Development of a wheat germ cell-free expression system for the production, the purification and the structural and functional characterization of eukaryotic membrane proteins : application to the preparation of hepatitis C viral proteins

Marie-Laure Fogeron

▶ **To cite this version:**

Marie-Laure Fogeron. Development of a wheat germ cell-free expression system for the production, the purification and the structural and functional characterization of eukaryotic membrane proteins : application to the preparation of hepatitis C viral proteins. Biochemistry, Molecular Biology. Université Claude Bernard - Lyon I, 2015. English. NNT : 2015LYO10081 . tel-01213970

HAL Id: tel-01213970

https://theses.hal.science/tel-01213970

Submitted on 9 Oct 2015

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.

N° d'ordre 81 - 2015 Année 2015

THESE DE L'UNIVERSITE DE LYON

Délivrée par

L'UNIVERSITE CLAUDE BERNARD LYON 1

Ecole Doctorale Interdisciplinaire Sciences-Santé (EDISS)

DIPLOME DE DOCTORAT
(arrêté du 7 août 2006)

soutenue publiquement le 30 juin 2015

par

Madame Marie-Laure FOGERON

Development of a wheat germ cell-free expression system for the production, the purification and the structural and functional characterization of eukaryotic membrane proteins

Application to the preparation of hepatitis C viral proteins

Directeurs de thèse : Madame Anja BÖCKMANN & Monsieur François PENIN

JURY

Monsieur Gilbert DELEAGE	*président du jury*
Monsieur Jean DUBUISSON	*rapporteur*
Monsieur Xavier HANOULLE	*rapporteur*
Monsieur Jean-Michel BETTON	*examinateur*
Monsieur François-Loic COSSET	*examinateur*
Madame Anja BÖCKMANN	*examinatrice*
Monsieur François PENIN	*invité*

REMERCIEMENTS

Je tiens tout d'abord à remercier Gilbert Deléage pour m'avoir accueillie dans l'unité BMSSI dont il est directeur.

Je souhaite également remercier Anja Böckmann et François Penin pour m'avoir donné l'opportunité de réaliser ce travail dans leur équipe, pour avoir partagé leur expérience et leur savoir, et pour m'avoir toujours encouragée à repousser mes limites.

Je tiens également à remercier Xavier Hanoulle et Jean Dubuisson d'avoir accepté d'être rapporteur de cette thèse. Un grand merci également à Jean-Michel Betton d'avoir accepté de faire partie de mon comité de suivi de thèse et de ce jury, mais aussi pour tous ses précieux conseils concernant la biochimie des protéines. Je souhaite également remercier François-Loïc Cosset d'avoir accepté de faire partie de ce jury, et Gilbert Deléage d'avoir accepté de le présider.

Un grand merci à Pierre Falson d'avoir accepté de faire partie de mon comité de suivi de thèse et pour tous ses précieux conseils sur les détergents.

I am very grateful to Prof. Yaeta Endo for his initial help during the development of the wheat germ cell-free system in the lab. His advices were unvaluable and I appreciated very much his great kindness. I would also like to thank Masaki Madono and Matthias Harbers from CellFree Sciences.

Of course I would like to thank our main collaborators for the great work over the last years. Many thanks to Beat Meier and his group from the ETH Zurich for the fruitful collaboration. Thank you to Alons Lends, Anne Schütz and Susanne Penzel for solid-state NMR measurements of our samples. Un grand merci à Simon Meister, Jérôme Gouttenoire et Darius Moradpour de l'Université de Lausanne. Many thanks also to David Paul, Vlastimil Jirasko and Ralf Bartenschlager from Heidelberg University. Rendez-vous au Schauinsland! Merci également à Célia Boukadida et Annette Martin de l'Institut Pasteur à Paris. Merci aussi à François-Loïc Cosset, Marlène Dreux, Sonia Assil, Christophe Ramière, Olivier Diaz, ainsi qu'à Patrice André du CIRI à Lyon. Merci également à Martin Baril et Daniel Lamarre de l'Université de Montréal. Je tiens enfin à remercier Philippe Roingeard et Sonia Georgeault de la Plate-Forme RIO des Microscopies à Tours pour avoir analysé nombre de nos échantillons en microscopie électronique, pour leurs conseils et leur enthousiasme. Merci également à Moreno Lelli du Centre de Résonance Magnétique Nucléaire à très hauts champs (CRMN) à Lyon pour les mesures de nos échantillons en RMN du liquide. I really enjoyed working with all of you. And I am happy to continue!

Un grand merci à toute l'équipe « RMN et virus de l'hépatite C » pour l'ambiance plus que sympathique et la participation sans faille aux séances de tri de germes de blé. Et plus particulièrement… Aurélie pour toutes nos discussions scientifiques et celles plus personnelles. Carole, vive le quick finish!!! Britta, es hat spaβ gemacht! Stéphane, jeune retraité bien occupé. Roland, alias DJ Roro lors des séances de tri, pour m'avoir appris à faire du CD et pour ta bonne humeur perpétuelle. Denis, arrête de courir, mais surtout pas de nous faire rire! Clément D., je suis très heureuse que tu sois arrivé dans l'équipe et je croise les doigts pour la thèse! Merci aussi à tous ceux que je n'ai cotoyé qu'un temps mais tout autant apprécié : Jenny, Nina, Birgit, Anaïs, Eve, Marie, Mihayl, Julie, Clément L., Loick, Thibault… et à tous ceux que j'oublie.

Merci à tous les IBCPien(ienne)s, plus particulièrement Dorothée, Christine et Souad qui rendent le quotidien agréablement plus facile. A Frédéric Delolme aussi, pour toutes les analyses en spectrométrie de masse et son dynamisme. Merci enfin à tous ceux que je ne peux citer ici faute de place.

Merci enfin à Ludovic Curabet du GAEC du Logis Neuf à Mornant et à l'abattoir de Roanne pour nous avoir fourni un foie de porc pour la préparation de lipides. En écrivant ces quelques mots, j'ai aussi une pensée pour les céréaliers qui sur leurs moissonneuses-batteuses sont sans doute loin d'imaginer qu'ils contribuent à des projets de recherche. Et pourtant!

Et parce qu'il y a eu une vie avant l'IBCP et les germes de blé :

Danke auch an alle meine ehemaligen Kollegen der Europroteome AG in Berlin, insbesondere Steffi, Silvia und Monika. Ich bin froh, dieses Abendteuer mit euch erlebt zu haben! Un grand merci également au Prof. Marc Reymond pour m'avoir accordé toute sa confiance au début de ma carrière.

Ein riesiges Dankeschön an die AG Lange! Bodo, Verena, Anne, Tanja, Seon-Hi, Karin, Hannah, Nicole, Armin, Karl und Ingrid, diese vieranhalb Jahre mit euch waren einfach der Hammer! Aus der Zeit am MPI behalte ich nur gute Errinnerungen. Tausend Dank Bodo, mir die Möglichkeit gegeben zu haben, mich beruflich so weiter zu entwickeln.

Merci enfin au « ex-NBB », cette équipe de filles attachantes qui m'a si bien accueillie à mon retour en France et tant encouragée pour le concours: Anne-Laure, ma binôme de choc lors de nos aventures au PBES (« je l'ai ma charlotte là??! »), Anne, Marie-Claude et Hélène pour les pauses Gerblé et toutes nos discussions, Héloise et Blandine pour être devenues plus que de simples collègues et enfin Nathalie pour m'avoir toujours soutenue.

Enfin, parce qu'il y a une vie à côté :

Merci à tous mes amis rencontrés à Berlin, ceux qui y sont restés et ceux qui sont maintenant aux quatre coins du monde. Les francophones bien sûr, Olivier, Véronique et Lina (que d'après-midis shopping, de soirées, de fous rires et autres galères!), mais aussi les autres. Herzlichen Dank auch an Karolin (bleib wie Du bist!) und Christopher. Bejta, ein zärtliches Gedanken habe ich ebenfalls für Dich und Deine Familie.

Merci à mes amis depuis le lycée : Karine, Jérôme, Fabrice et Anne-Lise. Que j'aime nos éclats de rire qui rendent la vie si légère! Merci aussi à Alexandra et Jean-Louis pour votre bonne humeur et les bonnes bouffes, en vous souhaitant le plus grand bonheur du monde.

Merci à mes beaux-parents, Marie-Monique et Pierre, pour leur grande gentillesse, ainsi qu'à mes belles-sœurs et beaux-frères, et à la bande de cousins : Killian, Célia, Maël, Augustin, Timothée… et les prochains! A la « maxoucoise » ou pas, que de belles vacances en perspective!

Merci à mon frère Christophe et à Marie-Christelle, ou devrais-je plutôt dire Miguel et Jeannie?! A mon « p'tit » frère Eric et à Catherine, loin des yeux peut-être mais toujours dans mon cœur. Vivement la prochaine fois! Merci aussi à mes nièces, Clémence, Juliette et Florence, et mes neveux, Gabriel et le petit canadien à venir. Je suis très heureuse que Flavien ait des cousins comme vous! Un grand merci à mes parents, Eliane et Robert, « fournisseurs officiels » de fruits et légumes frais et autres bons produits. Les petits week-ends en Ardèche sont toujours un grand bol d'oxygène.

Les deux hommes de ma vie, enfin! Antoine, merci pour TOUT. La vie n'est qu'un manège. Qu'il continue de nous faire tourner la tête! Mon petit Flavien, tes coups au creux de mon ventre d'abord, puis tes sourires, tes rires et enfin tes premiers pas m'ont accompagnée tout au long de cette thèse. Tu es mon plus grand bonheur.

This thesis has been carried out under the supervision of Anja Böckmann and François Penin, in the team « NMR and hepatitis C virus » that is part of the unit UMR5086 – BMSSI (Bases Moléculaires et Structurales des Systèmes Infectieux) at IBCP :

Institut de Biologie et Chimie des Protéines

7, passage du Vercors

F-69367 Lyon cedex7

Development of a wheat germ cell-free expression system for the production, the purification and the structural and functional characterization of eukaryotic membrane proteins

Application to the preparation of hepatitis C viral proteins

Summary

While 30% of the genome encodes for membrane proteins, less than 3% of protein structures in the Protein Data Bank correspond to such proteins. Due to their hydrophobic nature, membrane proteins are indeed notoriously difficult to express in classical cell-based protein expression systems. The structural study of the membrane proteins of hepatitis C virus (HCV) in their full-length and native form has therefore been for long time hampered. HCV is a positive-strand RNA virus building its replication complex on a specific membrane rearrangement (membranous web), which serves as a scaffold for the HCV replicase, and is induced by the concerted action of several HCV non-structural proteins including NS2, NS4B and NS5A. The knowledge of the three-dimensional structure of these proteins and their role in virus replication is still limited. To overcome the limitations that prevent the structural and functional studies of these proteins, a wheat germ cell-free protein expression system has been developed. A production protocol was designed which allows us to directly obtain membrane proteins in a soluble form by adding detergent during the *in vitro* protein synthesis. A large number of mainly viral proteins were successfully expressed, and full protocols were developed for the full-length NS2, NS4B and NS5A proteins. These membrane proteins were produced and purified by affinity chromatography using a *Strep*-tag II in the milligram range. These protein samples are homogenous, as shown by gel filtration analysis. Moreover, structural analyses by circular dichroism showed that the proteins produced in the wheat germ cell-free system are well folded. Reconstitution of these proteins in lipids is currently under optimization. The ultimate goal is to determine their structure by solid-state NMR in a native-like membrane lipids environment.

Keywords

Membrane protein biochemistry - Molecular Biology - Hepatitis C virus - Cell-free protein expression - Detergents - Protein purification - Lipid reconstitution - Nuclear magnetic resonance (NMR) - Protein structure and function

Discipline

Biochemistry

Développement d'un système d'expression acellulaire à base d'extrait de germe de blé pour la production, la purification et la caractérisation structurale et fonctionnelle de protéines membranaires eucaryotes

Application à la préparation des protéines du virus de l'hépatite C

Résumé substantiel

Les protéines membranaires jouent un rôle crucial dans de nombreux processus biologiques essentiels à la vie et sont donc des cibles importantes pour le développement de nouveaux médicaments. Alors qu'environ 30% du génome code pour des protéines membranaires, moins de 3% des structures de protéines dans la Protein Data Bank correspondent à ces protéines. En raison de leur nature hydrophobe, les protéines membranaires sont en effet très difficiles à exprimer dans des systèmes d'expression de protéines classiques en cellules, et notamment en bactéries. Souvent toxiques, les protéines membranaires subissent également parfois une dégradation dans les cellules ou s'agrègent dans des corps d'inclusion, rendant leur solubilisation et leur renaturation très délicates. Ainsi, non seulement l'expression, mais également la purification des protéines membranaires sont des étapes difficiles pour leur analyse structurale et fonctionnelle. C'est pourquoi l'étude structurale des protéines membranaires de virus de l'hépatite C, sous une forme native et entière, a été longtemps entravée.

Avec environ 160 millions de personnes chroniquement infectées dans le monde, l'hépatite C est un problème de santé mondial majeur. Le virus de l'hépatite C (VHC) est un virus à ARN positif dont le génome code pour une polyprotéine clivée de manière co- et post-traductionnelle en dix protéines qui sont toutes associées aux membranes. Bien que des stratégies thérapeutiques efficaces aient été mises au point récemment, le rôle des protéines du VHC dans la réplication du virus n'est toujours pas complètement élucidé. En particulier, le virus construit son complexe de réplication grâce à un réarrangement de la membrane du réticulum endoplasmique, appelé "membranous web". Ce réarrangement membranaire spécifique, qui sert de base pour la réplicase du VHC, est induit par l'action concertée de plusieurs protéines non structurales du VHC,

incluant NS2, NS4B et NS5A. La connaissance de la structure tridimensionnelle de ces protéines et leur rôle dans la réplication du virus reste limitée, principalement en raison du fait qu'elles sont des protéines membranaires, et de surcroît oligomériques: NS2 est composée d'un domaine transmembranaire et d'un ectodomaine, NS4B est une protéine membranaire intégrale et NS5A est ancrée à la membrane par une hélice amphipathique.

Dans ce contexte, le but de cette thèse a été de développer au laboratoire un système d'expression acellulaire à base de germe de blé, permettant l'expression et la purification de protéines membranaires du VHC pour leur caractérisation structurale et fonctionnelle. Le but ultime est la détermination de leur structure par RMN du solide dans un environnement lipidique semblable à l'environnement membranaire natif.

Dans la partie bibliographique ("Knowledge from literature"), des informations générales sur les protéines membranaires, et en particulier sur les systèmes d'expression classiques et les systèmes d'expression acellulaires, y compris le système acellulaire à base de germe de blé, sont présentées. Les connaissances actuelles sur le VHC sont également abordées. Dans la partie «Materiel and methods» qui suit, la préparation de constructions d'ADN pour l'expression acellulaire à base de germe de blé, la préparation des extraits de germe de blé, ainsi que l'ensemble du processus d'expression acellulaire sont détaillées. En outre, les procédures de purification et autres méthodes de caractérisation, ainsi que les techniques de reconstitution en lipides sont rapportées. Les résultats obtenus au cours de ce travail sont finalement présentés et discutés dans la dernière partie.

Au cours de cette thèse, un système d'expression de protéine acellulaire à base de germe de blé a été développé avec succès au laboratoire. La préparation d'extraits de germe de blé a été optimisée à partir de protocoles établis par le Prof. Y. Endo et ses collègues à l'Université d'Ehime à Matsuyama, Japon, conduisant à des extraits de très bonne qualité. De plus, la technique en bicouche s'est révélée être la plus appropriée par rapport à un système de dialyse. Les principaux paramètres d'expression, tels que la température ou la concentration en extrait de germe de blé, ont également été optimisés. L'expression de 75 constructions différentes, essentiellement des protéines membranaires, a ensuite été testée en mode précipité avec un succès supérieur à 82%.

En outre, l'expression des précurseurs NS3-4A et NS2-NS3^{pro} a mis en évidence que les protéines produites dans ce système sont actives. Le développement du système est donc un premier succès.

Un progrès majeur a été le choix des détergents pour exprimer les protéines membranaires directement sous forme solubilisée. Notamment le détergent MNG-3 a permis l'expression des protéines NS2, NS4B et NS5A entières sous forme solubilisée et bien repliée, comme l'ont montré des analyses de dichroïsme circulaire. L'expression directement sous forme solubilisée évite bien des difficultés liées à la purification à partir d'agrégats obtenus dans le mode de précipité ou bien de corps d'inclusion lorsque les protéines membranaires sont produites en bactéries, notamment la présence de formes mal repliées et la perte de protéine. Ces résultats représentent une avancée importante dans l'élaboration de stratégies pour l'expression acellulaire de protéines membranaires.

Jusqu'à présent, seuls certains domaines de NS2, NS4B et NS5A avaient pu être étudiés en utilisant différentes approches biophysiques telles que la cristallographie aux rayons X et la RMN en solution. Des études plus détaillées des formes complètes de ces protéines n'avaient en effet pas pu être réalisées en raison des difficultés rencontrées lors de leur surexpression. Dans le cas particulier de NS2, une protéine tronquée sans le domaine N-terminal a été utilisée pour la caractérisation fonctionnelle de l'activité protéasique. Alors que de nombreuses études sur les interactions entre NS2 et plusieurs autres protéines virales ont été réalisées *in cellulo*, aucun test *in vitro* n'a pu être réalisé sur ces interactions afin de mieux comprendre les mécanismes moléculaires mis en jeu. En effet, certaines de ces interactions ont lieu vraisemblablement via la région hydrophobe N-terminale insérée dans la membrane. En utilisant notre système d'expression acellulaire à base de germe de blé, les protéines NS2, NS4B et NS5A entières ont pu être produites et purifiées par chromatographie d'affinité grâce à une étiquette *Strep*II, sous forme solubilisée dans des quantités de l'ordre du milligramme. Des analyses en filtration sur gel ont montré que ces échantillons étaient homogènes et oligomériques. De plus, les structures secondaires déterminées par dichroïsme circulaire sont cohérentes avec les données structurales obtenues à partir des domaines isolés. La possibilité de produire les protéines NS2, NS4B et NS5A purifiées sous forme

solubilisée en présence de détergent, en quantités suffisantes, ouvre la voie à des études biochimiques structurales et fonctionnelles *in vitro* de ces protéines énigmatiques.

Le but ultime des travaux de recherche du laboratoire est de déterminer la structure de NS2, NS4B et NS5A par RMN du solide dans un environnement lipidique semblable à l'environnement membranaire natif. Bien que l'expression dans des systèmes acellulaires en présence de liposomes puisse permettre l'insertion directe des protéines produites dans la membrane, la solubilisation avec un détergent et la réinsertion ultérieure dans des membranes permet d'obtenir des quantités de protéines plus élevées et des préparations plus pures. Elle permet également de définir précisément la composition lipidique utilisée ainsi que le rapport lipide/protéine. La reconstitution des protéines NS2 et NS4B entières a donc été optimisée. Des analyses en microscopie électronique ont montré que les deux protéines étaient bien insérées dans les membranes. Les premiers spectres de RMN à l'état solide ont été obtenus pour la protéine NS4B dans un mélange de α-phosphatidylcholine et de cholestérol. Ces données préliminaires très encourageantes sont la preuve que le système mis en place permet de produire des échantillons de protéines marquées de qualité pour les études de RMN. L'insertion réussie dans des membranes est en effet un bon indicateur de l'intégrité structurale des protéines membranaires. Bien que les différents mélanges de lipides testés lors de ce travail n'ont pas montré de différences significatives en ce qui concerne l'efficacité de la reconstitution, il n'est pas exclu que la composition lipidique puisse influencer la qualité des spectres. L'aventure ne fait que commencer!

Le travail présenté ici a conduit à deux publications incluses dans ce manuscrit (cf section IV.2.2. et annexe 2, respectivement):

- Fogeron *et al.*, Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins**.** *Publiée dans Protein Expression and Purification* (Fogeron *et al*, 2015).

- Fogeron *et al.*, Functional expression, purification, characterization, and membrane reconstitution of nonstructural protein 2 from hepatitis C virus**.** *Soumise à Protein Expression and Purification (actuellement en révision)*.

Mots clés

Biochimie des protéines membranaires - Biologie moléculaire - Virus de l'hépatite C - Expression acellulaire de protéines - Détergents - Purification des protéines - Reconstitution en lipides - Résonance magnétique nucléaire (RMN) - Structure et fonction des protéines

Discipline

Biochimie

TABLE OF CONTENTS

LIST OF TABLES 18
LIST OF ILLUSTRATIONS 19
ABBREVIATIONS 21
I. INTRODUCTION 23
II. KNOWLEDGE FROM LITERATURE 27
1. Membrane proteins 29
1.1. General information on membrane proteins 29
1.2. Membrane protein expression systems 34
2. Wheat germ cell-free expression system 42
2.1. History 42
2.2. The wheat germ system 43
2.3. Isotopic labeling for NMR studies 47
3. Membrane protein expression using the wheat germ cell-free system 49
3.1. Expression modes 49
3.2. Advantages of the WGE-CF system over classical cell-based and other cell-free systems 54
4. Reconstitution of membrane proteins in lipids 57
4.1. Classical reconstitution protocols 57
4.2. GRecon 60
5. Hepatitis C virus 62
5.1. Hepatitis C 62
5.2. The virus 65
5.3. HCV life cycle 68
5.4. Model systems for the study of HCV 71
6. Hepatitis C virus membrane proteins 79
6.1. Core 79
6.2. E1 and E2 81
6.3. p7 82
6.4. NS2 83
6.5. NS3-4A 84
6.6. NS4B 85
6.7. NS5A 87
6.8. NS5B 91
III. MATERIAL AND METHODS 93
1. Plasmids 95
1.1. Primer design 95
1.2. Polymerase Chain Reaction (PCR) 95
1.3. Purification of PCR products 98
1.4. Digestion of PCR products and pEU-E01-MCS vector 98
1.5. Extraction of digested products from agarose gel 99
1.6. Ligation 99
1.7. Transformation of *E. coli* competent cells 99
1.8. Screening by colony PCR 100
1.9. Mini DNA preparation 100
1.10. Digestion of obtained constructs 101
1.11. Preparation of glycerol stocks 101

2. DNA production 102
2.1. Maxi DNA preparation 102
2.2. Phenol/chloroform extraction 103
3. Wheat germ extract preparation 104
3.1. Preparation of wheat germs 104
3.2. Preparation of wheat germ extracts 105
4. Wheat germ cell-free protein expression 108
4.1. Transcription 108
4.2. Translation 108
5. Protein sample preparation 112
6. Purification by affinity chromatography 113
7. Size exclusion chromatography 115
8. Circular dichroism 116
9. Lipid reconstitution 117
9.1. Preparation of lipids from pig liver 117
9.2. Preparation of Bio-Beads 118
9.3. Reconstitution by dialysis (classical protocol) 118
9.4. Reconstitution using the GRecon method 119
10. Membrane protein trapping with amphipol A8-35 121
11. SDS-PAGE and Western blotting 122
11.1. SDS-PAGE analysis 122
11.2. Western blotting analysis 122
12. Benzonase production 123
12.1. Bacterial cultures 123
12.2. Isolation of benzonase 123
12.3. Production yield and activity test 124
IV. RESULTS AND DISCUSSION 127
1. Implementation and optimization of the wheat germ cell-free system 129
1.1. Production of wheat germ extracts 129
1.2. Production of mRNA 132
1.3. Optimization of the expression conditions 134
1.4. Expression test of 75 different protein constructs 140
1.5. NS3-4A as an example to evidence activity of protein expressed using the WGE-CF system 144
1.6. Conclusion and perspectives 145
2. Wheat germ cell-free expression in the presence of detergents 146
2.1. Preliminary tests 146
2.2. MNG-3 and C12E8 detergents for the production of solubilized membrane proteins 149
2.3. Alternatives to commercial detergents 159
2.4. Conclusion and perspectives 162
3. NS2 163
3.1. Full-length NS2 expression test 163
3.2. Cotranslational protease activity of full-length NS2 164
3.3. Purification of full-length NS2 165
3.4. Analysis of full-length NS2 by circular dichroism 168
3.5. Protease activity of purified full-length NS2 169
3.6. Analysis of full-length NS2 by solution NMR 171
3.7. Reconstitution of full-length NS2 in lipids 173

3.8. Stabilization of full-length NS2 using amphipol A8-35 180
3.9. $NS2^{pro}$ 185
3.10. Conclusion and perspectives 187
4. NS4B 189
4.1. NS4B expression tests 189
4.2. Full-length NS4B purification 190
4.3. Analysis of full-length NS4B by circular dichroism 193
4.4. Analysis of full-length NS4B by solution NMR 193
4.5. Reconstitution of full-length NS4B in lipids 195
4.6. Analysis of full-length NS4B by solid-state NMR 198
4.7. Conclusion and perspectives 199
5. NS5A 201
5.1. NS5A expression test 201
5.2. NS5A purification 203
5.3. First solid-state NMR spectra of sedimented NS5A D1D2D3 204
5.4. Conclusion and perspectives 206
6. Core 208
6.1. Core expression test 208
6.2. Purification of capsid-like particles 209
6.3. First solid-state NMR spectra 211
6.4. Conclusion and perspectives 213
7. General conclusions and perspectives 214
REFERENCES 217
APPENDIX 1 - Amino acid sequences 249
APPENDIX 2 - NS2 manuscript 259
APPENDIX 3 - CV 279

LIST OF TABLES

Table II.1	Comparison of classical cell-based and main cell-free protein expression systems	56
Table III.1	Summary of cloned constructs	97
Table III.2	Preparation of the transcription mix	108
Table III.3	Preparation of the translation mix	110
Table III.4	Preparation of the feeding buffer	110
Table III.5	Summary of the *Strep*-Tactin column formats used in a standard way	114
Table IV.1	Expression of 75 different protein constructs using the WGE-CF system in the precipitate mode	141
Table IV.2	List of detergents tested for GFP expression using the WGE-CF system	147

LIST OF ILLUSTRATIONS

Figure II.1	The machinery of membrane protein assembly	31
Figure II.2	Types of single TM helix MPs	32
Figure II.3	Summary of various interactions stabilizing membrane proteins stably folded in fluid lipid bilayers	34
Figure II.4	Schematic representation of a wheat seed	42
Figure II.5	Schematic representation of the pEU vector	43
Figure II.6	Modes of expression for membrane proteins in cell-free systems	49
Figure II.7	Hepatitis C global prevalence 2010	62
Figure II.8	Evolution of HCV infection	63
Figure II.9	HCV lipoviroparticle (LVP)	66
Figure II.10	Genetic organization and polyprotein processing of HCV	68
Figure II.11	The hepatitis C virus replication cycle	69
Figure II.12	Important HCV cell culture systems	73
Figure II.13	Different approaches to create mouse models for the study of HCV	77
Figure II.14	Structure and membrane association of HCV proteins	79
Figure II.15	Schematic representation of the HCV core protein	81
Figure II.16	Model of NS2 membrane topology	83
Figure II.17	Dual membrane topology of HCV NS4B	87
Figure II.18	Schematic representation of the HCV NS5A	88
Figure II.19	Structure of the NS5A dimer	90
Figure III.1	Multiple Cloning Site (MCS) of the pEU-E01-MCS vector	95
Figure III.2	Anion-exchange chromatography column from the NucleoBond Xtra Maxi kit	102
Figure III.3	Preparation of wheat germs	104
Figure III.4	Example of the sample after the first 30,000 g centrifugation	106
Figure III.5	In vitro protein synthesis using the bilayer method	109
Figure III.6	Small scale in vitro protein synthesis using the dialysis method	111
Figure III.7	Schematic representation of protein sample preparation for small scale expression tests	112
Figure III.8	Schematic illustration of the *Strep*-tag purification cycle	113
Figure III.9	Schematic representation of the standard affinity purification process	114
Figure III.10	Far UV CD spectra associated with various types of secondary structure	116
Figure III.11	Preparation of lipids from pig liver	117
Figure III.12	Schematic representation of the trapping of NS2 with the amphipol A8/35	121
Figure III.13	Production of benzonase in M9 medium	124
Figure III.14	Quality control of home-made benzonase	125
Figure IV.1	Quality control of wheat germ extracts	131
Figure IV.2	Analysis of mRNA on agarose gel	132
Figure IV.3	Experimental conditions for GFP expression	136
Figure IV.4	Bilayer method versus dialysis method	139
Figure IV.5	Expression of 75 different protein constructs in the precipitate mode .	140

Figure IV.6 Expression of wild type and mutant NS3-4A precursors using the WGE-CF system in the precipitate mode 144
Figure IV.7 Expression of GFP in the presence of various detergents 148
Figure IV.8 Production of full-length NS2 in the presence of DDM, MNG-3 or calixarenes 160
Figure IV.9 Expression of GFP as well as full-length NS2, NS4B and NS5A proteins in the presence of amphipol A8-35 161
Figure IV.10 Small scale expression test of full-length NS2 constructs in the precipitate mode and in the presence of MNG-3 detergent 164
Figure IV.11 Cotranslational protease activity of full-length NS2 165
Figure IV.12 Purification of full-length NS2 expressed in the precipitate mode 166
Figure IV.13 Purification of full-length NS2 by affinity chromatography and analysis by size exclusion chromatography 167
Figure IV.14 Analysis of full-length NS2 by circular dichroism (CD) 169
Figure IV.15 Analysis of purified full-length NS2 protease activity 171
Figure IV.16 Analysis of full-length NS2 by solution NMR 173
Figure IV.17 Reconstitution of full-length NS2 with high LPR 175
Figure IV.18 Reconstitution of full-length NS2 in lipids by dialysis with a LPR of 1 177
Figure IV.19 Reconstitution of full-length NS2 in PC/Chol using the GRecon method 179
Figure IV.20 Characterization of NS2/A8-35 complexes 182
Figure IV.21 Precipitation tests of NS2/A8-35 complexes 184
Figure IV.22 Production of $NS2^{pro}$ 186
Figure IV.23 Small scale expression test of full-length NS4B constructs in the presence of MNG-3 detergent 190
Figure IV.24 Purification of full-length NS4B by affinity chromatography 191
Figure IV.25 Purification of full-length NS4B by affinity chromatography and analysis by size exclusion chromatography 192
Figure IV.26 Far UV circular dichroism (CD) spectrum of full-length NS4B after purification by affinity chromatography and size exclusion chromatography 193
Figure IV.27 Analysis of full-length NS4B by solution NMR 195
Figure IV.28 Reconstitution of full-length NS4B in lipids 197
Figure IV.29 Analysis of full-length NS4B by solid-state NMR 199
Figure IV.30 Small scale expression test of the NS5A constructs in the precipitate mode 202
Figure IV.31 Purified NS5A D1 and D1D2D3 203
Figure IV.32 Small scale purification of AH-D1 and full-length NS5A samples 204
Figure IV.33 Large scale purification of an uniformly $^{13}C/^{15}N$ labeled NS5A D1D2D3 for solid-state NMR analysis 205
Figure IV.34 First solid-state NMR spectra of sedimented NS5A D1D2D3 206
Figure IV.35 Small scale expression test of the core constructs in the precipitate mode 209
Figure IV.36 Examples of optimization steps for the isolation of core D1 nucleocapsid-like particles on Accudenz cushions 210
Figure IV.37 Production of an uniformly $^{13}C/^{15}N$ labeled core D1 sample for solid-state NMR analysis 211
Figure IV.38 First solid-state NMR spectra of core D1 212

ABBREVIATIONS

A	aa	amino acid residue
	ADP	Adenosine diphosphate
	AH	Amphipatic helix
	APR	Amphipol-to-protein ratio
	ARFP	Alternative reading frame protein
	ATP	Adenosine triphosphate
C	C12E8	Dodecyl octaethylene glycol ether
	CD	Circular dichroism
	CECF	Continuous exchange cell-free system
	CFCF	Continuous flow cell-free system
	CFS	Cell-free sample
	CKIα	Casein kinase I, isoform α
	CKII	Casein kinase II
	CLDN1	Claudin 1
	CMC	Critical micelle concentration
	CV	Column volume
	CypA	Cyclophilin A
D	DAA	Direct acting antiviral
	DDM	n-dodecyl-β-D-maltoside
	DM	n-decyl-β-D-maltoside
	DTT	Dithiothreitol
	DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
	DNA	Desoxyibonucleotidic acid
E	EDTA	Ethylenediaminetetraacetic acid
	EM	Electron microscopy
	ER	Endoplasmic reticulum
F	F	Frameshift protein
G	GAG	Glycosaminoglycan
	GDP	Guanosine diphosphate
	GFP	Green fluorescent protein
	GPCR	G protein-coupled receptor
	GST	Glutathione S-transferase
	GTP	Guanosine triphosphate
H	HAV	Hepatitis A virus
	HBV	Hepatitis B virus
	HCV	Hepatitis C virus
	HEV	Hepatitis E virus
	HTA	Host target antiviral
I	IFNα	Interferon α
	IRES	Internal ribosome entry site
	ISDR	Interferon sensitivity determining region
L	LCS	Low complexity sequence
	LD	Lipid droplet

	LDL	Low density lipoprotein
	LDL-R	Low density lipoprotein receptor
	LPR	Lipid-to-protein ratio
	LVP	Lipoviroparticle
M	MBP	Maltose-binding protein
	MNG-3	Lauryl maltose neopentyl glycol
	MP	Membrane protein
	mRNA	Messenger RNA
	MW	Molecular weight
	MWCO	Molecular weight cut-off
N	NI	Nucleoside inhibitor
	NMR	Nuclear magnetic resonance
	NNI	Non-nucleoside inhibitor
	NPC1L1	Niemann-Pick C1-like 1 cholesterol transporter
	NTP	Nucleoside triphosphate
	NTR	Non translated region
O	OCLN	Occludin
	OG	n-octyl-D-glucopyranoside
	OM	Outer membrane
	ORF	Open reading frame
P	P	Pellet
	PDB	Protein Data Bank
	PI4KIIIα	Phosphatidylinositol 4-kinase IIIα
	PI4P	Phosphatidylinositol 4-phosphate
R	RdRp	RNA-dependent RNA polymerase
	REM	Replication-enhancing mutation
	RNA	Ribonucleotidic acid
	RRL	Rabbit reticulocyte lysate
	RT	Room temperature
S	SDS	Sodium dodecyl sulfate
	SN	Supernatant
	SPR	Surface plasmon resonance
	SR	Signal recognition particle (SRP) receptor
	SRB1	Scavenger receptor class B type I
	SRP	Signal recognition particle
T	TM	Transmembrane domain
	tRNA	Transfer RNA
U	UPR	Unfolded protein response
V	VLDL	Very low density lipoprotein
W	WGE	Wheat germ extract
	WGE-CF	Cell-free system using wheat germ extract
Y	YFP	Yellow fluorescent protein

I. INTRODUCTION

Membrane proteins are essential in many biological processes crucial for life and are thus important targets for pharmaceutical drugs. While about 30% of the genome encodes for membrane proteins, less than 3% of protein structures in the Protein Data Bank correspond to such proteins. Due to their hydrophobic nature, membrane proteins are indeed notoriously difficult to express in classical cell-based protein expression systems. Often toxic, membrane proteins also undergo proteolysis in cells or aggregate in inclusion bodies, making delicate issues further solubilization and renaturation. Not only the expression, but also the purification of membrane proteins are major bottlenecks in their structural and functional analysis. This is why the structural study of the membrane proteins of hepatitis C virus in their full-length and native form has been for long time hampered.

With about 160 million people chronically infected worldwide, hepatitis C is a major global health problem. The hepatitis C virus (HCV) has a positive-strand RNA genome encoding for a polyprotein which is co-and posttranslationally cleaved into ten proteins, all being associated to membranes. Although numerous therapeutic strategies have been developed recently, the role of the HCV proteins remains not fully elucidated. In particular, the virus builds its replication complex on rearranged intracellular membranes, designated as membranous web. This specific membrane rearrangement, which serves as a scaffold for the HCV replicase, is induced by the concerted action of several HCV non-structural proteins including NS2, NS4B and NS5A. The knowledge of the three-dimensional structure of these proteins and their role in virus replication is limited, mainly due to the fact that they are oligomeric membrane proteins: NS2 is composed of a transmembrane domain and an ectodomain, NS4B is an integral membrane protein and NS5A is anchored to the membrane by an amphipathic helix.

In this context, the aim of this thesis was to develop a wheat germ cell-free protein expression allowing the expression and further purification of HCV membrane proteins for their structural and functional characterization. The ultimate goal is the determination of their structure by solid-state NMR in a native-like membrane lipids environment.

General information about membrane proteins, especially about the cell-based and cell-free expression systems available, including the wheat germ cell-free system, as well

as the current knowledge about HCV are first presented in this work (section « Knowledge from literature »). In the following « Material and methods » section, the preparation of DNA contructs for wheat germ cell-free expression, the preparation of the wheat germ extracts and the whole wheat germ cell-free expression process are detailed. In addition, the purification procedures and further characterization methods, as well as lipid reconstitution techniques are explained. The results obtained during this work are finally presented and discussed. A wheat germ cell-free expression system has been successfully developed in the lab. Full-length NS2, NS4B and NS5A have been produced in the milligram range directly in a detergent-solubilized form and purified by affinity chromatography using a *Strep*-tag II. Then, the homogeneity of these protein samples was checked by gel filtration analysis. Moreover, structural analyses by circular dichroism have been carried out and showed that the proteins produced in the wheat germ cell-free system are well folded. First reconstitution tests of these proteins in lipids using classical reconstitution protocols or a reconstitution protocol directly on sucrose gradient were developed too.

This work has led to two publications included in this manuscript :

- Fogeron *et al.*, Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins**.** *Published in Protein Expression and Purification* (Fogeron *et al*, 2015).

- Fogeron *et al.*, Functional expression, purification, characterization, and membrane reconstitution of nonstructural protein 2 from hepatitis C virus**.** *Under revision in Protein Expression and Purification.*

II. KNOWLEDGE FROM LITERATURE

1. Membrane proteins

Membrane proteins are key molecules in the cell and are important targets for pharmaceutical drugs (Arinaminpathy *et al*, 2009; Fagerberg *et al*, 2010). Although 25% to 30% of genes encode for membrane proteins, only 2.4% of protein structures in the Protein Data Bank (PDB) correspond to membrane proteins (http://www.rcsb.org). Membrane proteins are indeed notably difficult to produce, mainly due to their hydrophobic nature. In the first part of this section, general information is given about membrane protein topology, assembly and folding. In the second part, available cell-based and cell-free membrane protein expression systems are described.

1.1. General information on membrane proteins

1.1.1. Monotopic versus polytopic membrane proteins

Membrane proteins can be classified into two groups according to their topology, *i.e.* monotopic and polytopic membrane proteins (Blobel, 1980).

Monotopic membrane proteins

Monotopic membrane proteins interact with only one leaflet of the lipid bilayer and do not possess transmembrane spanning segments (Marcia *et al*, 2010). They can be hydrophobically associated with the membrane through either a hydrophobic loop, lipidation, electrostatic binding, or an in-plane α-helix like HCV core (Boulant *et al*, 2006) and NS5A (Brass *et al*, 2002; Penin *et al*, 2004a) proteins. Monotopic membrane proteins can be classified into two groups : amphitropic proteins and integral monotopic membrane proteins. The former bind to the membrane transiently, and are either soluble or dissociable from the membrane by chaotropic agents (Johnson & Cornell, 1999). The latter bind to the membrane constitutively, they can not be isolated in a water-soluble form and require detergent for solubilization (Lomize *et al*, 2007). Integral monotopic membrane proteins are involved in various physiological processes, including reactions with membrane-embedded substrates, interaction with other membrane components, assembly of transmembrane complexes, connection between membrane and cytoskeleton, endocytosis and other processes of membrane remodelling (Johnson & Cornell, 1999; Marcia *et al*, 2010).

Polytopic membrane proteins

Polytopic membrane proteins are characterized by at least one transmembrane spanning segment, and are therefore also called transmembrane proteins. Proteins crossing the membrane only once are called bitopic proteins. Polytopic membrane proteins are integral membrane proteins, *i.e.* their release from the membrane requires the disruption of the lipid bilayer. These proteins fall into two major structural classes, *i.e.* the first class consists of individual or bundled transmembrane α-helices and the second forms monomeric, dimeric or trimeric transmembrane β-barrels (Tamm *et al*, 2001). All membrane proteins of plasma and endoplasmic reticulum-derived membranes are α-helical, whereas the proteins of the outer membranes (OMs) of Gram-negative bacteria and likely a fair number of proteins of the OMs of mitochondria and chloroplasts are of the β-barrel type (Tamm *et al*, 2001).

1.1.2. Translocon and membrane protein assembly

In mammalian cells, protein targeting to the endoplasmic reticulum (ER) is mostly cotranslational (Walter & Johnson, 1994). Synchronization of translation with membrane targeting and insertion indeed prevents the membrane protein from premature folding or aggregation in the cytosol and also determines its topology (Goder & Spiess, 2001). The different steps of membrane protein assembly are summarized in **Figure II.1**. Translocon complexes, also called protein-conducting channels, work in concert with bound ribosomes to insert proteins into membranes during the first step of membrane protein assembly (Johnson & van Waes, 1999; White & Heijne, 2004).

The mammalian translocon consists of three to four copies of the heterotrimeric Sec61 αβγ complex forming a gated pore (Johnson & van Waes, 1999). When a hydrophobic signal sequence emerges from the ribosome, it is recognized by a signal recognition particle (SRP) in the context of the nascent chain-ribosome. This complex is targeted to the ER membrane by binding to the SRP receptor (SR), both SRP and SR being GTPases. The signal sequence then interacts with the Sec61α subunit of the translocon and initiates translocation of the polypeptide (Mothes *et al*, 1998; Goder & Spiess, 2001). Another component of the mammalian translocon is the translocation-associated membrane protein (or TRAM) which was shown to be in contact with signal sequences

and transmembrane segments early during translocation and to be required for efficient insertion of most, but not all, proteins (Johnson & van Waes, 1999; Goder & Spiess, 2001).

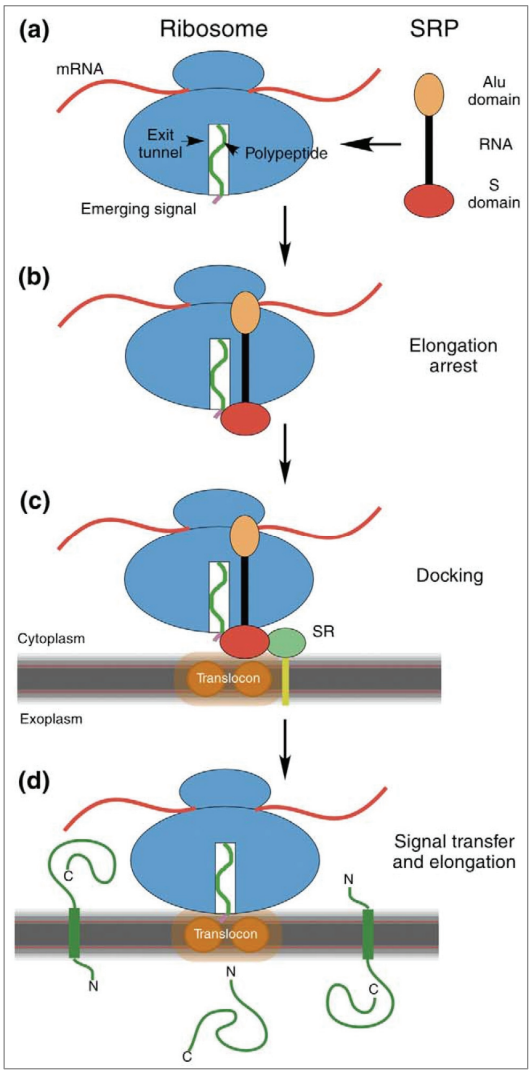


Figure II.1 ***The machinery of membrane protein assembly.*** From (White & Heijne, 2004). (a) A ribosome translating the mRNA of a protein targeted for export (translocation) across or insertion into membranes. A signal recognition particle (SRP), which is a GTPase, can bind to the ribosome and thereby arrest elongation. (b) The emerging signal (*purple*) is recognized by the ribosome and the SRP which bind to each other, causing the arrest of elongation. (c) The ribosome-SRP complex binds to the membrane-bound SRP receptor (SR, *green*), which is another GTPase and associates dynamically with the translocon (*orange*). The binding of the SRP to the SR causes the reciprocal stimulation of their GTPase activities, leading to the disengagement of the SRP from the SR and the ribosome. The nascent protein signal is then transferred to the translocon and elongation of the nascent peptide resumes. (d) Proteins targeted for translocation are secreted into the ER lumen, whereas the stop-transfer signals of membrane proteins are transferred to the membrane bilayer. The topology of the membrane protein is determined by the direction of insertion of the signal sequence N-terminus across the membrane.

1.1.3. Topogenesis of single- and multi-spanning membrane proteins

Single-spanning membrane proteins

Single-spanning membrane proteins present a final topology either with a cytoplasmic N- and an exoplasmic C-terminus (N_{cyt}/C_{exo}) or with an exoplasmic N- and a cytoplasmic C-terminus (N_{exo}/C_{cyt}). Single-spanning membrane proteins can be distinguished in three types, all inserted by the same machinery involving SRP, SR and the Sec61 translocon, and called type I, type II and type III (**Figure II.2**). With respect to topogenesis, there are two basic types of translocating signals, *i.e.* C-terminus-translocating signals (cleaved signals and signal-anchor) and N-terminus-translocating signals (reverse signal-anchor) (Goder & Spiess, 2001).

Type I membrane proteins are initially targeted to the ER by an N-terminal cleavable signal sequence which is a hydrophobic stretch of typically 7-15 predominantely apolar amino acid residues, and then anchored in the membrane by a subsequent stop-transfer sequence which is a segment of ~20 hydrophobic amino acid residues interrupting the further translocation of the polypeptide and acting as a transmembrane anchor (Goder & Spiess, 2001). This is for example the case of HCV E1 and E2 glycoproteins (Vieyres *et al*, 2014). Type II membrane proteins are both inserted and anchored by a signal-anchor sequence. Signal-anchor sequences span the lipid bilayer as a transmembrane helix and are generally longer than cleavable signals (~18-25 mostly apolar amino acid residues). Whereas they induce the translocation of their C-terminal end across the membrane like cleaved signals do, they lack a signal peptidase cleavage site and can be positioned internally within the polypeptide chain (Goder & Spiess, 2001). In contrast, type III membrane proteins are both inserted and anchored by a reverse signal-anchor which translocates their N-terminal end across the membrane (Goder & Spiess, 2001).

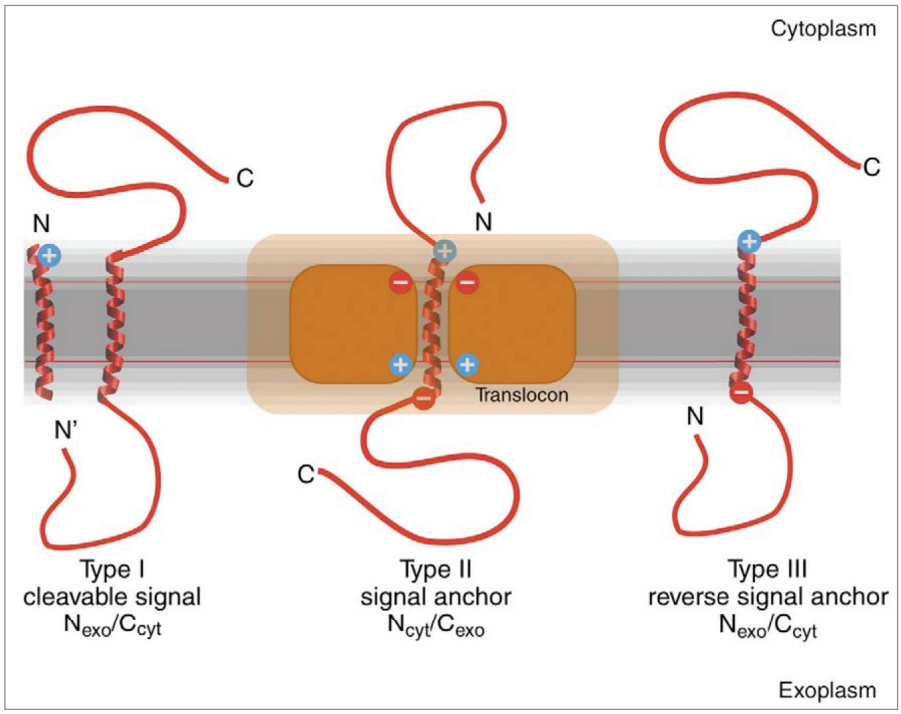


Figure II.2 ***Types of single TM helix MPs.*** From (White & Heijne, 2004)**.** The location of the N-terminus and the cleavability of the signal define three types of membrane proteins. A run of membrane-spanning hydrophobic residues flanked by charged residues generally constitutes the signals. In eukaryotes, the topology of the signal is determined by the balance of flanking charges. The flank with the greater positive charge is generally on the cytoplasmic side of the membrane.

There is an additional class of proteins that are predominantly exposed to the cytosol and anchored to the membrane by a very C-terminal signal sequence. Since the signal emerges from the ribosome only after translation has reached the stop codon, insertion of these proteins is necessarily posttranslational, targeting and insertion occuring in a SRP- and Sec61-independent manner (Goder & Spiess, 2001). This is for example the case of the HCV NS5B protein (Schmidt-Mende *et al*, 2001).

Multi-spanning membrane proteins

It is generally assumed that the first hydrophobic sequence is responsible for targeting the nascent protein to the ER and to initiate translocation and membrane insertion. Accordingly, these proteins could be classified as multi-spanning proteins of type I, type II or type III based on whether the most N-terminal apolar sequence is cleaved by signal peptidase or spans the membrane with an N_{cyt}/C_{exo} or N_{exo}/C_{cyt} orientation, respectively (Spiess, 1995; Goder & Spiess, 2001). The HCV p7 and NS2 proteins are multi-spanning membrane proteins of type III (Carrere-Kremer *et al*, 2002).

1.1.4. Folding of polytopic membrane proteins

Membrane secondary and tertiary structures can form at different stages, *i.e.* in the ribosome exit tunnel (Lu & Deutsch, 2005b), the ribosome exit vestibule (Tu *et al*, 2014), within the translocon channel (Lu & Deutsch, 2005a; Woolhead *et al*, 2006; Daniel *et al*, 2008), during membrane insertion (Ulmschneider *et al*, 2011; Ulmschneider *et al*, 2014) and within a membrane. For example, the stage of development of an α-helical structure can be important for the formation of tertiary structure because amino acid side chains are positioned on different faces of an α-helix, thereby allowing favorable alignment with residues in other transmembrane segments to form tertiary structure. Interactions between individual transmembranes helices determine the final fold and function of the mature protein (Cymer *et al*, 2014). They can consist of distinct interaction motifs such as the glycine zipper motif GxxxG. More generally, small residues separated by three residues (small-XXX-small) represent a prominent motif allowing transmembrane helices to approach each other and to pack tightly due to van der Waals attractions and steric constraints. Extensions of this motif (small-XXX-small-XXX-small) allows for flexibility of two helices since multiple small residues spaced this way can form a groove on the surface of α-helices (Kim *et al*, 2005; Cymer *et al*, 2014). Although such motifs are highly

abundant in membrane proteins, many interactions can not be explained by them (Li *et al*, 2012a). Indeed, many other types of interactions requiring more specificity can be formed between transmembrane helices, interactions with multiple transmembrane helices at the same time or a high degree of flexibility (Cymer *et al*, 2014).

Moreover, the structural stability of membrane proteins depends upon numerous physicochemical interactions which are summarized in a simplistic way in **Figure II.3** (Cymer *et al*, 2014).

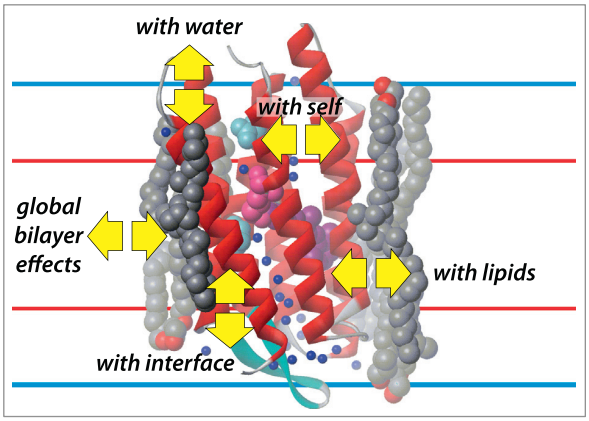


Figure II.3 ***Summary of various interactions stabilizing membrane proteins stably folded in fluid lipid bilayers.*** From (Cymer *et al*, 2014). The protein shown (*red helices*) is bR, determined to a resolution of 1.55 Å (PDB 1C3W). Interface boudaries are represented by *blue lines* and boundaries of the lipid hydrocarbon core by *red lines*. When the protein perturbs the lipid bilayer, « global bilayer effects » account for changes in the structure and stability of the bilayer. Both the tendency of the system to minimize exposure of the acyl chains (*grey*) to water (*blue*) and the tendency to maximize the distance between headgroups determine the free energy minimum of the bilayer. In order to minimize the free energy of the protein/bilayer complex, both the bilayer and the protein must adjust structurally.

1.2. Membrane protein expression systems

1.2.1. Cell-based systems

Expression systems based on bacteria, mammalian cells, insect cells and yeasts are described in this section. A comparison of these systems is given below in **Table II.1**. Although these cell-based expression systems are the most commonly used to produce membrane proteins, this list is not comprehensive and alternative hosts including various bacteria, archae, filamentous fungi, protozoa (*e.g. Leishmania tarentolae*) and plants (*e.g. Arabidopsis taliana*) are available, as described for example in (Bernaudat *et al*, 2011) and (Fernández & Vega, 2013).

Bacteria

- *Escherichia coli*

E. coli, a Gram-negative bacterium, is the most widely used expression host for the production of recombinant proteins. Indeed, more than 85% of all proteins in the PDB

with non-empty descriptor for « expression host » were produced in *E. coli* (Fernández & Vega, 2013). *E. coli* has been extensively characterized, is inexpensive and easy to use with a short generation time (Sahdev *et al*, 2008). However, the production of functional eukaryotic membrane proteins in this system is often limited by toxicity, formation of inclusion bodies containing not properly folded and aggregated proteins, low levels of expression and the lack of most of the eukaryotic posttranslational modifications (Sahdev *et al*, 2008). Changing the media composition, the carbon source or the growth temperature (Quick & Wright, 2002), but also using codon optimized constructs, different promoters or fusion partners (Weiss & Grisshammer, 2002; White *et al*, 2004), or even adding N-terminal signal peptides (Grisshammer *et al*, 1993) allow to some extent to overcome these limitations. A drawback however is the absence of cholesterol in bacterial membranes.

The strain *E. coli* BL21(DE3) is one of the most commonly used. This strain has a T7 expression system and protein production is induced by IPTG. However, uninduced expression has been observed, leading to cell death already at the transformation step if the gene product is toxic. The strains C41(DE3) and C43(DE3) are BL21(DE3) mutants which allow the expression of such proteins (Miroux & Walker, 1996). Based on the observation that these two strains carry mutations within the T7 promoter leading to a weaker T7 expression, the strain Lemo21, tunable to overexpress membrane proteins, has been developed (Wagner *et al*, 2008). Membrane protein production in *E. coli* can be improved by the introduction of various tags, such as GFP (Drew *et al*, 2001), MBP, GST, NusA (Junge *et al*, 2008) or Mistic (Roosild *et al*, 2005). The latter spontaneously associates with the inner membrane of *E. coli*, independently from the translocon machinery. Used as an N-terminal tag, Mistic, which is a 13 kDa protein from *Bacillus subtilis*, targets the proteins of interest to the membrane and facilitates their insertion (Roosild *et al*, 2005; Dvir & Choe, 2009; Nekrasova *et al*, 2010). In addition, in our institute, Pierre Falson and François Penin showed that the insertion of the Asp-Pro (DP) sequence dramatically reduced the toxicity of membrane proteins in *E. coli* and promoted their expression when produced as gluthatione S-transferase (GST) GST-DP-TM chimeras (Montigny *et al*, 2004).

- *Lactococcus lactis*

L. lactis is a Gram-positive lactic acid bacterium which can also be used for large scale production of membrane proteins (Kunji *et al*, 2005). *L. lactis* grows to high cell densities without the need for aeration. Protein expression in *L. lactis* is induced by nisin, an antimicrobial peptide, using the Nisin-Inducible Controlled gene Expression (NICE) system (de Ruyter *et al*, 1996; Zhou *et al*, 2006). This nisin-inducible promoter is a strong and tightly regulated promoter system allowing highly reproducible protein expression (Kunji *et al*, 2005). So far, no inclusion bodies have been observed in *L. lactis*, which is a major advantage over *E. coli* (Kunji *et al*, 2003). In addition, *L. lactis* has a single cell membrane allowing the direct use of ligands or inhibitors to study the activity of membrane proteins in whole cells (Bernaudat *et al*, 2011). The human KDEL receptor, the human hydrogenosomal ADP/ATP carrier and the human CFTR, a cystic fibrosis transmembrane conductance regulator, are examples of eukaryotic membrane proteins produced in *L. lactis* (Kunji *et al*, 2003; Steen *et al*, 2011).

Mammalian cells

Mammalian cells provide a near-native environment for overexpression of mammalian membrane proteins, with correct posttranslational modifications, translocation machinery and lipid environment. Human embryonic kidney cells (HEK293), Chinese hamster ovary cells (CHO), baby hamster kidney cells (BHK-21) and monkey kidney fibroblast cells (COS-7) are the most used cell lines for mammalian membrane protein expression (Andréll & Tate, 2013).

Expression of membrane proteins can be either transient or stable. Transient expression systems allow protein production within a few days whereas development of stable cell lines takes months (Andréll & Tate, 2013). Transient expression systems may rely on using electroporation or/and cationic compounds such as polyethylenimine (PEI) to facilitate the uptake of plasmid DNA into the cell. A successfull example of transient tranfection with a chemical reagent is the production of a thermally stable rhodopsin mutant whose structure was determined (Standfuss *et al*, 2007). Transient expression systems may also rely on recombinant non-replicative viruses that have evolved to efficiently enter mammalian cells (Andréll & Tate, 2013). An example is the Semliki Forest Virus (SFV) which has been used extensively to overexpress integral membrane proteins

such as G protein-coupled receptors (GPCRs) (Hassaine *et al*, 2006). Another example is the BacMan expression system in which insect cell viruses mediate gene transduction of mammalian cells (Dukkipati *et al*, 2008; Goehring *et al*, 2014).

Stable expression systems consist in transfection of cells followed by antibiotic selection of cells expressing stably the protein of interest under prolonged culture (Chaudhary *et al*, 2012). Membrane proteins can be expressed in such a system using either a constitutive or an inducible promoter (Andréll & Tate, 2013). Inducible promoters, like the tetracycline promoter (Yao *et al*, 1998), are usually prefered since they allow to produce 4- to 12-fold more membrane protein than constitutive promoters. Moreover, strong constitutive promoters often lead to the loss of expression under prolonged culture (Andréll & Tate, 2013). Two examples of membrane proteins successfully produced in stable cell lines using an inducible promoter are the rat serotonin transporter rSERT (Tate *et al*, 2003) and the human ammonia transporter RhCG (Chaudhary *et al*, 2012). Furthermore, the recently reported X-ray structures of HCV E1 and E2 glycoproteins have been done with protein transiently produced in HEK293 cells (Kong *et al*, 2013; Khan *et al*, 2014; Omari *et al*, 2014).

Insect cells

Insect cells, as a compromise between bacteria and mammalian cells, are widely used to express eukaryotic proteins. Indeed, more than 4.5% of all proteins in the PDB with non-empty descriptor for « expression host » were produced in insect cells (Fernández & Vega, 2013). Altough this system is more expensive and time-consuming than bacteria, it is easier and cheaper to handle than mammalian cells. Moreover, this system presents many advantages for eukaryotic proteins, like similar codon usage rules, better expression levels and fewer truncated proteins than in bacteria, as well as posttranslational modifications. However, posttranslational modifications are not all identical, like for example glycosylation. Indeed *N*-glycans of glycoproteins produced in insect cells are more uniform and less complex, which may affect the biological function of recombinant proteins (van Oers *et al*, 2014). Recent improvements regarding to glycoprotein expression are reviewed in (van Oers *et al*, 2014).

Commonly used insect cell lines are the Sf9, Sf21 and Schneider 2 (S2) cells. Protein expression is mediated either by transient transfection or using the baculovirus system.

Transient transfection of insect cells is performed in a similar way than for mammalian cells (described above). A protocol for « virus-free » transient transfection of suspension-adapted Sf9 insect cells has been reported recently (Shen *et al*, 2014). Two additional examples for transient transfection are the human ion channel ASIC3 and the human heterodimeric transporter SLC7A5/SLC3A2. Both were produced in Sf9 insect cells transfected using a commercial transfection reagent (Chen *et al*, 2013).

The baculovirus system relies on the infection of insect cells by recombinant viruses that encode the proteins of interest (Trometer & Falson, 2009; Contreras-Gómez *et al*, 2013; Hu *et al*, 2014). Promoters usually used are the polyhedrin promoter and the p10 late promoter (Bernaudat *et al*, 2011). As an example, the human membrane transporter protein OATP2B1 was produced in Sf9 insect cells using the baculovirus system (Tschantz *et al*, 2008). In addition, the crystal structures of the human β-2 adrenergic GPCR (Cherezov *et al*, 2007) and of two human ion channels (Jasti *et al*, 2007; Kawate *et al*, 2009) were determined after expression of these membrane proteins in Sf9 insect cells using the baculovirus system.

<u>*Yeasts*</u>

- *Saccharomyces cerevisiae*

S. cerevisiae is the famous baker's yeast. This budding yeast is a well established system for the production of eukaryotic integral membrane proteins (Bill, 2001; Griffith *et al*, 2003; Li *et al*, 2009; Hays *et al*, 2010). Its genome is sequenced and its mechanisms of protein quality control are well understood (Griffith *et al*, 2003; Hays *et al*, 2009). *S. cerevisiae* presents many advantages for the production of membrane proteins, such as proper membrane targeting and insertion machinery, posttranslational modifications and lower cost compared to mammalian or insect cells. In addition, *S. cerevisiae* is as easy to handle as bacteria (Hays *et al*, 2009).

The unfolded protein response (UPR) pathway of *S. cerevisiae*, which is an adpatative cellular mechanism monitoring the folding state of nascent proteins in the lumen of the ER, can be exploited to improve the expression of membrane proteins reaching their final cellular destination via the secretory pathway (Griffith *et al*, 2003). Moreover, the production of large quantities of eukaryotic membrane proteins in *S. cerevisiae* using a cleavable GFP-8His tag fused at the protein C-terminus has been recently described for

biochemical studies and determination of the crystal structure (Scharff-Poulsen & Pedersen, 2013). In addition, co-expression of chaperones in *S. cerevisiae* has also been described, for example for the rat MCT1 monocarboxylate transporter (Makuc *et al*, 2004), and chemical chaperones can also be used (Figler *et al*, 2000). High-throughput expression platforms have also been described (Newstead *et al*, 2007). However, the yeast *Pichia pastoris*, which is described below, is often prefered to *S. cerevisiae* (Bill, 2001).

- *Pichia pastoris*

Pichia pastoris is a methylotrophic yeast, reclassified as *Komagataella pastoris.* This microbial eukaryote is an established protein expression host mainly applied for the production of biopharmaceuticals and industrial enzymes (Ahmad *et al*, 2014). From a biological and technical point of view, *Pichia* is similar to the yeast *S. cerevisiae* (Higgins, 2001). It is as easy to manipulate as bacteria and has many of the advantages of eukaryotic expression, *i.e.* protein processing, folding and posttranslational modifications. Moreover, *Pichia* is faster, easier and cheaper to use than other eukaryotic expression systems such as baculovirus or mammalian tissue culture, and generally gives higher expression levels (Higgins, 2001).

With its strong AOX1 promoter and its capacity to grow at high density, *Pichia* stands as one of the most efficient hosts for mass production of eukaryotic membrane proteins (Bornert *et al*, 2001; Ramón & Marín, 2011; Goncalves, 2013). It has been applied routinely to produce affinity-tagged membrane proteins for purification and subsequent biochemical studies (Cohen *et al*, 2005; Lifshitz *et al*, 2007; Haviv *et al*, 2007). *Pichia* has also been the expression host of choice for elucidating the crystal structures of membrane proteins from diverse origins, even from higher eukaryotes, like for example the human aquaporin 4 (Ho *et al*, 2009), the human K2P TRAAK, a lipid- and mechano-sensitive K^+ ion channel (Brohawn *et al*, 2012) or the human A_{2A} adenosine receptor ($A_{2A}AR$) which is a G protein-coupled receptor (GPCR) (Hino *et al*, 2012).

1.2.2. Cell-free systems

Although some eukaryotic membrane proteins could be successfully expressed in cell-based expression systems, as described above, the production of such proteins is still challenging. Overexpression is often impaired by toxicity of the target proteins, and when

the protein of interest is expressed, its extraction from cellular membranes with detergent and its further purification in a functional form are notoriously known to be difficult. Cell-free expression systems allow to overcome these limitations. There are three major cell-free expression systems based on *E. coli* lysates, rabbit reticulocyte lysates (RRLs) and wheat germ extracts (WGEs) (Spirin, 2004). A comparison of these systems is given below in **Table II.1**. However, further cell-free expression systems are available, *e.g.* lysates from insect cells (Ezure *et al*, 2006; Sakurai *et al*, 2007), mammalian cells (HeLa) (Witherell, 2001; Mikami *et al*, 2006), Tobacco BY-2 lysates (Buntru *et al*, 2014), yeast (*S. cerevisiae*) (Edelmann *et al*, 2000; Wang *et al*, 2008; Wang *et al*, 2010; Gan & Jewett, 2014; Russ & Dueber, 2014) and *Leishmania tarentolae* (Kovtun *et al*, 2010; Kovtun *et al*, 2011).

E. coli lysates

This well-established system is probably the most popular. The *E. coli* S30 extract, obtained by centrifugation of lysed cells at 30,000 *g*, includes most enzymes required for protein synthesis (Zubay, 1973; Kigawa *et al*, 2004). To avoid background expression, endogenous mRNA and low molecular weight compounds are eliminated during the extract preparation. There are established protocols for protein expression using this system (Torizawa *et al*, 2004; Schwarz *et al*, 2007), but many commercial kits are also available. Reactions with *E. coli* S30 extracts are usually set up as coupled transcription/translation systems, and both plasmid DNA and linear DNA from PCR reactions can be used (Schwarz *et al*, 2010). High-throughput protein production is possible (Spirin, 2004; Aoki *et al*, 2009). The major drawback of this system is the lack of posttranslational modifications and specific eukaryotic folding systems (Schwarz *et al*, 2010). As other cell-free expression systems, *E. coli* S30 extracts allow selective labeling strategies for NMR studies (Vinarov *et al*, 2004; Schwarz *et al*, 2010). The production of proteins labeled with stable isotopes has been described for various protein targets (Vinarov *et al*, 2004; Abdine *et al*, 2010; Maslennikov *et al*, 2010; Abdine *et al*, 2011), including membrane proteins (Klammt *et al*, 2004; Miot & Betton, 2011).

An alternative to *E. coli* lysates is a minimal synthesis system which consists of a set of purified elements required for the translation reaction. This system is called PURE, for « protein synthesis using recombinant elements » (Shimizu *et al*, 2001).

Rabbit reticulocyte lysates

RRLs are a old and well-established system (Craig *et al*, 1992). The preparation of RRLs, from the blood of rabbits that have to be made anemic before, has been described more than thirty years ago (Pelham & Jackson, 1976; Jackson & Hunt, 1982). Commercial kits are available. Although this system provides more closely native conditions for the expression of eukaryotic proteins, especially posttranslational modifications, it is expensive and difficult to scale up. There are further limitations : since it is optimized for globin synthesis, the codon usage of RRLs is stongly biaised ; the amount of endogenous endonuclease is quite high in the lysate, which can be problematic ; and the high endogenous content of hemoglobin can be a problem during purification (Endo & Sawasaki, 2006; Schwarz *et al*, 2008). The major drawback of this system is the very low protein yield which does not allow structural studies. However, when glycosylation is required, RRLs seem to be the most suitable system to couple translation with glycosylation by adding microsomes, *e.g.* dog pancreas microsomes (Ghadessy & Holliger, 2004; Spirin, 2004). In addition, it was recently shown that target protein production could be increased over 10-fold by adding purified NS1 protein of influenza virus to the system and targeting mRNA with the encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES) (Anastasina *et al*, 2014).

Wheat germ extracts

The eukaryotic wheat germ cell-free (WGE-CF) system offers an interesting alternative to *E. coli* lysates and RRLs, combining posttranslation modifications with higher protein yields. Wheat germ embryos are surrounded by the endosperm which contains high levels of enzymes degrading proteins and nucleic acids, as well as translational inhibitors. This makes the WGE production difficult, but commercial kits are available (Spirin, 2004; Schwarz *et al*, 2008) and state-of-the-art protocols have been published (Takai *et al*, 2010). The WGE-CF system is described in detail in section II.2. The expression of membrane proteins using this system is described in section II.3.

2. Wheat germ cell-free expression system

2.1. History

The wheat germ cell-free system consists in using the translation machinery contained in wheat embryos (**Figure II.4**) to synthezise protein *in vitro*. This system is in fact quite old. Already 1973, both tobacco mosaic virus RNA and rabbit globin S9 RNA were efficiently translated using commercial wheat germ (Roberts & Paterson, 1973). However, this system has not been very popular due to the instability of WGEs (Roberts & Paterson, 1973; Marcu & Dudock, 1974; Scheele & Blackburn, 1979). About fifteen years ago, Yaeta Endo and colleagues from the Ehime University in Japan demonstrated that WGEs were contaminated by inhibitors during the extraction procedure, *e.g.* tritin (RNA N-glycosydase), thionins (small basic and cysteine rich proteins), ribonucleases, desoxyribonucleases (Madin *et al*, 2000), ribosomal RNA apurinic site specific lyase (Ogasawara *et al*, 1999), which were responsible for the WGE instability. It was shown that these inhibitors, which come from the endosperm (**Figure II.4**), could be eliminated by extensive washing of the wheat embryos before extract preparation, leading to an efficient and stable translation system (Madin *et al*, 2000). Further improvments were done few years later, especially with the use of the bilayer method (described below) (Sawasaki *et al*, 2002a). In addition, a high-throughput synthesis system based on PCR-generated DNA templates was developed (Sawasaki *et al*, 2002b). The applicability of the WGE-CF system for protein production to high-throughput structural analysis of proteins, especially by NMR, was then demonstrated (Morita *et al*, 2003; Morita *et al*, 2004). Additional work showed that proteins produced in the WGE-CF system are functional (Sawasaki *et al*, 2004). Moreover, Yaeta Endo and colleagues founded a company named CellFree Sciences which commercializes WGEs and further products related to protein expression using this system (http://www.cfsciences.com).

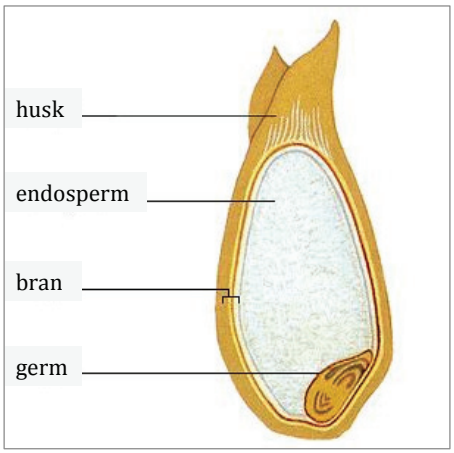


Figure II.4 ***Schematic representation of a wheat seed.*** Adapted from www.threes.com. The composition of a wheat seed can be simplistically described in four main components, *i.e.* the husk, the bran, the endosperm and the germ. The husk and the bran constitute the envelopes of the seed. They surround the endosperm which mainly contains starch and gluten and from which the flour is prepared. The endosperm contains inhibitors of protein synthesis and has to be carefully discarded during WGE preparation. The germ, or embryo, is the « living part » of the seed and contains the translation machinery, as well as lipids and vitamins. WGEs are prepared exclusively from the germ.

2.2. The wheat germ system

2.2.1. The wheat germ extracts

All the components for translation, *e.g.* ribosomes, tRNAs, aminoacyl-tRNA synthetases, initiation, elongation and termination factors, are found in the wheat embryos in a concentrated dried state and ready for protein synthesis after germination. A schematic representation of a wheat seed is shown in **Figure II.4**. Non treated durum wheat seeds, *i.e.* without pesticides or insecticides, are used for WGE preparation whose a detailed protocol has been published by Yaeta Endo and colleagues (Takai *et al*, 2010). As described above in section 2.1., the major point is the need to wash the wheat embryos extensively in order to remove residual endosperm containing inhibitors of protein synthesis (Madin *et al*, 2000). The WGE preparation takes about 4 days and despite the high cost of commercial WGEs, still today only few laboratories implemented in-house preparation. Since some proteins from the WGEs bind unspecifically to glutathione or metal-chelating resins, there are even commercial WGEs pretreated on such resins. These pretreatments lead to a higher purity of proteins with a His- or GST- affinity tag (Takai *et al*, 2010; Harbers, 2014). Proteins from the WGEs do not interfere with *Strep*-Tactin resins and no WGE pretreatment is required when proteins are expressed with a *Strep*-tag II (Schmidt & Skerra, 2007).

2.2.2. The pEU expression vector

The pEU vector is composed of a SP6 promoter, a translational enhancer and a multiple cloning site (**Figure II.5**). The translation enhancer E01 was optimized starting from the Ω sequence from tobacco mosaic virus, one of the most frequently used mRNA 5' leader sequences enhancing *in vivo* and *in vitro* translation (Kamura *et al*, 2005). Both the 5'-^{7}mGpppG (cap) and poly(A)-tail were eliminated to increase the translation initiation and the stability of the mRNA template (Sawasaki *et al*, 2002b; Endo & Sawasaki, 2006).

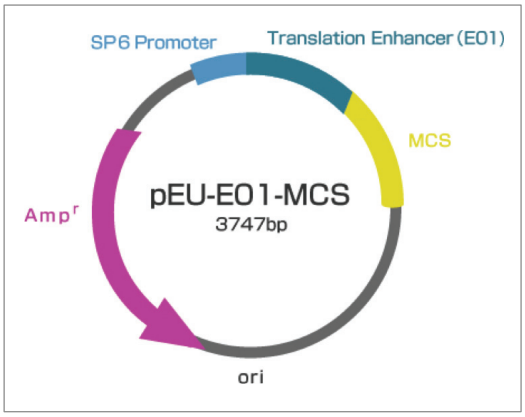


Figure II.5 ***Schematic representation of the pEU vector.*** From www.cfsciences.com. The SP6 promoter is represented in *blue*, the translation enhancer E01 in *dark blue* and the multiple cloning site (MCS) in *yellow*. The pEU vector also contains an ampicilline resistance gene (Amp^{r}) represented in *pink*.

2.2.3. Transcription from circular or linear DNA

Transcription can be performed starting from a plasmid containing the gene of interest. The DNA template can be used either directly in its circular form or linearized before transcription (Noirot *et al*, 2010; Takai *et al*, 2010). Although some commercial kit manufacturers recommend using linear DNA, both linear and circular DNA templates are equally effective and working with circular DNA is time and cost saving (Harbers, 2014).

A method for the preparation of PCR fragments which can be directly used for transcription has been described, with the aim to increase the throughput of the system (Takai *et al*, 2010). Effective translation requires neither 5'-capping nor a poly(A)-tail, but efficient translation depends on the length of the 3'-untranslated region which protects the RNA against degradation by 3' to 5' exonucleases (1,500 nucleotides are recommended). A 5'-enhancer sequence and a long 3' sequence were therefore optimized, allowing the amplification of many different cDNA clones in a two-step PCR method called the « split-primer » PCR (Sawasaki *et al*, 2002b; Takai *et al*, 2010). This method is not required if the cDNA clones are inserted into the pEU vector described above.

2.2.4. Transcription and translation : coupled or uncoupled

Transcription and translation can be performed either coupled in a single reaction step or separately. In the latter case, mRNA is prepared first and then added to the WGE before the translation step. Although coupled transcription/translation has been described for the WGE-CF system (Stueber *et al*, 1984), these reactions are usually done separately (Sawasaki *et al*, 2002a; Endo & Sawasaki, 2004; Endo & Sawasaki, 2006; Takai *et al*, 2010). Uncoupling transcription and translation allows for more flexibility to work under optimal reaction conditions (*e.g.* temperature, Mg^{2+} concentration), to use additives in the translation reaction without interfering with transcription, and to better identify and solve problems when they occur. In contrast, commercial kits for cell-free protein synthesis usually consist in coupled transcription/translation (so-called TNT® systems). Both methods can be applied to the different translation modes described below.

2.2.5. Translation modes

The main components required for translation are mRNA, WGE and a feeding buffer containing all substrates necessary for protein synthesis. There are four different modes

of translation : the batch mode (Kawasaki *et al*, 2003), the continuous-exchange system (CECF) (Katzen *et al*, 2005), the continuous-flow system (CFCF) (Spirin *et al*, 1988) and the bilayer method (Sawasaki *et al*, 2002a).

The batch reaction in which all reagents are mixed in a single compartment is the least complicated. However, mainly due to the accumulation of inhibitory byproducts in the single reaction compartment, the system works for only a few hours (Schwarz *et al*, 2008). Moreover, the amount of synthezised protein is usually not sufficient for biochemical analysis (Sawasaki *et al*, 2002a). The batch mode is however ideal for small scale high-throughput expression screening (Sawasaki *et al*, 2002b; Schwarz *et al*, 2008). In addition, an alternative to the batch mode is the so-called repeat-batch (or discontinuous batch) mode (Harbers, 2014). After a first incubation time, the batch reaction is concentrated on a dialysis membrane by centrifugation in order to remove the small molecular weight inhibitors from the translation reaction, and fresh reaction buffer is added to provide fresh substrates to the translation reaction. This process can be repeated several times and leads to better protein yields than the standard batch mode (Harbers, 2014).

The continuous-exchange cell-free protein synthesis system or CECF is in fact a dialysis system (Katzen *et al*, 2005), the translation mix being in the dialysis device and the dialysis buffer containing fresh substrates for translation. Mini- and maxi-CECF reactors as well as further CECF reactor designs are described with many details in (Schneider *et al*, 2009).

The continuous-flow cell-free protein synthesis system or CFCF was first described by Spirin (Spirin *et al*, 1988; Spirin, 2004). Similar to the CECF mode, this system provides an automated and continuous supply of substrates and removal of byproducts, the translation reaction proceeding for more than two days (*i.e.* more than ten times longer than the bacth mode) and yielding up to several milligrams of protein (Spirin *et al*, 1988; Morita *et al*, 2003). Although very efficient, this system is mechanically complicated and not suitable for high-throughput screening of proteins (Sawasaki *et al*, 2002a). Moreover, this system is not suitable for high molecular weigth proteins over 300 kDa (Spirin, 2004).

The bilayer method is a simplified version of CECF (Sawasaki *et al*, 2002a). Less expensive than CECF, it allows to synthezise higher protein amounts than the batch mode, yielding amounts of proteins sufficient for functional and even structural analysis. The substrate buffer is overlaid onto the translation mix, forming two separate layers which allow a diffusion controlled translation process (Sawasaki *et al*, 2002a; Takai *et al*, 2010). The bilayer method can be fully automated for large and efficient screening (Endo & Sawasaki, 2004). Moreover, its flexible format allows to screen different additives for the translation reaction, *e.g.* detergents or lipids for the expression of membrane proteins in a soluble form (Harbers, 2014), and the scale-up from 96-well plates to 6-well plates is easy. Over 13,000 human cDNA clones were already tested using the bilayer method (Goshima *et al*, 2008). Since it is easier to handle and less expensive than the CECF and CFCF modes but much more efficient than the batch mode, we implemented the bilayer method in the laboratory (refer to section III.4.2.1. for technical details).

2.2.6. Additives

Magnesium (Mg^{2+}) and potassium (K^{+}) ion concentrations have been shown to be critical and can be optimized for each protein. In addition, ATP and GTP provide the energy for mRNA translation. Since they are rapidly consumed, creatine phosphate, a high-energy phosphate donor, is also supplied to the translation mixture together with creatine kinase to recycle ADP and GDP to ATP and GTP (Torizawa *et al*, 2004). Since the excess of phosphate generated by the breakdown of energy sources has been identified as an inhibitor of translation, modified energy regeneration pathways have been proposed (Schwarz *et al*, 2008). Moreover, cofactors can be added, as described for example for the FMN-binding protein from *Desulfovibrio vulgaris* which was synthesized either in its apo- or holo-form, in absence or presence of the cofactor FMN, respectively (Abe *et al*, 2007). Finally, detergents or lipids can be added to the reaction mixture for the expression of membrane proteins directly in a soluble state. This aspect is detailed in section II.3.1.

2.2.7. Posttranslational modifications

Some posttranslational modifications can occur « spontaneously » in the WGE-CF system (*e.g.* phosphorylation), whereas others require additives such as canine microsomal membranes (*e.g.* glycosylation). Phosphorylation consists in the addition of phosphate groups to serine, threonine or tyrosine residues, altering the biological activity

of many proteins. This important posttranslational modification has been observed in the WGE-CF system (Nakamura, 1993; Langland *et al*, 1996). Myristoylation which consists in the linkage of myristoylate groups to proteins, playing a role in membrane localization, has also been observed in the WGE-CF system (Heuckeroth *et al*, 1988; Martin *et al*, 1997; Yamauchi *et al*, 2010). In addition, ubiquitination has been reported to occur in the WGE-CF system (Takahashi *et al*, 2009). Farnesylation, which is a kind of isoprenylation (linkage of isoprenoid groups to cysteine residues), is also possible in the WGE-CF system (Kong *et al*, 2000). Moreover, the N-terminal methionine is eliminated (Kanno *et al*, 2007). Glycosylation consists in the addition of oligosaccharides either to the NH_2 group of asparagine (N-linked glycosylation) or to the OH group of serine, threonine or hydroxylysine (O-linked glycosylation). There is a lack of glycosylation in the WGE-CF system which can be overcome by the addition of microsomal membranes (Rubenstein & Chappell, 1983; Kottler *et al*, 1989). WGEs lack the SRP required for membrane protein translocation (refer to section II.1.1.2.) and signal peptide cleavage in the WGE-CF system can only occur when the system is supplemented with canine microsomal membranes (Jackson & Blobel, 1977). Finally, disulfide bond formation might be compromised by the presence of dithiothreitol (DTT) in the translation reaction, but is possible while lowering DTT concentration, although this can lead to lower expression yields (Kawasaki *et al*, 2003), and eventually adding a protein disulfide isomerase to the translation reaction (Harbers, 2014). No information could be found about posttranslational modifications such as palmytoylation or N-acetylation.

2.3. Isotopic labeling for NMR studies

Only the *in vitro* synthezised proteins coded by the pEU vector are isotopically labeled, which is a major advantage of the WGE-CF system. Isotopic labeling using this system is very efficient and specific (Morita *et al*, 2004; Kohno, 2009). Applicability of this system to high-throughput structural analysis by NMR has been demonstrated, showing that isotopically labeled and biologically active proteins, having the expected folding as compared to previously determined structures, can be produced (Morita *et al*, 2003; Vinarov *et al*, 2004; Vinarov *et al*, 2006; Makino *et al*, 2009). Moreover, Morita and colleagues showed that WGEs contain two transaminases and one glutamine synthase mediating the interconversion between Ala and Glu, Glu and Asp, and Glu and Gln. These

enzymes are active and it is therefore necessary to use inhibitors for Ala, Glu and Asp specific labeling (Morita *et al*, 2004). For Ala residues, β-chloroalanine can be used to inhibit the activity of the alanine transaminase whereas for Asp residues, the activity of the aspartate transaminase can be inhibited by aminooxyacetic acid. For Glu residues, the activity of the aspartate transaminase and the glutamine synthase can be inhibited by aminooxyacetic acid or L-methionine sulfoximine. These inhibitors do not affect protein synthesis using this system (Morita *et al*, 2004). In addition, metabolic scrambling can be problematic when deuteration is seeked. Indeed, proton back exchange can occur when proteins are prepared from deuterated amino acids in H_2O, leading to contamination of ^{2}H sites by ^{1}H from the solvent. The use of aminooxyacetic acid and L-methionine sulfoximine has been reported to partially inhibit such exchange reactions during the production of ubiquitin using a WGE-CF system (Tonelli *et al*, 2011).

As an example, amino acid selective labeling with the WGE-CF system was applied to the structural analysis by NMR of the β2-microglobulin, a major structural component of amyloid fibrils (Kameda *et al*, 2009).

3. Membrane protein expression using the wheat germ cell-free system

3.1. Expression modes

There are three modes of expression for membrane proteins using the WGE-CF system, *i.e.* the so-called precipitate mode, the expression in presence of detergents and the expression in presence of lipids (**Figure II.6**).

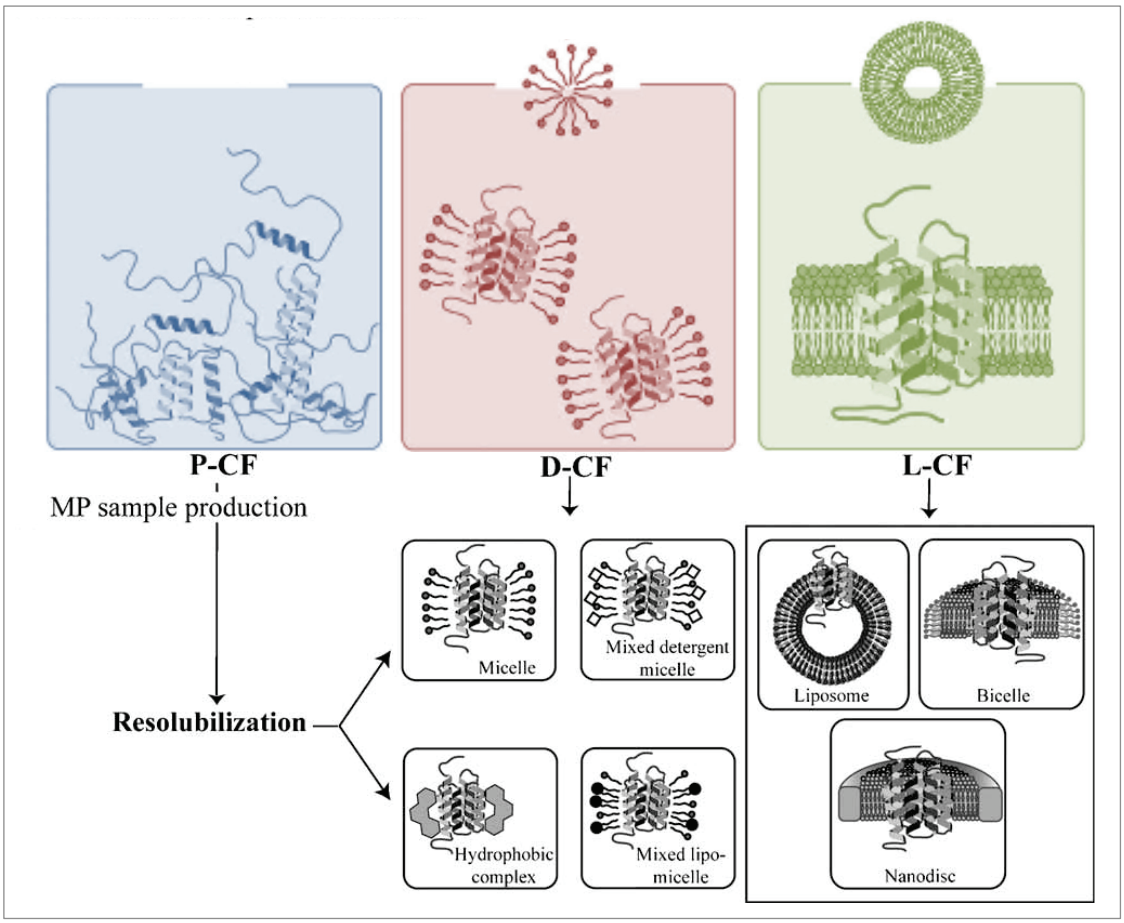


Figure II.6 ***Modes of expression for membrane proteins in cell-free systems.*** Adapted from (Junge *et al*, 2011). P-CF, precipitate mode (left panel, represented in *blue*) ; D-CF, expression in presence of detergents (middle panel, represented in *red*); L-CF, expression in presence of lipids (right panel, represented in *green*).

3.1.1. Precipitate mode

In the absence of detergent or lipids in the translation reaction, protein precipitates are formed during the synthesis of membrane proteins. These precipitates can be efficiently solubilized with detergents (Klammt *et al*, 2004), but functional solubilization is not a general process and proper folding can not be assured. On the one hand, there is evidence that the solubilization of membrane protein precipitates can result in

functionally folded proteins, as shown for the multidrug transporter EmrE (Klammt *et al*, 2004) {Danieli:2004up} or the human histamine-1 receptor (Sansuk *et al*, 2008) produced in *E. coli* cell-free systems. On the other hand, membrane proteins such as the nucleoside transporter Tsx (Klammt *et al*, 2005) or the human endothelin-B receptor, also produced in *E. coli* cell-free systems, were inactive after solubilization (Klammt *et al*, 2007). It is also worth to note that depletion of lipids from the WGE during preparation is not complete and WGEs might contain some residual lipids (Schwarz *et al*, 2008), which could explain why membrane proteins expressed in the precipitate mode are sometimes partially soluble.

3.1.2. Expression in presence of detergents

Detergents are amphipathic molecules consisting of a polar or charged head group and a hydrophobic tail. They can be classified into three groups depending on the type of the head group : ionic (cationic or anionic), nonionic and zwitterionic detergents. Whereas ionic detergents (*e.g.* SDS) disturb protein-protein interactions and therefore denaturate proteins, nonionic (*e.g.* DDM) and zwitterionic detergents (*e.g.* CHAPS) preserve protein-protein interactions and are thus considered as mild for membrane protein solubilization. Above a threshold of monomer concentration called the critical micelle concentration or CMC, detergents in aqueous solutions spontaneously form micellar, generally spherical, structures (Garavito & Ferguson-Miller, 2001; Seddon *et al*, 2004). The CMC is influenced by pH, ionic strength, temperature and also the presence of protein, lipid and other detergent molecules. Above the CMC, the detergent monomer concentration is constant and does not depend on the total detergent concentration (le Maire *et al*, 2000).

When expressed in the presence of detergents in cell-free expression systems, membrane proteins interact with detergent already during or shortly after translation, leading to the formation of proteomicelles. Mainly based on nonspecific hydrophobic interactions, this process does not use complex translocation machineries to facilitate the insertion of the proteins into membranes. Moreover, detergents are available instantly at the ribosomes, avoiding transport of synthesized proteins to membranes. Problems encountered regarding to the transport and translocation processes are therefore eliminated (Schwarz *et al*, 2008). However, the absence of translocon can be problematic for the topology of multi-spanning membrane proteins and proper protein folding is not

ensured. In addition, not all detergents are compatible with cell-free systems. The use of detergents in cell-free systems based on *E. coli* lysates has been broadly reported (Elbaz *et al*, 2004; Berrier *et al*, 2004; Klammt *et al*, 2005; Ishihara *et al*, 2005; Schwarz *et al*, 2007; Schwarz *et al*, 2008; Deniaud *et al*, 2010; Miot & Betton, 2011), suggesting that mild detergents with a low CMC, *e.g.* Brij (polyoxyethylene-lauryl-ether) or Tween (polyoxyethylene sorbitan monolaurate) derivatives, allow optimal solubilization yields without interfering with expression yields. However, the use of such detergents in the WGE-CF system has shown drawbacks, mainly affecting protein expression levels (Genji *et al*, 2010). Moreover, the available information on the best choice of detergent type and concentration for membrane protein expression in the WGE-CF system is still sparse (Genji *et al*, 2010; Chae *et al*, 2010; Nozawa *et al*, 2011; Carraher *et al*, 2013). In this work, we give new insight into the use of detergents for the expression of membrane proteins in the WGE-CF system (Fogeron *et al*, 2015) (refer to section IV.4.2.).

There are alternatives to traditional detergents, *e.g.* fluorinated surfactants and amphipols. Fluorinated or hemifluorinated, strongly hydrophobic but not lipophilic, surfactants (Park *et al*, 2007; Park *et al*, 2011; Blesneac *et al*, 2012) resemble detergents but interfere less than detergents do with the stabilization of protein/protein and protein/lipid interactions (Popot, 2010). Some amphipols, which are amphipathic polymers adsorbing onto the hydrophobic transmembrane surface of membrane proteins and stabilizing their native fold under particularly mild conditions, could also be used in cell-free systems. Amphipols and fluorinated surfactants have been reported to be compatible with lipid structures and to support direct membrane protein reconstitution into membranes (Nagy *et al*, 2001; Park *et al*, 2007). However, the only amphipol compatible with cell-free protein expression (NAPol) is not commercially available (Popot, 2010; Park *et al*, 2011).

In order to be analyzed in a native-like environment, membrane proteins expressed in the presence of detergents should be reconstituted in lipids after purification (refer to section II.5.).

3.1.3. Expression in presence of lipids

Although the ER is removed from the WGEs during preparation, it is possible to mimic the natural membrane environment while adding lipids to the translation reaction,

e.g. microsomes or liposomes. Cell-free systems tolerate relatively high concentrations of lipids and lipid mixtures, and even slightly beneficial effects have been observed on the expression efficiency (Klammt *et al*, 2004). Different protocols have been developed for the expression of membrane proteins in the presence of lipids (Shadiac *et al*, 2013).

Microsomes are membranous vesicles which are obtained from ER remaining pieces after cell lysis. Common sources for microsomes supplemented to eukaryotic systems are dog pancreas (Jackson & Blobel, 1977), oocytes (Kobilka, 1990; Lyford & Rosenberg, 1999) and oviduct cells (Rosenberg & East, 1992). Membrane proteins produced in the presence of microsomes are translocated through the translocon of the ER membrane with a signal peptide, and then glycosylated within the lumen of the membranes (Dobberstein & Blobel, 1977; Katz *et al*, 1977; Lingappa *et al*, 1978). Since the protein synthesis machinery is present only outside the vesicles, a prevalent inside-out orientation of membrane proteins can therefore be expected (Schwarz *et al*, 2008). However, the combination of cell-free protein expression and microsomal membranes is not widespread and only few publications report such approaches (Sachse *et al*, 2014).

Liposomes are artificial and spherical vesicles formed by lipid bilayers from either synthetic lipids or biological lipid extracts. They are most commonly prepared with asolectin from soybean which is a phospholipid mixture (Goren *et al*, 2009; Harbers, 2014) or using a mixture of defined phospholipids (*e.g.* phosphatidylcholine and cholesterol). Liposomes are called proteoliposomes after integration of membrane proteins into liposomal membranes. Various examples of membrane proteins produced using cell-free systems in presence of liposomes have already been published. The production of apo cytochrome b_5 in giant liposomes has been reported, followed by isolation of giant proteoliposomes by discontinuous density gradient centrifugation (Nomura *et al*, 2008). Moreover, the detergent-free integration of a mitochondrial inner membrane protein into liposomes using a WGE-CF system was reported the first time with the ADP/ATP carrier, showing that both the translation rate and the integration efficiency are lipid-dependent (Long *et al*, 2012). The curdlan synthase from *Agrobacterium* was inserted with a random orientation topology in DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine) liposomes during cell-free protein synthesis, followed by purification by density gradient flotation (Periasamy *et al*, 2013). *Shaker* potassium channels were expressed in the presence of liposomes and obtained

proteoliposomes were then injected in *Xenopus laevis* oocytes to test the functionality of the channels (Jarecki *et al*, 2013). Moreover, many other proteins with enzymatic activity have been prepared in presence of liposomes in the WGE-CF system (Goren & Fox, 2008; Sevova *et al*, 2010; Aly *et al*, 2012). In addition, cell-free translation and proteoliposome production systems have also been described for plant transporters (Nozawa *et al*, 2007; Nozawa *et al*, 2011; Nozawa & Tozawa, 2014). Examples of membrane proteins co-translationally inserted into lipid-based vesicular structures in cell-free systems in the absence of detergent are also presented in (Sachse *et al*, 2014). Finally, Nozawa and colleagues reported the use of liposomes formed from phospholipids in the presence of detergent for the rapid screening of various structurally divergent membrane proteins for their co-translational insertion in the WGE-CF system (Nozawa *et al*, 2011). Like for microsomes, the protein synthesis machinery is present only outside the vesicles and a prevalent inside-out orientation of membrane proteins can therefore be expected (Schwarz *et al*, 2008). Although co-translational insertion of membrane proteins into liposomes is in theory very attractive, not all proteins can be integrated into liposomes, some requiring a more complex lipid environment and others the translocon machinery (Sachse *et al*, 2014). In addition, this approach requires high lipid-to-protein ratios, suffers from poor yields and presents difficulties for purification.

There are alternatives to biological membrane vesicles and liposomes, *e.g.* bicelles and nanodiscs. Bicelles are assemblies of phospholipids and detergents in a flattened, disc-like shape, and represent an intermediate between micelles composed of pure detergents and lipidic membrane structures (Dürr *et al*, 2012; Lyukmanova *et al*, 2012; Sachse *et al*, 2014). Nanodiscs are nanoparticles composed by a discoidal phospholipid bilayer encircled in a non-covalent manner by two copies of a membrane scaffold protein, which is a modified apolipoprotein. The diameter of nanodiscs ranges from 10 to 20 nm, depending on the length and type of the membrane scaffold protein (Sachse *et al*, 2014). Nanodiscs have been reported to be an alternative to microsomes and liposomes (Ritchie *et al*, 2009; Bayburt & Sligar, 2010). During their synthesis, membrane proteins are incorporated in a passive manner into the nanodiscs and can later on be extracted from them in a native functional form (Ranaghan *et al*, 2011). A major advantage of nanodiscs is that they keep membrane proteins soluble in a detergent-free environment, possibly yielding monodisperse and homogenous samples (Borch & Hamann, 2009). Moreover, a

tag fused to the membrane scaffold protein allow for a simple purification procedure (Bayburt & Sligar, 2010).

3.2. Advantages of the WGE-CF system over classical cell-based and other cell-free systems

The WGE-CF system presents many advantages not only over classical cell-based protein expression systems but also over other cell-free systems. An overview of the main cell-based and cell-free expression systems available is given in **Table II.1**.

Like other cell-free protein expression systems, the WGE-CF system overcomes toxicity problems that often occur in classical cell-based expression systems when working with membrane proteins. Moreover, this system does not require any lysis procedure such as sample sonication or detergent lysis. Fewer process steps are therefore required, allowing automation of the system (www.cfsciences.com). In addition, this system is open, offering the possibility to work with additives such as chaperones, but also detergents and lipids which allow the expression of membrane proteins directly in a solubilized form. Moreover, cell-free protein expression systems are generally very efficient to produce isotopically labeled proteins for NMR studies. Indeed, only the synthesized protein is isotopically labeled and metabolic scrambling is negligible in comparison to cell-based systems.

In comparison to classical cell-based and prokaryotic cell-free expression systems, the WGE-CF system allows a better protein folding for eukaryotic proteins. Indeed, the rate of polypeptide growth on ribosomes differs considerably between eukaryotes and bacteria : it is five to ten times slower in eukaryotes, thus promoting the co-translational folding of proteins (Netzer & Hartl, 1997; Hartl & Hayer-Hartl, 2002). In addition, newly synthesized proteins are stabilized by eukaryotic chaperones which also promote folding. The eukaryotic translation machinery is even thought to have been optimized through evolution to facilitate co-translational domain folding (Endo & Sawasaki, 2006). In addition, translation can be performed in a wide temperature range between 4°C and 30°C, the optimal temperature being between 15°C and 26°C (Harbers, 2014), which supports a better protein folding too. Moreover, since the WGE-CF system contains only negligible amounts of endogenous proteases, synthesized proteins are stable.

The WGE preparation is much more critical and takes significantly longer than the preparation of *E. coli* lysates (Schwarz *et al*, 2008). However, prokaryotic cell-free systems neither provide specific eukaryotic folding systems nor posttranslational modifications (Schwarz *et al*, 2008). Although asparagine-linked glycosylation was recently reported in an *E. coli* cell-free system by coupling *in vitro* translation with an N-linked glycosylation pathway reconstituted from defined components (Guarino & DeLisa, 2012), glycosylation is not observed in prokaryotes. Addition of microsomal membranes to the WGE-CF system allows the production of glycosylated proteins (Lingappa *et al*, 1978) (refer to section II.3.1.3). Moreover, a study comparing *E. coli* and WGE-CF systems for the expression of eukaryotic proteins, including some membrane proteins, reported that the WGE-CF system was more efficient (Langlais *et al*, 2007). Finally, in contrast to prokaryotic systems, the WGE-CF system shows little codon usage preference and proteins encoded by AT-rich cDNAs can be synthesized without codon optimization.

In comparison to the RRLs, which is another eukaryotic cell-free system, the WGE-CF system offers many advantages (Endo & Sawasaki, 2006). Indeed, although the RRLs offer appropriate protein folding and posttranslational modifications too, codon preference is much looser in the WGE-CF system, expression yields are much higher and the scale-up is much easier as well as less expensive.

Table II.1 ***Comparison of classical cell-based and main cell-free protein expression systems.***
Adapted from {Endo:2006fk} and references in section II.1.2. and section II.2.2.

	Cell-based expression systems				Cell-free expression systems		
	bacteria	**mammalian cells**	**insect cells**	**yeast**	***E. coli***	**rabbit reticulocytes**	**wheat germ**
system	prokaryote	eukaryote	eukaryote	eukaryote	prokaryote	eukaryote	eukaryote
main posttranslational modifications							
signal peptide cleavage	no	yes	yes	yes	no	with microsomal membranes	with microsomal membranes
N-methionine removal	no	yes	yes	yes	no	yes	yes
phosphorylation	no	yes	yes	yes	no	yes	yes
disulfide bond formation	yes	yes	yes	yes	yes	n.a.	depends on DTT concentration
glycosylation	no	yes	more uniform, less complex N-glycans	yes	no	with microsomal membranes	with microsomal membranes
translation rate	high	low	low	high	high	low	low
folding	post-translation	co-translation	co-translation	co-translation	post-translation	co-translation	co-translation
codon preference	tight	loose	loose	loose	tight	tight (optimal only for globin synthesis)	loose
cost	low	expensive	expensive	low	low	expensive	expensive with commercial WGEs
scale-up	easy	difficult	difficult	easy	easy	difficult	quite easy
isotope labeling	metabolic scrambling	metabolic scrambling	metabolic scrambling	metabolic scrambling	efficient and specific	efficient and specific	efficient and specific
production range order	> 10 mg	mg	mg	> 10 mg	mg	µg	mg
membrane proteins							
production	often toxic	often toxic	often toxic	often toxic	easy	easy	easy
purification	difficult	difficult	difficult	difficult	easy	difficult (low amounts)	easy

4. Reconstitution of membrane proteins in lipids

Membrane proteins produced and purified in the presence of detergents have to be reconstituted in lipids to be analyzed in a native-like membrane environment (refer to section II.3.1.2. for details about the expression of membrane proteins in the presence of detergents). In this section, classical reconstitution protocols are described, as well as a new method called GRecon which consists in reconstituting detergent-solubilized membrane proteins on density gradients containing lipids and cyclodextrin for detergent removal.

4.1. Classical reconstitution protocols

4.1.1. Reconstitution strategies

There are four main strategies to reconstitute membrane proteins in phospholipid vesicles, *i.e.* the organic solvent-mediated reconstitution, mechanical methods, the direct incorporation into preformed liposomes and the detergent-mediated reconstitution (Rigaud *et al*, 1995; Rigaud & Lévy, 2003). Although bacteriorhodopsin was efficiently incorporated into liposomes by organic solvent-mediated reconstitution (Rigaud *et al*, 1983), most amphiphilic membrane proteins are denatured by organic solvents, thus limiting the use of this approach (Rigaud *et al*, 1995). Mechanical methods consist mainly in freeze-thawing of multilamellar vesicles formed by rehydration of dried lipids in aqueous buffer, sonication and extrusion (Rigaud *et al*, 1995). Direct incorporation into preformed liposomes requires « catalyzers » (Eytan *et al*, 1976; Enoch *et al*, 1979; Greenhut & Roseman, 1985) and lead to proteoliposomes with a wide size range and heterogeneous protein distribution among the liposomes (Rigaud *et al*, 1995). The most widely used strategy for membrane protein reconstitution is the detergent-mediated reconstitution because membrane proteins are purified and solubilized in detergents. In this approach, membrane proteins are first cosolubilized with phospholipids in the appropriate detergent in order to form an homogenous solution of lipid/protein/detergent and lipid/detergent micelles. Preformed liposomes can also be used (Rigaud & Lévy, 2003). The removal of the detergent, by methods described below, results in the formation of proteoliposomes which are bilayer vesicles with incorporated protein (Rigaud *et al*, 1995). Although the molecular mechanisms for the formation of proteoliposomes upon detergent removal from lipid/protein/detergent mixtures is not

fully understood, the different steps of proteoliposome formation and the mechanisms of lipid/protein association are described in detail in (Rigaud *et al*, 1995).

4.1.2. Detergent removal for detergent-mediated reconstitution

There are four different methods to achieve detergent removal : dilution, size exclusion chromatography, dialysis and hydrophobic adsorption onto polystyrene beads (Rigaud *et al*, 1995; Rigaud & Lévy, 2003). These methods take advantage of the properties of the detergent such as the CMC, the charge and/or the aggregation number (Seddon *et al*, 2004).

Detergent removal by dilution

Dilution with detergent-free buffer leads to the formation of proteoliposomes when the detergent concentration is under the CMC. However, this technique has some disadvantages, *i.e.* the addition of buffer leads to an important local heterogeneity, the detergent removal is not complete, and the sample has to be concentrated after dilution (Wang & Tonggu, 2015).

Detergent removal by size exclusion chromatography

This technique is based on the size differences between protein/detergent and lipid/detergent or detergent micelles. A detergent-free buffer or a buffer containing a further detergent below its CMC is used as eluent (Seddon *et al*, 2004). This method is fast and efficient, as shown for the human BK potassium channel (Wang & Sigworth, 2009) and the glycophorin A (Mimms *et al*, 1981). However, its main disadvantages are the dilution of the sample which might need to be concentrated after detergent removal and the possible retention of lipids or the aggregation of the protein on the gel filtration column (Wang & Tonggu, 2015).

Detergent removal by dialysis

Dialysis is one of the most common method for detergent removal. The lipid/protein/detergent mixture is dialyzed against a detergent-free buffer. Although this process is efficient for detergents with high CMCs and works best for those with low molecular weight or small cross-sectional area, it is not fully suitable for detergents with low CMCs (*e.g.* nonionic detergents) (Rigaud *et al*, 1995; Seddon *et al*, 2004). The dialysis

method has to be performed above the transition temperature of the lipids when synthetic lipids are used (Wang & Tonggu, 2015). Simple and not expensive, the dialysis procedure leads to homogeneously sized vesicles (Rigaud & Lévy, 2003). However, this method suffers from several drawbacks : this is a very long process (several days), the detergent removal rate can not be controlled and the final protein concentration is unknown since the osmotic pressure differences lead to volume changes. A technical trick to speed up the detergent removal process is to add Bio-Beads (see below) outside the dialysis bag to maintain the external concentration of dialyzed detergent at zero, therefore reducing the number of changes of buffer required during dialysis and decreasing the time of dialysis (Rigaud & Lévy, 2003; Kunert *et al*, 2014).

Detergent removal with polystyrene beads

Bio-Beads are hydrophobic polymer beads which bind rapidly detergents by their hydrophobic tail. Bio-Beads can be added directly to the protein/detergent/lipid mixture (batch procedure) and removed by centrifugation or filtration after detergent binding. A low bead-to-detergent ratio in the first step of reconstitution to slowly remove the detergent leads to homogenous liposomes (Rigaud *et al*, 1995). Size distribution of the proteoliposomes is determined by the detergent removal rate. Indeed, fast detergent removal leads to small liposomes whereas slow removal leads to large liposomes. Bio-Beads can remove almost all kinds of detergents but is especially suitable for detergents with a low CMC. This method was used to reconstitute efficiently into liposomes the Ca^{2+}-ATPase (Young *et al*, 1997), a type-IV P-type ATPase (Zhou & Graham, 2009), the rat heart mitochondrial F1F0 ATP synthase (Kim & Song, 2010) and the multidrug transporter LmrP (Schaedler *et al*, 2012). However, the hydrophobic nature of Bio-Beads can lead to protein adsorption (Levy *et al*, 1990) and an alternative is therefore to combine Bio-Beads with dialysis (Rigaud & Lévy, 2003; Kunert *et al*, 2014).

4.1.3. Reconstitution efficiency

The integrity and the activity of the protein of interest have to be preserved during reconstitution to allow further studies. Morphology and size of the proteoliposomes, the number of protein monomers incorporated, the final orientation of the incorporated transmembrane protein and the permeablity of the proteoliposomes are important criteria to take into account (Rigaud *et al*, 1995; Rigaud & Lévy, 2003). The efficiency of

reconstitution depends on the membrane protein of interest and several parameters such as the composition of the lipids, the lipid-to-protein ratio (LPR), the choice and the removal rate of detergent, the ionic strength and the conditions of the initial detergent solubilization (Wang & Tonggu, 2015). The efficiency of membrane protein incorporation can be accurately analyzed by density gradient, allowing the separation of the different populations after reconstitution, *i.e.* proteoliposomes, empty liposomes and the nonincorporated or aggregated protein (Cladera *et al*, 1997; Rigaud & Lévy, 2003). Freeze-fracture electron microscopy also allows to monitor the membrane protein incorporation (Gulik-Krzywicki *et al*, 1987).

4.1.4. Protein orientation

For membrane proteins with enzymatic activity, *e.g.* ATPases, protein orientation after reconstitution in the lipid bilayer is an important aspect. In detergent-mediated reconstitution, membrane proteins can be inserted in two orientations : inside-out (*i.e.* the protein cytoplasmic side faces the outside of the vesicle) and outside-out (*i.e.* the protein cytoplasmic side faces the inside of the vesicle). The insertion of a protein into preformed liposomes might lead to the formation of proteoliposomes with an unidirectional outside-out orientation whereas random orientations have been observed for samples reconstituted by detergent removal from lipid/protein/detergent mixtures (Eytan, 1982; Jain & Zakim, 1987; Rigaud *et al*, 1995). The protein orientation is additionally influenced by the size of the ectodomain. Indeed, due to steric hindrance, the bigger the ecdomain is, the higher is the inside-out orientation.

4.2. GRecon

The GRecon or gradient reconstitution method has been recently developed by Kühlbrandt and colleagues (Althoff *et al*, 2012). This approach combines detergent removal, lipid reconstitution and gradient centrifugation in a single step. The protein of interest is loaded on a sucrose gradient containing both solubilized lipids and cyclodextrin. Cyclodextrins are cyclic sugar oligomers of 950-1300 Da which have been used to remove detergents from membrane protein solutions (Degrip *et al*, 1998; Signorell *et al*, 2007). The detergent removal rate is a critical step for successful reconstitution and in the GRecon method, the cyclodextrin concentration in the gradient

and the steepness of the gradient itself allow to control this parameter. Kühlbrandt and colleagues claim that the GRecon works with a wide range of detergents including those with low CMC and for membrane proteins of all sizes and levels of complexity. They successfully applied this method to reconstitute in lipids the 1.7 MDa mitochondrial supercomplex $I_1III_2IV_1$, the 500 kDa ATP synthase from *Ilyobacter tartaricus*, the 56 kDa carnitine transporter CaiT from *E. coli*, the trimeric plant light-harvesting complex LHC-II and the 32 kDa monomeric β-barrel outer-membrane porin OmpG from *E. coli* with the aim to analyze these proteins by cryo-electron microscopy.

5. Hepatitis C virus

5.1. Hepatitis C

5.1.1. Virus discovery

Two viral pathogens were originally known to cause hepatitis : hepatitis A (HAV) and hepatitis B (HBV) viruses. Many patients with hepatitis however didn't harbor HAV or HBV infection. Their disease was referred as non-A non-B hepatitis and was formerly identified as a putative viral hepatitis occuring after transfusion of blood products or intravenous drug use (Westbrook & Dusheiko, 2014). Non-A non-B hepatitis was shown to lead to persistent infection in a high proportion of infected individuals and in some cases to progress to chronic liver disease, cirrhosis and hepatocellular carcinoma. The viral agent being the major cause non-A non-B hepatitis was identified in 1989 (Choo *et al*, 1989) and was named hepatitis C virus (HCV).

5.1.2. Epidemiology

Chronic infection with hepatitis C virus affects about 130-170 million individuals worldwide (Lavanchy, 2011) which represents 2% to 3% of world population. Transmission mainly occurs by parenteral routes associated with medical treatments, immunisation, blood transfusion and injecting drug use (Simmonds, 2013). The geographical distribution of HCV infection is heterogeneous (**Figure II.7**).

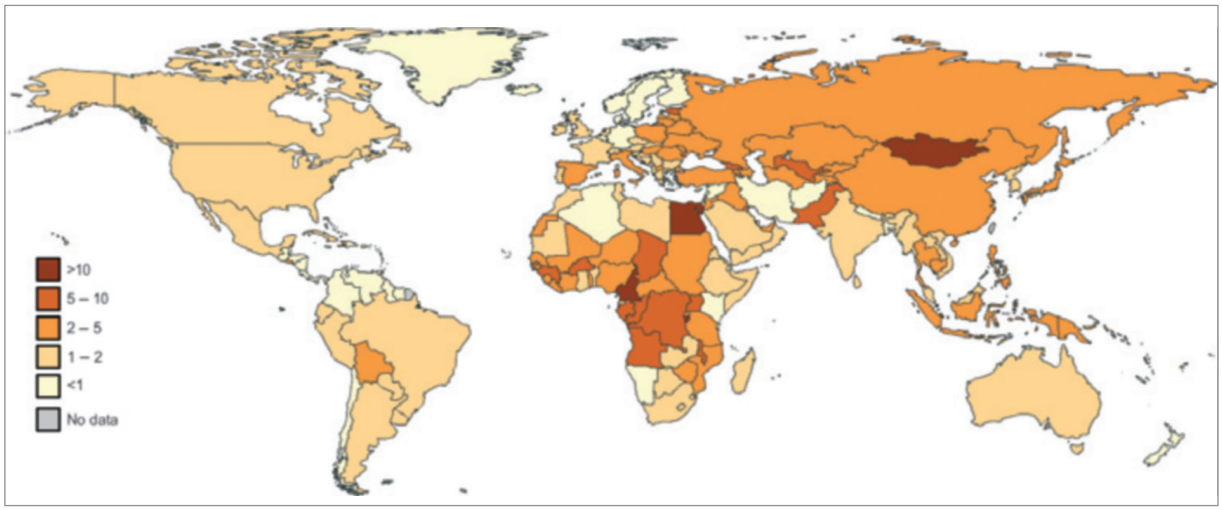


Figure II.7 ***Hepatitis C global prevalence 2010 (%)***. From (Lavanchy, 2011)

Countries in Western Europe and North America have low prevalence rates (< 2.5%) whereas countries in Africa and Asia have high prevalence rates (>5%). Egypt is the

country with the highest prevalence rate (14%), the main cause being the use of contaminated syringes in nationwide schistosomiasis treatments during the 1970s (Frank *et al*, 2000).

5.1.3. Pathogenesis

HCV infection leads first to acute hepatitis C which is clinically mild, typically unrecognised and thus infrequently diagnosed, particularly when the disease progresses to chronic hepatitis (Westbrook & Dusheiko, 2014). In very rare cases, HCV infection can be associated to fulminant hepatitis (Farci *et al*, 1996). Spontaneous clearance appears for approximately 25% of individuals with acute hepatitis C (**Figure II.8**). For ~ 75%, the infection is defined as being chronic after six months of persistance of HCV RNA within the blood (Westbrook & Dusheiko, 2014). The disease is stable for ~ 80% of individuals with chronic hepatitis. For ~ 20%, chronic hepatitis leads to cirrhosis within 20 years after HCV infection and each year in 1% to 4% of individuals with cirrhosis further to hepatocellular carcinoma (Di Bisceglie, 1998; Gordon *et al*, 1998). Hepatitis C is indeed the major cause of steatosis, liver cirrhosis and hepatocellular carcinoma (Yamane *et al*, 2013). As a consequence, chronic hepatitis C is one of the most frequent indications for liver transplantation in many countries (Brown, 2005).

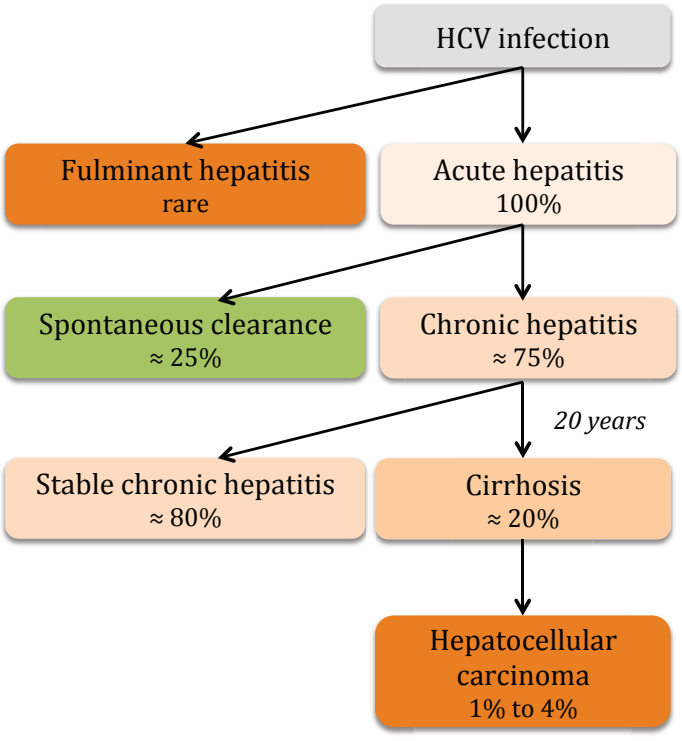


Figure II.8 ***Evolution of HCV infection.*** Adapted from (Westbrook & Dusheiko, 2014).

5.1.4. Therapeutic treatment

For decades the antiviral treatment of chronic hepatitis C has been based on the use of pegylated interferon α (IFNα) combined with ribavirin, a guanosine analogue with a broad spectrum of activity against DNA and RNA viruses (Pawlotsky, 2013). Due to the high rate of non responders and side effects, new therapeutic approaches targeting essential components of the HCV life cycle have been developped recently. They include two kinds of antivirals for treatment of hepatitis C : the direct acting antivirals (DAAs) and the host target antivirals (HTAs).

Key targets of DAAs so far have been the RNA-dependent RNA polymerase (RdRp) NS5B which is the catalytic center of HCV genome replication, and the viral serine protease complex consisting of NS3 and NS4A, responsible for all polyprotein cleavage steps downstream of NS3 (Pawlotsky, 2013; Gerold & Pietschmann, 2014). Inhibitors of the regulatory NS5A protein are becoming available too (*e.g.* daclatasvir) (Bartenschlager *et al*, 2013; Gerold & Pietschmann, 2014; Pawlotsky, 2014). However, DAAs have drawbacks like fast emergence of resistance mutations and virus genotype dependency. HTAs could overcome the limitations of DAA strategies.

Cyclophilin A (CypA), phosphatidylinositol 4-kinase IIIα (PI4KIIIα) and microRNA miR-122 are major HTAs implicated in HCV translation and polyprotein processing and currently in preclinical and clinical stages (Bartenschlager *et al*, 2013; Gerold & Pietschmann, 2014). CypA belongs to a highly conserved family of peptidyl-prolyl isomerases and interacts with NS5A. This could promote viral protein folding, regulate polyprotein processing and thereby facilitate RNA replication (Foster *et al*, 2011; Verdegem *et al*, 2011). PI4KIIIα is found at the plasma membrane and the ER and generates phosphoinositides at the cytosolic membrane leaflet. PI4KIIIα interacts with NS5A during HCV infection (Berger *et al*, 2011; Reiss *et al*, 2011; Reiss *et al*, 2013). MicroRNAs are non-coding nucleic acids that usually bind to host mRNA and block translation or target mRNAs for degradation (Wilson & Doudna, 2013). However, in the case of HCV infection, miR-122, which is mainly expressed in the liver, binds to the 5'-non-translated region of the HCV genome, increasing its stability and thus replication (Jopling *et al*, 2005; Jopling *et al*, 2008; Henke *et al*, 2008).

Whereas the standard-of-care treatment since 2011 is based on a triple combination of

pegylated IFNα, ribavirin and a DAA against NS3-4A (either telaprevir or boceprevir), new triple and quadriple combination therapies including pegylated IFNα, ribavirin and one or two DAAs or HTAs are currently being evaluated in clinical trials (Pawlotsky, 2013; Pawlotsky, 2014; Muir, 2014). The most promising DAA is probably the sofosbuvir which is a nucleotide HCV polymerase inhibitor with pangenotypic activity (Alberti & Piovesan, 2014; Kayali & Schmidt, 2014). Recent clinical trials carried out by the ANRS (France Recherche Nord&sud Sida-hiv Hépatites) on two large french cohorts demonstrated the high efficiency of the combination of sofosbuvir and daclatasvir (50th Annual Congress of The European Association for the Study of the Liver, 22-26 April 2015, Vienna, Austria).

5.2. The virus

5.2.1. Classification and genomic variability

Hepatitis C virus belongs, together with GBV-B, to the *Hepacivirus* genus of the family *Flaviviridae* which also includes the genera *Flavivirus* (*e.g.* yellow fever virus, West Nile virus and dengue virus), *Pestivirus* (*e.g.* bovine viral diarrhea virus) and *Pegivirus* (*e.g.* GBV-A, GBV-C) (Simmonds, 2013). The virus is classified into seven epidemiologically relevant genotypes (numbered 1 to 7) which differ from each other by more than 30% at the nucleotide level (Simmonds, 2004; Simmonds, 2013). Similarly, each genotype is divided into several subtypes (noted a, b, c, etc) which typically differ from each other by 20% to 25% in nucleotide sequences (Simmonds, 2004). The high degree of genetic variability is due to the combination of a lack of 5'-3' exonuclease proofreading activity by the RNA-dependent RNA polymerase and a high level of viral replication (about 10^{12} new virions each day) (Pawlotsky, 2003). A high degree of variability is also observed between viruses in a single patient, giving rise to quasispecies implicated in the lack of protective immunity against HCV and its persistence following infection (Forns *et al*, 1999).

5.2.2. Viral particle

HCV particles are tightly associated with cellular lipoproteins and lipids that determine both morphology and biophysical properties of the virions (Bartenschlager *et al*, 2011). In infected patients, these particules, called lipoviroparticles (LVP (Andre *et al*, 2002), **Figure II.9**), have an irregular and globular shape of about 100 nm. The density of LVP is very heterogenous, from 1 .25 g/mL to below 1.06 g/mL, and infectivity is

inversely correlated to density (Andre *et al*, 2002; Bartenschlager *et al*, 2011). The lipid bilayer composing the viral envelope comes from the host and contains apoB, apoE, apoC1, C2 and C3 proteins, as well as the viral glycoproteins E1 and E2 (Andre *et al*, 2002; Nielsen *et al*, 2006; Bartenschlager *et al*, 2011). Lipid composition of HCV virions in fact resembles the one of very low density (VLDL) and low-density (LDL) lipoproteins, with cholesteryl esters accounting for almost half of the total HCV lipids (Popescu *et al*, 2014; Dubuisson & Cosset, 2014). Inside the envelope, the nucleocapsid composed of the core protein contains a single positive strand RNA constituting the viral genome.

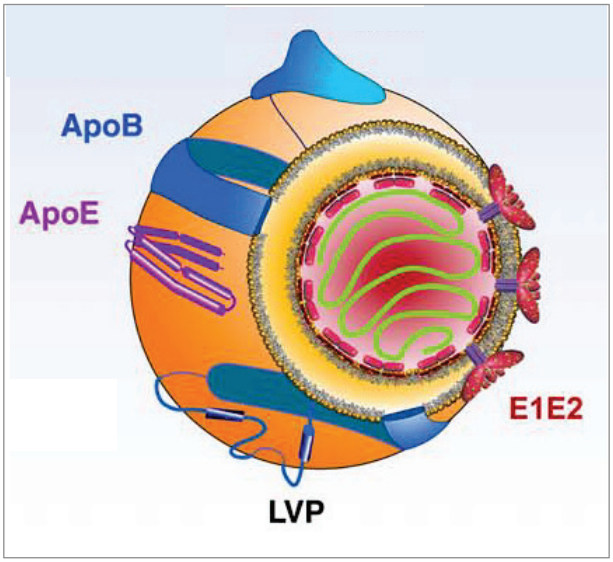


Figure II.9 ***HCV lipoviroparticle (LVP).*** From (Bartenschlager *et al*, 2011). Apolipoproteins ApoB and ApoE on the lipid bilayer are represented in *blue* and *purple*, respectively. Core dimers on the lipid bilayer inside the capsid is represented in *pink*. E1E2 heterodimers are represented in *red* and their transmembrane domains in *dark purple*. Positive-strand RNA inside the capsid is represented in *green*.

5.2.3. Genome structure and polyprotein processing

The HCV genome is composed by a ca. 9.6 kb single-stranded RNA of positive polarity containing a single open reading frame (ORF). This ORF encodes a polyprotein of 3,000 amino acid residues that is co- and posttranslationally cleaved by cellular and viral proteases into ten proteins. Highly conserved 5'- and 3'-non translated regions (NTRs) flank the ORF (Niepmann, 2013; Lohmann, 2013) (**Figure II.10**).

Non-translated regions

The 5' NTR plays a crucial role in replication and translation initiation. This highly conserved region is composed of the first 341 nucleotides of the viral genome (Penin *et al*, 2004b). This region is composed of four highly structured domains called I, II, III and IV (Lemon & Honda, 1997). It contains an internal ribosomal entry site (IRES) comprising domains II, III and IV (nucleotides 44 to 341) and the first 15 nucleotides of the ORF

(Niepmann, 2013; Pérard *et al*, 2013). The presence of a type III IRES ensures translation initiation in a cap-independent mechanism (Honda *et al*, 1996). With the recruitment of 40S ribosomal subunit, this region is essential for translation initiation. Moreover, two sites in the 5' NTR interact with the microRNA miR-122 (Jopling *et al*, 2005) which promotes accumulation of viral RNA by stimulation of IRES-mediated translation, enhancement of HCV replication and protecting viral RNA from degradation (Niepmann, 2013).

The 3' NTR is composed of the last 200 to 235 nucleotides of the viral genome. In contrast to the 5' NTR, this region is variable depending on the genotype and consists of three domains : a highly variable 40 nucleotides sequence, a poly-uracyl/pyrimidine motif of variable length, and a highly conserved 98 nucleotides sequence called X-tail and organized in three stem-loop structures (Kolykhalov *et al*, 2000). The poly-uracyl/pyrimidine motif together with the third stem-loop of the X-tail domain are essential for IRES activity stimulation and therefore translation (Song *et al*, 2006), whereas the second stem loop interacts with a stem loop within the RNA-dependent RNA polymerase coding sequence which is necessary for replication (Friebe *et al*, 2005).

HCV proteins

The polyprotein encoded by the viral genome is co- and posttranslationally processed by viral and host proteases into ten proteins : three structural proteins (Core, E1 and E2) and seven non-structural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B) (Moradpour & Penin, 2013). Cleavage of the structural proteins core, E1, E2 and the p7/NS2 junction is mediated by signal peptidase and signal peptide peptidase. The cleavage at the NS2/NS3 junction is mediated by the NS2 cysteine protease whose function is strongly enhanced by the N-terminal one-third of NS3. NS3 mediates the cleavage of the downstream proteins (Moradpour & Penin, 2013) : NS3 cleaves at the junctions NS3/NS4A and NS4A/NS4B, NS4A associates with the N-terminus of NS3 and the resulting NS3/4A complexes cleave at the junction NS4B/NS5A and NS5A/NS5B. All HCV proteins are associated to membranes. Refer to section II.6. for a detailed description of HCV proteins.

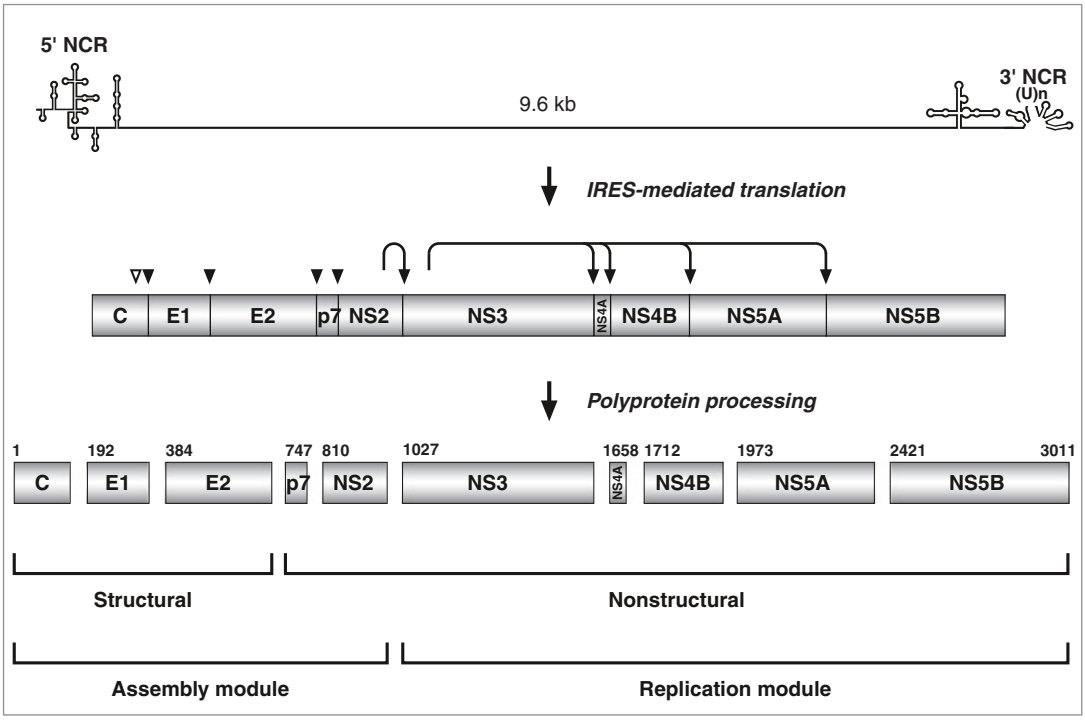


Figure II.10 ***Genetic organization and polyprotein processing of HCV.*** From (Moradpour & Penin, 2013). The 9.6 kb positive-strand RNA genome is schematically depicted at the *top*. Simplified RNA secondary structures in the 5' and 3' NTRs as well as the core and NS5B coding regions are shown. Internal ribosome entry site (IRES)-mediated translation yields a polyprotein precursor that is processed into the mature structural and non structural proteins. Amino acid numbers are shown above each protein (HCV H strain, genotype 1a, GenBank accession number AF009606). Cleavages by the endoplasmic reticulum signal peptidase are denoted by solid arrowheads. Further C-terminal processing of the core protein by signal peptide peptidase is indicated by an open arrowhead. Cleavages by the HCV NS2 and NS3-4A proteases are indicated by arrows. The polyprotein processing is illustrated here as a separate step for simplicity but occurs co- and posttranslationally. All components of the replication module (replicase) are also involved in assembly.

5.3. HCV life cycle

The HCV life cycle can be described in four steps : virus entry, genome translation and polyprotein processing, genome replication, and particule assembly and release (**Figure II.11**).

5.3.1. Virus entry

Initial attachment of HCV particles to hepatocytes is mediated by virion binding to glycosaminoglycans (GAGs) present at the surface of hepatocytes (Koutsoudakis *et al*, 2006). The viral particles can then interact with the LDL receptor (LDL-R) through its lipoprotein components (Agnello *et al*, 1999; Owen *et al*, 2009) and the virion is rapidely internalized and potentially sent to a degradation pathway (Popescu *et al*, 2014). This process is called the non-productive pathway. In contrast, the productive pathway is a

multistep process involving specific cellular entry factors : the lipoprotein receptor SRB1 (also known as scavenger receptor class B type I, SCARBI) (Scarselli *et al*, 2002), the tetraspanin molecule CD81 (Pileri *et al*, 1998) and the two tight-junction proteins Claudin-1 (CLDN1) and Occludin (OCLN) (Evans *et al*, 2007; Ploss *et al*, 2009; Liu *et al*, 2009a).

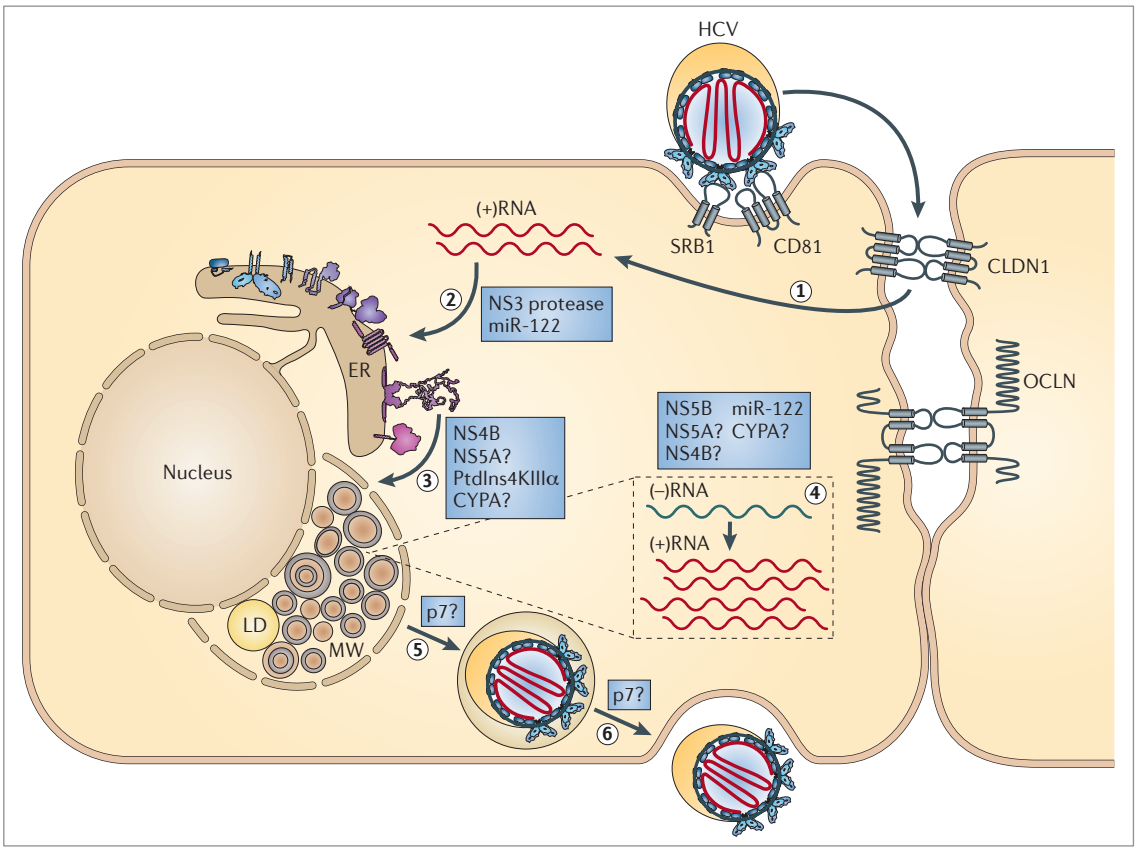


Figure II.11 ***The hepatitis C virus replication cycle.*** From (Bartenschlager *et al*, 2013). The HCV particle initially binds to scavenger receptor class B member 1 (SRB1), then engages in further interactions with the tight junction proteins claudin 1 (CLDN1) and occludin (OCLN) and finally enters cells by receptor-mediated endocytosis (step 1). After being released into the cytoplasm, the viral RNA genome is translated at the rough ER, giving rise to a polyprotein that is cleaved into mature proteins (step 2). The formation of the membranous web (MW), which is a membranous compartment composed of single-, double- and multi-membraned vesicles as well as lipid droplets (LDs), is induced by viral proteins, together with host cell factors (step 3). RNA replication occurs at an unspecified site within the membranous web and proceeds via a negative-sense copy ((-)RNA) that serves as a template for the production of excess amounts of positive sense RNA ((+)RNA) (step 4). Assembly of HCV particles probably initiates in close proximity to the ER and lipid droplets, where core protein and viral RNA accumulate. The viral envelope is acquired by budding through the ER membrane in a process that is linked to lipoprotein synthesis (step 5). HCV particles are thought to be released via the constitutive secretory pathway (step 6). Viral and host factors that are targeted by inhibitors are indicated in *blue boxes*. The steps of the replication cycle that are promoted by these factors are indicated.

In addition to these four essential entry factors, other host molecules such as the cholesterol transporter Niemann-Pick C1-like 1 (NPC1L1) (Sainz *et al*, 2012) or the

transferrin receptor 1 (Martin & Uprichard, 2013) support HCV cell invasion. After successive binding to these entry factors at the hepatocyte surface, the HCV particle is internalized by a clathrin-mediated endocytosis and low pH-dependent fusion takes place in early endosomes (Popescu *et al*, 2014).

5.3.2. Genome translation and polyprotein processing

The single-strand positive RNA of HCV is directly translated upon release of the viral genome into the cytoplasm. The host cell ribosomal machinery is used by the virus for translation which is cap-independent, and no viral factor is required. The ribosomal 40S subunit binds to HCV IRES domain II and initiation factor eIF3, tRNA, initiation factor eIF2 and GTP are recruited, giving ribosomal subunit 48S. Finally, the recruitement of ribosomal subunit 60S leads to a complete ribosomal complex that translates the viral genome into a polyprotein precursor (Hoffman & Liu, 2011). This polyprotein precursor is cleaved co- and posttranslationally by viral and cellular proteases into ten viral proteins (refer to section II.5.2.3. for details). Signal peptide of viral protein E1 leads the polyprotein precursor to ER and translation occurs in close association with ER membranes (Niepmann, 2013).

5.3.3. Genome replication

The viral genome is transcribed by the viral RdRp NS5B into a complementary RNA of negative polarity which serves as template for the synthesis of multiple viral RNAs of positive polarity (Lohmann *et al*, 1999; Lohmann, 2013). These RNAs of positive polarity serve either for viral RNA translation or for viral RNA encapsidation. Viral proteins NS3-4A, NS4B, NS5A and NS5B, together with both NTRs, are necessary and sufficient for the replication (Lohmann *et al*, 1999; Lohmann, 2013). HCV replication occurs in association with rearranged intracellular membranes, called membranous web (Gosert *et al*, 2003). Formation and activity of the membranous web are still poorly understood. Both NS4B and NS5A play a role in the induction of membrane rearrangements (Egger *et al*, 2002; Gosert *et al*, 2003) : NS4B induces single membrane vesicles, but the presence of NS5A is required for inducing double membrane vesicles (Romero-Brey *et al*, 2012; Paul *et al*, 2014). Membranes of HCV replicase could be derived from the ER membrane and biochemically modified to create a particular lipid environment required for the replicative activity of the complexes (Moradpour *et al*, 2007; Popescu *et al*, 2014).

5.3.4. Particle assembly and release

In order to assemble the RNA-containing capsid, the viral modules involved in the process have first to form and to localize in the proximity of the assembly site. The viral particle acquires then an envelope, budding in the ER lumen, and matures within the secretory pathway overlapping with the VLDL secretion pathway (Bartenschlager *et al*, 2013; Popescu *et al*, 2014). After cleavage from the polyprotein by the signal peptide peptidase, the core protein homodimerizes (Boulant *et al*, 2005) and interacts with the lipid droplets (LDs) (Barba *et al*, 1997). The level of core-LD accumulation inversely correlates with the efficiency of production of infectious viral particles, which could reflect the transient localization of core protein on LDs before it is transferred to ER-derived assembly sites (Shavinskaya *et al*, 2007; Dubuisson & Cosset, 2014). Besides cellular factors such as diacylglycerol acyltransferase-1 (DGAT1) or MAPK-regulated cytosolic phospholipase A2 (PLA2G4) (Dubuisson & Cosset, 2014), non-structural p7 and NS2 also regulate core traffiking to the site of virus assembly (Boson *et al*, 2011). The other non-structural proteins also seem to play a direct or indirect role in HCV morphogenesis (reviewed in (Dubuisson & Cosset, 2014)). After core protein, the replication complexes are recruited to the neighborhood of LDs, this step depending on NS5A protein and its phosphorylation level (Paul *et al*, 2014). Finally, enveloppe glycoproteins migrate, in interaction with p7, NS2 and NS3 (Jirasko *et al*, 2010; Stapleford & Lindenbach, 2011), to the neighborhood of LDs and arrive at the assembly site. After assembly, HCV particles are released from cells by transit through the secretory pathway, virions acquiring their low buoyant density characteristic during this process (Gastaminza *et al*, 2006).

5.4. Model systems for the study of HCV

5.4.1. Cell culture models

There are four major cell culture models to study HCV : subgenomic HCV replicons, retroviral HCV pseudotypes (called HCVpp), cell culture-derived HCV (called HCVcc), and HCV *trans*-complemented particles (called HCV_{TCP}). (**Figure II.12**).

Subgenomic HCV replicons

Subgenomic replicons are partial genomes of HCV encoding the non-structural proteins NS3 to NS5B as well as the 5'- and 3'- NTRs required for RNA replication. They became available in 1999 and provided the first step towards the development of tissue culture infectious models (Lohmann *et al*, 1999). Viral genomes with replication-enhancing mutations (REMs) and permissive clones of the human hepatoma cell line Huh-7 were first selected. Highly HCV-permissive Huh-7-derived cells lines Huh-7.5 and Huh-7-Lunet were then generated by curing of the HCV replicating cell clones with IFN (Blight *et al*, 2003; Friebe *et al*, 2005). These two cell lines are the gold standard in HCV research as they support all steps of the virus life cycle. Subgenomic replicon systems are used to test the effect of DAAs and HTAs on HCV RNA genome replication. However, predictive value when testing the efficacy of an antiviral compound is limited since REMs don't seem to increase viral fitness *in vivo* (Gerold & Pietschmann, 2014).

HCVpp

A few years after subgenomic HCV replicons, the establishment of retrovirus-based pseudoparticules systems allowed the entry of HCV into susceptible human hepatoma cells (Bartosch *et al*, 2003). The most widely used HCVpp system is based on retroviral particles decorated with the HCV E1 and E2 glycoproteins. Such HCVpp are generated by triple transfection of an E1E2 expression plasmid, a plasmid encoding the retroviral polymerase and capsid protein (gagpol), and a retroviral provirus into 293T cells. The proviral RNA is packaged into retroviral particles displaying E1E2 on their surface, containing a reporter gene (GFP or luciferase), and released in the cell culture supernatant. The generated pseudoparticles rely on interactions of HCV E1E2 with HCV-specific host cell surface proteins (entry factors) and follow a similar route of entry as infectious full length HCV. Upon entry into susceptible cells, the proviral RNA is reverse transcribed and integrated into the host cell genome allowing the expression of a given reporter gene and thus the quantification of successful entry events (Bartosch *et al*, 2003; Hsu *et al*, 2003; Gerold & Pietschmann, 2014). This model has allowed the identification of several cellular receptors essential for HCV entry into cells, such as Occludin (Ploss *et al*, 2009). On the one hand, HCVpp allow the isolated investigation of HCV entry and mimic well interactions of E2 with host entry factors (Bartosch *et al*, 2003). On the other hand, since HCVpp are produced in a non-hepatic cell line which has no VLDL pathway

and are therefore not associated to VLDL, the system can't predict lipoprotein-dependent interactions and poorly mimics fusion and uncoating events (Gerold & Pietschmann, 2014).

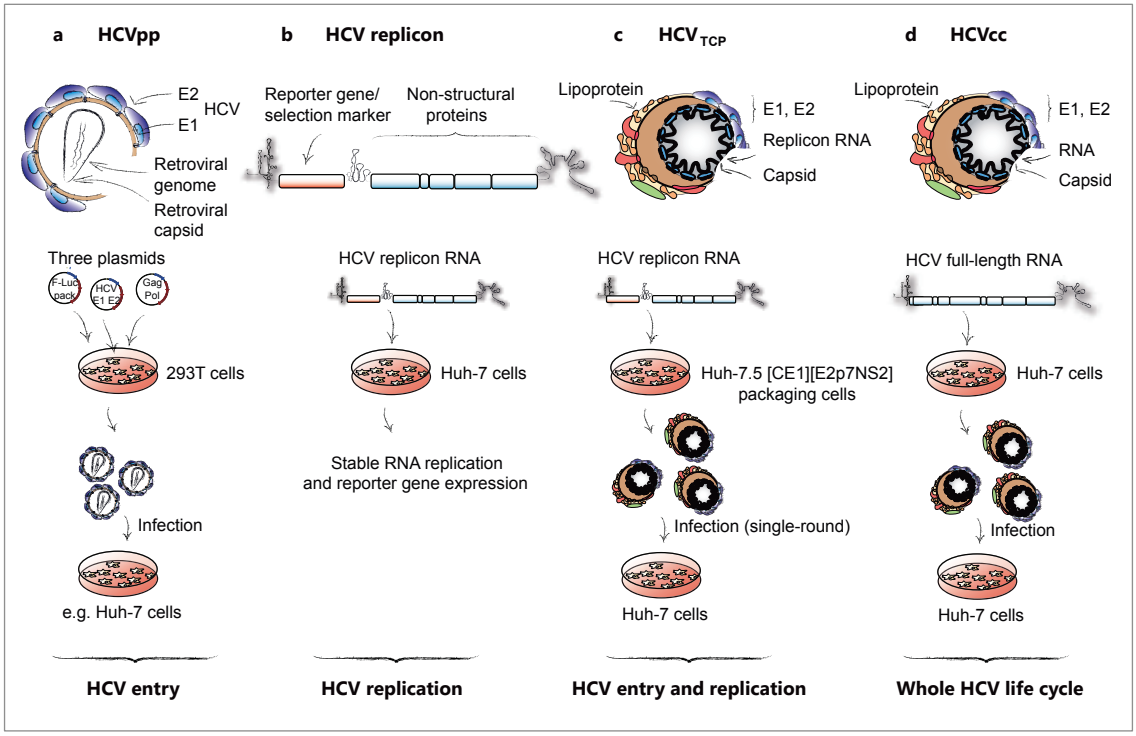


Figure II.12 ***Important HCV cell culture systems***. From (Gerold & Pietschmann, 2014). Cell-based HCV infection and replication models frequently utilized are depicted including (**a**) retroviral HCV pseudotypes (HCVpp), (**b**) subgenomic HCV replicons, (**c**) HCV transcomplemented particles (HCV_{TCP}), and (**d**) cell-culture-derived HCV (HCVcc).

HCVcc

A fully permissive cell culture model of HCV was established 2005 and enabled to adress viral assembly, egress and spread. Three independent laboratories described a recombinant genotype 2a HCV genome which could replicate and assemble virus particles in Huh-7.5 cells (Wakita *et al*, 2005; Lindenbach *et al*, 2005; Zhong *et al*, 2005). Most of the non-structural regions of those chimeric genomes originating from HCV of a Japanese patient with fulminant hepatitis, this HCV isolate was termed « Japanese fulminant hepatitis 1 » (JFH-1). JFH-1-based chimeric 2a genomes replicate without the need for adaptive mutations and assemble and release viral particules which are called cell-culture infectious HCV (HCVcc) (Wakita *et al*, 2005; Lindenbach *et al*, 2005). Such *in vitro* generated virions can infect chimpanzees and mice containing human liver grafts, two classical animal models for HCV infection (Lindenbach *et al*, 2006). The HCVcc system is the most widely used system in HCV research since it allows to study all aspects of the

viral life cycle *in vitro*. This system has however two major limitations : (i) HCVcc particles display a lower specific infectivity and higher buoyant density than serum-derived which may be due to impaired lipoprotein production in the context of Huh-7-derived cell clones (Lindenbach *et al*, 2006), and (ii) standard Huh-7-based cultures are non-polarized and therefore poorly reflect the highly polarized hepatocytes in the liver (Gerold & Pietschmann, 2014). Whereas the first HCVcc systems were genotype 2a-based, chimeric viruses encoding the structural proteins of genotypes 1 through 6 and the non-structural proteins from JFH-1 with certain adaptive mutations have been developed, allowing the analysis of genotype-specific entry and assembly events (Gottwein *et al*, 2009). Moreover, full-length infectious HCVcc were also generated for genotypes 1 and 2 (Yi & Lemon, 2009; Li *et al*, 2012b; Ramirez *et al*, 2014), allowing the dissection of all life cycle steps.

HCV_{TCP}

Trans-complementation systems consist in subgenomic replicons packed into single-round infectious particles. A JFH-1 subgenomic replicon providing the replication machinery of the virus and an expression system for the HCV proteins core, E1, E2, p7 and NS2 are required for HCV_{TCP} generation (Steinmann *et al*, 2008). HCV_{TCP} penetrate target cells in a CD81 receptor-dependent manner. This system can be helpful to decipher the mechanims of HCV assembly and to identify RNA elements and viral proteins involved in particle formation. This system is also a valuable tool for gene delivery or vaccination approaches (Steinmann *et al*, 2008). Although *trans*-complemented proteins can theoretically be derived from any patient isolate, providing isolate-specific information for HCV entry, replication and assembly, genetic incompatibility between the replication module and the packaging cassette is likely to limit efficacy of virus production when these units are derived from distinct genotypes (Gerold & Pietschmann, 2014).

5.4.2. Animals models

There are two categories of animal models : the naturally HCV permissive animal models and the mouse models.

Naturally HCV permissive animal models

So far, HCV was found to infect only few species other than human, *i.e*. chimpanzees and *Tupaia belangeri*.

❖ *Chimpanzees*

Chimpanzees can be chronically infected by HCV and experimental infection of these animals has played a crucial role in the discovery of the virus (Michael Houghton, 2009). Chimpanzees have also been very valuable for the evaluation of antiviral strategies (Vercauteren *et al*, 2014) and to date, chimpanzees are the only animals permitting extensive evaluation of the efficacy of potential vaccines against HCV (Houghton, 2011). Although studies of HCV infection in chimpanzees have led to a better understanding of innate and adaptive immune responses in the course of viral infection (Bowen & Walker, 2005; Rehermann, 2009), this model has several limitations : (i) biological variability between individual animals and the small animal cohorts lead to highly variable and difficult to interpret data (Mailly *et al*, 2013), (ii) chronic infection appears less frequently in infected chimpanzees than in humans (Lanford *et al*, 2001) , (iii) chronically infected animals don't readily develop cirrhosis or fibrosis and have much milder symptoms (Bukh *et al*, 2001), and (iv) hepatocellular carcinoma is rarely observed (Lanford *et al*, 2001). The limited avaibility of these animals, the cost for acquiring and maintaining them as well as ethic considerations have also hampered the use of this animal model. The use of chimpanzees for biomedical and behavioural research is now legally forbidden in Europe (Mailly *et al*, 2013).

❖ *Tupaia belangeri*

Tupaia belangeri, also called Northen treeshrew, is a non-rodent small mammal susceptible to HCV infection (Xie *et al*, 1998; Amako *et al*, 2010). Although this model has been used recently (Yang *et al*, 2013; Sun *et al*, 2013), the infection rate of tupaia is weak and viremia appears to be low and rarely sustained (Amako *et al*, 2010). The use of tupaia for the study of HCV pathogenesis and vaccine design is also hampered by the limited avaibility of these animals, their cost of housing and the absence of tupaia-specific reagents to assess HCV-host interactions in this model (Mailly *et al*, 2013).

Rodent models of HCV infection

The main rodent models are viral protein-HCV transgenic mice, the genetically humanized mouse model, human-liver xenograft mouse models (uPA-SCID and FRG mice with humanized liver) and immunocompetent xenograft models. These models are briefly described below and illustrated in **Figure II.13**.

❖ *Viral protein-HCV transgenic mice*

The most relevant transgenic mouse model is the FL-N/35 mouse, which expresses the complete viral polyprotein at physiological levels (Lerat *et al*, 2002). Hepatic steatosis, enhanced liver fibrosis and increased risk of HCC have been observed in this model (Lerat *et al*, 2002).

❖ *The genetically humanized mouse model*

This mouse model of HCV infection relies on genetic modifications of mice by introducing essential human specific factors for the viral life cycle. The first step of HCV infection is the viral entry which requires the presence of at least two human cell surface factors, CD81 and OCLN (Ploss *et al*, 2009). Adenoviral expression of human entry factors in mouse liver enables viral entry into murine hepatocytes *in vivo* (Dorner & Ploss, 2011). This humanized mouse model has limitations : (i) due to inefficient viral replication, there is no virus production by the infected mouse hepatocytes, which makes this model unsuitable for the evaluation of DAAs or antivirals targeting the assembly and egress steps of the viral life cycle, (ii) there is induction of an immune response against the adenoviral vectors used to introduce the human entry factors into mouse hepatocytes, which renders the study of HCV-induced immunopathogenesis in this model not possible (Mailly *et al*, 2013). However, it has been demonstrated that blunting the innate immune response allowed persistent infection and virus production (Dorner *et al*, 2013).

❖ *The uPA-SCID mouse with humanized liver*

Humanization of the mouse liver can be achieved by transplantation of primary human hepatocytes into immunodeficient mice that suffer from a constitutive or inducible liver injury (Mercer *et al*, 2001). The resulting human-liver chimeric mouse is currently the only reproducible system that supports robust infection of non-JFH based HCV clones and is the most widely used small animal model for the study of HCV (Vercauteren *et al*, 2014).

These immunodeficient (SCID) mice with hepatocyte-lethal phenotype due to the overexpression of the urokinase-type plasminogen activator (uPA) transgene in their liver can be efficiently engrafted with primary human hepatocytes in order to initiate infection with HCV (Mailly *et al*, 2013). Although immunodeficiency of the xenorecipient does not allow the study of host antiviral immune responses, the chimeric uPA-SCID

model has been efficiently used for the study of the basic aspects of the HCV life cycle, the evaluation of passive immunization strategies and the assessment of novel antiviral therapies (Vercauteren *et al*, 2014). This model has limitations : (i) these mice are very fragile and have to be engrafted within the first weeks of life (Meuleman *et al*, 2005), (ii) the uPA transgene can be deleted in some mice leading to the restoration of a wild type phenotype and thus loss of the human hepatocyte graft (Sandgren *et al*, 1991), and (iii) this model lacks a functional adaptive immune system and can't be used for the study of adaptive immune responses or for the evaluation of vaccines (Mailly *et al*, 2013).

❖ *The FRG mouse with humanized liver*

This model overcomes some of the limitations of the uPA-SCID mouse model : (i) these mice have a more profound immunodeficiency, (ii) the time of transplantation of human hepatocytes is easier to control upon drug withdrawal, and (iii) spontaneous reversion of the hepatocyte-lethal phenotype does not occur since there is a full deletion within the Fah (tyrosine catabolic enzyme fumarylacetoacetate hydrolase) encoding gene (Mailly *et al*, 2013).

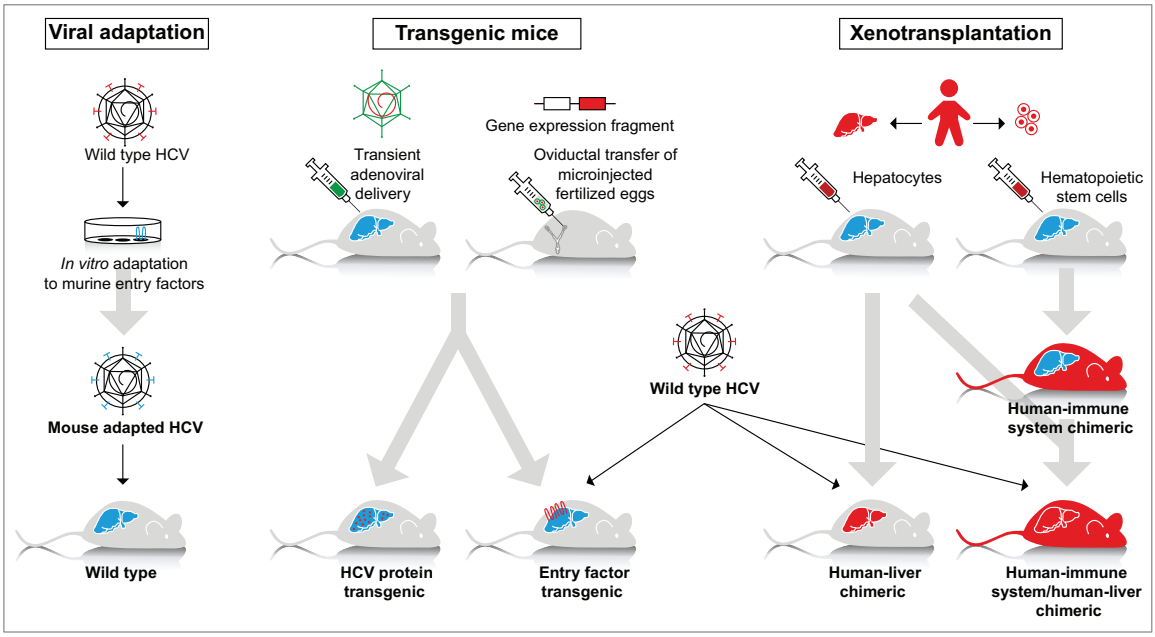


Figure II.13 ***Different approaches to create mouse models for the study of HCV.*** From (Vercauteren *et al*, 2014). Left panel : In vitro adaptation of HCV to mouse hepatoma cells may allow the isolation of viral variants that establish an infection in wild type mice. Middle panel : Transient or stable expression of viral genes may provide essential information on viral protein-host interactions. In addition, mice can be made transgenic for human factors that are essential to support infection of wild type HCV. Right panel : In xenotransplantation models, the genetic background of the host permits repopulation of the liver upon transplantation of humanhepatocytes. Additional transplantation of HLA-compatible hematopoietic stem cells may result in dually reconstituted mice.

❖ *Immunocompetent xenograft models*

One example of immunocompetent xenograft model is a rat model, in which primary human hepatocytes or Huh-7 human hepatoma cells are injected into the peritoneal cavity of fœtal rats when the rat fœtal system is still in development (Ouyang *et al*, 2001). However, the immune system of the immunocompetent rat model does not match the HLA molecules at the surface of the transplanted human hepatoma cells, hampering the study of adaptative rat immune responses against infected human hepatoma cells (Mailly *et al*, 2013; Vercauteren *et al*, 2014). An alternative approach is the development of mice engrafted with both human hepatocytes and human immune cells (Gutti *et al*, 2014).

6. Hepatitis C virus membrane proteins

As described previously in section II.5.2.3., the HCV genome encodes a polyprotein which is co- and posttranslationally cleaved into ten membrane-associated proteins (**Figure II.14**). However, the knowledge of the three-dimensional structure of these proteins and their role in virus replication is limited. Whereas cell culture and animals models (described previously in section II.5.4.) allow to better understand the viral life cycle and immunity responses, detailed information about the three-dimensional structure of HCV proteins is essential to better elucidate their respective function.

The current knowledge about the three-dimensional structure of HCV proteins and the relationship between their structure and function is described in this section. The current work aims to provide ways to get more insight into these aspects at the atomic level for core, NS2, NS4B and NS5A proteins.

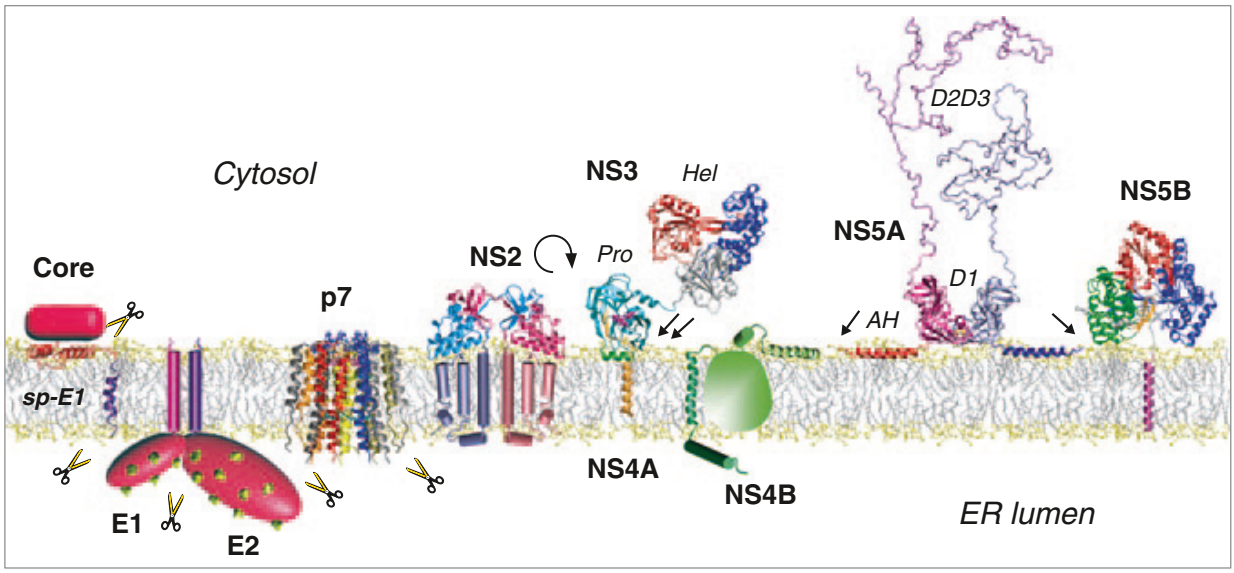


Figure II.14 ***Structure and membrane association of HCV proteins.*** From (Moradpour & Penin, 2013).

6.1. Core

The core protein is the first structural protein encoded by the HCV ORF and forms the viral nucleocapsid. Cleavage by signal peptidase at the junction core/E1 yields an immature core protein of 191 amino acid residues (**Figure II.15**). An internal signal sequence located between core and E1 targets the nascent polypeptide to the translocon of the ER membrane, followed by translocation of the E1 ectodomain into the ER lumen (Santolini *et al*, 1994; Okamoto *et al*, 2008). Further C-terminal processing by signal

peptide peptidase yields the mature 21 kDa core protein of approximately 177 amino acid residues (McLauchlan *et al*, 2002; Okamoto *et al*, 2008; Oehler *et al*, 2012). Although mutagenesis and *trans*-complementation experiments suggest that at least 177 amino acid residues are required for infectious particle production, the precise C-terminus of mature core has not yet been identified unequivocally (Kopp *et al*, 2010).

Mature core is a dimeric membrane protein stabilized through disulfide bond formation at Cys 128 (Boulant *et al*, 2005; Kushima *et al*, 2010). Core is composed of two domains : a N-terminal hydrophilic domain, D1 (aa 1 to 117), and a C-terminal hydrophobic domain, D2 (aa 117 to ~177) (Boulant *et al*, 2005). The signals and processes that mediate RNA packaging and the assembly of core into nucleocapsids are however largely unknown (Moradpour & Penin, 2013). D1 contains a high proportion of basic amino acid residues and has been implicated in both RNA binding and homo-oligomerization, promoting thereby nucleocapsid assembly. D1 exhibits RNA chaperone like activity, a general feature of nucleocapsid proteins required for the structural remodeling and packaging of the RNA genome in the viral particle (Cristofari, 2004). D1 may contribute to alterations of host cell functions upon HCV infection through interactions with numerous cellular factors (de Chassey *et al*, 2008). Although it contains several potential α-helices, isolated D1 behaves as an intrinsically unstructured protein (Boulant *et al*, 2005). Moreover, D1 folds upon interaction with the C-terminal hydrophobic domain D2, mediating association with LDs (Boulant *et al*, 2005; Boulant *et al*, 2006). D2 is composed of two amphipathic α-helices (aa 119 to 136 and aa 148 to 165) connected by a central hydrophobic loop and interacting in-plane with the LD phospholipid interface (Boulant *et al*, 2006). The trafficking of functional core to putative assembly sites and its interaction with viral proteins have been analyzed by fluorescent tagging and live cell imaging (Counihan *et al*, 2011; Coller *et al*, 2012). The association of core with LDs plays a central role in nucleocapsid assembly (Shavinskaya *et al*, 2007; Boulant *et al*, 2007; Miyanari *et al*, 2007). Its interaction with NS5A is also crucial for nucleocapsid assembly since NS5A is thought to deliver the HCV genome RNA to core protein. The membrane anchor of NS5A is indeed able to associate with the phospholipid monolayer of LDs and RNA transfer likely occurs on LDs or the ER-LD interface (Lindenbach, 2013). Due to the particular features of the core membrane-binding domain, the resulting RNA-containing nucleocapsid is expected to be surrounded by a lipid monolayer, allowing its full immersion into the hydrophobic core of LDs and ultimately of

LVPs (Bartenschlager *et al*, 2011).

The HCV genome presents alternative reading frames overlapping the core coding region as well as internal translation initiation events, leading to several F (frameshift) or ARFP (alternative reading frame protein) proteins (reviewed in (Branch *et al*, 2005)). However, these proteins do not seem to be essential for viral RNA replication (Vassilaki *et al*, 2008). In addition, a family of minicore proteins, ranging in size from 8 kDa to 14 kDa and whose role in the virus life cycle and pathogenesis remains to be elucidated, has also been detected (Eng *et al*, 2009).

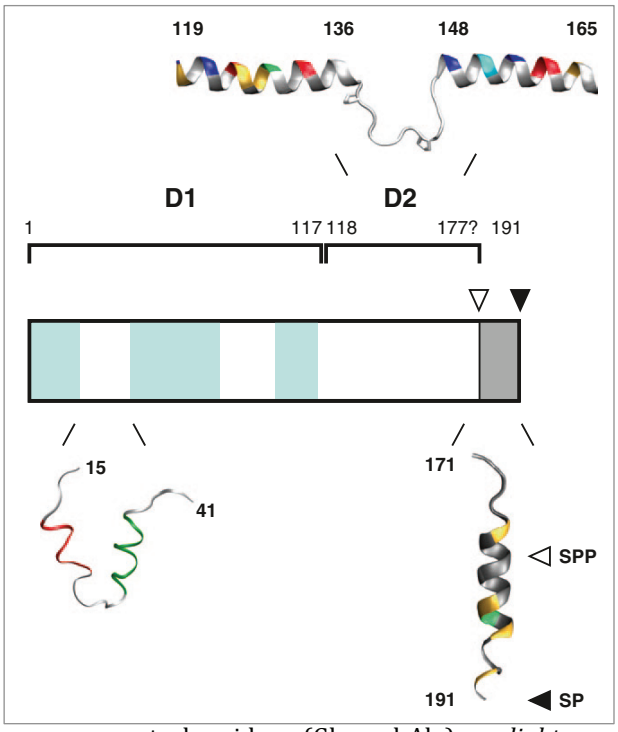


Figure II.15 ***Schematic representation of the HCV core protein.*** From (Moradpour & Penin, 2013). The two domains composing the mature core protein are represented here, together with the core-E1 signal peptide. The N-terminal domain D1 (aa 1 to 117) contains three hydrophilic, highly basic segments (*blue boxes*) which are separated by two more hydrophobic segments. When isolated, D1 is an intrinsically unstructured protein but includes several potential α-helices, notably a helix-loop-helix motif (aa 15 to 41, PDB entry ICWX, ribbon representation (Angus *et al*, 2011)). The α-helices 1 (aa 19 to 24) and 2 (aa 30 to 37) are *red* and *green*, respectively. The C-terminal domain D2 (aa 118 to ~177) consists of a central hydrophobic loop which connects two amphipathic α-helices (aa 119 to 136 and 148 to 165, ribbon representation (Boulant *et al*, 2006)). The core-E1 signal peptide at the C-terminus of core (PDB entry 2KQI (Oehler *et al*, 2012)) is cleaved by signal peptidase (SP, after aa 191, *solid arrowheads*) and signal peptide peptidase (SPP, after aa ~177, *open arrowheads*) to yield mature core protein. Hydrophobic residues are *grey*, neutral residues (Gly and Ala) are *light grey* and hydrophylic residues are colored accordingly to their physico-chemical properties : Ser, Thr, Asn and Gln in *yellow*, Asp and Glu in *red*, Arg and Lys in *blue*, His in *cyan* and Cys in *green*.

6.2. E1 and E2

The envelope glycoproteins E1 and E2 are released from HCV polyprotein by signal peptidase cleavages. Both proteins are essential for virus entry since they interact with entry receptors (Lavie *et al*, 2007). Whereas there is evidence for a direct interaction between E2 and CD81, E1 and E2 might interact with other entry factors via LVP (Lavie *et al*, 2015). E1 and E2 mediate the fusion with endosomal membranes (Cocquerel, 2006; Zeisel *et al*, 2013), and also play an essential role in virion assembly (Wakita *et al*, 2005).

E1 and E2 are type I membrane proteins with a highly glycosylated N-terminal ectodomain (composed of about 160 and 360 amino acid residues for E1 and E2, respectively) and a C-terminal hydrophobic anchor composed of about 30 amino acid residues (Lavie *et al*, 2007). Recently published crystal structures of the N-terminal part of E1 (Omari *et al*, 2014) and a large portion of E2 ectodomain (Kong *et al*, 2013; Khan *et al*, 2014) are helping to understand the functions of HCV envelope glycoprotein complex. E1 and E2 associate as non-covalent heterodimers (Bartosch *et al*, 2003; Op De Beeck *et al*, 2004) and heterogenous complexes formed by disulfide bonds (Vieyres *et al*, 2010). Moreover, their transmembrane domains play a major role in membrane anchoring, subcellular localization at the ER membrane and E1E2 heterodimer assembly which is believed to constitute the building block for the viral envelope (Op De Beeck *et al*, 2004; Lavie *et al*, 2015). It was recently demonstrated that functional E1 and E2 form transmembrane-dependent trimers at the surface of HCV particles (Falson *et al*, submitted).

E1 and E2 maturation and folding is a complex and interdependent process involving the ER chaperone machinery, disulfide bond formation (highly conserved cysteine residues in the ectodomains of E1 and E2 may form 4 and 9 disulfide bonds, respectively) and glycosylation (up to 6 and 11 glycosylation sites for E1 and E2, respectively) (Moradpour & Penin, 2013).

6.3. p7

p7 is a hydrophobic polypeptide of 63 amino acid residues, composed of two transmembrane α-helices linked by a small positively charged cytosolic loop, both N- and C-terminus being oriented towards the ER lumen (Carrere-Kremer *et al*, 2002). Since it forms hexamers or heptamers with cation channel activity, p7 is considered as a viroporin (Chandler *et al*, 2012; OuYang *et al*, 2013; Moradpour & Penin, 2013). A three-dimensional model of the monomer in a phospholipid bilayer was established based on combined NMR experiments and molecular dynamics simulations (Montserret *et al*, 2010). p7 is not required for viral RNA replication but plays an essential role in virion assembly and envelopment (Steinmann & Pietschmann, 2010; Gentzsch *et al*, 2013). It interacts with the viral proteins core, E1 and E2, as well as NS2 (Jirasko *et al*, 2010; Boson

et al, 2011; Moradpour & Penin, 2013; Atoom *et al*, 2014). The interaction between p7 and NS2 has been shown to be crucial for production of infectious HCV particules in cell culture (Vieyres *et al*, 2013).

6.4. NS2

NS2 is a 23 kDa protein of 217 amino acid residues, composed of a N-terminal membrane domain and a C-terminal cytosolic domain. The N-terminal membrane domain comprises three putative transmembrane segments (aa 4 to 23, 27 to 49 and 72 to 94) and a small α-helix (aa 61 to 70) whose atomic structures have been solved by NMR, allowing to propose a topoly model of full-length, membrane-associated NS2 (Jirasko *et al*, 2008; Jirasko *et al*, 2010) (**Figure II.16**).

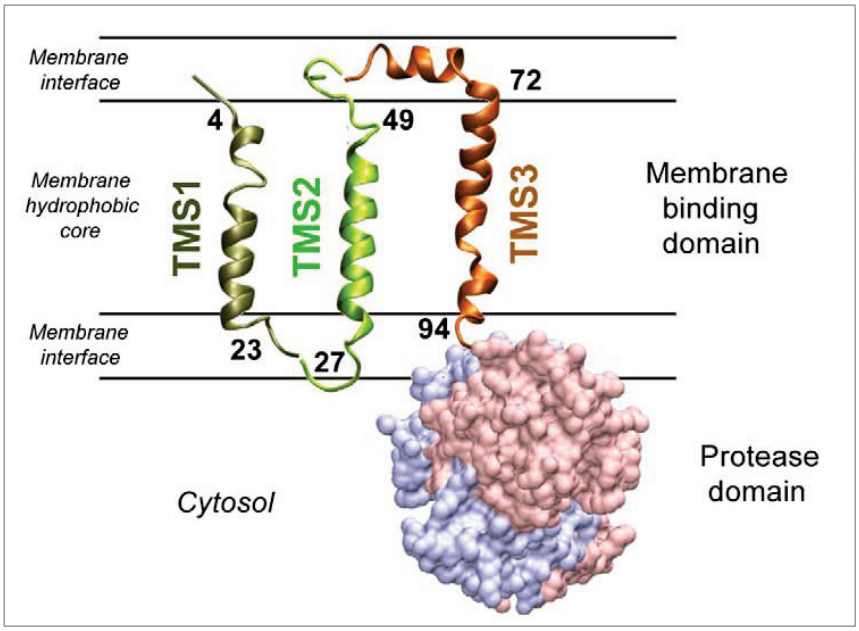


Figure II.16 ***Model of NS2 membrane topology*.** From (Jirasko *et al*, 2010). Transmembrane segments TMS1, TMS2 and TMS3 are shown in ribbon representation and colored *bronze*, *green* and *copper*, respectively. The TMS are tentatively positioned in the membrane and the limits of transmembrane helices are given (TMS1 : 4 to 23, TMS2 : 27 to 49, TMS3 : 72 to 94). The three TMS are represented as separated entities, since their intramolecular and/or intermolecular interactions are not known. The NS2 protease domain is shown in surface representation (side view) with dimer subunits (PDB entry 2HD0) shown in *light blue* and *pink*. For simplicity, the membrane domain is shown for only one protease subunit. The phospholipid bilayer is tentatively and schematically represented.

The C-terminal cytosolic domain encodes a cysteine protease which cleaves the polyprotein precursor at the NS2/NS3 junction and whose function is strongly enhanced by the N-terminal one-third of NS3 (Schregel *et al*, 2009). Protease activity itself is

dispensable for RNA replication but cleavage at the NS2/NS3 junction is essential to liberate fully functional NS3 protein, and thus promote viral RNA replication. The catalytic domain of NS2, called $NS2^{pro}$ (aa 94 to 217), involves a catalytic triad composed of His 143, Glu 163 and Cys 184. The crystal structure of $NS2^{pro}$ revealed a dimer with two composite active sites (Lorenz *et al*, 2006). Each active site is composed of residues from the two monomers : catalytic residues His 143 and Glu 163 are contributed by one monomer and Cys 184 by the other. Moreover, it has been shown recently that $NS2^{pro}$ itself associates with cellular membranes, a single charged residue in the second α-helix of $NS2^{pro}$ being required for proper membrane association, NS2 protein stability and efficient HCV polyprotein processing (Lange *et al*, 2014).

In addition to its implication in RNA replication, NS2 plays a central role in the organization of HCV infectious virus assembly. This function is independent of its protease activity and may involve a complex network of interactions with structural and other non-structural viral proteins : E1 (mediated by E2), E2, p7, NS3 and NS5A (Jirasko *et al*, 2010; Boson *et al*, 2011; Stapleford & Lindenbach, 2011; Ma *et al*, 2011; Popescu *et al*, 2011).

6.5. NS3-4A

NS3-4A is a non-covalent complex constituted of NS3 and the cofactor NS4A. NS3 is a multifunctional protein with a serine protease in the N-terminal one-third (aa 1 to 180) and an NTPase/RNA helicase in the C-terminal two-thirds (aa 181 to 631) (Moradpour & Penin, 2013). The NS3 NTPase/RNA helicase uses energy from ATP hydrolysis to unwind single- or double-stranded RNA regions with extensive secondary structures. The NS3 helicase is essential for RNA replication and also plays a role in viral particle assembly (Murray *et al*, 2008). NS4A is a 54-aa polypeptide which functions as a cofactor for the NS3 serine protease. Its N-terminal hydrophobic region forms a transmembrane α-helix required for the integral association of the NS3-4A complex (Brass *et al*, 2008). In addition, NS4A regulates positively NS5A hyperphosphorylation and thus viral replication (Lindenbach *et al*, 2007). NS4A interacts with other replicase components through a highly negatively charged α-helix in its C-terminal acidic region and contribute to RNA replication and virus particle assembly (Lindenbach *et al*, 2007; Moradpour & Penin, 2013).

The NS3-4A protease adopts a chymotrypsin-like fold with two β-barrel subdomains. A Zn^{2+}, coordinated by Cys 97, Cys 99, Cys 145 and His 149, stabilizes the structure. This Zn^{2+} binding site plays an important role in facilitating the processing of the NS2/NS3 junction by the NS2 protease (Schregel *et al*, 2009). The NS3 catalytic triad is composed by His 57, Asp 81 and Ser 139 (Moradpour & Penin, 2013). An in-plane amphipathic α-helix at the N-terminus of NS3 together with the transmembrane domain of NS4A ensure membrane association of the NS3-4A complex (Brass *et al*, 2008).

In addition to the viral proteins downstream of NS3, NS3-4A protease cleaves two crucial adaptor proteins in innate immune sensing : the RIG-I adaptor MAVS (Meylan *et al*, 2005; Tasaka *et al*, 2007), a mitochondrial host factor, and the TLR3 adaptor TRIF (Li *et al*, 2005). T cell protein tyrosine phosphatase, a modulator of the epidermal growth factor receptor, is cleaved by NS3-4A too (Brenndörfer *et al*, 2009). More recently, NS3-4A was also shown to interact with the membrane-associated peroxidase GPx8, revealing that GPx8 is involved in viral particle production but not in HCV entry or RNA replication (Morikawa *et al*, 2013).

NS3-4A protease is involved not only in the replication but also in the persistence and pathogenesis of HCV (Morikawa *et al*, 2011). It has therefore been identified as an important therapeutic target and two inhibitors of NS3 activity, telaprevir and boceprevir, have been licensed. They are currently used in triple therapy, combined with pegylated IFNα and ribavirin (Salloum & Tai, 2012; Pawlotsky, 2013). Second-generation NS3-4A protease inhibitors like faldaprevir, asunaprevir or danoprevir are currently in phase II or III clinical development (Pawlotsky, 2014).

6.6. NS4B

NS4B is a 27kDa, poorly characterized, integral membrane protein of 261 amino acid residues, with an N-terminal region (aa 1 to ~69), a central part harboring four predicted transmembrane passages (aa ~70 to ~190) and a C-terminal region (aa ~191 to 261) (Gouttenoire *et al*, 2010a). Membrane association of NS4B is mediated not only by transmembrane domains in its central part but also by determinants for membrane association in the N- and C-terminal regions (Moradpour & Penin, 2013). Indeed, the N-terminal region is composed of two amphipathic α-helices, AH1 (aa 4 to 32) and AH2 (aa

42 to 66). AH1, whose three-dimensional structure was reported recently, possesses a dual role in RNA replication and virus production, potentially governed by different topologies of the N-terminal part of NS4B (Gouttenoire *et al*, 2014). AH2 has the potential to traverse the phospholipid bilayer as a transmembrane segment, likely upon oligomerization, and plays an important role in the assembly of a functional replication complex (Gouttenoire *et al*, 2009a). The membrane topology of the N-terminal region of NS4B may be dynamic and modulated by protein-protein interactions within the HCV replication complex (Lundin *et al*, 2006; Gouttenoire *et al*, 2009a) (**Figure II.17**). The C-terminal region is composed of two amphipathic α-helices, H1 (aa 201 to 213) which is highly conserved, and H2 (aa 229 to 253) which is associated to the membrane (Gouttenoire *et al*, 2009b). The C-terminal region was also reported to comprise two palmitoylation sites (Yu *et al*, 2006) but this was not confirmed by other teams (Paul *et al*, 2015).

NS4B is a main inducer of the membranous web formation, a specific membrane alteration consisting of locally confined membranous vesicles that serves as a scaffold for the HCV replication complex (Egger *et al*, 2002; Gosert *et al*, 2003). NS4B interacts with other viral non-structural proteins, was reported to bind viral RNA (Einav *et al*, 2008), harbors NTPase activity (Einav *et al*, 2004; Thompson *et al*, 2009) and plays a role in virion assembly (Jones *et al*, 2009). NS4B interferes also with the RIG-I pathway (Tasaka *et al*, 2007; Nitta *et al*, 2013). Obtaining a three-dimensional structure of full-length NS4B would allow to gain further insight into the multiple functions of this protein which might be governed by distinct membrane topologies and/or interactions with other viral and cellular proteins (Moradpour & Penin, 2013).

Similar to the other HCV non-structural proteins, NS4B has been reported to form oligomers (Gouttenoire *et al*, 2010a). The formation of at least trimers was indeed observed in cross-linking studies (Yu *et al*, 2006). Moreover, oligomerization of NS4B in the membrane environment of intacts cells was observed in fluorescence resonance energy transfer and co-immunoprecipitation experiments (Gouttenoire *et al*, 2010b). Several conserved determinants were found to be involved in NS4B oligomerization through homotypic and heterotypic interactions. N-terminal amphipathic α-helix AH2 and C-terminal conserved elements were identified as major determinants for NS4B oligomerization (Gouttenoire *et al*, 2010b; Paul *et al*, 2011). Mutations affecting the

oligomerization of NS4B disrupt the membranous web formation and HCV RNA replication, which means that oligomerization of NS4B is required for the creation of a functional replication complex, likely through the induction of membrane curvature and vesicle formation (Moradpour & Penin, 2013).

As a master organizer of HCV replication complex formation, NS4B has recently emerged as a potential alternative target for DAAs therapeutic strategies (Rai & Deval, 2011).

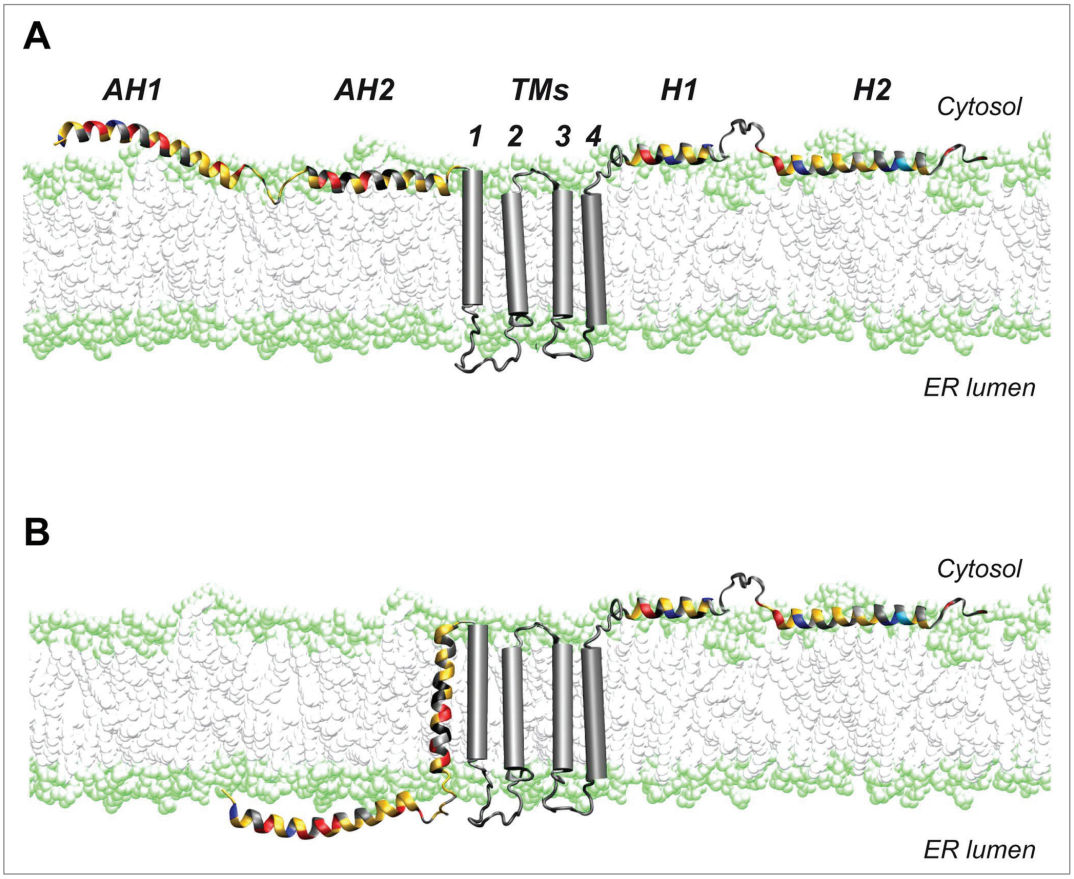


Figure II.17 ***Dual membrane topology of HCV NS4B.*** The N-terminal part of NS4B assumes a dual membrane topology : (A) cytosolic membrane topology, (B) luminal endoplasmic reticulum membrane topology. TMs, predicted transmembrane segments. From (Gouttenoire *et al*, 2014).

6.7. NS5A

NS5A is a monotopic membrane phosphoprotein, composed of 447 amino acid residues, which plays an important role in membranous web formation (Romero-Brey *et al*, 2012), modulation of HCV RNA replication and particle formation (Lindenbach, 2013).

NS5A is composed of a N-terminal membrane anchor (AH) (Brass *et al*, 2002) and three domains (D1, D2 and D3) separated by two low complexity sequences (LCS 1 and LCS 2) (Tellinghuisen *et al*, 2004) (**Figure II.18**). D1 (aa 36 to 231) and D2 (aa 250 to 342) are primarily involved in RNA replication whereas D3 (aa 356 to 447) is essential for viral assembly (Tellinghuisen *et al*, 2008; Appel *et al*, 2008; Kim *et al*, 2011). In addition, D1 is involved in LD binding (Miyanari *et al*, 2007) and D3 in interaction with the core protein (Masaki *et al*, 2008).

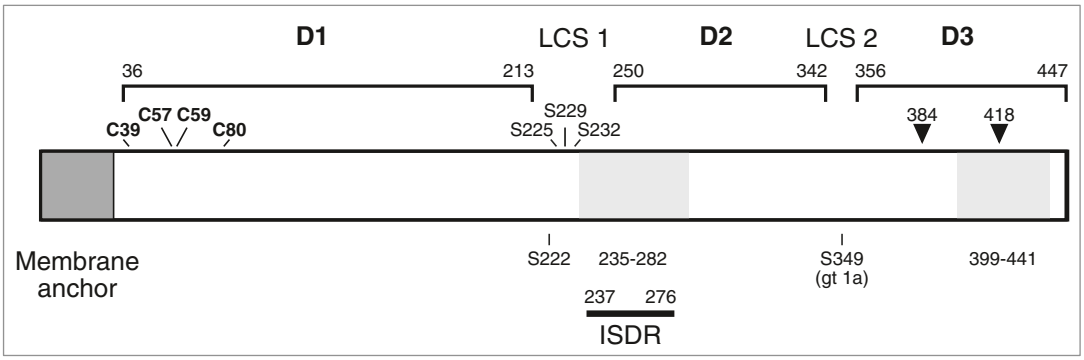


Figure II.18 ***Schematic representation of the HCV NS5A.*** From (Moradpour & Penin, 2013). NS5A is drawn to scale as a *box*, the domain organization shown being the one proposed by Tellinghuisen (Tellinghuisen *et al*, 2004). Amino acid positions relate to the HCV Con1 sequence (genoype 1b, GenBank accession number AJ238799). Domains D1, D2 and D3 are connected by low complexity sequences (LCS) 1 and 2. Cysteine residues 39, 57, 59 and 80 coordinate one zinc atom per NS5A protein. The N-terminal amphipathic α-helix mediating membrane association of NS5A is highlighted by a *grey box*. Phosphoacceptor sites mapped for genotype 1b (Ser 222 (Katze *et al*, 2000)) and genotype 1a HCV isolates (Ser 349 (Reed & Rice, 1999)), as well as serine residues 225, 229 and 232 affecting hyperphosphorylation of NS5A, are indicated. Two examples each of deletions (Blight *et al*, 2000) (Appel *et al*, 2005) and green fluorescent protein insertions (Moradpour *et al*, 2004b) found to be tolerated with respect to HCV RNA replication are indicated by *light gray boxes* and *arrowheads*, respectively. ISDR, interferon sensitivity determining region.

NS5A is anchored to the membrane by an N-terminal amphipathic α-helix embedded in-plane into the cytosolic leaflet of the membrane, allowing its association with a phospholipid monolayer and thus its interaction with core on LDs or the LD-ER interface (Penin *et al*, 2004b). D1 is a highly structured domain which forms stable homodimers. Two X-ray crystal structures of NS5A D1 revealed virtually identical conformations of the monomer, but distinct dimer organizations which have been proposed to form multimeric complexes (Tellinghuisen *et al*, 2005; Love *et al*, 2009; Lambert *et al*, 2014) (**Figure II.19**). Each monomer coordinates one zinc atom through the four fully conserved cysteine residues, Cys 39, Cys 57, Cys 59, and Cys 80. The first reported dimer organization is a « clam-like » structure with a groove that faces away from the

membrane and has been reported to bind either single- or double- stranded RNA (Tellinghuisen *et al*, 2005) (**Figure II.19c**). Multiple NS5A dimers of the « clam-like » structure may form a « basic railway » on intracellular membranes that would allow tethering as well as sliding of the viral RNA on intracellular membranes and coordination of its different fates during HCV replication (Moradpour *et al*, 2005). The second dimer organization is a « back-to-back » structure (Love *et al*, 2009) (**Figure II.19b**). The theoretical assembly of NS5A D1 based on simultaneous recruitment of the two monomer-monomer interfaces would results in a superhelical array exhibiting alternation between « clam-like » and « back-to-back » structures (Love *et al*, 2009). The recent description of « head-to-head » and « head-to-tail » dimeric structures suggested more complex assembly of NS5A D1. Such NS5A network could potentially form part of the membranous web and explain the multifaceted role of NS5A (Lambert *et al*, 2014).

D2 and D3 domains are natively unfolded monomeric conformers with limited sequence conservation between genotypes (Liang *et al*, 2007; Hanoulle *et al*, 2009; Verdegem *et al*, 2011). Their fast interconversion, with an intrinsic α-helical propensy, suggest that nascent secondary structures constitute molecular recognition elements and promote the interaction and stabilization of the conformations with specific viral or host proteins (Moradpour & Penin, 2013). The natively unfolded nature of D2 and D3 leads to a high number of viral and cellular interactants for NS5A which therefore appears to be a hub for protein interactions with high specificities and low affinities (de Chassey *et al*, 2008). For example, the interaction between NS5A and NS5B is critical for HCV RNA replication (Shirota *et al*, 2002; Shimakami *et al*, 2004). In addition, D2 and D3 were found to interact directly with the active site of cyclophilin A (CypA), proline residues in D2 and D3 serving as substrate for the peptidyl-prolyl *cis/trans* activity of CypA (Hanoulle *et al*, 2009; Verdegem *et al*, 2011) (**Figure II.19a**). Through interactions with NS5A, CypA seems to play an essential, although not exactly defined, role in HCV RNA replication and assembly (Kaul *et al*, 2009; Liu *et al*, 2009b). NS5A also interacts with phosphatidylinositol 4-kinase IIIα (PI4KIIIα), stimulating its activity and phosphatidylinositol 4-phosphate (PI4P) production which appears to be essential for HCV replication complex formation (Reiss *et al*, 2011; Berger *et al*, 2011; Lim & Hwang, 2011). Finally, NS5A has a potential role in modulating the response to IFNα therapy (Enomoto *et al*, 1996; Pawlotsky & Germanidis, 1999).

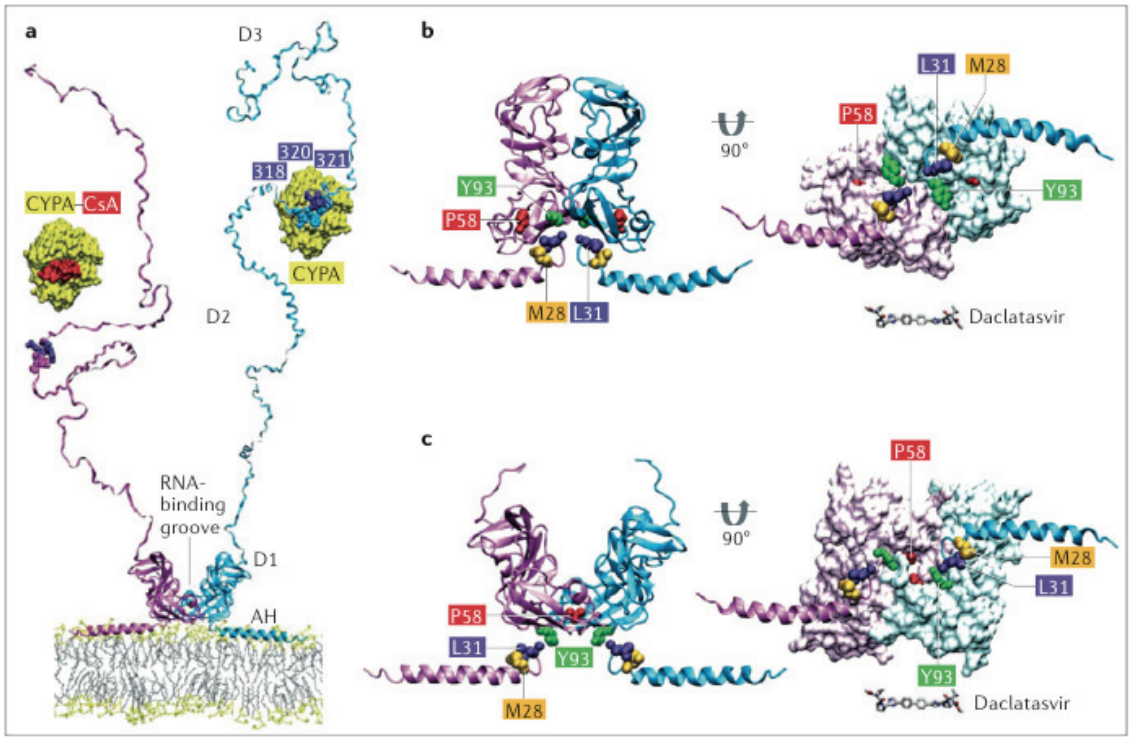


Figure II.19 ***Structure of the NS5A dimer*****.** From (Bartenschlager *et al*, 2013). (a) Ribbon diagram of the full-length NS5A dimer associated with a phospholipid membrane. Each subunit, in *lilac* and *cyan* respectively, is composed by the N-terminal amphipathic α-helix AH, the highly structured domain D1 and the intrinsically unfolded domains D2 and D3. A surface representation of CypA in complex with cyclosporin A is shown on the upper left, and a putative structure of CypA with its main binding site in NS5A D2 is shown on the upper right. All domains and structures are drawn to scale to illustrate the length and flexibility of the « arms » formed by D2 and D3 and the relative sizes of CypA, cyclosporin A and NS5A. (b) « back-to-back » and (c) « clam-like » structures of the NS5A dimer, respectively. The positions of mutations confering resistance to NS5A inbitors are indicated. The left panels show side views, the right panels views of the putative membrane-interacting surface. The position of AH relative to D1 in the right panels is arbitrary and assumes that resistance mutations observed in D1 and AH would be close to each other. The structure of daclatasvir is shown as stick representation, drawn to scale, in the right panels.

Although NS5A was for a long time thought to be not druggable because it does not possess any known enzymatic activity, it became a major target for the current development of many DAAs and HTAs. Indeed, interfering either with viral and cellular proteins recruited by NS5A or with their binding to NS5A could exert antiviral activity. As an example, daclatasvir or BMS-790052 is a compound defined as a highly potent inhibitor used in clinical trial for treatment of HCV infection (Bartenschlager *et al*, 2013). Recently, it was shown that daclatasvir-like inhibitors of NS5A block early biogenesis of HCV-induced membranous replication factories, independently of RNA replication (Berger *et al*, 2014).

NS5A is found in a basal phosphorylated form (56 kDa) and a hyperphosphorylated form (58 kDa). Basal phosphorylation occurs in central and C-terminal parts of NS5A while serine residues 225, 229, and 232 in LCS 1 are important for NS5A hyperphosphorylation. There is evidence that the phosphorylation state of NS5A modulates HCV RNA replication, possibly via regulating interactions with replication-*versus* assembly-specific host factors (Neddermann *et al*, 2004; Evans *et al*, 2004; Appel *et al*, 2005). Serine residues in LCS 1 can be phosphorylated by the α isoform of casein kinase I (CKIα) (Quintavalle *et al*, 2007) whereas Ser 457 in D3 (genotype 2 JFH-1 strain) can be phosphorylated by casein kinase II (CKII) (Tellinghuisen *et al*, 2008). CKII-mediated phosphorylation of Ser 457 was found to be essential for assembly of infectious virus particles in the case of JFH-1 isolate (Tellinghuisen *et al*, 2008). Additional kinases might be involved in generating the different phosphoforms of NS5A (Huang *et al*, 2007).

6.8. NS5B

The NS5B RNA-dependent RNA polymerase (RdRp) is responsible for both the synthesis of a complementary negative-strand RNA using as template genomic positive-strand RNA and the subsequent synthesis of genomic positive-strand RNA from this negative-strand RNA template. This enzyme, composed of 591 amino acid residues (68 kDa), has been extensively characterized and has emerged as a major target for antiviral strategies (Moradpour & Penin, 2013).

NS5B is a tail-anchored integral membrane protein whose the C-terminal 21 amino acid residues mediate posttranslational targeting to ER membrane, with a cytosolic orientation of the catalytic domain (Schmidt-Mende *et al*, 2001). Although the NS5B membrane anchor is not necessary for polymerase activity *in vitro*, it is required for RNA replication in cells (Moradpour *et al*, 2004a). The N-terminal first 530 amino acid residues form the catalytic domain which exhibits typical features of RdRps : fingers, palm and thumb subdomains (Bressanelli *et al*, 1999), as well as a Gly-Asp-Asp motif (Behrens *et al*, 1996). Fingers and thumb subdomains closely interact and encircle the NS5B active site which is occluded by a 40-aa linker between the catalytic domain and the C-terminal membrane anchor. NS5B requires only divalent ions as cofactors

(magnesium ou manganese) for template-directed RNA synthesis : two metal ions are chelated by the catalytic aspartic acid residues Asp 220 and Asp 318 located in the palm subdomain (Moradpour & Penin, 2013). The single-strand RNA template binds in a groove between the fingers and thumb subdomains which leads directly to the active site. NTPs access this site via a specific tunnel beginning at the backside periphery and extending into the active site located in the palm subdomain (Moradpour & Penin, 2013). Compared with other RdRps, NS5B has a rather closed structure of the active centre which is supposed to represent the initiation state of the polymerase. To switch from initiation to processive elongation, an opening of this structure is required to generate a large cavity capable of binding double-stranded RNA (Mosley *et al*, 2012).

NS5B has no proofreading activity which leads to the high genomic variability of HCV. NS5B interacts with the viral protein NS5A through the same binding site than CypA (Rosnoblet *et al*, 2012). NS5B also interacts with several host factors such as the lipid kinase PI4KIIIα which is essential for RNA replication (Reiss *et al*, 2011), or the retinoblastoma susceptibility protein pRb which blocks RdRp activity by overlaping with the polymerase active site of NS5B, leading to ubiquitination and subsequent degradation of pRb and probably contributing to the oncogenic property of HCV (Marascio *et al*, 2014).

NS5B inhibitors are divided into nucleoside inhibitors (NIs) and non-nucleoside inhibitors (NNIs). NIs, such as sofosbuvir which is the most powerful inhibitor of HCV replication, bind into the active site and are terminator substrates of the RNA polymerase activity. NNIs, such as BI-207127, bind into one of the allosteric sites of the polymerase (Marascio *et al*, 2014).

Recently, various crystal structures of stalled polymerase ternary complexes with enzymes, RNA templates, RNA primers, incoming nucleotides and catalytic metal ions, during both primed initiation and elongation of RNA synthesis were solved, allowing a better mechanistic understanding of NS5B (Appleby *et al*, 2015).

III. MATERIAL AND METHODS

1. Plasmids

During this work, I prepared 19 different constructs (**Table III.1**) according to the protocol detailed below. Further plasmids were provided by our collaborators, as indicated in **Table IV.1**.

1.1. Primer design

In order to efficiently express the target protein, CellFree Sciences recommends to select a restriction enzyme site as near as possible to the E01 translational enhancer. On the pEU-E01-MCS vector (3747 bp, **Figure II.5**), considering that the position 1 is located at the final G of the SP6 promoter, the SP6 promoter is located at position -17 to 1, the translation enhancer E01 at position 1 to 73 and the Multiple Cloning Site (MCS, **Figure III.1**) at position 74 to 193.

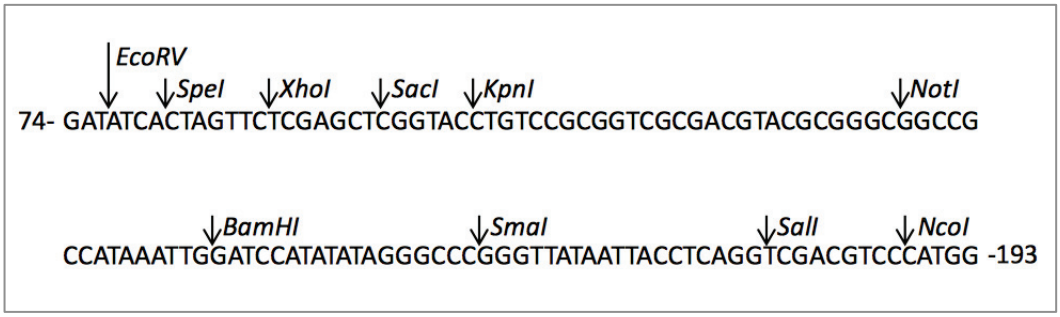


Figure III.1 ***Multiple Cloning Site (MCS) of the pEU-E01-MCS vector***

Since SpeI is near to the E01 translational enhancer and allows to have a good Kozak sequence, it was used for all constructs. The *Strep*-tag II sequence was added during PCR and primers were designed accordingly. Moreover, some nucleotides were added on both restriction sites in order to enhance the restriction enzyme activity. The primers used for this work are listed in **Table III.1** (bottom table).

1.2. Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) aims to amplify a gene of interest and consists of repeated cycles, each cycle comprising three temperature steps : denaturation, annealing and elongation. The denaturation step, usually performed at 95°C, leads to the disruption of hydrogen bonds between complementary bases, causing DNA melting of the template and thus yielding single-stranded DNA. During the

annealing step, the primers anneal to the single-stranded DNA template. The annealing temperature, usually 5°C below the Tm (melting temperature) of the primers, should be low enough to allow the hybridization of the primer to the single DNA strand, but high enough for specific hybridization. The polymerase binds then to the primer-template hybrid and starts DNA formation. During the elongation step, the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTPs that are complementary to the template in the 5' to 3' direction. The DNA polymerase can polymerize a thousand bases per minute during each elongation step in which the amount of DNA target is doubled, leading to exponential amplification of the specific DNA fragment. A single temperature step at a high temperature precedes the cycling. This initialization step aims to activate the DNA polymerase (heat activation). Moreover, a single elongation step is performed after the last cycle to ensure that any remaining single-stranded DNA is fully elongated. A final hold is also performed at 4°C for short-term storage.

Optimization of the PCR conditions was performed using the GoTaq DNA Polymerase (Promega, Ref. M300) while the *Pfu* DNA Polymerase (Promega, Ref. M7741) was used for amplification of the gene of interest and further cloning steps. Indeed, the *Pfu* DNA Polymerase has a proofreading activity which allows the excision of base misinsertions that may occur during polymerization.

Briefly, 1 ng/µL DNA template was mixed with 0.5 µM forward primer, 0.5 µM reverse primer, 0.2 mM of each dNTP and 1.25 unit DNA Polymerase in a final volume of 50 µL. The PCR was performed as follows (30 cycles) :

- initialization step : 95°C for 2 min
- denaturation : 95°C for 30 s
- annealing : 60°C for 30 s
- elongation : 72°C for 1 min 30 s
- final elongation : 72°C for 10 min
- final hold : 4°C for 10 min

Table III.1 ***Summary of cloned constructs.*** Top table : Overview of the various constructs. TCS, thrombine cleavage site. Bottom table : Forward and reverse primers. The sequences of the genes of interest are indicated in *blue*, the restriction sites in *green*, the *Strep*-tag II sequence in *red*, the linker sequence in *orange* and the start and stop codons in *bold black*.

#	Organism	Strain	Protein	Construct	Tag	Tag position	Restriction enzymes	DNA template	Destination vector
1	HCV	Con1	p7	full-length	StrepII	C-ter	SpeI/BamHI	pEU-6H-C2-p7	pEU-E01-MCS
2	HCV	J	Core	1-120	StrepII	C-ter	SpeI/BamHI	pEU-6H-N2-Core1b-E1	pEU-E01-MCS
3	HCV	J	Core	1-171	StrepII	C-ter	SpeI/BamHI	pEU-6H-N2-Core1b-E1	pEU-E01-MCS
4	HCV	J	Core	1-175	StrepII	C-ter	SpeI/BamHI	pEU-6H-N2-Core1b-E1	pEU-E01-MCS
5	HCV	J	Core	1-180	StrepII	C-ter	SpeI/BamHI	pEU-6H-N2-Core1b-E1	pEU-E01-MCS
6	HCV	J	Core	1-183	StrepII	C-ter	SpeI/BamHI	pEU-6H-N2-Core1b-E1	pEU-E01-MCS
7	HCV	J	Core	116-171	StrepII	C-ter	SpeI/BamHI	pEU-6H-N2-Core1b-E1	pEU-E01-MCS
8	HCV	Con1	NS2	94-217 ($NS2^{pro}$)	StrepII	N-ter	AgeI/BamHI	pEU-E01-StrepII-TCS-NS2fl-Con1	pEU-E01-StrepII-*AgeI*-TCS-*BspEI*
9	HCV	Con1	NS2	94-217 ($NS2^{pro}$)	StrepII-TCS	N-ter	BspEI/BamHI	pEU-E01-StrepII-TCS-NS2fl-Con1	pEU-E01-StrepII-*AgeI*-TCS-*BspEI*
10	HCV	JFH-1	NS2	94-217 ($NS2^{pro}$)	StrepII	N-ter	AgeI/BamHI	pEU-E01-StrepII-TCS-NS2fl-JFH1	pEU-E01-StrepII-*AgeI*-TCS-*BspEI*
11	HCV	JFH-1	NS2	94-217 ($NS2^{pro}$)	StrepII-TCS	N-ter	BspEI/BamHI	pEU-E01-StrepII-TCS-NS2fl-JFH1	pEU-E01-StrepII-*AgeI*-TCS-*BspEI*
12	HCV	Con1	NS5A	1-213 (AH-D1)	StrepII	C-ter	SpeI/BamHI	p7.7-NS5A-AH-D1D2D3-Con1	pEU-E01-MCS
13	HCV	Con1	NS5A	30-213 (D1)	StrepII	C-ter	SpeI/BamHI	p7.7-NS5A-AH-D1D2D3-Con1	pEU-E01-MCS
14	HCV	Con1	NS5A	1-447 (AH-D1D2D3)	StrepII	C-ter	SpeI/NotI	p7.7-NS5A-AH-D1D2D3-Con1	pEU-E01-MCS
15	HCV	Con1	NS5A	30-447 (D1D2D3)	StrepII	C-ter	SpeI/NotI	p7.7-NS5A-AH-D1D2D3-Con1	pEU-E01-MCS
16	Dengue virus	type 2	Capsid	capsid	-	-	SpeI/BamHI	pDVWSK601	pEU-E01-MCS
17	Dengue virus	type 2	Capsid	capsid	StrepII	C-ter	SpeI/BamHI	pDVWSK601	pEU-E01-MCS
18	Dengue virus	type 2	Capsid	capsid-sp	StrepII	C-ter	SpeI/BamHI	pDVWSK601	pEU-E01-MCS
19	Dengue virus	type 2	Capsid	capsid-sp-pr8	StrepII	C-ter	SpeI/BamHI	pDVWSK601	pEU-E01-MCS

#	Forward primer	Reverse primer
1	5'-ATCCAGACTAGT**ATG**GCCCTAGAGAACCTGGTG	5'-ATATATGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGAGGCGTATGCTCGTGGTGG
2	5'-ATCCAGACTAGT**ATG**AGCACGAATCCTAAA	5'-ATATATGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGAACCCAAATTGCGCGACCT
3	5'-ATCCAGACTAGT**ATG**AGCACGAATCCTAAA	5'-ATATATGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGAACCGGGCAGATTCCCTGT
4	5'-ATCCAGACTAGT**ATG**AGCACGAATCCTAAA	5'-ATATATGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGAAGAAAAGGAGCAACCGGG
5	5'-ATCCAGACTAGT**ATG**AGCACGAATCCTAAA	5'-ATATATGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGAAGCCAAAAGGAAGATAGA
6	5'-ATCCAGACTAGT**ATG**AGCACGAATCCTAAA	5'-ATATATGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGAGGACAGCAAAGCCAAAAG
7	5'-ATCCAGACTAGT**ATG**TCGCGCAATTTGGGTAAG	5'-ATATATGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGAACCGGGCAGATTCCCTGT
8	5'-ATCCAGACCGGTCAGGCTGGTATAACCAAAGTGCC	5'-ATATCTGGATCC**TCA**GAGGAGTCGCCACCCCTGCC
9	5'-ATCCAGACCGGTCAGGCTGGTATAACCAAAGTGCC	5'-ATATCTGGATCC**TCA**GAGGAGTCGCCACCCCTGCC
10	5'-ATCCAGACCGGTAGGGCCGCTTTGACACATGTG	5'-ATATCTGGATCC**TCA**AAGGAGCTTCCACCCCTTG
11	5'-ATCCAGACCGGTAGGGCCGCTTTGACACATGTG	5'-ATATCTGGATCC**TCA**AAGGAGCTTCCACCCCTTG
12	5'-GATATCACTAGT**ATG**GCATCCGGCTCGGCTAAGA	5'-ATATATGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGACGTCTCCGCCGTAATGTG
13	5'-GATATCACTAGT**ATG**GCACGATTGCCGGGAGTCCCC	5'-ATATATGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGACGTCTCCGCCGTAATGTG
14	5'-GATATCACTAGT**ATG**GCATCCGGCTCGGCTAAGA	5'-ATTTATGGCGGCCGC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGAGCAGCAGACGACGTCCTC
15	5'-GATATCACTAGT**ATG**GCACGATTGCCGGGAGTCCCC	5'-ATTTATGGCGGCCGC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGAGCAGCAGACGACGTCCTC
16	5'-GTCCAGACTAGT**ATG**AATAACCAACGAAAAAAGG	5'-ATATCTGGATCC**TCA**TCTGCGTCTCCTGTTCAAG
17	5'-GTCCAGACTAGT**ATG**AATAACCAACGAAAAAAGG	5'-ATATCTGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGATCTGCGTCTCCTGTTCAAG
18	5'-GTCCAGACTAGT**ATG**AATAACCAACGAAAAAAGG	5'-ATATCTGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGACGCCATCACTGTTGGAATC
19	5'-GTCCAGACTAGT**ATG**AATAACCAACGAAAAAAGG	5'-ATATCTGGATCC**TCA**TTTTTCAAATTGAGGATGAGACCAAGCAGATCCGTTACGTGTGGTTAAA

1.3. Purification of PCR products

PCR products were purified by phenol/chloroform extraction (refer to section III.2.2. for detailed explanations). Briefly, 40 µL of phenol/chloroform/isoamyl alcohol (25/24/1) (Sigma-Aldrich, Ref. P3803) were added to 40 µL of PCR product. The mixture was shortly vortexed and centrifuged for 5 min at 20,800 *g*, at room temperature (RT) using a rotor F45-30-11 (Eppendorf). The supernatant was transferred in a new tube and 40 µL chloroform (Sigma-Aldrich, Ref. 366919) were added. Another centrifugation step was done for 5 min at 20,800 *g*, RT, and the supernatant was isolated as previously. 100 µL 100% ethanol and 14 µL 3 M sodium acetate pH 5.2 were added to precipitate the PCR product. The mixture was incubated at -20°C for 10 min and centrifuged for 20 min at 20,800 *g*, 4°C. The supernatant was removed and the DNA precipitate was washed with 800 µL 70% ethanol and centrifuged for 10 min at 20,800 *g*, 4°C. The supernatant was removed and the pellet was dried under the hood before resuspension in 15 µL water.

1.4. Digestion of PCR products and pEU-E01-MCS vector

Digestion mixtures were prepared as follows :

- 15 µL PCR product or 5 µL pEU-E01-MCS (3 µg)
- 1 µL (10 units) restriction enzyme #1
- 1 µL (10 units) restriction enzyme #2
- 2 µL MultiCore 10X or Buffer D 10X (as recommended by Promega)
- water to a final volume of 20 µL

The mixture was incubated at 37°C for 2 h. Digestion products were stored at -20°C until use. Restriction enzymes were purchased from Promega : SpeI (Ref. R6591), BamH1 (Ref. R6021), NotI (Ref. R6431), BspEI (AccIII, Ref. R6581) and AgeI (Ref. R7251). Refer to **Table III.1** (top table) for details.

1.5. Extraction of digested products from agarose gel

The digested samples were loaded on a 1% agarose gel containing GelRed Nucleic Acid Gel Stain (Interchim, Ref. BY1740) for nucleic acid detection under UV light. In addition, 1 kb DNA Ladder (Promega, Ref. G694A) and 100 bp DNA Ladder (Promega, Ref. G2101) molecular weight markers were also loaded on the agarose gel for size calibration. The gel was run using TAE (Tris-Acetate-EDTA) buffer in a Sub-Cell® GT system (Bio-Rad, USA) at 150 V. The bands of interest were extracted from the agarose gel using the QIAquick Gel Extraction kit (Qiagen, Ref. 28704) according to the manufacturer's recommendations. DNA concentration was measured using the NanoDrop (2000c spectrophotometer, ThermoScientific).

1.6. Ligation

The amount of insert to use was calculated as follows :

$$ng\ \text{insert} = \frac{ng\ \text{vector} * \text{insert size}\ (kb)}{\text{vector size}\ (kb)} * \text{molar ratio}\ \frac{\text{insert}}{\text{vector}}$$

A molar ratio insert/vector of 3/1 was applied, and 100 ng vector were used. One unit T4 DNA Ligase (Promega, Ref. M1801) and 1 µL Buffer Ligase 10X (10 µL final volume) were added to the mixture which was incubated for 3 h at RT.

1.7. Transformation of *E. coli* competent cells

1.7.1. Preparation of E. coli competent cells

25 µL One Shot® TOP10 Chemically Competent *E. coli* (Life Technologies, Ref. C4040-06) were incubated at 37°C overnight, in 10 mL LB medium (Sigma-Aldrich, Ref. L3022) without antibiotic. LB medium contains 10 g/L tryptone (pancreatic digest of casein), 5 g/L yeast extract and 5 g/L NaCl. Then, 1 mL of this preculture was diluted in 100 mL LB medium without antibiotic and incubated at 37°C for about 4 h until a A_{600} of 0.4 was reached. The culture was incubated on ice for 10 min and centrifuged for 10 min at 3,000 *g*, 4°C. The pellet was resuspended in 10 mL of TSS buffer pH 6.5 containing

85% LB medium, 10% PEG 8000, 5% DMSO and 50 mM $MgCl_2$. Note that the buffer was beforehand filtered (0.2 µm filter). The competent cells were stored at -80°C.

1.7.2. Transformation

5 µL of ligation product were added to 50 µL *E. coli* TOP10 competent cells. The bacteria were incubated on ice for 30 min, then at 42°C for 45 s in a water bath and again on ice for 2 min. Finally, 250 µL LB medium (Sigma-Aldrich, Ref. L3022) were added and the bacteria were incubated at 37°C, 125 rpm for 1 h. Two different volumes (usually 50 µL and 200 µL) were spread out on agar plates containing 50 µg/mL ampicilline. The agar plates were incubated overnight at 37°C.

1.8. Screening by colony PCR

Ten colonies were picked up using a toothpick. Each colony was mixed with 0.5 µM forward primer, 0.5 µM reverse primer, 0.2 mM of each dNTP and 1.25 unit GoTaq DNA Polymerase (Promega, Ref. M300) in a final volume of 50 µL. The PCR was performed as described in section III.1.2. (bacteria were lyzed during the initialization step), and 20 µL of PCR products were analyzed on a 1% agarose gel containing GelRed Nucleic Acid Gel Stain (Interchim, Ref. BY1740) for nucleic acid detection under UV light. In parallel, each colony was pricked out on an agar plate containing 50 µg/mL of ampicilline and plates were incubated overnight at 37°C. For each construct, two positive colonies were selected and used to prepare 5 mL culture in LB medium (Sigma-Aldrich, Ref. L3022) containing 50 µg/mL of ampicilline. These cultures were incubated overnight at 37°C.

1.9. Mini DNA preparation

DNA was prepared from 2 mL culture (refer to section III.1.8.) using the QIAprep® Spin Miniprep Kit (Qiagen, Ref. 27106) according to the manufacturer's instructions. DNA concentration and purity (A_{260}/A_{280} ratio) were mesured using the NanoDrop (2000c spectrophotometer, ThermoScientific).

1.10. Digestion of obtained constructs

Plasmid DNA (refer to section III.1.9.) was digested with the restriction enzymes used for cloning. Digestion mixtures were prepared as follows :

- 1 µL plasmid DNA
- 0.5 µL (5 units) restriction enzyme #1
- 0.5 µL (5 units) restriction enzyme #2
- 1 µL MultiCore 10X or Buffer D 10X (as recommended by Promega)
- water to a final volume of 10 µL

The mixture was incubated at 37°C for 2 h and analyzed on agarose gel. Two bands were expected : one corresponding to the insert (gene of interest), another one corresponding to the vector (pEU-E01). When these two bands were obtained, sequencing was carried out (Biofidal, France) to complete the quality control of the constructs.

1.11. Preparation of glycerol stocks

600 µL of bacterial culture (refer to section III.1.8.) were mixed with 400 µL of 50% glycerol to obtain a 20% glycerol stock. Glycerol stocks were stored at -80°C and used for large DNA production (refer to section III.2.1.).

2. DNA production

2.1. Maxi DNA preparation

2.1.1. Bacterial cultures

A preculture was first prepared inoculating 5 mL LB medium (Sigma-Aldrich, Ref. L3022) containing 50 mg/mL ampicilline with either a glycerol stock or a fresh clone. The preculture was incubated for 8 h at 37°C, 125 rpm. A culture was then prepared inoculating 400 mL LB medium containing 50 mg/mL ampicilline with 4 mL preculture. The culture was incubated overnight at 37°C, 125 rpm.

2.1.2. DNA preparation

Plasmids were purified using a NucleoBond Xtra Maxi kit (Macherey-Nagel, Ref. 740414) according to the manufacturer's instructions. The principle is based on an anion-exchange chromatography. Indeed, interaction between negatively charged DNA backbone and positively charged anion-exchanger material leads to the binding of DNA to a macroporous silica gel at low pH. Contaminants are removed by stringent washing with increasing salt concentration. DNA is then eluted by high pH and finally alcohol precipitated. Briefly, the bacterial culture was centrifuged for 15 min at 6,000 *g*, 4°C. The pellet was resuspended in a buffer containing RNase and bacteria were then lysed in a buffer containing 2% NaOH and SDS. The mixture was neutralized with the corresponding buffer leading to the formation of a white precipitate (proteins and genomic DNA). The mixture was loaded on the column with filter paper on top (**Figure III.2**). The white precipitate was retained on the filter whereas nucleic acids interacted with the anion-exchange silice resin. The filter was removed, nucleic acids were eluted (15 mL), precipitated with 10.5 mL isopropanol and centrifuged for 30 min at 15,000 *g*, 4°C (rotor F34-6-38, Eppendorf).

Figure III.2 ***Anion-exchange chromatography column from the NucleoBond Xtra Maxi kit.*** Adapted from http://www.mn-net.com (Macherey-Nagel, France).

The pellet was washed with 5 mL 70% ethanol and centrifuged for 10 min at 15,000 *g*, 4°C. The pellet was dried under the hood and resuspended in 500 µL RNase free water (Water for molecular biology, Merck Millipore, Ref. H20MB). DNA concentration and purity (A_{260}/A_{280} ratio) were mesured using the NanoDrop (2000c spectrophotometer, ThermoScientific).

2.2. Phenol/chloroform extraction

Residual traces of RNAse from the resuspension buffer (refer to section III.2.1.) can be present in DNA samples and could interfer with *in vitro* protein synthesis. The aim of the phenol/chloroform extraction is to remove residual RNase. The phenol/chloroform extraction is indeed a liquid-liquid extraction method allowing the purification of nucleic acids and removal of proteins. The phenol/chloroform mixture is not miscible with water which leads to the formation of two distinct phases : the upper aqueous phase containing DNA and the lower organic phase containing proteins.

All tubes and tips used for the following steps were certified RNase free. One volume of phenol/chloroform/isoamyl alcohol (25/24/1) (Sigma-Aldrich, Ref. P3803) was added to DNA solution. The mixture was shortly vortexed and centrifuged for 5 min at 20,800 *g*, RT, using a rotor F45-30-11 (Eppendorf). The supernatant was transferred in a new tube and 1 volume chloroform (Sigma-Aldrich, Ref. 366919) was added. Another centrifugation step was done for 5 min at 20,800 *g*, RT and the supernatant was isolated as previously. Two and half volumes 100% ethanol and 1/10 3 M sodium acetate pH 5.2 were added to precipitate DNA. The mixture was incubated at -20°C for 10 min and centrifuged for 20 min at 20,800 *g*, 4°C. The supernatant was removed and the DNA precipitate was washed with 800 µL 70% ethanol and centrifuged for 10 min at 20,800 *g*, 4°C. The supernatant was removed and the pellet was dried under the hood before resuspension in 500 µL RNase free water (Water for molecular biology, Merck Millipore, Ref. H20MB). DNA concentration and purity (A_{260}/A_{280}) were mesured using the NanoDrop (2000c spectrophotometer, ThermoScientific). A A_{260}/A_{280} ratio between 1.70 and 1.85 was expected. If this was not the case, phenol/chloroform extraction was repeated. In order to standardize protocols, DNA preparations were diluted in RNase free water at a concentration of 1 µg/µL.

3. Wheat germ extract preparation

There are commercially available wheat germ extracts (WGEs, *e.g.* from CellFree Sciences or Promega) but in order to significantly reduce costs we prepare home-made extracts according to the protocol developped by Yaeta Endo and coworkers at Ehime University in Japan (Takai *et al*, 2010) with minor modifications. Refer to section II.2.1. for a schematic representation of a wheat seed (**Figure II.4**) and to section II.2.2.1. for general information about WGEs.

3.1. Preparation of wheat germs

Germs were isolated from non-treated durum wheat seeds (Sud Céréales). Seeds were first grinded using a mill (F100, SAMAP), then sieved using an industrial sieve. The germs were on the upper sieve (850 μm) together with light particules which were removed with cold air using a hair-drier. The germs were then further isolated in a solvent bath (carbone tetrachloride/cyclohexane, 600/240, v/v) : they float on the surface whereas the waste go down. This step was performed under the hood. The germs were quickly harvested using a spatula and stored under the hood till solvents were completely evaporated. Although these preliminary steps allowed an efficient isolation of wheat germs with 50% to 60% purity, it was necessary to sort them by eye selection in order to remove the ones with brownish particules, the ones with too much white matter coming from endosperm and the broken ones (**Figure III.3**). About 1 g sorted germs was obtained from 1 kg seeds. Although the wheat germs can in theory be stored for several months at 4°C, they were used directly for the preparation of the wheat germ extract.

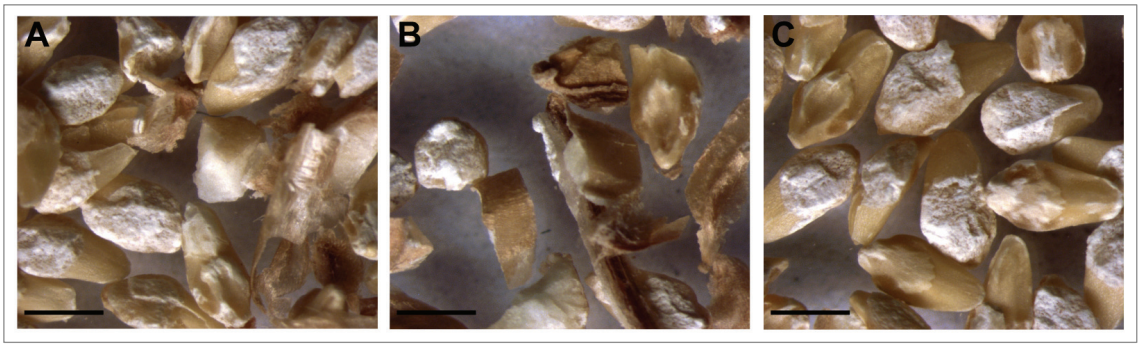


Figure III.3 ***Preparation of wheat germs.*** (A) Wheat germs before eye selection. (B) Waste, such as broken germs or brownish particles. (C) Wheat germs selected for wheat germ extract preparation. The white material around germs is endosperm which contains translation inhibitors. Endosperm is removed by extensive washing of the germs before extract preparation. Scale bar, ~ 1 mm.

3.2. Preparation of wheat germ extracts

The most optimized protocol is described here. Extracts were prepared starting from 20 g sorted germs (instead of 60 g in (Takai *et al*, 2010)). All buffers and water used during the extract preparation were cooled at 4°C. Moreover, all equipments and workbenches were cleaned with RNaseAway (Molecular BioProducts, Ref. 7002). Starting from the crushing of the wheat germs, only RNase free water was used (Water for molecular biology, Merck Millipore, Ref. H20MB).

3.2.1. Washing of the wheat germs

The washing steps described here are crucial because they allow the removal of residual endosperm that contains translation inhibitors such RIB (Ribosome Inactivating Proteins), thionin or DNases and RNases (Madin et al, 2000).

The germs were placed in a stocking for the washing steps during which they were gently kneaded. Several washing steps were first performed in 500 mL water till the water was clear. The germs were then washed twice in 500 mL 0.5% Nonidet P-40 for 5 min in a sonication bath (Elmasonic S30H, Elma, Germany), instead of only one detergent washing step in (Takai *et al*, 2010), and several times in 500 mL water until there was no bubbles caused by the detergent anymore. The germs were afterwards washed twice in 500 mL water for 2 min in a sonication bath. The germs were finally washed five additional times in 500 mL water and were dried carefully with laboratory paper towel.

3.2.2. Crushing of the wheat germs

The germs were crushed using the small bowl with four blades of a kitchen mixer (mini blender HR2860, Philips) in 40 mL buffer containing 80 mM Hepes-KOH pH 7.6, 200 mM potassium acetate, 10 mM magnesium acetate, 4 mM calcium chloride and 8 mM DTT. The blender was run 5 times 10 s (instead of 3 times 30 s in (Takai *et al*, 2010)). The mixture was centrifuged for 30 min at 30,000 *g*, 4°C (JA-20 fixed angle rotor, Beckman Coulter). Three phases were obtained : a precipitate at the bottom of the tube, a fatty fraction at the surface, and a supernatant (middle layer) (**Figure III.4**). The supernatant was further centrifuged for 15 min at 30,000 *g*, 4°C. There were still a precipitate and a fatty fraction after this centrifugation step. The middle layer was thus harvested as previously and, in contrast to (Takai *et al*, 2010), again centrifuged for 15

min at 30,000 *g*, 4°C in order to obtain a purer supernatant.

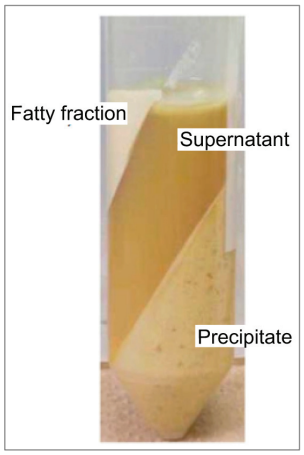


Figure III.4 ***Example of the sample after the first 30,000 g centrifugation.*** Adapted from (Takai *et al*, 2010).

Note that the use of a blender to crush the germs was a huge improvment in terms of time saving and crushing efficiency compared to the initial method which consisted in crushing the germs using a mortar and a pestle under liquid nitrogen.

3.2.3. Preparation of the Sephadex G25 columns

Sephadex G25 columns were home-made : 40 mL Sephadex G25 fine resin (GE Healthcare, Ref. 17-0032-01) resuspended in RNase free water were packed in a 50-mL syringe with a cotton wool filter at the bottom. Two columns were necessary for each desalting step (*i.e*. six columns for each 20 g wheat germ extract preparation). Columns were equilibrated by gravity with 3 volumes buffer and centrifuged for 5 min at 750 *g*, 4°C (JA-10 fixed angle rotor, Beckman Coulter). For centrifugation, the columns were hung in a 500-mL centrifuge bottle and capped with aluminum foil. After use, the columns were extensively washed with water, the resin was stored at 4°C in water containing 0.05% sodium azide and was reused several times.

3.2.4. Wheat germ extract desalting

The extract was loaded on a Sephadex G25 fine column (see previous section) equilibrated with a buffer containing 40 mM Hepes-KOH pH 7.6, 100 mM potassium acetate, 5 mM magnesium acetate, 2 mM calcium chloride and 4 mM DTT (20 mL extract on a 40-mL column). The columns were centrifuged for 5 min at 750 *g*, 4°C (JA-10 fixed angle rotor, Beckman Coulter) and the flowthrough was loaded on a new Sephadex G25 fine column equilibrated this time with a buffer containing 30 mM Hepes-KOH pH 7.6, 100 mM potassium acetate, 10.8 mM magnesium acetate, 0.4 mM spermidine, 2.5 mM

DTT, 1.4 mM ATP, 0.25 mM GTP and 16 mM creatine phosphate (20 mL extract on a 40-mL column). These columns were also centrifuged for 5 min at 750 *g*, 4°C. The desalting step on columns equilibrated with the second buffer was performed twice. Note that both buffers didn't contain any amino acids. Indeed, amino acids were added while preparing the translation reaction, which allows greater flexibility regarding to isotopic labeling strategies for NMR studies. Note also that at the beginning of this work, the second buffer was home-made. However, in order to save cost and to achieve a better reproducibility, we then used the commercial buffer called SUB-AMIX (CellFree Sciences, Ref. CFS-SUB-NAA). The same buffer was also used for translation.

3.2.5. Wheat germ extract concentration

The wheat germ extract was finally concentrated down to a final volume of 30 mL using Spin-X UF20 concentrators whose low binding polyethersulfone (PES) membrane has a MWCO of 10,000 Da (Corning, Ref. 431488). The concentrators were washed twice with RNase free water before used and were centrifuged at 8,000 *g*, 4°C (fixed angle rotor F34-6-38, Eppendorf) until the expected volume was reached. Note that A_{260} and A_{280} were measured at each step of the preparation to follow the extract concentration. According to (Takai *et al*, 2010), final A_{260} should be at least equal to 240. We observed that WGE efficiency could also be correlated to the A_{260}/A_{280} ratio (refer to section IV.1.1.). Furthermore, whereas according to (Takai *et al*, 2010) the WGEs should be concentrated between the two last desalting steps (see previous section), we concentrated WGEs at the end of the preparation to avoid any final adjustment. In addition, home-made WGEs were stored in liquid nitrogen, and not at -80°C as described in (Takai *et al*, 2010).

3.2.6. Quality control of wheat germ extracts

The efficiency of home-made wheat germ extracts was controlled through the expression of GFP which can be easily followed using a fluorimeter. Fluorescence intensity measurements were performed in 96-well black plates (Greiner Bio-One, Ref. 655900) with 150 µL sample, using a Infinite® M1000 microtiterplate reader (Tecan, Switzerland). An excitation wavelength of 488 nm and an emission wavelength of 509 nm, as well as a manual gain of 150 and a z-position of 21,500 were applied. Home-made

wheat germ extracts were compared to the commercial wheat germ extracts WEPRO1240 and WEPRO2240 (CellFree Sciences, Japan) as well as to each other.

4. Wheat germ cell-free protein expression

4.1. Transcription

Transcription was performed according to (Takai *et al*, 2010) in 1.5 mL Eppendorf tubes using 100 µg/mL plasmid, 2.5 mM NTP mix (Promega, Ref. E6000), 1U/µL SP6 RNA Polymerase (CellFree Sciences, Ref. CFS-TSC-ENZ) and 1U/µL RNase Inhibitor (CellFree Sciences, Ref. CFS-TSC-ENZ) in transcription buffer (CellFree Sciences, Ref. CFS-TSC-5TB), as detailed in **Table III.2**. Transcription buffer contains 80 mM Hepes-KOH pH 7.6, 16 mM magnesium acetate, 10 mM DTT and 2 mM spermidine. After incubation for 6 h at 37°C, mRNA was used directly for translation, but could also be stored at -80°C. Note that a white precipitate (magnesium pyrophosphate) appears during transcription.

Table III.2 ***Preparation of the transcription mix.*** Small scale, final volume of 50 µL for small scale protein expression in a 96-well plate. Large scale, final volume of 1500 µL for large scale protein expression in a 6-well plate (6 wells).

Reagent	Final concentration	Small scale (µL)	Large scale (µL)
Transcription buffer 5X	1X	10	300
NTP mix 25 mM	2.5 mM	5	150
RNAsin 80 U/µL	1 U/µL	0.6	18.8
SP6 RNA Polymerase 80 U/µL	1 U/µL	0.6	18.8
DNA 1 µg/µL	100 ng/µL	5	150
H_2O	-	28.8	862.5

4.2. Translation

4.2.1. Bilayer method

In a standard way, translation was performed using the so-called bilayer method according to (Takai *et al*, 2010) in flat bottom plates, either 96-well plates (Greiner Bio-One, Ref. 655101) for small scale expression tests or 6-well plates (Corning, Ref. 3471) for larger scale productions. The bottom layer, called translation mix, contains the

mRNA, the wheat germ extract and creatine kinase for energy regeneration. The upper layer corresponds to the feeding buffer which contains all substrates necessary for protein synthesis (see details below). There is no dialysis membrane between the two layers. Only the higher density of the WGE allows the formation of the bilayer. The feeding buffer is first added into the well and the reaction mix is then carefully filled at the bottom of the well. Translation is initiated in the reaction mix. Byproducts are then gradually diluted in the feeding buffer whereas fresh substrates gradually diffuse to the reaction mix (**Figure III.5**).

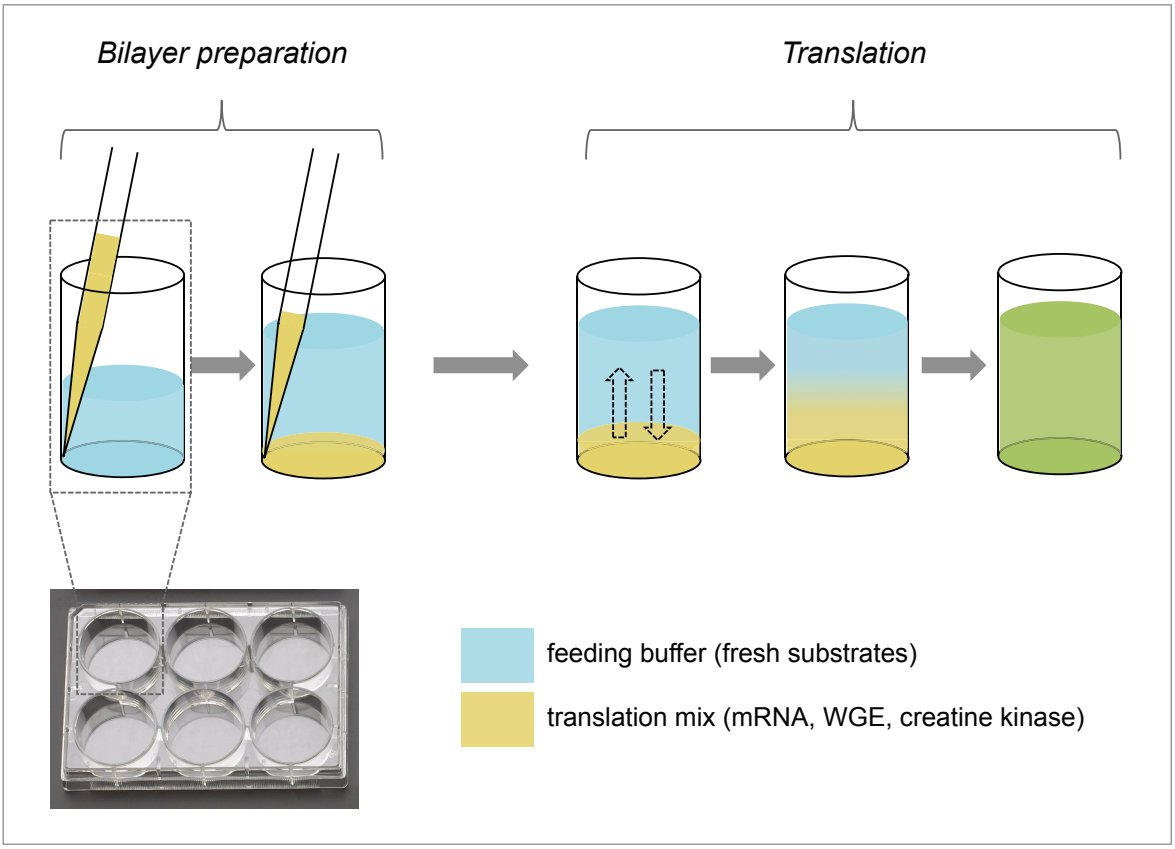


Figure III.5 ***In vitro protein synthesis using the bilayer method***

There is 20 µL translation mix as well as 200 µL feeding buffer per well in 96-well plates, and 500 µL translation mix as well as 5.5 mL feeding buffer per well in 6-well plates. The translation mix contains mRNA, WGE and creatine kinase (Roche, Ref. 10127566001), as detailed in **Table III.3**. The feeding buffer contains 30 mM Hepes-KOH pH 7.6, 100 mM potassium acetate, 2.7 mM magnesium acetate, 16 mM creatine phosphate, 0.4 mM spermidine, 1.2 mM ATP, 0.25 mM GTP, 4 mM DTT and 6 mM

(average concentration) amino acid mix. In the initial steps of this work, each stock solution was home-made. However, in order to save cost and to achieve a better reproducibility, we then used the commercial feeding buffer called SUB-AMIX (CellFree Sciences, Ref. CFS-SUB-NAA). The composition of this buffer is exactly the same than the one of the home-made feeding buffer but it doesn't contain any amino acid, allowing for more flexibility. The feeding buffer was prepared as detailed in **Table III.4**.

Table III.3 ***Preparation of the translation mix.*** Small scale, final volume of 20 µL per well for small scale protein expression in a 96-well plate. Large scale, final volume of 3 mL for large scale protein expression in a 6-well plate (6 wells). MNG-3 is added only for membrane proteins.

Reagent	Final concentration	Small scale (µL)	Large scale (µL)
mRNA	-	10	1500
WGE (A_{260} ≈ 240)	A_{260} ≈ 120	10	1500
Creatine kinase 4 mg/mL	40 ng/µL	0.2	30
Amino acid mix 12 mM	0.3 mM	0.5	75
MNG-3 5%	*0.3 mM*	*0.4*	*60*

Table III.4 ***Preparation of the feeding buffer.*** Small scale, final volume of 210 µL per well for small scale protein expression in a 96-well plate. Large scale, final volume of 35 mL for large scale protein expression in a 6-well plate (6 wells). MNG-3 is added only for membrane proteins.

Reagent	Final concentration	Small scale (µL)	Large scale (µL)
SUB-AMIX S1 40X	1X	5.25	875
SUB-AMIX S2 40X	1X	5.25	875
SUB-AMIX S3 40X	1X	5.25	875
SUB-AMIX S4 40X	1X	5.25	875
Amino acid mix 12 mM	0.3 mM	5.25	875
MNG-3 5%	*0.1%*	*4.2*	*700*
H_2O	-	179.5	29,925

An amino acid mixture was used standardly to produce unlabeled protein samples (Cambridge Isotope Laboratories, Ref. ULM-7891-1). It is worth to note that the proportion is different for each amino acid in this mixture (6% Ala, 6% Arg, 5% Asn, 8% Asp, 3% Cys, 9% Glu, 5% Gln, 5% Gly, 1% His, 3% Ile, 9% Leu, 12% Lys, 1% Met, 4% Phe, 5% Pro, 4% Ser, 4% Thr, 3% Trp, 3% Tyr and 4% Val). Therefore, depending on the amino acid composition of the protein of interest, preparing an optimized amino acid

mixture using individual amino acids (Fluka, Ref. 09416) with the same composition than the protein of interest could help to achieve better expression yields.

Note also that when expression was performed in the presence of detergent, the detergent was added in both the translation mix and the feeding buffer. In a standard way, the translation reaction was incubated at 22°C for 16 h without agitation (the bilayer would be disturbed by agitation). The plates were covered with an adhesive film to avoid both evaporation and potential contamination.

4.2.2. Dialysis method (CECF)

Translation using the dialysis method was performed in small dialysis cups fitting in a 24-well plate, as described in (Schneider *et al*, 2009) (**Figure III.6**). The dialysis cups were produced at the ETH Zurich according to the geometrical characteristics given in (Schneider *et al*, 2009). There were 50 µL translation mix per dialysis cup and 800 µL feeding buffer per well. The feeding buffer-to-translation mix ratio was therefore 16 (v/v) instead of 10 and 11 for translation using the bilayer method in 96-well and 6-well plates, respectively. The composition of the feeding buffer and the translation mix was exactly the same than the one described in the previous section for the bilayer method. A dialysis membrane with a MWCO of 12-14,000 Da was used (Spectra/Por, Ref. 132700). The plate was incubated overnight under gentle agitation at 60 rpm. As for the bilayer method, the plate was covered with an adhesive film to avoid both evaporation and potential contamination.

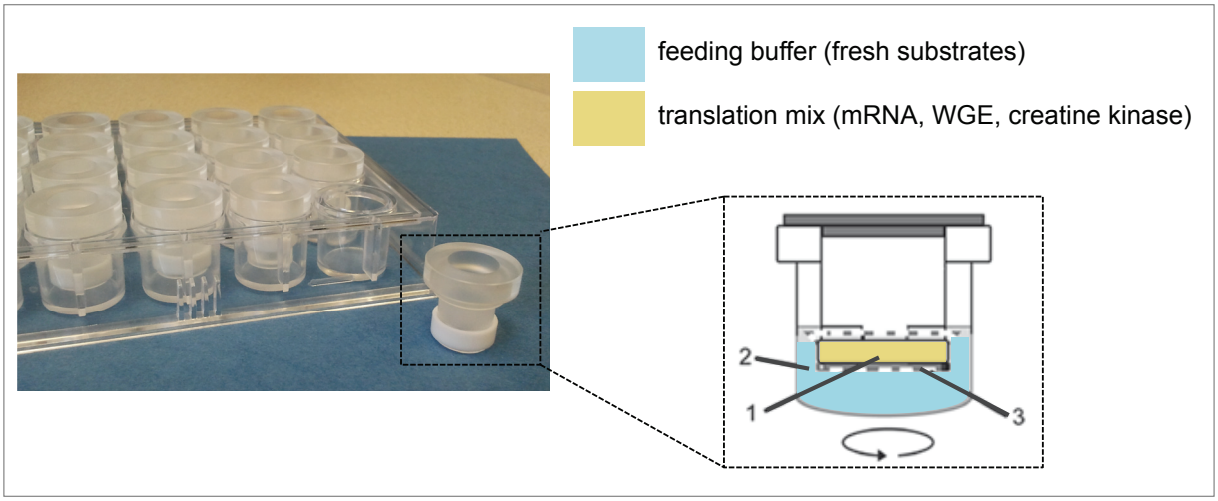


Figure III.6 ***Small scale in vitro protein synthesis using the dialysis method.*** Adapted from (Schneider *et al*, 2009). 1, translation mix; 2, feeding buffer; 3, dialysis membrane.

5. Protein sample preparation

Protein samples were prepared as follows for analysis of small scale expression tests (**Figure III.7**). After protein synthesis, 150 µL of cell-free sample (CFS) were incubated with 250 units of benzonase at RT on a rolling wheel for 30 min. CFS was then centrifuged at 20,000 *g*, 4°C for 30 min. The obtained supernatant was incubated with 5 µL *Strep*-Tactin magnetic beads (MagStrep "type2HC", IBA Lifesciences, Ref. 2-1612-002) at 4°C on the wheel for at least 30 min. According to the manufacturer, these beads have a binding capacity up to 1 nmol protein/mg beads. After three washing steps with 10 volumes of a buffer containing 100 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM EDTA, the magnetic beads were directly resuspended in 25 µL 1X SDS-PAGE loading buffer containing 62.5 mM Tris-HCl pH 6.8, 2% SDS (w/v), 10% glycerol (v/v), 5% β-mercaptoethanol (v/v) and 0.01% bromophenol blue (w/v). Theses samples were further designated as SN-beads. Note that this step aims to enrich the sample with the tagged protein and, when detergent is used, to assess whether it would interfere with the protein affinity capture. The pellet obtained after centrifugation of the CFS was resuspended in 25 µL 1X SDS-PAGE loading buffer. In addition, 20 µL CFS were mixed with 5 µL 5X loading buffer for SDS-PAGE analysis.

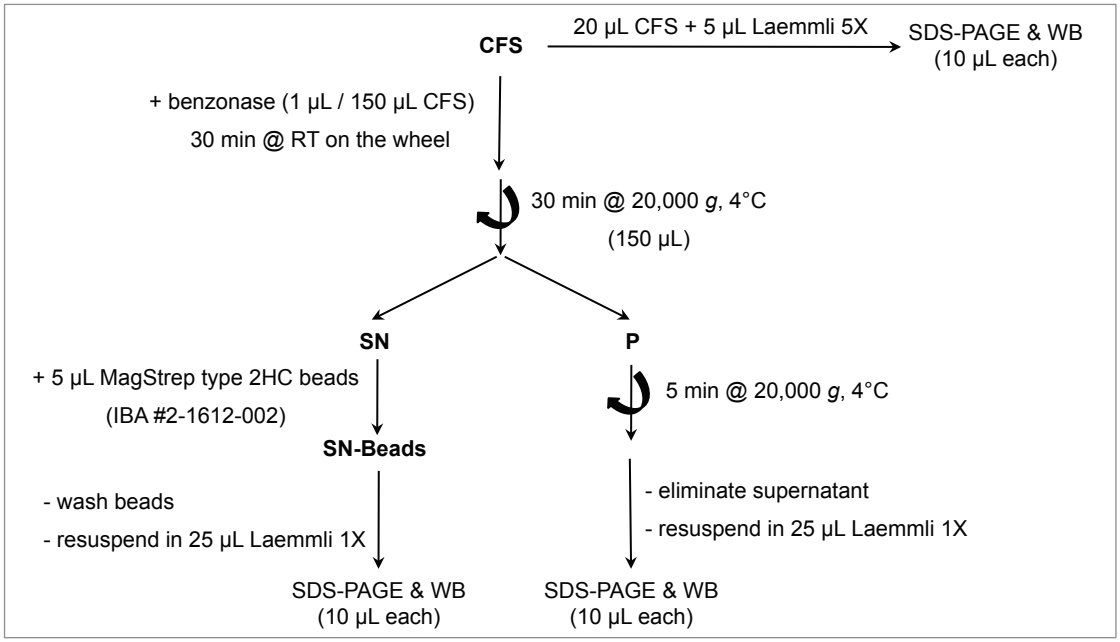


Figure III.7 ***Schematic representation of protein sample preparation for small scale expression tests.*** CFS, total cell-free sample; P, pellet; SN, supernatant; SN-Beads, supernatant incubated with *Strep*-tactin coated magnetic beads.

6. Purification by affinity chromatography

The principle of purification by affinity chromatography using *Strep*-Tactin columns is schematically described in **Figure III.8**.

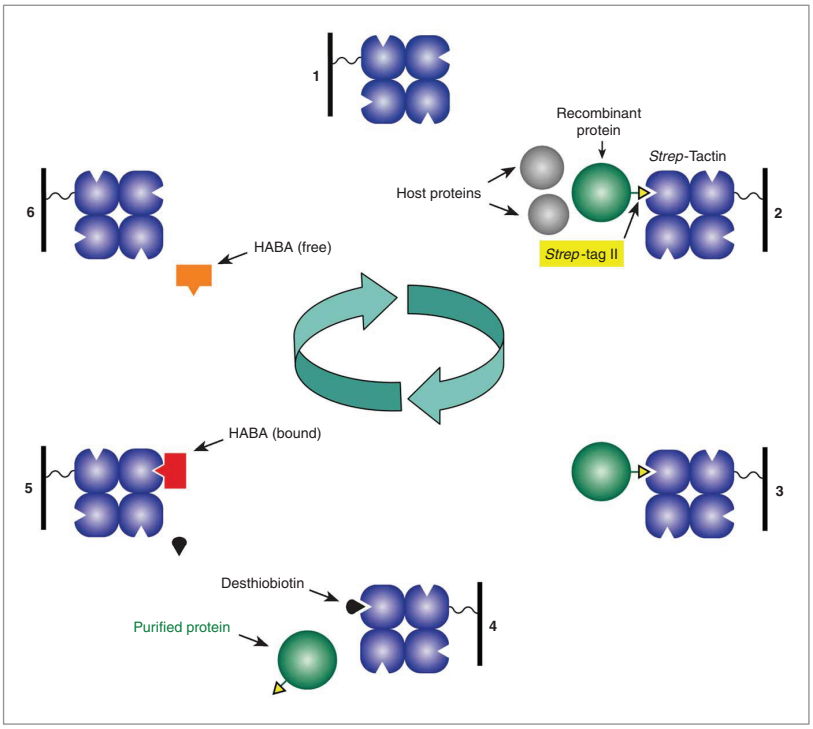


Figure III.8 ***Schematic illustration of the Strep-tag purification cycle.*** From (Schmidt & Skerra, 2007). The supernatant obtained after centrifugation of CFS and containing the *Strep*-tag II fusion protein is applied to a column with immobilized *Strep*-Tactin (steps 1 and 2). Contaminants from WGE are removed by washing with a buffer containing 100 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM EDTA (step 3) and the *Strep*-tag II fusion protein is specifically eluted by competition with 2.5 mM D-desthiobiotin (step 4). D-desthiobiotin is then removed by application of HABA solution and indicated via color change from yellow-orange to red (step 5). Finally, HABA is quickly removed with washing buffer (step 6), thus making the column ready for the next purification run (step 1).

After protein synthesis, the CFS was systematically incubated with benzonase at RT on a rolling wheel for 30 min. Unless specified, 2000 units benzonase per mL CFS were used. Moreover, for membrane proteins expressed in the presence of MNG-3 detergent, 0.25% DDM was added to the CFS together with benzonase, whereas 1% DDM was added for those expressed in the precipitate mode in order to solubilize them. CFS was then centrifuged at 20,000 *g*, 4°C for 30 min. The obtained supernatant was loaded on a *Strep*-Tactin gravity column (IBA Lifesciences, Germany). Purification was performed as described by the manufacturer as follows. The column was equilibrated with 2 CV

(column volumes) of a buffer containing 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 1 mM DTT as well as 0.1% DDM for membrane proteins. After sample loading, the column was washed 5 times with 1 CV of the same buffer. The protein of interest was eluted 6 x 0.5 CV of a buffer containing 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT and 2.5 mM desthiobiotin as well as 0.1% DDM for membrane proteins (**Figure III.9**). *Strep*-Tactin columns format used in a standard way is summarized in **Table III.5**. *Strep*-Tactin columns are regenerated with HABA (2-[4'-hydroxy-benzeneazo]benzoic acid) and used up to 5 times for the same protein. They are stored at 4°C in 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 0.05% NaN_3.

Table III.5 ***Summary of the Strep-Tactin column format used in a standard way.***

Plate type	Number of wells	CFS volume (mL)	*Strep*-Tactin column volume (mL)	Reference IBA
6-well	1	6	0.2	2-1207-550
6-well	6	36	1	2-1207-001
6-well	18	108	5	2-1207-051

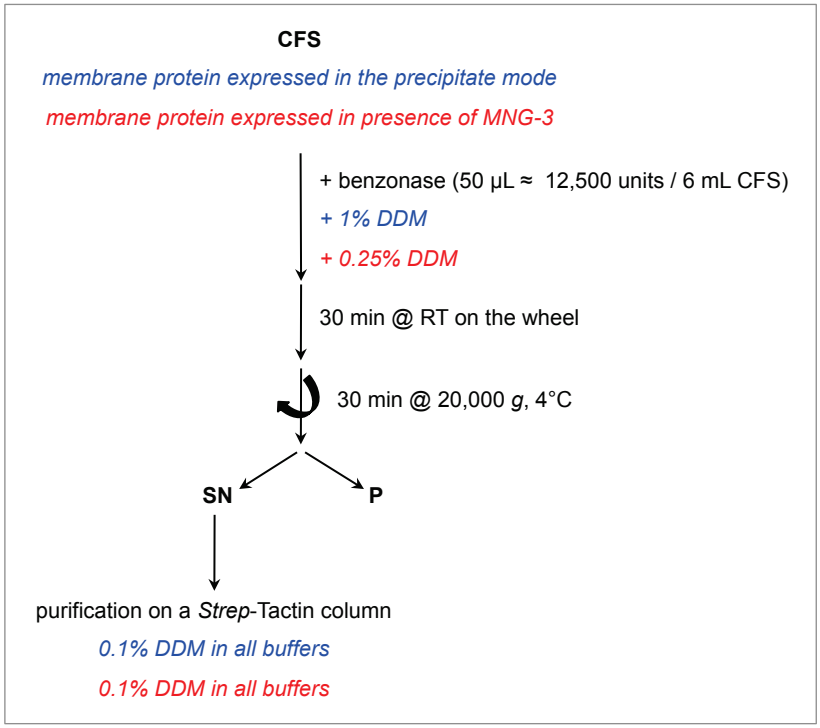


Figure III.9 ***Schematic representation of the standard affinity purification process.*** The protocole specific for membrane proteins expressed in the precipitate mode is indicated in *blue*, and in *red* for membrane proteins expressed in the presence of MNG-3. No detergent is used for soluble proteins.

7. Size exclusion chromatography

Size exclusion chromatography was performed using either a KW402.5-4F column (Shodex, USA) or a Superdex™ 200 Increase 3.2/100 column (GE Lifesciences, France) on the NGC Chromatography System (Bio-Rad, USA). A 50 mM phosphate buffer pH 7.4 containing 0.1% DDM was used as eluent. Affinity elution buffer was systematically loaded on the column as control. A flow rate of 150 µL/min was applied on the KW402.5-4F column whereas a flow rate of 75 µL/min was applied on the Superdex™ 200 Increase 3.2/100 column.

Calibration curve of the size exclusion chromatography column was established using protein standards described in (le Maire *et al*, 1986) : cytochrome C, 12.3 kDa (Sigma-Aldrich, Ref. C2506) ; myoglobin, 17.8 kDa (Sigma-Aldrich, Ref. M0630) ; trypsin inhibitor, 22.1 kDa (Sigma-Aldrich, Ref. T6522) ; β-lactoglobulin, 35 kDa (Sigma-Aldrich, Ref. L3908) ; transferrin, 81 kDa ; aldolase, 158 kDa ; ferritin, 440 kDa ; thyroglobulin, 669 kDa. Transferrin, aldolase, ferritin and thyroglobulin, as well as dextran blue, were part of the Gel Filtration Calibration Kit High Molecular Weight (GE Healthcare, Ref. 28-4038-42). A 50 mM phosphate buffer pH 7.4 containing 0.1% DDM was used as eluent and 100 µg of each protein standard were loaded separately on the column. In addition, a calibration curve without detergent was also done for the analysis of the NS2/A8-35 complexes.

8. Circular dichroism

Circular dichroism (CD) is based on the differential absorption of two circularly polarized lights, one left handed and the other right handed (Kelly *et al*, 2005). A CD signal is observed when a molecule is chiral, *i.e.* optically active. Far UV (ultraviolet) CD spectra allow the investigation of the secondary structure of proteins (**Figure III.10**).

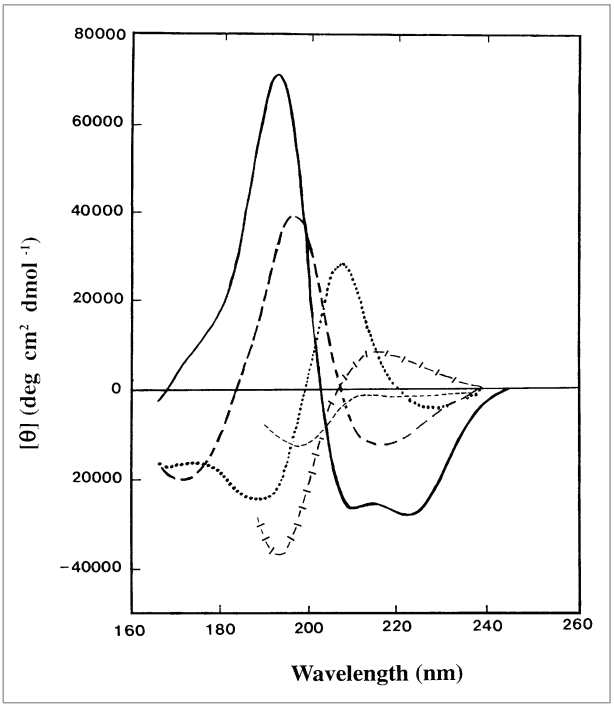


Figure III.10 ***Far UV CD spectra associated with various types of secondary structure.*** From (Kelly *et al*, 2005). Solid line, α-helix; long dashed line, anti-parallel β-sheet; dotted line, type I β-turn; cross dashed line, extended 3_1-helix or poly (Pro) II helix; short dashed line, irregular structure.

Spectra were recorded on a Chirascan spectrometer (Applied Photophysics) calibrated with 1S-(+)-10-camphorsulfonic acid. Measurements were carried out at 298 K in a 0.1-cm path length quartz cuvette. Spectra were measured in a 180–260 nm wavelength range with an increment of 0.2 nm, bandpass of 0.5 nm and integration time of 1 s. Spectra were processed, baseline corrected, smoothed and converted with the Chirascan software. Spectral units were expressed as the mean molar ellipticity per residue using the protein concentration determined at 280 nm using a NanoDrop spectrometer (2000c spectrophotometer, ThermoScientific). Estimation of the secondary structure content was carried out using the CDSSTR, CONTIN and SELCON3 approaches available on the DICHROWEB server (dichroweb.cryst.bbk.ac.uk/).

9. Lipid reconstitution

Either egg PC (L-α-phosphatidylcholine, Sigma-Aldrich, Ref. P3556), a mixture of egg PC and cholesterol (70/30, w/w) or home-made lipids from pig liver were used for lipid reconstitution which was performed by dialysis according to classical protocols or with the GRecon method (refer to section II.4).

9.1. Preparation of lipids from pig liver

Lipids were prepared from a fresh pig liver hold at the slaughterhouse directly after the animal death (**Figure III.11A**), together with Britta Kunert, post-doc in the lab. Briefly, 200 g of liver were cut into pieces using a kitchen knife and crushed using a blender. Considering that the liver was composed of about 160 mL water, 200 mL chloroform and 400 mL methanol were added (Folch *et al*, 1957; Bligh & Dyer, 1959). The mixture was incubated in a glass stoppered bottle (**Figure III.11B**) under agitation for 2 h at RT before being filtered using prepleated qualitative filter paper (Whatman, Ref. 1202-240) (**Figure III.11C**). Then, 200 mL chloroform and 200 mL water were added to the mixture in a separation funnel which was manually agitated, taking care to carefully degas the mixture. After overnight incubation under the hood (**Figure III.11D**), the bottom phase was harvested and solvent was evaporated for 15 min in a rotary evaporator (**Figure III.11E**).

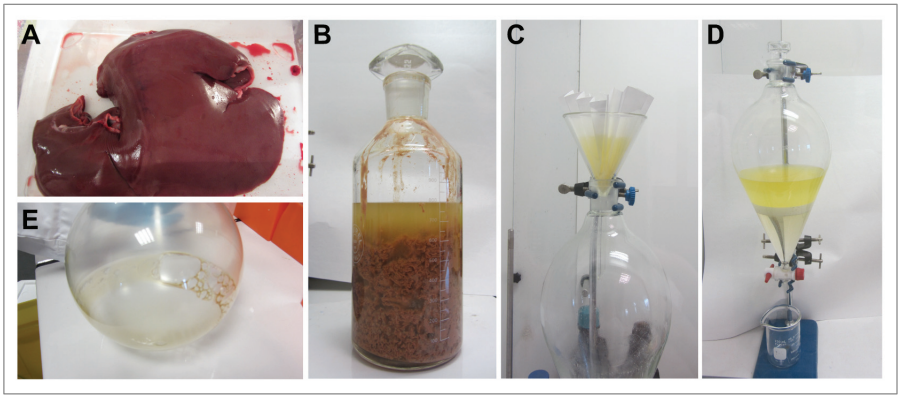


Figure III.11 ***Preparation of lipids from pig liver***. (**A**) Pig liver. (**B**) Crushed liver with chloroform and methanol. (**C**) Filtering of the liver extract obtained in (B). (**D**) Phase separation in a funnel. (**E**) Lipids after rotary evaporation.

The lipids were resuspended in a small volume of chloroform and aliquoted in glas tubes which were weighted beforehand. The chloroform was first evaporated under a liquid nitrogen flow and further in a lyophilizer for 2 h. The amount of lipids was

determined by weighting the glas tubes. The lipids were finally resuspended in chloroform and the desired amount of lipids was aliquoted in small glas bottles and stored at -20°C under nitrogen after solvent evaporation.

9.2. Preparation of Bio-Beads

1 to 2 g Bio-Beads (Bio-Beads® SM2 adsorbent, Bio-Rad, Ref. 152-8920) were incubated with 15 mL methanol in a 25-mL becher under gentle agitation for 30 min, under the hood. Agitation was then stopped to let the Bio-Bead settle at the bottom of the becher and the methanol was carefully removed. The Bio-Beads were then washed four times with 20 mL water under gentle agitation for 10 min. Note that the Bio-Beads need more time to settle at the bottom of the becher when they are in water. The Bio-Beads were stored at 4°C up to one month in water.

9.3. Reconstitution by dialysis (classical protocol)

Detergent-mediated membrane protein reconstitution is the most widely used strategy for reconstitution because membrane proteins are purified and solubilized in detergents. The reconstitution step consists in replacing the detergent by lipids. According to classical protocols, membrane proteins are first cosolubilized with phospholipids in the appropriate detergent in order to form an homogenous solution of lipid/protein/detergent and lipid/detergent micelles. This solution is dialyzed for several days against a buffer containing Bio-Beads. The removal of the detergent using Bio-Beads results in the formation of proteoliposomes which are bilayer vesicles with incorporated protein (Rigaud *et al*, 1995; Rigaud & Lévy, 2003).

The lipids were first solubilized either in DDM (Anatrace, Ref. D310) or Triton X-100 (Sigma-Aldrich, Ref. T8787) with a detergent-to-lipid ratio of 10 (mol/mol) (Kunert *et al*, 2014). In more detail, adequate amounts of lipids resuspended in chloroform and detergent resuspended in methanol were mixed. The solvents were first evaporated under a nitrogen flow while manually turning the glas tube to form a thin film on the wall of the tube. The solvents were further evaporated in a lyophilizer for at least 30 min. The dried lipid/detergent mixture was carefully resuspended in water (half of the

final volume) before the addition of twice concentrated buffer. Final concentration was 5 mg/mL in 100 mM Tris pH 8.0, 150 mM NaCl and 1 mM EDTA. Finally, the solubilized lipids were sonicated for 15 min in a sonication bath (Elmasonic S30H, Elma, Germany).

Purified protein, solubilized in 0.1% DDM, was mixed with the solubilized lipids at a given lipid-to-protein ratio (w/w). Protein/lipid/detergent samples were dialyzed at RT for 6 days against a buffer containing 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.025% NaN_3 and hydrophobic polystyrene beads (Bio-Beads SM-2 adsorbents, Bio-Rad, Ref. 152-3920). A dialysis membrane with a molecular weight cut-off of 6-8000 Da (Spectra/Por, Ref. 132645) was used. A ratio Bio-Beads-to-detergent of 100 (w/w) was applied. Detergent removal by dialysis and capture on the Bio-Beads leads to the formation of proteoliposomes (Rigaud & Lévy, 2003). Whereas small dialysis cups fitting in a 24-well plate (refer to section III.4.2.2.) were used for preliminary reconstitution tests, dialysis bags of the same cut-off were used for larger reconstitution scale.

9.4. Reconstitution using the GRecon method

The GRecon method consists in reconstituting membrane proteins in lipids directly on a density gradient (Althoff *et al*, 2012). Briefly, the membrane protein solubilized in detergent is loaded on a sucrose gradient containing detergent-solubilized lipids and α-cyclodextrin. During centrifugation, the protein goes through the gradient and meets α-cyclodextrin, which removes progessively the detergent, as well as solubilized lipids, leading to the formation of proteoliposomes.

The concentration of α-cyclodextrin required to remove the detergent and thus precipitate the protein of interest had to be determined beforehand. The protein sample in 0.1% DDM was incubated with a concentration range of α-cyclodextrin overnight at 26°C. The samples were then centrifuged for 1 h at 20,000 *g*, 26°C. The pellet and the supernatant obtained for each tested condition were analyzed by SDS-PAGE followed by Coomassie blue staining.

The lipids were solubilized in Triton X-100 (Sigma-Aldrich, Ref. T8787) with a detergent-to-lipid ratio of 1 (w/w) (Althoff *et al*, 2012). The preparation of the lipids

was otherwise performed as described in the previous section for the lipid reconstitution by dialysis.

Sucrose gradients were formed using a gradient maker (Gradient Master, Biocomp Instruments, Canada): 2 mL of a 60% sucrose solution containing detergent-solubilized lipids and α-cyclodextrin were overlaid with 2 mL a 30% sucrose solution. All sucrose solutions were in 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 1 mM DTT. The parameters for the gradient formation were determined experimentally by Denis Lacabanne, PhD student in the lab, and were the following: time = 1 min 5 s, rotation angle = 86°, speed = 18 rpm, cap = long. Note that lipids were omitted for a negative control with protein only.

The preformed gradients were overlaid with 200 µL detergent-solubilized protein (or only affinity elution buffer for a negative control with lipids only) and centrifuged at 200,000 *g*, 26°C for 17 h. After ultracentrifugation, 200 µL fractions were then harvested and analyzed by SDS-PAGE followed by Coomassie blue staining.

10. Membrane protein trapping with amphipol A8-35

The trapping consists in replacing the detergent with the amphipol, keeping the membrane protein soluble in a detergent-free aqueous solution (Tribet *et al*, 1996) (Workshop « Applications of Amphipols to Membrane Protein Studies, October 21-24, 2013, Paris). The membrane protein solubilized in 0.1% DDM was incubated, directly after the affinity purification step, with the amphipol A8-35 (Anatrace, Ref. A835) for 30 min on the wheel, at RT. An amphipol-to-protein ratio of 5 (w/w) was applied. For practical reasons, the protein/detergent/amphipol mixture was stored overnight at 4°C. The detergent was then removed while adding Bio-Beads directly to the protein/detergent/amphipol mixture, with a Bio-Beads-to-detergent ratio of 20 (w/w). Before weighting them, the Bio-Beads were carefully dried on a laboratory paper towel (refer to section III.9.2. for the preparation of the Bio-Beads). The mixture was then incubated for 2 h at RT, on the wheel. Agitation was stopped to let the Bio-Beads settle at the bottom of the tube. The supernatant was carefully harvested and centrifuged at 20,000 *g*, 4°C for 30 min to remove potential aggregates. In parallel, protein solubilized in 0.1% DDM was incubated with Bio-beads without A8-35 to verify that the detergent removal was complete. Protein/amphipol complexes were stored in liquid nitrogen. The whole procedure is summarized in **Figure III.12**.

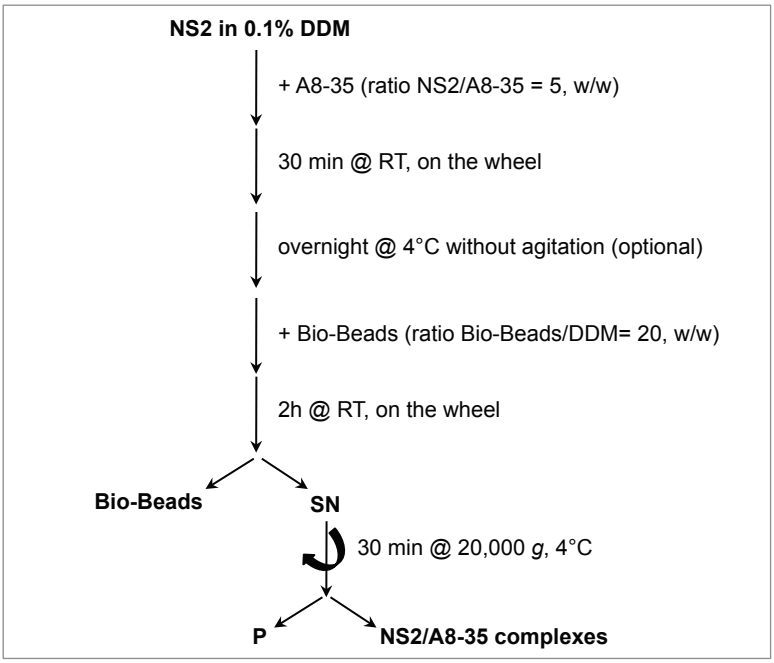


Figure III.12 Schematic representation of the *trapping of NS2 with the amphipol A8/35*.

11. SDS-PAGE and Western blotting

11.1. SDS-PAGE analysis

All expression experiments were assessed using 15% Coomassie blue stained SDS-PAGE gels. Samples were resuspended in a loading buffer containing 62.5 mM Tris-HCl pH 6.8, 2% SDS (w/v), 10% glycerol (v/v), 5% β-mercaptoethanol (v/v) and 0.01% bromophenol blue (w/v) and incubated at RT for 15 min before loading. For each sample, 10 µL were loaded on the SDS-PAGE gels. Moreover, 5 µL of Precision Plus Protein™ All Blue Standards (Bio-Rad, Ref. 161-0373) were systematically loaded on the gels for size calibration. The gels were run using Tris-Glycine-SDS buffer in a Mini-PROTEAN® Tetra Cell system (Bio-Rad, USA) at 80 V until the samples have reached the separation gel, and further at 120 V.

11.2. Western blotting analysis

Western-blotting analysis was carried out by protein transfer onto a nitrocellulose membrane using an iBlot® gel transfer device (Life Technologies). The nitrocellulose membrane was blocked with 5% non-fat milk powder in PBS-T buffer (12 mM sodium phosphate pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.05% Tween®20). In the case of the pellet and SN-beads samples, the nitrocellulose membrane was incubated with the StrepMAB-Classic, HRP conjugate antibody (IBA Lifesciences, Ref. 2-1509-001) for 1 h at RT (dilution 1/30,000). For a more sensitive detection of the protein of interest in the CFS, the nitrocellulose membrane was incubated with the StrepMAB-Classic primary antibody (IBA Lifesciences, Ref. 2-1507-001) for 1 h at RT (dilution 1/5,000) and further incubated with an anti-mouse IgG HRP-conjugated secondary antibody (Promega, Ref. W4021) for 1 h at RT (dilution 1/4,000). A primary mouse antibody against the C-terminal domain of NS2, kindly provided by Ralf Bartenschlager, was also used (dilution 1/1,000). In all cases, epitope-containing bands were visualized using the ECL Prime Western Blotting Detection Reagent (GE Healthcare, Ref. RPN2232).

12. Benzonase production

The benzonase is an endonuclease which degrades DNA and RNA independently of their shape (simple or double strand, circular or linear). In order to reduce very high costs when using commercial benzonase, I implemented and optimized the production of this enzyme in the laboratory. *E. coli* bacteria strain BLR transformed with the pMEssbenzonase plasmid were kindly provided by Jean-Michel Betton (Institut Pasteur, Paris). The benzonase is expressed in the periplasm and can be easily isolated by osmotic shock.

12.1. Bacterial cultures

A preculture was first prepared inoculating LB medium (Sigma-Aldrich, Ref. L3022) containing 100 µg/mL ampicillin with bacteria stored as glycerol stock. The preculture was incubated overnight at 37°C, 125 rpm. A culture was then prepared diluting the preculture in M9 medium at $A_{600} \approx 0.3$. The M9 medium contains 2 mM $MgSO_4$, 100 µM $CaCl_2$, 38 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 2 g/L glucose, 2 g/L NH_4Cl, trace elements, vitamines and 100 µg/mL ampicillin. The culture was incubated at 37°C, 125 rpm up to $A_{600} \approx 0.8$. Addition of 500 µM IPTG (Isopropyl β-D-1-thiopyranosid, Euromedex, Ref. EU0008-C) was performed to induce benzonase production and the culture was further incubated at 37°C, 125 rpm up to $A_{600} \approx 2.5$.

12.2. Isolation of benzonase

Isolation of benzonase was performed according to the protocol provided by Jean-Michel Betton. Unless precised, all following steps were performed at 4°C.

The bacterial culture was centrifuged for 10 min at 6,000 rpm ((JA-10 fixed angle rotor, Beckman Coulter). The pellet was resuspended in 1/10 volume of cold 30 mM Tris-HCl pH 7.8 buffer and centrifuged for 10 min at 5,500 rpm (fixed angle rotor F34-6-38, Eppendorf). The pellet was resuspended in TSE buffer containing 30 mM Tris-HCl pH 8.0, 20% sucrose and 1 mM EDTA. The volume of TSE buffer to use was calculated as follows :

$$X \text{ mL TSE} = (A_{600} \times \text{bacterial culture volume in mL} \times 25) / 1000$$

The suspension was incubated for 10 min at RT with regular shaking and centrifuged for 10 min at 8,000 rpm. The pellet was quickly resuspended in 1 volume ice chilled water, homogenized and centrifuged for 10 min at 9,000 rpm. The supernatant containing the benzonase was filtered (0.45 µm filter) and stored at -20°C in a buffer containing 20 mM Tris-HCl pH 8.0, 20 mM NaCl, 2 mM $MgCl_2$ and 50% glycerol.

12.3. Production yield and activity test

The production of benzonase was initially optimized in LB and M9 media, showing that M9 medium leads to a higher expression yield of benzonase (data not shown). Benzonase was produced here starting from 1 L culture in M9 medium. The amount of purified enzyme was estimated on Coomassie gel (**Figure III.13**) : approximately 30 mg benzonase were obtained from 1 L culture.

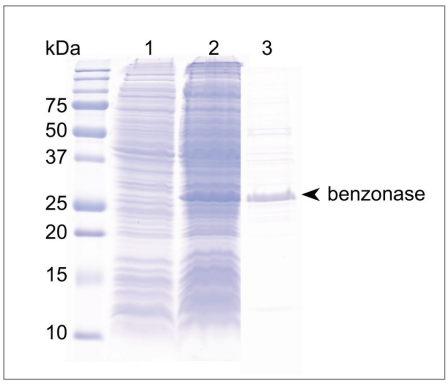


Figure III.13 ***Production of benzonase in M9 medium.*** 1, culture before induction ($A_{600} \approx 0.8$); 2, culture after induction with 500 µM IPTG ($A_{600} \approx 2.5$); 3, purified benzonase. 8 µL purified benzonase in 50% glycerol aqueous buffer were loaded on the gel and the amount was estimated to 1.5 µg. Since 160 mL benzonase were produced, the total amount was estimated to be approximately 30 mg.

The activity of benzonase was tested using low molecular weight DNA from salmon sperm (Sigma-Aldrich, Ref. 31149-10G-F). Moreover, the activity of home-made benzonase was compared to the activity of commercial Benzonase® Nuclease (Sigma-Aldrich, Ref. E1014, ≥ 250 units/µL). DNA from salmon sperm was diluted at different concentrations (20 mg/mL, 10mg/mL, 5 mg/mL, 2.5 mg/mL, 1 mg/mL, 0.5 mg/mL, 0.25 mg/mL and 0.02 mg/mL) in a buffer containing 30 mM Tris-HCl pH 8.8 and 2 mM $MgCl_2$. 1 µL of either commercial or home-made benzonase was added to 1 mL DNA solution and the mixture was incubated at 37°C for 30 min. Analysis of the samples on agarose gel (**Figure III.14A**) showed that home-made benzonase is slightly more efficient than the commercial one, 1 µL being enough to digest 2 mg DNA in 30 min at 37°C. Note that

this activity test is performed once a year in order to check the stability of the benzonase preparation.

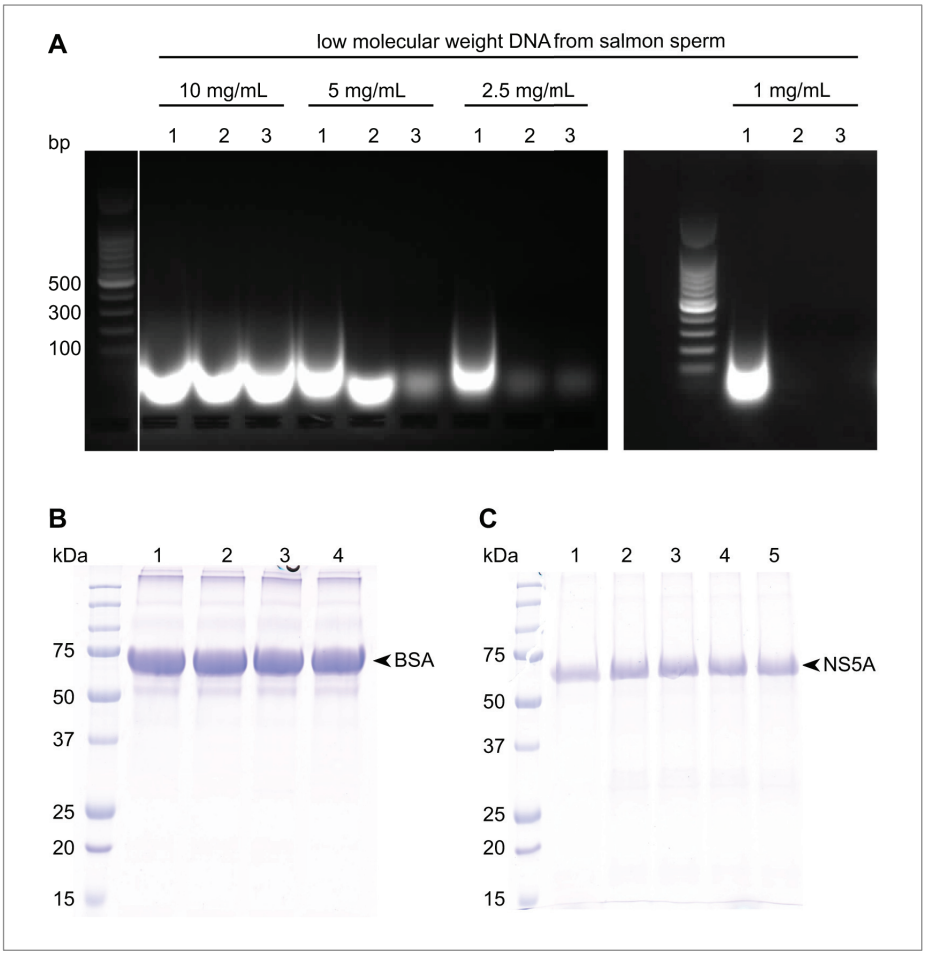


Figure III.14 ***Quality control of home-made benzonase.*** (A) Activity test using low molecular weight DNA from salmon sperm. The samples were loaded on a 1% agarose gel containing GelRed Nucleic Acid Gel Stain (Interchim, Ref. BY1740) for nucleic acid detection under UV light. BenchTop 100bp DNA Ladder (Promega, Ref. G8291) were also loaded on the gel for size calibration. 1, no benzonase; 2, commercial benzonase; 3, home-made benzonase. (B) and (C) Control of potential protease activity of the benzonase sample. Protein samples were analyzed by SDS-PAGE followed by Coomassie blue staining. A black arrowhead indicates the bands corresponding to the protein of interest. (B) Incubation of benzonase with BSA (bovine serum albumin). 1, BSA stored at -20°C without benzonase; 2, BSA without benzonase; 3, BSA incubated with commercial benzonase; 4, BSA incubated with home-made benzonase. Samples were incubated for 72 h at RT. (C) Incubation of home-made benzonase with full-length NS5A. NS5A was incubated either 1, without benzonase or with benzonase at 37°C for 2, 30 min; 3, 1 h; 4, 2 h; 5, 4 h.

Since some contaminants were detected on the Coomassie gel in the final benzonase sample, we verified whether there was any protease in this preparation. First, 500 µg BSA (bovine serum albumin) at a concentration of 1 mg/mL was incubated with 2.5 µL

of either commercial or home-made benzonase for 72 h at RT. Moreover, BSA stored at -20°C and BSA incubated for 72 h at RT without benzonase were analyzed as controls. As shown in **Figure III.14B**, the addition of benzonase had no effect on the integrity of BSA. In addition, 30 µg full-length NS5A protein (0.25 mg/mL) were incubated with 1 µL home-made benzonase at 37°C for 30 min, 1 h, 2 h and 4 h. No degradation of NS5A was observed (**Figure III.14C**), showing that the home-made benzonase preparation has no protease activity and can be used instead of the commercial benzonase.

IV. RESULTS AND DISCUSSION

1. Implementation and optimization of the wheat germ cell-free system

The development of the wheat germ cell-free system in the lab is described in this section. My work consisted in the optimization of the wheat germ extract preparation, of the production of good quality mRNA as well as of the various expression parameters. The screen of 75 protein constructs for their expression in this system as well as evidence for protein activity with a NS3-4A precursor construct are also presented in this section. All purification and characterization aspects will be detailed in the following sections.

1.1. Production of wheat germ extracts

Wheat germ extracts (WGEs) are prepared from durum wheat seeds, essentially according to the protocol developed by Yaeta Endo and coworkers at Ehime University in Japan (Takai *et al*, 2010) (refer to section III.3. for technical details). Briefly, the germs are isolated from the seeds by grinding and sieving, followed by solvent flotation. Time-consuming eye selection is then necessary to remove manually remaining brownish material, broken germs or germs with too much endosperm. The second critical step of WGE preparation is the efficient elimination of residual endosperm by extensive washing with water and detergent. Indeed, the endosperm contains translation inhibitors which substantially affect WGE quality. After washing, the germs are grinded using a kitchen blender. Several centrifugation steps allow to isolate the WGE from lipids and precipitated material. Moreover, several gel filtration steps on G25 columns allow the removal of small molecular weigth components such as amino acids from wheat germs, as well as buffer exchange. Improvments of the protocol of (Takai *et al*, 2010) are detailed in section III.3.

In the initial optimization phase of this work, home-made wheat germ extracts were systematically compared to the commercial WEPRO1240 from CellFree Sciences. The green fluorescent protein (GFP) was used for this purpose because its expression can be directly measured using a fluorimeter. After several optimization steps such as the grinding of the seeds using a blender or the upscaling of WGE preparation, the results obtained were very satisfactory. Indeed, the GFP expression yields with our home-made WGEs were similar to the commercial ones. However, a quality drift occured over time and lower GFP expression yields were obtained.

In order to understand this drift, 3 commercial (WEPRO1240 and WEPRO2240 from CellFree Sciences) and 8 home-made WGEs were tested in parallel (**Figure IV.1A**). Already the different commercial WGEs showed different efficiency. In line with the WEPRO2240 being an improved version of the WEPRO1240, the WEPRO2240 showed about 25% more efficiency than the WEPRO1240. Home-made WGEs prepared in 2012 (*e.g.* #120228 and #120423, note that the WGEs were annotated according to the nomenclature yymmdd) were about as efficient as the commercial WEPRO1240. However, WGEs prepared in 2013 and 2014 (*e.g.* #130926 and #140121) were about 50% less efficient than the former ones to express GFP. Among the various tested parameters such as the origin of the numerous reagents, the main reasons for this loss in efficiency were the conservation of wheat seeds over too long periods of time as well as the absence of the last concentration step in the protocol. Indeed, the WGE #140918D which was prepared with fresh wheat seeds and concentrated during WGE preparation (40 mL WGE were concentrated down to 30 mL) gave similar results than the older WEPRO1240 and older home-made WGEs (**Figure IV.1A**).

In addition, the WGE #140918D was compared to a subsequent home-made WGE (#141125D) prepared according to the same protocol and starting from the same wheat seeds. As illustrated in **Figure IV.1B**, both WGEs gave similar GFP expression levels, showing that the latest version of the WGE preparation protocol was reproducible. Moreover, the efficiency of the WGE #141125D was tested before and after the final concentration step, highlighting the necessity of this step. The WGE was indeed twice as efficient after concentration.

In order to verify whether there was a correlation between these values and the WGE efficiency, we also measured the absorbance at 260 nm (A_{260}) and 280 nm (A_{280}), and calculated the A_{260}/A_{280} ratio for all WGEs tested in **Figure IV.1A**. As shown in **Figure IV.1C,** two populations were observed on the plot. While WGEs with a A_{260}/A_{280} ratio lower than 1.5 were poorly efficient, commercial and good home-made WGEs had a A_{260}/A_{280} ratio between 1.5 and 2.25. Although information is missing for A_{260}/A_{280} ratios higher than 2.8 (WGE #141125D), the A_{260}/A_{280} ratio seems to be a good indicator of WGE quality.

To conclude, the home-made WGEs are on average 75% as efficient as the latest commercial WGEs. Despite this difference, it remains worth, from a financial point of view (1 mL WEPRO2240 costs around 800 €), to prepare home-made WGEs. Further improvement could come from testing different wheat strains.

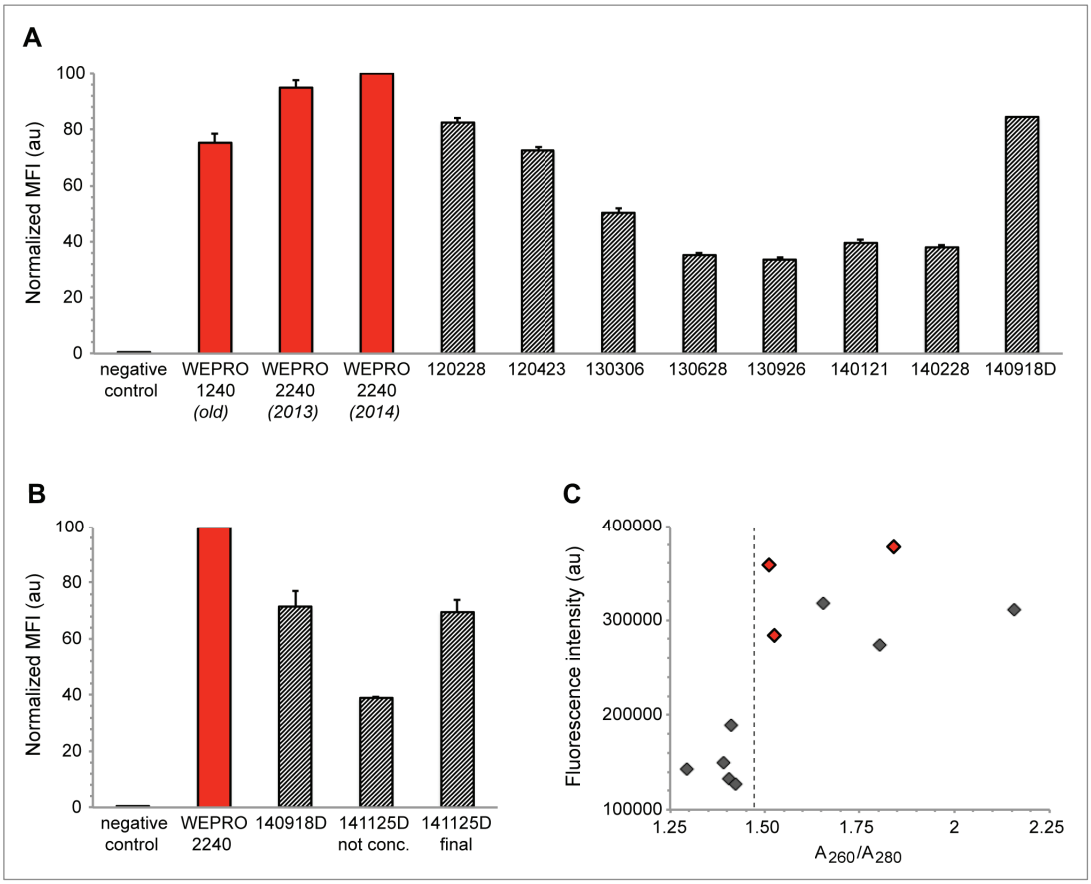


Figure IV.1 ***Quality control of wheat germ extracts.*** (**A**) Comparison of 3 commercial and 8 home-made WGEs for the expression of GFP. Commercial WGEs are in *red*, home-made WGEs in *grey*. Mean fluorescence intensity (MFI) was normalized compared to GFP expressed with the commercial WEPRO2240 from 2014. (**B**) Comparison of two recent home-made WGEs and impact of WGE final concentration step. Mean fluorescence intensity (MFI) was normalized compared to GFP expressed with the commercial WEPRO2240 from 2014. (**C**) Relationship between WGE efficiency for GFP expression and A_{260}/A_{280} ratio. Dots for commercial WGEs are in *red*, those of home-made WGEs in *grey*. A dotted line represents the threshold between high quality WGEs and less efficient WGEs.

1.2. Production of mRNA

The synthesis of mRNA is made by the SP6 RNA polymerase using the pEU-E01 plasmid which carries the gene of interest as template (refer to section III.4.1. for technical details). As shown in **Figure IV.2**, mRNAs synthesized from several constructs I prepared during this work (refer to section III.1.) were analyzed on agarose gel, showing their good quality. There is no SP6 terminator on the pEU-E01 plasmid and the SP6 RNA polymerase has a preference to fall off on the origin of replication, which leads to a kind of ladder of discrete mRNAs. There was no smear or ladder pattern of less base pairs than the gene of interest, and the mRNA bands on the gel were focused, showing that mRNA was not degraded.

The 12 constructs shown here correspond to the green fluorescent protein (GFP) as well as to various domains of p7, core and NS5A proteins. As detailed afterwards (**Table IV.1**), 10 of these contructs were successfully expressed while one (core 116-171) could not be expressed and the expression of another one (p7) was very low. However, there was no significant difference between their mRNA and the mRNA of successfully expressed constructs, suggesting that their weak expression was not due to inefficient transcription. Secondary structures of mRNA could for example compromise translation (Mureev *et al*, 2009). Changing the position of the tag could help to overcome such problem.

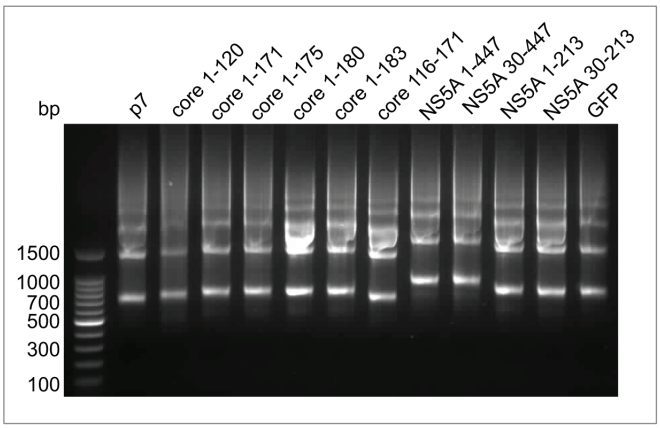


Figure IV.2 ***Analysis of mRNA on agarose gel.*** For each of the 12 constructs shown, 2 µL of transcription mix were loaded on a 1% agarose gel.

It is worth to note that in the initial steps of this work, mRNA was precipitated with ethanol after transcription, as described for example in (Morita *et al*, 2003; Kawasaki *et al*, 2003; Endo & Sawasaki, 2004; Takahashi *et al*, 2009). The mRNA pellet was then resuspended either in water or feeding buffer before translation. However, resuspension of the mRNA pellet was a delicate issue and due to often poorly solubilized mRNA, GFP expression yields were not reproducible. The mRNA was therefore used directly after transcription without further treatment as described in (Takai *et al*, 2010), leading to higher expression yield and better reproducibility. Note also that the amount of mRNA was not systematically quantified before translation. Indeed, residual free NTPs in the transcription mix hamper absorbance measurements and we failed quantifying mRNA using a Quant-iT™ RiboGreen® RNA Assay kit (Life Technologies) due to extremely high background. However, in a reasonable range, the mRNA amount seems not to influence too much translation efficiency (Masaki Madono, CellFree Sciences, personal communication). A white precipitate of magnesium pyrophosphate should appear during transcription. In the absence of this precipitate, which rarely occured, translation was not performed.

1.3. Optimization of the expression conditions

Main expression parameters such as temperature or wheat germ extract concentration were tested using GFP as model protein. Indeed, its expression can be followed directly by fluorescence intensity measurement, and the correspondance between given GFP amounts and fluorescence intensity allows to estimate the amount of GFP produced in the WGE-CF system (**Figure IV.3A**). Moreover, two modes of translation were compared for small scale expression : the bilayer method and the dialysis method (CECF).

1.3.1. Experimental conditions for GFP expression

All experiments described here were performed using the bilayer method (refer to section III.4.2.1. for technical details).

First of all, the kinetic of GFP expression was analyzed at 26°C in a 96-well plate. As shown in **Figure IV.3B**, after a short lag phase, GFP production was linear during the first 12 hours of translation and then much slower between 12 and 26 hours. A maximum was reached after 26 hours, followed by a decrease of fluorescence for longer incubation time, indicating some degradation process. In fact, the accumulation of byproducts in the reaction mix, especially phosphate, tends to inhibit protein synthesis (Schwarz *et al*, 2008). Moreover, the progressive decrease of available substrates such as amino acids, creatine phosphate, ATP and GTP slows the reaction down. Performing translation overnight, *i.e.* for 16 h-18 h, is therefore optimal. Longer incubation times would not significantly improve the expression yield (less than 10%) and might induce some degradation.

To evaluate the effect of temperature on protein expression, GFP was expressed in 96-well plates at three different temperatures (18°C, 22°C and 26°C) and fluorescence intensity was measured at three time points (4 h, 16 h and 20 h) for each condition. Results are shown in **Figure IV.3C**. Although GFP was expressed at 18°C, the expression yield was much lower than at higher temperatures. Indeed, expression yield after 20 hours reaction was more than 60% higher at 26°C than at 18°C. The difference between 22°C and 26°C was less significant : expression yield after 20 hours reaction was only 12% higher at 26°C than at 22°C. Considering these results and in order to get a better protein folding, we chose 22°C as a standard temperature to express membrane

proteins. Note that the kinetic of GFP expression at 26°C was similar to the one observed previously (**Figure IV.3B**). Concerning the estimation of GFP expression yield by referring to the standard curve (**Figure IV.3A**), one can calculate that on average 6 mg GFP are expressed per mL WGE at 26°C (*i.e.* 60 µg per well in a 96-well plate).

Furthermore, the effect of wheat germ extract concentration (expressed in A_{260}) on GFP expression was tested too (**Figure IV.3D**). Four WGE concentrations were tested to express GFP at 26°C in a black 96-well plate, directly in the fluorimeter. Translation was performed over 6 hours and fluorescence intensity was measured every 15 min. Since the reaction volume in the wells was higher than that used for standard fluorescence intensity measurements (220 µL instead of 150 µL), the values obtained can not be compared to those of panels A, B and C. However, they allowed us to compare the different conditions tested here. A lag phase was clearly visible, followed by a linear phase and a plateau phase which tends to be already reached after 5 hours. This early latter phase was due to the evaporation occuring in the fluorimeter, changing the volume and thus the z-position, thus leading to distorded values of fluorescence intensity. Moreover, the plateau phase was not observed when translation was performed according to the standard procedure (data not shown). Nevertheless, expression yield was 5 times higher when the WGE concentration was twice concentrated (A_{260} = 120 *vs* A_{260} = 60). Further tests, which were repeated either directly in the fluorimeter or according to the standard procedure, confirmed this result. There is thus clearly a synergic effect of the WGE concentration, at least during the first steps of translation. This observation is consistent with the fact that WGEs are more efficient when a concentration step is performed during WGE preparation (refer to section IV.1.1. for details).

We then verified whether working with larger volumes of feeding buffer would increase the expression yield of GFP. Translation was performed in a 48-well plate for 16 h at 26°C. Either 500 µL or 800 µL of feeding buffer were used with 50 µL translation mix (including 25 µL WGE). As shown in **Figure IV.3E**, a higher volume feeding buffer did not significantly increase the amount of GFP produced. A feeding buffer-to-translation mix ratio of 10 (v/v) seemed to be optimal for the bilayer method.

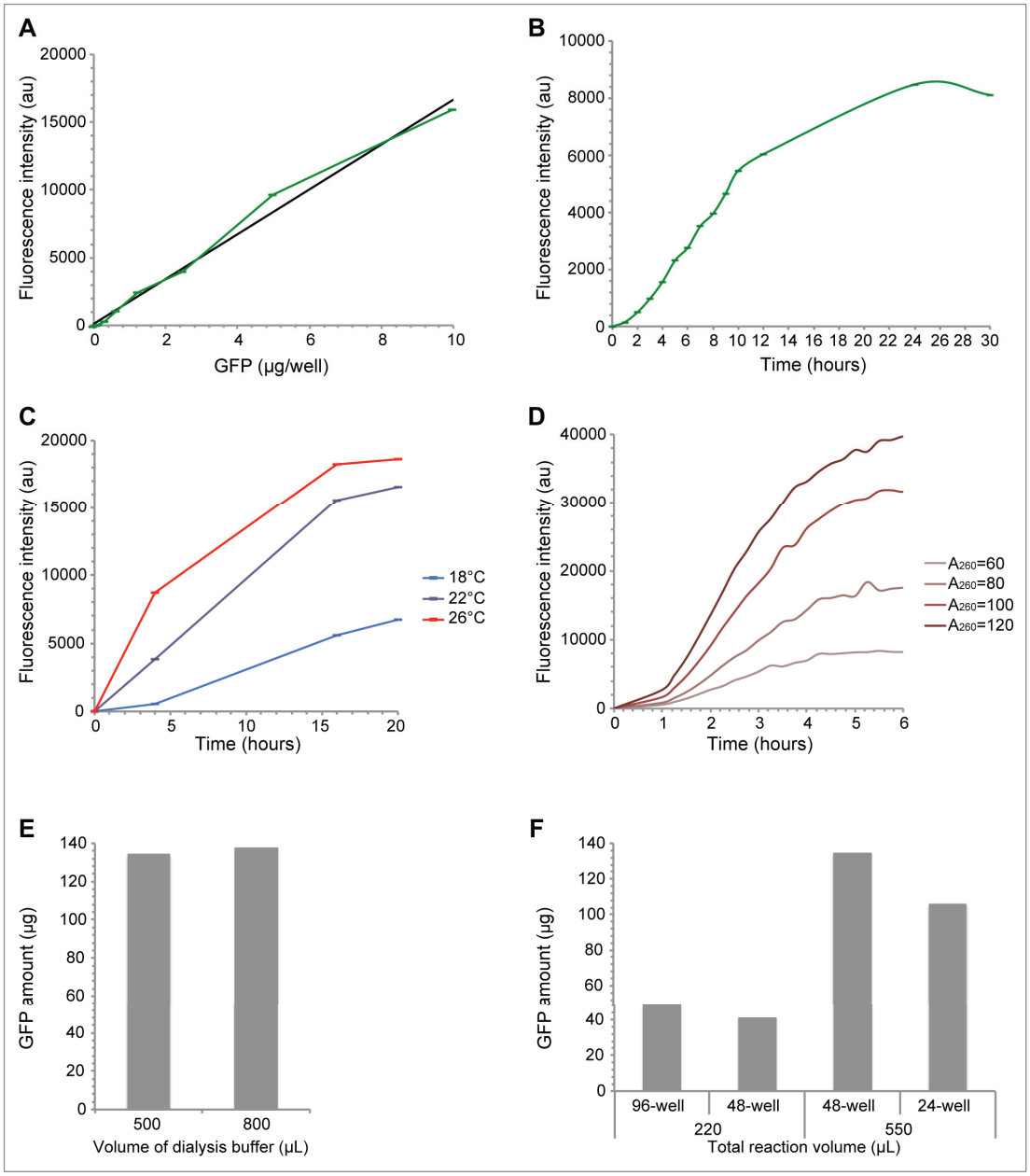


Figure IV.3 Experimental conditions for *GFP expression.* (**A**) GFP standard curve showing the correspondance between GFP amount and its fluorescence intensity (commercial GFP was from BPS Bioscience, Ref. 50277). (**B**) Kinetic of GFP expression at 26°C. Samples were diluted 1/8 for fluorescence intensity measurement. (**C**) Effect of temperature on GFP expression. GFP was produced at 18°C, 22°C and 26°C. Fluorescence intensity was measured after 4 h, 16 h and 20 h translation. Samples were diluted 1/4 for fluorescence intensity measurement. (**D**) Effect of WGE concentration (expressed in A_{260}) on GFP expression. Four final A_{260} were tested (60, 80, 100 and 120). Translation was performed directly in the fluorimeter at 26°C and fluorescence intensity was measured every 15 min over 6 h. (**E**) Effect of the feeding buffer volume on the expression yield. Translation was performed at 26°C in a 48-well plate with 50 µL translation mix (including 25 µL WGE) and either 500 µL or 800 µL feeding buffer. (**F**) Effect of the volume-to-surface ratio on the GFP expression yield. Translation was performed for 16 h at 26°C. Two reaction volumes were tested in parallel : Firstly, 20 µL translation mix (including 10 µL WGE) with 200 µL feeding buffer, either in a 96-well or in 48-well plate. Secondly, 50 µL translation mix (including 25 µL WGE) with 500 µL feeding buffer, either in a 48-well or in 24-well plate.

In addition, to evaluate the influence of geometric parameters, different plate formats were compared. Two reaction volumes were tested in parallel. Firstly, 20 µL translation mix (including 10 µL WGE) with 200 µL feeding buffer were incubated either in a 96-well or a 48-well plate. Secondly, 50 µL translation mix (including 25 µL WGE) with 500 µL feeding buffer were incubated either in a 48-well or a 24-well plate. Translation was performed for 16 h at 26°C in all cases. Note that the exchange surface is 0.34 cm^2 in a 96-well plate, 1 cm^2 in a 48-well plate and 1.9 cm^2 in a 24-well plate. As shown in **Figure IV.3F**, for both reaction volumes (220 µL and 550 µL), the expression yield starting from the same WGE amount was lower when the exchange surface was higher. The expression yield thus seemed to slightly increase with the total volume-to-surface ratio. Anyway, small scale expression tests were then performed in a standard way in 96-well plates using 20 µL translation mix (including 10 µL WGE) with 200 µL feeding buffer. For larger protein productions in 6-well plates, the volume-to-surface ratio was similar.

In addition, magnesium and potassium concentrations have been reported to be critical for cell-free systems, especially those based on *E. coli* lysates (Schwarz *et al*, 2008). The effect of potassium and magnesium concentrations was therefore also tested (data not shown). According to (Takai *et al*, 2010), standard concentrations are 100 mM potassium acetate and 3.3 mM magnesium acetate. We observed that 200 mM potassium acetate (with 3.3 mM magnesium acetate) strongly inhibited GFP expression whereas 15 mM magnesium acetate (with 100 mM potassium acetate) fully inhibited the system. Concentrations closer to the standard ones were therefore tested. Neither potassium acetate concentrations down to 50 mM or up to 150 mM (with 3.3 mM magnesium acetate) nor magnesium acetate concentrations down to 2 mM or up to 4.5 mM (with 100 mM potassium acetate) had significant effect on GFP expression. Potassium and magnesium concentrations thus seemed not to be as critical in the WGE-CF system as in cell-free systems based on *E. coli* lysates, and standard concentrations from (Takai *et al*, 2010) seemed to be optimal. Although this parameter is likely to be protein-dependent, the concentrations of potassium and magnesium were not further optimized for each protein tested in this work.

1.3.1. Bilayer method versus dialysis method (CECF)

The dialysis method is usually described in the literature as more efficient than the bilayer method (Harbers, 2014). We therefore compared the two approaches. The translation using the bilayer method was performed in a 48-well plate whereas the dialysis method was performed in a 24-well plate as described in section III.4.2.2. For both methods, 50 µL translation mix and 800 µL feeding buffer (*i.e.* feeding buffer-to-translation mix ratio of 16, according to (Klammt *et al*, 2007) and (Schneider *et al*, 2009) for the dialysis method) were incubated at 26°C. The same mRNA and feeding buffer were used for both methods. Two molecular weight cut-off (MWCO) were tested for the dialysis membrane (3,500 and 12-14,000 Da). The kinetic of GFP expression was analyzed over 24 hours with three time points for fluorescence intensity measurement at 16 h, 20 h and 24 h. As shown in **Figure IV.4A**, the GFP expression yield was similar with the two different MWCO used for the dialysis method. And although the expression yield of GFP was slightly higher with the bilayer method, the two methods led to similar GFP expression yields, this result being reproducible. In parallel, a shorter kinetic was also performed over 4 hours under the same experimental conditions and fluorescence intensity was measured after 1 h, 2 h and 4 h. This experiment showed that there was no significant difference between the three conditions tested in the early hours of the translation (data not shown). In the end, using a dialysis membrane did not seem to improve the translation efficiency and limiting diffusion between the translation mix and the feeding buffer could even lead to lower expression yields.

When the dialysis method is used, the addition of mRNA to the translation mix has been reported to improve the final expression yield (Madin *et al*, 2000; Endo & Sawasaki, 2004). We therefore expressed GFP over 40 hours using the dialysis method. The dialysis membrane used had a MWCO of 12-14,000 Da. As described above, 50 µL translation mix and 800 µL feeding buffer were incubated at 26°C. The addition of 12.5 µL mRNA and/or 800 µL fresh feeding buffer after 16 h or after 16 h and 24 h was tested. A negative control without addition of mRNA and/or feeding buffer was analyzed in parallel. Fluorescence intensity was measured at three different time points (16 h, 24 h and 40 h). As shown in **Figure IV.4B**, neither the addition of mRNA nor the exchange of feeding buffer allowed to improve significantly the GFP expression yield. Considering these results, we decided to work in a standard way with the bilayer method for small

scale expression tests. Indeed, in addition to similar expression yields in our hands, the bilayer method offers a much easier handling.

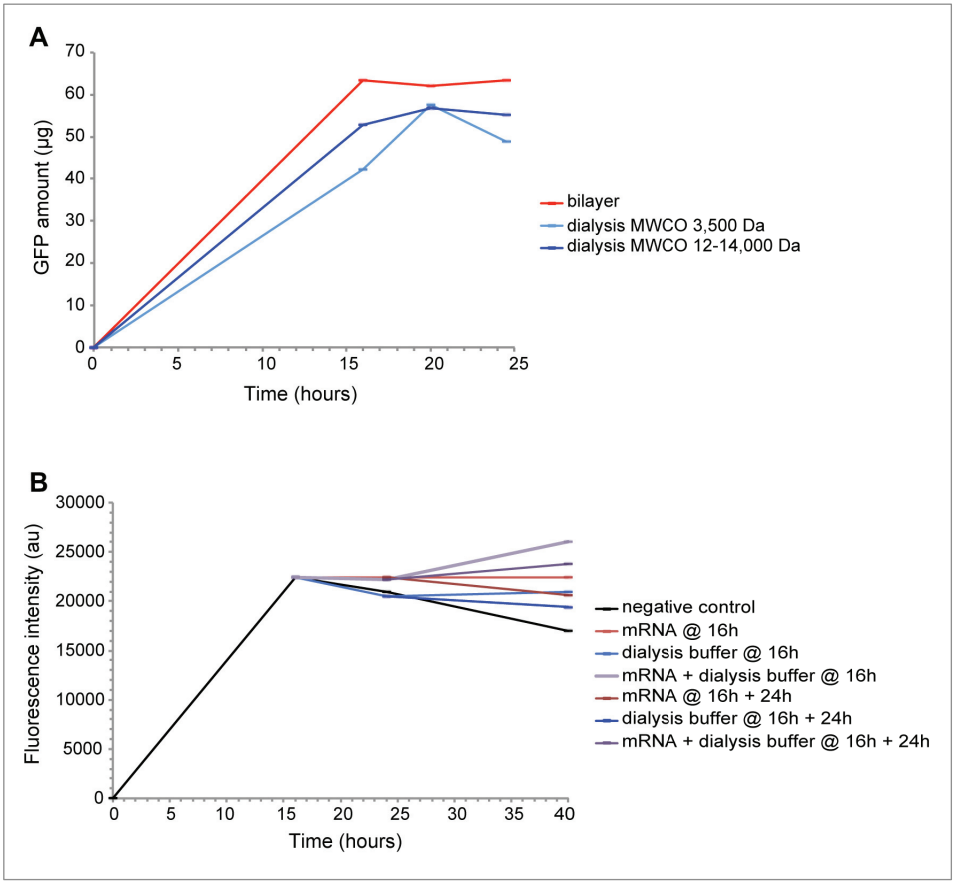


Figure IV.4 ***Bilayer method versus dialysis method.*** (**A**) Comparison of the bilayer and dialysis methods (as described in section III.4.2.). GFP was expressed over 24 h and fluorescence intensity was measured after 16 h, 20 h and 24 h. (**B**) Effect of the addition of mRNA or/and of feeding buffer on GFP expression using the dialysis method. GFP was expressed over 40 h. Addition of mRNA and/or feeding buffer was performed after 16 h or 16 h and 24 h. Fluorescence intensity was measured after 16 h, 24 h and 40 h.

1.4. Expression test of 75 different protein constructs

On the whole, 75 protein constructs corresponding mostly to HCV proteins but also to proteins from related viruses such as HBV, GBV-B or dengue virus as well as to various eukaryotic proteins were initially screened for expression in the precipitate mode (**Figure IV.5A**). Protein samples were prepared as summarized in **Figure IV.5B** for analysis. Out of these 75 constructs, there were 62 membrane proteins and 10 soluble proteins. More than 82% (62) of them were successfully expressed (**Figure IV.5C**). Detailed information on the 75 constructs is given in **Table IV.1**.

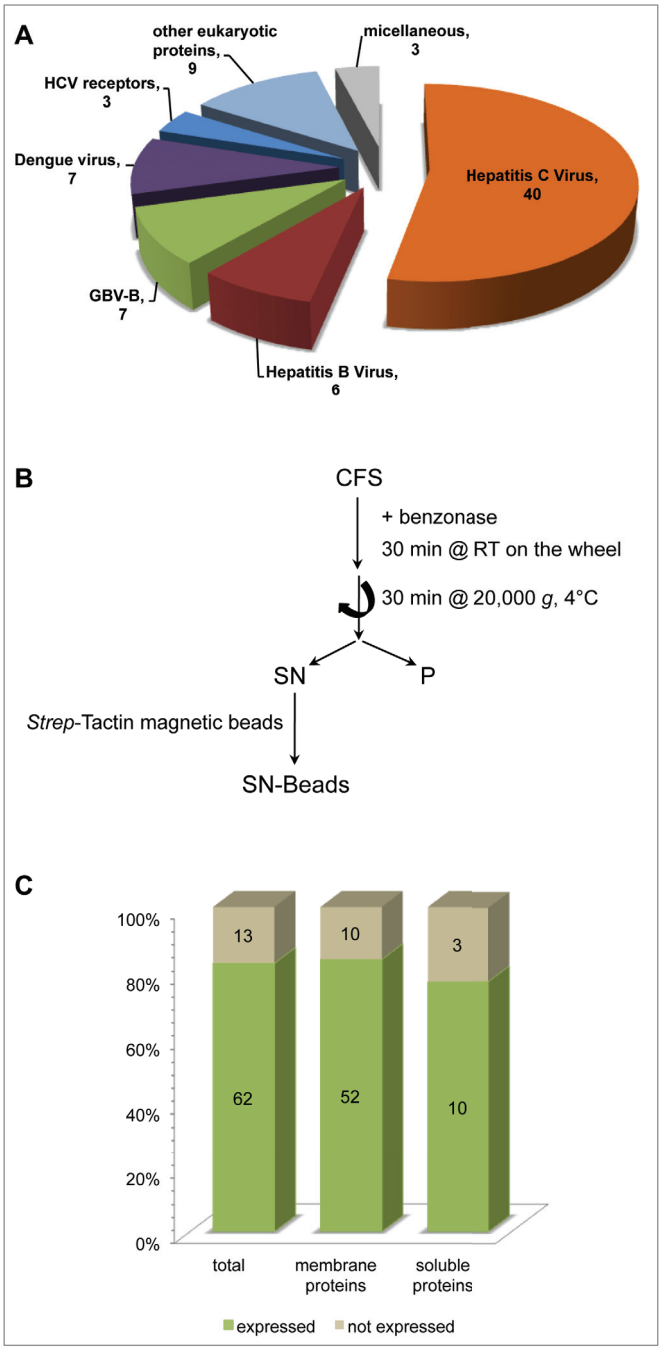


Figure IV.5 ***Expression of 75 different protein constructs using the WGE-CF system in the precipitate mode.*** (**A**) Distribution of constructs according to their origin organism. (**B**) Schematic representation of protein sample preparation for the analysis of small scale expression tests. CFS, cell-free sample; P, pellet obtained after centrifugation of CFS; SN, supernatant obtained after centrifugation of CFS; SN-Beads, sample enriched in protein of interest by incubation of supernatant with *Strep*-Tactin magnetic beads to capture tagged protein. CFS, P and SN-Beads were analyzed by SDS-PAGE followed by Coomassie blue staining and western blotting. Refer to sections III.5. and III.11, as well as Figure III.7 for technical details. (**C**) Results of the expression test of the 75 protein constructs in the precipitate mode (refer to Table IV.1 for details).

Table IV.1 (part 1/3) ***Expression of 75 different constructs using the WGE-CF system in the precipitate mode.*** -, no expression; +, good expression; ±, low expression level. TCS, thrombine cleavage site; YFP, yellow fluorescent protein; Trx, thioredoxine. Contructs were cloned by (a) M.-L. Fogeron, (b) V. Jirasko & R. Bartenschlager, (c) C. Boukadida & A. Martin, (d) M. Baril & D. Lamarre, (e) J. Gouttenoire & D. Moradpour, (f) S. Assil & M. Dreux, (g) D. Durantel, (h) J. Fresquet & C. Combet, (i) K. Mulder & R. Bartenschlager, (j) J. Molle & M.-L. Fogeron, (k) L. Gonzalez & P. Falson, (l) E. Benoit, (m) D. Paul & R. Bartenschlager, (n) M. Zayas & R. Bartenschlager, (o) L. Ballut, (p) Y. Endo, (q) Y. Jaillais

Organism	Protein	Strain	Construct	Clonage	Molecular mass (kDa)	Tag	Tag position	Soluble or membrane protein?	Expression (precipitate mode)
HCV	Core	J	1-120	(a)	13	StrepII	C-terminus	soluble	+
HCV	Core	J	1-171	(a)	20	StrepII	C-terminus	membrane	+
HCV	Core	J	1-175	(a)	20	StrepII	C-terminus	membrane	+
HCV	Core	J	1-180	(a)	20	StrepII	C-terminus	membrane	+
HCV	Core	J	1-183	(a)	20	StrepII	C-terminus	membrane	+
HCV	Core	J	116-171	(a)	6	StrepII	C-terminus	membrane	-
HCV	p7	Con1	full-length	(a)	7	StrepII	C-terminus	membrane	±
HCV	NS2	JFH-1	full-length	(b)	26	StrepII	N-terminus	membrane	+
HCV	NS2	JFH-1	full-length	(b)	26	StrepII + TCS	N-terminus	membrane	+
HCV	NS2	Con1	full-length	(b)	26	StrepII	N-terminus	membrane	+
HCV	NS2	Con1	full-length	(b)	26	StrepII + TCS	N-terminus	membrane	+
HCV	NS2	H77	full-length	(b)	26	StrepII	N-terminus	membrane	+
HCV	NS2	H77	full-length	(b)	26	StrepII + TCS	N-terminus	membrane	+
HCV	NS2	3a negro	full-length	(b)	26	StrepII	N-terminus	membrane	+
HCV	NS2	3a negro	full-length	(b)	26	StrepII + TCS	N-terminus	membrane	+
HCV	$NS2^{pro}$	Con1	94-217	(a)	14	StrepII	N-terminus	membrane-associated	+
HCV	$NS2^{pro}$	JFH-1	94-217	(a)	14	StrepII	N-terminus	membrane-associated	±
HCV	$NS2/NS3^{pro}$	JFH-1	1-217 / 1-213 (S1169)	(c)	46	Twin-StrepII	C-terminus	membrane	+
HCV	$NS2/NS3^{pro}$	JFH-1	1-217 (C997A) / 1-213 (S1169)	(c)	46	Twin-StrepII	C-terminus	membrane	+
HCV	$NS2^{pro}/NS3^{pro}$	JFH-1	94-217 / 1-213 (S1169)	(c)	36	Twin-StrepII	C-terminus	membrane	+
HCV	$NS2^{pro}/NS3^{pro}$	JFH-1	94-217 (C997A) / 1-213 (S1169)	(c)	36	Twin-StrepII	C-terminus	membrane	+
HCV	NS3/4A	Con1	full-length wild type	(d)	78	StrepII + TCS	N-terminus	membrane	+
HCV	NS3/4A	Con1	full-length S139A	(d)	78	StrepII + TCS	N-terminus	membrane	+
HCV	NS3/4A	Con1	full-length T631I	(d)	78	StrepII + TCS	N-terminus	membrane	+

Table IV.1 (part 2/3)

Organism	Protein	Strain	Construct	Clonage	Molecular mass (kDa)	Tag	Tag position	Soluble or membrane protein?	Expression (precipitate mode)
HCV	NS4B	JFH-1	1-260 (full-length)	(e)	27	His	C-terminus	membrane	+
HCV	NS4B	JFH-1	1-72	(e)	9	His	C-terminus	soluble	-
HCV	NS4B	JFH-1	40-130	(e)	10	His	C-terminus	membrane	+
HCV	NS4B	JFH-1	72-196	(e)	13	His	C-terminus	membrane	+
HCV	NS4B	JFH-1	129-196	(e)	8	His	C-terminus	membrane	+
HCV	NS4B	JFH-1	198-260	(e)	8	His	C-terminus	soluble	±
HCV	NS4B	Con1	full-length	(e)	27	StrepII	C-terminus	membrane	+
HCV	NS4B	H77	full-length	(e)	27	StrepII	C-terminus	membrane	+
HCV	NS4B	JFH-1	full-length	(e)	27	StrepII	C-terminus	membrane	+
HCV	NS4B	JFH-1	full-length	(e)	27	YFP-StrepII	C-terminus	membrane	-
HCV	NS5A	Con1	1-447 (full-length)	(a)	50	StrepII	C-terminus	membrane	+
HCV	NS5A	Con1	30-447	(a)	47	StrepII	C-terminus	soluble	+
HCV	NS5A	Con1	1-213	(a)	25	StrepII	C-terminus	membrane	+
HCV	NS5A	Con1	30-213	(a)	22	StrepII	C-terminus	soluble	+
HCV	NS5A	JFH-1	1-447 (full-length)	(f)	51	StrepII + TCS	N-terminus	membrane	-
HCV	NS5A	JFH-1	1-447 (full-length)	(f)	51	StrepII	C-terminus	membrane	+
HBV	Core	n.a.	full-length wt	(g)	20	no tag	-	membrane	+
HBV	Core	n.a.	delta NES	(g)	20	no tag	-	membrane	-
HBV	Core	n.a.	delta C	(g)	20	no tag	-	membrane	-
HBV	Core	n.a.	tetra HBc	(g)	20	no tag	-	membrane	-
HBV	pol	n.a.	full-length	(h)	94	StrepII	C-terminus	membrane	+
DHBV	pol	n.a.	full-length	(h)	94	StrepII	C-terminus	membrane	+
Dengue virus	Capsid	type 2	full-length	(a)	12	no tag	-	membrane	+
Dengue virus	Capsid	type 2	full-length	(a)	12	StrepII	C-terminus	membrane	+
Dengue virus	Capsid	type 2	full-length + sp	(a)	13	StrepII	C-terminus	membrane	+
Dengue virus	Capsid	type 2	full-length + sp + pr8	(a)	14	StrepII	C-terminus	membrane	+

Table IV.1 (part 3/3)

Organism	Protein	Strain	Construct	Clonage	Molecular mass (kDa)	Tag	Tag position	Soluble or membrane protein?	Expression (precipitate mode)
Dengue virus	NS4A	n.a.	full-length d2K	(i)	16	no tag	-	membrane	+
Dengue virus	NS4A	n.a.	full-length d2K	(i)	16	His	C-terminus	membrane	+
Dengue virus	NS4A	n.a.	full-length +2K	(i)	16	His	C-terminus	membrane	-
GBV-B	$NS2^{pro}/NS3^{1-5}$	n.a.	95-208 / 1-5	(c)	14	Twin-StrepII	N-terminus	membrane	+
GBV-B	$NS2^{pro}/NS3^{pro}$	n.a.	95-208 / 1-190	(c)	33	Twin-StrepII	N-terminus	membrane	+
GBV-B	$NS2^{pro}/NS3^{pro}$	n.a.	95-208 / 1-190	(c)	33	Twin-StrepII	C-terminus	membrane	+
GBV-B	$NS2/NS3^{pro}$	n.a.	1-208 / 1-190	(c)	44	Twin-StrepII	C-terminus	membrane	+
GBV-B	$NS2/NS3^{pro}$	n.a.	1-208 (C909A) / 1-190	(c)	44	Twin-StrepII	C-terminus	membrane	+
GBV-B	$NS2^{pro}/NS3^{pro}$	n.a.	87-208 / 1-190	(c)	34	Twin-StrepII	C-terminus	membrane	+
GBV-B	$NS2^{pro}/NS3^{pro}$	n.a.	87-208 (C909A) / 1-190	(c)	34	Twin-StrepII	C-terminus	membrane	+
HCV receptors	CLDN1	-	full-length	(j)	23	StrepII	C-terminus	membrane	+
HCV receptors	SRB1	-	full-length	(j)	61	StrepII	C-terminus	membrane	±
HCV receptors	CD81	-	full-length	(j)	26	StrepII	C-terminus	membrane	-
other eukaryotic proteins	ABCG2	-	full-length	(k)	72	StrepII	C-terminus	membrane	+
other eukaryotic proteins	ANXA2	-	full-length	(f)	39	StrepII + TCS	N-terminus	soluble	+
other eukaryotic proteins	VKORC1	-	full-length	(l)	18	StrepII	C-terminus	membrane	-
other eukaryotic proteins	OCIA	-	full-length	(e)	28	StrepII	C-terminus	membrane	+
other eukaryotic proteins	TMEM147	-	full-length	(m)	25	StrepII	C-terminus	membrane	+
other eukaryotic proteins	TMEM147	-	full-length	(m)	25	His + GFP + TCS / StrepII	N-ter / C-ter	membrane	+
other eukaryotic proteins	TMEM147	-	full-length	(m)	25	His + Trx + TCS / StrepII	N-ter / C-ter	membrane	+
other eukaryotic proteins	NAP1L1	-	full-length	(n)	45	StrepII	C-terminus	soluble	+
other eukaryotic proteins	NAP1L4	-	full-length	(n)	43	StrepII	C-terminus	soluble	+
micellaneous	Chi (*Plasmodium*)	-	full-length	(o)	25	StrepII	C-terminus	soluble	+
micellaneous	SP6 RNA polymerase	-	full-length	(p)	96	no tag	-	soluble	+
micellaneous	BKI1 (*A. thaliana*)	-	full-length	(q)	37	StrepII	C-terminus	soluble	+

1.5. NS3-4A as an example to evidence activity of protein expressed using the WGE-CF system

Three NS3-4A precursors (HCV strain Con1, GenBank accession number AJ238799) provided by Daniel Lamarre (CRCHUM, University of Montreal) were expressed in the precipitate mode : a wild type construct and two mutants, one carrying a mutation within the NS3 protease ($NS3^{pro}$) catalytic site (S139A), the other one carrying a mutation within the cleavage site (T631I). The collaboration with Daniel Lamarre initially aimed to produce NS3-4A to analyze protein-inhibitor interactions using surface plasmon resonance (SPR) by Stéphane Sarrazin during his post-doc in the lab.

As described in section II.6.5., NS3 protease domain cleaves at the junction between NS3 and NS4A. Whereas a band corresponding to cleaved NS3 was observed on the Coomassie gel for the wild type NS3-4A precursor, a mutation within the $NS3^{pro}$ catalytic site (S139A) almost fully inhibited the cleavage, while a mutation within the NS3-4A cleavage site (T631I) inhibited it only partially (**Figure IV.6**). These results provided evidence that proteins expressed in the WGE-CF system can be active. In addition, activity of proteins produced in the WGE-CF system was also highlighted with NS2-$NS3^{pro}$ precursors, as detailed in section IV.3.2.

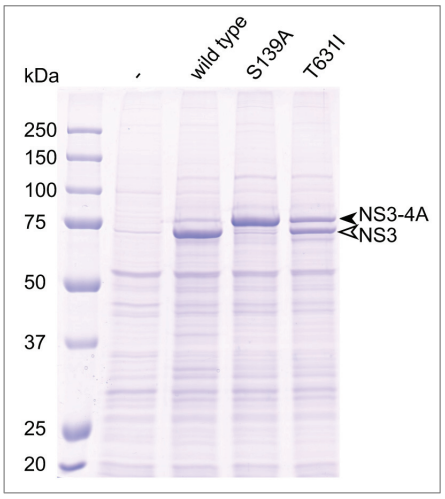


Figure IV.6 ***Expression of wild type and mutant NS3-4A precursors using the WGE-CF system in the precipitate mode.*** Pellets obtained after centrifugation of the CFS were analyzed by SDS-PAGE. -, negative control (no NS3-4A mRNA); wild type, wild type NS3-4A precursor; S139A, NS3-4A precursor with a mutation within the $NS3^{pro}$ catalytic site; T631I, NS3-4A precursor with a mutation within the NS3-4A cleavage site. A black arrowhead indicates the bands corresponding to NS3-4A precursor and an empty arrowhead indicates the bands corresponding to cleaved NS3. Note that NS4A is too small to be detected by SDS-PAGE in the used conditions (54 aa).

In the particular case of NS3-4A, solubilization of the protein after synthesis in the precipitate mode was possible using DDM. However, purification was delicate due to strong autoproteolysis. This aspect will not be further developped in this work.

1.6. Conclusion and perspectives

The initial phase of this thesis concerned the development and implementation of the wheat germ cell-free system in the lab. It was found that the bilayer method where mRNA is directly used for translation was best suited. Moreover, wheat germ extract preparation as well as main expression parameters were optimized, allowing the successful expression of more than 82% of the 75 protein constructs initially tested, including mainly membrane proteins. Since then, further constructs were expressed in this system with the same efficiency. Whereas the first 75 constructs had a molecular weight between 6 kDa and 96 kDa, we showed with the PI4KIIIα (230 kDa, collaboration with Volker Lohmann, Heidelberg) that even bigger proteins can be successfully expressed. In addition, Aurélie Badillo, manager of the recombinant protein unit at RD Biotech, successfully applies the techniques developed in the lab for clients from the industry or academic research. Up to now, Aurélie tested 10 proteins and could successfully express 9 of them. Three proteins were inflammatory factors whose activity has been proved in bioassays. Her work underlines the efficiency and utility of the WGE-CF system implemented and optimized during this work.

Recently, Matthias Harbers from CellFree Sciences kindly shared with us a protocol optimized by the company to produce proteins in the CECF mode (refer to section IV.1.3.1.). In the lab, Joanna Bons (trainee at RD Biotech) and Aurélie Badillo tested the expression of GFP using this protocol. They observed that the expression yield was about 3 times higher than using the bilayer method. However, larger volumes of feeding buffer are required, leading to a cost which is also 3 times higher. This cost would be even higher when working with isotopically labeled amino acids. From a financial point of view, this protocol is therefore not attractive. Nevertheless, this protocol has two main advantages : only 500 µL WGE are required for a total reaction mix volume of 3 mL and the volume used for further purification is lower (only the translation mix is loaded on the column instead of the total reaction volume with the bilayer method). This technique could therefore be interesting for the production of soluble proteins which do not require any isotopic labeling. However, when working with membrane proteins, an important optimization work would have to be done for the expression in the presence of detergent, as the scale-up to dialysis would produce more concentrated protein which would require ill-defined additional amounts of detergent.

2. Wheat germ cell-free expression in the presence of detergents

Previously, in section IV.1.4, we showed that expression of membrane proteins in the precipitate mode was mostly successfull with over 80% positive results. However, expression of membrane proteins in a precipitate form is not the preferred alternative. Indeed, although membrane proteins can be solubilized in a folded form using mild detergents, irreversible aggregation or solublization in a misfolded form could occur. While expression of membrane proteins in the presence of liposomes or nanodiscs is in principle very attractive, it requires high lipid-to-protein ratios which are not suitable for structural studies by solid-state NMR. For these reasons, we optimized the expression of membrane proteins in the presence of detergents with the aim to reconstitute purified proteins in lipids with low lipid-to-protein ratios for solid-state NMR. The challenge while using detergents for cell-free membrane protein expression is to find a compromise between good expression level and large extent of solubilization. Preliminary tests were performed with various detergents using GFP as model protein since fluorescence measurement allows a fast and easy read-out of protein expression. The results are presented in the first part of this section. The second part of this section describes the identification of two detergents allowing to produce HCV membrane proteins directly in a solubilized form without affecting their expression level is described. Finally, alternatives to commercial detergents such as calixarenes or amphipols were tested, as described in the third part of this section.

2.1. Preliminary tests

In order to identify compounds that are compatible with protein expression using the WGE-CF system, 18 different detergents were initially tested for GFP expression (**Table IV.2**).

All detergents were used at a standard concentration of 0.1% (w/v) which is above their CMC, except for DHPC, CHAPS, Cholate, OG and deoxycholate (**Table IV.2**). In these preliminary tests, detergents were added only to the feeding buffer (upper layer). The expression level was analyzed by measurement of GFP fluorescence intensity and results are presented in **Figure IV.7A**. Note that the potential effect of the addition of detergent stock solutions on the pH of the feeding buffer was verified using pH-paper. Indeed, pH modification could impact GFP fluorescence intensity without changes in

GFP expression level. At 0.1%, only LDAO slightly increased the pH of the feeding buffer. Detergents such as DPC, CTAB, LDAO, deoxycholate or FC-14 seemed to fully inhibit GFP expression. Others such as Tween-20 or Brij-35 for example inhibited GFP expression only partially. Finally, detergents such as CHAPS or OG had no significant effect on GFP expression level.

Detergent	Charge	MW (Da)	CMC (%)
DHPC	Zwitterionic (lipid)	453	0.679
DPC	Zwitterionic	351	0.032
CHAPS	Zwitterionic	615	0.492
Tween-20	Nonionic	1225	0.006
NP-40	Nonionic	617	0.0179
Triton X-100	Nonionic	647	0.0155
DM	Nonionic	483	0.087
DDM	Nonionic	511	0.009
Lysolecithin	Zwitterionic	n.a.	n.a.
Cholate	Anionic	431	0.388
OG	Nonionic	292	0.350
CTAB	Cationic	364	0.0364
Brij-35	Nonionic	1225	0.011
LDAO	Zwitterionic	229	0.023
C12E8	Nonionic	539	0.005
Deoxycholate	Anionic	414	0.248
FC-14	Zwitterionic	379	0.0046
Digitonin	Nonionic	1230	< 0.006

Table IV.2 ***List of detergents tested for GFP expression using the WGE-CF system.*** MW, molecular weight (Da); CMC, critical micelle concentration (%, w/v). Note that the CMC is dependent on the physico-chemical conditions. DHPC, 1,2-diheptanoyl-sn-glycero-3-phosphatidylcholine; DPC (FC-12), dodecyl-phosphocholine; CHAPS, 3-((3-cholamidopropyl)dimethylammonio)-1-propansulfonat; Tween-20, polyoxyethylene (20) sorbitan monolaurate; NP-40, nonyl-phenoxypolyethoxylethanol; Triton X-100, polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether; DM, n-decyl-β-D-maltoside; DDM, n-dodecyl-β-D-maltoside; Lysolecithin, L-α-lysophosphatidylcholine; OG, n-octyl-D-glucopyranoside; CTAB, hexadecyltrimethylammonium bromide; Brij-35, polyethylene glycol dodecyl ether; LDAO, lauryl dimethyl amide oxide; C12E8, dodecyl octaethylene glycol ether; Sodium deoxycholate, 3α,12α-dihydroxy-5β-cholan-24-oic acid sodium salt; FC-14, n-tetradecyl-phosphocholine; Digitonin, steroid glycoside.

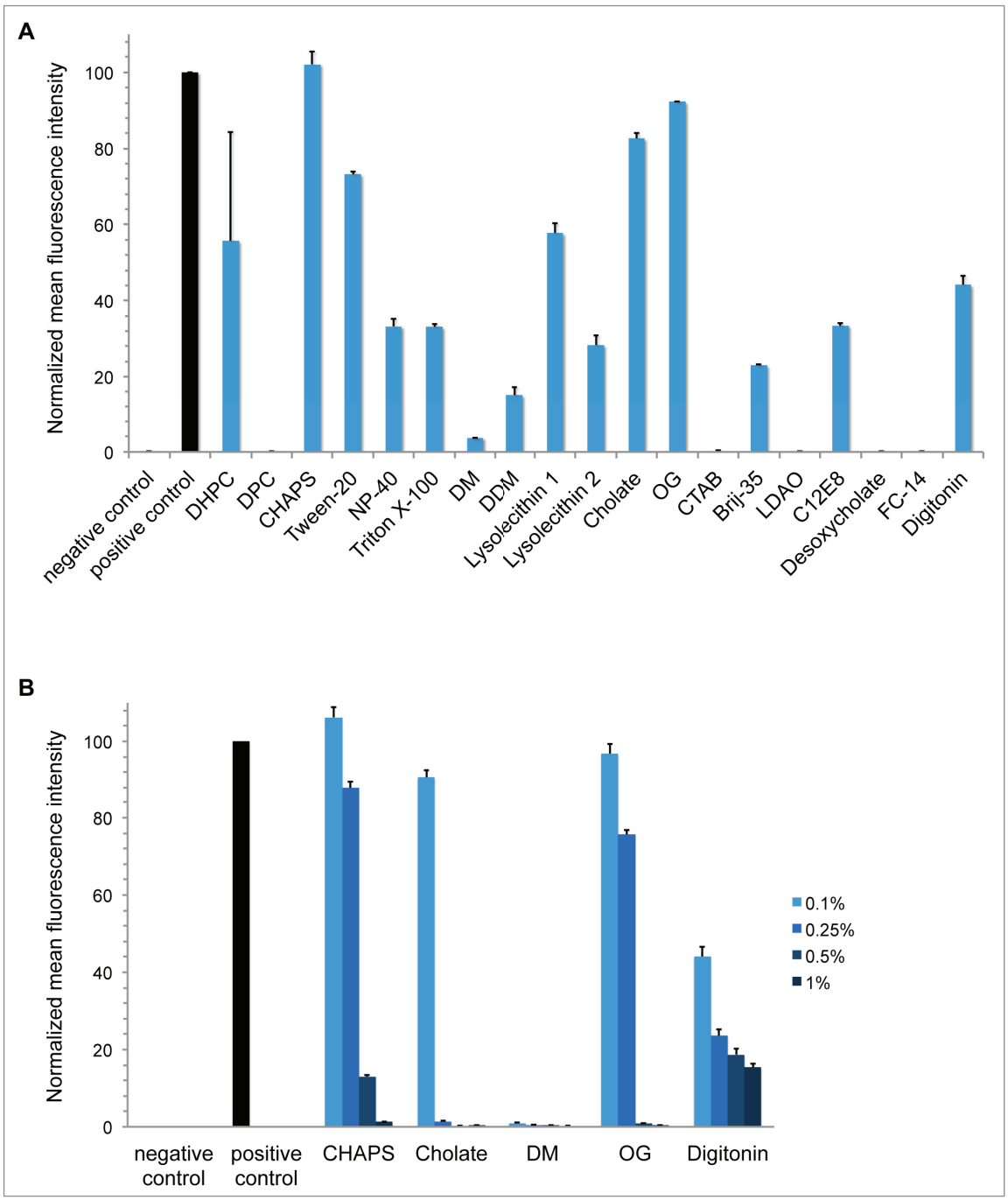


Figure IV.7 ***Expression of GFP in the presence of various detergents.*** (**A**) 18 detergents were tested at a standard concentration of 0.1% (w/v). Two different lot numbers were used for lysolecithin, giving different results. (**B**) Expression of GFP in the presence of different concentrations of CHAPS, cholate, DM or OG. Detergent concentrations are given in % (w/v). (**A**) and (**B**) Parameters for fluorescence measurements : excitation wavelength : 488 nm, emission wavelength : 509 nm. Negative control, no GFP mRNA; positive control (in *black*), GFP expression in absence of detergent. Values of fluorescence intensity were normalized compared to GFP expressed in the absence of detergent (positive control).

In a second time, CHAPS, cholate, DM, OG and digitonin were selected to be tested at four different concentrations : 0.1%, 0.25%, 0.5% and 1% (w/v). Whereas all tested

concentrations are above the CMC for DM and digitonin, the CMC of CHAPS, cholate and OG is between 0.25% and 0.5% (**Table IV.2**). Starting from 0.5%, all detergents slightly but not significantly increased the pH of the feeding buffer. The results presented in **Figure IV.7B** are reproducible compared to those obtained previously at a concentration of 0.1% (for DM inhibition was however slightly stronger than previously). CHAPS and OG started to partially inhibit GFP expression only at 0.25%, whereas cholate almost fully inhibited GFP expression at this concentration. Higher concentrations of digitonin led to stonger but not complete inhibition (85% inhibition at 1%).

As a potential interference of detergents with fluorescence measurements can not be excluded, the tests performed with the GFP are only indicative and only detergents compatible with the WGE-CF system are validated by this approach. In addition, the expression of full-length NS4B (with a His-tag fused at its C-terminus) was tested in the presence of CHAPS, cholate, lysolecithin, OG and digitonin at a standard concentration of 0.1% (w/v). Only digitonin allowed to partially express NS4B in a solubilized form but NS4B expression level seemed to be slightly reduced. This very preliminary test showed that detergent should be tested directly on the proteins of interest in order to determine if the detergent allows to efficiently solubilize them without affecting their expression level (refer to section IV.2.2. (Fogeron *et al*, 2015)).

2.2. MNG-3 and C12E8 detergents for the production of solubilized membrane proteins

Mild detergents with a low critical micelle concentration have been reported to allow the successful expression of directly solubilized membrane proteins in cell-free systems based on *E. coli* lysates (Klammt *et al*, 2005; Schwarz *et al*, 2007; Schwarz *et al*, 2008). However, the use of these detergents in a WGE-CF system has shown drawbacks, mainly interfering with expression yields, and only few information on detergent type and concentration for membrane protein expression in such system is avaible (Genji *et al*, 2010). Considering the preliminary results described in section IV.2.1. and based on fruitful discussion with Pierre Falson, we therefore screened 9 detergents (DPC, DM, DDM, MNG-3, C12E8, LDAO, CHAPS, cholate and OG) for the expression of full-length

NS2, using a standard concentration of 0.1%. Whereas several detergents showed a clear inhibitory effect on NS2 expression, two of them, MNG-3 and C12E8, allowed the expression of NS2 directly in a solubilized form without affecting its expression level.

For both MNG-3 and C12E8, the optimal concentration ranges were established for the expression of either full-length NS2 or $NS2^{pro}$ directly in a solubilized form. Since MNG-3 has a structure very similar to that of DDM, which a standard detergent used for reconstitution of proteoliposomes in the lab, we further focused on this detergent. We could establish that MNG-3 allows the expression of full-length NS4B and NS5A proteins directly in solubilized forms too and determined the optimal concentration range for the expression of these two proteins.

We observed that high detergent monomer concentrations rather than high micelle concentrations might be responsible for the inhibitory effect occuring when detergent with high CMC are used. Indeed, whereas DM and DDM had a strong negative effect on NS2 expression level at relatively low micelle concentration around 2 µM and 8 µM, C12E8 did not affect protein expression at micelle concentrations up to 49 µM. Moreover, a detergent micelle-to-protein ratio (mol/mol) around one order of magnitude could be taken into account for future detergent screens.

These results were recently published in Protein Expression and Purification (Fogeron *et al*, 2015). The publication is included in the following.

Protein Expression and Purification 105 (2015) 39–46

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins

Marie-Laure Fogeron [a], Aurélie Badillo [a,b], Vlastimil Jirasko [c,d], Jérôme Gouttenoire [e], David Paul [c], Loick Lancien [a], Darius Moradpour [e], Ralf Bartenschlager [c,d], Beat H. Meier [f], François Penin [a,*], Anja Böckmann [a,*]

[a] Institut de Biologie et Chimie des Protéines, Bases Moléculaires et Structurales des Systèmes Infectieux, Labex Ecofect, UMR 5086 CNRS, Université de Lyon, 7 passage du Vercors, 69367 Lyon, France
[b] RD-Biotech, Recombinant Protein Unit, 3 rue Henri Baigue, 25000 Besançon, France
[c] Department of Infectious Diseases, Molecular Virology, Heidelberg University, Im Neuenheimer Feld 345, 69120 Heidelberg, Germany
[d] German Centre for Infection Research (DZIF), Partner Site Heidelberg, Heidelberg, Germany
[e] Division of Gastroenterology and Hepatology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland
[f] Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 27 June 2014
and in revised form 26 September 2014
Available online 12 October 2014

Keywords:
Cell-free protein expression
Detergents
Hepatitis C virus proteins

ABSTRACT

Membrane proteins are notoriously difficult to express in a soluble form. Here, we use wheat germ cell-free expression in the presence of various detergents to produce the non-structural membrane proteins 2, 4B and 5A of the hepatitis C virus (HCV). We show that lauryl maltose neopentyl glycol (MNG-3) and dodecyl octaethylene glycol ether (C12E8) detergents can yield essentially soluble membrane proteins at detergent concentrations that do not inhibit the cell-free reaction. This finding can be explained by the low critical micelle concentration (CMC) of these detergents, which keeps the monomer concentrations low while at the same time providing the necessary excess of detergent concentration above CMC required for full target protein solubilization. We estimate that a tenfold excess of detergent micelles with respect to the protein concentration is sufficient for solubilization, a number that we propose as a guideline for detergent screening assays.

© 2014 Elsevier Inc. All rights reserved.

Introduction

Cell-free protein expression is an alternative approach to protein production when classical methods using cell-based systems fail (for a recent review see [1,2] and for a recent book [3,4]). As an open system, which allows addition of solubilizing agents, it is highly promising notably for the expression of membrane proteins. Furthermore, the wheat germ system is particularly favorable for the expression of eukaryotic proteins, as the slower ribosomal translation of wheat germ ribosomes better mimics eukaryotic systems than their prokaryotic counterpart. This often enables correct co-translational protein folding [5]. Three expression modes for membrane proteins have been developed in cell-free systems: as precipitate (in the absence of lipid/detergent molecules), solubilized by the presence of detergents, or directly inserted into model membranes [6,7]. While co-translational insertion of the protein into liposomes [8,9] is in principle very attractive, with proteoliposome recovery directly possible by simple centrifugation, this strategy is often difficult to follow because many proteins do not spontaneously insert into the liposome membrane and/or require specific lipid compositions for insertion [8,9]. In addition, this strategy requires high lipid-protein ratios, suffers from poor yields, and presents difficulties for purification. Therefore, protein expression in the presence of detergent [9–14] remains the most convenient approach, since proteins may be expressed in their folded, functional, and solubilized form, with high yield, and straightforward downstream purification. Expression in the precipitated form is the least preferred alternative because of often irreversible aggregation, although proteins can sometimes be resolubilized in a folded form using mild detergents [10,15].

* Corresponding authors.
E-mail addresses: f.penin@ibcp.fr (F. Penin), a.bockmann@ibcp.fr (A. Böckmann).

http://dx.doi.org/10.1016/j.pep.2014.10.003
1046-5928/© 2014 Elsevier Inc. All rights reserved.

When using cell-free expression in the presence of detergent, key parameters to be optimized are the expression yield and, more importantly, the extent of the membrane protein solubilization. While solubilization can be compromised by low detergent concentrations, the expression yields can be dramatically lowered by high detergent concentrations. Both effects depend on the detergent used. Concerning the detergent type, it is well known that charged detergents, such as sodium dodecyl sulfate (SDS) and cetyltrimethylammonium bromide (CTAB), are efficient for solubilization but generally inhibit protein activity to a large extent. In contrast, non-ionic and some zwitterionic detergents are considered as mild detergents, maintaining native protein structure and activity [16]. Best choices are thus non-ionic or some zwitterionic detergents, and furthermore those which allow complete solubilization before reaching concentrations affecting the translation machinery.

For membrane protein expression using bacterial cell extracts, previous results suggest that mild detergents with a low critical micelle concentration, such as polyoxyethylene-lauryl-ether (Brij) and polyoxyethylene sorbitan monolaurate (TWEEN) derivatives allow successful production and have proven to be optimal with respect to the solubilized protein yield [10,14,15,17–22]. However, their use in wheat germ cell-free expression has shown drawbacks, as evaluated by expression of green fluorescent protein (GFP) as a model to evaluate expression yield [23]. Additionally, for membrane protein expression in the wheat germ system, the available information on the best choice of detergent type and concentration for membrane protein expression is still sparse [9,12,23].

A new class of low CMC detergents are the maltose-neopentyl glycol (MNG) amphiphiles, which have been described recently in the literature [24]. MNG amphiphiles feature a central quaternary carbon derived from neopentyl glycol, two hydrophilic groups derived from maltose, as well as two *n*-decyl lipophilic tails. They were recently developed for solubilization, stabilization and crystallization of membrane proteins, and have shown promising results in the expression of bacteriorhodopsin in a wheat germ cell-free system [24].

We present here a detergent screen to establish which detergents are compatible with the wheat germ cell-free reaction. Subsequently, we use then the two most promising reagents, C12E8 and MNG-3, for a detailed analysis of the optimal concentration range for protein solubilization. We test this using the hepatitis C virus (HCV) nonstructural proteins 2, 4B and 5A (NS2, NS4B and NS5A), which are membrane proteins notoriously difficult to overexpress due to their toxicity in cell-based systems. The current knowledge about these HCV proteins has been reviewed in [25–27].

Results

Screening of detergents for compatibility with the cell-free reaction

Detergent screening in the cell-free expression reaction was carried out using the HCV NS2 integral membrane protein fused to a *Strep*-tag II at the N-terminus, in presence of the following detergents: DPC, DM, DDM, MNG-3, C12E8, LDAO, CHAPS, cholate and OG (listed in Table 1 with the full name). Each detergent was tested at a concentration of 0.05% (w/v) and 0.1% (w/v), i.e. above their CMC (given in Table 1), except for CHAPS, cholate and OG which have a CMC > 0.5% (w/v). For the latter detergents, an additional concentration above their CMC of 1% (w/v) was tested. Three different fractions were analysed by SDS–PAGE and western blotting using monoclonal antibody (mAb) against *Strep*-tag II: first, the total reaction mixture resulting from the cell-free reaction (cell-free sample, CFS, see Materials and Methods); second the pellet (P) resulting from centrifugation of this solution at 20,000*g* for 30 min; and third the supernatant, referred to as SN-beads, as the protein in the supernatant fraction was captured using affinity *Strep*-Tactin coated magnetic beads.

The results for the HCV NS2 expressed in the presence of the different detergents are shown in Fig. 1 and Table 1. When expressed in the absence of detergent, NS2 is mainly expressed as a precipitate and detected in the pellet. The presence of a small portion of NS2 in the soluble fraction, despite the absence of detergent, could be explained by the residual presence of lipids from the wheat germ extract (WGE) or some low molecular weight, non-native but soluble, oligomers. Fig. 1 shows that expression yield (as seen in the gels corresponding to the total cell-free sample, CFS) is only maintained at a comparable level to the reaction in the absence of detergent for MNG-3, C12E8, CHAPS, Cholate and OG at 0.1%. At 1%, which was tested for CHAPS, Cholate and OG in order to increase concentration above the CMC, all three fail to support protein expression. As can be seen in Table 1, at 0.05%, MNG-3, C12E8, and DM (which is then below CMC), as well as CHAPS, Cholate and OG (also below CMC) also yield comparable expression. DDM shows reduced expression under all conditions used, and DPC and LDAO fail in all cases.

In order to determine the amount of protein expressed in a soluble state, a comparison between the bands for the pellet (P) and the supernatant (SN-beads) (Fig. 1) has to be made. It can be seen

Table 1
Expression level of full-length NS2 in a cell-free system using wheat germ extract in presence of nine different detergents.

Detergent	Charge	MW	Aggregation	CMC		Detergent concentration 0.05%				Detergent concentration 0.1%			
		(Da)	Number	μM	%	Det. conc. (μM)	Micelle conc. (μM)	CFS	SN-Beads	Det. conc. (μM)	Micelle conc. (μM)	CFS	SN-Beads
Control								+++	+			+++	+
DPC	Zwitterionic	351	50–60	900	0.032	1425	9–10	–	–	2849	32–39	–	–
DM	Nonionic	483	110–140	1800	0.087	1035	–	+++	+	2070	1.9–2.5	+	–
DDM	Nonionic	511	98	170	0.009	978	8	++	+	1957	18	+	+
MNG-3	Nonionic	1005	?*	10	0.001	498	?*	+++	+++	995	?*	+++	+++
C12E8	Nonionic	539	90–120	90	0.005	928	7–9	+++	+++	1855	15–20	+++	+++
LDAO	Zwitterionic	229	74	1000	0.023	2183	16	–	–	4367	45	–	–
CHAPS	Zwitterionic	615	4–14	8000	0.492	813	–	+++	+	1626	–	+++	++
cholate	Anionic	431	4	9000	0.388	1160	–	+++	+	2320	–	+++	+
OG	Nonionic	292	900–1300	12000	0.350	1712	–	+++	+	3425	–	+++	+

DPC, dodecyl-phosphocholine; DM, n-decyl–D-maltoside; DDM, n-dodecyl–D-maltoside; MNG-3, lauryl maltose neopentyl glycol; C12E8, dodecyl octaethylene glycol ether; LDAO, lauryl dimethyl amide oxide; CHAPS, 3-((3-cholamidopropyl)dimethylammonio)-1-propansulfonat; cholate, 3, 7, 12-trihydroxy-5-cholan-24-oic acid, monosodium salt; OG, n-octyl–D-glucopyranoside. CFS, cell-free sample; SN-Beads, supernatant obtained after centrifugation of CFS and incubated with *Strep*-Tactin magnetic beads to capture *Strep*-tag II tagged proteins. Expression level is indicated as follow: +++, high; ++, medium; +, low; –, no expression. *No data is available for the aggregation number of MNG-3. Considering protein yields after purification by affinity chromatography of around 0.5 mg NS2 per milliliter of wheat germ extract, +++ represents at least 5 μg of protein in the cell-free sample containing 10 μl of wheat-germ extract.

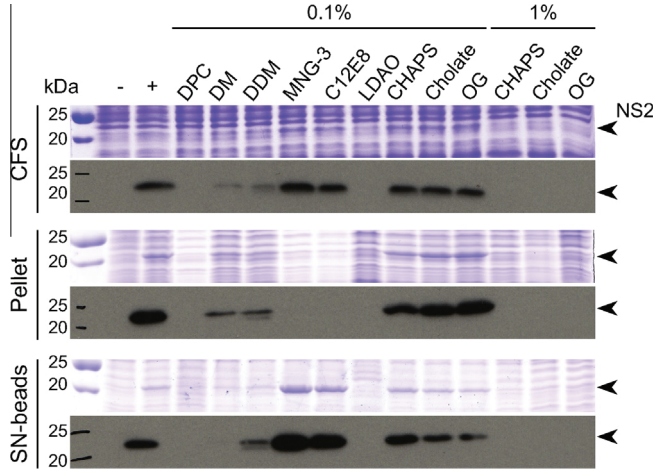


Fig. 1. Production of full-length NS2 using the wheat germ cell-free expression system in the presence of nine different detergents. Protein samples were analysed by SDS–PAGE followed by Coomassie blue staining (upper panels) and western blotting (lower panels). CFS, cell-free sample; pellet, pellet obtained after centrifugation of CFS; SN-beads, supernatant obtained after centrifugation of CFS and incubated with *Strep*-Tactin magnetic beads to capture *Strep*-tag II-tagged NS2 protein; –, negative control (no NS2); +, positive control (NS2 expressed in the absence of detergent). The black arrowheads indicate NS2. CFS – 3.6% of total sample (220 µl reaction mix); Pellet and SN beads – 27% of the total sample (220 µl reaction mix) loaded on the gel.

that only MNG-3 and C12E8 allow to obtain the protein in a fully soluble form, without any compromise in yield. CHAPS, Cholate and OG do not allow for better solubility than observed for the reference experiment with no detergent added (column denoted by +). This cannot be improved by adding more detergent, as above the CMC, at 1%, all three detergents inhibit the reaction completely. Table 1 indicates that neither DM nor DDM is efficient in solubilizing the protein at 0.05% or 0.1%. Thus, in order to express NS2 with high yield and in a soluble state, only the two nonionic, low CMC detergents MNG-3 and C12E8 were fully successful.

Expression of NS2 in presence of DDM, MNG-3 or C12E8

In order to determine the minimal amount of detergent required to produce soluble NS2, as well as to assess if higher concentrations affect yield, we tested a concentration range from 0 to 4500/5000 µM for DDM, MNG-3, and C12E8 (Fig. 2 and Table S1). Although DDM showed a partial inhibitory effect in the preliminary screen, it was tested for comparison, since it is often used as a standard detergent for solubilization and purification of membrane proteins and reconstitution in lipids, and also for its chemical likeness when compared to MNG-3 (Fig. 2D).

As shown in Fig. 2A (CFS, top), with increasing concentrations the presence of DDM progressively alters NS2 expression yields, and fully inhibits it around 3000 µM. Soluble protein is partially observed at all concentrations below 3000 µM, but inhibition of expression occurs before complete solubilization can be reached. In contrast, MNG-3 does not significantly inhibit NS2 expression (Fig. 2B, CFS) and starts to solubilize NS2 around a concentration of 200 µM (SN-beads), and allows to express NS2 in a completely soluble form at a concentration of 650 µM, where all the protein is found in the supernatant fraction (Fig. 2B). The apparent decrease of NS2 present in the supernatant above 3000 µM (Fig. 2B, bottom) is probably due to the fact that MNG-3 starts to inhibit the interaction of the *Strep*-tag II with the *Strep*-Tactin coated magnetic beads. Indeed, detergents can affect affinity tag purification, and while we do not observe this for C12E8 even when exceeding the reported limits for this detergent (90 µM, IBA technical info) by a factor of 50, we do observe an effect of MNG-3 when exceeding 2000 µM. Fig. 2C shows that C12E8 has no inhibiting effect on NS2 expression up to a detergent concentration of 4500 µM and allows NS2 expression in a soluble form at 1300 µM, with solubilization starting around 300 µM. The detergent concentration above CMC, as an indicator for the concentration of detergent monomers susceptible to form micelles, is given in Table S1. One can see that soluble protein is obtained around 600 µM detergent in micelle form, while for DDM, the reaction is inhibited before this concentration is reached. Unlike MNG-3, C12E8 does not interfere with affinity capture using *Strep*-Tactin coated magnetic beads.

Expression of NS2^{pro} in presence of MNG-3 or C12E8

Like the full-length NS2 protein considered above, its cytosolic protease domain denoted NS2^{pro} is expressed as a precipitate when protein synthesis occurs without addition of a solubilizing agent (Fig. 3). Indeed, this domain, for which the 3D structure is known [28], was recently shown to associate with the membrane surface [29]. The expression of NS2^{pro} (N-terminally fused to *Strep*-tag II) in the presence of MNG-3 or C12E8 is shown in Fig. 3. For each detergent, a concentration range from 0 to 2000 µM was tested. NS2^{pro} expression is not inhibited by MNG-3 or C12E8 up to the maximum concentration used. MNG-3 allows expression of NS2^{pro} in a fully soluble form around 2000 µM (Fig. 3A). C12E8 is slightly more efficient, with NS2^{pro} solubilization being complete at 1300 µM (Fig. 3B).

Expression of NS4B, NS5A and NS5A-ΔAH in presence of MNG-3

Having identified MNG-3 as an efficient solubilizing agent for the NS2 protein, we further investigated if it allows as well the expression of other HCV proteins in a soluble form. Results for the NS4B and NS5A membrane proteins are shown in Fig. 4, with tested concentration ranges from 0 to 1300 µM (Table S1 and Fig. 4). It can be seen that for NS4B, which is an integral membrane protein (reviewed in [30]), solubilization is complete at 650 µM (Fig. 4A), which is similar to the results observed for the full-length NS2 protein. In contrast to NS2 and NS4B, NS5A is a so-called monotopic membrane protein predicted to anchor in-plane to the membrane interface by its N-terminal amphipathic alpha-helix denoted "AH" [31]. As shown in Fig. 4B and C respectively, for both NS5A and NS5A-ΔAH, solubilization reaches a high level around 300–650 µM, but only shows slightly higher values at the maximum concentration of detergent used. However, the solubilization of NS5A and NS5A-ΔAH is not complete, even at 1300 µM. The residual insoluble forms of NS5A and NS5A-ΔAH are likely due to the tendency of this protein to associate in large assemblies [32], which however remain uncharacterized to date.

Analysis of NS2 and NS4B produced and purified in the presence of MNG-3 by exclusion chromatography

Soluble NS2 and NS4B expressed in presence of MNG-3 were purified by affinity chromatography using a *Strep*-Tactin column (Fig. 5A and B). While the purity level achieved for NS4B is quite high, some contaminants are observed in the NS2 sample. Better purity could be achieved by using a different detergent during purification, e.g., DDM (data not shown). Both proteins were analyzed by size exclusion chromatography (Fig. 5C and D). NS2 was eluted in a single peak at 3.2 ml (Fig. 5B), and no aggregate (which would elute at the void volume at 2.6 ml) was observed. The following three additional peaks come from the reagents present in the *Strep*-Tactin elution buffer. The elution peak corresponding to NS2 is rather asymmetric, which might be due to the oligomeric nature of NS2 established previously [28] and/or interactions with

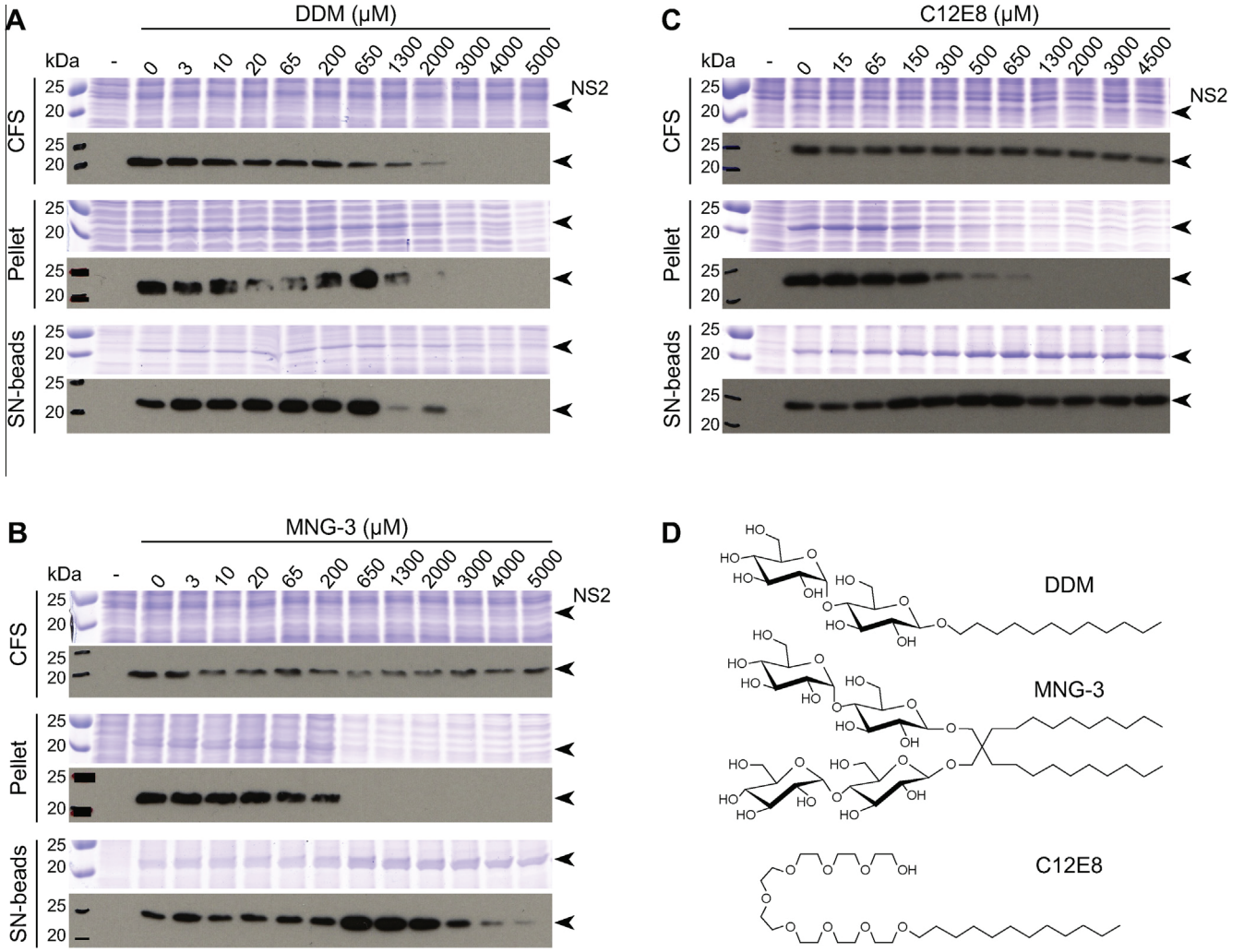


Fig. 2. Determination of the optimal DDM, MNG-3 and C12E8 detergent concentration for the expression and solubilization of the NS2 protein. NS2 was expressed in presence of the indicated concentrations of DDM (A), MNG-3 (B) or C12E8 (C). For conditions and labels, see Fig. 1. The black arrow indicates the band corresponding to NS2 protein. Detergent concentrations are given in μM. (D) Chemical structures of DDM, MNG-3 and C12E8.

the column chromatographic media. The same observations can be made for NS4B, which is eluted at 3.25 ml, and which is also expected to form oligomers [33]. For both proteins, superimposition of size exclusion profiles at 280 nm and 220 nm confirm that only the first elution peak corresponds to protein (Fig. S1). These results indicate that for both NS2 and NS4B, expression in MNG-3 yields monodisperse sample.

Discussion

We here tested different detergents for their dual ability to allow efficient expression of membrane proteins in a wheat-germ cell-free reaction and at the same time lead to the production of solubilized proteins. While several detergents tested clearly show an inhibitory effect on the expression, two of them, MNG-3 and C12E8, allowed us to obtain the protein in a soluble, monodispersed form without significant inhibition of the expression yield. With respect to protein yield in cell-free reactions in the presence of detergents, it is unclear mechanistically if it is micelles or detergent monomers that are detrimental for the correct functioning of the cell-free expression. However, we observe that the high detergent monomer concentrations, which occur when detergents with high CMCs are used, cause inhibition of the cell-free reaction. Indeed, for DM and DDM, a relatively low micelle concentration around 2 and 8 μM (0.05% and 0.1%, respectively) already severely inhibits the expression. In contrast, C12E8 concentrations do not inhibit the cell-free reaction up to the highest ones tested (4500 μM, corresponding to a micelle concentration of 37–49 μM). This supports the idea that high monomer concentrations rather than high micelle concentrations inhibit the system. This might explain why low-CMC detergents, which by definition show low concentrations of detergent monomers, are often more successfully used in cell-free expression.

The relevant variables with respect to the detergent concentration are the number of detergent monomers bound to the protein in its fully solubilized state, and the critical micelle concentration. The latter is often well-defined, and should be exceeded in order to yield detergent molecules with an affinity such that they interact with each other (which is the dominating factor [34]), and also with the protein [35]. The former parameter is more difficult to assess, as the number of detergent molecules needed to solubilize a protein depends on the hydrophobic sector of the membrane protein to be covered, as well as the cross-sectional area of the detergent and the hydrophobic chain length. Based on calculations, it has been suggested that the protein–detergent complexes might have physicochemical properties quite different from those of detergent micelles. Also, the degree of detergent binding per unit of hydrophobic surface area can vary for different membrane proteins [35]. Empirically, a ratio between 1.0 and 1.4 for the number

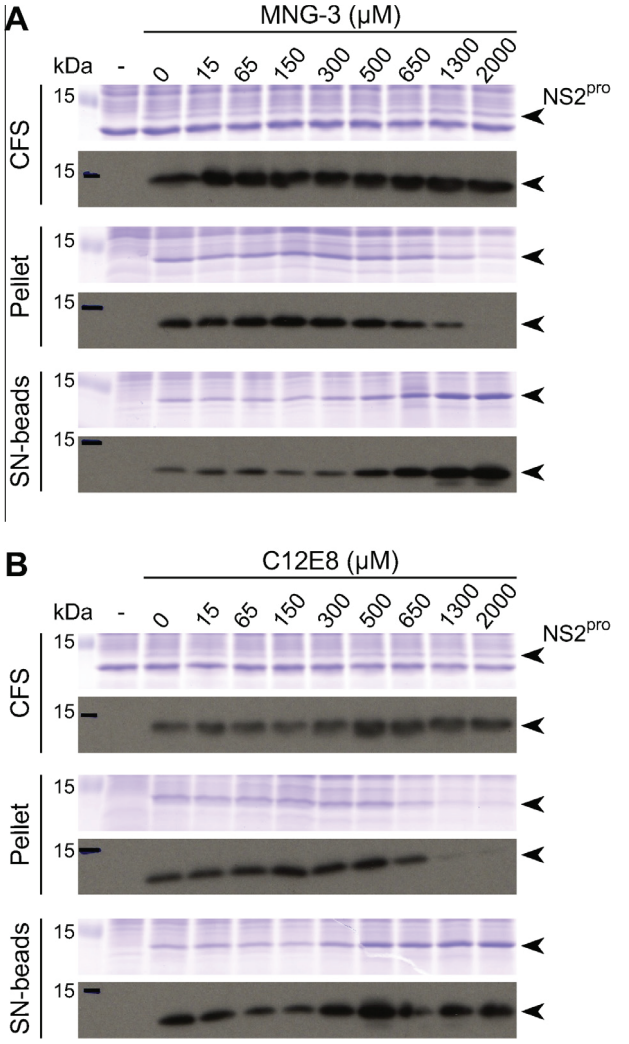


Fig. 3. Determination of the optimal concentration of MNG-3 and C12E8 detergents for the expression of $NS2^{pro}$. For conditions and labels, see Fig. 1. The black arrow indicates the band corresponding to $NS2^{pro}$. Detergent concentrations are given in μM.

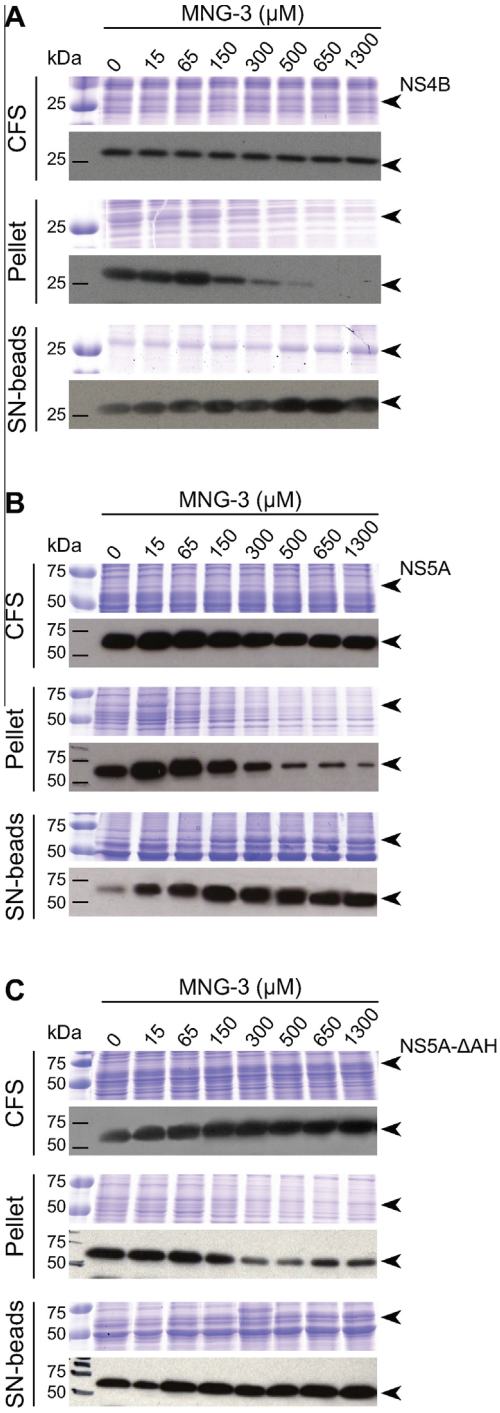


Fig. 4. Determination of the optimal concentration of MNG-3 for the expression of the NS4B, NS5A and NS5A-ΔAH proteins. For conditions and labels, see Fig. 1. The black arrow indicates the band corresponding to the protein of interest. Detergent concentrations are given in μM.

of detergent molecules in the protein–detergent complex versus the free detergent micelle has been observed for Aac3p solubilized in different detergents [34]. This indicates that the increase in the number of detergent molecules in the protein–detergent complex compared to the free detergent micelle might represent only a small factor [34], and that the micelle concentration can serve as a good approximation.

The concentration of micelles at which solubilization occurs can be estimated for C12E8 to 10–13 μM (Table S1) for NS2 and $NS2^{pro}$, for about 1 μM of protein produced in the reaction mixture. Therefore, under these conditions, about ten detergent micelles per protein molecule produced are available, which is notably comparable to ratios calculated from data revealing the detergent concentration for complete solubilization in the literature. Indeed, three GPCRs [18] as well as an olfactory receptor [36] were successfully solubilized in cell-free reactions mixture using detergent-micelle-to-protein ratios between 5 and 8. Thus, in general the ratio seems to be around one order of magnitude, which can be taken into account for future studies to more efficiently screen the range of tested detergent concentrations in cell-free reactions. The observed excess of detergent required for full protein solubilization could be explained by either (i) the detergent interacting as well with other components of the wheat germ extract (the total wheat germ endogenous protein concentration in the reaction is about 2 mg/ml), (ii) the need of more availability of micelles at the exit of the ribosome, or by (iii) the need of more detergent monomers in the mixed protein–detergent micelle compared to the free detergent micelle. Also, one should keep in mind that both detergent CMC and aggregation numbers are dependent on the physico-chemical conditions, and are thus difficult to determine precisely.

For MNG-3, the micelle size is strongly concentration-dependent, and no aggregation number has been determined so far (P.

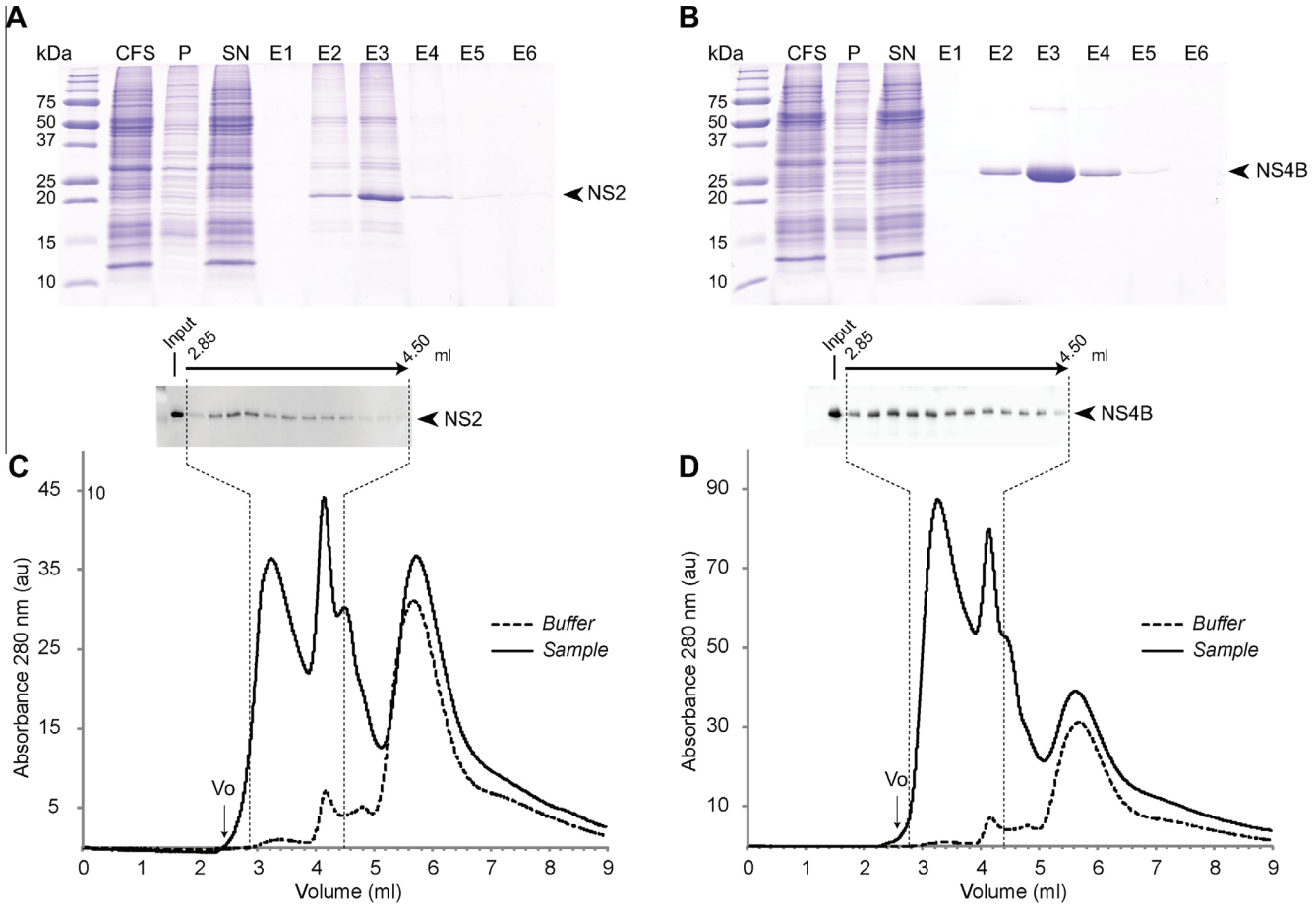


Fig. 5. Purification of NS2 and NS4B by affinity chromatography, and analysis by size exclusion chromatography. NS2 (A) and NS4B (B) purified by affinity chromatography were analyzed by SDS–PAGE followed by Coomassie blue staining. CFS, cell-free sample – 0.02% of total sample (36 ml reaction mix) loaded on the gel; P, pellet obtained after centrifugation of CFS – 0.2% of total sample loaded on the gel; SN, supernatant obtained after centrifugation of CFS – 0.02% of total sample loaded on the gel; E1 to E6, elution fractions – 1.6% of total fraction loaded on the gel. The black arrow indicates the band corresponding to the protein of interest. Size exclusion chromatography profiles of NS2 (C) and NS4B (D). The profile of the protein sample of interest is drawn in continuous line and the one of the affinity elution buffer in dotted line. Western blotting analyses of the collected fractions are shown above the chromatographic profiles.

S. Chae, personal communication). The results for NS2 show however that soluble protein is obtained at about the same detergent concentration as for C12E8. MNG-3 has been used in expression of all proteins tested here. It can be seen in Figs. 3A and 4 that, even if the order of magnitude remains comparable, not all proteins become soluble at the same concentration threshold. This means that there is probably a dependence on the nature of the protein used, and we describe in the following how the structural features of the different proteins could account for this observation.

$NS2^{pro}$ is the globular domain of the NS2 protein, which is anchored to the membrane mainly by the N-terminal integral membrane domain [37]. The crystal structure of $NS2^{pro}$ has been resolved in the presence of detergents and reveals its ability to dimerize [28]. It exhibits two helices with a hydrophobic surface that has been shown to interact with membranes [29]. This rationalizes why $NS2^{pro}$ is expressed as a precipitate in the cell-free reaction in the absence of detergents. With $NS2^{pro}$ expressed in somewhat smaller amounts than the full-length protein, protein-specific factors must be the cause for the higher micelle concentrations required for solubilization. It is possible for example that the absence of a membrane bilayer during the translation of $NS2^{pro}$ can lower the efficiency of the self dimerization process and prolong the time during which hydrophobic interaction interfaces are exposed which demand higher amount of the detergent to keep the protein in the soluble state. In case of the NS2 full-length protein, the N-terminal domain could contribute to the interaction between monomers and facilitate native protein folding.

The integral membrane protein NS4B (reviewed in [30]) expresses very well in the presence of MNG-3, and is solubilized at approximately the same concentration of detergent as NS2. Of note, these two viral proteins both form oligomers and contain several transmembrane passages responsible for their integral membrane character [25–27].

NS5A is a globular protein displaying an α-helical membrane anchor at its N-terminal end, and a large disordered portion, including the C-terminal D2 and D3 domains (reviewed in [25]), which might also interact with detergent. NS5A is also known to bind to nucleic acids in its hypothesized role of transporting the HCV viral RNA genome to the core protein for encapsidation. Both NS5A and NS5A-ΔAH are progressively solubilized by the addition of more MNG-3 to the reaction; however, at around 500–650 μM of MNG-3, the pellet fraction seems to remain stable, and no further solubilization is observed. One hypothesis explaining the fact that part of it is invariably found in the pellet, independent of the detergent concentration, might be that the treatment with the endonuclease benzonase is not efficient in accessing the entire amount of nucleic acids present in a particular form of dense NS5A/RNA/DNA assemblies which pellet during centrifugation.

Conclusion

Even if the general use of C12E8 and MNG-3 detergents for a successful expression and solubilization of membrane proteins in the wheat germ cell-free system remains to be established, our results present a further step in devising strategies for the expression of membrane proteins in this cell-free system. Proteins expressed in their detergent-solubilized form can subsequently be used for studies by NMR in solution state and X-ray crystallography, as well as, after reconstitution into membranes, by solid-state NMR and electron microscopy.

Materials & methods

Plasmids

cDNA of NS2 and NS2^{pro} (HCV strain JFH1; GenBank accession number AB047639), NS4B, NS5A and NS5A-ΔAH (HCV strain Con1; GenBank accession number AJ238799) were PCR amplified and cloned into pEU-E01-MCS vector (CellFree Sciences, Japan). For full-length NS2 construct, the amino acid (aa) sequence Met-Ala-Ser precedes a *Strep*-tag II [38] followed by a thrombin cleavage site downstream of a Ser-Gly linker, yielding to the final aa sequence MASWSHPQFEKTGLVPRGSG fused at the N terminus of full-length NS2, For the NS2^{pro} construct, no thrombine cleavage site has been inserted yielding to the aa sequence MASWSHPQFEKSG. For NS4B, NS5A and NS5A-ΔAH constructs, the *Strep*-tag II was fused at the C-terminal end of the protein. For NS4B, a thrombin cleavage site followed by a Ser-Ala linker was inserted upstream of the *Strep*-tag II, yielding to the C-terminal aa sequence LVPRGSAWSHPQFEK. For NS5A and NS5A-ΔAH, no thrombin cleavage site has been inserted yielding to the aa sequence SAWSHPQFEK. The resulting plasmids were transformed into *Escherichia coli* TOP10 chemically competent cells (Life Technologies) and DNA was prepared using a NucleoBond Xtra Maxi kit (Macherey-Nagel, France). Plasmids were further purified by a phenol/chloroform extraction according to CellFree Sciences (Yokohama, Japan) recommendations.

Wheat germ cell-free protein expression

Home-made wheat germ extracts were prepared from non-treated durum wheat seeds (Sud Céréales, France) according to Takai et al. [39]. Transcription and translation were carried out separately, as described in [25–27,39] and in Noirot et al. [40]. Transcription was performed in Eppendorf tubes using 100 μg/ml plasmid, 2.5 mM NTP mix (Promega, France), 1 U/μl SP6 RNA Polymerase (CellFree Sciences, Japan) and 1 U/μl RNase Inhibitor (CellFree Sciences, Japan) in transcription buffer (CellFree Sciences, Japan). After incubation for 6 h at 37 °C, mRNA was used directly for translation. Translation was performed in 96-well plates using the so-called bilayer method. The bottom layer, called reaction mix, contains 10 μl mRNA, 10 μl wheat germ extract and 40 μg/ml creatine kinase. The upper layer corresponds to the feeding buffer (200 μl per well) that contains all substrates necessary for protein synthesis: 30 mM Hepes-KOH pH7.6, 100 mM potassium acetate, 2.7 mM magnesium acetate, 16 mM creatine phosphate, 0.4 mM spermidine, 1.2 mM ATP, 0.25 mM GTP, 4 mM DTT and 6 mM amino acids mix. When expression was performed in presence of detergent, the detergent was added in both the reaction mix and the feeding buffer. The reaction sample was incubated at 22 °C for 16 h.

Sample preparation

After protein synthesis, 150 μl of cell free sample (CFS) were incubated with 250 units of benzonase at room temperature (RT) on a rolling wheel for 30 min. CFS was then centrifuged at 20,000*g*, 4 °C for 30 min. The obtained supernatant was incubated with 5 μl *Strep*-Tactin magnetic beads (MagStrep "type2HC": IBA Lifesciences, Germany) at 4 °C on the wheel for at least 30 min. After three washing steps with 10 volumes of a buffer containing 100 mM Tris–HCl pH8, 150 mM NaCl and 1 mM EDTA, the magnetic beads were directly resuspended in 25 μl 1× SDS–PAGE loading buffer (62.5 mM Tris–HCl pH6.8, 2% SDS (w/v), 10% glycerol (v/v), 5% β-mercaptoethanol (v/v) and 0.01% bromophenol blue (w/v)). Theses samples were further designated as SN-beads. Note that this step aimed to enrich the sample with the tagged protein and to assess whether the used detergent would interfere with the protein affinity capture. The pellet obtained after centrifugation of the CFS was resuspended in 25 μl 1× SDS–PAGE loading buffer. In addition, 20 μl CFS were mixed with 5 μl 5X loading buffer for SDS–PAGE analysis.

SDS–PAGE and Western blotting analysis

All expression experiments were assessed using 15% Coomassie blue stained SDS–PAGE gels. Samples are resuspended in a loading buffer containing 62.5 mM Tris–HCl pH6.8, 2% SDS (w/v), 10% glycerol (v/v), 5% β-mercaptoethanol (v/v) and 0.01% bromophenol blue (w/v) and incubated at RT for 15 min before loading. For each sample, 10 μl were loaded on the SDS–PAGE gels. Western-blotting analysis was carried out by protein transfer onto a nitrocellulose membrane using an iBlot® gel transfer device (Life Technologies). The nitrocellulose membrane was blocked with 5% non-fat milk powder in PBS-T buffer (12 mM sodium phosphate pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.05% Tween® 20) and, in the case of the pellet and SN-beads samples, incubated with the StrepMAB-Classic, HRP conjugate antibody (IBA Lifesciences, Germany) for 1 h at RT. For a more sensitive detection of the protein of interest in the CFS, the nitrocellulose membrane is incubated with the StrepMAB-Classic primary antibody (IBA Lifesciences, Germany) for 1 h at RT and further incubated with an anti-mouse IgG HRP-conjugated secondary antibody (Promega, France) for 1 h at RT. In both cases, epitope-containing bands are visualized using the ECL Prime Western Blotting Detection Reagent (GE Healthcare, France).

Purification of NS2 and NS4B by affinity chromatography

Both NS2 and NS4B were expressed in presence of 1300 μM (0.065%) MNG-3 using the bilayer method as described above. Protein synthesis was upscaled in a 6-well plate (500 μl reaction mix and 5.5 ml feeding buffer per well). After protein synthesis, the cell free sample (CFS) was incubated with benzonase at room temperature (RT) on a rolling wheel for 30 min. MNG-3 concentration was adjusted to 0.1%. CFS was then centrifuged at 20,000*g*, 4 °C for 30 min. The obtained supernatant was loaded on a Strep-Tactin gravity column (IBA Lifesciences, Germany). Purification was performed as described by the manufacturer, all buffers containing 0.1% MNG-3. Proteins of interest were eluted in 100 mM Tris–HCl pH8, 150 mM NaCl, 1 mM EDTA, 0.1% MNG-3 and 2.5 mM desthiobiotin.

Size exclusion chromatography

For both NS2 and NS4B, 200 μl fractions of protein purified by affinity chromatography were loaded on a KW402.5-4F column (Shodex, USA) using the NGC Chromatography System (Bio-Rad, USA). The column was equilibrated with 50 mM phosphate buffer pH 7.4 containing 0.1% MNG-3 and a flow rate of 150 μl/min was applied. 200 μl affinity elution buffer were also loaded on the column as control.

Acknowledgments

This work was supported in part by a grant from the French ANRS (France Recherche, Nord & Sud, Sida-HIV et Hépatites), an autonomous agency at INSERM, France, and the Swiss National Science Foundation (grant 31003A-138484 to DM). R.B. was supported by the Deutsche Forschungsgemeinschaft (TRR 83, TP13). We acknowledge the platforms Centre Commun de Microanalyse des Protéines and Production et Analyse de Protéines from the UMS 3444 BioSciences Gerland-Lyon Sud. We wish to express our thanks to Marc Le Maire for critical reading of the manuscript and to Pierre Falson for fruitful discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.pep.2014.10.003.

References

[1] F. Bernhard, Y. Tozawa, Cell-free expression – making a mark, Curr. Opin. Struct. Biol. 23 (2013) 374–380, http://dx.doi.org/10.1016/j.sbi.2013.03.012.
[2] G. Rosenblum, B.S. Cooperman, Engine out of the chassis: cell-free protein synthesis and its uses, FEBS Lett. 588 (2014) 261–268, http://dx.doi.org/10.1016/j.febslet.2013.10.016.
[3] T. Terada, T. Murata, M. Shirouzu, S. Yokoyama, Cell-free expression of protein complexes for structural biology, Methods Mol. Biol. 1091 (2014) 151–159, http://dx.doi.org/10.1007/978-1-62703-691-7_10.
[4] S.-I. Makino, E.T. Beebe, J.L. Markley, B.G. Fox, Cell-free protein synthesis for functional and structural studies, Methods Mol. Biol. 1091 (2014) 161–178, http://dx.doi.org/10.1007/978-1-62703-691-7_11.
[5] Y. Endo, T. Sawasaki, Cell-free expression systems for eukaryotic protein production, Curr. Opin. Biotechnol. 17 (2006) 373–380, http://dx.doi.org/10.1016/j.copbio.2006.06.009.
[6] J.R. Swartz (Ed.), Cell-Free Protein Expression, Springer, Berlin, 2013.
[7] Y. Endo, K. Takai, T. Ueda (Eds.), Cell-free Expression Systems: Methods and Protocols, Humana Press, 2009.
[8] A.R. Long, C.C. O'Brien, N.N. Alder, The cell-free integration of a polytopic mitochondrial membrane protein into liposomes occurs cotranslationally and in a lipid-dependent manner, PLoS ONE 7 (2012) e46332, http://dx.doi.org/10.1371/journal.pone.0046332.g005.
[9] A. Nozawa, T. Ogasawara, S. Matsunaga, T. Iwasaki, T. Sawasaki, Y. Endo, Production and partial purification of membrane proteins using a liposome-supplemented wheat cell-free translation system, BMC Biotechnol. 11 (2011) 35, http://dx.doi.org/10.1186/1472-6750-11-35.
[10] C. Klammt, D. Schwarz, K. Fendler, W. Haase, V. Dötsch, F. Bernhard, Evaluation of detergents for the soluble expression of α-helical and β-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system, FEBS J. 272 (2005) 6024–6038, http://dx.doi.org/10.1111/j.1742-4658.2005.05002.x.
[11] C. Klammt, I. Maslennikov, M. Bayrhuber, C. Eichmann, N. Vajpai, E.J.C. Chiu, et al., Facile backbone structure determination of human membrane proteins by NMR spectroscopy, Nat. Methods 9 (2012) 834–839, http://dx.doi.org/10.1038/nmeth.2033.
[12] C. Carraher, A.R. Nazmi, R.D. Newcomb, A. Kralicek, Protein expression and purification, Protein Expr. Purif. 90 (2013) 160–169, http://dx.doi.org/10.1016/j.pep.2013.06.002.
[13] B.L. Cook, D. Steuerwald, L. Kaiser, J. Graveland-Bikker, M. Vanberghem, A.P. Berke, et al., Large-scale production and study of a synthetic G protein-coupled receptor: human olfactory receptor 17-4, PNAS 106 (2009) 11925–11930, http://dx.doi.org/10.1073/pnas.0811089106.
[14] C. Berrier, K.-H. Park, S. Abes, A. Bibonne, J.-M. Betton, A. Ghazi, Cell-free synthesis of a functional ion channel in the absence of a membrane and in the presence of detergent, Biochemistry 43 (2004) 12585–12591, http://dx.doi.org/10.1021/bi049049y.
[15] D. Schwarz, F. Junge, F. Durst, N. Frölich, B. Schneider, S. Reckel, et al., Preparative scale expression of membrane proteins in *Escherichia coli*-based continuous exchange cell-free systems, Nat. Protoc. 2 (2007) 2945–2957, http://dx.doi.org/10.1038/nprot.2007.426.
[16] A.M. Seddon, P. Curnow, P.J. Booth, Membrane proteins, lipids and detergents: not just a soap opera, Biochim. Biophys. Acta 1666 (2004) 105–117, http://dx.doi.org/10.1016/j.bbamem.2004.04.011.
[17] D. Schwarz, V. Dötsch, F. Bernhard, Production of membrane proteins using cell-free expression systems, Proteomics 8 (2008) 3933–3946, http://dx.doi.org/10.1002/pmic.200800171.
[18] G. Ishihara, M. Goto, M. Saeki, K. Ito, T. Hori, T. Kigawa, et al., Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors, Protein Expr. Purif. 41 (2005) 27–37, http://dx.doi.org/10.1016/j.pep.2005.01.013.
[19] Y. Elbaz, S. Steiner-Mordoch, T. Danieli, S. Schuldiner, In vitro synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 1519–1524, http://dx.doi.org/10.1073/pnas.0306533101.
[20] I. Blesneac, S. Ravaud, C. Juillan-Binard, L.-A. Barret, M. Zoonens, A. Polidori, et al., Production of UCP1 a membrane protein from the inner mitochondrial membrane using the cell free expression system in the presence of a fluorinated surfactant, Biochim. Biophys. Acta 2012 (1818) 798–805, http://dx.doi.org/10.1016/j.bbamem.2011.12.016.
[21] A. Deniaud, L. Liguori, I. Blesneac, J.L. Lenormand, E. Pebay-Peyroula, Crystallization of the membrane protein hVDAC1 produced in cell-free system, Biochim. Biophys. Acta 1798 (2010) 1540–1546, http://dx.doi.org/10.1016/j.bbamem.2010.04.010.
[22] M. Miot, J.-M. Betton, Reconstitution of the Cpx signaling system from cell-free synthesized proteins, N. Biotechnol. 28 (2011) 277–281, http://dx.doi.org/10.1016/j.nbt.2010.06.012.
[23] T. Genji, A. Nozawa, Y. Tozawa, Biochemical and biophysical research communications, Biochem. Biophys. Res. Commun. 400 (2010) 638–642, http://dx.doi.org/10.1016/j.bbrc.2010.08.119.
[24] P.S. Chae, S.G.F. Rasmussen, R.R. Rana, K. Gotfryd, R. Chandra, M.A. Goren, et al., Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins, Nat. Methods 7 (2010) 1003–1008, http://dx.doi.org/10.1038/nmeth.1526.
[25] D. Moradpour, F. Penin, Hepatitis C virus proteins: from structure to function, Curr. Top. Microbiol. Immunol. 5 (2013) 113–142, http://dx.doi.org/10.1007/978-3-642-27340-7_5.
[26] D. Moradpour, F. Penin, C.M. Rice, Replication of hepatitis C virus, Nat. Rev. Microbiol. 5 (2007) 453–463, http://dx.doi.org/10.1038/nrmicro1645.
[27] V. Lohmann, Hepatitis C virus RNA replication, Curr. Top. Microbiol. Immunol. 369 (2013) 167–198, http://dx.doi.org/10.1007/978-3-642-27340-7_7.
[28] I.C. Lorenz, J. Marcotrigiano, T.G. Dentzer, C.M. Rice, Structure of the catalytic domain of the hepatitis C virus NS2-3 protease, Nature 442 (2006) 831–835, http://dx.doi.org/10.1038/nature04975.
[29] C.M. Lange, P. Bellecave, V.L. Dao Thi, H.T. Tran, F. Penin, D. Moradpour, et al., Determinants for membrane association of the hepatitis C virus NS2 protease domain, J. Virol. 88 (2014) 6519–6523, http://dx.doi.org/10.1128/JVI.00224-64814.
[30] J. Gouttenoire, F. Penin, D. Moradpour, Hepatitis C virus nonstructural protein 4B: a journey into unexplored territory, Rev. Med. Virol. 20 (2010) 117–129, http://dx.doi.org/10.1002/rmv.640.
[31] F. Penin, V. Brass, N. Appel, S. Ramboarina, R. Montserret, D. Ficheux, et al., Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A, J. Biol. Chem. 279 (2004) 40835–40843, http://dx.doi.org/10.1074/jbc.M404761200.
[32] R.A. Love, O. Brodsky, M.J. Hickey, P.A. Wells, C.N. Cronin, Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus, J. Virol. 83 (2009) 4395–4403, http://dx.doi.org/10.1128/JVI.02352-08.
[33] G.-Y. Yu, K.-J. Lee, L. Gao, M.M.C. Lai, Palmitoylation and polymerization of hepatitis C virus NS4B protein, J. Virol. 80 (2006) 6013–6023, http://dx.doi.org/10.1128/JVI.00053-06.
[34] E.R.S. Kunji, M. Harding, P.J.G. Butler, P. Akamine, Determination of the molecular mass and dimensions of membrane proteins by size exclusion chromatography, Methods 46 (2008) 62–72, http://dx.doi.org/10.1016/j.ymeth.2008.10.020.
[35] M. le Maire, P. Champeil, J.V. Moller, Interaction of membrane proteins and lipids with solubilizing detergents, Biochim. Biophys. Acta 1508 (2000) 86–111.
[36] L. Kaiser, J. Graveland-Bikker, D. Steuerwald, M. Vanberghem, K. Herlihy, S. Zhang, Efficient cell-free production of olfactory receptors: detergent optimization, structure, and ligand binding analyses, PNAS 105 (2008) 15726–15731, http://dx.doi.org/10.1073/pnas.0804766105.
[37] V. Jirasko, R. Montserret, J.Y. Lee, J. Gouttenoire, D. Moradpour, F. Penin, et al., Structural and functional studies of nonstructural protein 2 of the hepatitis C virus reveal its key role as organizer of virion assembly, PLoS Pathog. 6 (2010) e1001233, http://dx.doi.org/10.1371/journal.ppat.1001233.
[38] T.G.M. Schmidt, A. Skerra, The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins, Nat. Protoc. 2 (2007) 1528–1535, http://dx.doi.org/10.1038/nprot.2007.209.
[39] K. Takai, T. Sawasaki, Y. Endo, Practical cell-free protein synthesis system using purified wheat embryos, Nat. Protoc. 5 (2010) 227–238, http://dx.doi.org/10.1038/nprot.2009.207.
[40] C. Noirot, B. Habenstein, L. Bousset, R. Melki, B.H. Meier, Y. Endo, et al., Wheat-germ cell-free production of prion proteins for solid-state NMR structural studies, N. Biotechnol. 28 (2011) 232–238, http://dx.doi.org/10.1016/j.nbt.2010.06.016.

2.3. Alternatives to commercial detergents

Two alternatives to commercial detergents were tested during this work : calixarenes and the amphipol A8-35.

2.3.1. Expression of full-length NS2 in the presence of calixarenes

Calixarenes are anionic based detergents (C4Cn, n=1-12) which have been designed to structure the membrane domains through hydrophobic interactions and a network of salt bridges with the basic residues found at the cytosol-membrane interface of membrane proteins (Matar-Merheb *et al*, 2011) (**Figure IV.8B**).

Expression of full-length NS2 (HCV strain JFH-1) was tested in the presence of four different calixarenes : C4C3, C4C7, C4C10 and C4C12. Calixarenes were kindly provided by Pierre Falson (IBCP, Lyon). Whereas C4C3 and C4C7 were tested at concentrations of 0.01%, 0.1% and 1% (w/v), C4C10 and C4C12 were tested at concentrations of 0.001%, 0.01% and 0.1% (w/v), depending on their CMC. Moreover, expression of full-length NS2 was tested in parallel in the presence of DDM or MNG-3, at the same concentrations than C4C10 and C4C12. Unfortunately, analysis of CFS by western blotting using an antibody against the *Strep*-tag II gave poor signal quality and thus did not allow us to conclude directly on the potential effect of calixarenes on full-length NS2 expression level. However, analysis of the pellets and the supernatants obtained after centrifugation of CFS showed that 0.01% C4C3 allowed to slightly increase the solubilization level of NS2 (data not shown). At higher concentration, expression level seemed to be decreased. In the same way, C4C7 seemed to strongly inhibit protein expression already at 0.01%. Concerning C4C10 or C4C12, as shown in **Figure IV.8A**, a concentration of 0.001% allowed the production of NS2 in a partially solubilized form. Expression level seemed to decrease significantly at higher concentrations. In fact, C4C10 and C4C12 seemed to give similar results than DDM, however with a stronger inhibition effect on protein expression level. In contrast, MNG-3 allowed to better solubilize NS2 without affecting its expression level. These results are very preliminary and although calixarenes could have been promising, their use in the WGE-CF system was not followed up. We indeed decided to focus on the commercial detergent MNG-3 which gave much better results in this test.

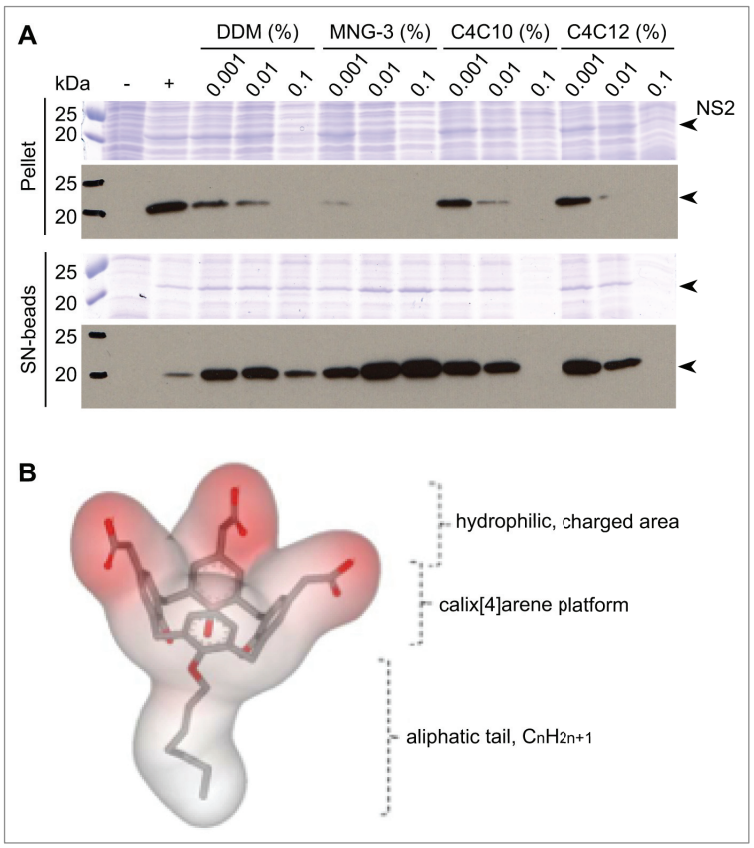


Figure IV.8 ***Production of full-length NS2 in the presence of DDM, MNG-3 or calixarenes.*** (**A**) Protein samples were analyzed by SDS-PAGE (upper panels) and western blotting (lower panels). Pellet, pellet obtained after centrifugation of CFS; SN-beads, sample enriched in NS2 by incubation of CFS supernatant with *Strep*-Tactin magnetic beads to capture tagged NS2. -, negative control (no NS2); +, positive control (NS2 expressed in absence of detergent). Comparable amounts were loaded on the gel for Pellet and SN-beads. The black arrowheads indicate the bands corresponding to NS2. (**B**) Chemical structure of calixarenes (C4Cn). Three aromatic rings are substituted by a methylene carboxyl group, -CH_2COOH, at the *para* position. An aliphatic chain R, $O(CH2)_{0\text{-}11}CH3$, is grafted onto the fourth phenolic group. The resulting 3D structure is modelled from the crystal structure of nitrile derivatives, substituting the CN groups with carboxyl groups (Matar-Merheb *et al*, 2011).

2.3.2. Expression of GFP and full-length NS2, NS4B and NS5A proteins in the presence of amphipol A8-35

Amphipols are amphipathic polymers adsorbing onto the hydrophobic transmembrane surface of membrane proteins and represent an alternative to detergents (Popot, 2010). The amphipol A8-35 is the only commercially available one (Anatrace, USA). This poly(sodium acrylate)-based amphipol comprises 35% of free carboxylates, 25 % of octyl chains and 40 % of isopropyl groups (**Figure IV.9B**). Its average molar mass is 4.3 kDa (Zoonens & Popot, 2014). Amphipol A8-35 has been reported to block *in vitro* expression using *E. coli* lysates, presumably as a consequence of binding to hydrophobic segments as they exit the ribosome tunnel (Park *et al*, 2011). Although a nonionic amphipol (NAPol) was reported to be compatible with cell-free protein expression, it is not commercially available (Bazzacco *et al*, 2012). As the use of A8-35 in WGE-CF systems had not been reported, we tested the expression of GFP and full-length NS2, NS4B and NS5A proteins in the presence of 3 g/L A8-35. As shown in **Figure IV.9A**, western blotting analysis using an antibody against the *Strep*-tag II revealed that A8-35 fully inhibited the expression of both soluble (GFP) and membrane proteins (NS2, NS4B and NS5A). Moreover, fluorescence intensity measurements confirmed the inhibition of GFP expression by A8-35. In contrast, GFP expression is not inhibited by A8-35 in an *E. coli* cell-free system (Workshop « Applications of Amphipols to Membrane Protein Studies, October 21-24, 2013, Paris). Recently, a study reported the inhibitory effect of A8-35 on the expression of curdlan synthase in a WGE-CF system, supporting the present results (Periasamy *et al*, 2013).

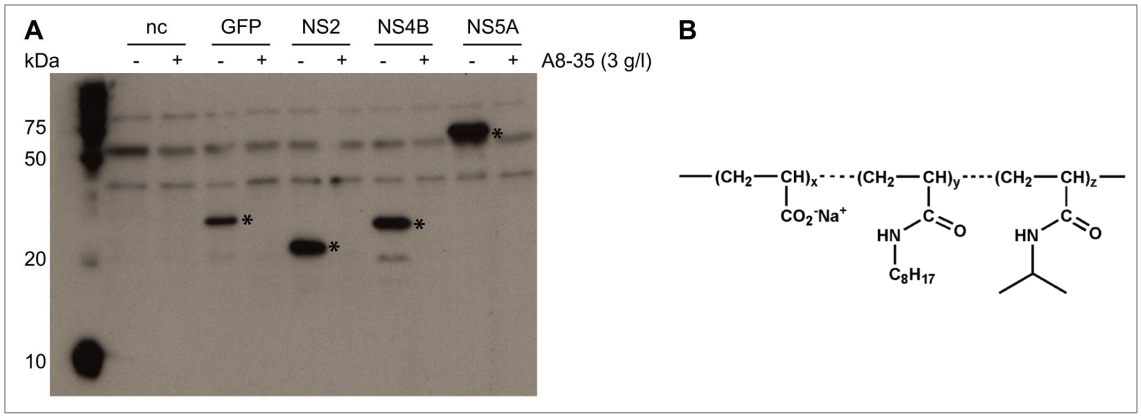


Figure IV.9 ***Expression of GFP as well as full-length NS2, NS4B and NS5A proteins in the presence of amphipol A8-35.*** (**A**) CFS was analyzed by western blotting using an antibody against the *Strep*-tag II fused to each protein. The bands corresponding to the protein of interest are indicated by an asterix. nc, negative control (no mRNA). (**B**) Chemical structure of A8-35 (Bazzacco *et al*, 2012).

2.4. Conclusion and perspectives

Here the aim was to optimize the expression of solubilized membrane proteins in the presence of detergent for efficient further purification and reconstitution in lipids in view of structural studies by solid-state NMR. Preliminary tests were performed using 18 detergents as well as calixarenes and amphipols. Whereas the only amphipol commercially available was not compatible with the WGE-CF system, the use of calixarenes instead of commercial detergents proved not to be very attractive. We finally identified two detergents with a low critical micelle concentration, MNG-3 and C12E8, which allowed us to efficiently express HCV membrane proteins directly in a solubilized form.

Recently, the company CellFree Sciences performed tests for the expression of six membrane proteins in the presence of various detergents (Matthias Harbers, personal communication). In keeping with our results, they found that digitonin increased both the yield and the solubility (tested concentrations : 0.25% or 0.5%, w/v). They also found that Brij-35 increased the solubility without decreasing the expression yield (tested concentrations : 0.025% or 0.05%, w/v). In our case, inhibition of GFP expression by Brij-35 could be due to the quality of the detergent used, and it might be worth to repeat this test with high quality detergent. Nevertheless, accordingly to our results, they found that Tween 20, Triton X-100 and NP-40 tend to reduce expression yields, whereas their effect on protein solubility was protein-dependent, and that CHAPS decreased the expression yield at concentrations higher than 0.25%. In addition, they also observed a decrease of the expression yield with DDM (tested concentrations : 0.03% and 0.06%, w/v). These results underline the fact that the effects of detergents are mainly protein-dependent and that the most suitable detergent has to be found for each membrane protein. Although the general use of C12E8 and MNG-3 detergents for the successful expression of solubilized membrane proteins remains to be established, MNG-3 proved in our hands to be the detergent of choice for the expression of several other membrane proteins, and also for soluble proteins which tend to aggregate. The expression of membrane proteins in a detergent-solubilized form opens the way for functional and structural studies, especially by solid-state NMR and electron microscopy after reconstitution into membranes. The progress made in this direction for the study of NS2, NS4B, NS5A and core HCV proteins are described in the following sections.

3. NS2

NS2 is a 23 kDa integral membrane protein essential for HCV polyprotein processing and virion assembly. NS2 is composed of a N-terminal membrane domain and a cytosolic C-terminal domain. The N-terminal domain is believed to comprise 3 transmembrane segments, whereas the C-terminal domain, called $NS2^{pro}$, displays protease activity and cleaves at the junction between NS2 and NS3, thus promoting HCV RNA replication. The x-ray crystal structure of $NS2^{pro}$ revealed a domain-swapped homodimer with composite catalytic triads including residues from the two chains. $NS2^{pro}$ was moreover shown to contribute to membrane association via two α-helices (refer to section II.6.4. for details). The aim of this work is to produce purified isotopically labeled full-length NS2 and $NS2^{pro}$ samples for structural studies.

3.1. Full-length NS2 expression test

Eight different full-length NS2 constructs from different HCV strains (JFH-1, Con1, H77 and 452, GenBank accession numbers AB047639, AJ238799, AF009606 and DQ437509, respectively) were cloned by Vlastimil Jirasko, post-doc in Ralf Bartenschlager's lab in Heidelberg, with either a *Strep*-tag II or both a *Strep*-tag II and a thrombine cleavage site fused at the N-terminus. The amino acid sequences of the latter constructs are given in Appendix 1.

All these constructs were successfully expressed using the WGE-CF system in the precipitate mode. Moreover, the expression of the four constructs with both a *Strep*-tag II and a thrombine cleavage site was also successfully tested in the presence of MNG-3. The CFS as well as the pellet and the supernatant obtained after centrifugation of the CFS were analyzed by SDS-PAGE. Before analysis, the supernatant was incubated with magnetic beads coated with *Strep*-Tactin to capture the *Strep*-tag II fused at the N-terminus of each construct. This step aims to enrich the sample in the protein of interest since it is usually not or hardly detected in the CFS and the supernatant on Coomassie gels.

As expected due to its hydrophobic nature, the NS2 protein is mainly insoluble in the absence of detergent and was therefore detected in the pellet, independently from the strain (**Figure IV.10**). Only a minor portion of NS2 was found in the supernatant, possibly due to the presence of some residual lipids in the WGE. In contrast, in the

presence of 0.1 % MNG-3, fully solubilized JFH-1 and Con1 NS2 were obtained (see also (Fogeron *et al*, 2015)) while H77 and 452 NS2 partially remained in the pellet (**Figure IV.10**). At least in the case of 452 NS2, this could be explained by the fact that the expression level of this construct seemed to be higher than that of the JFH-1 and Con1 constructs. It is thus possible that higher amounts of MNG-3 could allow the expression of 452 NS2 in a fully solubilized form. This remains to be checked.

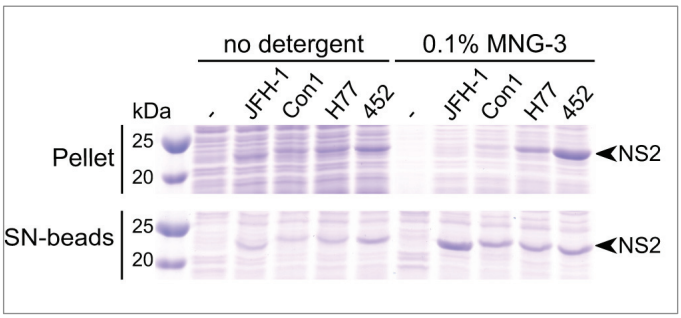


Figure IV.10 ***Small scale expression test of full-length NS2 constructs in the precipitate mode and in the presence of MNG-3 detergent.*** Full-length NS2 from different HCV strains were expressed in the absence of detergent (left) or in the presence of 0.1% MNG-3 (right). Protein samples were analyzed by SDS-PAGE. -, negative control (no NS2 mRNA); Pellet, pellet obtained after centrifugation of CFS; SN-beads, sample enriched in NS2 by incubation of CFS supernatant with *Strep*-Tactin magnetic beads to capture tagged NS2. Comparable amounts were loaded on the gel for Pellet and SN-beads. Black arrowheads indicate the bands corresponding to full-length NS2.

It should be mentioned that full-length NS2 (HCV strain JFH-1) was also expressed using various isotopically labeled amino acid mixtures : $^{13}C/^{15}N$, $^{2}H/^{15}N$ and $^{2}H/^{13}C/^{15}N$. None of these labeled aa mixtures seemed to affect the expression level of NS2 compared to a standard unlabeled aa mixture. Note also that in the following sections, the full-length NS2 construct from the HCV strain JFH-1 with both a *Strep*-tag II and a thrombine cleavage site fused at the N-terminus is systematically used.

3.2. Cotranslational protease activity of full-length NS2

In order to evaluate the functionality of NS2 expressed in the WGE-CF system we monitored its cysteine-protease activity. To this end, we expressed a NS2-NS3^{pro} *Strep*-tagged precursor construct from the HCV strain JFH-1 including the N-terminal protease domain of NS3 (NS3^{pro}) and denoted NS2-NS3^{pro}-ST. The cleavage products were detected by western blotting using either an antibody against the NS3^{pro} C-terminal *Strep*-tag II or a conformational antibody against NS2. As shown in **Figure IV.11**, both cleavage products, being NS3^{pro}-ST (left panel) and NS2 (right panel) could be detected in the reaction mixture. Cleavage specificity was confirmed by using the inactive NS2

mutant bearing the C184A mutation in its catalytic site, for which accumulation of uncleaved NS2-NS3^{pro} precursor was observed.

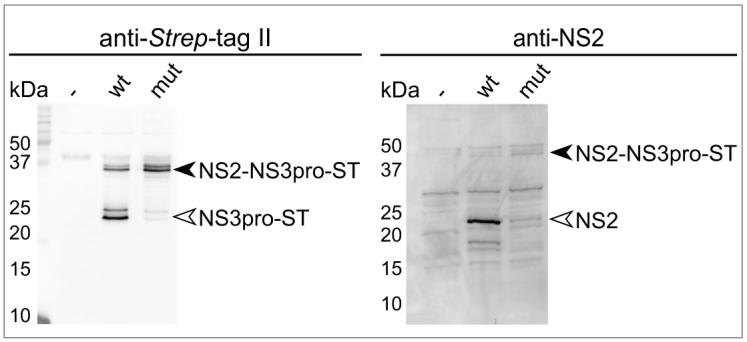


Figure IV.11 ***Protease activity of full-length NS2.*** Wild type and mutant NS2-NS3^{pro}-ST precursors (HCV strain JFH-1) were expressed in the presence of 0.1% MNG-3. Note that NS3^{pro} is inactive (S139A mutation) to avoid unspecific cleavages by this protease. CFS was analyzed by immunoblotting using antibodies, either against the *Strep*-tag II fused at the C-terminus of NS3^{pro} (left panel) or against the C-terminal domain of NS2 (right panel). -, negative control (no NS2-NS3^{pro} mRNA); wt, wild type NS2; mut, NS2 with C184A mutation. Black arrowheads and empty arrowheads indicate the bands corresponding to uncleaved NS2-NS3^{pro}-ST protein precursor and cleaved NS2 and NS3^{pro}-ST, respectively.

3.3. Purification of full-length NS2

3.3.1. Purification of full-length NS2 expressed in the precipitate mode

Before protein expression in the presence of detergent was optimized (refer to section IV.2.), full-length NS2 was successfully expressed in the precipitate mode. Solubilization tests were performed after translation with either 1% DDM, digitonin, OG, DPC or Brij-35 (**Table IV.2**). Note that all detergent concentrations are given in w/v. Detergents were added directly to the CFS after 16 hours incubation and before its centrifugation. Whereas OG did not allow to solubilize NS2, protein solubilization was incomplete with digitonin. In contrast, DDM, Brij-35 and DPC allowed to almost completely solubilize the protein (data not shown). However, in the presence of DPC, NS2 could not be detected in the SN-Beads sample, indicating that this detergent disturbed the interaction with the magnetic beads coated with *Strep*-Tactin (refer to section III.7. for technical details).

In addition, NS2 was produced in the precipitate mode using 250 µL WGE, and a small scale purification was performed after solubilization with 1% either DDM or Brij-35 added directly to the CFS. A detergent concentration of 0.1% was used during the whole purification process. The purification yield was 4 times higher with DDM than with Brij-35 (**Figure IV.12A**). The question was then to determine whether centrifuging the CFS and resuspending NS2 found in the pellet with 1% DDM (*i.e.* « indirect » solubilization) instead of adding 1% DDM directly to the CFS (*i.e.* « direct »

solubilization) would improve the purification process. Indeed, the centrifugation of NS2 in the absence of detergent could have been an additional purification step. As shown in **Figure IV.12B**, purification yield was 30% higher with the direct solubilization compared to the indirect solubilization. Moreover, indirect solubilization did not allow to achieve higher purity levels (**Figure IV.12C**).

Although expression in the precipitate mode followed by solubilization with DDM gave satisfactory results, expression in presence of MNG-3 and further purification in presence of DDM allowed to achieve much higher production yields, as described in the following section.

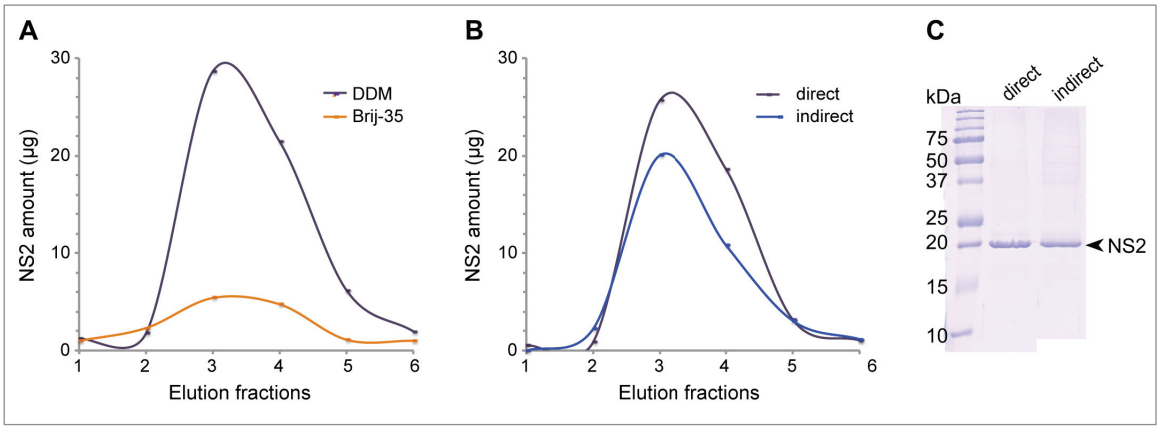


Figure IV.12 ***Purification of full-length NS2 expressed in the precipitate mode.*** (**A**) Affinity chromatography elution profiles of NS2 (HCV strain JFH-1) solubilized in 1% DDM (*purple*) or Brij-35 (*orange*). Purification was performed starting from 6 mL supernatant on a 200-µL *Strep*-Tactin column, in presence of 0.1% detergent. (**B**) Affinity chromatography elution profiles of NS2 (HCV strain JFH-1) solubilized directly (*purple*) or indirectly (*blue*) in 1% DDM. Purification was performed in the presence of 0.1% DDM. (**C**) Purified NS2 solubilized directly or indirectly in DDM was analyzed by SDS-PAGE.

3.3.2. Purification of full-length NS2 expressed in the presence of detergent

A major challenge in the expression of membrane proteins is to isolate the protein in a pure form, and to avoid aggregate formation. The N-terminal *Strep*-tag II allowed us to obtain pure NS2 protein using a single affinity-purification step starting from the supernatant fraction containing the solubilized protein (**Figure IV.13A**). It is worth to mention that the addition of 0.25% DDM to the CFS before the affinity purification step allowed us to exchange MNG-3 to DDM and to achieve a high NS2 purity level with a good yield. Indeed, the purity of NS2 obtained in the presence of MNG-3 alone during this step was not satisfactory (see (Fogeron *et al*, 2015)), as well as in the presence of

lower amount of DDM (data not shown). In addition, while pure NS2 was obtained in the presence of higher amounts of DDM (0.5% and 1%), the protein yield was seriously reduced because of the decrease of NS2-*strep*-tag II binding strength to *Strep*-Tactin column. The yield of purified NS2 was about 1 mg/mL WGE. To further characterize purified NS2 protein, we performed size exclusion chromatography. A major peak containing NS2 was observed (**Figure IV.13B**), while the following minor peak could be assigned to reagents present in the *Strep*-Tactin elution buffer. NS2 protein-detergent complexes eluted with an apparent molecular mass of about 150 kDa (**Figure IV.13C**), indicating the oligomeric nature of recombinant full-length NS2 in detergent micelles.

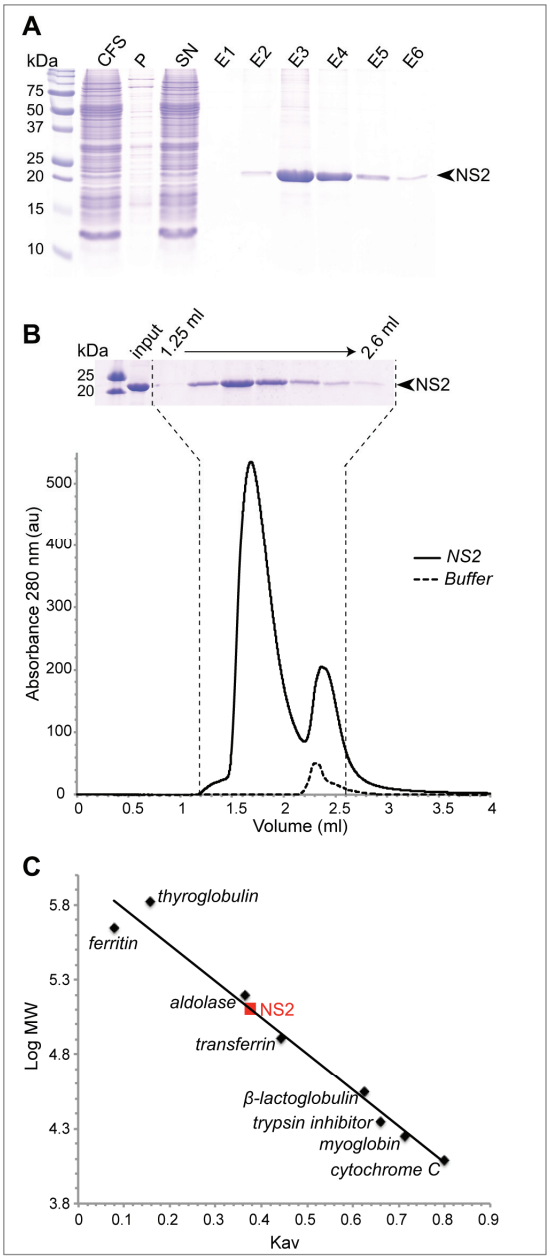


Figure IV.13 ***Purification of full-length NS2 by affinity chromatography and analysis by size exclusion chromatography.*** (**A**) Full-length N-terminally *Strep*-tag II tagged NS2 (HCV strain JFH-1) was expressed using 4.5 mL WGE in the presence of 0.1% MNG-3 and purified on a 5-mL *Strep*-Tactin column in the presence of 0.1% DDM. Protein samples were analyzed by SDS-PAGE. CFS, total cell-free sample (8 µL from 108 mL loaded on the gel); P, pellet obtained after centrifugation of CFS (equivalent to 80 µL from 108 mL loaded on the gel); SN, supernatant obtained after centrifugation of CFS and loaded on the affinity column (8 µL from 108 mL loaded on the gel); E1 to E6, affinity elution fractions (8 µL from 2.5 mL loaded on the gel). The black arrowhead indicates the bands corresponding to NS2. (**B**) Size exclusion chromatography of E3 elution fraction using a Superdex™ 200 Increase 3.2/100 column. A 50 mM phosphate buffer pH 7.4 containing 0.1% DDM was used as eluent and a flow rate of 75 µL/min was applied. The elution profile of NS2 is drawn in a solid line, and that of the affinity elution buffer in a dotted line. SDS-PAGE and Comassie staining analysis of the collected fractions is shown above the chromatographic profile. Elution volume of NS2 is 1.65 mL. (**C**) Calibration curve of the size exclusion chromatography column established using protein standards (V_0 is 1.25 mL). NS2 is indicated in *red* on the standard curve.

Note that we also tried to purify full-length NS2 expressed in the presence of C12E8 following the same purification process. Although C12E8 did not disturb the binding of NS2 on *Strep*-Tactin coated magnetic beads (Fogeron *et al*, 2015), this detergent interfered with the purification on *Strep*-Tactin columns, hampering the binding of the tagged protein to the *Strep*-Tactin.

3.4. Analysis of full-length NS2 by circular dichroism

The far UV circular dichroism (CD) spectrum of full-length NS2 eluted from size exclusion chromatography in 0.1% DDM (refer to previous section) is typical of a well-folded protein (**Figure IV.14A**). Additionally, the molar ellipticity per residue is in the range that can be expected for such a protein. The two minima at 208 and 222 nm, together with a maximum at 192 nm, are typical of α-helical folding. Spectral deconvolution indicated an α-helix content of 52-58%, while turns represent 12-17%, and β-sheet content is limited to 8-9%. The latter value is consistent with the content in β-sheet deduced from the crystal structure of the cytosolic domain of NS2 (Lorenz *et al*, 2006). The limited number of residues folded into α-helices in this crystal structure indicates that the main contribution to helix content in full-length NS2 is due to its membrane domain. This is in agreement with the previous structural analyses of this membrane domain using synthetic peptides, which mostly revealed the presence of α-helical segments (Jirasko *et al*, 2008; Jirasko *et al*, 2010).

In addition, in view of structural studies by NMR, we analyzed the stability of purified full-length NS2 in a large temperature range. Thermal denaturation of full-length NS2 purified by affinity chromatography and size exclusion chromatography was performed between 20°C and 90°C with 5°C incrementation steps, showing that the protein is stable up to 40°C (**Figure IV.14B**).

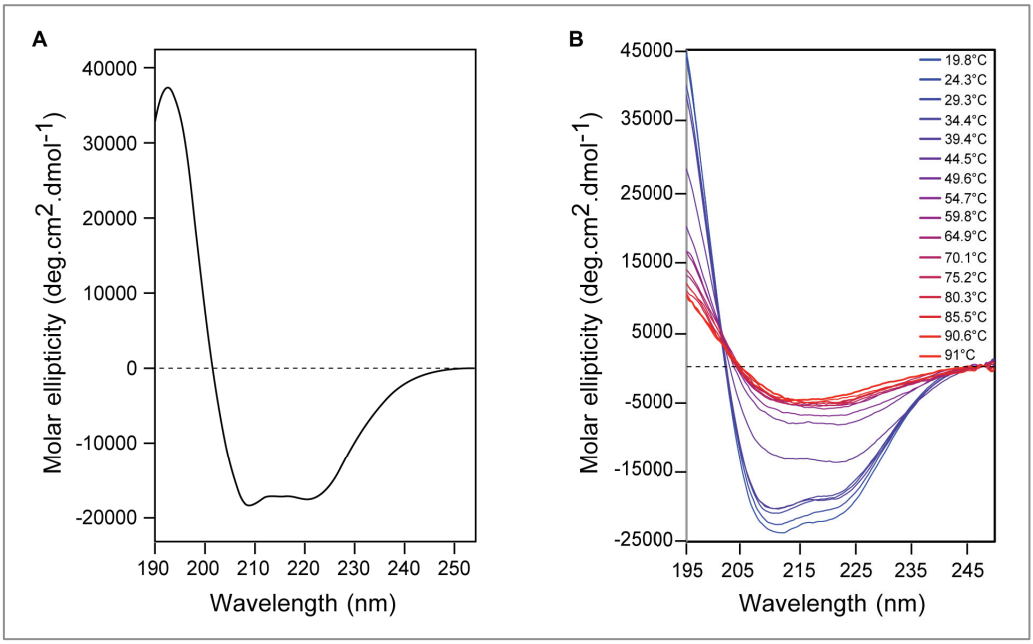


Figure IV.14 ***Analysis of full-length NS2 by circular dichroism (CD).*** (**A**) Far UV CD spectrum of full-length NS2 (HCV strain JFH-1) after purification by affinity chromatography and size exclusion chromatography. The NS2 sample was in 50 mM phosphate buffer pH 7.4 containing 0.1% DDM. Estimation of the secondary structure content resulted in an α-helix content of 52-58%, 12-17% turns, and 8-9% β-sheet. (**B**) Thermal denaturation of full-length NS2.

3.5. Protease activity of purified full-length NS2

As described in section IV.3.2., full-length NS2 is active during *in vitro* protein synthesis. We then raised the question whether purified full-length NS2 is also active. Although there is no enzymatic test available, we observed autoproteolysis, as detailed below.

The total mass of purified full-length NS2 was analyzed by mass spectrometry by Frédéric Delolme, engineer at the Protein Science Facility at IBCP, Lyon (**Figure IV.15A and 15B**). The expected mass of the protein is 25,865 Da. Two peaks were obtained for NS2 purified by affinity chromatography followed by size exclusion chromatography in phosphate buffer pH 7.4 containing 0.1% DDM without any further additives (**Figure IV.15A**). Whereas the peak at 25,738 Da corresponds to the full-length protein after N-terminal methionine removal, another peak was observed at 24,938 Da. This peak probably corresponds to the protein after deletion of the *Strep*-tag II by autoproteolysis, which might occur after the affinity chromatography purification step on the *Strep*-Tactin column (as reminder, there is a thrombine cleavage site between the tag and the

protein). When NS2 was analyzed directly after purification by affinity chromatography, only one peak corresponding to the full-length protein after N-terminal methionine removal was observed at 25,760 Da (**Figure IV.15B**). Note that this sample was in Tris buffer pH 8.0 containing 0.1% DDM and 1 mM EDTA. In addition, the NS2 sample purified by affinity chromatography followed by size exclusion chromatography was incubated at RT for 72 h and analyzed by SDS-PAGE after 24 h, 48 h and 72 h (**Figure IV.15C**). Two bands were already observed on the gel before incubation, most probably corresponding to the two peaks detected by mass spectrometry, namely full-length NS2 and NS2 deleted from the tag. After 24 h, only the band corresponding to NS2 deleted from the tag was detected on the gel. This band was weaker, indicating that the protein was partially degraded. Moreover, only a very weak band was observed after 48 h and the sample was fully degraded after 72 h. Moreover, a NS2 sample was incubated at RT for 72 h directly after purification by affinity chromatography. At this step, the protein was in Tris buffer pH 8.0 containing 0.1% DDM and 1 mM EDTA. As shown in **Figure IV.15D**, no degradation was observed after 72 h. Furthermore, a NS2 sample obtained after purification by affinity chromatography and dialysis against phosphate buffer pH 6.0 containing chelating resin (Chelex 100®, Sigma-Aldrich) was incubated in the same way for 72 h. The chelating resin is a styrene divinylbenzene copolymer binding polyvalent metal ions. No degradation was observed in this case (**Figure IV.15E**).

Autoproteolysis seemed therefore not to be buffer- or pH-dependent but chelation of metal ions with EDTA or chelating resin was able to prevent it. Although it is important to keep this in mind while working with full-length NS2, this is the proof that purified full-length NS2 is active.

Alternatively, one option could be to work with the construct without the thrombine cleavage site. However, we observed that production yields were lower for NS2 constructs lacking this sequence. An even better option would be to work with inactive mutants of in the future.

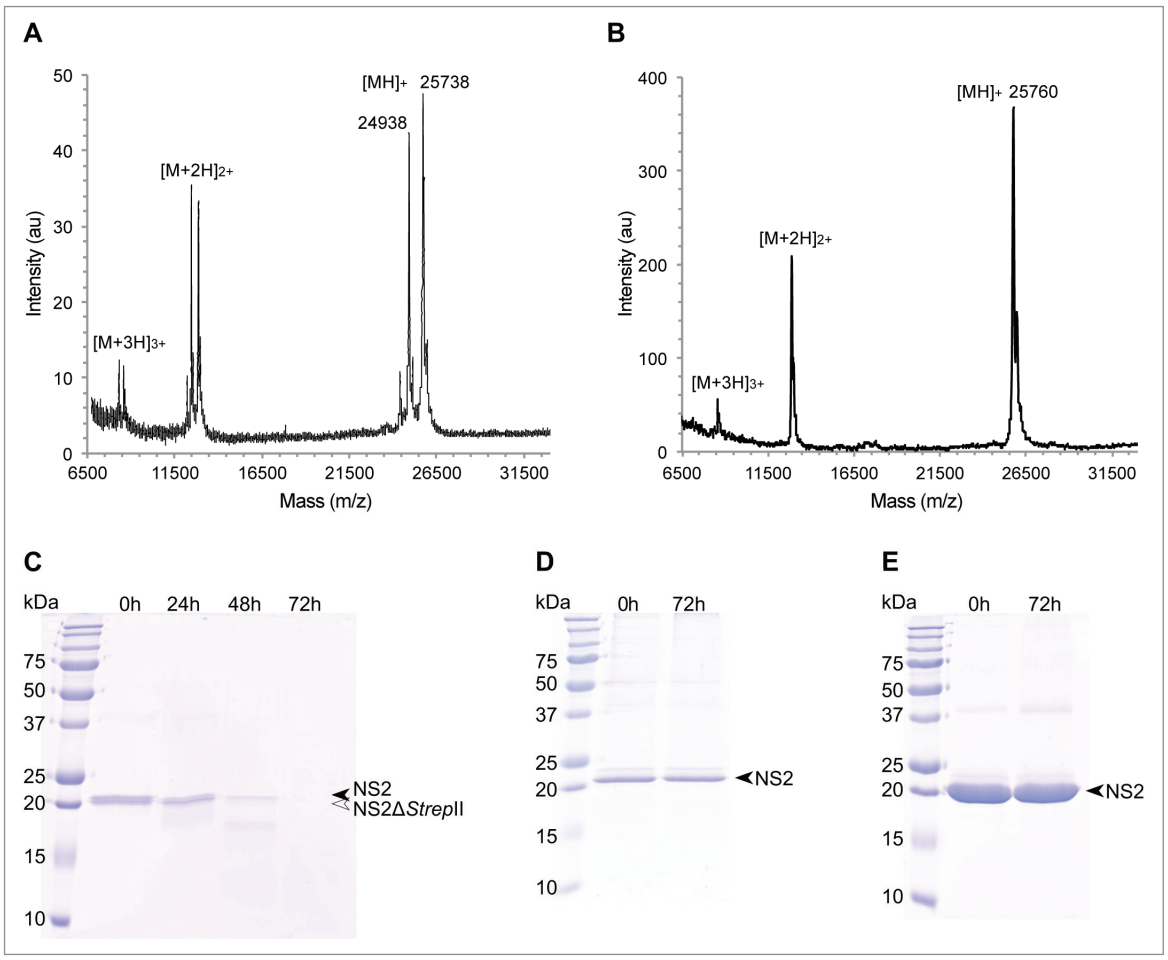


Figure IV.15 ***Analysis of purified full-length NS2 protease activity.*** (**A**) Total mass analysis by mass spectrometry (MALDI-TOF) after purification by affinity chromatography and size exclusion chromatography. The protein sample was in 50 mM phosphate buffer pH 7.4 and 0.1% DDM. (**B**) Total mass analysis by mass spectrometry (MALDI-TOF) directly after purification by affinity chromatography. The protein sample was in 100 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin and 0.1% DDM. (**C**) Incubation of NS2 in phosphate buffer pH 7.4 containing 0.1% DDM for 72 h at RT after purification by affinity chromatography and size exclusion chromatography (sample from panel A). A black arrowhead indicates full-length NS2 and an empty arrowhead indicates NS2 deleted from the *Strep*-tag II. (**D**) Incubation of NS2 in Tris buffer pH 8.0 containing 0.1% DDM for 72 h at RT directly after purification by affinity chromatography (sample from panel B). Note that the affinity elution buffer contains 1 mM EDTA. A black arrowhead indicates full-length NS2. (**E**) Incubation of NS2 in phosphate buffer pH 6.0 containing 0.1% DDM for 72 h at RT after purification by affinity chromatography followed by dialysis against phosphate buffer pH 6.0 for solution NMR analysis and further concentration. Note that chelating resin was added to the dialysis buffer to bind divalent cations. A black arrowhead indicates full-length NS2.

3.6. Analysis of full-length NS2 by solution NMR

Uniformly $^{13}C/^{15}N$ labeled full-length NS2 was produced in the presence of MNG-3 using 3 mL WGE to prepare a solution NMR sample. The CFS was treated with benzonase and 0.25% non-deuterated DDM were added. NS2 was purified from the supernatant obtained after centrifugation of the CFS on four 1-mL *Strep*-Tactin columns.

The whole purification process was performed in the presence of 0.1% DDM and tail deuterated DDM was used starting from the *Strep*-Tactin washing steps. In total, 2 mg purified NS2 were obtained (*i.e.* purified protein yield of 0.67 mg/mL WGE). The most concentrated elution fractions were pooled (1.2 mg NS2 in 1 mL) and dialyzed against 50 mM phosphate buffer pH 6.0. A dialysis membrane with a MWCO of 6-8 kDa was used and dialysis was performed over 3 h, the dialysis buffer being changed every hour. In these conditions, the loss of detergent by dialysis is limited. Note that chelating resin was added to the dialysis buffer to catch metal ions in order to avoid autoproteolysis (refer to section IV.3.5.). The protein was then concentrated overnight at 4°C, directly in the dialysis bag packed in Sephadex G25 powder for liquid adsorption. After concentration, the NS2 sample was collected and centrifuged at 20,000 *g*, 4°C for 10 min (a precipitate appeared during the third dialysis step, likely due to some loss of detergent). At the end, 290 µL supernatant were available to which 5 mM (8 µL) tail deuterated DDM and 5% (15 µL) D_2O were added. Final NS2 sample concentration was 2.28 mg/mL, there was therefore 710 µg protein in this sample (**Figure IV.16A**).

Measurements were done on a 1 GHz spectrometer at the CRMN (Centre de Résonance Magnétique Nucléaire à très hauts champs) in Lyon together with Moreno Lelli, engineer at the CRMN and Roland Montserret, engineer in the lab. A 2D TROSY (Transverse Relaxation Optimized SpectroscopY) spectrum is shown in **Figure IV.16B**. Whereas the protein is composed, together with the tag, of 237 amino acids (refer to Appendix 1 for full amino acid sequence), only few residues were detected. Moreover, the proton dispersion was very narrow between 7 and 8.5 ppm. The few peaks observed likely correspond to the unfolded, flexible tag (19 residues in total). NS2/detergent micelles are most probably too large to tumble sufficiently fast to yield narrow NMR signals and thus can not be analyzed by solution NMR. Solid-state NMR represents thus the alternative of choice. In order to analyze NS2 in a native-like membrane lipids environment, reconstitution of the protein in lipids has been optimized, as described in the following section.

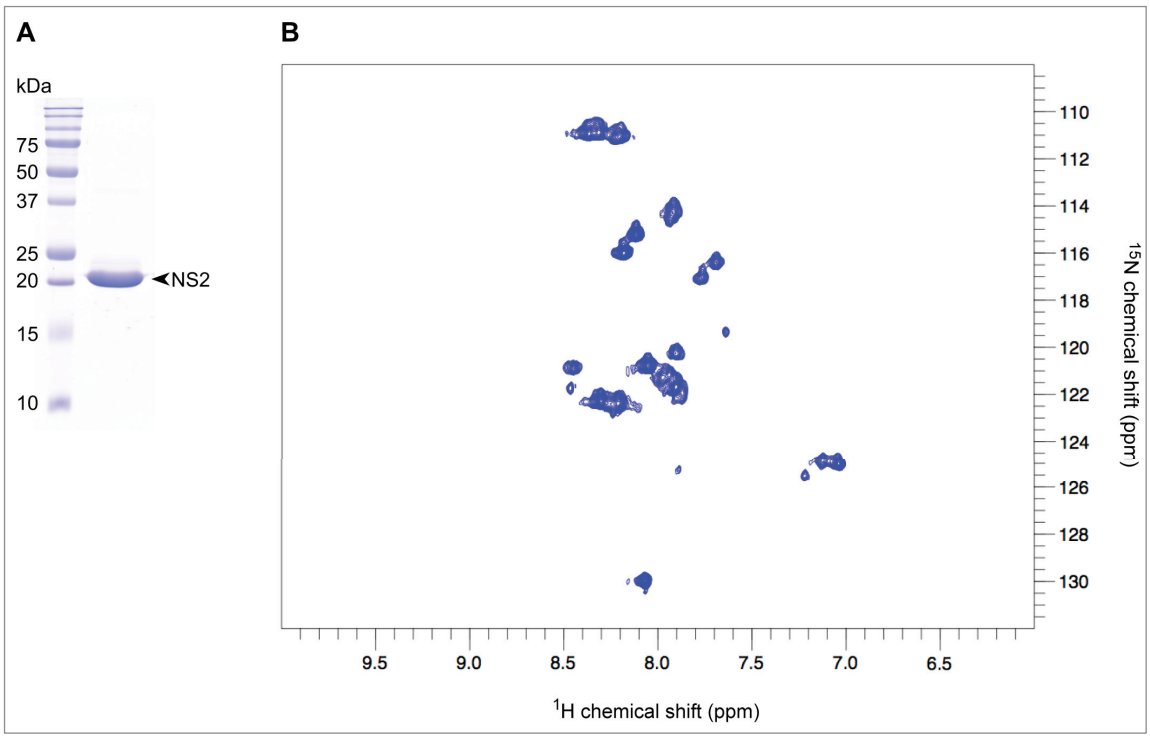


Figure IV.16 ***Analysis of full-length NS2 by solution NMR.*** (**A**) Purified uniformly $^{13}C/^{15}N$ labeled NS2 (HCV strain JFH-1) for solution NMR analysis. The protein (2.28 mg/mL) was in 50 mM phosphate buffer pH 6.0 containing 5 mM tail deuterated DDM and 5% D_2O. (**B**) TROSY spectrum recorded on a 1GHz spectrometer at the CRMN in Lyon (TGIR-RMN-THC Fr3050 CNRS). Number of scans : NS=16, recycling delay : D1=1s, acquisition time : AQ=0.08s/scan, total acquisition time : 1h15.

3.7. Reconstitution of full-length NS2 in lipids

Full-length NS2 was reconstituted in lipids using either a classical reconstitution protocol including a long detergent-dialysis step or the GRecon method which consists in reconstituting proteoliposomes directly on a density gradient containing lipids and detergent (refer to section III.9.).

3.7.1. Classical reconstitution protocol

Preliminary small scale reconstitution tests

A first reconstitution test was performed with three different types of lipids : egg L-α -phosphatidylcholine (PC), a mixture of egg PC and cholesterol (PC/Chol, 70/30, w/w) and home-made lipids from pig liver. Lipid-to-protein ratios (LPRs, mol/mol) of 10, 20, 100 and 800 were tested. Bio-Beads were progressively added directly to the protein/lipid/detergent mixtures to remove the detergent. Reconstituted samples were analyzed by electron microscopy (EM) at the Centre d'Imagerie Quantitative Lyon Est

(CIQLE). Negative staining using phosphotungstic acid was done by Elisabeth Errazuriz, engineer at the CIQLE. Unfortunately, EM analysis was not concluding. One hypothesis was that the protein could have aggregated on the Bio-Beads. Therefore, further tests (detailed below) were performed using the dialysis method in which the Bio-Beads are added to the dialysis buffer. Note that this technique takes several days instead of a few hours when adding the Bio-Beads directly to the sample. Small dialysis cups fitting in a 24-well plate were used for this purpose. Several reconstitution tests were performed this way with PC, PC/Chol and home-made lipids from pig liver at LPR 0.5, 1 and 2 (w/w). However, SDS-PAGE analysis showed that a large portion of NS2 was systematically lost during the reconstitution process (data not shown). Since this was not the case for NS4B which was tested in parallel, one could think that it was not possible to reconstitute NS2 by this approach and/or that most of NS2 was lost because of irreversible binding to the walls of the dialysis device. Nevertheless, EM analysis of the NS2 samples, which was performed by Philippe Roingeard and Sonia Georgeault at the Plate-Forme RIO des Microscopies in Tours, showed that multilayers dense to electrons were present for reconstituted samples whereas only structures clear to electrons were present for control samples containing only lipids, independently from the lipids used for reconstitution. Although no immunogold labeling was performed, the multilayers dense to electrons were a good indication for a at least partially successfull reconstitution. These results therefore encouraged us to go forward with the optimization of the reconstitution and we continued using a larger format, directly in dialysis bags.

Reconstitution tests with high LPR

Considering the previous results, we performed reconstitution at larger scale using the dialysis method. In short, 100 µg (235 µL at 0.42 mg/mL) of NS2 purified by affinity chromatography in 0.1% DDM were reconstituted in either PC or PC/Chol (70/30, w/w) using a LPR of 20 (w/w). The lipids were solubilized with DDM using a detergent-to-lipid ratio of 10 (mol/mol). The protein/lipid/detergent mixture was dialyzed at RT for 6 days (MWCO 6-8 kDa). Bio-Beads were added to the dialysis buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT and 0.025% NaN_3), a ratio Bio-Beads-to-detergent of 100 (w/w) being applied. Reconstituted samples were centrifuged and the pellet and the supernatant obtained were analyzed by SDS-PAGE. Due to the high

amount of lipids in the samples, the protein was mainly detected in the supernatant and the bands were broader and more diffuse than usual (**Figure IV.17A**). Nevertheless, NS2 was not lost during the reconstitution process, in contrast to the preliminary small scale reconstitution tests. The problem in the preliminary tests was thus not the reconstitution itself but the material used on which the protein most probably aggregated. Reconstituted samples were analyzed by EM by David Paul, post-doc in Ralf Bartenschlager's lab in Heidelberg. As already indicated by size exclusion chromatography (refer to section IV.3.3.2.), NS2 solubilized in detergent forms multimers, which can be visualized by EM using immunolabeling of the N-terminal *Strep*-tag II fusion protein produced in the presence of detergent (**Figure IV.17B**). Absence of gold particles on negative control grids lacking NS2 confirmed labeling specificity (data not shown). When detergent-solubilized lipids were added to detergent-solubilized NS2, the protein bound to or integrated into the lipid bilayer upon detergent removal, as shown in **Figure IV.17C** and **Figure IV.17D** for PC and a PC/Chol mixture, respectively. No dispersed labeling of NS2 protein was found, and gold-decorated liposomes were observed, suggesting the tight association/insertion of NS2 into model membranes.

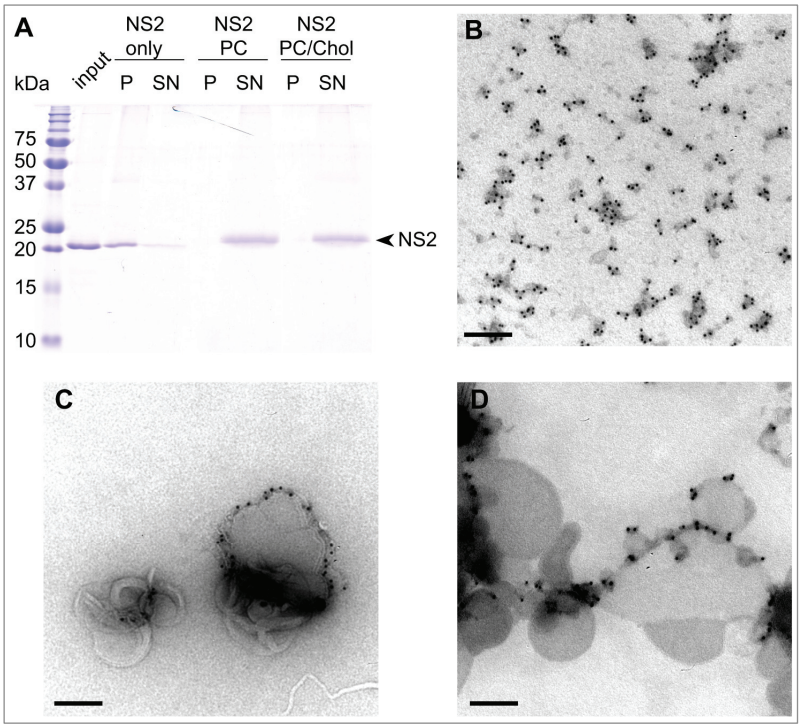


Figure IV.17 ***Reconstitution of full-length NS2 with high LPR.*** (**A**) P, pellet and SN, supernatant obtained after centrifugation of reconstituted samples were analyzed by SDS-PAGE. A black arrowhead indicates the bands corresponding to NS2 (HCV strain JFH-1). (**B**, **C** and **D**) Electron microscopy analysis. Negative staining was performed after immunogold labeling (10 nm colloidal gold suspension) with an antibody against the *Strep*-tag II. (**B**) NS2 in 0.1% DDM after purification by affinity chromatography and size exclusion chromatography, (**C**) NS2 reconstituted in egg PC (L-α-phosphatidylcholine) and (**D**) NS2 reconstituted in a mixture of egg PC and cholesterol (70/30, w/w). Scale bar, 200 nm.

Reconstitution tests with low LPR

The amount of lipids for reconstitution should be as low as possible in view of analysis by solid-state NMR. Considering the previous results and in view of testing reconstitution with the GRecon method detailed below, reconstitution was performed at larger scale using the dialysis method followed by analysis on a discontinuous sucrose gradient. A PC/Chol mixture was used since cholesterol was observed to accumulate at the membranous web (Ralf Bartenschlager, personal communication). In short, 165 µg (300 µL at 0.55 mg/mL) of NS2 purified by affinity chromatography in 0.1% DDM were reconstituted in PC/Chol (70/30, w/w) using a LPR of 1 (w/w). The PC/Chol mixture was solubilized either with DDM or with Triton X-100 using in both cases a detergent-to-lipid ratio of 10 (mol/mol). The protein/lipid/detergent mixture was dialyzed at RT for 6 days (MWCO 6-8 kDa). Bio-Beads were added to the dialysis buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT and 0.025% NaN_3), a ratio Bio-Beads-to-detergent of 100 (w/w) being applied. Reconstituted samples were centrifuged and the pellet and the supernatant obtained were analyzed by SDS-PAGE. **Figure IV.18A** shows that NS2 was not lost during the reconstitution process and that the reconstitution yield was higher when the lipids were solubilized in Triton X-100.

Reconstituted samples were also loaded on a discontinuous gradient from 5% to 80% sucrose (**Figure IV.18B**). Gradients fractions were analyzed by SDS-PAGE and western blotting using an antibody against the *Strep*-tag II (**Figure IV.18C**), as well as by electron microscopy (**Figure IV.18D**). While protein aggregates (NS2 without lipids) were detected around 50% sucrose, NS2 reconstituted in PC/Chol/DDM or PC/Chol/Triton was detected around 5-10% and 20-30% sucrose, respectively (**Figure IV.18B and 18C**). An additional band, which was not visible on the sucrose gradient after centrifugation, was observed on the Coomassie gel and the western blot around 80% sucrose for NS2 reconstituted in PC/Chol/Triton. Like previously, EM analysis was performed by David Paul (post-doc in Ralf Bartenschlager's lab in Heidelberg). Reconstituted samples were analyzed before and after the analysis on density gradient by immunogold labeling using an antibody against the *Strep*-tag II (**Figure IV.18D**). Whereas the samples with only lipids showed membranous structures without gold coverage, the sample with only NS2 and the corresponding 50% sucrose fraction showed, as expected, big and electron dense protein aggregates. Moreover, for the NS2

sample reconstituted in PC/Chol/DDM and the corresponding 5% sucrose gradient, many liposomes without label, some membranous structures with few gold particles and liposomes with heavy gold coverage were observed, the latter clearly corresponding to reconstituted NS2. Furthermore, the NS2 sample reconstituted in PC/Chol/Triton and the corresponding 30% sucrose gradient showed protein/lipid structures which still might contain detergent. In contrast, in the corresponding 80% sucrose gradient, although gold particles were observed, the presence of membranes was not obvious.

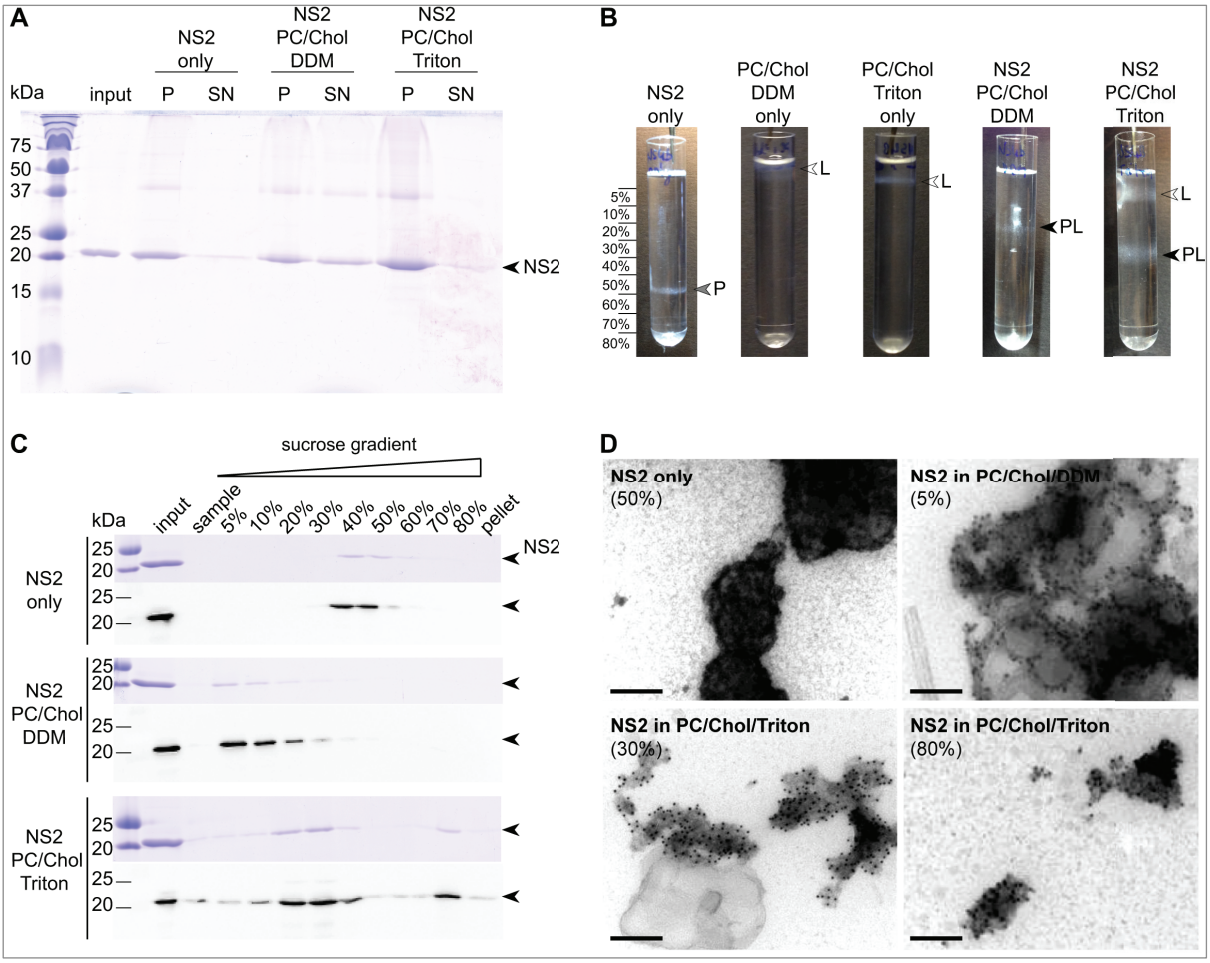


Figure IV.18 ***Reconstitution of full-length NS2 in lipids by dialysis with a LPR of 1.*** (**A**) P, pellet and SN, supernatant obtained after centrifugation of reconstituted samples were analyzed by SDS-PAGE. A black arrowhead indicates the bands corresponding to NS2 (HCV strain JFH-1). (**B**) Analysis of reconstituted samples on a sucrose discontinuous gradient. Sucrose solutions were prepared in 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 1 mM DTT. A grey arrowhead indicates the band corresponding to aggregated NS2 (annotated P), empty arrowheads indicate the bands corresponding to lipids only (annotated L) and black arrowheads indicate the bands correponding to proteoliposomes (annotated LP). (**C**) Fractions from the sucrose gradient shown in (B) were analyzed by SDS-PAGE (top panels) and western blotting with an antibody against the *Strep*-tag II (bottom panels). Black arrowheads indicate the bands corresponding to NS2. (**D**) Electron microscopy analysis of fractions from the sucrose gradient shown in (B). Negative staining was performed after immunogold labeling (10 nm colloidal gold suspension) with an antibody against the *Strep*-tag II. Scale bar, 200 nm.

3.7.2. GRecon method

The GRecon (gradient reconstitution) method consists in reconstituting membrane proteins in lipids directly on a density gradient (Althoff *et al*, 2012). Briefly, the membrane protein solubilized in detergent is loaded on a sucrose gradient containing detergent-solubilized lipids and α-cyclodextrin. During centrifugation, the protein goes through the gradient and meets α-cyclodextrin, which removes progressively the detergent, as well as solubilized lipids, leading to the formation of proteoliposomes (refer to section II.4.2. and III.10.3. for technical details).

First, the optimal α-cyclodextrin concentration for the precipitation of NS2 had to be determined. Twelve α-cyclodextrin concentrations between 0.05% and 0.6% were tested as described in section III.9.4., showing that a concentration of 0.55% α-cyclodextrin was optimal to precipitate the protein in the sample (**Figure IV.19A**). Considering the previous results showing that aggregated NS2 and NS2/PC/Chol proteoliposomes (with PC/Chol solubilized in Triton X-100) obtained by dialysis had a density around 50% and 30%, respectively, a gradient from 30% to 60% sucrose was chosen for the GRecon test. Reconstitution was performed with PC/Chol (70/30, w/w) at a LPR of 1 (w/w). Two controls were prepared in parallel : a gradient without PC/Chol (NS2 only) and a gradient without protein (PC/Chol only). The gradients obtained after centrifugation are shown in **Figure IV.19B**. Gradient fractions were analyzed by SDS-PAGE (**Figure IV.19C**) and by electron microscopy (**Figure IV.19D**). Surprisingly, in the absence of lipids, only part of the sample was aggregated on the gradient, the other part remaining on the top of the gradient in an apparent soluble form (**Figure IV.19C**, top panel). One hypothesis was that the α-cyclodextrin concentration used to remove the DDM was ultimately not optimal. Therefore, such a gradient without lipids was repeated with the same 0.55% α-cyclodextrin concentration and two additional concentrations of 0.65% and 0.75% (**Figure IV.19E**). Although aggregated protein seemed to be more focused in the gradient with higher α-cyclodextrin concentrations, there was still a part of the sample on the top of the gradient. Our hypothesis is that NS2 could possibly be associated with some residual lipids from the wheat germ extract which would interfere with the reconstitution process.

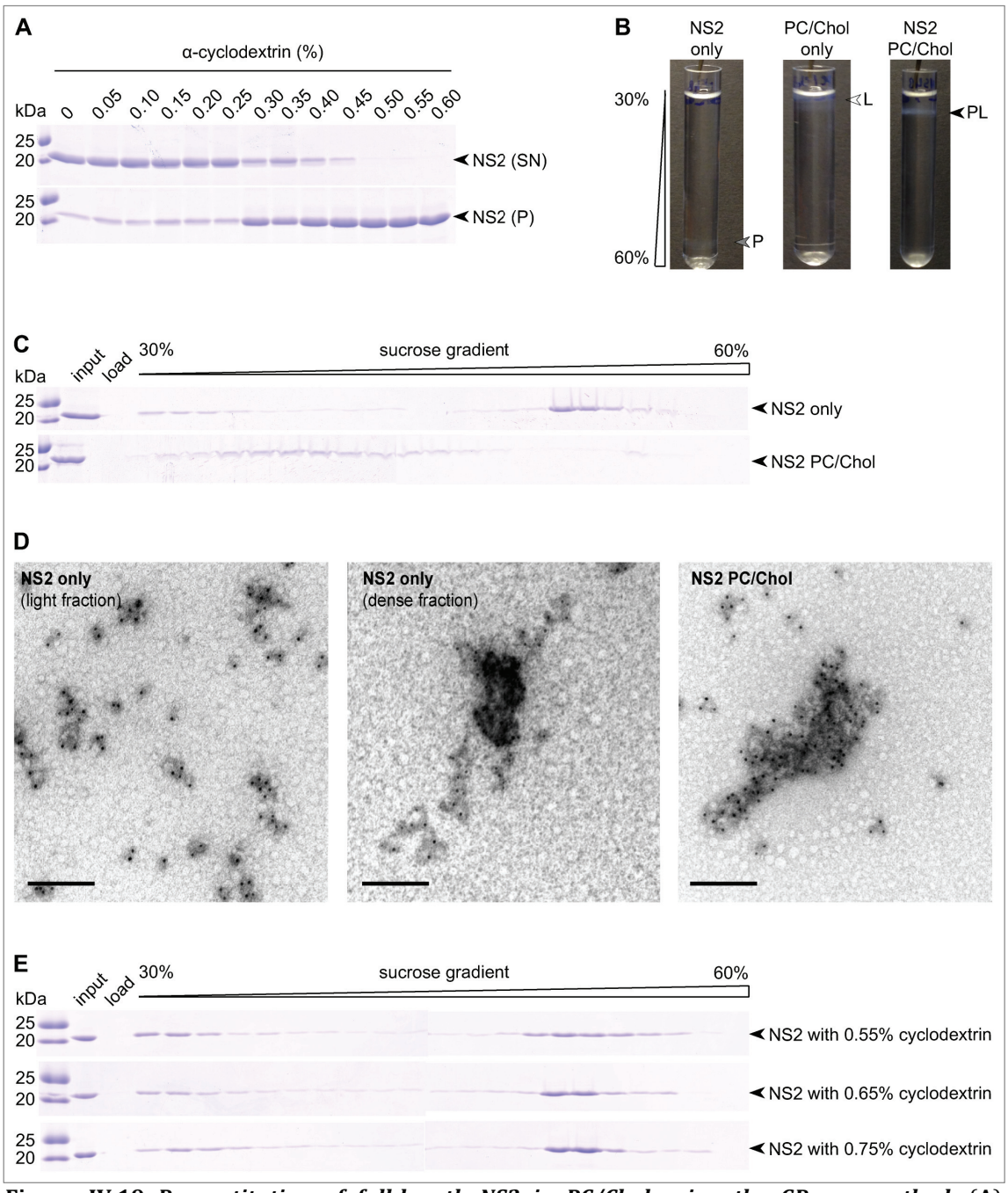


Figure IV.19 ***Reconstitution of full-length NS2 in PC/Chol using the GRecon method.*** (**A**) Determination of the optimal α-cyclodextrin concentration to precipitate NS2 (HCV strain JFH-1) by removal of DDM. The supernatant (SN, top panel) and the pellet (P, bottom panel) obtained after centrifugation of the precipitated samples were analyzed by SDS-PAGE. Black arrowheads indicate the bands corresponding to NS2. (**B**) GRecon gradients after centrifugation at 200,000 *g* for 16 h. Sucrose solutions were prepared in 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 1 mM DTT. A grey arrowhead indicates the bands corresponding to aggregated NS2 (annotated P), an empty arrowhead indicates the band corresponding to PC/Chol only (annotated L) and a black arrowhead indicates the band correponding to proteoliposomes (annotated LP). (**C**) Fractions from the GRecon gradients shown in (B) were analyzed by SDS-PAGE. Black arrowheads indicate the bands corresponding to NS2. (**D**) Electron microscopy analysis of fractions from the GRecon gradients shown in (B). Negative staining was performed after immunogold labeling (10 nm colloidal gold suspension) with an antibody against the *Strep*-tag II. Scale bar, 200 nm. (**E**) Precipitation test of NS2 directly on GRecon gradients using different α-cyclodextrin concentrations. Samples were analyzed by SDS-PAGE. Black arrowheads indicate the bands corresponding to NS2.

Like previously, samples were analyzed by EM by David Paul (post-doc in Ralf Bartenschlager's lab, Heidelberg University). Immunogold labeling was performed using an antibody against the *Strep*-tag II. As expected, the PC/Chol only sample did not show gold labeling. The lighter fraction of NS2 only (without lipids) exhibited a high degree of labeling which was kind of evenly dispersed all over the grid (**Figure IV.19D**, left panel). In contrast, the denser fraction of NS2 only showed material condensed in some electron dense structures which stained positive and are likely protein aggregates (**Figure IV.19D**, middle panel). The NS2 PC/Chol sample showed many dispersed gold particles like the NS2 lighter fraction and in addition some bigger clusters of gold with more electron density which looked like some protein/lipid/detergent mixtures. No well-defined liposomes or membrane structures could be observed (**Figure IV.19D**, right panel).

Although these results are encouraging, an important optimization work still has to be done to reconstitute membrane proteins in lipids using this method. Denis Lacabanne, PhD student in the lab, and Clément Danis, master student of whom I am in charge of, currently optimize the reconstitution of the ABC transporter BmrA using the GRecon method. BmrA was chosen as a model protein for optimization because it can be easily produced in large amounts in bacteria, and a well-established protocol allowing reproducible and functional membrane reconstitution by classical methods is available. The insight gained from the pilot study on BmrA will then be used to guide the parameter choice for the GRecon of other proteins of interest.

3.8. Stabilization of full-length NS2 using amphipol A8-35

Amphipols are amphipathic polymers with both hydrophobic and hydrophilic moieties. They are poorly efficient for direct solubilization of membrane proteins for which detergents must be used, but they allow to stabilize membrane proteins in detergent-free solutions. Moreover, because of their acidic nature, amphipols can be used to co-precipitate intact proteins when the pH of buffer solution is lowered. Indeed, when the carboxylate groups get protonated, a sharp decrease in the solubility and precipitation of amphipols is observed (Ning *et al*, 2014). In addition, divalent cations such as Ca^{2+} can lead to amphipol precipitation and represent therefore another way to

precipitate protein/amphipol complexes for solid-state NMR studies (Picard *et al*, 2006; Diab *et al*, 2007). Amphipols could thus be an interesting alternative to detergents for structural NMR studies.

3.8.1. Full-length NS2 trapping using A8-35

The trapping procedure consists in replacing the detergent by an amphipol, here the commercial amphipol A8-35. Briefly, an amphipol-to-protein ratio (APR) of 5 (w/w) was applied and detergent was removed using Bio-Beads with a ratio Bio-Beads-to-detergent of 20 (w/w) (refer to section III.10 for technical details). After trapping, the samples were centrifuged for 30 min at 20,000 *g*, 4°C and analyzed by SDS-PAGE. As shown in **Figure IV.20A**, NS2 in 0.1% DDM without addition of Bio-Beads remained in the supernatant after centrifugation, confirming that the protein was well solubilized. In contrast, NS2 in 0.1% DDM incubated with Bio-Beads without A8-35 fully precipitated and aggregated on the Bio-Beads, showing that the detergent was correctly removed by the Bio-Beads. Moreover, when NS2 in 0.1% DDM was incubated with A8-35 and further with Bio-Beads, NS2 was found in the supernatant, meaning that the trapping was successful, protein/amphipol complexes being soluble.

Preliminary analyses were performed in order to characterize the NS2/A8-35 complexes. Size exclusion chromatography showed a major peak containing NS2 (**Figure IV.20B**), while the following minor peak could be assigned to reagents present in the *Strep*-Tactin elution buffer, as already described for NS2 in 0.1% DDM. The NS2/A8-35 complexes eluted with an apparent molecular mass of about 173 kDa (**Figure IV.20D**), indicating the oligomeric nature of full-length NS2 in complex with the amphipol A8-35. In contrast to full-length NS2 in 0.1% DDM, the peak is large and asymmetric which means that the NS2/A8-35 oligomers are somehow heterogenous.

Furthermore, the far UV CD spectrum of NS2/A8-35 complexes after size exclusion chromatography is typical of a well-folded protein and is similar to the one obtained for NS2 in 0.1% DDM. In **Figure IV.20C**, the CD spectrum of the NS2/A8-35 complexes is represented in red, whereas the one of NS2 in 0.1% DDM is drawn in black for comparison. Although the shape of the two spectra is similar, the molar ellipticity per residue is different. At this point, it is unclear whether this difference is due to the amphipol or simply to some problem with the protein concentration measurement.

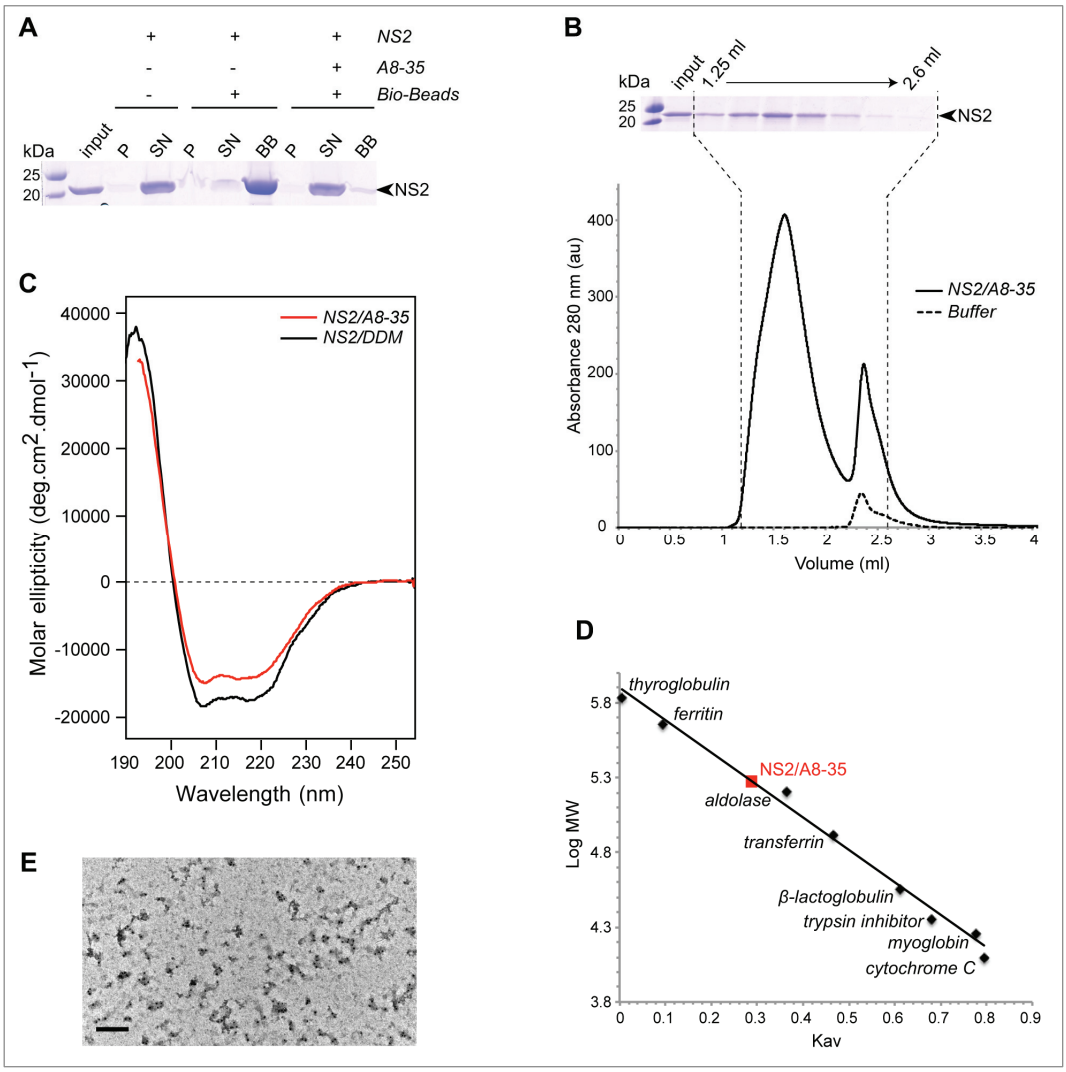


Figure IV.20 ***Characterization of NS2/A8-35 complexes.*** (**A**) Trapping of full-length NS2 (HCV strain JFH-1) using A8-35. Protein samples were analyzed by SDS-PAGE. P, pellet; SN, supernatant; BB, Bio-Beads. A black arrowhead indicates the bands corresponding to NS2. (**B**) Size exclusion chromatography of NS2/A8-35 complexes (conditions as detailed in the caption of Figure IV.13). The elution profile of NS2/A8-35 is drawn in a solid line, and that of the affinity elution buffer in a dotted line. SDS-PAGE and Comassie staining analysis of the collected fractions is shown above the chromatographic profile. Elution volume of NS2/A8-35 is 1.59 mL. (**C**) Far UV circular dichroism (CD) spectrum of NS2/A8-35 complexes is shown in *red line*. For comparison, a far UV spectrum of NS2 in DDM is shown in *black line*. (**D**) Calibration curve of the size exclusion chromatography column established using protein standards without detergent (V_0 is 1.25 mL). NS2/A8-35 complexes are indicated in *red* on the standard curve. (**E**) Transmission electron microscopy analysis of NS2/A8-35 after size exclusion chromatography. Negative staining was performed after immunogold labeling (10 nm colloidal gold suspension) with an antibody against the *Strep*-tag II. Scale bar, 200 nm.

Indeed, an error of 1% for the concentration leads already to important changes in the molar ellipticity per residue. Circular dichroism analysis of the NS2/A8-35 complexes should therefore be repeated to better characterize them. Like NS2 solubilized in detergent, NS2/A8-35 complexes form multimers, which can be visualized by EM using

immunolabeling of the N-terminal *Strep*-tag II fusion protein (David Paul, post-doc in Ralf Bartenschlager's lab in Heidelberg, **Figure IV.20E**).

3.8.2. Precipitation of NS2/A8-35 complexes

Both NS2 in 0,1% DDM and NS2/A8-35 complexes were first centrifuged for 15 h at 200,000 *g*, 4°C. The pellet and the supernatant obtained by ultracentrifugation were analyzed by SDS-PAGE (**Figure IV.21A**). In the presence of detergent, NS2 remained soluble and was found almost exclusively in the supernatant. In complex with the amphipol A8-35, NS2 also remained soluble, although a small portion of the sample was detected in the pellet, possibly corresponding to aggregates which might be avoided while working with a higher APR. Therefore, ultracentrifugation did not allow to sediment the NS2/A8-35 complexes. Nevertheless, an ultracentrifugation step would allow to remove the few aggregates present in the sample before precipitation by addition of $CaCl_2$ or pH decrease, allowing the preparation of better quality samples for solid-state NMR.

The addition of either 5 mM, 10 mM or 20 mM was tested to precipitate the NS2/A8-35 complexes : 5 µL of a ten times concentrated $CaCl_2$ solution were added to 25 µL NS2/A8-35. The samples were incubated on ice for 30 min without agitation and centrifuged for 1 h at 20,000 *g*, 4°C. The pellet and the supernatant obtained by centrifugation were analyzed by SDS-PAGE (**Figure IV.21B**). Only a minor part of the sample was detected in the pellet after addition of 10 mM or 20 mM $CaCl_2$ and the protein remained soluble at all tested $CaCl_2$ concentrations. For comparison, 12 mM $CaCl_2$ were reported to lead to the precipitation of a large portion of bacteriorhodopsin/A8-35 complexes (about 80% after 15 min and over 90% after 4 h) (Diab *et al*, 2007). At this point, precipitation tests by addition of $CaCl_2$ were not followed up. Not only higher $CaCl_2$ concentrations could be tested, but also longer incubation times and higher centrifugation speeds. Moreover, high salt concentrations could also lead to protein/amphipol complex precipitation. For example, amphipol tend to precipitate above 300 mM NaCl (Workshop « Applications of Amphipols to Membrane Protein Studies, October 21-24, 2013, Paris), but such high salt concentrations could be detrimental for solid-state NMR studies.

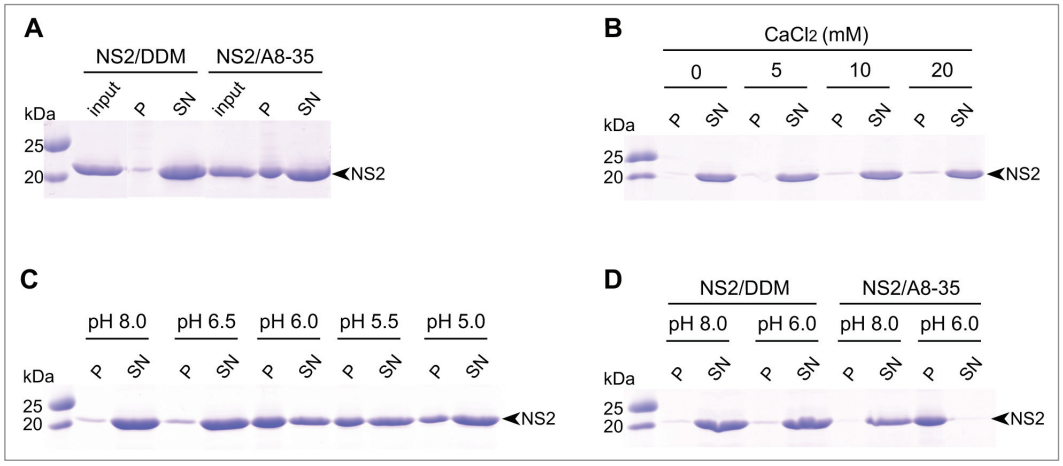


Figure IV.21 ***Precipitation tests of NS2/A8-35 complexes.*** (**A**) Sedimentation by ultracentrifugation of NS2 in 0.1% DDM (left) and NS2/A8-35 complexes with an APR of 5 (right) at 200,000 *g* for 14 h. (**B**) Precipitation of NS2/A8-35 complexes by addition of $CaCl_2$. (**C**) Precipitation of NS2/A8-35 complexes by changing the pH through addition of HCl. (**D**) Effect of lower pH on NS2 in 0.1% DDM (left) and NS2/A8-35 complexes (right). All panels : protein samples were analyzed by SDS-PAGE. P, pellet; SN, supernatant. A black arrowhead indicates the bands corresponding to NS2.

As described above, lowering the pH could lead to the precipitation of the NS2/A8-35 complexes which are initially in a Tris buffer at pH 8.0. The pH was modified by addition of 0.37% HCl to 50 µL NS2/A8-35. The obtained pH was estimated using pH-paper, the samples were incubated on ice for 30 min without agitation and centrifuged for 1 h at 20,000 *g*, 4°C, and the pellet and the supernatant obtained were analyzed by SDS-PAGE, as described previously. As shown in **Figure IV.21C**, NS2 remained fully soluble at pH 6.5, whereas it precipitated partially at lower pH, precipitation being better at pH 6.0 than at pH 5.5 and pH 5.0. Further tests showed that pH 6.0 seemed to be optimal for NS2/A8-35 precipitation. In order to verify that the precipitation was mediated by the amphipol, the pH of both NS2 in 0.1% DDM and NS2/A8-35 complexes was lowered to 6.0 (instead of pH 8.0). The pH was modified by addition of 3.7% HCl (instead of 0.37% HCl previously) to 50 µL NS2/A8-35. Moreover, the samples were incubated at room temperature for 30 min on the wheel and centrifuged for 1 h at 20,000 *g*, 4°C. The pellet and the supernatant obtained were analyzed by SDS-PAGE (**Figure IV.21D**). Whereas NS2 remained soluble in the presence of DDM at pH 6.0, the protein in complex with the amphipol fully precipitated at this pH, showing that the precipitation was amphipol-mediated.

Further tests should be carried out. The amphipol-to-protein ratio could for example be optimized to have as less amphipol as possible in the NMR rotor, while

keeping the protein in a fully soluble state. Nevertheless, these results, as an alternative to protein reconstitution in lipids, open a new way for the preparation of labeled samples for solid-state NMR. It is also worth to note that protein/amphipol complexes can be used for further lipid reconstitution when the protein is unstable in detergent (Workshop « Applications of Amphipols to Membrane Protein Studies, October 21-24, 2013, Paris).

3.9. NS2^{pro}

NS2^{pro} is the cytosolic protease domain of NS2 (refer to section II.6.4. for details). Two different NS2^{pro} constructs (HCV strains Con1 and JFH-1, GenBank accession numbers AJ238799 and AB047639, respectively) were successfully cloned in the pEU-E01-MCS vector with a *Strep*-tag II fused at the N-terminus, as described in section III.1. The amino acid sequences of these constructs are given in Appendix 1.

Cell-free expression of these two constructs was first tested at small scale in the precipitate mode. The CFS as well as the pellet and the supernatant obtained after centrifugation of the CFS were analyzed by both SDS-PAGE and western blotting with an antibody against the *Strep*-tag II. Before analysis, the supernatant was incubated with magnetic beads coated with *Strep*-Tactin to capture the *Strep*-tag II fused to each construct. Whereas NS2^{pro} Con1 was successfully expressed, the expression level of NS2^{pro} JFH-1 was very low and sometimes even nil (**Figure IV.22A**). In the precipitate mode, both constructs were mainly detected in the pellet. This was not surprising since NS2^{pro} was shown to associate with the membrane surface (Lange *et al*, 2014). In addition, a weak band was also observed in the supernatants, probably due to the presence of some residual lipids in the WGE. NS2^{pro} Con1 produced in the precipitate mode could be successfully solubilized with 1% DDM for purification by affinity chromatography. However, purified protein yield was extremely low (0.05 mg/mL WGE).

In addition, expression of these two constructs was also tested in the presence of detergent (MNG-3 or C12E8). NS2^{pro} Con1 was successfully expressed directly in a solubilized form (Fogeron *et al*, 2015). In contrast, the expression level of NS2^{pro} JFH-1

could not be improved by the use of detergent. Although purification of NS2^{pro} Con1 expressed in the presence of MNG-3 by affinity chromatography allowed us to improve final yield compared to NS2^{pro} Con1 expressed in the precipitate mode, purified protein yield was disappointing (0.16 mg/mL WGE). And more problematic, bands which could correspond to contaminants but most likely to degradation products were observed on the Coomassie gel (**Figure IV.22B**). Since NS2^{pro} has a protease activity, autoproteolysis could indeed occur. Taking this into account, expression of these two contructs was tested in the presence of MNG-3 and protease inhibitor (Complete EDTA-free, Roche, Ref. 04693152001). However, for both NS2^{pro} Con1 and NS2^{pro} JFH-1, expression level was approximately half lower in the presence of protease inhibitor (data not shown). Moreover, NS2^{pro} Con1 was expressed in the presence of MNG-3 and the addition of either 5 mM EDTA or protease inhibitor (Complete EDTA-free, Roche, Ref. 04693152001) to the CFS prior purification by affinity chromatography was tested. Compared to the standard condition, comparable purified protein yields were obtained (data not shown). Although slightly less degradation products seemed to appear on the Coomassie gel while using protease inhibitors, final yields were low and the bands on the Coomassie gels thus weak, making any conclusion hazardous.

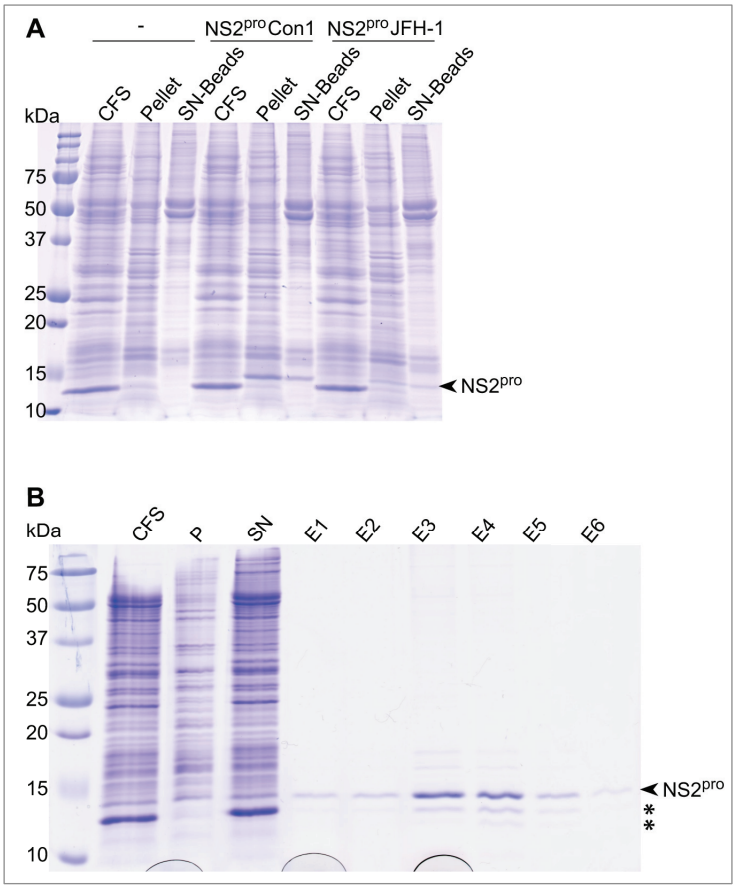


Figure IV.22 ***Production of NS2^{pro}.*** (**A**) Expression test of NS2^{pro} Con1 and NS2^{pro} JFH-1 in the precipitate mode. CFS, total cell-free sample; Pellet, pellets obtained after centrifugation of the CFS; SN-Beads, sample enriched in NS2^{pro} by incubation of CFS supernatant with *Strep*-Tactin magnetic beads to capture tagged NS2^{pro}. Comparable amounts were loaded on the gel for Pellet and SN-beads. A black arrowhead indicates the bands corresponding to NS2^{pro}. (**B**) Purification of NS2^{pro} Con1 expressed using 250 µL WGE in the presence of MNG-3. CFS, total cell-free sample**;** P, pellet obtained after centrifugation of CFS; SN, supernatant obtained after centrifugation of CFS and loaded on a 200-µL *Strep*-Tactin column; E1 to E6, elution fractions. All buffers contained 0.1% DDM. Protein samples were analyzed by SDS-PAGE. A black arrowhead indicates NS2^{pro}. Additional bands which could correspond to degradation products are indicated by an asterix.

At this step, the study of NS2^{pro} was not followed up. Adding a thrombine cleavage site between the *Strep*-tag II and NS2^{pro} might allow to achieve higher expression yields, as it seems to be the case for NS2, but for unknown reason, cloning of such constructs failed to date. Moreover, working with mutants inhibiting protease activity would avoid autoproteolysis.

3.10. Conclusion and perspectives

NS2 is an important protein in the life cycle of the hepatitis C virus, and is believed to play manifold roles in the progress of virus assembly. Parts of its structure could be studied using different biophysical approaches such as x-ray crystallography for NS2^{pro} and solution-state NMR of synthetic peptides for its membrane binding domain. However, more detailed studies of the full-length protein are missing, as active full-length NS2 has not been available due to the difficulties encountered for overexpression. So far, N-terminally truncated protein has been used to functionally characterize the protease activity *in vitro*, and since the protein was produced in bacteria, refolding from inclusion bodies was mandatory (Thibeault *et al*, 2001). Here, full-length NS2 contructs from the HCV strains JFH-1, Con1, H77 and 452 were successfully expressed in the WGE-CF system, both in the precipitate mode and in the presence of MNG-3. In particular, full-length NS2 (HCV strain JFH-1) could be expressed in a functional detergent-solubilized form. It is worth to note that purified full-length NS2 was active and underwent autoproteolysis in the absence of EDTA or chelating resin. Electron microscopy and size exclusion chromatography analyses showed that the protein is oligomeric. In addition, circular dichroism showed that it is well folded, with a structural content in agreement with previous structural analyses, especially of the membrane domain using synthetic peptides, which mostly revealed the presence of α-helical segments.

Although the analysis of the cytosolic domain NS2^{pro} has proved to be more difficult than expected, probably mainly due to autoproteolysis, full-length NS2 is very stable as long as the sample contains a chelating compound such as EDTA. Preliminary reconstitution tests could thereby be performed giving very encouraging results and opening the way for the preparation of isotopically labeled samples in view of solid-state NMR studies. Indeed, while expression in cell-free systems in the presence of liposomes

might allow for direct membrane insertion of the produced proteins, detergent solubilization and subsequent reinsertion into membranes allowed us to obtain higher protein yields and purer preparations. In addition, this approach allows the control of the lipid-to-protein ratio that should be as low as possible with the aim to have much more protein than lipids in the NMR rotor. Successful insertion into membranes is also a good indicator for the structural integrity of the membrane protein.

The direct expression in soluble form circumvents many difficulties linked to purification from inclusion bodies, notably the presence of non-native forms upon refolding and protein loss. While many studies have addressed interactions between NS2 with several other viral proteins *in cellulo*, no *in vitro* assays on these interactions could be carried out in order to gain knowledge on the molecular mechanisms. Indeed, these interactions presumably take place *via* the hydrophobic membrane-inserted N-terminal region (Jirasko *et al*, 2010; Popescu *et al*, 2011), and their occurrence might as well rely on the correct interaction of the protein complex with the membrane. An experimental set-up targeting the identification of these interactions at a molecular level is thus difficult with truncated forms, and as well might be unsuccessful in detergent-solubilized states. The availability of full-length NS2 in a purified, detergent-solubilized or membrane-bound form, or even in complex with amphipol, in sufficient quantities, opens the perspective for biochemical *in vitro* structural and functional studies of this enigmatic protein, including structural studies which aim at analyzing the protein in its membrane environment by means of electron microscopy and solid-state NMR. Furthermore, it will allow the probing of interactions with other viral proteins which can also be successfully produced using the WGE-CF system (Fogeron *et al*, 2015).

A publication including some of the results presented in this section, and entitled "Functional expression, purification, characterization, and membrane reconstitution of nonstructural protein 2 from hepatitis C virus" has been submitted to Protein Expression and Purification and is currently under review. The manuscript is presented in Appendix 2.

4. NS4B

NS4B is a 27 kDa integral membrane protein involved in HCV replication complex formation : it is thought to induce the formation of the membranous web, a specific membrane alteration that serves as a scaffold for the replication complex. NS4B is predicted to be composed of four transmembrane passages, as well as amphipathic helices interacting in-plane of the membrane interface (refer to section II.6.6. for details). The aim of this work is to produce purified isotopically labeled full-length NS4B samples for structural studies.

4.1. NS4B expression tests

Six different NS4B constructs (HCV strain JFH-1, GenBank accession number AB047639) were initially cloned by Jérôme Gouttenoire, researcher in Darius Moradpour's lab in Lausanne, with a His tag fused at the C-terminus. These constructs correspond to the full-length protein (aa 1-260), the two N-terminal amphipathic α-helices (aa 1-72), the second N-terminal amphipathic α-helix with the first two predicted transmembrane segments (aa 40-130), the four predicted transmembrane segments (aa 72-196), the last two transmembrane segments (aa 129-196) and the two C-terminal α-helices (aa 198-260). Neither the construct corresponding to the two N-terminal amphipathic α-helices (aa 1-72) nor the one correponding to the two C-terminal α-helices (aa 198-260) could be expressed in the precipitate mode. In contrast, the expression of the four other constructs was successful (**Table IV.1**, 2/3). Although the expression of full-length NS4B was tested in the presence of detergents and lipids, the work on this construct was not followed up since the His tag does not allow an efficient purification of tagged proteins expressed in the WGE-CF system because of unspecific binding of several WGE proteins to nickel columns.

Three additional full-length NS4B constructs (HCV strains JFH-1, Con1 and H77, GenBank accession numbers AB047639, AJ238799 and AF009606, respectively) with both a *Strep*-tag II and a thrombine cleavage site fused at the C-terminus, as well as a full-length NS4B construct (HCV strain JFH-1) with the YFP (Yellow Fluorescent Protein), a *Strep*-tag II and a thrombine cleavage site fused at the C-terminus, were then cloned by Jérôme Gouttenoire. The expression of these constructs was initially tested in the precipitate mode. The CFS as well as the pellet and the supernatant obtained after

centrifugation of the CFS were analyzed by SDS-PAGE. Before analysis, the supernatant was incubated with magnetic beads coated with *Strep*-Tactin to capture the *Strep*-tag II fused at the C-terminus of each construct. Whereas the first three constructs were successfully expressed, full-length NS4B Con1 showing the highest expression level, no expression was detected for the construct with the YFP. In the absence of detergent, as expected due to its hydrophobic nature, NS4B is mainly insoluble and was therefore detected in the pellet, independently from the strain. As for NS2, a minor portion was found in the supernatant, possibly due to the presence of some residual lipids in the WGE. Moreover, NS4B from HCV strains Con1 and JFH-1 were successfully expressed directly in a solubilized form in the presence of MNG-3 (**Figure IV.23**, see also (Fogeron *et al*, 2015)). The expression of full-length NS4B from HCV strain H77 in the presence of MNG-3 has not been tested yet. The amino acid sequences of these three constructs are given in Appendix 1.

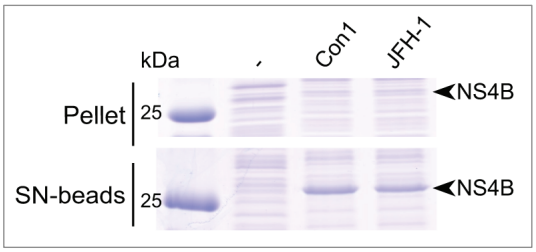


Figure IV.23 ***Small scale expression test of full-length NS4B constructs in the presence of MNG-3 detergent.*** Full-length NS4B contructs from HCV strains Con1 and JFH-1 were expressed in the presence of 0.1% MNG-3. Protein samples were analyzed by SDS-PAGE. -, negative control (no NS4B mRNA); Pellet, pellet obtained after centrifugation of CFS; SN-beads, sample enriched in NS4B by incubation of CFS supernatant with *Strep*-Tactin magnetic beads to capture tagged NS4B. Comparable amounts were loaded on the gel for Pellet and SN-beads. Black arrowheads indicate the bands corresponding to NS4B.

4.2. Full-length NS4B purification

4.2.1. Purification of full-length NS4B expressed in the precipitate mode

Before the expression in the presence of detergent had been optimized (refer to section IV.2.), full-length NS4B was successfully expressed in the precipitate mode. Briefly, full-length NS4B Con1 was solubilized after translation while adding 1% (w/v) DDM directly to the CFS treated with benzonase, as optimized for full-length NS2 (refer to section IV.3.3.1.), and purified by affinity chromatography on a *Strep*-tactin column (**Figure IV.24**). On average, the purified NS4B yield was around 0.5 mg/mL WGE. This work was in part performed together with Loick Lancien, master student of whom I was in charge of.

Size exclusion chromatography showed that the purified NS4B sample was homogenous without aggregates and the apparent molecular mass was estimated to be around 58 kDa (data not shown). Moreover, the total mass was analyzed by mass spectrometry by Frédéric Delolme, engineer at the Protein Science Facility at IBCP, Lyon. While the expected mass was 29,139 Da, a peak at 29,147 Da was obtained. Given the precision of the method, these values are consistent. Furthermore, full-length NS4B Con1 purified by affinity chromatography and size exclusion chromatography was also analyzed by circular dichroism, showing that the protein was well folded with a large proportion of α-helical content. Thermal denaturation was also carried out, showing that the protein is stable up to 40°C (data not shown).

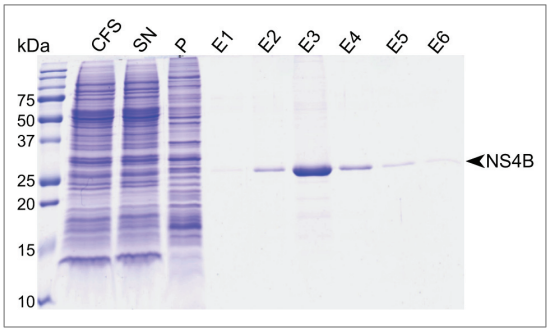


Figure IV.24 ***Purification of full-length NS4B by affinity chromatography.*** Full-length NS4B (HCV strain Con1) was expressed in the precipitate mode using 1.5 mL WGE, solubilized with 1% DDM and purified on a 1-mL *Strep*-Tactin column in the presence of 0.1% DDM. Protein samples were analyzed by SDS-PAGE. CFS, total cell-free sample; P, pellet obtained after centrifugation of CFS; SN, supernatant obtained after centrifugation of CFS and loaded on the affinity column; E1 to E6, affinity elution fractions. The black arrowhead indicates the bands corresponding to NS4B.

4.2.2. Purification of full-length NS4B expressed in the presence of MNG-3 detergent

As for NS2, after expression in the presence of MNG-3, the *Strep*-tag II allowed us to obtain pure NS4B protein using a single affinity purification step starting from the supernatant fraction containing the solubilized protein (**Figure IV.25A**). Although full-length NS4B could be efficiently purified in the presence of MNG-3 (Fogeron *et al*, 2015), the addition of 0.25% DDM to the CFS treated with benzonase before the affinity purification step allowed us to exchange MNG-3 to DDM. Indeed, DDM is the detergent standardly used in the lab for further reconstitution in lipids, and there is moreover no commercially available deuterated MNG-3. On average, the purified protein yield was about 1 mg/mL WGE, with a maximum yield of 1.35 mg NS4B per mL WGE being reached recently. To further characterize purified NS4B protein, we performed size exclusion chromatography. A first peak containing NS4B was observed (**Figure IV.25B**), while the following peak is due to reagents present in the *Strep*-Tactin elution buffer. NS4B protein-detergent complexes eluted with an apparent molecular mass of about 50

kDa (**Figure IV.25C**), indicating that NS4B could be a monomer. The results presented here were obtained together with Clément Danis, master student of whom I am in charge of, and are consistent with results obtained previously.

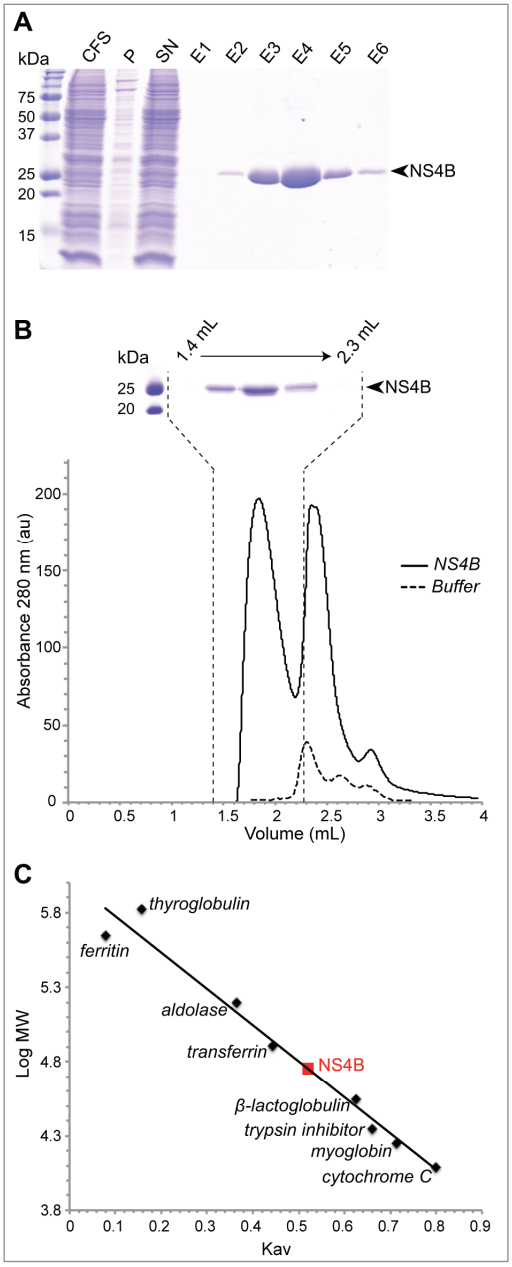


Figure IV.25 ***Purification of full-length NS4B by affinity chromatography and analysis by size exclusion chromatography.*** (**A**) Full-length C-terminally *Strep*-tag II tagged NS4B (HCV strain JFH-1) was expressed using 4.5 mL WGE in the presence of 0.1% MNG-3 and purified on a 5-mL *Strep*-Tactin column in the presence of 0.1% DDM. Protein samples were analyzed by SDS-PAGE (conditions as detailed in the caption of Figure IV.13). CFS, total cell-free sample; P, pellet obtained after centrifugation of CFS; SN, supernatant obtained after centrifugation of CFS and loaded on the affinity column; E1 to E6, affinity elution fractions. The black arrowhead indicates the bands corresponding to NS4B. (**B**) Size exclusion chromatography of E4 elution fraction (conditions as detailed in the caption of Figure IV.13). The elution profile of NS4B is drawn in a solid line, and that of the affinity elution buffer in a dotted line. SDS-PAGE and Comassie staining analysis of the collected fractions is shown above the chromatographic profile. Elution volume of NS4B is 1.85 mL. (**C**) Calibration curve of the size exclusion chromatography column established using protein standards (V_0 is 1.25 mL). NS4B is indicated in *red* on the standard curve.

Moreover, the total mass of the NS4B JFH-1 sample after purification by affinity chromatography was analyzed by mass spectrometry by Frédéric Delolme, engineer at the Protein Science Facility at IBCP, Lyon. While the expected mass was 29,229 Da, a peak at 29,109 Da was obtained, which corresponds to full-length NS4B after N-terminal methionine removal. Furthermore, an additional band with a molecular weight around

60 kDa was often, but not systematically, observed on Coomassie gels after purification and/or lipid reconstitution. This band was identified by mass spectrometry after tryptic digestion as being NS4B, and corresponds therefore to a NS4B dimer (data not shown).

4.3. Analysis of full-length NS4B by circular dichroism

The far UV circular dichroism (CD) spectrum of full-length NS4B eluted from size exclusion chromatography in 0.1% DDM (refer to previous section) is typical of a well-folded protein, as shown in **Figure IV.26** (done together with Loick Lancien). Additionally, as for NS2, the molar ellipticity per residue is in the range that can be expected for such a protein. The two minima at 208 and 222 nm, together with a maximum at 192 nm, are typical of α-helical folding. Spectral deconvolution indicated an α-helix content of 72-79%, while turns represent 7-10%, and β-sheet content is almost nil (0-1%). The high α-helical content is consistent with the presence of two amphipathic α-helices at the N-terminus, two α-helices at the C-terminus, as well as four predicted transmembrane segments, which most probably adopt a α-helical structure.

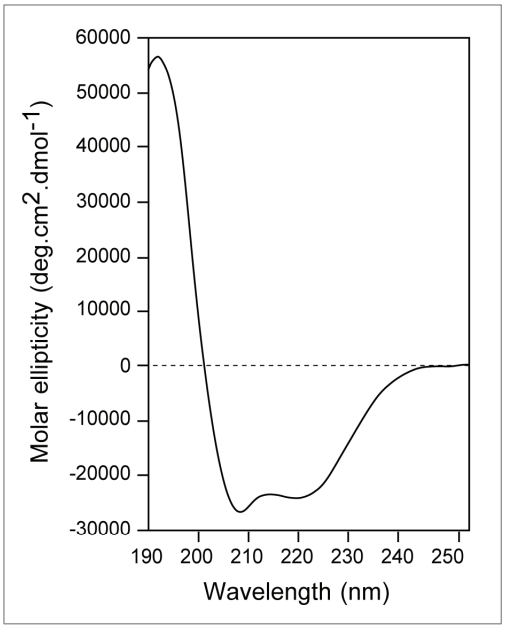


Figure IV.26 ***Far UV circular dichroism (CD) spectrum of full-length NS4B after purification by affinity chromatography and size exclusion chromatography*** (conditions as detailed in the caption of Figure IV.14A). Estimation of the secondary structure content resulted in an α-helical content of 72-79%, 7-10% turns, and 0-1% β-sheet.

4.4. Analysis of full-length NS4B by solution NMR

As described for full-length NS2 in section IV.3.4., uniformly $^{13}C/^{15}N$ labeled full-length NS4B (HCV strain Con1) was produced in the presence of MNG-3 using 3 mL WGE

to prepare a solution NMR sample (done together with Loick Lancien). In contrast to NS2, NS4B was expressed in the precipitate mode because its expression in the presence of detergent had not been optimized yet. The CFS was treated with benzonase and 1% non-deuterated DDM was added to solubilize the protein. NS4B was purified from the supernatant obtained after centrifugation of the CFS on four 1-mL *Strep*-Tactin columns. The whole purification process was performed in the presence of 0.1% DDM and tail deuterated DDM was used starting from the *Strep*-Tactin washing steps. In total, 1.7 mg purified NS4B were obtained (*i.e.* purified protein yield of 0.57 mg/mL WGE). The most concentrated elution fractions were pooled (690 µg NS4B in 1 mL) and dialyzed against 50 mM phosphate buffer pH 6.0. A dialysis membrane with a MWCO of 6-8 kDa was used and dialysis was performed over 3 h, the dialysis buffer being changed every hour. In these conditions, the loss of detergent by dialysis is limited. The protein was then concentrated overnight at 4°C, directly in the dialysis bag packed in Sephadex G25 powder for liquid adsorption. After concentration, the NS4B sample was collected and centrifuged at 20,000 *g*, 4°C for 10 min (a precipitate appeared during the third dialysis step, likely due to some loss of detergent). At the end, 275 µL supernatant were available to which 5 mM (7 µL) tail deuterated DDM and 5% (14 µL) D_2O were added. Final sample concentration was 1.69 mg/mL, there was therefore 500 µg protein in this sample (**Figure IV.27A**).

Measurements were done on a 1 GHz spectrometer at the CRMN (Centre de Résonance Magnétique Nucléaire à très hauts champs) in Lyon by Moreno Lelli, engineer at the CRMN and Roland Montserret, engineer in the lab. A 2D TROSY (Transverse Relaxation Optimized SpectroscopY) spectrum is shown in **Figure IV.27B**. Whereas the protein is composed, together with the tag, of 277 amino acids (refer to Appendix 1 for full amino acid sequence), only few residues were detected. Moreover, as for NS2, the proton dispersion was very narrow between 7 and 8.5 ppm and the few peaks observed likely correspond to the unfolded, flexible tag (19 residues in total). As for NS2, NS4B/detergent micelles are most probably too large to get a sufficient tumbling and thus can not be successfully analyzed by solution NMR. In order to analyze NS4B in a native-like membrane lipids environment using solid-state NMR methods, reconstitution of the protein in lipids has been optimized, as described in the following section.

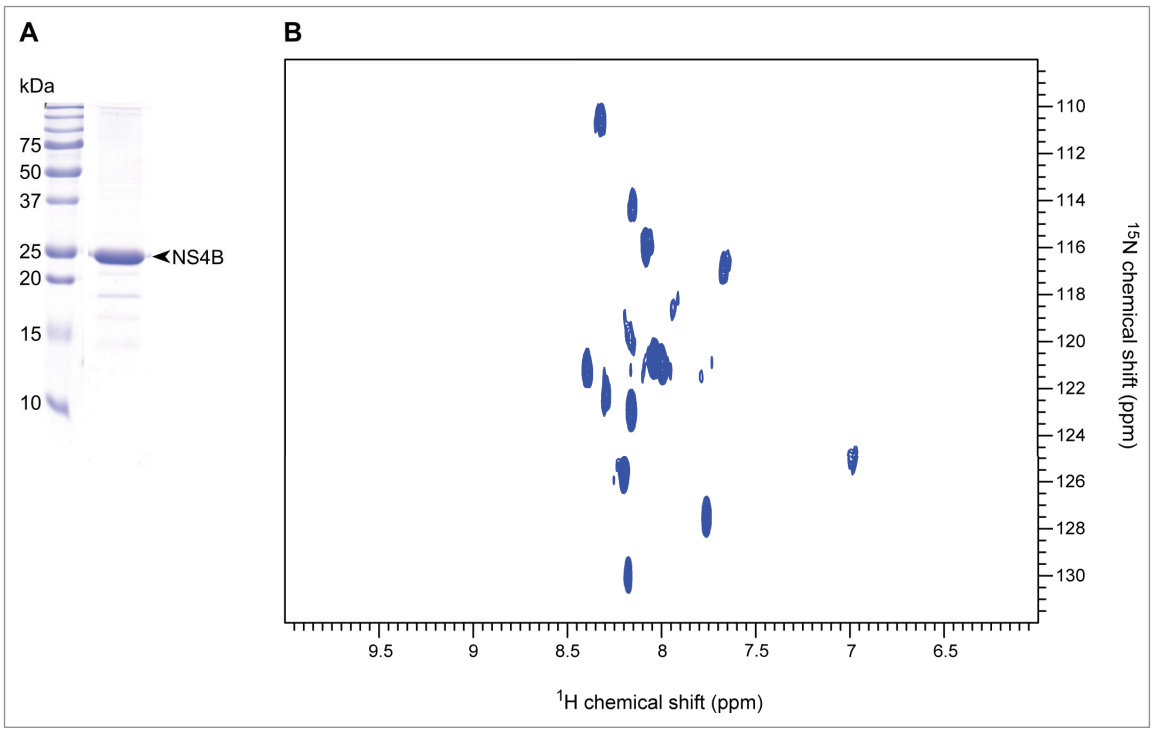


Figure IV.27 ***Analysis of full-length NS4B by solution NMR.*** (**A**) Purified uniformly $^{13}C/^{15}N$ labeled NS4B (HCV strain Con1) for solution NMR analysis. The protein (1.69 mg/mL) was in 50 mM phosphate buffer pH 6.0 containing 5 mM tail deuterated DDM and 5% D_2O. (**B**) TROSY spectrum recorded on a 1GHz spectrometer at the CRMN in Lyon (TGIR-RMN-THC Fr3050 CNRS). Number of scans : NS=24, recycling delay : D1=1s, acquisition time : AQ=0.08s/scan, total acquisition time : 1h50.

4.5. Reconstitution of full-length NS4B in lipids

Preliminary small scale reconstitution tests

As described previously for NS2 (refer to section IV.3.7.1.), a first reconstitution test was performed (together with Loick Lancien) with three different types of lipids : egg L-α -phosphatidylcholine (PC), a mixture of egg PC and cholesterol (PC/Chol, 70/30, w/w) and home-made lipids from pig liver, at LPRs of of 10, 20, 100 and 800 (mol/mol). Bio-Beads were progressively added directly to the protein/lipid/detergent mixtures. As for NS2, EM analysis was not concluding, suggesting that the protein could have aggregated on the Bio-Beads (data not shown). Further tests were therefore performed using the dialysis method in small dialysis cups fitting in a 24-well plate, and adding the Bio-Beads in the dialysis buffer. Several reconstitution tests were performed this way with PC, PC/Chol and home-made lipids from pig liver at LPR 0.5, 1 and 2 (w/w). In contrast to NS2, NS4B was not lost during the reconstitution process. Moreover, EM analysis of

these samples, which was performed by Philippe Roingeard and Sonia Georgeault at the Plate-Forme RIO des Microscopies in Tours, showed that multilayers dense to electrons were present for reconstituted samples whereas only structures clear to electrons were observed for lipids alone (data not shown). Moreover, the reconstitution seemed to be better with the PC/Chol mixture. Although no immunogold labeling was performed, the multilayers dense to electrons, together with the SDS-PAGE analysis (**Figure IV.28A**, top panel), were a good indication for a at least partially successful reconstitution. These results therefore encouraged us to go on with the optimization of the reconstitution using a PC/Chol mixture.

Reconstitution tests with high LPR

Reconstitution with high LPR was tested by David Paul, post-doc in Ralf Bartenschlager's lab in Heidelberg. Indeed, David came to Lyon to learn how to produce detergent-solubilized NS4B in the WGE-CF system and reconstituted the protein in lipids in Heidelberg following protocols given by François Penin. Briefly, David worked with full-length NS4B from the strain Con1 and used egg L-α-phosphatidylcholine (PC) at a LPR of 20 (w/w). Reconstitution was performed either with preformed liposomes or with detergent-solubilized PC. In both cases, electron microscopy using immunolabeling of the C-terminal *Strep*-tag II fusion protein showed branching tubular structures, clearly pointing out a NS4B-mediated membrane rearrangement (data not shown). These results were fully consistent with the role of NS4B in the formation of the membranous web.

Reconstitution tests with low LPR

Although highly encouraging results were obtained with high LPR, solid-state NMR studies ask for a minimum amount of lipids used for reconstitution in view of maximizing the signal-to-noise ratio. Considering the previous results, a PC/Chol mixture was used and reconstitution was performed at larger scale using the dialysis method. In short, 160 µg (285 µL at 0.35 mg/mL) of NS4B Con1 purified by affinity chromatography in 0.1% DDM were reconstituted in PC/Chol (70/30, w/w) using a LPR of either 0.5, 1 or 2 (w/w). The PC/Chol mixture was solubilized with DDM, a detergent-to-lipid ratio of 10 (mol/mol) being used. The protein/lipid/detergent mixture was dialyzed at RT for 6 days (MWCO 6-8 kDa). Bio-Beads were added to the dialysis buffer,

with a Bio-Beads-to-detergent ratio of 100 (w/w). Reconstituted samples were centrifuged and the pellet and the supernatant obtained were analyzed by SDS-PAGE, showing that the protein was not lost during the reconstitution process (**Figure IV.28A**, bottom panel), and that reconstitution efficiency appeared to be higher than at small scale (preliminary tests). Electron microscopy analysis was performed by Philippe Roingeard and Sonia Georgeault at the Plate-Forme RIO des Microscopies in Tours. Reconstituted samples were pelleted by centrifugation and embedded in resin prior to negative staining on sections. As expected, NS4B alone showed aggregates after reconstitution (**Figure IV.28B**, left panel), whereas PC/Chol alone showed structures clear to electrons (**Figure IV.28B**, middle panel). Multilayers dense to electron were observed for NS4B reconstituted in PC/Chol (**Figure IV.28B**, right panel). Although no immunogold labeling was done on the samples, the multilayers dense to electron were a good indicator for efficient reconstitution. This reconstitution process was therefore applied for the preparation of a first NS4B sample for solid-state NMR (refer to the following section).

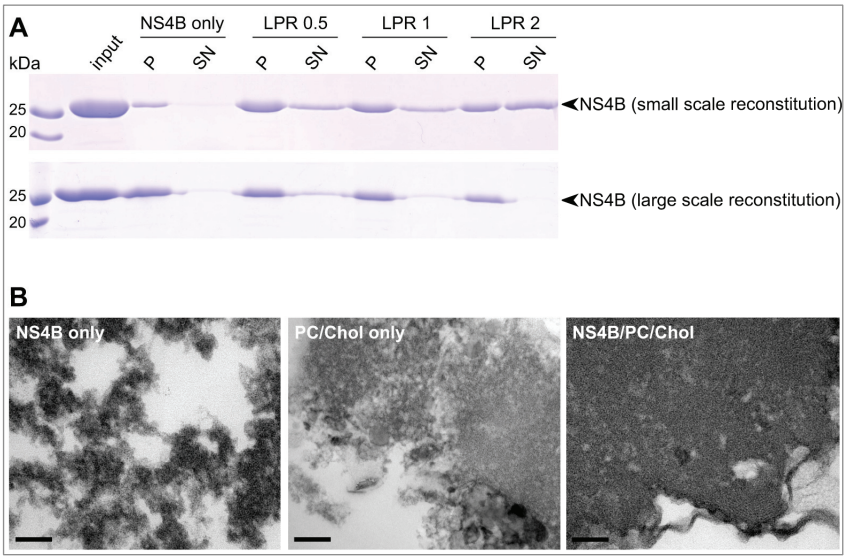


Figure IV.28 ***Reconstitution of full-length NS4B in lipids.*** (**A**) Full-length NS4B (HCV strain Con1) was reconstituted in PC/Chol (70/30, w/w). Either small scale (top panel) or large scale (bottom panel) reconstitution tests were performed. (**B**) Electron microscopy analysis of reconstituted samples. Negative staining was performed using uranyl acetate (no immunogold labeling). Scale bar, 100 nm.

4.6. Analysis of full-length NS4B by solid-state NMR

Uniformly $^{2}H/^{13}C/^{15}N$ labeled full-length NS4B (HCV strain JFH-1) was produced in the presence of MNG-3 using 6 mL WGE for solid-state NMR analysis. This sample was prepared together with Vlastimil Jirasko (post-doc in Ralf Bartenschlager's lab in Heidelberg) who came to Lyon to learn how to produce isotopically labeled membrane proteins using the WGE-CF system for solid-state NMR studies. Briefly, the CFS was treated with benzonase and 0.25% DDM were added. NS4B was purified from the supernatant obtained after centrifugation of the CFS on a 5-mL *Strep*-Tactin column. The whole purification process was performed in presence of 0.1% DDM. At the end, 1 mg purified NS4B was obtained in the most concentrated elution fraction, which corresponds to an unusually low purified protein yield of 0.16 mg/mL WGE. Nevertheless, NS4B was reconstituted in PC/Chol (70/30, w/w) with a LPR of 0.5 (w/w) using a classical dialysis protocol, as described previously. After reconstitution, the sample was centrifuged at 20,000 *g*, 4°C for 1 h. The pellet and the supernatant were analyzed by SDS-PAGE, showing that there was no significant protein loss during the reconstitution process and that most of the sample was in the pellet, which is a good indicator for a high reconstitution yield. The reconstituted sample which was filled in a 0.8 mm NMR rotor is shown in **Figure IV.29A**.

NMR measurements were done on a 850 MHz wide bore spectrometer at the ETH Zurich by Susanne Penzel, PhD student in Beat Meier's lab. A 2D HSQC ^{1}H-^{15}N correlation spectrum is shown in **Figure IV.29B**. The spectrum shows a signal dispersion which corresponds to a well-folded α-helical protein. The overlap is still important, which can be expected for such a large protein. The obtained signal-to-noise ratio is promising.

Vlastimil Jirasko then prepared another uniformly $^{2}H/^{13}C/^{15}N$ labeled full-length NS4B sample with some modifications. Indeed, reconstitution was performed with a LPR of 0.25 (instead of 0.5) and proteoliposomes were isolated by ultracentrifugation on two OptiPrep™ cushions (30% and 90%). Whereas the empty lipids float on the 30% cushion and the aggregated protein sediments under the 90% cushion, the liposomes are found between these two cushions. The proteoliposomes were pelleted in a larger 1.3 mm NMR rotor. The HSQC spectrum, which was also recorded on a 850 MHz wide bore spectrometer by Susanne Penzel at the ETH Zurich, was very similar to the one presented here, with a slightly worse resolution since the sample was spun at 60 kHz

instead of 90 kHz with the 0.8 mm NMR rotor. The use of a larger rotor allowed the recording of a 2D DREAM spectrum, showing isolated resonances with linewidths below 1 ppm. Intense signals corresponding to the major amino acids, Ala, Val, Leu and Ile which make up nearly 40 % of the protein in α-helical conformation dominate the spectrum (data not shown). The very good quality of this spectrum evidences that the obtained sample is homogenous and clearly indicates that the work described in this section has the potential to become an important venue in solid-state NMR sample preparation.

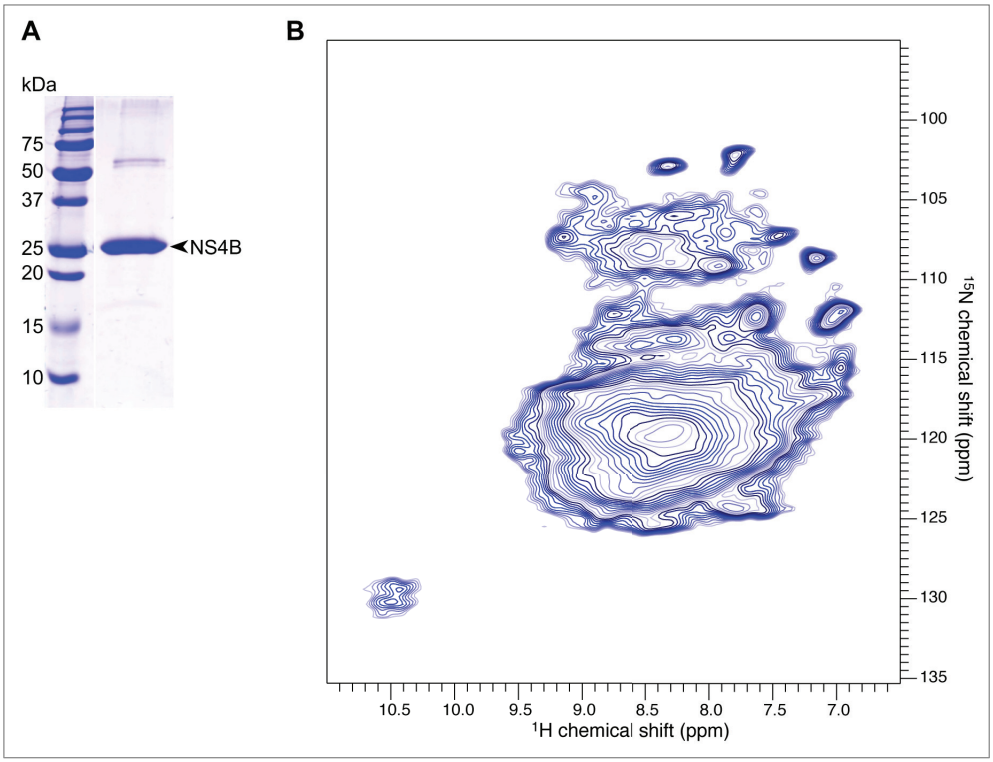


Figure IV.29 ***Analysis of full-length NS4B by solid-state NMR.*** (**A**) Purified uniformly $^{2}H/^{13}C/^{15}N$ labeled NS4B (HCV strain JFH-1) was reconstituted in PC/Chol (70/30, w/w) with a LPR of 1 (w/w) for solid-state NMR analysis. (**B**) HSQC spectrum recorded on a 850 MHz spectrometer at the ETH Zurich. Number of scans : 448, recycling delay : 0.8s, acquisition time : 70ms/scan, total acquisition time : 32 h.

4.7. Conclusion and perspectives

During this thesis, full-length NS4B constructs from HCV strains JFH-1, Con1 and H77 were successfully expressed in the WGE-CF system in the precipitate mode, with the possibility to solubilize them with DDM for further purification by affinity chromatography. Moreover, full-length NS4B constructs from HCV strains JFH-1 and

Con1 could be expressed directly in a MNG-3 detergent-solubilized form for further purification by affinity chromatography. Very high purity levels could be achieved in only one purification step. Moreover, size exclusion chromatography analyses showed that the protein is homogenous. In addition, circular dichroism showed that NS4B is well folded, with a high α-helical content in agreement with available structural analyses. Reconstitution of NS4B in lipids proved to be efficient and first solid-state NMR spectra are very promising.

The work on NS4B was performed from the beginning in collaboration with David Paul and Vlastimil Jirasko, post-docs in Ralf Bartenschlager's lab in Heidelberg, and Jérôme Gouttenoire, researcher in Darius Moradpour's lab in Lausanne. David, Vlastimil and Jérôme came to Lyon to learn with me how to produce NS4B in the WGE-CF system and to implement this system in their respective laboratories. Thanks to this approach, Jérôme could produce soluble NS4B from HCV strain JFH-1 for immunisation and an antibody against this protein is now available. David performed reconstitution tests and showed clearly that NS4B rearranged the membranes. Vlastimil is now going to prepare further NS4B samples with different isotopic labeling strategies to improve the quality of the NMR spectra.

The successful expression of the protein as described in this work thus resulted in a better characterization of full-length NS4B and opens the way towards the resolution of its three-dimensional structure which would allow to gain more insight into NS4B still unelucidated multiple functions that might be governed by distinct membrane topologies and/or interactions with other viral and cellular proteins.

A publication including the results presented here is currently in preparation.

5. NS5A

NS5A is a 56 kDa membrane-associated phosphoprotein exhibiting RNA-binding properties and interacting with the NS5B polymerase and the core protein. Although its function in the HCV life cycle remains to be elucidated, current evidence indicates that NS5A likely functions as a « molecular switch » between replication and assembly. NS5A is anchored to the membrane via an amphipathic N-terminal α-helix (AH) and composed of three domains denoted D1, D2 and D3. While D1 is well structured, recent structural studies demonstrated that D2 and D3 are intrinsically unfolded (refer to section II.6.7. for details). The aim of this work is to produce purified isotopically labeled NS5A samples for structural NMR studies.

5.1. NS5A expression test

Four different NS5A constructs (HCV strain Con1, GenBank accession number AJ238799) were successfully cloned in the pEU-E01-MCS vector with a *Strep*-tag II fused at the C-terminus (as described in section III.1.). These constructs correspond to D1 (aa 30 to 213), D1D2D3 (aa 30 to 447), AH-D1 (aa 1 to 213) and AH-D1D2D3 (aa 1 to 447, full-length NS5A). The amino acid sequences of these constructs are given in Appendix 1.

Cell-free expression of these constructs was tested at small scale in the precipitate mode. The CFS as well as the pellet and the supernatant obtained after centrifugation of the CFS were analyzed by both SDS-PAGE and western blotting with an antibody against the *Strep*-tag II. Before analysis, the supernatant was incubated with magnetic beads coated with *Strep*-Tactin to capture the *Strep*-tag II fused to each construct. All four constructs were successfully expressed.

As expected due to their hydrophobic nature, AH-D1 and full-length NS5A were mainly detected in the pellet as shown in **Figure IV.30A**. However, a band was also observed in the supernatants (**Figure IV.30B**). The presence of some residual lipids in the WGE could explain why these constructs are partially expressed in a solubilized form. The constructs D1 and D1D2D3 lack the amphipathic N-terminal α-helix and are therefore expected to be expressed in a soluble form. Indeed, they were mainly detected in the supernatant (**Figure IV.30B**). However, a weak band was detected in the pellet for NS5A D1D2D3 on the Coomassie gel, and for both constructs in western blotting (data not shown). This expression test was done at the very beginning of this work and no

benzonase treatment was done on these samples. Since NS5A is known to interact with nucleic acids, these two constructs could partially form large and insoluble RNA/protein complexes. However, latter work showed that expression of D1D2D3 in the presence of MNG-3 combined with benzonase treatment could not completely avoid the formation of such complexes (Fogeron *et al*, 2015).

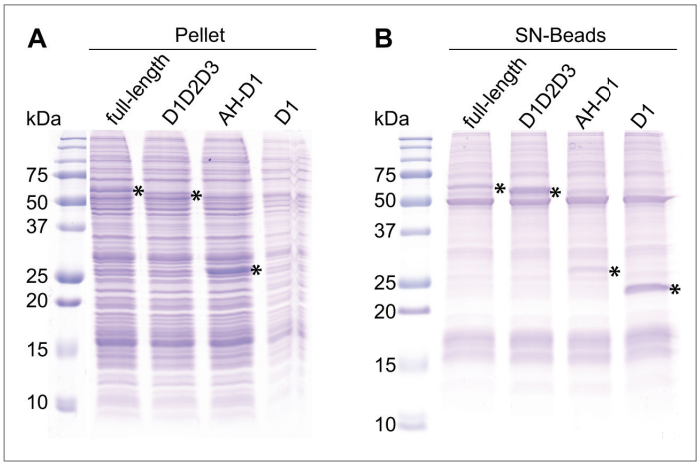


Figure IV.30 ***Small scale expression test of the NS5A constructs in the precipitate mode.*** (**A**) Pellets obtained after centrifugation of the CFS. (**B**) Samples enriched in NS5A by incubation of CFS supernatant with *Strep*-Tactin magnetic beads to capture tagged NS5A constructs (SN-Beads). Comparable amounts were loaded on the gel for Pellet and SN-beads. Protein samples were analyzed by SDS-PAGE. The bands corresponding to the proteins of interest are indicated by an asterix.

As NS2 (refer to section IV.3.1.), D1 and D1D2D3 were expressed using various isotopically labeled amino acid mixtures : $^{13}C/^{15}N$, $^{2}H/^{15}N$ and $^{2}H/^{13}C/^{15}N$. Whereas $^{13}C/^{15}N$ and $^{2}H/^{15}N$ labeled aa mixtures seemed not to affect the expression level of these constructs compared to a standard unlabeled aa mixture, yield was slightly lower with the $^{2}H/^{13}C/^{15}N$ labeled aa mixture. AH-D1 and full-length NS5A were also expressed using $^{13}C/^{15}N$ labeled amino acids, without affecting the expression yield (data not shown).

Finally, full-length NS5A from HCV strain JFH-1, with a *Strep*-tag II fused at its C-terminus, was successfully expressed too and mainly detected in the pellet in the precipitate mode (collaboration with Sonia Assil and Marlène Dreux, ENS Lyon). Expression of full-length NS5A JFH-1 with a *Strep*-tag II fused at the N-terminus was also tested, but without success.

5.2. NS5A purification

NS5A D1 and D1D2D3 are soluble constructs which can be directly purified by affinity chromatography after their synthesis in the precipitate mode (**Figure IV.31**). However, we experienced that adding 1% DDM to the CFS before purification leads to higher purity levels. Purified protein yields are usually between 0.5 and 1 mg/mL WGE for these constructs. Moreover, both purified D1 and D1D2D3 can be partially sedimented by overnight ultracentrifugation, allowing their study by solid-state NMR (refer to section IV.5.3). Indeed, D1 and D1D2D3 have been reported to form large soluble assemblies, experiments such as dynamic light scattering (DLS) or small angle X-ray scattering (SAXS) indicating a heterogenous oligomeric state (Love *et al*, 2009).

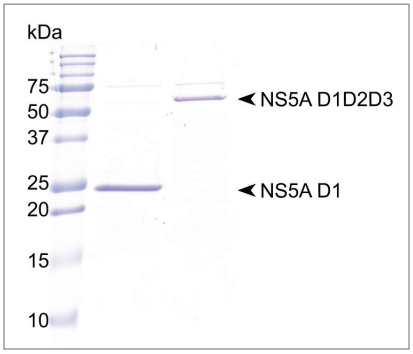


Figure IV.31 ***Purified NS5A D1 and D1D2D3.*** Both constructs were expressed in the precipitate mode. Supernatant obtained after centrifugation of CFS was loaded on a *Strep*-Tactin column. Protein samples were analyzed by SDS-PAGE. Black arrowheads indicate the bands corresponding to the proteins of interest.

AH-D1 and full-length NS5A were mainly detected in the pellet after expression in the precipitate mode. These two proteins can however be solubilized with detergent after their synthesis. Indeed, addition of 1% DDM to the CFS allows for solubilization. Although solubilization is not complete, using higher DDM amounts could interfere with further purification, leading to lower purified protein yields. Purification of these two constructs by affinity chromatography is shown in **Figure IV.32**. DDM was added in washing and elution buffers at a concentration of 0.1%. Purified protein yields up to 0.30 and 0.40 mg/mL WGE for AH-D1 and full-length NS5A, respectively, have already been obtained. However, the purified protein yield is on average ~ 0.25 mg/mL WGE for both constructs, which is lower than for the corresponding soluble constructs. Although parameters such as WGE quality for example can impact purified protein yields, especially protein expression directly in a solubilized form in the presence of detergent (*e.g.* MNG-3) might improve them significantly. In contrast to D1 and D1D2D3, AH-D1 and full-length NS5A seem very difficult to sediment, probably due to the presence of

detergent. In addition, sedimentation most probably depends on the concentration of the samples.

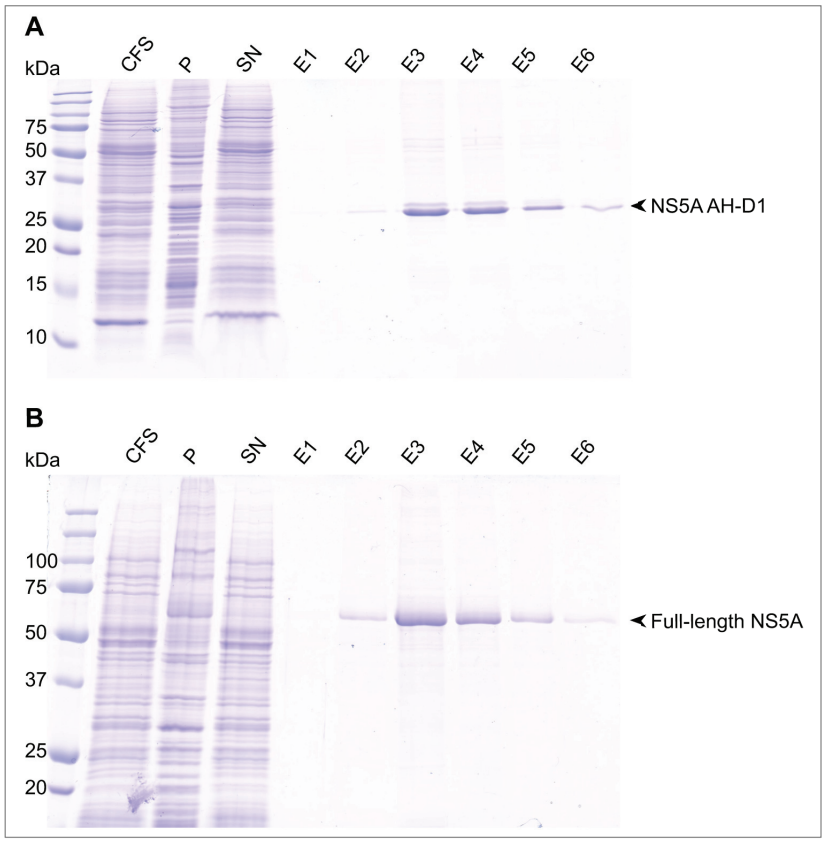


Figure IV.32 ***Small scale purification of AH-D1 and full-length NS5A samples.*** (**A**) NS5A AH-D1 and (**B**) Full-length NS5A were expressed using 250 µL WGE in the precipitate mode, solubilized with 1% DDM and purified on a 200-µL *Strep*-Tactin column in the presence of 0.1% DDM. These samples are uniformly $^{13}C/^{15}N$ labeled. CFS, total cell-free sample**;** P, pellet obtained after centrifugation of CFS; SN, supernatant obtained after centrifugation of CFS and loaded on the affinity column; E1 to E6, elution fractions. All buffers contained 0.1% DDM. Protein samples were analyzed by SDS-PAGE. A black arrowhead indicates the band corresponding to the protein of interest.

5.3. First solid-state NMR spectra of sedimented NS5A D1D2D3

As a proof-of-concept, about 5 mg uniformly $^{13}C/^{15}N$ labeled NS5A D1D2D3 were produced for analysis by solid-state NMR. D1D2D3 was expressed in the precipitate mode using 9 mL WGE, in the presence of an algal $^{13}C/^{15}N$ labeled amino acid mixture. The CFS was treated with benzonase. The supernatant obtained after centrifugation of the CFS was loaded on a *Strep*-tactin column for purification by affinity chromatography (**Figure IV.33**). Purified protein yield was ~ 0.55 mg/mL WGE.

Ultracentrifugation of elution fractions for 16 h at 200,000 *g* allowed us to sediment the protein. Elution fractions E2 to E5 were used for sedimentation, giving a total protein amount of ~ 4.7 mg. After ultracentrifugation, the supernatants were analyzed by SDS-PAGE and absorbance at 280 nm was measured. The protein amount in the supernatants was estimated to be ~ 1.2 mg. Deductively, the protein amount in the pellets was estimated to be ~ 3.5 mg. The pellets were then filled in a 3.2 mm NMR rotor by centrifugation using a filling tool. This step was a delicate issue because the pellets tended to stick to the wall of the tubes and were therefore very difficult to transfer into the NMR rotor. A large amount of NS5A D1D2D3 was probably lost during this process. The NMR rotor was stored at 4°C until the NMR measurements.

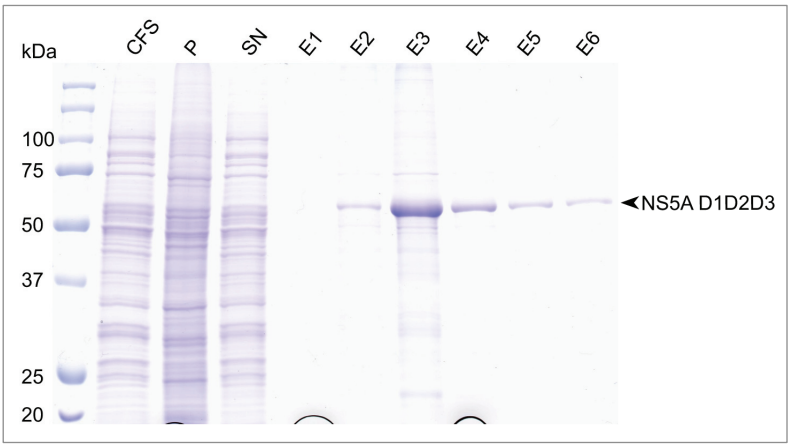


Figure IV.33 ***Large scale purification of an uniformly $^{13}C/^{15}N$ labeled NS5A D1D2D3 for solid-state NMR analysis.*** NS5A D1D2D3 was expressed using 9 mL WGE in the precipitate mode and purified on a 5-mL *Strep*-Tactin column in the absence of DDM. CFS, total cell-free sample, treated with benzonase (420 units/mL, incubation for 1 h at room temperature); P, pellet obtained after centrifugation of CFS; SN, supernatant obtained after centrifugation of CFS and loaded on the affinity column; E1 to E6, elution fractions. Protein samples were analyzed by SDS-PAGE. A black arrowhead indicates the band corresponding to NS5A D1D2D3.

NMR measurements were done on a 800 MHz wide bore spectrometer shared between the IBCP and the CRMN (Centre de Résonance Magnétique Nucléaire à très hauts champs) in Lyon by Carole Gardiennet, researcher in the lab. Both direct pulsed and cross-polarized (CP) spectra were recorded (**Figure IV.34**). The 1D CP spectrum of NS5A D1D2D3 showed promising line-widths (0.3-0.4 ppm) and a good dispersion, typical for a folded protein. The presence of additional narrow resonances in the directly pulsed spectrum pointed to the presence of mobile parts in the protein. In contrast to

the core protein (refer to section IV.6.3.), no large amounts of RNA were observed in the directly pulsed spectrum, indicating that the treatment with benzonase allowed an efficient removal of excess RNA.

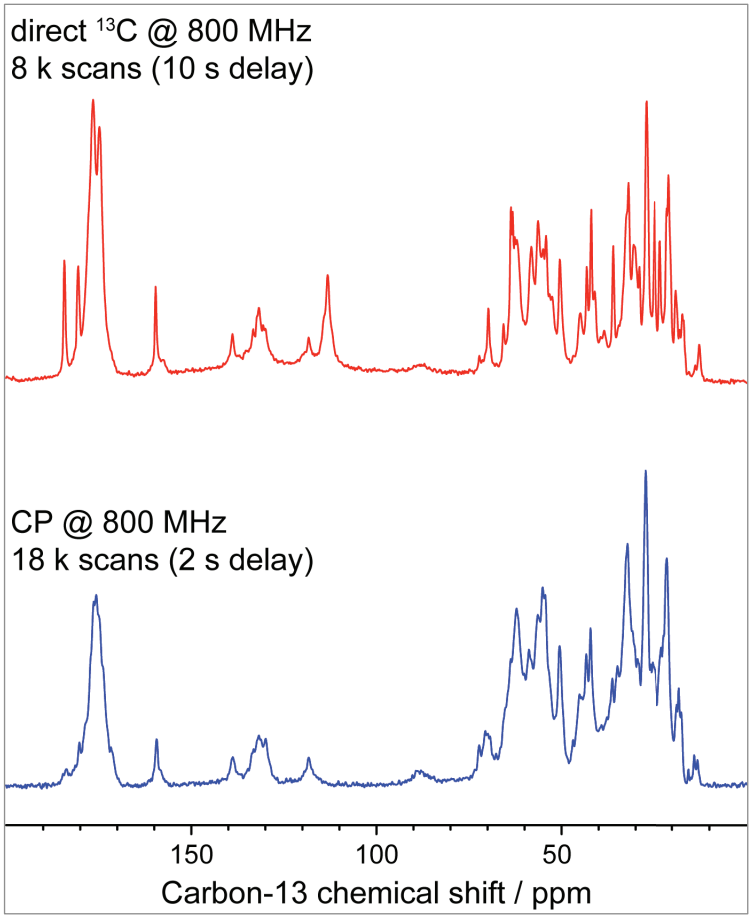


Figure IV.34 ***First solid-state NMR spectra of sedimented NS5A D1D2D3.*** Direct pulsed (top panel, in *red*) and cross-polarized (bottom panel, in *blue*) spectra were recorded on a 800 MHz wide bore spectrometer at the CRMN in Lyon.

5.4. Conclusion and perspectives

In summary, four NS5A contructs from the HCV strain Con1, corresponding to D1, D1D2D3, AH-D1 and full-length NS5A, are available, as well as a construct for full-length NS5A from the HCV strain JFH-1. All these constructs were successfully expressed in the WGE-CF system. D1 and D1D2D3 are soluble constructs, whereas the others have to be solubilized in detergent before purification when they are produced in the precipitate mode. Moreover, we showed that full-length NS5A can be expressed directly in a detergent-solubilized form (Fogeron *et al*, 2015).

D1 and D1D2D3 can be partially sedimented by ultracentrifugation which allow a simple way for sample preparation for their analysis by solid-state NMR. A first uniformly $^{13}C/^{15}N$ labeled NS5A D1D2D3 sample was prepared which gave promising results. In parallel, NS5A D1D2D3 was also analyzed by solution NMR. Samples were prepared by Stéphane Sarrazin, post-doc in the lab, and measurements were done in Lille by Xavier Hanoulle, researcher in Guy Lippens' lab. Interestingly, solution NMR revealed that NS5A D1D2D3 was phosphorylated. This was the first proof in our hands that phosphorylation of the synthesized proteins is possible in the WGE-CF system.

Sedimentation of AH-D1 and full-length NS5A for solid-state NMR studies seems to be hampered by the presence of detergent and the lower concentration of protein samples compared to D1 and D1D2D3. Reconstitution of these two constructs in lipids is therefore under optimization. I trained Alons Lends, PhD student in Beat Meier's lab, who is currently working on the production of reconstituted NS5A samples for solid-state NMR analysis.

Moreover, NS5A constructs produced in the lab using the WGE-CF system were used for further applications. For example, I trained a DUT student, Clément Lagrange, with whom I produced NS5A AH-D1 and full-length NS5A in the precipitate mode followed by solubilization in DDM and further purification, with the aim to analyze potential interactions between NS5A and annexin A2 by SPR. This project was carried out in collaboration with Sonia Assil and Marlène Dreux (ENS Lyon). SPR analyses were performed by Stéphane Sarrazin. More recently, AH-D1 and full-length NS5A were provided to Olivier Diaz and Patrice André (CIRI, INSERM, Lyon) for enzyme activity assay. With these samples, they could show that NS5A enhances the activity of hexokinase 2, leading to an increased glycolysis rate during HCV infection (Ramière *et al*, 2014).

6. Core

Mature HCV core is a 19 kDa protein composed of two domains: the N-terminus hydrophilic domain 1 (D1) which promotes viral nucleocapsid formation, and the C-terminus hydrophobic domain 2 (D2) which mediates association to the membrane interface of cellular lipid droplets (refer to section II.6.1. for details). The aim of this work is to produce purified isotopically labeled core samples (D1, D2 and full-length core) for structural studies.

6.1. Core expression test

Six different core constructs (HCV strain J, GenBank accession number D90208) were successfully cloned in the pEU-E01-MCS vector with a *Strep*-tag II fused at the C-terminus (as described in section III.1.). A first construct corresponds to D1 (aa 1 to 120), a second one to D2 (aa 116 to 171) and four others to the expected full-length mature core protein (aa 1 to 171, 1 to 175, 1 to 180 and 1 to 183). Indeed, as described in section II.6.1., the C-terminus of mature core after cleavage of the signal sequence between core and E1 by the signal peptide peptidase has not been identified precisely. This is the reason why four different constructs were prepared and analyzed for the full-length protein. The amino acid sequences of the constructs corresponding to D1 (aa 1 to 120) and full-length mature core (aa 1-175) are given in Appendix 1.

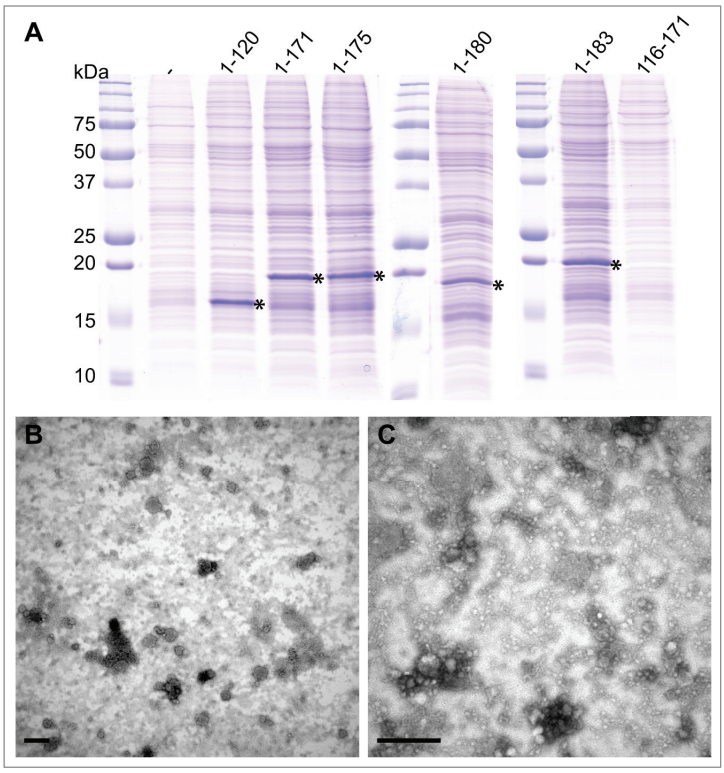


Figure IV.35 ***Small scale expression test of the core constructs in the precipitate mode.*** (**A**) Pellets obtained after centrifugation of the CFS were analyzed by SDS-PAGE. No benzonase treatment was done on these samples. -, negative control (no core mRNA). The bands corresponding to the proteins of interest are indicated by an asterix. (**B**, **C**) Transmission electron microscopy analysis of core 1-120 (**B**, core D1) and core 1-175 (**C**, full-length core). Negative staining was performed using phosphotungstic acid. Bar scale, 500 nm.

Cell-free expression of these constructs was tested at small scale in the precipitate mode. The CFS as well as the pellet and the supernatant obtained after centrifugation of the CFS were analyzed by both SDS-PAGE and western blotting with an antibody against the *Strep*-tag II. Constructs 1-120, 1-171, 1-175, 1-180 and 1-183 were successfully expressed. They were mainly detected in the pellet as shown in **Figure IV.35A**. In contrast, no expression of core 116-171 could be observed, neither on the Coomassie gel (**Figure IV.35A**) nor in western blotting.

The pellets obtained after centrifugation of the CFS were resuspended in buffer for their analysis by electron microscopy (EM) analysis. EM was performed at the Centre d'Imagerie Quantitative Lyon Est (CIQLE), together with Elisabeth Errazuriz, engineer at the CIQLE, and Stéphane Sarrazin, post-doc in the lab. Negative staining using phosphotungstic acid revealed that core expressed in the WGE-CF system spontaneously assembles into nucleocapsid-like particles, as shown for core 1-120 (D1) and core 1-175 (full-length core) in **Figure IV.35 B and C**, respectively.

6.2. Purification of capsid-like particles

Capsid-like particles formed during expression of core constructs can be isolated according to their density. Thanks to the specific isotopic labeling of expressed proteins, high purity of core samples might not be essential for NMR analysis. Part of this work consisted in optimizing the isolation of nucleocapsid-like particles on density gradients. Accudenz, a non-ionic tri-iodinated derivative of benzoic acid with three aliphatic hydrophilic side chains, was prefered to sucrose which could be conducive to bacterial growth in the NMR rotor upon long term storage. Moreover, a single Accudenz cushion was prefered to an Accudenz gradient for easier handling and better reproducibility, with the aim to pellet the nucleocapsid-like particles while contaminants would not go through the cushion. Accudenz was prepared in a buffer containing 10 mM Tris-Acetate pH 7.4, 100 mM NaCl, 0.62% NP-40, 50 mM potassium acetate and 4 mM magnesium acetate. The pellets obtained after centrifugation of the CFS were resuspended in this buffer (Pres) before being loaded on the Accudenz cushion. Not all optimization steps are presented here. As an example, different densities of Accudenz were tested for core D1 (1-120) (**Figure IV.36A**), showing that nucleocapsid-like particles go through a 30%

Accudenz cushion whereas many contaminants are eliminated. There are more contaminants left in the final sample (Pc) when lower Accudenz densities are used and the nucleocapsid-like particles float on the Accudenz cushion at higher densities. Treatment of CFS with benzonase was also tested, but nucleocapsid-like particles tend to be disrupted when nucleic acids are degraded, as shown in **Figure IV.36B**.

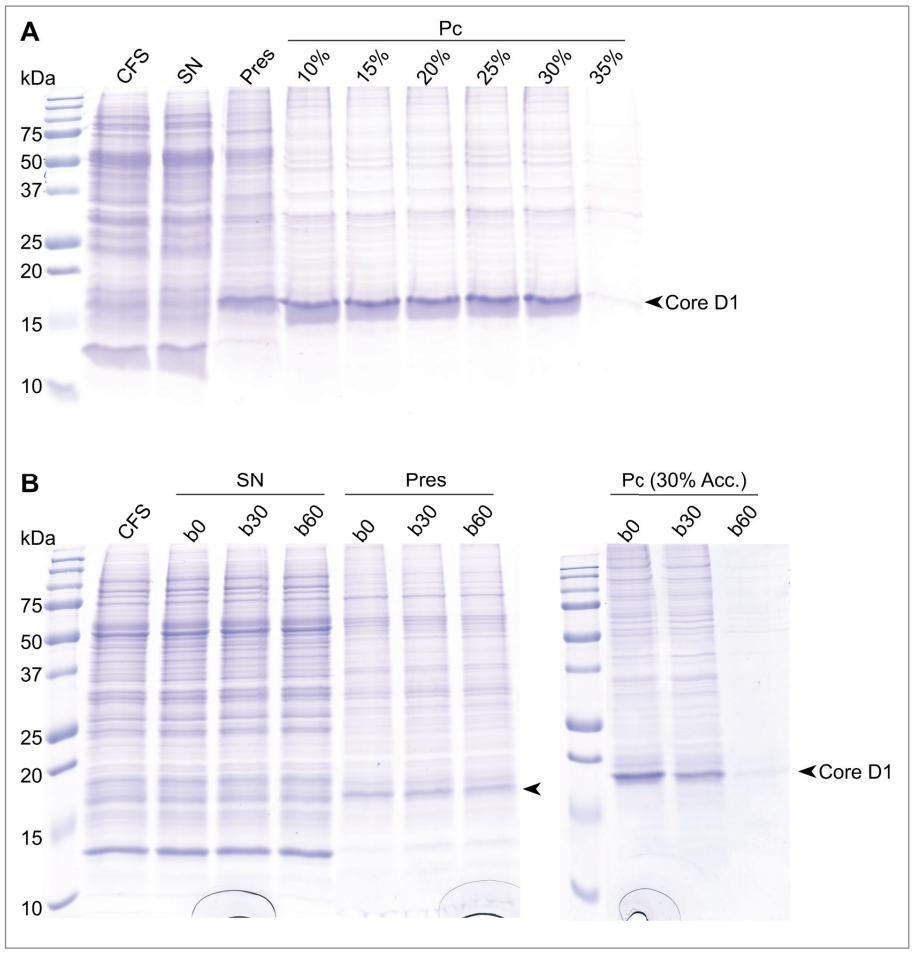


Figure IV.36 ***Examples of optimization steps for the isolation of core D1 nucleocapsid-like particles on Accudenz cushions.*** (**A**) Six different Accudenz cushion densities were tested (10%, 15%, 20%, 25%, 30% and 35%), 30% Accudenz giving the best result. (**B**) Analysis of the effect of benzonase treatment on the isolation of nucleocapsid-like particles. b0, no benzonase treatment; b30 and b60, addition of 625 units benzonase/mL and incubation at room temperature for 30 min and 60 min, respectively. (**A**) and (**B**) Protein samples were analyzed by SDS-PAGE. CFS, total cell-free sample; SN, supernatant obtained after centrifugation of CFS; Pres, pellet obtained after centrifugation of CFS and resuspended in Tris-Acetate buffer; Pc, pellet obtained after centrifugation of Pres loaded on an Accudenz cushion at 200,000 *g* (**A**) or 20,000 *g* (**B**) for 1 h. A black arrowhead indicates the band corresponding to core D1.

In addition, centrifugation of Accudenz cushions was performed at either 20,000 *g* or 200,000 *g*, showing that ultracentrifugation is not necessary to pellet nucleocapsid-like particles. In addition, loading directly the CFS on an Accudenz cushion resulted in much more contaminants in the final sample.

Isolation of nucleocapsid-like particles was optimized for full-length core (1-175) too. In addition, nucleocapsid-like particles formed by the capsid protein from dengue virus were analysed in parallel, given similar but slightly less convincing results (data not shown).

6.3. First solid-state NMR spectra

As a proof-of-concept, about 3 mg uniformly $^{13}C/^{15}N$ labeled core D1 were produced and analyzed by solid-state NMR. D1 was expressed in the precipitate mode using 9 mL WGE, in the presence of an algal $^{13}C/^{15}N$ labeled amino acid mixture. Core D1 was then partially purified by centrifugation on a 30% Accudenz cushion, as shown on **Figure IV.37**. No benzonase treatment was done on this sample.

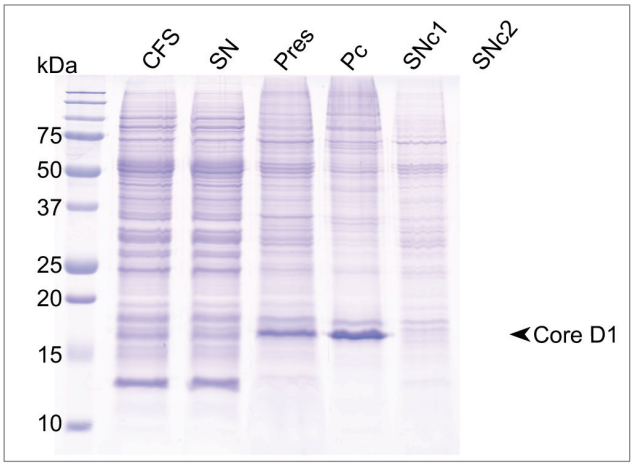


Figure IV.37 ***Production of an uniformly $^{13}C/^{15}N$ labeled core D1 sample for solid-state NMR analysis.*** CFS, total cell-free sample; SN, supernatant obtained after centrifugation of CFS; Pres, pellet obtained after centrifugation of CFS and resuspended in 10 mL buffer (*i.e.* 1/20 of initial CFS volume); Pc, pellet obtained after centrifugation of Pres loaded on a 30% Accudenz cushion at 20,000 *g* for 1 h; SNc1 and SNc2, supernatants obtained after centrifugation of Pres loaded on a 30% Accudenz cushion at 20,000 *g* for 1 h (SNc1, top corresponding to Pres and SNc2, bottom corresponding to the Accudenz cushion). Protein samples were analyzed by SDS-PAGE. A black arrowhead indicates the band corresponding to core D1.

The pellet obtained was shortly « washed » in Tris-Acetate buffer, this step removing part of the contaminants (data not shown) and then filled in a 3.2 mm NMR rotor by centrifugation at 4,500 *g* for 20 min, using a special filling tool. This rotor was stored at 4°C until the NMR measurements.

NMR measurements were done on a 850 MHz wide bore spectrometer at the ETH Zurich by Anne Schütz, PhD student in Beat Meier's lab. Both direct pulsed and CP spectra were recorded (**Figure IV.38**). The 1D CP spectrum is typical for a protein sample. The protein amount in the NMR rotor was estimated between 3 mg and 4 mg (*i.e.* protein yield between 0.3 mg and 0.4 mg/mL WGE). The directly pulsed spectrum presents a broad peak between 90 and 140 ppm, showing that large amounts of RNA were co-sedimented. 2D DARR spectra were also recorded (data not shown) but the signal-to-noise ratio was very low. One reason for this could be that the NMR rotor was not full. Also, the cross-polarization process seemed not very efficient, most probably because of protein dynamics. Lowering the temperature shall be an efficient approach to test this hypothesis.

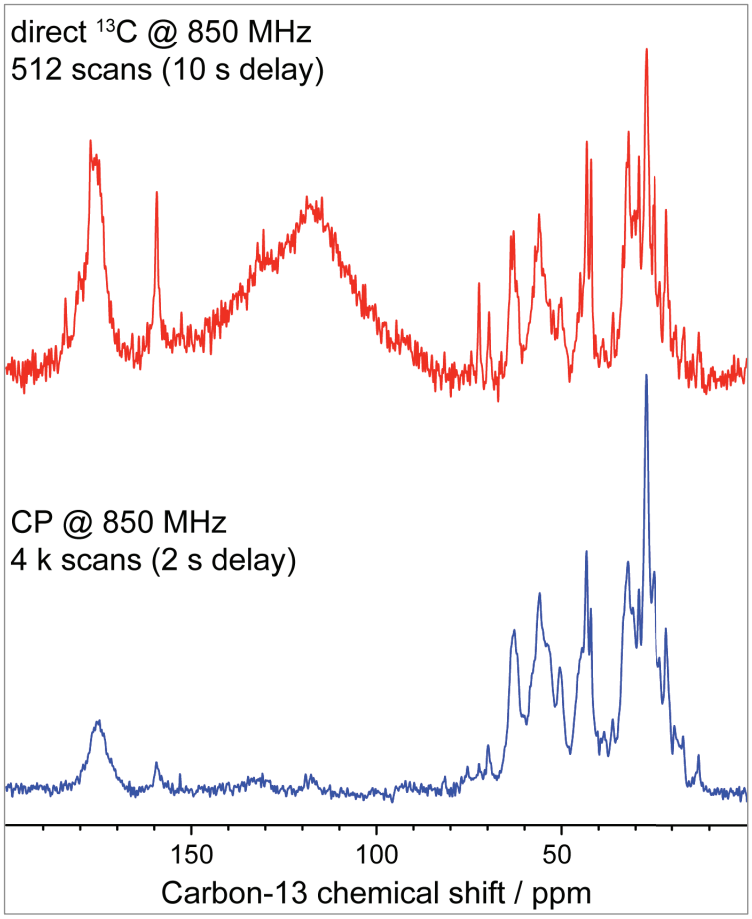


Figure IV.38 ***First solid-state NMR spectra of core D1.*** Direct pulsed (top panel, in *red*) and cross-polarized (bottom panel, in *blue*) spectra were recorded on a 850 MHz wide bore spectrometer at the ETH Zurich.

6.4. Conclusion and perspectives

Five constructs corresponding to either core D1 or full-length core were successfully expressed in the WGE-CF system (precipitate mode). As shown by electron microscopy, both D1 and the full-length protein spontaneouly assemble into nucleocapsid-like particles which can be isolated on Accudenz cushions. Although the purity level of these isolated nucleocapsid-like particles is not optimal, it should not hamper NMR studies since only the synthesized protein is isotopically labeled. However, a problem emerged from the first solid-state NMR spectra : large amounts of RNA are co-sedimented with the nucleocapsid-like particles, maybe in addition to that included in the capsid. The current challenge is thus the removal of excess RNA co-sedimenting with core nucleocapid-like particles in order to increase the protein amount in the NMR rotor and thus enhance signal-to-noise ratio, enabling 2D spectroscopy. Although RNA can be removed while treating the CFS samples with benzonase, there are indications that benzonase treatment would lead to the disruption of the nucleocapsid-like particles.

The expression of either D1 (aa 1 to 124) or full-length core protein (aa 1 to 179) in bacteria followed by *in vitro* self-assembly of nucleocapsid-like particles has been reported (Kunkel *et al*, 2001), showing that the 124 N-terminal amino acid residues are sufficient for self-assembly. This is in agreement with the fact that we observed nucleocapsid-like particles when expressing core D1 in the WGE-CF system. Moreover, self-assembly of D1 and full-length core has been reported to require structured RNA molecules and no self-assembly was observed in the absence of nucleotides (Kunkel *et al*, 2001). This is coherent with the observation that degradation of nucleotides from CFS seems to lead to nucleocapsid-like particle disruption. The difficulty lies now in the optimization of the optimal amount of benzonase and incubation time which would allow to degrade the RNA outside the nucleocapsid-like particles without interfering with the integrity of these particles.

Finally, the expression of core D2 was not successful in the precipitate mode. This domain interacts with ER membrane and LDs. With hindsight, the expression of core D2 should therefore be tested in the presence of detergent, although we experienced that zero expression in the precipitate mode is usually not promising. Positioning the *Strep*-tag II at the N-terminus might also help.

7. General conclusions and perspectives

During this work, a wheat germ cell-free protein expression system has been successfully developed. Based on protocols established by Yaeta Endo and coworkers at Ehime University in Japan, and with generous initial help of Prof. Endo, the preparation of wheat germ extracts (WGEs) has been optimized, leading to high quality WGEs as efficient as commercial ones. Moreover, protein translation using the bilayer method has proved to be the most suitable approach for the production of isotopically labeled proteins compared to dialysis systems. Main expression parameters such as temperature or WGE concentration have also been optimized. After implementation and optimization, the system has been tested for the expression of 75 different protein constructs, including mainly membrane proteins, in the precipitate mode with a success over 82%. Furthermore, the expression of NS3-4A and NS2-NS3^{pro} precursor constructs has evidenced that these proteins produced in the WGE-CF system are active. The development of the system is therefore a first success.

A major progress was the use of detergents to express membrane proteins directly in a soluble form. Especially the MNG-3 allowed the expression of the HCV non-structural proteins NS2, NS4B and NS5A in a full-length, detergent-solubilized and well-folded form. Also the C12E8 allowed the expression of full-length NS2 and NS2^{pro} in a solubilized form. This work was published in Protein Expression and Purification (Fogeron *et al*, 2015). The direct expression in a solubilized form circumvents many difficulties linked to purification from aggregates obtained in the precipitate mode and even further inclusion bodies while expressing membrane proteins in bacteria, notably the presence of non-native forms on refolding and protein loss. Although the general use of MNG-3 and C12E8 for a successful expression and solubilization of membrane proteins remains to be established, these results present a further step in devising strategies for the wheat germ cell-free expression of membrane proteins. Moreover, alternatives to MNG-3 and C12E8 could be considered such as the use of further detergents, detergent or detergent/lipid mixtures, or even nonionic amphipols (NApols) if they would become commercially available.

So far, only parts of NS2, NS4B and NS5A structures could be studied using different biophysical approaches such as x-ray crystallography and solution-state NMR. More detailed studies of these proteins were indeed missing since active full-length proteins

were not available due to the difficulties encountered for their expression in cellular systems. While many studies have addressed interactions between NS2 with several other viral proteins *in cellulo*, no *in vitro* assays on these interactions could be carried out in order to gain knowledge on the underlying molecular mechanisms. Indeed, some of these interactions presumably take place *via* the hydrophobic membrane-inserted N-terminal region, and their occurrence might as well rely on the correct interaction of the protein complex with the membrane. By using our wheat germ cell-free expression system, full-length NS2, NS4B and NS5A were produced and purified in milligram amounts in a homogenous detergent-solubilized form with expected secondary structure content. In the particular case of NS2, the purified full-length protein produced in the WGE-CF system was active in absence of EDTA or chelating resin. To our knowledge, this is the first time that protease activity of full-length NS2 has been evidenced *in vitro*. The availability of full-length NS2, NS4B and NS5A in a detergent-solubilized or membrane-bound form, in sufficient quantities, thus opens the perspective for biochemical *in vitro* structural and functional studies of these enigmatic proteins. While size exclusion chromatography and electron microscopy analyses showed that NS2, NS4B and NS5A are oligomeric proteins, the number of monomers remains to be determined. Analytical ultracentrifugation and size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) experiments would help to gain more insight in the oligomerization process of these proteins.

The ultimate goal of the research projects in the laboratory is to determine the structure of NS2, NS4B and NS5A by solid-state NMR in a native-like membrane lipids environment. Although expression in cell-free systems in the presence of liposomes might allow the direct membrane insertion of the produced proteins, detergent solubilization and subsequent reinsertion into membranes allow us to obtain higher protein yields and purer preparations. They also allow us to precisely determine the lipid composition used as well as the lipid-to-protein ratio. While lipid reconstitution of purified full-length NS2 or NS4B has been optimized, optimal reconstitution conditions have to be determined for each membrane protein. First solid-state NMR spectra of NS4B in a mixture of egg α-phosphatidylcholine and cholesterol have been obtained. These very encouraging preliminary data are the proof that our WGE-CF system allows the production of good quality labeled protein samples for NMR studies. Successful

insertion into membranes is indeed a good indicator for the structural integrity of the membrane protein, and homogenous membrane insertion is a prerequisite for further structural studies of the protein. However, neither the orientation of NS2 and NS4B in membranes after reconstitution nor the potential effect of inside-out or outside-out orientations on NMR spectra quality have been determined yet. In addition, although the different lipid mixtures tested here did not show significant differences regarding to the reconstitution efficiency, it can't be excluded that the lipid composition could influence the quality of the spectra. The adventure is just beginning!

Although the high cost of commercial WGEs or the tedious preparation of home-made WGEs might dishearten, wheat germ cell-free expression has undeniable advantages over classical cell-based expression systems and even other cell-free systems, especially for membrane proteins. The techniques developed during this work are now successfully applied by Aurélie Badillo at RD Biotech, as part of a cooperation agreement between this company and CNRS. Furthermore, I trained several collaborators who could thus successfully established the system in their own laboratory, allowing them to move forward with their projects. The development of a wheat germ cell-free system presented in this work thus opens the way to the structural and functional analyses of not only HCV membrane proteins, but also further viral and eukaryotic membranes proteins. While the work on HCV membrane proteins will be thorough, the scope of this project will be expanded, through collaborations with virologists and molecular biologists, to proteins from related viruses, such as dengue, hepatitis B (HBV) or hepatitis E (HEV) viruses. Finally, a close collaboration with solid-state NMR spectroscopists is essential. Indeed, Beat Meier and coworkers at the ETH Zurich develop central solid-state NMR theoretical, methodological and technological concepts and approaches, as well as dispose of state-of-the-art high-field magnets and probes enabling these new methodological approaches. The synergy generated by the complementarity of wheat germ cell-free protein expression and new solid-state NMR approaches is thus expected to lead to new insights into the structural features of the NS2, NS4B and NS5A membrane proteins and their role in HCV replication to better understand the viral HCV life cycle.

REFERENCES

Abdine A, Verhoeven MA & Warschawski DE (2011) Cell-free expression and labeling strategies for a new decade in solid-state NMR. *New BIOTECHNOLOGY* **28:** 272–276

Abdine A, Verhoeven MA, Park K-H, Ghazi A, Guittet É, Berrier C, Van Heijenoort C & Warschawski DE (2010) Structural study of the membrane protein MscL using cell-free expression and solid-state NMR. *Journal of Magnetic Resonance* **204:** 155–159

Abe M, Ohno S, Yokogawa T, Nakanishi T, Arisaka F, Hosoya T, Hiramatsu T, Suzuki M, Ogasawara T, Sawasaki T, Nishikawa K, Kitamura M, Hori H & Endo Y (2007) Detection of structural changes in a cofactor binding protein by using a wheat germ cell-free protein synthesis system coupled with unnatural amino acid probing. *Proteins* **67:** 643–652

Agnello V, Abel G, Elfahal M, Knight GB & Zhang QX (1999) Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *PNAS* **96:** 12766–12771

Ahmad M, Hirz M, Pichler H & Schwab H (2014) Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol* **98:** 5301–5317

Alberti A & Piovesan S (2014) The evolution of the therapeutic strategy in hepatitis C: Features of sofosbuvir and indications. *Digestive and Liver Disease* **46:** S174–S178

Althoff T, Davies KM, Schulze S, Joos F & Kühlbrandt W (2012) GRecon: A Method for the Lipid Reconstitution of Membrane Proteins. *Angew. Chem. Int. Ed.* **51:** 8343–8347

Aly KA, Beebe ET, Chan CH, Goren MA, Sepúlveda C, Makino S-I, Fox BG & Forest KT (2012) Cell-free production of integral membrane aspartic acid proteases reveals zinc-dependent methyltransferase activity of the Pseudomonas aeruginosaprepilin peptidase PilD. *MicrobiologyOpen* **2:** 94–104

Amako Y, Tsukiyama-Kohara K, Katsume A, Hirata Y, Sekiguchi S, Tobita Y, Hayashi Y, Hishima T, Funata N, Yonekawa H & Kohara M (2010) Pathogenesis of hepatitis C virus infection in Tupaia belangeri. *Journal of Virology* **84:** 303–311

Anastasina M, Terenin I, Butcher SJ & Kainov DE (2014) A technique to increase protein yield in a rabbit reticulocyte lysate translation system. *BioTechniques* **56:** 36–39

Andre P, Komurian-Pradel F, Deforges S, Perret M, Berland JL, Sodoyer M, Pol S, Brechot C, Paranhos-Baccala G & Lotteau V (2002) Characterization of Low- and Very-Low-Density Hepatitis C Virus RNA-Containing Particles. *Journal of Virology* **76:** 6919–6928

Andréll J & Tate CG (2013) Overexpression of membrane proteins in mammalian cells for structural studies. *Mol Membr Biol* **30:** 52–63

Angus AGN, Loquet A, Stack SJ, Dalrymple D, Gatherer D, Penin F & Patel AH (2011) Conserved Glycine 33 Residue in Flexible Domain I of Hepatitis C Virus Core Protein Is Critical for Virus Infectivity. *Journal of Virology* **86:** 679–690

Aoki M, Matsuda T, Tomo Y, Miyata Y, Inoue M, Kigawa T & Yokoyama S (2009) Automated system for high-throughput protein production using the dialysis cell-free method. *Protein Expr. Purif.* **68:** 128–136

Appel N, Pietschmann T & Bartenschlager R (2005) Mutational Analysis of Hepatitis C Virus Nonstructural Protein 5A: Potential Role of Differential Phosphorylation in RNA Replication and Identification of a Genetically Flexible Domain. *Journal of Virology* **79:** 3187–3194

Appel N, Zayas M, Miller S, Krijnse-Locker J, Schaller T, Friebe P, Kallis S, Engel U & Bartenschlager R (2008) Essential Role of Domain III of Nonstructural Protein 5A for Hepatitis C Virus Infectious Particle Assembly. *PLoS Pathog* **4:** e1000035

Appleby TC, Perry JK, Murakami E, Barauskas O, Feng J, Cho A, Fox D, Wetmore DR, McGrath ME, Ray AS, Sofia MJ, Swaminathan S & Edwards TE (2015) Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* **347:** 771–775

Arinaminpathy Y, Khurana E, Engelman DM & Gerstein MB (2009) Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discovery Today* **14:** 1130–1135

Atoom AM, Taylor NGA & Russell RS (2014) The elusive function of the hepatitis C virus p7 protein. *Virology* **462-463:** 377–387

Barba G, Harper F, Harada T, Kohara M, Goulinet S, Matsuura Y, Eder G, Schaff Z, Chapman MJ, Miyamura T & Brechot C (1997) Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets. *PNAS* **94:** 1200–1205

Bartenschlager R, Lohmann V & Penin F (2013) The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. *Nature Publishing Group* **11:** 482–496

Bartenschlager R, Penin F, Lohmann V & André P (2011) Assembly of infectious hepatitis C virus particles. *Trends in Microbiology* **19:** 95–103

Bartosch B, Dubuisson J & Cosset FL (2003) Infectious Hepatitis C Virus Pseudo-particles Containing Functional E1-E2 Envelope Protein Complexes. *Journal of Experimental Medicine* **197:** 633–642

Bayburt TH & Sligar SG (2010) Membrane protein assembly into Nanodiscs. *FEBS LETTERS* **584:** 1721–1727

Bazzacco P, Billon-Denis E, Sharma KS, Catoire LJ, Mary S, Le Bon C, Point É, Banères J-L, Durand G, Zito F, Pucci B & Popot J-L (2012) Nonionic Homopolymeric Amphipols: Application to Membrane Protein Folding, Cell-Free Synthesis, and Solution Nuclear Magnetic Resonance. *Biochemistry* **51:** 1416–1430

Behrens SE, Tomei L & De Francesco R (1996) Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J* **15:** 12–22

Berger C, Romero-Brey I, Radujkovic D, Terreux R, Zayas M, Paul D, Harak C, Hoppe S, Gao M, Penin F, Lohmann V & Bartenschlager R (2014) Daclatasvir-Like Inhibitors of NS5A Block Early Biogenesis of Hepatitis C Virus-Induced Membranous Replication Factories, Independent of RNA Replication. *Gastroenterology* **147:** 1094–1105.e25

Berger KL, Kelly SM, Jordan TX, Tartell MA & Randall G (2011) Hepatitis C Virus Stimulates the Phosphatidylinositol 4-Kinase III Alpha-Dependent Phosphatidylinositol 4-Phosphate Production That Is Essential for Its Replication. *Journal of Virology* **85:** 8870–8883

Bernaudat F, Frelet-Barrand A, Pochon N, Dementin S, Hivin P, Boutigny S, Rioux J-B, Salvi D, Seigneurin-Berny D, Richaud P, Joyard J, Pignol D, Sabaty M, Desnos T, Pebay-Peyroula E, Darrouzet E, Vernet T & Rolland N (2011) Heterologous Expression of Membrane Proteins: Choosing the Appropriate Host. *PloS One* **6:** e29191

Berrier C, Park K-H, Abes S, Bibonne A, Betton J-M & Ghazi A (2004) Cell-Free Synthesis of a Functional Ion Channel in the Absence of a Membrane and in the Presence of Detergent †. *Biochemistry* **43:** 12585–12591

Bill R (2001) Yeast - a panacea for the structure-function analysis of membrane proteins? *Current Genetics* **40:** 157–171

Blesneac I, Ravaud S, Juillan-Binard C, Barret L-A, Zoonens M, Polidori A, Miroux B, Pucci B & Pebay-Peyroula E (2012) Production of UCP1 a membrane protein from the inner mitochondrial membrane using the cell free expression system in the presence of a fluorinated surfactant. *BBA - Biomembranes* **1818:** 798–805

Bligh EG & Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* **37:** 911–917

Blight KJ, Kolykhalov AA & Rice CM (2000) Efficient initiation of HCV RNA replication in cell culture. *Science* **290:** 1972–1974

Blight KJ, McKeating JA, Marcotrigiano J & Rice CM (2003) Efficient replication of hepatitis C virus genotype 1a RNAs in cell culture. *Journal of Virology* **77:** 3181–3190

Blobel G (1980) Intracellular protein topogenesis. *PNAS* **77:** 1496–1500

Borch J & Hamann T (2009) The nanodisc: a novel tool for membrane protein studies. *Biol. Chem.* **390:** 805–814

Bornert O, Alkhalfioui F, Logez C & Wagner R (2001) Overexpression of Membrane Proteins Using Pichia pastoris Hoboken, NJ, USA: John Wiley & Sons, Inc.

Boson B, Granio O, Bartenschlager R & Cosset F-L (2011) A Concerted Action of Hepatitis C Virus P7 and Nonstructural Protein 2 Regulates Core Localization at the Endoplasmic Reticulum and Virus Assembly. *PLoS Pathog* **7:** e1002144

Boulant S, Montserret R, Hope RG, Ratinier M, Targett-Adams P, Lavergne J-P, Penin F & McLauchlan J (2006) Structural Determinants That Target the Hepatitis C Virus Core Protein to Lipid Droplets. *Journal of Biological Chemistry* **281:** 22236–22247

Boulant S, Targett-Adams P & McLauchlan J (2007) Disrupting the association of hepatitis C virus core protein with lipid droplets correlates with a loss in production of infectious virus. *Journal of General Virology* **88:** 2204–2213

Boulant S, Vanbelle C, Ebel C, Penin F & Lavergne J-P (2005) Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *Journal of Virology* **79:** 11353–11365

Bowen DG & Walker CM (2005) Adaptive immune responses in acute and chronic hepatitis C virus infection. *Nature* **436:** 946–952

Branch AD, Stump DD, Gutierrez JA, Eng F & Walewski JL (2005) The hepatitis C virus alternate reading frame (ARF) and its family of novel products: the alternate reading frame protein/F-protein, the double-frameshift protein, and others. *Semin Liver Dis* **25:** 105–117

Brass V, Berke JM, Montserret R, Blum HE, Penin F & Moradpour D (2008) Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex. *PNAS* **105:** 14545–14550

Brass V, Bieck E, Montserret R, Wölk B, Hellings JA, Blum HE, Penin F & Moradpour D (2002) An amino-terminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *Journal of Biological Chemistry* **277:** 8130–8139

Brenndörfer ED, Karthe J, Frelin L, Cebula P, Erhardt A, Schulte am Esch J, Hengel H, Bartenschlager R, Sällberg M, Häussinger D & Bode JG (2009) Nonstructural 3/4A protease of hepatitis C virus activates epithelial growth factor-induced signal transduction by cleavage of the T-cell protein tyrosine phosphatase. *Hepatology* **49:** 1810–1820

Bressanelli S, Tomei L, Roussel A, Incitti I, Vitale RL, Mathieu M, De Francesco R & Rey FA (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *PNAS* **96:** 13034–13039

Brohawn SG, del Mármol J & MacKinnon R (2012) Crystal structure of the human K2P TRAAK, a lipid- and mechano-sensitive K+ ion channel. *Science* **335:** 436–441

Brown RS (2005) Hepatitis C and liver transplantation. *Nature* **436:** 973–978

Bukh J, Forns X, Emerson SU & Purcell RH (2001) Studies of hepatitis C virus in chimpanzees and their importance for vaccine development. *Intervirology* **44:** 132–142

Buntru M, Vogel S, Spiegel H & Schillberg S (2014) Tobacco BY-2 cell-free lysate: an alternative and highly-productive plant-based in vitro translation system. *BMC Biotechnology* **14:** 1–11

Carraher C, Nazmi AR, Newcomb RD & Kralicek A (2013) Recombinant expression, detergent solubilisation and purification of insect odorant receptor subunits. *Protein Expr. Purif.* **90:** 160–169

Carrere-Kremer S, Montpellier-Pala C, Cocquerel L, Wychowski C, Penin F & Dubuisson J (2002) Subcellular Localization and Topology of the p7 Polypeptide of Hepatitis C Virus. *Journal of Virology* **76:** 3720–3730

Chae PS, Rasmussen SGF, Rana RR, Gotfryd K, Chandra R, Goren MA, Kruse AC, Nurva S, Loland CJ, Pierre Y, Drew D, Popot J-L, Picot D, Fox BG, Guan L, Gether U, Byrne B, Kobilka B & Gellman SH (2010) Maltose–neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat Meth* **7:** 1003–1008

Chandler DE, Penin F, Schulten K & Chipot C (2012) The p7 Protein of Hepatitis C Virus Forms Structurally Plastic, Minimalist Ion Channels. *PLoS Comput Biol* **8:** e1002702

Chaudhary S, Pak JE, Gruswitz F, Sharma V & Stroud RM (2012) Overexpressing human membrane proteins in stably transfected and clonal human embryonic kidney 293S cells. *Nat Protoc* **7:** 453–466

Chen H, Shaffer PL, Huang X & Rose PE (2013) Rapid screening of membrane protein expression in transiently transfected insect cells. *Protein Expr. Purif.* **88:** 134–142

Cherezov V, Rosenbaum DM & Hanson MA (2007) High-resolution crystal structure of an engineered human β2-adrenergic G protein–coupled receptor. *Science*

Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW & Houghton M (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244:** 359–362

Cladera J, Rigaud JL, Villaverde J & Duñach M (1997) Liposome solubilization and membrane protein reconstitution using Chaps and Chapso. *Eur. J. Biochem.* **243:** 798–804

Cocquerel L (2006) Hepatitis C virus entry: potential receptors and their biological functions. *Journal of General Virology* **87:** 1075–1084

Cohen E, Goldshleger R, Shainskaya A, Tal DM, Ebel C, le Maire M & Karlish SJD (2005) Purification of Na+,K+-ATPase expressed in Pichia pastoris reveals an essential role of phospholipid-protein interactions. *Journal of Biological Chemistry* **280:** 16610–16618

Coller KE, Heaton NS, Berger KL, Cooper JD, Saunders JL & Randall G (2012) Molecular Determinants and Dynamics of Hepatitis C Virus Secretion. *PLoS Pathog* **8:** e1002466

Contreras-Gómez A, Sánchez-Mirón A, García-Camacho F, Molina-Grima E & Chisti Y (2013) Protein production using the baculovirus-insect cell expression system. *Biotechnol Progress* **30:** 1–18

Counihan NA, Rawlinson SM & Lindenbach BD (2011) Trafficking of Hepatitis C Virus Core Protein during Virus Particle Assembly. *PLoS Pathog* **7:** e1002302

Craig D, Howell MT, Gibbs CL, Hunt T & Jackson RJ (1992) Plasmid cDNA-directed protein synthesis in a coupled eukaryotic in vitro transcription-translation system. *Nucleic Acids Research* **20:** 4987–4995

Cristofari G (2004) The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro. *Nucleic Acids Research* **32:** 2623–2631

Cymer F, Heijne von G & White SH (2014) Mechanisms of Integral Membrane Protein Insertion and Folding. *Journal of Molecular Biology***:** 1–24

Daniel CJ, Conti B, Johnson AE & Skach WR (2008) Control of Translocation through the Sec61 Translocon by Nascent Polypeptide Structure within the Ribosome. *Journal of Biological Chemistry* **283:** 20864–20873

de Chassey B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaugué S, Meiffren G, Pradezynski F, Faria BF, Chantier T, Le Breton M, Pellet J, Davoust N, Mangeot PE, Chaboud A, Penin F, Jacob Y, Vidalain PO, Vidal M, Andre P, et al (2008) Hepatitis C virus infection protein network. *Mol Syst Biol* **4:** 230

de Ruyter PG, Kuipers OP & de Vos WM (1996) Controlled gene expression systems for Lactococcus lactis with the food-grade inducer nisin. *Appl. Environ. Microbiol.* **62:** 3662–3667

Degrip WJ, Vanoostrum J & Bovee-Geurts PH (1998) Selective detergent-extraction from mixed detergent/lipid/protein micelles, using cyclodextrin inclusion compounds: a novel generic approach for the preparation of proteoliposomes. *Biochem. J.* **330 (Pt 2):** 667–674

Deniaud A, Liguori L, Blesneac I, Lenormand JL & Pebay-Peyroula E (2010) Crystallization of the membrane protein hVDAC1 produced in cell-free system. *BBA - Biomembranes* **1798:** 1540–1546

Di Bisceglie AM (1998) Hepatitis C. *The Lancet* **351:** 351–355

Diab C, Tribet C, Gohon Y, Popot JL & Winnik FM (2007) Complexation of integral membrane proteins by phosphorylcholine-based amphipols. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1768:** 2737–2747

Dobberstein B & Blobel G (1977) Functional interaction of plant ribosomes with animal microsomal membranes. *Biochemical and Biophysical Research Communications* **74:** 1675–1682

Dorner M & Ploss A (2011) Deconstructing hepatitis C virus infection in humanized mice. *Ann. N. Y. Acad. Sci.* **1245:** 59–62

Dorner M, Horwitz JA, Donovan BM, Labitt RN, Budell WC, Friling T, Vogt A, Catanese MT, Satoh T, Kawai T, Akira S, Law M, Rice CM & Ploss A (2013) Completion of the entire hepatitis C virus life-cycle in genetically humanized mice. *Nature* **501:** 237–241

Drew DE, Heijne von G, Nordlund P & de Gier J (2001) Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS LETTERS* **507:** 220–224

Dubuisson J & Cosset F-L (2014) Virology and cell biology of the hepatitis C virus life cycle – An update. *Journal of Hepatology* **61:** S3–S13

Dukkipati A, Park HH, Waghray D, Fischer S & Garcia KC (2008) BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. *Protein Expr. Purif.* **62:** 160–170

Dürr UHN, Gildenberg M & Ramamoorthy A (2012) The Magic of Bicelles Lights Up Membrane Protein Structure. *Chem. Rev.* **112:** 6054–6074

Dvir H & Choe S (2009) Bacterial expression of a eukaryotic membrane protein in fusion to various Mistic orthologs. *Protein Expr. Purif.* **68:** 28–33

Edelmann A, Kirchberger J, Naumann M & Kopperschläger G (2000) Generation of catalytically active 6-phosphofructokinase from Saccharomyces cerevisiae in a cell-free system. *Eur. J. Biochem.* **267:** 4825–4830

Egger D, Wolk B, Gosert R, Bianchi L, Blum HE, Moradpour D & Bienz K (2002) Expression of Hepatitis C Virus Proteins Induces Distinct Membrane Alterations Including a Candidate Viral Replication Complex. *Journal of Virology* **76:** 5974–5984

Einav S, Elazar M, Danieli T & Glenn JS (2004) A nucleotide binding motif in hepatitis C virus (HCV) NS4B mediates HCV RNA replication. *Journal of Virology* **78:** 11288–11295

Einav S, Gerber D, Bryson PD, Sklan EH, Elazar M, Maerkl SJ, Glenn JS & Quake SR (2008) Discovery of a hepatitis C target and its pharmacological inhibitors by microfluidic affinity analysis. *Nat Biotechnol* **26:** 1019–1027

Elbaz Y, Steiner-Mordoch S, Danieli T & Schuldiner S (2004) In vitro synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state. *PNAS* **101:** 1519–1524

Endo Y & Sawasaki T (2004) High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. *J. Struct. Funct. Genomics* **5:** 45–57

Endo Y & Sawasaki T (2006) Cell-free expression systems for eukaryotic protein production. *Curr. Opin. Biotechnol.* **17:** 373–380

Eng FJ, Walewski JL, Klepper AL, Fishman SL, Desai SM, McMullan LK, Evans MJ, Rice CM & Branch AD (2009) Internal Initiation Stimulates Production of p8 Minicore, a Member of a Newly Discovered Family of Hepatitis C Virus Core Protein Isoforms. *Journal of Virology* **83:** 3104–3114

Enoch HG, Fleming PJ & Strittmacher P (1979) The binding of cytochrome b_5 to phospholipid vesicles and biological membranes. *The Journal of Biological Chemistry* **254:** 6483–6488

Enomoto N, Sakuma I & Asahina Y (1996) Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *The New England Journal of Medecine* **334:** 77–82

Evans MJ, Hahn von T, Tscherne DM, Syder AJ, Panis M, Wölk B, Hatziioannou T, McKeating JA, Bieniasz PD & Rice CM (2007) Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature* **446:** 801–805

Evans MJ, Rice CM & Goff SP (2004) Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *PNAS* **101:** 13038–13043

Eytan GD (1982) Use of liposomes for reconstitution of biological functions. *Biochim. Biophys. Acta* **694:** 185–202

Eytan GD, Matheson MJ & Racker E (1976) Incorporation of mitochondrial membrane proteins into liposomes containing acidic phospholipids. *Journal of Biological Chemistry* **251:** 6831–6837

Ezure T, Suzuki T, Higashide S, Shintani E, Endo K, Kobayashi SI, Shikata M, Ito M, Tanimizu K & Nishimura O (2006) Cell-Free Protein Synthesis System Prepared from Insect Cells by Freeze-Thawing. *Biotechnol Progress* **22:** 1570–1577

Fagerberg L, Jonasson K, Heijne von G, Uhlén M & Berglund L (2010) Prediction of the human membrane proteome. *Proteomics* **10:** 1141–1149

Falson P, Bartosch B, Alsaleh K, Tews BA, Loquet A, Ciczora Y, Riva L, Montigny C, Duverlie G, Pêcheur EI, le Maire M, Cosset FL, Dubuisson J & Penin F (*Submitted*) Hepatitis C virus envelope glycoprotein E1 forms trimers at the surface of the virion.

Farci P, Alter HJ, Shimoda A, Govindarajan S, Cheung LC, Melpolder JC, Sacher RA, Shih JW & Purcell RH (1996) Hepatitis C virus-associated fulminant hepatic failure. *N. Engl. J. Med.* **335:** 631–634

Fernández FJ & Vega MC (2013) Technologies to keep an eye on: alternative hosts for protein production in structural biology. *Current Opinion in Structural Biology* **23:** 365–373

Figler RA, Omote H, Nakamoto RK & Al-Shawi MK (2000) Use of Chemical Chaperones in the Yeast Saccharomyces cerevisiae to Enhance Heterologous Membrane Protein Expression: High-Yield Expression and Purification of Human P-Glycoprotein. *Archives of Biochemistry and Biophysics* **376:** 34–46

Fogeron M-L, Badillo A, Jirasko V, Gouttenoire J, Paul D, Lancien L, Moradpour D, Bartenschlager R, Meier BH, Penin F & Böckmann A (2015) Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins. *Protein Expr. Purif.* **105:** 39–46

Fogeron M-L, Paul D, Jirasko V, Montserret R, Lacabanne D, Molle J, Badillo A, Boukadida C, Georgeault S, Roingeard P, Martin A, Bartenschlager R, Penin F & Böckmann A (*Submitted*) Functional expression, purification, characterization, and membrane reconstitution of nonstructural protein 2 from hepatitis C virus. *Under review in Protein Expr. Purif.*

Folch J, Lees M & Sloane Stanley GH (1957) A simple method for the isolation and purification of total lipides from animal tissues. *Journal of Biological Chemistry* **226:** 497–509

Forns X, Purcell RH & Bukh J (1999) Quasispecies in viral persistence and pathogenesis of hepatitis C virus. *Trends in Microbiology* **7:** 402–410

Foster TL, Gallay P, Stonehouse NJ & Harris M (2011) Cyclophilin A Interacts with Domain II of Hepatitis C Virus NS5A and Stimulates RNA Binding in an Isomerase-Dependent Manner. *Journal of Virology* **85:** 7460–7464

Frank C, Mohamed MK, Strickland GT, Lavanchy D, Arthur RR, Magder LS, Khoby TE, Abdel-Wahab Y, Ohn ESA, Anwar W & Sallam I (2000) The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *The Lancet* **355:** 887–891

Friebe P, Boudet J, Simorre J-P & Bartenschlager R (2005) Kissing-loop interaction in the 3' end of the hepatitis C virus genome essential for RNA replication. *Journal of Virology* **79:** 380–392

Gan R & Jewett MC (2014) A combined cell-free transcription-translation system from Saccharomyces cerevisiae for rapid and robust protein synthe. *Biotechnology Journal*

Garavito RM & Ferguson-Miller S (2001) Detergents as tools in membrane biochemistry. *Journal of Biological Chemistry* **276:** 32403–32406

Gastaminza P, Kapadia SB & Chisari FV (2006) Differential Biophysical Properties of Infectious Intracellular and Secreted Hepatitis C Virus Particles. *Journal of Virology* **80:** 11074–11081

Genji T, Nozawa A & Tozawa Y (2010) Efficient production and purification of functional bacteriorhodopsin with a wheat-germ cell-free system and a combination of Fos-choline and CHAPS detergents. *Biochemical and Biophysical Research Communications* **400:** 638–642

Gentzsch J, Brohm C, Steinmann E, Friesland M, Menzel N, Vieyres G, Perin PM, Frentzen A, Kaderali L & Pietschmann T (2013) Hepatitis C Virus p7 is Critical for Capsid Assembly and Envelopment. *PLoS Pathog* **9:** e1003355

Gerold G & Pietschmann T (2014) The HCV Life Cycle: In vitro Tissue Culture Systems and Therapeutic Targets. *Dig Dis* **32:** 525–537

Ghadessy FJ & Holliger P (2004) A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein Engineering Design and Selection* **17:** 201–204

Goder V & Spiess M (2001) Topogenesis of membrane proteins: determinants and dynamics. *FEBS LETTERS* **504:** 87–93

Goehring A, Lee C-H, Wang KH, Michel JC, Claxton DP, Baconguis I, Althoff T, Fischer S, Garcia KC & Gouaux E (2014) Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nature Protocols* **9:** 2574–2582

Goncalves AM (2013) Pichia pastoris: A Recombinant Microfactory for Antibodies and Human Membrane Proteins. *J. Microbiol. Biotechnol.* **23:** 587–601

Gordon SC, Bayati N & Silverman AL (1998) Clinical outcome of hepatitis C as a function of mode of transmission. *Hepatology* **28:** 562–567

Goren MA & Fox BG (2008) Wheat germ cell-free translation, purification, and assembly of a functional human stearoyl-CoA desaturase complex. *Protein Expr. Purif.* **62:** 171–178

Goren MA, Nozawa A, Makino S & Wrobel RL (2009) Cell-free translation of integral membrane proteins into unilamelar liposomes. *Methods in Enzymology* **463:** 647–673

Gosert R, Egger D, Lohmann V, Bartenschlager R, Blum HE, Bienz K & Moradpour D (2003) Identification of the Hepatitis C Virus RNA Replication Complex in Huh-7 Cells Harboring Subgenomic Replicons. *Journal of Virology* **77:** 5487–5492

Goshima N, Kawamura Y, Fukumoto A, Miura A, Honma R, Satoh R, Wakamatsu A, Yamamoto J-I, Kimura K, Nishikawa T, Andoh T, Iida Y, Ishikawa K, Ito E, Kagawa N, Kaminaga C, Kanehori K-I, Kawakami B, Kenmochi K, Kimura R, et al (2008) Human protein factory for converting the transcriptome into an in vitro–expressed proteome. *Nat Meth* **5:** 1011–1017

Gottwein JM, Scheel TKH, Jensen TB, Lademann JB, Prentoe JC, Knudsen ML, Hoegh AM & Bukh J (2009) Development and characterization of hepatitis C virus genotype 1-7 cell culture systems: role of CD81 and scavenger receptor class B type I and effect of antiviral drugs. *Hepatology* **49:** 364–377

Gouttenoire J, Castet V, Montserret R, Arora N, Raussens V, Ruysschaert J-M, Diesis E, Blum HE, Penin F & Moradpour D (2009a) Identification of a novel determinant for membrane association in hepatitis C virus nonstructural protein 4B. *Journal of Virology* **83:** 6257–6268

Gouttenoire J, Montserret R, Kennel A, Penin F & Moradpour D (2009b) An amphipathic alpha-helix at the C terminus of hepatitis C virus nonstructural protein 4B mediates membrane association. *Journal of Virology* **83:** 11378–11384

Gouttenoire J, Montserret R, Paul D, Castillo R, Meister S, Bartenschlager R, Penin F & Moradpour D (2014) Aminoterminal Amphipathic α-Helix AH1 of Hepatitis C Virus Nonstructural Protein 4B Possesses a Dual Role in RNA Replication and Virus Production. *PLoS Pathog* **10:** e1004501

Gouttenoire J, Penin F & Moradpour D (2010a) Hepatitis C virus nonstructural protein 4B: a journey into unexplored territory. *Rev. Med. Virol.* **20:** 117–129

Gouttenoire J, Roingeard P, Penin F & Moradpour D (2010b) Amphipathic alpha-helix AH2 is a major determinant for the oligomerization of hepatitis C virus nonstructural protein 4B. *Journal of Virology* **84:** 12529–12537

Greenhut SF & Roseman MA (1985) Distribution of cytochrome b5 between sonicated phospholipid vesicles of different size. *Journal of Biological Chemistry* **260:** 5883–5886

Griffith DA, Delipala C, Leadsham J, Jarvis SM & Oesterhelt D (2003) A novel yeast expression system for the overproduction of quality-controlled membrane proteins. *FEBS LETTERS* **553:** 45–50

Grisshammer R, Duckworth R & Henderson R (1993) Expression of a rat neurotensin receptor in Escherichia coli. *Biochem. J.*

Guarino C & DeLisa MP (2012) A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. *Glycobiology* **22:** 596–601

Gulik-Krzywicki T, Seigneuret M & Rigaud JL (1987) Monomer-oligomer equilibrium of bacteriorhodopsin in reconstituted proteoliposomes. A freeze-fracture electron microscope study. *Journal of Biological Chemistry* **262:** 15580–15588

Gutti TL, Knibbe JS, Makarov E, Zhang J, Yannam GR, Gorantla S, Sun Y, Mercer DF, Suemizu H, Wisecarver JL, Osna NA, Bronich TK & Poluektova LY (2014) Human Hepatocytes and Hematolymphoid Dual Reconstitution in Treosulfan-Conditioned uPA-NOG Mice. *The American Journal of Pathology* **184:** 101–109

Hanoulle X, Badillo A, Wieruszeski J-M, Verdegem D, Landrieu I, Bartenschlager R, Penin F & Lippens G (2009) Hepatitis C virus NS5A protein is a substrate for the peptidyl-prolyl cis/trans isomerase activity of cyclophilins A and B. *Journal of Biological Chemistry* **284:** 13589–13601

Harbers M (2014) Wheat germ systems for cell-free protein expression. *FEBS LETTERS* **17:** 2762–2773

Hartl FU & Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295:** 1852–1858

Hassaine G, Wagner R, Kempf J, Cherouati N, Hassaine N, Prual C, André N, Reinhart C, Pattus F & Lundstrom K (2006) Semliki Forest virus vectors for overexpression of 101 G protein-coupled receptors in mammalian host cells. *Protein Expr. Purif.* **45:** 343–351

Haviv H, Cohen E, Lifshitz Y, Tal DM, Goldshleger R & Karlish SJD (2007) Stabilization of Na +,K +-ATPase Purified from Pichia pastorisMembranes by Specific Interactions with Lipids †. *Biochemistry* **46:** 12855–12867

Hays FA, Roe-Zurz Z & Stroud RM (2010) Chapter 29 - Overexpression and Purification of Integral Membrane Proteins in Yeast 2nd ed. Elsevier Inc

Hays FA, Roe-Zurz Z, Li M, Kelly L, Gruswitz F, Sali A & Stroud RM (2009) Ratiocinative screen of eukaryotic integral membrane protein expression and solubilization for structure determination. *J. Struct. Funct. Genomics* **10:** 9–16

Henke JI, Goergen D, Zheng J, Song Y, Schüttler CG, Fehr C, Jünemann C & Niepmann M (2008) microRNA-122 stimulates translation of hepatitis C virus RNA. *EMBO J* **27:** 3300–3310

Heuckeroth RO, Towler DA, Adams SP, Glaser L & Gordon JI (1988) 11-(Ethylthio)undecanoic acid. A myristic acid analogue of altered hydrophobicity which is functional for peptide N-myristoylation with wheat germ and yeast acyltransferase. *Journal of Biological Chemistry* **263:** 2127–2133

Higgins DR (2001) Overview of Protein Expression in Pichia pastoris Hoboken, NJ, USA: John Wiley & Sons, Inc.

Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, Kusano-Arai O, Weyand S, Shimamura T, Nomura N, Cameron AD, Kobayashi T, Hamakubo T, Iwata S & Murata T (2012) G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* **482:** 237–240

Ho JD, Yeh R, Sandstrom A, Chorny I, Harries WEC, Robbins RA, Miercke LJW & Stroud RM (2009) Crystal structure of human aquaporin 4 at 1.8 A and its mechanism of conductance. *Proc. Natl. Acad. Sci. U.S.A.* **106:** 7437–7442

Hoffman B & Liu Q (2011) Hepatitis C viral protein translation: mechanisms and implications in developing antivirals. *Liver International* **31:** 1449–1467

Honda M, Ping LH, Rijnbrand RC, Amphlett E, Clarke B, Rowlands D & Lemon SM (1996) Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. *Virology* **222:** 31–42

Houghton M (2011) Prospects for prophylactic and therapeutic vaccines against the hepatitis C viruses. *Immunol. Rev.* **239:** 99–108

Hsu M, Zhang J, Flint M, Logvinoff C, Cheng-Mayer C, Rice CM & McKeating JA (2003) Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. *PNAS* **100:** 7271–7276

Hu N-J, Rada H, Rahman N, Nettleship JE, Bird L, Iwata S, Drew D, Cameron AD & Owens RJ (2014) GFP-Based Expression Screening of Membrane Proteins in Insect Cells Using the Baculovirus System. In *Liposomes*, Weissig V (ed) pp 197–209. New York, NY: Springer New York

Huang Y, Staschke K, De Francesco R & Tan S-L (2007) Phosphorylation of hepatitis C virus NS5A nonstructural protein: A new paradigm for phosphorylation-dependent viral RNA replication? *Virology* **364:** 1–9

Ishihara G, Goto M, Saeki M, Ito K, Hori T, Kigawa T, Shirouzu M & Yokoyama S (2005) Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors. *Protein Expr. Purif.* **41:** 27–37

Jackson RC & Blobel G (1977) Post-translational cleavage of presecretory proteins with an extract of rough microsomes from dog pancreas containing signal peptidase activity. *PNAS* **74:** 5598–5602

Jackson RJ & Hunt T (1982) Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. *Methods in Enzymology* **96:** 50–74

Jain MK & Zakim D (1987) The spontaneous incorporation of proteins into preformed bilayers. *Biochim. Biophys. Acta* **906:** 33–68

Jarecki BW, Makino S-I, Beebe ET, Fox BG & Chanda B (2013) Function of Shaker potassium channels produced by cell-free translation upon injection into Xenopus oocytes. *Sci. Rep.* **3:**

Jasti J, Furukawa H, Gonzales EB & Gouaux E (2007) Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. *Nature* **449:** 316–323

Jirasko V, Montserret R, Appel N, Janvier A, Eustachi L, Brohm C, Steinmann E, Pietschmann T, Penin F & Bartenschlager R (2008) Structural and Functional Characterization of Nonstructural Protein 2 for Its Role in Hepatitis C Virus Assembly. *Journal of Biological Chemistry* **283:** 28546–28562

Jirasko V, Montserret R, Lee J-Y, Gouttenoire J, Moradpour D, Penin F & Bartenschlager R (2010) Structural and Functional Studies of Nonstructural Protein 2 of the Hepatitis C Virus Reveal Its Key Role as Organizer of Virion Assembly. *PLoS Pathog* **6:** e1001233

Johnson AE & van Waes MA (1999) The translocon: a dynamic gateway at the ER membrane. *Annu. Rev. Cell Dev. Biol.* **15:** 799–842

Johnson JE & Cornell RB (1999) Amphitropic proteins: regulation by reversible membrane interactions (review). *Mol Membr Biol* **16:** 217–235

Jones DM, Patel AH, Targett-Adams P & McLauchlan J (2009) The hepatitis C virus NS4B protein can trans-complement viral RNA replication and modulates production of infectious virus. *Journal of Virology* **83:** 2163–2177

Jopling CL, Schütz S & Sarnow P (2008) Position-Dependent Function for a Tandem MicroRNA miR-122-Binding Site Located in the Hepatitis C Virus RNA Genome. *Cell Host and Microbe* **4:** 77–85

Jopling CL, Yi M, Lancaster AM, Lemon SM & Sarnow P (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309:** 1577–1581

Junge F, Haberstock S, Roos C, Stefer S, Proverbio D, Dötsch V & Bernhard F (2011) Advances in cell-free protein synthesis for the functional and structural analysis of membrane proteins. *New BIOTECHNOLOGY* **28:** 262–271

Junge F, Schneider B, Reckel S, Schwarz D, Dötsch V & Bernhard F (2008) Large-scale production of functional membrane proteins. *Cell. Mol. Life Sci.* **65:** 1729–1755

Kameda A, Morita EH, Sakurai K, Naiki H & Goto Y (2009) NMR-based characterization of a refolding intermediate of β2-microglobulin labeled using a wheat germ cell-free system. *Protein Sci.* **18:** 1592–1601

Kamura N, Sawasaki T, Kasahara Y, Takai K & Endo Y (2005) Selection of 5′-untranslated sequences that enhance initiation of translation in a cell-free protein synthesis system from wheat embryos. *Bioorganic & Medicinal Chemistry Letters* **15:** 5402–5406

Kanno T, Kitano M, Kato R, Omori A, Endo Y & Tozawa Y (2007) Sequence specificity and efficiency of protein N-terminal methionine elimination in wheat-embryo cell-free system. *Protein Expr. Purif.* **52:** 59–65

Katz FN, Rothman JE, Lingappa VR, Blobel G & Lodish HF (1977) Membrane assembly in vitro: Synthesis, glycosylatio, and asymmetric insertion of a transmembrane protein. *PNAS* **74:** 3278–3282

Katze MG, Kwieciszewski B, Goodlett DR, Blakely CM, Neddermann P, Tan S-L & Aebersold R (2000) Ser2194 Is a Highly Conserved Major Phosphorylation Site of the Hepatitis C Virus Nonstructural Protein NS5A. *Virology* **278:** 501–513

Katzen F, Chang G & Kudlicki W (2005) The past, present and future of cell-free protein synthesis. *Trends in Biotechnology* **23:** 150–156

Kaul A, Stauffer S, Berger C, Pertel T, Schmitt J, Kallis S, Zayas Lopez M, Lohmann V, Luban J & Bartenschlager R (2009) Essential Role of Cyclophilin A for Hepatitis C Virus Replication and Virus Production and Possible Link to Polyprotein Cleavage Kinetics. *PLoS Pathog* **5:** e1000546

Kawasaki T, Gouda MD, Sawasaki T, Takai K & Endo Y (2003) Efficient synthesis of a disulfide-containing protein through a batch cell-free system from wheat germ. *European Journal of Biochemistry* **270:** 4780–4786

Kawate T, Michel JC, Birdsong WT & Gouaux E (2009) Crystal structure of the ATP-gated P2X4 ion channel in the closed state. *Nature* **460:** 592–598

Kayali Z & Schmidt WN (2014) Finally sofosbuvir: an oral anti-HCV drug with wide performance capability. *Pharmacogenomics and Personalized Medecine* **7:** 387–398

Kelly SM, Jess TJ & Price NC (2005) How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1751:** 119–139

Khan AG, Whidby J, Miller MT, Scarborough H, Zatorski AV, Cygan A, Price AA, Yost SA, Bohannon CD, Jacob J, Grakoui A & Marcotrigiano J (2014) Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. *Nature* **509:** 381–384

Kigawa T, Yabuki T, Matsuda N, Matsuda T, Nakajima R, Tanaka A & Yokoyama S (2004) Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *J. Struct. Funct. Genomics* **5:** 63–68

Kim M & Song E (2010) Iron transport by proteoliposomes containing mitochondrial F1Fo ATP synthase isolated from rat heart. *Biochimie* **92:** 333–342

Kim S, Jeon T-J, Oberai A, Yang D, Schmidt JJ & Bowie JU (2005) Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *PNAS* **102:** 14278–14283

Kim S, Welsch C, Yi M & Lemon SM (2011) Regulation of the production of infectious genotype 1a hepatitis C virus by NS5A domain III. *Journal of Virology* **85:** 6645–6656

Klammt C, Löhr F, Schafer B, Haase W, Dötsch V, Ruterjans H, Glaubitz C & Bernhard F (2004) High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.* **271:** 568–580

Klammt C, Schwarz D, Fendler K, Haase W, Dötsch V & Bernhard F (2005) Evaluation of detergents for the soluble expression of α-helical and β-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS Journal* **272:** 6024–6038

Klammt C, Srivastava A, Eifler N, Junge F, Beyermann M, Schwarz D, Michel H, Doetsch V & Bernhard F (2007) Functional analysis of cell-free-produced human endothelin B receptor reveals transmembrane segment 1 as an essential area for ET-1 binding and homodimer formation. *FEBS Journal* **274:** 3257–3269

Kobilka BK (1990) The role of cytosolic and membrane factors in processing of the human beta-2 adrenergic receptor following translocation and glycosylation in a cell-free system. *Journal of Biological Chemistry* **265:** 7610–7618

Kohno T (2009) NMR Assignment Method for Amide Signals with Cell-Free Protein Synthesis System. In *Liposomes*, Weissig V (ed) pp 113–126. Totowa, NJ: Humana Press

Kolykhalov AA, Mihalik K, Feinstone SM & Rice CM (2000) Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo. *Journal of Virology* **74:** 2046–2051

Kong AM, Speed CJ, O'Malley CJ, Layton MJ, Meehan T, Loveland KL, Cheema S, Ooms LM & Mitchell CA (2000) Cloning and Characterization of a 72-kDa Inositol-polyphosphate 5-Phosphatase Localized to the Golgi Network. *Journal of Biological Chemistry* **275:** 24052–24064

Kong L, Giang E, Nieusma T, Kadam RU, Cogburn KE, Hua Y, Dai X, Stanfield RL, Burton DR, Ward AB, Wilson IA & Law M (2013) Hepatitis C Virus E2 Envelope Glycoprotein Core Structure. *Science* **342:** 1090–1094

Kopp M, Murray CL, Jones CT & Rice CM (2010) Genetic analysis of the carboxy-terminal region of the hepatitis C virus core protein. *Journal of Virology* **84:** 1666–1673

Kottler ML, Counis R & Degrelle H (1989) Sex steroid-binding protein: identification and comparison of the primary product following cell-free translation of human and monkey (Macaca fascicularis) liver RNA. *J. Steroid Biochem.* **33:** 201–207

Koutsoudakis G, Kaul A, Steinmann E, Kallis S, Lohmann V, Pietschmann T & Bartenschlager R (2006) Characterization of the Early Steps of Hepatitis C Virus Infection by Using Luciferase Reporter Viruses. *Journal of Virology* **80:** 5308–5320

Kovtun O, Mureev S, Johnston W & Alexandrov K (2010) Towards the Construction of Expressed Proteomes Using a Leishmania tarentolae Based Cell-Free Expression System. *PLoS ONE* **5:** e14388

Kovtun O, Mureev S, Jung W, Kubala MH, Johnston W & Alexandrov K (2011) Leishmania cell-free protein expression system. *Methods* **55:** 58–64

Kunert B, Gardiennet C & Lacabanne D (2014) Efficient and stable reconstitution of the ABC transporter BmrA for solid-state NMR studies. *Frontiers in Molecular Biosciences* **5:** 1–11

Kunji ERS, Chan KW, Slotboom D-J, Floyd S, O'Connor R & Monné M (2005) Eukaryotic membrane protein overproduction in Lactococcus lactis. *Curr. Opin. Biotechnol.* **16:** 546–551

Kunji ERS, Slotboom D-J & Poolman B (2003) Lactococcus lactis as host for overproduction of functional membrane proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1610:** 97–108

Kunkel M, Lorinczi M, Rijnbrand R, Lemon SM & Watowich SJ (2001) Self-Assembly of Nucleocapsid-Like Particles from Recombinant Hepatitis C Virus Core Protein. *Journal of Virology* **75:** 2119–2129

Kushima Y, Wakita T & Hijikata M (2010) A Disulfide-Bonded Dimer of the Core Protein of Hepatitis C Virus Is Important for Virus-Like Particle Production. *Journal of Virology* **84:** 9118–9127

Lambert SM, Langley DR, Garnett JA, Angell R, Hedgethorne K, Meanwell NA & Matthews SJ (2014) The crystal structure of NS5A domain 1 from genotype 1a reveals new clues to the mechanism of action for dimeric HCV inhibitors. *Protein Sci.* **23:** 723–734

Lanford RE, Bigger C, Bassett S & Klimpel G (2001) The chimpanzee model of hepatitis C virus infections. *ILAR J* **42:** 117–126

Lange CM, Bellecave P, Dao Thi VL, Tran HTL, Penin F, Moradpour D & Gouttenoire J (2014) Determinants for membrane association of the hepatitis C virus NS2 protease domain. *Journal of Virology* **88:** 6519–6523

Langlais C, Guilleaume B, Wermke N, Scheuermann T, Ebert L, LaBaer J & Korn B (2007) A systematic approach for testing expression of human full-length proteins in cell-free expression systems. *BMC Biotechnol* **7:** 64

Langland JO, Langland LA, Browning KS & Roth DA (1996) Phosphorylation of plant eukaryotic initiation factor-2 by the plant-encoded double-stranded RNA-dependent protein kinase, pPKR, and inhibition of protein synthesis in vitro. *Journal of Biological Chemistry* **271:** 4539–4544

Lavanchy D (2011) Evolving epidemiology of hepatitis C virus. *Clin. Microbiol. Infect.* **17:** 107–115

Lavie M, Goffard A & Dubuisson J (2007) Assembly of a functional HCV glycoprotein heterodimer. *Curr Issues Mol Biol* **9:** 71–86

Lavie M, Penin F & Dubuisson J (2015) HCV envelope glycoproteins in evirion assembly and entry. *Future Virology* **10:** 297–312

le Maire M, Aggerbeck LP, Monteilhet C & Andersen JP (1986) The use of high-performance liquid chromatography for the determination of size and molecular weight of proteins: a caution and a list of membrane proteins suitable. *Analytical Biochemistry* **154:** 525–535

le Maire M, Champeil P & Møller JV (2000) Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim. Biophys. Acta* **1508:** 86–111

Lemon SM & Honda M (1997) Internal ribosome entry sites within the RNA genomes of hepatitis C virus and other flaviviruses. *Seminars in virology* **8:** 274–288

Lerat H, Honda M, Beard MR, Loesch K, Sun J, Yang Y, Okuda M, Gosert R, Xiao SY, Weinman SA & Lemon SM (2002) Steatosis and liver cancer in transgenic mice expressing the structural and nonstructural proteins of hepatitis C virus. *Gastroenterology* **122:** 352–365

Levy D, Bluzat A, Seigneuret M & Rigaud JL (1990) A systematic study of liposome and proteoliposome reconstitution involving Bio-Bead-mediated Triton X-100 removal. *Biochim. Biophys. Acta* **1025:** 179–190

Li E, Wimley WC & Hristova K (2012a) Transmembrane helix dimerization: Beyond the search for sequence motifs. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1818:** 183–193

Li K, Foy E, Ferreon JC & Nakamura M (2005) Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. In

Li M, Hays FA, Roe-Zurz Z, Vuong L, Kelly L, Ho C-M, Robbins RM, Pieper U, O'Connell JD III, Miercke LJW, Giacomini KM, Sali A & Stroud RM (2009) Selecting Optimum Eukaryotic Integral Membrane Proteins for Structure Determination by Rapid Expression and Solubilization Screening. *Journal of Molecular Biology* **385:** 820–830

Li YP, Ramirez S & Jensen SB (2012b) Highly efficient full-length hepatitis C virus genotype 1 (strain TN) infectious culture system. In

Liang Y, Ye H, Kang CB & Yoon HS (2007) Domain 2 of Nonstructural Protein 5A (NS5A) of Hepatitis C Virus Is Natively Unfolded †. *Biochemistry* **46:** 11550–11558

Lifshitz Y, Petrovich E, Haviv H, Goldshleger R, Tal DM, Garty H & Karlish SJD (2007) Purification of the Human α2 Isoform of Na,K-ATPase Expressed in Pichia pastoris.Stabilization by Lipids and FXYD1 †. *Biochemistry* **46:** 14937–14950

Lim Y-S & Hwang SB (2011) Hepatitis C virus NS5A protein interacts with phosphatidylinositol 4-kinase type IIIalpha and regulates viral propagation. *J. Biol. Chem.* **286:** 11290–11298

Lindenbach BD (2013) Virion Assembly and Release. In *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*, Bartenschlager R (ed) pp 199–218. Berlin, Heidelberg: Springer Berlin Heidelberg

Lindenbach BD, Evans MJ, Syder AJ, Wölk B, Tellinghuisen TL, Liu CC, Maruyama T, Hynes RO, Burton DR, McKeating JA & Rice CM (2005) Complete replication of hepatitis C virus in cell culture. *Science* **309:** 623–626

Lindenbach BD, Meuleman P, Ploss A, Vanwolleghem T, Syder AJ, McKeating JA, Lanford RE, Feinstone SM, Major ME, Leroux-Roels G & Rice CM (2006) Cell culture-grown hepatitis C virus is infectious in vivo and can be recultured in vitro. *PNAS* **103:** 3805–3809

Lindenbach BD, Prágai BM, Montserret R, Beran RKF, Pyle AM, Penin F & Rice CM (2007) The C terminus of hepatitis C virus NS4A encodes an electrostatic switch that regulates NS5A hyperphosphorylation and viral replication. *Journal of Virology* **81:** 8905–8918

Lingappa VR, Lingappa JR, Prasad R, Ebner KE & Blobel G (1978) Coupled cell-free synthesis, segregation, and core glycosylation of a secretory protein. *PNAS* **75:** 2338–2342

Liu S, Yang W, Shen L, Turner JR, Coyne CB & Wang T (2009a) Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection. *Journal of Virology* **83:** 2011–2014

Liu Z, Yang F, Robotham JM & Tang H (2009b) Critical Role of Cyclophilin A and Its Prolyl-Peptidyl Isomerase Activity in the Structure and Function of the Hepatitis C Virus Replication Complex. *Journal of Virology* **83:** 6554–6565

Lohmann V (2013) Hepatitis C Virus RNA Replication. In *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*, Bartenschlager R (ed) pp 167–198. Berlin, Heidelberg: Springer Berlin Heidelberg

Lohmann V, Körner F, Koch J, Herian U, Theilmann L & Bartenschlager R (1999) Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285:** 110–113

Lomize AL, Pogozheva ID, Lomize MA & Mosberg HI (2007) The role of hydrophobic interactions in positioning of peripheral proteins in membranes. *BMC Struct Biol* **7:** 44

Long AR, O'Brien CC & Alder NN (2012) The Cell-Free Integration of a Polytopic Mitochondrial Membrane Protein into Liposomes Occurs Cotranslationally and in a Lipid-Dependent Manner. *PLoS ONE* **7:** e46332

Lorenz IC, Marcotrigiano J, Dentzer TG & Rice CM (2006) Structure of the catalytic domain of the hepatitis C virus NS2-3 protease. *Nature* **442:** 831–835

Love RA, Brodsky O, Hickey MJ, Wells PA & Cronin CN (2009) Crystal Structure of a Novel Dimeric Form of NS5A Domain I Protein from Hepatitis C Virus. *Journal of Virology* **83:** 4395–4403

Lu J & Deutsch C (2005a) Secondary Structure Formation of a Transmembrane Segment in Kv Channels †. *Biochemistry* **44:** 8230–8243

Lu J & Deutsch C (2005b) Folding zones inside the ribosomal exit tunnel. *Nat Struct Mol Biol* **12:** 1123–1129

Lundin M, Lindström H, Grönwall C & Persson MAA (2006) Dual topology of the processed hepatitis C virus protein NS4B is influenced by the NS5A protein. *Journal of General Virology* **87:** 3263–3272

Lyford LK & Rosenberg RL (1999) Cell-free expression and functional reconstitution of homo-oligomeric alpha7 nicotinic acetylcholine receptors into planar lipid bilayers. *Journal of Biological Chemistry* **274:** 25675–25681

Lyukmanova EN, Shenkarev ZO, Khabibullina NF, Kopeina GS, Shulepko MA, Paramonov AS, Mineev KS, Tikhonov RV, Shingarova LN, Petrovskaya LE, Dolgikh DA, Arseniev AS & Kirpichnikov MP (2012) Lipid–protein nanodiscs for cell-free production of integral membrane proteins in a soluble and folded state: Comparison with detergent micelles, bicelles and liposomes. *BBA - Biomembranes* **1818:** 349–358

Ma Y, Anantpadma M, Timpe JM, Shanmugam S, Singh SM, Lemon SM & Yi M (2011) Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins. *Journal of Virology* **85:** 86–97

Madin K, Sawasaki T, Ogasawara T & Endo Y (2000) A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *PNAS* **97:** 559–564

Mailly L, Robinet E, Meuleman P, Baumert TF & Zeisel MB (2013) Hepatitis C virus infection and related liver disease: the quest for the best animal model. *Front Microbiol* **4:** 213

Makino S-I, Goren MA, Fox BG & Markley JL (2009) Cell-Free Protein Synthesis Technology in NMR High-Throughput Structure Determination. In *Liposomes*, Weissig V (ed) pp 127–147. Totowa, NJ: Humana Press

Makuc J, Cappellaro C & Boles E (2004) Co-expression of a mammalian accessory trafficking protein enables functional expression of the rat MCT1 monocarboxylate transporter in Saccharomyces cerevisiae. *FEMS Yeast Res.* **4:** 795–801

Marascio N, Torti C, Liberto MC & Focà A (2014) Update on different aspects of HCV variability: focus on NS5B polymerase. *BMC Infectious Diseases* **14:** S1

Marcia M, Langer JD, Parcej D, Vogel V, Peng G & Michel H (2010) Characterizing a monotopic membrane enzyme. Biochemical, enzymatic and crystallization studies on Aquifex aeolicus sulfide:quinone oxidoreductase. *BBA - Biomembranes* **1798:** 2114–2123

Marcu K & Dudock B (1974) Characterization of a highly efficient protein synthesizing system derived from commercial wheat germ. *Nucleic Acids Research* **1:** 1385–1397

Martin DN & Uprichard SL (2013) Identification of transferrin receptor 1 as a hepatitis C virus entry factor. *PNAS* **110:** 10777–10782

Martin KH, Grosenbach DW, Franke CA & Hruby DE (1997) Identification and analysis of three myristylated vaccinia virus late proteins. *Journal of Virology* **71:** 5218–5226

Masaki T, Suzuki R, Murakami K, Aizaki H, Ishii K, Murayama A, Date T, Matsuura Y, Miyamura T, Wakita T & Suzuki T (2008) Interaction of hepatitis C virus nonstructural protein 5A with core protein is critical for the production of infectious virus particles. *Journal of Virology* **82:** 7964–7976

Maslennikov I, Klammt C & Hwang E (2010) Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. *PNAS* **107:** 10902–10907

Matar-Merheb R, Rhimi M, Leydier A, Huché F, Galián C, Desuzinges-Mandon E, Ficheux D, Flot D, Aghajari N, Kahn R, Di Pietro A, Jault J-M, Coleman AW & Falson P (2011) Structuring Detergents for Extracting and Stabilizing Functional Membrane Proteins. *PLoS ONE* **6:** e18036

McLauchlan J, Lemberg MK, Hope G & Martoglio B (2002) Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. *EMBO J* **21:** 3980–3988

Mercer DF, Schiller DE, Elliott JF, Douglas DN, Hao C, Rinfret A, Addison WR, Fischer KP, Churchill TA, Lakey JR, Tyrrell DL & Kneteman NM (2001) Hepatitis C virus replication in mice with chimeric human livers. *Nat Med* **7:** 927–933

Meuleman P, Libbrecht L, De Vos R, de Hemptinne B, Gevaert K, Vandekerckhove J, Roskams T & Leroux-Roels G (2005) Morphological and biochemical characterization of a human liver in a uPA-SCID mouse chimera. *Hepatology* **41:** 847–856

Meylan E, Curran J, Hofmann K, Moradpour D, Binder M, Bartenschlager R & Tschopp J (2005) Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437:** 1167–1172

Michael Houghton (2009) The long and winding road leading to the identification of the hepatitis C virus. *Journal of Hepatology* **51:** 939–948

Mikami S, Masutani M, Sonenberg N, Yokoyama S & Imataka H (2006) An efficient mammalian cell-free translation system supplemented with translation factors. *Protein Expr. Purif.* **46:** 348–357

Mimms LT, Zampighi G, Nozaki Y, Tanford C & Reynolds JA (1981) Phospholipid vesicle formation and transmembrane protein incorporation using octyl glucoside. *Biochemistry* **20:** 833–840

Miot M & Betton J-M (2011) Reconstitution of the Cpx signaling system from cell-free synthesized proteins. *New BIOTECHNOLOGY* **28:** 277–281

Miroux B & Walker JE (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology* **260:** 289–298

Miyanari Y, Atsuzawa K, Usuda N, Watashi K, Hishiki T, Zayas M, Bartenschlager R, Wakita T, Hijikata M & Shimotohno K (2007) The lipid droplet is an important organelle for hepatitis C virus production. *Nat Cell Biol* **9:** 1089–1097

Montigny C, Penin F, Lethias C & Falson P (2004) Overcoming the toxicity of membrane peptide expression in bacteria by upstream insertion of Asp-Pro sequence. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1660:** 53–65

Montserret R, Saint N, Vanbelle C, Salvay AG, Simorre J-P, Ebel C, Sapay N, Renisio J-G, Böckmann A, Steinmann E, Pietschmann T, Dubuisson J, Chipot C & Penin F (2010) NMR structure and ion channel activity of the p7 protein from hepatitis C virus. *J. Biol. Chem.* **285:** 31446–31461

Moradpour D & Penin F (2013) Hepatitis C Virus Proteins: From Structure to Function. In *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*, Bartenschlager R (ed) pp 113–142. Berlin, Heidelberg: Springer Berlin Heidelberg

Moradpour D, Brass V & Penin F (2005) Function follows form: The structure of the N-terminal domain of HCV NS5A. *Hepatology* **42:** 732–735

Moradpour D, Brass V, Bieck E, Friebe P, Gosert R, Blum HE, Bartenschlager R, Penin F & Lohmann V (2004a) Membrane Association of the RNA-Dependent RNA Polymerase Is Essential for Hepatitis C Virus RNA Replication. *Journal of Virology* **78:** 13278–13284

Moradpour D, Evans MJ, Gosert R, Yuan Z, Blum HE, Goff SP, Lindenbach BD & Rice CM (2004b) Insertion of green fluorescent protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus replication complexes. *Journal of Virology* **78:** 7400–7409

Moradpour D, Penin F & Rice CM (2007) Replication of hepatitis C virus. *Nat Rev Micro* **5:** 453–463

Morikawa K, Gouttenoire J, Hernandez C, Dao Thi VL, Tran HTL, Lange CM, Dill MT, Heim MH, Donzé O, Penin F, Quadroni M & Moradpour D (2013) Quantitative proteomics identifies the membrane-associated peroxidase GPx8 as a cellular substrate of the hepatitis C virus NS3-4A protease. *Hepatology* **59:** 423–433

Morikawa K, Lange CM, Gouttenoire J, Meylan E, Brass V, Penin F & Moradpour D (2011) Nonstructural protein 3-4A: the Swiss army knife of hepatitis C virus. *Journal of Viral Hepatitis* **18:** 305–315

Morita EH, Sawasaki T, Tanaka R, Endo Y & Kohno T (2003) A wheat germ cell-free system is a novel way to screen protein folding and function. *Protein Sci.* **12:** 1216–1221

Morita EH, Shimizu M, Ogasawara T, Endo Y, Tanaka R & Kohno T (2004) A novel way of amino acid-specific assignment in (1)H-(15)N HSQC spectra with a wheat germ cell-free protein synthesis system. *J. Biomol. NMR* **30:** 37–45

Mosley RT, Edwards TE, Murakami E, Lam AM, Grice RL, Du J, Sofia MJ, Furman PA & Otto MJ (2012) Structure of Hepatitis C Virus Polymerase in Complex with Primer-Template RNA. *Journal of Virology* **86:** 6503–6511

Mothes W, Jungnickel B, Brunner J & Rapoport TA (1998) Signal sequence recognition in cotranslational translocation by protein components of the endoplasmic reticulum membrane. *J. Cell Biol.* **142:** 355–364

Muir AJ (2014) The Rapid Evolution of Treatment Strategies for Hepatitis C. *The American Journal of Gastroenterology* **109:** 628–635

Mureev S, Kovtun O, Nguyen UTT & Alexandrov K (2009) Species-independent translational leaders facilitate cell-free expression. *Nat Biotechnol* **27:** 747–752

Murray CL, Jones CT & Rice CM (2008) Architects of assembly: roles of Flaviviridae non-structural proteins in virion morphogenesis. *Nat Rev Micro* **6:** 699–708

Nagy JK, Kuhn Hoffmann A, Keyes MH, Gray DN, Oxenoid K & Sanders CR (2001) Use of amphipathic polymers to deliver a membrane protein to lipid bilayers. *FEBS LETTERS* **501:** 115–120

Nakamura S (1993) Possible role of phosphorylation in the function of chicken MyoD1. *Journal of Biological Chemistry* **268:** 11670–11677

Neddermann P, Quintavalle M, Di Pietro C, Clementi A, Cerretani M, Altamura S, Bartholomew L & De Francesco R (2004) Reduction of hepatitis C virus NS5A hyperphosphorylation by selective inhibition of cellular kinases activates viral RNA replication in cell culture. *Journal of Virology* **78:** 13306–13314

Nekrasova OV, Wulfson AN, Tikhonov RV, Yakimov SA, Simonova TN, Tagvey AI, Dolgikh DA, Ostrovsky MA & Kirpichnikov MP (2010) A new hybrid protein for production of recombinant bacteriorhodopsin in Escherichia coli. *Journal of Biotechnology* **147:** 145–150

Netzer WJ & Hartl FU (1997) Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388:** 343–349

Newstead S, Kim H, Heijne von G, Iwata S & Drew D (2007) High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae. *PNAS* **104:** 13936–13941

Nielsen SU, Bassendine MF, Burt AD, Martin C, Pumeechockchai W & Toms GL (2006) Association between hepatitis C virus and very-low-density lipoprotein (VLDL)/LDL analyzed in iodixanol density gradients. *Journal of Virology* **80:** 2418–2428

Niepmann M (2013) Hepatitis C Virus RNA Translation. In *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*, Bartenschlager R (ed) pp 143–166. Berlin, Heidelberg: Springer Berlin Heidelberg

Ning Z, Hawley B, Seebun D & Figeys D (2014) APols-Aided Protein Precipitation: A Rapid Method for Concentrating Proteins for Proteomic Analysis. *J Membrane Biol* **247:** 941–947

Nitta S, Sakamoto N, Nakagawa M, Kakinuma S, Mishima K, Kusano-Kitazume A, Kiyohashi K, Murakawa M, Nishimura-Sakurai Y, Azuma S, Tasaka-Fujita M, Asahina Y, Yoneyama M, Fujita T & Watanabe M (2013) Hepatitis C virus NS4B protein targets STING and abrogates RIG-I-mediated type I interferon-dependent innate immunity. *Hepatology* **57:** 46–58

Noirot C, Habenstein B, Bousset L, Melki R, Meier BH, Endo Y, Penin F & Böckmann A (2010) Wheat-germ cell-free production of prion proteins for solid-state NMR structural studies. *New BIOTECHNOLOGY***:** 1–7

Nomura S-IM, Kondoh S, Asayama W, Asada A, Nishikawa S & Akiyoshi K (2008) Direct preparation of giant proteo-liposomes by in vitro membrane protein synthesis. *Journal of Biotechnology* **133:** 190–195

Nozawa A & Tozawa Y (2014) Incorporation of adenine nucleotide transporter, Ant1p, into proteoliposomes facilitates ATP translocation and activation of encapsulated luciferase. *JBIOSC* **118:** 130–133

Nozawa A, Nanamiya H, Miyata T, Linka N, Endo Y, Weber APM & Tozawa Y (2007) A cell-free translation and proteoliposome reconstitution system for functional analysis of plant solute transporters. *Plant and Cell Physiology* **48:** 1815–1820

Nozawa A, Ogasawara T, Matsunaga S, Iwasaki T, Sawasaki T & Endo Y (2011) Production and partial purification of membrane proteins using a liposome-supplemented wheat cell-free translation system. *BMC Biotechnol* **11:** 35

Oehler V, Filipe A, Montserret R, da Costa D, Brown G, Penin F & McLauchlan J (2012) Structural Analysis of Hepatitis C Virus Core-E1 Signal Peptide and Requirements for Cleavage of the Genotype 3a Signal Sequence by Signal Peptide Peptidase. *Journal of Virology* **86:** 7818–7828

Ogasawara T, Sawasaki T, Morishita R, Ozawa A, Madin K & Endo Y (1999) A new class of enzyme acting on damaged ribosomes: ribosomal RNA apurinic site specific lyase found in wheat germ. *EMBO J* **18:** 6522–6531

Okamoto K, Mori Y, Komoda Y, Okamoto T, Okochi M, Takeda M, Suzuki T, Moriishi K & Matsuura Y (2008) Intramembrane processing by signal peptide peptidase regulates the membrane localization of hepatitis C virus core protein and viral propagation. *Journal of Virology* **82:** 8349–8361

Omari El K, Iourin O, Kadlec J, Sutton G, Harlos K, Grimes JM & Stuart DI (2014) Unexpected structure for the N-terminal domain of hepatitis C virus envelope glycoprotein E1. *Nature Communications* **5:** 1–5

Op De Beeck A, Voisset C, Bartosch B, Ciczora Y, Cocquerel L, Keck Z, Foung S, Cosset FL & Dubuisson J (2004) Characterization of Functional Hepatitis C Virus Envelope Glycoproteins. *Journal of Virology* **78:** 2994–3002

OuYang B, Xie S, Berardi MJ, Zhao X, Dev J, Yu W, Sun B & Chou JJ (2013) Unusual architecture of the p7 channel from hepatitis C virus. *Nature* **498:** 521–525

Ouyang EC, Wu CH, Walton C, Promrat K & Wu GY (2001) Transplantation of human hepatocytes into tolerized genetically immunocompetent rats. *WJG* **7:** 324–330

Owen DM, Huang H, Ye J & Gale M Jr. (2009) Apolipoprotein E on hepatitis C virion facilitates infection through interaction with low-density lipoprotein receptor. *Virology* **394:** 99–108

Park K-H, Berrier C, Lebaupain F, Pucci B, Popot J-L, Ghazi A & Zito F (2007) Fluorinated and hemifluorinated surfactants as alternatives to detergents for membrane protein cell-free synthesis. *Biochem. J.* **403:** 183

Park K-H, Billon-Denis E, Dahmane T, Lebaupain F, Pucci B, Breyton C & Zito F (2011) In the cauldron of cell-free synthesis of membrane proteins: playing with new surfactants. *New BIOTECHNOLOGY* **28:** 255–261

Paul D, Bartenschlager R & McCormick C (2015) The predominant species of nonstructural protein 4B in hepatitis C virus-replicating cells is not palmitoylated. *Journal of General Virology*

Paul D, Madan V & Bartenschlager R (2014) Hepatitis C Virus RNA Replication and Assembly: Living on the Fat of the Land. *Cell Host and Microbe* **16:** 569–579

Paul D, Romero-Brey I, Gouttenoire J, Stoitsova S, Krijnse-Locker J, Moradpour D & Bartenschlager R (2011) NS4B Self-Interaction through Conserved C-Terminal Elements Is Required for the Establishment of Functional Hepatitis C Virus Replication Complexes. *Journal of Virology* **85:** 6963–6976

Pawlotsky J-M (2013) Hepatitis C virus: standard-of-care treatment. *Adv. Pharmacol.* **67:** 169–215

Pawlotsky J-M (2014) New Hepatitis C Virus (HCV) Drugs and the Hope for a Cure: Concepts in Anti-HCV Drug Development. *Semin Liver Dis* **34:** 022–029

Pawlotsky JM (2003) Hepatitis C virus genetic variability: pathogenic and clinical implications Clinics in liver disease

Pawlotsky JM & Germanidis G (1999) The non-structural 5A protein of hepatitis C virus. *Journal of Viral Hepatitis* **6:** 343–356

Pelham HR & Jackson RJ (1976) An efficient mRNA-dependent translation system from reticulocyte lysates. *Eur. J. Biochem.* **67:** 247–256

Penin F, Brass V, Ramboarina S, Montserret R, Ficheux D, Blum HE, Bartenschlager R & Moradpour D (2004a) Structure and Function of the Membrane Anchor Domain of Hepatitis C Virus Nonstructural Protein 5A. *Journal of Biological Chemistry* **279:** 40835–40843

Penin F, Dubuisson J, Rey FA, Moradpour D & Pawlotsky J-M (2004b) Structural biology of hepatitis C virus. *Hepatology* **39:** 5–19

Periasamy A, Shadiac N, Amalraj A, Garajová S, Nagarajan Y, Waters S, Mertens HDT & Hrmova M (2013) Cell-free protein synthesis of membrane (1,3)-β-d-glucan (curdlan) synthase: Co-translational insertion in liposomes and reconstitution in nanodiscs. *BBA - Biomembranes* **1828:** 743–757

Pérard J, Leyrat C, Baudin F, Drouet E & Jamin M (2013) Structure of the full-length HCV IRES in solution. *Nature Communications* **4:** 1612

Picard M, Dahmane T, Garrigos M, Gauron C, Giusti F, le Maire M, Popot J-L & Champeil P (2006) Protective and Inhibitory Effects of Various Types of Amphipols on the Ca 2+-ATPase from Sarcoplasmic Reticulum: A Comparative Study †. *Biochemistry* **45:** 1861–1869

Pileri P, Uematsu Y, Campagnoli S, Galli G, Falugi F, Petracca R, Weiner AJ, Houghton M, Rosa D, Grandi G & Abrignani S (1998) Binding of hepatitis C virus to CD81. *Science* **282:** 938–941

Ploss A, Evans MJ, Gaysinskaya VA, Panis M, You H, de Jong YP & Rice CM (2009) Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* **457:** 882–886

Popescu C-I, Callens N, Trinel D, Roingeard P, Moradpour D, Descamps V, Duverlie G, Penin F, Héliot L, Rouillé Y & Dubuisson J (2011) NS2 Protein of Hepatitis C Virus Interacts with Structural and Non-Structural Proteins towards Virus Assembly. *PLoS Pathog* **7:** e1001278

Popescu C-I, Riva L, Vlaicu O, Farhat R, Rouillé Y & Dubuisson J (2014) Hepatitis C Virus Life Cycle and Lipid Metabolism. *Biology* **3:** 892–921

Popot J-L (2010) Amphipols, Nanodiscs, and Fluorinated Surfactants: Three Nonconventional Approaches to Studying Membrane Proteins in Aqueous Solutions. *Annu. Rev. Biochem.* **79:** 737–775

Quick M & Wright EM (2002) Employing Escherichia coli to functionally express, purify, and characterize a human transporter. *PNAS* **99:** 8597–8601

Quintavalle M, Sambucini S, Summa V, Orsatti L, Talamo F, De Francesco R & Neddermann P (2007) Hepatitis C virus NS5A is a direct substrate of casein kinase I-alpha, a cellular kinase identified by inhibitor affinity chromatography using specific NS5A hyperphosphorylation inhibitors. *Journal of Biological Chemistry* **282:** 5536–5544

Rai R & Deval J (2011) New opportunities in anti-hepatitis C virus drug discovery: targeting NS4B. *Antiviral Res.* **90:** 93–101

Ramière C, Rodriguez J, Enache LS, Lotteau V, André P & Diaz O (2014) Activity of hexokinase is increased by its interaction with hepatitis C virus protein NS5A. *Journal of Virology* **88:** 3246–3254

Ramirez S, Li Y-P, Jensen SB, Pedersen J, Gottwein JM & Bukh J (2014) Highly efficient infectious cell culture of three hepatitis C virus genotype 2b strains and sensitivity to lead protease, nonstructural protein 5A, and polymerase inhibitors. *Hepatology* **59:** 395–407

Ramón A & Marín M (2011) Advances in the production of membrane proteins in Pichia pastoris. *Biotechnology Journal* **6:** 700–706

Ranaghan MJ, Schwall CT, Alder NN & Birge RR (2011) Green Proteorhodopsin Reconstituted into Nanoscale Phospholipid Bilayers (Nanodiscs) as Photoactive Monomers. *J. Am. Chem. Soc.* **133:** 18318–18327

Reed KE & Rice CM (1999) Identification of the major phosphorylation site of the hepatitis C virus H strain NS5A protein as serine 2321. *Journal of Biological Chemistry* **274:** 28011–28018

Rehermann B (2009) Hepatitis C virus versus innate and adaptive immune responses: a tale of coevolution and coexistence. *J. Clin. Invest.* **119:** 1745–1754

Reiss S, Harak C, Romero-Brey I, Radujkovic D, Klein R, Ruggieri A, Rebhan I, Bartenschlager R & Lohmann V (2013) The Lipid Kinase Phosphatidylinositol-4 Kinase III Alpha Regulates the Phosphorylation Status of Hepatitis C Virus NS5A. *PLoS Pathog* **9:** e1003359

Reiss S, Rebhan I, Backes P, Romero-Brey I, Erfle H, Matula P, Kaderali L, Poenisch M, Blankenburg H, Hiet M-S, Longerich T, Diehl S, Ramirez F, Balla T, Rohr K, Kaul A, Bühler S, Pepperkok R, Lengauer T, Albrecht M, et al (2011) Recruitment and Activation of a Lipid Kinase by Hepatitis C Virus NS5A Is Essential for Integrity of the Membranous Replication Compartment. *Cell Host and Microbe* **9:** 32–45

Rigaud J-L & Lévy D (2003) Reconstitution of membrane proteins into liposomes. *Meth. Enzymol.* **372:** 65–86

Rigaud JL, Bluzat A & Buschlen S (1983) Incorporation of bacteriorhodopsin into large unilamellar liposomes by reverse phase evaporation. *Biochemical and Biophysical Research Communications* **111:** 373–382

Rigaud JL, Pitard B & Levy D (1995) Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. *Biochimica et Biophysica Acta* **1231:** 223–246

Ritchie TK, Grinkova YV, Bayburt TH, Denisov IG, Zolnerciks JK, Atkins WM & Sligar SG (2009) Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. In *Liposomes, Part F* pp 211–231. Elsevier

Roberts BE & Paterson BM (1973) Efficient Translation of Tobacco Mosaic Virus RNA and Rabbit Globin 9S RNA in a Cell-Free System from Commercial Wheat Germ. *PNAS* **70:** 2330–2334

Romero-Brey I, Merz A, Chiramel A, Lee J-Y, Chlanda P, Haselman U, Santarella-Mellwig R, Habermann A, Hoppe S, Kallis S, Walther P, Antony C, Krijnse-Locker J & Bartenschlager R (2012) Three-Dimensional Architecture and Biogenesis of Membrane Structures Associated with Hepatitis C Virus Replication. *PLoS Pathog* **8:** e1003056

Roosild TP, Greenwald J, Vega M, Castronovo S, Riek R & Choe S (2005) NMR Structure of Mistic, a Membrane-Integrating Protein for Membrane Protein Expression. *Science* **307:** 1317–1321

Rosenberg RL & East JE (1992) Cell-free expression of functional Shaker potassium channels. *Nature* **360:** 166–169

Rosnoblet C, Fritzinger B, Legrand D, Launay H, Wieruszeski JM, Lippens G & Hanoulle X (2012) Hepatitis C Virus NS5B and Host Cyclophilin A Share a Common Binding Site on NS5A. *Journal of Biological Chemistry* **287:** 44249–44260

Rubenstein JL & Chappell TG (1983) Construction of a synthetic messenger RNA encoding a membrane protein. *J. Cell Biol.* **96:** 1464–1469

Russ ZN & Dueber JE (2014) Cell-free protein synthesis: Search for the happy middle. *Biotechnology Journal*

Sachse R, Dondapati SK, Fenz SF, Schmidt T & Kubick S (2014) Membrane protein synthesis in cell-free systems: From bio-mimetic systems to bio-membranes. *FEBS LETTERS* **588:** 2774–2781

Sahdev S, Khattar SK & Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol. Cell. Biochem.* **307:** 249–264

Sainz B, Barretto N, Martin DN, Hiraga N, Imamura M, Hussain S, Marsh KA, Yu X, Chayama K, Alrefai WA & Uprichard SL (2012) Identification of the Niemann-Pick C1–like 1 cholesterol absorption receptor as a new hepatitis C virus entry factor. *Nat Med* **18:** 281–285

Sakurai N, Moriya K, Suzuki T, Sofuku K, Mochiki H, Nishimura O & Utsumi T (2007) Detection of co- and posttranslational protein N-myristoylation by metabolic labeling in an insect cell-free protein synthesis system. *Anal. Biochem.* **362:** 236–244

Salloum S & Tai AW (2012) Treating hepatitis C infection by targeting the host. *Translational Research* **159:** 421–429

Sandgren EP, Palmiter RD, Heckel JL, Daugherty CC, Brinster RL & Degen JL (1991) Complete hepatic regeneration after somatic deletion of an albumin-plasminogen activator transgene. *Cell* **66:** 245–256

Sansuk K, Balog CIA, van der Does AM, Booth R, de Grip WJ, Deelder AM, Bakker RA, Leurs R & Hensbergen PJ (2008) GPCR Proteomics: Mass Spectrometric and Functional Analysis of Histamine H 1Receptor after Baculovirus-Driven and in VitroCell Free Expression. *J. Proteome Res.* **7:** 621–629

Santolini E, Migliaccio G & La Monica N (1994) Biosynthesis and biochemical properties of the hepatitis C virus core protein. *Journal of Virology* **68:** 3631–3641

Sawasaki T, Hasegawa Y, Morishita R, Seki M, Shinozaki K & Endo Y (2004) Genome-scale, biochemical annotation method based on the wheat germ cell-free protein synthesis system. *Phytochemistry* **65:** 1549–1555

Sawasaki T, Hasegawa Y, Tsuchimochi M, Kamura N, Ogasawara T, Kuroita T & Endo Y (2002a) A bilayer cell-free protein synthesis system for high-throughput screening of gene products. *FEBS LETTERS* **514:** 102–105

Sawasaki T, Ogasawara T, Morishita R & Endo Y (2002b) A cell-free protein synthesis system for high-throughput proteomics. *PNAS* **99:** 14652–14657

Scarselli E, Ansuini H, Cerino R, Roccasecca RM, Acali S, Filocamo G, Traboni C, Nicosia A, Cortese R & Vitelli A (2002) The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J* **21:** 5017–5025

Schaedler TA, Tong Z & van Veen HW (2012) The multidrug transporter LmrP protein mediates selective calcium efflux. *J. Biol. Chem.* **287:** 27682–27690

Scharff-Poulsen P & Pedersen PA (2013) Saccharomyces cerevisiaeâ€“Based Platform for Rapid Production and Evaluation of Eukaryotic Nutrient Transporters and Transceptors for Biochemical Studies and Crystallography. *PLoS ONE* **8:** e76851

Scheele G & Blackburn P (1979) Role of mammalian RNase inhibitor in cell-free protein synthesis. *PNAS* **76:** 4898–4902

Schmidt TG & Skerra A (2007) The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* **2:** 1528–1535

Schmidt-Mende J, Bieck E, Hugle T, Penin F, Rice CM, Blum HE & Moradpour D (2001) Determinants for Membrane Association of the Hepatitis C Virus RNA-dependent RNA Polymerase. *Journal of Biological Chemistry* **276:** 44052–44063

Schneider B, Junge F, Shirokov VA, Durst F, Schwarz D, Dötsch V & Bernhard F (2009) Membrane Protein Expression in Cell-Free Systems. In *Liposomes*, Weissig V (ed) pp 165–186. Totowa, NJ: Humana Press

Schregel V, Jacobi S, Penin F & Tautz N (2009) Hepatitis C virus NS2 is a protease stimulated by cofactor domains in NS3. *Proc. Natl. Acad. Sci. U.S.A.* **106:** 5342–5347

Schwarz D, Daley D, Beckhaus T, Dötsch V & Bernhard F (2010) Cell-free expression profiling of E. coli inner membrane proteins. *Proteomics* **10:** 1762–1779

Schwarz D, Dötsch V & Bernhard F (2008) Production of membrane proteins using cell-free expression systems. *Proteomics* **8:** 3933–3946

Schwarz D, Junge F, Durst F, Frölich N, Schneider B, Reckel S, Sobhanifar S, Dötsch V & Bernhard F (2007) Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat Protoc* **2:** 2945–2957

Seddon AM, Curnow P & Booth PJ (2004) Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1666:** 105–117

Sevova ES, Goren MA, Schwartz KJ, Hsu F-F, Turk J, Fox BG & Bangs JD (2010) Cell-free synthesis and functional characterization of sphingolipid synthases from parasitic trypanosomatid protozoa. *J. Biol. Chem.* **285:** 20580–20587

Shadiac N, Nagarajan Y, Waters S & Hrmova M (2013) Close allies in membrane protein research: Cell-free synthesis and nanotechnology. *Mol Membr Biol* **30:** 229–245

Shavinskaya A, Boulant S, Penin F, McLauchlan J & Bartenschlager R (2007) The Lipid Droplet Binding Domain of Hepatitis C Virus Core Protein Is a Major Determinant for Efficient Virus Assembly. *Journal of Biological Chemistry* **282:** 37158–37169

Shen X, Hacker DL, Baldi L & Wurm FM (2014) Virus-free transient protein production in Sf9 cells. **171:** 61–70 Available at: http://dx.doi.org/10.1016/j.jbiotec.2013.11.018

Shimakami T, Hijikata M, Luo H, Ma YY, Kaneko S, Shimotohno K & Murakami S (2004) Effect of interaction between hepatitis C virus NS5A and NS5B on hepatitis C virus RNA replication with the hepatitis C virus replicon. *Journal of Virology* **78:** 2738–2748

Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K & Ueda T (2001) Cell-free translation reconstituted with purified components. *Nat Biotechnol* **19:** 751–755

Shirota Y, Luo H, Qin W, Kaneko S, Yamashita T, Kobayashi K & Murakami S (2002) Hepatitis C Virus (HCV) NS5A Binds RNA-dependent RNA Polymerase (RdRP) NS5B and Modulates RNA-dependent RNA Polymerase Activity. *Journal of Biological Chemistry* **277:** 11149–11155

Signorell GA, Kaufmann TC, Kukulski W, Engel A & Rémigy H-W (2007) Controlled 2D crystallization of membrane proteins using methyl-β-cyclodextrin. *J. Struct. Biol.* **157:** 321–328

Simmonds P (2004) Genetic diversity and evolution of hepatitis C virus - 15 years on. *Journal of General Virology* **85:** 3173–3188

Simmonds P (2013) The Origin of Hepatitis C Virus. In *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*, Bartenschlager R (ed) pp 1–15. Berlin, Heidelberg: Springer Berlin Heidelberg

Song Y, Friebe P, Tzima E, Junemann C, Bartenschlager R & Niepmann M (2006) The Hepatitis C Virus RNA 3'-Untranslated Region Strongly Enhances Translation Directed by the Internal Ribosome Entry Site. *Journal of Virology* **80:** 11579–11588

Spiess M (1995) Heads or tails--what determines the orientation of proteins in the membrane. *FEBS LETTERS* **369:** 76–79

Spirin AS (2004) High-throughput cell-free systems for synthesis of functionally active proteins. *Trends in Biotechnology* **22:** 538–545

Spirin AS, Baranov VI, Ryabova LA, Ovodov SY & Alakhov YB (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **242:** 1162–1164

Standfuss J, Xie G, Edwards PC, Burghammer M, Oprian DD & Schertler GFX (2007) Crystal structure of a thermally stable rhodopsin mutant. *Journal of Molecular Biology* **372:** 1179–1188

Stapleford KA & Lindenbach BD (2011) Hepatitis C virus NS2 coordinates virus particle assembly through physical interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes. *Journal of Virology* **85:** 1706–1717

Steen A, Wiederhold E, Gandhi T, Breitling R & Slotboom DJ (2011) Physiological Adaptation of the Bacterium Lactococcus lactis in Response to the Production of Human CFTR. *Molecular & Cellular Proteomics* **10:** M000052–MCP200–M000052–MCP200

Steinmann E & Pietschmann T (2010) Hepatitis C Virus P7—A Viroporin Crucial for Virus Assembly and an Emerging Target for Antiviral Therapy. *Viruses* **2:** 2078–2095

Steinmann E, Brohm C, Kallis S, Bartenschlager R & Pietschmann T (2008) Efficient trans-Encapsidation of Hepatitis C Virus RNAs into Infectious Virus-Like Particles. *Journal of Virology* **82:** 7034–7046

Stueber D, Ibrahimi I, Cutler D, Dobberstein B & Bujard H (1984) A novel in vitro transcription-translation system: accurate and efficient synthesis of single proteins from cloned DNA sequences. *EMBO J* **3:** 3143–3148

Sun H, Zhang A, Yan G, Piao C, Li W, Sun C, Wu X, Li X, Chen Y & Wang X (2013) Metabolomic analysis of key regulatory metabolites in hepatitis C virus-infected tree shrews. *Mol. Cell Proteomics* **12:** 710–719

Takahashi H, Nozawa A, Seki M, Shinozaki K, Endo Y & Sawasaki T (2009) A simple and high-sensitivity method for analysis of ubiquitination and polyubiquitination based on wheat cell-free protein synthesis. *BMC Plant Biol* **9:** 39

Takai K, Sawasaki T & Endo Y (2010) Practical cell-free protein synthesis system using purified wheat embryos. *Nat Protoc* **5:** 227–238

Tamm LK, Arora A & Kleinschmidt JH (2001) Structure and Assembly of -Barrel Membrane Proteins. *Journal of Biological Chemistry* **276:** 32399–32402

Tasaka M, Sakamoto N, Itakura Y, Nakagawa M, Itsui Y, Sekine-Osajima Y, Nishimura-Sakurai Y, Chen C-H, Yoneyama M, Fujita T, Wakita T, Maekawa S, Enomoto N & Watanabe M (2007) Hepatitis C virus non-structural proteins responsible for suppression of the RIG-I/Cardif-induced interferon response. *Journal of General Virology* **88:** 3323–3333

Tate CG, Haase J, Baker C, Boorsma M, Magnani F, Vallis Y & Williams DC (2003) Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1610:** 141–153

Tellinghuisen TL, Foss KL & Treadaway J (2008) Regulation of Hepatitis C Virion Production via Phosphorylation of the NS5A Protein. *PLoS Pathog* **4:** e1000032

Tellinghuisen TL, Marcotrigiano J & Rice CM (2005) Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature* **435:** 374–379

Tellinghuisen TL, Marcotrigiano J, Gorbalenya AE & Rice CM (2004) The NS5A protein of hepatitis C virus is a zinc metalloprotein. *Journal of Biological Chemistry* **279:** 48576–48587

Thibeault D, Maurice R, Pilote L, Lamarre D & Pause A (2001) In vitro characterization of a purified NS2/3 protease variant of hepatitis C virus. *Journal of Biological Chemistry* **276:** 46678–46684

Thompson AA, Zou A, Yan J, Duggal R, Hao W, Molina D, Cronin CN & Wells PA (2009) Biochemical Characterization of Recombinant Hepatitis C Virus Nonstructural Protein 4B: Evidence for ATP/GTP Hydrolysis and Adenylate Kinase Activity †. *Biochemistry* **48:** 906–916

Tonelli K, Singarapu KK & Markley JL (2011) Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J. Biomol. NMR* **51:** 467–476

Torizawa T, Shimizu M, Taoka M, Miyano H & Kainosho M (2004) Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol. *J. Biomol. NMR* **30:** 311–325

Tribet C, Audebert R & Popot JL (1996) Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *PNAS* **93:** 15047–15050

Trometer C & Falson P (2009) Mammalian Membrane Protein Expression in Baculovirus-Infected Insect Cells. In *Liposomes*, Weissig V (ed) pp 105–117. Totowa, NJ: Humana Press

Tschantz WR, Pfeifer ND, Meade CL, Wang L, Lanzetti A, Kamath AV, Berlioz-Seux F & Hashim MF (2008) Expression, purification and characterization of the human membrane transporter protein OATP2B1 from Sf9 insect cells. *Protein Expr. Purif.* **57:** 163–171

Tu L, Khanna P & Deutsch C (2014) Transmembrane Segments Form Tertiary Hairpins in the Folding Vestibule of the Ribosome. *Journal of Molecular Biology* **426:** 185–198

Ulmschneider JP, Smith JC, White SH & Ulmschneider MB (2011) In Silico Partitioning and Transmembrane Insertion of Hydrophobic Peptides under Equilibrium Conditions. *J. Am. Chem. Soc.* **133:** 15487–15495

Ulmschneider MB, Ulmschneider JP, Schiller N, Wallace BA, Heijne von G & White SH (2014) Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion. *Nature Communications* **5:** 4863

van Oers MM, Pijlman GP & Vlak JM (2014) Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *Journal of General Virology* **96:** 6–23

Vassilaki N, Friebe P, Meuleman P, Kallis S, Kaul A, Paranhos-Baccalà G, Leroux-Roels G, Mavromara P & Bartenschlager R (2008) Role of the hepatitis C virus core+1 open reading frame and core cis-acting RNA elements in viral RNA translation and replication. *Journal of Virology* **82:** 11503–11515

Vercauteren K, de Jong YP & Meuleman P (2014) HCV animal models and liver disease. *Journal of Hepatology* **61:** S26–S33

Verdegem D, Badillo A, Wieruszeski JM, Landrieu I, Leroy A, Bartenschlager R, Penin F, Lippens G & Hanoulle X (2011) Domain 3 of NS5A Protein from the Hepatitis C Virus Has Intrinsic -Helical Propensity and Is a Substrate of Cyclophilin A. *Journal of Biological Chemistry* **286:** 20441–20454

Vieyres G, Brohm C, Friesland M, Gentzsch J, Wolk B, Roingeard P, Steinmann E & Pietschmann T (2013) Subcellular Localization and Function of an Epitope-Tagged p7 Viroporin in Hepatitis C Virus-Producing Cells. *Journal of Virology* **87:** 1664–1678

Vieyres G, Dubuisson J & Pietschmann T (2014) Incorporation of Hepatitis C Virus E1 and E2 Glycoproteins: The Keystones on a Peculiar Virion. *Viruses* **6:** 1149–1187

Vieyres G, Thomas X, Descamps V, Duverlie G, Patel AH & Dubuisson J (2010) Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. *Journal of Virology* **84:** 10159–10168

Vinarov DA, Lytle BL, Peterson FC, Tyler EM, Volkman BF & Markley JL (2004) Cell-free protein production and labeling protocol for NMR-based structural proteomics. *Nat Meth* **1:** 149–153

Vinarov DA, Newman CLL & Markley JL (2006) Wheat germ cell-free platform for eukaryotic protein production. *FEBS Journal* **273:** 4160–4169

Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Högbom M, van Wijk KJ, Slotboom DJ, Persson JO & de Gier J-W (2008) Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci. U.S.A.* **105:** 14371–14376

Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, Murthy K, Habermann A, Kräusslich H-G, Mizokami M, Bartenschlager R & Liang TJ (2005) Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature* **11:** 791–796

Walter P & Johnson AE (1994) Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10:** 87–119

Wang L & Sigworth FJ (2009) Structure of the BK potassium channel in a lipid membrane from electron cryomicroscopy. *Nature* **461:** 292–295

Wang L & Tonggu L (2015) Membrane protein reconstitution for functional and structural studies. *Sci. China Life Sci.* **58:** 66–74

Wang X, Liu J, Zhao WM & Zhao K-N (2010) Translational comparison of HPV58 long and short L1 mRNAs in yeast (Saccharomyces cerevisiae) cell-free system. *JBIOSC* **110:** 58–65

Wang X, Liu J, Zheng Y, Li J, Wang H, Zhou Y, Qi M, Yu H, Tang W & Zhao WM (2008) An optimized yeast cell-free system: Sufficient for translation of human papillomavirus 58 L1 mRNA and assembly of virus-like particles. *Journal of Bioscience and Bioengineering* **106:** 8–15

Weiss HM & Grisshammer R (2002) Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli. *Eur. J. Biochem.* **269:** 82–92

Westbrook RH & Dusheiko G (2014) Natural history of hepatitis C. *Journal of Hepatology* **61:** S58–S68

White JF, Trinh LB, Shiloach J & Grisshammer R (2004) Automated large-scale purification of a G protein-coupled receptor for neurotensin. *FEBS LETTERS* **564:** 289–293

White SH & Heijne GV (2004) The machinery of membrane protein assembly. *Current Opinion in Structural Biology* **14:** 397–404

Wilson RC & Doudna JA (2013) Molecular Mechanisms of RNA Interference. *Annu. Rev. Biophys.* **42:** 217–239

Witherell G (2001) In vitro translation using HeLa extract. *Curr Protoc Cell Biol* **Chapter 11:** Unit 11.8–11.8.10

Woolhead CA, Johnson AE & Bernstein HD (2006) Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Molecular Cell* **22:** 587–598

Xie ZC, Riezu-Boj JI, Lasarte JJ, Guillen J, Su JH, Civeira MP & Prieto J (1998) Transmission of hepatitis C virus infection to tree shrews. *Virology* **244:** 513–520

Yamane D, McGivern DR, Masaki T & Lemon SM (2013) Liver Injury and Disease Pathogenesis in Chronic Hepatitis C. In *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*, Bartenschlager R (ed) pp 263–288. Berlin, Heidelberg: Springer Berlin Heidelberg

Yamauchi S, Fusada N, Hayashi H, Utsumi T, Uozumi N, Endo Y & Tozawa Y (2010) The consensus motif for N-myristoylation of plant proteins in a wheat germ cell-free translation system. *FEBS Journal* **277:** 3596–3607

Yang M, Li N, Li F, Zhu Q, Liu X, Han Q, Wang Y, Chen Y, Zeng X, Lv Y, Zhang P, Yang C & Liu Z (2013) Xanthohumol, a main prenylated chalcone from hops, reduces liver damage and modulates oxidative reaction and apoptosis in hepatitis C virus infected Tupaia belangeri. *International Immunopharmacology* **16:** 466–474

Yao F, Svensjö T, Winkler T, Lu M, Eriksson C & Eriksson E (1998) Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. *Hum. Gene Ther.* **9:** 1939–1950

Yi M & Lemon SM (2009) Genotype 1a HCV (H77S) infection system. *Methods Mol. Biol.* **510:** 337–346

Young HS, Rigaud JL, Lacapere JJ, Reddy LG & Stokes DL (1997) How to make tubular crystals by reconstitution of detergent-solubilized Ca2(+)-ATPase. *Biophysj* **72:** 2545–2558

Yu G-Y, Lee K-J, Gao L & Lai MMC (2006) Palmitoylation and polymerization of hepatitis C virus NS4B protein. *Journal of Virology* **80:** 6013–6023

Zeisel MB, Felmlee DJ & Baumert TF (2013) Hepatitis C Virus Entry. In *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*, Bartenschlager R (ed) pp 87–112. Berlin, Heidelberg: Springer Berlin Heidelberg

Zhong J, Gastaminza P, Cheng G, Kapadia S, Kato T, Burton DR, Wieland SF, Uprichard SL, Wakita T & Chisari FV (2005) Robust hepatitis C virus infection in vitro. *PNAS* **102:** 9294–9299

Zhou X & Graham TR (2009) Reconstitution of phospholipid translocase activity with purified Drs2p, a type-IV P-type ATPase from budding yeast. *Proc. Natl. Acad. Sci. U.S.A.* **106:** 16586–16591

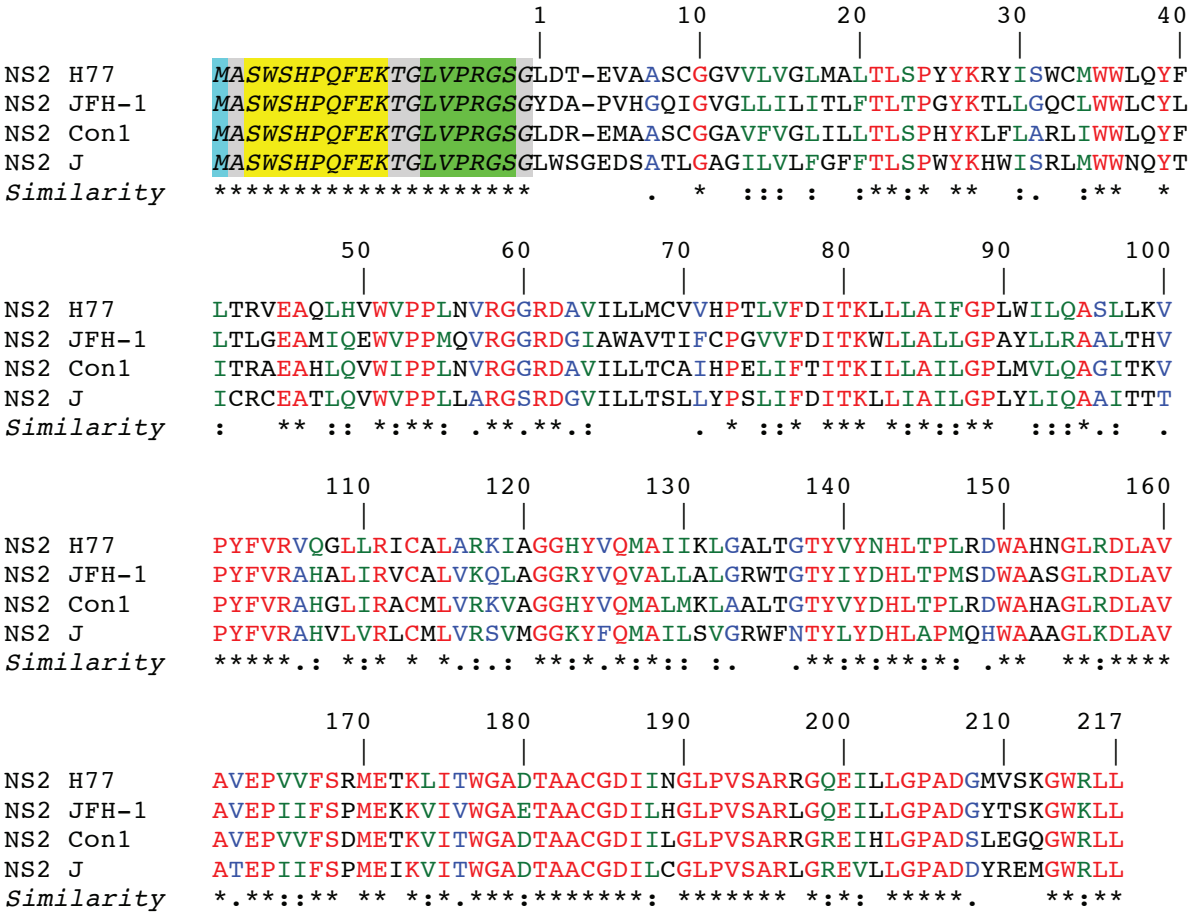
Zhou XX, Li WF, Ma GX & Pan YJ (2006) The nisin-controlled gene expression system: Construction, application and improvements. *Biotechnology Advances* **24:** 285–295

Zoonens M & Popot J-L (2014) Amphipols for Each Season. *J Membrane Biol* **247:** 759–796

Zubay G (1973) In vitro synthesis of protein in microbial systems. *Annu. Rev. Genet.* **7:** 267–287

APPENDIX 1 - Amino acid sequences

Sequences of full-length NS2 constructs



Sequences of full-length NS2 proteins from HCV strains H77, JFH-1, Con1 and 452 (subtypes 1a, 2a, 1b and 3a, GenBank accession numbers AF009606, AB047639, AJ238799 and DQ437509, respectively) were aligned with Clustal W (https://npsa-prabi.ibcp.fr/cgi-bin/align_clustalw.pl). Similarity : identical, highly similar and similar residues at each position of NS2 protein are indicated with asterisks, colons and dots, respectively, according to the Clustal W conventions. Amino acids are numbered with respect to their position within NS2 protein of strain H77. The N-terminal methionine is highlighted in *blue*, the *Strep*-tag II sequence at the N-terminus in *yellow*, the thrombine cleavage site sequence in *green* and linker sequences in *grey*.

Sequences of NS2^{pro} constructs

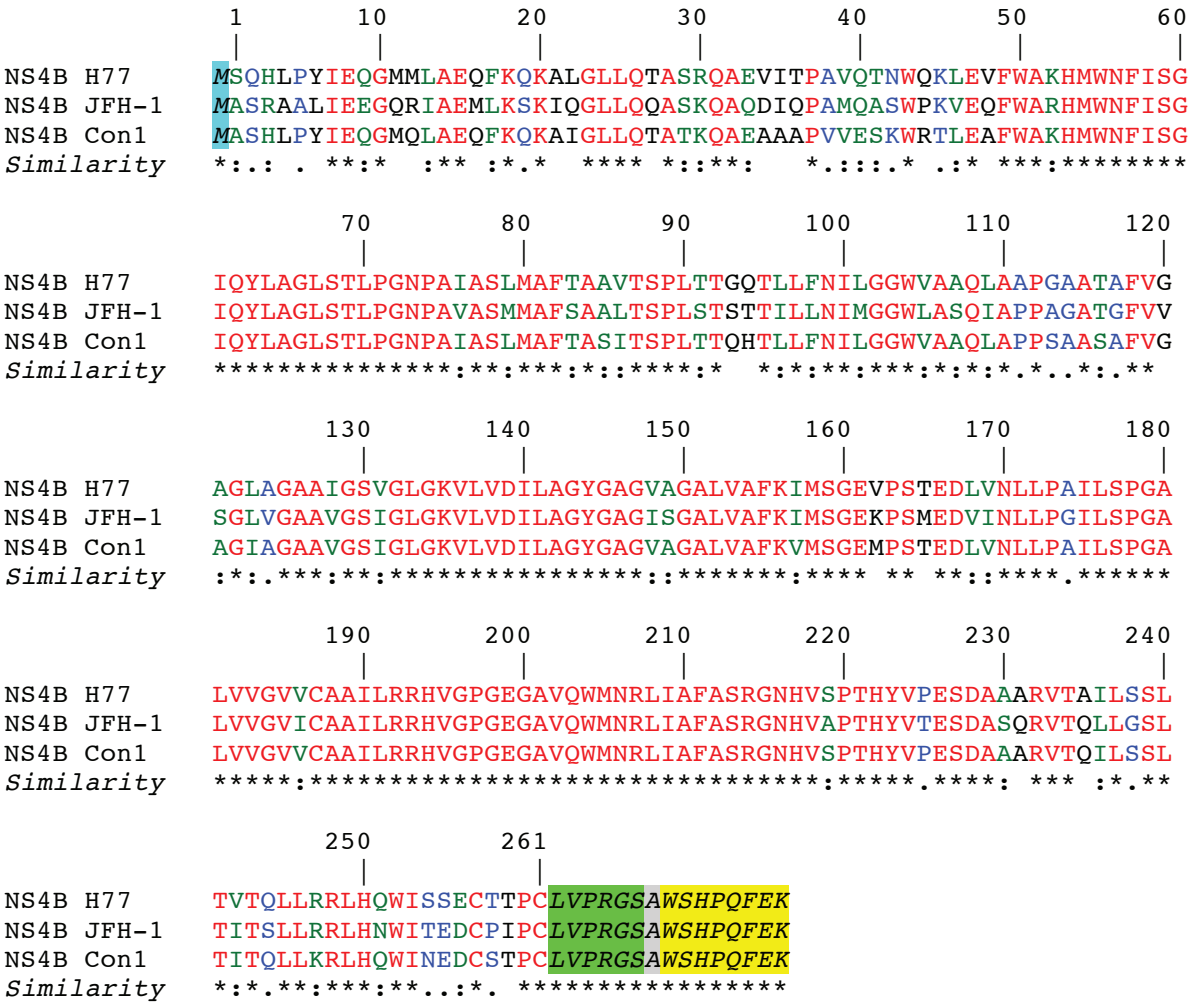
```
                          94    100        110         120         130      140
                          |     |          |           |           |        |
NS2 JFH-1       MASWSHPQFEKTGRAALTHVPYFVRAHALIRVCALVKQLAGGRYVQVALLALGRWTGTYI
NS2 Con1        MASWSHPQFEKTGQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYV
Similarity      **************:*.:*:********.***.* **:::***:***:**: *.  ****:

                        150        160        170        180        190        200
                          |          |          |          |          |          |
NS2 JFH-1       YDHLTPMSDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGQE
NS2 Con1        YDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGRE
Similarity      ******: *** :************::** **.***.***:*******: ******* *:*

                        210    217
                          |      |
NS2 JFH-1       ILLGPADGYTSKGWKLL
NS2 Con1        IHLGPADSLEGQGWRLL
Similarity      * *****.  .:**:**
```

Sequences of NS2^{pro} proteins from HCV strains JFH-1 and Con1 (subtypes 2a and 1b, GenBank accession numbers AB047639 and AJ238799, respectively) were aligned with Clustal W (https://npsa-prabi.ibcp.fr/cgi-bin/align_clustalw.pl). Similarity : identical, highly similar and similar residues at each position of NS2^{pro} protein are indicated with asterisks, colons and dots, respectively, according to the Clustal W conventions. Amino acids are numbered with respect to their position within NS2^{pro} protein. The N-terminal methionine is highlighted in *blue*, the *Strep*-tag II sequence at the N-terminus in *yellow* and linker sequences in *grey*.

Sequences of full-length NS4B constructs



Sequences of full-length NS4B proteins from HCV strains H77, JFH-1 and Con1 (subtypes 1a, 2a and 1b, GenBank accession numbers AF009606, AB047639 and AJ238799, respectively) were aligned with Clustal W (https://npsa-prabi.ibcp.fr/cgi-bin/align_clustalw.pl). Similarity : identical, highly similar and similar residues at each position of NS4B protein are indicated with asterisks, colons and dots, respectively, according to the Clustal W conventions. Amino acids are numbered with respect to their position within NS4B protein. The N-terminal methionine is highlighted in *blue*, the *Strep*-tag II sequence at the C-terminus in *yellow*, the thrombine cleavage site sequence in *green* and the linker sequence in *grey*.

Sequence of full-length NS5A construct

```
                1       10        20        30        40        50
                |       |         |         |         |         |
NS5A FL      MASGSWLRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCP

               60        70        80        90       100       110
                |        |         |         |         |         |
NS5A FL      CGAQITGHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEY

              120       130       140       150       160       170
                |        |         |         |         |         |
NS5A FL      VEVTRVGDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEVTFLVG

              180       190       200       210       220       230
                |        |         |         |         |         |
NS5A FL      LNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPS

              240       250       260       270       280       290
                |        |         |         |         |         |
NS5A FL      LKATCTTRHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFEPLQAEEDEREVSV

              300       310       320       330       340       350
                |        |         |         |         |         |
NS5A FL      PAEILRRSRKFPRAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPAKAPPIPPPRRK

              360       370       380       390       400       410
                |        |         |         |         |         |
NS5A FL      RTVVLSESTVSSALAELATKTFGSSESSAVDSGTATASPDQPSDDGDAGSDVESYSSMPP

              420       430       440    447
                |        |         |      |
NS5A FL      LEGEPGDPDLSDGSWSTVSEEASEDVVCCSAWSHPQFEK
```

Sequence of full-length (FL) NS5A (1-447) protein from HCV strain Con1 (subtype 1b, GenBank accession number AJ238799). Amino acids are numbered with respect to their position within NS5A protein. The N-terminal methionine is highlighted in *blue*, the *Strep*-tag II sequence at the C-terminus in *yellow* and linker sequences in *grey*. Note that an alanine was added between the N-terminal methionine and the sequence of the protein to have a better Kozak sequence.

Sequence of NS5A D1D2D3 construct

```
                30        40        50        60        70        80
                |         |         |         |         |         |
NS5A D1D2D3   MARLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPRTCSNTWHGT

                90       100       110       120       130       140
                |         |         |         |         |         |
NS5A D1D2D3   FPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNVKCPCQVPAP

               150       160       170       180       190       200
                |         |         |         |         |         |
NS5A D1D2D3   EFFTEVDGVRLHRYAPACKPLLREEVTFLVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPS

               210       220       230       240       250       260
                |         |         |         |         |         |
NS5A D1D2D3   HITAETAKRRLARGSPPSLASSSASQLSAPSLKATCTTRHDSPDADLIEANLLWRQEMGG

               270       280       290       300       310       320
                |         |         |         |         |         |
NS5A D1D2D3   NITRVESENKVVILDSFEPLQAEEDEREVSVPAEILRRSRKFPRAMPIWARPDYNPPLLE

               330       340       350       360       370       380
                |         |         |         |         |         |
NS5A D1D2D3   SWKDPDYVPPVVHGCPLPPAKAPPIPPPRRKRTVVLSESTVSSALAELATKTFGSSESSA

               390       400       410       420       430       440    447
                |         |         |         |         |         |      |
NS5A D1D2D3   VDSGTATASPDQPSDDGDAGSDVESYSSMPPLEGEPGDPDLSDGSWSTVSEEASEDVVCC


NS5A D1D2D3   SAWSHPQFEK
```

Sequence of NS5A D1D2D3 (30-447) protein from HCV strain Con1 (subtype 1b, GenBank accession number AJ238799). Amino acids are numbered with respect to their position within NS5A protein. The N-terminal methionine is highlighted in *blue*, the *Strep*-tag II sequence at the C-terminus in *yellow* and linker sequences in *grey*. Note that an alanine was added between the N-terminal methionine and the sequence of the protein to have a better Kozak sequence.

Sequence of NS5A AH-D1 construct

```
                1        10         20         30         40         50
                |        |          |          |          |          |
NS5A AH-D1      MASGSWLRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCP

               60        70         80         90        100        110
                |        |          |          |          |          |
NS5A AH-D1      CGAQITGHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEY

              120       130        140        150        160        170
                |        |          |          |          |          |
NS5A AH-D1      VEVTRVGDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEVTFLVG

              180       190        200        210  213
                |        |          |          |   |
NS5A AH-D1      LNQYLVGSQLPCEPEPDVAVLTSMLTDPSHITAETSAWSHPQFEK
```

Sequence of NS5A AH-D1 (1-213) protein from HCV strain Con1 (subtype 1b, GenBank accession number AJ238799). Amino acids are numbered with respect to their position within NS5A protein. The N-terminal methionine is highlighted in *blue*, the *Strep*-tag II sequence at the C-terminus in *yellow* and linker sequences in *grey*. Note that an alanine was added between the N-terminal methionine and the sequence of the protein to have a better Kozak sequence.

Sequence of NS5A D1 construct

```
                30         40         50         60         70         80
                 |         |          |          |          |          |
NS5A D1         MARLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQITGHVKNGSMRIVGPRTCSNTWHGT

                90        100        110        120        130        140
                 |         |          |          |          |          |
NS5A D1         FPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRVGDFHYVTGMTTDNVKCPCQVPAP

               150        160        170        180        190        200
                 |         |          |          |          |          |
NS5A D1         EFFTEVDGVRLHRYAPACKPLLREEVTFLVGLNQYLVGSQLPCEPEPDVAVLTSMLTDPS

               210  213
                 |   |
NS5A D1         HITAETSAWSHPQFEK
```

Sequence of NS5A D1 (30-213) protein from HCV strain Con1 (subtype 1b, GenBank accession number AJ238799). Amino acids are numbered with respect to their position within NS5A protein. The N-terminal methionine is highlighted in *blue*, the *Strep*-tag II sequence at the C-terminus in *yellow* and linker sequences in *grey*. Note that an alanine was added between the N-terminal methionine and the sequence of the protein to have a better Kozak sequence.

Sequence of full-length core construct (1-175)

```
                1       10        20        30        40        50        60
                |       |         |         |         |         |         |
Core J 1-175    MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG

                        70        80        90       100       110       120
                        |         |         |         |         |         |
Core J 1-175    RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG

                      130       140       150       160       170  175
                        |         |         |         |         |    |
Core J 1-175    KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSSAWSH

Core J 1-175    PQFEK
```

Sequence of full-length core (1-175) protein from HCV strain J (subtype 1b, GenBank accession number D90208). Amino acids are numbered with respect to their position within core protein. The *Strep*-tag II sequence at the C-terminus is highlighted in *yellow* and the linker sequence in *grey*.

Sequence of core D1 construct (1-120)

```
                1       10        20        30        40        50        60
                |       |         |         |         |         |         |
Core J 1-120    MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG

                        70        80        90       100       110       120
                        |         |         |         |         |         |
Core J 1-120    RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG

Core J 1-120    SAWSHPQFEK
```

Sequence of core D1 (1-120) protein from HCV strain J (subtype 1b, GenBank accession number D90208). Amino acids are numbered with respect to their position within core protein. The *Strep*-tag II sequence at the C-terminus is highlighted in *yellow* and the linker sequence in *grey*.

APPENDIX 2 - NS2 manuscript

Functional expression, purification, characterization, and membrane reconstitution of nonstructural protein 2 from hepatitis C virus

Marie-Laure Fogeron[1], David Paul[2], Vlastimil Jirasko[2], Roland Montserret[1], Denis Lacabanne[1], Jennifer Molle[1], Aurélie Badillo[1,3], Célia Boukadida[4], Sonia Georgeault[5], Philippe Roingeard[5,6], Annette Martin[4], Ralf Bartenschlager[2], François Penin[1#] & Anja Böckmann[1#]

[1]*Institut de Biologie et Chimie des Protéines, Bases Moléculaires et Structurales des Systèmes Infectieux, Labex Ecofect, UMR 5086 CNRS, Université de Lyon, Lyon, France*

[2]*Department of Infectious Diseases, Molecular Virology, Heidelberg University, Heidelberg, Germany and German Centre for Infection Research (DZIF), partner site Heidelberg, Heidelberg, Germany.*

[3]*RD-Biotech, Recombinant Protein Unit, Besançon, France*

[4] *Institut Pasteur, CNRS UMR 3569 Université Paris Diderot – Sorbonne Paris Cité Paris, France*

[5] *Plate-Forme RIO des Microscopies, PPF ASB, Université François Rabelais and CHRU de Tours, Tours, France*

[6] *INSERM U966, Universite François Rabelais and CHRU de Tours, Tours, France*

Corresponding authors: a.bockmann@ibcp.fr; f.penin@ibcp.fr

Abstract

Non-structural protein 2 (NS2) of the hepatitis C virus (HCV) is an integral membrane protein that contains a cysteine protease and that plays a central organizing role in assembly of infectious progeny virions. While the crystal structure of the protease domain has been solved, the NS2 full-length form remains biochemically and structurally uncharacterized because recombinant NS2 could not be prepared in sufficient quantities from cell-based systems. By using a wheat germ cell-free expression system, we successfully produced and purified milligram amounts of a detergent-solubilized form of full-length NS2 exhibiting protease activity and expected secondary structure content. Furthermore, immuno-electron microscopy analyses of reconstituted proteoliposomes demonstrated NS2 association with model membranes.

Keywords: hepatitis C virus, non-structural protein 2, cell-free protein expression, membrane protein, lipid reconstitution

Introduction

The HCV non-structural protein 2 (NS2) is critical for both HCV polyprotein processing and virus particle formation [1-3]. The C-terminal domain functions as a cysteine protease [4] that cleaves the NS2-NS3 junction to liberate fully functional NS3 thus promoting HCV RNA replication [5-7]. In addition, NS2 is thought to play a central organizing role during HCV assembly, likely involved in a complex network of interactions including E2, p7, NS3, and NS5A [8-12]. NS2 is a 217-amino-acid (aa) integral membrane protein that is bound to intracellular membranes *via* its hydrophobic N-terminal integral membrane domain (residues 1 to about 100). This domain is believed to comprise 3 transmembrane segments whose atomic structures have been studied separately by NMR (nuclear magnetic resonance) spectroscopy, allowing us to propose a topology model of full-length, membrane-associated NS2 [8, 13]. The x-ray crystal structure of the C-terminal protease domain displays a domain-swapped homodimer with composite catalytic triads including residues from the two chains [4]. This domain was shown to contribute to membrane association via two α-helices [14]. Although a wealth of *in cellulo* and functional studies has been carried out for NS2 [8, 13, 14], its biochemical analysis is hampered by the very low expression level of recombinant full-length NS2 using classical cell-based expression approaches [15]. We here present successful cell-free expression of the protein using wheat germ extracts (WGE) in the presence of detergent [16-20]. Moreover, we demonstrate that the protein can be obtained in a detergent-solubilized, well-folded, homogenous and functional form, and that it interacts with model membranes.

Results

NS2 can be efficiently produced in an active form using a wheat germ extract (WGE) cell-free expression system

Expression of NS2 with a *Strep*-tag II (ST) fused at its N-terminus in the WGE cell-free expression system relies on the production of synthetic, *in vitro* transcribed RNA encoding the protein sequence, which is added to a reaction mixture containing the WGE, together with energy-sources, energy-recycling enzymes and aa needed for protein synthesis [16, 21, 22]. Using this approach, NS2 from different HCV genotypes comprising isolates Con1 (genotype 1b), H77 (genotype1a), JFH-1 (genotype 2a) and 452 (genotype 3a) was successfully expressed (Figure 1A). In the absence of detergent, the protein is mainly insoluble as indicated by its localization in the pellet fraction. Only a minor portion of NS2 protein was found in the supernatant (denoted "SN-beads" since the tagged protein is concentrated by binding of the NS2 N-terminal *Strep*-tag II to magnetic *Strep*-Tactin beads). However, in the presence of 0.1 % lauryl maltose neopentyl glycol (MNG-3) detergent, fully solubilized JFH-1 and Con1 NS2 were obtained (see also reference [16]) while H77 and 452 NS2 partially remained in the pellet. In order to evaluate NS2 functionality in our expression system, we monitored its cysteine-protease activity. To this end we expressed a JFH-1 NS2-NS3pro *Strep*-tagged precursor construct (denoted NS2-NS3pro-ST) that include the 213 N-terminal aa of NS3, and detected the cleavage products using antibodies directed either against the NS3pro C-terminal *Strep*-tag II (denoted NS3pro-ST), or a conformational NS2 antibody. Note that NS3pro-ST was inactivated (S139A mutation) to avoid unspecific cleavages by this protease. The immunoblot displayed in Figure 1B shows that both cleavage products, being NS3pro-ST and NS2 could be detected in the reaction mixture. Cleavage specificity was confirmed by using the inactive NS2 mutant bearing the C184A mutation in its catalytic site, for which accumulation of uncleaved NS2-NS3pro-ST precursor was observed.

Solubilized NS2 is obtained in a pure and homogenous form

A major challenge in the expression of membrane proteins is to isolate the protein in a pure form, and to avoid formation of aggregates. The N-terminal *Strep*-tag II allowed us to obtain pure NS2 protein using a single affinity-purification step starting from the supernatant fraction containing the solubilized protein (Figure 2A). It is worth mentioning that the addition of 0.25% DDM before the affinity-purification step not only allowed buffer exchange from MNG-3 to DDM, but resulted in improved NS2 purity and yield. Indeed, the purity of NS2 obtained in the presence of MNG-3 during this step was not satisfactory [16], as also observed in the presence of lower amounts of DDM (data not shown). With respect to yield, while purity was good in the presence of even higher amounts of DDM (0.5 - 1%), the protein yield was substantially reduced under this condition since the high detergent concentration decreases NS2-*strep*-tag II binding to the *strep*-Tactin column. The yield of purified NS2 was about 1 mg per mL of WGE. To further characterize purified NS2 protein, we performed size exclusion chromatography. A major peak containing NS2 was observed (Fig. 2B), while the following minor peak is due to reagents present in the *Strep*-Tactin elution buffer. NS2 protein-detergent complexes eluted with an apparent molecular mass of about 150 kDa (Fig. 2C), indicating the oligomeric nature of recombinant full-length NS2 in detergent micelles.

NS2 is well folded

The far UV circular dichroism (CD) spectrum of full-length NS2 eluted from size exclusion chromatography in 0.1% n-dodecyl β-D-maltoside (DDM) is typical of a well-folded protein (Figure 3). Additionally, the mean molar ellipticity per residue is in the range that can be expected for a protein (few 10,000's). The two minima at 208 and 222 nm, together with a maximum at 192 nm, are typical of α-helical folding. Spectral deconvolution indicated an α-helix content of 52-58 %, while turns represent 12-17%, and β-sheet content is limited to 8-9

%. The latter value is consistent with the content in β-sheet deduced from the crystal structure of the cytosolic domain of NS2 [4]. The limited number of residues folded into α-helices in this crystal structure indicates that the main contribution to helix content in full-length NS2 is due to its membrane domain. This is in agreement with our previous structural analyses of this membrane domain using synthetic peptides, which mostly revealed the presence of α-helical segments [8, 13].

NS2 can be reconstituted into model membranes

While expression in cell-free systems in the presence of liposomes might allow for direct membrane insertion of the produced proteins, detergent solubilization and subsequent reinsertion into membranes allows for higher protein yields and purer preparations. Successful insertion into membranes is also a good indicator for the structural integrity of the membrane protein. Moreover, homogenous membrane insertion is a prerequisite for further structural studies of the protein by electron microscopy or solid-state NMR. To this end, we analyzed our protein preparations and the possibility to reconstitute NS2 proteoliposomes by electron microscopy. As already indicated by size exclusion chromatography, NS2 solubilized in detergent forms multimers, which can be visualized by electron microscopy using immunolabeling of the N-terminal *Strep*-tag II fusion protein produced in the presence of detergent (Figure 4A). Absence of gold particles on negative control grids lacking NS2 confirmed labeling specificity (data not shown). When detergent-solubilized lipids are added to detergent-solubilized NS2, the protein binds to or integrates into the lipid bilayer upon detergent removal, as shown in Figures 4B and 4C for L-α-phosphatidylcholine (PC) and a PC/cholesterol mixture (70/30 w/w) respectively. No dispersed labeling of NS2 protein is found, and gold-decorated liposomes are observed, suggesting the tight association/insertion of NS2 into model membranes.

Discussion

NS2 is an important protein in the life cycle of the virus, and is believed to play manifold roles in the progress of virus assembly. Parts of its structure could be studied using different biophysical approaches as x-ray crystallography and solution-state NMR [4, 8, 13]. However, more detailed studies of the protein are missing, as active full-length NS2 has not been available due to the difficulties encountered for overexpression. So far, N-terminally truncated protein has been used to functionally characterize the protease activity, and refolding from inclusion bodies was mandatory [15]. The direct expression in soluble form as presented here circumvents many difficulties linked to purification from inclusion bodies, notably the presence of non-native forms on refolding and protein loss. While many studies have addressed interactions between NS2 with several other viral proteins *in cellulo*, no *in-vitro* assays on these interactions could be carried out in order to gain knowledge on the molecular mechanisms. Indeed, these interactions presumably take place *via* the hydrophobic membrane-inserted N-terminal region [9, 13], and their occurrence might as well rely on the details of the interaction of the protein complex with the membrane. An experimental set-up targeting the identification of these interactions at a molecular level is thus difficult with truncated forms of the protein, and as well might be unsuccessful in detergent-solubilized states. The availability of full-length NS2 in a purified, detergent-solubilized or membrane-bound forms, in sufficient quantities, opens the perspective for biochemical *in vitro* structural and functional studies of this enigmatic protein, including structural studies which aim at analyzing the protein in its membrane environment by means of electron microscopy and solid-state NMR. Furthermore, it will allow the probing of interactions with other viral proteins which can also be successfully produced using WGE cell-free expression system [16].

Material and Methods

Plasmids

cDNA of full-length NS2 from four different HCV strains was PCR amplified and cloned into the pEU-E01-MCS vector (CellFree Sciences, Japan). HCV strains, genotypes (in parenthesis), and their respective GenBank accession numbers are as follows: JFH-1 (2a), AB047639; Con1 (1b), AJ238799; H77 (1a), AF009606; strain 452 (3a), DQ437509. The aa sequence Met-Ala-Ser precedes a *Strep*-tag II followed by a thrombin cleavage site downstream of a Ser-Gly linker, yielding to the final aa sequence MASWSHPQFEKTGLVPRGSG fused at the N-terminus of the various full-length NS2 proteins [16]. In addition, cDNA of the two following NS2-NS3pro-ST precursors (strain JFH-1) was PCR amplified and cloned into the pEU-E01-MCS vector too. The first construct comprises wild type full-length NS2, and the second the full-length NS2 with a mutation within its catalytic site (C184A mutation, C997A accordingly to HCV polyprotein numbering). Note that, in both constructs, NS3pro serine protease activity was inactived by the S139A mutation (S1169A according to HCV polyprotein numbering) to avoid unspecific cleavages by this protease. The amino acid linker sequence Ser-Gly-Gly precedes a Twin *Strep*-tag II fused at the C-terminus of NS3pro, yielding to the final aa sequence SGGWSHPQFEK$(GGGS)_2$GGSAWSHPQFEK. As detailled in [16], the resulting plasmids were transformed into *Escherichia coli* TOP10 chemically competent cells (Life Technologies) and DNA was prepared using a NucleoBond Xtra Maxi kit (Macherey-Nagel, France). Plasmids were further purified by a phenol/chloroform extraction according to CellFree Sciences (Yokohama, Japan) recommendations.

Wheat germ cell-free protein expression and sample preparation

Wheat germ cell-free protein expression and sample preparation were performed as described in [16]. In short, home-made wheat germ extracts [22] were prepared according to Takai et al.

[39]. Transcription and translation were carried out separately, as described in [25–27, 39] and in Noirot et al. [40], and translation was performed using the so-called bilayer method. When present, MNG-3 detergent was added in both the reaction mix and the feeding buffer. The reaction mixture was incubated at 22 °C for 16 h. Translation was performed either in 96-well plates for small-scale expression test or 6-well plates for larger production.

SDS-PAGE and Western blotting analysis

SDS-PAGE and Western blotting analysis was performed as described in [16]. For activity test using the NS2-NS3pro-ST precursor, the CFS was analyzed by western blotting using antibodies either against the *Strep*-tag II fused at the C-terminus of NS3pro (StrepMAB-Classic, IBA Lifesciences, Germany) or against the C-terminal domain of NS2.

Purification of full-length NS2 by affinity chromatography

NS2 was expressed in presence of 0.1% MNG-3 using the bilayer method as described previously [16] adapted to wells of 6-well plates (i.e., 0.5 mL of reaction mix containing 0.25 mL mRNA and 40 μg/ml creatine kinase and 0.25 mL of wheat germ extract for the bottom layer, and 5.5 mL of feeding buffer for the upper layer in each well). After protein synthesis, the CFS (total cell-free sample) of 18 wells was pooled (108 mL) and incubated with benzonase at RT on a rolling wheel for 30 min. In addition, DDM was added to CFS at a final concentration of 0.25%, leading to a higher purity level than MNG-3. CFS was then centrifuged at 20.000 *g*, 4°C for 30 min and the supernatant was loaded on a 5 mL *Strep*-Tactin high capacity gravity column (IBA Lifesciences, Germany). Purification was performed as specified by the manufacturer, all buffers containing 0.1% DDM. Full-length NS2 was eluted in 100 mM Tris-HCl pH8, 150 mM NaCl, 1 mM EDTA, 0.1% DDM and 2.5 mM desthiobiotin.

Size exclusion chromatography

Size exclusion chromatography of full-length NS2 previously purified by affinity chromatography was performed using a Superdex^{TM} 200 Increase 3.2/100 column (GE Lifesciences). A 50 mM phosphate buffer pH 7.4 containing 0.1% DDM was used as eluent and 150 μg protein were loaded on the column. Calibration curve of the size exclusion chromatography column was established according to le Maire et al. [23] using the following protein standards: cytochrome C, 12.3 kDa ; myoglobin, 17.8 kDa, trypsin inhibitor, 22.1 kDa ; β-lactoglobulin, 35 kDa ; transferrin, 81 kDa ; aldolase, 158 kDa ; ferritin, 440 kDa ; thyroglobulin, 669 kDa. A 50 mM phosphate buffer pH 7.4 containing 0.1% DDM was used as eluent and 100 μg of each protein standard were run separately on the column.

Circular dichroism

Spectra were recorded on a Chirascan spectrometer (Apllied Photophysics) calibrated with 1S-(+)-10-camphorsulfonic acid. Measurements were carried out at 298 K in a 0.1-cm path length quartz cuvette. Spectra were measured in a 180–260 nm wavelength range with an increment of 0.2 nm, bandpass of 0.5 nm and integration time of 1 s. Spectra were processed, baseline corrected, smoothed and converted with the Chirascan software. Spectral units were expressed as the mean molar ellipticity per residue using the protein concentration determined at 280 nm using a NanoDrop spectrometer. Estimation of the secondary structure content was carried out using the CDSSTR, CONTIN and SELCON3 approaches available on the DICHROWEB server (dichroweb.cryst.bbk.ac.uk/).

Lipid reconstitution

Either egg PC (L-α-phosphatidylcholine, 99% pure) or a mixture of egg PC and cholesterol (99% pure, 70/30, w/w) used for lipid reconstitution were purchased from Sigma. The lipids were first solubilized in DDM with a detergent-to-lipid ratio of 10 (mol/mol) [24]. Purified full-length NS2, solubilized in 0.1% DDM, was then mixed with the lipids at a lipid-to-

protein ratio of 20 (w/w). Protein-lipid-detergent samples were dialyzed at room temperature (RT) for 6 days against a buffer containing 100 mM Tris-HCl pH8, 150 mM NaCl, 1 mM EDTA, 1 mM DTT and hydrophobic polystyrene beads (Bio-Beads SM-2 adsorbents, Bio-Rad). A dialysis membrane with a molecular weight cut-off of 6-8000 Da (Spectra/Por) was used. A ratio Bio-Beads-to-detergent of 100 (w/w) was applied. Detergent removal by the Bio-Beads leads to the formation of proteoliposomes [25].

Electron microscopy

The material was adsorbed to carbon- and pioloform-coated 300 mesh copper grids (Science Services GmbH) for 10 min at RT, and remaining liquid was drained using a whatman paper. For immunolabeling, all incubation steps were performed by floating the grids on top of drops. Prior to immunogold labeling, grids were incubated with blocking solution (0.8% bovine serum albumin, 0.1% fish skin gelatin, 50 mM glycine in PBS). Primary mouse anti-*Strep*-tag II antibody was diluted 1:100 and incubated for 20 min in blocking solution. After five 2-min washes with PBS, grids were incubated with rabbit anti-mouse bridging antibody (Dako Cytomation) diluted 1:150 in blocking solution for 30 min. After five additional wash steps in PBS, bound antibodies were labeled with protein A coupled to 10 nm colloidal gold suspension (Cell Microscopy Center) diluted 1:60 in blocking solution for 30 min. Grids were washed five times with PBS and fixed with 1% glutaraldehyde (Science Services GmbH; Munich, Germany) in PBS for 5 min. Subsequently, grids were washed five times with PBS followed by 5 wash steps in water and negatively stained as follows: grids were washed 3 times for 2 min with water on ice and subsequently, structures were negatively stained using a mixture of 3% aqueous uranylacetate and 2% methyl-cellulose (40:60, v/v) for 8 min on ice. Grids were looped out and dried for 5 min at RT. Samples were examined by using a Zeiss EM-10 transmission electron microscope (Zeiss, Goettingen, Germany) with a built in Mega View camera (Olympus; Tokyo, Japan).

Figures

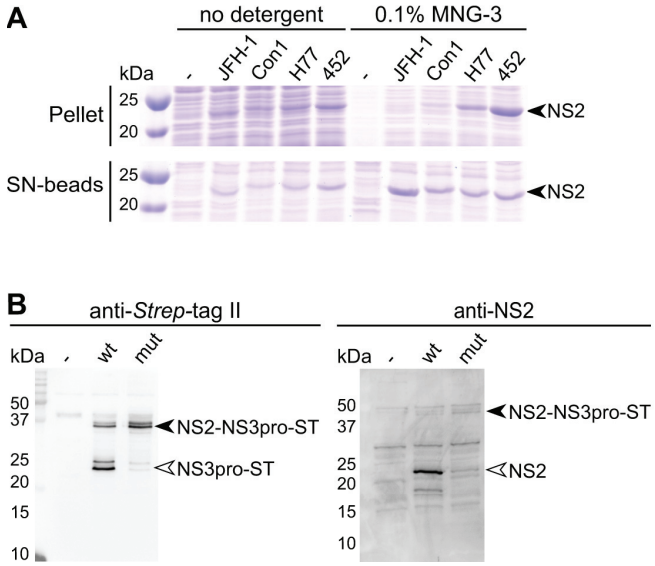


Figure 1: Expression and protease activity of full-length NS2 in the wheat germ extract cell-free system. (A) Full-length NS2 from different HCV strains were expressed in the absence of detergent, (left) or in the presence of 0.1% MNG-3 (right) as described previously [16]. Protein samples were analyzed by SDS-PAGE followed by Coomassie blue staining. -, negative control (no NS2 expression); Pellet, pellet obtained after centrifugation of total cell-free sample (CFS); SN-beads, supernatant obtained after centrifugation of CFS and incubated with *Strep*-Tactin magnetic beads to capture tagged NS2. Comparable amounts were loaded on the gel for Pellet and SN-beads. Bands corresponding to full-length NS2 are indicated by black arrowheads. (B) Wild type and mutant NS2-NS3pro-ST precursors (strain JFH-1) were expressed in the presence of 0.1% MNG-3. Total CFS were analyzed by immunoblotting using antibodies either against the *Strep*-tag II fused at the C-terminus of NS3pro-ST (left panel) or against the C-terminal domain of NS2 [8] (right panel). -, negative control (no NS2-NS3pro-ST expression); wt, wild type NS2; mut, NS2 with C184A mutation. Bands corresponding to uncleaved NS2-NS3pro-ST protein precursor are indicated by black arrowheads while cleaved NS2 and NS3pro-ST are indicated by empty arrowheads.

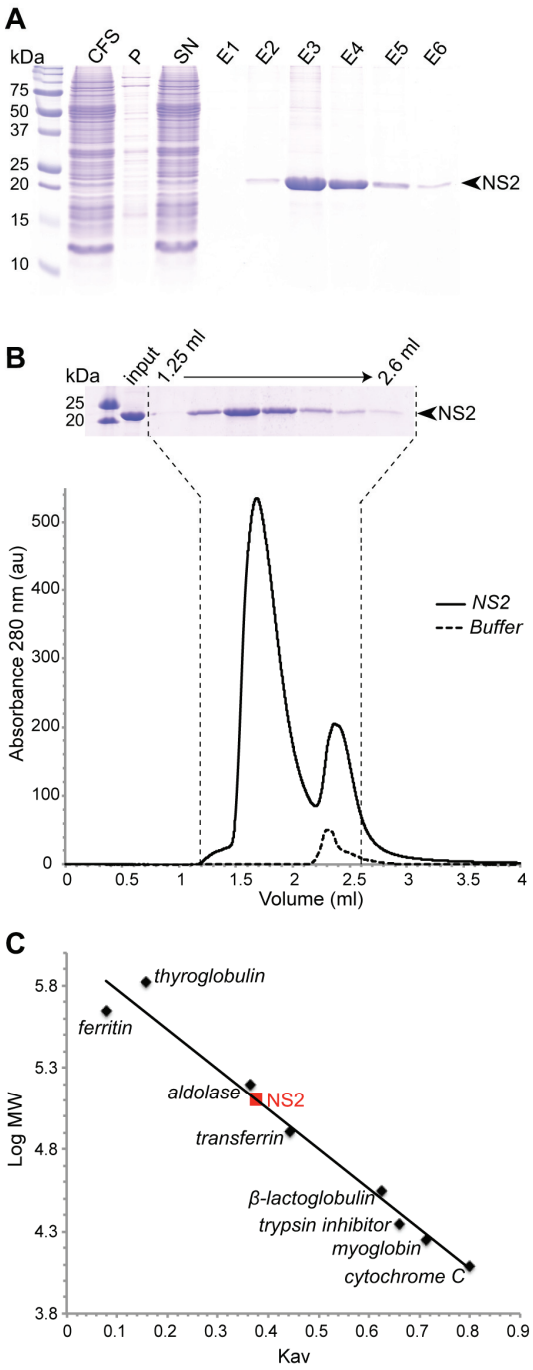


Figure 2: Purification of full-length NS2 (strain JFH-1 of HCV genotype 2a) by affinity chromatography and analysis by size exclusion chromatography. (A) Full- length N-terminally *Strep*-tag II tagged NS2 was expressed in presence of 0.1% MNG-3, complemented with 0.25% DDM before purification on a *Strep*-Tactin column in presence of 0.1% DDM. Protein samples were analyzed by SDS-PAGE followed by Coomassie blue staining. CFS, total cell-free sample (8 μL from 108 mL loaded on the gel); P, pellet obtained after centrifugation of CFS (equivalent to 80 μL from 108 mL loaded on the gel); SN, supernatant obtained after centrifugation of CFS and loaded on the affinity column (8 μL from 108 mL loaded on the gel); E1 to E6, affinity elution fractions (8 μL from 2.5 mL loaded on the gel). The black arrowhead indicates the band corresponding to NS2. (B) Size exclusion chromatography of E3 elution fraction using Superdex^{TM} 200 Increase 3.2/100 column. The elution profile of NS2 is drawn in a solid line, and that of the affinity elution buffer in a dotted line. SDS-PAGE and Comassie staining analysis of the collected fractions is shown above the chromatographic profile. Elution volume of NS2 is 1.65 mL. (C) Molecular weight calibration curve of the size exclusion chromatography column established using protein standards (V_0 is 1.25 mL). NS2 Kav is indicated in red on the standard curve.

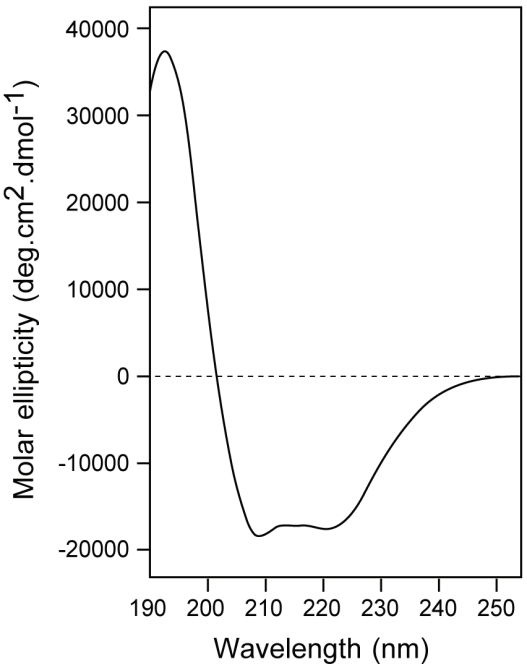


Figure 3: Far UV circular dichroism (CD) spectrum of NS2 after purification by affinity chromatography and size exclusion chromatography. Estimation of the secondary-structure content resulted in an α-helix content of 52-58 %, in 12-17% turns, and in 8-9 % β-sheet.

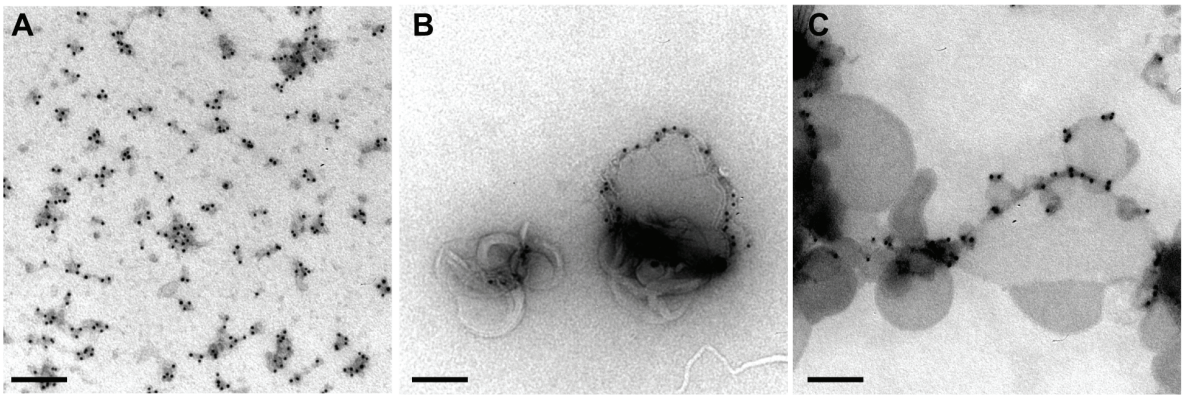


Figure 4: Transmission electron microscopy analysis of full-length NS2 using negative staining and immunogold labeling. (A) NS2 in 0.1% DDM after purification by affinity chromatography and size exclusion chromatography. (B) NS2 reconstituted in egg PC (L-α-phosphatidylcholine). (C) NS2 reconstituted in a mixture of egg PC and cholesterol (70/30, w/w). The scale bar corresponds to 200 nm in all panels.

Acknowledgements

This work was supported by a grant from the French ANRS (France Recherche, Nord & Sud, Sida-HIV et Hépatites), an autonomous agency at INSERM, France, and by the ANR (ANR-14-CE09-0024B and ANR-11-BINF-003 Mapping project). R.B. is supported by the Deutsche Forschungsgemeinschaft (TRR 83, TP13). We would like to thank Yaeta Endo for the interesting discussions on protein cell-free expression.

References

[1] R. Bartenschlager, V. Lohmann, F. Penin, The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection, Nat Rev Micro. 11 (2013) 482–496. doi:10.1038/nrmicro3046.

[2] B.D. Lindenbach, C.M. Rice, The ins and outs of hepatitis C virus entry and assembly, Nat Rev Micro. 11 (2013) 688–700. doi:10.1038/nrmicro3098.

[3] D. Moradpour, F. Penin, Hepatitis C Virus Proteins: From Structure to Function, in: Current Topics in Microbiology and Immunology, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013: pp. 113–142. doi:10.1007/978-3-642-27340-7_5.

[4] I.C. Lorenz, J. Marcotrigiano, T.G. Dentzer, C.M. Rice, Structure of the catalytic domain of the hepatitis C virus NS2-3 protease, Nature. 442 (2006) 831–835. doi:10.1038/nature04975.

[5] S. Welbourn, R. Green, I. Gamache, S. Dandache, V. Lohmann, R. Bartenschlager, et al., Hepatitis C Virus NS2/3 Processing Is Required for NS3 Stability and Viral RNA Replication, J. Biol. Chem. 280 (2005) 29604–29611. doi:10.1074/jbc.M505019200.

[6] T. Phan, R.K.F. Beran, C. Peters, I.C. Lorenz, B.D. Lindenbach, Hepatitis C virus NS2 protein contributes to virus particle assembly via opposing epistatic interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes, J. Virol. 83 (2009)

8379–8395. doi:10.1128/JVI.00891-09.

[7] D. Paul, V. Madan, R. Bartenschlager, Hepatitis C Virus RNA Replication and Assembly: Living on the Fat of the Land, Cell Host Microbe. 16 (2014) 569–579. doi:10.1016/j.chom.2014.10.008.

[8] V. Jirasko, R. Montserret, J.Y. Lee, J. Gouttenoire, D. Moradpour, F. Penin, et al., Structural and functional studies of nonstructural protein 2 of the hepatitis C virus reveal its key role as organizer of virion assembly, PLoS Pathog. 6 (2010) e1001233. doi:10.1371/journal.ppat.1001233.

[9] C.-I. Popescu, N. Callens, D. Trinel, P. Roingeard, D. Moradpour, V. Descamps, et al., NS2 protein of hepatitis C virus interacts with structural and non-structural proteins towards virus assembly, PLoS Pathog. 7 (2011) e1001278. doi:10.1371/journal.ppat.1001278.

[10] B. Boson, O. Granio, R. Bartenschlager, F.-L. Cosset, A concerted action of hepatitis C virus p7 and nonstructural protein 2 regulates core localization at the endoplasmic reticulum and virus assembly, PLoS Pathog. 7 (2011) e1002144. doi:10.1371/journal.ppat.1002144.

[11] Y. Ma, M. Anantpadma, J.M. Timpe, S. Shanmugam, S.M. Singh, S.M. Lemon, et al., Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins, J. Virol. 85 (2011) 86–97. doi:10.1128/JVI.01070-10.

[12] K.A. Stapleford, B.D. Lindenbach, Hepatitis C virus NS2 coordinates virus particle assembly through physical interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes, J. Virol. 85 (2011) 1706–1717. doi:10.1128/JVI.02268-10.

[13] V. Jirasko, R. Montserret, N. Appel, A. Janvier, L. Eustachi, C. Brohm, et al., Structural and functional characterization of nonstructural protein 2 for its role in

hepatitis C virus assembly, J. Biol. Chem. 283 (2008) 28546–28562. doi:10.1074/jbc.M803981200.

[14] C.M. Lange, P. Bellecave, V.L. Dao Thi, H.T.L. Tran, F. Penin, D. Moradpour, et al., Determinants for Membrane Association of the Hepatitis C Virus NS2 Protease Domain, J. Virol. 88 (2014) 6519–6523. doi:10.1128/JVI.00224-14.

[15] D. Thibeault, R. Maurice, L. Pilote, D. Lamarre, A. Pause, In Vitro Characterization of a Purified NS2/3 Protease Variant of Hepatitis C Virus, J. Biol. Chem. 276 (2001) 46678–46684. doi:10.1074/jbc.M108266200.

[16] M.-L. Fogeron, A. Badillo, V. Jirasko, J. Gouttenoire, D. Paul, L. Lancien, et al., Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins, Protein Expression and Purification. 105 (2015) 39–46. doi:10.1016/j.pep.2014.10.003.

[17] L.A. Isaksson, J. Enberg, R. Neutze, B.G. Karlsson, A. Pedersen, Expression screening of membrane proteins with cell-free protein synthesis, Protein Expression and Purification. 82 (2012) 218–225. doi:10.1016/j.pep.2012.01.003.

[18] D. Schwarz, F. Junge, F. Durst, N. Frölich, B. Schneider, S. Reckel, et al., Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems, Nat Protoc. 2 (2007) 2945–2957. doi:10.1038/nprot.2007.426.

[19] C. Berrier, K.-H. Park, S. Abes, A. Bibonne, J.-M. Betton, A. Ghazi, Cell-free synthesis of a functional ion channel in the absence of a membrane and in the presence of detergent, Biochemistry. 43 (2004) 12585–12591. doi:10.1021/bi049049y.

[20] C. Klammt, F. Löhr, B. Schafer, W. Haase, V. Dötsch, H. Ruterjans, et al., High level cell-free expression and specific labeling of integral membrane proteins, Eur J

Biochem. 271 (2004) 568–580.

[21] K. Takai, T. Sawasaki, Y. Endo, Practical cell-free protein synthesis system using purified wheat embryos, Nat Protoc. 5 (2010) 227–238. doi:10.1038/nprot.2009.207.

[22] C. Noirot, B. Habenstein, L. Bousset, R. Melki, B.H. Meier, Y. Endo, et al., Wheat-germ cell-free production of prion proteins for solid-state NMR structural studies, N Biotechnol. 28 (2011) 232–238. doi:10.1016/j.nbt.2010.06.016.

[23] M. LEMAIRE, L.P. Aggerbeck, C. Monteilhet, J.P. Andersen, J.V. Moller, The Use of High-Performance Liquid-Chromatography for the Determination of Size and Molecular-Weight of Proteins - a Caution and a List of Membrane-Proteins Suitable as Standards, Analytical Biochemistry. 154 (1986) 525–535. doi:10.1016/0003-2697(86)90025-4.

[24] B. Kunert, C. Gardiennet, D. Lacabanne, D. Calles-Garcia, P. Falson, J.-M. Jault, et al., Efficient and stable reconstitution of the ABC transporter BmrA for solid-state NMR studies, Front. Mol. Biosci. 1 (2014) 1–11. doi:10.3389/fmolb.2014.00005.

[25] J.-L. Rigaud, D. Lévy, Reconstitution of membrane proteins into liposomes, Methods in Enzymology. 372 (2003) 65–86. doi:10.1016/S0076-6879(03)72004-7.

APPENDIX 3 - CV

Marie-Laure Fogeron

Personal data

Name	Marie-Laure Fogeron
Birth	August 3rd 1976, Annonay, France
Nationality	French
Work address	7, passage du Vercors F-69367 Lyon Cedex 7
Telephone	04 37 65 29 33
E-Mail	marie-laure.fogeron@ibcp.fr

Education

1998-1999	**Institut National Polytechnique de Lorraine**, Nancy, France Diplôme d'Etudes Supérieures Spécialisées (DESS) Génie protéique
1994-1998	**Joseph Fourier University**, Grenoble, France Maîtrise de Biochimie

Professional experience

Since 2010/12	Ingénieur d'Etudes, **Institute of Biology and Chemistry of Proteins**, Lyon, France *Groups leaders*: Dr. Anja Böckmann & Dr. François Penin *Project*: Development of a cell-free system for the expression, purification and characterization of eukaryotic and viral membrane proteins.
2009/10- 2010/11	Ingénieur, **INSERM U851**, Lyon, France *Group leader*: Dr. Nathalie Bonnefoy-Bérard *Project*: Characterization of small molecules inhibiting the anti-apoptotic protein Bfl-1.
2005/04- 009/09	Scientist, **Max Planck Institute for Molecular Genetic**, Berlin, Germany *Group leader*: Dr. Bodo Lange *Project*: Isolation and characterization of protein complexes by tandem affinity purification (TAP).
2003/10- 2005/03	Scientist, **Proteome Factory AG**, Berlin, Germany *Group leader*: Dr. Christian Scheler *Projects*: Planning, coordination and realisation of contract research projects in the field of the differential proteome analysis. Development of new procedures for proteome analysis.
2000/09- 2003/09	Scientist, **Europroteome AG**, Berlin, Germany *Groups leaders*: Prof. Marc Reymond & Dr. Dirk Sawitzky *Projects*: Establishment and direction of the two-dimensional electrophoresis lab. Realisation of different screening projects on human samples with the aim to discover new cancer markers for diagnosis and therapy. Development and optimization of a fractionation protocol for the purification of membrane proteins.

Languages

French	mother tongue
English	fluent
German	fluent (Kleines Deutsches Sprachdiplom, Goethe Institut)

Publications

Badillo A, Receveur-Bréchot V, Hanoulle X, **Fogeron ML**, Miron S, Molle J, Montserret R, Böckmann A, Tellinghuisen TL, Bartenschlager R, Lippens G, Ricard-Blum S, Penin F (*in preparation*) Overall structural model of NS5A protein from hepatitis C virus and modulation by mutations confering resistance of HCV replication to cyclosporin A.

Fogeron ML, Paul D, Jirasko V, Montserret R, Lacabanne D, Molle J, Badillo A, Boukadida C, Georgeault S, Roingeard P, Martin A, Bartenschlager R, Penin F, Böckmann A. Functional expression, purification, characterization, and membrane reconstitution of nonstructural protein 2 from hepatitis C virus. *Submitted to Protein Expr Purif. (currently under review).*

Fogeron ML, Badillo A, Jirasko V, Gouttenoire J, Paul D, Lancien L, Moradpour D, Bartenschlager R, Meier BH, Penin F, Böckmann A (2015) Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins. *Protein Expr Purif.* **105:** 39-46.

Mathieu AL, Sperandio O, Pottiez V, Balzarin S, Herlédan A, Elkaïm JO, **Fogeron ML**, Piveteau C, Dassonneville S, Deprez B, Villoutreix BO, Bonnefoy N, Leroux F (2014) Identification of Small Inhibitory Molecules Targeting the Bfl-1 Anti-Apoptotic Protein That Alleviates Resistance to ABT-737. *J Biomol Screen* **19:** 1035-1046.

Fogeron ML, Müller H, Schade S, Dreher F, Lehmann V, Kühnel A, Scholz AK, Kashofer K, Zerck A, Fauler B, Lurz R, Herwig R, Zatloukal K, Lehrach H, Gobom J, Nordhoff E, Lange BMH (2013) LGALS3BP regulates centriole biogenesis and centrosome hypertrophy in cancer cells. *Nat. Comm.* **4:** 1531.

Mercier BC, Ventre E, **Fogeron ML**, Debaud AL, Tomkowiak M, Marvel J, Bonnefoy N (2012) NOD1 cooperates with TLR2 to enhance T cell receptor-mediated activation in CD8 T cells. *PloS one* **7:** e42170.

Broch S, Henon H, Debaud AL, **Fogeron ML**, Bonnefoy-Berard N, Anizon F, Moreau P (2010) Synthesis and biological activities of new di- and trimeric quinoline derivatives. *Bioorganic & medicinal chemistry* **18:** 7132-7143.

Fogeron ML, Müller H, Lehmann V, Habermann K, Kurtz T, Scholz AK, Siebert S, Rak KH, Gobom J, Lange BMH (2007) Protein Complex Composition and Function in Health and Disease. *Science Inside* **2**: 66-67.

Müller H, **Fogeron ML**, Lehmann V, Lange BMH (2007) Das andere Ende des Spindelcheckpoints. *Biospektrum* **13**: 32-34.

Müller H, **Fogeron ML**, Lehmann V, Lehrach H, Lange BM. (2006) A centrosome-independent role for gamma-TuRC proteins in the spindle assembly checkpoint. *Science* **314** : 654-657.

Ebert MP, Kruger S, **Fogeron ML**, Lamer S, Chen J, Pross M, Schulz HU, Lage H, Heim S, Roessner A, Malfertheiner P, Rocken C (2005) Overexpression of cathepsin B in gastric cancer identified by proteome analysis. *Proteomics* **5:** 1693-1704.

Patents

Seibert V, Lamer S, Rothmann-Cosic K, Tortola Perez S, Ilyina T, Heim S, Meuer J, Bushmann T, **Fogeron ML** and Van der Linden M. Tumor marker proteins for diagnosis and therapy of cancer and cancer risk assessment. WO2004040298, published May 13th 2004.

Seibert V, Lamer S, **Fogeron ML**, Tortola Perez S, Rothmann-Cosic K, Bushmann T, Meuer J, Ilyina T, Van der Linden M and Heim S. Agents and methods for diagnosis and therapy of cancer and cancer risk assessment. WO2004005928, published January 15th 2004.

Lamer S, **Fogeron ML**, Lage H and Kellner U. Tumor markers and the use thereof for the diagnosis and treatment of tumor diseases. WO2004003564, published January 8th 2004.

Oral presentations

« Wheat germ cell-free expression, purification and characterization of NS2, NS4B and NS5A proteins of Hepatitis C virus», 1ère Réunion Plénière du GDR « Protéines Membranaires : Aspects Moléculaires et Cellulaires, Praz-sur-Arly, France, March 19th 2015

« Wheat germ cell-free expression, purification and characterization of NS2, NS4B and NS5A proteins of Hepatitis C virus», 15ème Réunion du Réseau National Hépatites (ANRS), Paris, France, January 30th 2015

« Expression, purification and characterisation of the non structural protein 2 (NS2) of Hepatitis C virus», 18ème Journée Scientifique de l'EDISS, Lyon, France, October 3rd 2013

« Cell-free expression of HCV proteins: the characterisation of NS2 as an example», séminaire d'unité BMSSI, Lyon, France, September 30th 2013

« Expression, purification and characterisation of HCV NS2 full length protein and its ectodomain NS2^{pro}», 12th Schauinsland Retreat, Freiburg, Germany, June 7th 2013

« Cell-free expression of membrane proteins using wheat germ extract», Ecole Thématique CNRS « Expression, purification and stabilization of membrane proteins for structural analysis », Grasse, France, October 2nd 2012

« Cell-free protein expression using wheat germ extract », 11th Schauinsland Retreat, Freiburg, Germany, June1st 2012

« Cell-free protein expression using wheat germ extract », in parallel to the Bio-NMR Symposium « Towards structural studies of membrane proteins », Zurich, Switzerland, January 19th 2012

« Cell-free protein expression using wheat germ extract », séminaire d'unité BMSSI, Aussois, France, May 25th 2011

Recent posters

Fogeron ML, Gardiennet C, Sarrazin S, Badillo A, Schütz A, Endo Y, Meier BH, Penin F and Böckmann A *« Cell-free expression of viral membrane proteins for their structural analysis by solid-state NMR»*, 17ème Journée Scientifique de l'EDISS, Lyon, France, October 18th 2012

Fogeron ML, Gardiennet C, Sarrazin S, Schütz A, Badillo A, Endo Y, Meier BH, Penin F and Böckmann A *« Cell-free expression of viral membrane proteins for their structural analysis by solid-state NMR»*, XXVth ICMRBS, Lyon, France, August 19th-24th 2012

Fogeron ML, Sarrazin S, Schütz A, Endo Y, Meier BH, Penin F and Böckmann A *« Cell-free expression of viral membrane proteins for their structural analysis by solid-state NMR»*, XIIIème Journée Rhône-Alpes de RMN, Grenoble, France, June 22nd 2012

Other communications

Jirasko V, **Fogeron ML**, Penzel S, Paul D, Lends A, Penin F, Meier BH, Böckmann A and Bartenschalger R *« Preparation of Hepatitis C and Dengue virus membrane proteins in a form of proteoliposomes for solid-state NMR analysis»*, EBSA Solid-State NMR School, Munich, Germany, November 17th-22nd 2014 – POSTER

Böckmann A, **Fogeron ML** and Penin F *«Cell-free protein expression using wheat germ extracts»*, Advanced School on Biological Solid-State NMR, Brno, Czech Republic, October 7th-12th 2012 – ORAL PRESENTATION

Boukadida C, Marnata C, Cohen L, **Fogeron ML**, Gouttenoire J, Moradpour D, Penin F and Martin A *«GB virus B NS2 is an auto-protease with similar membrane topology as HCV NS2»*, 19th International Hepatitis C Virus & Related Diseases (Molecular Virology and Pathogenesis), Venice, Italy, October 5th-9th 2012 – POSTER

Böckmann A, Meier B, Melki R, Habenstein B, Luckgei N, Schütz A, Gath J, Bousset L, **Fogeron ML** and Penin F *«Studies of protein fibrils and membrane proteins by solid-state NMR»*, Protein Island Matsuyama International Symposium, Matsuyama, Japan, September 21st-23rd 2011 – ORAL PRESENTATION