

JEAN-BAPTISTE DUVIGNAUD

**ÉTUDE STRUCTURALE ET FONCTIONNELLE DE
LA PROTÉINE CORE DU VIRUS DE L'HÉPATITE C.**

Thèse présentée
à la Faculté des études supérieures et postdoctorales de l'Université Laval
dans le cadre du programme de doctorat en Biochimie
pour l'obtention du grade de philosophiae doctor (Ph.D)

DÉPARTEMENT DE BIOCHIMIE, DE MICROBIOLOGIE ET
DE BIO-INFORMATIQUE
FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

2012

Résumé

Le virus de l'hépatite C (VHC) représente un problème majeur de santé publique avec au dessus de 120 millions de personnes infectées et à ce jour aucun traitement capable de contrer de manière efficace ce virus. Le VHC a été découvert en 1989, classé dans la famille des *flaviviridae* et est le seul représentant du genre *hepacivirus*. Il s'agit d'un virus enveloppé, formé donc d'une bicouche lipidique, d'une capsid, et d'un acide nucléique (ARN). La capsid virale est formée d'une seule et unique protéine appelée « Core ». Il s'agit d'une protéine de 177 acides aminés sous sa forme mature. Il a été démontré au sein du laboratoire du Professeur Denis Leclerc que la première moitié de la protéine (C82) était suffisante pour générer à la fois *in vivo* et *in vitro* la formation de la capsid virale. Ayant comme objectif de comprendre les mécanismes régissant l'assemblage viral, nous avons étudié cette protéine d'un point de vue structural.

Suite à la mise au point d'un protocole robuste d'expression, de marquage isotopique et de purification, nous avons étudié la Core C82 par dichroïsme circulaire et résonance magnétique nucléaire. Les données expérimentales obtenues suggèrent que la Core C82 est non structurée et flexible, confirmant ainsi l'analyse de la séquence primaire et les prédictions de structure secondaire de la Core. Nous avons ainsi mis en évidence que la moitié N-terminale de la Core du VHC appartenait à la famille des protéines intrinsèquement non structurées (IUP). Les IUPs sont des protéines peu ou très peu structurées et parfois capables de se structurer en liaison à un partenaire (protéine, acide nucléique, petite molécule). Nous avons donc, dans le but de structurer la Core C82, testé de multiples conditions de sel, détergent, agent lipomimétique et un partenaire protéique (protéine p53). Seul l'ajout d'un agent lipomimétique (2,2,2-trifluoroéthanol, TFE) à la Core C82 est venu modifier sa structure et sa dynamique. Nous avons ainsi pu démontrer que la Core C82 pouvait adopter une structure en hélice α . Nous avons, de plus, confirmé que ce repliement était également présent dans des versions tronquées plus longues de la Core (formes C124 et C170).

Enfin, nous avons, parallèlement à son étude structurale, mis au point un test *in vitro* d'inhibition de l'assemblage. Cet outil nous a permis de découvrir plusieurs peptides dérivés de la séquence protéique de la Core et de la protéine NS5A du VHC, ayant un effet inhibiteur sur l'assemblage viral *in vitro* de la Core mature (C170). Ces travaux innovateurs représentent une avenue encourageante vers la découverte d'un moyen efficace de soigner l'infection au VHC.

Au final, même si la structure 3D de la Core mature reste non déterminée, notre étude structurale de la Core C82 est la plus complète réalisée à ce jour à l'échelle atomique. Par ailleurs, nos travaux préliminaires sur la forme mature de la Core (C170) semblent être très prometteurs et permettraient de déterminer la structure 3D de la protéine Core mature.

Abstract

The hepatitis C virus (HCV) is a major public health problem with more than 120 million infected people, and to date no efficient treatment is available. HCV was discovered in 1989, classified in the *flaviviridae* family, and is the only member of the *hepacivirus* genus. It is an enveloped virus, consisting in a lipid bilayer, a capsid, and a ribonucleic acid (RNA). The viral capsid is composed by only one protein called “Core” protein. It is a 177 amino acid long protein in its mature form. It was demonstrated in our laboratory that the first half of this protein (C82) was sufficient to generate, both *in vivo* and *in vitro*, the assembly of the viral capsid. In order to understand the mechanism controlling the capsid assembly, we have studied the structural aspect of this protein.

After developing robust protocols of overexpression, isotope labelling and purification of the protein C82, we studied this truncated form using circular dichroism and nuclear magnetic resonance. The experimental data obtained suggest that Core C82 is an unstructured and very flexible protein, thus confirming primary sequence analysis and secondary structure predictions. We established that the N-terminal half of the Core protein is a member of the intrinsically unstructured protein family (IUP). IUPs are proteins which are totally or partially unstructured, but can sometimes undergo a structural change induced by the binding of a partner (protein, nucleic acid, small molecule). In order to induce a structured form of the C82 protein, we have tested a large range of conditions of salt, detergent, lipomimetic solvent, as well as interaction with a proteic partner (p53). Only the addition of a lipomimetic agent (2,2,2-trifluoroethanol, TFE) to the Core C82 protein resulted in a structural and a dynamical change. In this special condition, we were able to demonstrate that the Core C82 can adopt an α -helix conformation. We have, moreover, confirmed that this conformation was also present in longer truncated form of the Core protein (C124 and C170).

We also, along its structural analysis, developed an *in vitro* test to inhibit the assembly of the Core protein. This tool allowed us to discover several small peptides derived from the HCV Core and NS5A proteins amino acid sequences, with an inhibitory

effect on the viral assembly of the mature Core protein (C170). This innovative work is potentially an interesting step towards the development of an efficient treatment against HCV.

Ultimately, even if the 3D structure of the mature Core protein remains unsolved, our structural study of the C82 truncated form is the most comprehensive study to date at an atomic resolution. Moreover, our preliminary works on the mature form of the Core protein (C170) are very promising and should eventually lead to the three dimensional structure.

Remerciements

Je tiens dans un premier temps à remercier mon directeur de thèse Stéphane Gagné, qui a été d'une grande aide dans la réalisation de cette thèse. Il a su être inspirant et de bon conseil. Je tiens à remercier également, mon co-directeur de thèse, Denis Leclerc, sans qui cette thèse n'aurait pas été possible.

Il est bien évident que je veux ici remercier mes collègues et amis de paillasses du CHUL. Merci à Christian Savard (co-fondateur de la Mafia HCV), merci pour le temps, les niaiseries et autres moments partagés aux CHUL. Je remercie aussi Rémi Fromentin, qui par son bref passage au sein du laboratoire a su y mettre son œil critique et plein de bon sens. Merci à Nathalie Majeau (chef de projet HCV), Marilène Bolduc, Marie-Eve « Freedom Win » Laliberté-Gagné, et tous les autres membres du laboratoire Leclerc.

Un grand merci aux membres du groupe Gagné. Merci à Pierre-Yves Savard, Sébastien Morin, pour leur aide sur l'aimant RMN. Un spécial à Olivier Fisette, sans qui mon ordinateur au laboratoire n'aurait sûrement jamais pu fonctionner. Merci à Jérémie Leclerc, pour ses goûts musicaux et sa protéine aussi capricieuse que la mienne.

Je voudrais remercier aussi les membres du département de biochimie, microbiologie et bio-informatique, pour leur support dans cette fin de thèse.

Mon avant dernier message s'adresse à ma famille, mes parents qui ont su me soutenir et m'encourager à faire des études. Je suis certain qu'aujourd'hui, ils sont fiers de moi. Un merci tout spécial à ma sœur Élise, qui est loin de moi.

Et enfin, un énorme merci à Laurie, avec qui je partage ma vie depuis quelques années déjà, qui a été présente dans les meilleurs et moins bons moments de cette thèse.

Avant-Propos

La présente thèse est une thèse « mixte ». Ceci signifie qu'un certain nombre des chapitres présentés sont des articles publiés dans des revues spécialisées (voir les chapitres 3 et 4) ou en soumission (voir le chapitre 7) tandis que les autres chapitres feront état de l'avancement de divers travaux réalisés au sein des laboratoires des Drs. Stéphane M. Gagné et Denis Leclerc.

Contribution des Auteurs

Articles publiés :

Chapitre 3

Jean-Baptiste Duvignaud, Christian Savard, Rémi Fromentin, Nathalie Majeau, Denis Leclerc and Stéphane M. Gagné. Structure and dynamics of the N-terminal half of hepatitis C virus core protein : an intrinsically unstructured protein. *Biochem. Biophys. Res. Commun.* 2009 Jan 2 ; 378(1) :27-31.

Jean-Baptiste Duvignaud : J'ai réalisé toutes les expériences présentées dans cet article, de la production, purification, aux analyses bioinformatiques et biophysiques. J'ai également rédigé l'article sous sa forme publiée.

Christian Savard, Rémi Fromentin : Ils ont été d'une aide logistique précieuse lors des tests de production et purification de la protéine Core C82.

Nathalie Majeau : Le Dr. Majeau a cloné le fragment N-terminal de la protéine Core (fragment C82) dans le vecteur pET-3d et optimisé sa séquence pour une expression dans *Escherichia Coli*.

Denis Leclerc, Stéphane M. Gagné : Ils ont supervisé les travaux et ont fait une lecture critique du manuscrit.

Chapitre 4

Jean-Baptiste Duvignaud, Denis Leclerc and Stéphane M. Gagné. Structure and dynamics changes induced by 2,2,2-trifluoro-ethanol (TFE) on the N-terminal half of hepatitis C virus core protein. *Biochem. Cell Biol.* 2010 Apr; 88(2): 315-323.

Jean-Baptiste Duvignaud : L'ensemble des expériences et la rédaction de l'article ont été réalisés par moi-même, Jean-Baptiste Duvignaud.

Denis Leclerc, Stéphane M. Gagné : Ils ont supervisé les travaux et ont fait une lecture critique du manuscrit.

Article soumis :

Chapitre 7

Jean-Baptiste Duvignaud, Nathalie Majeau, Priscilla Delisle, Normand Voyer, Stéphane M. Gagné and Denis Leclerc. Interfering with HCV assembly *in vitro* using affinity peptides directed towards core protein. (Soumis dans *Canadian Journal of Microbiology*).

Jean-Baptiste Duvignaud : J'ai réalisé la production et la purification de la Core C170, et l'ensemble des tests d'assemblage *in vitro* et des gels de mobilité.

Nathalie Majeau : Le Dr. Majeau a fait une lecture critique du manuscrit.

Priscilla Delisle : Priscilla a été stagiaire au sein du laboratoire du Dr N. Voyer, où elle a synthétisé un grand nombre de peptides pour nos tests d'inhibition.

Normand Voyer, Stéphane M. Gagné et Denis Leclerc : Ils ont supervisé les travaux et ont fait une lecture critique du manuscrit.

Autres contributions :

J'ajouterai que durant cette thèse, j'ai tenu à être impliqué de proche ou de loin à d'autres projets de recherche, que ce soit par curiosité, ou pour offrir les compétences que j'ai acquises au cours de ce doctorat. Le résultat de ces diverses collaborations est la publication de trois articles et l'acceptation d'un manuscrit de Jessica Godin-Ethier (étudiante au doctorat au sein du laboratoire de Réjean Lapointe de l'université de Montréal), ainsi que l'écriture prochaine d'un second article relié aux travaux du Dr Réjean Lapointe.

Accepté dans Molecular Immunology (Août 2011):

Godin-Ethier J., Hanafi L.-A., Duvignaud J.-B., Leclerc D. and Lapointe R. IDO expression by human B lymphocytes in response to T lymphocyte stimuli and TLR engagement is biologically inactive.

Publiés:

Fromentin R., Majeau N., Laliberté Gagné M.-E., Boivin A., Duvignaud J.-B. and Leclerc D. A method for *in vitro* assembly of hepatitis C virus core protein and for screening of inhibitors. *Anal Biochem* 2007 Jul 1;366(1):37-45.

Majeau N., Bolduc M., Duvignaud J.-B., Fromentin R. and Leclerc D. Effect of cAMP-dependent protein kinase A (PKA) on HCV nucleocapsid assembly and degradation. *Biochem. Cell Biol.* 2007 Feb; 85(1): 78-87.

Tremblay M.-H., Majeau N., Gagné M.-E., Lecours K., Morin H., Duvignaud J.-B., Bolduc M., Chouinard N., Paré C., Gagné S.-M., and Leclerc D. Effect of mutations K97A and E128A on RNA binding and self assembly of papaya mosaic potexvirus coat protein. *FEBS J.* 2006 Jan; 273(1): 14-25.

*À ma famille qui a toujours été présente pour
m'encourager à poursuivre mes études, et à
Laurie qui partage ma vie.*

Table des matières

Résumé.....	i
Abstract.....	iii
Remerciements.....	v
Avant-Propos.....	vii
Table des matières.....	xiii
Liste des tableaux.....	xvii
Liste des figures.....	xix
Liste des abréviations.....	xxi
Chapitre 1.....	1
Introduction.....	1
1.1 Rappels historiques.....	2
1.2 Phylogénie du VHC.....	4
1.3 Le VHC, un virus génétiquement très diversifié.....	6
1.4 Épidémiologie et transmission du VHC.....	7
1.5 Physiopathologie et traitement du VHC.....	9
1.6 Organisation génétique.....	10
1.7 Le cycle réplicatif du VHC.....	11
1.8 La particule virale : Structure.....	14
1.9 Les protéines du VHC.....	16
1.9.1 Les protéines structurales.....	16
1.9.2 Les protéines non structurales.....	18
1.10 La protéine Core du VHC.....	26
1.10.1 Organisation.....	26
1.10.2 Maturation et localisation.....	27
1.10.3 Structure.....	29
1.10.4 Fonctions.....	32
1.10.5 Autres rôles.....	39
1.11 Objectif général.....	42
1.11.1 Objectifs spécifiques.....	42
Chapitre 2.....	45
Les méthodes biophysiques utilisées.....	45
2.1 Le dichroïsme circulaire.....	47
2.1.1 Principe.....	47
2.1.2 Les applications du dichroïsme circulaire de protéine.....	48
2.2 La résonance magnétique nucléaire.....	53
2.2.1 Principes de la RMN.....	53
2.2.2 L'analyse structurale de protéines.....	55
Chapitre 3.....	67
Structure and dynamics of the N-terminal half of hepatitis C virus Core protein: An intrinsically unstructured protein.....	67
3.1 Résumé.....	68
3.2 Abstract.....	68
3.3 Introduction.....	69
3.4 Materials and methods.....	70

3.4.1 Sequence analysis and secondary structure predictions	70
3.4.2 Expression and purification of C82	70
3.4.3 Trypsin Digestion	71
3.4.4 In vitro assembly reactions, density gradient centrifugation and electron microscopy.....	71
3.4.5 CD spectrum experiments.....	71
3.4.6 NMR samples	71
3.4.7 NMR data collection and data processing	72
3.4.8 Backbone assignments and relaxation analysis	72
3.5 Results and discussion	73
3.5.1 Secondary structure prediction and sequence analysis of C82	73
3.5.2 Expression, purification and structural characterization of C82 by CD and NMR spectroscopy.....	75
3.5.3 Backbone dynamics of C82 by NMR	79
3.5.4 HCV Core protein: an IUP.....	81
3.6 Acknowledgments	82
3.7 Appendix A. Supplementary data	82
3.8 References.....	83
Chapitre 4.....	87
Structure and dynamics changes induced by TFE on the N-terminal half of hepatitis C virus Core protein.	87
4.1 Résumé.....	88
4.2 Abstract.....	88
4.3 Introduction.....	89
4.4 Materials and methods	91
4.4.1 Purification of truncated hepatitis Core protein C82	91
4.4.2 Circular dichroism experiments.....	91
4.4.3 Nuclear magnetic resonance samples	92
4.4.4 NMR data collection and data processing	92
4.4.5 Backbone assignment and relaxation analysis.....	92
4.5 Results.....	93
4.5.1 Secondary structure of C82 under various conditions: a CD study.....	93
4.5.2 NMR characterization of C82 in 90 % TFE	96
4.5.3 NMR assignments and secondary structure analysis of C82 in 90 % TFE	98
4.5.4 Backbone dynamics of C82 by NMR	100
4.6 Discussion	103
4.7 Acknowledgments	107
4.8 Supplementary materials.....	107
4.9 References.....	108
Chapitre 5.....	113
Analyse structurale des diverses formes tronquées de la Core du VHC	113
5.1 Introduction.....	114
5.2 Matériels et méthodes	115
5.2.1 Protocole de production et purification de C124 et C170	115
5.2.2 Liaison croisée au glutaraldéhyde.....	115
5.2.3 Analyse Structurale.....	116
5.3 Résultats.....	118

5.3.1 Production et purification des formes C82, C124 et C170 de la Core du VHC	118
5.3.2 La protéine Core : une protéine multimérique ?	121
5.3.3 Analyse structurale par dichroïsme circulaire, fluorescence et RMN	125
5.3.4 Effet du TFE et du SDS sur la structure de la Core C124 et C170	127
5.3.5 Conclusion	134
Chapitre 6	135
Interaction des protéines Core (C82) et p53	135
6.1 Introduction	136
6.2 Matériels et méthodes	138
6.2.1 Protocole de production et purification de C82	138
6.2.2 Protocole de production et purification des diverses formes de p53	138
6.2.3 Dichroïsme circulaire	139
6.2.4 RMN	139
6.3 Résultats	140
6.4 Conclusion	151
Chapitre 7	153
Interfering with HCV assembly <i>in vitro</i> using affinity peptides directed towards Core protein	153
7.1 Résumé	154
7.2 Abstract	154
7.3 Introduction	155
7.4 Materials and methods	156
7.4.1 Expression and purification of C170	156
7.4.2 Kinetics of C170 <i>in vitro</i> assembly reactions	157
7.4.3 Peptide library	157
7.4.4 Electrophoresis mobility shift assay	157
7.4.5 Electron microscopy	158
7.5 Results & Discussion	159
7.5.1 <i>In vitro</i> assembly of hepatitis Core protein C170	159
7.5.2 Inhibition of assembly using peptides derived from the HCV Core protein sequence	160
7.5.3 Inhibition of assembly with peptides derived from NS5A, p21 and p73	165
7.6 Conclusion	168
7.7 Acknowledgments	169
7.8 Supplementary materials	169
7.9 References	170
Chapitre 8	173
Conclusion et perspectives	173
Annexes	181
Annexe 1	182
Structure and dynamics of the N-terminal half of hepatitis C virus Core protein: an intrinsically unstructured protein: Supplementary Data	182
Annexe 2	189
Structure and dynamics changes induced by TFE on of the N-terminal half of hepatitis C virus Core protein: Supplementary Data	189
Annexe 3	194

Interfering with HCV assembly <i>in vitro</i> using affinity peptides directed towards Core protein : Supplementary data	194
Annexe 4	200
A method for in vitro assembly of hepatitis C virus core protein and for screening of inhibitors	200
Bibliographie	210

Liste des tableaux

Tableau 1.1 : Liste de protéines cellulaires qui lient la protéine Core.	41
Tableau 5.1 : Résumé des valeurs de rayon hydrodynamique (R_h) mesuré pour les trois formes tronquées de la Core de l'hépatite C.	123
Tableau 5.2 : Valeurs calculées du rayon hydrodynamique (R_h) pour diverses formes multimérique de la Core de l'hépatite C.	124
Tableau 7.1 : Sequences of HCV Core peptides used in this study.	163
Tableau 7.2 : Sequences of diverse peptides tested as potential assembly inhibitors.	166

Liste des figures

Figure 1.1 : Schéma du modèle expérimental utilisé pour le clonage du virus de l'hépatite C.....	3
Figure 1.2 : Arbre phylogénétique de la famille des <i>flaviviridae</i>	5
Figure 1.3 : Carte de distribution géographique du VHC.....	8
Figure 1.4 : Représentation du génome du virus de l'hépatite et des protéines produites après divers clivages.....	11
Figure 1.5 : Modèles d'entrée du virus.....	14
Figure 1.6 : Structure de la particule virale du VHC.....	15
Figure 1.7 : Structure de la protéine p7 au sein de la membrane.....	18
Figure 1.8 : Structure de la protéine NS3 en complexe avec un court fragment N-terminal de NS4.....	20
Figure 1.9 : Structure sous forme dimérique du domaine I de la protéine NS5A.....	23
Figure 1.10 : Structure de la protéine NS5B.....	24
Figure 1.11 : Représentation schématique de la protéine Core.....	27
Figure 1.12 : Structure de la protéine Core du virus de la dengue.....	32
Figure 2.1 : Spectre de dichroïsme circulaire typique des trois types de repliement principaux.....	51
Figure 2.2 : RMN monodimensionnelle.....	57
Figure 2.3 : Représentation schématique d'une expérience 2D homonucléaire.....	58
Figure 2.4 : Spectre 2D ¹ H- ¹⁵ N HSQC de la protéine TEM-1.....	60
Figure 2.5 : Stratégie d'attribution de la chaîne principale.....	62
Figure 3.1 : Secondary structure predictions for C82.....	74
Figure 3.2 : Purification profile of the C82.....	77
Figure 3.3 : Circular dichroism and 2D ¹ H- ¹⁵ N HSQC spectra of C82.....	78
Figure 3.4 : Sequential NMR dynamic data.....	80
Figure 4.1 : Circular dichroism experiments on C82.....	95
Figure 4.2 : 2D ¹ H- ¹⁵ N HSQC and 2D ¹⁵ N NOESY-HSQC spectra of C82 in 90 % TFE.....	97
Figure 4.3 : Sequential NMR dynamic data for C82 in 90 % TFE.....	99
Figure 4.4 : Comparison of the distribution of heteronuclear NOEs measured.....	102
Figure 4.5 : Representation of C82 NMR dynamics data on the structure 1CWV.....	105
Figure 4.6 : Superposition of three 3D structures of peptide 1-25 from HCV Core protein.....	106
Figure 5.1 : Profil d'expression des diverses formes tronquées de la Core.....	119
Figure 5.2 : Profil de purification des diverses formes tronquées de la Core.....	120
Figure 5.3 : Résultats de liaison croisée au glutaraldéhyde pour les clones C82, C124 et C170.....	123
Figure 5.4 : Spectres de dichroïsme circulaire et de fluorescence des formes tronquées C124 et C170.....	126
Figure 5.5 : Spectres de dichroïsme circulaire des formes tronquées C124 et C170 en présence de quantité croissante de TFE et SDS.....	128
Figure 5.6 : Spectres de dichroïsme circulaire de la Core C170 en présence de TFE, SDS et de DM.....	129

Figure 5.7 : Spectres d'émission de fluorescence des tryptophanes pour les formes C124 et C170 en absence et en présence de quantité croissante de TFE et SDS.....	132
Figure 5.8 : Spectre 1D du proton de la protéine Core C170 en absence et en présence de 30 % TFE.....	133
Figure 6.1 : Profil de purification des formes monomérique, dimérique et tétramérique de la protéine p53.....	141
Figure 6.2 : Spectres de dichroïsme circulaire des protéines p53 monomérique, dimérique et tétramérique.....	142
Figure 6.3 : Représentation graphique des ratios utilisés lors des expériences RMN d'interaction Core C82 – p53.....	143
Figure 6.4 : Résumé des données d'interaction entre la Core C82 et p53 monomérique obtenues par résonance magnétique nucléaire.....	147
Figure 6.5 : Résumé des données d'interaction entre la Core C82 et p53 dimérique obtenues par résonance magnétique nucléaire.....	150
Figure 7.1 : HCV C170 purification and assembly.....	160
Figure 7.2 : Assembly kinetics of C170 with the addition of diverse peptides derived from HCV Core amino acid sequence.....	165
Figure 7.3 : Assembly kinetics of C170 with the addition of diverse peptides selected from three interacting proteins (NS5A, p21 and p73).....	168

Liste des abréviations

ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
ALAT	Alanine amino-tranf�rase
ARFP	Alternate Reading Frame Protein
ARN	Acide ribonucl�ique
ARN _t ou tRNA	ARN de transfert
BMRB	Biological magnetic resonance data bank
BVDV	Virus de la diarrh�e bovine
C	Cyst�ine
CSI	Chemical shift index (index de d�placement chimique)
Cys	Cyst�ine
CD	Circular dichroism
D	Acide aspartique
DDL	Diffusion dynamique de la lumi�re
DM	<i>n</i> -dodecyl- β-D-maltoside
DMSO	Dimethyl sulfoxide
DSS	Sodium 2,2-dimethyl-2-silapentane-5-sulfonate
DTT	Dithiothreitol
DV	Dengue virus ou Virus de dengue
EDTA	Ethylenediamine tetra-acetic acid
EGTA	Ethyleneglycol tetra-acetic acid
FID	Free induction decay
G	Glycine
GBV	GB virus ou Virus GB
H	Histidine
HChV	Virus de la peste porcine
HCV	Hepatitis C virus
HSQC	Heteronuclear single-quantum coherence
HVR	Highly variable region
IDP	Intrinsically disordered protein
IPTG	Isopropyl-1-thio-β-D-galactoside
IRES	Internal ribosome entry site
IUP	Intrinsically unstructured protein
IUPAC	International union of pure and applied chemistry
JEV	Virus de l'enc�phalite japonaise
K	Lysine
Met	Methionine
MW	Membrane Web
NDP	Natively disordered protein
NLP	Nucleocapsid-like particle
NMR	Nuclear magnetic resonance
NOE	Nuclear overhauser effect

NOESY	Nuclear overhauser effect spectroscopy
PDB	Protein data bank
Phe	Phenylalanine
PONDR	Predictor of natural disordered regions
PPII	Poly-proline de type II
PPV	Pseudo-particule virale ou Particule pseudo-virale
R	Arginine
Q	Glutamine
RdRp	RNA-dependent RNA polymerase
RE	Réticulum endoplasmique
R _h	Rayon hydrodynamique
RMN	Résonance magnétique nucléaire
RX	Cristallographie des rayons X
S	Serine
SDS	Sodium dodécyl sulphate
SDS-PAGE	Sodium dodécyl sulphate – polyacrylamide gel electrophoresis
SFV	Semliki Forest Virus
SP	Signal peptidase
SPP	Signal peptide peptidase
SSP	Secondary structure propensity
TFE	2,2,2-trifluoroéthanol
TOCSY	Total correlated spectroscopy
Tyr	Tyrosine
Trp	Tryptophane
UTR	Untranslated region
UV	Ultraviolet
VHA	Virus de l'hépatite A
VHB	Virus de l'hépatite B
VHC	Virus de l'hépatite C
VIH	Virus d'immunodéficience humaine
W	Tryptophane
WHO	Worldwide Health Organisation
WNV	West Nile virus (Virus du Nile)
YFV	Yellow fever virus (Virus de la fièvre jaune)

Chapitre 1

Introduction

1.1 Rappels historiques

Les avancées de la médecine ont été nombreuses lors du XX^{ème} siècle. La démocratisation de la transfusion sanguine a révolutionné le monde de la médecine et permis de soigner un grand nombre de patients. Par contre, d'un point de vue de santé publique, cette technique a créé une nouvelle voie de transmission d'agents infectieux tels que les virus, les bactéries, les protozoaires et les prions [62, 217, 234].

C'est au milieu des années 60 et début 70, que les virus de l'hépatite B (VHB) [24] et de l'hépatite A (VHA) [76] ont été découverts. Dès 1970, Alter et coll. recommandent que les échantillons sanguins soient testés pour exclure tout risque de propagation du virus et ce, peu après la mise au point d'un test de détection de VHB [8, 204]. L'application des tests de dépistage anti-VHA [142] et anti-VHB a permis dès 1975 de révéler l'existence d'autres hépatites post-transfusionnelles dites « non A et non B » [7].

Les premiers travaux concernant l'agent infectieux dit « non-A, non-B » ont permis de montrer : premièrement que celui-ci était transmissible au chimpanzé [9, 102, 241], qu'il était sensible à un traitement au chloroforme, laissant penser à un virus enveloppé [33], et vraisemblablement à ARN, dû à la méthode commune de purification [33], et enfin que sa taille était inférieure à 80 nm [34] et de l'ordre de 30-60 nm [97]. Malgré tous ces efforts, le mystère « non-A, non-B » reste non résolu, et ce, pour plusieurs raisons dont l'absence d'antigène spécifique, l'absence d'un système de culture cellulaire et l'absence d'anticorps détectables chez les patients.

À la fin des années 80, les chercheurs de l'entreprise « Chiron Corp. » ont cloné le génome de l'agent étiologique désormais appelé « Virus de l'Hépatite C » (VHC) [50]. La découverte de ce virus est le résultat d'un travail colossal utilisant des techniques de biologie moléculaire (Figure 1.1). Le projet ambitieux a nécessité une très grande quantité de plasma de chimpanzé, pour en extraire les ADNs et ARNs totaux. À partir de ceux-ci une banque d'ADN complémentaire a été générée, clonée au sein du bactériophage λ -gt11 et les produits surexprimés dans *E.coli* ont été criblés contre le sérum de patient infecté. Un

seul clone, parmi des millions testés, a donné un résultat, et un an plus tard, le génome complet était séquencé, identifiant ainsi le VHC comme l'agent responsable des hépatites dites non-A, non-B [6].

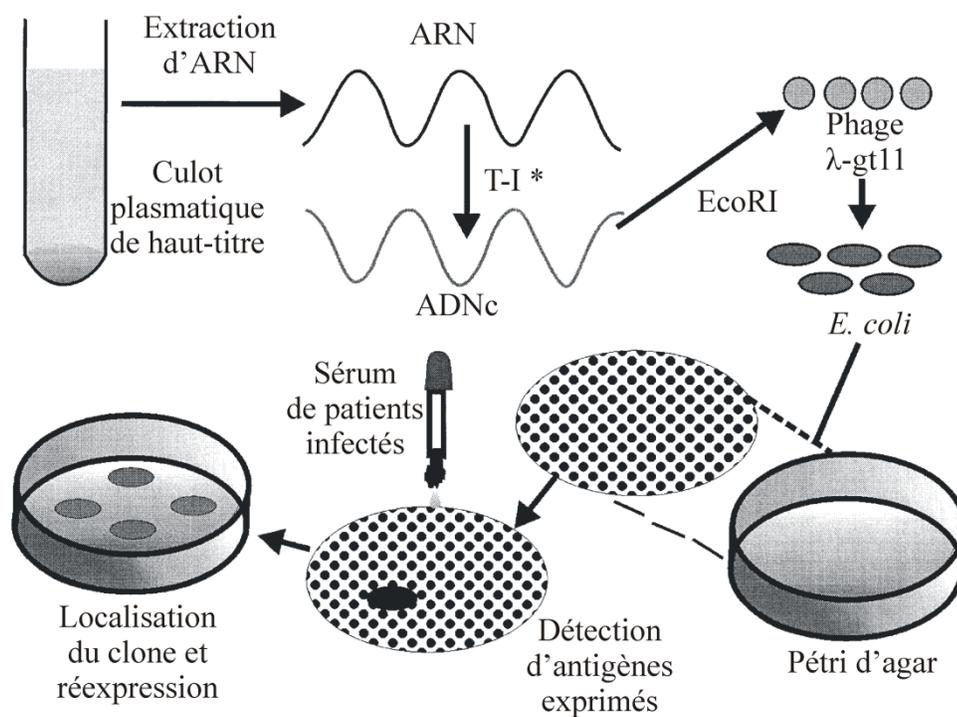


Figure 1.1 : Schéma du modèle expérimental utilisé pour le clonage du virus de l'hépatite C. (Figure adaptée de [6], * T-I signifie Transcription inverse)

1.2 Phylogénie du VHC

Pour des raisons de similitudes génétiques et de morphologie, le virus de l'hépatite C, nouvellement découvert, a été classé au sein de la famille des *Flaviviridae* [188]. Cette famille est composée de trois genres : les *flavivirus*, les *pestivirus* et les *hepacivirus* (Figure 1.2).

Le genre *flavivirus* est le plus représenté au sein de cette famille, avec plus de 50 représentants, qui pour la plupart sont des virus transmissibles par des insectes (virus de la fièvre jaune (YFV), virus de la dengue (DV), virus de l'encéphalite japonaise (JEV) et le virus du Nil (WNV), etc.).

Le genre *pestivirus*, avec lequel le VHC partage un degré élevé de similarité de séquence, est, quant à lui, constitué principalement de virus animaux. Les virus de la diarrhée virale bovine (BVDV) et de la peste porcine (HChV) en sont les principaux membres.

Le VHC constitue à lui seul le genre *hepacivirus* [65]. Mais de nouveaux prétendants existent. En effet, des études phylogénétiques récentes démontrent que les virus dits GBV (A, B et C) possèdent une grande similitude génétique avec le VHC [77, 212, 236], mais aucune décision concernant leur classification finale n'a été prise. Seul le virus GBV-B démontre un rapprochement possible au genre *hepacivirus* [77]. L'intérêt apparu au milieu des années 90 pour ces virus était dû au fait qu'ils représentaient un système alternatif potentiel d'étude du VHC jusqu'à la découverte récente d'un système de culture cellulaire efficace du VHC [160, 251, 278].

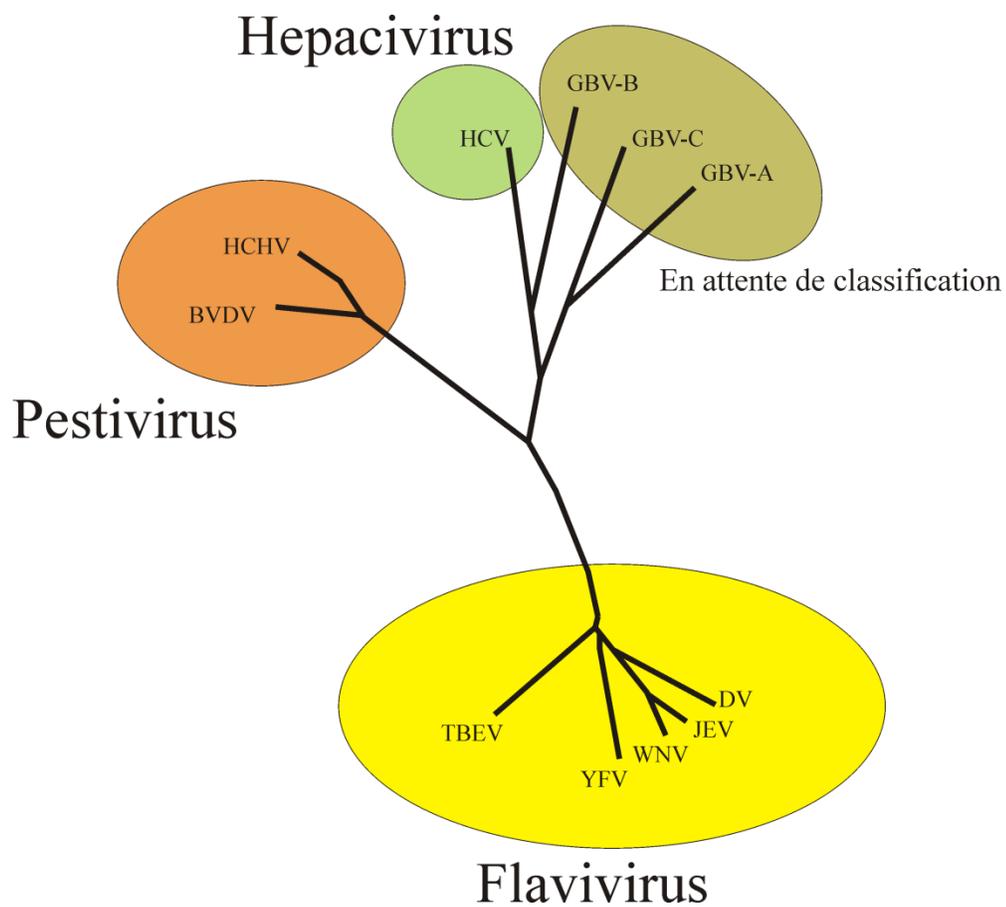


Figure 1.2 : Arbre phylogénétique de la famille des *flaviviridae*, basé sur l'analyse comparée de la protéine NS5B (Figure adaptée de [77]).

1.3 Le VHC, un virus génétiquement très diversifié

Le virus de l'hépatite C possède une grande diversité génétique. La comparaison de séquences de divers échantillons de VHC indique que celui-ci peut être classé en une série de génotypes bien distincts, qui diffèrent les uns des autres de par leurs séquences nucléotidiques, et montrent aussi une distribution géographique et épidémiologique variée [65, 144]. Six génotypes principaux ont ainsi été établis et divergent les uns des autres par environ 30 % de leurs séquences nucléotidiques [235]. Il existe aussi quelques génotypes moins connus et moins répandus, amenant ainsi le nombre de génotypes actuels à 11 génotypes. Les génotypes 7 à 11 sont principalement localisés en Asie (Thaïlande, Vietnam et Indonésie) [27].

Au sein d'un même génotype, il existe là encore une grande diversité génétique, on parle alors de sous-groupes ou sous-types, et ceux-ci sont identifiés par l'ajout d'une lettre minuscule au chiffre définissant leur génotype. On parlera alors de génotype 1a, 1b, 1c, 2a, 2b, etc., pour un total d'un peu moins de 50 sous-types. Pour que deux virus appartiennent au même sous-type, ceux-ci doivent avoir une identité de séquences de l'ordre de 70 % à 80 %. Pour des séquences ayant une identité supérieure à 90 %, on parlera alors de « quasi-espèces » au sein d'un sous-groupe. Cette large diversité génétique est certainement due à des erreurs accumulées lors de la réplication du génome viral par l'ARN polymérase du VHC.

Globalement, le VHC possède une diversité génétique importante. Mais de manière surprenante, les régions 5' et 3' du génome du VHC sont beaucoup plus conservées que le reste du génome.

1.4 Épidémiologie et transmission du VHC

Le virus de l'hépatite C est un problème majeur de santé publique. Les estimations les plus optimistes font état de 100 à 120 millions de cas, tandis que les plus pessimistes annoncent au dessus de 200 millions de personnes infectées. Le chiffre souvent évoqué de 170 millions date de 1999, et représente près de 3 % de la population mondiale (WHO, <http://www.who.int/csr/disease/hepatitis/Hepc.pdf>). Ces chiffres ne sont pas anodins, en comparaison le virus d'immunodéficience humaine (VIH) dénombre 33,3 millions de personnes infectées, ce qui est cinq fois moins que le VHC (rapport WHO 2009, http://www.who.int/hiv/data/2009_global_summary.png).

L'ensemble de la planète est touché par le VHC, et sa prévalence est géographiquement très différente. Les pays industrialisés de l'ouest (Amérique du Nord et Europe, principalement) sont de nos jours les moins touchés avec une prévalence souvent inférieure à 1-2 % (Figure 1.3). Ceci a été rendu possible grâce au dépistage systématique du VHC dans les banques de sang, rendant ainsi la transmission par transfusion quasi-nulle. Au Canada, les cas déclarés sont d'environ 40 personnes infectées pour 100 000 personnes (< 0,1 %). Notons qu'il s'agit du plus faible taux d'infection mondiale [27]. Dans la province de Québec, ~ 20 personnes sont infectées pour 100 000 personnes, alors que le territoire du Yukon montre le nombre le plus élevé de personnes infectées avec ~ 90 cas pour 100 000 personnes (Agence de la santé publique du Canada et Statistiques Canada).

L'Asie et l'Afrique présentent les taux les plus élevés (> 2 %). Notons le cas exceptionnel de l'Égypte et ses ~ 20 % de la population infectée au VHC, qui sont dus à une contamination iatrogénique massive lors d'une campagne de prévention contre le schistosome [81].

La transmission du VHC se fait par contact direct avec le sang et ses produits dérivés. La transfusion sanguine a été longtemps considérée comme un des principaux facteurs de la propagation du VHC. De nos jours, le faisceau s'est élargi et l'on sait que la transmission se fait principalement par transfusion sanguine non contrôlée, l'usage de

drogues intraveineuses, par transmission nosocomiale (matériel médical mal ou non stérilisé), lors de transplantations d'organes, et par des pratiques sociales et culturelles telles que le piercing et le tatouage. Les transmissions sexuelles et de mère à enfant sont des cas rares et isolés [27].



Figure 1.3 : Carte de distribution géographique du VHC (Figure tirée de [227]).

1.5 Physiopathologie et traitement du VHC

L'infection au VHC se distingue par une phase aiguë et une phase chronique. La phase aiguë de l'infection par le VHC est quasi-asymptomatique. Quelques sujets présenteront les signes d'une grippe ou jaunisse qui ne persistera pas. Par contre, les marqueurs biochimiques tels que la mesure de l'ARN viral, du taux d'alanine aminotransférase (ALAT) et de la présence d'anticorps anti-VHC [206] sont facilement mesurables et quantifiables, permettant ainsi de poser un diagnostic précis. Après généralement 3-4 mois (durée de la phase aiguë), 20 % des personnes infectées élimineront le virus de manière naturelle, et 80 % des personnes développeront une hépatite dite chronique. L'hépatite chronique est caractérisée d'un point de vue biochimique par la persistance d'ARN viral dans le sang, d'anticorps anti-VHC et de taux variables d'ALAT qui définissent une chronicité évolutive ou non. Si la chronicité est non évolutive, le patient sera soumis à des contrôles réguliers pour vérifier l'avancement de la maladie. 60 % des personnes infectées développent une hépatite chronique évolutive qui dans ~25 % des cas mènera au développement d'une cirrhose dans les 20 à 30 années suivantes, valeur qui est estimée à 45 % dans la perspective des années 2030 [57]. Un peu moins de 10 % des cas évolueront vers un cancer du foie, qui est aujourd'hui la principale cause des transplantations de foie.

Les traitements du VHC sont actuellement restreints à l'utilisation d'interféron alpha pégylé en association ou non à la ribavirine. Ce traitement n'est malheureusement pas des plus efficaces, car seulement ~50 % des patients élimineront le virus. Les recherches actuelles se dirigent vers des vaccins prophylactiques ou divers inhibiteurs du cycle cellulaire du virus (inhibiteurs de protéase, de polymérase, d'assemblage viral). Les besoins s'orientent vers une médication plus efficace et présentant moins d'effets secondaires que le traitement actuel [22].

1.6 Organisation génétique

Le virus de l'hépatite C est un virus enveloppé possédant un ARN simple brin de polarité positive, de taille avoisinant les 9600 paires de base et codant une seule polyprotéine d'environ 3000 résidus d'acides aminés (Figure 1.4). Cet ARN simple brin possède à ses extrémités 5' et 3' des régions non codantes, de 341 et ~230 nucléotides respectivement, extrêmement conservées et impliquées dans la réplication du génome, la synthèse de la polyprotéine [248], dans la dimérisation du génome [55] et dans l'assemblage viral [145]. La région 5', aussi appelée IRES pour « Internal Ribosome Entry Site », est responsable de la traduction coiffe-indépendante du génome viral [248, 253].

La polyprotéine produite va subir plusieurs clivages successifs impliquant des protéases cellulaires et virales menant ainsi à la production de 10 protéines que l'on peut scinder en deux groupes : les protéines structurales et les protéines non-structurales. Le premier groupe est constitué de la protéine Core, impliquée dans la formation de la capsid, et des protéines d'enveloppe E1 et E2. On retrouve à l'interface de ces deux groupes, la protéine p7 dont le rôle supposé est d'être une viroporine. Le deuxième groupe est constitué de 6 protéines impliquées principalement dans la maturation, la réplication du génome viral, ainsi que dans la formation de la particule virale (Figure 1.4). Notons l'existence d'une 11^e protéine, issue d'un changement de cadre de lecture au sein de la séquence de la protéine Core, que l'on nomme protéine F (ou ARFP, ou Core+1), dont le rôle reste à éclaircir [252].

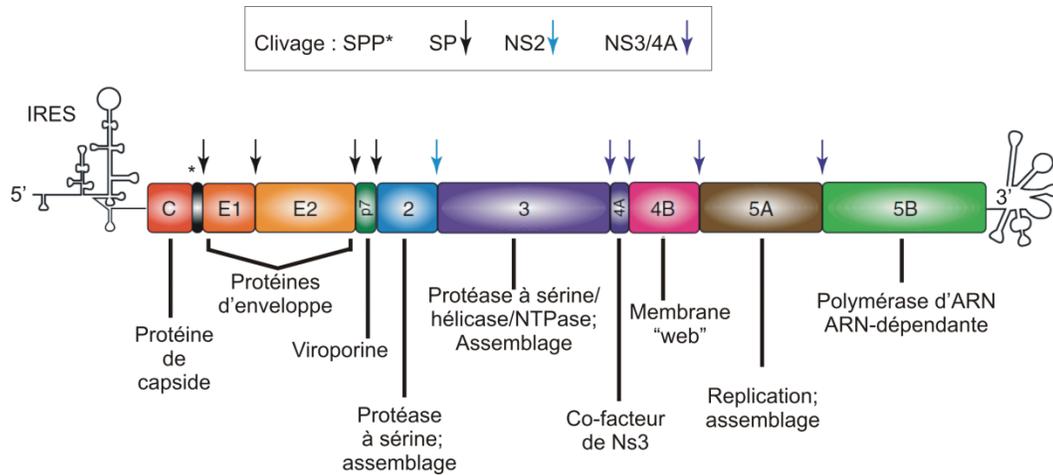


Figure 1.4 : Représentation du génome du virus de l'hépatite et des protéines produites après divers clivages. La maturation de la Core est abordée à la section 1.10.2 (page 27) (Figure adaptée de [17]).

1.7 Le cycle répliatif du VHC

Le cycle répliatif peut être divisé en quatre parties : l'entrée du virus au sein de la cellule, la traduction, suivie de la répliation du génome viral, et enfin l'évasion virale.

L'entrée du VHC se fait par l'entremise de protéines membranaires présentes à la surface de la cellule hôte. Deux modèles similaires sont souvent évoqués et ne varient que par le simple fait que le virion se présente, soit seul, soit en association avec des lipoprotéines pour infecter la cellule (pour revue : [17, 18, 40, 51]) (Figure 1.5 A et B). Dans le 2^e cas, des récepteurs aux lipoprotéines (le récepteur LDL et des glycosaminoglycanes) faciliteront le rapprochement de la particule virale à ses récepteurs spécifiques. Les protéines impliquées sont les récepteurs SR-BI et CD81 qui lient le virion via une interaction avec la glycoprotéine d'enveloppe E2. Le virion est ensuite transféré

aux protéines Claudin-1 et Occludine, présentes au niveau des jonctions serrées, et est, par la suite, internalisé par un mécanisme d'endocytose dépendant des clathrines (pour revue : [18, 40, 51, 245]). Suite à la diminution du pH intra-endosomal, la nucléocapside contenant l'ARN viral est libérée au sein du cytoplasme par fusion des membranes virale et endosomale, suivi d'une étape de décapsidation pour relâcher l'ARN viral. Une fois libéré, l'ARN est traduit au niveau du réticulum endoplasmique granuleux (REG). Cette étape est contrôlée par la région 5' non codante (NC, ou UTR (pour « untranslated region »)) de l'ARN viral. Cette région 5'-UTR est constituée de 341 nucléotides extrêmement conservés formant quatre domaines repliés en tige boucle numérotés de I à IV incluant le codon AUG initiateur [213]. Les domaines II à IV constituent l'IRES, celui-ci va recruter la sous-unité 40S du ribosome, les facteurs de traduction eIF2 et eIF3 et la sous-unité 60S facilitant la formation du ribosome essentiel à la traduction (pour revue [228]). L'ARN viral sera traduit donnant lieu à une longue polyprotéine d'environ 3000 résidus d'acides aminés, qui une fois maturée par des protéases cellulaires et virales mènera à la production de 10 protéines. La synthèse des protéines du VHC a pour effet de modifier de manière importante la distribution interne des lipides via l'augmentation des gouttelettes lipidiques, une dilatation du RE, et une augmentation des membranes internes de la cellule, favorisant la formation d'une toile membranaire appelée « web » [69]. C'est dans cet environnement lipidique que le complexe répliatif incluant les protéines non structurales du VHC et des protéines de l'hôte va se former et recruter l'ARN viral [69, 85, 191]. De faibles quantités d'ARN à polarité négative sont produites pour, à leur tour, servir de matrice à la synthèse d'ARN à polarité positive. L'initiation de l'assemblage viral s'effectue au niveau des gouttelettes lipidiques et requiert la présence de protéines non structurales pour permettre la production de virus infectieux, faisant en sorte que les sites de réplication et d'assemblage soient intimement reliés. Le virion se forme ensuite par bourgeonnement au niveau du RE au niveau duquel les protéines E1 et E2 sont ancrées. La maturation et l'évasion sont des processus qui sont encore obscurs du simple fait que les systèmes de production de virus infectieux sont peu efficaces pour le relargage d'une grande quantité de virus [17, 119].

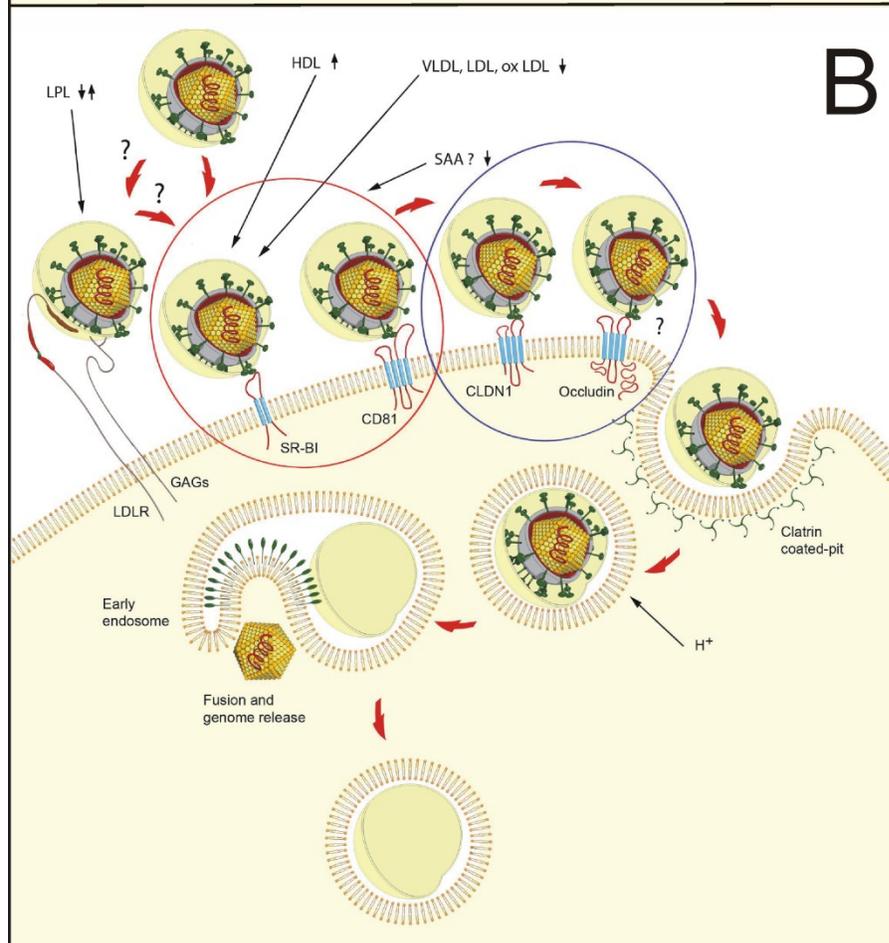
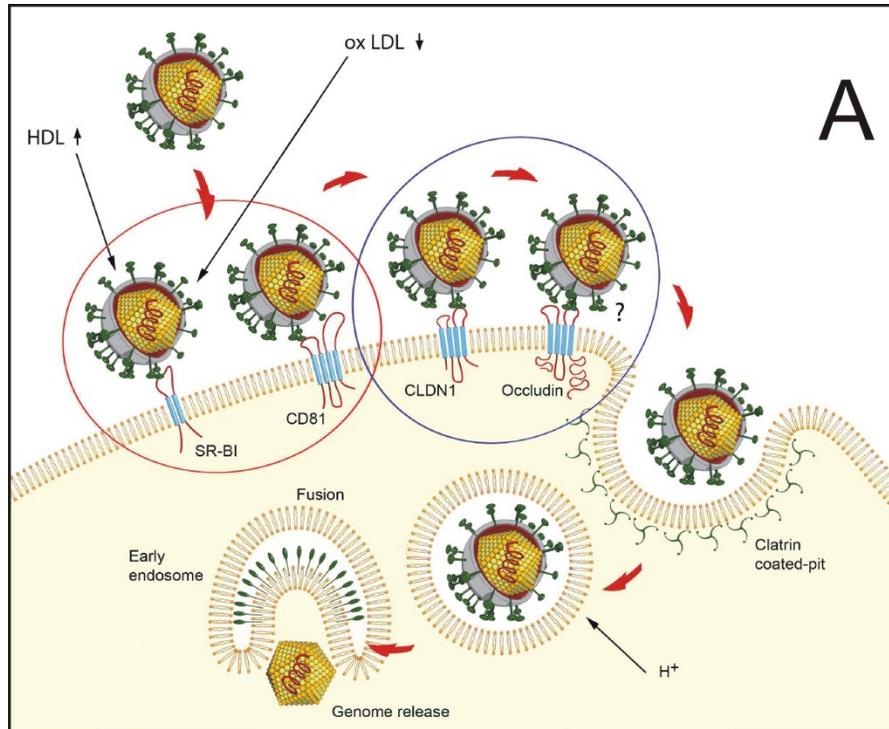


Figure 1.5 : Modèles d'entrée du virus. A : virion seul, B : Virion associé aux lipides et lipoprotéines (Figures tirées de [40]) (Voir page précédente).

1.8 La particule virale : Structure

La famille des *flaviviridae* est composée de petits virus d'environ 50 nm de diamètre. Les structures des virus de la dengue (DV) et du virus du Nile (WNV) ont été résolues [143, 194, 277]. Les images de cryo-microscopie montrent de manière claire l'organisation ultra-structurale de ces virus. À la surface, on retrouve les protéines d'enveloppe ancrées dans une bicouche lipidique de 35 à 40 Å [194] et enfin, la capsidie contenant l'ARN viral. De manière non surprenante, la comparaison de la structure des virions montre qu'ils sont très similaires, on remarquera une légère divergence au niveau de l'organisation des protéines d'enveloppe.

Pour ce qui est du VHC, l'absence d'un système efficace de production de virus a longtemps été un frein à l'étude structurale de la particule virale. Cependant, alors que le VHC n'avait jamais été observé, les travaux des docteurs Bradley [34] et He [97] ont montré que le diamètre du VHC était inférieur à 80 nm et compris entre 30 et 60 nm. Les études subséquentes produites au cours des années 1990 et 2000 ont confirmé ces résultats et ont permis de différencier les formes enveloppées du virus (~50-65 nm) des formes non-enveloppées (~30-45 nm) [84, 122, 175, 275]. La structure 3D d'un pseudo-virion a été obtenue par cryo-microscopie. Les images obtenues à une résolution de 30 Å confirment la ressemblance du VHC aux autres *flavivirus* [274] (Figure 1.6). De plus, des études de microscopie électronique à rotation optique ont révélé que la capsidie (particule non-enveloppée) possédait une structure icosaédrique et que la longueur de l'arrête mesurait ~20 nm [110].

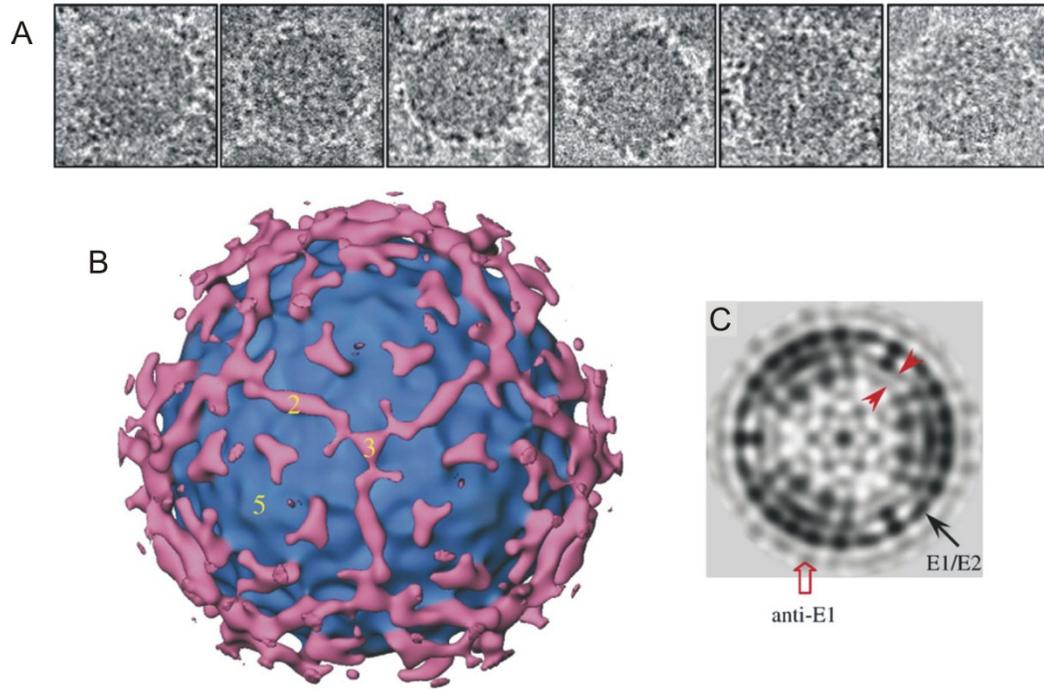


Figure 1.6 : Structure de la particule virale du VHC a) images de cryo-microscopie de particules virales du VHC, b) représentation de la reconstruction 3D de particules virales du VHC recouverte d'anticorps (en violet) c) coupe de densité d'une particule virale obtenue par cryo-microscopie (les pointes de flèches rouges indiquent la bicouche lipidique, la flèche bleue indique les protéines E1/E2, et la flèche rouge indique les anticorps anti-E1) (Figure tirée de [274]).

1.9 Les protéines du VHC

1.9.1 Les protéines structurales

1.9.1.1 La protéine Core

La protéine Core du VHC étant le principal champ d'étude de ce doctorat, une section détaillée lui sera consacrée (pages 26 à 41).

1.9.1.2 Les protéines E1 et E2

Les protéines E1 et E2 sont les deux autres protéines dites structurales du VHC. En effet, elles constituent la surface du virion du VHC, où elles sont implantées dans la bicouche lipidique présentant à l'extérieur du virion leurs chaînes polypeptidiques permettant l'adhésion du virion sur les cellules à infecter. Les protéines E1 et E2 possèdent des régions très peu conservées au niveau de leur séquence. Ces régions sont appelées HVR1 et HVR2, pour « Highly Variable Region ». En effet, ces deux protéines sont à l'origine de l'interaction du virus avec la membrane cellulaire. Elles sont aussi les protéines de surface qui sont en contact avec le système immunitaire. Les régions HVR permettent ainsi au virus d'échapper à une reconnaissance rapide du virus par le système immunitaire.

D'un point de vue biochimique et structural, les protéines E1 et E2 sont clivées de la polyprotéine par la « signal peptidase ». Les deux protéines possèdent un ancrage au sein de la membrane du RE. Elles possèdent un nombre important de modifications post-traductionnelles. Au niveau structural, elles forment un hétérodimère [63, 223] et l'association des deux protéines permettrait leur bon repliement. À ce jour, la structure d'aucune de ces deux protéines n'est disponible, mais la région d'ancrage de E1 dans la membrane a été déterminée par RMN [238], ainsi qu'un court segment de E2 [4].

1.9.1.3 La protéine p7

À la jonction entre les protéines structurales et les protéines non-structurales, on retrouve une petite protéine hydrophobe de 63 acides aminés [157]. Ce court segment protéique provient d'un clivage par la « signal peptidase » (SP) entre la protéine E2 et NS2. La protéine p7 est une protéine membranaire intégrale possédant deux passages transmembranaires dont l'orientation reste à ce jour controversée [52, 53, 190]. On la retrouve principalement au réticulum endoplasmique [42]. La protéine s'assemblerait en hexamère ou en heptamère formant un pore [52, 53, 190]. La fonction supposée de cette protéine est d'être une viroporine. Ce type de protéine est retrouvé dans des virus apparentés (tel BVDV), mais aussi au sein du virus VIH (avec la protéine VPU) et au sein du virus influenza (avec la protéine M2). Ces pores auraient pour fonction de contrôler l'acidification facilitant l'entrée et/ou la sortie du virus de la cellule hôte. Par ailleurs, l'activité inhibitrice de composés inhibiteurs de viroporine vient confirmer sa potentielle fonction au sein du VHC. D'un point de vue structural, des travaux récents sur la structure de la protéine p7 en solution sous forme monomérique sont venus confirmer la structure prédite en épingle à cheveux, c'est-à-dire, deux passages transmembranaires reliés par une courte boucle contenant deux acides aminés très conservés (K33 et R35). Cependant, ces travaux complémentaires viennent de manière étonnante complexifier la nature de la structure tridimensionnelle de la protéine p7. Chacun des passages transmembranaires serait formé en réalité de deux hélices alpha, reliés par de très courtes régions flexibles (Figure 1.7).

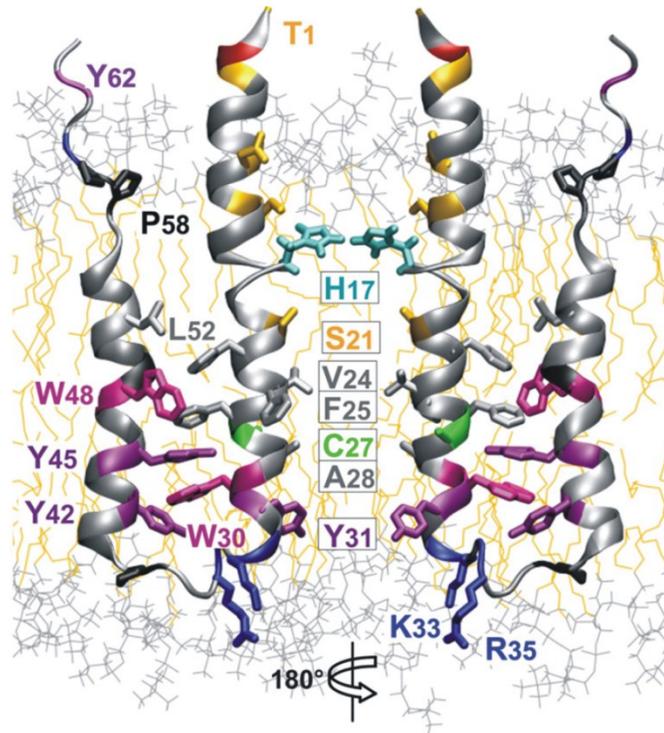


Figure 1.7 : Structure de la protéine p7 au sein de la membrane. Le modèle représenté tente de simuler l'orientation des unités monomériques au sein du pore. (Figure tirée de [190]) (PDB ID : 2K8J).

1.9.2 Les protéines non structurales

1.9.2.1 La protéine NS2

NS2 est une protéine membranaire intégrale de 23 kDa (217 acides aminés) non glycosylée. Elle posséderait trois passages transmembranaires et un domaine cytosolique. Longtemps, son rôle au sein du cycle viral est resté inconnu. Aujourd'hui, sa participation au clivage NS2/NS3 est bien décrite. La protéine formerait un dimère au niveau de la membrane du réticulum endoplasmique (RE) où le clivage NS2/NS3 serait initié. Ce clivage est facilité par la présence de membranes, les régions C-terminale de NS2 et N-

terminale de NS3 sont indispensables pour le clivage. De plus, il a été démontré que ce clivage serait dépendant du zinc. NS2 au sein de la polyprotéine NS2-NS5 serait ainsi une métalloprotéase à cystéine constituée d'une diade catalytique impliquant les résidus H952 et C993. Toutefois, son rôle dans sa forme mature (clivée) est resté longtemps inconnu. Des travaux récents impliqueraient NS2 dans l'assemblage et le relargage du virus, lui attribuant un rôle de plateforme permettant de recruter tous les acteurs de l'assemblage viral [61, 116, 173, 203, 271].

1.9.2.2 La protéine NS3

Située en C-terminal de NS2 avec qui elle collabore pour engendrer le clivage NS2/NS3, NS3 est une protéine multifonctionnelle. Il s'agit d'une protéine de 631 résidus, que l'on peut diviser en deux domaines majeurs ayant des activités différentes. Le domaine 1, allant des résidus 1 à 180, possède une activité protéase à sérine [16, 68, 89, 246] tandis que le domaine 2, allant de 181 à 631, montre une activité hélicase et NTPase [94, 123, 129]. Plusieurs structures 3D ont été obtenues, que ce soit de la protéine pleine longueur, de chaque domaine indépendant, en association ou non avec NS4A ou des ARNs.

Le domaine protéase à sérine présente un repliement de type chymotrypsine et est composé de deux tonneaux bêta. Le bon repliement de NS3 se fait via l'interaction avec NS4A. La triade catalytique est composée des résidus H57, D81 et S139 dont les mutations abolissent l'activité de NS3 [16] (Figure 1.8). Lorsque correctement repliée en présence de NS4A, NS3 procède au clivage en *cis* NS3/NS4A et engendre par la suite en *trans* les clivages permettant de produire les protéines NS4A, NS4B, NS5A et NS5B.

Le domaine C-terminal possède, quant à lui, une activité hélicase et NTPase. Ce domaine possède trois sous-domaines séparés par des cavités [269] (Figure 1.8) . Le domaine hélicase-NTPase de NS3 se lie de manière efficace à l'ARN [94], mais pour effectuer son rôle, il est nécessaire qu'elle se retrouve sous la forme d'un dimère [48, 128].

Les différents domaines ont souvent été étudiés de manière indépendante, mais l'activité est maximale lorsque NS3 est complète et que son cofacteur (NS4A) est présent [134].

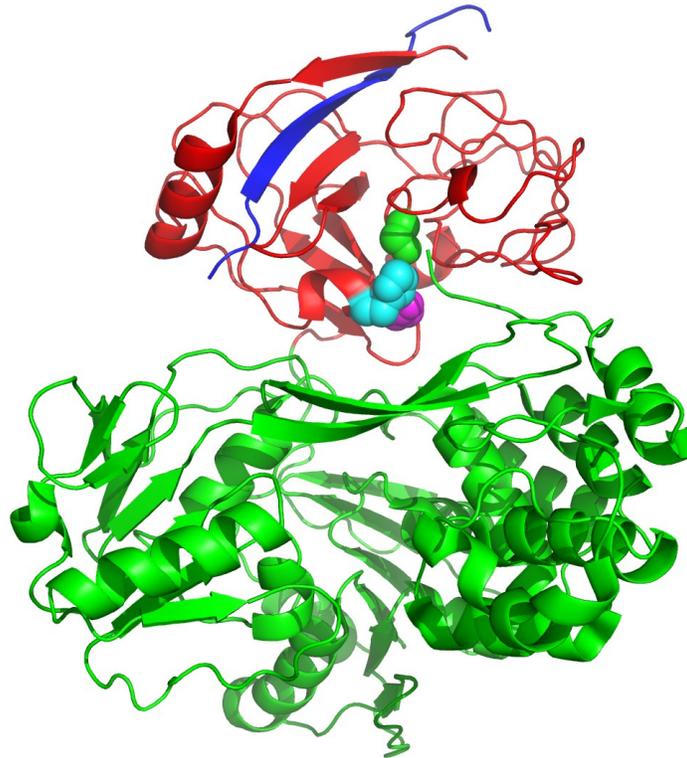


Figure 1.8 : Structure de la protéine NS3 en complexe avec un court fragment N-terminal de NS4. Le domaine protéase à serine est un rouge et la triade catalytique est indiquée au centre (H57 (bleu turquoise), D81 (magenta) et S139 (vert)). Le domaine hélicase/NTPase est en vert. Et le court fragment N-terminal de NS4 est en bleu. (PDB ID : 1CU1).

1.9.2.3 La protéine NS4A

NS4A est une protéine hydrophobe de 54 acides aminés (8 kDa). Elle possède en N-terminal un segment transmembranaire, suivi d'une courte région qui serait impliquée dans

l'interaction avec la protéine NS3 [72, 158]. L'interaction permettrait le repliement de la région N-terminale de NS3 [247, 268]. Cette interaction rendrait NS4A indispensable au bon fonctionnement de NS3, c'est pour cette raison que NS4A est considérée comme un cofacteur. NS4A aurait un rôle dans la régulation de la phosphorylation de NS5A via sa région C-terminale [13, 159, 161].

1.9.2.4 La protéine NS4B

La protéine NS4B est une longue protéine hydrophobe de 261 résidus. Il s'agit d'une protéine intégrale possédant plusieurs passages transmembranaires. La topologie exacte de NS4B au niveau du RE reste à confirmer, mais celle-ci serait dynamique, en particulier pour sa région N-terminale composée de deux hélices (résidus 1 à 69) qui pourraient être orientées, soit vers le cytosol, soit vers la lumière du RE [88], impliquant qu'une des deux hélices puisse traverser la membrane, ce qui a été démontré par les travaux de Gouttenoire et coll. [86]. La région centrale (résidus 70 à 190) serait composée de 4 hélices transmembranaires et la queue C-terminale (résidus 191 à 261) serait composée de deux hélices amphipathiques orientées vers le cytosol interagissant avec les membranes [87].

NS4B serait à l'origine de la redistribution des lipides sous forme de « toiles membranaires » ou en anglais « Membrane Web » (MW) qui s'accumuleraient près des gouttelettes lipidiques, lieu privilégié de la réplication du virus [85, 191]. Toutefois, le fonctionnement de NS4B pour cette fonction spécifique reste à ce jour inconnu. Récemment, il a été reporté que NS4B posséderait une activité enzymatique de type NTPase [70], mais cette fonction reste controversée [88].

1.9.2.5 La protéine NS5A

NS5A est une protéine de 448 résidus qui s'associe aux membranes par l'entremise d'une hélice amphipathique présente à son extrémité N-terminale [36]. On retrouve NS5A

sous 2 formes présentant des poids moléculaires de 56 et 58 kDa. Cette différence, originellement attribuée à une variation du site de clivage par la NS3, est en réalité due à des formes phosphorylées et hyperphosphorylées [211]. En effet, cette protéine possède un grand nombre de résidus sérine et thréonine qui sont des sites de phosphorylation potentielle. NS5A est composée de trois domaines, en plus de son domaine d'ancrage aux membranes, séparés par de courtes régions peu complexes. Le domaine I est le plus grand des trois domaines. Il possède la capacité de lier un atome de Zn^{2+} [244]. La détermination de sa structure 3D a permis de mettre en évidence la formation d'un homodimère (Figure 1.9). Les domaines II et III sont riches en proline et leurs analyses structurales respectives montrent que ces domaines sont intrinsèquement non structurés [95, 96, 154, 156].

Son rôle reste certes peu connu, mais il s'agit d'une protéine indispensable pour la réplication et l'assemblage viral. NS5A possède un grand nombre d'interactions avec des protéines du virus (Core, NS5B) et de l'hôte [98] régulées par sa phosphorylation, mais la pertinence de ces interactions reste à clarifier. Par ailleurs, il a été montré que NS5A possède une affinité pour les ARNs et plus particulièrement pour les régions poly (U/UC) présentes à l'extrémité 3' UTR du génome de HCV [106]. Les trois domaines de NS5A seraient impliqués dans cette interaction [80, 108]. NS5A jouerait ainsi un rôle important dans la régulation de la réplication, et serait un acteur clé de l'assemblage viral grâce à son interaction avec la protéine Core [182].

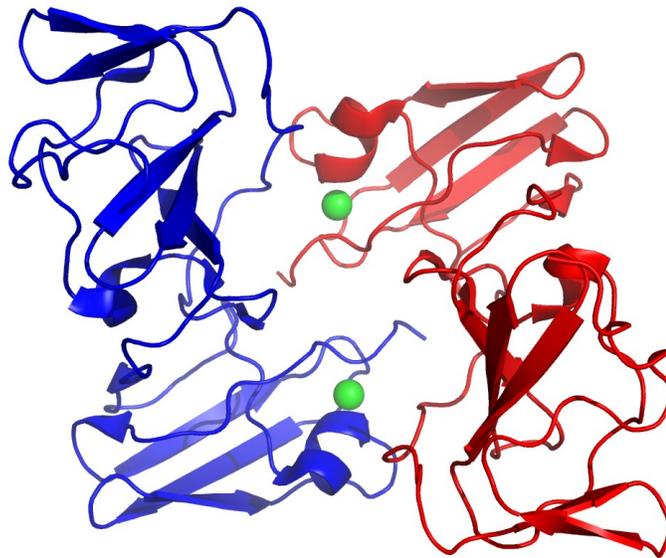


Figure 1.9 : Structure sous forme dimérique du domaine I de la protéine NS5A lié à un atome de Zinc (en vert) (PDB ID : 1ZH1).

1.9.2.6 La protéine NS5B

NS5B, comme la plupart des protéines du VHC, est une protéine associée aux membranes via un court domaine transmembranaire dans sa région C-terminale [220]. NS5B est une protéine de 591 résidus et correspond à l'ARN polymérase ARN dépendante (RdRp) [20, 166]. Elle a pour fonction de synthétiser de l'ARN à partir d'ARN. La structure 3D de NS5B montre un repliement comparable à ceux d'autres protéines RdRp connues, présentant la forme d'une main que l'on divise en trois sous-domaines appelés : paume, doigts et pouce [2, 37, 152] (Figure 1.10). Le site actif se situe au niveau de la paume, tandis que les doigts et le pouce favorisent l'interaction avec l'ARN. L'initiation de

la réplication nécessite la présence du complexe de réplication impliquant d'autres protéines non structurales (NS3, NS5A). Les RdRps sont des enzymes dépourvues d'activité correctrice, ce qui explique la grande diversité génétique du VHC au sein d'un même individu.

Notons toutefois que NS5B fait l'objet de recherches particulièrement intenses, car elle est une cible intéressante et prometteuse afin de développer des inhibiteurs efficaces de la réplication virale. L'obtention de ces inhibiteurs associés à d'autres moyens de thérapies permettrait certainement de supprimer le virus présent chez le patient.

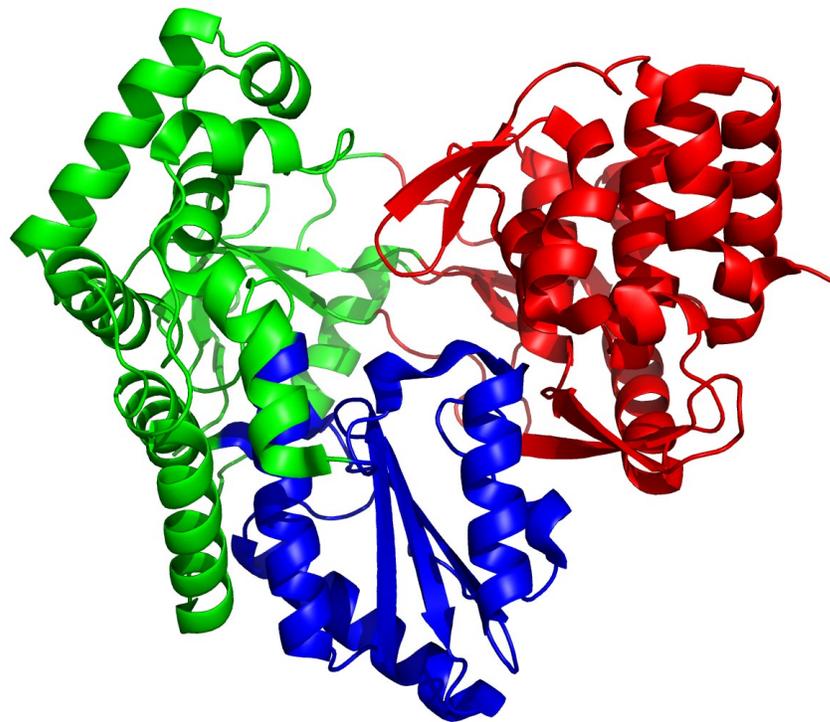


Figure 1.10 : Structure de la protéine NS5B, formée de la paume (en bleu), du pouce (en rouge) et des doigts (en vert). (PDB ID : 1GX6).

1.9.2.7 Les protéines ARFP/F/Core+1

En 2001, Walewski et coll. ont montré par des études bio-informatiques l'éventualité d'une 11^e protéine [252]. Celle-ci se situerait dans la région codante de la protéine Core. Son existence a été confirmée plus tard par des études de traduction *in vitro* et dans des cellules transfectées. Le mécanisme d'obtention de cette protéine reste controversé et serait dépendant du génotype [250], mais on dénombre 3 mécanismes principaux (pour revue : [35, 250]) : 1- un changement du cadre de lecture +1 [28, 49, 265], 2- l'utilisation d'un codon AUG initiateur interne [32, 249] et 3- une initiation de la traduction à un codon dit « non-AUG » [15, 32]. Les protéines produites possèdent plusieurs noms : ARFP (pour « alternative reading frame protein »), F (pour « frameshift ») ou encore Core+1/F/S (pour indiquer le changement de cadre de lecture, et les formes longue (F) et courte (S)). Diverses formes ont ainsi été mises en évidence, ayant des tailles variées allant de 38 à 162 résidus [32, 250]. Toutes possèdent un caractère basique important [250], et sont peu stables [264]. D'un point de vue structural, deux formes de Core+1/S ont été étudiées et celles-ci ne présentent pas de structure secondaire stable, mais possèdent la capacité de se multimériser [31].

Son rôle n'est pas clair. Elle n'est pas indispensable pour la production d'une infection [186], ce qui est appuyé par le fait que la protéine Core régule de manière négative la production de la protéine ARFP/F/Core+1 [261]. Son rôle serait peut-être en lien avec la réponse immunitaire ou le développement du cancer [56, 79, 224, 262].

1.10 La protéine Core du VHC

1.10.1 Organisation

La Core du VHC est une petite protéine de 191 acides aminés (~21 kDa), elle possède un point isoélectrique élevé proche de 12, et il s'agit de la protéine la plus conservée au sein du génome viral. L'analyse de sa structure primaire permet de la diviser en trois domaines distincts (Figure 1.11). Le domaine D1 (aa 1-120) est le domaine présumé d'interaction avec l'acide nucléique. Il possède un caractère hydrophile fort et est composé d'un très grand nombre d'acides aminés chargés (3 régions), espacés par deux segments hydrophobes (Figure 1.11). Le domaine D2 (aa 121-177), que l'on peut qualifier de central, est globalement hydrophobe. Ce domaine permettrait l'interaction de la Core avec les membranes et les gouttelettes lipidiques. Enfin, le troisième domaine dit D3 (aa 178-191) est un segment très hydrophobe, il serait intégré dans la membrane du réticulum et permettrait le clivage Core/E1.

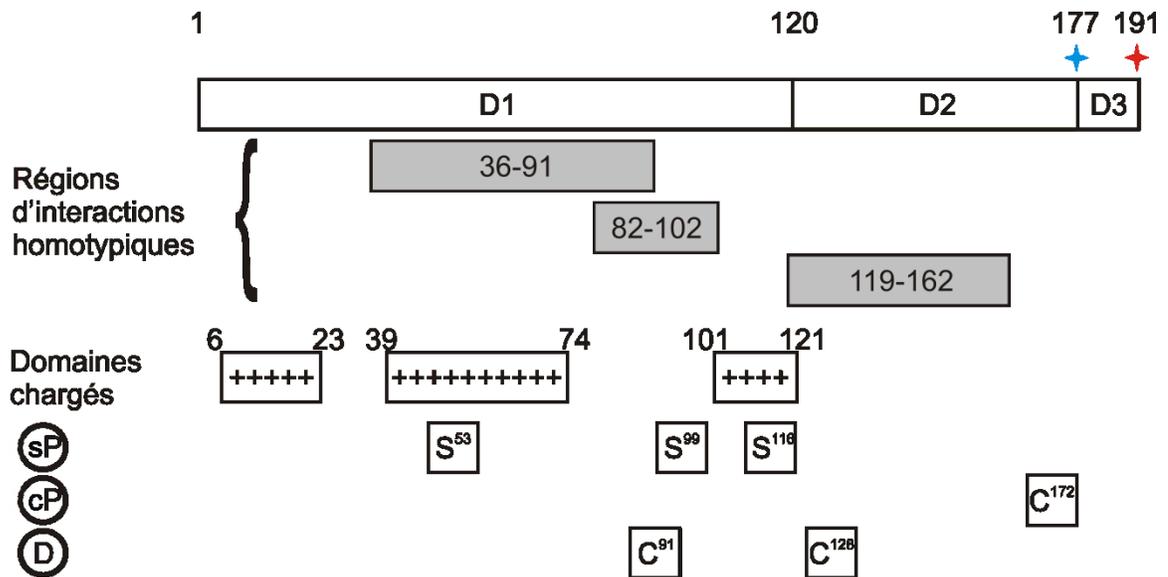


Figure 1.11 : Représentation schématique de la protéine Core. Les sites de clivages par les protéines « Signal Peptidase » et « Signal Peptide Peptidase » sont indiqués par des étoiles orange et bleue. Les régions d'interactions homotypiques cartographiées sont indiquées, de même que les trois sous-domaines chargés positivement. Les sites de modifications post-traductionnelles sont indiqués (phosphorylation (sP), de palmitoylation (cP)) ainsi que les cystéines C⁹¹ et C¹²⁸ potentiellement impliquées dans la formation d'un pont disulfure inter-moléculaire (D) permettant la dimérisation de la protéine.

1.10.2 Maturation et localisation

La protéine Core se situe en N-terminal de la polyprotéine produite, et est la première protéine extraite de celle-ci. Son clivage est orchestré en deux temps au niveau du RE par deux protéases de l'hôte. Le premier clivage implique la protéine « signal peptidase » (SP) [100, 218] et permet le relâchement d'une première forme immature de la Core (p21) de 191 acides aminés du reste de la polyprotéine. Cette forme de la Core reste

prise au niveau du RE grâce à l'ancrage du domaine D3 et à l'interaction du domaine D2 aux membranes [29, 200]. Le deuxième clivage implique la protéine « signal peptide peptidase » (SPP), qui va produire un clivage au niveau de l'ancrage à la membrane pour relâcher la forme mature de la protéine Core (p19) [107, 185]. L'efficacité du clivage par la SPP est inversement proportionnelle à la longueur du segment C-terminal, signifiant donc que le clivage par la SP précède le clivage par la SPP [171]. Le domaine D3 d'ancrage n'est pas le seul déterminant au clivage par la SPP [151], car la substitution de certains acides aminés (L139, V140 et L144) dans le domaine D2 ou même l'absence du domaine D2 inhibe totalement le clivage [171, 200]. La position exacte de la réaction de clivage par la SPP est restée longtemps incertaine. Les études menées sur ce second clivage ont pour la plupart été réalisées dans divers systèmes d'expression et modèles, permettant d'identifier plusieurs sites regroupés entre les acides aminés 173 et 181 [107, 162, 198, 218, 270]. Les travaux d'Okamoto et coll. en 2008, et plus récemment Kopp et coll. en système cellulaire humain ont démontré que le clivage aurait lieu entre les acides aminés 177 et 178 menant ainsi à une forme mature de 177 acides aminés [139, 199].

Le processus de maturation de la Core s'effectue au niveau du RE où la Core, sous forme immature et mature, s'accumule de manière transitoire. Une fois clivée, la Core semble pouvoir se diriger vers divers compartiments cellulaires. On retrouve parmi ceux-ci, les gouttelettes lipidiques [14, 103], le noyau [43, 130, 163, 231] et la mitochondrie [222], où la Core pourra alors jouer un de ces nombreux rôles.

Les gouttelettes lipidiques sont un lieu de stockage de lipides intracellulaires. Elles sont constituées principalement de triglycérides et d'esters de cholestérols [83], entourées d'une monocouche de phospholipides à laquelle sont associées diverses lipo-protéines. La Core se dirige et s'accumule au niveau des gouttelettes lipidiques dans un étroit complexe avec le RE, et ceci a été observé dans diverses lignées cellulaires hépatiques et non hépatiques, ainsi que dans des biopsies de patients [83]. La région impliquée dans l'interaction aux gouttelettes lipidiques (ainsi que le RE et la mitochondrie) réside au niveau du domaine D2 de la Core grâce à la présence de deux hélices α amphipathiques. Boulant et coll. ont même mis en évidence que le domaine D2 seul pouvait se rendre aux

gouttelettes lipidiques [29]. Dans le cas d'une infection au VHC, les gouttelettes lipidiques en lien étroit avec le RE vont être le siège de la réplication virale, et de la production de virus infectieux.

La protéine Core, et en particulier des formes tronquées (qui pourrait être aussi la protéine ARFP/F/Core+1) ont aussi été détectées dans le noyau [43, 45, 111, 130, 162, 231, 240, 266]. Cette localisation serait due à la présence de 3 signaux de localisation nucléaire (région riche en acides aminés basiques et proline) dans la région N-terminale de la Core [43]. Le rôle de la Core, et/ou de ses formes tronquées, ou même de la protéine ARFP/F/Core+1 au noyau serait potentiellement lié à la régulation de gènes de l'hôte.

Enfin, la Core et une forme de la protéine ARFP (26-161) ont été également observées au niveau des membranes de la mitochondrie et proche du RE [126, 140, 208, 222, 258]. Le rôle de la Core, au niveau des membranes de la mitochondrie, serait d'être impliquée dans la régulation du stress oxydatif [136, 138, 140].

L'importance de cette relocalisation au noyau et à la mitochondrie reste encore aujourd'hui peu comprise, mais certaines études sur le rôle de la Core semblent impliquer celle-ci dans le développement de symptômes et complications du VHC (stéatose, fibrose, hépatocarcinome).

1.10.3 Structure

1.10.3.1 Structure de la Core du VHC

La structure tridimensionnelle de la Core du VHC était au début de mon travail de recherche non résolue. Diverses équipes de par le monde se sont intéressées d'un point de vue structural à cette protéine particulièrement importante du virus. Diverses formes tronquées ont été étudiées en absence et en présence d'ARN (sous forme encapsidée ou

non), et en absence et en présence de détergents lipomimétiques, menant à des résultats parfois divergents [30, 41, 66, 67, 146, 147, 214-216]. Aussi, deux courts segments de la Core ont été étudiés par résonance magnétique nucléaire (RMN) et cristallographie des rayons X (RX), en présence de 2,2,2-trifluoroéthanol (TFE), ou lié à un anticorps ([149] (PDB ID : 1CWX) et [187] (PDB IDs : 1NLB et 1N64)).

Au sein, de notre laboratoire, nous nous sommes intéressés à la plus petite forme tronquée de la Core dite C82, montrant la capacité de former des nucléocapsides à la fois *in vivo* et *in vitro* (capacité que l'on utilisera au chapitre 7 de cette thèse et à l'annexe 4 [82], pour la recherche d'inhibiteurs de la formation de la capsid). Les résultats de nos travaux sur l'étude structurale de la Core sont présentés dans les chapitres 3, 4 et 5 de cette thèse.

L'étude structurale de la protéine Core faisant partie intégrante de ce projet de thèse, les aspects structuraux de ces différents travaux seront présentés dans la section « résultats » et ils seront comparés et discutés en conclusion.

1.10.3.2 Structure des protéines Core d'autres virus de la famille des *flaviviridae*.

Si la structure de la Core de VHC n'est pas connue, l'étude structurale des protéines Core de virus apparentés tels que le virus de la « Dengue » (DV) et du « West Nile » (WNV), peuvent peut-être apporter un éclairage significatif sur la structure potentiellement adoptée de la Core du VHC. Les structures de ces deux protéines ont été obtenues respectivement par résonance magnétique nucléaire et cristallographie par diffraction des rayons X [118, 172]. Elles arborent toutes les deux un repliement en hélice α et se présentent sous la forme d'un dimère. Celui-ci est formé de deux monomères orientés de manière antiparallèle. Chacun des monomères est composé de 4 hélices α nommées de $\alpha 1$ à $\alpha 4$. Les hélices $\alpha 2$ et $\alpha 4$ permettent la formation du dimère (Figure 1.12). La comparaison de ces deux structures ne fait état que de très peu de différences, seul l'orientation de l'hélice $\alpha 1$ semble varier. Une asymétrie des charges est observée avec une face plus hydrophobe (au dos de l'hélice $\alpha 4$) qui interagirait avec l'enveloppe, et une face chargée basique qui permettrait l'interaction avec l'acide nucléique. On notera que dans les deux

cas des sections flexibles non-structurées en N- et/ou C-terminal ont été clivées, pour faciliter la production et l'analyse structurale.

La Core du VHC possède elle aussi un certain degré d'asymétrie au niveau de sa séquence primaire. La région N-terminale (ou domaine D1) possède un très grand nombre de résidus chargés positivement (28 acides aminés dans les 120 premiers), et la région dite centrale (ou domaine D2) possède un caractère hydrophobe fort. Toutefois, cette simple comparaison ne permet pas de faire une analogie structurale évidente. Les protéines Core des *flavivirus*, tels que les virus de la dengue et du nil, sont des protéines d'une centaine d'acides aminés comparativement aux 177 de la Core mature du VHC. Les protéines Core des *flavivirus* et des *pestivirus* ne possèderaient pas le domaine D2. Ceci signifie donc que l'organisation structurale des protéines Core de ces virus est différente de celle de la Core du VHC. Par contre, la Core du virus GB (appartenant peut-être à la famille des *flaviviridae*) présenterait une organisation structurale similaire à celle du VHC [30, 104, 114], mais sa structure n'est pas connue.

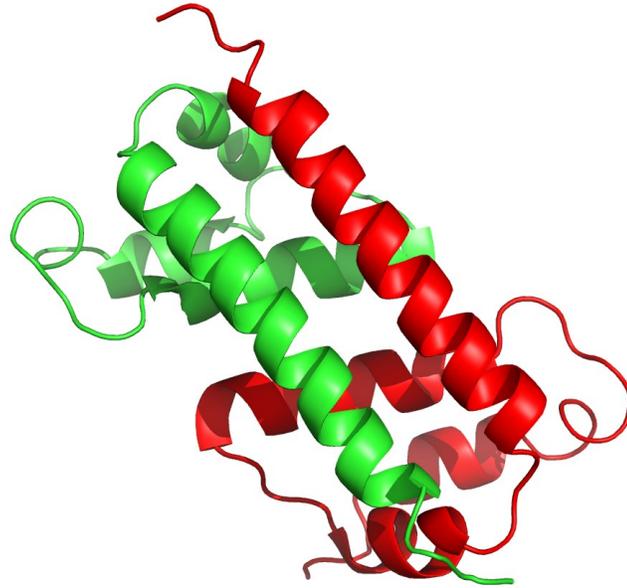


Figure 1.12 : Structure de la protéine Core du virus de la dengue, ici représentée sous la forme d'un dimère. (PDB ID : 1R6R).

1.10.4 Fonctions

1.10.4.1 Formation de la capsid virale

Les travaux sur la compréhension de l'assemblage viral du VHC ont longtemps été ralentis par le très bas niveau de particules virales chez les individus infectés, de l'incapacité de propager efficacement le virus en cellules cultivées, et du manque d'un modèle animal. Le premier système mis au point pour l'étude de l'assemblage viral a été décrit par Baumert et coll. en 1998 [19]. Ils ont développé un système de production de pseudo-particules virales en cellules d'insectes utilisant un *Baculovirus* recombinant codant

pour les protéines structurales du VHC (C, E1, E2). Les pseudo-particules purifiées possèdent des tailles de 40 à 60 nm, comparables à celles décrites dans la littérature (voir Section 1.8 (page 14)). D'autres systèmes, utilisant des virus recombinants (Vesicular Stomatitis Virus (VSV) ou Semliki Forest Virus (SFV)), ont été développés, démontrant la capacité de produire des pseudo-particules ou des assemblages abortifs dans des cellules de mammifères [23, 71]. Ces travaux ont permis de confirmer que l'assemblage viral et le bourgeonnement s'effectuaient au niveau du RE. D'autres systèmes eucaryotes et procaryotes ont aussi été développés (*Pichia pastoris*, *E.coli*) montrant la capacité de former des pseudo-particules enveloppées et non enveloppées [1, 74, 167]. La production de pseudo-particules a, dans le cas de la levure *Pichia Pastori*, été observée au niveau du noyau, et la même équipe aurait également observé, par microscopie électronique, des capsides au niveau du noyau dans des cellules de patients infectés au VHC [1, 73]. Cette observation n'a jamais été confirmée dans d'autres modèles.

Pour contourner les problèmes liés à l'étude de l'assemblage et du bourgeonnement du VHC énoncés plus haut, des systèmes *in vitro* [145] et acellulaires [133] et de production de pseudo-particules ou de pseudo-capsides ont été développés. Le principe de l'auto-assemblage *in vitro* mis en place par Kunkel et coll. est relativement simple et inspiré des travaux de Wengler et coll. sur l'auto assemblage du virus Sinbis [145, 254]. Pour cela, il suffit de produire la protéine d'intérêt en bactérie, de la purifier et d'ajouter des molécules d'ARN pour déclencher le phénomène d'encapsidation [145]. Les capsides obtenues sont comparables à celles non enveloppées isolées de patients infectés. La liaison à l'ARN et la formation de la capside protègent de manière efficace la dégradation de la Core par des protéases [146] suggérant que cette interaction apporte des changements importants à la protéine Core, qui dans sa forme mature non assemblée, est dégradée par des protéases. Ces plateformes d'assemblage *in vitro* ou acellulaire ont servi à déterminer la plus petite portion de la Core permettant l'assemblage viral [145, 178]. Les travaux de Nathalie Majeau, au sein du laboratoire du Dr. Denis Leclerc, ont permis de mettre en évidence que des formes tronquées de la Core allant des acides aminés 1 à 75 jusqu'à 1 à 170 engendrent la formation de capside. Dans le cas d'une production en levure *Pichia pastoris*, les 82 premiers acides aminés (C82) auxquels est ajoutée la séquence

d'ancrage au RE (acides aminés 170 à 191), forment également des particules virales non enveloppées. La protéine C82 représente donc la plus petite portion de la Core ayant la capacité de former des capsides *in vitro* et *in vivo*.

Ce n'est que 16 ans après la découverte du virus VHC qu'un système de culture cellulaire permettant la propagation virale a été décrit [160, 251, 278], ouvrant ainsi une grande porte pour l'élucidation de divers processus du cycle du VHC et notamment ceux d'assemblage, de bourgeonnement et de relargage du virus à l'extérieur de la cellule. L'assemblage viral est dépendant du bon adressage des protéines Core matures et NS5A au niveau des gouttelettes lipidiques, et de la présence de protéines de l'hôte souvent impliquées dans la biosynthèse des lipides supportant ainsi l'idée que le VHC extracellulaire est associé à des vésicules lipidiques (pour revue : [17]). Les progrès récents sur l'assemblage laissent toutefois de nombreuses zones grises (initiation de l'assemblage, lieu de l'assemblage, relargage du virion). Un système plus efficace de production de virus serait certainement nécessaire pour compléter nos connaissances.

1.10.4.2 Interaction protéine Core – protéine Core

La protéine Core est une protéine capable de s'assembler pour former une capsid. Plusieurs études de cartographie des sites d'interaction de la protéine Core ont été publiées, menant à des résultats contradictoires.

En l'espace de trois années, les régions 36-91 [183], 82-102 (région riche en tryptophane) [197] et 119-162 [267] ont été montrées comme impliquées dans les interactions homotypiques de la Core par des expériences de double-hybride. Une étude plus récente décrit la région 72-91 comme étant le site d'interaction homotypique [196]. Par ailleurs, Nolandt et coll. ont montré l'existence d'une interaction hétérotopique entre le C-terminal et le N-terminal de la Core [197].

Divers résultats en lien avec la multimérisation de la Core ont été publiés, mais là encore, une grande hétérogénéité dans les résultats est observée. Certains affirment que la

région N-terminale arborant des tailles allant de 102 à 124 résidus peut interagir avec elle-même pour former un dimère [3, 141, 147, 202], d'autres tendent à dire que cette interaction n'est possible qu'après l'ajout d'un ARN [131, 183]. Dans le cas de versions longues de la protéine Core (169 ou 179 acides aminés de long), la région riche en tryptophane (82-102) est parfois impliquée dans l'interaction, ceci a été démontré par des travaux sur l'accessibilité au solvant [30], alors que d'autres obtiennent le contraire [147]. Par contre, tous s'accordent pour dire que dans des versions courtes de la protéine, cette région ne semble pas impliquée dans l'auto-multimérisation. L'implication des cystéines dans la formation d'un dimère est aussi controversée. Kunkel et coll. montrent que la dimérisation de la forme C1-124 est due à un pont disulfure créé au niveau de la cystéine 91, or cette cystéine ne semble pas conservée au sein des diverses souches de VHC [147]. En revanche, Yan et coll. (1998) ainsi que Boulant et coll. (2005) montrent que l'interaction de la Core n'implique pas la formation d'un pont disulfure au niveau de la cystéine conservée C128 (pas de cystéine en position 91 dans leur construction) [30, 267]. Or récemment, la cystéine 128 a été impliquée dans la formation d'un dimère de la Core. En absence de cette cystéine, la propagation virale est rendue nulle, néanmoins, la Core conserve sa capacité à lier l'ARN et interagit toujours avec les gouttelettes lipidiques [148].

L'ensemble de ces résultats ne nous permet pas tirer des conclusions claires sur les régions d'interaction de la Core avec elle-même. Néanmoins, la présence de la région C-terminale semble favoriser la multimérisation de la Core, mais n'est pas indispensable. Par ailleurs, la présence d'un acide nucléique (ARN) serait un facteur important pour l'association de la protéine avec elle-même, entraînant potentiellement un remaniement structural permettant la multimérisation de la Core.

1.10.4.3 Interaction avec l'ARN

La protéine Core se divise en trois domaines. Le domaine N-terminal (D1) est extrêmement basique et hydrophile (28 charges positives) dans lequel on distingue trois sous-domaines chargés (aa 6-23, 39-74 et 101-121). Les résidus basiques sont fréquemment impliqués dans des interactions avec des acides nucléiques, car la charge positive qu'ils

exposé peut interagir avec la charge négative du phosphate présent sur la chaîne principale d'un acide nucléique. En effet, la protéine Core possède la capacité de lier divers ARNs : ARN ribosomal [218], ARN de transferts [82, 145, 178], et l'ARN viral du VHC [55, 75, 82, 145, 233, 242, 276] et son site d'interaction a été délimité à la région N-terminale de la protéine Core (domaine D1) [145, 178, 218]. Néanmoins, une controverse persiste sur la spécificité de liaison. Alors que certains estiment que l'interaction n'est pas ARN spécifique [218], d'autres affirment le contraire et montrent que la Core possède une affinité marquée pour la région 5'UTR du génome du VHC [75, 233, 276] et plus précisément au niveau de la boucle IIIId [242], ainsi que la région 3'UTR [55, 113, 273].

Ces diverses interactions sont impliquées dans diverses fonctions de la protéine Core. En lien avec sa fonction d'encapsidation, Kunkel et coll. ont démontré que la Core pouvait former des capsides en présence de la région 5'UTR du génome du VHC ou d'un ARN de transfert (Yeast ARN_t^{Phe}) [145]. La spécificité serait due au repliement « tige-boucle » de l'ARN, car en absence de repliement, ces mêmes ARNs n'entraînent plus la formation de capsides [145]. Par ailleurs, les recherches se sont portées sur les régions spécifiques de la Core permettant la liaison à l'ARN et d'entraîner la formation de capsides. Santolini et coll. ont démontré que les 75 premiers acides aminés de la Core permettaient la liaison à l'ARN, éliminant donc le troisième sous-domaine chargé d'un rôle important dans ces fonctions [133, 145, 178, 218]. Des deux sous-domaines chargés restant (aa 6-23 et aa 39-74), le deuxième domaine semble essentiel pour l'interaction avec l'ARN et la formation de capsides *in vitro* [82, 132]. Par ailleurs, des résultats similaires en culture cellulaire ont été observés récemment, confirmant l'importance de la région 39-74 de la Core (en particulier les quatre résidus chargés suivants, R50, K51, R59 et R62) pour générer un virus infectieux [5]. Néanmoins, il apparaît que malgré des substitutions de ces résidus, l'interaction de la Core et de l'ARN n'est pas abolie, la multimérisation de la Core n'est pas affectée, ni son interaction aux membranes, suggérant donc qu'aucun motif particulier sur la Core n'est absolument nécessaire pour la formation de capside. La seule contrainte est la présence d'un ARN structuré arborant une charge globale négative.

L'interaction de la Core avec son propre ARN n'a pas que pour seule fonction d'engendrer la formation de la capsid virale. En effet, la protéine Core possède la capacité de réguler sa propre traduction grâce à son interaction avec la région 5'UTR (IRES) du génome viral. En effet, Shimoike et coll. ont suggéré que la protéine Core pouvait contrôler la traduction de l'ARN viral [233]. Ces résultats ont, par la suite, été confirmés par plusieurs équipes [26, 153, 232, 276] apportant néanmoins un éclaircissement sur les régions impliquées dans l'interaction de la Core et de la région 5'UTR (IRES) du génome viral. En 2000, Tanaka et coll. rapportent que la Core possède une affinité importante pour la boucle IIIId de l'IRES [242]. De plus, Shimoike et coll. (2006), montrent que la région de l'IRES permettant le contrôle de la traduction est portée par la boucle IIIId [232]. En ce qui concerne la cartographie de la région de la Core impliquée dans cette fonction, les résultats sont encore une fois contradictoires. Zhang et coll. impliquent les résidus 33 à 44 de la Core [276], Li et coll. impliquent les 20 premiers résidus de la Core [153], tandis que Shimoike et coll. affirment qu'aucun des motifs chargés positivement (aa 5-13, 38-43, 58-71) n'est essentiel, et que, la seule présence de charge à l'un de ces trois motifs permet le contrôle de la traduction [232]. Encore une fois, il semblerait que la seule présence de l'extrémité N-terminale fortement chargée soit le déterminant fonctionnel de la protéine Core. Cette activité de contrôle de la traduction aurait un rôle charnière dans le cycle viral du VHC permettant le passage d'un mode de traduction à celui de réplication virale [276]. Ceci est appuyé par le fait que l'action de la Core est dépendante de sa concentration [26].

Enfin, la Core posséderait une activité chaperonne d'acide nucléique. Le domaine D1 de la Core favoriserait l'association de brins complémentaires d'ADN, et posséderait une activité d'échange de brin dans le cas de mésappariement [55]. De plus, la dimérisation de l'ARN viral au niveau de l'extrémité 3'-UTR a été démontrée. La boucle SL2 de la région dite « X-ARN » (cette région à l'extrême 3'-UTR est très conservée et indispensable au cycle viral) serait la cible de la Core pour engendrer un dimère d'ARN [55, 113, 225] via un réarrangement structural de l'X-ARN. Une fois de plus, le domaine N-terminal (D1) de la Core est le siège de cette nouvelle propriété. En accord avec l'idée que la charge globale de la protéine Core est le facteur déterminant de sa capacité à lier l'ARN pour former des capsides ou pour contrôler sa propre traduction, Ivanyi-Nagy et coll. démontrent

que des constructions peptidiques issues de la juxtaposition des domaines chargés de la Core (les régions hydrophobes ou neutres ont été tronquées) sont également capables d'engendrer la dimérisation de l'ARN viral [113]. Aussi étonnant que cela puisse être, la Core ou des fragments de celle-ci possède une grande variété de fonctions différentes. La dimérisation de l'ARN viral aurait potentiellement un rôle dans le contrôle de la réplication et dans la formation de la capsid. Il est à noter que ce rôle de chaperonne est également porté par d'autres protéines Core de virus apparentés [112, 114].

1.10.4.4 Modifications post-traductionnelles

La protéine Core est la cible de plusieurs modifications post-traductionnelles qui peuvent avoir un rôle potentiellement important sur ses fonctions, sa localisation et sa dégradation.

La Core du VHC est la cible des protéines kinases PKA et PKC. Les acides aminés ciblés sont les serines en position 53, 99 et 116 [168, 230]. La phosphorylation de la Core semble être essentielle pour la suppression de l'expression et la réplication du virus de l'hépatite B en cellule [221, 230]. Par ailleurs, la phosphorylation de la sérine 116 aurait un rôle répresseur sur la transcription du gène p21^{waf1} [121] et la localisation favorisée de celle-ci serait au noyau [168]. Aoki et coll. suggèrent que l'interaction de la Core et la protéine 14-3-3 serait dépendante de la phosphorylation de la sérine 53 [11].

Deux études se sont attardées sur les effets de la phosphorylation dans les événements d'assemblage, l'un en système acellulaire et l'autre en levures [132, 176]. Klein et coll. concluent que la phosphorylation n'a aucun effet sur l'assemblage viral après avoir étudié les mutants SxA (x étant les résidus 53, 99 ou 116) (mutants inhibant la phosphorylation) et le mutant S53D (mutant mimant la phosphorylation) [132]. Majeau et coll. confirment ces résultats et ajoutent que le mutant S116D n'a pas d'effet sur l'assemblage. Toutefois, la phosphorylation de la sérine 116 aurait un effet sur la stabilité et la dégradation de la protéine [176].

Récemment, il a été démontré que la protéine Core du VHC pouvait être palmitoylée et le principal résidu visé serait la cystéine 172. Cette modification post-traductionnelle n'aurait pas d'influence sur la maturation de la Core, ni sur sa localisation aux gouttelettes lipidiques. Toutefois, la palmitoylation de la Core jouerait un rôle important dans le rapprochement des gouttelettes lipidiques et du réticulum endoplasmique, ce qui a pour conséquence de diminuer l'assemblage viral et l'infectivité du virus [177].

1.10.5 Autres rôles

1.10.5.1 Interactions de la Core avec d'autres protéines du virus

La protéine Core est essentielle pour l'interaction avec l'ARN viral, l'assemblage via ces interactions homotypiques. Néanmoins, elle possède des interactions avec d'autres membres protéiques du VHC.

La protéine E1 est ancrée au niveau du RE et interagit avec E2 [63, 272], et se retrouve à la surface du virion. Une interaction entre la protéine E1 et la Core a été mise en évidence, impliquant le domaine D2 de la Core [164, 170]. Cette interaction serait favorisée par la multimérisation de la Core et de son interaction avec l'ARN [196]. Une certaine controverse persiste sur le domaine d'interaction de E1 [164, 170, 196]. Les travaux de Nakai et coll. semblent montrer que E1 peut alterner entre deux conformations, et que l'une d'entre elles impliquant un domaine cytosolique favoriserait l'interaction avec la protéine Core [196]. Le rôle de cette interaction serait le recrutement de la capsid (formée ou en formation) au niveau du RE pour favoriser le bourgeonnement.

Les protéines Core et NS5A démontrent une co-localisation au niveau des gouttelettes lipidiques [229]. La protéine NS5A serait essentielle à l'assemblage et la production de particule virale infectieuse [12, 182, 243]. En effet, en absence d'interaction entre ces deux protéines, aucun relargage de virus ne serait observé. L'interaction de la Core et de NS5A a été identifiée au niveau du domaine III de NS5A et, en particulier, dans

une région riche en résidus sérine, qui serait impliquée dans la phosphorylation. La mutation des résidus S2428 et S2430 en alanine conduit à une production virale nulle; par contre, les mutations en acide glutamique (qui miment la présence d'une phosphosérine) restaurent l'interaction avec la Core et rétablissent la production de virus. Ces travaux suggèrent que la protéine NS5A joue un rôle majeur dans la régulation de l'assemblage viral.

Enfin deux études récentes ont rapporté une interaction de la Core avec les protéines NS5B et NS3. Dans le cas de NS5B, l'interaction se ferait via les 75 premiers acides aminés de la Core et les premiers 371 acides aminés de NS5B. L'interaction de la Core avec la protéine NS5B aurait pour effet de supprimer la réplication virale. Dans le cas de NS3, une co-localisation au niveau des gouttelettes lipidiques a également été observée [174, 189]. L'interaction impliquerait les 106 premiers acides aminés de la Core. Toutefois, la multimérisation de la Core (C106 ou C169) favoriserait son interaction avec le domaine hélicase de NS3 (aa 167-631) [193]. La démonstration de cette interaction impliquerait potentiellement la protéine NS3 dans l'assemblage viral.

1.10.5.2 Interactions avec des protéines de l'hôte

La protéine Core jouerait un rôle important dans la pathogénèse du virus VHC. À cet effet, la Core possède un grand nombre de rôles, d'activités et de partenaires protéiques au sein de la cellule. En plus de son rôle principal qu'est la formation de la capsid virale, la Core semble perturber, moduler le fonctionnement de la cellule. Elle serait impliquée dans diverses fonctions cellulaires, telles que l'apoptose, le contrôle du cycle cellulaire, l'accumulation de lipides intracellulaires (stéatose) ainsi que dans les processus de réponse immunitaire. Son implication dans ces processus a deux origines principales. 1- La protéine Core module la transcription, la traduction et l'expression de certains gènes et protéines. 2- Ses activités sont aussi dues à des interactions directes et indirectes avec des protéines en modifiant de manière positive ou négative leurs activités. (Pour revue : [109, 135, 137, 184, 192, 209]). Le tableau suivant représente quelques unes des interactions de la Core avec des protéines de l'hôte menant à un dysfonctionnement de la cellule.

Tableau 1.1 : Liste de protéines cellulaires qui lient la protéine Core.

Protéines en interaction avec la Core	rôle	références
TNF-R1	apoptose	Zhu et coll., 1998
hnRNP K	contrôle de la transcription transduction du signal	Hsieh et coll., 1998
Dead Box protein (DDX3/DBX, Cap-Rf)	hélicase à ARN	Mamiya et coll., 1999 Owsianka et coll., 1999 You et coll., 1999
Lymphotoxin β receptor	apoptose	Chen et coll., 1997 Matsumoto et coll., 1997
Fas	apoptose	Moorman et coll., 2003
Apolipoprotein AII	métabolisme des lipides	Sabile et coll., 1999
p53	contrôle de la transcription et régulation du cycle cellulaire	Ray et coll., 1997 Lu et coll., 1999
p73	contrôle de la transcription et régulation du cycle cellulaire	Alisi et coll., 2003
14-3-3	transduction du signal apoptose	Aoki et coll., 2000 Lee et coll., 2007
p21 waf1/cip1/sdi1	régulation du cycle cellulaire	Wang et coll., 2000 Yoshida et coll., 2001
gC1qR	immunité	Yao et coll., 2001 Yao et coll., 2007 Yao et coll., 2008
retinoid X receptor X Smad3	régulation de la transcription et prolifération cellulaire	Tsutsumi et coll., 2002
	contrôle de la transcription	Cheng et coll., 2004
IKappa B kinase β	transduction du signal	Joo et coll., 2005
protein kinase PKR	apoptose prolifération cellulaire	Yan et coll., 2007
TLR2	immunité	Hoffman et coll., 2009
Hsp 60	apoptose	Kang et coll., 2009
Tubulin	transport	Roovhand et coll., 2009
ACY3	?	Chen et coll., 2009
HCBP6 et 12	développement de cancer	Chen et coll., 2009
prohibitin	stress oxydatif	Tsutsumi et coll., 2009
Mcl-1	apoptose	Mohd-Ismail et coll., 2009
fibrinogen β	inflammation	Ait-Goughoulte et coll., 2010

1.11 Objectif général

Le sujet de doctorat, que l'on m'a confié, s'inscrit dans le cadre d'un projet de recherche plus large concernant la compréhension de la formation de la capsid virale par la protéine Core et la recherche d'un traitement et/ou d'un vaccin contre le virus de l'hépatite C [101]. Aussi le but principal de ce projet de recherche est la détermination de la structure tridimensionnelle de la protéine Core du virus de l'hépatite C, et plus particulièrement de sa forme tronquée C82, qui a été montrée comme étant la plus petite partie de la protéine Core du VHC à posséder la capacité d'engendrer une capsid virale *in vivo* et *in vitro*.

1.11.1 Objectifs spécifiques

Dans le but de répondre au mieux à l'objectif principal, nous avons subdivisé le projet en plusieurs objectifs intermédiaires dont la liste suit :

- 1- L'expression, le marquage et la purification de la protéine Core (C82 et autres formes tronquées si possible).
- 2- La détermination de la structure tridimensionnelle de la protéine Core (C82 et autres formes tronquées si possible) par résonance magnétique nucléaire (RMN).
- 3- L'étude la dynamique de la protéine Core (C82 et autres formes tronquées si possible).
- 4- La détermination des interactions protéine/protéine, protéine/acide nucléique ou protéine/inhibiteur d'assemblage.

Les objectifs ont été atteints et sont présentés dans la section « résultats » de cette thèse. Les objectifs 1, 2 et 3 sont abordés et décrits aux chapitres 3, 4 et 5. L'objectif 4, s'intéressant aux interactions de la Core avec des partenaires protéiques et peptidiques, est quant à lui présenté dans les chapitres 6 et 7. Les chapitres 3, 4 ont été publiés dans des

revues scientifiques. Le chapitre 7 est en cours de révision et les chapitres 5 et 6 sont des résultats préliminaires.

Chapitre 2

Les méthodes biophysiques utilisées

Durant les 20 dernières années, le nombre de structures tridimensionnelles de protéines déterminées à haute résolution a augmenté de manière exponentielle. Avec plus de 67 000 structures de protéines présentes dans la PDB en 2011 contre moins de 500 en 1990, le constat est clair, ces deux dernières décennies ont été riches grâce à la démocratisation de techniques biophysiques autrefois réservées à des chimistes et/ou des physiciens. Si nous prenons pour exemple la résonance magnétique nucléaire, trois prix Nobel ont été décernés dans ce laps de temps. R. Ernst a reçu, en 1991, le prix Nobel de chimie pour le développement de la RMN par transformée de Fourier et le développement d'expériences multidimensionnelles. Le prix Nobel de chimie 2002 a été décerné, en partie, à K. Wüthrich pour le développement de la spectroscopie de résonance magnétique nucléaire pour la détermination de structure tridimensionnelle de macromolécules biologiques en solution. Et enfin, le prix Nobel 2003 de physiologie et médecine a été attribué aux docteurs Paul C. Lanterbur et Sir Peter Mansfield, pour leurs découvertes concernant l'imagerie par résonance magnétique (<http://nobelprize.org>).

La résonance magnétique nucléaire n'est qu'une des nombreuses techniques biophysiques utilisées par les biochimistes pour étudier la structure des protéines. On peut citer la diffraction des rayons X, le dichroïsme circulaire, les spectroscopies infrarouge, raman et de fluorescence, la diffusion des rayons X ou des neutrons aux petits angles, la cryomicroscopie, la microscopie à force atomique, la calorimétrie, la modélisation par bio-informatique, les outils de prédiction de structure secondaire et tertiaire... L'ensemble de ces techniques a permis et permet d'étudier la structure de protéines et de complexes protéiques de taille croissante en lien avec leur rôle biologique (exemple : virus, ribosome etc.).

Parmi ces méthodes, deux ont été principalement utilisées durant cette thèse : le dichroïsme circulaire et la résonance magnétique nucléaire. Dans ce chapitre, les deux méthodes seront décrites en abordant succinctement la théorie et leurs diverses applications.

2.1 Le dichroïsme circulaire

2.1.1 Principe

Le dichroïsme circulaire repose sur l'utilisation d'une source lumineuse polarisée dirigée sur un échantillon. Une lumière polarisée plane peut être décrite comme étant la somme de deux composantes lumineuses circulaires polarisées d'égale amplitude, l'une étant dans le sens antihoraire (gauche, L) et l'autre étant dans le sens horaire (droite, R). Le dichroïsme circulaire se définit alors comme la différence d'absorption de ces deux composantes [21]. Pour mesurer un signal de dichroïsme circulaire, il faut étudier une substance qui doit être optiquement active, c'est-à-dire une substance qui absorbe différemment la lumière selon sa polarisation droite ou gauche, on parle alors de substances chirales ou de chromophores. Si après le passage à travers la substance étudiée, les composantes L et R ne sont pas absorbées ou si les deux composantes le sont de manière égale, alors aucun dichroïsme ne sera mesuré. Au contraire, si une seule des deux composantes est affectée ou si les deux composantes sont affectées de manière différente, alors un signal dichroïque sera mesuré. L'appareil de dichroïsme circulaire mesure et enregistre une variation d'absorbance entre les composantes gauche et droite de la lumière résultante, définie par l'équation suivante (éq.1.1) :

$$\Delta A = A_L - A_D \quad (\text{équation 1.1})$$

Fréquemment, le signal mesuré sera affiché en ellipticité (Θ , en millidegré). Il existe une relation simple reliant le ΔA et le θ définie par l'équation 1.2 :

$$\theta = 32\,980 \times \Delta A \quad (\text{équation 1.2})$$

Cette absorbance ou ellipticité est dépendante de la longueur d'onde (λ) de la lumière polarisée dirigée sur la substance étudiée. Ainsi en faisant varier la λ , nous obtenons un spectre de dichroïsme circulaire.

2.1.2 Les applications du dichroïsme circulaire de protéine.

Le dichroïsme circulaire est une méthode très fréquemment utilisée pour étudier la structure des protéines. En effet, une protéine possède plusieurs centres chiraux ou chromophores parmi lesquels on retrouve le lien peptidique dont la contribution est mesurable entre 180 et 260 nm (UV lointain), les chaînes latérales des résidus Tyr, Trp, Phe, His et Met et les ponts disulfures (Cys) qui eux ont une contribution entre 250 et 350 nm (UV proche). Ces deux fenêtres spectrales nous informeront à la fois sur les structures secondaire, tertiaire et parfois quaternaire adoptées par la protéine étudiée. Il faut toutefois noter dès à présent que cette technique nous fournit des informations globales contrairement à la résonance magnétique nucléaire et à la diffraction des rayons X. Néanmoins, le dichroïsme circulaire est utilisé à plusieurs fins pour par exemple :

- Estimer le contenu en structure secondaire d'une protéine (hélice α , brin β , etc.)
- Détecter des changements structuraux
- Étudier la stabilité des protéines
- Mesurer des interactions protéines-ligands

2.1.2.1 Préparation d'un échantillon

Afin d'obtenir les meilleurs résultats possibles, la condition *sine qua non* est de travailler avec un échantillon le plus pur possible (> 95 %). En effet, la présence de faible quantité de contaminants peut mener à l'obtention de résultats erronés.

De nos jours, les protéines étudiées proviennent quasiment toutes d'un système de surproduction dans un organisme procaryote tel que la bactérie *E. coli*. Ce système de surproduction permet d'obtenir généralement une grande quantité de protéines; de plus il est fréquent qu'une séquence poly His-tag clivable ou non soit ajoutée pour faciliter la purification de la protéine. En l'absence de séquence tag, un éventail large de colonnes chromatographiques est disponible, utilisant différentes caractéristiques chimique,

physique ou biologique de la protéine. On peut nommer quatre types de chromatographie typiquement utilisés :

- Chromatographie échangeuse ionique (anionique ou cationique)
- Chromatographie hydrophobe
- Chromatographie d'exclusion ou tamisage moléculaire
- Chromatographie d'affinité

Une fois purifiée, la protéine doit être dosée avec une grande précision et disposée dans un tampon adéquat pour une étude en dichroïsme circulaire. La majorité des tampons permettent d'enregistrer des données fiables, mais il faut toutefois les utiliser à des concentrations faibles pour éviter toute interférence. L'utilisation de certains solvants (dioxane, DMSO), agents chélateurs (EDTA, EGTA) ou réducteurs (DTT, β -mercaptoéthanol) doit être évitée.

Enfin, la quantité requise d'échantillon est typiquement de moins de 250 μ l à une concentration de 0,1 à 0,4 mg/ml lorsque l'on utilise une cuvette ayant pour trajet optique 1 mm ou de moins de 50 μ l à une concentration de 1 à 1,5 mg/ml pour un trajet optique de 0,1 mm, et ce pour des études en UV lointains. Dans le cas des études en UV proches, la concentration des échantillons est plus grande (de 1 à 10 mg/ml). Si cela n'est pas possible (par exemple à cause de l'agrégation), l'alternative est d'augmenter la longueur du trajet optique (de 0,1 cm à 1 cm).

2.1.2.2 Estimer le contenu en structures secondaires

Le dichroïsme circulaire est une méthode simple et rapide pour approximer le contenu en structure secondaire d'une protéine. Pour ce faire, il faut enregistrer une série de spectres en UV lointains (entre 180 et 250 nm, selon le type d'appareil et la qualité de l'échantillon), région spectrale dans laquelle les structures secondaires typiques des protéines vont donner lieu à un signal dichroïque (Figure 2.1). D'un simple coup d'œil, il

est facile d'évaluer qualitativement la structure globale d'une protéine. On distingue trois signaux simples caractéristiques des hélices α , des brins β et des régions non structurées.

- Les protéines repliées en hélice α vont posséder un spectre typique présentant deux minima à 208 et 222 nm, et un pic positif vers 190-195 nm.
- Les protéines repliées en brin β vont présenter un spectre moins intense avec un seul minimum que l'on retrouve généralement entre 210 et 225 nm, avec également un pic positif vers 190-200 nm.
- Les protéines non structurées, dénaturées, ont quant à elles un seul et intense minimum entre 195 et 200 nm. On peut parfois observer des variantes avec un pic positif proche de 215 nm, dans ce cas là, on pourra soupçonner la présence d'hélice 3_{10} ou de région en poly (Pro) II (PPII).

Il va de soi que pour une protéine possédant plusieurs types de structures secondaires, le signal enregistré sera alors une combinaison des signaux précédents. Il existe plusieurs algorithmes capables d'estimer le contenu en structures secondaires d'un échantillon protéique inconnu. Pour cela, ces algorithmes utilisent une banque de protéines dont les structures cristallographiques et les spectres de dichroïsme circulaire sont connus. Voici une liste non exhaustive d'algorithmes fréquemment utilisés : CONTIN, SVD, VARSLC, CDSSTR, SELCON, K2d, CDNN [10, 25, 99, 179, 205, 239]. La plupart de ces algorithmes ont été implémentés sur le serveur DICHROWEB [165, 255], avec lequel nous avons travaillé durant cette thèse. Il faut être conscient que les résultats obtenus sont des estimations et les méthodes utilisées fournissent de meilleurs résultats pour les structures secondaires hélicoïdales [256].

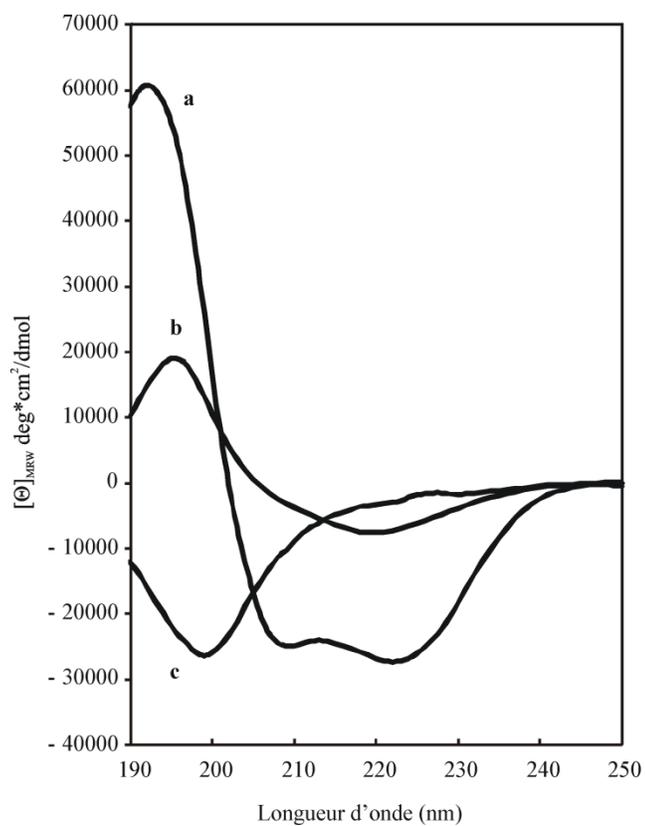


Figure 2.1 : Spectre de dichroïsme circulaire typique des trois types de repliement principaux. a) hélice α , b) feuillet β et c) désordonné.

2.1.2.3 Structure tertiaire et quaternaire

L'obtention d'informations sur les structures tertiaires et quaternaires découle d'observables dans la région des UV proches (250 à 350 nm). Les chromophores qui génèrent un signal à ces longueurs d'ondes sont principalement les acides aminés aromatiques (tyrosine (Tyr), tryptophane (Trp), phénylalanine (Phe)) et les ponts disulfures (entre deux cystéines (Cys)). Les phénylalanines possèdent des maximum d'absorption généralement entre 255 et 270 nm, les tyrosines entre 275 et 282 nm, les tryptophanes entre 288 et 295 nm et les ponts disulfures impliquant des cystéines vers 260 nm.

Le spectre obtenu ne donne que peu d'informations sur la structure tridimensionnelle de la protéine, mais il représente une « empreinte » qui peut servir de référence dans le cadre d'étude de la stabilité thermodynamique de la structure tertiaire et/ou quaternaire. Cette empreinte sera caractéristique d'un repliement et tout changement pourra alors être mesuré et sera interprété comme un changement de la structure tertiaire ou quaternaire de la protéine.

2.1.2.4 Autres applications

Les autres applications du dichroïsme circulaire sont nombreuses : on peut compter sur l'étude de la stabilité thermodynamique des protéines, sur le repliement et la dénaturation, et sur l'interaction des protéines avec divers ligands [90-93, 127]. Certains appareils permettent de faire du « stopped-flow » rendant la technique très intéressante dans le cas du repliement des protéines et des interactions avec de petits ligands.

2.2 La résonance magnétique nucléaire

La résonance magnétique nucléaire est certainement la technique biophysique pour l'étude structurale des protéines la plus versatile. En effet, elle permet à la fois d'étudier la structure tridimensionnelle, la dynamique, les interactions avec divers ligands, le repliement et la dénaturation d'une protéine. Ceci ressemble étrangement à ce que le dichroïsme circulaire peut faire, à ceci près que la RMN possède la capacité de nous donner des informations à une résolution atomique.

2.2.1 Principes de la RMN

Les atomes couramment étudiés en RMN des protéines sont l'hydrogène (^1H), le carbone (^{13}C) et l'azote (^{15}N). Les noyaux de ces atomes possèdent une charge électrique en rotation qui va engendrer un moment magnétique (μ) non nul dans la direction de l'axe de rotation. Ces noyaux possèdent un nombre quantique de spins (I) de $\frac{1}{2}$ et possèdent deux états distincts α ($I=+\frac{1}{2}$) et β ($I=-\frac{1}{2}$), faisant en sorte qu'ils se comportent comme des dipôles (comparables à des petits aimants) qui peuvent interagir avec un champ magnétique. En absence de champ magnétique externe, les spins n'adopteront aucune orientation préférentielle, donnant lieu ainsi à une aimantation macroscopique nulle. Lorsque qu'un champ magnétique intense (B_0) est appliqué, la dégénérescence des spins va être levée et les spins vont s'orienter parallèlement ou antiparallèlement au champ magnétique. Les états de spins α (de basse énergie) et β (de haute énergie) seront différenciés d'un point de vue énergétique par la relation suivante (éq. 2.1) :

$$\Delta E = \frac{\gamma h B_0}{2\pi} = h\nu \quad (\text{équation 2.1})$$

où γ est le rapport gyromagnétique (constante caractéristique de chaque noyau), h la constante de Planck et B_0 le champ magnétique appliqué.

Cette différence énergétique, définie par la relation de Boltzmann (éq. 2.2), crée une variation au sein de la population donnant lieu à une aimantation résultante macroscopique totale du moment magnétique (M) parallèle au champ magnétique.

$$\frac{N_{\beta}}{N_{\alpha}} = e^{-\frac{\Delta E}{kT}} \quad (\text{équation 2.2})$$

où N_{α} et N_{β} représentent les populations de spins, ΔE , la différence d'énergie des deux états, k la constante de Boltzmann, et T la température.

Chaque spin possède un mouvement de précession autour de l'axe du champ magnétique (B_0) à une fréquence (ν) caractéristique (éq. 2.3) reliée au rapport gyromagnétique du noyau observé et au champ magnétique externe ressenti influencé directement par son environnement proche.

$$\nu = \frac{\gamma B_0(1-\sigma)}{2\pi} \quad (\text{équation 2.3})$$

où ν est la fréquence de résonance d'un noyau; γ le rapport gyromagnétique associé au noyau; B_0 , le champ magnétique externe et σ la constante d'écran (caractérisant l'effet de l'environnement sur le champ magnétique externe B_0 ressenti par le noyau observé) [105]. Ainsi chaque noyau dans son environnement présentera une fréquence de résonance qui lui sera propre, c'est-à-dire un déplacement chimique (δ) particulier.

Le déplacement chimique (δ) est obtenu par la différence de fréquence du noyau observé (ν) et d'une référence (ν_{ref}) selon l'équation suivante (éq. 2.4).

$$\delta = 10^6 \frac{\nu - \nu_{ref}}{\nu_{ref}} \quad (\text{équation 2.4})$$

L'application d'un champ magnétique supplémentaire B_1 (ou impulsion) perpendiculaire à l'axe du champ va avoir pour effet de créer une composante transversale de l'aimantation (M_{xy}). Une fois l'impulsion finie, l'aimantation macroscopique M_{xy} va revenir à l'état d'équilibre (M_z) en oscillant autour de l'axe de B_0 . Le signal de retour à l'équilibre est appelé FID pour « free induction decay ». La décroissance du signal M_{xy} est régie par des processus de relaxation (temps de relaxation magnétique (T1) et nucléaire (T2)). La transformée de Fourier de cette FID convertit la dimension de temps en une dimension de fréquence résultant en un spectre d'amplitude en fonction de la fréquence de résonance (voir Figure 2.2 F).

2.2.2 L'analyse structurale de protéines

L'obtention de la structure tridimensionnelle d'une protéine est un processus long et fastidieux qui nécessite l'obtention d'un échantillon protéique de haute qualité, l'enregistrement de nombreux spectres (1D, 2D et 3D), l'attribution des résonances et la mesure de contraintes structurales.

2.2.2.1 Production d'un échantillon protéique marqué

Les protéines sont principalement constituées d'atomes d'hydrogène (^1H), de carbone (^{12}C), d'azote (^{14}N) et d'oxygène (^{16}O). Nous avons vu plus tôt que les atomes possédant un nombre quantique de spins de $\frac{1}{2}$ sont plus facilement étudiables en RMN. Pour travailler avec des protéines possédant des atomes de ^{15}N et ^{13}C ($I = \frac{1}{2}$), il faut alors les produire dans un milieu particulier dans lequel sont ajoutés des substrats de croissance bactérienne marqués à l'azote-15 ($^{15}\text{NH}_4\text{Cl}$) et au carbone-13 (glucose, $^{13}\text{C}_6\text{H}_{12}\text{O}_6$), car l'abondance naturelle de ces isotopes est trop basse ($\sim 0,4\%$ pour le ^{15}N et $\sim 1,1\%$ pour le ^{13}C).

Les étapes de production et de purification sont remises à l'épreuve lorsque nous devons marquer les protéines au ^{15}N et au ^{13}C . En effet, le milieu de culture utilisé

permettant le marquage isotopique est très différent du milieu traditionnellement utilisé (exemple milieu LB, ou 2XYT). Cette variation, qui peut sembler anodine, est en réalité un défi important. En effet, le profil d'expression protéique se voit affecté et modifié du simple fait que la bactérie se retrouve dans un milieu différent. Les milieux couramment utilisés sont soit des milieux minimums (composés principalement d'eau, de sels, et de glucose) soit des milieux riches marqués (utilisés plus sporadiquement car très onéreux). Comme nous l'avons vu dans le chapitre précédent sur le dichroïsme circulaire, les échantillons protéiques utilisés pour des analyses structurales doivent être particulièrement purs. Cette condition sine qua non s'avère également vraie dans le cadre des études par RMN. Il faut alors adapter toutes les étapes de la production jusqu'à la purification pour obtenir un échantillon marqué et pur (idéalement > 95 %) (Exemple de protocole mis au point, [181]).

2.2.2.2 La RMN mono- vs multidimensionnelle

Une expérience monodimensionnelle typique peut être schématisée comme suit : il y a, tout d'abord, une période de préparation durant laquelle le système rejoint l'équilibre, suivie d'une excitation (impulsion 90°), pour mener l'aimantation sur le plan perpendiculaire à l'aimantation, et enfin, l'acquisition du signal durant laquelle la FID est enregistrée. Une fois, la transformée de Fourier appliquée, nous obtenons un spectre de fréquence.

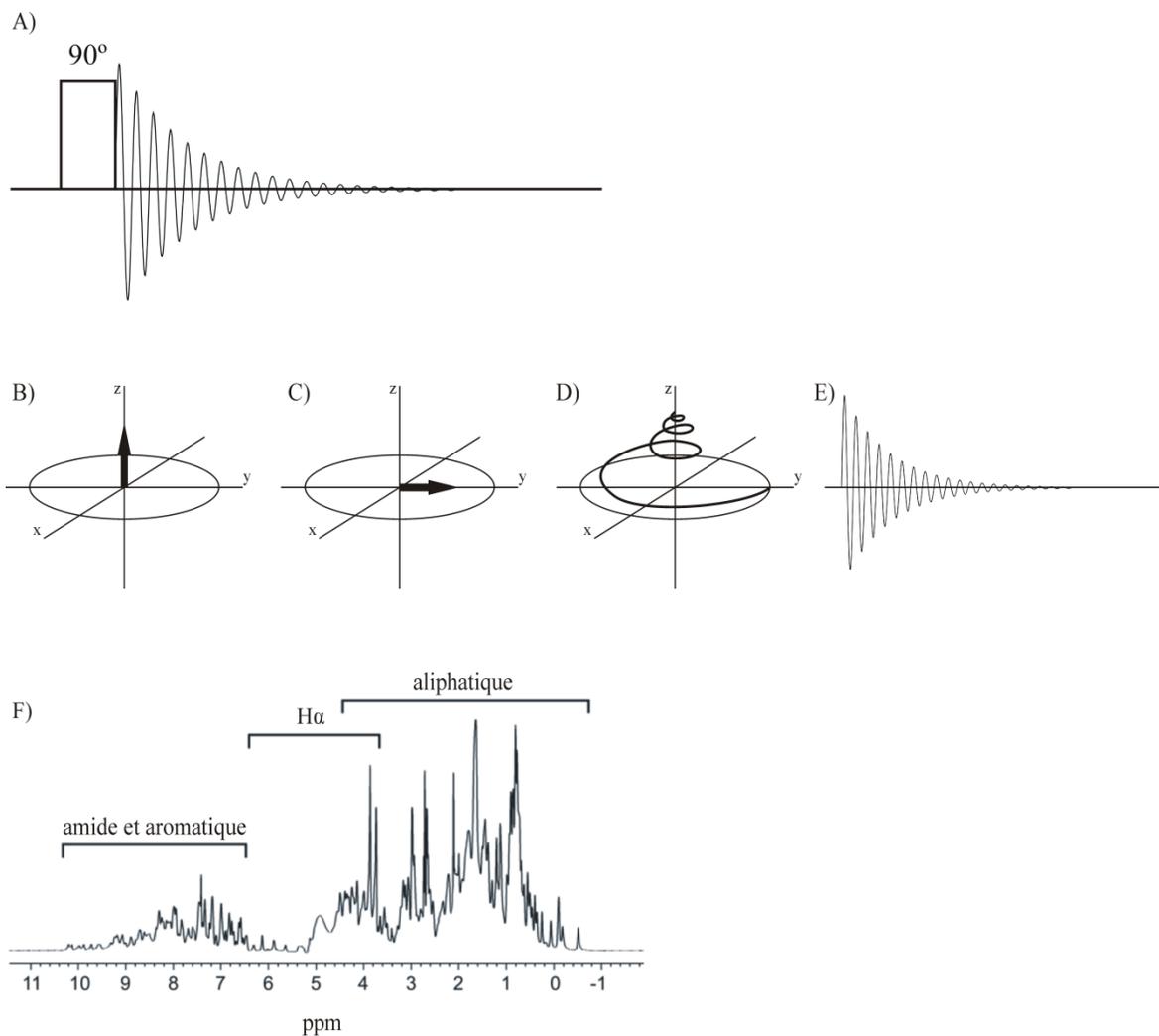


Figure 2.2 : RMN monodimensionnelle A) Séquence de pulse 1D ^1H typique, B) Aimantation à l'équilibre aligné dans le champ B_0 C) Aimantation après impulsion de 90° D et E) Retour à l'équilibre et enregistrement de la FID F) Spectre 1D après transformée de Fourier.

La RMN monodimensionnelle du proton ne permet pas d'obtenir une grande quantité d'information pour de larges molécules en raison de la superposition des signaux, mais elle reste néanmoins une expérience importante, car elle est le reflet de l'état d'une protéine dans un contexte particulier (tampon, pH, température). La dispersion des signaux dans certaines régions caractéristiques (amides, protons des chaînes latérales) peut tout de même nous donner des indices de l'état replié ou non de la protéine.

Une expérience multidimensionnelle (Figure 2.3) est conçue comme une succession d'expériences monodimensionnelles durant lesquelles la période d'excitation est séparée de la période d'acquisition (t_2) par une succession de périodes intermédiaires dites d'évolution (t_1) et de mélange (τ_m) durant lesquelles on laisse évoluer le système et transfère l'aimantation d'un atome vers un autre (au besoin).

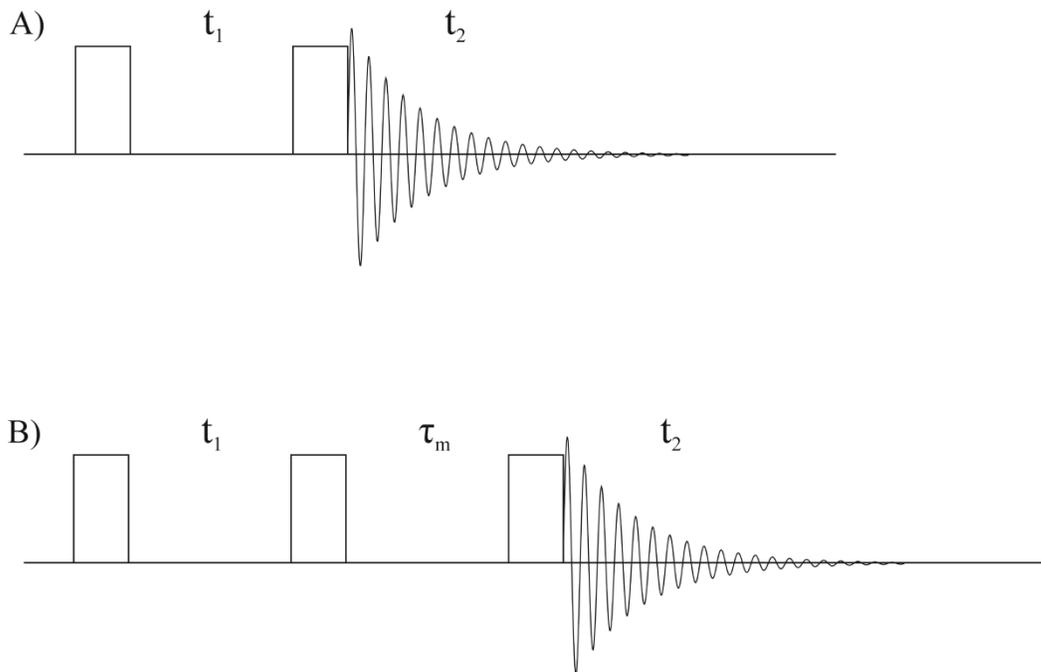


Figure 2.3 : A) représentation schématique d'une expérience 2D homonucléaire, B) représentation schématique d'une expérience 2D avec transfert d'aimantation (τ_m).

La répétition de cette séquence où l'on fera varier le t_1 d'un incrément Δt_1 , permet d'obtenir une série de spectres 1D dont les FIDs auront une amplitude modulée par le délai t_1 . Pour obtenir au final un spectre de résonance à x dimensions, il faudra alors appliquer autant de transformées de Fourier que de t_x .

Des techniques multidimensionnelles mononucléaires permettent de lever la dégénérescence observée en RMN 1D du ^1H rendant l'attribution des fréquences de résonances possible. Dans le cas des protéines ayant un poids moléculaire supérieur à 7-8 kDa, il est avantageux d'enregistrer des spectres hétéro-nucléaires multidimensionnels (2, 3, voire 4 dimensions) permettant une fois de plus de contourner cette dégénérescence en apportant une ou deux dimensions supplémentaires (c'est-à-dire le δ des atomes de ^{13}C et/ou ^{15}N) permettant ainsi d'attribuer non seulement les atomes de ^1H , mais aussi les ^{13}C et ^{15}N , et de mesurer des contraintes structurales.

Le spectre présenté en Figure 2.4 est un spectre 2D ^1H - ^{15}N HSQC qui corrèle les protons liés à un atome d'azote. Ce spectre représente en quelque sorte l'empreinte digitale d'une protéine, car chaque pic, corrélation ou résonance représente le lien amide de la chaîne principale de chacun des acides aminés de la protéine (sauf proline pas de lien N-H) et certains liens amides de chaînes latérales (glutamine, asparagine).

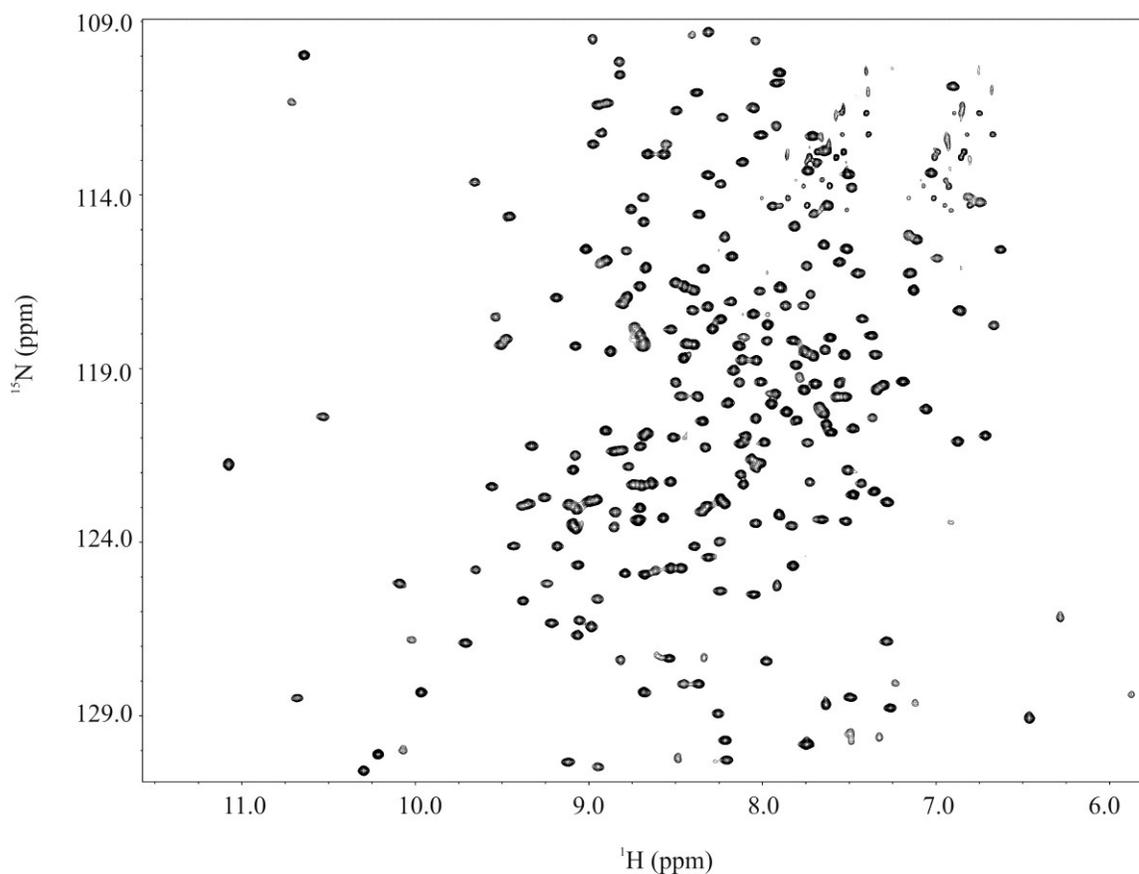


Figure 2.4 : Spectre 2D ^1H - ^{15}N HSQC de la protéine TEM-1 (Figure obtenue de P.-Y. Savard, [219]).

2.2.2.3 Stratégie d'attribution

L'attribution consiste à assigner chacune des résonances observées à un noyau particulier dans la molécule étudiée. Pour les petites molécules, l'attribution des résonances peut être envisagée grâce à des expériences mononucléaires multidimensionnelles tel que proposé par le Dr. K Wüthrich [263]. Dans le cas de plus grosses molécules (comme les protéines), nous prenons avantage de la RMN hétéronucléaire multidimensionnelle pour attribuer les résonances des acides aminés (protons et hétéro-noyaux) de la chaîne principale. La stratégie employée est l'attribution séquentielle grâce à l'enregistrement de spectres tridimensionnels qui permettent de relier des atomes entre eux selon leur position

au sein du polypeptide [263]. Certains spectres relieront des atomes au sein du même acide aminé, tandis que d'autres relieront des atomes de deux acides aminés voisins dans la séquence primaire de la protéine [263]. Ainsi un spectre HNCO permet de mettre en évidence les corrélations du lien NH de l'acide aminé *i* avec le CO de l'acide aminé *i-1*. Une série de spectres dits HNCO, HNCA, HNCACB, CBCACONH permettent alors de glisser sur la séquence et d'attribuer chacune des résonances des noyaux mentionnés (Figure 2.5). Les chaînes latérales sont attribuées via des expériences TOCSY (ce spectre corrèle des atomes au sein du même système de spins ou acides aminé), et NOESY (ce spectre corrèle des atomes distants dans l'espace ($<5 \text{ \AA}$)). Il existe un outil informatique utilisé au laboratoire permettant de faciliter l'étape d'attribution du squelette, il s'agit du programme Smartnotebook [237].

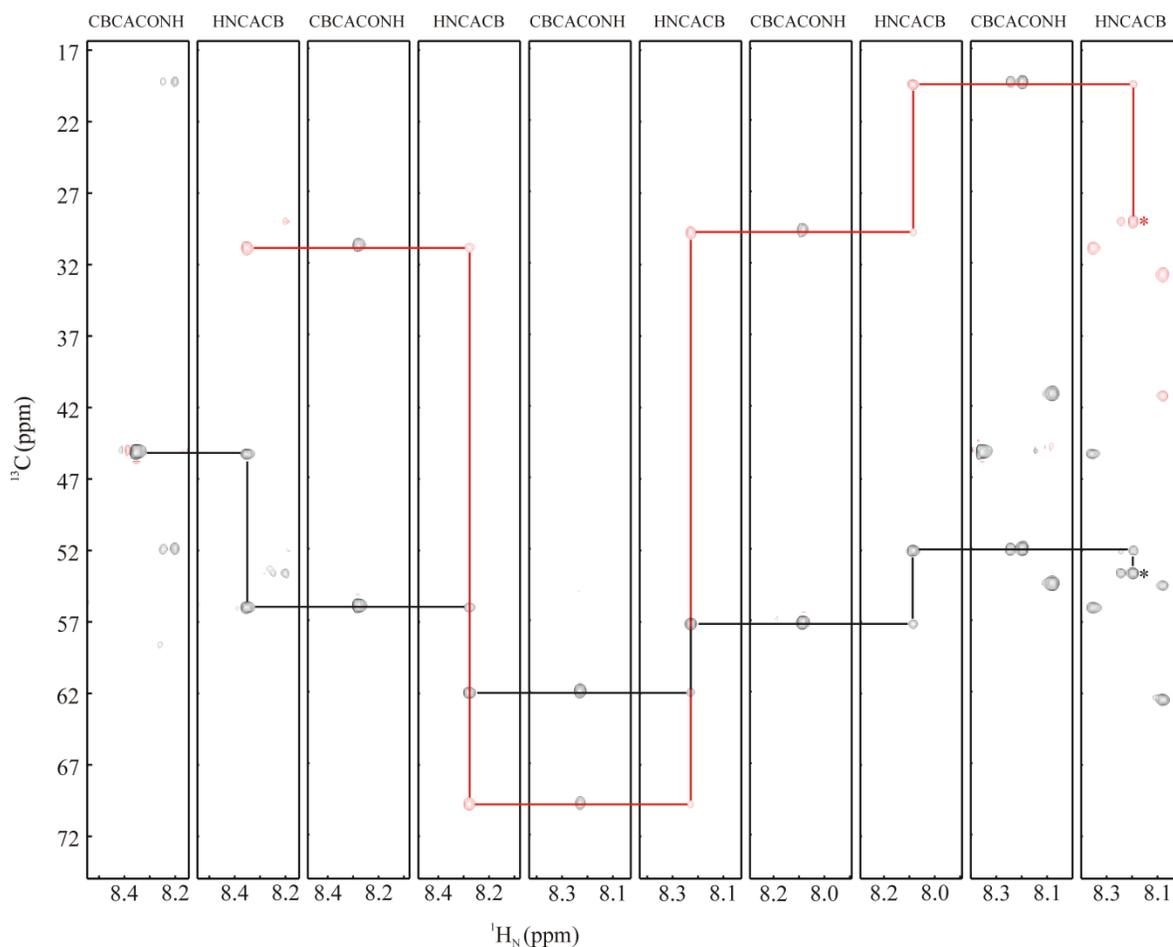


Figure 2.5 : Stratégie d'attribution de la chaîne principale. Les astérisques représentent les $C\alpha$ (en noir) et $C\beta$ (en rouge) de l'acide aminé Q78. Ainsi en suivant la méthode d'attribution, nous pouvons attribuer les résonances des $C\alpha$ et $C\beta$ des acides aminés voisins. Cet exemple utilisant les spectres 3D-HNCACB et 3D-CBCACONH montre le chemin d'attribution de l'acide aminé Q78 (à droite) jusqu'à l'acide aminé R74 (à gauche) qui précède l'acide aminé G73 (observé dans le premier panneau à gauche) qui ne possède pas de $C\beta$ (expliquant l'interruption de la chaîne).

2.2.2.4 La détermination de la structure 3D d'une protéine

Bien que cette partie de la RMN n'ait pas été abordée au cours de cette thèse, on ne peut vouloir décrire au mieux l'utilisation de la RMN sans au moins évoquer le cheminement nécessaire à l'obtention de la structure 3D d'une protéine. Une fois l'attribution complétée ($\geq 90\%$, chaîne principale et chaînes latérales), il est nécessaire d'obtenir des contraintes structurales.

Ces contraintes structurales sont :

- Les NOEs (« Nuclear Overhauser Effect ») : ceux-ci permettent d'obtenir des contraintes de distances entre des protons à une distance maximale de 5-6 Å dans l'espace.
- Les constantes de couplage (J) vont nous permettre de mesurer les angles dièdres (φ, χ, ψ).
- Les déplacements chimiques nous donneront des indications structurales :
 - de repliement ou non et de structure secondaire via l'index de déplacement chimique (CSI, [259, 260])
 - et d'angles dièdres (via TALOS(+), [54, 226])
- Les orientations de vecteurs N-H, C-H, C-C, C-N via l'enregistrement de spectres mesurant les couplages dipolaires résiduels.

Ces diverses contraintes sont obtenues grâce à l'acquisition de différentes expériences RMN (NOESY, HNHA...). Pour calculer une famille de structures 3D de protéines, il est nécessaire d'avoir en main le plus grand nombre possible de contraintes structurales. Pour cela, une dynamique moléculaire de type « recuit simulé » est réalisée. Celle-ci consiste dans un premier temps à mener la simulation à haute température à partir d'une structure initiale aléatoire étendue à laquelle sont ajoutées les contraintes structurales obtenues. Par la suite, la température est lentement diminuée afin de contraindre le système à occuper les états de plus basse énergie et d'obtenir une structure de bonne géométrie. Les outils informatiques de calcul de famille de structures de protéines peuvent être X-PLOR, CNS ou CYANA [38, 39, 115]. La dernière étape consiste à valider les modèles produits en

lien avec les contraintes mesurées. Les calculs seront répétés plusieurs fois dans le but d'affiner le modèle structural calculé.

2.2.2.5 Autres Applications

2.2.2.5.1 Étude de la dynamique de la chaîne principale

On ne peut imaginer les protéines comme une unité structurale totalement rigide. Les protéines sont intrinsèquement des molécules dynamiques (liaison, repliement, formation de lien hydrogène...) et la RMN est la technique idéale pour étudier d'un point de vue atomique ces mouvements. L'analyse de la dynamique est basée sur l'enregistrement de trois types d'expériences qui mesurent la relaxation du lien H-N. La relaxation du vecteur N-H sert à décrire la dynamique de la chaîne principale de la protéine. Ainsi, une région flexible comme une boucle sera plus dynamique qu'une région engagée dans la formation d'une hélice α ou d'un brin β .

Les expériences de dynamique mesurent les taux de relaxation longitudinaux (R_1) et transversaux (R_2) ainsi que les NOE hétéronucléaires associés à chaque vecteur ou lien N-H. Le taux de relaxation longitudinal (R_1) est une mesure du retour de la magnétisation le long de l'axe du champ externe après une perturbation, quant au taux de relaxation transverse (R_2), il est une mesure du retour de la magnétisation le long de l'axe perpendiculaire au champ externe. La mesure des taux se fait grâce à l'enregistrement d'une série de spectres 2D ^1H - ^{15}N HSQC, qui varieront les uns des autres par la durée du délai de relaxation. Les NOEs hétéronucléaires se définissent comme étant une mesure de la relaxation entre deux spins couplés dans l'espace. Ils sont obtenus en enregistrant deux spectres HSQC avec ou sans saturation de la magnétisation durant le délai de relaxation et les valeurs de NOEs seront obtenues en faisant le ratio des intensités de chaque corrélation du spectre HSQC avec ou sans saturation. Toutes les données obtenues sont analysées en utilisant le programme ModelFree [180, 201]. Celui-ci permet d'extraire, à partir des données brutes, certains paramètres physiques décrivant la dynamique d'un lien N-H tels

que : le S^2 (paramètre d'ordre (0 (complètement désordonné) $< S^2 < 1$ (ordonné et rigide)), le τ_m (temps de corrélation global), le τ_e (temps de corrélation effectif) et le R_{ex} (paramètre d'échange conformationnel) associé à chaque vecteur. Si nous prenons l'exemple de la β -lactamase TEM-1 étudiée au sein du laboratoire du Dr S. Gagné, l'analyse de la dynamique de la chaîne principale a permis de définir que TEM-1 était peu flexible sur une échelle de temps de la pico à la nanoseconde, et que la boucle- Ω ainsi que le site actif possédaient des mouvements mesurables sur une échelle de temps de la micro à la milliseconde compatible avec le rôle biologique de cette protéine [219].

2.2.2.5.2 Interaction protéine-ligand

Une des autres applications très intéressantes de la RMN, est la mesure d'interaction d'une protéine avec un ligand, que celui-ci soit une autre protéine, un peptide, un acide nucléique ou autres. Ces expériences sont basées une nouvelle fois sur un spectre HSQC de protéine, qui correspond à son empreinte digitale. Pour cela, il faut que seule la protéine étudiée soit marquée et que le spectre HSQC soit attribué. La liaison du ligand sur la protéine va engendrer une variation des déplacements chimiques des acides aminés impliqués dans la liaison (lien N-H dans le cas du ^{15}N -HSQC). Il est possible alors de cartographier la surface d'interaction sur un modèle structural (si la structure 3D est connue). En procédant par titrage, il est possible d'obtenir la constante de dissociation (K_D) du complexe et de définir l'affinité de la protéine pour son ligand, qui dans une situation d'échange rapide se fera par une mesure de la variation des déplacements chimiques en fonction du ratio protéine / ligand.

Au final, cette méthode permet de manière simple de mettre en évidence les acides aminés impliqués dans l'interaction et de mesurer une constante de dissociation entre les deux partenaires. D'autres méthodes de mesure d'interaction en RMN sont envisageables et peuvent être consultées dans la revue du Dr. Fielding [78].

Chapitre 3

Structure and dynamics of the N-terminal half of hepatitis C virus Core protein: An intrinsically unstructured protein

Jean-Baptiste Duvignaud, Christian Savard, Rémi Fromentin, Nathalie Majeau, Denis Leclerc and Stéphane M. Gagné.

3.1 Résumé

La protéine Core du virus de l'hépatite C joue un rôle important dans l'assemblage et l'emballage du génome viral. Nous avons étudié la structure de la moitié N-terminale de la protéine Core (C82) qui s'est révélée être le plus court fragment permettant la formation de particules pseudo-virales (PPV) *in vitro* et en levures. L'analyse structurale bioinformatique de la Core C82 suggère que celle-ci est principalement non structurée. Les données de dichroïsme circulaire et de résonance magnétique nucléaire indiquent que la Core C82 ne possède pas de structures secondaires. De plus, les données de relaxation (obtenues par RMN) montrent que la Core C82 est extrêmement désordonnée et flexible. Ces résultats tendent à inclure la moitié N-terminale de la protéine Core du VHC à la famille grandissante des protéines intrinsèquement non-structurées (IUP), ce qui explique la tendance de la protéine Core du VHC à interagir avec un large spectre de protéines de l'hôte, une caractéristique bien connue des IUPs.

3.2 Abstract

Hepatitis C virus Core protein plays an important role in the assembly and packaging of the viral genome. We have studied the structure of the N-terminal half of the Core protein (C82) which was shown to be sufficient for the formation of nucleocapsid-like particle (NLP) *in vitro* and in yeast. Structural bioinformatics analysis of C82 suggests that it is mostly unstructured. Circular dichroism and structural NMR data indicate that C82 lacks secondary structure. Moreover, NMR relaxation data shows that C82 is highly disordered. These results indicate that the N-terminal half of the HCV Core protein belongs to the growing family of intrinsically unstructured proteins (IUP). This explains the tendency of the hepatitis C virus Core protein to interact with several host proteins, a well-documented characteristic of IUPs.

3.3 Introduction

Hepatitis C virus (HCV) is the only member representing the *hepacivirus* genus of the *flaviviridae*. Flaviviruses have a positive-sense, single-stranded RNA genome of 9.6 kb [5] composed of one large open reading frame coding for a 3010 amino acid polyprotein [32]. HCV Core protein (HCV-C) is located at the N-terminus of the polyprotein [5, 12] which is cleaved by host-encoded proteinases to generate an immature and a mature Core protein of 191 and 179 amino acids, respectively [13]. The Core protein of HCV is larger than other *flaviviridae* which are approximately 100 amino acids long. The N-terminal half of HCV-C was shown to be sufficient for assembly in nucleocapsid-like-particles (NLPs) in presence of structured RNA [25]. Structures of Core proteins from two others flaviviruses, the dengue and West Nile viruses, have been characterised by cryomicroscopy, NMR and X-ray crystallography [8, 17, 24]. These studies revealed that the structure of these proteins is mostly α -helical, although the N-terminal 20 amino acids of dengue and West Nile viruses were unstructured as shown by protease digestion, X-ray and NMR [8, 17]. In contrast with the Core of other *flaviviridae*, HCV-C is predicted to contain only few structural elements [21]. Although the three-dimensional structure of HCV-C is still unsolved, recent studies have provided insights into the structure of HCV-C [2, 31]. Boulant et al. [2] have shown that the whole protein (HCV-C 2-169) adopts an α -helical conformation for nearly 50 % of the protein in presence of detergents or lipids. In this study, we have investigated the structure and dynamics of the 82-residues N-terminal half of the HCV-C (C82) in solution. C82 was shown to be sufficient for the formation of nucleocapsid-like particle (NLP) *in vitro* and in yeast [25]. We bring experimental evidence that C82 is in a highly disordered form in aqueous solution. The possibility that the Core or the N-terminal half of the protein belongs to the family of intrinsically unstructured proteins (IUP) [30, 34] will be discussed.

3.4 Materials and methods

3.4.1 Sequence analysis and secondary structure predictions

Amino acid composition of C82 was studied with the ProtParam Tool (www.expasy.org). The mean amino acid composition of 356 194 proteins was obtained from the Swiss-Prot database (UniProtKB/Swiss-Prot protein knowledgebase release 55.0 statistics). We used five algorithms to predict the secondary structure of the HCV-C. The PONDR software was used to predict disorder [9].

3.4.2 Expression and purification of C82

The C82 clone was optimized with the most representative codon for translation in *Escherichia coli* and fused to a C-terminal His₆-tag. This optimized C82 was cloned in a pET3d expression vector and overexpressed in the *E. coli* strains BL21 (DE3) Star (Invitrogen) as already described [25]. The harvested cells were resuspended in a 30 mL ice cold lysis buffer of 50 mM phosphate, 300 mM NaCl at pH 12 with 1 x cocktail of protease inhibitors (Roche Diagnostics GmbH), then lysed by sonication. The lysate was centrifuged at 27,000 x g for 30 min. The filtered supernatant was added to Ni-NTA resin (QIAGEN). After 90 min at 4°C, the beads were washed with 50 mL of three buffers with increasing concentrations of imidazol (up to 20 mM) and NaCl (up to 750 mM). C82 was eluted in either an assembly buffer [25] or an elution buffer (50 mM, pH 8, phosphate buffer, 300 mM NaCl, 500 mM imidazol). A reversed phase HPLC method was used as the second purification step using a VYDAC C4 column. A step gradient was used as follows: 1 min, 10 % acetonitrile; 1-15 min, 20 % acetonitrile; 15-35 min, 30 % acetonitrile; 35-45 min, 100 % acetonitrile. C82 eluted between 15-35 min and was lyophilised. The purity of the protein was estimated by SDS-PAGE and confirmed by Western blot.

3.4.3 Trypsin Digestion

Trypsin digestion of pure C82 was as previously described [20]. Trypsin digestion was done at 4°C with a sample concentrated at ~200 µM. At each interval of the reaction, 10 µl were collected, 1 x cocktail of protease inhibitors and SDS were added to stop the reaction, and the sample was heated at 95°C for 5 min.

3.4.4 In vitro assembly reactions, density gradient centrifugation and electron microscopy

All experiments were carried as previously described [25].

3.4.5 CD spectrum experiments

C82 protein was lyophilised and resuspended in the experimental buffer (10 mM phosphate, pH 7.2). CD measurements were carried on an Jasco Model J-710 instrument at room temperature, using 0.1 cm quartz cuvettes. Data were collected and processed using the Jasco software. Deconvolution was achieved with DICHROWEB [23].

3.4.6 NMR samples

Uniform labelling of C82 was done in either U-¹⁵N and U-¹⁵N-¹³C BioExpress Cell Growth Media (CIL, Andover, MA). Lyophilised C82 was resuspended in the NMR buffer

(25 mM, pH 6.6, phosphate, 50 mM NaCl, 0.2 mM DSS, 0.1 % NaN₃, and 1 x cocktail of protease inhibitors). NMR experiments were carried at a protein concentration of 0.4 mM.

3.4.7 NMR data collection and data processing

All experiments were carried out at 278 K on a 600 MHz VARIAN INOVA spectrometer equipped with Z-axis pulsed field gradient and triple resonance cryogenic probe. For protein assignment, 2D ¹⁵N-HSQC, 3D CBCA(CO)NNH, 3D HNCACB and 3D HNCO were recorded. For relaxation studies, ¹⁵N-R₁, ¹⁵N-R₂ and {¹H}-¹⁵N NOE were recorded (details in suppl. mat.). Chemical shift referencing method is based on IUPAC recommendations using DSS [28]. All NMR data were processed using NMRPipe [7].

3.4.8 Backbone assignments and relaxation analysis

NMR spectra were analyzed with NMRView [16] and assignments (BMRB entry 15767) were done using Smartnotebook [33]. Determination of R₁ and R₂ relaxation rates was done using CURVEFIT (A.G. Palmer, Columbia University, New York, NY). {¹H}-¹⁵N NOE values were obtained directly from the HetNOE analysis function in NMRView. Analysis of relaxation data was performed using the extended model-free formalism [29], using the statistical approach of Mandel et al. [27]. Values for the ¹⁵N gyromagnetic ratio, H-N bond length, and chemical shift anisotropy were $-2.712 \times 10^7 \text{ rad T}^{-1} \text{ s}^{-1}$, 1.02 Å, and -172 ppm, respectively. Residues with significant overlap or poor signal-to-noise ratio were discarded, therefore allowing for the characterization of 58 residues.

3.5 Results and discussion

3.5.1 Secondary structure prediction and sequence analysis of C82

The first half of the protein is predicted to be quasi-depleted of secondary structure elements. In order to assess whether the lack of predicted secondary structure corresponds to disorder, we used the disorder predictor PONDR. The combination of PONDR and secondary structure predictions strongly suggest that the N-terminal region of HCV-C is largely unstructured and highly disordered (Figure 3.1A). Moreover, it is known that amino acid composition in proteins differs in disordered compared to ordered regions [22]. We therefore analyzed the amino acid composition of C82 and compared it to the mean composition of proteins obtained from Swiss-Prot database to verify if it fits this model (Figure 3.1B). Amino acids promoting disorder are overrepresented with a value of 71 % in C82 versus 48.6 % in the databank. As suggested by Uversky et al. [36], a charge-hydrophobicity plot is a useful tool to predict whether a protein is unstructured. The calculated mean net charge of C82 is 0.23 and its mean hydrophobicity is 0.33. The comparison of C82 values with the data bank of known unstructured proteins clearly shows that C82 is in the region of unstructured proteins (Figure 3.1C). Moreover unfolded proteins have a tendency to extreme pI (<5 and >9) [36]; C82 has a pI of 12.2. All predictions strongly suggest that C82 is unstructured, but we nevertheless need experimental validation.

3.5.2 Expression, purification and structural characterization of C82 by CD and NMR spectroscopy

To improve the expression yield in *E. coli*, we optimized the codons of C82, and overproduction was done in the *E. coli* strain BL21 (DE3) Star (Invitrogen) (Figure 3.2A, lanes 1 and 2). We used a simple and efficient two-step purification protocol. The quality of the purified samples was adequate for structural analysis by CD spectroscopy and NMR (Figure 3.2A, lane 4). To confirm our final product, we performed an SDS-PAGE electrophoresis (Figure 3.2A) and an *in vitro* assembly assay showing that the C82 was able to assemble into NLPs in presence of tRNA as seen by electron microscopy (Figure 3.2B). A trypsin digestion test indicates that C82 is significantly sensitive to degradation (Figure 3.2C).

The C82 CD spectrum showed a minimum near 200 nm indicating that this protein is a random coil that lacks apparent secondary structure (Figure 3.3A). However, C82 is able to organize itself *in vitro* into NLPs in the presence of structured RNA. The encapsidation phenomenon most likely induces a structural change in C82 enabling it to adopt a stable secondary structure. The RNA/C82 mixture created turbidity, which is an indication of the encapsidation phenomenon [11] but these samples cannot be used neither for CD nor NMR spectroscopy. We therefore tried several conditions that could induce a structural change; salt concentration, detergent, and pH. Unfortunately, none of these conditions affected the CD signal (data not shown).

Two dimensional ^{15}N -HSQC (Figure 3.3B) and 2D ^{15}N -NOESY-HSQC spectra (suppl. mat.) were recorded. Although there is a poor $^1\text{H}_\text{N}$ chemical shift distribution in the HSQC of C82 (0.95 ppm), which is typical of proteins that possess little tertiary structure and are disordered, we observed 63 of the 69 expected peaks. Moreover, the NOESY spectrum of C82 contains few NOE correlations which is indicative of the absence of secondary structure.

Backbone assignments of $^1\text{H}_\text{N}$, ^{13}C , and ^{15}N resonances resulted in 93 % of non-proline backbone $^1\text{H}_\text{N}$ and ^{15}N , 95 % of all $^{13}\text{C}_\alpha$, 94 % of all $^{13}\text{C}_\beta$, and 93 % of all $^{13}\text{C}'$ assignments being assigned.

The chemical shift index (CSI) [39] was used to assess secondary structure information. As shown in Figure 3.4 (D), CSI C_α of C82 is oscillating around zero, suggesting that C82 exists in an unstructured conformation.

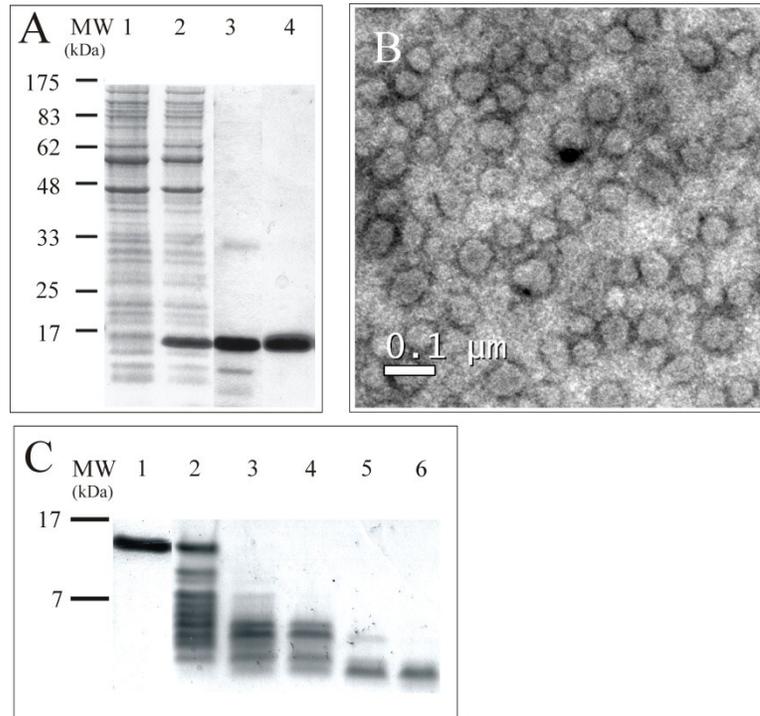


Figure 3.2 : A) Purification profile of the C82 on a 10 % SDS PAGE. MW, molecular weight; column 1, before induction; column 2 after 3h induction; column 3, first elution after IMAC; and column 4, elution after reverse phase HPLC column. It should be noted that C82 migrates at an abnormal molecular weight of ~15 kDa instead of the expected ~10 kDa. B) Electron micrograph of nucleocapsid like particles (NLPs) after *in vitro* assembly with the 5' IRES (nt 1- 372) of the HCV genome. The bar corresponds to 0.1 μm. C) Trypsin digestion of C82 on a 10 % SDS PAGE. Left to right is the reaction time : T = 0 sec, 30 sec, 1 min, 5 min, 10 min, and 20 min.

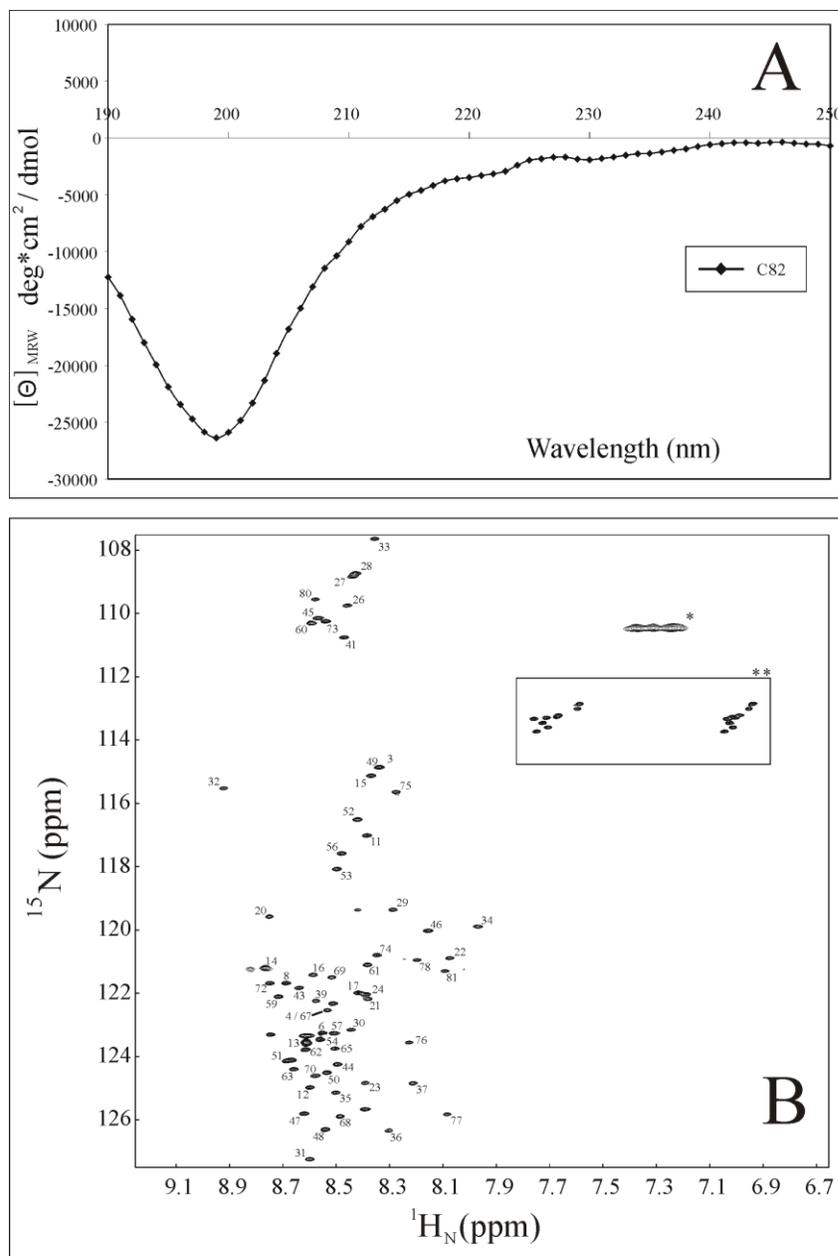


Figure 3.3 : A) CD spectra of C82. The mean residue ellipticity is calculated as follows: $[\Theta]_{MRW} = (\delta \times MRW) / (10 \times c \times l)$, where δ is the ellipticity in degrees, MRW is the mean residue weight, c is the concentration of the sample in g/ml; and l is the path length in cm. B) 2D ^1H - ^{15}N HSQC spectra of C82, with assignments (large spectra in suppl.mat.). Not shown in this expansion is the NH of the lateral chain of the tryptophan residue at 10.2 ppm (H_N) and 129.66 ppm (N_H). * represent the correlation of an arginine side chain residue. ** represent correlations for the side chain of Gln and Asn amino acids.

3.5.3 Backbone dynamics of C82 by NMR

Relaxation data allow us to determine whether C82 is ordered without regular secondary structure or is simply disordered, as both situations are potentially compatible with the CD data. Taking into account its molecular weight, C82 has R_1 and R_2 data that are quite typical of an unstructured protein with average values of $1.67 \pm 0.06 \text{ s}^{-1}$ and $4.96 \pm 0.54 \text{ s}^{-1}$, respectively. The NOE values, with an average of 0.19 ± 0.10 , are characteristic of an unfolded protein with unrestricted fast dynamics. This was confirmed by a comparison of the measured NOE with other well ordered proteins in the BMRB chemical shift databank (<http://www.bmrwisc.edu>) (suppl. mat.). The model-free approach was used to characterize the global and internal motions of the two samples. Acceptable fit could not be extracted with the use of a global correlation time (τ_m) and motional parameters were extracted with a local correlation time τ_{loc} for each residue. 47 residues were better fitted with model 2, the others did not fit with either models 1 and 3, and models 4 and 5 were not considered as they implicate calculations with four variables and only three experimental values are available. The mean τ_{loc} is $4.76 \pm 0.54 \text{ ns}$ (from 3.1 to 5.6 ns) (Figure 3A). The optimized dynamics parameters are summarized in Figure 3B and C. If the previous results were suggesting that C82 is unstructured, the order parameter (S^2) values unambiguously confirm that C82 is highly disordered. It is interesting to note that the dynamical parameters (S^2 and τ_e , Figure 3B and C) obtained are exactly in the range of values observed for unstructured proteins [3]. The dynamical study of the unfolded form of the pro-peptide of subtilisin results in a S^2 of 0.57 ± 0.06 and a τ_e of $74 \pm 11 \text{ ps}$; corresponding values for C82 are 0.59 ± 0.04 and $112 \pm 16 \text{ ps}$.

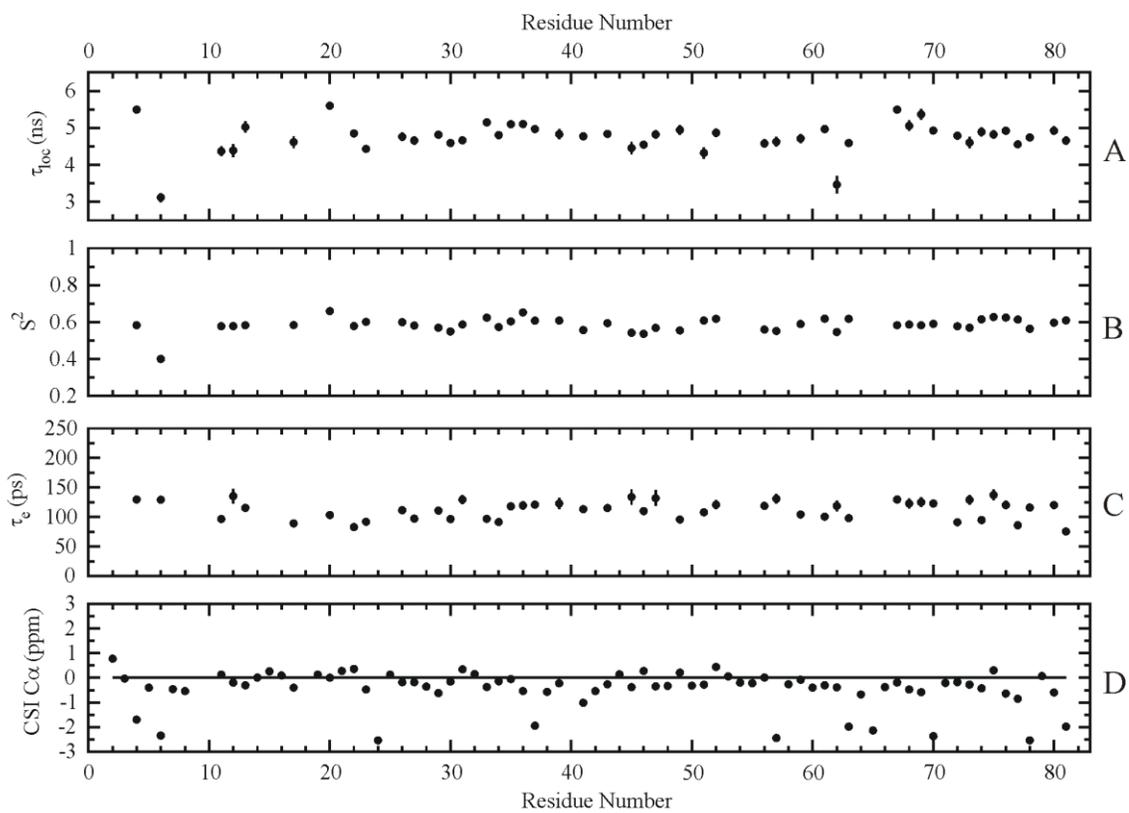


Figure 3.4 : Sequential NMR dynamic data. A) local correlation time (τ_{loc}), B) order parameters (S^2) obtained with model-free, C) effective correlation time for internal motions (τ_e), and D) chemical shift index plot for $C\alpha$ for C82.

3.5.4 HCV Core protein: an IUP

In agreement with predicted data on whole HCV Core, experimental data of isolated C82 indicate that the N-terminal half of HCV Core protein is disordered in aqueous solution. We therefore conclude that C82, and hence the N-terminal half of whole HCV Core, is a member of the IUP family.

IUPs are an important family of proteins that may represent 30 % of the proteins of the human genome [10]. IUPs are known to interact with different partners and the disorder and flexibility of these proteins are important to their function. IUPs are generally involved in binding nucleic acids, other proteins, membranes or a large range of small ligands. More than 30 functions have been assigned to IUPs. They can be chaperones, regulatory proteins, proteins involved in cell signalling and, consequently, can be involved in the pathogenesis of cancers [10, 14, 34, 35, 37].

HCV-C binds to structured RNA and assembles into NLPs. Moreover HCV-C was shown in several reports to interact with several host proteins: the C-terminus of p53 [18], the intracellular domain of lymphotoxin β receptor [4], the DEAD-box protein (DDX3, CAP-Rf) [26], the 14-3-3 protein [1], and the p21Waf1/Cip1/Sid1 [38]. In addition, Cristofari et al. [6] has proposed that the Core protein and its truncated forms at the C-terminus are potent nucleic acid chaperones. Consistently, Tompa & Csermely [35] have suggested that structural disorder is present in RNA and protein chaperones. It may be that HCV-C, through its unstructured domain at the N-terminus, induces a defined structure to the viral RNA for packaging of the genomic RNA. Or inversely, the binding of RNA may confer structure of the N-terminal half of the Core protein permitting interaction with RNA and homotypic interactions as suggested by Kim et al. [19]. Recently, Ivanyi-Nagy et al. [15] have reported that a few other Core proteins of the flavivirus family have this ability to be chaperone, and interestingly, all of these have predicted disordered regions. The dengue and West Nile viruses Core proteins were shown to be structured and rich in α -helices with only short sequence (10-20 amino acids) shown to be unstructured [8, 17]. Other Core

proteins of flavivirus family have large unstructured extremities like HCV-C or can be totally unstructured as predicted for BVDV Core protein. This is the first experimental report of structural and dynamical studies that shows that a flavivirus nucleocapsid is an IUP.

3.6 Acknowledgments

Thanks to P.-Y. Savard and S. Morin, for their advice and technical assistance with NMR studies. This project was funded by a CIHR operating grant, two FCI infrastructure grants (Innovation and New-Opportunity) and the “Réseau SIDAMI” of the Fonds de la Recherche en Santé du Québec (FRSQ).

3.7 Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.bbrc.2008.10.141](https://doi.org/10.1016/j.bbrc.2008.10.141). (Voir Annexe 1)

3.8 References

1. Aoki, H., Hayashi, J., Moriyama, M., Arakawa, Y., and Hino, O., *Hepatitis C virus core protein interacts with 14-3-3 protein and activates the kinase Raf-1*. J Virol, 2000. **74**(4): p. 1736-41.
2. Boulant, S., Vanbelle, C., Ebel, C., Penin, F., and Lavergne, J.P., *Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features*. J Virol, 2005. **79**(17): p. 11353-65.
3. Buevich, A.V., and Baum, J., *Dynamics of Unfolded Proteins: Incorporation of Distributions of Correlation Times in the Model Free Analysis of NMR Relaxation Data*. J Am Chem Soc, 1999. **121**: p. 8671-8672.
4. Chen, C.M., You, L.R., Hwang, L.H., and Lee, Y.H., *Direct interaction of hepatitis C virus core protein with the cellular lymphotoxin-beta receptor modulates the signal pathway of the lymphotoxin-beta receptor*. J Virol, 1997. **71**(12): p. 9417-26.
5. Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., and Houghton, M., *Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome*. Science, 1989. **244**(4902): p. 359-62.
6. Cristofari, G., Ivanyi-Nagy, R., Gabus, C., Boulant, S., Lavergne, J.P., Penin, F., and Darlix, J.L., *The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro*. Nucleic Acids Res, 2004. **32**(8): p. 2623-31.
7. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes*. J Biomol NMR, 1995. **6**(3): p. 277-93.
8. Dokland, T., Walsh, M., Mackenzie, J.M., Khromykh, A.A., Ee, K.H., and Wang, S., *West Nile virus core protein; tetramer structure and ribbon formation*. Structure (Camb), 2004. **12**(7): p. 1157-63.
9. Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradovic, Z., *Intrinsic disorder and protein function*. Biochemistry, 2002. **41**(21): p. 6573-82.
10. Fink, A.L., *Natively unfolded proteins*. Curr Opin Struct Biol, 2005. **15**(1): p. 35-41.
11. Fromentin, R., Majeau, N., Laliberte Gagne, M.E., Boivin, A., Duvignaud, J.B., and Leclerc, D., *A method for in vitro assembly of hepatitis C virus core protein and for screening of inhibitors*. Anal Biochem, 2007. **366**(1): p. 37-45.

12. Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., and Shimotohno, K., *Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis*. Proc Natl Acad Sci U S A, 1991. **88**(13): p. 5547-51.
13. Hussy, P., Langen, H., Mous, J., and Jacobsen, H., *Hepatitis C virus core protein: carboxy-terminal boundaries of two processed species suggest cleavage by a signal peptide peptidase*. Virology, 1996. **224**(1): p. 93-104.
14. Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z., and Dunker, A.K., *Intrinsic disorder in cell-signaling and cancer-associated proteins*. J Mol Biol, 2002. **323**(3): p. 573-84.
15. Ivanyi-Nagy, R., Lavergne, J.P., Gabus, C., Ficheux, D., and Darlix, J.L., *RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae*. Nucleic Acids Res, 2008. **36**(3): p. 712-25.
16. Johnson, B.A.a.B., R.A., *NMR View - computer program for the visualization and analysis of NMR data*. J. Biomol. NMR, 1994. **4**: p. 603-614.
17. Jones, C.T., Ma, L., Burgner, J.W., Groesch, T.D., Post, C.B., and Kuhn, R.J., *Flavivirus capsid is a dimeric alpha-helical protein*. J Virol, 2003. **77**(12): p. 7143-9.
18. Kao, C.F., Chen, S.Y., Chen, J.Y., and Wu Lee, Y.H., *Modulation of p53 transcription regulatory activity and post-translational modification by hepatitis C virus core protein*. Oncogene, 2004. **23**(14): p. 2472-83.
19. Kim, M., Ha, Y., and Park, H.J., *Structural requirements for assembly and homotypic interactions of the hepatitis C virus core protein*. Virus Res, 2006.
20. Kunkel, M. and Watowich, S.J., *Conformational changes accompanying self-assembly of the hepatitis C virus core protein*. Virology, 2002. **294**(2): p. 239-45.
21. Kunkel, M. and Watowich, S.J., *Biophysical characterization of hepatitis C virus core protein: implications for interactions within the virus and host*. FEBS Lett, 2004. **557**(1-3): p. 174-80.
22. Lise, S. and Jones, D.T., *Sequence patterns associated with disordered regions in proteins*. Proteins, 2005. **58**(1): p. 144-50.
23. Lobley, A., Whitmore, L., and Wallace, B.A., *DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra*. Bioinformatics, 2002. **18**(1): p. 211-2.
24. Ma, L., Jones, C.T., Groesch, T.D., Kuhn, R.J., and Post, C.B., *Solution structure of dengue virus capsid protein reveals another fold*. Proc Natl Acad Sci U S A, 2004. **101**(10): p. 3414-9.

25. Majeau, N., Gagne, V., Boivin, A., Bolduc, M., Majeau, J.A., Ouellet, D., and Leclerc, D., *The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation*. J Gen Virol, 2004. **85**(Pt 4): p. 971-81.
26. Mamiya, N. and Worman, H.J., *Hepatitis C virus core protein binds to a DEAD box RNA helicase*. J Biol Chem, 1999. **274**(22): p. 15751-6.
27. Mandel, A.M., Akke, M., and Palmer, A.G., 3rd, *Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme*. J Mol Biol, 1995. **246**(1): p. 144-63.
28. Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., and Wuthrich, K., *Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy*. J Biomol NMR, 1998. **12**(1): p. 1-23.
29. Palmer, A.G., Rance, M., and Wright, P.E., *Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-12 Heteronuclear Nmr-Spectroscopy*. Journal of the American Chemical Society, 1991. **113**(12): p. 4371-4380.
30. Receveur-Brechot, V., Bourhis, J.M., Uversky, V.N., Canard, B., and Longhi, S., *Assessing protein disorder and induced folding*. Proteins, 2006. **62**(1): p. 24-45.
31. Rodriguez-Casado, A., Molina, M., and Carmona, P., *Spectroscopic study of conformational changes accompanying self-assembly of HCV core protein*. Proteins, 2007. **66**(1): p. 110-7.
32. Schulze zur Wiesch, J., Schmitz, H., Borowski, E., and Borowski, P., *The proteins of the Hepatitis C virus: their features and interactions with intracellular protein phosphorylation*. Arch Virol, 2003. **148**(7): p. 1247-67.
33. Slupsky, C.M., Boyko, R.F., Booth, V.K., and Sykes, B.D., *Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments*. J Biomol NMR, 2003. **27**(4): p. 313-21.
34. Tompa, P., *Intrinsically unstructured proteins*. Trends Biochem Sci, 2002. **27**(10): p. 527-33.
35. Tompa, P. and Csermely, P., *The role of structural disorder in the function of RNA and protein chaperones*. Faseb J, 2004. **18**(11): p. 1169-75.
36. Uversky, V.N., Gillespie, J.R., and Fink, A.L., *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins, 2000. **41**(3): p. 415-27.

37. Vucetic, S., Brown, C.J., Dunker, A.K., and Obradovic, Z., *Flavors of protein disorder*. *Proteins*, 2003. **52**(4): p. 573-84.
38. Wang, F., Yoshida, I., Takamatsu, M., Ishido, S., Fujita, T., Oka, K., and Hotta, H., *Complex formation between hepatitis C virus core protein and p21Waf1/Cip1/Sdi1*. *Biochem Biophys Res Commun*, 2000. **273**(2): p. 479-84.
39. Wishart, D.S. and Sykes, B.D., *The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data*. *J Biomol NMR*, 1994. **4**(2): p. 171-80.

Chapitre 4

Structure and dynamics changes induced by TFE on
the N-terminal half of hepatitis C virus Core protein.

Jean-Baptiste Duvignaud, Denis Leclerc and Stéphane M. Gagné.

4.1 Résumé

Outre son rôle dans l'encapsulation de l'ARN viral, la protéine « Core » du virus de l'hépatite C (VHC) est impliquée dans plusieurs autres interactions. Nous avons récemment proposé que ceci soit relié au fait que la moitié N-terminale de cette protéine (C82) consiste en un domaine intrinsèquement non structuré (IUP). Les domaines IUP peuvent adopter une structure secondaire lorsqu'ils interagissent avec une autre molécule, notamment les acides nucléiques ou une autre protéine. Il est possible de mimer ces conditions en modifiant l'environnement de la protéine. Nous avons examiné sa propension à se replier en fonction de la concentration en sels, de la présence de détergents, du pH et de l'ajout de 2,2,2-trifluoro-éthanol (TFE); seul l'ajout de TFE a résulté en un changement structural. L'effet de l'ajout de TFE a été étudié par dichroïsme circulaire, et par des données structurales et dynamiques obtenues par RMN. Ces données indiquent que C82 peut adopter une structure en hélice α ; cette confirmation est probablement pertinente à l'un des rôles fonctionnels de la protéine « Core » du VHC.

[Traduit par la rédaction].

4.2 Abstract

The Core protein of hepatitis C virus is involved in several interactions other than the encapsidation of viral RNA. We recently proposed that this is related to the fact that the N-terminal half of this protein (C82) is an intrinsically unstructured protein (IUP) domain. IUP domains can adopt a secondary structure when they are interacting with another molecule, such as a nucleic acid or a protein. It is also possible to mimic these conditions by modifying the environment of the protein. We investigated the propensity of this protein to fold as a function of salt concentration, detergent, pH, and 2,2,2-trifluoro-ethanol (TFE); only the addition of TFE resulted in a structural change. The effect of TFE addition was studied by circular dichroism, structural, and dynamic data obtained by NMR. The data

indicate that C82 can adopt an α -helical structure; being accessible, this conformation is likely relevant to one of the functional roles of the HCV Core protein.

4.3 Introduction

Hepatitis C virus (HCV) is a highly invasive pathogen that is responsible for infecting more than 170 million people worldwide [17]. Affected patients can develop cirrhosis and hepatocarcinoma, and current therapies are limited in efficiency [36]. The HCV genome is composed of a single positive-strand RNA of 9.6 kb, with two non-coding regions, 5' and 3' untranslated regions (UTR), at its extremities [5]. The genome encodes a polyprotein composed of 10 proteins divided in 2 categories : structural (Core, E1, E2, and p7) and non-structural (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) [20]. The Core protein is positioned at the N-terminal of the polyprotein, which is processed by two consecutive cleavage reactions at the endoplasmic reticulum (ER) membrane by host-encoded proteases, leading to a ~179 amino acid mature protein [14]. The major role of HCV Core protein is to encapsulate the virus genome, and Majeau et al. (2004) demonstrated that the N-terminal half of the Core (C82) was sufficient to form nucleocapsid-like particles (NLPs) with the addition of structured RNA. HCV Core protein is a multifunctional protein known to have a wide range of interactions involving its N- to C-terminal regions [1, 16, 25, 32]. Recently, it has been shown that the Core protein is one of the proteins of the HCV genome that interacts with several cellular proteins within human cells [6]. Furthermore, the N-terminal half of the HCV Core protein (C82) was shown to be an intrinsically unstructured protein (IUP) [9]. IUPs are an important family of proteins that may represent 30 % of the proteins of the human genome [10]. IUPs are known to interact with different partners and the disorder and (or) order of these proteins are important to their function. IUPs are generally involved in binding nucleic acids, other proteins, membranes, or a large range of small ligands. More than 30 functions have been assigned to IUPs. They can be chaperones, regulatory proteins, proteins involved in cell signalling and, consequently, can be involved in the pathogenesis of cancers. Similar to other IUPs, the Core protein and particularly the

N-terminal region has several interacting partners [4, 9, 16, 21, 35]. Consequently, this protein may adopt different conformations depending on the absence or presence of partners. In this study, we investigated the ability of the first half of the HCV Core protein to fold. We therefore characterized the structure and dynamics of C82 in different conditions. We focused our study of the C82 protein in 2,2,2-trifluoro-ethanol (TFE). The relevance of TFE is multiple. First, TFE is a lipomimetic solvent [30] and the HCV Core protein is known to interact with the ER membrane and lipid droplets [28]. Second, TFE can mimic the environment found in protein-protein interactions [13, 29] and HCV Core protein has several known protein partners. Finally, TFE is widely used in structural studies to stabilize potent conformation [30, 34]. In this context, we show that C82 can fold with the addition of TFE, going from a highly flexible unfolded form to a more rigid α -helical conformation. These results confirm data from Boulant et al. (2005), which proposed that the α -helical structure may be present in the N-terminus of the HCV Core protein when detergents or lipids are added, and from Rodriguez-Casado et al. (2007), which proposed an α -helical or β -sheet content depending on the composition of the RNA partner.

4.4 Materials and methods

4.4.1 Purification of truncated hepatitis Core protein C82

A two-step purification protocol was used as described previously [9]. At the end of the protocol, the protein was lyophilized and the purity and identity of the protein were estimated by SDS-PAGE and Western blot (data not shown) after resuspension in the appropriate buffer. In order to confirm the functionality of the truncated protein with a C-terminal His-tag, encapsidation tests were performed as previously described [19]. Furthermore, C82 purified with both protocols described in Duvignaud et al. (2009) and Majeau et al. (2004) resulted in well-formed NLPs.

4.4.2 Circular dichroism experiments

Circular dichroism (CD) measurements were done as described previously [9]. Experiments were performed on C82 in the experimental buffer (10 mmol·L⁻¹ phosphate, pH 7.2) as a control, with pH varying from 2 to 12, with addition of salt ([KCl] up to 1 mol·L⁻¹), with addition of detergent ([SDS] up to 300 mmol·L⁻¹), and with 2,2,2-trifluoroethanol (TFE) up to 90 %. The deconvolution was achieved with DICHROWEB [18].

4.4.3 Nuclear magnetic resonance samples

Uniform labelling of C82 was done in U-¹⁵N-¹³C BioExpress Cell Growth Media (Cambridge Isotope Laboratories, Andover, Mass.). C82 protein was purified as previously described [9] and resuspended in 90 % deuterated TFE (2,2,2-trifluoro-ethanol-D₂), 10 % H₂O, pH 6.6, 0.2 mmol·L⁻¹ DSS. Nuclear magnetic resonance experiments were carried out at a protein concentration of 0.4 mmol·L⁻¹.

4.4.4 NMR data collection and data processing

All experiments were carried out at 5°C on a Varian Inova spectrometer operating at a ¹H resonance frequency of 600 MHz equipped with a *z*-axis pulsed-field gradient and triple-resonance cryogenic probe. This particular temperature was used in accordance with our previous work [9]. For protein assignment, a set of two- and three-dimensional (2D and 3D) experiments were recorded: 2D ¹H-¹⁵N HSQC, 3D CBCA(CO)NNH, 3D HNCACB and 3D HNCO. For relaxation studies, ¹⁵N-R₁, ¹⁵N-R₂ and {¹H}-¹⁵N NOE were recorded (details in suppl. mat). Chemical shift referencing follows IUPAC recommendations using DSS [23]. All NMR data were processed using NMRPipe software [7] and analyzed with NMRView [15] and Smartnotebook [33].

4.4.5 Backbone assignment and relaxation analysis

Experiments were done and analyzed as previously described [9]. Assignments were deposited in the Biological Magnetic Resonance data bank (BMRB; entry No. 15768). Residues with significant overlap or poor signal-to-noise ratio were discarded, therefore allowing for the dynamic characterization of 39 residues.

4.5 Results

4.5.1 Secondary structure of C82 under various conditions: a CD study

C82 is an intrinsically unstructured protein [9] able to self-assemble *in vitro* into NLPs (Nucleocapsid-like particles) upon addition of structured RNA [19]. Unfortunately, NLPs can not be studied by NMR owing to their size and the turbidity induced by the self-assembly reaction [11]. Considering that C82 is an IUP, we have tested several conditions to screen for secondary structure accessible to C82 in specific environments or when in complex with one of its partners. These tests included increasing salt concentration (KCl) from 0 to 1 mol·L⁻¹, addition of detergent (SDS) from 0 to 0.3 mol·L⁻¹, and changing the pH of the solution to range between 2 and 12. CD was used to rapidly screen for global structural content as a function of sample conditions. It was previously shown using CD that C82 does not contain regular secondary structure in 10 mmol·L⁻¹ phosphate at pH 7.2 [9]. In Figures 4.1A-D, we plotted the molar ellipticity at 222 nm as a function of the different conditions tested: salt, detergent, pH, and TFE, respectively. The molar ellipticity values at 222 nm reflect the change in helical composition within the protein, whereas 218 nm reflects the β -sheet and 210 nm the extended structure [12]. Increasing salt or SDS concentrations do not induce noticeable changes in ellipticity at 222 nm (Figures 4.1A and 4.1B) or at any other characteristic wavelengths. Variation of pH from 2 to 12 in the protein solution also had no effect on the molar ellipticity (Figure 4.1C), and no structural changes were observed. Finally, in order to mimic a hydrophobic environment similar to the one found in the membrane [30] or in protein-protein interactions [13, 29], we tested the ability of C82 to form secondary structures in the presence of TFE. Increasing amounts of TFE (from 0 % to 90 %) were tested to seek structural changes (Figures 4.1D and 4.1E). At 50 % TFE, the CD data at 222 nm decreases, which is an indication that C82 acquired some structure (Figure 4.1D) and the CD spectrum shows evidence of an early formation of α -helices (Figure 4.1E). At 90 % TFE, C82 shows a typical CD spectrum of an α -helical structure with minima at 208 and 222 nm (Figure 4.1E). Using deconvolution software

[18], 40 % of helical content was estimated for C82 in 90 % TFE. We also noticed the presence of an isodichroic point at 205 nm, which is a strong evidence of a structural transition between two-conformational states (unstructured to structured in this case) [8].

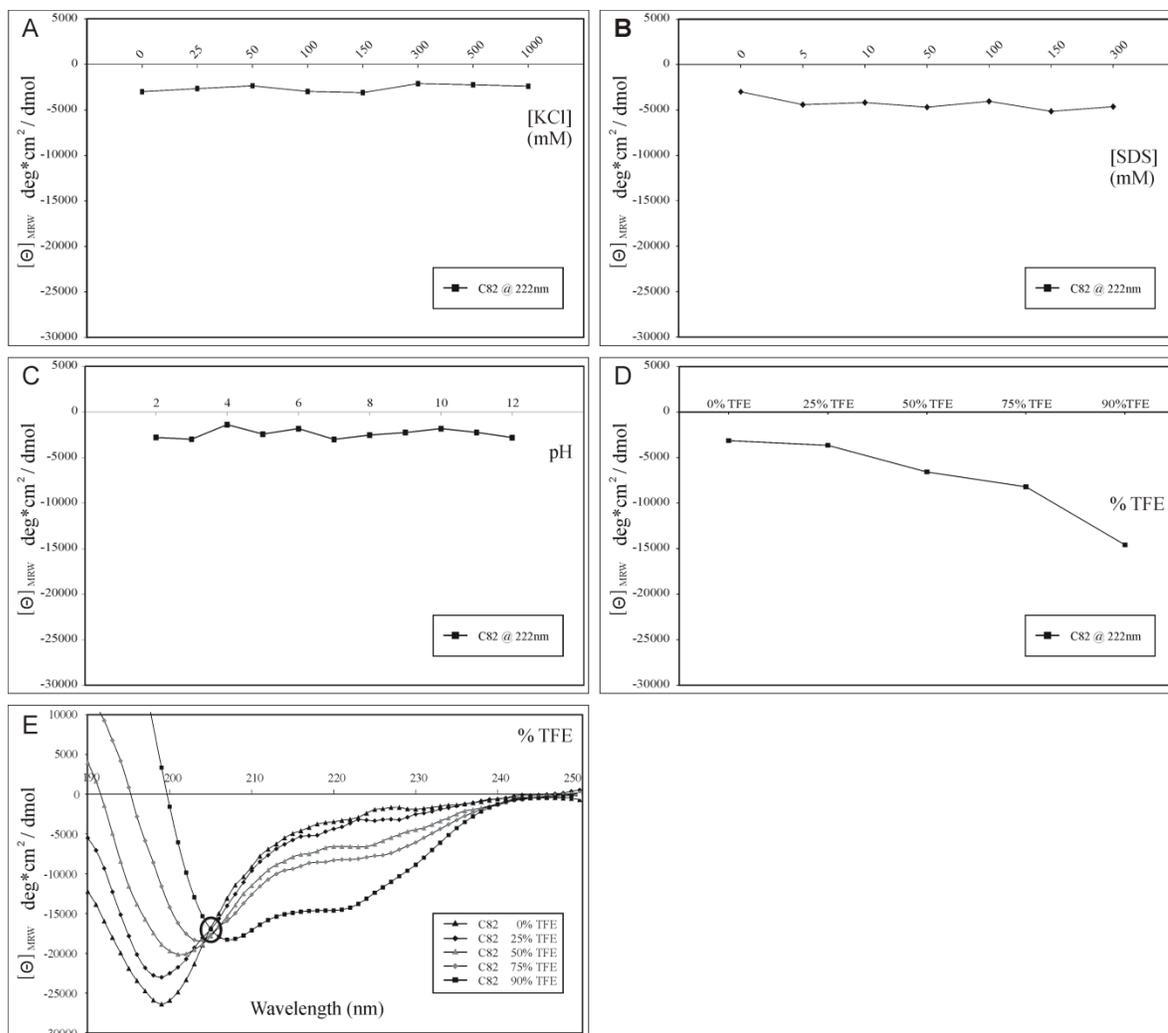


Figure 4.1 : Circular dichroism experiments on C82. A) CD plot at 222 nm of C82 with different concentrations of KCl ($0 \text{ mmol} \cdot \text{L}^{-1} \leq [\text{KCl}] \leq 1 \text{ mol} \cdot \text{L}^{-1}$). B) CD plot at 222 nm of C82 with different concentration of SDS ($0 \text{ mmol} \cdot \text{L}^{-1} \leq [\text{SDS}] \leq 300 \text{ mmol} \cdot \text{L}^{-1}$). C) CD plot at 222 nm of C82 at pH range of 2-12. D) CD plot at 222 nm of C82 with TFE varying from 0 % to 90 %. E) CD spectra of C82 with TFE varying from 0 % to 90 %. The black circle represents an isodichroic point. The mean residue ellipticity is calculated as follows: $[\Theta]_{MRW} = (\delta \times MRW) / (10 \times c \times l)$, where δ is the ellipticity in degrees, MRW is the mean residue weight, c is the concentration of the sample in grams per millilitre ($\text{g} \cdot \text{mL}^{-1}$); and l is the path length in centimetres (cm).

4.5.2 NMR characterization of C82 in 90 % TFE

NMR was used to evaluate secondary and tertiary structure in C82 in 90 % TFE. 2D ^1H - ^{15}N HSQC (Figure 4.2A) and 2D ^{15}N NOESY-HSQC (Figure 4.2B) spectra were recorded at 5°C. The 2D ^1H - ^{15}N HSQC spectrum correlates covalently linked $^1\text{H}_\text{N}$ and ^{15}N . Poorly structured proteins have $^1\text{H}_\text{N}$ chemical shift dispersion of less than 1 ppm which is the case for isolated C82 in aqueous solution [9]. The C82 HSQC spectrum in 90 % TFE is, however, significantly different; we observe a larger $^1\text{H}_\text{N}$ chemical shift distribution in the ^1H - ^{15}N HSQC spectrum (from 7.4 to 9.2 ppm), which is more typical of a structured protein (Figure 4.2A). Forty-eight intense peaks were readily detectable, which is below the expected 69 correlations. This could be explained by resonance broadening due to conformational exchange on the microsecond to millisecond timescale. We also noticed that correlation peaks have a much larger line width in the TFE sample. This can be explained by an increased viscosity, as the protein is dissolved in 90 % deuterated TFE. The 2D ^{15}N NOESY-HSQC spectrum shows several HN-HN correlations giving strong evidence for the presence of α -helices (Figure 4.2B). Given the results of these 2D spectra, which are in agreement with the CD data, the addition of TFE appears to stabilize the structure of C82, which undergoes an unstructured to α -helical transition.

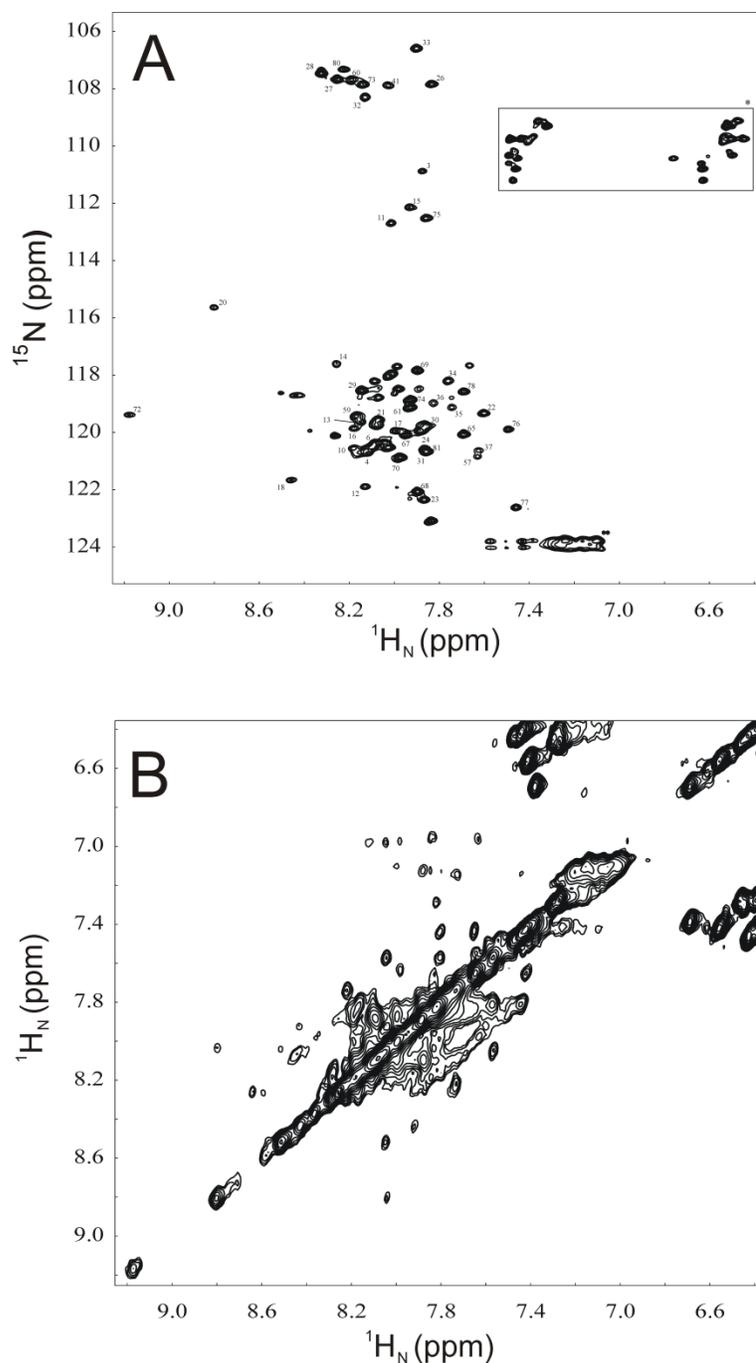


Figure 4.2 : (A) Two-dimensional ^1H - ^{15}N HSQC spectra of C82 in 90 % TFE with assignments (large spectra in suppl. mat). Not shown in this expansion is the NH of the lateral chain of a tryptophan residue at 9.68 ppm (H_N) and 126.23 ppm (N_H). *, correlations for the side chain of Gln and Asn amino acids; **, correlations of the side chain of Arg amino acid. (B) Two-dimensional ^{15}N NOESY-HSQC spectra of C82 in 90 % TFE showing the d_{NN} region. All spectra were recorded at 5°C.

4.5.3 NMR assignments and secondary structure analysis of C82 in 90 % TFE

2D ^1H - ^{15}N HSQC spectra of C82 in 90 % TFE exhibits an adequate dispersion of the proton and nitrogen resonances to permit sequential backbone chemical shift assignments (Figure 4.2A). Backbone $^1\text{H}_\text{N}$, ^{13}C (C_α , C_β , and C'), and ^{15}N resonances were assigned using 3D heteronuclear NMR spectroscopy; 70.6 % of non-proline backbone $^1\text{H}_\text{N}$ and ^{15}N , 73.8 % of all $^{13}\text{C}_\alpha$, 71.4 % of all $^{13}\text{C}_\beta$, and 70.6 % of all $^{13}\text{C}'$ assignments were obtained. As indicated by the backbone dynamic data, missing assignments (most of them between residues 37 and 58) can be explained by resonance broadening due to conformational exchange. The C-terminal His₆-tag was not assigned owing to significant chemical shift overlap.

The C_α and C_β chemical shifts were used to assess secondary structure information by comparing these with random coil values using the secondary structure propensity method (SSP, [24]). As shown in figure 4.3A, the $\text{C}_\alpha\text{C}_\beta$ -based SSP of C82 in 90 % TFE indicates several positive regions suggesting that these parts of the protein can fold as α -helices. We have defined 4 regions corresponding to potential α -helices (with SSP values > 0.1), 3 in the first 40 residues of C82 and 1 in the last 15 residues (see schematic representation under figure 4.3A). The N- and C-terminal α -helices of the protein represent 38 % of the protein sequence.

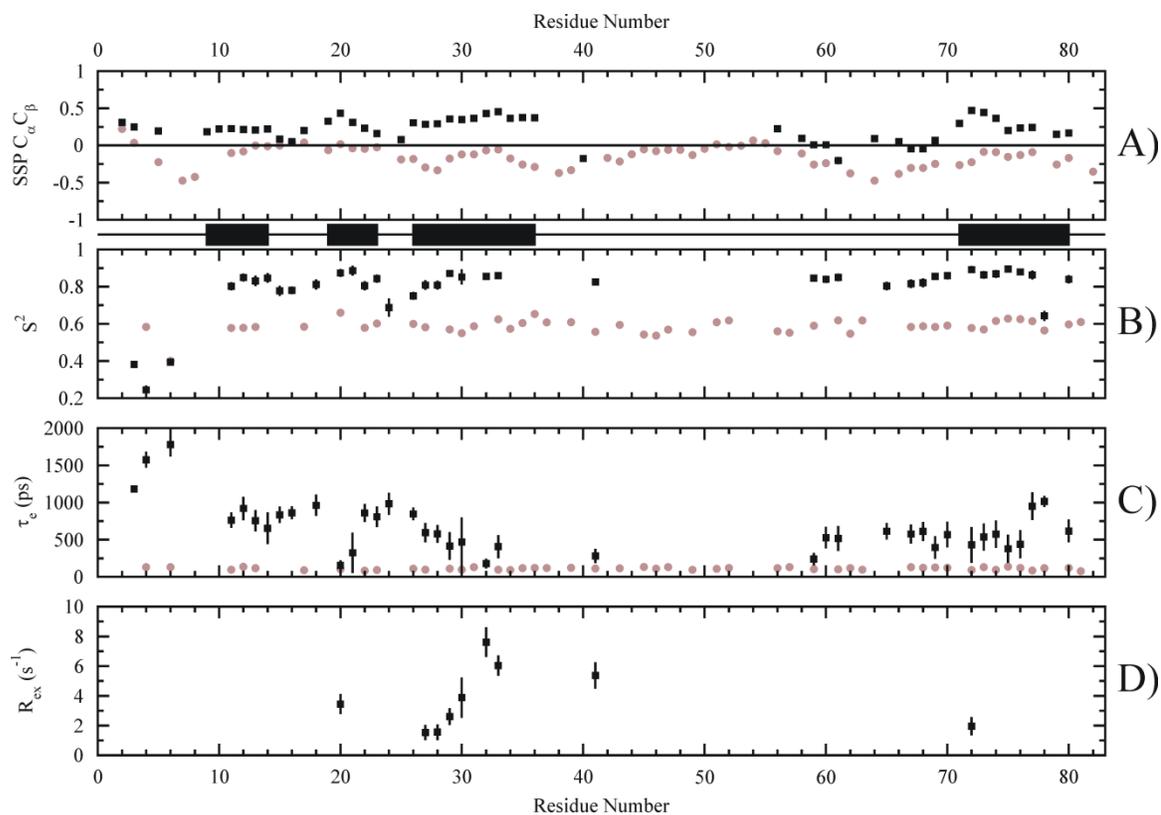


Figure 4.3 : Sequential NMR dynamic data for C82 in 90 % TFE (black square) compared with C82 in H₂O (grey circle) : A) SSP plot of C_α and C_β for C82 in 90 % TFE and H₂O. A positive SSP value indicate an α -helix propensity; the inverse is true for β -sheet. B) Order parameters (S^2) obtained with model-free. (C) Effective correlation time for internal motions (τ_c). (D) Chemical exchange term R_{ex} . Between panels A and B, a secondary structure representation of C82 in 90 % TFE is depicted by black squares for α -helices with a high SSP score.

4.5.4 Backbone dynamics of C82 by NMR

Proteins are intrinsically dynamic, and NMR is a well-established technique to study the dynamics of proteins in solution at atomic resolution. The timescale of protein motions ranges from picoseconds to seconds, and NMR allows for the characterization of motions across this range of timescales. CD experimental data does not discriminate between an helical structure that is rigid on the experimental timescale and a large number of residues that adopt a partially disordered helical conformation. To better characterize the helical content and understand the dynamics of C82 in 90 % TFE, ^{15}N relaxation parameters (^{15}N - R_1 , ^{15}N - R_2 , and $\{^1\text{H}\}$ - ^{15}N -NOE) were recorded at 600 MHz ^1H frequency. Residues with significant overlap or poor signal-to-noise ratio were discarded, therefore allowing the characterization of 39 residues.

The model-free approach was used to characterize the global and internal motions of the C82 protein in 90 % TFE. Dynamic parameters such as τ_m , τ_{loc} , S^2 , τ_e , and R_{ex} describe global and local dynamic motions. τ_m represents the global correlation time, τ_{loc} the local correlation time, S^2 the order parameter for internal motions (S^2_f for the fast timescale and S^2_s for the slow timescale), τ_e represents the effective correlation time for internal motions, and R_{ex} represents the conformational exchange contribution to R_2 from micro to milliseconds timescale motions. Five different models were used to fit experimental data: model 1: S^2 ; model 2 : S^2 , τ_e ; model 3 : S^2 , R_{ex} ; model 4 : S^2 , τ_e , R_{ex} ; model 5 : S_s^2 , τ_s , S_f^2 .

In 90 % TFE, the observed R_1 , R_2 , and NOE values are similar to those expected for a folded protein. Although the R_1 values are as expected for a folded protein of this size, several R_2 's are much larger than expected. The NOE's, with an average of 0.51, are significantly higher than those observed for C82 in water [9] (Figure 4.4) and indicate that C82 gains significant order in TFE, although probably not as ordered as a typical globular protein (Figure 4.4).

Residues selected to determine the τ_m of C82 included only those with an R_1 within $+R_1 \pm \sigma_{R1}$ and an R_2 within $+R_2 \pm \sigma_{R2}$. Additional residues were discarded based on the selection method used by Mandel et al. [22]. Using model 2 with the remaining 23 residues, we obtained an initial estimated τ_m of 10.3 ns which is significantly longer than expected for an 82-residues protein in aqueous solution [3]. Three factors can contribute to this high value, as τ_m depends on the hydrated radius of the protein, the viscosity of the solution and the temperature. The first two factors would be affected in a 90 % TFE solution compared with a normal aqueous solution and can contribute to a longer τ_m . Several studies have shown that TFE:H₂O mixtures are more viscous than water alone. Using the work of Kaiser & al. (1989), and Olivé & al. (1996), we have extrapolated the theoretical viscosity of a 90:10 mixture of TFE:H₂O, hence we would expect a τ_m that is about 2.2 times longer than a sample in aqueous solution at 278 K. Using the average τ_{loc} obtained for C82 in our previous study (4.77 ns), we would expect a τ_m on the order of 10.5 ns in 90 % TFE at 278 K, which is very close to the value obtained.

Using 10.3 ns as the global correlation time, 25 residues were best fitted with model 2, 9 with model 4 and 5 with model 5. After a final optimization of the dynamical parameters (Figure 4.3), the average S^2 for the whole protein is 0.79 ± 0.14 using all residues, and 0.84 ± 0.04 when discarding 4 outsiders (3, 4, 6, 78). Residues between 10 and 34 are well ordered and are folded in α -helix conformations according to the SSP (3 helices with $\langle S^2 \rangle$ of 0.82 ± 0.05). The third helix (residues 26-36) has R_{ex} terms that suggest that there is conformational exchange occurring on the microsecond to millisecond timescale (Figure 4.3D). The R_{ex} terms increase from residues 27 to 33, which is the likely reason we do not have relaxation information for the following central portion of C82 (residue 37-58). Finally, the C-terminal portion (residues 59-81) of C82 is well ordered with an S^2 of 0.84 ± 0.06 , and according to SSP, one α -helix from residue 71 to 80, which is highly ordered with a S^2 of 0.85 ± 0.08 (Figures 4.3A and B).

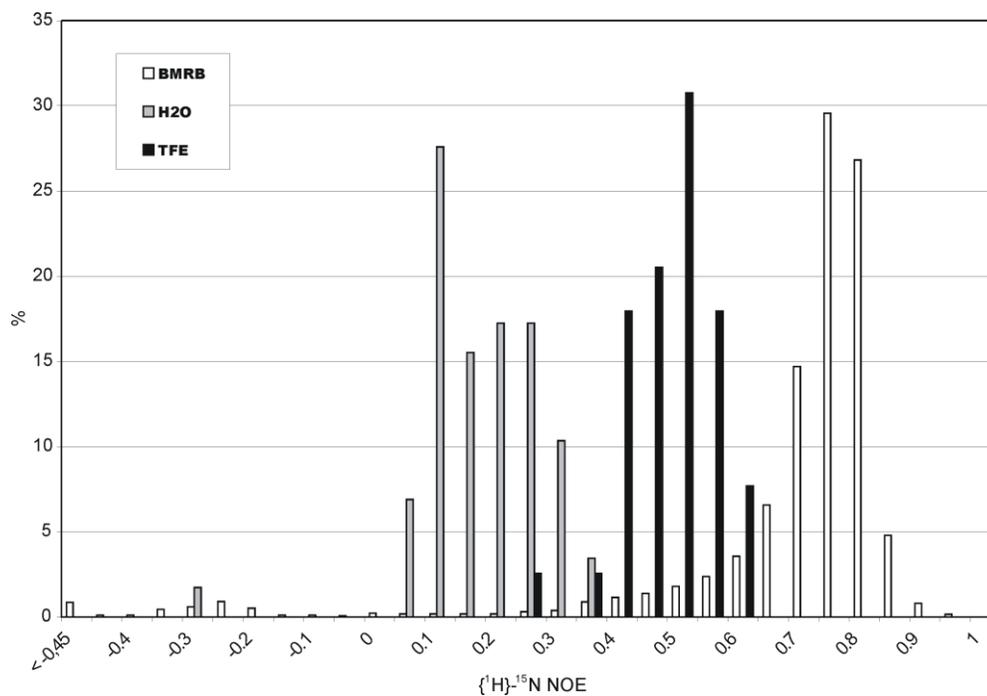


Figure 4.4 : Comparison of the distribution of heteronuclear NOEs measured for C82 in 90 % TFE (grey bar) and in “water condition” (black bar) [9] to the distribution of heteronuclear NOEs found in the BMRB (white bar). Data from the BMRB corresponds to 4509 heteronuclear NOEs taken from 29 depositions. We selected BMRB data that was coherent (no heteronuclear NOE > 1) and only heteronuclear NOE data from 600MHz experiments.

4.6 Discussion

Using both CD and NMR methods, we show that C82 can partially fold into an α -helical structure in the presence of TFE. CD experiments show that C82 has a 40 % α -helix content. Meanwhile, the NMR SSP data indicates that 38 % of observed residues adopt an α -helical conformation. Despite these values being similar, one must bear in mind that the NMR data characterizes only 60 % of the residues of C82. It is also important to remember that SSP indicates a secondary structure propensity. The SSP results suggest that the 4 helices are present \sim 30 % of the time, which raises a significant difference between CD and NMR results. A potential explanation for this apparent discrepancy in helical content is that residues in the central region (37-58) fold in an α -helix and would contribute to an additional percentage if they could be observed by NMR. It should also be noted that the dynamics data suggest significant flexibility within the 4 helices on the CD timescale, which would reduce the contribution of these residues to the observed CD ellipticity. This could account for the difference in α -helix content determined by CD and NMR SSP data.

We plotted the $C_{\alpha}C_{\beta}$ -based SSP and dynamics data on a deposited structure (PDB ID : 1CWX, model 3) of a N-terminal peptide of HCV Core protein (residues 2 to 45) in 40 % TFE by NMR (Figure 4.5). It is interesting to see that at 40 % TFE a small content of α -helix is present as also suggested by our result in CD at 50 % and 75 % TFE. Looking at this peptide structure, it appears that the third helix is the first to fold when TFE is added to the C82. The other two helices were not observed in the 1CWX structure and amino acids with a positive SSP are shown in Figure 4.5A. The three helices have high S^2 (Figure 4.5B) and only the third helix and a portion of its linker to helix 2 have conformational exchange (R_{ex}) (Figure 4.5C). The secondary structure discrepancy between C82 and the 2-45 peptide may be due to the different TFE concentration, and (or) the different experimental temperature (5°C in this study and 20°C with the 1CWX structure), which may cause less pronounced secondary structure in the 2-45 peptide. It should, however, be noted that the 2D 1H - ^{15}N HSQC of C82 at 20°C conserves the chemical shift dispersion observed at 5°C, indicating that the structure at higher temperature is maintained (data not shown). The

conformational difference may also simply be a consequence of the shorter nature of the peptide in the 1CWX structure.

In this report, we do not claim that C82 is natively folded in α -helices, but the results presented here indicate that C82 is able to adopt an α -helix conformation. Despite our best effort, the NMR data of C82 do not allow for complete atomic resolution structure determination, mostly owing to a structure that remains relatively dynamic and to the prevalent presence of conformational exchange. However, de novo structure prediction of a short peptide of HCV Core protein comprising residues 1-25 also confirms that the first half of the protein can adopt an α -helical conformation under hydrophilic condition as supported by data obtained by PEP-FOLD [26, 27] (Figure 4.6). De novo structures obtained from PEP-FOLD show the presence of two α -helices positioned exactly as observed by NMR (amino acids 6-14 and 19-22). Moreover, several reports [2, 31] suggest that the first 120 amino acids of the HCV Core protein can fold in either α -helices or β -sheets. Our experimental data suggest that a hydrophobic environment will be favourable to trigger an α -helix conformation rather than a β -sheet. It is also likely, depending of the environment or the ligand, that other conformations can be adopted [31]. Considering that HCV Core protein has several potential partners [6], it is reasonable to speculate that several different structures can be adopted. All of these contributions are opening interesting avenues for the study of the different structures that HCV Core protein can adopt and new conformations may be revealed in the near future.

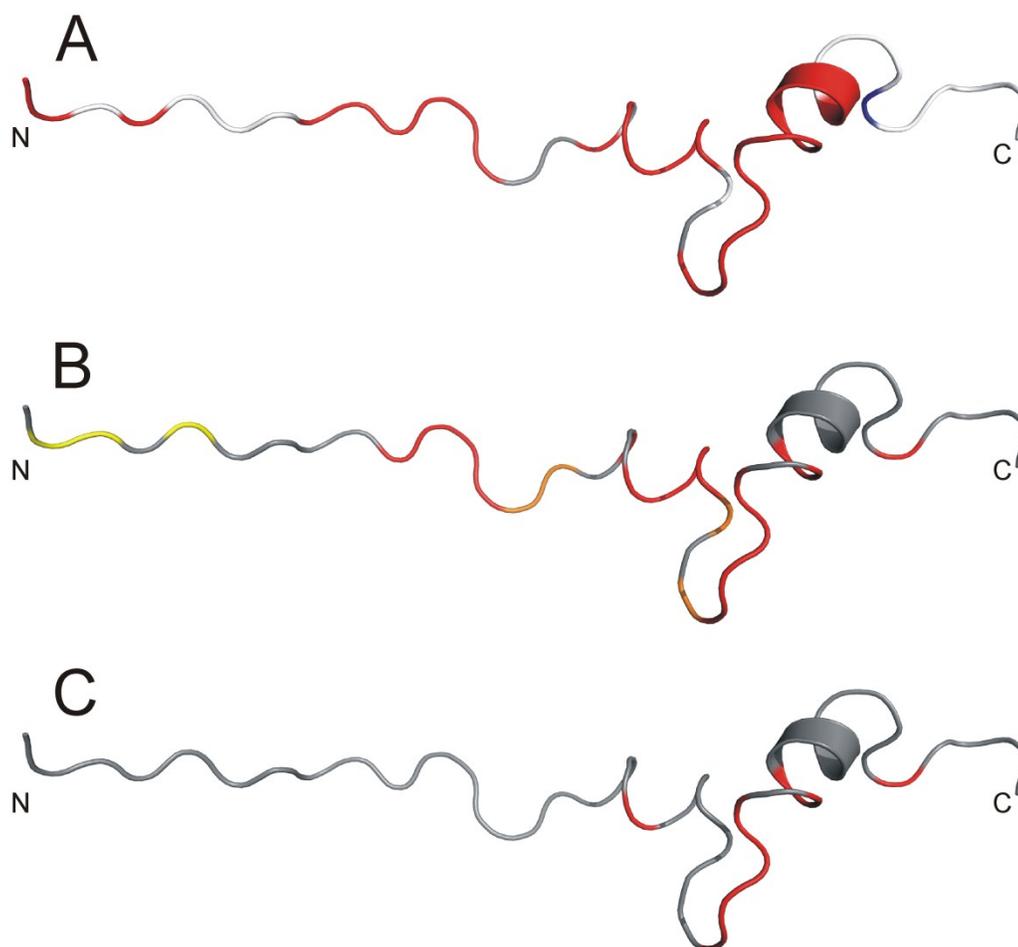


Figure 4.5 : Representation of C82 NMR dynamics data on the structure 1CWX (003) obtained from PDB databank, which represent the first 45 amino acids in 40 % TFE. The colour version of this figure is available at <http://bcb.nrc.ca>. (A) Superposition of SSP $C_{\alpha}C_{\beta}$ data on the structure : red, amino acids with $SSP \geq 0.1$ (helix propensity); blue, amino acids with $SSP \leq -0.1$ (β -sheet propensity); grey, amino acids with SSP within -0.1 and 0.1 (undetermined propensity); white, unassigned amino acids and those preceding a proline residue. (B) Order parameters (S^2), amino acids with high S^2 (>0.7) are in red, S^2 between 0.6 and 0.7 are in orange, and those under 0.6 are in yellow. (C) R_{ex} , amino acids with an R_{ex} term are coloured in red.

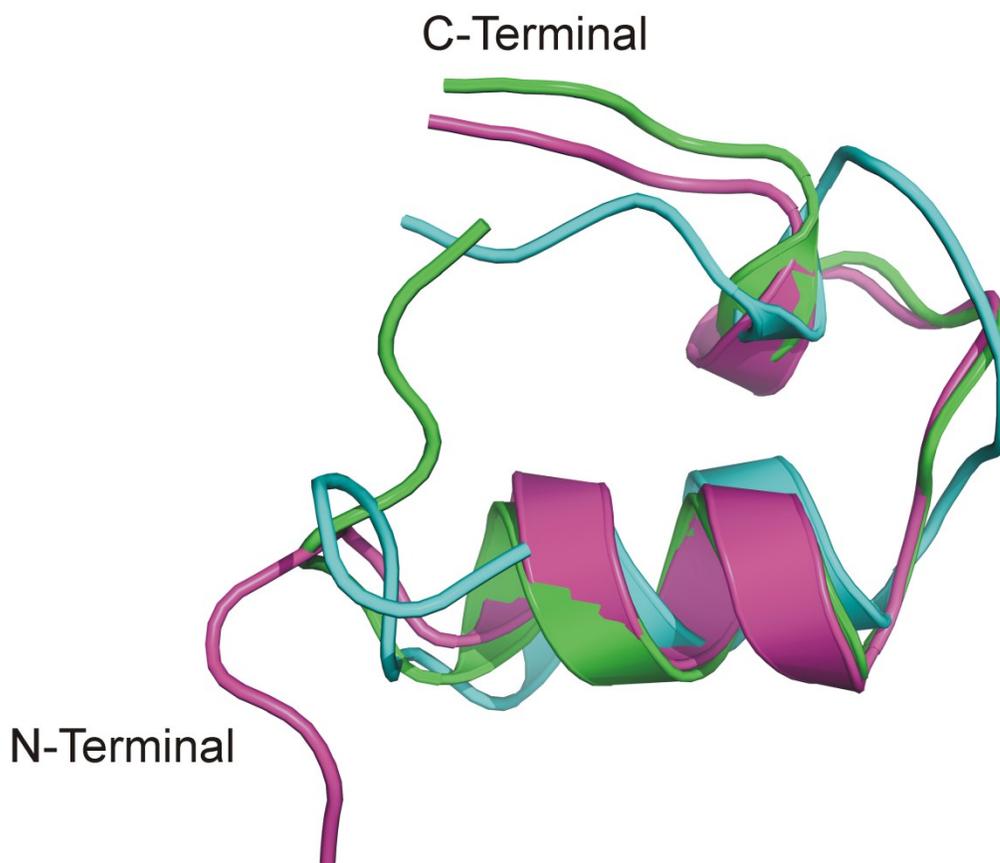


Figure 4.6 : Superposition of three 3D structures of peptide 1-25 from HCV Core protein. These structures were obtained using the PEP-FOLD *de novo* structure prediction protocol [26, 27]. They were simulated in a hydrophylic environment, providing evidence that HCV Core protein can fold in α -helices, as observed in the TFE study. Amino acids 6 to 14 and 19 to 22 are folded in α -helices. These structures were selected from 18 structures obtained with PEP-FOLD, which returns the most representative conformations identified in terms of energy and population [26, 27].

4.7 Acknowledgments

Thanks to C. Savard, R. Fromentin, N. Majeau and the bees for stimulating discussions on HCV. This project was funded by a CIHR operating grant, FQRNT grant 113287, and the “Réseau SIDAMI” of the Fonds de la Recherche en Santé du Québec (FRSQ).

4.8 Supplementary materials

Voir Annexe 2.

4.9 References

1. Aoki, H., Hayashi, J., Moriyama, M., Arakawa, Y., and Hino, O., *Hepatitis C virus core protein interacts with 14-3-3 protein and activates the kinase Raf-1*. J Virol, 2000. **74**(4): p. 1736-41.
2. Boulant, S., Vanbelle, C., Ebel, C., Penin, F., and Lavergne, J.P., *Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features*. J Virol, 2005. **79**(17): p. 11353-65.
3. Brokx, R.D., Scheek, R.M., Weljie, A.M., and Vogel, H.J., *Backbone dynamic properties of the central linker region of calcium-calmodulin in 35% trifluoroethanol*. J Struct Biol, 2004. **146**(3): p. 272-80.
4. Chen, C.M., You, L.R., Hwang, L.H., and Lee, Y.H., *Direct interaction of hepatitis C virus core protein with the cellular lymphotoxin-beta receptor modulates the signal pathway of the lymphotoxin-beta receptor*. J Virol, 1997. **71**(12): p. 9417-26.
5. Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., and Houghton, M., *Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome*. Science, 1989. **244**(4902): p. 359-62.
6. de Chasse, B., Navratil, V., Tafforeau, L., Hiet, M.S., Aublin-Gex, A., Agaogue, S., Meiffren, G., Pradezynski, F., Faria, B.F., Chantier, T., Le Breton, M., Pellet, J., Davoust, N., Mangeot, P.E., Chaboud, A., Penin, F., Jacob, Y., Vidalain, P.O., Vidal, M., Andre, P., Rabourdin-Combe, C., and Lotteau, V., *Hepatitis C virus infection protein network*. Mol Syst Biol, 2008. **4**: p. 230.
7. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes*. J Biomol NMR, 1995. **6**(3): p. 277-93.
8. Dettin, M., Scarinci, C., Zanotto, C., Roncon, R., De Rossi, A., and Di Bello, C., *Biological and conformational studies on analogues of a synthetic peptide enhancing HIV-1 infection*. J Pept Sci, 1998. **4**(7): p. 436-48.
9. Duvignaud, J.B., Savard, C., Fromentin, R., Majeau, N., Leclerc, D., and Gagne, S.M., *Structure and dynamics of the N-terminal half of hepatitis C virus core protein: an intrinsically unstructured protein*. Biochem Biophys Res Commun, 2009. **378**(1): p. 27-31.
10. Fink, A.L., *Natively unfolded proteins*. Curr Opin Struct Biol, 2005. **15**(1): p. 35-41.

11. Fromentin, R., Majeau, N., Laliberte Gagne, M.E., Boivin, A., Duvignaud, J.B., and Leclerc, D., *A method for in vitro assembly of hepatitis C virus core protein and for screening of inhibitors*. *Anal Biochem*, 2007. **366**(1): p. 37-45.
12. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure*. *Nature Protocols*, 2006. **1**(6): p. 2876-2890.
13. Hua, Q.X., Jia, W.H., Bullock, B.P., Habener, J.F., and Weiss, M.A., *Transcriptional activator-coactivator recognition: nascent folding of a kinase-inducible transactivation domain predicts its structure on coactivator binding*. *Biochemistry*, 1998. **37**(17): p. 5858-66.
14. Hussy, P., Langen, H., Mous, J., and Jacobsen, H., *Hepatitis C virus core protein: carboxy-terminal boundaries of two processed species suggest cleavage by a signal peptide peptidase*. *Virology*, 1996. **224**(1): p. 93-104.
15. Johnson, B.A.a.B., R.A., *NMR View - computer program for the visualization and analysis of NMR data*. *J. Biomol. NMR*, 1994. **4**: p. 603-614.
16. Kao, C.F., Chen, S.Y., Chen, J.Y., and Wu Lee, Y.H., *Modulation of p53 transcription regulatory activity and post-translational modification by hepatitis C virus core protein*. *Oncogene*, 2004. **23**(14): p. 2472-83.
17. Lauer, G.M. and Walker, B.D., *Hepatitis C virus infection*. *N Engl J Med*, 2001. **345**(1): p. 41-52.
18. Lobley, A., Whitmore, L., and Wallace, B.A., *DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra*. *Bioinformatics*, 2002. **18**(1): p. 211-2.
19. Majeau, N., Gagne, V., Boivin, A., Bolduc, M., Majeau, J.A., Ouellet, D., and Leclerc, D., *The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation*. *J Gen Virol*, 2004. **85**(Pt 4): p. 971-81.
20. Major, M.E. and Feinstone, S.M., *The molecular virology of hepatitis C*. *Hepatology*, 1997. **25**(6): p. 1527-38.
21. Mamiya, N. and Worman, H.J., *Hepatitis C virus core protein binds to a DEAD box RNA helicase*. *J Biol Chem*, 1999. **274**(22): p. 15751-6.
22. Mandel, A.M., Akke, M., and Palmer, A.G., 3rd, *Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme*. *J Mol Biol*, 1995. **246**(1): p. 144-63.
23. Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., and Wuthrich, K., *Recommendations for the presentation of NMR structures of proteins and nucleic acids*. *IUPAC-IUBMB-IUPAB Inter-Union Task Group on the*

- Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy.* J Biomol NMR, 1998. **12**(1): p. 1-23.
24. Marsh, J.A., Singh, V.K., Jia, Z., and Forman-Kay, J.D., *Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation.* Protein Sci, 2006. **15**(12): p. 2795-804.
 25. Masaki, T., Suzuki, R., Murakami, K., Aizaki, H., Ishii, K., Murayama, A., Date, T., Matsuura, Y., Miyamura, T., Wakita, T., and Suzuki, T., *Interaction of hepatitis C virus nonstructural protein 5A with core protein is critical for the production of infectious virus particles.* J Virol, 2008. **82**(16): p. 7964-76.
 26. Maupetit, J., Derreumaux, P., and Tuffery, P., *A fast method for large-scale De Novo peptide and miniprotein structure prediction.* J Comput Chem, 2009.
 27. Maupetit, J., Derreumaux, P., and Tuffery, P., *PEP-FOLD: an online resource for de novo peptide structure prediction.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W498-503.
 28. McLauchlan, J., Lemberg, M.K., Hope, G., and Martoglio, B., *Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets.* Embo J, 2002. **21**(15): p. 3980-8.
 29. Moncoq, K., Broutin, I., Craescu, C.T., Vachette, P., Ducruix, A., and Durand, D., *SAXS study of the PIR domain from the Grb14 molecular adaptor: a natively unfolded protein with a transient structure primer?* Biophys J, 2004. **87**(6): p. 4056-64.
 30. Muller, I., Sarramegna, V., Milon, A., and Talmont, F.J., *The N-Terminal End Truncated Mu-Opioid Receptor: from Expression to Circular Dichroism Analysis.* Appl Biochem Biotechnol, 2009.
 31. Rodriguez-Casado, A., Molina, M., and Carmona, P., *Core protein-nucleic acid interactions in hepatitis C virus as revealed by Raman and circular dichroism spectroscopy.* Appl Spectrosc, 2007. **61**(11): p. 1219-24.
 32. Sabile, A., Perlemuter, G., Bono, F., Kohara, K., Demaugre, F., Kohara, M., Matsuura, Y., Miyamura, T., Brechot, C., and Barba, G., *Hepatitis C virus core protein binds to apolipoprotein AII and its secretion is modulated by fibrates.* Hepatology, 1999. **30**(4): p. 1064-76.
 33. Slupsky, C.M., Boyko, R.F., Booth, V.K., and Sykes, B.D., *Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments.* J Biomol NMR, 2003. **27**(4): p. 313-21.

34. Vamvakas, S.S., Leondiadis, L., Pairas, G., Manessi-Zoupa, E., Spyroulias, G.A., and Cordopatis, P., *Folding in solution of the C-catalytic protein fragment of angiotensin-converting enzyme*. J Pept Sci, 2009. **15**(8): p. 504-10.
35. Wang, F., Yoshida, I., Takamatsu, M., Ishido, S., Fujita, T., Oka, K., and Hotta, H., *Complex formation between hepatitis C virus core protein and p21Waf1/Cip1/Sdi1*. Biochem Biophys Res Commun, 2000. **273**(2): p. 479-84.
36. Yuan, H.J. and Lee, W.M., *Nonresponse to treatment for hepatitis C: current management strategies*. Drugs, 2008. **68**(1): p. 27-42.

Chapitre 5

Analyse structurale des diverses formes tronquées de
la Core du VHC

5.1 Introduction

La structure tridimensionnelle de la Core du virus de l'hépatite C reste à ce jour un mystère non résolu. De nombreuses équipes de par le monde se sont penchées sur la question et plusieurs publications font état de résultats sur la structure adoptée de la Core. Les premiers travaux publiés sont l'œuvre de Megan Kunkel [147], suivi de ceux de Steve Boulant [30] sans toutefois apporter d'informations structurales à l'échelle atomique. Par contre, certains blocs structuraux ont été analysés en profondeur par résonance magnétique nucléaire tels que les 82 premiers acides aminés de la Core présentés dans cette thèse aux chapitres 3 et 4, mais aussi deux peptides de la région C-terminale représentant les deux hélices amphipatiques prédites en position 118-134 et 144-165 (Figure 3.1 chapitre 3) interagissant avec les lipides du réticulum endoplasmique et des gouttelettes lipidiques [29].

À l'époque où ont été réalisés la grande majorité des travaux présentés dans ce chapitre, l'obtention de protéine Core pleine longueur était un défi important, et aucune étude structurale de la Core n'avait été encore publiée. Ce chapitre vous présentera donc les résultats que nous avons obtenus sur l'analyse structurale de deux formes tronquées de la Core (C124 et C170) dans des conditions non explorées dans les travaux des auteurs susmentionnés.

5.2 Matériels et méthodes

5.2.1 Protocole de production et purification de C124 et C170

Tout comme le clone C82 utilisé aux chapitres 3 et 4 de cette thèse, les clones C124 et C170 ont été optimisés pour une expression protéique bactérienne. Les constructions C124 et C170 ont été clonées au sein du plasmide pET3d et ont été utilisées pour transformer *E. coli* BL21(DE3) pLysS. La culture bactérienne a été effectuée en flasque de 2L à 22°C ou 37°C dans un milieu 2XYT (pour 1 L, 16 g de tryptone, 10 g d'extrait de levure, 5 g de NaCl, et le pH ajusté à 7.2). Une pré-culture durant la nuit en milieu 2XYT a été utilisée pour ensemercer la culture principale en présence de 50 µg/ml d'ampicilline. Une fois que la D.O.₆₀₀ a atteint 0,6, l'induction a été amorcée à l'aide de 1mM d'IPTG pour une durée de trois à quatre heures. Une fois la surexpression protéique terminée, le milieu de culture a été centrifugé à 4000 x g pendant 30 minutes. Le culot a été resuspendu dans le tampon dit de « lyse » (dont la composition est la suivante : 8 M Urée, 20 mM Tris, 300 mM NaCl, pH 7.4, auquel un cocktail d'inhibiteurs de protéase est ajouté) et soumis à une lyse bactérienne par sonication. Une chromatographie d'affinité au Ni²⁺ via le His-tag présent en C-terminal sur les diverses constructions (C82, C124 ou C170) a été effectuée suivis d'une colonne de dessalage pour renaturer et équilibrer les protéines produites dans un tampon d'analyse standardisé (« tampon RMN » : 25 mM Phosphate, 50 mM NaCl, pH 6.6). Au final, plusieurs milligrammes de protéines pures ont été obtenus, et au besoin, les protéines ont été lyophilisées après dialyse contre de l'eau.

5.2.2 Liaison croisée au glutaraldéhyde

Une faible quantité de protéine est nécessaire pour ce type d'expérience. Les protéines C82, C124 et C170 étaient à des concentrations de l'ordre de 0,5 à 1 mg/ml dans le « tampon RMN ». La réaction a lieu sur une dizaine de minutes à température ambiante, une faible quantité de glutaraldéhyde est ajouté à la solution protéique (0.1 % final), et le

tout est placé à l'abri de la lumière pour la durée de l'expérience. La réaction a pour effet de créer un lien covalent entre les partenaires protéiques proches dans l'espace (c'est-à-dire des protéines sous forme de multimères, ou des protéines qui ont tendance à interagir entre elles). L'expérience prend fin par l'ajout du tampon de chargement contenant du SDS et du β -mercaptoethanol, les échantillons sont bouillis et ensuite déposés sur un gel SDS-PAGE pour migration.

5.2.3 Analyse Structurale

5.2.3.1 Diffusion dynamique de la lumière (DDL)

Les expériences ont été enregistrées sur l'appareil « Zetasizer nano series » dans une cuvette de 1 cm x 1 cm de parcours optique. Les protéines C82, C124 et C170 avaient pour concentration respective 0,62 mg/ml, 0,79 mg/ml et 1,01 mg/ml, les trois protéines étant resuspendues dans le « tampon RMN » et filtré à 0,22 μ m. Les expériences ont été enregistrées à 25°C. La viscosité du tampon a été approximée à celle de l'eau soit 0,8872 cP. L'analyse a été effectuée avec le logiciel « Zetasizer ».

5.2.3.2 Dichroïsme circulaire

Les spectres de dichroïsme circulaire ont été mesurés sur un spectropolarimètre Jasco J-710 à température ambiante. Les expériences pour l'ensemble des trois constructions ont été réalisées dans le dit « tampon RMN » à des concentrations variant de 0,2 mg/ml à 1,3 mg/ml. Les enregistrements ont été faits en utilisant une cuvette de quartz possédant un trajet optique de 1 mm. Les spectres finaux consistent en la moyenne de 10 enregistrements indépendants ensuite convertis en ellipticité molaire moyenne par résidu ($[\theta]_{MRW}$ (voir définition légende Figure 3.3)). Les expériences où l'ajout de dodécylsulfate de sodium (SDS) ou de 2,2,2-trifluoroéthanol (TFE) a été nécessaire ont été réalisées par

dilution extemporanée. Un spectre en présence de *n*-dodécyl β -D-maltoside (DM) a été réalisé comme publié par Boulant et coll. [30], à des fins de comparaison.

5.2.3.3 Fluorescence de C124 et C170

Les mesures de fluorescence ont été effectuées sur l'appareil « Cary Eclipse Fluorescence Spectrophotometer » à température ambiante dans une cuvette de 0,5 cm de parcours optique, utilisant le logiciel fourni avec l'appareil. La longueur d'onde d'excitation utilisée a été ajustée à 295 nm, pour se limiter seulement à la fluorescence des tryptophanes. Les spectres d'émission ont été récoltés de 300 à 400 nm. Les échantillons protéiques (C124 et C170) ont été préparés à une concentration de 10 μ M dans le « tampon RMN » avec ou sans ajout de TFE ou de SDS. La Core C82 a été omise pour la simple raison qu'elle ne possède pas de résidu tryptophane.

5.2.3.4 RMN

Toutes les expériences ont été enregistrées à l'université Laval sur un appareil Varian INOVA 600 à 25°C. Les spectres ont été transformés et analysés en utilisant le logiciel VNMRJ (Varian Inc.). Un spectre 1D proton de référence a été enregistré, pour la Core C170 à une concentration de \sim 1 mg/ml, et un spectre en présence de 30 % de TFE à une concentration de \sim 0,7 mg/ml. Les échantillons avec TFE ont été obtenus par dilution. Notons également, que 5 % de D₂O et 1 mM de DSS ont été ajoutés à tous les échantillons pour des fins de calibration et référencement. Pour créer les représentations graphiques, la suite logicielle Thermo Grams a été utilisée.

5.3 Résultats

5.3.1 Production et purification des formes C82, C124 et C170 de la Core du VHC

Tout projet d'analyse structurale de protéine requiert un ajustement très pointu des protocoles de production et de purification des protéines étudiées. C'est pourquoi tous les clones, avec lesquels nous avons travaillé, ont été optimisés pour une expression bactérienne chez *E. Coli*. Cette modification a permis d'approximativement doubler la quantité de protéine produite. Nous nous sommes également attardés à l'influence de la température sur la production protéique. Plusieurs températures ont été alors testées, et seulement deux ont été retenues : 22°C et 37°C. En effet, les clones C82 et C124 montrent une bonne surexpression à 22°C, alors que le clone C170 ne présente aucune surexpression notable (Figure 5.1, colonne 2, 5 et 8 respectivement). Notons par contre qu'à 37°C, les trois constructions montrent un haut niveau de surexpression (Figure 5.1, colonne 3, 6 et 9). C'est pourquoi les cultures bactériennes ont été alors réalisées à 37°C.

Un protocole de purification a été développé en parallèle pour les formes C82, C124 et C170 de la Core. Le protocole utilisé et publié [67] (voir chapitre 3) pour la forme C82 utilisait un tampon présentant un pH élevé (de 12). L'idée était d'utiliser un pH proche du point isoélectrique (pI=12,23, pour la forme C82) pour diminuer la charge de la protéine et ainsi minimiser les interactions possibles de la Core avec des contaminants cellulaires (protéines et acides nucléiques). Le protocole conçu ayant porté ses fruits, une approche similaire a été utilisée pour les clones C124 (pI=12,01) et C170 (pI=11,64), avec des tampons utilisés à pH 11.5 et 11 respectivement. Au cours de mon doctorat, j'ai eu plusieurs fois l'occasion de travailler à l'amélioration de ce protocole, en raison du besoin croissant de quantité de protéines pour moi-même et mes collègues pour nos projets respectifs. Ce processus aboutira finalement à la mise au point d'un protocole à deux étapes impliquant une colonne de nickel (Ni^{2+}) en condition dénaturante (8 M urée) et une colonne

de dessalage pour « renaturer / replier » la protéine, permettant au final d'obtenir des protéines pures (Figure 5.2) et fonctionnelles comme en atteste le test d'assemblage *in vitro* développé au sein du laboratoire (voir chapitre 7 + annexe 4).

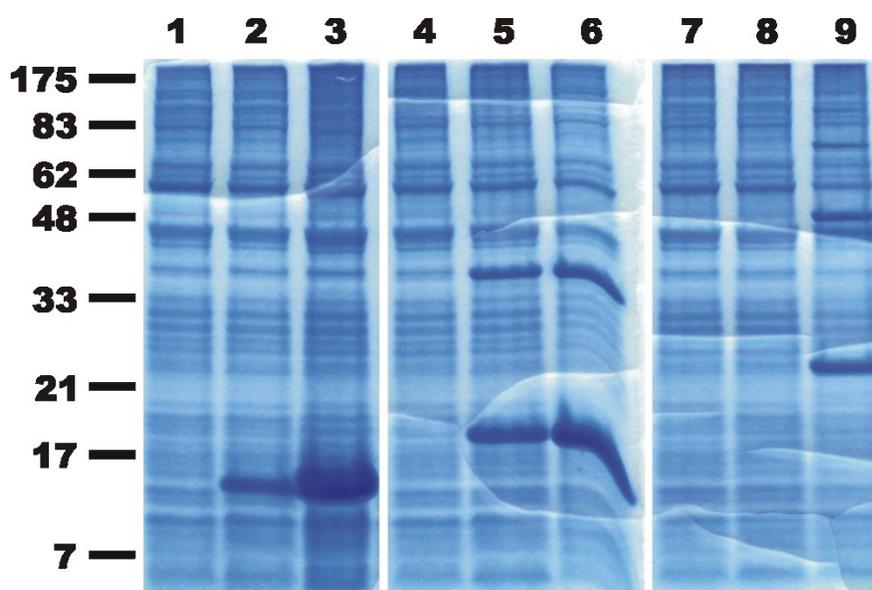


Figure 5.1 : Profil d'expression des diverses formes tronquées de la Core avant (1, 4, 7) et après induction à 22°C (2, 5, 8) et 37°C (3, 6, 9). Colonne 1 à 3 pour la Core C82, 4 à 6 pour la C124 et 7 à 9 pour la C170.

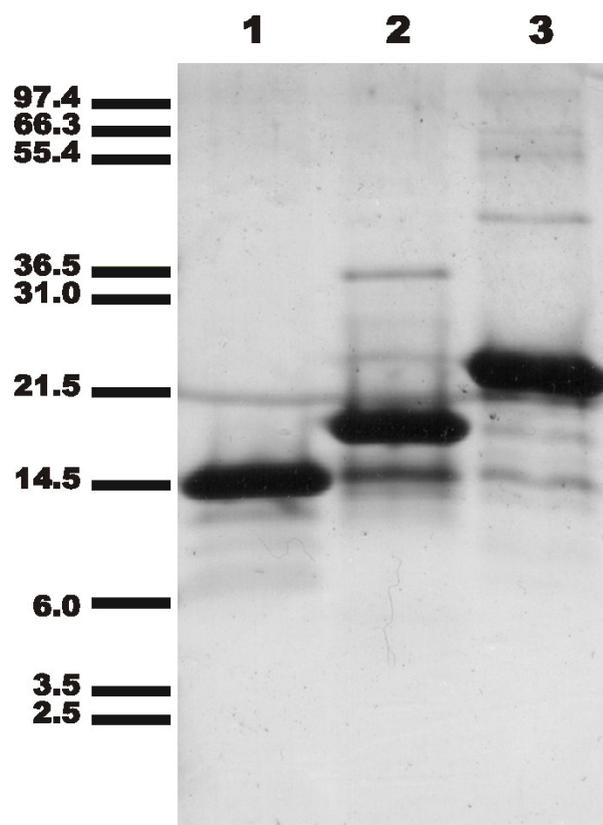


Figure 5.2 : Profil de purification après deux étapes, pour les formes tronquées de la Core, 1- C82, 2- C124 et 3- C170.

5.3.2 La protéine Core : une protéine multimérique ?

La protéine Core est une protéine qui a pour principale fonction l'encapsulation du génome viral. Il s'agit donc d'une protéine capable de s'associer à elle-même pour former la capsid. Nous avons donc vérifié l'état multimérique de chacune des versions tronquées de la Core. Pour cela, nous avons effectué une expérience dite de liaison croisée (cross-link) à l'aide d'un agent de liaison : le glutaraldéhyde. Les résultats présentés en Figure 5.3 sont sans équivoque, la Core C82 est monomérique, tandis que la forme C124 se présente sous la forme d'un dimère et enfin, la Core C170 semble quant à elle se comporter comme un multimère, dont la taille est supérieure à 200 kDa.

Des expériences de diffusion dynamique de la lumière (DDL) sont venues confirmer ces résultats. Nous avons mesuré les rayons hydrodynamiques (R_h) des trois formes tronquées de la Core que l'on retrouve dans le Tableau 5.1 et nous les avons comparés à des valeurs théoriques de R_h pour des protéines repliées ou non [257] (voir Tableau 5.2). La Core C82 présente de manière non surprenante un R_h très comparable à celui d'une protéine monomérique et non repliée, ce qui appuie nos précédents travaux publiés et présentés au chapitre 3 de cette thèse. La Core C124 possède quant à elle, un R_h de 56 Å comparable à celui d'une protéine dénaturée et dimérique. Par contre, le R_h mesuré pour le Core C170 ne nous permet pas d'affirmer de manière claire comme pour la C82 et la C124, de l'état multimérique de la protéine lorsque que l'on compare le R_h mesuré aux valeurs théoriques calculées. En effet si la C170 est une protéine totalement repliée, alors le complexe protéique serait de ~ 66 unités monomériques, dans le cas d'une protéine non repliée, le R_h mesuré se situe entre les valeurs calculées d'un dimère et d'un trimère. Cette dernière hypothèse semble improbable au vu des résultats de liaison croisée obtenus plus haut. Malgré tout, nous pouvons imaginer que la Core C170 multimérique est partiellement repliée, ce qui expliquerait de manière simple notre résultat.

Globalement, ces résultats sont comparables à ceux obtenus et publiés ultérieurement par M. Kunkel [147]. Il faut noter toutefois que dans cette étude, les protéines étudiées étaient la Core C124 et C179. Pour cette dernière, leurs travaux

suggèrent que la protéine serait sous la forme d'un multimère d'environ 24-26 unités monomériques. Néanmoins, pour le clone C124, nos résultats corroborent ceux publiés par M. Kunkel, sur l'état dimérique de la Core C124 en absence d'agent réducteur. Il est à noter que M. Kunkel a démontré que la formation de ce dimère était très certainement dû à un pont disulfure, qui impliquerait la cystéine en position 91, la seule incluse dans le clone C124. Par contre, Boulant et coll. [30], quant à eux, ne suggèrent pas l'implication d'un pont disulfure dans la formation d'un dimère ni dans le cas de la forme tronquée C124 ni C169. Or de récents travaux proposent la formation d'un dimère de le Core de l'hépatite C impliquant la cystéine en position 128 [148].

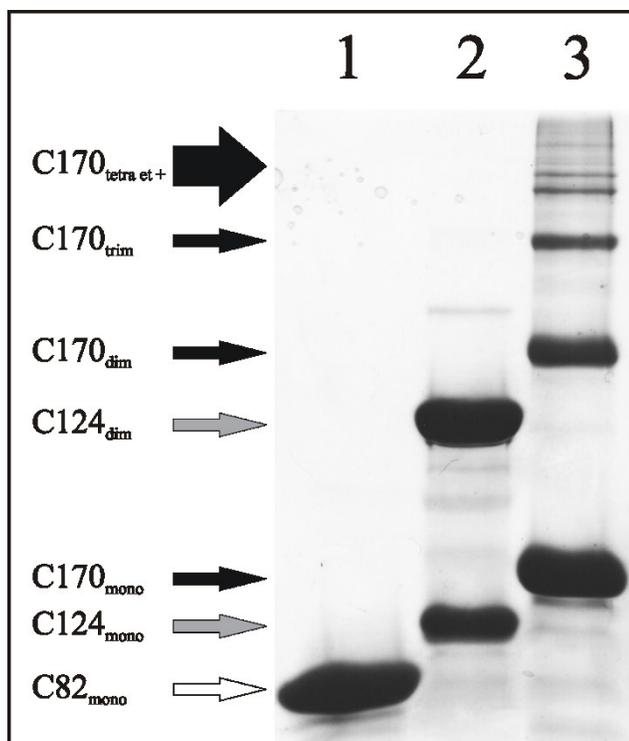


Figure 5.3 : Résultats de liaison croisée au glutaraldéhyde pour les clones C82 (colonne 1), C124 (colonne 2) et C170 (colonne 3).

Tableau 5.1 : Résumé des valeurs de rayon hydrodynamique (R_h) mesuré pour les trois formes tronquées de la Core de l'hépatite C.

Protéines	R_h mesurés par Zétasizer (Å)
C82	28
C124	56
C170	72

Tableau 5.2 : Valeurs calculées du rayon hydrodynamique (R_h) pour diverses formes multimérique de la Core de l'hépatite C en fonction de son état de repliement selon les travaux de Wilkins et coll. [257]. Les cases colorées correspondent aux états multimériques suspectés pour les trois constructions d'après les valeurs mesurées répertoriées au Tableau 5.1.

C82	Nombre de résidus	R_h théorique (Å) si natif	R_h théorique (Å) si dénaturé
1	89	17,46	28,55
2	178	21,35	42,38
3	267	24,01	53,40
4	356	26,10	62,91

C124	Nombre de résidus	R_h théorique (Å) si natif	R_h théorique (Å) si dénaturé
1	131	19,53	35,58
2	262	23,88	52,82
3	393	26,86	66,56
4	524	29,19	78,42

C170	Nombre de résidus	R_h théorique (Å) si natif	R_h théorique (Å) si dénaturé
1	177	21,31	42,24
2	354	26,06	62,71
3	531	29,31	79,01
4	708	31,86	93,09
6	1062	35,83	117,30
8	1416	38,95	138,20
44	7788	63,86	365,18
66	11682	71,82	460,13

5.3.3 Analyse structurale par dichroïsme circulaire, fluorescence et RMN

La forme C82 faisant l'objet des chapitres 3 et 4, nous nous concentrerons dans les paragraphes à venir sur les formes tronquées C124 et C170.

Les protéines C124 et C170 purifiées ont été étudiées par dichroïsme circulaire pour en déterminer leurs contenus en structure secondaire. La protéine C124 présente un spectre typique de protéine non repliée (Figure 5.4A), avec un minimum à 200 nm, similaire à celui obtenu pour la Core C82 (voir chapitre 3). Le signal CD de la Core C170 quant à lui arbore un spectre d'une protéine agrégée et/ou repliée partiellement avec un minimum proche de 203 nm et un épaulement vers 220 nm. Le contenu en structure secondaire estimé est de 26 % d'hélice alpha, 33 % de brins beta et 42 % d'autres structures. Ces résultats sont comparables à ceux obtenus et publiés par M. Kunkel [147], pour des protéines de taille similaire (C124 et C179).

Nous avons par ailleurs enregistrés des données de fluorescence. En effet, cette technique est utilisée couramment pour étudier la structure, le repliement des protéines, par l'entremise de résidus aromatiques tels que le tryptophane. La Core du VHC possède cinq tryptophanes situés dans sa région centrale en position : W₇₆, W₈₃, W₉₃, W₉₆ et W₁₀₇. Cette partie de la Core (de l'acide aminé 76 à 107), serait une des régions impliquées dans la dimérisation de la Core [141]. Le résidu tryptophane possède la caractéristique d'émettre un signal qui varie en fonction de son environnement, qu'il soit hydrophile ou hydrophobe, ce que l'on peut traduire par accessible ou non au solvant. Le tryptophane seul en solution émet à une longueur d'onde d'émission maximale (λ_{\max}) de ~ 350 nm. Cette valeur de λ_{\max} est significative d'un tryptophane totalement accessible au solvant. Inversement, un tryptophane enfoui dans un milieu hydrophobe aura une λ_{\max} proche de 333 nm [64].

Dans les cas qui nous concernent, nous pouvons affirmer que la Core C124 possède des tryptophanes largement exposés au solvant avec une λ_{\max} de 352 nm (Figure 5.4B). Ce résultat paraît un peu étonnant, car si la région riche en tryptophane était le centre principal de l'interaction protéine - protéine, alors nous aurions dû obtenir une λ_{\max} qui aurait dû être

déplacée vers le bleu (longueur d'onde plus petite), car ceux-ci auraient été impliqués dans la formation du dimère, et donc moins accessibles au solvant. Dans le cas de la Core C170, la λ_{max} mesurée est de 342 nm (Figure 5.4B). Sachant, que la Core C170 forme un large multimère, nous nous serions attendus là encore à ce que les tryptophanes soient cachés du solvant. Ces résultats quelque peu surprenants sont tout à fait comparables à ceux obtenus par M. Kunkel et sont une indication que la protéine Core dans les conditions d'étude utilisées n'est peut-être pas repliée convenablement.

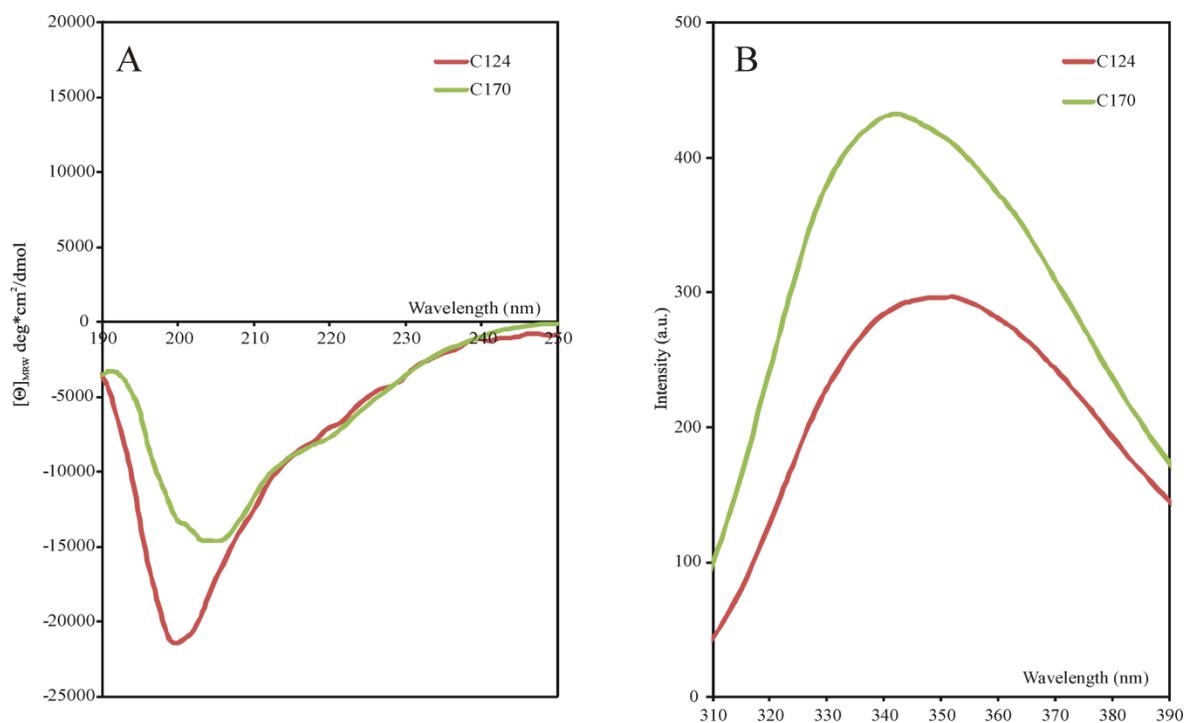


Figure 5.4 : Spectres des formes C124 et C170. A : Spectre de dichroïsme circulaire, B : Spectre d'émission de fluorescence des tryptophanes.

5.3.4 Effet du TFE et du SDS sur la structure de la Core C124 et C170

Dans le but de trouver des conditions d'études structurales adéquates, nous avons en parallèle vérifié l'effet du TFE ou de micelles de SDS sur la structure adoptée de la protéine et tenté de vérifier l'effet de ces additifs sur l'état multimérique de la protéine par dichroïsme circulaire, fluorescence et RMN. En effet, ces deux produits sont fréquemment utilisés dans des études structurales comme lipomimétique, mais également pour mimer des environnements favorables aux interactions protéine - protéine.

Pour le clone C124, nous avons ajouté jusqu'à 50 % de TFE ou jusqu'à 150 mM de SDS (Figure 5.5, panneaux A et B). À la fois, l'ajout de TFE et de SDS ont tendance à entraîner un léger déplacement du minimum vers 203 nm et une légère accentuation du creusement présent à 220 nm, significatif d'un potentiel début de repliement en hélice α . Mais dans les deux cas, aucun changement drastique de la structure secondaire n'est observé. Au contraire, l'ajout de TFE ou de SDS à la Core C170 entraîne un changement significatif des spectres CD (Figure 5.5, panneaux C et D). Ceci est attribuable à une augmentation de la proportion de structure secondaire au sein de la C170 et notamment d'hélices α , car les spectres possèdent deux minimums locaux, un à 208 nm et un autre à 222 nm. Que ce soit avec le TFE ou le SDS, les changements sont observables dès les plus faibles ajouts, ce qui signifie très clairement que ces produits viennent stabiliser des éléments structuraux présents dans la structure native de la protéine. Ces résultats viennent corroborer ceux présentés par S. Boulant et coll. [30] avec la Core C169 dans 0,1 % de *n*-dodécyl β -D-maltoside (DM), que l'on a reproduit avec le clone C170 (Figure 5.6). Notons, qu'en présence de 0,1 % DM, la Core C169 contient ~50 % d'hélice α et est sous forme dimérique, ce qui est très comparable aux ~46-48 % d'hélice α estimés pour la C170 dans 30 % de TFE ou 50 mM SDS. Il reste néanmoins à vérifier l'état multimérique de la Core C170 en présence de ces additifs.

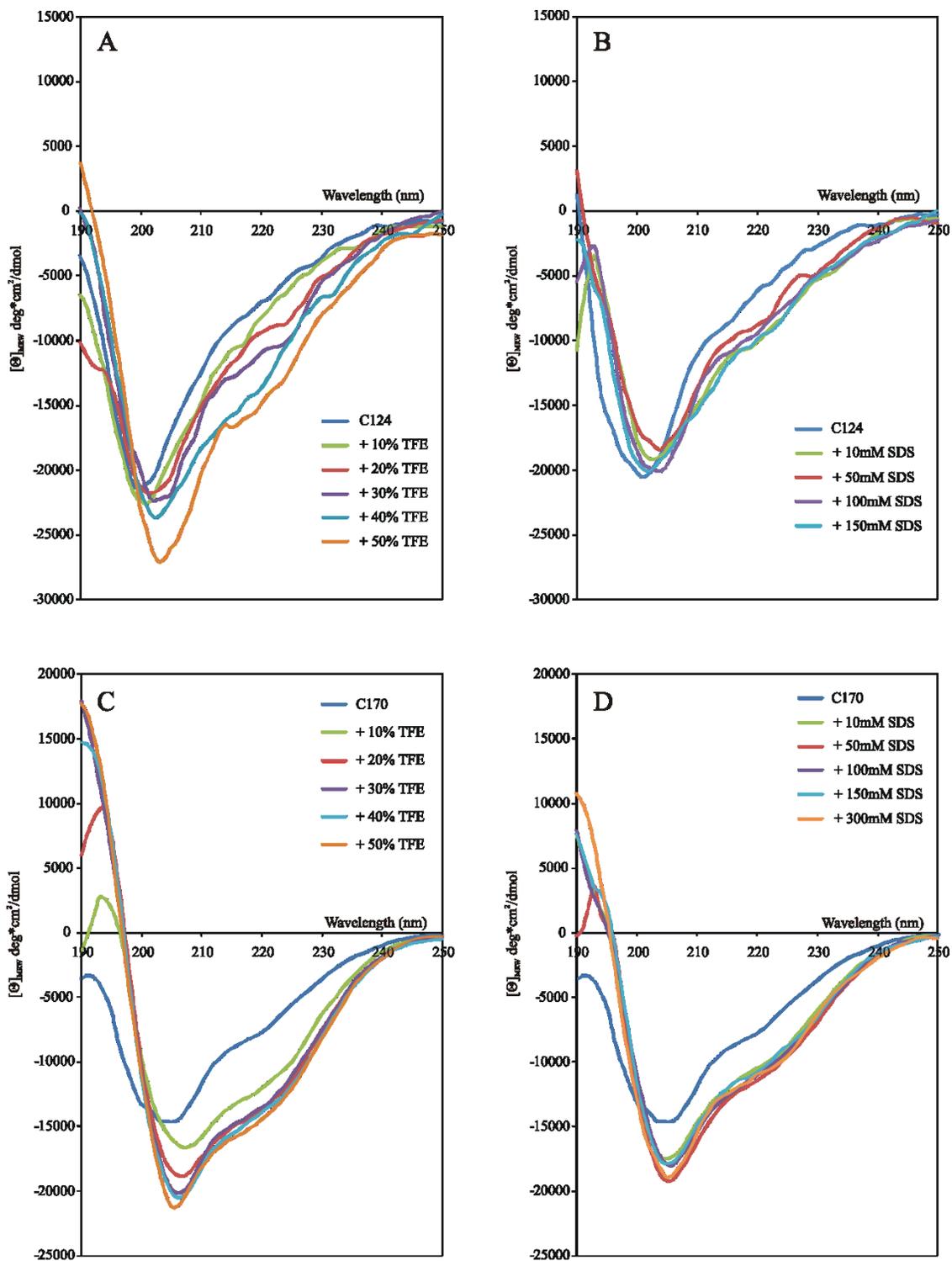


Figure 5.5 : Spectres de dichroïsme circulaire des formes tronquées C124 (A et B), et C170 (C et D), en présence de quantité croissante de TFE (A et C) et SDS (B et D).

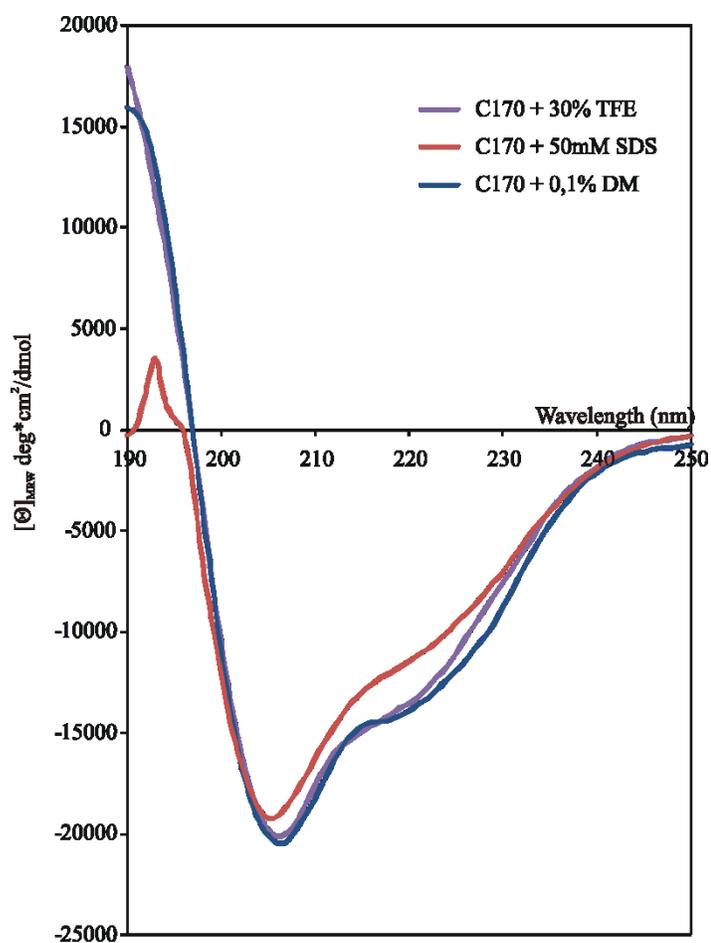


Figure 5.6 : Spectres de dichroïsme circulaire de la Core C170 en présence de 30 % TFE (violet), SDS (en rouge) et de DM (en bleu).

Si l'on s'attarde aux données de fluorescence, on observe de manière intéressante que les deux produits ont des effets opposés. En effet, que ce soit pour la Core C124 ou C170, le SDS engendre un déplacement de la λ_{\max} vers une valeur commune de 338 nm, quelle que soit sa concentration (Figure 5.7 B et D). Ceci a pour signification, que les tryptophanes se retrouvent dans un milieu plus hydrophobe. Ceci peut être dû au fait que le SDS favorise la formation d'un dimère ou que les micelles de SDS viennent se « lier » à la Core au niveau de cette région. Il en ressort donc que la région riche en tryptophanes est une clé maîtresse au bon repliement de la Core, lorsque que le C-terminal est présent.

Tout comme dans le cas du SDS, l'ajout de TFE aux formes tronquées C124 et C170 montrent une nouvelle fois que les tryptophanes de la Core réagissent de la même façon à cet ajout. Au fur et à mesure des additions de TFE, la λ_{\max} tend à se rendre vers une valeur commune proche de 347-348 nm. Il paraît évident que le TFE n'a pas le même effet sur la région riche en tryptophanes que le SDS malgré un effet similaire sur le repliement pour la Core C170 en particulier. Contrairement au SDS, l'ajout de TFE à la C170 semble rendre les tryptophanes plus accessibles au solvant, avec un déplacement de la λ_{\max} vers des longueurs d'ondes plus élevées et proches de la valeur critique de 350 nm, où l'on suppose les tryptophanes dans un environnement plus hydrophile. Bien que ces expériences ne puissent nous donner avec certitude l'effet de ces divers produits sur l'état multimérique de la C170, nous pouvons imaginer à partir des valeurs de λ_{\max} obtenues et des spectres CD, qu'en plus de replier la Core en hélice α , le TFE possède certainement la capacité de briser le multimère pour engendrer une protéine peut-être monomérique. Cette spéculation pourrait être également vraie dans le cas du SDS. C'est en tout cas ce que nous suggèrent les spectres 1D du ^1H de la C170 en absence et en présence de 30 % de TFE (Figure 5.8 A et B). Dans un premier temps, nous remarquons un élargissement de la dispersion spectrale dans la dimension ^1H , significatif d'un gain de structure secondaire de la Core, et une diminution de la largeur des raies, relatif à une diminution de la taille de la protéine en solution. D'autre part, il est important de noter que la concentration de la Core C170 dans 30 % de TFE est de 0,7 mg/ml vs. 1 mg/ml pour la référence. Au vu des spectres 1D, il paraît évident malgré une concentration plus faible, que le signal global de la Core C170 dans 30 % TFE est aussi, voire plus intense que celui de la référence, nous amenant une

fois de plus à supposer que la Core C170 se trouve dans un état multimérique moindre en présence de TFE.

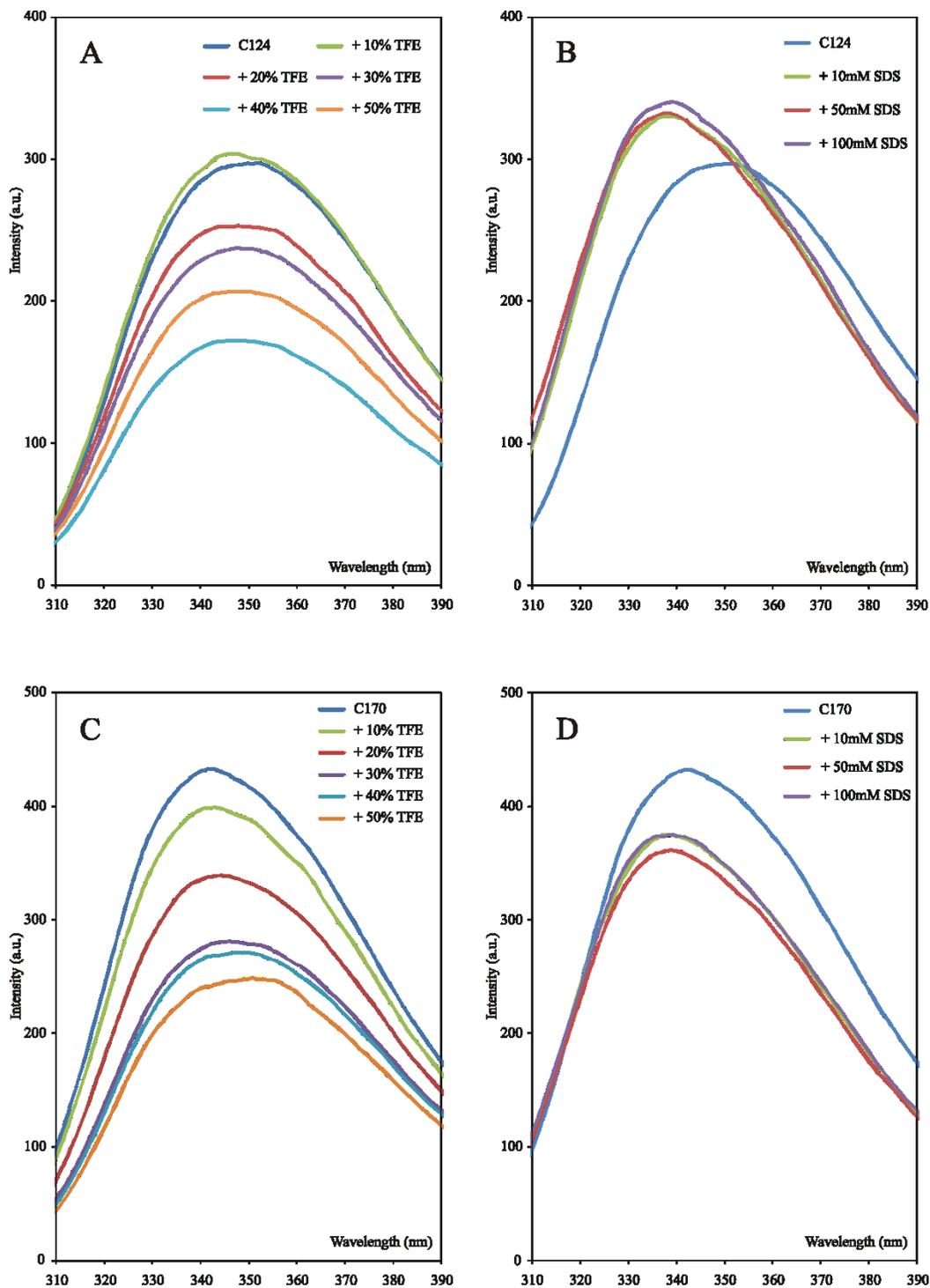


Figure 5.7 : Spectres d'émission de fluorescence des tryptophanes pour les formes C124 (A et B) et C170 (C et D) en absence et en présence de quantité croissante de TFE (A et C) et SDS (B et D).

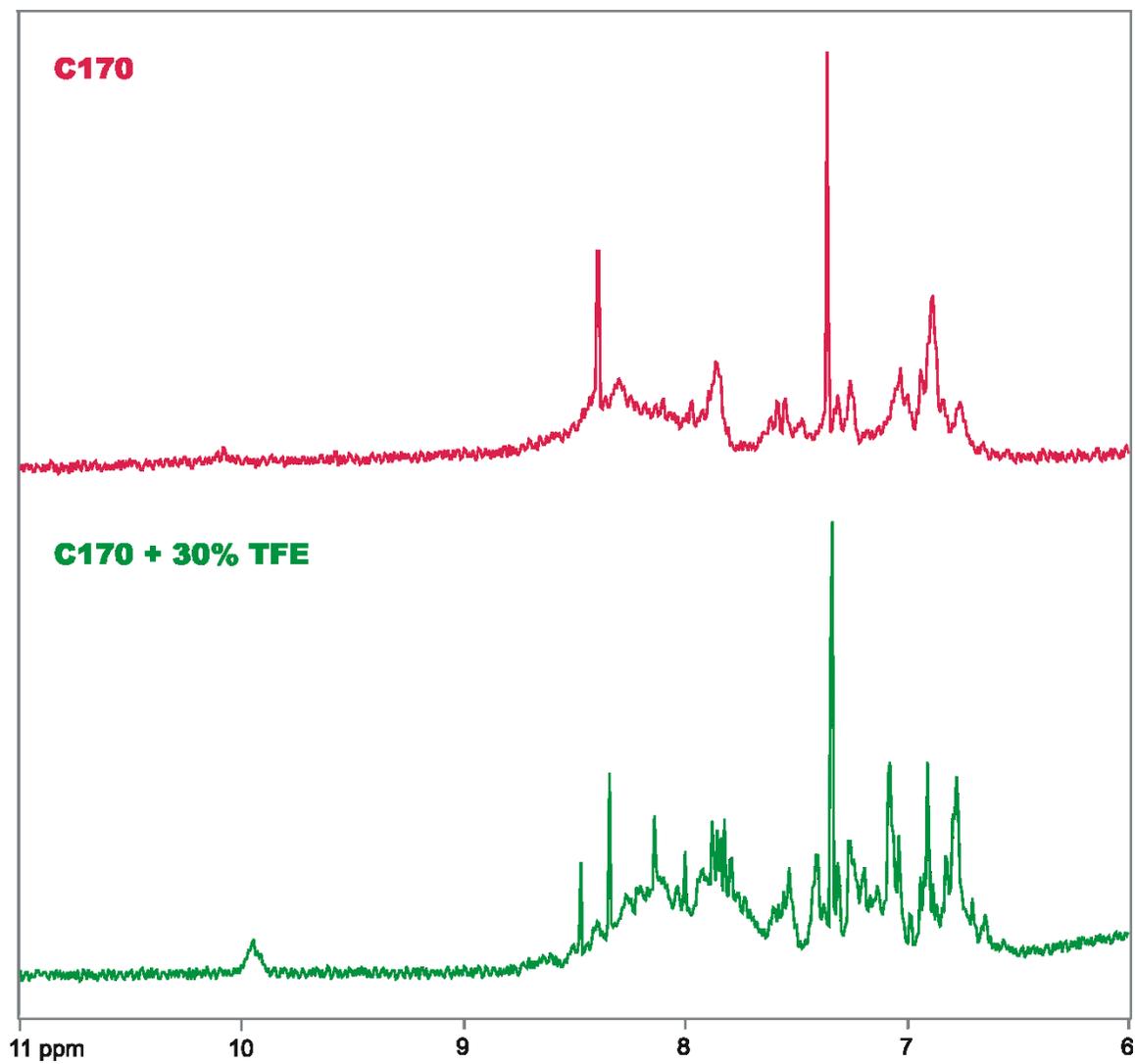


Figure 5.8 : Spectre 1D du proton de la protéine Core C170 en absence et en présence de 30 % TFE.

5.3.5 Conclusion

En conclusion, ces travaux préliminaires sont très encourageants. En effet, nous avons mis en place un protocole robuste de production et purification des diverses formes tronquées de la Core, avec l'obtention, en routine, de plusieurs milligrammes de protéines pures (10 à 20 mg par litre de culture selon la construction). Les résultats structuraux obtenus et particulièrement ceux pour la forme tronquée « C170 » sont très prometteurs. En effet, la Core adopte une structure en hélice α que ce soit en présence de micelles de SDS ou en présence de faible quantité de TFE. La Core est une protéine que l'on retrouve principalement au réticulum endoplasmique et aux gouttelettes lipidiques ce qui rend le choix de ces agents lipomimétiques (SDS et TFE) très intéressant. La poursuite de ce projet nécessitera d'approfondir la compréhension de l'effet de ces agents lipomimétiques sur la structure et l'état multimérique de la Core. Une fois, les réponses obtenues, la détermination de la structure tridimensionnelle de la protéine pourra être envisagée, que ce soit par résonance magnétique nucléaire ou par diffraction des rayons X.

Chapitre 6

Interaction des protéines Core (C82) et p53

6.1 Introduction

Nous avons vu au chapitre 3, que la moitié N-terminale de la Core du virus de l'hépatite C appartenait à la famille des protéines non structurée de manière native ou « Intrinsically Unstructured Protein ». Toutefois, la protéine Core C82 est capable d'adopter une structure en hélice α comme nous l'avons vu au chapitre 4. En effet, les « IUPs » sont des protéines qui au contact d'un ligand peuvent acquérir un certain degré de repliement. C'est dans cette optique, que nous avons dans un premier temps voulu étudier l'effet de l'ajout d'un acide nucléique (partenaire indispensable de la Core) sur le repliement adopté par la Core C82. Or il s'est avéré que le mélange *in vitro* d'un acide nucléique et de la Core de l'hépatite C était accompagné de l'apparition de turbidité, rendant l'étude structurale impossible. Toutefois, nous utiliserons cette propriété particulière à des fins d'étude d'inhibition de l'encapsidation *in vitro* (voir chapitre 7 et annexe 4 [82]). Nous nous sommes donc intéressé dans un deuxième temps aux divers partenaires protéiques de la Core. Le choix s'est porté sur la protéine p53, pour plusieurs raisons. La première raison était bien évidemment le fait que p53 interagit avec la Core dans sa région N-terminale, comme l'ont démontré Kao et coll. [124]. La deuxième est l'importance de cette protéine dans le cycle cellulaire et le rôle de cette interaction dans la pathogénicité du VHC. En effet, p53 est une protéine humaine d'importance capitale car impliquée dans la régulation de nombreux gènes et particulièrement ceux dirigeant la division cellulaire, et l'apoptose [59, 210]. De plus, l'inactivation de p53 par des délétions, mutations, ou interactions avec des protéines virales ou cellulaires est un facteur clé du développement de près de 50 % des cancers humains [47]. En effet, plusieurs protéines du VHC ont un potentiel oncogénique, et la Core est l'une d'entre elles. Il a été démontré que celle-ci pouvait avoir un effet inhibiteur ou activateur sur la fonction de p53; malgré tout, il reste que l'effet de la Core sur le fonctionnement de p53 est à ce jour peu connu et semble controversé [125].

p53 est une protéine de 393 résidus, que l'on peut diviser en trois domaines principaux, le domaine de transactivation (TAD) (aa 1-93), le domaine de liaison à l'ADN (DBD) (aa 94-312), et le domaine de tétramérisation (TD) (aa 313-393). Le domaine de

tétramérisation peut à son tour être divisé en deux sous-parties, la région strictement impliquée dans la multimérisation (aa 325-355) possédant un brin β et une hélice α et le domaine C-terminal (aa 360-393) non structuré étant le siège de diverses interactions et modifications post-traductionnelles [44, 124].

Lors de l'élaboration de ce projet, nous avons contacté la Dre Cheryl Arrowsmith de l'Université de Toronto (collaboratrice du Dr. Stéphane Gagné sur le projet de protéomique structurale de « *Methanobacterium Thermoautotrophicum* » [120]) qui possédait les clones du domaine de tétramérisation nécessaires au projet, qu'elle nous a fourni gracieusement.

Je tiens à préciser que les travaux présentés dans ce chapitre sont préliminaires.

6.2 Matériels et méthodes

6.2.1 Protocole de production et purification de C82

Les échantillons de Core C82 uniformément marqués au ^{15}N utilisés pour l'enregistrement des spectres RMN ont été exprimés, purifiés et préparés de la même façon que décrit précédemment [67].

6.2.2 Protocole de production et purification des diverses formes de p53

E. coli BL21(DE3) pLysS (Invitrogen) a été transformée par chacun des trois clones de p53 (monomère, dimère et tétramère) obtenus de la Dre Cheryl Arrowsmith. Les cultures ont été effectuées en flasque de 2 litres à 37°C dans un milieu 2XYT (pour 1 L, 16 g de tryptone, 10 g d'extrait de levure, 5 g de NaCl, et le pH ajusté à 7.2) en présence de 50 µg/ml d'ampicilline. La surproduction protéique a été induite par l'ajout de 1 mM d'IPTG (isopropyl-1-thio-β-D-galactopyranoside) lorsqu'une D.O.₆₀₀ de 0,6 a été atteinte. Après trois heures d'induction, les cultures ont été centrifugées à 4000 x g pendant 30 minutes. Une fois le culot bactérien resuspendu dans un tampon 8 M Urée, 20 mM Tris, 300 mM NaCl, pH 7,4 + 1x cocktail d'inhibiteurs, celui-ci a été soumis à une étape de sonication dans le but de briser les bactéries. Le lysat a été par la suite clarifié par centrifugation à 27000 x g pendant 20 min à 4°C et mélangé à une résine de nickel (Ni-NTA resin (QIAGEN)) pour y fixer les protéines produites via leurs queues d'histidines. Dans le but de se séparer de contaminants liés de manière non spécifique à la résine, nous avons effectué des lavages avec les tampons suivants (tampon 1 : 50 mL de 8 M Urée, 20 mM Tris, 300 mM NaCl, pH 7,4 ; tampon 2 : 50 mL de 4 M Urée, 20 mM Tris, 300 mM NaCl, 25 mM Imidazole pH 7,4 ; tampon 3 : 15 mL de 2 M Urée, 20 mM Tris, 300 mM NaCl, 25 mM Imidazole pH 7,4). Les protéines ainsi lavées ont été éluées dans un tampon 2 M Urée, 20 mM Tris, 300 mM NaCl, 500 mM Imidazole, pH 7,4 et appliquées sur une colonne de dessalage équilibrée dans un tampon 25 mM phosphate, 50 mM NaCl, pH 6.6 (tampon

« RMN »). La queue de poly-histidine a été ensuite clivée par digestion à la thrombine (1 U/mg à température ambiante pendant trois heures). Les protéines ainsi clivées ont été séparées de la thrombine et de la queue de poly-histidine par un passage sur des colonnes de benzamidine et de nickel mises en série. Les protéines ainsi purifiées ont été concentrées dans une cellule à agitation (Amicon Stirred-cell). La pureté des protéines a été vérifiée par électrophorèse sur gel de poly-acrylamide (SDS-PAGE). Au final, quelques milligrammes de protéines pures ont été obtenus, et au besoin, les protéines ont été lyophilisées après dialyse contre de l'eau.

6.2.3 Dichroïsme circulaire

Les spectres de dichroïsme circulaire ont été enregistrés sur un spectropolarimètre Jasco J-710 à température ambiante. Les expériences pour l'ensemble des trois protéines ont été réalisées dans le dit « tampon RMN » à des concentrations variant de 0,8 mg/ml à 1,1 mg/ml. Les enregistrements ont été faits en utilisant une cuvette de quartz possédant un trajet optique de 1 mm. Les spectres finaux consistent en la moyenne de dix enregistrements indépendants ensuite convertis en ellipticité molaire moyenne par résidu ($[\theta]_{MRW}$).

6.2.4 RMN

Toutes les expériences ont été enregistrées à l'université Laval sur un appareil Varian INOVA 600 à 5°C. Des spectres 2D ^{15}N -HSQC de la Core C82 marquée au ^{15}N ont été enregistrés en absence (référence) et en présence de p53 monomérique ou dimérique avec un ratio C82 : p53 de 1 : 4 (mM : mM) pour le monomère et 1 : 8 (mM : mM) pour le dimère. Les spectres HSQC avaient pour largeur spectrale 11990 Hz dans la dimension ^1H et 1944 Hz dans la dimension ^{15}N pour les expériences avec le monomère de p53 et 12001 Hz en ^1H et 1944 Hz en ^{15}N pour les expériences avec le dimère. Les spectres ont été transformés dans NMRPipe [60] et analysés dans NMRView [117].

6.3 Résultats

Dans le but de tester l'hypothèse que la Core C82 peut adopter une structure secondaire au contact de la région C-terminale de p53, nous avons produit et purifié le domaine C-terminal de la protéine p53 sous ses formes monomérique, dimérique et tétramérique. Les formes dimériques et monomériques ont été obtenues par des mutations spécifiques au sein de la forme native tétramérique [58]. Le dimère possède ainsi deux mutations en position 340 (méthionine pour glutamine) et 344 (leucine pour arginine), quant au monomère, lui, ne présente qu'une seule mutation en position 344 (leucine pour proline). La purification des trois formes de p53 a été très efficace, comme en témoigne le gel en Figure 6.1. Nous avons également évalué et contrôlé par dichroïsme circulaire le repliement des trois protéines (p53 mono, p53 dim et p53 tet). Alors que les formes, dimérique et tétramérique, montrent des spectres typiques de protéines repliées, le monomère (clone LP344) arbore quant à lui un spectre CD typique d'une protéine non structurée avec un minimum à 200 nm (Figure 6.2). Ces résultats viennent donc confirmer le repliement adopté et attendu par nos trois protéines [59, 150, 169].

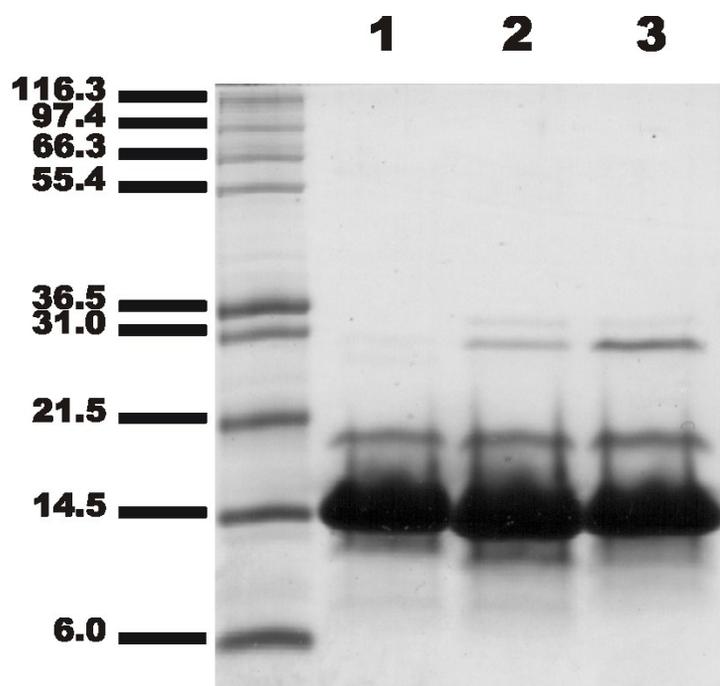


Figure 6.1 : Profil de purification des formes monomérique (colonne 1), dimérique (colonne 2) et tétramérique (colonne 3) de la protéine p53 après deux étapes de purification

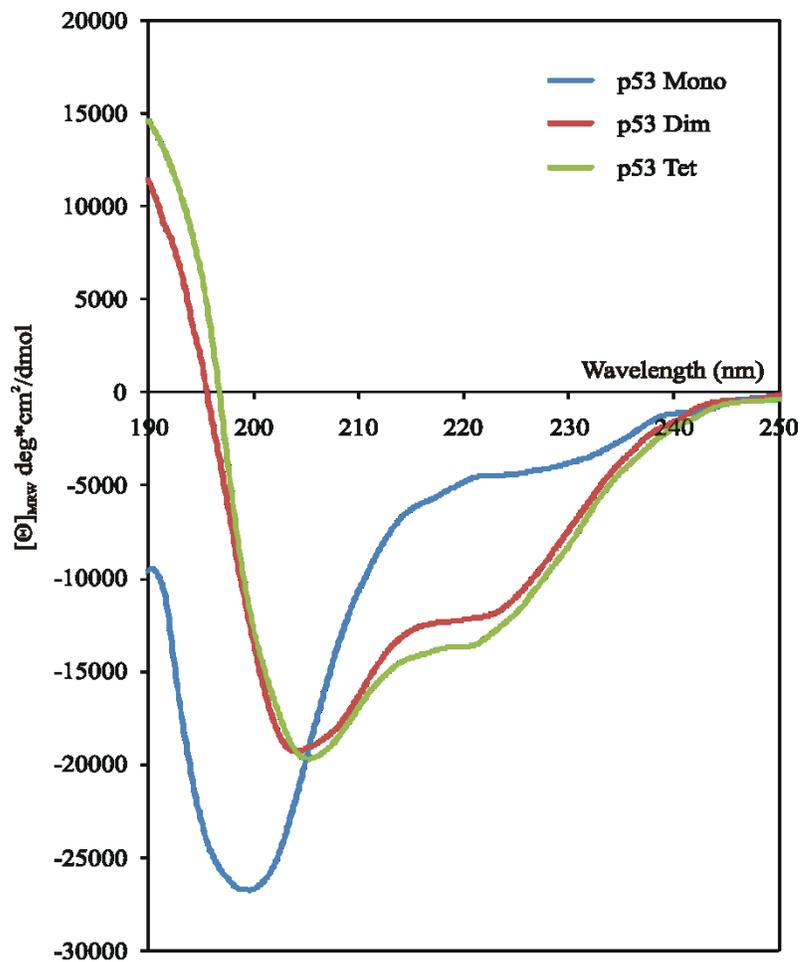


Figure 6.2 : Spectres de dichroïsme circulaire des protéines p53 monomérique, dimérique et tétramérique. La protéine monomérique possède un spectre typique de protéine non repliée, contrairement aux deux autres constructions qui ont un spectre typique de protéine repliée en hélice alpha.

Une fois les diverses protéines purifiées, et dans le but de vérifier notre hypothèse, nous avons enregistré une série de spectre 2D ^1H , ^{15}N -HSQC de la Core C82 marquée au ^{15}N , en absence et en présence de p53 monomérique ou p53 dimérique non marquée. En effet, comme nous l'avons vu précédemment le spectre 2D ^1H , ^{15}N -HSQC est l'outil de choix en RMN permettant à la fois de mettre en évidence le repliement d'une protéine ainsi que son interaction avec un partenaire, par la simple comparaison de deux spectres. Nous avons ainsi choisi de travailler en excès de p53 (ratio C82 : p53 de 1 : 4) (Figure 6.3), pour être certain d'observer le contact entre les deux partenaires.

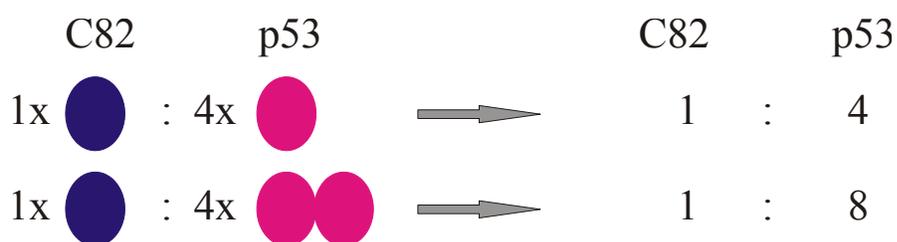


Figure 6.3 : Représentation graphique des ratios utilisés lors des expériences RMN d'interaction Core C82 – p53.

Le spectre de la Core C82 seule témoigne de l'absence de repliement car il exhibe une dispersion spectrale réduite (inférieure à 1 ppm) dans la dimension du ^1H comme nous l'avons vu au chapitre 3 (Figure 6.4 A). Le repliement de la Core au contact d'un partenaire devrait normalement être accompagné d'un élargissement de la dispersion spectrale dans la dimension ^1H . Or si l'on compare les dispersions spectrales dans la dimension du ^1H en absence et en présence de p53 (monomérique ou dimérique), nous ne remarquons aucune différence (Figure 6.4 AB et 6.5 AB). Cela signifie donc que la Core C82 ne gagne vraisemblablement pas de structures secondaires en présence de p53. Par contre, ces expériences nous ont permis de prouver que l'interaction physique Core – p53 se faisait de manière directe entre les deux protéines, car jusqu'à ce jour aucune preuve de cette interaction directe n'avait été apportée, car seules des techniques indirectes avaient été réalisées [124].

La comparaison des spectres en absence et en présence de p53 (monomérique ou dimérique) dévoile plusieurs caractéristiques permettant d'affirmer l'existence d'une interaction physique directe telle qu'une diminution globale du signal RMN, la disparition ou l'apparition de certaines résonances et la perturbation de déplacements chimiques. Les figures 6.4 C et 6.5 C résument, respectivement, les variations observées après l'ajout de p53 monomérique et p53 dimérique. Nous avons mesuré une diminution moyenne d'intensité des spectres d'un facteur 2,5 dans le cas de la p53 monomérique et d'un facteur 3,3 pour la p53 dimérique. Cette diminution importante de signal et la disparition de certaines résonances sont dues très certainement au fait que nous nous trouvons dans une situation d'interaction faible entre les partenaires protéiques (interaction de l'ordre du micromolaire [207] et que la C82 et la p53 (monomère ou dimère) sont en échange entre des formes libre et complexée. Par ailleurs, la formation du complexe C82-p53 (monomérique ou dimérique) va diminuer les mouvements de rotation globaux de la C82 et ainsi affecter sa relaxation transversale (T_2) résultant en une diminution du signal RMN, et c'est ce que l'on observe dans les deux cas.

L'interaction C82 : p53 monomérique a lieu au cœur de la région hydrophobe de la Core C82. En effet, cinq résonances disparaissent du spectre (proton HN des acides aminés G26, I30, G32, G33 et L36), sept sont affectées de manière importante présentant une diminution de leur signal supérieure à la moyenne (2,2) (Figure 6.4). Cinq parmi ces sept présentent une légère variation de leur déplacement chimique. Par ailleurs nous noterons un effet moindre au niveau des régions flanquantes en N et C de la région d'interaction. En effet, les résidus 20 à 24 et 39 à 70 montrent un grand nombre de résidus dont l'intensité du signal diminue singulièrement (diminution supérieure à la médiane (~2)).

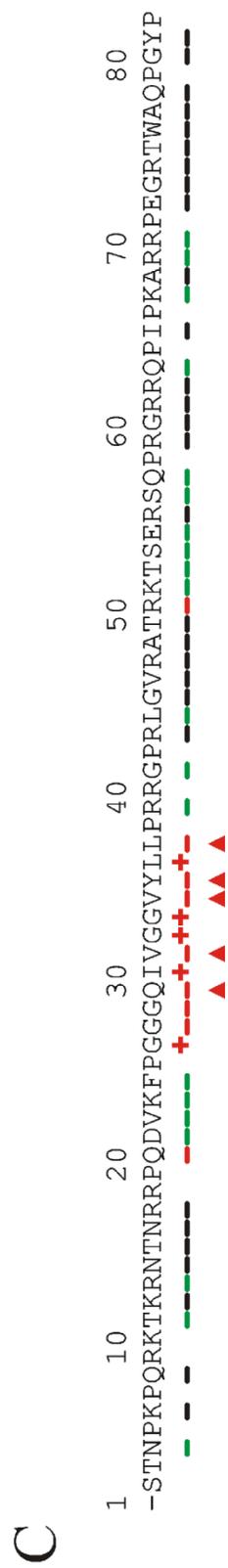
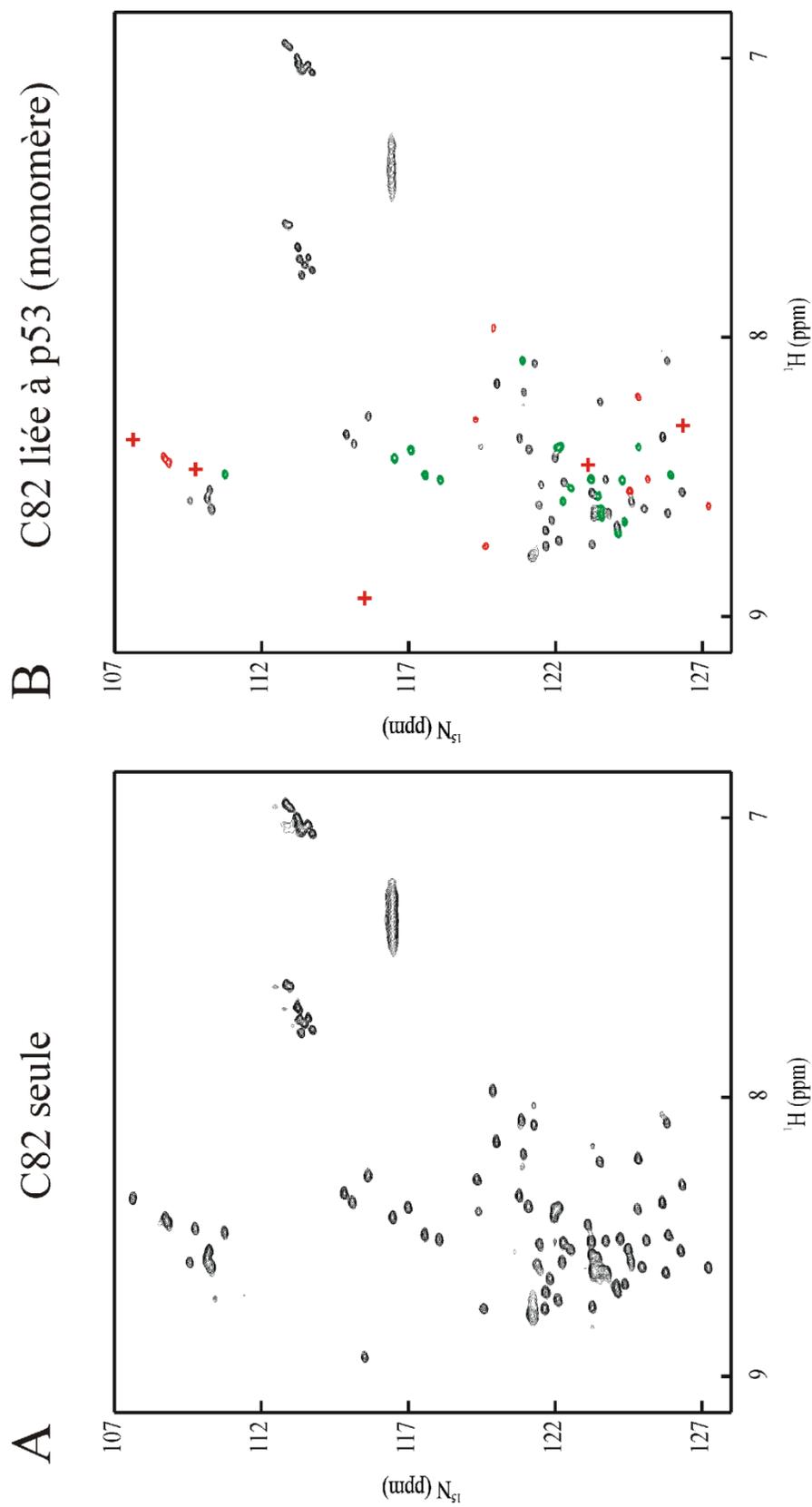


Figure 6.4 : Résumé des données d'interaction entre la Core C82 et p53 monomérique obtenues par résonance magnétique nucléaire. A) Spectre ^1H , ^{15}N -HSQC de la Core C82 seule. B) Spectre ^1H , ^{15}N -HSQC de la Core C82 en présence de p53-mono (ratio 1 :4). Un symbole « + » est utilisé pour indiquer les résonances ayant disparu. De plus, un choix de couleur a été apposé pour mettre en valeur l'effet de l'interaction sur l'intensité des résonances. Le rouge représente les résonances dont la diminution de l'intensité est supérieure à la moyenne, le vert quant à lui reflète celles dont la diminution est supérieure à la médiane et en noir, celles dont la diminution est inférieure à la médiane. C) Séquence de la Core C82 sur laquelle les effets mesurés et observés ont été reportés. Les triangles indiquent des résonances dont le déplacement chimique a changé en raison de la présence de p53-mono. Le code couleur reste identique à celui de la figure B).

Quant au dimère de p53, il lie la Core C82 dans sa région C-terminale, comme en atteste la Figure 6.5. En effet, c'est entre les résidus 41 et 82 que la plus grande partie des changements ont lieu. Quatre résidus disparaissent (résidus 67 à 70), quinze voient leur intensité diminuer drastiquement (valeur supérieure à la moyenne (3,3)), dix-sept présentent une modification de leur déplacement chimique, six sont en processus de dédoublement (acides aminés 51, 52, 53, 55, 56, 57) et neuf nouvelles résonances apparaissent. Ces résultats indiquent clairement que nous sommes dans une situation où les partenaires oscillent entre des formes libre et complexée. Cette idée est appuyée par le dédoublement clair des résonances des résidus 51 à 57, signifiant qu'ils sont dans deux environnements différents. Encore une fois, ces variations en présence de p53 dimérique témoignent d'une interaction faible entre les deux protéines. On pourrait se risquer à dire que la diminution du signal RMN mesuré est relative à l'augmentation de taille de la protéine C82 qui passerait de 10 kDa (libre) à 30 kDa lorsque en complexe avec un dimère de p53 (facteur de 3). Cette remarque est tout aussi valable pour l'interaction de la Core avec la p53 monomérique où le poids moléculaire du complexe serait d'environ 20 kDa (facteur de 2).

Figure 6.5 : Résumé des données d'interaction entre la Core C82 et p53 dimérique obtenues par résonance magnétique nucléaire. A) Spectre ^1H , ^{15}N -HSQC de la Core C82 seule. B) Spectre ^1H , ^{15}N -HSQC de la Core C82 en présence de p53-dim (ratio 1 :4). Un symbole « + » est utilisé pour indiquer les résonances ayant disparu. De plus, un choix de couleur a été apposé pour mettre en valeur l'effet de l'interaction sur l'intensité des résonances. Le rouge représente les résonances dont la diminution de l'intensité est supérieure à la moyenne, le vert reflète celles dont la diminution est supérieure à la médiane, en noir, celles dont la diminution est inférieure à la médiane et en violet, les nouvelles résonances. C) Séquence de la Core C82 sur laquelle les effets mesurés et observés ont été reportés. Les triangles indiquent des résonances dont le déplacement chimique a changé en raison de la présence de p53-dim, et les cercles indiquent les résonances en cours de dédoublement. Le code couleur reste identique à celui de la figure B).

Il est surprenant d'observer que l'interaction Core : p53 est différente selon l'état multimérique de cette dernière, même si la région 50-70 semble affectée dans les 2 cas. Rappelons-nous que la p53 monomérique n'arbore au vu de son spectre de dichroïsme circulaire aucune structure secondaire, contrairement à la forme dimérique qui possède un brin β et une hélice α [59]. Cette remarque nous permet donc de supposer qu'en absence de repliement la p53 ne proposerait pas la même surface d'interaction à la Core C82, ce qui expliquerait simplement cette différence marquante d'interactions.

6.4 Conclusion

En conclusion, ces travaux préliminaires nous ont permis de mettre en évidence que la Core C82 n'acquière pas de structure secondaire au contact de la protéine p53 (monomérique ou dimérique). Par contre, l'interaction physique directe entre la Core de l'hépatite C et la protéine p53 a été démontrée pour la première fois.

Ces résultats ouvrent grand la porte à la poursuite de ce projet. En effet, même si la Core ne se structure pas au contact de la région C-terminale de la p53, il reste important de pouvoir mesurer et apprécier cette interaction dans le contexte de la Core et de la p53, car rappelons-nous que le domaine C-terminal de la p53 est également non structuré. Peut-être que celui-ci adopte un repliement au contact de la Core. D'autre part, il serait important de cartographier les sites précis de liaison des deux protéines dans une optique thérapeutique, en développant des inhibiteurs spécifiques de cette interaction.

Chapitre 7

Interfering with HCV assembly *in vitro* using affinity peptides directed towards Core protein

Jean-Baptiste Duvignaud, Nathalie Majeau, Priscilla Delisle, Normand Voyer, Stéphane M. Gagné and Denis Leclerc.

Soumis dans Canadian Journal of Microbiology.

7.1 Résumé

L'assemblage est une étape clé du cycle de vie d'un virus. Dans le cas du virus de l'hépatite C, la protéine Core est la seule protéine dite structurale à interagir avec le génome viral. La protéine Core produite en bactéries et purifiée est capable de s'auto-assembler *in vitro* pour former des pseudo-particules virales, grâce à l'ajout d'un ARN structuré, ce qui constitue une expérience précieuse pour étudier l'assemblage du VHC. L'inhibition de l'auto-assemblage de la Core C170 a été testée en utilisant de courts peptides dérivés de la séquence peptidique de la Core, de la protéine NS5A du VHC et de diverses autres protéines dont l'interaction avec la Core a été démontrée (p21, p73). De manière intéressante, les peptides dérivés de la séquence de la Core se sont montrés les plus efficaces. Ceux arborant le meilleur pouvoir inhibiteur sont dérivés de régions impliqués dans l'interaction Core-Core facilitant l'assemblage viral. Nous avons aussi mis en évidence l'effet inhibiteur d'un peptide dérivé de la région C-terminale de la protéine NS5A du VHC sur l'assemblage viral.

7.2 Abstract

Viral assembly is a crucial key step in the life cycle of every virus. In the case of hepatitis C virus (HCV), the Core protein is the only structural protein to interact directly with the viral genomic RNA. Purified recombinant Core protein is able to self-assemble *in vitro* into nucleocapsid-like particles (NLPs) upon addition of a structured RNA, providing a robust assay with which to study HCV virus assembly. Inhibition of self-assembly of the C170 Core protein (first 170 amino acids) was tested using small peptides derived from the HCV Core, HCV NS5A and from diverse proteins known to interact with HCV Core protein (p21, p73). Interestingly, peptides derived from Core were the best inhibitors. These peptides are derived from regions of Core predicted to be involved in the interaction between Core subunits during viral assembly. We also demonstrate that a peptide derived from the C-terminal end of NS5A protein moderately inhibits the assembly process.

7.3 Introduction

Hepatitis C virus (HCV) is one of the most devastating infectious diseases affecting humans, with more than 170 million people infected worldwide [17]. HCV infection is the major cause of chronic liver disease, which includes chronic hepatitis, cirrhosis and hepatocellular carcinoma. No vaccine exists and treatments are limited in both variety and efficiency [31]. Several avenues are currently being investigated in order to develop new treatments or vaccines [25], and the quest for inhibitors is one of the most promising [7]. Inhibitors of HCV protease [21] and the development of nucleotide analogs [12] that interfere with viral replication have already yielded encouraging results [26]. However, such inhibitors are also known to induce the appearance of resistant strains [32]. It is therefore important to develop inhibitors to other highly conserved protein targets of the virus to decrease the risk of emergence of resistance. The Core protein of HCV, which is considered the most conserved protein throughout all strains of HCV, is therefore an excellent choice for the development of a new class of inhibitor [11]. The Core protein is located at the N-terminal end of the polyprotein and is cleaved by two host proteases—signal peptidase (sp) and signal peptide peptidase (spp)—leading to production of the mature form of the protein [23, 24]. Core—the only structural protein of the capsid—triggers assembly and packaging of the HCV genome [18]. The self-assembly of HCV Core protein is activated by the addition of a structured RNA (tRNA and HCV-IRES) to the purified protein [14, 18]. We have developed a simple and rapid *in vitro* assay in which the progress of assembly is monitored by measuring an increase in turbidity, thereby allowing the kinetics of assembly to be determined [8]. Most previous data on Core–Core protein interaction or nucleocapsid like particle (NLP) formation have been generated using truncated Core (first 120 amino acids or less) [8, 14, 18]. These previous studies have clarified conditions for HCV assembly *in vitro* (salt concentration, basic residues involved, types and concentration of nucleic acids) and helped identify some peptide inhibitors of this process.

To improve our understanding of HCV assembly and to develop a more efficient tool to study its inhibition using peptides, we have extended our *in vitro* assembly studies to

a Core protein (C170) that is closer in size to the protein that shapes virions (177 amino acids) [23, 24]. We investigated the inhibition of assembly using small peptides derived from HCV protein sequences, or from sequences of diverse proteins known to interact with the Core protein. We succeeded in identifying several peptides that interfere with the assembly process: three peptide sequences from Core and one from NS5A were shown to be the most efficient.

7.4 Materials and methods

7.4.1 Expression and purification of C170

A clone of HCV Core containing amino acids 1–170 (C170) followed by a His₆-tag at the C-terminus was optimized for translation in *E. coli* and cloned into the vector pET-3D. The protein was expressed in *E. coli* strain BL21 (DE3) Star (Invitrogen) as described previously [5, 6]. After 3 hours of expression, the bacterial pellet was resuspended in 30 mL of lysis buffer [8 M urea, 20 mM Tris, 300 mM NaCl, pH 7.4, with the addition of 1x cocktail of protease inhibitors (Roche Diagnostics GmbH)], and lysed by sonication. The lysate was centrifuged at 27,000 x g for 30 min. Protein purification consisted of a two-step protocol: a Ni-NTA resin column (Ni-nitrilotriacetic acid resin, Qiagen) followed by a desalting column. The Ni-NTA beads were washed with decreasing concentrations of urea, and C170 proteins were eluted at a final concentration of 2 M urea. The eluted proteins were then diluted twice with PBS and centrifuged at 4000 x g to remove Ni-NTA beads. C170 protein was then injected into a 5 mL desalting column (HiTrapTM Desalting, GE HealthCare) previously equilibrated with PBS (1.76 mM KH₂PO₄, 10 mM Na₂HPO₄, 137 mM NaCl, 2.7 mM KCl at pH 7.4). Acetate or phosphate buffer at neutral pH support self-assembly of HCV Core protein as previously demonstrated by Kunkel et al. [14]. Finally, the efficiency of purification was confirmed by SDS-PAGE, and C170 protein was quantified using a BCA protein assay kit (Pierce).

7.4.2 Kinetics of C170 in vitro assembly reactions

C170 protein (8 μg) and HCV-IRES (first 374 nucleotides of strain H77c) (2.4 μg) were diluted in 60 μl and 40 μl of PBS, respectively, and then mixed in a microcuvette (Uvette Eppendorf) to initiate the assembly reaction. The molar ratio of Protein:HCV-IRES was 20:1, in a final reaction volume of 100 μl . For the assay with the C82 truncated Core protein, conditions were as described previously [8]. Peptide inhibitors were added to the protein mixture, maintaining a volume of 60 μl , to yield a final molar ratio of Protein:Peptide:HCV-IRES of 20:1000:1. The assembly reaction was monitored at 350 nm in a UV/visible spectrophotometer (Biochrom Ltd, Ultrospec 2100 pro) at room temperature. Absorbance was recorded every two seconds for at least ten minutes. Initial rate (first 8 seconds) and maximum absorbance were deduced from the resulting spectrum.

7.4.3 Peptide library

HCV peptides used in this work covering Core protein and NS5A come from the library obtained from the NIH AIDS Research and Reference Reagent Program. We used 23 synthetic peptides covering the first 170 amino acids of the Core protein, and three synthetic peptides covering a C-terminal region of NS5A. Others peptides were purchased from GenScript (NJ, USA).

7.4.4 Electrophoresis mobility shift assay

Mixtures of Peptide:HCV-IRES were incubated for 10 minutes at room temperature. The ratios used for these experiments were exactly the same as those used in kinetics experiments, in a final volume of 25 μl . Loading dye (5 μl) was added to the samples before loading them onto a 0.8 % agarose gel. Electrophoresis was performed in Tris/acetic acid/EDTA buffer for 30 min at 90 V. The gel was stained with ethidium bromide and photographed using a digital camera.

7.4.5 Electron microscopy

Samples collected after in vitro assembly were directly absorbed on 400-mesh carbon-Formvar grids (Canemco) for 5 min. The grids were washed once with filtered phosphate-buffered saline (PBS) and stained for 10 min with filtered 2 % (w/v) uranyl acetate. Grids were then dried on filter paper before being examined under an electron microscope with an acceleration voltage of 60 kV at a magnification of 150,000. Images were captured and treated with Digital Micrograph (version 3.8.2, Gatan).

7.5 Results & Discussion

7.5.1 In vitro assembly of hepatitis Core protein C170

We have developed an easy and fast method to monitor the progress of HCV NLP formation by recording the increase in turbidity when the protein and RNA components are mixed [8]. The formation of NLPs was initiated by mixing the purified recombinant Core with a structured RNA (tRNA or HCV-IRES) followed by measuring the optical density (OD) at a wavelength of 350 nm [8].

In this study, the assay was performed with HCV-IRES RNA (first 374 nucleotides of the HCV RNA genome) and a HCV Core protein consisting of amino acids 1–170 (C170), which is very close in size to the mature Core protein observed *in vivo* (177 amino acids) [23, 24]. Recombinant C170 was purified using a two-step purification protocol. The quality of the purified protein was confirmed by SDS-PAGE electrophoresis (Figure 1A), and an *in vitro* assembly test was performed to confirm self-assembly into NLPs upon addition of a structured RNA (HCV-IRES) (Figure 1C). *In vitro* assembly of a molar ratio of Protein:HCV-IRES of 20:1 was plotted (Figure 1B). Upon addition of RNA, assembly is extremely fast and is almost complete after the first minute of reaction. The initial rate was evaluated from the first eight seconds of the reaction. The initial values generated from different experiments were remarkably constant (4.81 OD/min) and reproducible (standard deviation = 0.4). In order to verify if a longer Core protein sequence is better for NLP assembly, we compared our results with data from a shorter truncated form of the HCV Core protein called C82 that was used previously in a similar assay [8]. C170 was seen to initiate the start of the reaction faster than C82 (more than twice as fast) (Figure 1B). This result suggests that C-terminal domains of Core are involved in the assembly process, either through an increased affinity for RNA or through a stronger Core–Core protein interaction.

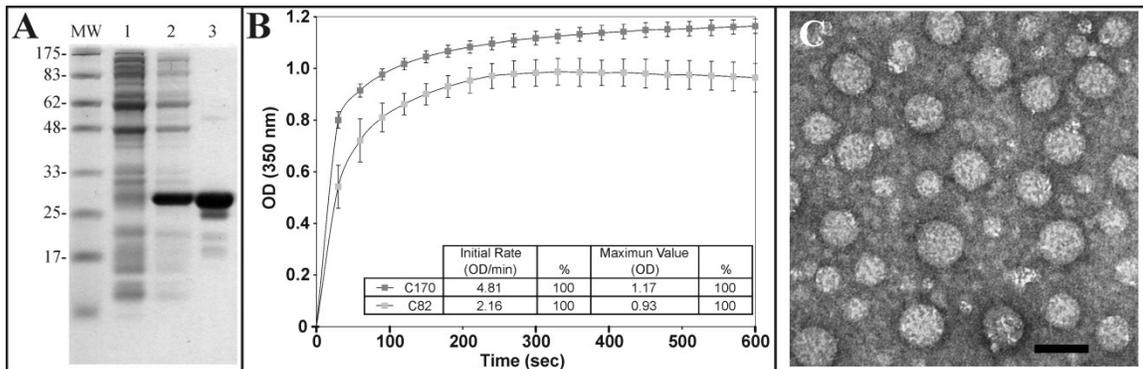


Figure 7.1 : HCV C170 purification and assembly. A) HCV Core protein C170 was expressed in *E. coli* and purified under denaturing conditions on a Ni-affinity column. After desalting on a Hi-Trap desalting column, the protein was analysed by 10 % SDS-Page. Lanes: MW, Molecular weight; 1, before induction; 2, after induction; 3, C170 after a two step purification protocol. B) Purified protein C170 and truncated protein C82 were mixed with HCV-IRES RNA at a molar ratio of protein : RNA of 20:1 and assembly was monitored by recording optical density at a wavelength of 350 nm every 2 s for 10 min. Standards errors were calculated from five independent experiments. Initial rates (OD/min) were calculated over the first 8 sec of the reaction, and maximum values (OD) were taken directly from recorded data. C) Electron micrograph of C170 nucleocapsid like particles (NLPs). Bar = 50 nm.

7.5.2 Inhibition of assembly using peptides derived from the HCV Core protein sequence

The HCV Core sequence from amino acid 1 to 170 can be divided into 2 parts: (1) a positively charged domain involved in RNA binding (amino acids 1–120) [18] that includes a tryptophan-rich domain that appears to be involved in the interaction between Core subunits (amino acids 76–107) [13, 14]; and (2) the C-terminal part of the protein interacting with lipids (ER, lipid droplets), which is located between amino acids 121 and

170 [9] where protein–protein interactions have been described [22]. To screen potential inhibitors from the HCV Core sequence, we tested peptides of 18 amino acids in length, overlapping by 11 amino acids, covering amino acids 1–172 of the HCV Core sequence (Table I). We evaluated the effect of each peptide in inhibiting assembly by recording the initial rate and the maximum OD of assembly reactions (Figures 2A and B). We found 7 peptides covering 4 regions of the Core that showed an inhibitory effect on NLP formation: 22-39, 43-60, 78-95, 85-102, 92-109, 120-137 and 127-144. The peptides were tested individually for their potential to bind the HCV-IRES RNA; this interaction is derived from a non-specific interaction between opposite charges. Three domains were identified: 1–25, 36–74 and 92–130, covering a region described previously as having RNA affinity [27] (Figure 2C). Interestingly, most of these peptides had little effect on NLP formation. Peptide 22–39 was shown to inhibit NLP formation with an initial rate reduced to 58 % of the reference value. Interestingly, this peptide had been reported previously to inhibit NLP formation with C82 with an initial rate of 0.32 OD/min [8]. A recent study reported that a peptide derived from this region (16–34) can interfere with HCV replication [28]. Therefore, our results further strengthen the potential of this peptide for use in therapy since it interferes with two important aspects of the HCV life cycle: assembly and replication. In order to design a smaller molecule to inhibit the assembly process, we created a bank of smaller peptides derived from this sequence. Unfortunately, as soon as the peptides reached a critical size of 16 amino acids and shorter, the inhibitory effect was lost (see supporting information). This result suggests that 16 amino acids is a critical size for efficacy of binding to Core and triggering inhibition of assembly.

Peptide 120-137 also showed an inhibitory effect on oligomerisation by reducing the initial rate of the reaction to 60 % of its starting value. It was reported that the C-terminal domain of the Core is involved in protein–protein interactions [22, 30]. This peptide includes cysteine 128—described recently to be involved in dimerization of the Core protein [16]. It is possible that this amino acid participates in NLP formation through the creation of a disulfide bound. However, this result is controversial since others have reported that Core–Core interaction is not dependent on the oxidation state of the protein [2, 15, 30]. The most inhibitory peptides were found in the tryptophan-rich domain located

between amino acids 78 and 109. Peptides 78-95 and 92-109 reduced the initial rate of assembly by 54 % and 77 %, respectively, with the final OD reading being decreased by ~15 % (Figure 3A). These two peptides overlap with peptide 85–102, which interferes the most with NLP formation—a very low initial rate (5 % of the reference) and an OD decreased to 16 % of the reference value. Interestingly, this peptide corresponds to the main homotypic interaction domain [22], which also inhibits dimerization of a truncated form of HCV Core protein [13].

Tableau 7.1 : Sequences of HCV Core peptides used in this study

	Position (aa)	HCV Core peptide sequence
1	1-18	MSTNPKPQRKTKRNTNRR
2	8-25	QRKTKRNTNRRPQDVKFP
3	15-32	TNRRPQDVKFPGGGQIVG
4	22-39	VKFPGGGQIVGGVYLLPR
5	29-46	QIVGGVYLLPRRGPRLGV
6	36-53	LLPRRGPRLGVRATRKTS
7	43-60	RLGVRATRKTSERSQPRG
8	50-67	RKTSERSQPRGRRQPIPK
9	57-74	QPRGRRQPIPKARRPEGR
10	64-81	PIPKARRPEGRTWAQPGY
11	71-88	PEGRTWAQPGYPWPLYGN
12	78-95	QPGYPWPLYGNEGCGWAG
13	85-102	LYGNEGCGWAGWLLSPRG
14	92-109	GWAGWLLSPRGSRPSWGP
15	99-116	SPRGSRPSWGPTDPRRRS
16	106-123	SWGPTDPRRRSRNLGKVI
17	113-130	RRRSRNLGKVIDTLTCGF
18	120-137	GKVIDTLTCGFADLMGYI
19	127-144	TCGFADLMGYIPLVGAPL
20	134-151	MGYIPLVGAPLGGAARAL
21	141-158	GAPLGGAARALAHGVRVL
22	148-165	ARALAHGVRVLEDGVNYA
23	155-172	VRVLEDGVNYATGNLPGC

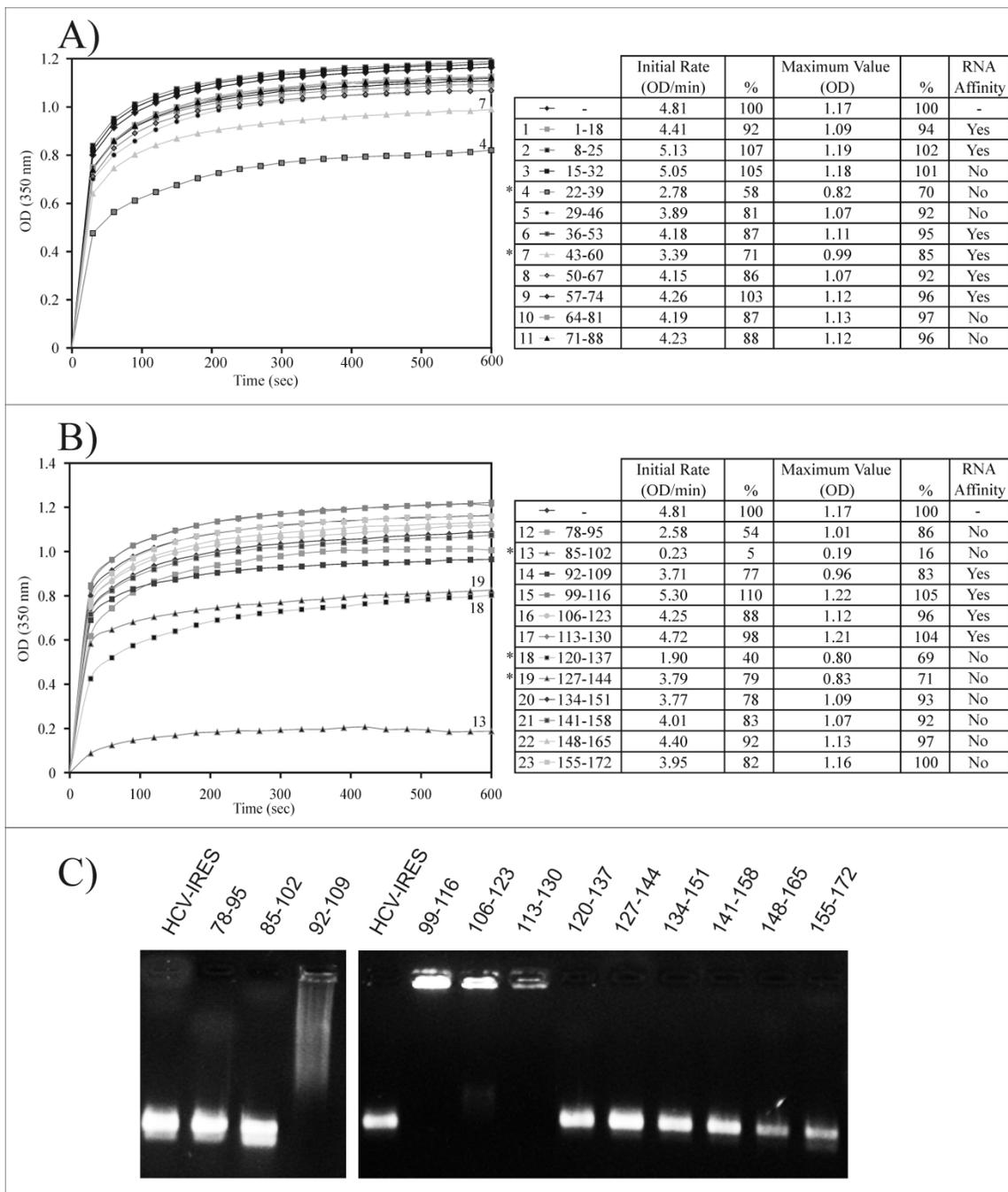


Figure 7.2 : Assembly kinetics of C170 with the addition of diverse peptides derived from HCV Core amino acid sequence. Core protein C170 was first incubated with peptides (18 amino acids long) before mixing with HCV-IRES at a molar ratio of 20:1000:1 (protein:peptide:RNA). A) Peptides covering the Core amino acid sequence from 1 to 88 and B) from 78 to 172. C) Each peptide alone (peptides 12 to 23) was mixed with the HCV-IRES RNA at a molar ratio of 1000:1 and the RNA affinity to the peptide was evaluated by electrophoresis mobility shift assay on a 0.8 % agarose gel. Peptides 1 to 11 were already tested and the gel shift mobility assay can be seen in Fromentin et al. [8]. (See above)

7.5.3 Inhibition of assembly with peptides derived from NS5A, p21 and p73

HCV Core protein is known to contain an intrinsically disordered domain [5, 6], which confers on this protein a large spectrum of interacting partners, including other HCV proteins and host proteins [1, 3, 4, 10, 19, 29]. From the growing list of HCV Core interacting partners, three proteins were selected to generate candidate inhibitory peptides. The three proteins are NS5A from HCV, and two host proteins: p21 and p73. The choice of NS5A is of particular interest since this protein is implicated in the assembly process with the Core protein [9, 20]. The choice of p21 and p73 proteins was done, as their interactions to the Core protein were demonstrated to be on short regions and not dependent on the tridimensional folding of those proteins. Their interactions were mapped to amino acids 139 to 164 for p21 and amino acids 321 to 353 for p73. Peptides of fifteen amino acids long were designed to cover the length of these regions. Finally, we constructed a bank of additional peptides varying in length from 13 to 18 amino acids long (Table II).

From the three peptides derived from the C-terminal end of NS5A, one (peptide 2407–2419) exhibits an interesting inhibiting effect (Figures 3A & 3C). Both the initial rate and maximum OD decrease strongly to levels as low as 20 % and 50 %, respectively, of the

reference values, which confirmed the interaction of the C-terminal end of NS5A with the Core protein, and its potential role in the regulation of HCV assembly at the interface of ER/lipid droplets [9]. The assay was also tested in the presence of the selected peptides derived from the p21 and p73 protein sequence. However, none of these peptides exhibited an inhibitory effect (Figure 3B), suggesting that these regions do not interfere with the multimerization of Core subunits.

Tableau 7.2 : Sequences of diverse peptides tested as potential assembly inhibitors.

Position (aa)	HCV NS5A peptide sequence
2395-2412	PGDPDLSDGSWSTVSSGA
2402-2416	DGSWSTVSSGADTED
2407-2419	TVSSGADTEDVVC
	p21 peptide sequence
139-153	GRKRRQTSMTDFYHS
146-160	SMTDFYHSKRRLIFS
150-164	FYHSKRRLIFSKRKP
	p73 peptide sequence
321-335	AFKQSPPAIPALGTN
336-350	VKKRRHGDEDMFYMH
324-338	QSPPAIPALGTNVKK
328-342	AIPALGTNVKKRRHG
339-353	RRHGDEDMFYMHVIRG

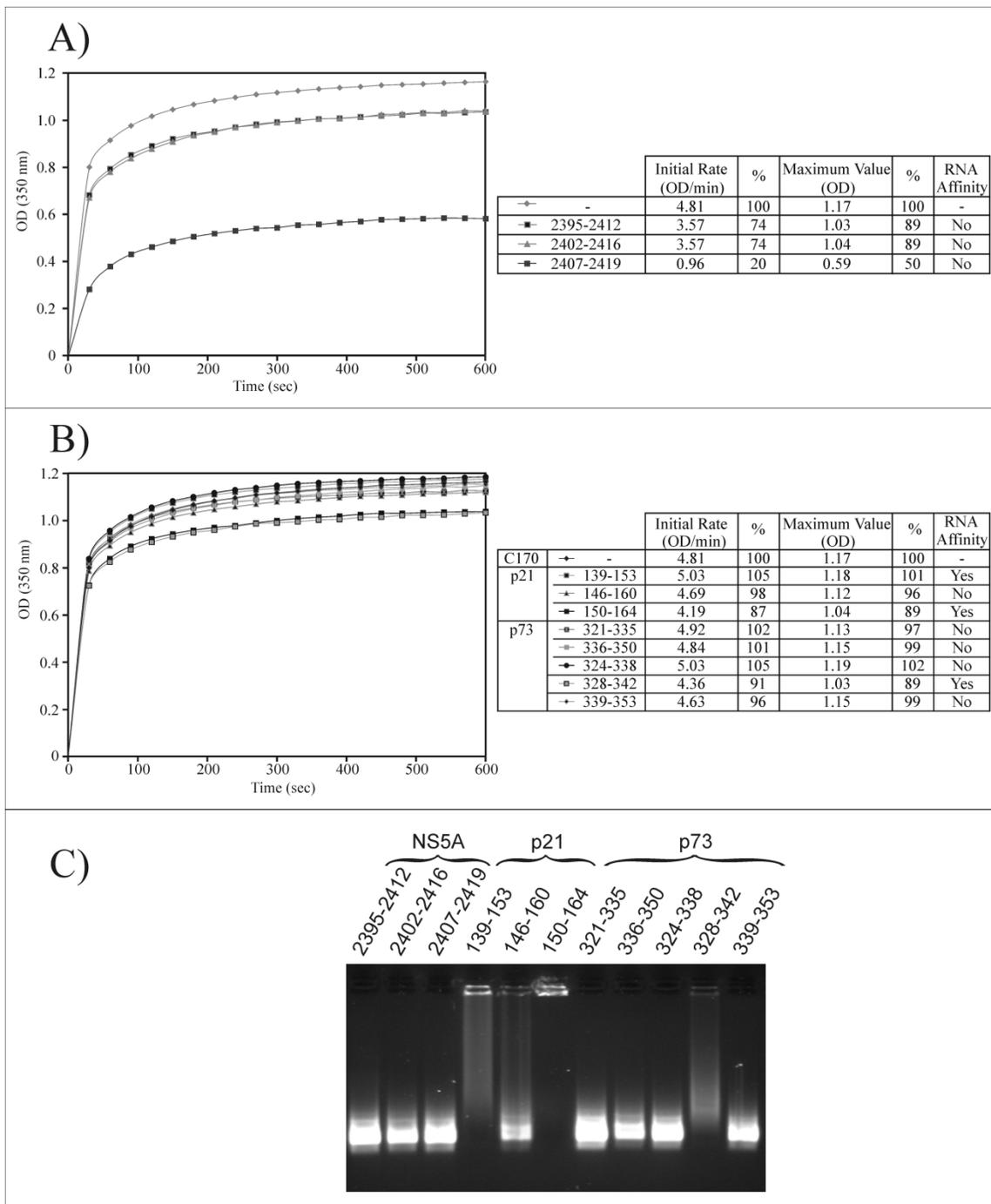


Figure 7.3 : Assembly kinetics of C170 with the addition of diverse peptides selected from three interacting proteins (NS5A, p21 and p73). Peptides of 13 or 18 amino acid long were mixed with the Core protein C170 in a molar ratio 20:1000:1 (protein:peptide:RNA). RNA affinity was evaluated by electrophoresis mobility shift assay as in Figure 7.2. A) Assembly kinetics with peptides derived from NS5A and B) p21 and p73 proteins. C) Peptides derived from NS5A, p21 and p73 proteins were mixed with the HCV-IRES RNA at a molar ratio of 1000:1 and the RNA affinity to the peptide was evaluated by electrophoresis mobility shift assay on a 0.8 % agarose gel. (Page précédente)

7.6 Conclusion

To conclude, we have succeeded in developing an *in vitro* assembly assay for an HCV Core protein of 170 amino acids in length. Interestingly, this protein has a higher initial rate of assembly and final OD than C82, since C170 possesses more interaction domains to itself and to HCV-IRES, allowing a more efficient and rapid assembly process. A large range of peptides were tested in this study, making it possible to highlight seven inhibitors of HCV Core assembly, confirming the ease and efficiency of this *in vitro* test for screening for inhibitors of the assembly process of the HCV. The most potent inhibitory peptides were derived from HCV protein sequences (Core protein, and NS5A protein). Indeed, peptides derived from Core protein sequence highlight some regions known to be involved in Core–Core interaction, i.e. the homotypic Core interaction domain, also known as the “tryptophan-rich domain” (85–102), and the heterotypic Core interaction domain (120–144), which was shown recently to contain a cysteine (Cys 128) implicated in the formation of a disulfide-bonded dimer Core protein [16]. The result obtained for the “2407-2419” peptide derived from NS5A strengthens the role of the interaction between NS5A and the Core protein for the assembly of the HCV virus at the lipid droplet location.

The results obtained in this study are the beginning of a great story and future works will include *in vivo* test of those peptides to measure their potentials as therapeutics. The seven peptides identified here represent as many interesting avenues, and are a good starting point for improving their affinity to the HCV Core protein and their inhibitory effect on HCV Core assembly by rational design.

7.7 Acknowledgments

Thanks to the ‘Département de Biochimie, Microbiologie et Bioinformatique’ of Laval University, for financial support awarded to J.-B. Duvignaud. We thank Mélanie Tremblay for help for the synthesis of the “6929” peptide library. This project was funded by a CIHR operating grant, two FCI infrastructure grants (Innovation and New-Opportunity) and the “Réseau SIDAMI” of the Fonds de la Recherche en Santé du Québec (FRSQ) and by a PROTEO New Initiative Grant (FQRNT).

7.8 Supplementary materials

Voir Annexe 3

7.9 References

1. Alisi, A., Giambartolomei, S., Cupelli, F., Merlo, P., Fontemaggi, G., Spaziani, A., and Balsano, C., *Physical and functional interaction between HCV core protein and the different p73 isoforms*. *Oncogene*, 2003. **22**(17): p. 2573-80.
2. Boulant, S., Vanbelle, C., Ebel, C., Penin, F., and Lavergne, J.P., *Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features*. *J Virol*, 2005. **79**(17): p. 11353-65.
3. Chen, C.M., You, L.R., Hwang, L.H., and Lee, Y.H., *Direct interaction of hepatitis C virus core protein with the cellular lymphotoxin-beta receptor modulates the signal pathway of the lymphotoxin-beta receptor*. *J Virol*, 1997. **71**(12): p. 9417-26.
4. de Chasse, B., Navratil, V., Tafforeau, L., Hiet, M.S., Aublin-Gex, A., Agaogue, S., Meiffren, G., Pradezynski, F., Faria, B.F., Chantier, T., Le Breton, M., Pellet, J., Davoust, N., Mangeot, P.E., Chaboud, A., Penin, F., Jacob, Y., Vidalain, P.O., Vidal, M., Andre, P., Rabourdin-Combe, C., and Lotteau, V., *Hepatitis C virus infection protein network*. *Mol Syst Biol*, 2008. **4**: p. 230.
5. Duvignaud, J.B., Leclerc, D., and Gagne, S.M., *Structure and dynamics changes induced by 2,2,2-trifluoro-ethanol (TFE) on the N-terminal half of hepatitis C virus core protein*. *Biochem Cell Biol*, 2010. **88**(2): p. 315-23.
6. Duvignaud, J.B., Savard, C., Fromentin, R., Majeau, N., Leclerc, D., and Gagne, S.M., *Structure and dynamics of the N-terminal half of hepatitis C virus core protein: an intrinsically unstructured protein*. *Biochem Biophys Res Commun*, 2009. **378**(1): p. 27-31.
7. Flisiak, R. and Parfieniuk, A., *Investigational drugs for hepatitis C*. *Expert Opin Investig Drugs*, 2010. **19**(1): p. 63-75.
8. Fromentin, R., Majeau, N., Laliberte Gagne, M.E., Boivin, A., Duvignaud, J.B., and Leclerc, D., *A method for in vitro assembly of hepatitis C virus core protein and for screening of inhibitors*. *Anal Biochem*, 2007. **366**(1): p. 37-45.
9. Jones, D.M. and McLauchlan, J., *Hepatitis C virus: assembly and release of virus particles*. *J Biol Chem*, 2010.
10. Kao, C.F., Chen, S.Y., Chen, J.Y., and Wu Lee, Y.H., *Modulation of p53 transcription regulatory activity and post-translational modification by hepatitis C virus core protein*. *Oncogene*, 2004. **23**(14): p. 2472-83.

11. Karthe, J., Tessmann, K., Li, J., Machida, R., Daleman, M., Haussinger, D., and Heintges, T., *Specific targeting of hepatitis C virus core protein by an intracellular single-chain antibody of human origin*. *Hepatology*, 2008. **48**(3): p. 702-12.
12. Klumpp, K., Leveque, V., Le Pogam, S., Ma, H., Jiang, W.R., Kang, H., Granycome, C., Singer, M., Laxton, C., Hang, J.Q., Sarma, K., Smith, D.B., Heindl, D., Hobbs, C.J., Merrett, J.H., Symons, J., Cammack, N., Martin, J.A., Devos, R., and Najera, I., *The novel nucleoside analog R1479 (4'-azidocytidine) is a potent inhibitor of NS5B-dependent RNA synthesis and hepatitis C virus replication in cell culture*. *J Biol Chem*, 2006. **281**(7): p. 3793-9.
13. Kota, S., Coito, C., Mousseau, G., Lavergne, J.P., and Strosberg, A.D., *Peptide inhibitors of hepatitis C virus core oligomerization and virus production*. *J Gen Virol*, 2009. **90**(Pt 6): p. 1319-28.
14. Kunkel, M., Lorinczi, M., Rijnbrand, R., Lemon, S.M., and Watowich, S.J., *Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein*. *J Virol*, 2001. **75**(5): p. 2119-29.
15. Kunkel, M. and Watowich, S.J., *Biophysical characterization of hepatitis C virus core protein: implications for interactions within the virus and host*. *FEBS Lett*, 2004. **557**(1-3): p. 174-80.
16. Kushima, Y., Wakita, T., and Hijikata, M., *A disulfide-bonded dimer of the core protein of hepatitis C virus is important for virus-like particle production*. *J Virol*, 2010. **84**(18): p. 9118-27.
17. Lauer, G.M. and Walker, B.D., *Hepatitis C virus infection*. *N Engl J Med*, 2001. **345**(1): p. 41-52.
18. Majeau, N., Gagne, V., Boivin, A., Bolduc, M., Majeau, J.A., Ouellet, D., and Leclerc, D., *The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation*. *J Gen Virol*, 2004. **85**(Pt 4): p. 971-81.
19. Mamiya, N. and Worman, H.J., *Hepatitis C virus core protein binds to a DEAD box RNA helicase*. *J Biol Chem*, 1999. **274**(22): p. 15751-6.
20. Masaki, T., Suzuki, R., Murakami, K., Aizaki, H., Ishii, K., Murayama, A., Date, T., Matsuura, Y., Miyamura, T., Wakita, T., and Suzuki, T., *Interaction of hepatitis C virus nonstructural protein 5A with core protein is critical for the production of infectious virus particles*. *J Virol*, 2008. **82**(16): p. 7964-76.
21. McHutchison, J.G., Manns, M.P., Muir, A.J., Terrault, N.A., Jacobson, I.M., Afdhal, N.H., Heathcote, E.J., Zeuzem, S., Reesink, H.W., Garg, J., Bsharat, M., George, S., Kauffman, R.S., Adda, N., and Di Bisceglie, A.M., *Telaprevir for previously treated chronic HCV infection*. *N Engl J Med*, 2010. **362**(14): p. 1292-303.

22. Nolandt, O., Kern, V., Muller, H., Pfaff, E., Theilmann, L., Welker, R., and Krausslich, H.G., *Analysis of hepatitis C virus core protein interaction domains*. J Gen Virol, 1997. **78 (Pt 6)**: p. 1331-40.
23. Ogino, T., Fukuda, H., Imajoh-Ohmi, S., Kohara, M., and Nomoto, A., *Membrane binding properties and terminal residues of the mature hepatitis C virus capsid protein in insect cells*. J Virol, 2004. **78(21)**: p. 11766-77.
24. Okamoto, K., Mori, Y., Komoda, Y., Okamoto, T., Okochi, M., Takeda, M., Suzuki, T., Moriishi, K., and Matsuura, Y., *Intramembrane processing by signal peptide peptidase regulates the membrane localization of hepatitis C virus core protein and viral propagation*. J Virol, 2008. **82(17)**: p. 8349-61.
25. Romero-Lopez, C., Diaz-Gonzalez, R., Barroso-delJesus, A., and Berzal-Herranz, A., *Inhibition of hepatitis C virus replication and internal ribosome entry site-dependent translation by an RNA molecule*. J Gen Virol, 2009. **90(Pt 7)**: p. 1659-69.
26. Seden, K., Back, D., and Khoo, S., *New directly acting antivirals for hepatitis C: potential for interaction with antiretrovirals*. J Antimicrob Chemother, 2010. **65(6)**: p. 1079-85.
27. Shimoike, T., Koyama, C., Murakami, K., Suzuki, R., Matsuura, Y., Miyamura, T., and Suzuki, T., *Down-regulation of the internal ribosome entry site (IRES)-mediated translation of the hepatitis C virus: critical role of binding of the stem-loop IIIId domain of IRES and the viral core protein*. Virology, 2006. **345(2)**: p. 434-45.
28. Sun, C., Pager, C.T., Luo, G., Sarnow, P., and Cate, J.H., *Hepatitis C Virus Core-Derived Peptides Inhibit Genotype 1b Viral Genome Replication via Interaction with DDX3X*. PLoS One, 2010. **5(9)**: p. e12826.
29. Wang, F., Yoshida, I., Takamatsu, M., Ishido, S., Fujita, T., Oka, K., and Hotta, H., *Complex formation between hepatitis C virus core protein and p21Waf1/Cip1/Sdi1*. Biochem Biophys Res Commun, 2000. **273(2)**: p. 479-84.
30. Yan, B.S., Tam, M.H., and Syu, W.J., *Self-association of the C-terminal domain of the hepatitis-C virus core protein*. Eur J Biochem, 1998. **258(1)**: p. 100-6.
31. Yuan, H.J. and Lee, W.M., *Nonresponse to treatment for hepatitis C: current management strategies*. Drugs, 2008. **68(1)**: p. 27-42.
32. Zhou, Y., Bartels, D.J., Hanzelka, B.L., Muh, U., Wei, Y., Chu, H.M., Tigges, A.M., Brennan, D.L., Rao, B.G., Swenson, L., Kwong, A.D., and Lin, C., *Phenotypic characterization of resistant Val36 variants of hepatitis C virus NS3-4A serine protease*. Antimicrob Agents Chemother, 2008. **52(1)**: p. 110-20.

Chapitre 8

Conclusion et perspectives

La protéine Core est une des protéines les plus importantes du virus de l'hépatite C. Elle constitue à elle-seule la capsid virale, en interagissant avec l'ARN génomique viral, les lipides et les protéines de surface. Elle interagit intimement avec d'autres protéines virales (NS3, NS5A et NS5B) et avec un grand nombre de protéines de l'hôte. Cette dernière capacité fait de la protéine Core un sujet d'étude important pour la compréhension de son rôle non seulement au sein du cycle viral, mais aussi dans la pathogénicité de la maladie. La caractérisation de ses rôles et de ses implications au sein de la cellule infectée est en émergence grâce au système de culture cellulaire récemment développé [160, 251, 278]. Si ces divers rôles et fonctions commencent à être mieux compris et détaillés, sa caractérisation structurale précise reste, quant à elle, à améliorer.

L'objectif de cette thèse était de déterminer la structure tridimensionnelle et la dynamique de la moitié N-terminale de la protéine Core (C82), dans le but de mettre en évidence les aspects structuraux de celle-ci impliqués dans la liaison à l'ARN, et à sa multimérisation. Le choix d'étudier la région N-terminale de la Core (C82) a été motivé par le fait que cette courte section de la protéine conservait la capacité de lier l'ARN et de s'assembler pour former une capsid à la fois *in vivo* et *in vitro*.

Pour répondre au mieux à l'objectif, nous avons décidé de diviser le projet en sous-objectifs. Le premier a consisté à mettre au point un protocole efficace de production, purification et de marquage de la protéine C82. Nous avons alors utilisé de diverses stratégies pour augmenter le rendement de production de la protéine Core en bactérie avec succès (optimisation du clone pour production en bactérie, température, souche bactérienne). Le deuxième moment clé a été la mise au point d'un protocole efficace de purification. Cette tâche ardue a été couronnée de succès dès lors que nous avons décidé de travailler en condition de purification extrême (pH 12, sels, et agents chao-tropiques). Le protocole de purification en deux étapes, mis en place nous a permis d'obtenir une protéine extrêmement pure, condition *sine qua non* à une étude structurale. Enfin, le marquage sélectif aux ^{15}N et ^{13}C , pour une étude en RMN, a pu être obtenu grâce à l'utilisation d'un milieu riche marqué. Ces résultats positifs nous ont donné les moyens de poursuivre ce projet et de tenter de résoudre la structure tridimensionnelle de la Core C82.

Les 2^{ème} et 3^{ème} objectifs se sont intéressés à la caractérisation structurale et dynamique de la Core par résonance magnétique nucléaire. Nos travaux nous ont menés à conclure que la région N-terminale de la Core (C82) était non repliée et flexible à l'état natif, et qu'elle appartenait à la famille des protéines dites intrinsèquement non-structurées (« IUP » (Intrinsically Unstructured Protein), « IDP » (Intrinsically Disordered Protein), ou encore « NDP » (Natively Disordered Protein)).

En effet, l'analyse bio-informatique de la structure primaire et les prédictions de structures secondaires montrent clairement que la protéine Core ne présente que peu d'éléments de structure secondaire (Figure 3.1A). Les données expérimentales de dichroïsme circulaire et de résonance magnétique nucléaire sont venues confirmer cette observation, avec des spectres typiques de protéine non-structurée (minimum vers 200 nm au niveau du spectre CD et une largeur spectrale inférieure à 1 ppm dans la dimension ^1H du spectre RMN ^1H - ^{15}N -HSQC, voir Chapitre 3). Toutefois, la présence potentielle de courts fragments en brin β (comme suggéré par les prédictions) n'a pu être exclue que par l'obtention des déplacements chimiques du squelette de la Core C82 et l'analyse de ceux-ci via le « Chemical Shift Index » (CSI). En effet, les déplacements chimiques des atomes composant le squelette d'une protéine sont influencés par leur environnement et les éléments de structures secondaires auxquels ils appartiennent. Dans le cas de la protéine Core C82 monomérique, les valeurs de CSI associées nous ont permis de confirmer son état non structuré. L'étude des propriétés dynamiques de la Core C82 est venue, quant à elle, indiquer qu'elle était non structurée et extrêmement flexible, et ce sur toute sa longueur. En effet, l'analyse des valeurs de NOEs (moyenne de $0,19 \pm 0,10$) et des valeurs de S^2 (moyenne de $0,58 \pm 0,05$) apporte une fois de plus une preuve de l'absence d'éléments de structures secondaires et de l'extrême flexibilité de la protéine C82. En comparaison, les valeurs typiques de NOEs pour les hélices α et brins β sont supérieures à 0,5 alors que pour les S^2 , des valeurs proches de 0,8 sont généralement observées.

Enfin, la grande diversité de fonctions et d'interactions protéiques de la Core du VHC a été une autre indication qui nous a incités à proposer que la région N-terminale de la Core soit associée à la famille des « IUPs ». Par ailleurs, nos travaux sur la forme C124

(Chapitre 5) et les études publiées par Kunkel et coll. [147], Boulant et coll. [30] et Rodriguez-Casado et coll. [214] corroborent nos dires sur l'état non-structuré de la région N-terminale de la Core (domaine D1).

De manière étonnante, la Core du VHC n'est pas la seule protéine de ce virus à démontrer ces caractéristiques physiques. Les protéines NS5A (domaine 2 et 3, en particulier) et les diverses formes de ARFP/F/Core+1 démontrent, elles aussi, une appartenance aux IUPs [31, 95, 96, 155, 156]. Toutefois, aucune autre similarité d'un point de vue fonctionnel et structural ne rapproche ces trois protéines. Les protéines Core et NS5A possèdent, toutes les deux, une affinité aux ARNs et sont impliquées dans la régulation et la formation de la capsid alors que pour la protéine ARFP/F/Core+1, aucune fonction claire ne lui est attribuée, et elle n'est pas indispensable pour la production d'une infection [186].

Par ailleurs, cette observation ne semble pas être un fait unique au virus de l'hépatite C. En effet, les protéines Core de plusieurs autres membres de la famille des *flaviviridae* possèdent des régions plus ou moins longues considérées comme des « IUPs » [114, 195] et toutes auraient une activité de chaperonne d'ARN [112].

Les « IUPs » sont des protéines qui ne possèdent pas ou peu de structures secondaire et tertiaire à l'état natif, mais sont capables de se replier au contact d'un partenaire pour effectuer leur rôle. Dans le but de vérifier le potentiel repliement de la Core C82, nous avons décidé de l'étudier, d'un point de vue structural, dans diverses conditions de sel, pH, détergent et de TFE (2,2,2-trifluoroéthanol). Seul l'ajout de TFE a démontré la capacité d'induire un repliement à la Core C82, comme en témoignent les spectres de dichroïsme circulaire et les spectres RMN (Chapitre 4). L'attribution des déplacements chimiques des atomes de la chaîne principale a été effectuée (~71 %), nous permettant ainsi de confirmer par analyse SSP la présence d'hélices α . En effet, trois courtes hélices ont ainsi été mises en évidence dans les 40 premiers acides aminés de la Core et une quatrième hélice à son extrémité C-terminale. La présence d'hélices dans cette région avait été suspectée par une étude similaire sur un très court peptide représentant une région

antigénique (aa 2-45). D'un point de vue dynamique, l'ajout de TFE vient diminuer de manière significative la flexibilité de la protéine, ce qui est en accord avec le gain de structure secondaire. Bien que ces travaux aient été effectués dans des conditions extrêmes (90 % TFE), ils contribuent de manière significative aux connaissances sur le repliement potentiellement adopté de la région N-terminale de la Core.

Par ailleurs, nous nous sommes aussi intéressés à la structure de deux formes tronquées plus longues de la Core dites C124 et C170 (Chapitre 5). La forme C124 représente le domaine D1 de la Core. Ce domaine contient l'ensemble des régions chargées de la Core potentiellement impliquées dans son interaction avec l'ARN viral et certains domaines d'interactions homotypiques. Quant à la forme C170, cette construction est très proche de la forme mature produite (C177). Elle est constituée des domaines D1 et D2. C'est ce dernier domaine qui permet le maintien de la Core maturée au niveau des gouttelettes lipidiques.

Nos travaux sur la forme C124 libre nous ont menés aux mêmes conclusions que pour le fragment C82. Cette forme représentant le domaine D1 ne possède aucun repliement en absence de ligands ou autres agents lipomimétiques (détergents ou TFE) (chapitre 5), ce qui est en accord avec les travaux de Kunkel et coll. [147], Boulant et coll. [30] et Rodriguez-Casado et coll. [214]. Cependant, notons que la Core C124, contrairement au fragment C82, possède la capacité de former des dimères, comme nous avons pu l'observer grâce à nos tests de liaisons croisées au glutaraldéhyde. Ceci nous permet donc de penser que bien que le fragment C82 soit le plus petit fragment permettant la formation de capsid *in vitro*, la présence des résidus 83 à 124 ajoute un segment favorisant l'interaction de la protéine avec elle-même en l'absence d'ARN. Ceci est confirmé par les travaux de Rodriguez-Casado et coll. [214] qui montrent que la Core C120 s'auto-assemble à très haute concentration induisant par là-même un repliement partiel de la Core en brin β . La même équipe aurait réussi à étudier l'interaction de cette forme tronquée avec divers ARNs et démontré plusieurs repliements possibles (présence d'hélice α et principalement de brin β) en fonction du type d'ARN lié. Dans le cas de la Core C82, le mélange de la protéine et de l'ARN (IRES ou ARN^{t^{phe}}) n'a pu être étudié d'un point de

vue structural en raison de l'obtention d'une solution turbide accompagnant le phénomène d'assemblage viral [82] (chapitre 7). Cette solution s'est avérée non étudiable en RMN et CD du fait de la grosseur de l'assemblage protéique obtenu et de la non homogénéité de la solution finale. Au final, les résultats obtenus par Rodriguez-Casado et coll. tendent à montrer que la région N-terminale de la Core est capable d'adopter principalement un repliement en brin β .

Concernant la version longue de la protéine Core (C170), aucune étude structurale en présence d'ARN n'a été réalisée et publiée à ce jour. Toutefois, nos travaux et ceux publiés par Boulant et coll. [30] démontrent que la Core mature adopte principalement un repliement en hélice α comme les prédictions de structure secondaire le suggèrent. L'étude de Boulant et coll. a été réalisée en présence de détergents et dans ces conditions la protéine formerait un dimère [30]. Dans notre étude, à la fois l'ajout de SDS (sous forme micellaire) ou de TFE favorise le repliement en hélice α , cependant l'état multimérique de la Core reste indéterminé. Ce repliement en hélice α ne semble pas seulement se confiner au domaine D2, mais aucune évidence d'un point de vue atomique n'a été apporté dans l'étude de Boulant et coll. [30]. Par contre nos travaux sur la Core C82 en présence de TFE viennent ici confirmer que la région N-terminale de la Core possède la capacité d'adopter ce type de structure secondaire.

Dans la lignée de nos travaux sur le repliement adopté de la Core C82, nous avons également voulu vérifier l'interaction de celle-ci avec l'un de ces nombreux partenaires protéiques et étudier l'effet de cette interaction sur la structure adoptée par la Core C82 (Chapitre 6). Nous avons ainsi choisi la protéine p53 connue pour interagir avec la protéine Core dans sa région N-terminale. L'ajout de celle-ci (sous forme monomérique et dimérique) à la Core C82 ne semble pas induire de changement structural important. Toutefois, l'interaction directe de la Core et de p53 a été démontrée comme en témoignent les spectres HSQC de la Core C82 en présence de p53 monomérique ou dimérique. De manière surprenante, nos travaux montrent que les sites d'interactions de Core et de p53 varient selon l'état de multimérisation de p53. Cette observation pourrait peut-être avoir une incidence biologique, car p53 serait dans des états multimériques différents selon sa

localisation intracellulaire. Ces travaux préliminaires pourraient être poursuivis et permettraient d'apporter une réponse quant au rôle de cette interaction *in vivo*.

Le dernier chapitre de cette thèse (chapitre 7), bien que très différent dans l'approche expérimentale, répond au 4^{ième} objectif de cette thèse qui était d'étudier, l'interaction de la Core avec des partenaires protéines, des acides nucléiques ou des inhibiteurs. Nous avons mis en place au laboratoire un test simple nous permettant de suivre la cinétique d'assemblage de la Core (C82) du VHC. Il consiste à mélanger la protéine Core (C82) du VHC à son ARN cible (IRES ou ARN_t^{phe}), et de mesurer l'absorbance créée par l'assemblage viral à une longueur d'onde déterminée (350 nm). Cette méthode spectroscopique s'est très vite avérée un outil exceptionnel, pour cribler des inhibiteurs d'assemblage potentiels. Rémi Fromentin a mis en place ce test pour la version C82 de la Core et a testé des peptides dérivés de la séquence de la Core C82 contre l'assemblage viral [82]. Un seul peptide a démontré la capacité d'inhiber l'assemblage viral (aa 22-39). Nous avons par la suite transféré ce test à une version plus longue de la Core (C170) et testé une banque de peptides dérivés de diverses régions de la séquence du virus de l'hépatite C et d'autres dérivés de régions de protéines dont l'interaction avec la Core du VHC a été démontrée. Nous avons donc dans un premier temps confirmé l'efficacité du peptide inhibiteur mis en évidence par Rémi Fromentin, auquel nous avons ajouté 4 autres peptides candidats, tous issus du virus de l'hépatite C (séquence de la Core et de la protéine NS5A). Notre hypothèse concernant les peptides dérivés de protéines interagissant avec la Core s'est avérée non efficace dans la recherche d'inhibiteurs, mais ceci ne signifie pas qu'il n'interagisse pas avec la protéine Core. Par contre, le choix des peptides dérivés du VHC a porté ses fruits, nous permettant par là-même de confirmer les sites d'interaction de la Core avec elle-même dictant son assemblage viral, et de mettre en avant le rôle majeur de la protéine NS5A dans le contrôle de l'assemblage viral.

En conclusion, la structure tridimensionnelle de la protéine Core n'a toujours pas été élucidée et ce malgré de nombreuses tentatives par plusieurs équipes, dont la nôtre. L'étude structurale du fragment N-terminal de la protéine Core par RMN est, à ce jour, la plus

importante réalisée à résolution atomique. Nos travaux sur la Core C82 ont démontré que celle-ci était non structurée à l'état natif et que le potentiel repliement adopté par la Core mature semble faire l'objet de divergence surtout en ce qui à trait au domaine D1. Nos travaux et ceux de Boulant et coll. semblent favoriser la présence d'hélices α alors que les travaux de Rodriguez-Casado et coll. favorisent la présence de brin β . La grande difficulté d'étude de cette protéine à la fois sous sa forme légère (monomère, dimère) et sous sa forme encapsidée ou en interaction avec l'ARN ne fait qu'augmenter l'attente d'une réponse pouvant nous éclaircir sur la réelle structure adoptée par le Core. Nous avons mis en place un protocole de production et de purification efficace de la protéine C170, et effectué des expériences préliminaires laissant présager une élucidation prochaine de la structure 3D de la protéine C170. Par ailleurs, nos travaux sur l'inhibition de l'assemblage viral pourraient être exploités pour l'étude structurale de la protéine Core en association avec l'ARN. En effet, l'ajout d'un court peptide à la solution pourrait peut-être permettre l'obtention d'un complexe protéine-ARN étudiable par RMN et ou cristallographie des rayons X. Concernant, l'interaction de la Core avec des partenaires, des études par RMN pourraient être menées avec plusieurs protéines cibles que l'on a citées en introduction de cette thèse (Tableau 1.1). Enfin, l'outil développé au sein du laboratoire pour mesurer l'assemblage viral, associé à des tests *in vivo* (système cellulaire, [160, 251, 278]) pourrait être utilisé à grande échelle pour le criblage de banques de produits naturels ou autres (peptides dérivés d'études d'interaction, de produits chimiques), dans le but de trouver de potentiels inhibiteurs de l'assemblage viral. De plus, la présence d'une région IUP au sein de la Core du VHC serait selon Chen et coll. [46], un avantage pour la découverte de molécules liant la Core et ayant potentiellement un effet inhibiteur sur ses divers rôles au sein du cycle viral.

Annexes

Annexe 1

Structure and dynamics of the N-terminal half of hepatitis C virus Core protein: an intrinsically unstructured protein: Supplementary Data

Authors : Jean-Baptiste Duvignaud, Christian Savard, Remi Fromentin, Nathalie Majeau, Denis Leclerc and Stéphane M. Gagné

Relaxation analysis

For relaxation studies, three experiments were carried: ^{15}N -T₁, ^{15}N -T₂ and $\{^1\text{H}\}$ - ^{15}N NOE. ^{15}N -T₁ delay times of 10.9, 21.8, 43.5, 87.0, 173.9, 347.8, 695.7, 1391.3, and 1989.1 ms were used. ^{15}N -T₂ delay times of 10, 30, 50, 70, 90, 130, 170, 210, and 250 ms were recorded. For $\{^1\text{H}\}$ - ^{15}N NOE experiments, a pulse sequence obtained from L.E. Kay was used, where one spectrum was acquired with a 3 s ^1H saturation time for the NOE to build up and another spectrum was acquired with a 3 s recycle delay.

Determination of ^{15}N R₁ and ^{15}N R₂ relaxation rates was accomplished using CURVEFIT (A. G. Palmer, Columbia University, New-York, NY). $\{^1\text{H}\}$ - ^{15}N NOE values were obtained directly from the HetNOE analysis function in NMRView. A noise floor of 2.2 %, 2.6 % and 4.0 % was used in the calculation of ^{15}N R₁, ^{15}N R₂ and $\{^1\text{H}\}$ - ^{15}N NOE. Analysis of relaxation data was performed using the extended model-free formalism, using the statistical approach of Mandel & al. No global correlation time (τ_m) was assumed in the model-free analysis, as C82 is not globular under the experimental conditions. We analyzed the dynamical motions of this sample using a local correlation time (τ_{loc}) for each backbone amide independently. Values for the ^{15}N gyromagnetic ratio, H-N bond length, and chemical shift anisotropy were $-2.712 * 10^7 \text{ rad T}^{-1} \text{ s}^{-1}$, 1.02 Å, and -172 ppm, respectively

Tableau A1.1 : Relaxation data of HCV C82 protein.

#	AA	$R_{1(600\text{MHz})} (\text{s}^{-1})$	$R_{1(600\text{MHz})} (\text{s}^{-1})$	$\text{NOE}_{(600\text{MHz})}$	S^2	τ_c (ps)	τ_{loc} (ns)
1	M						
2	S						
3	T						
4	N	1,575 ± 0,021	5,508 ± 0,062	0,103 ± 0,009	0,584 ± 0,010	129,973 ± 6,602	5,496 ± 0,110
5	P						
6	K	1,527 ± 0,040	2,752 ± 0,043	-0,277 ± 0,006	0,401 ± 0,008	129,437 ± 5,482	3,117 ± 0,120
7	P						
8	Q	1,684 ± 0,016	3,536 ± 0,136	0,188 ± 0,003			
9	R						
10	K						
11	T	1,683 ± 0,019	4,560 ± 0,173	0,257 ± 0,004	0,578 ± 0,011	96,765 ± 4,646	4,367 ± 0,133
12	K	1,753 ± 0,032	4,651 ± 0,215	0,145 ± 0,004	0,579 ± 0,014	135,426 ± 12,087	4,385 ± 0,177
13	R	1,618 ± 0,041	5,124 ± 0,198	0,173 ± 0,004	0,583 ± 0,014	115,581 ± 6,815	5,021 ± 0,155
14	N						
15	T	1,687 ± 0,114	4,949 ± 0,204	0,261 ± 0,005			
16	N	1,786 ± 0,108	5,555 ± 0,273	0,209 ± 0,004			
17	R	1,636 ± 0,045	4,756 ± 0,194	0,289 ± 0,007	0,585 ± 0,013	89,332 ± 4,954	4,615 ± 0,158
18	R						
19	P						
20	Q	1,640 ± 0,028	6,201 ± 0,128	0,342 ± 0,007	0,660 ± 0,012	103,556 ± 6,067	5,599 ± 0,106
21	D						
22	V	1,574 ± 0,017	4,885 ± 0,085	0,297 ± 0,007	0,579 ± 0,010	83,028 ± 3,653	4,849 ± 0,108
23	K	1,708 ± 0,021	4,756 ± 0,030	0,315 ± 0,005	0,602 ± 0,011	91,984 ± 4,319	4,429 ± 0,101
24	F						
25	P						
26	G	1,689 ± 0,046	5,048 ± 0,114	0,233 ± 0,005	0,600 ± 0,012	111,624 ± 6,640	4,758 ± 0,113
27	G	1,642 ± 0,017	4,809 ± 0,143	0,252 ± 0,002	0,582 ± 0,011	97,386 ± 4,645	4,657 ± 0,109
28	G						
29	Q	1,624 ± 0,017	4,873 ± 0,106	0,170 ± 0,004	0,570 ± 0,010	111,049 ± 5,156	4,812 ± 0,106
30	I	1,590 ± 0,023	4,518 ± 0,035	0,193 ± 0,005	0,550 ± 0,010	96,604 ± 3,946	4,587 ± 0,102
31	V	1,705 ± 0,020	4,907 ± 0,039	0,155 ± 0,006	0,587 ± 0,011	129,692 ± 7,276	4,662 ± 0,107
32	G	1,756 ± 0,051	6,169 ± 0,142	0,203 ± 0,005			
33	G	1,638 ± 0,010	5,501 ± 0,154	0,312 ± 0,005	0,624 ± 0,012	97,242 ± 5,150	5,149 ± 0,109
34	V	1,592 ± 0,013	4,832 ± 0,047	0,252 ± 0,004	0,574 ± 0,011	91,488 ± 3,831	4,806 ± 0,105
35	Y	1,648 ± 0,010	5,338 ± 0,048	0,206 ± 0,004	0,604 ± 0,011	118,266 ± 6,401	5,097 ± 0,107
36	L	1,735 ± 0,022	5,724 ± 0,063	0,301 ± 0,005	0,653 ± 0,012	119,581 ± 7,275	5,102 ± 0,111
37	L	1,678 ± 0,032	5,267 ± 0,063	0,211 ± 0,005	0,608 ± 0,011	120,991 ± 6,841	4,968 ± 0,110
38	P						
39	R	1,712 ± 0,057	5,188 ± 0,059	0,216 ± 0,004	0,609 ± 0,013	123,189 ± 9,484	4,828 ± 0,135
40	R						
41	G	1,605 ± 0,020	4,730 ± 0,101	0,134 ± 0,004	0,557 ± 0,011	113,247 ± 5,275	4,772 ± 0,104

... (suite)

#	AA	$R_{1(600\text{MHz})} (\text{s}^{-1})$	$R_{1(600\text{MHz})} (\text{s}^{-1})$	$\text{NOE}_{(600\text{MHz})}$	S^2	τ_c (ps)	τ_{loc} (ns)
42	P						
43	R	1,669 ± 0,019	5,068 ± 0,110	0,208 ± 0,004	0,595 ± 0,011	115,212 ± 5,959	4,837 ± 0,100
44	L	1,693 ± 0,073	4,746 ± 0,099	0,133 ± 0,006			
45	G	1,668 ± 0,059	4,436 ± 0,156	0,056 ± 0,006	0,543 ± 0,013	134,160 ± 12,692	4,452 ± 0,172
46	V	1,599 ± 0,026	4,435 ± 0,065	0,114 ± 0,005	0,537 ± 0,010	110,203 ± 4,789	4,545 ± 0,108
47	R	1,647 ± 0,015	4,900 ± 0,060	0,104 ± 0,003	0,569 ± 0,013	132,121 ± 13,247	4,819 ± 0,112
48	A	1,666 ± 0,020	4,925 ± 0,120	0,071 ± 0,003			
49	T	1,542 ± 0,014	4,792 ± 0,160	0,190 ± 0,005	0,555 ± 0,011	96,042 ± 4,118	4,943 ± 0,127
50	R	1,721 ± 0,047	4,979 ± 0,082	0,112 ± 0,004			
51	K	1,768 ± 0,040	4,755 ± 0,217	0,276 ± 0,002	0,609 ± 0,013	108,091 ± 6,357	4,320 ± 0,153
52	T	1,712 ± 0,048	5,257 ± 0,112	0,236 ± 0,004	0,618 ± 0,012	120,898 ± 7,688	4,867 ± 0,115
53	S	1,701 ± 0,076	5,311 ± 0,168	0,129 ± 0,004			
54	E	1,716 ± 0,046	5,233 ± 0,168	0,120 ± 0,006			
55	R						
56	S	1,661 ± 0,032	4,642 ± 0,134	0,132 ± 0,004	0,560 ± 0,010	119,019 ± 5,577	4,577 ± 0,110
57	Q	1,663 ± 0,045	4,639 ± 0,118	0,076 ± 0,003	0,552 ± 0,012	131,118 ± 8,486	4,623 ± 0,127
58	P						
59	R	1,663 ± 0,011	4,913 ± 0,173	0,237 ± 0,004	0,590 ± 0,012	104,454 ± 5,497	4,711 ± 0,124
60	G	1,801 ± 0,149	4,736 ± 0,163	0,086 ± 0,008			
61	R	1,668 ± 0,040	5,340 ± 0,145	0,299 ± 0,004	0,619 ± 0,012	100,865 ± 5,654	4,966 ± 0,111
62	R	1,791 ± 0,060	3,775 ± 0,211	0,134 ± 0,007	0,547 ± 0,013	118,933 ± 8,692	3,463 ± 0,235
63	Q	1,728 ± 0,029	5,009 ± 0,090	0,318 ± 0,004	0,618 ± 0,012	98,081 ± 5,225	4,588 ± 0,109
64	P						
65	I	1,594 ± 0,046	5,259 ± 0,070	0,107 ± 0,004			
66	P						
67	K	1,575 ± 0,021	5,508 ± 0,062	0,103 ± 0,009	0,584 ± 0,010	129,973 ± 6,602	5,496 ± 0,110
68	A	1,638 ± 0,057	5,200 ± 0,038	0,158 ± 0,006	0,587 ± 0,012	123,041 ± 8,491	5,056 ± 0,144
69	R	1,584 ± 0,044	5,405 ± 0,189	0,125 ± 0,007	0,583 ± 0,012	125,258 ± 8,077	5,367 ± 0,142
70	R	1,663 ± 0,014	5,114 ± 0,094	0,170 ± 0,004	0,591 ± 0,011	123,014 ± 6,324	4,926 ± 0,107
71	P						
72	E	1,603 ± 0,029	4,857 ± 0,050	0,261 ± 0,003	0,578 ± 0,010	91,437 ± 3,841	4,788 ± 0,102
73	G	1,688 ± 0,044	4,738 ± 0,189	0,121 ± 0,007	0,570 ± 0,013	129,315 ± 8,133	4,601 ± 0,148
74	R	1,666 ± 0,016	5,239 ± 0,172	0,316 ± 0,004	0,616 ± 0,013	94,984 ± 5,206	4,891 ± 0,125
75	T	1,767 ± 0,038	5,353 ± 0,151	0,218 ± 0,004	0,628 ± 0,012	137,655 ± 8,912	4,821 ± 0,113
76	W	1,720 ± 0,020	5,378 ± 0,125	0,250 ± 0,005	0,625 ± 0,012	120,525 ± 7,026	4,922 ± 0,108
77	A	1,701 ± 0,013	4,939 ± 0,119	0,354 ± 0,009	0,614 ± 0,011	86,366 ± 4,372	4,551 ± 0,102
78	Q	1,633 ± 0,035	4,774 ± 0,051	0,145 ± 0,008	0,565 ± 0,011	116,134 ± 5,646	4,738 ± 0,108
79	P						
80	G	1,670 ± 0,024	5,171 ± 0,148	0,192 ± 0,006	0,597 ± 0,011	120,286 ± 6,538	4,923 ± 0,116
81	Y	1,658 ± 0,012	4,955 ± 0,130	0,388 ± 0,008	0,610 ± 0,012	75,506 ± 3,703	4,654 ± 0,109
82	P						

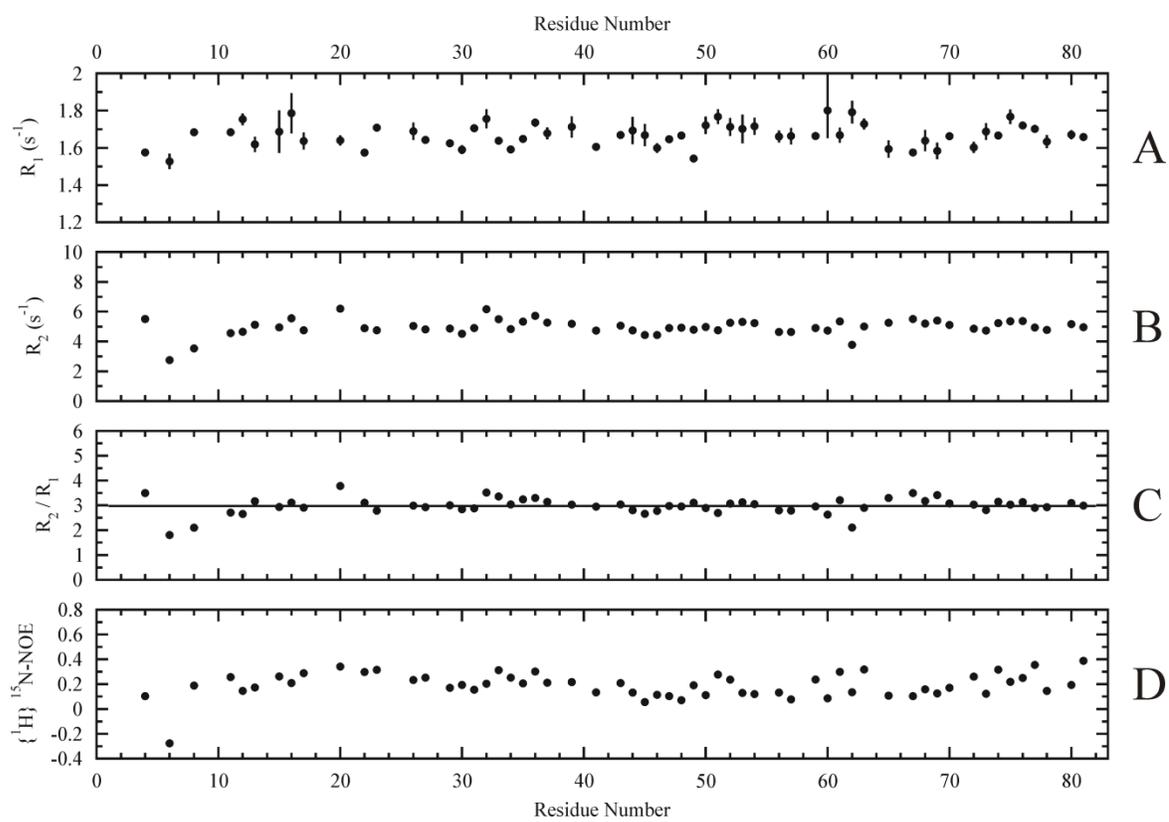


Figure A1.1 : Sequential representation of NMR dynamic data obtained for C82. A) ^{15}N R_1 , B) ^{15}N R_2 , C) R_2/R_1 ratio, D) $\{^1\text{H}\}$ - ^{15}N NOE

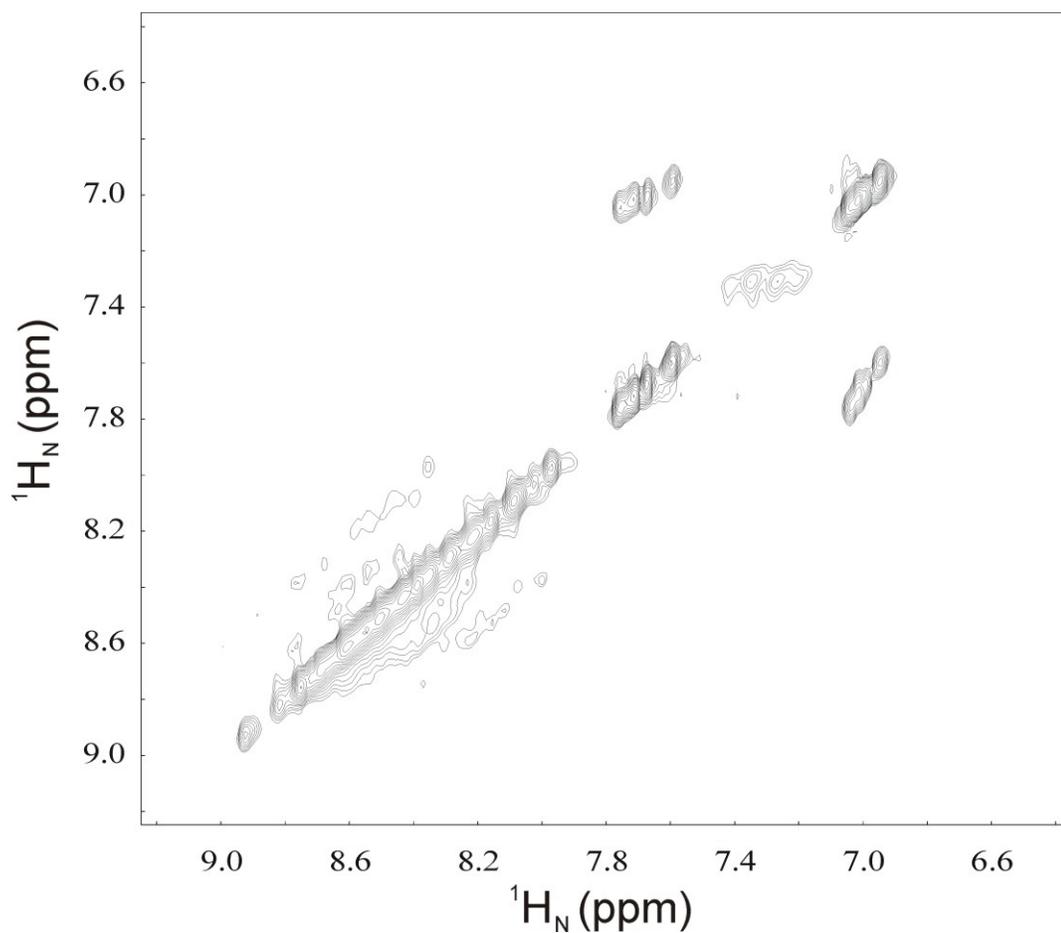


Figure A1.2 : 2D ^{15}N -NOESY spectra of hepatitis Core protein : C82.

NMR was used to evaluate tertiary and secondary structure of hepatitis Core protein C82 in an aqueous solution. 2D ^{15}N -NOESY-HSQC spectra were recorded at 5°C. The 2D NOESY spectrum correlates protons that are within a distance of 5 Å in the structure. Some of these NOE correlations are characteristic of secondary structure in folded proteins (e.g. strong HN-HN correlations are expected in α -helical structure, weaker ones in β -sheet structures, and typically none in disordered proteins). The 2D NOESY spectra of C82 contains very few to no NOE correlations which is indicative of the absence of secondary structure.

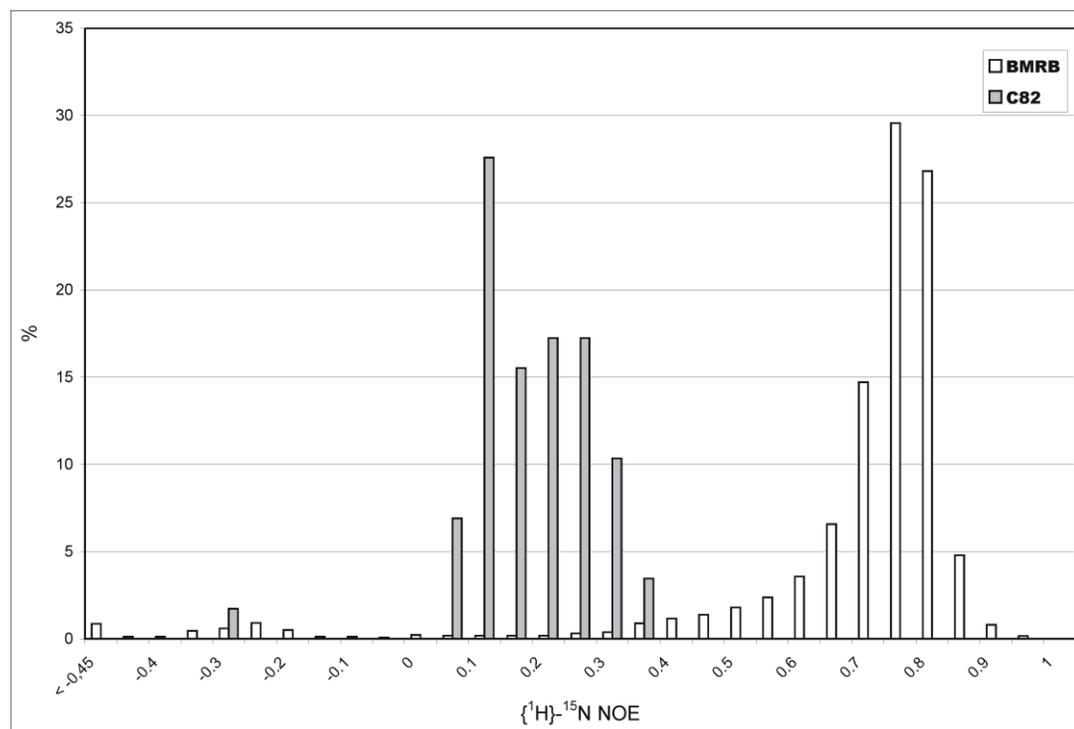


Figure A1.3 : Comparison of the distribution of heteronuclear NOEs measured for C82 in “water condition” (grey bars) to the distribution of heteronuclear NOEs found in the BMRB (white bars).

For $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE experiments, a pulse sequence obtained from L.E. Kay was used. To measure the $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOEs, one spectrum was acquired with a 3 sec ^1H saturation time for the NOE to build up and another spectrum was acquired with a 3 sec recycle delay. $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE values were obtained directly from the HetNOE analysis function in NMRView. The NOE values clearly reflect an unfolded protein with unrestricted fast dynamics. In order to compare these NOE values to that of other proteins, we have analyzed the BMRB chemical shift databank (<http://www.bmrwisc.edu>) and extracted statistics for heteronuclear NOEs at 600 MHz. It should be noted that proteins found in the BMRB are almost entirely well-structured proteins. Comparison of the measured $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE for C82 with those of the BMRB (white) indicates important differences. The heteronuclear

NOEs for C82 are all very low with an average of 0.19 ± 0.10 . In comparison, heteronuclear NOEs of structured proteins in the BMRB have high values with an average of 0.70 ± 0.31 (n=4509) using all data and 0.75 ± 0.09 (n=4231) using data within one standard deviation. These results are an unambiguous experimental evidence of the high level of disorder on the ns-ps time scale for the C82 sample.

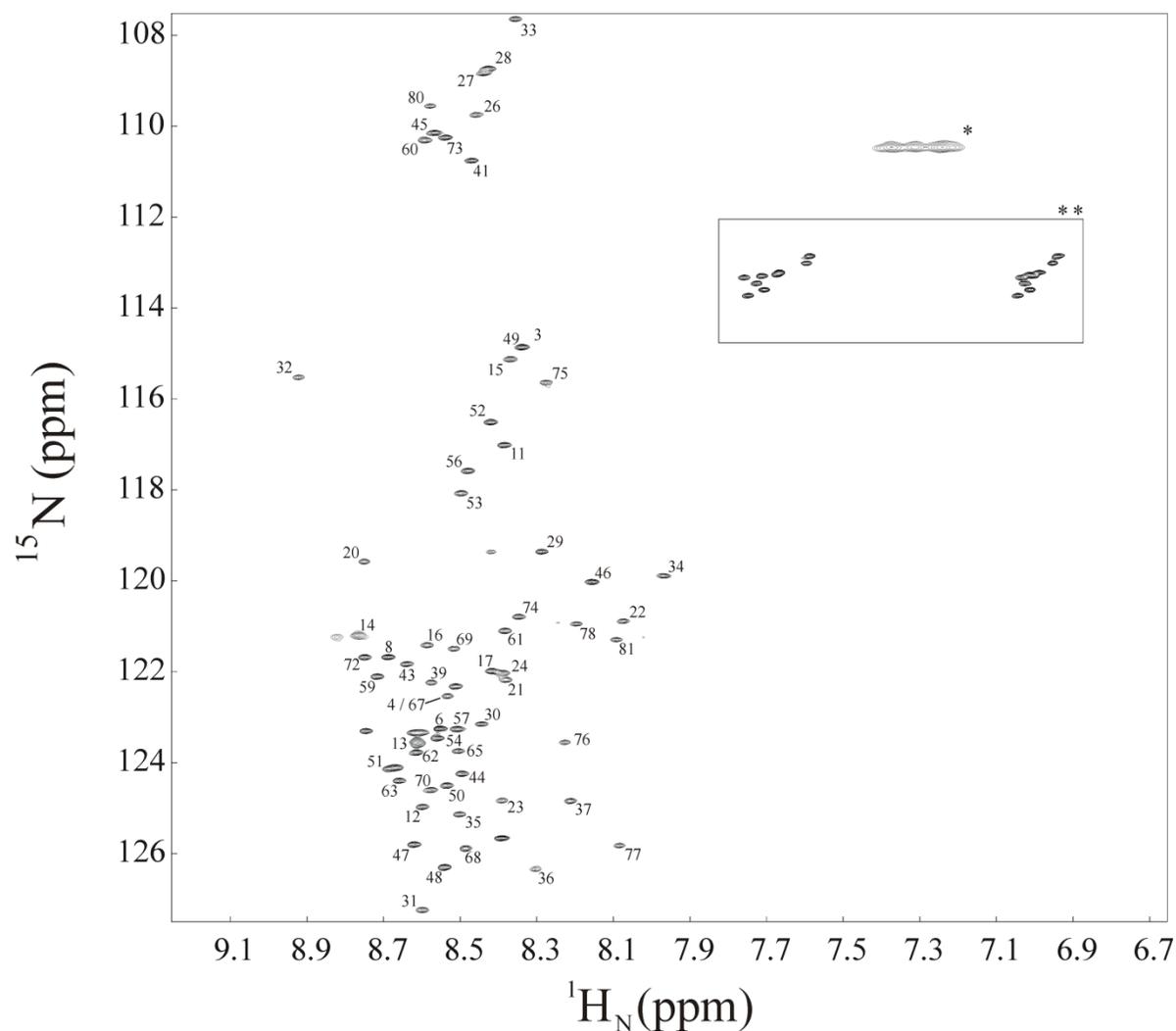


Figure A1.4 : 2D ^1H - ^{15}N HSQC spectrum of HCV C82 protein.

Annexe 2

Structure and dynamics changes induced by TFE on of the N-terminal half of hepatitis C virus Core protein: Supplementary Data

Authors : Jean-Baptiste Duvignaud, Denis Leclerc and Stéphane M. Gagné

Relaxation analysis

For relaxation studies, three experiments were carried: ^{15}N - T_1 , ^{15}N - T_2 and $\{^1\text{H}\}$ - ^{15}N NOE. ^{15}N - T_1 delay times of 10.9, 21.8, 43.5, 87.0, 173.9, 347.8, 695.7, 1391.3, and 1989.1 ms were used. ^{15}N - T_2 delay times of 10, 30, 50, 70, 90, 130, 170, 210, and 250 ms were recorded. For $\{^1\text{H}\}$ - ^{15}N NOE experiments, a pulse sequence obtained from L.E. Kay was used, where one spectrum was acquired with a 3 sec ^1H saturation time for the NOE to build up and another spectrum was acquired with a 3 s recycle delay.

Determination of ^{15}N - R_1 and ^{15}N - R_2 relaxation rates was accomplished using CURVEFIT (A. G. Palmer, Columbia University, New-York, NY). $\{^1\text{H}\}$ - ^{15}N NOE values were obtained directly from the HetNOE analysis function in NMRView. A noise floor of 2.5 %, 3.5 % and 5.5 % was used in the calculation of ^{15}N - R_1 , ^{15}N - R_2 and $\{^1\text{H}\}$ - ^{15}N NOE. Analysis of relaxation data was performed using the extended model-free formalism, using the statistical approach of Mandel & al [180]. As C82 is evidently folded, an initial global correlation time (τ_m) was derived using an S^2 - τ_e model. Values for the ^{15}N gyromagnetic ratio, H-N bond length, and chemical shift anisotropy were $-2.712 * 10^7 \text{ rad T}^{-1} \text{ s}^{-1}$, 1.02 Å, and -172 ppm, respectively.

Tableau A2.2 : Relaxation data of HCV C82 protein in 90 % TFE.

#	AA	$R_{1(600\text{MHz})} (\text{s}^{-1})$	$R_{1(600\text{MHz})} (\text{s}^{-1})$	$\text{NOE}_{(600\text{MHz})}$	S^2	$\tau_c (\text{ps})$	$R_{ex} (\text{s}^{-1})$
1	M						
2	S						
3	T	1,696 ± 0,019	7,299 ± 0,177	0,294 ± 0,036	0,382 ± 0,019	1181,028 ± 47,832	
4	N	1,809 ± 0,149	5,684 ± 0,272	0,423 ± 0,047	0,245 ± 0,023	1576,368 ± 107,355	
5	P						
6	K	1,653 ± 0,056	7,557 ± 0,185	0,553 ± 0,053	0,394 ± 0,019	1780,169 ± 162,663	
7	P						
8	Q						
9	R						
10	K						
11	T	1,411 ± 0,013	12,118 ± 0,333	0,508 ± 0,035	0,802 ± 0,022	763,680 ± 102,015	
12	K	1,422 ± 0,018	12,689 ± 0,188	0,628 ± 0,025	0,849 ± 0,021	920,992 ± 157,457	
13	R	1,424 ± 0,038	12,528 ± 0,893	0,540 ± 0,049	0,831 ± 0,028	755,504 ± 142,093	
14	N	1,454 ± 0,055	13,638 ± 0,500	0,533 ± 0,021	0,847 ± 0,026	653,695 ± 214,150	
15	T	1,402 ± 0,063	12,015 ± 0,195	0,487 ± 0,019	0,778 ± 0,027	834,696 ± 109,603	
16	N	1,436 ± 0,023	11,818 ± 0,181	0,509 ± 0,020	0,781 ± 0,021	862,585 ± 87,255	
17	R						
18	R	1,463 ± 0,050	12,661 ± 0,404	0,584 ± 0,041	0,812 ± 0,027	962,978 ± 143,264	
19	P						
20	Q	1,279 ± 0,038	16,971 ± 0,338	0,566 ± 0,039	0,874 ± 0,022	155,949 ± 69,327	3,443 ± 0,664
21	D	1,312 ± 0,091	13,911 ± 0,657	0,597 ± 0,076	0,886 ± 0,027	324,431 ± 271,785	
22	V	1,457 ± 0,045	12,632 ± 0,307	0,531 ± 0,017	0,805 ± 0,025	858,969 ± 122,011	
23	K	1,409 ± 0,027	12,688 ± 0,404	0,585 ± 0,017	0,843 ± 0,022	808,244 ± 140,104	
24	F	1,374 ± 0,045	11,039 ± 0,796	0,511 ± 0,049	0,688 ± 0,049	982,126 ± 147,816	
25	P						
26	G	1,450 ± 0,017	11,455 ± 0,184	0,456 ± 0,012	0,751 ± 0,021	847,694 ± 83,332	
27	G	1,408 ± 0,015	14,263 ± 0,299	0,427 ± 0,013	0,809 ± 0,026	594,749 ± 131,314	1,529 ± 0,512
28	G	1,402 ± 0,019	14,279 ± 0,212	0,420 ± 0,010	0,809 ± 0,024	579,366 ± 120,708	1,552 ± 0,529
29	Q	1,359 ± 0,020	16,198 ± 0,361	0,528 ± 0,016	0,871 ± 0,018	414,787 ± 186,432	2,607 ± 0,567
30	I	1,355 ± 0,100	17,159 ± 1,528	0,556 ± 0,081	0,852 ± 0,041	469,562 ± 329,541	3,877 ± 1,357
31	V						
32	G	1,281 ± 0,022	20,846 ± 1,091	0,491 ± 0,019	0,856 ± 0,019	178,921 ± 63,891	7,609 ± 1,001
33	G	1,356 ± 0,009	19,383 ± 0,548	0,485 ± 0,015	0,859 ± 0,018	406,845 ± 154,650	6,036 ± 0,682
34	V						
35	Y						
36	L						
37	L						
38	P						
39	R						
40	R						
41	G	1,314 ± 0,037	18,212 ± 1,093	0,371 ± 0,019	0,826 ± 0,012	282,184 ± 96,754	5,360 ± 0,882

... (suite)

	AA	$R_{1(600\text{MHz})}(\text{s}^{-1})$	$R_{1(600\text{MHz})}(\text{s}^{-1})$	$\text{NOE}_{(600\text{MHz})}$	S^2	$\tau_c(\text{ps})$	$R_{ex}(\text{s}^{-1})$
42	P						
43	R						
44	L						
45	G						
46	V						
47	R						
48	A						
49	T						
50	R						
51	K						
52	T						
53	S						
54	E						
55	R						
56	S						
57	Q						
58	P						
59	R	1,297 ± 0,019	17,159 ± 2,314	0,433 ± 0,042	0,846 ± 0,016	238,136 ± 83,802	
60	G	1,376 ± 0,028	13,024 ± 0,141	0,476 ± 0,011	0,840 ± 0,020	527,304 ± 145,967	
61	R	1,389 ± 0,016	13,427 ± 0,170	0,493 ± 0,010	0,850 ± 0,020	515,627 ± 166,508	
62	R						
63	Q						
64	P						
65	I	1,411 ± 0,033	12,718 ± 0,161	0,425 ± 0,011	0,804 ± 0,023	614,574 ± 112,919	
66	P						
67	K	1,409 ± 0,020	12,943 ± 0,246	0,432 ± 0,014	0,816 ± 0,023	576,629 ± 128,847	
68	A	1,429 ± 0,025	13,293 ± 0,255	0,456 ± 0,010	0,821 ± 0,024	611,794 ± 129,170	
69	R	1,376 ± 0,035	14,243 ± 0,223	0,464 ± 0,014	0,856 ± 0,015	395,399 ± 151,146	
70	R	1,388 ± 0,020	13,372 ± 0,335	0,531 ± 0,012	0,859 ± 0,020	568,678 ± 173,122	
71	P						
72	E	1,355 ± 0,050	15,840 ± 0,382	0,603 ± 0,030	0,892 ± 0,020	429,770 ± 238,691	1,958 ± 0,617
73	G	1,408 ± 0,015	14,000 ± 0,459	0,524 ± 0,015	0,864 ± 0,020	536,481 ± 183,382	
74	R	1,414 ± 0,014	14,101 ± 0,355	0,546 ± 0,010	0,869 ± 0,021	574,249 ± 186,187	
75	T	1,354 ± 0,012	14,375 ± 0,324	0,576 ± 0,016	0,895 ± 0,014	377,602 ± 189,868	
76	W	1,403 ± 0,042	14,422 ± 0,323	0,539 ± 0,019	0,879 ± 0,017	438,990 ± 189,239	
77	A	1,442 ± 0,043	13,373 ± 0,262	0,638 ± 0,024	0,863 ± 0,024	951,363 ± 185,761	
78	Q	1,450 ± 0,014	10,546 ± 0,150	0,449 ± 0,014	0,643 ± 0,027	1014,180 ± 74,450	
79	P						
80	G	1,383 ± 0,017	12,824 ± 0,392	0,512 ± 0,024	0,840 ± 0,022	617,288 ± 154,142	
81	Y						
82	P						

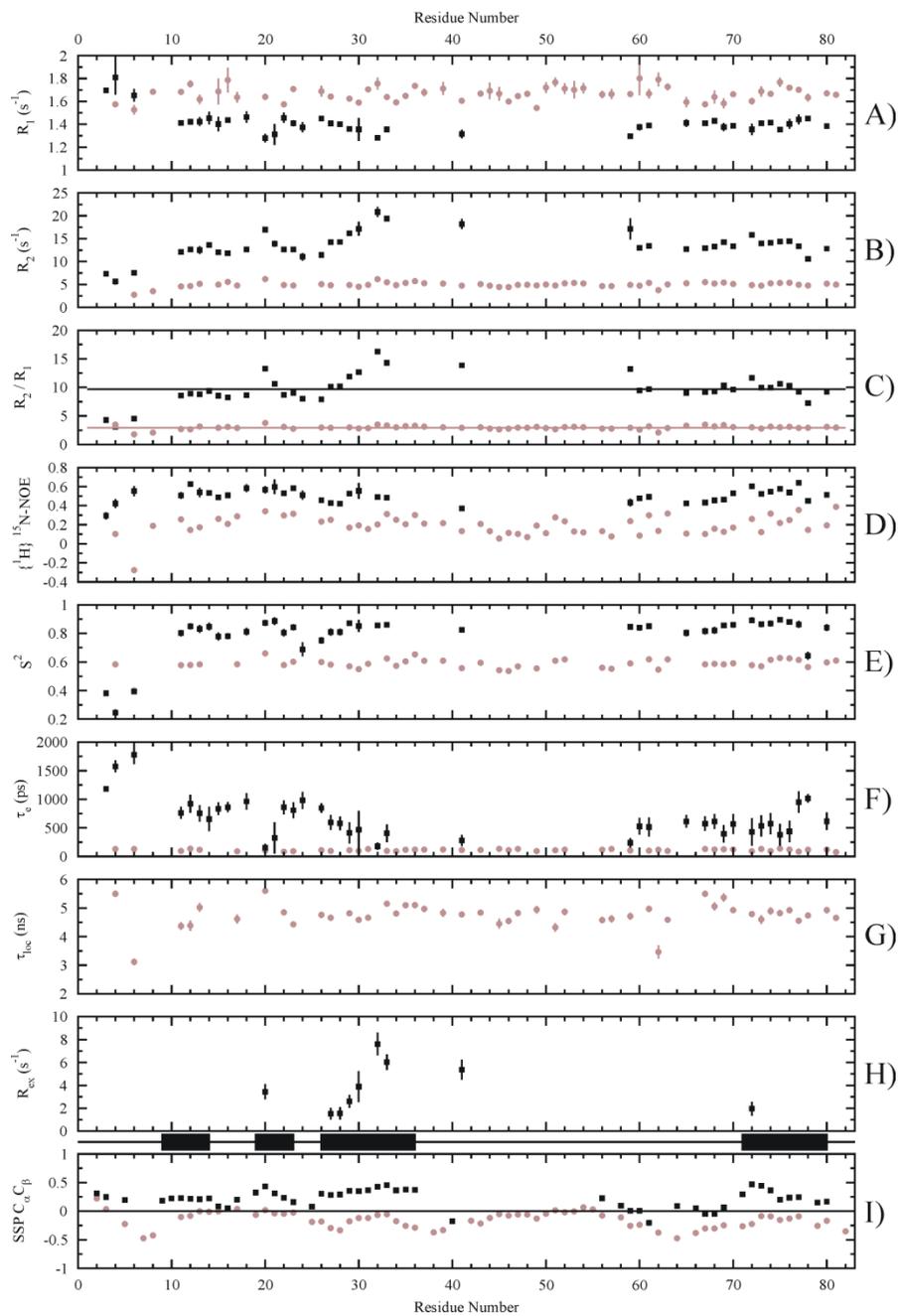


Figure A2.1 : Sequential representation of NMR dynamic data obtained for C82 in H₂O (grey) compared to 90 % TFE (black).

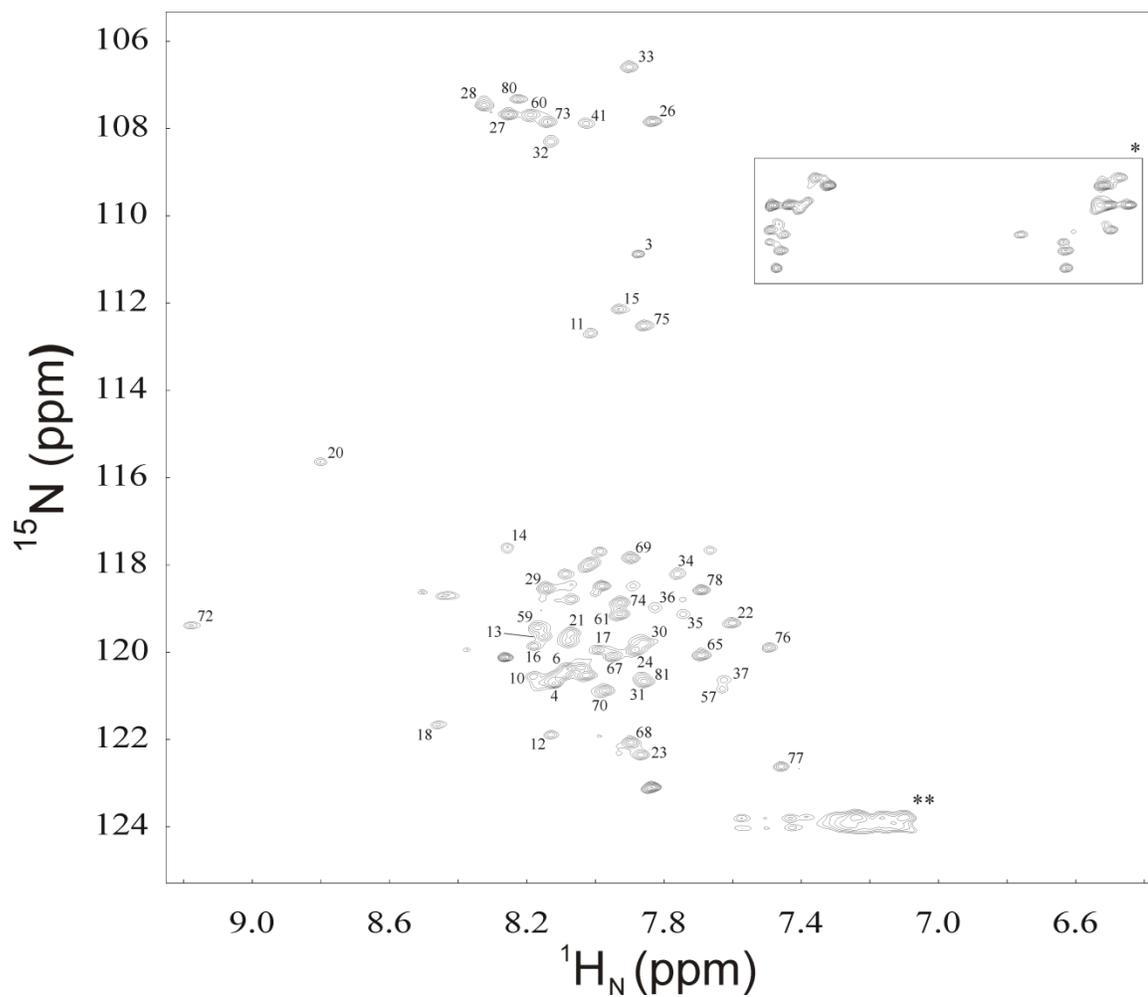


Figure A2.2 : Large 2D ^1H - ^{15}N HSQC spectra of C82 in 90 % TFE with assignments. Not shown in this expansion is the NH of the lateral chain of a tryptophan residue at 9.68 ppm (HN) and 126.23 ppm (NH). * Correlations for the side chain of Gln and Asn amino acids. ** Correlations of the side chain of Arg amino acid.

Annexe 3

Interfering with HCV assembly *in vitro* using affinity peptides directed towards Core protein : Supplementary data

Authors : Jean-Baptiste Duvignaud, Nathalie Majeau, Priscilla Delisle, Normand Voyer, Stéphane M. Gagné and Denis Leclerc

Inhibitory test for the “6929” derived peptides.

The result concerning the small peptide “6929” derived from the HCV core protein sequence obtained in the work of Fromentin et al. [82], gives us the opportunity to establish a strategy to downsize it, in order to find the minimum length of the peptide implicated in the inhibition of the assembly. So we designed 41 smaller peptides (see list figures below). The results are presented here. It clearly shows that downsizing the length of the peptide do not conserve the inhibitory effect awaited.

Materials and methods

Kinetics of C170 *in vitro* assembly reactions

In this particular test, C170 protein (8 μg) and HCV-IRES were diluted in PBS in respectively 60 μl and 40 μl , then mixed in a microcuvette (Uvette eppendorf) to initiate the assembly reaction with a ratio protein:HCV-IRES of **10:1** ($\mu\text{g}/\mu\text{g}$), giving a final reaction volume of 100 μl . Peptides were, as described in Fromentin et al., mixed to the HCV-IRES prior the addition of the core protein resulting at the end to a ratio Protein:Peptide:Ires of **10:1000:1** ($\mu\text{g}/\mu\text{g}/\mu\text{g}$). The assembly reaction was monitored at 350 nm in a UV/visible spectrophotometer (biochrom Ltd, Ultrospec 2100 pro) at room temperature. Absorbance was recorded every two seconds for at least ten minutes. Initial rate (first 8 seconds) and maximum absorbance was deduced of the resulted spectrum.

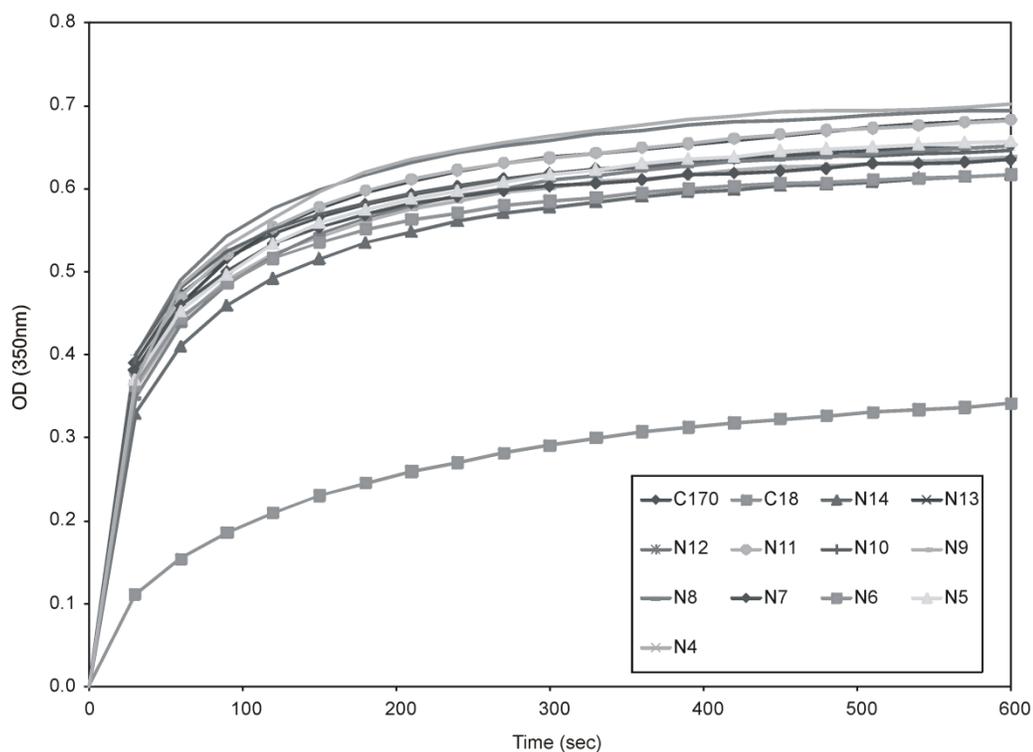
Figure Legends:

Figure S1: Assembly kinetics of C170 with the addition of peptides downsized at the C-terminal end derived from peptide “6929”. Assembly tests were done with a ratio 10:1000:1 (C170:”inhibitor”:HCV-IRES).

Figure S2: Assembly kinetics of C170 with the addition of peptides downsized at the N-terminal end derived from peptide “6929”. Assembly tests were done with a ratio 10:1000:1 (C170:”inhibitor”:HCV-IRES).

Figure S3: Assembly kinetics of C170 with the addition of peptides made of six amino acids long derived from peptide “6929”. Assembly tests were done with a ratio 10:1000:1 (C170:”inhibitor”:HCV-IRES).

Figure S1

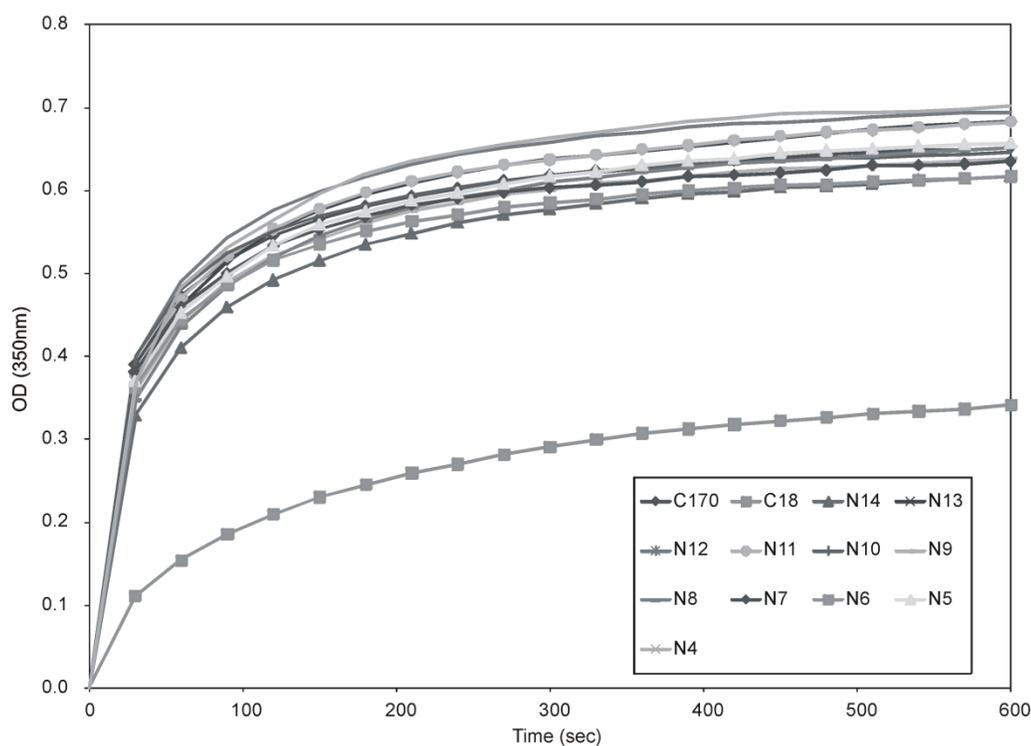


Reference			Initial Rate (OD/min)	%	Maximum Value (OD)	%
C170			1.956	100	0.652	100
From the C-terminal			Initial Rate (OD/min)	%	Maximum Value (OD)	%
C18	VKFPGGGQIVGGVYLLPR	T	0.414	21	0.342	52
C17	KFPGGGQIVGGVYLLPR	T	1.134	58	0.496	76
C16	FPGGGQIVGGVYLLPR	T	1.806	92	0.603	92
C15	PGGGQIVGGVYLLPR	NT ^{††}	-	-	-	-
C14	GGGQIVGGVYLLPR	NT ^{††}	-	-	-	-
C13	GGQIVGGVYLLPR	NT ^{††}	-	-	-	-
C12	GQIVGGVYLLPR	NT ^{††}	-	-	-	-
C11	QIVGGVYLLPR	NT ^{††}	-	-	-	-
C10	IVGGVYLLPR	NT ^{††}	-	-	-	-
C9	VGGVYLLPR	T	1.98	101	0.631	97
C8	GGVYLLPR	NT ^{††}	-	-	-	-
C7	GVYLLPR	NT ^{††}	-	-	-	-
C6	VYLLPR	T	1.95	100	0.655	100
C5	YLLPR	T	2.16	110	0.695	107
C4	LLPR	NT ^{††}	-	-	-	-
C3	LPR	NT ^{††}	-	-	-	-

T = Tested

NT^{††} = Not tested due to solubility issuesNT[†] = Not synthesised

Figure S2

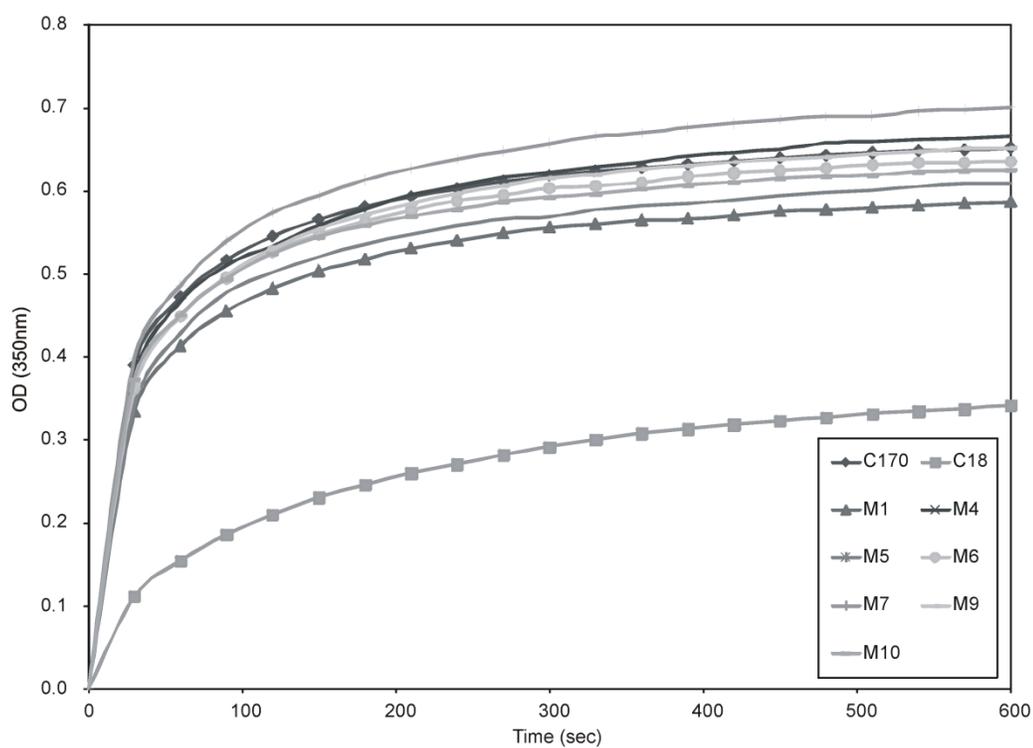


Reference			Initial Rate (OD/min)	%	Maximum Value (OD)	%
C170			1.956	100	0.652	100
From the N-terminal			Initial Rate (OD/min)	%	Maximum Value (OD)	%
N17	VKFPGGGQIVGGVYLLP	NT [†]	-	-	-	-
N16	VKFPGGGQIVGGVYLL	NT ^{††}	-	-	-	-
N15	VKFPGGGQIVGGVYL	NT ^{††}	-	-	-	-
N14	VKFPGGGQIVGGVY	T	1.578	81	0.618	95
N13	VKFPGGGQIVGGV	T	1.86	95	0.684	105
N12	VKFPGGGQIVGG	T	1.986	102	0.696	107
N11	VKFPGGGQIVG	T	1.902	97	0.683	105
N10	VKFPGGGQIV	T	2.064	106	0.646	99
N9	VKFPGGGQI	T	1.776	91	0.638	98
N8	VKFPGGGQ	T	1.68	86	0.652	100
N7	VKFPGGG	T	1.938	99	0.636	98
N6	VKFPGG	T	1.782	91	0.618	95
N5	VKFPG	T	1.836	94	0.659	101
N4	VKFP	T	1.848	94	0.703	108
N3	VKF	NT [†]	-	-	-	-

T = Tested

NT^{††} = Not tested due to solubility issuesNT[†] = Not synthesised

Figure S3



Reference			Initial Rate (OD/min)	%	Maximum Value (OD)	%
C170			1.956	100	0.652	100
Six amino acids long			Initial Rate (OD/min)	%	Maximum Value (OD)	%
M1	KFPGGG	T	1.608	82	0.587	90
M2	FPGGGQ	NT ^{††}	-	-	-	-
M3	PGGGQI	NT ^{††}	-	-	-	-
M4	GGGGIV	T	1.866	95	0.666	102
M5	GGQIVG	T	1.464	75	0.610	93
M6	GQIVGG	T	1.764	90	0.636	98
M7	QIVGGV	T	1.926	98	0.701	107
M8	IVGGVY	NT ^{††}	-	-	-	-
M9	VGGVYL	T	1.728	88	0.654	100
M10	GGVYLL	T	1.782	91	0.625	96
M11	GVYLLP	NT [†]	-	-	-	-

T = Tested

NT^{††} = Not tested due to solubility issuesNT[†] = Not synthesized

Annexe 4

A method for in vitro assembly of hepatitis C virus core protein and for screening of inhibitors

Authors : Rémi Fromentim, Nathalie Majeau, Marie-Ève Laliberté-Gagné, Annie Boivin, Jean-Baptiste Duvignaud, and Denis Leclerc

Available online at www.sciencedirect.com

Analytical Biochemistry 366 (2007) 37–45

ANALYTICAL
BIOCHEMISTRYwww.elsevier.com/locate/yabio

A method for in vitro assembly of hepatitis C virus core protein and for screening of inhibitors

Rémi Fromentin, Nathalie Majeau, Marie-Eve Laliberté Gagné, Annie Boivin,
Jean-Baptiste Duvignaud, Denis Leclerc*

Centre de Recherche en Infectiologie, Université Laval, Qué., Canada QC G1V 4G2

Received 26 January 2007
Available online 2 April 2007

Abstract

The assembly of hepatitis C virus (HCV) is not well understood. We investigated HCV nucleocapsid assembly in vitro and the role of electrostatic/hydrophobic interactions in this process. We developed a simple and rapid in vitro assay in which the progress of assembly is monitored by measuring an increase in turbidity, thereby allowing the kinetics of assembly to be determined. Assembly is performed using a truncated HCV core (C1–82), containing the minimal assembly domain, purified from *Escherichia coli*. The increase in turbidity is linked to the formation of nucleocapsid-like particles (NLPs) in solution, and nucleic acids are essential to initiate nucleocapsid assembly under the experimental conditions used. The sensitivity of NLP formation to salt strongly suggests that electrostatic forces govern in vitro assembly. Mutational analysis of C1–82 demonstrated that it is the global positive charge of C1–82 rather than any specific basic residue that is important for the assembly process. Our in vitro assembly assay provides an easy and efficient means of screening for assembly inhibitors, and we have identified several inhibitory peptides that could represent a starting point for drug design.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Hepatitis C virus; Nucleocapsid; In vitro assembly; Kinetics; Absorbance spectrometry

Hepatitis C virus (HCV)¹ is a major public health concern worldwide because more than 3% of the world's population is infected with this virus [1]. HCV is a plus-strand RNA virus that causes acute and chronic liver disease and is associated with 95% of cases of posttransfusion hepatitis [2] and more than 50% of cases of non-A, non-B hepatitis. Current therapy is the use of pegylated interferon and ribavirin, but outcomes are unsatisfactory given that only 42% of patients infected with HCV genotype 1 (the most common genotype in North America) respond positively to

treatment [3]. Therefore, there is an urgent need to identify new targets for the development of drugs to both cure and prevent the spread of the disease.

HCV core (HCV-C) protein, a 191-amino acid protein located at the N terminus of the HCV polyprotein, is the largest core protein in the family of *Flaviviridae*. The C-terminal region of HCV-C is highly hydrophobic and targets the protein to the endoplasmic reticulum (ER) membranes. This hydrophobic region is further cleaved by a host protease, the signal peptide peptidase, to generate the mature protein of 177 or 179 amino acids [4,5]. The core protein is also submitted to posttranslational modifications such as ubiquitinylation [6] and phosphorylation [7,8], which can potentially regulate its degradation in the infected cell. Little is known about HCV nucleocapsid assembly; however, based on what is known in other RNA viruses, assembly is expected to be initiated by the interaction between core protein and viral RNA. Electrostatic forces appear to be important in this phenomenon [9].

* Corresponding author. Fax: +1 418 654 2715.

E-mail address: denis.leclerc@crchul.ulaval.ca (D. Leclerc).

¹ Abbreviations used: HCV, hepatitis C virus; HCV-C, HCV core; ER, endoplasmic reticulum; 5' UTR, 5' untranslated region; IRES, internal ribosome entry site; NLP, nucleocapsid-like particle; NTA, nitrotriacetic acid; BCA, bicinchoninic acid; tRNA, transfer RNA; Bis-ANS, 4,4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid dipotassium salt; EM, electron microscopy; ssDNA, single-stranded DNA; ssRNA, single-stranded RNA.

The core protein has been shown to package the viral RNA [10], possibly through specific interactions with the 5' untranslated region (5' UTR) of the RNA genome [11,12]. The region around the 5' UTR is characterized by both secondary and tertiary structures [13–16], contributing to the internal ribosome entry site (IRES) required for initiation of cap-independent translation [17]. The N-terminal half (N-terminal 82 amino acids) of the protein is rich in positively charged residues (K and R). This domain is sufficient to trigger the formation of nucleocapsid-like particles (NLPs) *in vitro* when structured RNA is added to the purified protein [18].

A more effective assay for the detection of nucleocapsid formation is required to improve our understanding of the biochemistry of this process and to allow screening for inhibitors of assembly. Here we describe the development of a method that allows the kinetics of assembly to be followed *in vitro* by measurement of solution turbidity using a spectrophotometer. This system has revealed the critical role of electrostatic interactions in particle formation and has identified a domain within the core that could be sensitive to assembly inhibitors. This novel method can be used to perform high-throughput screening of potential assembly inhibitors such as inhibiting compounds or peptides.

Materials and methods

Cloning and expression of HCV-C proteins in Escherichia coli

The nucleotide sequence of HCV-C protein was optimized with most abundant codons for translation in bacteria [19]. This optimized sequence was used to generate the construct C1–82 as well as the other mutant forms (see Fig. 5A later). All forms of C1–82 were cloned into the pET3d expression vector (New England Biolabs), which harbors a 6-histidine tag at the C-terminal end to ease the purification process on a nickel affinity column (Qiagen). The mutated forms RR39, 40AA, RR43, 47AA, RK50, 51AA, RR55, 59AA, E54A, RR61, 62AA, Δ8–23, Δ39–62, and C1–71 were amplified from the C1–82 clone with the following primers: K23A (forward), 5', and K23A (reverse), 5'; RR39, 40AA (forward), 5'-GCGGG TCCGCGTCTGGGTGTTTCG-3', and RR39, 40AA (reverse), 5'-CGCCGGCAGCAGGTAAACACCACC-3'; RR43, 47AA (forward), 5'-GGTGTTCGCGCACCCG TAAAACCTCTGAAC-3', and RR43, 47AA (reverse), 5'-CAGCGCCGGACCACGACGCGGCAGCAGG-3'; RK50, 51AA (forward), 5'-GCGACCTCTGAACGTTTCAGCCG-3', and RK50, 51AA (reverse), 5'-CGCGGTC GCACGAACACCCAGACG-3'; E54A (forward), 5'-CG TTCTCAGCCGCGTGGTTCGTC-3', and E54A (reverse), 5'-CGCAGAGGTTTTACGGGTCGCAC-3'; RR55, 59AA (forward), 5'-CAGCCGGCGGGTCGTCGTCAGCCGATCCCG-3', and RR55, 59AA (reverse), 5'-AGACGCTT CAGAGTTTTACGGGTCGC-3'; RR61, 62AA (forward), 5'-GCGCAGCCGATCCCGAAAGCGCG-3', and RR61, 62AA (reverse), 5'-CGCACCCACGCGGCTGAGA

ACGTTC-3'; Δ8–23 (forward), 5'-TTCCCGGGTGGCG GTCAG-3', and Δ8–23 (reverse), 5'-CTGCGGTTTCGG GTTGG-3'; Δ39–62 (forward), 5'-CAGCCGATCCCGA AAGCG-3', and Δ39–62 (reverse), 5'-CGGCAGCAGGT AAACACC-3'; C1–71 8h (forward), 5'-CACCACCATCA CCACCATCACCCTAA-3', and C1–71 (reverse), 5'-CG GACGACGCGCTTTCGGGATCGGCTG-3'. Clones were circularized by ligation of the PCR products. The sequences of the clones were confirmed by DNA sequencing.

Purification of HCV-C proteins in E. coli

The expression and purification of the recombinant proteins were performed using an Ni-nitrilotriacetic acid (NTA) resin (Qiagen) as described previously [18]. HCV-C proteins were eluted in assembly buffer (1.7 mM magnesium acetate, 100 mM potassium acetate, 25 mM Hepes [pH 7.4], and 500 mM imidazol). To eliminate NLPs pre-assembled in *E. coli*, purified HCV-C proteins were ultracentrifuged for 3 h at 100,000g and 4 °C (Optima L-90K ultracentrifuge, Beckman Coulter). The purity of the protein was estimated by 10% SDS-PAGE and confirmed by Western blot. Core proteins were quantified using a bicinchoninic acid (BCA) protein assay kit (Pierce).

RNA transcripts

The plasmid used for the transcription of the IRES, corresponding to the full-length 5' UTR of HCV H77c (nt 1–374, genotype 1a generously provided by J. Bukh, National Institutes of Health) [20], was amplified by PCR with primers 5' end IRES (5'-ACTGAACTAGTGCCAGCCCCCT GATGGGGGCG-3') and 3' end IRES and cloned downstream of a T7 promoter of pBlueScript II KS (±) vector (Stratagene). PCR product with primers 5' end T7 promoter (5'-AGTGAGCGCGCGTAATACGACTCA-3') and 3' end IRES 5'-CGATGGATCCCGGTTTTCTTGA GGT TTAGGTATCGTGCTCATGGTGCACGGTGT ACGAGACC-3' was used as a template for *in vitro* transcription (T7 RiboMAX Express Large Scale RNA Production System, Promega) of IRES. Ribosomal NTP (Promega), yeast transfer RNA (tRNA, Sigma), and polynucleotides (polyU, polyA, and polyC, Amersham) were commercial preparations. All nucleic acids were dissolved in assembly buffer.

Kinetic analysis of in vitro assembly reactions

For kinetic analysis, 8 μg (760 pmol) of C1–82 was diluted in 50 μl of assembly buffer in a microcuvette (Eppendorf). Optical density was monitored at 350 nm in a spectrometer (Biochrom Ultraspec 2100pro UV/Vis) at 25 °C. Different amounts of nucleic acid in a volume of 50 μl were added to the solution and mixed vigorously in the microcuvette. Assembly was monitored with the spectrophotometer. Approximately 6 s elapsed before the first

time point was measured. Optical density was recorded by Swift II 2.0 software every 2 s for 10 min. The maximum value represents the maximum optical density obtained during this 10-min analysis. The initial speed constant was deduced from the slope of the kinetic curve for the first 10 s of the reaction. Each slope is drawn based on five points measured at 2-s intervals.

When potassium chloride (1 M), potassium nitrate (1 M), potassium acetate (1 M), sodium chloride (1 M, Sigma), 4,4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid dipotassium salt (Bis-ANS, Sigma), ethanol, methanol, or glycerol was added for kinetic analysis, each compound was mixed with C1–82 (760 pmol) in the microcuvette (maximum volume of 70 μ l) prior to the addition of RNA. The optimal amount of IRES (38 pmol) in a final reaction volume of 100 μ l was then added as described previously.

Electron microscopy

Samples collected after *in vitro* assembly were directly absorbed on 400-mesh carbon-Formvar grids (Canemco) for 5 min. The grids were washed once with filtered phosphate-buffered saline (PBS) and stained for 10 min with filtered 2% (w/v) uranyl acetate. Grids were then dried on filter paper before being examined under an electron microscope with an acceleration voltage of 60 kV at a magnification of 150,000 \times . Images were captured and treated with Digital Micrograph (version 3.8.2, Gatan).

Northwestern dot-blot

The recombinant protein C1–82 was deposited on a nitrocellulose membrane (Bio-Rad) using a vacuum to ensure homogeneity of the dots. The membrane was blocked with 5% milk in assembly buffer at room temperature for 1 h. The membrane was washed three times with assembly buffer, followed by the addition of 2.5 μ g of radio-labeled IRES. The blots were incubated overnight at room temperature. After three washes, data acquisition was performed on a PhosphoImager (ImageQuant).

Electrophoresis mobility shift assay

The RNA (IRES) was incubated with recombinant proteins at room temperature for 10 min. We used 16 pmol of RNA and 16 nmol of peptides for each reaction in *in vitro* assembly buffer. The final volume of the reaction was 20 μ l, and 4 μ l of loading dye was added to the sample before loading onto a 0.8% agarose gel. Electrophoresis was performed in Tris/acetic acid/EDTA buffer for 30 min at 90 V. The gel was stained with ethidium bromide and photographed with a digital camera (Bio-Rad).

Peptide library

The 11 peptides of the tested library were 18 amino acids in length. They were synthetic peptides corresponding to

the first 88 amino acids of HCV strain H77 (AIDS Research and Reference Reagent Program, National Institutes of Health). Each peptide was added to IRES solution at a ratio of 1000 peptides to 1 molecule of IRES. This solution was then mixed vigorously with C1–82 solution in the microcuvette to initiate assembly. The C1–82/IRES ratio was 20:1.

Results and discussion

In vitro assembly of HCV NLPs as monitored by an increase in solution turbidity

The region of the core protein corresponding to amino acids 1–82 has been shown to be the minimal domain required for assembly [18]; we have shown previously that C1–82 protein forms NLPs when tRNA is added to the protein [18]. Recombinant C1–82 protein was purified by affinity chromatography (Fig. 1A) using a 6 \times H tag located at the C terminus of the protein. The RNA used in the assembly assay was derived from the full-length 5' UTR (nt 1–374) of HCV genome H77c [20] (hereafter referred to as IRES). Purified C1–82 and the IRES RNA were mixed at a molar ratio of 20:1 (protein/RNA) with 760 pmol of protein (8 μ g) and 38 pmol of RNA (4.4 μ g) in 100 μ l of low-salt buffer. The absorbance of the two constituents and newly formed NLPs was measured using a spectrophotometer at different wavelengths after 10 min of reaction. Unlike the unassembled constituents (RNA and protein), the turbid assembly solution absorbs light at 350 nm (Fig. 1B). Thus, the assembly process was monitored at 350 nm for 10 min. As RNA was added to the solution, we observed a rapid increase in absorbance associated with the turbidity of the sample. One-half of the final absorbance value was reached in less than 30 s, indicating that assembly of C1–82 is very rapid and efficient. The *in vitro* assembly assay was repeated five times with different preparations of proteins and RNA. The average value of the plateau was 0.926 OD with a standard deviation of 0.070 OD, showing that the assay is reliable and reproducible (Fig. 1C). As expected, neither the IRES nor the C1–82 protein alone exhibited any increase in turbidity under the experimental conditions used.

To measure the speed at which the assembly reaction occurs, we determined the slope of the curve during the first 10 s of the reaction. The mean value for this constant (referred to as initial speed constant) for five repetitions using separate batches of purified C1–82 protein was estimated at 2.16 OD/min with a standard deviation of only 0.09 OD/min. The standard deviation between experiments was acceptable, confirming that this is a robust and reliable method for measurement of NLP formation. The initial speed is a good indicator of the efficiency of *in vitro* assembly. The *in vitro* assembly of alphavirus core protein, a virus structurally similar to HCV, was also shown to be a rapid process [21].

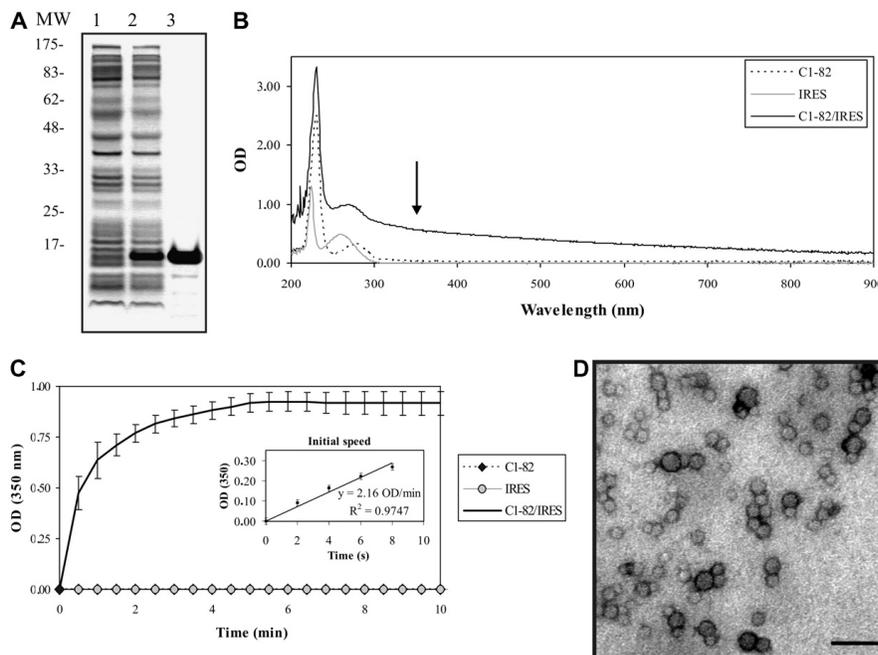


Fig. 1. (A) HCV C1-82 expression and purification: SDS PAGE (10%) of lysed bacterial cell pellets before induction (lane 1) and after 3 h induction (lane 2) with 1 mM isopropyl- β -D-thiogalactoside (IPTG) and C1-82 purified on an Ni^{2+} affinity column (lane 3). Molecular markers (in kDa) are shown on the left. (B) Absorbance spectra of C1-82 (760 pmol), IRES (38 pmol), and a mixed solution of C1-82/IRES (20:1 ratio). The arrow indicates the wavelength selected for the kinetic assay. (C) Kinetics of in vitro assembly. Data points are the means (\pm SD) of five independent experiments performed with 760 pmol C1-82 mixed with 38 pmol IRES. The kinetics of the first 10 s of the reaction and the initial speed estimation are shown in the inset. (D) Electron micrograph of negatively stained NLPs produced in the in vitro assembly assay. Bar = 100 nm.

Influence of NLP formation on increase in turbidity during assembly

To confirm that the increase in turbidity is linked to the formation of NLPs, we observed the assembly product after 20 min of reaction by electron microscopy (EM). Particles 25 ± 7 nm in diameter and uniform in appearance were produced (Fig. 1D), as expected and as reported previously [18]. No particles were seen under EM in grids with core protein only (data not shown), consistent with the previously published observation that recombinant C1-82 is a soluble and monodispersed protein (J.-B. Duvignaud et al., unpublished).

Effect of salt on in vitro assembly of HCV NLPs

To assess the role of electrostatic forces in NLP assembly, we measured the influence of increasing the concentration of potassium chloride in our assay on the initial speed constant (Fig. 2A). We observed that the increase in absorbance induced by the formation of NLPs decreases propor-

tionally with the increase in potassium chloride concentration, reaching complete inhibition with 250 mM KCl. Other salts, such as potassium nitrate, potassium acetate, and sodium chloride, also affected in vitro assembly in a concentration-dependent manner (data not shown). The inhibition of in vitro assembly of NLPs by all of these monovalent salts illustrates the fundamental role played by electrostatic forces in the assembly of HCV NLPs.

It is likely that this inhibition is caused by neutralization of positively charged residues found in C1-82 that are involved in the interaction with nucleic acids. To test this hypothesis, we performed a Northwestern dot-blot assay using different concentrations of C1-82 proteins incubated with radiolabeled IRES in different salt concentrations. As expected, we observed that increasing the salt concentration efficiently reduced the interaction between the C1-82 protein and the IRES (Fig. 2B). As proposed by van der Schoot and Bruinsma [9], electrostatic and nonspecific attractions between RNA (negatively charged) and protein (positively charged) provide the dynamic driving force for viral assembly.

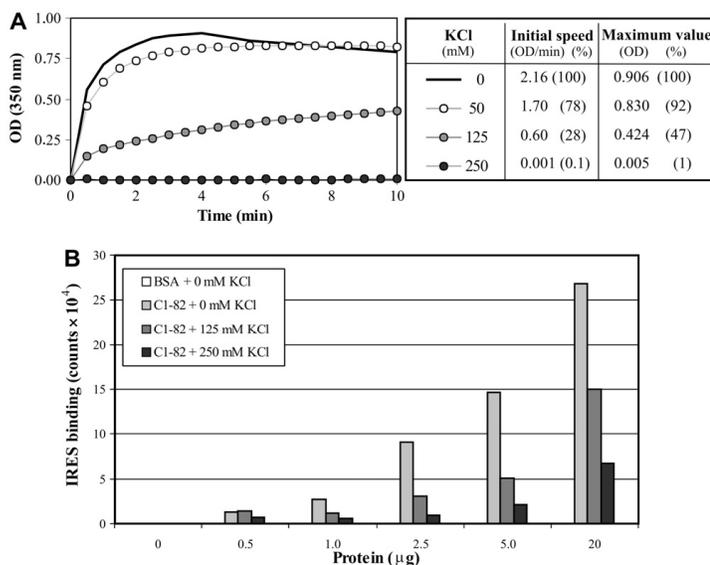


Fig. 2. Effect of salt on in vitro assembly of C1-82. (A) Kinetics of in vitro assembly of C1-82 (760 pmol) mixed with IRES RNA (38 pmol) in the presence of increasing concentrations of KCl. (B) Different quantities of C1-82 were blotted on a nitrocellulose membrane and incubated with 2.5 µg of radiolabeled IRES in the presence of increasing concentrations of KCl. Radioactivity associated with each dot was analyzed using a PhosphorImager (Typhoon 9200).

Influence of nucleic acid concentration and species on in vitro assembly of HCV NLPs

The optimal C1-82/IRES ratio for in vitro assembly was evaluated. As shown in Fig. 3A, the kinetics of assembly was optimal when a protein/RNA molar ratio of 20:1 was maintained. With higher amounts of proteins (40:1, 80:1, and 160:1 ratios) or RNA (1:1, 5:1, and 10:1 ratios), assembly was less efficient. These results are in accordance with a previous report in which a similar optimized ratio for in vitro formation of NLPs with truncated core protein (1–124) was observed [10].

We next examined the capacity of different templates to induce in vitro assembly. The following templates were tested: ribonucleotides, structured yeast tRNA, unstructured polynucleotides (200–260mer of polyU, 405–660mer of polyA, and 290–434mer of polyC), and single-stranded DNA (ssDNA) of various sizes (117 and 37 nt) (Fig. 3B). In all cases, we maintained the same protein/RNA ratio (20:1) that previously had been shown to be optimal with the IRES. All of the RNA types tested, and even ssDNA, could initiate NLP assembly. Neither monovalent cations (potassium and sodium) in any concentration (data not shown) nor ribonucleotides can act as nucleation agents, as described previously for viruses of the *Bromoviridae* family [22]. EM analysis of tRNA NLPs and polynucleotide NLPs revealed that they were similar in size and shape to IRES NLPs (data not shown). Kunkel and coworkers

[10] proposed that structured RNA is required for NLP formation. We showed that single-stranded RNA (ssRNA) can be a substrate for formation of NLP but that structured RNA appeared to be more efficient to trigger the assembly process.

Although tRNA previously has been shown to be efficient in triggering in vitro assembly of HCV-C [10,18], we found it to be less efficient than the IRES in triggering the assembly process in vitro at a 20:1 ratio (Fig. 3C); we found a 1:1 ratio to be optimal for this RNA. Therefore, if only 5 IRES (373 nt) molecules are necessary for 100 C1-82 molecules, 100 tRNAs of 75 nt are necessary for optimal assembly for the same amount of C1-82 protein. This result suggests that the size of the RNA is critical in the initiation of assembly. In summary, for a given number of C1-82 subunits, the optimal number of RNA molecules necessary to complete assembly decreases with the length of the RNA. A similar observation was made with alphavirus nucleocapsid protein [21], where the optimal molar ratio of protein/RNA varied with the length of RNA used for assembly. The initial speed constant of the optimal protein/tRNA assay was lower than that observed for the IRES RNA. It is possible that the protein has a lower affinity for tRNA than for IRES. The addition of structured or unstructured RNA, but not salt or ribonucleotides, permits electrostatic repulsion between core subunits to be overcome, allowing NLP assembly.

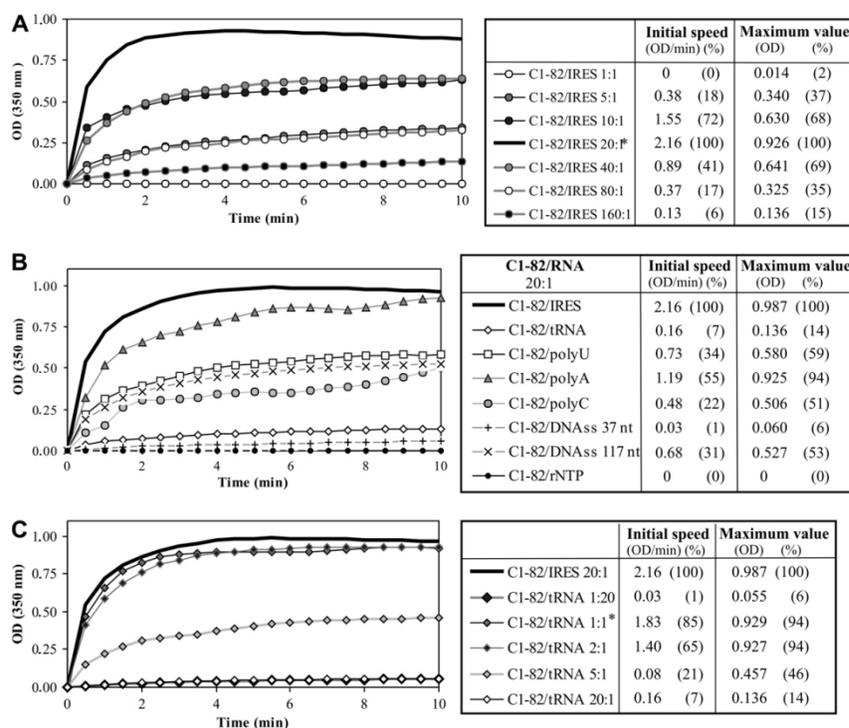


Fig. 3. Influence of nucleic acid species on in vitro assembly of C1-82: Kinetics of in vitro assembly of C1-82 (760 pmol) mixed with different concentrations of IRES RNA (A), different RNAs (38 pmol) to maintain a protein/RNA ratio of 20:1 (B), and different concentrations of tRNA (C). Asterisk (*) indicates the optimal ratio for each RNA.

Identification of basic residues critical for in vitro assembly of HCV NLPs

We showed that salt can have a detrimental effect on in vitro assembly, suggesting that charged residues of the protein are important in the recognition of the RNA to trigger the assembly process. To further investigate the importance of charged residues in NLP formation in vitro, we introduced mutations in the two positively charged regions (8–23 and 39–62) of C1-82, which are separated by a glycine-rich domain (Fig. 4A). We deleted each of these regions individually and determined the impact on assembly. The recombinant proteins were expressed and purified from *E. coli* as before (Fig. 4B). As shown in Fig. 4C, deletion of the 8–23 fragment had little effect on NLP formation if we consider only the maximum value obtained for this mutant. However, the initial speed was affected, showing a decrease of 26% as compared with that for C1-82. This suggests that deletion in the first charged amino acid cluster can impair the initial rate of particle formation but not the multimerization process itself.

In contrast, deletion of residues 39–62 drastically affected the kinetics of assembly, reducing the maximum value by half and the initial speed by a factor of 4 (Fig. 4C). This result is consistent with the previously reported observation that particle formation is decreased with a Δ 39–64 mutant in a cell-free in vitro assembly system [23]. To investigate whether specific regions in the 39–62 fragment are more important for assembly, we mutated several clusters of charged residues. The concentration and purity of the mutant proteins produced were roughly the same for all proteins (Fig. 4B). Most of the mutants were as capable of in vitro assembly as was C1-82 (Fig. 4C). However, the mutant proteins RR43–47AA and RR61–62AA showed an 11% decrease in efficiency of NLP formation and decreases of 27 and 11%, respectively, in the initial speed of assembly (Fig. 4C). None of these specific mutations was as drastic as the Δ 39–64 deletion mutant. Given that no specific region playing a key role in NLP assembly could be identified, the global charge of the protein is probably the driving force for multimerization and the assembly process.

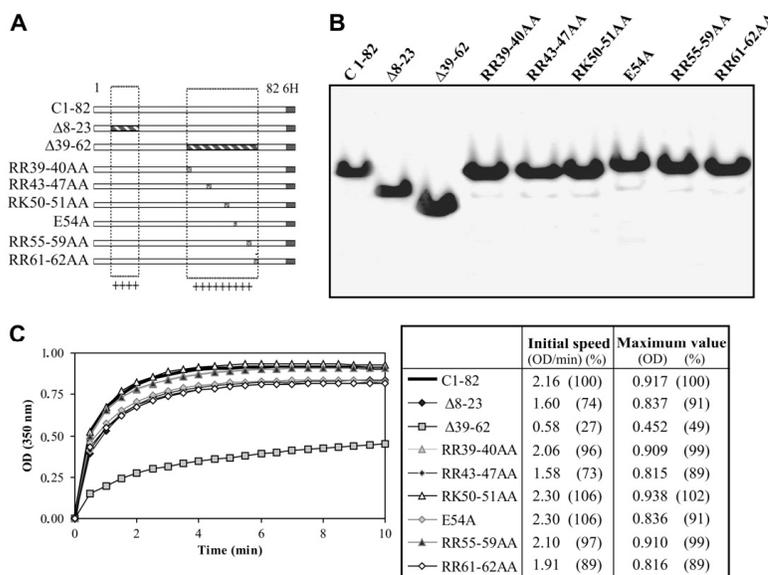


Fig. 4. Influence of positively charged residues on in vitro assembly of C1-82. (A) Schematic representation of C1-82 and mutated derivatives. The boxed regions represent the two positively charged regions (8–23 and 39–62), which are separated by a glycine-rich domain. The deleted regions in $\Delta 8-23$ and $\Delta 39-62$ are shaded, and the position of amino acid substitutions in the 39–62 stretch in the other constructs is indicated. (B) SDS PAGE of purified C1-82 and mutated derivatives. (C) Kinetics of in vitro assembly of the mutants (760 pmol) mixed with IRES RNA (38 pmol) (20:1 ratio).

Inhibition of in vitro assembly of HCV NLPs

The core protein is one of the most conserved proteins of all HCV viral genotypes sequenced to this point. For this reason, it is an excellent target for the development of viral inhibitors. Recently, based on the same rationale, peptides inhibiting the in vitro assembly of the human immunodeficiency virus (HIV-1) have been developed [24]. Although the efficacy of such peptides has not yet been validated in vivo, they were shown to be a good template for the development of assembly inhibitors.

We selected 11 peptides of 18 amino acids in length, corresponding to fragments of the HCV-C protein, and tested each one independently for the ability to inhibit NLP assembly in vitro. The peptide was added to the IRES prior to the addition of the protein so as to give a final molar ratio of C1-82/peptide/IRES of 20:1000:1. The influences of the different peptides on assembly varied (Fig. 5A). In fact, 4 of the 11 peptides tested (15–32, 43–60, 64–81, and 71–88) had no effect on in vitro assembly. Among the others, 5 peptides (1–18, 8–25, 22–39, 29–46, and 36–53) partially inhibited assembly by less than 50%, and 2 peptides (50–67 and 57–74) reduced assembly by factors of 4.5- and 3.3-fold, respectively. Interestingly, the two most efficient inhibitory peptides include amino acids R61–R62 and are rich in basic residues, again highlighting the important role of electrostatic forces in HCV NLP

assembly. The peptides most likely inhibit assembly by interacting with the IRES, thereby interfering with C1–82 binding to RNA. To test this hypothesis, we performed a gel retardation assay with IRES RNA and the peptides used in the assembly assay (Fig. 5B). As expected, peptides 50–67 and 57–74 bound the IRES very efficiently. This result suggests that the positively charged residues of these peptides play an important role in inhibition of assembly. However, the number of basic residues in peptides 50–67 and 57–74 are not the sole determinant for the inhibition given that other basic peptides, such as 1–18, 8–25, and 36–53, failed to inhibit the assembly process despite the fact that they could interact efficiently with the IRES and induce a shift of the RNA in the gel retardation assay (Fig. 5B).

Peptide 22–39 was the most efficient one in affecting the initial speed of assembly. This peptide was derived from the hydrophobic glycine-rich cluster (25–38) located between the two basic clusters. This peptide does not interact with nucleic acids (Fig. 5C) but can efficiently inhibit in vitro assembly by a factor of 2. Peptides 15–32 and 29–46, each comprising only one-half of the hydrophobic domain, did not affect in vitro NLP assembly. A β -sheet structure was predicted in the 29–38 region, which overlaps with the sequence of the inhibitory peptide (22–39) (J-B. Duvignaud et al., unpublished). Also, β -sheet in the HCV-C 1–120 recently was shown to be important for the interaction

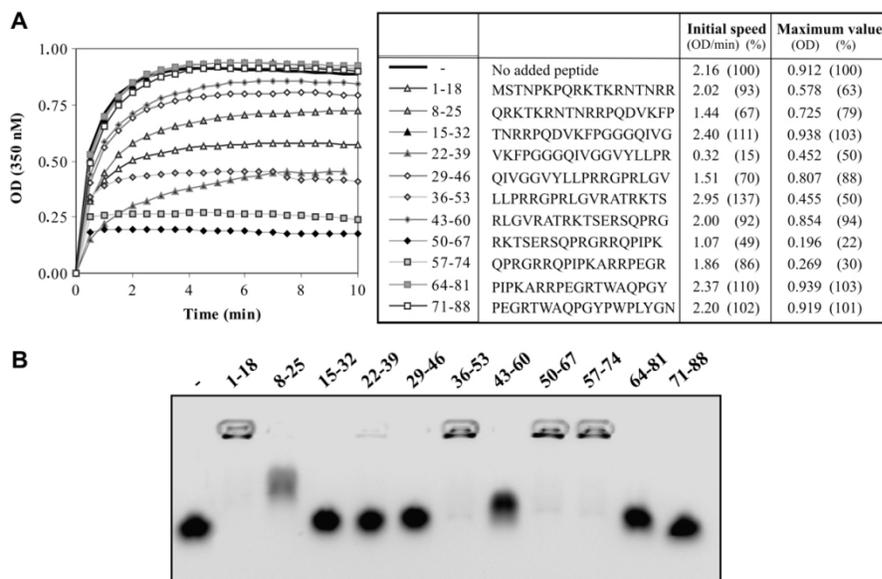


Fig. 5. Peptide-based inhibition of in vitro assembly of C1 82. (A) Kinetics of in vitro assembly with C1 82 (760 pmol) initiated with 38 pmol of IRES alone or together with 3.8 nmol of various peptides. (B) Electrophoresis mobility shift assay (0.8% agarose) of 16 nmol of each peptide mixed with 16 pmol of IRES. Mixing was performed 10 min prior to loading the gel. The gel was stained with ethidium bromide and photographed using a digital camera under UV_{254nm}.

between the subunits [25]. Therefore, the peptide (22–39) probably interferes with the subunit interactions by binding on the predicted β -sheet structure.

Because the hydrophobic region of core appears to be the most promising target for inhibition of assembly, we

next tested chemicals known to interact with hydrophobic clusters. Bis-ANS, a fluorescent probe that binds hydrophobic residues on protein, has been shown to inhibit in vitro HBV capsid assembly [26]. We showed that this compound also inhibits C1–82 assembly in vitro in a

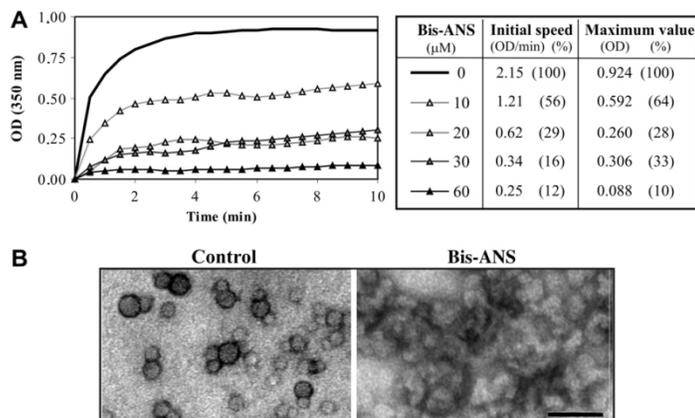


Fig. 6. Inhibition by Bis-ANS of in vitro assembly of C1 82. (A) Kinetics of in vitro assembly of C1 82 (760 pmol) mixed with IRES RNA (38 pmol) in the presence of increasing amounts of Bis-ANS. (B) Electron micrograph of negatively stained NLPs harvested following in vitro assembly. The left panel shows control NLPs, and the right panel shows in vitro assembly in the presence of 60 μ M of Bis-ANS. Bar = 50 nm.

dose-dependent manner (Fig. 6A). As little as 20 μM of Bis-ANS reduced both the initial speed and global in vitro assembly by more than threefold. Bis-ANS clearly interfered with assembly and led to formation of nonspecific aggregates that could be observed under EM (Fig. 6B).

It is not clear whether inhibitors (Bis-ANS or peptides) will efficiently inhibit assembly using in vivo systems. Bis-ANS has been shown to trigger detrimental secondary effects in cells and, therefore, is not suitable for therapy in its current form [26]. However, the inhibitory effects of this compound confirm that targeting hydrophobic clusters of the core is relevant to the development of assembly inhibitors for HCV NLPs. Peptide inhibitors are small peptides that usually have short half-lives and poor cell permeability. However, peptides could be a valuable starting point for drug design of assembly inhibitors.

In conclusion, we have developed a simple, robust, and efficient method to perform extensive studies of HCV nucleocapsid assembly in vitro. We showed that electrostatic interactions with the RNA template govern NLP formation in vitro, and we identified a region of the protein (22–39) that could be important for capsid assembly. This method could be expanded and used to perform large-scale screening of assembly inhibitors such as inhibiting peptides and banks of chemicals.

References

- [1] B.J. Thomson, R.G. Finch, Hepatitis C virus infection, *Clin. Microbiol. Infect.* 11 (2005) 86–94.
- [2] Y. Matsuura, T. Miyamura, The molecular biology of hepatitis C virus, *Semin. Virol.* 4 (1993) 297–304.
- [3] G. Saracco, A. Olivero, A. Ciancio, S. Carezzi, M. Rizzetto, Therapy of chronic hepatitis C: A critical review, *Curr. Drug Targets Infect. Disord.* 3 (2003) 25–32.
- [4] J. McLauchlan, M.K. Lemberg, G. Hope, B. Martoglio, Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets, *EMBO J.* 21 (2002) 3980–3988.
- [5] T. Ogino, H. Fukuda, S. Imajoh-Ohmi, M. Kohara, A. Nomoto, Membrane binding properties and terminal residues of the mature hepatitis C virus capsid protein in insect cells, *J. Virol.* 78 (2004) 11766–11777.
- [6] R. Suzuki, K. Tamura, J. Li, K. Ishii, Y. Matsuura, T. Miyamura, T. Suzuki, Ubiquitin-mediated degradation of hepatitis C virus core protein is regulated by processing at its carboxyl terminus, *Virology* 280 (2001) 301–309.
- [7] C. Shih, C. Chen, S. Chen, Y. Lee, Modulation of the trans-suppression activity of hepatitis C virus core protein by phosphorylation, *J. Virol.* 69 (1995) 1160–1171.
- [8] N. Majeau, M. Bolduc, J.-B. Duvignaud, R. Fromentin, D. Leclerc, Effect of cAMP-dependent protein kinase A (PKA) on HCV nucleocapsid assembly and degradation, *Biochem. Cell Biol.* 85 (2007) 78–87.
- [9] P. van der Schoot, R. Bruinsma, Electrostatics and the assembly of an RNA virus, *Phys. Rev. E* 71 (2005) 061928.
- [10] M. Kunkel, M. Lorinczi, R. Rijnbrand, S.M. Lemon, S.J. Watowich, Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein, *J. Virol.* 75 (2001) 2119–2129.
- [11] T. Shimoike, S. Mimori, H. Tani, Y. Matsuura, T. Miyamura, Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation, *J. Virol.* 73 (1999) 9718–9725.
- [12] Y. Tanaka, T. Shimoike, K. Ishii, R. Suzuki, T. Suzuki, H. Ushijima, Y. Matsuura, T. Miyamura, Selective binding of hepatitis C virus core protein to synthetic oligonucleotides corresponding to the 5' untranslated region of the viral genome, *Virology* 270 (2000) 229–236.
- [13] J. Bukh, R.H. Purcell, R.H. Miller, Sequence analysis of the 5' noncoding region of hepatitis C virus, *Proc. Natl. Acad. Sci. USA* 89 (1992) 4942–4946.
- [14] M. Honda, M.R. Beard, L.-H. Ping, S.M. Lemon, A phylogenetically conserved stem loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation, *J. Virol.* 73 (1999) 1165–1174.
- [15] L.P. Beales, D.J. Rowlands, A. Holzenburg, The internal ribosome entry site (IRES) of hepatitis C virus visualized by electron microscopy, *RNA* 7 (2001) 661–670.
- [16] M. Piron, N. Beguiristain, A. Nadal, E. Martinez-Salas, J. Gomez, Characterizing the function and structural organization of the 5' tRNA-like motif within the hepatitis C virus quasispecies, *Nucleic Acids Res.* 33 (2005) 1487–1502.
- [17] K. Tsukiyama-Kohara, N. Izuka, M. Kohara, A. Nomoto, Internal ribosome entry site within hepatitis C virus RNA, *J. Virol.* 66 (1992) 1476–1483.
- [18] N. Majeau, V. Gagne, A. Boivin, M. Bolduc, J.-A. Majeau, D. Ouellet, D. Leclerc, The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation, *J. Gen. Virol.* 85 (2004) 971–981.
- [19] T. Ikemura, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system, *J. Mol. Biol.* 151 (1981) 389–409.
- [20] R. Rijnbrand, P.J. Bredenbeek, P.C. Haasnoot, J.S. Kieft, W.J. Spaan, S.M. Lemon, The influence of downstream protein-coding sequence on internal ribosome entry on hepatitis C virus and other flavivirus RNAs, *RNA* 7 (2001) 585–597.
- [21] T.L. Tellinghuisen, A.E. Hamburger, B.R. Fisher, R. Ostendorp, R.J. Kuhn, In vitro assembly of alphavirus cores by using nucleocapsid protein expressed in *Escherichia coli*, *J. Virol.* 73 (1999) 5309–5319.
- [22] J.B. Bancroft, E. Hiebert, C.E. Bracker, The effects of various polyanions on shell formation of some spherical viruses, *Virology* 39 (1969) 924–930.
- [23] K. Klein, S. Dellos, J. Lingappa, Identification of residues in the hepatitis C virus core protein that are critical for capsid assembly in a cell-free system, *J. Virol.* 79 (2005) 6814–6826.
- [24] J. Sticht, M. Humbert, S. Findlow, J. Bodem, B. Muller, U. Dietrich, J. Werner, H.-G. Krausslich, A peptide inhibitor of HIV-1 assembly in vitro, *Nat. Struct. Mol. Biol.* 12 (2005) 671–677.
- [25] M.M.P.C.A. Rodriguez-Casado, Conformational features of truncated hepatitis C virus core protein in virus-like particles, *Biopolymers* 82 (2006) 334–338.
- [26] A. Zlotnick, P. Ceres, S. Singh, J.M. Johnson, A small molecule inhibits and misdirects assembly of hepatitis B virus capsids, *J. Virol.* 76 (2002) 4848–4854.

Bibliographie

1. Acosta-Rivero, N., Aguilar, J.C., Musacchio, A., Falcon, V., Vina, A., de la Rosa, M.C., and Morales, J., *Characterization of the HCV core virus-like particles produced in the methylotrophic yeast Pichia pastoris*. Biochem. Biophys. Res. Commun., 2001. **287**(1): p. 122-5.
2. Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K., and Miyano, M., *Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus*. Structure, 1999. **7**(11): p. 1417-26.
3. Ai, L.S., Lee, Y.W., and Chen, S.S., *Characterization of hepatitis C virus core protein multimerization and membrane envelopment: revelation of a cascade of core-membrane interactions*. J. Virol., 2009. **83**(19): p. 9923-39.
4. Albecka, A., Montserret, R., Krey, T., Tarr, A.W., Diesis, E., Ball, J.K., Descamps, V., Duverlie, G., Rey, F., Penin, F., and Dubuisson, J., *Identification of new functional regions in hepatitis C virus envelope glycoprotein E2*. J. Virol., 2011. **85**(4): p. 1777-92.
5. Alsaleh, K., Delavalle, P.Y., Pillez, A., Duverlie, G., Descamps, V., Rouille, Y., Dubuisson, J., and Wychowski, C., *Identification of basic amino acids at the N-terminal end of the core protein that are crucial for hepatitis C virus infectivity*. J. Virol., 2010. **84**(24): p. 12515-28.
6. Alter, H.J., *Discovery of Non-A, Non-B Hepatitis and Identification of Its Etiology*. Am. J. Med, 1999. **107**(6B): p. 16S-20S.
7. Alter, H.J., Holland, P.V., Morrow, A.G., Purcell, R.H., Feinstone, S.M., and Moritsugu, Y., *Clinical and serological analysis of transfusion-associated hepatitis*. Lancet, 1975. **2**(7940): p. 838-41.
8. Alter, H.J., Holland, P.V., and Schmidt, P.J., *Hepatitis-associated antigen. To test or not to test?* Lancet, 1970. **2**(7664): p. 142-3.
9. Alter, H.J., Purcell, R.H., Holland, P.V., and Popper, H., *Transmissible agent in non-A, non-B hepatitis*. Lancet, 1978. **1**(8062): p. 459-63.
10. Andrade, M.A., Chacon, P., Merelo, J.J., and Moran, F., *Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network*. Protein Eng, 1993. **6**(4): p. 383-90.

11. Aoki, H., Hayashi, J., Moriyama, M., Arakawa, Y., and Hino, O., *Hepatitis C virus core protein interacts with 14-3-3 protein and activates the kinase Raf-1*. J. Virol., 2000. **74**(4): p. 1736-41.
12. Appel, N., Zayas, M., Miller, S., Krijnse-Locker, J., Schaller, T., Friebe, P., Kallis, S., Engel, U., and Bartenschlager, R., *Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly*. PLoS Pathog., 2008. **4**(3): p. e1000035.
13. Asabe, S.I., Tanji, Y., Satoh, S., Kaneko, T., Kimura, K., and Shimotohno, K., *The N-terminal region of hepatitis C virus-encoded NS5A is important for NS4A-dependent phosphorylation*. J. Virol., 1997. **71**(1): p. 790-6.
14. Barba, G., Harper, F., Harada, T., Kohara, M., Goulinet, S., Matsuura, Y., Eder, G., Schaff, Z., Chapman, M.J., Miyamura, T., and Brechot, C., *Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets*. Proc. Natl. Acad. Sci. U S A, 1997. **94**(4): p. 1200-5.
15. Baril, M. and Brakier-Gingras, L., *Translation of the F protein of hepatitis C virus is initiated at a non-AUG codon in a +1 reading frame relative to the polyprotein*. Nucleic Acids Res., 2005. **33**(5): p. 1474-86.
16. Bartenschlager, R., Ahlborn-Laake, L., Mous, J., and Jacobsen, H., *Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions*. J. Virol., 1993. **67**(7): p. 3835-44.
17. Bartenschlager, R., Penin, F., Lohmann, V., and Andre, P., *Assembly of infectious hepatitis C virus particles*. Trends Microbiol., 2011. **19**(2): p. 95-103.
18. Barth, H., Liang, T.J., and Baumert, T.F., *Hepatitis C virus entry: molecular biology and clinical implications*. Hepatology, 2006. **44**(3): p. 527-35.
19. Baumert, T.F., Ito, S., Wong, D.T., and Liang, T.J., *Hepatitis C virus structural proteins assemble into viruslike particles in insect cells*. J. Virol., 1998. **72**(5): p. 3827-36.
20. Behrens, S.E., Tomei, L., and De Francesco, R., *Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus*. EMBO J., 1996. **15**(1): p. 12-22.
21. Berova, N., Nakanishi, K., and Woody, R.W., *Circular Dichroism: Principles and Applications, 2nd Edition*. 2nd ed. 2000: Wiley-VCH. 912.
22. Birerdinc, A. and Younossi, Z.M., *Emerging therapies for hepatitis C virus*. Expert Opin. Emerg. Drugs, 2010. **15**(4): p. 1-10.

23. Blanchard, E., Brand, D., Trassard, S., Goudeau, A., and Roingeard, P., *Hepatitis C virus-like particle morphogenesis*. J. Virol., 2002. **76**(8): p. 4073-9.
24. Blumberg, B.S., Alter, H.J., and Visnich, S., *A "New" Antigen in Leukemia Sera*. JAMA, 1965. **191**: p. 541-6.
25. Bohm, G., Muhr, R., and Jaenicke, R., *Quantitative analysis of protein far UV circular dichroism spectra by neural networks*. Protein Eng, 1992. **5**(3): p. 191-5.
26. Boni, S., Lavergne, J.P., Boulant, S., and Cahour, A., *Hepatitis C virus core protein acts as a trans-modulating factor on internal translation initiation of the viral RNA*. J. Biol. Chem., 2005. **280**(18): p. 17737-48.
27. Bostan, N. and Mahmood, T., *An Overview about hepatitis C: A devastating virus*. Critical Reviews in Microbiology, 2010. **36**(2): p. 91-133.
28. Boulant, S., Becchi, M., Penin, F., and Lavergne, J.P., *Unusual multiple recoding events leading to alternative forms of hepatitis C virus core protein from genotype 1b*. J. Biol. Chem., 2003. **278**(46): p. 45785-92.
29. Boulant, S., Montserret, R., Hope, R.G., Ratinier, M., Targett-Adams, P., Lavergne, J.P., Penin, F., and McLauchlan, J., *Structural determinants that target the hepatitis C virus core protein to lipid droplets*. J. Biol. Chem., 2006. **281**(31): p. 22236-47.
30. Boulant, S., Vanbelle, C., Ebel, C., Penin, F., and Lavergne, J.P., *Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features*. J. Virol., 2005. **79**(17): p. 11353-65.
31. Boumlic, A., Nomine, Y., Charbonnier, S., Dalagiorgou, G., Vassilaki, N., Kieffer, B., Trave, G., Mavromara, P., and Orfanoudakis, G., *Prevalence of intrinsic disorder in the hepatitis C virus ARFP/Core+1/S protein*. FEBS J., 2010. **277**(3): p. 774-89.
32. Boumlic, A., Vassilaki, N., Dalagiorgou, G., Kochlios, E., Kakkanas, A., Georgopoulou, U., Markoulatos, P., Orfanoudakis, G., and Mavromara, P., *Internal translation initiation stimulates expression of the ARF/core+1 open reading frame of HCV genotype 1b*. Virus Res., 2011. **155**(1): p. 213-20.
33. Bradley, D.W. and Maynard, J.E., *Etiology and natural history of post-transfusion and enterically-transmitted non-A, non-B hepatitis*. Semin. Liver Dis., 1986. **6**(1): p. 56-66.
34. Bradley, D.W., McCaustland, K.A., Cook, E.H., Schable, C.A., Ebert, J.W., and Maynard, J.E., *Posttransfusion non-A, non-B hepatitis in chimpanzees. Physicochemical evidence that the tubule-forming agent is a small, enveloped virus*. Gastroenterology, 1985. **88**(3): p. 773-9.

35. Branch, A.D., Stump, D.D., Gutierrez, J.A., Eng, F., and Walewski, J.L., *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., 2005. **25**(1): p. 105-17.
36. Brass, V., Bieck, E., Montserret, R., Wolk, B., Hellings, J.A., Blum, H.E., Penin, F., and Moradpour, D., *An amino-terminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A*. J. Biol. Chem., 2002. **277**(10): p. 8130-9.
37. Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R.L., Mathieu, M., De Francesco, R., and Rey, F.A., *Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus*. Proc. Natl. Acad. Sci. U S A, 1999. **96**(23): p. 13034-9.
38. Brunger, A.T., *Version 1.2 of the Crystallography and NMR system*. Nat Protoc, 2007. **2**(11): p. 2728-33.
39. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L., *Crystallography & NMR system: A new software suite for macromolecular structure determination*. Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.
40. Burlone, M.E. and Budkowska, A., *Hepatitis C virus cell entry: role of lipoproteins and cellular receptors*. J. Gen. Virol., 2009. **90**(Pt 5): p. 1055-70.
41. Carmona, P. and Molina, M., *Interactions of protein and nucleic acid components of hepatitis C virus as revealed by Fourier transform infrared spectroscopy*. Biochemistry, 2010. **49**(23): p. 4724-31.
42. Carrere-Kremer, S., Montpellier-Pala, C., Cocquerel, L., Wychowski, C., Penin, F., and Dubuisson, J., *Subcellular localization and topology of the p7 polypeptide of hepatitis C virus*. J. Virol., 2002. **76**(8): p. 3720-30.
43. Chang, S.C., Yen, J.H., Kang, H.Y., Jang, M.H., and Chang, M.F., *Nuclear localization signals in the core protein of hepatitis C virus*. Biochem. Biophys. Res. Commun., 1994. **205**(2): p. 1284-90.
44. Chen, J., *Intrinsically disordered p53 extreme C-terminus binds to S100B(beta-beta) through "fly-casting"*. J Am Chem Soc, 2009. **131**(6): p. 2088-9.
45. Chen, L., Chen, P., Fan, G., Li, L., and Liu, C., *Localization of hepatitis C virus core protein in the nucleus of peripheral blood mononuclear cells of hepatitis C patients*. Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi, 2002. **16**(1): p. 37-9.

46. Cheng, Y., LeGall, T., Oldfield, C.J., Mueller, J.P., Van, Y.Y., Romero, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K., *Rational drug design via intrinsically disordered protein*. Trends Biotechnol., 2006. **24**(10): p. 435-42.
47. Chitayat, S. and Arrowsmith, C.H., *Four p(53)s in a pod*. Nat Struct Mol Biol, 2010. **17**(4): p. 390-1.
48. Cho, H.S., Ha, N.C., Kang, L.W., Chung, K.M., Back, S.H., Jang, S.K., and Oh, B.H., *Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA*. J. Biol. Chem., 1998. **273**(24): p. 15045-52.
49. Choi, J., Xu, Z., and Ou, J.H., *Triple decoding of hepatitis C virus RNA by programmed translational frameshifting*. Mol. Cell Biol., 2003. **23**(5): p. 1489-97.
50. Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., and Houghton, M., *Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome*. Science, 1989. **244**(4902): p. 359-62.
51. Cocquerel, L., Voisset, C., and Dubuisson, J., *Hepatitis C virus entry: potential receptors and their biological functions*. J. Gen. Virol., 2006. **87**(Pt 5): p. 1075-84.
52. Cook, G.A. and Opella, S.J., *NMR studies of p7 protein from hepatitis C virus*. Eur. Biophys. J., 2010. **39**(7): p. 1097-104.
53. Cook, G.A. and Opella, S.J., *Secondary structure, dynamics, and architecture of the p7 membrane protein from hepatitis C virus by NMR spectroscopy*. Biochim. Biophys. Acta, 2010.
54. Cornilescu, G., Delaglio, F., and Bax, A., *Protein backbone angle restraints from searching a database for chemical shift and sequence homology*. J Biomol NMR, 1999. **13**(3): p. 289-302.
55. Cristofari, G., Ivanyi-Nagy, R., Gabus, C., Boulant, S., Lavergne, J.P., Penin, F., and Darlix, J.L., *The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro*. Nucleic Acids Res., 2004. **32**(8): p. 2623-31.
56. Dalagiorgou, G., Vassilaki, N., Foka, P., Boumlic, A., Kakkanas, A., Kochlios, E., Khalili, S., Aslanoglou, E., Veletza, S., Orfanoudakis, G., Vassilopoulos, D., Hadziyannis, S., Koskinas, J., Mavromara, P., *High Levels of HCV core+1 Antibodies in HCV Patients with Hepatocellular Carcinoma* J. Gen. Virol., 2011.
57. Davis, G.L., Alter, M.J., El-Serag, H., Poynard, T., and Jennings, L.W., *Aging of hepatitis C virus (HCV)-infected persons in the United States: a multiple cohort model of HCV prevalence and disease progression*. Gastroenterology, 2010. **138**(2): p. 513-21, 521 e1-6.

58. Davison, T.S., Nie, X., Ma, W., Lin, Y., Kay, C., Benchimol, S., and Arrowsmith, C.H., *Structure and functionality of a designed p53 dimer*. J. Mol. Biol., 2001. **307**(2): p. 605-17.
59. Davison, T.S., Nie, X., Ma, W., Lin, Y., Kay, C., Benchimol, S., and Arrowsmith, C.H., *Structure and functionality of a designed p53 dimer*. J Mol Biol, 2001. **307**(2): p. 605-17.
60. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes*. J Biomol NMR, 1995. **6**(3): p. 277-93.
61. Dentzer, T.G., Lorenz, I.C., Evans, M.J., and Rice, C.M., *Determinants of the hepatitis C virus nonstructural protein 2 protease domain required for production of infectious virus*. J. Virol., 2009. **83**(24): p. 12702-13.
62. Dodd, R.Y., *Current risk for transfusion transmitted infections*. Curr. Opin. Hematol., 2007. **14**(6): p. 671-6.
63. Dubuisson, J., Hsu, H.H., Cheung, R.C., Greenberg, H.B., Russell, D.G., and Rice, C.M., *Formation and intracellular localization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia and Sindbis viruses*. J. Virol., 1994. **68**(10): p. 6147-60.
64. Dufourcq, J. and Faucon, J.F., *Intrinsic fluorescence study of lipid-protein interactions in membrane models. Binding of melittin, an amphipathic peptide, to phospholipid vesicles*. Biochim. Biophys. Acta, 1977. **467**(1): p. 1-11.
65. Dusheiko, G. and Simmonds, P., *Sequence variability of hepatitis C virus and its clinical relevance*. J. Viral Hepat., 1994. **1**(1): p. 3-15.
66. Duvignaud, J.B., Leclerc, D., and Gagne, S.M., *Structure and dynamics changes induced by 2,2,2-trifluoro-ethanol (TFE) on the N-terminal half of hepatitis C virus core protein*. Biochem. Cell Biol., 2010. **88**(2): p. 315-23.
67. Duvignaud, J.B., Savard, C., Fromentin, R., Majeau, N., Leclerc, D., and Gagne, S.M., *Structure and dynamics of the N-terminal half of hepatitis C virus core protein: an intrinsically unstructured protein*. Biochem. Biophys. Res. Commun., 2009. **378**(1): p. 27-31.
68. Eckart, M.R., Selby, M., Masiarz, F., Lee, C., Berger, K., Crawford, K., Kuo, C., Kuo, G., Houghton, M., and Choo, Q.L., *The hepatitis C virus encodes a serine protease involved in processing of the putative nonstructural proteins from the viral polyprotein precursor*. Biochem. Biophys. Res. Commun., 1993. **192**(2): p. 399-406.

69. Egger, D., Wolk, B., Gosert, R., Bianchi, L., Blum, H.E., Moradpour, D., and Bienz, K., *Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex*. J. Virol., 2002. **76**(12): p. 5974-84.
70. Einav, S., Elazar, M., Danieli, T., and Glenn, J.S., *A nucleotide binding motif in hepatitis C virus (HCV) NS4B mediates HCV RNA replication*. J. Virol., 2004. **78**(20): p. 11288-95.
71. Ezelle, H.J., Markovic, D., and Barber, G.N., *Generation of hepatitis C virus-like particles by use of a recombinant vesicular stomatitis virus vector*. J. Virol., 2002. **76**(23): p. 12325-34.
72. Failla, C., Tomei, L., and De Francesco, R., *Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins*. J. Virol., 1994. **68**(6): p. 3753-60.
73. Falcon, V., Acosta-Rivero, N., Chinae, G., de la Rosa, M.C., Menendez, I., Duenas-Carrera, S., Gra, B., Rodriguez, A., Tsutsumi, V., Shibayama, M., Luna-Munoz, J., Miranda-Sanchez, M.M., Morales-Grillo, J., and Kouri, J., *Nuclear localization of nucleocapsid-like particles and HCV core protein in hepatocytes of a chronically HCV-infected patient*. Biochem. Biophys. Res. Commun., 2003. **310**(1): p. 54-8.
74. Falcon, V., Garcia, C., de la Rosa, M.C., Menendez, I., Seoane, J., and Grillo, J.M., *Ultrastructural and immunocytochemical evidences of core-particle formation in the methylotrophic Pichia pastoris yeast when expressing HCV structural proteins (core-E1)*. Tissue Cell, 1999. **31**(2): p. 117-25.
75. Fan, Z., Yang, Q.R., Twu, J.S., and Sherker, A.H., *Specific in vitro association between the hepatitis C viral genome and core protein*. J. Med. Virol., 1999. **59**(2): p. 131-4.
76. Feinstone, S.M., Kapikian, A.Z., and Purceli, R.H., *Hepatitis A: detection by immune electron microscopy of a viruslike antigen associated with acute illness*. Science, 1973. **182**(116): p. 1026-8.
77. Ferron, F., Bussetta, C., Dutartre, H., and Canard, B., *The modeled structure of the RNA dependent RNA polymerase of GBV-C virus suggests a role for motif E in Flaviviridae RNA polymerases*. BMC Bioinformatics, 2005. **6**: p. 255.
78. Fielding, L., *NMR methods for the determination of protein-ligand dissociation constants*. Progress in nuclear magnetic resonance spectroscopy, 2007. **51**: p. 219-242.
79. Fiorucci, M., Boulant, S., Fournillier, A., Abraham, J.D., Lavergne, J.P., Paranhos-Baccala, G., Inchauspe, G., and Bain, C., *Expression of the alternative reading*

- frame protein of Hepatitis C virus induces cytokines involved in hepatic injuries. J. Gen. Virol.*, 2007. **88**(Pt 4): p. 1149-62.
80. Foster, T.L., Belyaeva, T., Stonehouse, N.J., Pearson, A.R., and Harris, M., *All three domains of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. J. Virol.*, 2010. **84**(18): p. 9267-77.
 81. Frank, C., Mohamed, M.K., Strickland, G.T., Lavanchy, D., Arthur, R.R., Magder, L.S., El Khoby, T., Abdel-Wahab, Y., Aly Ohn, E.S., Anwar, W., and Sallam, I., *The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. Lancet*, 2000. **355**(9207): p. 887-91.
 82. Fromentin, R., Majeau, N., Laliberte Gagne, M.E., Boivin, A., Duvignaud, J.B., and Leclerc, D., *A method for in vitro assembly of hepatitis C virus core protein and for screening of inhibitors. Anal. Biochem.*, 2007. **366**(1): p. 37-45.
 83. Fukasawa, M., *Cellular lipid droplets and hepatitis C virus life cycle. Biol. Pharm. Bull.*, 2010. **33**(3): p. 355-9.
 84. Gastaminza, P., Dryden, K.A., Boyd, B., Wood, M.R., Law, M., Yeager, M., and Chisari, F.V., *Ultrastructural and biophysical characterization of hepatitis C virus particles produced in cell culture. J. Virol.*, 2010. **84**(21): p. 10999-1009.
 85. Gosert, R., Egger, D., Lohmann, V., Bartenschlager, R., Blum, H.E., Bienz, K., and Moradpour, D., *Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. J. Virol.*, 2003. **77**(9): p. 5487-92.
 86. Gouttenoire, J., Castet, V., Montserret, R., Arora, N., Raussens, V., Ruyschaert, J.M., Diesis, E., Blum, H.E., Penin, F., and Moradpour, D., *Identification of a novel determinant for membrane association in hepatitis C virus nonstructural protein 4B. J. Virol.*, 2009. **83**(12): p. 6257-68.
 87. Gouttenoire, J., Montserret, R., Kennel, A., Penin, F., and Moradpour, D., *An amphipathic alpha-helix at the C terminus of hepatitis C virus nonstructural protein 4B mediates membrane association. J. Virol.*, 2009. **83**(21): p. 11378-84.
 88. Gouttenoire, J., Penin, F., and Moradpour, D., *Hepatitis C virus nonstructural protein 4B: a journey into unexplored territory. Rev. Med. Virol.*, 2010. **20**(2): p. 117-29.
 89. Grakoui, A., McCourt, D.W., Wychowski, C., Feinstone, S.M., and Rice, C.M., *Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. J. Virol.*, 1993. **67**(5): p. 2832-43.
 90. Greenfield, N.J., *Circular dichroism analysis for protein-protein interactions. Methods Mol Biol*, 2004. **261**: p. 55-78.

91. Greenfield, N.J., *Analysis of the kinetics of folding of proteins and peptides using circular dichroism*. Nat Protoc, 2006. **1**(6): p. 2891-9.
92. Greenfield, N.J., *Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism*. Nat Protoc, 2006. **1**(6): p. 2733-41.
93. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure*. Nat Protoc, 2006. **1**(6): p. 2876-90.
94. Gwack, Y., Kim, D.W., Han, J.H., and Choe, J., *Characterization of RNA binding activity and RNA helicase activity of the hepatitis C virus NS3 protein*. Biochem. Biophys. Res. Commun., 1996. **225**(2): p. 654-9.
95. Hanouille, X., Badillo, A., Verdegem, D., Penin, F., and Lippens, G., *The domain 2 of the HCV NS5A protein is intrinsically unstructured*. Protein Pept. Lett., 2010. **17**(8): p. 1012-8.
96. Hanouille, X., Verdegem, D., Badillo, A., Wieruszeski, J.M., Penin, F., and Lippens, G., *Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded*. Biochem. Biophys. Res. Commun., 2009. **381**(4): p. 634-8.
97. He, L.F., Alling, D., Popkin, T., Shapiro, M., Alter, H.J., and Purcell, R.H., *Determining the size of non-A, non-B hepatitis virus by filtration*. J Infect Dis, 1987. **156**(4): p. 636-40.
98. He, Y., Staschke, K.A., and Tan, S.L., *HCV NS5A: A Multifunctional Regulator of Cellular Pathways and Virus Replication*, in *Hepatitis C Viruses: Genomes and Molecular Biology*, S.L. Tan, Editor. 2006: Norfolk (UK).
99. Hennessey, J.P., Jr. and Johnson, W.C., Jr., *Information content in the circular dichroism of proteins*. Biochemistry, 1981. **20**(5): p. 1085-94.
100. Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., and Shimotohno, K., *Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis*. Proc. Natl. Acad. Sci. U S A, 1991. **88**(13): p. 5547-51.
101. Hitomi, Y., McDonnell, W.M., Killeen, A.A., and Askari, F.K., *Sequence analysis of the hepatitis C virus (HCV) core gene suggests the core protein as an appropriate target for HCV vaccine strategies*. J. Viral. Hepat., 1995. **2**(5): p. 235-41.
102. Hollinger, F.B., Gitnick, G.L., Aach, R.D., Szmunes, W., Mosley, J.W., Stevens, C.E., Peters, R.L., Weiner, J.M., Werch, J.B., and Lander, J.J., *Non-A, non-B hepatitis transmission in chimpanzees: a project of the transfusion-transmitted viruses study group*. Intervirology, 1978. **10**(1): p. 60-8.

103. Hope, R.G. and McLauchlan, J., *Sequence motifs required for lipid droplet association and protein stability are unique to the hepatitis C virus core protein*. J. Gen. Virol., 2000. **81**(Pt 8): p. 1913-25.
104. Hope, R.G., Murphy, D.J., and McLauchlan, J., *The domains required to direct core proteins of hepatitis C virus and GB virus-B to lipid droplets share common features with plant oleosin proteins*. J. Biol. Chem., 2002. **277**(6): p. 4261-70.
105. Hore, P.J., *Nuclear Magnetic Resonance*. fifth ed. Oxford Chemistry Primers, ed. S.G. Davies, et al. Vol. 32. 1995, Oxford: Oxford science publications. 90.
106. Huang, L., Hwang, J., Sharma, S.D., Hargittai, M.R., Chen, Y., Arnold, J.J., Raney, K.D., and Cameron, C.E., *Hepatitis C virus nonstructural protein 5A (NS5A) is an RNA-binding protein*. J. Biol. Chem., 2005. **280**(43): p. 36417-28.
107. Hussy, P., Langen, H., Mous, J., and Jacobsen, H., *Hepatitis C virus core protein: carboxy-terminal boundaries of two processed species suggest cleavage by a signal peptide peptidase*. Virology, 1996. **224**(1): p. 93-104.
108. Hwang, J., Huang, L., Cordek, D.G., Vaughan, R., Reynolds, S.L., Kihara, G., Raney, K.D., Kao, C.C., and Cameron, C.E., *Hepatitis C virus nonstructural protein 5A: biochemical characterization of a novel structural class of RNA-binding proteins*. J. Virol., 2010. **84**(24): p. 12480-91.
109. Irshad, M. and Dhar, I., *Hepatitis C virus core protein: an update on its molecular biology, cellular functions and clinical implications*. Med. Princ. Pract., 2006. **15**(6): p. 405-16.
110. Ishida, S., Kaito, M., Kohara, M., Tsukiyama-Kohora, K., Fujita, N., Ikoma, J., Adachi, Y., and Watanabe, S., *Hepatitis C virus core particle detected by immunoelectron microscopy and optical rotation technique*. Hepatol. Res., 2001. **20**(3): p. 335-347.
111. Isoyama, T., Kuge, S., and Nomoto, A., *The core protein of hepatitis C virus is imported into the nucleus by transport receptor Kap123p but inhibits Kap121p-dependent nuclear import of yeast API-like transcription factor in yeast cells*. J. Biol. Chem., 2002. **277**(42): p. 39634-41.
112. Ivanyi-Nagy, R. and Darlix, J.L., *Intrinsic disorder in the core proteins of flaviviruses*. Protein Pept. Lett., 2010. **17**(8): p. 1019-25.
113. Ivanyi-Nagy, R., Kanevsky, I., Gabus, C., Lavergne, J.P., Ficheux, D., Penin, F., Fosse, P., and Darlix, J.L., *Analysis of hepatitis C virus RNA dimerization and core-RNA interactions*. Nucleic Acids Res., 2006. **34**(9): p. 2618-33.

114. Ivanyi-Nagy, R., Lavergne, J.P., Gabus, C., Ficheux, D., and Darlix, J.L., *RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae*. *Nucleic Acids Res.*, 2008. **36**(3): p. 712-25.
115. Jee, J. and Guntert, P., *Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment*. *J Struct Funct Genomics*, 2003. **4**(2-3): p. 179-89.
116. Jirasko, V., Montserret, R., Lee, J.Y., Gouttenoire, J., Moradpour, D., Penin, F., and Bartenschlager, R., *Structural and functional studies of nonstructural protein 2 of the hepatitis C virus reveal its key role as organizer of virion assembly*. *PLoS Pathog.*, 2010. **6**(12): p. e1001233.
117. Johnson, B.A., *NMR View - computer program for the visualization and analysis of NMR data*. *J. Biomol. NMR*, 1994. **4**: p. 603-614.
118. Jones, C.T., Ma, L., Burgner, J.W., Groesch, T.D., Post, C.B., and Kuhn, R.J., *Flavivirus capsid is a dimeric alpha-helical protein*. *J. Virol.*, 2003. **77**(12): p. 7143-9.
119. Jones, D.M. and McLauchlan, J., *Hepatitis C virus: assembly and release of virus particles*. *J. Biol. Chem.*, 2010. **285**(30): p. 22733-9.
120. Julien, O., Gignac, I., Hutton, A., Yee, A., Arrowsmith, C.H., and Gagne, S.M., *MTH187 from Methanobacterium thermoautotrophicum has three HEAT-like repeats*. *J Biomol NMR*, 2006. **35**(2): p. 149-54.
121. Jung, E.Y., Lee, M.N., Yang, H.Y., Yu, D., and Jang, K.L., *The repressive activity of hepatitis C virus core protein on the transcription of p21(waf1) is regulated by protein kinase A-mediated phosphorylation*. *Virus Res.*, 2001. **79**(1-2): p. 109-15.
122. Kaito, M., Watanabe, S., Tsukiyama-Kohara, K., Yamaguchi, K., Kobayashi, Y., Konishi, M., Yokoi, M., Ishida, S., Suzuki, S., and Kohara, M., *Hepatitis C virus particle detected by immunoelectron microscopic study*. *J. Gen. Virol.*, 1994. **75** (Pt 7): p. 1755-60.
123. Kanai, A., Tanabe, K., and Kohara, M., *Poly(U) binding activity of hepatitis C virus NS3 protein, a putative RNA helicase*. *FEBS Lett.*, 1995. **376**(3): p. 221-4.
124. Kao, C.F., Chen, S.Y., Chen, J.Y., and Wu Lee, Y.H., *Modulation of p53 transcription regulatory activity and post-translational modification by hepatitis C virus core protein*. *Oncogene*, 2004. **23**(14): p. 2472-83.
125. Kasprzak, A. and Adamek, A., *Role of hepatitis C virus proteins (C, NS3, NS5A) in hepatic oncogenesis*. *Hepato Res*, 2008. **38**(1): p. 1-26.

126. Kasprzak, A., Seidel, J., Biczysko, W., Wysocki, J., Spachacz, R., and Zabel, M., *Intracellular localization of NS3 and C proteins in chronic hepatitis C*. *Liver Int.*, 2005. **25**(4): p. 896-903.
127. Kelly, S.M. and Price, N.C., *Circular dichroism to study protein interactions*. *Curr Protoc Protein Sci*, 2006. **Chapter 20**: p. Unit 20 10.
128. Khu, Y.L., Koh, E., Lim, S.P., Tan, Y.H., Brenner, S., Lim, S.G., Hong, W.J., and Goh, P.Y., *Mutations that affect dimer formation and helicase activity of the hepatitis C virus helicase*. *J. Virol.*, 2001. **75**(1): p. 205-14.
129. Kim, D.W., Gwack, Y., Han, J.H., and Choe, J., *C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity*. *Biochem. Biophys. Res. Commun.*, 1995. **215**(1): p. 160-6.
130. Kim, D.W., Suzuki, R., Harada, T., Saito, I., and Miyamura, T., *Trans-suppression of gene expression by hepatitis C viral core protein*. *Jpn J. Med. Sci. Biol.*, 1994. **47**(4): p. 211-20.
131. Kim, M., Ha, Y., and Park, H.J., *Structural requirements for assembly and homotypic interactions of the hepatitis C virus core protein*. *Virus Res.*, 2006. **122**(1-2): p. 137-43.
132. Klein, K.C., Dellos, S.R., and Lingappa, J.R., *Identification of residues in the hepatitis C virus core protein that are critical for capsid assembly in a cell-free system*. *J. Virol.*, 2005. **79**(11): p. 6814-26.
133. Klein, K.C., Polyak, S.J., and Lingappa, J.R., *Unique features of hepatitis C virus capsid formation revealed by de novo cell-free assembly*. *J. Virol.*, 2004. **78**(17): p. 9257-69.
134. Koch, J.O., Lohmann, V., Herian, U., and Bartenschlager, R., *In vitro studies on the activation of the hepatitis C virus NS3 proteinase by the NS4A cofactor*. *Virology*, 1996. **221**(1): p. 54-66.
135. Koike, K., *Molecular basis of hepatitis C virus-associated hepatocarcinogenesis: lessons from animal model studies*. *Clin. Gastroenterol. Hepatol.*, 2005. **3**(10 Suppl 2): p. S132-5.
136. Koike, K., *Oxidative stress and apoptosis in hepatitis C: the core issue*. *J. Gastroenterol.*, 2006. **41**(3): p. 292-4.
137. Koike, K., *Hepatitis C virus contributes to hepatocarcinogenesis by modulating metabolic and intracellular signaling pathways*. *J. Gastroenterol. Hepatol.*, 2007. **22 Suppl 1**: p. S108-11.

138. Koike, K. and Miyoshi, H., *Oxidative stress and hepatitis C viral infection*. Hepatol. Res., 2006. **34**(2): p. 65-73.
139. Kopp, M., Murray, C.L., Jones, C.T., and Rice, C.M., *Genetic analysis of the carboxy-terminal region of the hepatitis C virus core protein*. J. Virol., 2010. **84**(4): p. 1666-73.
140. Korenaga, M., Okuda, M., Otani, K., Wang, T., Li, Y., and Weinman, S.A., *Mitochondrial dysfunction in hepatitis C*. J. Clin. Gastroenterol., 2005. **39**(4 Suppl 2): p. S162-6.
141. Kota, S., Coito, C., Mousseau, G., Lavergne, J.P., and Strosberg, A.D., *Peptide inhibitors of hepatitis C virus core oligomerization and virus production*. J. Gen. Virol., 2009. **90**(Pt 6): p. 1319-28.
142. Krugman, S., Friedman, H., and Lattimer, C., *Viral hepatitis, type A. Identification by specific complement fixation and immune adherence tests*. N. Engl. J. Med., 1975. **292**(22): p. 1141-3.
143. Kuhn, R.J., Zhang, W., Rossmann, M.G., Pletnev, S.V., Corver, J., Lenches, E., Jones, C.T., Mukhopadhyay, S., Chipman, P.R., Strauss, E.G., Baker, T.S., and Strauss, J.H., *Structure of dengue virus: implications for flavivirus organization, maturation, and fusion*. Cell, 2002. **108**(5): p. 717-25.
144. Kuiken, C. and Simmonds, P., *Nomenclature and numbering of the hepatitis C virus*. Methods Mol. Biol., 2009. **510**: p. 33-53.
145. Kunkel, M., Lorinczi, M., Rijnbrand, R., Lemon, S.M., and Watowich, S.J., *Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein*. J. Virol., 2001. **75**(5): p. 2119-29.
146. Kunkel, M. and Watowich, S.J., *Conformational changes accompanying self-assembly of the hepatitis C virus core protein*. Virology, 2002. **294**(2): p. 239-45.
147. Kunkel, M. and Watowich, S.J., *Biophysical characterization of hepatitis C virus core protein: implications for interactions within the virus and host*. FEBS Lett., 2004. **557**(1-3): p. 174-80.
148. Kushima, Y., Wakita, T., and Hijikata, M., *A disulfide-bonded dimer of the core protein of hepatitis C virus is important for virus-like particle production*. J. Virol., 2010. **84**(18): p. 9118-27.
149. Ladaviere, L., Deleage, G., Montserret, R., Dalbon, P., Jolivet, M., and Penin, F., *Structural Analysis of the Immunodominant Antigenic Region of the Hepatitis C Virus Capsid Protein by NMR*. Unpublished, 1999.

150. Lee, W., Harvey, T.S., Yin, Y., Yau, P., Litchfield, D., and Arrowsmith, C.H., *Solution structure of the tetrameric minimum transforming domain of p53*. *Nat Struct Biol*, 1994. **1**(12): p. 877-90.
151. Lemberg, M.K. and Martoglio, B., *Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis*. *Mol. Cell.*, 2002. **10**(4): p. 735-44.
152. Lesburg, C.A., Cable, M.B., Ferrari, E., Hong, Z., Mannarino, A.F., and Weber, P.C., *Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site*. *Nat. Struct. Biol.*, 1999. **6**(10): p. 937-43.
153. Li, D., Takyar, S.T., Lott, W.B., and Gowans, E.J., *Amino acids 1-20 of the hepatitis C virus (HCV) core protein specifically inhibit HCV IRES-dependent translation in HepG2 cells, and inhibit both HCV IRES- and cap-dependent translation in HuH7 and CV-1 cells*. *J. Gen. Virol.*, 2003. **84**(Pt 4): p. 815-25.
154. Liang, Y., Kang, C.B., and Yoon, H.S., *Molecular and structural characterization of the domain 2 of hepatitis C virus non-structural protein 5A*. *Mol Cells*, 2006. **22**(1): p. 13-20.
155. Liang, Y., Kang, C.B., and Yoon, H.S., *Molecular and structural characterization of the domain 2 of hepatitis C virus non-structural protein 5A*. *Mol. Cells*, 2006. **22**(1): p. 13-20.
156. Liang, Y., Ye, H., Kang, C.B., and Yoon, H.S., *Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded*. *Biochemistry*, 2007. **46**(41): p. 11550-8.
157. Lin, C., Lindenbach, B.D., Pragai, B.M., McCourt, D.W., and Rice, C.M., *Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini*. *J. Virol.*, 1994. **68**(8): p. 5063-73.
158. Lin, C., Thomson, J.A., and Rice, C.M., *A central region in the hepatitis C virus NS4A protein allows formation of an active NS3-NS4A serine proteinase complex in vivo and in vitro*. *J. Virol.*, 1995. **69**(7): p. 4373-80.
159. Lin, C., Wu, J.W., Hsiao, K., and Su, M.S., *The hepatitis C virus NS4A protein: interactions with the NS4B and NS5A proteins*. *J. Virol.*, 1997. **71**(9): p. 6465-71.
160. Lindenbach, B.D., Evans, M.J., Syder, A.J., Wolk, B., Tellinghuisen, T.L., Liu, C.C., Maruyama, T., Hynes, R.O., Burton, D.R., McKeating, J.A., and Rice, C.M., *Complete replication of hepatitis C virus in cell culture*. *Science*, 2005. **309**(5734): p. 623-6.
161. Lindenbach, B.D., Pragai, B.M., Montserret, R., Beran, R.K., Pyle, A.M., Penin, F., and Rice, C.M., *The C terminus of hepatitis C virus NS4A encodes an electrostatic*

- switch that regulates NS5A hyperphosphorylation and viral replication.* J. Virol., 2007. **81**(17): p. 8905-18.
162. Liu, Q., Tackney, C., Bhat, R.A., Prince, A.M., and Zhang, P., *Regulated processing of hepatitis C virus core protein is linked to subcellular localization.* J. Virol., 1997. **71**(1): p. 657-62.
163. Lo, S.Y., Masiarz, F., Hwang, S.B., Lai, M.M., and Ou, J.H., *Differential subcellular localization of hepatitis C virus core gene products.* Virology, 1995. **213**(2): p. 455-61.
164. Lo, S.Y., Selby, M.J., and Ou, J.H., *Interaction between hepatitis C virus core protein and E1 envelope protein.* J. Virol., 1996. **70**(8): p. 5177-82.
165. Lobley, A., Whitmore, L., and Wallace, B.A., *DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra.* Bioinformatics, 2002. **18**(1): p. 211-2.
166. Lohmann, V., Korner, F., Herian, U., and Bartenschlager, R., *Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity.* J. Virol., 1997. **71**(11): p. 8416-28.
167. Lorenzo, L.J., Duenas-Carrera, S., Falcon, V., Acosta-Rivero, N., Gonzalez, E., de la Rosa, M.C., Menendez, I., and Morales, J., *Assembly of truncated HCV core antigen into virus-like particles in Escherichia coli.* Biochem. Biophys. Res. Commun., 2001. **281**(4): p. 962-5.
168. Lu, W. and Ou, J.H., *Phosphorylation of hepatitis C virus core protein by protein kinase A and protein kinase C.* Virology, 2002. **300**(1): p. 20-30.
169. Lubin, D.J., Butler, J.S., and Loh, S.N., *Folding of tetrameric p53: oligomerization and tumorigenic mutations induce misfolding and loss of function.* J Mol Biol, 2010. **395**(4): p. 705-16.
170. Ma, H.C., Ke, C.H., Hsieh, T.Y., and Lo, S.Y., *The first hydrophobic domain of the hepatitis C virus E1 protein is important for interaction with the capsid protein.* J. Gen. Virol., 2002. **83**(Pt 12): p. 3085-92.
171. Ma, H.C., Ku, Y.Y., Hsieh, Y.C., and Lo, S.Y., *Characterization of the cleavage of signal peptide at the C-terminus of hepatitis C virus core protein by signal peptide peptidase.* J. Biomed. Sci., 2007. **14**(1): p. 31-41.
172. Ma, L., Jones, C.T., Groesch, T.D., Kuhn, R.J., and Post, C.B., *Solution structure of dengue virus capsid protein reveals another fold.* Proc. Natl. Acad. Sci. U S A, 2004. **101**(10): p. 3414-9.

173. Ma, Y., Anantpadma, M., Timpe, J.M., Shanmugam, S., Singh, S.M., Lemon, S.M., and Yi, M., *Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins*. J. Virol., 2011. **85**(1): p. 86-97.
174. Ma, Y., Yates, J., Liang, Y., Lemon, S.M., and Yi, M., *NS3 helicase domains involved in infectious intracellular hepatitis C virus particle assembly*. J. Virol., 2008. **82**(15): p. 7624-39.
175. Maillard, P., Krawczynski, K., Nitkiewicz, J., Bronnert, C., Sidorkiewicz, M., Gounon, P., Dubuisson, J., Faure, G., Crainic, R., and Budkowska, A., *Nonenveloped nucleocapsids of hepatitis C virus in the serum of infected patients*. J. Virol., 2001. **75**(17): p. 8240-50.
176. Majeau, N., Bolduc, M., Duvignaud, J.B., Fromentin, R., and Leclerc, D., *Effect of cAMP-dependent protein kinase A (PKA) on HCV nucleocapsid assembly and degradation*. Biochem. Cell. Biol., 2007. **85**(1): p. 78-87.
177. Majeau, N., Fromentin, R., Savard, C., Duval, M., Tremblay, M.J., and Leclerc, D., *Palmitoylation of hepatitis C virus core protein is important for virion production*. J. Biol. Chem., 2009. **284**(49): p. 33915-25.
178. Majeau, N., Gagne, V., Boivin, A., Bolduc, M., Majeau, J.A., Ouellet, D., and Leclerc, D., *The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation*. J. Gen. Virol., 2004. **85**(Pt 4): p. 971-81.
179. Manavalan, P. and Johnson, W.C., Jr., *Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra*. Anal Biochem, 1987. **167**(1): p. 76-85.
180. Mandel, A.M., Akke, M., and Palmer, A.G., 3rd, *Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme*. J Mol Biol, 1995. **246**(1): p. 144-63.
181. Marley, J., Lu, M., and Bracken, C., *A method for efficient isotopic labeling of recombinant proteins*. J Biomol NMR, 2001. **20**(1): p. 71-5.
182. Masaki, T., Suzuki, R., Murakami, K., Aizaki, H., Ishii, K., Murayama, A., Date, T., Matsuura, Y., Miyamura, T., Wakita, T., and Suzuki, T., *Interaction of hepatitis C virus nonstructural protein 5A with core protein is critical for the production of infectious virus particles*. J. Virol., 2008. **82**(16): p. 7964-76.
183. Matsumoto, M., Hwang, S.B., Jeng, K.S., Zhu, N., and Lai, M.M., *Homotypic interaction and multimerization of hepatitis C virus core protein*. Virology, 1996. **218**(1): p. 43-51.

184. McLauchlan, J., *Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes*. J. Viral. Hepat., 2000. **7**(1): p. 2-14.
185. McLauchlan, J., Lemberg, M.K., Hope, G., and Martoglio, B., *Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets*. EMBO J., 2002. **21**(15): p. 3980-8.
186. McMullan, L.K., Grakoui, A., Evans, M.J., Mihalik, K., Puig, M., Branch, A.D., Feinstone, S.M., and Rice, C.M., *Evidence for a functional RNA element in the hepatitis C virus core gene*. Proc. Natl. Acad. Sci. U S A, 2007. **104**(8): p. 2879-84.
187. Menez, R., Bossus, M., Muller, B.H., Sibai, G., Dalbon, P., Ducancel, F., Jolivet-Reynaud, C., and Stura, E.A., *Crystal structure of a hydrophobic immunodominant antigenic site on hepatitis C virus core protein complexed to monoclonal antibody 19D9D6*. J. Immunol., 2003. **170**(4): p. 1917-24.
188. Miller, R.H. and Purcell, R.H., *Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups*. Proc. Natl. Acad. Sci. U S A, 1990. **87**(6): p. 2057-61.
189. Miyanari, Y., Atsuzawa, K., Usuda, N., Watashi, K., Hishiki, T., Zayas, M., Bartenschlager, R., Wakita, T., Hijikata, M., and Shimotohno, K., *The lipid droplet is an important organelle for hepatitis C virus production*. Nat. Cell. Biol., 2007. **9**(9): p. 1089-97.
190. Montserret, R., Saint, N., Vanbelle, C., Salvay, A.G., Simorre, J.P., Ebel, C., Sapay, N., Renisio, J.G., Bockmann, A., Steinmann, E., Pietschmann, T., Dubuisson, J., Chipot, C., and Penin, F., *NMR structure and ion channel activity of the p7 protein from hepatitis C virus*. J. Biol. Chem., 2010. **285**(41): p. 31446-61.
191. Moradpour, D., Gosert, R., Egger, D., Penin, F., Blum, H.E., and Bienz, K., *Membrane association of hepatitis C virus nonstructural proteins and identification of the membrane alteration that harbors the viral replication complex*. Antiviral Res., 2003. **60**(2): p. 103-9.
192. Mori, Y., Moriishi, K., and Matsuura, Y., *Hepatitis C virus core protein: its coordinate roles with PA28gamma in metabolic abnormality and carcinogenicity in the liver*. Int. J. Biochem. Cell. Biol., 2008. **40**(8): p. 1437-42.
193. Mousseau, G., Kota, S., Takahashi, V., Frick, D.N., and Strosberg, A.D., *Dimerization-driven interaction of hepatitis C virus core protein with NS3 helicase*. J. Gen. Virol., 2011. **92**(Pt 1): p. 101-11.
194. Mukhopadhyay, S., Kim, B.S., Chipman, P.R., Rossmann, M.G., and Kuhn, R.J., *Structure of West Nile virus*. Science, 2003. **302**(5643): p. 248.

195. Murray, C.L., Marcotrigiano, J., and Rice, C.M., *Bovine viral diarrhea virus core is an intrinsically disordered protein that binds RNA*. J. Virol., 2008. **82**(3): p. 1294-304.
196. Nakai, K., Okamoto, T., Kimura-Someya, T., Ishii, K., Lim, C.K., Tani, H., Matsuo, E., Abe, T., Mori, Y., Suzuki, T., Miyamura, T., Nunberg, J.H., Moriishi, K., and Matsuura, Y., *Oligomerization of hepatitis C virus core protein is crucial for interaction with the cytoplasmic domain of E1 envelope protein*. J. Virol., 2006. **80**(22): p. 11265-73.
197. Nolandt, O., Kern, V., Muller, H., Pfaff, E., Theilmann, L., Welker, R., and Krausslich, H.G., *Analysis of hepatitis C virus core protein interaction domains*. J. Gen. Virol., 1997. **78 (Pt 6)**: p. 1331-40.
198. Ogino, T., Fukuda, H., Imajoh-Ohmi, S., Kohara, M., and Nomoto, A., *Membrane binding properties and terminal residues of the mature hepatitis C virus capsid protein in insect cells*. J. Virol., 2004. **78**(21): p. 11766-77.
199. Okamoto, K., Mori, Y., Komoda, Y., Okamoto, T., Okochi, M., Takeda, M., Suzuki, T., Moriishi, K., and Matsuura, Y., *Intramembrane processing by signal peptide peptidase regulates the membrane localization of hepatitis C virus core protein and viral propagation*. J. Virol., 2008. **82**(17): p. 8349-61.
200. Okamoto, K., Moriishi, K., Miyamura, T., and Matsuura, Y., *Intramembrane proteolysis and endoplasmic reticulum retention of hepatitis C virus core protein*. J. Virol., 2004. **78**(12): p. 6370-80.
201. Palmer, A.G., Rance, M., and Wright, P.E., *Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-12 Heteronuclear Nmr-Spectroscopy*. Journal of the American Chemical Society, 1991. **113**(12): p. 4371-4380.
202. Perez-Berna, A.J., Veiga, A.S., Castanho, M.A., and Villalain, J., *Hepatitis C virus core protein binding to lipid membranes: the role of domains 1 and 2*. J. Viral. Hepat., 2008. **15**(5): p. 346-56.
203. Phan, T., Beran, R.K., Peters, C., Lorenz, I.C., and Lindenbach, B.D., *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., 2009. **83**(17): p. 8379-95.
204. Prince, A.M., *An antigen detected in the blood during the incubation period of serum hepatitis*. Proc. Natl. Acad. Sci. U S A, 1968. **60**(3): p. 814-21.
205. Provencher, S.W. and Glockner, J., *Estimation of globular protein secondary structure from circular dichroism*. Biochemistry, 1981. **20**(1): p. 33-7.

206. Racanelli, V. and Rehermann, B., *Hepatitis C virus infection: when silence is deception*. Trends Immunol., 2003. **24**(8): p. 456-64.
207. Rajagopalan, S., Andreeva, A., Rutherford, T.J., and Fersht, A.R., *Mapping the physical and functional interactions between the tumor suppressors p53 and BRCA2*. Proc Natl Acad Sci U S A, 2010. **107**(19): p. 8587-92.
208. Ratinier, M., Boulant, S., Crussard, S., McLauchlan, J., and Lavergne, J.P., *Subcellular localizations of the hepatitis C virus alternate reading frame proteins*. Virus Res., 2009. **139**(1): p. 106-10.
209. Ray, R.B. and Ray, R., *Hepatitis C virus core protein: intriguing properties and functional relevance*. FEMS Microbiol. Lett., 2001. **202**(2): p. 149-56.
210. Ray, R.B., Steele, R., Meyer, K., and Ray, R., *Transcriptional repression of p53 promoter by hepatitis C virus core protein*. J Biol Chem, 1997. **272**(17): p. 10983-6.
211. Reed, K.E., Gorbalenya, A.E., and Rice, C.M., *The NS5A/NS5 proteins of viruses from three genera of the family flaviviridae are phosphorylated by associated serine/threonine kinases*. J. Virol., 1998. **72**(7): p. 6199-206.
212. Reshetnyak, V.I., Karlovich, T.I., and Ilchenko, L.U., *Hepatitis G virus*. World J. Gastroenterol., 2008. **14**(30): p. 4725-34.
213. Rijnbrand, R.C. and Lemon, S.M., *Internal ribosome entry site-mediated translation in hepatitis C virus replication*. Curr. Top. Microbiol. Immunol., 2000. **242**: p. 85-116.
214. Rodriguez-Casado, A., Molina, M., and Carmona, P., *Conformational features of truncated hepatitis C virus core protein in virus-like particles*. Biopolymers, 2006. **82**(4): p. 334-8.
215. Rodriguez-Casado, A., Molina, M., and Carmona, P., *Core protein-nucleic acid interactions in hepatitis C virus as revealed by Raman and circular dichroism spectroscopy*. Appl. Spectrosc., 2007. **61**(11): p. 1219-24.
216. Rodriguez-Casado, A., Molina, M., and Carmona, P., *Spectroscopic study of conformational changes accompanying self-assembly of HCV core protein*. Proteins, 2007. **66**(1): p. 110-7.
217. Rouzioux, C., *Le risque viral en transfusion sanguine*. Journal de pédiatrie et de puériculture, 1995. **5**: p. 273-276.
218. Santolini, E., Migliaccio, G., and La Monica, N., *Biosynthesis and biochemical properties of the hepatitis C virus core protein*. J. Virol., 1994. **68**(6): p. 3631-41.

219. Savard, P.Y. and Gagné, S.M., *Backbone dynamics of TEM-1 determined by NMR: evidence for a highly ordered protein*. *Biochemistry*, 2006. **45**(38): p. 11414-24.
220. Schmidt-Mende, J., Bieck, E., Hugle, T., Penin, F., Rice, C.M., Blum, H.E., and Moradpour, D., *Determinants for membrane association of the hepatitis C virus RNA-dependent RNA polymerase*. *J. Biol. Chem.*, 2001. **276**(47): p. 44052-63.
221. Schuttler, C.G., Fiedler, N., Schmidt, K., Repp, R., Gerlich, W.H., and Schaefer, S., *Suppression of hepatitis B virus enhancer 1 and 2 by hepatitis C virus core protein*. *J. Hepatol.*, 2002. **37**(6): p. 855-62.
222. Schwer, B., Ren, S., Pietschmann, T., Kartenbeck, J., Kaehlcke, K., Bartenschlager, R., Yen, T.S., and Ott, M., *Targeting of hepatitis C virus core protein to mitochondria through a novel C-terminal localization motif*. *J. Virol.*, 2004. **78**(15): p. 7958-68.
223. Seong, Y.R., Lee, C.H., and Im, D.S., *Characterization of the structural proteins of hepatitis C virus expressed by an adenovirus recombinant*. *Virus Res.*, 1998. **55**(2): p. 177-85.
224. Shao, S.W., Wu, W.B., Bian, Z.Q., Yu, J.G., Zhao, P., Zhao, L.J., Zhu, S.Y., and Qi, Z.T., *Hepatitis C virus F protein inhibits cell apoptosis by activation of intracellular NF-kappaB pathway*. *Hepatol Res.*, 2009. **39**(3): p. 282-9.
225. Sharma, K., Didier, P., Darlix, J.L., de Rocquigny, H., Bensikaddour, H., Lavergne, J.P., Penin, F., Lessinger, J.M., and Mely, Y., *Kinetic analysis of the nucleic acid chaperone activity of the hepatitis C virus core protein*. *Nucleic Acids Res.*, 2010. **38**(11): p. 3632-42.
226. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A., *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts*. *J Biomol NMR*, 2009. **44**(4): p. 213-23.
227. Shepard, C.W., Finelli, L., and Alter, M.J., *Global epidemiology of hepatitis C virus infection*. *Lancet Infect. Dis.*, 2005. **5**(9): p. 558-67.
228. Shi, S.T. and Lai, M.M.C., *HCV 5' and 3'UTR: When Translation Meets Replication.*, in *Hepatitis C Viruses: Genomes and Molecular Biology*, S.L. Tan, Editor. 2006, Horizon Bioscience: Norfolk (UK). p. 49-87.
229. Shi, S.T., Polyak, S.J., Tu, H., Taylor, D.R., Gretch, D.R., and Lai, M.M., *Hepatitis C virus NS5A colocalizes with the core protein on lipid droplets and interacts with apolipoproteins*. *Virology*, 2002. **292**(2): p. 198-210.
230. Shih, C.M., Chen, C.M., Chen, S.Y., and Lee, Y.H., *Modulation of the trans-suppression activity of hepatitis C virus core protein by phosphorylation*. *J. Virol.*, 1995. **69**(2): p. 1160-71.

231. Shih, C.M., Lo, S.J., Miyamura, T., Chen, S.Y., and Lee, Y.H., *Suppression of hepatitis B virus expression and replication by hepatitis C virus core protein in HuH-7 cells*. J. Virol., 1993. **67**(10): p. 5823-32.
232. Shimoike, T., Koyama, C., Murakami, K., Suzuki, R., Matsuura, Y., Miyamura, T., and Suzuki, T., *Down-regulation of the internal ribosome entry site (IRES)-mediated translation of the hepatitis C virus: critical role of binding of the stem-loop IIIId domain of IRES and the viral core protein*. Virology, 2006. **345**(2): p. 434-45.
233. Shimoike, T., Mimori, S., Tani, H., Matsuura, Y., and Miyamura, T., *Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation*. J. Virol., 1999. **73**(12): p. 9718-25.
234. Sibinga, C.T.S., Cash, J.D., and Corash, L.M., *Inactivation of Viruses, Bacteria, Protozoa, and Leukocytes in Labile Blood Components by Using Nucleic Acid Targeted Methods*, in *Transfusion Medicine: Quo Vadis? What Has Been Achieved, What Is To Be Expected*. 2001, Springer US. p. 113-123.
235. Simmonds, P., Alberti, A., Alter, H.J., Bonino, F., Bradley, D.W., Brechot, C., Brouwer, J.T., Chan, S.W., Chayama, K., Chen, D.S., and et al., *A proposed system for the nomenclature of hepatitis C viral genotypes*. Hepatology, 1994. **19**(5): p. 1321-4.
236. Simons, J.N., Leary, T.P., Dawson, G.J., Pilot-Matias, T.J., Muerhoff, A.S., Schlauder, G.G., Desai, S.M., and Mushahwar, I.K., *Isolation of novel virus-like sequences associated with human hepatitis*. Nat. Med., 1995. **1**(6): p. 564-9.
237. Slupsky, C.M., Boyko, R.F., Booth, V.K., and Sykes, B.D., *Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments*. J Biomol NMR, 2003. **27**(4): p. 313-21.
238. Spadaccini, R., D'Errico, G., D'Alessio, V., Notomista, E., Bianchi, A., Merola, M., and Picone, D., *Structural characterization of the transmembrane proximal region of the hepatitis C virus E1 glycoprotein*. Biochim. Biophys. Acta, 2010. **1798**(3): p. 344-53.
239. Sreerama, N. and Woody, R.W., *A self-consistent method for the analysis of protein secondary structure from circular dichroism*. Anal Biochem, 1993. **209**(1): p. 32-44.
240. Suzuki, R., Sakamoto, S., Tsutsumi, T., Rikimaru, A., Tanaka, K., Shimoike, T., Moriishi, K., Iwasaki, T., Mizumoto, K., Matsuura, Y., Miyamura, T., and Suzuki, T., *Molecular determinants for subcellular localization of hepatitis C virus core protein*. J. Virol., 2005. **79**(2): p. 1271-81.

241. Tabor, E., Gerety, R.J., Drucker, J.A., Seeff, L.B., Hoofnagle, J.H., Jackson, D.R., April, M., Barker, L.F., and Pineda-Tamondong, G., *Transmission of non-A, non-B hepatitis from man to chimpanzee*. *Lancet*, 1978. **1**(8062): p. 463-6.
242. Tanaka, Y., Shimoike, T., Ishii, K., Suzuki, R., Suzuki, T., Ushijima, H., Matsuura, Y., and Miyamura, T., *Selective binding of hepatitis C virus core protein to synthetic oligonucleotides corresponding to the 5' untranslated region of the viral genome*. *Virology*, 2000. **270**(1): p. 229-36.
243. Tellinghuisen, T.L., Foss, K.L., and Treadaway, J., *Regulation of hepatitis C virion production via phosphorylation of the NS5A protein*. *PLoS Pathog.*, 2008. **4**(3): p. e1000032.
244. Tellinghuisen, T.L., Marcotrigiano, J., Gorbalenya, A.E., and Rice, C.M., *The NS5A protein of hepatitis C virus is a zinc metalloprotein*. *J. Biol. Chem.*, 2004. **279**(47): p. 48576-87.
245. Timpe, J.M. and McKeating, J.A., *Hepatitis C virus entry: possible targets for therapy*. *Gut*, 2008. **57**(12): p. 1728-37.
246. Tomei, L., Failla, C., Santolini, E., De Francesco, R., and La Monica, N., *NS3 is a serine protease required for processing of hepatitis C virus polyprotein*. *J. Virol.*, 1993. **67**(7): p. 4017-26.
247. Tomei, L., Failla, C., Vitale, R.L., Bianchi, E., and De Francesco, R., *A central hydrophobic domain of the hepatitis C virus NS4A protein is necessary and sufficient for the activation of the NS3 protease*. *J. Gen. Virol.*, 1996. **77** (Pt 5): p. 1065-70.
248. Tsukiyama-Kohara, K., Iizuka, N., Kohara, M., and Nomoto, A., *Internal ribosome entry site within hepatitis C virus RNA*. *J. Virol.*, 1992. **66**(3): p. 1476-83.
249. Vassilaki, N. and Mavromara, P., *Two alternative translation mechanisms are responsible for the expression of the HCV ARFP/F/core+1 coding open reading frame*. *J. Biol. Chem.*, 2003. **278**(42): p. 40503-13.
250. Vassilaki, N. and Mavromara, P., *The HCV ARFP/F/core+1 protein: production and functional analysis of an unconventional viral product*. *IUBMB Life*, 2009. **61**(7): p. 739-52.
251. Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H.G., Mizokami, M., Bartenschlager, R., and Liang, T.J., *Production of infectious hepatitis C virus in tissue culture from a cloned viral genome*. *Nat. Med.*, 2005. **11**(7): p. 791-6.

252. Walewski, J.L., Keller, T.R., Stump, D.D., and Branch, A.D., *Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame*. RNA, 2001. **7**(5): p. 710-21.
253. Wang, C., Sarnow, P., and Siddiqui, A., *Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism*. J. Virol., 1993. **67**(6): p. 3338-44.
254. Wengler, G., Boege, U., Bischoff, H., and Wahn, K., *The core protein of the alphavirus Sindbis virus assembles into core-like nucleoproteins with the viral genome RNA and with other single-stranded nucleic acids in vitro*. Virology, 1982. **118**(2): p. 401-10.
255. Whitmore, L. and Wallace, B.A., *DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W668-73.
256. Whitmore, L. and Wallace, B.A., *Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases*. Biopolymers, 2008. **89**(5): p. 392-400.
257. Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A., and Smith, L.J., *Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques*. Biochemistry, 1999. **38**(50): p. 16424-31.
258. Williamson, C.D. and Colberg-Poley, A.M., *Access of viral proteins to mitochondria via mitochondria-associated membranes*. Rev. Med. Virol., 2009. **19**(3): p. 147-64.
259. Wishart, D.S. and Sykes, B.D., *The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data*. J Biomol NMR, 1994. **4**(2): p. 171-80.
260. Wishart, D.S., Sykes, B.D., and Richards, F.M., *The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy*. Biochemistry, 1992. **31**(6): p. 1647-51.
261. Wolf, M., Dimitrova, M., Baumert, T.F., and Schuster, C., *The major form of hepatitis C virus alternate reading frame protein is suppressed by core protein expression*. Nucleic Acids Res., 2008. **36**(9): p. 3054-64.
262. Wu, W.B., Shao, S.W., Zhao, L.J., Luan, J., Cao, J., Gao, J., Zhu, S.Y., and Qi, Z.T., *Hepatitis C virus F protein up-regulates c-myc and down-regulates p53 in human hepatoma HepG2 cells*. Intervirology, 2007. **50**(5): p. 341-6.
263. Wüthrich, K., *NMR of proteins and nucleic acids*. 1986: Wiley.

264. Xu, Z., Choi, J., Lu, W., and Ou, J.H., *Hepatitis C virus f protein is a short-lived protein associated with the endoplasmic reticulum*. J. Virol., 2003. **77**(2): p. 1578-83.
265. Xu, Z., Choi, J., Yen, T.S., Lu, W., Strohecker, A., Govindarajan, S., Chien, D., Selby, M.J., and Ou, J., *Synthesis of a novel hepatitis C virus protein by ribosomal frameshift*. EMBO J., 2001. **20**(14): p. 3840-8.
266. Yamaguchi, R., Momosaki, S., Gao, G., Hsia, C.C., Kojiro, M., Scudamore, C., and Tabor, E., *Truncated hepatitis C virus core protein encoded in hepatocellular carcinomas*. Int. J. Mol. Med, 2004. **14**(6): p. 1097-100.
267. Yan, B.S., Tam, M.H., and Syu, W.J., *Self-association of the C-terminal domain of the hepatitis-C virus core protein*. Eur. J. Biochem., 1998. **258**(1): p. 100-6.
268. Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J.L., Sardana, M., Steinkuehler, C., Tomei, L., De Francesco, R., Kuo, L.C., and Chen, Z., *Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form*. Protein Sci., 1998. **7**(4): p. 837-47.
269. Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A.D., Le, H.V., and Weber, P.C., *Structure of the hepatitis C virus RNA helicase domain*. Nat. Struct. Biol., 1997. **4**(6): p. 463-7.
270. Yasui, K., Wakita, T., Tsukiyama-Kohara, K., Funahashi, S.I., Ichikawa, M., Kajita, T., Moradpour, D., Wands, J.R., and Kohara, M., *The native form and maturation process of hepatitis C virus core protein*. J. Virol., 1998. **72**(7): p. 6048-55.
271. Yi, M., Ma, Y., Yates, J., and Lemon, S.M., *Trans-complementation of an NS2 defect in a late step in hepatitis C virus (HCV) particle assembly and maturation*. PLoS Pathog., 2009. **5**(5): p. e1000403.
272. Yi, M., Nakamoto, Y., Kaneko, S., Yamashita, T., and Murakami, S., *Delineation of regions important for heteromeric association of hepatitis C virus E1 and E2*. Virology, 1997. **231**(1): p. 119-29.
273. Yu, K.L., Jang, S.I., and You, J.C., *Identification of in vivo interaction between Hepatitis C Virus core protein and 5' and 3' UTR RNA*. Virus Res., 2009. **145**(2): p. 285-92.
274. Yu, X., Qiao, M., Atanasov, I., Hu, Z., Kato, T., Liang, T.J., and Zhou, Z.H., *Cryo-electron microscopy and three-dimensional reconstructions of hepatitis C virus particles*. Virology, 2007. **367**(1): p. 126-34.
275. Yuasa, T., Ishikawa, G., Manabe, S., Sekiguchi, S., Takeuchi, K., and Miyamura, T., *The particle size of hepatitis C virus estimated by filtration through microporous regenerated cellulose fibre*. J. Gen. Virol., 1991. **72** (Pt 8): p. 2021-4.

276. Zhang, J., Yamada, O., Yoshida, H., Iwai, T., and Araki, H., *Autogenous translational inhibition of core protein: implication for switch from translation to RNA replication in hepatitis C virus*. *Virology*, 2002. **293**(1): p. 141-50.
277. Zhang, W., Chipman, P.R., Corver, J., Johnson, P.R., Zhang, Y., Mukhopadhyay, S., Baker, T.S., Strauss, J.H., Rossmann, M.G., and Kuhn, R.J., *Visualization of membrane protein domains by cryo-electron microscopy of dengue virus*. *Nat. Struct. Biol.*, 2003. **10**(11): p. 907-12.
278. Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D.R., Wieland, S.F., Uprichard, S.L., Wakita, T., and Chisari, F.V., *Robust hepatitis C virus infection in vitro*. *Proc. Natl. Acad. Sci. U S A*, 2005. **102**(26): p. 9294-9.

