350 mM NaCl, 25 mM imidazole, 1 mM DTT).

The protein was eluted by the elution buffer B (100 mM Tris-Cl, pH 7.5, 350 mM NaCl, 500 mM imidazole, 1 mM DTT). The eluted PRO was further purified by high-resolution Superdex S-75 gel filtration column (Amersham) with Buffer C (10 mM Tris-HCl pH 8, 350 mM Ammonium Acetate, 1mM DTT).

All through these steps, a contaminating band of low apparent molecular weight was detected when fractions were analyzed by 15 % Tris-Tricine SDS-PAGE (Fig. 1). We therefore pooled and concentrated only those fractions that were >95 % pure as judged by Coomassie blue staining. The purified protein samples were concentrated to 5-34 mg/ml, frozen and stored at -80°C. Concentration was estimated from OD280 nm assuming absorption of 0.54 $cm^{-1}$ for a 1 mg/ml solution, as calculated from the PRO construct sequence.

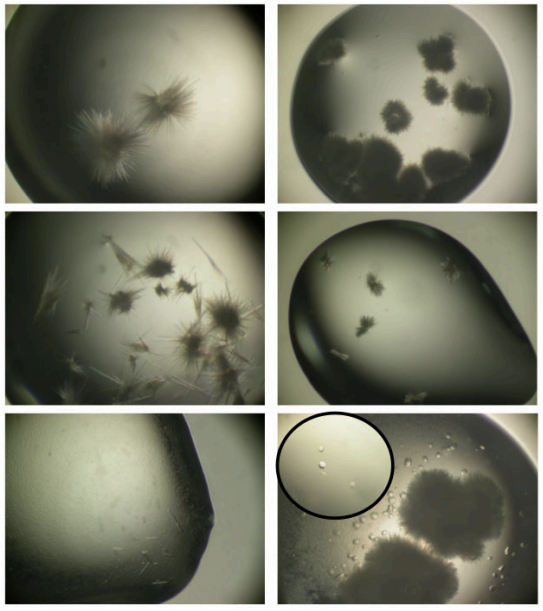
**Fig. 1 :** SDS-PAGE analysis by 15% Tris-Tricine gels of PRO purification**.**



### 2.2. Crystallization

Screening for crystallization conditions was performed by robotics with commercial screens using the sitting drop vapour diffusion method. Equal volumes (100 nl) of protein solution in buffer C were mixed with well solutions. With the most concentrated sample tried (34 mg/ml) showers of needles were obtained within a few days with a well solution of 0.1 M Hepes pH 7.5, 2 M Ammonium formate at 293 K. Fine screening around this condition was performed in larger drop volumes (1 µl protein solution plus 1 µl crystallization reagent equilibrated against a 0.5 ml reservoir volume) using the hanging-drop vapour-diffusion setup with and without microseeding. We easily reproduced the needles (Fig. 2) but failed to obtain diffraction-quality crystals (see below). The exception was for a preparation where all fractions of the gel filtration step had been pooled by mistake before concentrating to 39 mg/ml as judged by OD280 nm. In one single drop (lower right panel of Fig. 2, well solution 0.1 M Hepes pH 7.5, 2.5 M Ammonium formate), hexagonal crystals of up to 50 x 50 x 40 $\mu m^3$ grew after a few days of the needles' appearance.

**Fig. 2** : Crystals of PRO. Needles in the first five panels were obtained with highly pure fractions of PRO with (fifth panel) or without seeding. The sixth (lower right panel) depicts crystals obtained once when all fractions of the S-75 purification step of Fig. 1 were pooled.



### 2.3. Data collection and MIRAS phasing from hexagonal crystals in a single drop

Both the needles and the hexagons were tested for diffraction with synchrotron sources (SOLEIL Proxima 1 and ESRF ID14-1). Prior to testing, crystals were transferred for ~30 s in 0.1 M Hepes pH 7.5, 4 M Ammonium formate, 16 % glycerol and flash frozen by plunging into liquid nitrogen. The needles did not show any diffraction. On the other hand, the hexagons diffracted to close to 2 Å resolution. All data were processed with the XDS package (Kabsch, 2010). Further data analysis was performed with the CCP4 suite (The CCP4 suite: programs for protein crystallography, 1994) except where otherwise indicated.

#### 2.3.1. Native data

We collected a first native dataset to 2.1 Å resolution (Table 1). Space group was assigned as either $P3_121$ or $P3_221$ with no indication of twinning. Given the point group and cell parameters, a PRO molecule occupies some 20 % of the asymmetric unit (asu). We therefore expected at least 2, and given diffraction quality likely 3, PRO molecules per asu. Both the self rotation and the native Patterson functions however were featureless (not shown).

#### 2.3.2. Derivative data

PRO displays no detectable sequence identity to any other protein of known structure. In the absence of any methionine in PRO's sequence and given the lack of reproducibility of the hexagonal crystals, we sought experimental phases by heavy atom soaks. We focused on mercury compounds to take advantage of the single (catalytic) cysteine (best result: 1 mM TCEP, 4 M Ammonium formate and 0,1 M Hepes pH 7.5 for 1 hour, then 10 mM $HgAc_2$, 4 M Ammonium formate and 0,1 M Hepes pH 7.5 for 24 hours, then 4 M Ammonium formate, 0.1 M Hepes pH 7.5 and 12 % glycerol for 1 min.). We also performed short (2 min.) NaI or CsCl cryosoaks (2 M NaI or CsCl, 2 M Ammonium formate, 0.1 M Hepes pH 7.5 and 16 % glycerol).

Derivative data quality was assessed and heavy atom positions sought with SHELXC/D/E (Sheldrick, 2008) through the hkl2map interface (Pape & Schneider, 2004). The $HgAc_2$ derivative was isomorphous with the native and showed strong anomalous signal, but only to low resolution (up to ~5.5 Å).

Anomalous signal extended to beyond 4 Å for both NaI and CsCl cryosoaks, but these displayed nonisomorphism to the native. Furthermore, they suffered from radiation damage that set off quickly at the energy at which we collected the CsCl and NaI data (8 keV). This problem was alleviated somewhat in the case of a 0.8M NaI cryosoak, for which we could get a complete dataset (Table 1).

Of the solutions found by SHELXD independently for the $HgAc_2$, CsCl and NaI derivatives, none was clearly correct as assessed by comparison of SHELXE density modification results for the direct and enantiomorph hands after either SAD or SIRAS phasing.

**Table 1** : Phasing statistics for the first native and the three best derivative datasets

| Data-processing software | XDS / XSCALE | | | |
|---|---|---|---|---|
| Phasing method | MIRAS | | | |
| No. of derivatives | 3 | | | |
| Solution software | Initial heavy atom locations: SHELXC/D (Sheldrick, 2008)<br>Refinement, completion and pruning of heavy atom substructure, phasing: autoSHARP (Vonrhein et al, 2007) | | | |
| Phasing set | HgCl2 | NaI | CsCl | Native |
| Radiation source | SOLEIL - PX1 | SOLEIL - PX1 | SOLEIL - PX1 | ESRF – ID14-1 |
| Wavelength (Å) | 1.008 | 1.55 | 1.55 | 0.934 |
| Temperature (K) | 100 | 100 | 100 | 100 |
| Resolution range (Å) | 44 – 2.60 | 44 – 3.14 | 44 – 3.07 | 44.2 – 2.10 |
| N° of sites (after completion and pruning by autoSHARP) | 4 | 14 | 6 | - |
| Phasing resolution range (Å) | 44.2 – 2.10 | | | |
| Figure of merit (all data) | 0.146 | | | |

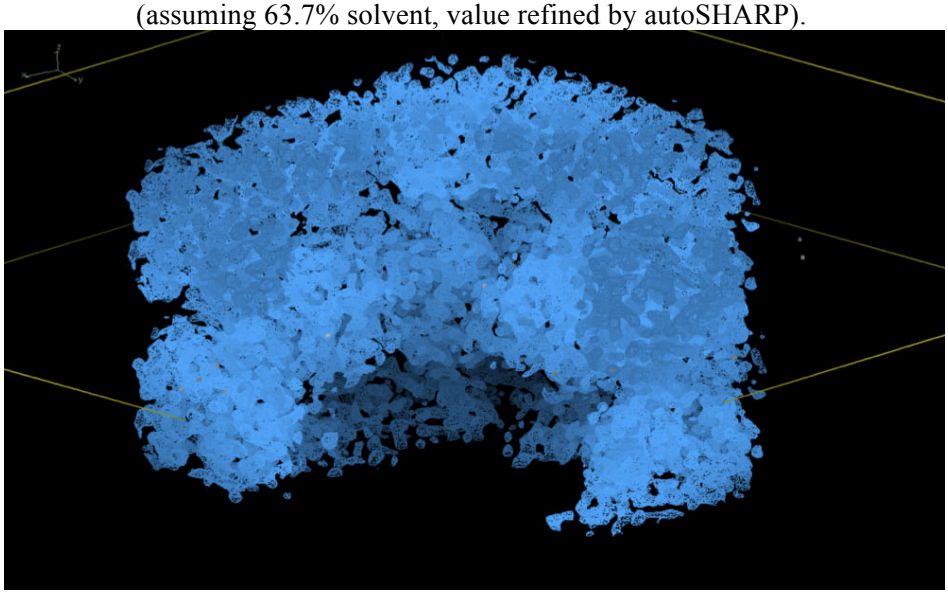
#### 2.3.3. MIRAS phasing and density modification with autoSHARP

The SHELXD $HgAc_2$ best SIRAS solution after 1000 trials with data cut off at 5 Å resolution showed reasonable contrast to other solutions. We used it as a starting heavy atom model for 3-derivative MIRAS phasing by autoSHARP (Vonrhein et al, 2007). We input parameters derived from our prior belief of the asu content to the autoSHARP scripts. In particular, after heavy atom location, refinement and phasing we started solvent flipping by assuming 40% solvent content. However, autoSHARP refined solvent content in subsequent density modification up to 63.7%.

The solvent-flipped map assuming space group $P3_121$ and the original heavy atom hand showed significantly better map statistics than $P3_221$ and the inverse hand.

At this stage it was assumed that there were two molecules per asu and 60% solvent content. The map was interpretable (Fig. 3) and indeed cycling between solvent flipping and automated building with ARP_wARP (Perrakis et al, 1999) unambiguously succeeded (final $R_{free}$: 0.27) in building most of a PRO molecule and several extra helices.

**Fig. 3** : Initial electron density map to 2.1 Å resolution after MIRAS phasing and solvent flipping (assuming 63.7% solvent, value refined by autoSHARP).



### 2.4. Model rebuilding and reassessment of the asu content

We rebuilt the PRO molecule manually with Coot (Emsley et al, 2010) and refined the model with phenix.refine (Adams et al, 2010). Extra helices outside the PRO molecule clearly did not match the PRO sequence. The electron density of the helices was of sufficient quality for us to assign sequence in places. This partial sequence matches that of *E. coli* 30S ribosomal S15 protein (10.1kDa). We could indeed confirm by mass spectrometry that the contaminant detected by 15 % Tris-Tricine SDS-PAGE (Fig. 1) is S15. We completed model building and refinement of the model with one PRO and one S15 in the asu (Table 2).

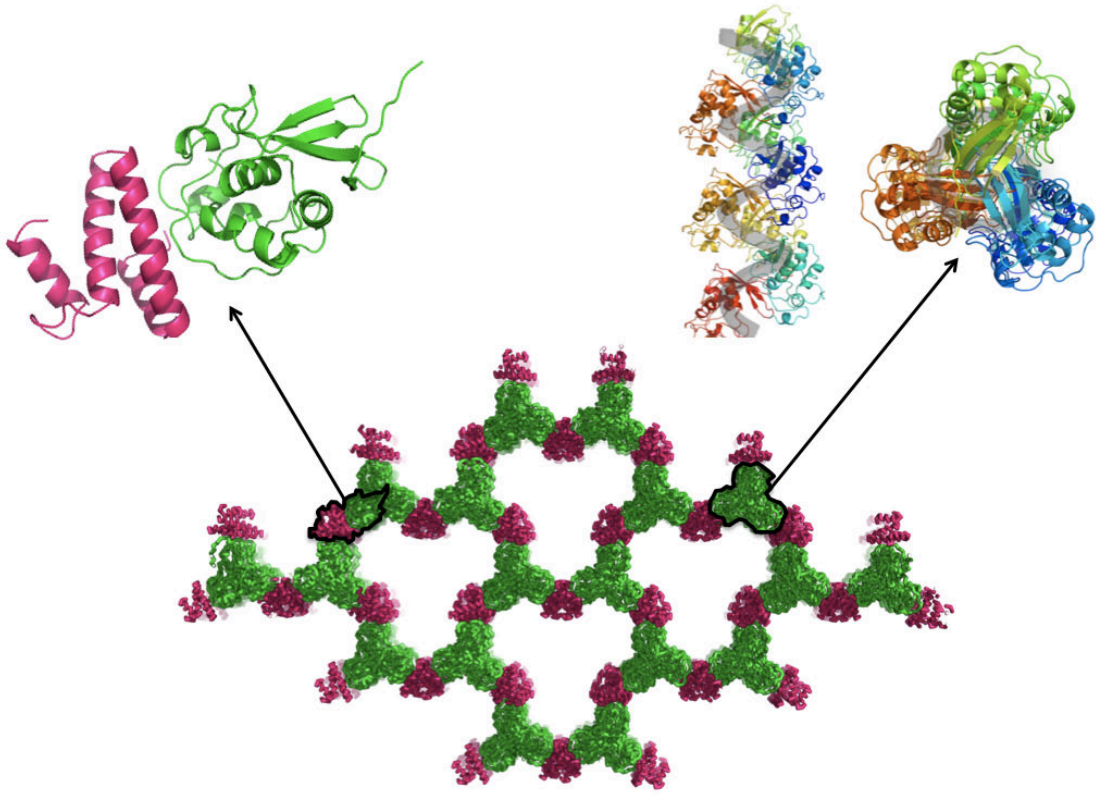
**Table 2** : Values for the outer shell are given in parentheses.

| | |
|---|---|
| Diffraction source | SOLEIL - PX1 |
| Wavelength (Å) | 0.8856 |
| Temperature (K) | 100 |
| Resolution range (Å) | 39.5–2.000 (2.05–2.00) |
| No. of unique reflections | 29610 (2175) |
| Completeness (%) | 99.04 (99.5) |
| Redundancy | 3.72 (3.74) |
| $< I/\sigma(I)>$ | 14.8 (2.2) |
| $R_{merge}$ | 6.4 (61.2) |
| Data-processing software | XDS XSCALE |
| Refinement software | phenix.refine |
| Refinement on | *F* |
| σ cutoff | none |
| Resolution range (Å) | 39.029–2.000 (2.07–2.00) |
| No. of reflections used in refinement | 29610 (2719) |
| Atomic displacement model | TLS + restrained individual *B* factor |
| Overall average *B* factor (Å$^2$) | 33.23 |
| Final $R_{work}$ | 0.1678 (0.2224) |
| No. of reflections for $R_{free}$ | 2275 (220) |
| Final $R_{free}$ | 0.2011 (0.2769) |
| Ramachandran plot analysis | |
| Favoured regions (%) | 99.12 |
| Allowed regions (%) | 0.88 |
| Outliers regions (%) | 0 |

### 3. Conclusion

Crystal packing gives clues as to the sequence of events that led to diffraction-quality crystals (Fig. 4). PRO packs as infinite helices along the crystallographic $3_1$ axis. This is due to one molecule inserting its C-terminus into the catalytic cleft of the next and is actually a relevant interaction in TYMV polyprotein processing (Robin et al, submitted). We therefore surmise that this tends to happen whenever PRO is sufficiently concentrated.

**Fig. 4** : Crystal packing. By itself, PRO packs as infinite helices (upper right panel, each PRO molecule is depicted with a distinct color. It is colored green in the other panels). When PRO is associated with S15 (magenta), the PRO helices are bridged across large solvent channels



Recombinant PRO helices however do not tend to form ordered three-dimensional crystals for want of connecting interactions in the orthogonal plane. Fortuitously, S15 provided such bridging interactions between PRO helices. S15 is a common contaminant in immobilised metal affinity chromatography from *E. coli* cytoplasmic extracts (Bolanos-Garcia & Davies, 2006).

Analysis of the S15/PRO interface with PISA (Krissinel & Henrick, 2007) shows that it is a typical crystal contact burying 740Å$^2$ of solvent-accessible surface area per molecule. The resulting packing however is a robust three-dimensional grid despite a 69% solvent content.

With hindsight, the presence of S15 in the crystals combined with the high solvent content and resolution of the data would have allowed phasing by molecular replacement with Phaser (McCoy et al, 2007). Even falsely assuming several S15 molecules in the asymmetric unit, there is an unambiguous solution (translation Z-score of 23) for a single S15 molecule in the correct spacegroup. Automated model rebuilding and refinement with phenix.autobuild (Adams et al, 2010) subsequently builds a mostly correct model ($R_{free}$= 0.28) comprising 133 out of 148 ordered PRO residues in addition to S15.

Finally, a search of the PDB indicates that the highest resolution so far for a structure of *E. coli* ribosomal S15 subunit was 3 Å, since this protein is usually reported in complexes of much larger size (and more biological relevance). A windfall therefore of the present crystal structure is that it provides, for who may be interested, a higher resolution view of *E. coli* S15.

**Acknowledgements** We acknowledge the Structural biology and Proteomics pole of the IMAGIF integrated platform (https://www.imagif.cnrs.fr/?nlang=en) for access to Crystallization and Mass Spectrometry services. We thank Synchrotrons ESRF (European Synchrotron Radiation Facility) and SOLEIL (beamline PROXIMA 1) for generous allocation of beamtime and gratefully acknowledge Pierre Legrand's help in data collection and processing.

This work was supported in part by the Agence nationale de la recherche (contract ANR-06-BLAN-0062) to IJ and SB. CR was the recipient of a MESR (French ministry for education and research) predoctoral fellowship.

**References**


Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr* **66:** 213-221

Bolanos-Garcia VM & Davies OR (2006) Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochim. Biophys. Acta* **1760:** 1304-1313

Camborde L, Planchais S, Tournier V, Jakubiec A, Drugeon G, Lacassagne E, Pflieger S, Chenon M & Jupin I (2010) The Ubiquitin-Proteasome System Regulates the Accumulation of Turnip yellow mosaic virus RNA-Dependent RNA Polymerase during Viral Infection. *The Plant Cell Online* **22:** 3142 -3152

Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr* **66:** 486-501

Jakubiec A, Drugeon G, Camborde L & Jupin I (2007) Proteolytic processing of turnip yellow mosaic virus replication proteins and functional impact on infectivity. *J. Virol* **81:** 11402-11412

Jakubiec A, Notaise J, Tournier V, Héricourt F, Block MA, Drugeon G, van Aelst L & Jupin I (2004) Assembly of turnip yellow mosaic virus replication complexes: interaction between the proteinase and polymerase domains of the replication proteins. *J. Virol* **78:** 7945-7957

Kabsch W (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr* **66:** 125-132

Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol* **372:** 774-797

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40:** 658-674

Pape T & Schneider TR (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *Journal of Applied Crystallography* **37:** 843-844

Perrakis A, Morris R & Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol* **6:** 458-463

Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr., A, Found. Crystallogr* **64:** 112-122

The CCP4 suite: programs for protein crystallography (1994) *Acta Crystallogr. D Biol. Crystallogr* **50:** 760-763

Vonrhein C, Blanc E, Roversi P & Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol* **364:** 215-230

# Complément aux résultats de l'article 1 : phasage alternatif

Dans cet article, la cristallogenèse et la résolution de la structure vous ont été exposées. Nous avons constaté ensuite, que nous aurions pu procéder différemment pour résoudre le problème des phases. En effet, si nous avions caractérisé la contamination qui nous accompagnait au cours de nos étapes de purification et de cristallisation, nous aurions pu accéder à la structure plus aisément. Cette alternative vous est ici présentée.

### Purification partielle de TYMVpro

Lorsque les cristaux sont apparus, nous avons eu des difficultés sur un projet annexe au niveau de la résolution des gels de polyacrylamide. Les gels que nous utilisions pour des protéines de petites tailles, nous donnaient en effet des informations de mauvaises qualités. Les bandes correspondantes à des protéines inférieures à 10 kDa environ étaient floues voire quasi invisibles. Nous sommes alors passé à des gels de type tris-tricine afin de lever ce problème et, content de leur qualité résolutive, nous les avons utilisé alors pour toutes les protéines étudiées de petite taille. Ainsi les expériences concernant la protéase de TYMV ont été analysées sur ces gels à compter de ce moment. La première purification complète du domaine TYMVpro étudiée sur ces nouveaux gels nous a révélé que nous n'étions pas aussi purs que nous le pensions et la purification de TYMVpro semblait inachevée. Un contaminant d'environ 10 kDa était retrouvé dans plusieurs fractions chevauchantes à celles contenant la protéase de TYMV en sortie de gel filtration. Nous avons donc analysé sur ces gels la sortie de la gel filtration ayant servi aux essais de cristallisation afin de voir s'il en était de même à ce moment là (Figure 1 article 1). A notre grande surprise, la contamination semblait beaucoup plus importante et nous nous sommes aperçu que lors de la réunification des fractions de purification, nous avions également pris par erreur les fractions contaminées aux alentours de celles contenant la protéase (toutes les fractions présentes sur le gel Figure 1).

Cependant, nous avions déjà concentré nos efforts sur l'incorporation d'atomes lourds pour résoudre les phases par MIRAS. Enthousiasmés par la structure se faisant de plus en plus proche nous avons donc cherché à utiliser les données, et ce n'est que lors des difficultés à retrouver les phases que nous nous sommes intéressé de plus près à cette contamination. Et si nous avions cristallisé ce contaminant pour le moment inconnu et non la protéase ?

La protéase est certe majoritaire mais, l'on sait qu'il est possible de cristalliser un contaminant présent même à moins de 5 % dans la solution finale de cristallisation (Veesler 2008). Pour s'assurer de cela, en parallèle de nos tentatives de phasage, nous avons pêché plusieurs cristaux, que nous avons lavé dans la solution de cristallisation et resuspendu dans du bleu de dépôt pour gel tris-tricine. Nous avons ensuite fait migrer cette solution sur un gel 16,5 % et avons constaté que la solution contenait bien la protéase mais aussi une bande de plus faible poids moléculaire aux alentours de 10 kDa (Figure 17). La bande correspondante a été donnée à la plateforme de spectrométrie de masse afin de l'identifier. Elle pouvait soit correspondre à un peptide de dégration de la protéase, soit au contaminant vu lors de la purification. Seulement, une panne technique de la plateforme a retardé de plusieurs semaines ce résultat et deux jours avant son obtention, nous avons finalement résolu les phases par MIRAS nous confirmant la présence de S15 dans l'empilement cristallin comme l'explique l'article 1.

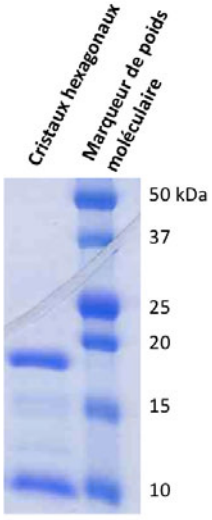



**Figure 17 : Analyse de la composition des cristaux hexagonaux sur gel Tris –tricine**
Après lavage dans la solution de cristallisation, les cristaux ont été déposés sur gel tris-tricine et le résultat de la migration montre la présence de deux espèces protéiques.

Malgré tout, dans un but purement informatif, nous avons cherché à savoir si nous aurions pu résoudre le problème des phases en partant de la structure résolue de S15, par remplacement moléculaire.

### Remplacement moléculaire à partir de la structure résolue du contaminant

Après analyse sur gel des cristaux, nous savions qu'il y avait au moins une molécule de S15 et une molécule de TYMVpro associées dans l'empilement cristallin. A partir des données du cristal natif, l'on sait également que le groupe d'espace est $P3_121$ ou $P3_221$. Compte tenu de la masse moléculaire du domaine TYMVpro (18 kDa) et de celle du contaminant S15 (10 kDa), nous avons décidé d'utiliser le module de remplacement moléculaire automatique de Phaser 2.3.0 inclus dans la suite phenix avec deux molécules de S15 ainsi que 2 molécules de TYMVpro possibles dans l'unité asymétrique.

Nous avons utilisé la chaîne O de la structure portant le code pdb 2QAL afin que le modèle de S15 utilisé corresponde à la sous unité ribosomale présente chez *E. coli*. Avec S15 comme seul modèle et avec la séquence de TYMVpro donnée comme information supplémentaire, nous avons utilisé phaser en gardant tous les paramètres par défaut. Le programme a pu sans ambiguïté retrouver le groupe d'espace $P3_121$ et assigner le modèle à 31% de la densité électronique de l'unité asymétrique. A l'issue de l'assignement de phaser, les 78 résidus de S15 suffisamment définis étaient correctement placés dans la densité électronique calculée.

A partir de la carte calculée et de la structure partiellement reconstruite de S15, nous avons alors utilisé la fonction autobuild de la suite phenix afin de voir si le domaine TYMVpro aurait pu lui aussi être reconstruit aisément et correctement. En utilisant là encore les paramètres par défaut, la structure primaire de TYMVpro et les fichiers de sortie de phaser, une solution unique de sa structure tridimensionnelle a été trouvée. Après 4 cycles de reconstruction, l'accord entre la structure reconstruite et les données de diffraction nous donnaient un R égal à 0.27 et un Rfree égal à 0.29, ne laissant donc aucun doute sur l'exactitude de cette structure.

## Conclusion

La contamination, redoutée généralement par les biocristallographes, montre ici qu'elle peut, dans notre cas, aider la cristallisation et également la résolution de la structure. Bien sûr, nous reconnaissons que ce phénomène de très faible probabilité n'est pas un cas généralisable et il n'est donc pas possible, avec cette seule étude, d'en faire une méthode alternative pour l'obtention de cristaux en cas d'optimisation problématique.

Au contraire, dans la plupart des cas de contamination relatés, on constate qu'elle a été plutôt handicapante pour la pousuite du projet initial. Le contaminant décrit dans ces études a en effet cristallisé seul, au détriment de la protéine d'intérêt (Kefala 2010, Cámara-Artigas 2006). Les cas sont nombreux et tous ne sont bien évidemment pas vantés dans une publication. Parfois, un contaminant ayant le même poids moléculaire que la protéine étudiée, peut nous tromper en amont de la cristallisation. On peut aussi être faussé par les méthodes de caractérisation non satisfaisantes comme pour nos premiers gels d'identification des fractions d'élution. La résolution analytique ici, n'était alors pas suffisante. On peut également tomber sur un contaminant plus « subtil » comme celui référencé par Veesler en 2008. Ce contaminant non détectable par les méthodes classiques d'identification et présent à moins de 5% dans la solution finale a pourtant cristallisé seul. Dans le cas où l'on ne possède pas d'homologue nous permettant de nous rendre compte rapidement de cette erreur, on peut alors perdre temps et énergie à résoudre cette structure non désirée.

Il existe désormais un cas plus heureux de contamination puisque sans celle-ci, nous serions peut-être encore en phase d'optimisation cristalline. Cela m'a d'ailleurs appris à quel point la cristallographie est une science surprenante voir amusante. Il n'y a pas de protocole pré-établi et l'on ne peut parfois comprendre les propriétés bienfaisantes de celui appliqué qu'une fois la structure finale résolue.

# PARTIE II

# Etude structurale de la protéase du TYMV (Article 2)

# Résumé

Afin d'étudier la régulation du cycle de multiplication du virus modèle TYMV au niveau structurale, nous avons produit en *E. coli*, purifié et cristallisé son domaine protéase (TYMVpro). Ce domaine est en effet un acteur clé dans la régulation de la réplication. Tout d'abord, il est responsable de la maturation des enzymes virales impliquées dans le complexe de réplication grâce à son activité protéase à cystéine. Puis, il permet le recrutement à la membrane de l'ARN polymérase ARN-dépendante et assure sa stabilisation grâce à son activité déubiquitinase (DUB). Pour pouvoir comparer les différences structurales qui peuvent exister lors de la réalisation de ses différentes fonctions, nous avons cherché à résoudre la structure du domaine protéase faisant 150 résidus, soit environ 18 kDa. Pour cela, des cristaux hexagonaux de petite taille mais diffractant à 2 Å de résolution ont été obtenus et la structure a été résolu par cristallographie aux rayons X en utilisant la méthode MIRAS pour résoudre le problème des phases.

L'étude de cette structure nous a permis d'avoir une vue de l'activité protéase de TYMVpro. En effet, l'empilement cristallin est tel qu'une molécule de protéase insère son C-terminus dans la crevasse catalytique de la suivante. Cette configuration correspond alors à la maturation protéolytique du dernier produit de clivage de la 206K au cours de laquelle la protéase réalise une coupure en *trans* au niveau C-terminal d'une autre protéase. Dans ce cas de figure, la protéase est à la fois son propre substrat et l'enzyme.

Nous avons pu mettre en évidence les déterminants structuraux impliqués dans la fonction protéase et notamment, expliquer les résultats de précédentes études concernant les sites P et P' du substrat. La terminologie P et P' a été utilisée pour la première fois par Schechter et Berger pour décrire la spécificité de la papaïne qu'ils étudiaient alors (Schechter 1967) (Figure 18). Dans les études concernant la protéase du TYMV, il avait été déterminé que seuls les sites situés en amont de la coupure, soit les sites P, avaient une réelle importance pour la reconnaissance et l'activité de la protéase. Le fait que le site actif soit directement exposé au solvant justifie ces observations. En effet, seul la partie du substrat située en amont du site de clivage établit des interactions fortes avec la protéase enzyme. A l'inverse, les résidus P' situés en aval sont directement relargués vers l'extérieur et n'établiront probablement que peu voire aucun contact avec l'enzyme.

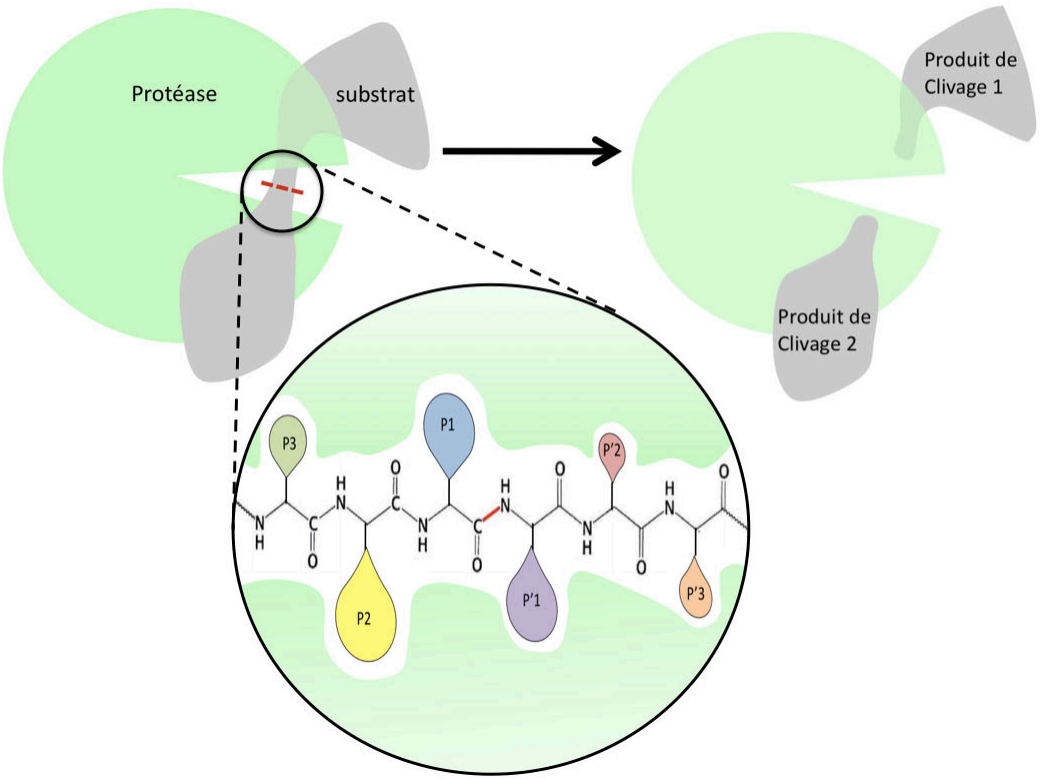



**Figure 18 : Nomenclature des sites P**
Les résidus du substrat reconnus par la protéase au niveau du domaine catalytique sont appelés P. La reconnaissance de ces sites est fonction de la spécificité de la protéase. En amont du site de coupure représenté en rouge, les sites sont notés P et, en aval, ils sont notés P'. Ainsi, la coupure par la protéase (verte) du substrat (gris), va aboutir à deux produits de clivages. Le premier terminant par P1 et le second commençant par P'1.

L'examen plus poussé des sites P nous a permis finalement d'établir un profil de séquence impliquée dans la reconnaissance de la protéase pour son substrat. Compte tenu de l'étroitesse du canal menant à la diade catalytique, il est indispensable que les sites P1 et P2 soient composés de petits acides aminés. Les sites P3 et P4 quant à eux, sont situés au niveau d'interfaces hydrophobes. Finalement on a pu étendre l'implication des sites P jusqu'au 5ème résidu qui, comme le montre la structure, établit de nombreux ponts salins et doit donc être positivement chargé.

L'analyse du site catalytique nous a montré qu'il est formé grâce à l'arrangement de deux lobes. Cette architecture nous a permis de remonter à des homologues structuraux de type OTU et plus précisément à des ubiquitines hydrolases cellulaires et virales. En revanche, aucune endopeptidase ne semble avoir un repliement de ce type. Cette découverte reste en accord avec le fait que TYMVpro possède également une activité DUB et que les OTU protéines homologues trouvées sont toutes des protéases à cystéine appartenant au clan CA. Cette homologie a par la même occasion résolue l'ambiguité d'assignation à un clan qui existait concernant la protéase de TYMV en la rangeant avec ses homologues dans le clan CA.

La structure de TYMVpro nous renseigne enfin sur un lobe N-terminal, spécifique des tymovirus et directement impliqué dans la reconnaissance par la protéase de son substrat. Deux cis-prolines en amont de l'histidine catalytique participent doublement à cette reconnaissance. En effet, elles permettent la bonne orientation du site catalytique dans la molécule enzyme et forment dans la molécule substrat une surface de contact importante avec le lobe N-terminal de l'enzyme. Par comparaison avec les homologues structuraux de type OTU, résolus seuls ou en complexe avec une molécule d'ubiquitine, on a pu émettre l'hypothèse que ce lobe N-terminal est sans doute impliqué également dans la reconnaissance de l'ubiquitine lorsque TYMVpro agit en tant que déubiquitinase.

# Article 2 (en cours de soumission)

# Crystal structure of a viral multifunctional proteinase/ubiquitin hydrolase in polyprotein processing mode

Charlotte Robin[1*], Lionel Beaurepaire[1*], Mélanie Chenon[2], Isabelle Jupin[2] and Stéphane Bressanelli[1§]

From the Laboratoire de Virologie Moléculaire et Structurale, Centre de Recherche de Gif, CNRS (UPR 3296), Avenue de la Terrasse, 91198 Gif sur Yvette Cedex, France[1], Laboratoire de Virologie Moléculaire, CNRS - Univ Paris Diderot, Sorbonne Paris Cité, Institut Jacques Monod, UMR 7592, 15 rue Hélène Brion, 75205 Paris Cedex 13, France[2]
*These authors contributed equally to this work
§To whom correspondence should be addressed: Stéphane Bressanelli, Virologie Moléculaire et Structurale, CNRS UPR3296, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette cedex, France, Tel +33 (0)169823852 Fax +33 (0)169824308 email stephane.bressanelli@vms.cnrs-gif.fr

Running title: Structure of a viral PRO/DUB.

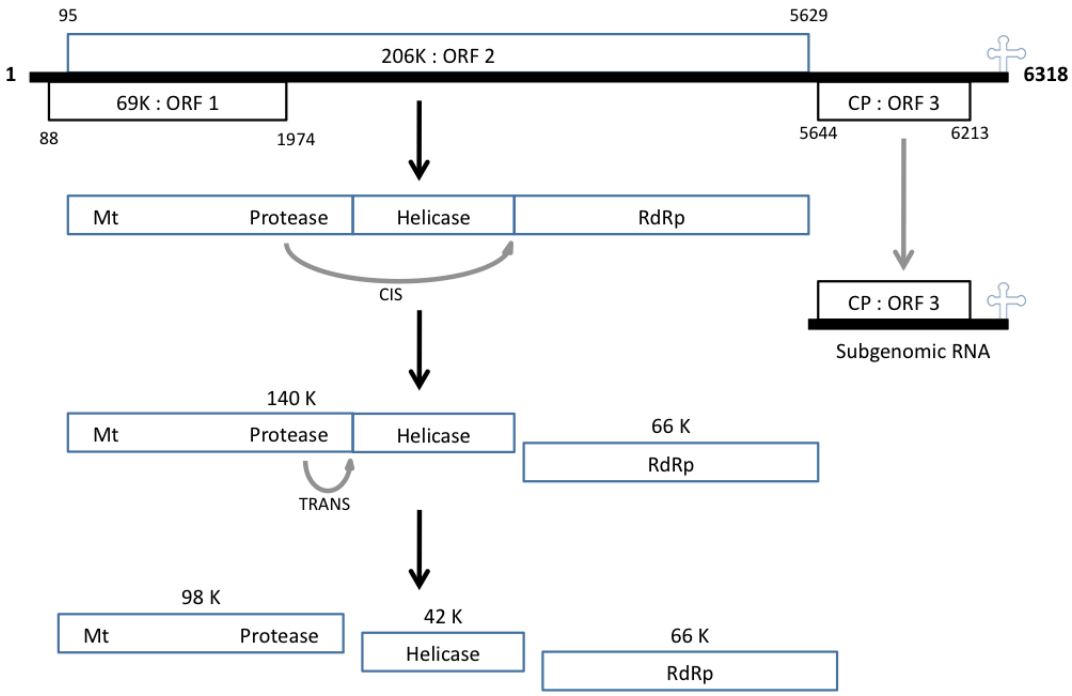
## Abstract

**Turnip yellow mosaic virus (TYMV) serves as a model system for positive-stranded RNA virus replication. Key events in TYMV replication complex formation and regulation are two proteolytic cuts performed by a cysteine proteinase domain (PRO) encoded within the replication polyprotein. We report here the crystal structure of the 150-residue PRO. In the crystal, one molecule's C-terminus inserts into the catalytic cleft of the next, providing a view of the N-terminal product complex in replication polyprotein processing. This allows us to locate the specificity determinants of PRO for its proteinase substrates. An N-terminal lobe that is unique to tymoviruses provides crucial interaction surfaces in molecular recognition. This lobe is missing from the closest structural homologs of PRO, ubiquitin hydrolases (DUBs) from the Ovarian tumor (OTU) family. We recently reported that PRO is a functional DUB *in vitro* and *in vivo* (Chenon et al, in press). OTU DUBs of known structures (the DUB domains from Crimean Congo hemorrhagic fever virus and from yeast ovarian tumour-domain containing protein 1) were reported as complexes with ubiquitin. Superposition of PRO with these complexes suggests that the N-terminal lobe is also involved in recognition of ubiquitin. Thus, we propose that the compact PRO switches between its diverse substrates thanks to a dedicated specificity lobe apposed to a minimal cysteine proteinase domain.**

## Introduction

Plus-strand RNA (RNA+) viruses are the largest class of eukaryotic viruses. They include significant pathogens of humans, animals and plants (1). From the sequencing of their genomes, it has become clear that despite a huge diversity, these viruses possess high similarities at the molecular level (2). Indeed, common strategies and regulatory mechanisms have been uncovered in the replication of RNA+ viruses (3). Thus, all RNA+ viruses studied to date subvert the host cell to synthesize new viral genomes at an intracellular membrane. There, synthesis of the viral progeny requires the establishment of specific and regulated interactions between viral proteins and different cellular factors, assembled within a replication complex. There are two further very common, if not universal, features of RNA+ virus replication. The first one is that the viral components of the replication complex are synthesized as part of a polyprotein that is subsequently processed by viral (and possibly also cellular) proteinases. The second one is that the highly compact viral genome codes for usually multifunctional proteins. This latter feature is connected to the former, as intermediate maturation products of the polyprotein may have different (or differently tuned) functions than the final products. Turnip yellow mosaic virus (TYMV) is a model RNA+ virus whose replication is particularly well characterized at the molecular and cellular levels (4)(5)(6). Its 6.3kb genome codes for only three proteins, the largest of which (206K) is a polyprotein of 206kDa that contains all the viral components of the replication machinery. From N- to C-terminus, 206K harbors a methyltransferase, a cysteine proteinase (PRO), a helicase (HEL or 42K) and an RNA-dependent RNA polymerase (POL or 66K). Precise temporal regulation of 66K activity is achieved by a series of steps: First, 66K is cleaved from 206K by PRO. The two products (66K and 140K) remain associated through an interaction that has been mapped to the PRO region and 140K is targeted to the chloroplast, forming the nascent replication complex (6). Later in the viral cycle, a second cleavage by PRO in 140K occurs at the PRO-HEL junction, triggering the regulation of synthesis of the positive rather than the negative strand (7) (Figure 1). Recently, we reported that the ubiquitin-proteasome system regulates availability of 66K during infection of plant cells (8) and that PRO is a functional ubiquitin hydrolase in vitro and in vivo (Chenon et al, in press).

**Figure 1**: Genomic organization of TYMV RNA and proteolytic processing by PRO of the 206K polyprotein.




Here we describe the crystal structure of the recombinant PRO. This structure sheds light on both of the two documented functions of PRO: First, one PRO's C-terminus inserts into the catalytic cleft of the next in the crystal. This provides us with a detailed view of the N-terminal product complex in replication polyprotein processing. Furthermore, this structure establishes PRO as an ubiquitin hydrolase of the Ovarian tumor (OTU) family. Comparison with the available complexes of OTU DUBs with ubiquitin provides a rationale for the interaction of PRO with ubiquitin. In both processing proteinase and DUB functions, a key element in molecular recognition by PRO seems to be a tymovirus-specific lobe N-terminal to the catalytic domain.

**Experimental procedures**

**Protein expression, purification, and crystallization.**

The production, purification and crystallization of an N-terminally 6-histidine tagged PRO domain are described in details elsewhere (Robin et al, submitted). Briefly, the coding sequence of the PRO domain (residues 728-879 of 206K) was inserted into a modified pGEX plasmid with an in-frame N-terminal 6His-tag. This plasmid was transformed into *Escherichia coli* BL21 rosetta (DE3). An overnight culture was used to inoculate 1-L of LB media containing 50 μg/L carbenicillin and 25 μg/L chloramphenicol. This culture was grown at 37°C to optical density (OD600) 0.6. Expression was induced by the addition of 0.5 mM isopropyl-b-D-thiogalactopyranoside (IPTG) and the cells were grown for 4 hours at 30°C. The cell pellet was harvested, frozen and stored at -20°C.

The pellet was resuspended in 20 mL of lysis buffer (100 mM Tris-HCl pH 7.5, 350 mM NaCl, 25 mM imidazole, 1 mM DTT, 0.5 % Triton X-100, 2 mg/ml Lysozyme, and 1 U/ml Benzonase) and incubated for 60 min on ice. Lysis was completed by 5 freeze/thaw cycles (70 K/303 K).

The disrupted cell lysate was centrifuged at 8000 xg during 30 min and the supernatant was loaded onto a 1 mL Ni2+-NTA agarose column (Qiagen) preequilibrated with buffer A (100 mM Tris-HCl pH 7.5, 350 mM NaCl, 25 mM imidazole, 1 mM DTT). The column was washed with 50 mL of buffer A, followed by 10 mL of washing buffer $A^2$ (100 mM Tris-HCl, pH 6.0, 350 mM NaCl, 25 mM imidazole, 1 mM DTT). The protein was eluted by the elution buffer B (100 mM Tris-Cl, pH 7.5, 350 mM NaCl, 500 mM imidazole, 1 mM DTT).

The eluted PRO was further purified by high-resolution Superdex S-75 gel filtration column (Amersham) with Buffer C (10 mM Tris-HCl pH 8, 350 mM Ammonium Acetate, 1mM DTT).

A pool from all fractions of the gel filtration step in buffer C was concentrated to 39 mg/ml as judged by OD280 nm. Showers of needles and hexagonal crystals of up to 50 x 50 x 40 $\mu m^3$ grew in a single vapor diffusion drop where 1 μl protein solution plus 1 μl well solution (0.1 M Hepes pH 7.5, 2.5 M Ammonium formate) was equilibrated against a 0.5 ml reservoir volume. Prior to testing, crystals were transferred for ~ 30 s in 0.1 M Hepes pH 7.5, 4 M Ammonium formate, 16 % glycerol and flash frozen by plunging into liquid nitrogen. The hexagons diffracted to close to 2 Å resolution on synchrotron beamlines.

**MIRAS phasing, model building and refinement.**

Details of the structure determination are given elsewhere (Robin et al, submitted). Briefly, the structure was solved by MIRAS from three poor derivatives thanks to the high (69%) solvent content

of the crystals. Heavy atom derivatives ($HgAc_2$, NaI and CsCl) were obtained by soaking. Data were processed with the XDS package (9). Initial heavy atom sites were located with SHELXD (10). This first heavy atom model was refined, completed and pruned and initial phases were computed and improved with autoSHARP (11). The resulting map was interpretable and a first model was built with phenix.autobuild (12). The model was manually rebuilt with COOT (13) and refined with phenix.refine (12). Data processing and refinement statistics are collated in Table 1.

**Structure analysis.**

Interfaces in the crystal were assessed using the PISA server (14) (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Homologs of PRO were sought and superimposed with the DALI server (15) (http://ekhidna.biocenter.helsinki.fi/dali_server). Structures were displayed and figures 2, 4 and 5 were prepared with Pymol (www.pymol.org). Figure 3 was generated with *ALINE* (16).

**Table1 : Data processing statistics and final structure refinement and model validation**
Values for the outer shell are given in parentheses.

| | |
|---|---|
| Diffraction source | SOLEIL - PX1 |
| Wavelength (Å) | 0.8856 |
| Temperature (K) | 100 |
| Resolution range (Å) | 39.5–2.000 (2.05–2.00) |
| No. of unique reflections | 29610 (2175) |
| Completeness (%) | 99.04 (99.5) |
| Redundancy | 3.72 (3.74) |
| $< I/\sigma(I) >$ | 14.8 (2.2) |
| $R_{merge}$ | 6.4 (61.2) |
| Data-processing software | XDS XSCALE |
| Refinement software | phenix.refine |
| Refinement on | *F* |
| σ cutoff | none |
| Resolution range (Å) | 39.029–2.000 (2.07–2.00) |
| No. of reflections used in refinement | 29610 (2719) |
| Atomic displacement model | TLS + restrained individual *B* factor |
| Overall average *B* factor (Å$^2$) | 33.23 |
| Final $R_{work}$ | 0.1678 (0.2224) |
| No. of reflections for $R_{free}$ | 2275 (220) |
| Final $R_{free}$ | 0.2011 (0.2769) |
| Ramachandran plot analysis | |
| Favoured regions (%) | 99.12 |
| Allowed regions (%) | 0.88 |
| Outliers regions (%) | 0 |

## Results

### Overall protein architecture.

We report here the structure of the TYMV polyprotein processing proteinase domain to a resolution of 2 Å with a final $R_{free}$ of 20.1% (Table 1). All data including 3 derivative datasets obtained by heavy atom soaks were from crystals grown in a single crystallization drop. The asymmetric unit contains a single PRO molecule packing as infinite helices along the crystallographic $3_1$ axis and an *E. coli* contaminant (ribosomal protein S15). S15 bridges the PRO helices, explaining why diffraction-quality crystals only grew from a PRO preparation heavily contaminated by S15 (Robin et al, submitted). PRO displays a three-lobed architecture (Fig. 2A). The N-terminal lobe comprises two short helices flanking a two-stranded, distorted b-sheet. The catalytic domain is made up by the central and C-terminal lobes (a bundle of five helices and a four-stranded b-sheet, respectively). The catalytic dyad Cys783-His869 (TYMV polyprotein numbering) is completely solvent exposed and lies at the interface between helix a3 (the first helix in the central lobe) and strand b6 (the last strand of the C-terminal lobe). Indeed, Cys783 is the first residue of helix a3 and His869 the first residue of strand b6. Remarkably, the five-residue loop between strands b5 and b6 contains two successive cis-prolines 865-Gly-Pro-Pro-867 (Fig. 2B). Downstream of strand b6, the main chain makes a sharp turn so that the C-terminal residues of PRO 874-Lys-Arg-Leu-Leu-Gly-Ser-879 point away from the a3–b6 interface. Similarly, the long loop (residues 771-782) connecting the N-terminal lobe to the central lobe comes to helix a3 from the other side of the a3–b6 interface. This feature is unusual in the closest relatives of known structure of PRO (see below) and contributes to the exposure of the catalytic dyad at the exit of the catalytic site. The long b2-a3 loop (residues 771-782) connecting the N-terminal to the central lobe threads through a cleft between helices a4 and a5 at the back of the central lobe. This results in the N-terminal lobe being apposed to the catalytic domain but on the other side from the catalytic dyad. Important residues in this positioning of loop b2-a3 are Arg769, that participates in an extended network of interactions, including a salt bridge to Asp809 and a hydrogen bond to Pro808 at the base of a5; Pro777, that positions the main chain to make two hydrogen bonds to the indole ring nitrogen of Trp800 in a4; and Pro779, that inserts into a hydrophobic pocket lined by Trp800, Leu785 and Leu822. The pattern of conservation among *Tymoviridae* proteinases (Fig. 3) indicates that this arrangement, and consequently the position of the N-terminal lobe, are very likely conserved in the family.

**Figure 2: Structure of PRO.**
*A*, Overall structure in ribbon representation. The N-terminal lobe is colored in blue, the central lobe in green and the C-terminal lobe in red. Secondary structure elements are labeled : α for helices and β for β-sheets with numbers from N to C-terminal for each kind of element.
*B*, Final 2Fo-Fc electron density map for the two cis-prolines 865-Gly-Pro-Pro-867 directly upstream strand b6. Also visible and labeled is the catalytic dyad Cys783-His869.

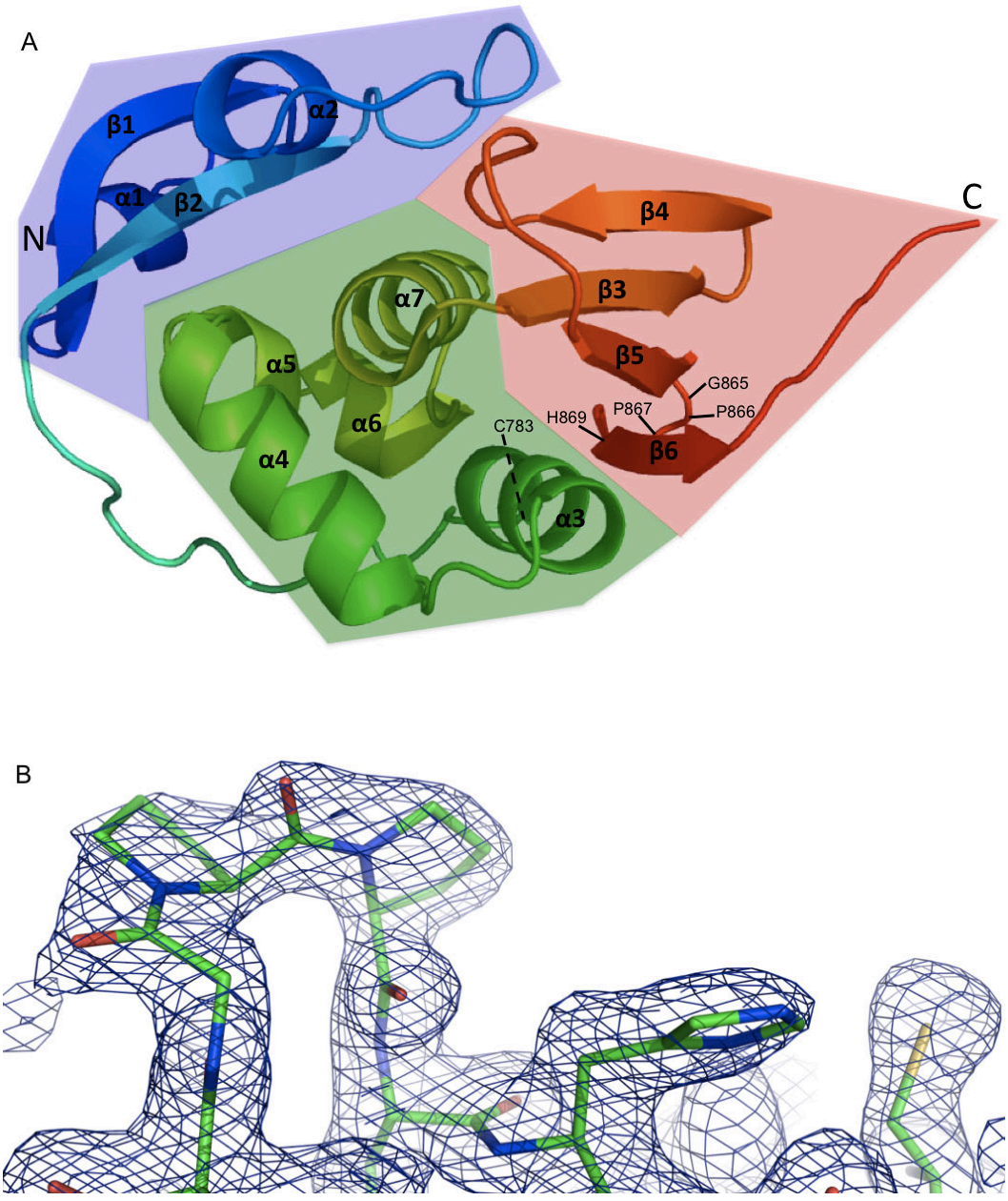
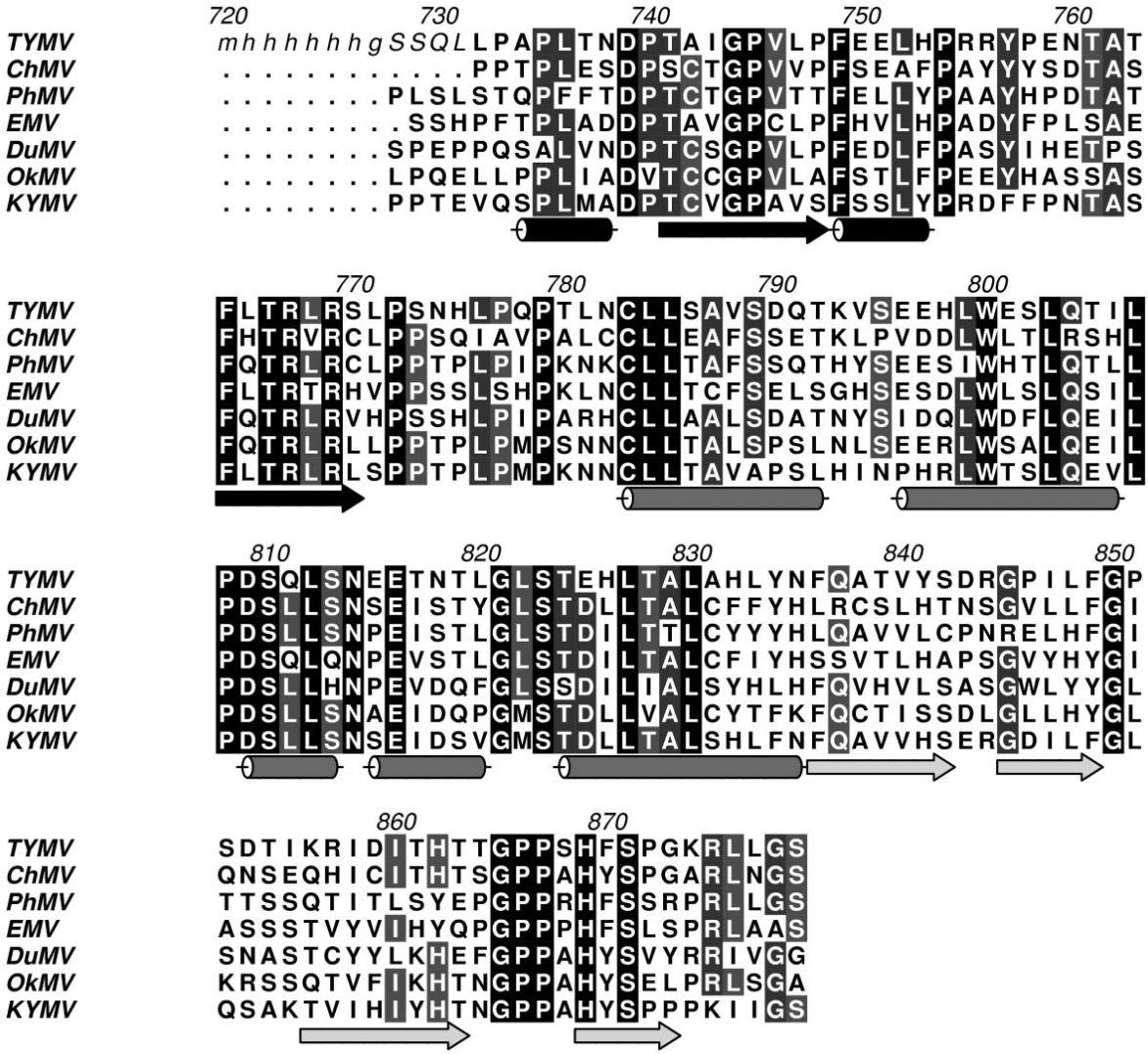


**Figure 3: Amino acid sequence alignment of polyprotein processing endopeptidases belonging to the tymovirus genus and with at least 65% sequence identity with PRO.**
PRO sequences are displayed for TYMV (Turnip yellow mosaic virus), ChMV (Chayote mosaic virus), PHyMV (Physalis mottle virus), EMV (Eggplant mosaic virus), DuMV (Dulcamara mottle virus), OkMV (Okra mosaic virus) and KYMV (Kennedya yellow mosaic virus). Alignment was obtained from the MEROPS database (http://merops.sanger.ac.uk) and edited and displayed using the Aline programme (Bond, C.S. and Schüttelkopf, A.W. (2009), Acta cryst. D65, 510-512). The numbering of TYMV polyprotein is used here and the N-terminal part indicated by italics is the part not visible in the electron density. The secondary structure of PRO is reported below the alignment. Structural elements belonging to the N-terminal lobe appear in black, those in the central lobe appear in gray and finally those in the C-terminal lobe in light gray

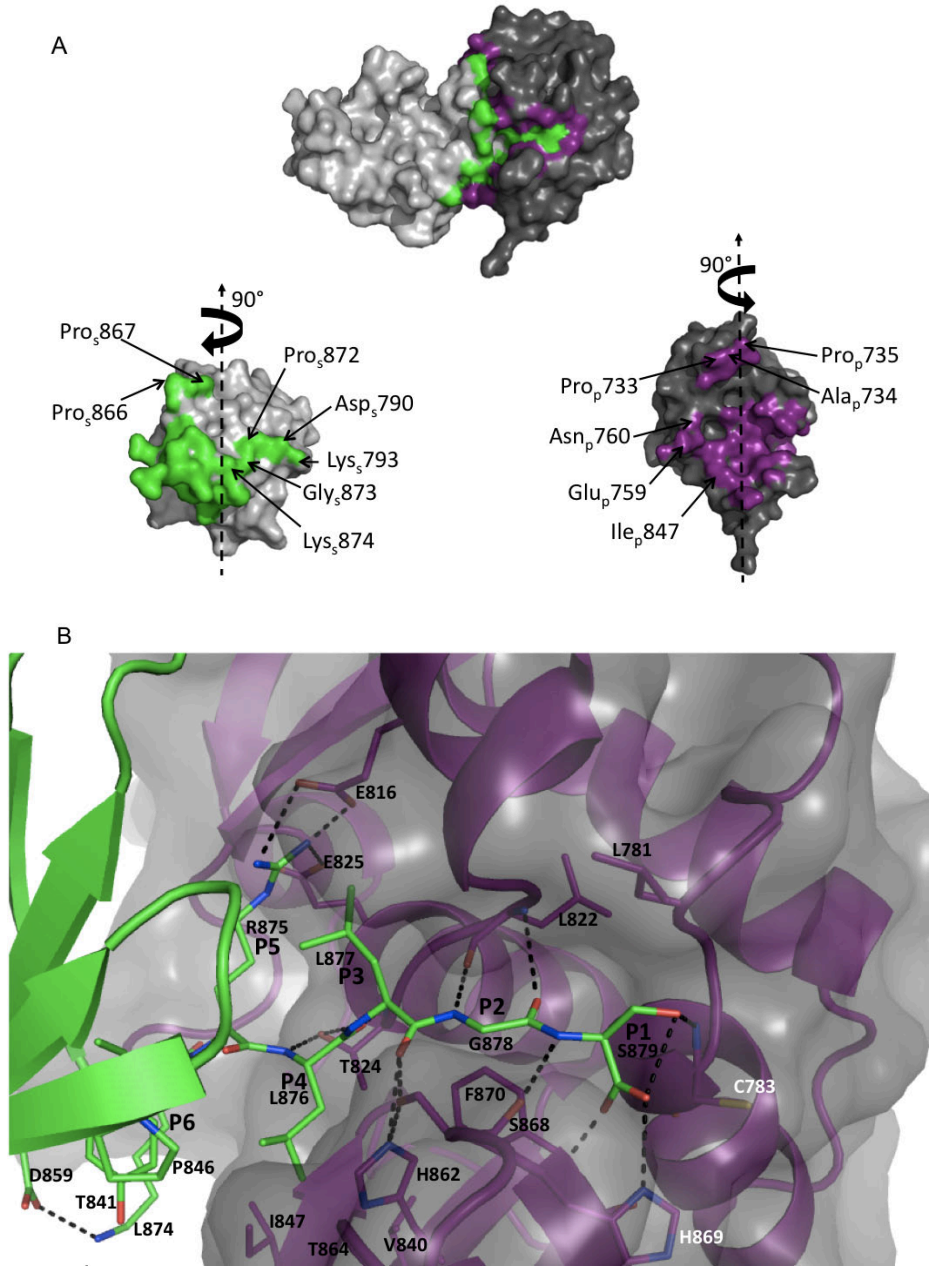


**A view of the N-terminal product complex in viral polyprotein processing.**

The infinite helices of PRO in the crystal are formed by the insertion of the C-terminus of one molecule into the catalytic site of the next. Thus, we have captured the N-terminal product complex resulting from the self-cleavage in *trans* of a viral polyprotein by its resident proteinase. The specificity of PRO is on the N-terminal (P) side of the scissile bond, while the C-terminal (P') side is not important as defined by mutagenesis studies (17)(7). This structure thus reveals the molecular determinants of PRO specificity in its processing proteinase function (Fig. 4). The specificity of PRO is defined as P5-(K/R)LX(G/A/S)(G/A/S)-P1 (7). The molecular determinants for this are now readily assigned by examining the interactions between one PRO molecule (hereafter called "substrate" with relevant residues with an "s" subscript) and the next (hereafter called "peptidase" with relevant residues with a "p" subscript).

Analysis of the peptidase-substrate interface using the PISA server (14) (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) shows that 940 $Å^2$ (11.9%) and 825 $Å^2$ (10.5%) of solvent-accessible surface area are buried in the complex for the substrate and peptidase, respectively. This interface is not expected to be stable in solution. Accordingly, we find that PRO solutions up to 10mg/ml are monodisperse as measured by dynamic light scattering (not shown), with an hydrodynamic radius of 2.3nm very close to the one calculated from the crystal structure for the monomer (2.14nm). The residues involved in the interface are mostly, but not exclusively (Fig. 4A) in and around the entry to the catalytic cleft of the peptidase and in the C-terminus of the substrate. The last five residues of the substrate 875-$Arg_s$-$Leu_s$-$Leu_s$-$Gly_s$-$Ser_s$-879 are funneled in an extended beta conformation towards the catalytic dyad of the peptidase $Cys_p$783-$His_p$869 (Fig. 4B). Indeed one of the carboxyterminal oxygens and the main chain nitrogen of $Ser_s$879 are hydrogen-bonded to the main chain nitrogen and carbonyl, respectively, of $Phe_p$870. Likewise, the main chain polar atoms of $Gly_s$878 are hydrogen-bonded to the main chain polar atoms of $Leu_p$822 in the connection between helices a6 and a7 on the other side of the catalytic cleft. The net effect is a small three-stranded intermolecular beta-sheet firmly holding 878-$Gly_s$-$Ser_s$-879 in place. Further upstream the substrate, the extended conformation of the main chain is maintained by hydrogen bonds from side chains of the peptidase. Key side chains are those of $His_p$862 (that also participates in the P4 and P2 specificity, see below) and $Thr_p$824 (Fig. 4B).

**Figure 4: Analysis of the PRO peptidase-substrate interaction.**
*A*, Surface representation of the interface. The peptidase molecule is colored grey and substrate molecule light grey. Atoms involved in the interface are colored magenta for the peptidase and green for the substrate. Top panel: Two PRO molecules as in the crystal. Bottom panels: the two molecules have been rotated to display the surfaces of interaction. Residues involved are labeled.
*B*, Representation of interactions between P sites of the substrate and catalytic cleft of the peptidase.
Secondary structure elements and carbons are colored magenta for the peptidase and green for the substrate. Other atoms are colored red for oxygens, blue for nitrogens and yellow for sulfurs. Hydrogen bonds and salt bridges are indicated by black dotted lines. The peptidase's catalytic dyad is labeled in white and other displayed residues in black.



**Structural basis of PRO proteinase sequence specificity.**

In our crystal structure, the side chain hydroxyl of $Ser_s879$ hydrogen bonds to the main chain nitrogen of the catalytic $Cys_p783$. The relaxed G/A/S specificity at P1 stems largely from the solvent exposure of the exit of the catalytic cleft (see above). Still, the side chain of $Leu_p781$ caps the S1 site, precluding the presence of a large side chain at position P1. Similarly, the constriction of the active site cleft at the S2 pocket is less pronounced than in the closest relatives of known structure (see below). The S2 pocket, lined by $His_p862$, $Phe_p870$ and $Ser_p868$ and occupied by $Gly_s878$ in our crystal structure, could readily accommodate a small side chain. At first sight, the sharp difference in specificity (7) between P3 (essentially no specificity) and P4 (strict specificity for a hydrophobic amino acid with a strong preference for Leu) is somewhat surprising. Both $Leu_s877$ (P3) and $Leu_s876$ (P4) make extensive contacts to conserved, shallow hydrophobic pockets at the surface of the peptidase's central and C-terminal lobes, respectively. These contacts bury respectively 89% of $Leu_s876$'s and 96% of $Leu_s877$'s solvent-accessible surface area, as reported by PISA. A major difference though is that $Leu_s877$ (P3) is buttressed only on one side, with a cavity in the interface between substrate and peptidase on the other side. In contrast, $Leu_s876$ (P4) is sandwiched between the two C-terminal lobes of the peptidase and substrate, with $His_p862$, $Phe_p870$, $Val_p840$, $Ile_p847$ and $Ser_p842$ on one side and $Pro_s846$ and $Tyr_s841$ on the other side. This likely accounts for the specificity at position P4. As for the strict P5 specificity for a positively charged amino acid, it is readily explained by salt bridges of $Arg_s875$ to two conserved glutamates that protrude from the central lobe of the peptidase, $Glu_p816$ in the small helix a6 overhanging the S3 pocket and $Glu_p825$ in helix a7.

**Peptidase-substrate recognition patches**

Molecular recognition of the substrate by the peptidase further involves patches recessed from the C-terminus of the substrate (Fig. 4A). Thus, there is a prominent hydrophobic contact between the double cis-proline of the substrate $Pro_s866$-$Pro_s867$ and $Pro_p733$-$Ala_p734$-$Pro_p735$ at the base of helix a1 in the N-terminal lobe of the peptidase. A second patch in this lobe is centered on $Asn_p760$ at the tip of the extended a2-b2 loop. $Asn_p760$ makes both a hydrogen bond to $Asp_s790$ and a stacking interaction to the $Pro_s872$-$Gly_s873$ motif that makes up the sharp turn to the C-terminal residues mentioned above. This contact is completed on one side by an electrostatic interaction between $Glu_p759$ and $Lys_s793$.

On the other side, the hydrophobic contact is taken up between the aliphatic part of the side chain of $Lys_s874$ and $Ile_p847$, an interaction that is continuous with the S4 pocket.

**PRO has an ubiquitin hydrolase (DUB) fold distantly related to the Ovarian Tumor (OTU) domain class of DUB enzymes.**

It was previously remarked that PRO shares limited sequence similarities around the two catalytic residues with the OTU domain class of DUB enzymes and we recently reported that PRO is a functional DUB (18) (Chenon et al, in press). We used the DALI server (19) (http://ekhidna.biocenter.helsinki.fi/dali_server) to seek the closest homologs of PRO with published structures. Although no close homolog is available in the Protein Data Bank (PDB) and the N-terminal lobe cannot be matched at all, the fold of the PRO catalytic domain is clearly the same as the core fold of the OTU1 cellular DUB (cOTU, *Saccharomyces cerevisiae*, DALI Z-score 7.5, 91 residues matched) (20) and nairovirus DUB (vOTU, *Bunyaviridae*, DALI Z-score 6.8, 90 residues matched) (21)(22). These two DUBs are assigned to clan CA of papain-like proteinases in the MEROPS peptidase database scheme (23) (http://merops.sanger.ac.uk/). Clan CA contains numerous peptidases including viral processing proteinases from *Picornaviridae* and *Coronaviridae*, but only DUBs have substantial DALI Z-scores in comparisons with PRO. Indeed, the nearest homolog of PRO with reported endopeptidase activity is the bacterial Staphopain (Z-score 3.4).

A DALI superposition of cOTU and vOTU yields a Z-score of 12.2. PRO is thus the more structurally divergent of the three homologous DUBs. This is due to cOTU and vOTU being structurally superimposable in places both downstream of the catalytic histidine and upstream of the catalytic cystein (126 residues matched by DALI). In the latter case, the segment directly upstream of the homolog of helix a3 covers the exit from the active site, making it less solvent-exposed than PRO's, in both cOTU and vOTU (Fig. 5A).

**Comparison with available OTU DUB substrate complexes.**

The structures of both cOTU and vOTU have been reported as covalent complexes with ubiquitin-derived probes (20)(21)(22). These works have shown that both DUBs recognize the b-grasp fold of ubiquitin by specific recognition of the same surfaces. These surfaces include the ubiquitin C-terminal residues and a remote hydrophobic patch centered on ubiquitin Ile44.

This patch is recognized by nonhomologous regions of the two DUBs due to a 75° rotation of ubiquitin around an axis defined by the main chain of the 5 C-terminal ubiquitin residues. In both cases though, these DUB regions are in positions occupied by the N-terminal lobe in PRO (Fig. 5A).

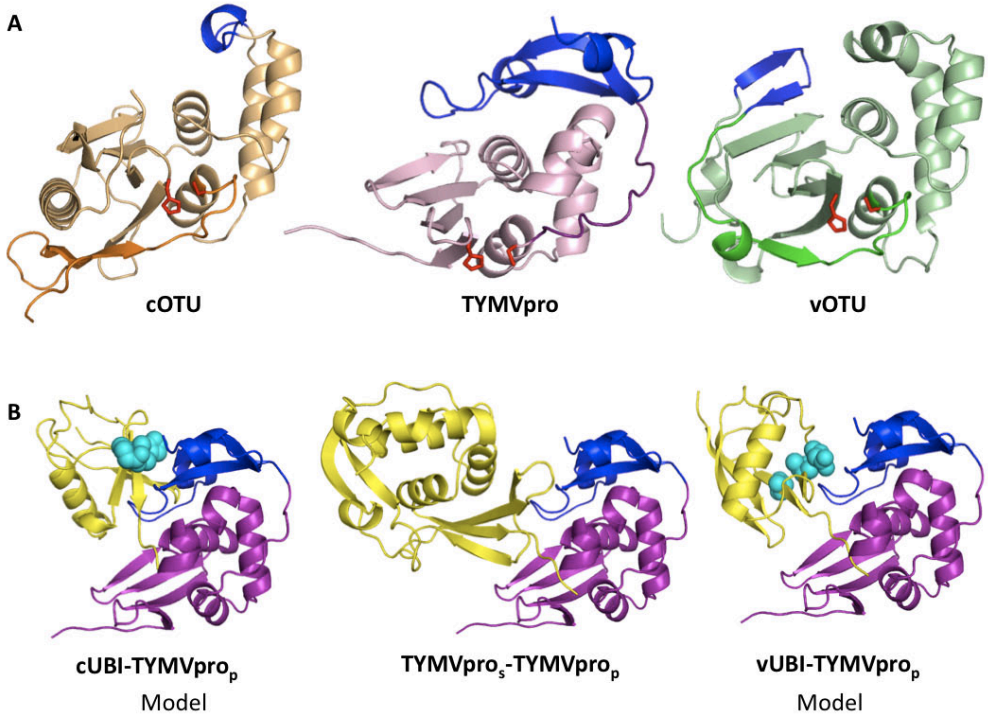
Indeed, when the three complexes of $PRO_p$–$PRO_s$, cOTU–ubiquitin and vOTU–ubiquitin are superimposed, the ubiquitin Ile44 patch is turned towards the surface of the N-terminal lobe, either the base of the b1-b2 sheet (cOTU–ubiquitin) or the tip of the extended a2-b2 loop (vOTU-ubiquitin) of $PRO_p$ (Fig. 5B). Furthermore, the orientation of the substrates' C-termini is much more similar in the two viral DUBs and the ubiquitin orientation in the cOTU complex would lead to clashes with the tip of the a2-b2 loop.

**Figure 5: Comparison with available structures of OTU DUBs**

Molecules are displayed in ribbon representation. The three OTU DUBs PRO, cOTU in complex with ubiquitin (cellular DUB, PDB code 3C0R) and vOTU in complex with ubiquitin (*Bunyaviridae* DUB, PDB code 3PHW) have been superimposed.

*A*, only the DUBs are displayed and colored light red (PRO, middle panel), light orange (cOTU, left panel) and light green (vOTU, right panel). The segments leading to the catalytic cysteine are in darker colors and the catalytic dyads are in red. Elements involved in the recognition of a hydrophobic patch centered on ubiquitin Ile44 are in blue in cOTU and vOTU. The N-terminal lobe of PRO is also in blue.

*B*, PRO in complex with itself (middle panel) and superimposed on complexes with ubiquitin (cOTU complex, left panel; vOTU complex, right panel. Only the ubiquitins are displayed). The PRO enzyme is in magenta with the N-terminal lobe in blue. The substrates are in yellow. Residues making up the ubiquitin hydrophobic patch (Ile44, Gln 49 and His68) are displayed as light blue spheres.



**Discussion**

In the present work, we provide structural insights into viral polyprotein processing by a viral proteinase that cleaves in *trans* at its own C-terminus. Although such an event is common enough in the viral world, particularly among positive-stranded RNA viruses, it is to our knowledge only the second time that the crystal structure of such an N-terminal product complex is reported (24). Our structure precludes the possibility of the C-terminus of PRO looping back towards the entry to the catalytic cleft. Therefore, cleavage of the TYMV replication polyprotein at the PRO/HEL junction (Fig.1) occurs only in *trans*. This cleavage is a regulatory event in the replication of the TYMV RNA genome. It occurs in the replication complex comprising the two products of the first cleavage: the 66K RdRp and the 140K protein. 140K harbors PRO and localization determinants to the chloroplast envelope, where it recruits 66K. There, cleavage of 140K at the PRO/HEL junction into 98K and the 42K helicase contributes to the switch to synthesis of the + strand (6)(7). As for other positive-stranded RNA viruses, a strictly *trans* cleavage likely takes part in the regulation by requiring a sufficient concentration of 140K at the chloroplast membrane and/or remodeling of this membrane into a special compartment for viral replication.

The interface in the crystal of the N-terminal complex of this *trans* cleavage reveals the molecular determinants for the peptidase sequence specificity. The fact that PRO proteinase specificity is confined to the P side is readily explained by the fact that the catalytic cleft ends abruptly at the catalytic dyad, leaving it completely solvent-exposed. Thus, residues on the P' side of the substrates will have little or no contact with PRO. In contrast, there are extensive interactions with the P-side residues up to P5 from both lobes making up the catalytic proteinase domain. The C-terminal b-sheet lobe thus provides a hydrophobic pocket for P4 and the central a-helical lobe salt bridges to the positively charged P5 and a hydrophobic patch for P3. The constriction of the active site cleft at the interface between the two lobes ensures that only small residues (but not necessarily glycines) can be at positions P2 and P1.

The fold of this two-lobed catalytic domain is clearly a more compact version of the OTU domain fold of ubiquitin hydrolases. The least dissimilar OTU domains to PRO are those of the cellular OTU1 DUB (cOTU) (20) and the recently reported viral OTU domain encoded in the L protein of Crimean–Congo haemorrhagic fever virus (vOTU) (21)(22). Thus, PRO is closest to enzymes with no endopeptidase activity. cOTU is a strict ubiquitin hydrolase, but vOTU also has interferon-stimulated gene 15 (ISG15) removal activity. We recently reported that 98K has DUB activity *in vivo* and *in vitro* and that this DUB activity is localized in PRO (Chenon et al, in press). This diversity of functions is remarkable in a proteinase domain that is smaller (148 ordered residues, Fig. 3) than either vOTU (162 ordered residues) or cOTU (170 ordered residues). Furthermore, it implies that PRO must recognize substrates that differ at positions P1 (G for ubiquitin, but A and S for the HEL/66K and PRO/HEL junctions, respectively) as well as cleave either endopeptide bonds (HEL/66K, PRO/HEL) or isopeptide bonds (ubiquitin, ISG15, Nedd8). Our crystal structure shows that the latter problem is neatly solved by the solvent exposure of the active site of PRO on the P' / lysine sidechain side of the cleavage, while cOTU and vOTU have an extra segment partially occluding the exit (Fig. 5A). The former problem of relaxed sequence specificity is solved by a lesser constriction of the PRO active site cleft at the P1 and P2 positions compared to cOTU and vOTU, where bulky residues strictly restrict specificity to GG.

Among peptidases that process polyproteins from RNA viruses with a Cys/His catalytic dyad, there are at least two structural clans with unrelated folds. The first is clan CA, that comprises cOTU and vOTU. Another is clan CN, whose type is the nsP2 proteinase of alphaviruses (25). Alphaviruses, including Sindbis virus, Semliki Forest Virus and Chikungunya virus, are animal relatives of tymoviruses. The two families share many features in their replication strategies, including successive cleavages of the replication polyprotein by the resident proteinase regulating RNA + vs – strand synthesis (7). Nevertheless, our data clearly assign PRO to clan CA. Although not unexpected (18) (Chenon et al, in press), this result could not be firmly established by sequence comparisons alone (http://merops.sanger.ac.uk/). The two other families of processing proteinases assigned to clan CA are also from positive-stranded RNA viruses: They are the coronavirus papain-like proteinases PLP1 and PLP2 (26)(27) and the picornavirus leader proteinase (28).

Strikingly, these proteinases have also been reported to be ubiquitin hydrolases (29). Examination of their structures shows that their active sites are less exposed than PRO's and more similar in this respect to cOTU's and vOTU's (not shown).

**A substrate recognition lobe unique to *Tymoviridae***

Together with the active site's exposure, the relaxed P1-P2 specificity of PRO allows it to hydrolyze diverse substrates comprising at least two nonidentical protein sequences and an ubiquitin conjugate. However, these features imply that PRO needs strong recognition elements away from the active site as extra specificity determinants. For the PRO/HEL cleavage, we identify two such elements in the N-terminal lobe that is specific to PRO and conserved among *Tymoviridae* both in sequence (Fig. 3) and in orientation (from our structural analysis): One element lies at the base of helix a1 and the other at the tip of the a2-b2 loop (Fig. 4A). The first element in particular recognizes a signature bulge made of two successive cis-prolines just upstream of the catalytic His in the PRO substrate molecule. Whether the N-terminal lobe is also prominently used in recognition of the other protein substrate, namely the HEL/66K junction, is unknown at present. However, superpositions with ubiquitin complexes reported for cOTU and vOTU (Fig. 5) lead us to propose that the N-terminal lobe will also be a major recognition module for this substrate of PRO. cOTU and vOTU are too divergent from PRO for us to be specific about any residues involved. Further work will be required to clarify these points and to assess whether PRO recognizes distinct target molecules in the same orientation (as vOTU does (21)(22)), or whether PRO uses distinct orientations of the targets and distinct recognition patches on the N-terminal lobe.

**Footnotes**

We thank the Structural biology and Proteomics pole of the IMAGIF integrated platform (https://www.imagif.cnrs.fr/) for access to Crystallization and Mass Spectrometry services and Synchrotrons SOLEIL (beamline PROXIMA 1) and ESRF (European Synchrotron Radiation Facility) for generous allocation of beamtime.

This work was supported in part by the Agence nationale de la recherche (contract ANR-06-BLAN-0062). CR was supported by a Ph.D fellowship from the French Ministry of Education.

Coordinates and structure factors will be available from the Protein Data Bank upon acceptance of this manuscript.

## References

1. Regenmortel, M. H. V., Fauquet, C., Bishop, D. H. L., Carstens, E., Estes, M., Lemon, S., Maniloff, J., Mayo, M., McGeoch, D., Pringle, C., and others (2000) *Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses.*, Academic Press
2. Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R., and Kaesberg, P. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4358-4362
3. Buck, K. W. (1996) *Adv. Virus Res.* **47**, 159-251
4. Prod'homme, D., Le Panse, S., Drugeon, G., and Jupin, I. (2001) *Virology* **281**, 88-101
5. Prod'homme, D., Jakubiec, A., Tournier, V., Drugeon, G., and Jupin, I. (2003) *J. Virol* **77**, 9124-9135
6. Jakubiec, A., Notaise, J., Tournier, V., Héricourt, F., Block, M. A., Drugeon, G., van Aelst, L., and Jupin, I. (2004) *J. Virol* **78**, 7945-7957
7. Jakubiec, A., Drugeon, G., Camborde, L., and Jupin, I. (2007) *J. Virol* **81**, 11402-11412
8. Camborde, L., Planchais, S., Tournier, V., Jakubiec, A., Drugeon, G., Lacassagne, E., Pflieger, S., Chenon, M., and Jupin, I. (2010) *The Plant Cell Online* **22**, 3142 -3152
9. Kabsch, W. (2010) *Acta Crystallogr. D Biol. Crystallogr* **66**, 125-132
10. Sheldrick, G. M. (2008) *Acta Crystallogr., A, Found. Crystallogr* **64**, 112-122
11. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) *Methods Mol. Biol* **364**, 215-230
12. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallogr. D Biol. Crystallogr* **66**, 213-221
13. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) *Acta Crystallogr. D Biol. Crystallogr* **66**, 486-501
14. Krissinel, E., and Henrick, K. (2007) *J. Mol. Biol* **372**, 774-797
15. Dali server: conservation mapping in 3D [online] http://nar.oxfordjournals.org.gate1.inist.fr/content/38/suppl_2/W545 (Accessed June 10, 2011).
16. Bond, C. S., and Schüttelkopf, A. W. (2009) *Acta Crystallogr D Biol Crystallogr* **65**, 510-512
17. BRANSOM, K. L., WALLACE, S. E., and DREHER, T. W. (1996) *Virology* **217**, 404-406
18. Makarova, K. S., Aravind, L., and Koonin, E. V. (2000) *Trends in Biochemical Sciences* **25**, 50-52
19. Holm, L., and Rosenstrom, P. (2010) *Nucleic Acids Research* **38**, W545-W549
20. Messick, T. E., Russell, N. S., Iwata, A. J., Sarachan, K. L., Shiekhattar, R., Shanks, J. R., Reyes-Turcu, F. E., Wilkinson, K. D., and Marmorstein, R. (2008) *J. Biol. Chem* **283**, 11038-11049
21. Akutsu, M., Ye, Y., Virdee, S., Chin, J. W., and Komander, D. (2011) *Proc. Natl. Acad. Sci. U.S.A* **108**, 2228-2233
22. James, T. W., Frias-Staheli, N., Bacik, J.-P., Levingston Macleod, J. M., Khajehpour, M., García-Sastre, A., and Mark, B. L. (2011) *Proc. Natl. Acad. Sci. U.S.A* **108**, 2222-2227
23. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010) *Nucleic Acids Res* **38**, D227-233
24. Lee, J., Feldman, A. R., Delmas, B., and Paetzel, M. (2007) *Journal of Biological Chemistry* **282**, 24928 -24937
25. Russo, A. T., White, M. A., and Watowich, S. J. (2006) *Structure* **14**, 1449-1458
26. Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C., and Mesecar, A. D. (2006) *Proceedings of the National Academy of Sciences* **103**, 5717 -5722
27. Wojdyla, J. A., Manolaridis, I., van Kasteren, P. B., Kikkert, M., Snijder, E. J., Gorbalenya, A. E., and Tucker, P. A. (2010) *J. Virol.* **84**, 10063-10073
28. Guarné, A., Tormo, J., Kirchweger, R., Pfistermueller, D., Fita, I., and Skern, T. (1998) *EMBO J* **17**, 7469-7479
29. Wang, D., Fang, L., Li, P., Sun, L., Fan, J., Zhang, Q., Luo, R., Liu, X., Li, K., Chen, H., Chen, Z., and Xiao, S. (2011) *J. Virol.* **85**, 3758-3766

# Compléments aux résultats de l'article : vue structurale de la fonction DUB

## Obtention d'un complexe stable entre TYMVpro et une ubiquitine suicide

Les déterminants structuraux concernant l'activité endopeptidase de TYMVpro étant à présent étudiés, nous avons cherché à identifier et comparer ceux impliqués dans l'activité ubiquitine hydrolase de cette protéase à cystéine. Pour cela, nous avons cherché à obtenir cette fois un complexe cristallographique de la protéase avec une molécule d'ubiquitine.

Plusieurs possibilités ont été envisagées alors. Tout d'abord, puisque nous disposions encore de cristaux robustes dans LA goutte ayant abouti aux petits hexagones, nous avons réalisé des trempages en présence d'ubiquitine. Les cristaux étant constitués de 69 % de solvant, la possibilité que l'ubiquitine s'y insère jusqu'aux molécules de TYMVpro était raisonnable. Après alignement structural avec des homologues résolus en présence d'une molécule d'ubiquitine (Figure 5 de l'Article 2), nous avons simulé l'attachement d'une molécule d'ubiquitine dans le réseau cristallin et avons supposé alors qu'elle pouvait accéder au site catalytique de la protéase sans encombre (Figure 19). En revanche, le site actif concernant la fonction endopeptidase et la fonction déubiquitinase étant identiques, il faudrait que les molécules d'ubiquitines déplacent les C-termini de TYMVpro pour s'insérer dans le canal menant au site actif. Cette opération étant en revanche beaucoup moins probable et risquant de dégrader les cristaux dans les cas les plus favorables, nous ne nous sommes pas attardé sur ces expériences de trempage après les premières analyses de diffraction ne révélant aucune formation de complexe.

Nous avons finalement préféré une méthode plus « raisonnable ». En effet, nous avons ensuite cherché à former un complexe stable entre l'ubiquitine et TYMVpro afin de le purifier et le cristalliser. Ces expériences ont été réalisées grâce à des ubiquitines suicides auxquelles sont greffés des groupes fonctionnels pouvant se lier de façon irréversible aux cystéines actives de certaines DUBs. Ces dérivés sont commercialement disponibles (http://www.enzolifesciences.com) et des protocoles pour la formation de complexes avec des DUBs ont déjà été publiés (Capodagli 2011, Akutsu 2011, Boudreaux 2010, Edelmann 2009, Hemelaar 2004, Borodovsky 2002). Cependant, la DUB de TYMV est assez éloignée de celles utilisées dans ces cas et une mise au point est inévitable.

De plus, ces composés sont destinés à l'identification de la fonction DUB, et une étude par cristallographie aux rayons X demande beaucoup de matériel biologique pour les essais de cristallogenèse.

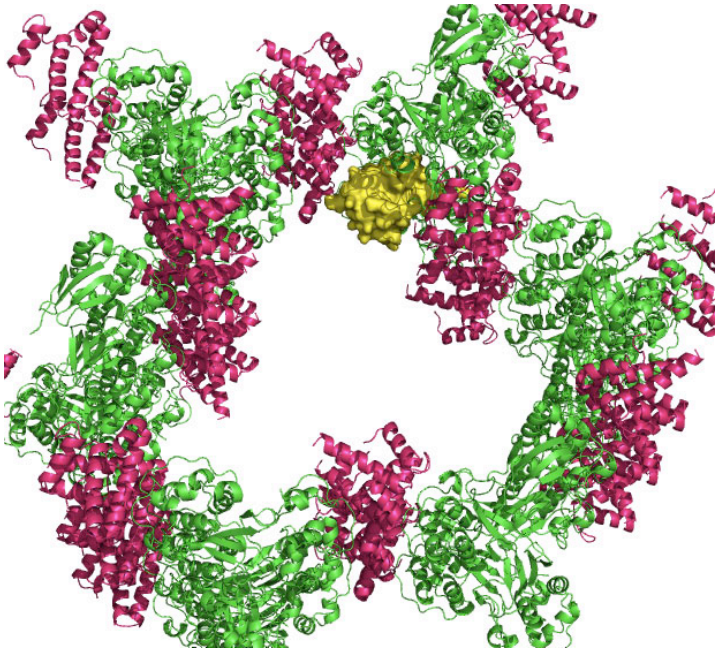


**Figure 19 : Simulation d'accession d'une molécule d'ubiquitine dans l'empilement cristallin**
L'empilement cristallin est représenté en cartoon dans pymol. Les molécules de TYMVpro sont en vert et les molécules de S15 en rose. Une molécule d'ubiquitine est finalement représentée en surface jaune. Elle est issue de l'alignement DALI de TYMVpro avec vOTU portant le code PDB 3PHW et vu dans la Figure 5 de l'Article 2. Seule la chaine B correspondante à l'ubiquitine est alors visible.

Ces composés sont alors trop coûteux dans le cadre envisagé ici. Nous avons donc cherché à synthétiser directement au laboratoire les dérivés afin d'envisager par la suite des essais de cristallisation avec suffisamment de matériel protéique pour cela. Trois composés étaient réalisables simplement et nous avons donc commencé nos investigations par ceux-ci et l'un de ces trois composés a d'ailleurs été cristallisé en présence de l'homologue structuralement le plus proche de TYMVpro. Ces ubiquitines modifiées sont en fait greffées à des groupements électrophiles pour les cystéines qui forment une liaison covalente en présence de certaines protéases. Ces groupements sont introduits chimiquement en remplaçant la glycine C-terminale de l'ubiquitine par un analogue portant le groupement désiré.

Pour cela, une protéine comprenant les résidus 1-75 de l'ubiquitine en fusion avec un domaine intéine est exprimée (Figure 20). Cette protéine chimérique comprend aussi un domaine de liaison à la chitine, ce qui permet de la purifier simplement. Une série de réactions chimiques permet alors de substituer au domaine intéine un analogue de glycine portant le groupement électrophile analogue à une glycine désiré (Figure 20).

Les trois groupements dont nous nous sommes servis sont le bromoéthylamine (Br2), le bromopropylamine (Br3) et le chloroéthylamine (Cl). Après chaque synthèse, tous les substrats suicides formés ont bien entendu été vérifié par spectrométrie de masse. Puis, ils ont été mis en contact avec la protéase purifiée. Malheureusement, aucune interaction covalente n'a pu être détectée après purification et dépôt sur gel de polyacrylamide.

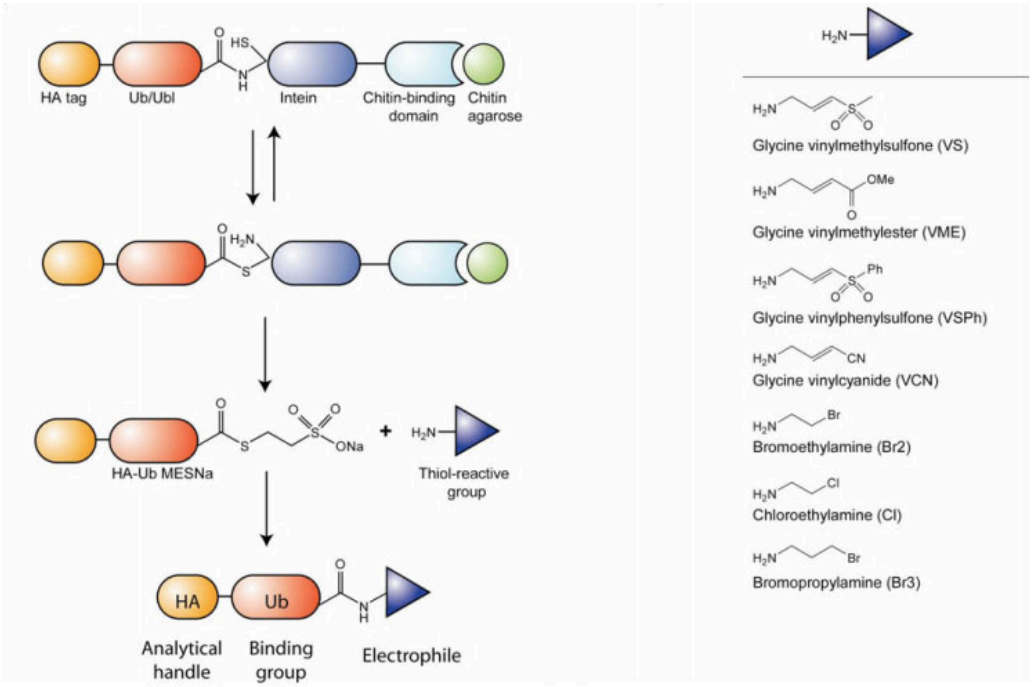



**Figure 20 : Synthèse des dérivés HAUb**
Représentation des réactions nécessaires à la synthèse des HAUb suicides par ligation reposant sur une intéine. A gauche la cascade réactionnelle et à droite les groupements possibles pour la ligation C-terminale sur l'ubiquitine étiquettée HA (Love 2007).

Nos collaborateurs ont montré que la protéase était en revanche capable de se lier à l'ubiquitine vinyl sulfone (Chenon 2011). Nous avons donc testé quelques nanogrammes de composés commerciaux afin de s'assurer de cette interaction dans nos conditions et avons décidé d'entamer une collaboration avec le groupe du Dr. Ploegh (Whitehead Institute for Biomedical Research - Nine Cambridge Center - Cambridge MA, 02142) afin qu'ils nous fournissent des quantités plus importantes de ce substrat suicide pour pouvoir par la suite envisager la cristallisation du complexe. Ce laboratoire a répondu favorablement à notre demande, mais ne pouvait nous fournir que de l'ubiquitine vinyl méthyl ester. Nous avons donc de nouveau caractérisé l'interaction de TYMVpro avec ce substrat suicide et après constatation de son interaction avec notre déubiquitinase, nous avons initié la mise au point d'un protocole visant à former plusieurs milligrammes de complexe. Une fois le protocole optimisé, nous avons cherché à le purifier avant de nous lancer enfin dans les essais de cristallisation.

## Matériel et méthode

*Expression/Production/purification de TYMVpro*

Le gène codant pour la protéase PRO a été inséré dans un plasmide dérivé du vecteur pGex permettant d'étiqueter PRO au niveau de son extrémité N-terminale avec 6 histidines (Figure 21). Le plasmide est ensuite transformé dans *Escherichia coli* Rosetta (DE3) et une pré-culture est réalisée sur la nuit à 37°C. Cette pré-culture est alors utilisée pour inoculer 1L de milieu Luria Broth (LB) contenant 50 µg/L carbenicillin et 25 µg/L chloramphénicol. La culture est mise à 37°C et lorsque la densité optique mesurée à la longueur d'onde de 600 nm atteint la valeur de 0.6, l'expression est induite par addition de 0.5 mM isopropyl-b-D-thiogalactopyranoside (IPTG). Les cellules sont ensuite remises en culture pendant 4 heures à 30°C et à l'issue de cette incubation, le culot bactérien est collecté par centrifugation, congelé et stocké à -20°C.

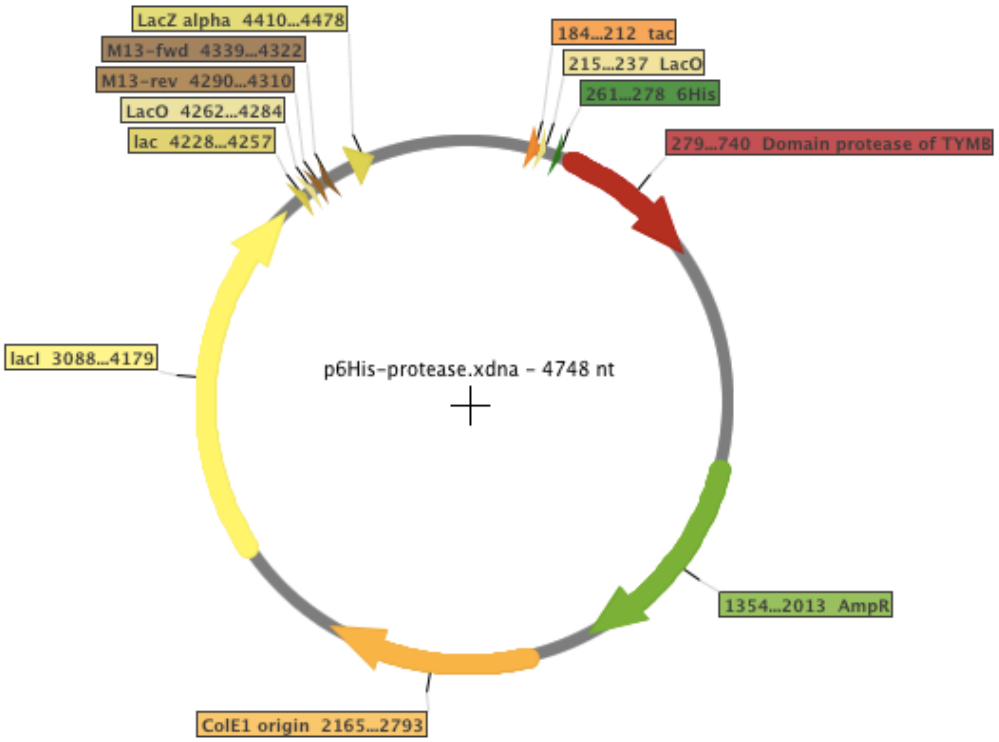



**Figure 21 : Carte plasmidique de la construction utilisée pour l'expression du domaine TYMVpro**
Le gène codant pour le domaine TYMVpro a été inséré dans un vecteur dérivé du pGex et portant un gène de résistance à l'ampiciline.

Approximativement 5 g de culot sont resuspendus dans 20 mL de tampon de lyse (100 mM Tris-HCl pH 7.5, 350 mM NaCl, 25 mM imidazole, 1 mM DTT, 0.5 % Triton X-100, 2 mg/ml Lysozyme, et 1 U/ml Benzonase) et incubés 60 min dans la glace. La lyse est ensuite complétée par 5 cycles de congélation/décongélation dans l'azote liquide (70 K/303 K).

Le lysat est centrifugé à 8000 g durant 30 min et le surnageant est injecté sur une colonne contenant 1 mL de résine $Ni^{2+}$-NTA agarose (Qiagen). La colonne est initialement pré-équilibrée en tampon A (100 mM Tris-HCl pH 7.5, 350 mM NaCl, 25 mM imidazole et 1 mM DTT). Elle est, après injection, lavée avec 50 mL de ce même tampon, puis 10 mL de Tampon A[2] (100 mM Tris-HCl, pH 6.0, 350 mM NaCl, 25 mM imidazole et 1 mM DTT). TYMVpro est ensuite éluée grâce au tampon d'élution B (100 mM Tris-HCl, pH 7.5, 350 mM NaCl, 500 mM imidazole et 1 mM DTT).

Les fractions éluées sont finalement purifiées par gel filtration à l'aide d'une colonne Superdex S-75 (Amersham) en présence de tampon C (10 mM Tris-HCl pH 8, 350 mM Ammonium Acétate et 1 mM DTT). Les fractions finales de TYMVpro sont identifiées sur gel de polyacrylamide 16,5 % puis congelés dans l'azote liquide et stockés à -80°C (Figure 22).

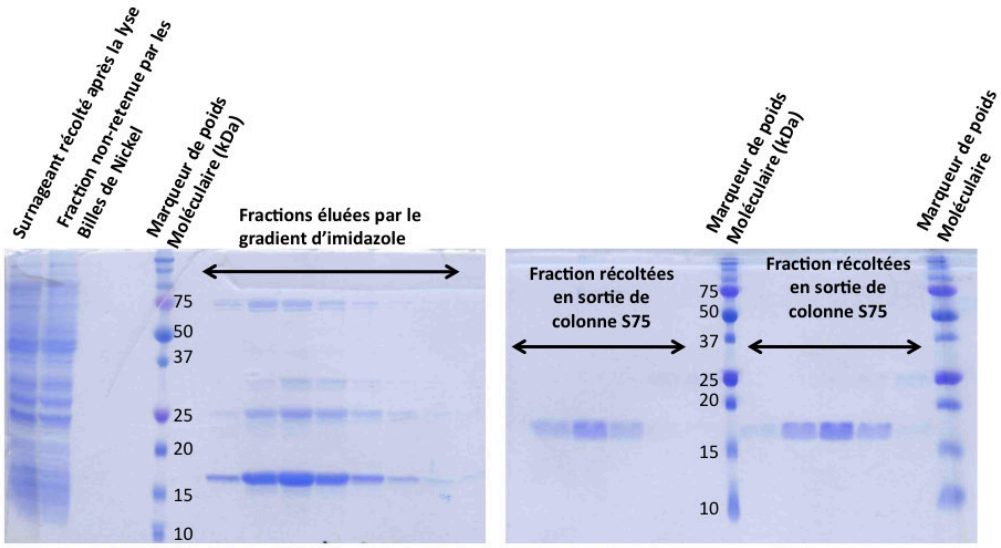



**Figure 22 : Analyse des étapes de purification du domaine TYMVpro**
Les différentes fractions protéiques des étapes de purification sont déposées sur gel de polyacrylamide SDS-page 16,5 % et l'on peut observer ici le résultat de la migration.

*Cristallisation de TYMVpro*

Les fractions d'élution sont rassemblées et concentrées entre 5 et 34 mg/mL. A partir de ces différentes concentrations, l'analyse robotisée de centaines de conditions est réalisée grâce à la plateforme dédiée imagif (www.imagif.cnrs.fr). Des kits commerciaux sont utilisés et la méthode de la goutte assise réalisée avec 100 nL de chaque solution. En effet, pour la formation de la goutte, le robot utilise 100 nL de protéase concentrée dans le tampon C et 100 nL de la solution de cristallisation commerciale. Une goutte, réalisée avec la solution concentrée à 34 mg/mL et une solution de cristallisation qui contient 0,1 M Hepes à pH 7,5 et 2,0 M d'ammonium formate, a permit la formation de petites aiguilles en moins de 5 jours.

A partir de ces informations, des tests plus fins et en plus gros volumes ont été réalisés manuellement par goutte suspendue avec 1 µl de chaque solution. A diverses reprises les aiguilles ont été reproduites mais aucune diffraction n'a été observée. Après de nombreux essais, une solution concentrée à 39 mg/mL et fortement contaminée (Article 2) a permis dans une unique goutte l'obtention de petit hexagones de 50 x 50 x 40 $\mu m^3$, après la formation préalable des aiguilles habituelles (Article 1). Ces hexagones ont a leur tour été testés par diffraction aux rayons X et cette fois, des données allant jusqu'à 2 Å de résolution ont pu être observées. La goutte n'a par la suite jamais pu être reproduite.

*Collecte des données et phasage par MIRAS*

Toutes les données cristallographiques ont été obtenues à partir de cristaux issus de cette unique goutte et le détail de la résolution de la structure fait l'objet de l'Article 1 exposé dans la Partie précédente.

*Expression et production de l'HAUb-activée*

Le gène codant pour l'ubiqutine humaine portant une étiquette HA (HAUb) en 5' et délétée de sa glycine 76, a été inséré dans le plasmide pTYB2. Ce vecteur permet de fusionner l'ubiquitine avec l'intéine et un domaine de liaison à la chitine. Ce plasmide nous a été fournit directement par le Dr. Ploegh (Whitehead Institute for Biomedical Research - Nine Cambridge Center - Cambridge MA, 02142).

Ce plasmide est transformé dans *Escherichia coli* BL21 et une pré-culture est réalisée sur la nuit à 37°C. Cette pré-culture est utilisée ensuite pour inoculer 1 L de milieu LB et cette culture est mise à 37°C. Lorsque la densité optique mesurée à la longueur d'onde de 600 nm atteint la valeur de 0.6, l'expression est induite par addition de 0.5 mM d'IPTG et transférée à 18°C durant la nuit. Le matin, le culot bactérien est collecté par centrifugation, congelé et stocké à -20°C.
Un culot d'environ 7 g est repris dans 25 mL de Tp A (50 mM Hepes pH 6,5, 50 mM NaOAc, 100 mM NaCl) et la lyse est réalisée par sonication (3 cycles de 1 minutes toutes les 2 minutes). Le lysat est finalement centrifugé à 8000 xg pendant 20 min.

*Purification d'Ubi-MESNa*

Le surnageant de lyse est injecté avec un débit sur une colonne contenant 5 mL de résine de chitine (Biolabs) initialement pré-équilibrée en tampon A, avec un débit de 0,5 mL/min. La colonne est ensuite lavée, toujours avec le Tampon A, puis équilibrée en Tampon B (Tp A + 100 mM de β-mercaptoethanesulfonic acid sodium salt (MESNa)). Dans cet état, la colonne est incubée sur la nuit à 37°C afin d'induire le clivage de la résine. Le lendemain, l'élution de l'ubiquitine MESNa thioester est réalisée avec 25 mL supplémentaires de tampon B.

*Synthèse et purification d'HAUb-Br2, d'HAUb-Br3 ou HAUb-Cl*

Les fractions d'élution sont concentrées dans un volume final de 500 µL soit environ 2 mg/mL de protéine. A cette solution concentrée en HAUb-MESNa, une solution de 0,2 mmol de 2-bromoethyl amine hydrobromide (Br2) ou de chloroéthyl amine (Cl) est ajoutée ainsi que 100 µL de solution de NaOH à 2 M. Ce mélange est vortexé et incubé 20 minutes à température ambiante puis, 100 µL de solution d'HCl à 2 M sont finalement ajoutés. Cette solution est dialysée contre 50 mM NaOAc de pH 4,5 pendant 2 heures dans une cassette Pierce de 3,5 mL et de cut-off 3,5 K. Des aliquots sont réalisés et congelés à la température de -80°C et quelques microlitres sont déposés sur gel de polyacrylamide afin d'analyser les bandes correspondantes par spectrométrie de masse (Figure 23).

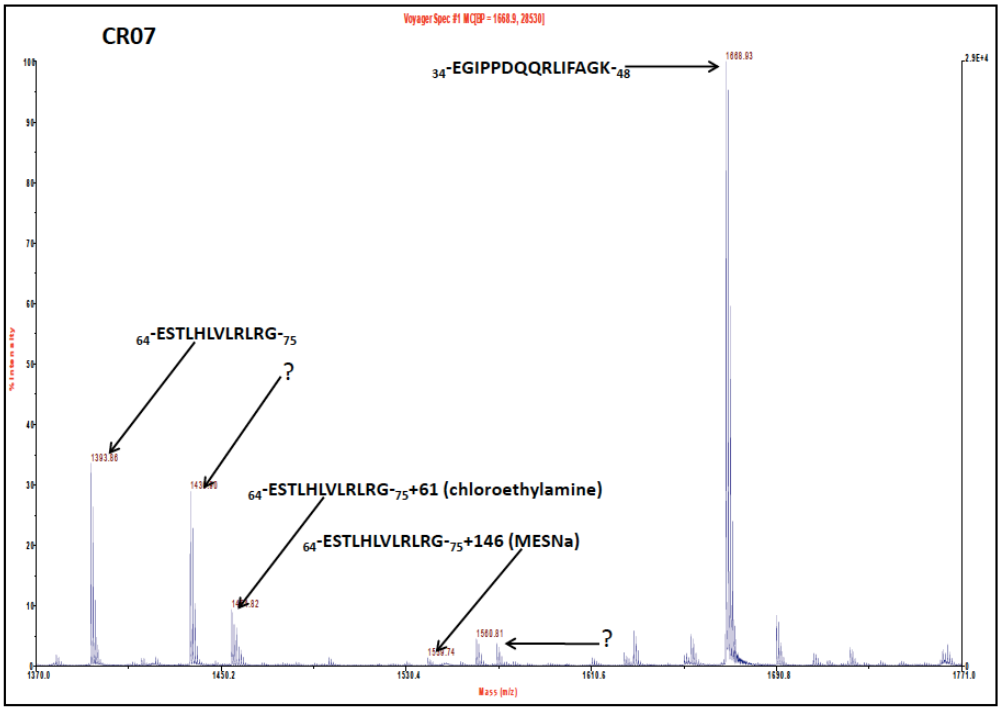



**Figure 23 : Spectre de masse du produit de synthèse de l'HAUb-CL**
La bande correspondante au produit de synthèse de l'HAUb-CL a été découpée sur gel tris-tricine et digéré par l'enzyme lysC. Une empreinte peptidique MALDI et une recherche ciblée de peptides C-terminal modifiés par un groupement MESNa ou CL a été réalisé. Un signal a été détecté à un rapport m/z=1393.8 et correspondant au peptide C-terminal non modifié, avec une précision de massecomprise entre 13 et 24 ppm. Un second pic à m/z=1454.82 et ayant une différence de masse de +61 a été détecté et correspond donc au peptide C-ter modifié par un groupement CL. On note aussi la présence d'un pic à m/z=1539.74 ayant une différence de masse de +146 et correspondant donc à l'HAUB-MESNa. La synthèse n'est donc pas totale mais cela a déjà été rapporté. On retient donc principalement que de l'HAUB-CL est bien synthétisé et que sa purification utlérieure sera suffisante.

*Mise en contact pour la formation du complexe*

Dans le cas des dérivés synthétisés par nous même, de nombreux protocoles ont été testés. Tout d'abord, nous appuyant sur les publications Akutsu 2011, Edelmann 2009 et Borodovsky 2002, nous avons réalisé nos premiers tests. Aucun complexe n'étant visible, nous avons essayé le tampon utilisé par nos collaborateurs lors des essais avec l'ubiquitine VS (Chenon 2011). Là encore, aucun complexe ne s'est formé.

Dans le cas de l'HAUb-VME reçue, le protocole final après optimisation est le suivant. L'ubiquitine suicide est concentrée à 2 mg/mL et congelée à -80°C. Les fractions purifiées de TYMVpro sont quant à elles concentrées à 15 mg/mL et congelées égalements. Afin de former le complexe, les deux protéines pures sont décongelées et diluées à 0,8 mg/mL pour HAUb-VME et 0,4 mg/mL pour TYMVpro dans une solution contenant 50 mM Tris HCl pH 8, 25 mM $MgCl_2$, 2,5 mM EDTA ajusté à pH 8, 150 mM NaCl et 1 mM de DTT. Après incubation 15 minutes à température ambiante, les deux solutions diluées de protéines sont additionnées en volumes équivalents et incubées 1 h à température ambiante. La formation du complexe est vérifiée sur gel Tris-Tricine 16,5 % avant sa purification (Figure 24).

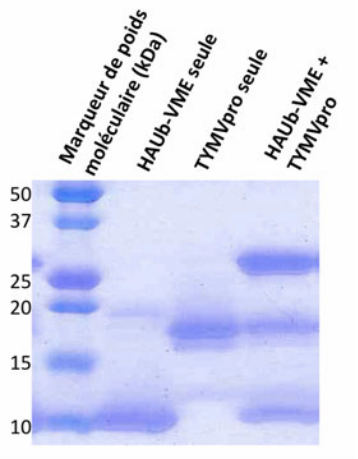



**Figure 24: Analyse sur gel du complexe HAUb-VME / TYMVpro**
Si l'on compare la piste correspondante à la protéase seule et celle de l'HAUb-VME seule à celle correspondante à la mise en contact de la protéase avec l'ubiquitine suicide, 1 h à 4°C, on constate que du complexe se forme bien. On remarque aussi que les quantité formées sont raisonnables pour une étude cristallographique à suivre puisqu'ici 5 microlitre ont été déposé et la concentration initiale en protéase est de 0,4 mg/mL et le double pour l'ubiquitine suicide HAUb-VME.

# Poursuite du projet

Le complexe formé et purifié est maintenant en cours de cristallisation. Nous en sommes à la recherche de signes cristallins en testant un large spectre de solutions de cristallisation de façon robotisée. Les signes éventuels seront utilisés pour la cristallogenèse manuelle en plus gros volume et nous espérons alors obtenir des cristaux exploitables pour la diffraction des rayons X afin de pouvoir accéder à la structure de ce complexe.

Avec cette structure nous souhaitons réaliser des comparaisons structurales avec le dimère de protéase-enzyme / protéase-substrat marquant la fonction endopeptidase, et observer les différences qui existent avec la fonction déubiquitinase mise en évidence par ce nouveau complexe. Le site actif étant le même et le mécanisme catalytique identique nous souhaitons essentiellement observer des différences dans les déterminants impliqués dans la reconnaissance des substrats. Nous aimerions également déterminer l'implication du lobe N-terminal dans la fonction déubiquitinase. Tout cela dans le but de comprendre comment, au cours du cycle viral, TYMVpro peut discerner ces deux fonctions. Ce travail pourra ensuite être supplémenté par de nouveaux complexe avec comme substrat potentiels : des chaines d'ubiquitine ou la polymérase du TYMV ubiquitinylée ou non.

# PARTIE III

# Obtention d'un fragment soluble de la RdRp du TYMV

# Introduction

L'ARN-polymérase dépendante de l'ARN (66K) est également un acteur clé de la réplication virale. Elle synthétise l'ARN complémentaire du génome de polarité négative, puis, en utilisant ce brin comme matrice, synthétise de nouveaux génomes, ainsi qu'un ARN sous-génomique codant pour la protéine de capside chez le TYMV.

La régulation de tous ces événements de synthèse d'ARN est un phénomène complexe, qui implique des interactions entre la 66K et d'autres protéines virales ainsi que des protéines cellulaires. On sait notamment que, comme pour tous les virus à ARN positif, le complexe de réplication est associé à une membrane cellulaire. Dans le cas de TYMV, c'est l'enveloppe des chloroplastes qui est concernée.

La régulation précise des différentes étapes de la réplication implique en premier lieu le clivage de la 66K, initialement incluse dans un précurseur polyprotéique, puis son recrutement à la membrane par l'intermédiaire d'une interaction avec le domaine protéase. La synthèse sous forme de précurseur polypeptidique implique que les protéines nécessaires à la réplication sont synthétisées en quantité équimolaire. Afin de réguler leur quantité ainsi que leur maturation, plusieurs moyens vont être engagés. Un mode de régulation de la 66K a été mit en évidence récemment par nos collaborateurs. En effet, la phosphorylation réversible de la 66K sur trois sites spécifiques a été identifiée (Jakubiec 2006) et l'on sait que deux premiers sites conduisent, lorsqu'ils sont phosphorylés, à la dégradation de la 66K et régulent ainsi sa quantité disponible dans la cellule. Le troisième site quant à lui, constitue un événement de régulation plus tardif en touchant directement la synthèse d'ARN.

Localisées dans la séquence PEST N-terminale, la thréonine 64 et la sérine 80 sont, dans leur état phosphorylé, un signal pour l'ubiquitination de la 66K. La polymérase, une fois polyubiquitinylée, est ensuite adressée au protéasome en vue de sa dégradation. De façon intéressante, la protéase peut, via différents moyens, intervenir pour la protéger de ce processus d'adressage et de dégradation. Lors de son interaction directe avec la polymérase, PRO peut venir masquer les sites de phosphorylation en question et empêcher le signal de se faire. Dans le cas où la polyubiquitination a déjà eu lieu, la protéase peut intervenir cette fois en tant que déubiquitinase pour cliver l'étiquette d'adressage constituée au minimum de quatre ubiquitines.

Le troisième site quant à lui, concerne la sérine 326, et permet de moduler directement la synthèse ARN au niveau du ratio entre les brins (-) et (+). Ce site est localisé dans la partie C-terminale du motif A de la paume. Ce motif, retrouvé chez toutes les polymérases, est très conservé chez les virus. Du fait que d'autres polymérases virales soient connues comme étant phosphorylées et compte tenu de la conservation de ce résidu au sein du site catalytique, on peut penser que le mécanisme de régulation qui lui est associé, peut être retrouvé similaire chez d'autres virus à ARN simple brin de polarité positive. On sait notamment que la polymérase du virus de la dengue de type 2 (Kapoor 1995), celle du virus de la mosaique du concombre (Kim 2002) et du norovirus (Eden 2011) sont phosphorylées. Nous nous sommes donc intéressé à la RdRp du TYMV afin d'en faire un modèle d'étude conformationnel. En effet, en résolvant dans un premier temps sa structure native puis, celle de mutants constitutifs touchant la sérine 326, par cristallographie aux rayons X, nous espérions observer les éventuels changements conformationnels induits par une telle modification. Ces mutants mimeraient ainsi la phosphorylation qui a lieu au cours du cycle viral grâce à la charge d'un aspartate ou, au contraire, en seraient déficients grâce à une alanine remplaçant la sérine 326. Grâce à la caractérisation avancée au niveau mécanistique cellulaire et à ces structures tridimensionnelles, nous espérons ainsi comprendre plus en détails le mode de régulation par phosphorylation qui touche cette polymérase et relier ces structures à sa fonction.

Pour étudier les trois sites de phosphorylation mis en évidence par nos collaborateurs, nous avons cherché à cloner un fragment commençant par la séquence PEST et englobant le domaine polymérase. Ainsi, le domaine « TYMVpol » choisis est constitué des 585 derniers résidus de la polyprotéine et correspond à la 66K dans son ensemble. Cependant, les séquences PEST sont généralement non structurées et constituent un signal pour la dégradation de la protéine en contexte cellulaire (Rechsteiner 1996). Nous n'étions donc pas certains d'obtenir des productions stables de ce domaine complet. L'expression d'un fragment plus court, délété de la partie N-terminale a aussi été envisagé et dans ce cas, seule la sérine 326 serait étudiée d'un point de vue structural. Cependant, ne sachant pas où est délimitée précisément la séquence PEST, et les alignements de séquences réalisés avec d'autres polymérases virales ne nous éclairant pas davantage, nous avons commencé par l'expression du fragment complet de 585 résidus. De la protéolyse ménagée sur les productions purifiées de ce domaine TYMVpol ont ensuite été réalisées pour résoudre l'incertitude concernant la délétion de la partie N-terminale.

Pour en arriver là, nous nous sommes heurtés à de nombreux problèmes à commencer par l'expression, initiée il y a plusieurs années par nos collaborateurs. Cette étape a été en effet la plus limitante pour ce projet. De nombreuses constructions ont été essayées ainsi que divers systèmes d'expression avant d'obtenir enfin le fragment désiré. Lorsque celui-ci a pu être exprimé, sa solubilité a également posé des problèmes pour la poursuite du projet. Nous avons là encore tenté de nombreux protocoles afin d'accéder finalement à une production soluble, pure et en quantité suffisante. Malheureusement, bien que le fragment TYMVpol semblait replié, il ne présentait aucune activité. Des essais de cristallisation ont malgré tout été réalisés sans succès et des expériences de protéolyse ménagée n'ont pas permis d'en savoir plus sur la délétion éventuelle de la séquence PEST. Un changement radical de stratégie a finalement été engagé afin d'obtenir cette RdRp soluble et active.

# Résultats

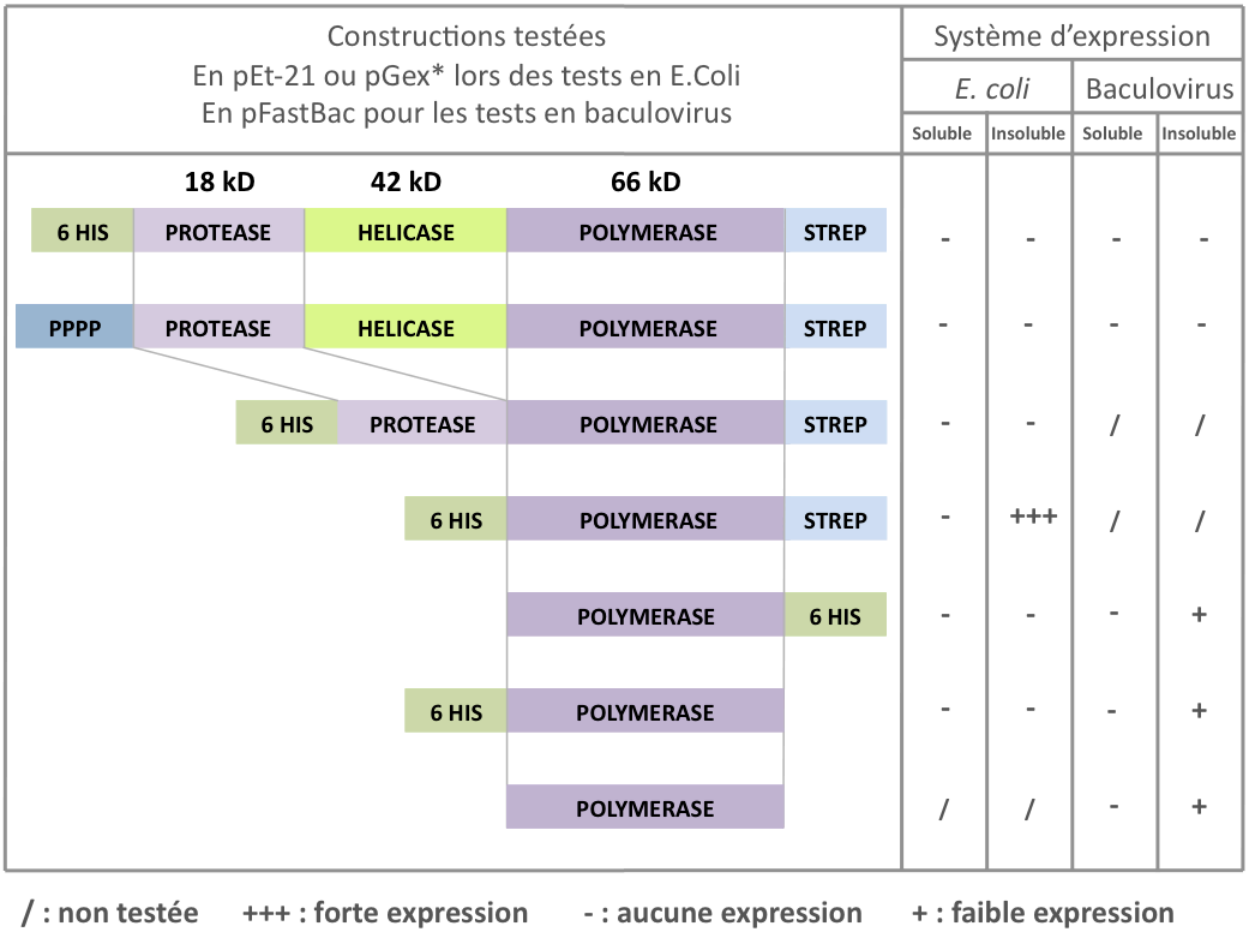
## Expression du domaine TYMVpol

Dans un premier temps, des tests avec différents plasmides, dont celui dérivé du pGex utilisé pour l'expression de TYMVpro, ont été réalisés. Aucune expression n'a cependant été détectée et ce quelle que soit la souche d'*E. coli* et la construction testée par nos collaborateurs (Table 2). Plusieurs étiquettes ont été fusionnées au niveau C-terminal et/ou N-terminal et, le domaine d'expression a également été étendu en amont afin de reconstituer intégralement ou partiellement la polyprotéine et ainsi associer à l'expression de la RdRp, celles de ses partenaires viraux.
Un changement de système d'expression a alors été envisagé. On savait déjà que la production en plante n'aboutissait pas à des quantités produites suffisantes de 66K pour la cristallographie. Nous avons donc essayé une autre alternative : l'expression en cellules d'insecte, plus proche du système eucaryote naturel du TYMV que ne l'est *E. coli*. Précedemment, l'expression en baculovirus de la 66K seule aboutissait à des productions majoritairement ou totalement insolubles. Nous avons donc essayé de cloner en pFastbac les constructions non testées jusqu'alors, soit un fragment plus étendu ou la 66K portant une double étiquette au niveau N-terminal et C-terminal (Table 2). Dans tous les cas, aucune expression n'a pu être détectée.

Finalement, c'est un changement de plasmide bactérien qui a été la clé à ce problème, associé à la construction 6His-66K-ST uniquement (Table 2). Effectivement, avec le gène du domaine TYMVpol doublement étiquetté et inséré dans le pET21-b(+), tous les tests d'expression réalisés ont permis d'observer une forte expression et ce dans diverses souches d'*E. coli*. Malheureusement, la totalité de la polymérase de TYMV exprimée se trouvait dans la fraction insoluble.

Afin de corriger ce problème de solubilité, nous avons essayé de jouer sur les conditions de production et/ou de lyse bactérienne.

Table 2 : Différents tests d'expression réalisés




## Production et lyse bactérienne

Afin de solubiliser une partie ou la totalité de TYMVpol lors de sa production, nous avons tout d'abord modulé les paramètres d'expression. Lors des cultures nous avons ainsi fait varier la température, le volume de culture, le milieu de culture, le temps d'induction, la concentration en IPTG, et/ou l'ajout ou non d'Ethanol ou de sorbitol et de bétaïne dans la culture. Aucun de ces paramètres ne nous a apporté de solution quant à la solubilisation de la 66K et l'intégralité était toujours insoluble (Table 3). Nous avons ensuite essayé de changer les paramètres de lyse en modulant la composition du tampon (Table 4) ainsi que la méthode en elle même (Table 5). Là encore, aucune variation n'a permit de solubiliser ne serait-ce qu'une partie de la protéine d'intérêt.

Finalement, nous sommes repartis de la meilleure production insoluble afin de forcer la solubilisation du domaine TYMVpol par dénaturation puis repliement. Après traitement par l'urée du culot issu de la lyse et dépôt sur gel, les quantités obtenues étaient importantes et plus que suffisantes pour poursuivre par des tests de renaturation (Figure 25).

La bande d'un poids moléculaire de 66kDa a été analysée par spectrométrie de masse préalablement afin de s'assurer qu'elle correspondait bien au domaine d'intérêt avant de poursuivre par des tests de repliement.

Table 3 : Conditions de production testées

Les tampons de lyse et les méthodes de lyse sont à reliées aux Tables 4 et 5

| Volume de culutre | Volume pour le test | Température | Ethanol | Temps d'induction | IPTG | Milieu de culture | Contact bille | Tampon de lyse | Méthode de lyse |
|---|---|---|---|---|---|---|---|---|---|
| 20 mL | 1 mL | 23 °C | Oui | 4H | 0,5 mM | LB | Ni-NTA | 1 | 1 |
| 20 mL | 1 mL | 30 °C | Oui | 4H | 0,5 mM | LB | Ni-NTA | 1 | 1 |
| 20 mL | 1 mL | 23 °C | Non | 4H | 0,5 mM | LB | Ni-NTA | 1 | 1 |
| 20 mL | 1 mL | 30 °C | Non | 4H | 0,5 mM | LB | Ni-NTA | 1 | 1 |
| 20 mL | 1 mL | 23 °C | Oui | 4H | 0,5 mM | LB | Ni-NTA | 2 | 1 |
| 20 mL | 1 mL | 30 °C | Oui | 4H | 0,5 mM | LB | Ni-NTA | 2 | 1 |
| 20 mL | 1 mL | 23 °C | Non | 4H | 0,5 mM | LB | Ni-NTA | 2 | 1 |
| 20 mL | 1 mL | 30 °C | Non | 4H | 0,5 mM | LB | Ni-NTA | 2 | 1 |
| 100 mL | 100 mL | 30 °C | Oui | 4H | 0,5 mM | LB | Ni-NTA | 1 | 1 |
| 100 mL | 100 mL | 23 °C | Non | 4H | 0,5 mM | LB | Ni-NTA | 1 | 1 |
| 20 mL | 1 mL | 30 °C | Oui | 4H | 0,5 mM | LB | Ni-NTA | 3 | 1 |
| 20 mL | 1 mL | 30 °C | Non | 4H | 0,5 mM | LB | Ni-NTA | 3 | 1 |
| 20 mL | 1 mL | 30 °C | Oui | 4H | 0,5 mM | LB | ST | 3 | 1 |
| 20 mL | 1 mL | 30 °C | Non | 4H | 0,5 mM | LB | ST | 3 | 1 |
| 20 mL | 1 mL | 30 °C | Oui | 4H | 0,5 mM | LB | ST | 1 | 1 |
| 20 mL | 1 mL | 30 °C | Non | 4H | 0,5 mM | LB | ST | 1 | 1 |
| 20 mL | 1 mL | 23 °C | Oui | 4H | 0,5 mM | LB | Ni-NTA | 4 | 1 |
| 25 mL | 25 mL | 30 °C | Oui | 4H | 0,5 mM | LB | ST | 1 | 1 |
| 20 mL | 1 mL | 30 °C | Oui | 4H | 0,5 mL | LB | Ni-NTA | 2 | 2 |
| 25 mL | 25 mL | 30 °C | Non | 4H | 0,5 mM | LB | ST IJM | 5 | 1 |
| 20 mL | 1 mL | 30 °C | Non | 4H | 0,5 mM | LB | Ni-NTA | 1 | 2 |
| 20 mL | 1 mL | 30 °C | Non | 4H | 0,5 mM | 2YT | Ni-NTA | 1 | 1 |
| 20 mL | 1 mL | 18 °C | Non | ON | 0,5 mM | LB | Non | 1 | 1 |
| 20 mL | 1 mL | 18 °C | Non | ON | 0,1 mM | LB | Ni-NTA | 1 | 2 |

Table 4 : Tampons de lyse utilisés

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 20 mM Tris pH 7,5<br>1% Triton<br>500 mM NaCl<br>20 % glycérol<br>50 mM Imidazole | 50 mM tris ph8<br>200 mM Li2SO4<br>50 mM arginine<br>50 mM glutamate<br>5 % glycérol<br>0 ,2 % CHAPS<br>0,1 mM TCEP | 20 mM tris pH 8,8<br>1% Triton<br>500 mM NaCl<br>20 % glycérol<br>50 mM Imidazole | 1 mM EDTA<br>1 mM DTT<br>20 mM Tris pH 7,5<br>1% triton<br>500 mM NaCl<br>20 % glycérol<br>50 mM Imidazole | 100 mM Tris pH 8<br>150 mM NaCl<br>1 mM EDTA<br>0 ,1 % Tween 20 |

Table 5 : Méthodes de lyse utilisées

| 1 | 2 |
|---|---|
| Lysozyme + 5 cycles de congélations/décongélations | Sonication : 6 sec – 6 sec – 300 W – 3 fois |

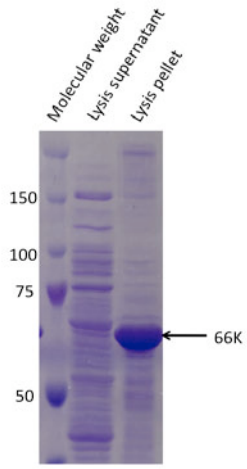



Figure 25 : Analyse de la solubilisation de la 66K sur gel de polyacrylamide

## Repliement de la 66K

De nombreux protocoles sont disponibles pour cette étape et la plupart sont regroupés dans une base de données spécifique nommée refold (http://refold.med.monash.edu.au). En faisant un bilan des techniques mises au point dans cette base de données, nous nous sommes rendu compte que, même pour des protéines homologues, les techniques de repliement référencées pouvaient être très divergentes. Nous ne pouvions donc pas déterminer un protocole de départ par ce biais et nous avons alors testé en parallèle plusieurs façons de replier les protéines. Ainsi, le repliement par dialyse, par dilutions successives ou encore sur colonne ont été les premiers employés. Trois approches globales résultantes peuvent être distinguées. Dans la première, nous purifions préalablement le surnageant de dénaturation sur billes de Nickel et nous avons parfaitement obtenu une protéine d'environ 66 kDa, en quantité importante et pure après dialyse visant à ôter l'urée. Après analyse par spectrométrie de masse de la bande correspondante, nous avons confirmé notre capacité à produire et purifier une chaperonne d'*E. coli* : la DnaK.

La deuxième approche, testée en parallèle, a été le repliement directement sur colonne de Nickel. Au cours du lavage de la résine, après injection du surnageant de dénaturation, nous ôtons progressivement l'urée, sur des temps plus ou moins long. Dans tous les cas, la 66K de TYMV, restait invariablement accrochée à la résine même au cours de l'élution.

Nous nous sommes demandé alors si l'étiquette histidine n'étais pas enfouie dans la protéine lorsque celle-ci était repliée ou si le repliement de la 66K ne nécéssite par des partenaires, et nous nous sommes dirigé vers la troisième approche.

Gardant à l'esprit que les six histidines N-terminales pouvaient être inaccessibles une fois la protéine repliée et sachant également que la résine de streptactine ne tolère pas les doses d'urée présentes au cours de la dénaturation, nous avons finalement tenté de replier la 66K directement et de la purifier dans un second temps. Ainsi, on pouvait également ajouter plus ou moins de facteurs pouvant aider son repliement avant sa purification. Par dilution successive, nous ne sommes pas parvenus à obtenir de la 66K stable en fin de renaturation. En revanche, nous sommes parvenus à trouver des conditions de dialyse permettant la resolubilisation complète de la RdRp et, avant purification, plusieurs milligrammes de protéine ont pu être obtenus à partir d'un culot issu d'un litre de culture. La purification ultérieure sur résine de streptactine n'a pas posé de problèmes particuliers et en fin d'expérience, nous étions en présence de quantités importantes de 66K pure et stable dans le temps.

Par précaution, nous avons comme toujours analysé la bande résultante par spectrométrie de masse et validé le protocole. Afin de s'assurer du bon repliement de la protéine, nous avons finalement réalisé une expérience de dichroïsme circulaire ainsi que des tests d'activité.

### Dichroïsme circulaire et tests d'activité

En sortie de purification, les fractions d'élution ont été concentrées à 6 mg/mL (Figure 26) et des analyses de repliement par dichroïsme circulaire ont été réalisées sur des échantillons plus ou moins dilués à partir de cette solution mère. Les spectres obtenus sont caractéristiques d'une protéine repliée et composée principalement d'hélices α (Figure 27) et nous avons donc transmit nos échantillons à nos collaborateurs afin qu'ils vérifient l'activité de la RdRp. Malheureusement, toutes nos préparations se sont révélées inactives et l'étude des mutants touchant la phosphorylation n'a pu être entreprise. Des essais de cristallisation ont malgré tout été tenté mais aucun signe de croissance cristalline n'a été obtenu dans les conditions testées de façon robotisées ou manuelles. Nous avons donc entreprit des essais de protéolyse ménagée afin de trouver un fragment stable qui engloberait le domaine polymérase C-terminal.

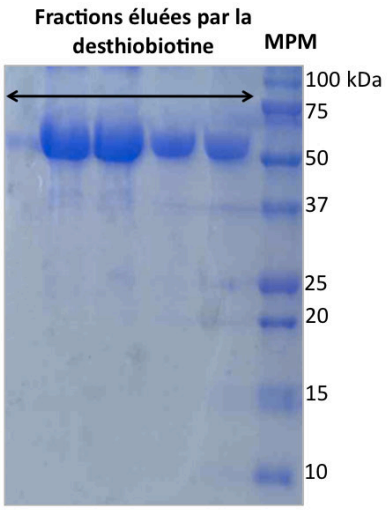



Figure 26 : Analyse sur gel de polyacrylamide 10% des fractions issus de la purification de TYMVpol

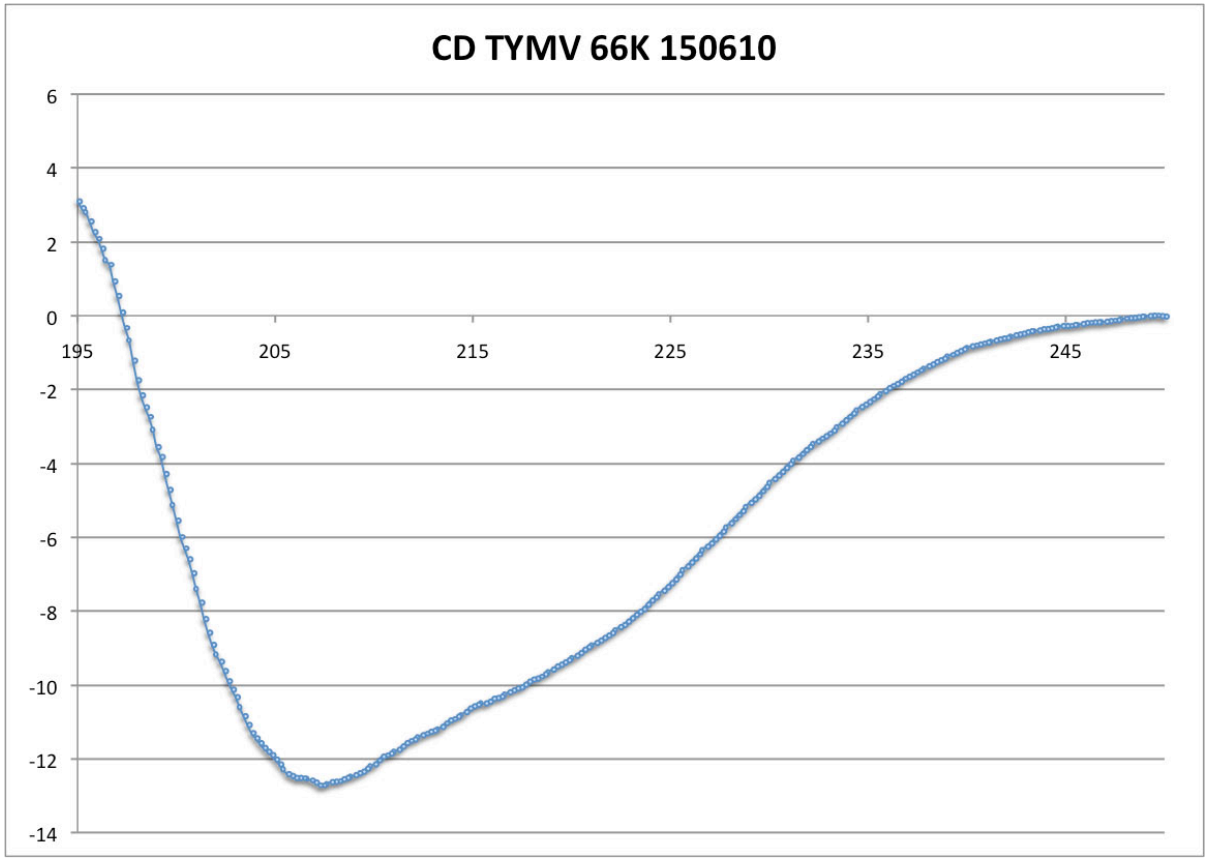



**Figure 27 : spectre de dichroïsme circulaire de la 66K solubilisée et purifiée**
Le spectre obtenu ici est celui résultant de la soustraction du signal correspondant au tampon. On constate qu'il y a deux minimas négatifs aux alentours de 208nm et 222 nm. Ces profil est caractéristique d'une protéine repliée et majoritairement composée d'hélices alpha (Wallace 2001)

## Protéolyse ménagée et identification du domaine polymérase stable

Plusieurs endopeptidases ont été utilisées, dans différentes conditions, afin d'identifier un domaine stable constitué au minimum des motifs conservés A à E de la RdRp. Chaque expérience a abouti à l'identification de fragments sur gel, mais aucun ne semblait correspondre à l'ensemble du domaine polymérase ni ne se retrouvait dans les différentes expériences (Figure 28). Nous avons donc décidé de changer d'approches pour identifier un sous fragment de TYMVpol qui serait exprimable de façon soluble et stable.

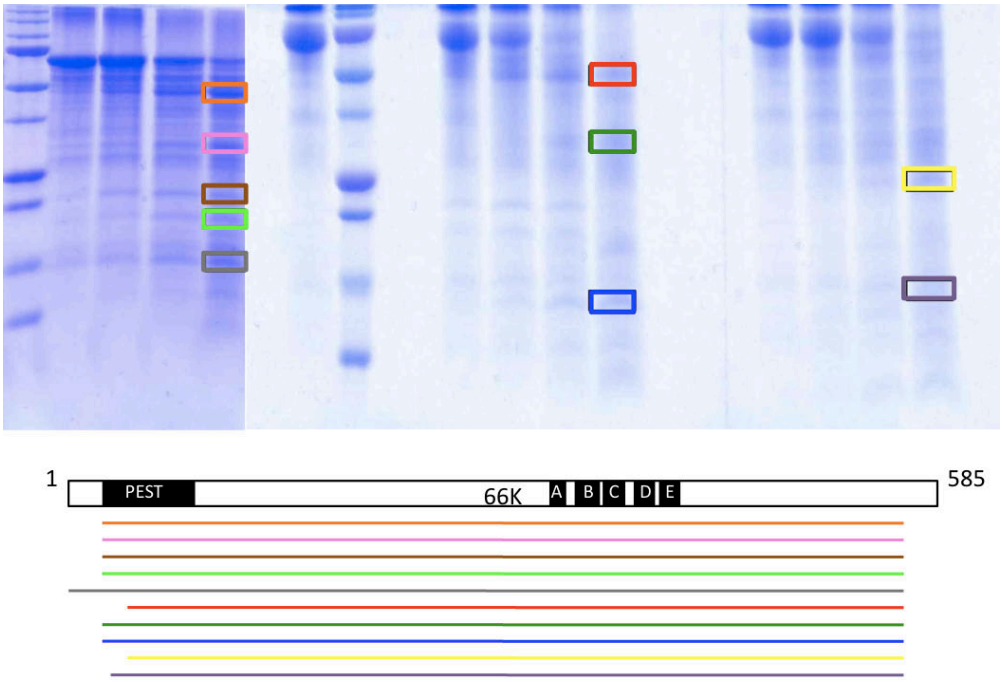



**Figure 28 : Analyse par spectrométrie de masse des produits de protéolyse ménagée**
Le domaine TYMVpol purifié a été digéré par différentes enzymes et le détail est fourni dans la partie « matériel et méthode » à suivre. Ici sont représenté par des lignes colorées, sous le schéma du domaine TYMVpol, la séquence recouverte par les peptides identifiés par spectrométrie de masse. Les bandes d'où sont tirées ces analyses sont colorées de façons identiques aux lignes de couleur.

## Expression d'un fragment soluble

Pour cela, nous nous sommes adressé à l'équipe de clonage et d'expression pESPRIT de la plateforme $P^3$ : Protein Production Plateform (http://www.p-cube.eu). En effet, cette équipe se concentre sur les protéines dont l'expression s'avère difficile (Yumerefendi 2010).

Dans le cas de notre polymérase, plusieurs stratégies vont être employées afin d'obtenir une construction exprimable de façon soluble et englobant au minimum le domaine RdRp.

Tout d'abord, en partant de notre gène codant pour les 585 derniers résidus de la polyprotéine, nous allons réaliser des sous clonages, de façon robotisée, et tester leur expression. Ces clones vont être obtenus grâce à l'action de l'exonucléase III, qui va dégrader de façon aléatoire la partie 5' du gène et fournir alors une protéine plus ou moins délétée en N-terminal de la séquence PEST. Les fragments suffisamment longs pour contenir le domaine RdRp, quelques milliers, vont alors être insérés dans un plasmide d'expression optimisé pour les tests ultérieurs. En effet, en amont du site d'insertion, une étiquette histidine est présente afin de faciliter la purification et, en aval, un peptide accepteur de biotine est présent afin de réaliser un test de biotinylation permettant de vérifier l'état de solubilité de la protéine. Les protéines exprimées dans la fraction soluble vont en effet être biotinylée à un niveau significativement supérieur à celles exprimées dans la fraction insoluble et vont pouvoir être détectées, en colonie, grâce à la streptavidine flourescente se liant à la biotine. Celles qui posséderont également une étiquette histidine seront alors doublement détectées grâce à un anticorps fluorescent anti-histidine. Ces protéines seront ensuite testées en milieu liquide où, là encore, par robotisation assistée, l'expression et la purification sur billes de nickel seront vérifiées. Les fragments finaux seront finalement utilisés pour tenter d'obtenir une production à plus grande échelle suivie d'une purification en vue de réaliser des essais de cristallisation.

D'autres approches vont également être réalisées par la plateforme pESPRIT telles que la co-expression avec le domaine protéase muté au niveau de sa cystéine catalytique afin de le rendre inactif.

# Matériel et méthodes

## Essais sur la production insoluble

*Expression de la 66K*

Le gène codant pour la polymérase 66K a été inséré dans le plasmide pET21 permettant d'étiqueter 66K avec 6 histidines au niveau de son extrémité N-terminal et avec un strep-tag en C-terminal. Le plasmide est ensuite transformé dans *Escherichia coli* BL21 (DE3) et une pré-culture est réalisée sur la nuit à 37°C. Cette pré-culture est alors utilisée pour inoculer 1L de milieu Luria broth (LB) contenant 50 µg/L carbenicillin. La culture est mise à 37°C et lorsque la densité optique mesurée à la longueur d'onde de 600 nm atteint la valeur de 0.6, l'expression est induite par addition de 0.5 mM isopropyl-b-D-thiogalactopyranoside (IPTG). Les cellules sont ensuite remisent en culture pendant 4 heures à 30°C et à l'issu de cette incubation, le culot bactérien est collecté par centrifugation et stocké à -20°C.

*Solubilistation de la 66K*

Approximativement 5 g de culot sont resuspendus dans 20 mL de tampon de lyse (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 1 % Triton X-100, 20 % glycérol, 2 mg/ml Lysozyme, 1 pastille d'antiprotéase complète mini EDTA free (Roche) et 1 U/ml Benzonase) et incubés 60 min dans la glace. La lyse est complétée par 5 cycles de congélation/décongélation dans l'azote liquide (70 K/303 K).
Le lysat est alors centrifugé à 8000 g durant 30 min et le culot est remit en suspension dans un tampon dénaturant (50 mM Tris-HCl pH 8.0, 8 M Urée et 2 mM DTT) durant 1 heure à température ambiante. Le surnageant de solubilisation est récupéré après centrifugation à 8000 xg pendant 30 min et injecté dans une cassette de dialyse PIERCE de 30mL et de cut-off 20 kDa. La cassette et plongée dans 5 L de tampon de renaturation (100 mM tris pH 8 et 2 mM DTT) à 4 °C sur la nuit. Un second bain identique de renaturation est réalisé sur 4 h à la suite du premier. Le contenu de la dialyse est finalement collecté et centrifugé à 8000 xg pendant 20 min et le surnageant est stocké à 4 °C.

*Purification de la 66K*

Le surnageant est alors injecté sur une colonne contenant 2 mL de résine streptactin superflow (Qiagen). La colonne est initialement pré-equilibrée avec la solution du deuxième bain de renaturation. Elle est ensuite lavée avec 50 mL de ce même tampon et la 66K est ensuite éluée grâce au tampon d'élution (100 mM tris pH 8, 2 mM DTT et 2.5 mM de desthiobiotine). La pureté est évaluée sur gel tris-tricine 12 % puis, les fractions satisfaisantes sont congelées dans l'azote liquide et stockées à -80 °C.

*Protéolyse ménagée*

Quatre protéases ont été testées : la chymotrypsine, la trypsine, la papaïne et la thermolysine. Dans un volume total de 10 µL, un ratio de 1/100 de 66K par rapport à chacune des protéases a été mit à incuber pendant 5, 15, 30 ou 75 minutes à température ambiante. En fin d'incubation, les protéases sont inactivées avec 1 µL d'une solution d'antiprotéase réalisée par dilution d'une pastille de complete mini EDTA free (Roche) dans 2 mL d'eau distillée. Puis 4 µL de bleu de dépôt 2X pour gel de polyacrylamide sont ajoutés et le tout est chauffé 5 minutes à 96 °C avant d'être intégralement déposé sur gel.

## Production soluble de 66K : pESPRIT

*Préparation de constructions utilisées par la plateforme*

Trois constructions différentes de la 66K non étiquettée vont être réalisées. Elles vont différer au niveau de leur terminaison 3' et seront distinguées par les appellations AITS, IEVS ou full length. Sur la séquence du gène complet de la 66K, ces différences de constructions 3' sont représenté de façon colorée comme leurs appellations correspondantes.

Séquence du gène de la 66K clonée

acccctccgcgagtccgacccatcggagccctaactttcacctgccgccacacattccgctatcgtacgaccgtgactttgtaactgtcaaccccactttacctgat
aggggccggagacgagattggatacccatttttaccgcctagccgtctgcccttacattttgatttacctccggccatcactcctccgccagtttctacatctgttgac
cgcctcaggctaaagcatcgccggtctatcctggtgaatttttcgatagcctggcggccttttttcttgcctgctcatgatccatcaactcgcgagatcctgcataaagat
agtcaagcaaccagtttccttggttcgaccgcccgttttccctgagttgtcaaccaagttctctgattagtgccaaacacgcaccgaaccatgaccccacgctgctac
ggctagtatcaataaacgcttgagatttcgaccttcggacagtccgcatcagattaccgcggatgacgttgtcctcggtttgcaactctttcactctctctgtcgcgcct
ttcccgtcaaccaaactcgaccgtgccattcaatcctgagctgtttgcagaatgtatctccctgaatgaatatgctcagctttcatccaaaacccagtccacgattgtag
aaacgcgtctcgcagcgatccggattggcgacacaccactgtcaagatctttgcaaaagcccagcataaagtgaatgatggttcgatatttgggtcatggaaagcg
gtcagacgctggccctgatgcatgattatgtgatcctggttctgggtccagtaaagaaatatcagcgtatattcgataacgcggatcggccacccaatatatatagtc
ctgcggcaaaacgccaaaccaattacgggattggtgccaagagcatctgacacattcgacgccaaaaattgctaatgattacaccgcgttcgaccagtcacaaca
ggcgaatctgttgtgcttgaggccttgaaaatgaagcgtctgaatattccgtctcatctcatacagctccatgtacacctgaagactaatgtttccacccaattcggcc
cttaacttgtatgcgccttaccggcgaaccgggcacgtatgatgacaacaccgattataatctggcggtcatttacagccagtacgacgtgggctcatgcccgatca
ggtttcaggagacgatagtctcatcgatcatccgctgcccacacgtcatgattggccatcggtgttgaagcggctgcacctgcgttttaaattggaactgacaagtca
ccgctgttctgcggatactatgtaggtccggctgggtgcattcgcaatccgctggcattattctgcaaactgatgatcgccgttgacgatgacgcgctcgatgaccg
agactgagctatttgacagaatttaccacaggtcaccttctgggcgagagcttatggcacctgcttcccgaaacccatgtgcaataccagagcgcttgtttcgacttt
ctgccgtaggtgtccgcgtcacgaaaagatgcttctggatgattctacgccggcactgtcccttctggagcgtattacctcatctccgcgctggcttacgaaaaatgc
atgtacttactaccggcgaagttgaggcttgcgattacctcttaagccagacacagagctttccagaatcaatcgaagtgagccatgccgaaagcgagctccttca
tacgtgcaa

Les trois gènes plus ou moins délétés en 3' ont été amplifié par PCR et inséré dans le vecteur pESPRIT002 (Figure 29). Ce vecteur contient un gène de résistance à la kanamycine, le gène du peptide accepteur de biotine et de l'étiquette histidine nécessaire à la technique de purification et de détection présenté auparavant.

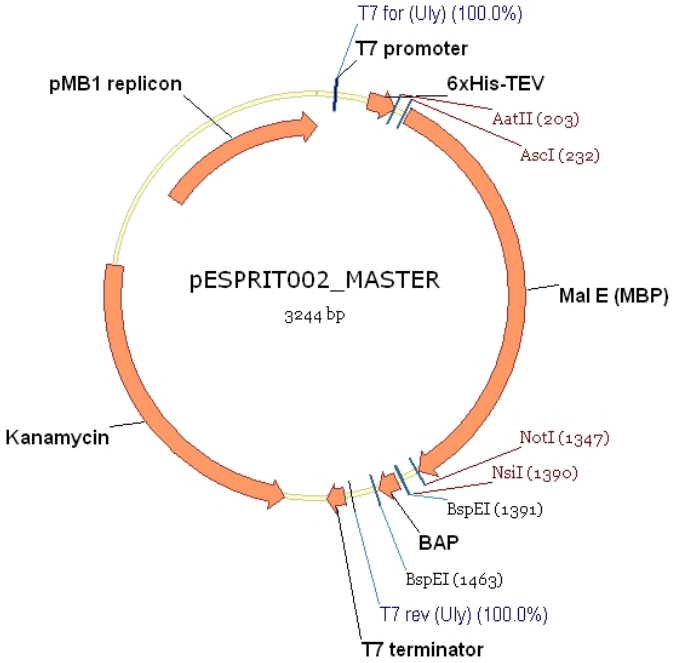



Figure 29 : Carte du plasmide pESPRIT002

La 66K, une fois clonée, sera par la suite co-exprimée avec le domaine TYMVpro. Pour cela, ce domaine a été inséré dans un nouveau plasmide pACYC (Figure 30).

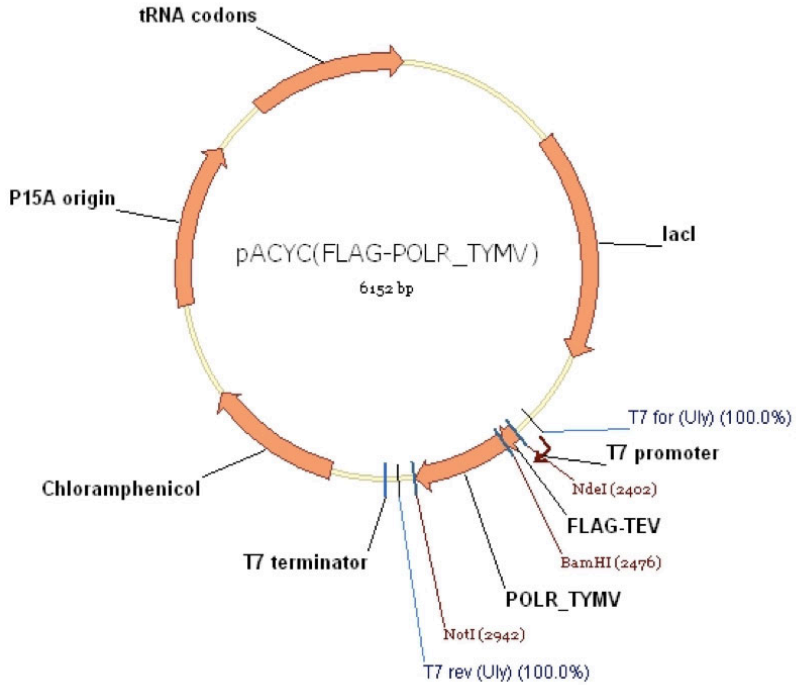



Figure 30 : Carte du plasmide pACYC pour l'expression du domaine TYMVpro

# Discussion

Les difficultés que nous avons rencontrées durant les tentatives de production de la polymérase du TYMV, s'expliquent par la biologie liée au virus. En effet, un des phénomènes de régulation que nous essayons de mettre en évidence, au niveau structural, est associé à la stabilisation et à la dégradation de la RdRp. Vouloir exprimer un fragment portant une région flexible constituant un signal de dégradation n'est pas une base favorable pour l'obtention d'une production stable.

Les séquences PEST sont principalement présentes chez les protéines dont le temps de demi vie est relativement court, et servent à réguler rapidement l'activité de certaines enzymes en constituant un signal de dégradation. De plus, il apparaît que ce sont des boucles ou des extensions flexibles, exposées au solvant et de nature hydrophobe (Rechsteiner 1996). Or, de nombreuses expériences et revues exposent le fait que de telles régions hydrophobes et désordonnées sont néfastes pour l'expression des protéines (Esnouf 2006, Dyson 2010).

Nous avons naturellement tenté de caractériser la séquence désordonnée en N-terminal de notre domaine TYMVpro afin de la déléter. Différents algorythmes de prédiction ont été utilisés mais aucun n'a abouti à une solution claire. Si l'on s'intéresse par exemple à la sortie de l'interface Globplot (http://globplot.embl.de), on observe que la région désordonnée s'étend au delà du 150$^{ième}$ résidu pour la His-66K-ST (Figure 31). Après inspection du gène en question, nous n'avons cependant pas observé de site clair nous permettant de délimiter la partie désordonnée en vu de sa délétion pour ne garder que le domaine polymérase. De plus, des alignemements de séquence réalisés sur d'autres RdRp de virus à ARN(+), ne nous ont pas permis non plus de délimiter précisément ce domaine fonctionnel au sein de la 66K du TYMV.

Nos collaborateurs ayant auparavant testé de nombreuses possibilités d'expression pour cette polymérase, nous avons finalement fait appel à la plateforme ESPRIT. Cette plateforme qui réalise des sous fragments aléatoires de la protéine d'intérêt, finit par trouver, parmi les milliers générées, une construction soluble. Par ce moyen, l'expression et la cristallisation d'un domaine d'une polymérase virale a par exemple été rendu possible (Tarendeau 2008 et 2007, Guilligay 2008).

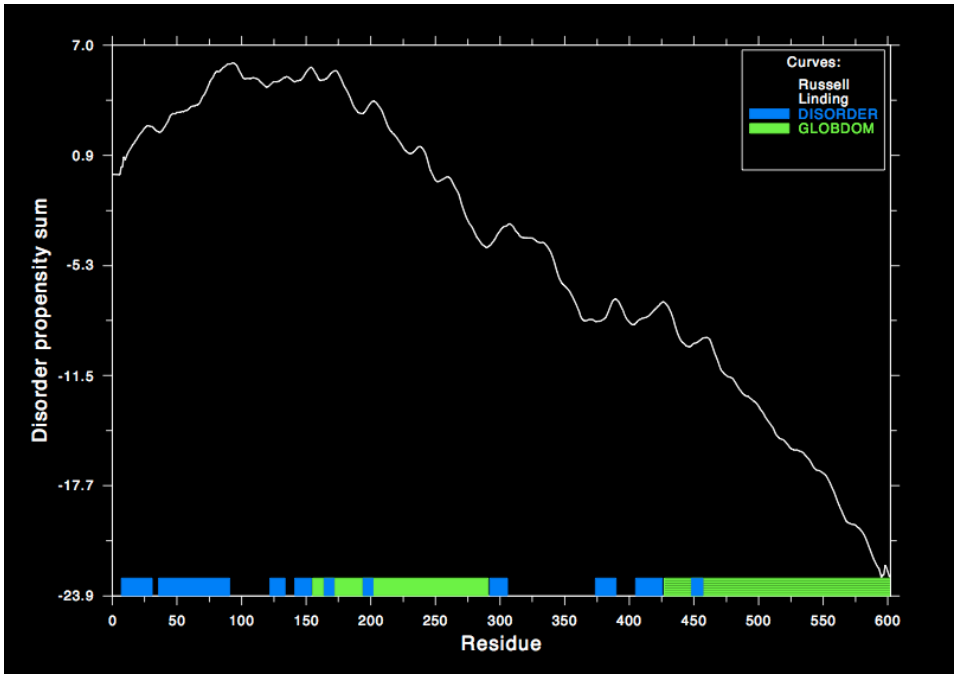



**Figure 31 : Prédiction de désordre à partir de la séquence de la 66K**
Les régions ascendantes de la courbe correspondent aux régions prédites comme étant désordonnées à l'inverse des régions descendantes. Ces régions sont alors représentées en bleu pour les désordonnées et en vert pour les ordonnées sur l'axe des résidus en abscisse.

L'approche envisagée pour le moment est un peu différente de celle utilisée dans l'exemple cité précédemment. Dans notre cas, nous connaissons en effet un partenanaire de la polymérase du TYMV permettant sa stabilisation : la protéase. De plus, nous connaissons une construction soluble de ce partenaire, lors de sa production en bactérie. Des tentatives de co-expression vont ainsi directement être réalisées lors des essais d'obtention de sous fragments solubles du domaine polymérase. La méthode générale « CoESPRIT » est décrite plus en détail dans la publication de An en 2011.

Le projet venant de commencer, nous espérons à présent obtenir rapidement une voire plusieurs constructions nous permettant d'étudier enfin par cristallographie ce domaine polymérase et le mécanisme de phosphorylation associée à la sérine 326 de son motif A.

# Perspectives et discussion

## Découvertes récentes

Actuellement, nous savons que l'ARN polymérase ARN dépendante du virus de la mosaïque jaune du navet est polyubiquitinylée au cours du cycle viral. Cette poly-ubiquitynation est réalisée par la cellule hôte en vue de l'adressage au protéasome de cette polymérase où elle sera dégradée. On sait aussi que la protéase du TYMV est capable de la maturer, puis de la protéger de la phosphorylation constituant le signal pour l'ubiquitination. Cette protection se fait via une interaction directe entre ces deux protéines essentielles à la réplication du génome viral. On sait également que cette protéase a une activité ubiquitine hydrolase et qu'elle peut ainsi cliver l'étiquette poly-ubiquitine accrochée à la polymérase. Ce mécanisme a été parfaitement démontré, très récemment, au niveau cellulaire par nos collaborateurs (Chenon 2011). Nous avons donc cherché à relier ce mécanisme de régulation du cycle viral à la structure de ces deux partenaires clés.

En résolvant tout d'abord la structure tridimensionnelle de la protéase, nous avons pu avoir une vue structurale de l'activité protéolytique nécessaire à la maturation des protéines de réplication. L'étude de cette dernière nous a permis de mieux comprendre comment se déroulait cette maturation au cours de la réplication virale. En effet, les sites de reconnaissance de la protéase pour ses substrats ont pu être caractérisés ainsi que les sites P caractéristiques des protéases. De plus, nous avons levé l'ambiguité qui régnait sur l'assignation de la protéase à un clan en montrant qu'elle était homologue aux déubiquitinases de type OTU par comparaison structurale. Nous avons donc caractérisé le commencement de ce mécanisme de régulation au niveau viral et souhaitons à présent étudier les étapes plus tardives et les interactions impliquées entre la cellule hôte et le virus.

## Etude structurale de la fonction DUB

Afin d'étudier cette fois les caractéristiques structurales liées à la fonction ubiquitine hydrolase, nous essayons maintenant d'obtenir la structure d'un complexe avec une molécule d'ubiquitine suicide. A partir de la protéase produite et purifiée, nous avons mis au point la formation d'un tel complexe ainsi que sa purification. Des essais de cristallisation doivent prochainement être initiés.

Nous espérons observer d'éventuelles différences de conformation ou de contacts impliquées dans la reconnaissance avec le substrat en comparant ces deux activités : endopeptidase et déubiquitinase. On suspecte en effet des changements de conformation et plus particulièrement au niveau du site actif de cette enzyme suite à la publication de données structurales de DUB seules et en complexe avec une ubiquitine (Kommander 2009, Sulea 2006, Sato 2008). La comparaison de ces différentes structures montre en effet que des réarrangements ont lieu au niveau du site actif essentiellement, lorsqu'une molécule d'ubiquitine s'y fixe. En revanche, la comparaison de ces complexes avec le domaine TYMVpro que nous avons résolu, ne nous permet pas de spéculer sur son cas et l'obtention de la structure du complexe TYMVpro/ubiquitine est donc essentielle.

Dans un second temps, nous aimerions poursuivre cette caractérisation structurale dans un contexte plus proche de celui du cycle viral en obtenant cette fois un complexe entre la déubiquitinase du TYMV et de courtes chaines de deux à quatre molécules d'ubiquitine. En effet, la protéase reconnaît l'étiquette polyubiquitine associé à la 66K qui est constituée d'au moins quatre ubiquitines. Des tests in vitro, réalisés par nos collaborateurs, ont déjà permis de montrer l'affinité de la DUB du TYMV pour de tels substrats. Pour obtenir ce complexe, nous utiliserons bien entendu un mutant inactif de la protéase.

Finalement, un complexe ternaire entre TYMVpro, l'ubiquitine et un peptide dérivé de la polymérase du TYMV sera étudié dans le but d'observer le substrat naturel de cette DUB et le mécanisme plus précis de déubiquitination engagée lors du cycle viral. Cette étude est d'autant plus intéressante que c'est la première DUB dont le substrat naturel soit identifié.

Pour que cette étude cristallographique soit complète, nous essayons également de résoudre la structure de ce substrat naturel qu'est la polymérase.

## Résolution de la structure de la 66K

Actuellement, la plateforme ESPRIT travaille sur l'expression d'un fragment soluble du domaine polymérase. L'obtention de celui-ci pourra alors nous permettre de produire la polymérase du TYMV, la purifier et réaliser des tests de cristallisation.

L'obtention de la structure tridimensionnelle nous intéresse avant tout pour l'étude de la phosphorylation constitutive impliquée dans la régulation de la réplication. Cette polymérase est d'ailleurs la première dont les sites de phosphorylation ont été précisément identifiés et caractérisés.
Ainsi, lorsque sa structure native sera obtenue, nous pourrons poursuivre par la cristallisation d'un mutant mimant la phosphorylation. Ce mutant sera généré par mutation du site phosphorylé, en acide aspartique puisque la charge de ce résidu permet de mimer l'état de phosphorylation. Par comparaison structurale de ces deux structures de polymérase, nous pourrons ainsi observer d'éventuels changements de conformation et caractériser plus en profondeur le mécanisme de régulation associé.
Cette polymérase étant également le substrat de la protéase à deux reprises au cours du cycle viral, la modélisation de l'interaction entre ces deux partenaires peut également nous permettre de mieux comprendre les découvertes fonctionnelles exposées précédemment. Aucun homologue de structure résolue n'est pour le moment connu. Son étude par cristallographie est donc là encore essentielle.

## Rôle de l'ubiquitination réversible du TYMV

Le rôle de l'ubiquitination réversible de la polymérase et l'action de protection de la protéase au cours du cycle viral, ne sont pas encore clairs. On suppose tout d'abord que l'ubiquitination est un moyen de défense de la cellule hôte. En dégradant la RdRp, enzyme clé du cycle viral, par la voie ubiquitine/protéasome, la cellule peut inhiber la multiplication du virus et limiter ainsi son infection. D'un autre côté, le fait que cette régulation intervienne à un stade tardif de l'infection, peut signifier que c'est une stratégie plutôt engagée par le virus pour réguler la quantité de polymérase présente aux différentes étapes de la réplication.

## Caractérisation de ce mécanisme de régulation au niveau cellulaire

En parallèle des études cristallographiques, il serait intéressant de caractériser plus en profondeur cette régulation par ubiquitination réversible. Cela pourrait notamment nous permettre de déterminer son rôle exact au cours de la réplication et comprendre comment et pourquoi le cycle viral est ainsi affecté. D'autres cycles de réplication de virus à ARN(+) sont connus comme étant régulés également par des phénomènes liés à l'ubiquitination (Banks 2003, Shackelford 2005).

Ainsi, toujours dans l'optique de faire de TYMV un virus modèle, nous pourrions utiliser notre caractérisation globale de cette régulation à d'autres virus. Concernant les étapes initiales, il serait donc intéressant d'identifier les kinases à l'origine de la phosphorylation de la 66K et responsables du signal d'ubiquitination ainsi que les ubiquitines ligases de la cellule hôtes impliquées. Toujours dans le cadre des interconnexions hôte/pathogène, la recherche de substrats cellulaires de la protéase du TYMV pourrait permettre de trouver d'autres rôles de cette enzyme au cours du cycle infectieux. Enfin, des expériences in vivo sur tous les acteurs déterminés ainsi pourraient nous fournir la vue globale et précise souhaitée.

# Bibliographie

# Bibliographie


**An Y**, Meresse P, Mas PJ, Hart DJ. *CoESPRIT: a library-based construct screening method for identification and expression of soluble protein complexes*. PLoS ONE (**2011**) vol. 6 (2) pp. e16261

**Azzo A**, Bongiovanni A, Nastasi T. *E3 ubiquitin ligases as regulators of membrane protein trafficking and degradation*. Traffic (**2005**) vol. 6 (6) pp. 429-41

**Banks** L, Pim D, Thomas M. *Viruses and the 26S proteasome: hacking into destruction*. Trends Biochem Sci (**2003**) vol. 28 (8) pp. 452-9

**Barretto** N, Jukneliene D, Ratia K, Chen Z, Mesecar AD, Baker SC. *Deubiquitinating activity of the SARS-CoV papain-like protease*. Adv Exp Med Biol (**2006**) vol. 581 pp. 37-41

**Borodovsky** A, Ovaa H, Kolli N, Gan-Erdene T, Wilkinson KD, Ploegh HL, Kessler BM. *Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family*. Chem Biol (**2002**) vol. 9 (10) pp. 1149-59

**Böttcher** S, Maresch C, Granzow H, Klupp BG, Teifke JP, Mettenleiter TC. *Mutagenesis of the active-site cysteine in the ubiquitin-specific protease contained in large tegument protein pUL36 of pseudorabies virus impairs viral replication in vitro and neuroinvasion in vivo*. Journal of Virology (**2008**) vol. 82 (12) pp. 6009-16

**Boudreaux** DA, Maiti TK, Davies CW, Das C. *Ubiquitin vinyl methyl ester binding orients the misaligned active site of the ubiquitin hydrolase UCHL1 into productive conformation*. Proc Natl Acad Sci USA (**2010**) vol. 107 (20) pp. 9117-22

**Bozarth** CS, Weiland JJ, Dreher TW. *Expression of ORF-69 of turnip yellow mosaic virus is necessary for viral spread in plants*. Virology (**1992**) vol. 187 (1) pp. 124-30

**Bransom** KL, Wallace SE, Dreher TW. *Identification of the cleavage site recognized by the turnip yellow mosaic virus protease*. Virology (**1996**) vol. 217 (1) pp. 404-6

**Bransom** KL, Weiland JJ, Tsai CH, Dreher TW. *Coding density of the turnip yellow mosaic virus genome: roles of the overlapping coat protein and p206-readthrough coding regions*. Virology (**1995**) vol. 206 (1) pp. 403-12

**Bransom** KL, Weiland JJ, Dreher TW. *Proteolytic maturation of the 206-kDa nonstructural protein encoded by turnip yellow mosaic virus RNA*. Virology (**1991**) vol. 184 (1) pp. 351-8

**Burmeister**. *Structural changes in a cryo-cooled protein crystal owing to radiation damage*. Acta Crystallographica Section D (**2000**) D56, 328-341

**Camborde** L, Planchais S, Tournier V, Jakubiec A, Drugeon G, Lacassagne E, Pflieger S, Chenon M, Jupin I. *The Ubiquitin-Proteasome System Regulates the Accumulation of Turnip yellow mosaic virus RNA-Dependent RNA Polymerase during Viral Infection*. The Plant cell (**2010**)

**Camborde** L, Tournier V, Noizet M, Jupin I. *A Turnip yellow mosaic virus infection system in Arabidopsis suspension cell culture*. FEBS Letters (**2007**) vol. 581 (2) pp. 337-41

**Canady** MA, Larson SB, Day J, McPherson A. *Crystal structure of turnip yellow mosaic virus*. Nature Structural Biology (**1996**) vol. 3 pp. 771-10

**Capodagli** GC, McKercher MA, Baker EA, Masters EM, Brunzelle JS, Pegan SD. *Structural analysis of a viral ovarian tumor domain protease from the Crimean-Congo hemorrhagic fever virus in complex with covalently bonded ubiquitin*. Journal of Virology (**2011**) vol. 85 (7) pp. 3621-30

**Chenon** M, Camborde L, Jupin I. *A viral deubiquitylating enzyme stabilizes viral RNA-dependent RNA polymerase and influences viral infectivity*. (**2011**) EMBO

**Clague** et Urbé. *Ubiquitin: same molecule, different degradation pathways*. Cell (**2010**) vol. 143 (5) pp. 682-5

**Dougherty** et Semler. *Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes*. Microbiological Reviews (**1993**) vol. 57 (4) pp. 781-822

**Dubochet** J, Adrian M, Chang JJ, Homo JC, Lepault J, Alasdair W. McDowall AW and Schultz P. *Cryo-electron microscopy of vitrified specimens*. Quaterly reviews of biophysics (**1988**) vol. 21 (2) pp. 129-101

**Dyson**. *Selection of soluble protein expression constructs: the experimental determination of protein domain boundaries*. Biochem Soc Trans (2010) vol. 38 (4) pp. 908-13

**Eden** J, Sharpe L, White P, Brown A. *Norovirus RNA-dependent RNA-polymerase is phosphorylated by an important survival kinase, Akt*. Journal of Virology (**2011**) pp. 1-18

**Esnouf** RM, Hamer R, Sussman JL, Silman I, Trudgian D, Yang ZR, Prilusky J. *Honing the in silico toolkit for detecting protein disorder*. Acta Crystallogr D Biol Crystallogr (**2006**) vol. 62 (Pt 10) pp. 1260-6

**Evans** PC, Ovaa H, Hamon M, Kilshaw PJ, Hamm S, Bauer S, Ploegh HL, Smith TS. *Zinc-finger protein A20, a regulator of inflammation and cell survival, has de-ubiquitinating activity*. Biochem J (**2004**) vol. 378 (Pt 3) pp. 727-34

**Ghosh** AK, Takayama J, Rao KV, Ratia K, Chaudhuri R, Mulhearn DC, Lee H, Nichols DB, Baliji S, Baker SC, Johnson ME, Mesecar AD. *Severe acute respiratory syndrome coronavirus papain-like novel protease inhibitors: design, synthesis, protein-ligand X-ray structure and biological evaluation*. J Med Chem (**2010**) vol. 53 (13) pp. 4968-79

**Glenn** et Searles. *Distinct domains mediate the early and late functions of the Drosophila ovarian tumor proteins*. Mech Dev (**2001**) vol. 102 (1-2) pp. 181-91

**Gorbalenya** et Koonin. *Viral proteins containing the purine NTP-binding sequence pattern*. Nucleic Acids Research (**1989**) vol. 17 (21) pp. 8413-40

**Guilligay** D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, Sehr P, Lewis J, Ruigrok R, Ortin J, Hart DJ, Cusack S. *The structural basis for cap binding by influenza virus polymerase subunit PB2*. Nat Struct Mol Biol (**2008**) vol. 15 (5) pp. 500-6

**Haglund** et Dikic. *Ubiquitylation and cell signaling*. EMBO J (**2005**) vol. 24 (19) pp. 3353-9

**Harcourt** BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, Smith CM, Rota PA, Baker SC. *Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity*. Journal of Virology (**2004**) vol. 78 (24) pp. 13600-12

**Harker**. *The Application of the Three-Dimensional Patterson Method and the Crystal Structures of Proustite, AgAsS, and Pyrargyrite, AgSbS*. The Journal of Chemical Physics (**1936**)

**Hemelaar** J, Borodovsky A, Kessler BM, Reverter D, Cook J, Kolli N, Gan-Erdene T, Wilkinson KD, Gill G, Lima CD, Ploegh HL, Ovaa H. *Specific and covalent targeting of conjugating and deconjugating enzymes of ubiquitin-like proteins*. Mol Cell Biol (**2004**) vol. 24 (1) pp. 84-95

**Hendrickson**. *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation*. Science (**1991**)

**Jakubiec** A, Drugeon G, Camborde L, Jupin I. *Proteolytic processing of turnip yellow mosaic virus replication proteins and functional impact on infectivity*. Journal of Virology (**2007**) vol. 81 (20) pp. 11402-12

**Jakubiec** A, Tournier V, Drugeon G, Pflieger S, Camborde L, Vinh J, Héricourt F, Redeker V, Jupin I. *Phosphorylation of viral RNA-dependent RNA polymerase and its role in replication of a plus-strand RNA*

*virus*. J Biol Chem (**2006**) vol. 281 (30) pp. 21236-49

**James** TW, Frias-Staheli N, Bacik JP, Levingston Macleod JM, Khajehpour M, García-Sastre A, Mark BL. *Structural basis for the removal of ubiquitin and interferon-stimulated gene 15 by a viral ovarian tumor domain-containing protease*. Proc Natl Acad Sci USA (**2011**) vol. 108 (6) pp. 2222-7

**Jarosinski** K, Kattenhorn L, Kaufer B, Ploegh H, Osterrieder N. *A herpesvirus ubiquitin-specific protease is critical for efficient T cell lymphoma formation*. Proc Natl Acad Sci USA (**2007**) vol. 104 (50) pp. 20025-30

**Jiang** et Tang. *Mechanism of inhibiting type I interferon induction by hepatitis B virus X protein*. Protein Cell (**2010**) vol. 1 (12) pp. 1106-17

**Kadaré** G, David C, Haenni AL. *ATPase, GTPase, and RNA binding activities associated with the 206-kilodalton protein of turnip yellow mosaic virus*. Journal of Virology (**1996**) vol. 70 (11) pp. 8169-74

**Kapoor** M, Zhang L, Ramachandra M, Kusukawa J, Ebner KE, Padmanabhan R. *Association between NS3 and NS5 proteins of dengue virus type 2 in the putative RNA replicase is linked to differential phosphorylation of NS5*. J Biol Chem (**1995**) vol. 270 (32) pp. 19100-6

**Karpe** et Lole. *Deubiquitination activity associated with hepatitis E virus putative papain like cysteine protease*. J Gen Virol (**2011**) pp.

**Kattenhorn** LM, Korbel GA, Kessler BM, Spooner E, Ploegh HL. *A deubiquitinating enzyme encoded by HSV-1 belongs to a family of cysteine proteases that is conserved across the family Herpesviridae*. Mol Cell (**2005**) vol. 19 (4) pp. 547-57

**Kim** SH, Palukaitis P, Park YI. *Phosphorylation of cucumber mosaic virus RNA polymerase 2a protein inhibits formation of replicase complex*. The EMBO journal (**2002**) vol. 21 (9) pp. 2292-300

**Komander** D, Clague M, Urbé S. *Breaking the chains: structure and function of the deubiquitinases*. Nat Rev Mol Cell Biol (**2009**) vol. 10 (8) pp. 550-63

**Koonin**. *The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses*. J Gen Virol (**1991**) vol. 72 ( Pt 9) pp. 2197-206

**Liao** T, Wu C, Su W, Jeng K, Lai MMC. *Ubiquitination and deubiquitination of NP protein regulates influenza A virus RNA replication*. EMBO J (**2010**) vol. 29 (22) pp. 3879-90

**Lindner** HA, Fotouhi-Ardakani N, Lytvyn V, Lachance P, Sulea T, Ménard R. *The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme*. Journal of Virology (**2005**) vol. 79 (24) pp. 15199-208

**Lodish**. *Molecular cell biology*. (**2008**) pp. 1150

**Love** KR, Catic A, Schlieker C, Ploegh HL. *Mechanisms, biology and inhibitors of deubiquitinating enzymes*. Nat Chem Biol (**2007**) vol. 3 (11) pp. 697-705

**Lwoff A**. *From protozoa to bacteria and viruses. Fifty years with microbes (André Lwoff)*. Annu. Rev. Microbiol. (**1971**) vol. 25 pp. 1-26

**Makarova** KS, Aravind L, Koonin EV. *A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and Chlamydia pneumoniae*. Trends Biochem Sci (**2000**) vol. 25 (2) pp. 50-2

**Mayer E**. *New method for vitrifying water and other liquids by rapid cooling of their aerosols*. Journal of applied physics (**1985**)

**McPherson A**. Increasing the size of microcrystals by fine sampling of pH limits. Journal of applied crystallography (**1995**)

**Morch** MD, Boyer JC, Haenni AL. *Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA*. Nucleic Acids Research (**1988**) vol. 16 (13) pp. 6157-73

**Navaza J**. *AMoRe: an automated package for molecular replacement*. Acta Crystallographica Section A : Foundations of crystallography (**1994**) vol.50 pp. 157-6

**Passmore LA** et Barford D. *Getting into position: the catalytic mechanisms of protein ubiquitylation*. Biochem J (**2004**) vol. 379 (Pt 3) pp. 513-25

**Patterson AL**. *A Fourier series method for the determination of the components of interatomic distances in crystals*. Physical Review (**1934**)

**Pickart CM** et Eddins MJ. *Ubiquitin: structures, functions, mechanisms*. Biochim Biophys Acta (**2004**) vol. 1695 (1-3) pp. 55-72

**Poch** O, Sauvaget I, Delarue M, Tordo N. *Identification of four conserved motifs among the RNA-dependent polymerase encoding elements*. The EMBO journal (**1989**) vol. 8 (12) pp. 3867-74

**Prod'homme** D, Jakubiec A, Tournier V, Drugeon G, Jupin I. *Targeting of the turnip yellow mosaic virus 66K replication protein to the chloroplast envelope is mediated by the 140K protein*. Journal of Virology (**2003**) vol. 77 (17) pp. 9124-35

**Prod'homme** D, Le Panse S, Drugeon G, Jupin I. *Detection and subcellular localization of the turnip yellow mosaic virus 66K replication protein in infected cells*. Virology (**2001**) vol. 281 (1) pp. 88-101

**Ravelli RB** et McSweeney SM. *The fingerprint'that X-rays can leave on structures*. Structure (**2000**)

**Rechsteiner M** et Rogers SW. *PEST sequences and regulation by proteolysis*. Trends Biochem Sci (**1996**) vol. 21 (7) pp. 267-71

**Reyes-Turcu** FE, Ventii KH, Wilkinson KD. *Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes*. Annu Rev Biochem (**2009**) vol. 78 pp. 363-97

**Rhodes N**, Clark DE et Willet P. *Similarity searching in databases of flexible 3D structures using autocorrelation vectors derived from smoothed bounded distance matrices*. Journal of chemical information and modeling (**2006**)

**Rossmann MG**. *The single isomorphous replacement method*. A correction. Acta Crystallographica (**1962**)

**Sato** Y, Yoshikawa A, Yamagata A, Mimura H, Yamashita M, Ookata K, Nureki O, Iwai K, Komada M, Fukai S. *Structural basis for specific cleavage of Lys 63-linked polyubiquitin chains*. Nature (**2008**) vol. 455 (7211) pp. 358-62

**Shackelford** J, Maier C, Pagano JS. *Epstein-Barr virus activates beta-catenin in type III latently infected B lymphocyte lines: association with deubiquitinating enzymes*. Proc Natl Acad Sci USA (**2003**) vol. 100 (26) pp. 15572-6

**Sulea** T, Lindner HA, Ménard R. *Structural aspects of recently discovered viral deubiquitinating activities*. Biol Chem (**2006**) vol. 387 (7) pp. 853-62

**Sulea** T, Lindner HA, Purisima EO, Ménard R. *Binding site-based classification of coronaviral papain-like proteases*. Proteins (**2006**) vol. 62 (3) pp. 760-75

**Tarendeau** F, Crepin T, Guilligay D, Ruigrok R, Cusack S, Hart DJ. *Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit*. PLoS Pathog (**2008**) vol. 4 (8) pp. e1000136

**Tarendeau** F, Boudet J, Guilligay D, Mas PJ, Bougault C, Boulo S, Baudin F, Ruigrok R, Daigle N, Ellenberg J, Cusack S, Simorre J, Hart DJ. *Structure and nuclear import function of the C-terminal domain of influenza*

*virus polymerase PB2 subunit*. Nat Struct Mol Biol (**2007**) vol. 14 (3) pp. 229-33

**Van Regenmortel** MH, Mayo MA, Fauquet CM, Maniloff J. *Virus nomenclature: consensus versus chaos*. Archives of virology (**2000**) vol. 145 (10) pp. 2227-32

**Wang** D, Fang L, Li P, Sun L, Fan J, Zhang Q, Luo R, Liu X, Li K, Chen H, Chen Z, Xiao S. *The leader proteinase of foot-and-mouth disease virus negatively regulates the type I interferon pathway by acting as a viral deubiquitinase*. Journal of Virology (**2011**) vol. 85 (8) pp. 3758-66

**Wang** BC. *Resolution of phase ambiguity in macromolecular crystallography*. Methods in Enzymology (**1985**)

**Weiland** JJ, Dreher TW. *Cis-preferential replication of the turnip yellow mosaic virus RNA genome*. Proc Natl Acad Sci USA (**1993**) vol. 90 (13) pp. 6095-9

**Weiland** JJ, Dreher TW. *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 Research (**1989**) vol. 17 (12) pp. 4675-87

**Welchman** R, Gordon C, Mayer RJ. *Ubiquitin and ubiquitin-like proteins as multifunctional signals*. Nat Rev Mol Cell Biol (**2005**) vol. 6 (8) pp. 599-609

**Whitehurst** CB, Ning S, Bentz GL, Dufour F, Gershburg E, Shackelford J, Langelier Y, Pagano JS. *The Epstein-Barr virus (EBV) deubiquitinating enzyme BPLF1 reduces EBV ribonucleotide reductase activity*. Journal of Virology (**2009**) vol. 83 (9) pp. 4345-53

**Wojdyla** JA, Manolaridis I, van Kasteren PB, Kikkert M, Snijder EJ, Gorbalenya AE, Tucker PA. *Papain-like protease 1 from transmissible gastroenteritis virus: crystal structure and enzymatic activity toward viral and cellular substrates*. Journal of Virology (**2010**) vol. 84 (19) pp. 10063-73

**Yumerefendi** H, Tarendeau F, Mas PJ, Hart DJ. *ESPRIT: An automated, library-based method for mapping and soluble expression of protein domains from challenging targets*. Journal of Structural Biology (**2010**) vol. 172 (1) pp. 66-74

**Zhang** ZY, Langret DC et Perdrew JP. Planar-surface charge densities and energies beyond the local-density approximation. Physical Review B Condens Matter (**1990**) Vol. 41 pp. 5674-10

# Résumé

Le virus de la mosaïque jaune du navet (TYMV) est un excellent modèle pour la réplication des virus à ARN simple brin de polarité positive. Ce petit virus de plante code l'ensemble des protéines nécessaires à sa réplication sous forme d'un précurseur polyprotéine (206K). Du N- au C-terminus, celui-ci porte une activité méthyltransférase, protéase à cystéine (PRO), hélicase et ARN-polymérase ARN dépendante (POL). Comme tous les virus à ARN(+) connus, son complexe de réplication est étroitement lié à des membranes cellulaires. Dans le cas du TYMV, c'est l'enveloppe des chloroplastes qui est impliquée. Un des acteurs clés de la réplication est la polymérase qui permet la synthèse de nouveaux génomes. La régulation de son activité implique dans un premier temps son clivage de la 206K par PRO. Ainsi libérée, elle est recrutée aux chloroplastes par une interaction directe avec le domaine PRO du produit de clivage 140K, afin de former le complexe de réplication. Un second clivage par PRO contenu dans 140K, à la jonction PRO-hélicase, permet de poursuivre le cycle de réplication. Récemment, il a également été montré que l'activité POL était régulée par la voie ubiquitine-protéasome durant le cycle viral. Ubiquitinilée par la cellule hôte, elle est adressée au protéasome où elle sera dégradée. Cependant, PRO, grâce à sa seconde fonction ubiquitine hydrolase, est capable de la protéger de cette dégradation. Afin de caractériser d'un point de vue structural ce mécanisme de régulation de la réplication, nous avons cristallisé, à l'aide d'un contaminant, PRO et avons résolu sa structure. L'empilement cristallin est tel que le C-terminus d'un domaine PRO est inséré dans la crevasse catalytique du domaine PRO suivant, nous fournissant ainsi des informations structurales sur son activité endopeptidase. Dans un second temps, afin d'avoir des informations sur sa seconde activité, nous avons réalisé un complexe stable entre PRO et une molécule d'ubiquitine afin de le cristalliser et résoudre sa structure. Enfin, nous avons initié l'étude cristallographique de la polymérase.



**Mots clés** : Tymovirus, réplication, polymérase, protéase, ubiquitination, cristallographie



Laboratoire de Virologie Moléculaire et structurale
Equipe de biocristallographie
UPR CNRS 3296
1 avenue de la terrasse
91198 Gif sur Yvette cedex

Université Paris-sud 11
Pôle Ingenierie des protéines et cibles thérapeuthiques
UFR «Faculté de Pharmacie de Chatenay Malabry»
5, rue Jean Baptiste Clément
92296 Châtenay-Malabry Cedex