



**HAL**  
open science

# Étude de la duplication du domaine V3 de la région NS5A du Virus de l'Hépatite C de génotype 1b : Épidémiologie clinique et moléculaire et aspects fonctionnels

Hélène Le Guillou-Guillemette

► **To cite this version:**

Hélène Le Guillou-Guillemette. Étude de la duplication du domaine V3 de la région NS5A du Virus de l'Hépatite C de génotype 1b : Épidémiologie clinique et moléculaire et aspects fonctionnels. Médecine humaine et pathologie. Université d'Angers, 2014. Français. NNT : . tel-01784327

**HAL Id: tel-01784327**

**<https://theses.hal.science/tel-01784327>**

Submitted on 3 May 2018

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

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

# Thèse de Doctorat

Hélène LE GUILLOU-GUILLEMETTE

*Mémoire présenté en vue de l'obtention du  
grade de Docteur de l'Université d'Angers  
sous le label de L'Université Nantes Angers Le Mans*

École doctorale : *Biologie Santé*

Discipline : *Santé*

Spécialité : *Microbiologie*

Unité de recherche : *Equipe HIFIH-UPRES EA3859*

Soutenue le 8 juillet 2014

Thèse N° : 1405

## Étude de la duplication du domaine V3 de la région NS5A du Virus de l'Hépatite C de génotype 1b : Épidémiologie clinique et moléculaire et aspects fonctionnels

### JURY

Rapporteurs :	<b>Jacque IZOPET</b> , PU-PH, Université Paul Sabatier Toulouse III, CHU Toulouse <b>Cécile HENQUELL</b> , PH, Université d'Auvergne, Clermont-Ferrand
Examineurs :	<b>Pascale JEANNIN</b> , PU-PH, Université d'Angers <b>Catherine GAUDY-GRAFFIN</b> , MCU-PH, Université François Rabelais, Tours <b>Christophe LEMAIRE</b> , MCU, Université d'Angers
Présidente de jury	<b>Pascale JEANNIN</b> , PU-PH, Université d'Angers
Directeur de Thèse :	<b>Françoise LUNEL-FABIANI</b> , PU-PH, Université d'Angers

**A Monsieur le Professeur Jacques Izopet** : Je vous remercie de me faire l'honneur et le plaisir de rapporter cette thèse. Je vous remercie d'avoir soutenu ce projet au sein de l'AC11, puis de m'avoir encouragée régulièrement. Je vous remercie enfin de me faire partager vos connaissances et votre expérience en virologie à chacune de nos rencontres.

**A Madame le Docteur Cécile Henquell** : je te remercie de me faire l'honneur de juger ce travail et d'avoir accepté la charge de rapporteur. Je n'oublie surtout pas nos collaborations au sein de l'AC11, tes encouragements, ta disponibilité et ton aide au cours de ce projet, enfin je t'adresse un remerciement tout particulier pour la transmission des protocoles E1.

**A Madame le Professeur Pascale Jeannin** : je te remercie sincèrement d'avoir accepté de présider le jury de ma thèse, je te remercie aussi de nos riches collaborations, de ta disponibilité et de ton accueil toujours enthousiaste.

**A Madame le Docteur Catherine Gaudy-Graffin** : je te suis très reconnaissante d'avoir accepté d'examiner ce travail de thèse, je te remercie aussi de nos échanges scientifiques et de nos collaborations, parfois plus pratiques et plus « 15189 », toujours chaleureuses et sympathiques !

**A Monsieur le Docteur Christophe Lemaire** : je te remercie d'avoir accepté cette collaboration « virale » et de m'avoir initiée à la génétique microbienne, son langage et ses principes ; merci d'avoir abandonné parfois *Venturia inaequalis* et ses cousins pour ce projet.

**A Madame le Professeur Françoise Lunel-Fabiani** : je vous remercie vivement de la confiance que vous m'accordez depuis plusieurs années au sein de votre laboratoire, depuis mon pré-internat jusqu'au poste de praticien hospitalier, me confiant ce projet mais aussi beaucoup d'autres missions.

**A Sandrine Bertrais** : je te suis infiniment reconnaissante d'avoir accepté de prendre ce projet à bras le corps. La mission était délicate, tu n'as pas laissé tomber l'affaire et tu ne m'as surtout jamais laissé tomber. Je te remercie, de l'immense travail fourni, de tes questions et propositions statistiques, de ta disponibilité et de ta patience à mon égard.

Ce travail est aussi le fruit de nombreuses collaborations : que les virologues, cliniciens, ARC et TEC des 20 centres participant soient remerciés ici de leur partenariat si précieux ; le **Pr P. Calès et l'équipe HIFIH**, le **Pr C. Payan**, le **Dr E. Meurs** et son équipe, le **Dr D. Vitour** et **S. Dabo**, qui m'ont accueillie très simplement dans leur prestigieux institut ; le **Dr V. Tafforeau**, un grand merci pour les analyses de double-hybride, **l'équipe Inserm 892 et le Dr Y. Delneste** qui nous accueillies et largement guidées et conseillées en production de protéines recombinantes ; merci à **ML Gervais**. Un remerciement tout particulier à **Pascal Veillon** pour son aide et son data management ! Ma reconnaissance va aussi aux très nombreux patients qui ont acceptés de participer à cette étude.

Ce projet n'aurait pu aboutir sans le dynamisme et l'enthousiasme de mes co-équipières investies dans ce projet : **Adeline Pivert et Elise Bouthry**. Je suis profondément reconnaissante de votre ténacité et de votre engagement, du temps passé ensemble et de vos encouragements aussi. Un immense merci ! Elise, la direction de ta thèse et de ton M2 a été pour moi un très grand plaisir. Que votre avenir à toutes les deux soit plein de succès professionnels et personnels !

Je te remercie très sincèrement **Alexandra**, le travail à tes côtés est un plaisir et une richesse au quotidien, ton soutien et ton amitié, ton dynamisme et ton écoute me sont très précieux, je n'oublie pas aussi ton aide dans la dernière ligne droite ; puissions-nous continuer sur ce même chemin !

Un remerciement particulier à mes collègues virologues d'ARGO et d'ailleurs, devenues mes chères amies : **Sophie** (merci de ton aide si précieuse), **Anne, Julia, Agnès, Florence, Catherine et Babeth**. Merci de vos soutiens et amitiés indéfectibles.

Mes remerciements vont aussi à **l'équipe de virologie du CHU d'Angers**, votre dynamisme et votre investissement quotidien dans les projets sont un puissant moteur de motivation ; merci de votre confiance toujours, de votre bonne humeur très souvent, de votre patience aussi...Je te remercie, **Véronique**, de ta collaboration, de nos échanges virologiques ou plus personnels, de ton amitié et de tes encouragements ! Merci à **Axelle** et à **Marlène** pour votre soutien et votre compréhension ces derniers mois en particulier ; merci à **Christine, Maryline, Isabelle, Dominique, Sylvain, Aude, Corinne et Estelle** de ce que vous m'apportez au quotidien, merci à **Patricia, Céline et Karine**.

Que mes collègues angevins, cliniciens, biologistes ou qualitiens, soient aussi remerciés ici pour la richesse des échanges, la confiance réciproque et la passion partagée du travail hospitalier, merci à **M. Eric Pichard et M. Jean-Marie Chennebault**, mes maîtres infectiologues, à l'équipe VIH : **Pascale, Pierre, Yves-Marie**, mon guide bamakois, **Paule, Nicolas et Valérie** ; aux équipes du PBH, à mes amis microbiologistes : **Marie**, ma jumelle d'assistantat, **Bernard, Matthieu** et **Mme Catherine Varache, Anne et Valérie**, mes amies des RCEB/PTA, **Elise et Gilles**, mes guides Cofrac, **Charline**, superwoman de l'immunologie du VHC et bientôt du VIH...

A mes amis d'ici et d'ailleurs, toujours présents et fidèles : **Virginie et Nicolas, Cécile et Damien, Anne et Aymeric, Christophe et Cathy, Stéphanie et Alban, Dorothée et Hugues, Anne et Pierre, Christophe et Christelle**, à vos pioupioups aussi qui égaient tellement nos vies !

A ma famille, à **mes parents**, à **ma sœur Julia**, à **David** et à la petite **Mélissa**, à **mon frère Romain** et à **Clélia** ; merci de votre soutien et de vos encouragements, de votre aide et de votre écoute.

**A Thomas, avec tout mon amour, merci de ta présence à mes côtés chaque jour,  
A Colin et Timothée,  
Ce travail vous est dédié.**

# SOMMAIRE

ABREVIATIONS .....	9
INTRODUCTION .....	12
CHAPITRE 1 : INTRODUCTION BIBLIOGRAPHIQUE .....	14
<b>I. Le virus de l'Hépatite C .....</b>	<b>14</b>
<u>1. Caractéristiques du virus de l'hépatite C .....</u>	<u>14</u>
1.1. Taxonomie .....	14
1.2. Structure .....	16
1.3. Organisation génomique .....	17
1.3.1. Les régions non codantes (5'NC et 3'NC) .....	17
1.3.2. La région codante .....	19
1.3.2.1. Les protéines structurales .....	20
1.3.2.2. Les protéines non structurales .....	21
1.3.2.3. La protéine F .....	26
1.4. Cycle viral .....	26
1.4.1. Modèles d'étude .....	26
1.4.2. Etapes du cycle viral .....	31
<u>2. Diversité génétique du VHC .....</u>	<u>35</u>
2.1 <i>Épidémiologie</i> .....	35
2.1.1. Modes de transmission .....	36
2.1.2. Prévalence de l'infection .....	38
2.1.3. Epidémiologie moléculaire .....	41
2.2. Variabilité génétique .....	42
2.2.1. Génotypes et sous-types (échelle inter-individuelle) .....	43
2.2.2. Les quasi-espèces (échelle intra-individuelle) .....	44
2.2.3. Les recombinaisons .....	46
<u>3. Prise en charge de la maladie à VHC .....</u>	<u>47</u>
3.1. <i>Physiopathologie de l'infection par le VHC</i> .....	47
3.1.1. L'hépatite aiguë .....	47
3.1.2. L'hépatite chronique .....	47
3.1.3. La réponse immunitaire au cours de l'infection par le VHC .....	49

3.1.4. Les manifestations extra-hépatiques .....	53
3.2. <i>Prise en charge virologique de l'infection par le VHC</i> .....	54
3.2.1. Circonstances de découverte .....	54
3.2.2. Le diagnostic sérologique .....	54
3.2.3. Outils moléculaires de prise en charge et de suivi virologique .....	55
<u>3.2.3.1. Détection et quantification de l'ARN du VHC</u> .....	55
<u>3.2.3.2. Génotypage du VHC</u> .....	56
3.3. <i>Stratégies thérapeutiques</i> .....	56
3.3.1. Indications thérapeutiques .....	56
3.3.2. Arsenal thérapeutique .....	57
<u>3.3.2.1. La bithérapie conventionnelle</u> .....	57
<u>3.3.2.2. Les inhibiteurs de NS3/4A</u> .....	58
3.3.2.2.1. Les trithérapies de première génération .....	58
3.3.2.2.2. Les anti-protéases de deuxième génération .....	59
<u>3.3.2.3. Les nouvelles classes thérapeutiques</u> .....	59
3.3.3.3.1. Les inhibiteurs de la protéine NS5A.....	60
3.3.3.3.2. Les inhibiteurs de la polymérase NS5B .....	61
<u>3.3.2.4. Les autres molécules en développement</u> .....	64
3.3.2.4.1. Les autres inhibiteurs spécifiques .....	64
3.3.2.4.2. Les inhibiteurs ciblant la cellule hôte (HTAs : Host Targeting Antiviral) .....	64
<b>II. La protéine NS5A</b> .....	67
<u>1. Structure de la protéine NS5A et fonctions dans le cycle viral</u> .....	69
1.1. <i>Le domaine d'ancrage</i> .....	69
1.2. <i>Le domaine I</i> .....	70
1.3. <i>Le domaine II</i> .....	71
1.4. <i>Le domaine III</i> .....	72
<u>2. Protéine NS5A et système immunitaire de l'hôte</u> .....	74
2.1. <i>Interaction avec la réponse immunitaire innée</i> .....	74
2.2. <i>Interaction avec la réponse immunitaire adaptative</i> .....	75
<u>3. Protéine NS5A et oncogénèse</u> .....	76
3.1. <i>Interactions avec la régulation du cycle cellulaire</i> .....	76

3.2. Interactions avec les mécanismes cellulaires d'apoptose et de stress oxydatif .....	77
<b>III. Le polymorphisme génétique viral : origine et évolution .....</b>	<b>79</b>
1. Bases biologiques du polymorphisme viral: <u>mutation, indels, recombinaison et duplication</u> .....	79
2. Evolution des génomes viraux à ARN .....	83
2.1 Principe.....	83
2.2 Intérêt pratique.....	84
<b>IV. Place de l'étude du domaine V3 .....</b>	<b>87</b>
<b>CHAPITRE 2 : EPIDEMIOLOGIE CLINIQUE DE LA DUPLICATION V3 .....</b>	<b>89</b>
<b>I. Introduction .....</b>	<b>89</b>
<b>II. Article 1:</b> “Identification of a duplicated V3 domain in NS5A associated with cirrhosis and hepatocellular carcinoma in HCV-1b patients” .....	91
<b>CHAPITRE 3 : EPIDEMIOLOGIE MOLECULAIRE DES SOUCHES PORTANT UNE INSERTION DU DOMAINE V3.....</b>	<b>116</b>
<b>I. Introduction .....</b>	<b>116</b>
<b>II. Article 2:</b> “Evolutionary history of V3 duplication in HCV 1b strains” .....	118
<b>CHAPITRE 4 : ASPECTS FONCTIONNELS, RESULTATS PRELIMINAIRES....</b>	<b>142</b>
<b>I. Introduction .....</b>	<b>142</b>
<b>II. Matériel et Méthodes .....</b>	<b>144</b>
1. <u>Criblage en double hybride levure,</u> (Collaboration Dr L. Tafforeau, CIRI U1111/UMR5308, Lyon) .....	144
1.1. Principe du double-hybride en système levure .....	144
1.2. Constructions pour crible en double-hybride levure .....	146
1.3. Crible .....	146

1.4. Analyses .....	149
<u>2. Production de protéines et expériences de pull-down</u> .....	149
2.1. Principe du pull-down.....	149
2.2. Production en cellules de mammifères .....	150
2.2.1. Constructions des plasmides recombinants NS5A (Collaboration Dr E. Meurs, Unité Hépacivirus et immunité innée, Institut Pasteur, Paris) .....	150
2.2.2. Transfections.....	151
2.2.3. Immunofluorescence .....	152
2.2.4. Extraction des protéines totales .....	152
2.3. Production en bactéries .....	153
2.3.1. Construction du plasmide recombinant pFLAG-CTC <sup>TM</sup> .....	153
2.3.2. Purification des plasmides .....	154
2.3.3. Culture et expression de la protéine NS5A.....	154
2.3.4. Extraction des protéines totales .....	154
2.4. Purification de la protéine NS5A .....	155
2.5. Pull-down.....	156
<u>3. Analyses des protéines produites</u> .....	156
3.1. Gel SDS-PAGE coloré au bleu de Coomassie .....	156
3.2. Analyse par western blot.....	157
<b>III. Résultats</b> .....	158
<u>1. Crible en double-hybride levure</u> .....	158
<u>2. Expression de la protéine NS5A en cellules de mammifères</u> .....	162
2.1. Immunofluorescence en lignées cellulaires .....	162
2.2. Etablissement d'une lignée stable.....	164
2.3. Phase de production.....	165
2.4. Optimisation de la quantité de protéine NS5A obtenue .....	166
<u>3. Expression de la protéine NS5A en système bactérien</u> .....	167
3.1. Construction du plasmide recombinant pFLAG-CTC <sup>TM</sup> .....	167
3.2. Production de la protéine NS5A .....	168
3.3. Optimisation de la quantité de protéine NS5A obtenue.....	169



4. Interaction protéine-protéine : pull-down .....	170
4.1. Optimisation des conditions d'extraction de la protéine NS5A.....	170
4.2. Purification par affinité de la protéine NS5A : optimisation des conditions d'élution .....	170
4.3. Quantification de la protéine NS5A .....	171
4.4. Essais de pull-down .....	172
4.3.1. Essai sur la production cellulaire en Huh7.....	172
4.3.2. Essai sur la production bactérienne.....	173
<b>IV. Discussion .....</b>	<b>174</b>
<b>CHAPITRE 5 : CONCLUSION ET PERSPECTIVES .....</b>	<b>181</b>
<b>RÉFÉRENCES .....</b>	<b>185</b>
<b>ANNEXE .....</b>	<b>208</b>
<b>1. Genetic diversity of the Hepatitis C virus: impact and issues in the antiviral therapy. Highlight Topic 2007.</b>	
Le Guillou-Guillemette H, Vallet S, Gaudy-Graffin C, Payan C, Pivert A, Goudeau A and Lunel-Fabiani F. <i>World Journal of Gastroenterology</i> 2007, Vol 13, 17: 2416-24.....	208
<b>2. Duplication of the V3 domain in the genotype 1b hepatitis C virus NS5A protein: clonal analysis and physicochemical properties related to hepatocellular carcinoma occurrence.</b>	
Petsaris O, Vallet S, Le Guillou-Guillemette H., Veillon P., S. Gouriou S., Barbier G., Nousbaum JB., Saliou P., NKontchou G., Trinchet JC., Lunel-Fabiani F., Payan C. <i>Manuscrit soumis à J Viral Hepatitis</i> .....	220

## ABBREVIATIONS

3'NC	région 3' Non Codante
5'NC	région 5' Non Codante
ANRS	Agence Nationale de Recherche sur le Sida et les Hépatites
ADN	acide désoxyribonucléique
ADNc	ADN complémentaire
AMM	autorisation de mise sur le marché
ApoE	apolipoprotéine E
ATP	adénosine triphosphate
ATU	autorisation temporaire d'utilisation
ARFP	alternative reading frame protein
ARN	acide ribonucléique
ARN hélicase CAP-Rf	ARN hélicase cap-release factor
ASPM	mitotic spindle protein M
BDV	border disease virus
BVDV	bovine viral diarrhea virus
CCL5	chemokine ligand 5
CDK	cycline-dépendante kinase
CHC	carcinome hépato-cellulaire
CIDE-B	cell death-inducing DFFA-like effector B
CK2	caséine kinase cellulaire de type 2
CLDN-1	claudin-1 protein
CM	cryoglobulinémie mixte
CSFV	classical swine fever virus
CYPA	cyclophiline A
DAA	direct acting antiviral
DC-SIGN	dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DGV	diagnostic génomique viral
DHV	duck hepatitis virus
DV	dengue virus
eIF	eukaryotic initiation factor
ELISA	enzyme-linked immunosorbent assay
EMCV	encephalomyocarditis virus
EPgV	equine pegivirus
ERK	extracellular signal regulated kinase
Fas	apoptosis stimulating fragment
FASN	fattic acid synthase
GAG	glycosaminoglycans
GBV-B	hepatitis GB virus B
Grb2	growth factor receptor-bound protein 2
GTP	guanosine triphosphate
HNANB	hépatite non A non B
HnRNP L	heterogeneous nuclear Ribonucleoprotein L
HTA	host targeting antiviral
HTLV	human T cell leukemia virus
hVAP-33	human vesicle-associated membrane protein A of 33 kDa

HVR	highly variable region
ICTV	international committee on taxonomy of viruses
IGF2BP1	insulin-like growth factor 2 mRNA-binding protein 1
INF	interferon
IRES	internal ribosome entry site
ISGs	interferon stimulated genes
ISDR	interferon sensibility determining region
JEV	japanese encephalitis virus
JFH1	japanese fulminant hepatitis clone 1
LCS	low complexity sequences
LD	lipid droplets
LDL(-R)	low density lipoprotein (-receptor)
Ac anti-LKM1	anticorps anti-microsome de foie et de rein (liver kidney microsomal antibodies)
LNH	lymphoma non-hodgkinien
L-SIGN	liver/lymph node-specific intercellular adhesion molecule-3-grabbing integrin
MAPK	mitogen-activated protein kinase
MEPS	measurably evolving populations
Met-tRNA	complexe Methionine-ARN de transfert
MLV	murine leukemia virus
NFAR	nuclear factor associated with double-stranded RNA
NF- $\kappa$ B	nuclear factor-kappa B
NLS	nuclear localisation signal
NNI	non nucleos(t)ide polymerase inhibitor
NPHV	non primate hepacivirus
NTP	nucleoside triphosphate
OCLN	occluding protein
PBMC	peripheral blood mononuclear cell
PI3K	phosphatidylinositol-4,5-bisphosphate 3-kinase
PI4K $\alpha$	phosphatidylinositol-4 kinase III alpha
PCBP	poly(C)-binding protein
PKR-bd	protein kinase R binding region
PTB	polypyrimidine tract-binding protein
RdRp	ARN polymerase ARN dépendante
ROS	reactive oxygen signaling
RT-PCR	reverse transcription-polymerase chain reaction
SAP/JNK	stress activated protein kinase/c-Jun N-terminal kinase
SNP	single nucleotide polymorphism
SR-BI	scavenger receptor-class B type 1
STAT3	Signal transducer and activator of transcription 3
TBP	TATA binding protein
TLR	Toll-like receptor
TMA	transcription mediated amplification
TNF $\alpha$	tumor necrosis factor alpha
TRAIL	TNF-related apoptosis-inducing ligand
uPA-SCID	urokinase-type plasminogen activator-severe combined immunodeficiency

VLDL	very low density lipoprotein
VHB	virus de l'hépatite B
VHC	virus de l'hépatite C
VIH	virus de l'immunodéficience humaine acquise
V3 domain	variable domain 3
YFV	yellow fever virus

## INTRODUCTION

L'infection par le Virus de l'Hépatite C (VHC) concerne environ 180 millions d'individus dans le monde [1] et reste une préoccupation majeure de santé publique à cause de l'évolution silencieuse et péjorative de cette maladie. En effet, près de 80% des sujets infectés n'éliminent pas le virus après la phase aiguë et évoluent vers une infection chronique. Pendant une vingtaine d'année en moyenne, l'infection est asymptomatique, malheureusement environ 20% des patients vont développer une fibrose hépatique, puis une cirrhose, et parmi ces cas, 5% par an développeront un carcinome hépatocellulaire (CHC) [2].

Le VHC a été découvert en 1989. Il s'agit d'un virus de petite taille, enveloppé, avec un génome constitué d'un ARN positif simple brin. Ce virus, strictement humain, est caractérisé par sa grande diversité génétique. En effet, l'absence d'activité correctrice de son ARN polymérase ARN dépendante alliée à un niveau de réplication très élevé sont à l'origine d'un polymorphisme viral majeur. Ainsi, la classification du VHC revue en 2014 décrit 7 génotypes (notés de 1 à 7) partageant au moins 70% d'homologie de séquence, et 67 sous-types [3], et cette diversité s'exprime également à l'échelle de l'individu avec la notion de quasiespèce. Cette grande diversité génétique conduit à des difficultés dans la prise en charge de l'infection. D'une part au plan préventif, comme pour le VIH, la variabilité virale du VHC ne permet pas pour l'instant l'obtention d'un vaccin efficace malgré de nombreux essais [4, 5]. D'autre part, au plan curatif, plusieurs traitements ont déjà été proposés avec des succès importants. En effet, après l'utilisation de l'interféron standard, les traitements ont longtemps reposé sur la bithérapie associant l'interféron pégylé et la ribavirine. Cette ligne thérapeutique a permis l'éradication virale chez 80 à 90% des patients infectés par les génotypes 2 et 3 mais seulement chez 50% des patients infectés par les autres génotypes, en particulier le génotype 1 qui est un des plus prévalent en Europe et en Amérique du Nord [6]. Cette différence de sensibilité au traitement selon le génotype a donc conduit à de très nombreuses études sur les facteurs prédictifs de réponse au traitement. Il est ainsi apparu que l'âge, le sexe, la charge virale initiale, le génotype viral et très récemment plusieurs polymorphismes génétiques (SNP : single nucleotide polymorphism) situés autour et dans le gène codant l'interleukine 28 étaient des facteurs prédictifs de la réponse au traitement [7]. Concernant la diversité génétique virale, plusieurs régions du génome pourraient être impliquées dans l'échappement thérapeutique ; la région de l'enveloppe et plus particulièrement HVR1 et la région NS5A avec plusieurs domaines PKR, ISDR et V3 ont été particulièrement suggérées [8]. L'arrivée en 2011 des premières molécules inhibitrices de protéase et en 2014 des molécules de seconde

génération ainsi que de nouveaux inhibiteurs de polymérase et de NS5A avec une activité pangénotypique représentent une véritable révolution thérapeutique et fait naître l'espoir d'une éradication de cette infection [9].

Une précédente étude, menée au laboratoire de Virologie du CHU d'Angers chez des patients infectés par le VHC de génotype 1b, avait montré une diversité génétique supérieure et statistiquement significative de la quasiespèce virale chez les patients répondeurs à une bithérapie par rapport aux patients non répondeurs, la région V3 du gène NS5A étant prédictive de la réponse au traitement [10]. Au cours de ce travail, nous avons mis en évidence chez 5 patients une insertion dans le domaine V3. Après séquençage et alignement avec la souche de référence japonaise HCV-J, nous avons identifié cette insertion comme étant une probable duplication du domaine V3 [11]. Dans ce contexte, nous avons proposé une étude multicentrique afin d'identifier les circonstances épidémiologiques et l'éventuel rôle physiopathologique de cette duplication. Nous avons obtenu en 2006 le soutien du groupe AC11-VHC de l'ANRS pour mener ce travail. Ainsi, nous avons pu réaliser une étude épidémio-clinique en France, suivie d'un travail de phylogénie moléculaire inférant l'histoire évolutive de ces souches, des travaux préliminaires d'analyse fonctionnelle ont été aussi initiés.

## CHAPITRE 1 : INTRODUCTION BIBLIOGRAPHIQUE

### I. Le virus de l'Hépatite C

#### 1. Caractéristiques du virus de l'hépatite C

##### *1.1. Taxonomie*

Le virus de l'hépatite C a été découvert en 1989 et est considéré comme l'agent responsable de la majorité des hépatites de transmission parentérale et post transfusionnelle auparavant appelées hépatites non A non B (HNANB). En effet, malgré l'identification du virus de l'hépatite A en 1973 et de l'hépatite B en 1967, de nombreuses hépatites survenant après transfusion restaient sans étiologie et étaient alors appelées HNANB. L'agent suspecté semblait d'origine virale, il était transmissible par le sang au chimpanzé. C'est à partir de sérum de chimpanzé infecté par un concentré de facteur VIII ayant transmis une HNANB à des patients hémophiles que Houghton *et al.* ont pu identifier le VHC [12]. Cette équipe a construit une banque d'ADNc clonés en système bactérien à partir des ADN et ADNc extraits du sérum précédent. Les protéines exprimées à partir de cette banque ont été soumises à un immunoscreening (par technique d'immunoblot) en utilisant les sérums des patients souffrant d'une HNANB comme source d'anticorps [13]. Un premier complexe antigène-anticorps spécifique de l'agent des HNANB a été identifié (clone 5-1-1). Le clone correspondant à cette protéine a été utilisé ensuite comme sonde d'hybridation pour tester la banque. Ce travail de génétique inverse a permis l'identification d'un nouveau clone (c100). Cette protéine, fusionnée à la superoxyde dismutase, a ensuite permis la mise au point des premiers tests sérologiques. Le virus a alors été très rapidement caractérisé [12, 14].

Le VHC appartient à la famille des Flaviviridae caractérisée par la présence d'une enveloppe et d'un génome constitué d'un ARN monocaténaire de polarité positive [15]. Les virus de cette famille présentent des similarités d'organisation génomique et des homologies de séquences, ils sont répartis en quatre genres [16] actuellement répertoriés dans la classification taxonomique internationale (Figure 1):

- le genre *Flavivirus* qui compte près de 53 espèces parmi lesquelles des virus d'importance en pathologie humaine comme le virus de la fièvre jaune [YFV], le virus de la Dengue [DV] et le virus de l'encéphalite japonaise [JEV], tous responsables d'arboviroses,

- le genre *Pestivirus*, comprend les virus de la diarrhée bovine 1 et 2 [BVDV], le virus de la peste porcine classique [CSFV] et le virus de la peste ovine ou Border disease virus [BDV], responsables de pathologies animales,
- le genre *Hepacivirus* qui comprend le virus de l'hépatite C, et auquel est parfois ajouté le virus GBV-B découvert en 1995 [17, 18]. L'ajout de la nouvelle espèce NPHV (non primate hepacivirus), découvert chez le cheval, a été proposé en 2012 [19].
- la création d'un quatrième genre appelé *Pegivirus* a été récemment proposée. Le genre *Pegivirus A* regroupe les virus identifiés chez les primates, dont les virus GBV-A et GBV-C identifiés chez l'homme en 1995 et 1996 respectivement, et le genre *Pegivirus B* incluant le virus GBV-D découvert plus récemment en 2010 chez la chauve-souris [20] et l'EPgV (equine pegivirus) identifié chez le cheval en 2013 [21].

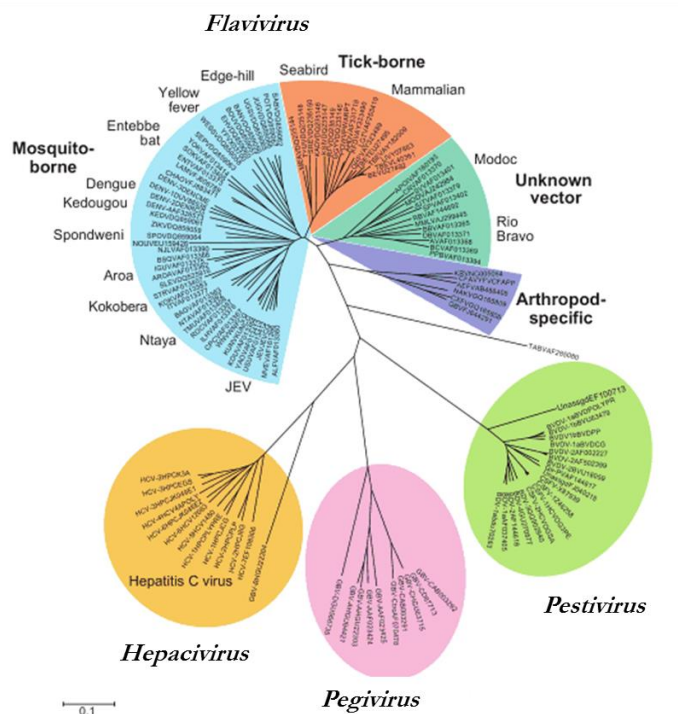


Figure 1. Organisation taxonomique : phylogénie des séquences NS5/NS5B des membres de la famille des Flaviviridae, d'après Simmonds, 2013 [22].



## 1.2. Structure

La structure virale du VHC a été caractérisée en 1994 par immunomicroscopie électronique [23]. Il s'agit d'un virus de petite taille (55 à 65nm de diamètre) enveloppé à capsid icosaédrique (Figure 2). L'enveloppe est constituée d'une double couche phospholipidique d'origine cellulaire dans laquelle sont enchâssées les glycoprotéines virales E1 (gp31) et E2 (gp70) organisées en hétérodimères. La capsid est composée de la protéine de core C (p21). Le génome est constitué d'une molécule d'ARN monocaténaire, linéaire, non segmenté, de polarité positive, monocistronique, d'environ 9,6 kb.

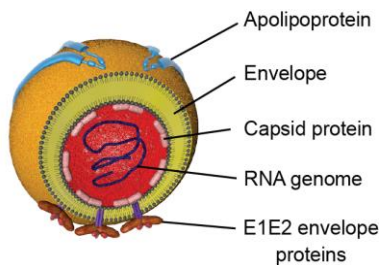


Figure 2. Schéma de la particule virale du VHC, d'après Fénéant, 2014 [24].

Le génome comprend un cadre de lecture ouvert unique codant une polyprotéine immature d'environ 3000 acides aminés (aa), qui est flanqué de régions non codantes en 5' et 3' (figure 3) [25]. Le clivage de la polyprotéine par des peptidases cellulaires et virales produit les protéines de structures (capside, enveloppe E1 et E2 et viroprotéine p7) et les protéines non structurales (NS2 à NS5B).

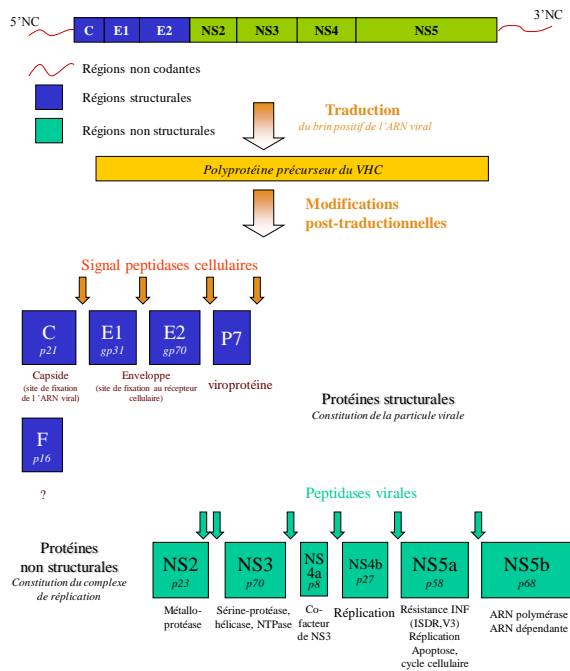


Figure 3. Organisation génomique du VHC adapté de Moradpour, 2001 [25]

### 1.3. Organisation génomique

#### 1.3.1. Les régions non codantes (5'NC et 3'NC)

Les régions très conservées 5'NC et 3'NC interviennent à la fois lors de la traduction et de la réplication virale. La région 5'NC est la région la plus conservée du génome du VHC, rendant son utilisation attractive en diagnostic moléculaire.

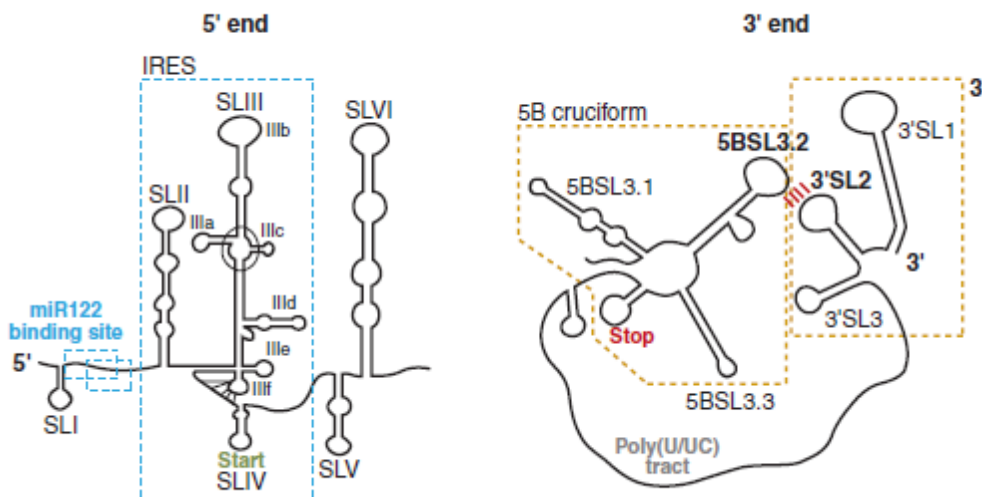


Figure 4. Schémas des régions 5' et 3' non codantes du VHC, d'après Dustin, 2007 [26].

La région 5'NC est constituée d'une séquence de 341 nucléotides divisée en quatre domaines avec des structures secondaires de type tige-boucle (I à IV) (Figure 4). Les domaines II à IV forment une structure servant de site d'entrée interne du ribosome (IRES). Un codon initiateur AUG au sein du domaine IV permet l'initiation de la traduction de la polyprotéine virale sans l'intermédiaire d'une coiffe méthylée. La sous-unité ribosomale 40S se fixe au niveau du domaine III constituant un premier complexe avec l'IRES puis eIF3 et le complexe (eIF2/GTP/Met-tRNA) sont recrutés pour former le complexe de pré-initiation. Pour terminer, survient l'hydrolyse du GTP, le relargage du facteur eIF2, l'arrivée de la sous-unité ribosomale 60S, ces étapes conduisent à la formation de l'unité ribosomale 80S par l'intervention des domaines II et III. Il a par ailleurs été démontré que les domaines II et IV interagissaient avec la sous unité 40S [27, 28]. L'IRES peut aussi permettre l'initiation de la traduction par une voie indépendante du facteur eIF2 [29, 30]. La région 5'NC interagit aussi avec de nombreuses protéines cellulaires telles que la protéine La autoantigène ou la protéine hnRNP L (heterogeneous nuclear Ribonucleoprotein L) ou PCBP (poly(C)-binding protein) qui stimulent la traduction, l'intervention du facteur PTB (polypyrimidine tract-binding protein) reste controversée [31-36]. Les domaines I et II sont impliqués dans la réplication virale [37].

La courte région 3'NC est structurée en trois parties : un domaine variable court, une région poly-uridylée (poly U-C) et enfin une région hautement conservée appelée région 3'X (Figure 4). Cette région est constituée de trois structures tige-boucle et est indispensable pour la réplication virale en favorisant l'initiation de la synthèse des brins d'ARN négatifs [38]. La

région 3' intervient aussi lors de la traduction. Ainsi le facteur IGF2BP1 (insulin-like growth factor 2 mRNA-binding protein 1) se lie au domaine IV en 5'NC et aux domaines variable court et polyU-C en 3'NC, ces interactions entraînent une stimulation de la traduction dirigée par l'IRES. D'autres facteurs, comme le complexe heterotrimeric NFAR (nuclear factor associated with double-stranded RNA) ou la protéine La, sont capables de se lier aux deux régions non codantes, permettant leur rapprochement et favorisant la traduction [39-41].

### 1.3.2. La région codante

L'unique cadre de lecture ouvert en aval de la région 5'NC code une polyprotéine précurseur clivée en 10 protéines virales matures par les protéases cellulaires et virales (Figure 5a). L'ensemble de ces protéines est localisé au niveau du réticulum endoplasmique (Figure 5b).

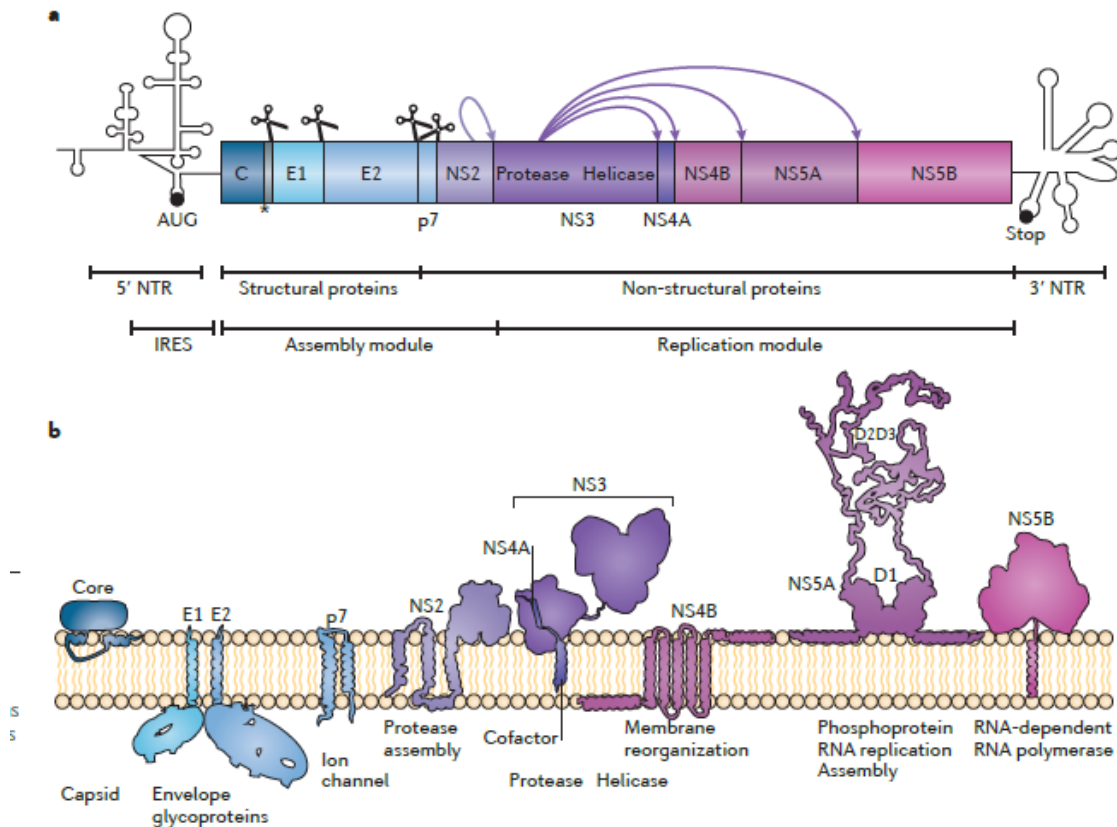


Figure 5. (a) Représentation des domaines et fonctions principales du génome du VHC ; (b) topologie membranaire des protéines virales clivées et de leur fonction, d'après Bartenschlager, 2013 [42]

### 1.3.2.1. Les protéines structurales

Dans la partie N-terminale du génome se trouvent les domaines codant les 4 protéines structurales : la protéine de capsid C, les deux glycoprotéines d'enveloppe E1 et E2 et la protéine p7. Elles sont clivées par des peptidases cellulaires dans le réticulum endoplasmique [43].

La protéine de capsid C (core) est clivée d'abord par une signal peptidase cellulaire dans la lumière du RE puis par une seconde signal peptide peptidase cellulaire qui permet la maturation définitive de la protéine core et son transfert du RE vers les gouttelettes lipidiques (LD ou lipid droplets). Cette protéine mature est constitué de deux domaines ; le domaine I en N-terminal avec une majorité de résidus basiques et le domaine II en C-terminal plus hydrophobe et probablement impliqué dans l'attachement aux LD. Il a été démontré que le complexe de réplication, composé des protéines non structurales et de l'ARN viral, est recruté à proximité de LD chargées de protéine core. En effet, l'absence de protéine core au niveau des LD empêche l'association entre le complexe de réplication et les LD. En outre, la protéine core interagit avec le domaine III de NS5A et permet alors l'assemblage viral [44, 45]. La présence de la protéine Core au niveau des LD est l'objet d'intenses recherches : cette protéine pourrait être un modulateur du métabolisme lipidique, elle pourrait ainsi favoriser l'accumulation de lipides au niveau hépatique [46]. Cette hypothèse expliquerait en partie les troubles métaboliques et en particulier la stéatose, qui accompagnent l'infection par le VHC. Cependant, le mécanisme conduisant à cette pathologie est encore mal connu. Pour terminer, de nombreux travaux ont mis en évidence l'implication de la protéine core dans le processus d'hépatocarcinogénèse. Elle est capable d'interagir avec de nombreux facteurs de transcription tels que la ribonucléoprotéine nucléaire K, l'ARN hélicase CAP-Rf, p53, p21 et NF- $\kappa$ B, ces interactions permettant une modulation des voies de signalisation cellulaire qui contrôlent le cycle cellulaire. La protéine core est aussi impliquée dans le contrôle de la croissance des hépatocytes, la survenue du stress oxydatif et l'apoptose [47].

Les glycoprotéines E1 et E2 sont insérées dans la bicouche lipidique provenant de la cellule infectée et l'ensemble constitue l'enveloppe virale. Elles sont clivées par des signal peptidases cellulaires à partir de la polyprotéine immature. Ces protéines constituées d'un grand domaine extra membranaire en N-terminal et d'un court domaine transmembranaire en C-terminal sont très fortement glycosylées, ce qui participe à une conformation adéquate. Ces

glycosylations permettent aussi de masquer les épitopes immunogènes de l'enveloppe et protègent le virus de la réponse immunitaire de l'hôte. Les protéines E1 et E2 s'assemblent de façon covalente en hétérodimères stables grâce à des liaisons disulfures [48]. La protéine E2 présente plusieurs domaines hypervariables (HVR) [49]. Le domaine HVR1, situé en N-terminal et d'une taille de 27aa, permettrait au virus d'échapper à la réponse immunitaire anticorps de par son apparente variabilité, réalisant ainsi un épitope neutralisant. Cependant les propriétés physico-chimiques des résidus à chaque position sont relativement conservées entre les génotypes, permettant une conservation de la conformation. Enfin, la région HVR1 et une autre région hypervariable, HVR2, décrite dans E2, sont impliquées dans le processus d'entrée du virus, en se liant à différents récepteurs [50].

La protéine p7 est une petite protéine hydrophobe de 63aa constituée de deux domaines transmembranaires en hélice  $\alpha$  (TM1 et TM2) reliés par une boucle ; elle est clivée par les signal-peptidases cellulaires du réticulum endoplasmique. Il s'agit d'une viroprotéine assurant une fonction de canal ionique cationique, elle s'assemble en hexamères ou heptamères et permet le relargage des particules virales infectieuses. La protéine p7 n'est pas indispensable à la réplication virale mais est absolument requise pour l'infectiosité du virus [51]. Elle modifie la perméabilité membranaire pour permettre l'assemblage viral et le relargage des virions. Il a été démontré que la protéine p7 était inhibée par l'amantadine [52, 53], molécule qui a été utilisée dans le traitement du VHC avec un succès modeste. Cette protéine reste néanmoins un candidat attractif pour le développement de nouveaux antiviraux [54].

#### 1.3.2.2. Les protéines non structurales

La seconde partie du cadre de lecture ouvert comprend les domaines codant les 6 protéines non structurales [43].

La protéine NS2 (p23) est une protéine hydrophobe transmembranaire dont la fonction principale est le clivage protéolytique NS2/NS3. Associée à la partie N-terminale de NS3, elle forme une métallo-protéase zinc-dépendante, qui réalise l'autoclivage de leur propre jonction. La protéine NS2 est une cystéine protéase avec une structure dimérique comprenant deux sites actifs [55]. La protéine NS2 est aussi le lieu privilégié des recombinaisons intergénomiques que ce soit chez des variants naturels ou des chimères construites en laboratoire [56-58]. Bien

que la fonction de NS2 au cours du cycle viral ne soit pas clairement établie, cette protéine ne semble pas indispensable à la réplication virale, mais plusieurs études tendent à lui attribuer un rôle lors de la phase finale de ce cycle. En effet, la capacité des génomes chimériques à produire des virions infectieux est hautement dépendante de la position de la jonction chimérique dans NS2 [57], les délétions et substitutions dans NS2 apparaissent délétères pour cette production virale, enfin certaines mutations entraînent une absence d'assemblage viral [59-61]. Il semblerait ainsi que NS2 coordonne l'assemblage viral via des interactions avec les protéines d'enveloppe E1 et E2, et les protéines non structurales NS3 et NS5A [62, 63]. La protéine NS2 interagit avec de nombreux facteurs cellulaires et entraîne une inhibition de l'expression de promoteurs de gènes cellulaires, codant notamment des cytokines (INF $\beta$ , CCL5/Rantes). D'autres analyses ont aussi rapporté la capacité de NS2 à se lier à des facteurs pro-apoptotiques, soit en association avec les protéines virales Core, E2 et NS5A (voie Fas), soit par liaison directe (inhibition du facteur CIDE-B). Cependant, l'ensemble de ces hypothèses doivent être confirmées en système cellulaire infectieux [64].

Les protéines NS3/NS4A (p70 et p8) : La protéine NS3 est multifonctionnelle car elle porte plusieurs activités enzymatique, NS4A est son cofacteur. La structure tridimensionnelle en feuillet  $\beta$  de la protéine NS3 a été déterminée en 1996 et a permis l'identification de quatre domaines majeurs : la triade catalytique, la poche de liaison au substrat, le site de liaison à NS4A et le domaine de liaison au zinc (Figure 6). Le co-facteur NS4A se compose de deux régions, l'une est hydrophobe en N-terminal et la seconde est hydrophile en C-terminal. Outre l'activité autocatalytique cystéine-protéase exercée conjointement avec NS2 et précédemment décrite, la protéine NS3 possède en N-terminal une activité sérine-protéase chymotrypsine-like permettant le clivage des protéines en aval (NS3/NS4A, NS4A/NS4B, NS4B/NS5A, NS5A/NS5B) et en C-terminal des activités NTPase/hélicase [65, 66]. La protéine NS4A est indispensable à l'activité protéase. En effet, la protéine NS3 est dépourvue de domaine transmembranaire mais son association de façon non covalente avec NS4A qui est une protéine transmembranaire permet un ancrage aux membranes cellulaires. La liaison à NS4A entraîne un changement de conformation de NS3, elle passe d'une structure en feuillet constituée de deux barils  $\beta$  contenant chacun 6 brins  $\beta$  à deux barils  $\beta$  dont l'un devient constitué de 8 feuillet ; cette liaison permet le rapprochement des résidus de la triade catalytique autorisant ainsi l'activité de protéolyse.

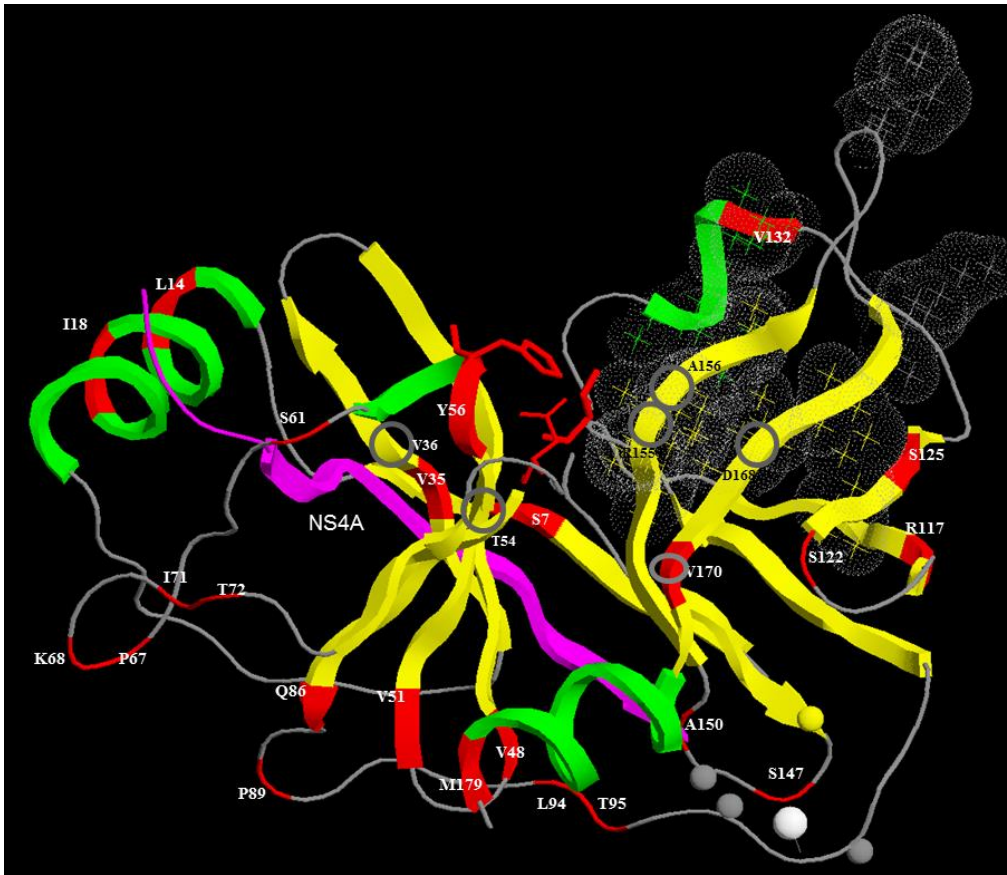


Figure 6. Structure tri-dimensionnelle du complexe NS3/NS4A. Le co-facteur NS4A est en violet. Les résidus formant la triade catalytique (H57, D81, S139) représentés en rouge ont une configuration boules et bâtonnets ; les principaux résidus polymorphes sont en rouges, les résidus impliqués dans la résistance sont entourés par des cercles gris, l'atome Zinc est représenté par une sphère blanche avec ses sites de liaisons (sphères grises et jaune), les nuages de points gris représentent les sites de liaison au substrat (S1 à S6) [8].

L'activité NS3/NS4A est essentielle à la formation des constituants matures du complexe de réplication du VHC. Le complexe NS3/NS4A serait aussi impliqué dans le contrôle de la réponse immunitaire de l'hôte, plusieurs interactions avec les voies de signalisation stimulées par l'ARN double brin ont été mises en évidence, suggérant une inhibition de la voie de production de l'interféron [67, 68]. L'activité hélicase/NTPase semblerait intervenir au moment de l'initiation de la réplication en séparant les brins positifs et négatifs de l'ARN du VHC. Plusieurs études ont rapporté un rôle potentiel de la protéine NS3 lors de la carcinogénèse [69, 70], NS3 peut ainsi induire la transformation de lignée cellulaire et se lier à des facteurs pro-apoptotique en inhibant leur activité [71]. Enfin, NS3 a été la première cible des nouvelles thérapies anti-VHC avec les deux molécules commercialisées en 2011 : le bocéprévir et le télaprévir. Elles ont un mode d'action identique,



elles inhibent de façon directe la protéase virale en se liant de façon covalente au site actif, ce sont des inhibiteurs compétitifs du substrat de l'enzyme [72]. Ces traitements utilisés en trithérapies avec l'interféron pégylé et la ribavirine ont conduit à une amélioration considérable des taux d'éradication virale chez les patients infectés par des virus de génotype 1. Au cours des études de phases 2 et 3, la réponse virologique soutenue (RVS) variait de 67 à 75% chez les patients naïfs de traitements [73-75] soulignant un gain majeur par rapport au succès thérapeutique observé précédemment avec la bithérapie (RVS de 50% en moyenne). D'autres molécules, inhibiteurs de seconde génération, sont en phase finale d'essai ou viennent d'obtenir une autorisation temporaire d'utilisation (ATU) comme le siméprévir.

La protéine NS4B (p27) est une petite protéine hydrophobe constituée de quatre domaines transmembranaires, elle est présente au niveau du réticulum endoplasmique au moment de la réplication avec les autres protéines non structurales. Elle possède aussi deux sites de palmitoylation en C-terminal permettant des ancrages supplémentaires à la membrane et facilitant l'oligomérisation de NS4B. Enfin plusieurs hélices amphiphiles ont été identifiées en N- et C-terminal au niveau cytosolique. Le principal rôle de NS4B est d'entraîner une réorganisation des membranes intracellulaires permettant la formation de nouvelles structures membranaires qui deviennent le site de réplication de la cellule infectée. Outre cette altération des membranes cellulaires, NS4B est aussi capable de moduler la production de particules infectieuses jouant ainsi un rôle dans l'assemblage viral [76]. NS4B supporte aussi une activité ATP/GTPase [77] et reconnaît l'extrémité 3' du brin ARN négatif du VHC [78] ; ces deux propriétés pourraient suggérer un rôle au cours de la réplication du génome du VHC. NS4B reste encore mal caractérisée ; cependant elle apparaît comme une cible thérapeutique émergente, de nombreux inhibiteurs anti-NS4B appartenant à différentes classes chimiques sont ainsi en cours d'essai [79].

La protéine NS5A est une phosphoprotéine existant sous deux formes, hypophosphorylée (p56) et hyperphosphorylée (p58). Cette protéine n'est pas totalement caractérisée, sa structure n'est pour l'instant que partiellement résolue. Cependant plusieurs fonctions dans le cycle viral ont été identifiées à la fois dans le processus de réplication virale mais aussi lors de la maturation des particules. Elle intervient aussi dans de nombreuses interactions avec les protéines cellulaires ce qui lui confère une implication majeure dans le contrôle de la réponse immunitaire de l'hôte, dans la résistance au traitement par l'interféron

et dans les processus d'oncogenèse. Cette protéine étant l'objet de notre travail, sa structure et ses fonctions sont plus longuement rapportées dans le chapitre suivant.

Enfin, la protéine NS5B (p68) est une ARN polymérase ARN dépendante (RdRp) indispensable à la réplication du virus. Elle assure la synthèse de l'ARN génomique en utilisant une matrice ARN ; elle s'associe aux autres protéines non structurales au sein du complexe de réplication [80]. Elle possède la structure classique en « main droite semi-ouverte » des polymérases avec la paume contenant le site actif, les doigts et le pouce qui interagissent et créent un tunnel au sein duquel passe la matrice ARN (Figure 7). Un second tunnel est formé pour le passage des NTP. La liaison à la matrice et l'initiation de la synthèse de l'ARN semblent régularisées par une boucle  $\beta$  en épingle à cheveux située dans le domaine du pouce et orientée en direction du site actif [81]. Il est important de souligner que cet enzyme est dépourvu d'activité correctrice et est à l'origine de l'importante variabilité génétique inter- et intra-génomique ( $10^{-3}$  à  $10^{-4}$  taux de substitution par site et par an), cela correspond à une mutation par nouveau génome produit. Pour terminer, il s'agit d'une cible thérapeutique prometteuse, plusieurs classes d'inhibiteurs incluant des inhibiteurs nucléos(t)iques et des inhibiteurs non nucléosidiques ont été développés et sont à différentes phases d'essai [82]. Le sofosbuvir (Sovaldi®) est le premier inhibiteur de polymérase à être commercialisé avec l'obtention de son AMM américaine en 2013 et son AMM européenne en janvier 2014.

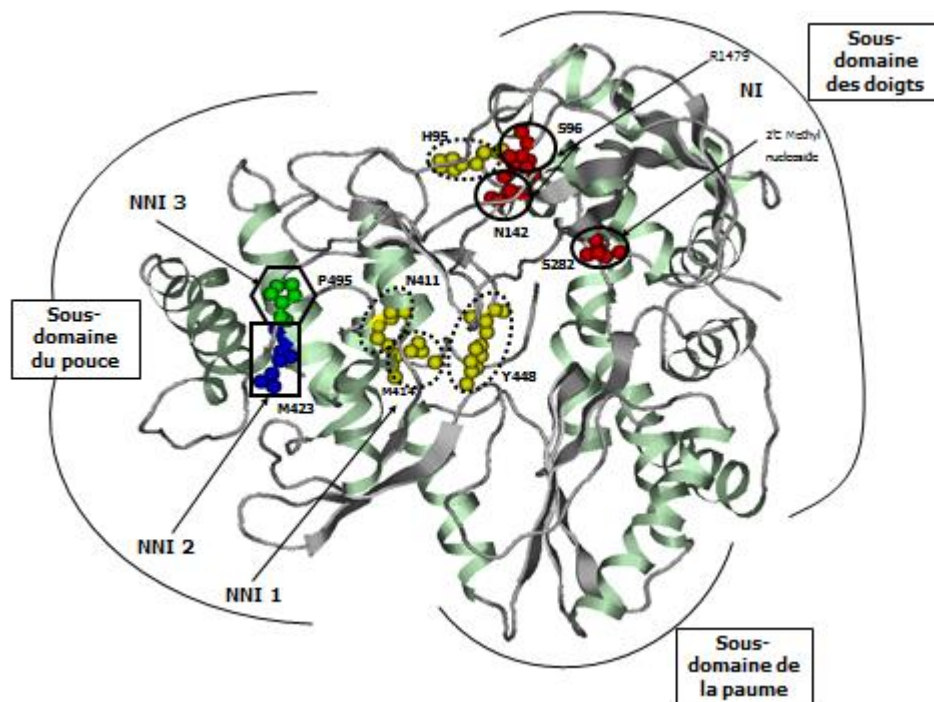


Figure 7. Structure de l'ARN polymérase du VHC [8]

### 1.3.2.3. La protéine F

Une protéine particulière a aussi été identifiée et appelée protéine F (pour Frameshift), ARFP (pour Alternative Reading Frame Protein) ou core+1 (pour indiquer sa position). En effet, la protéine F est codée à partir d'un cadre de lecture alternatif chevauchant le gène C et démarrant au nucléotide core+1. Le rôle de cette protéine reste difficile à caractériser car les modèles de culture du VHC et les modèles animaux (souris uPA-SCID) portent des virus mutés abolissant la production de la protéine F mais permettant une production virale efficace [83]. Cette protéine non conventionnelle est exprimée *in vivo* et semble stimuler la réponse immunitaire spécifique. Ainsi, la prévalence des anticorps anti-protéine F s'avère plus élevée chez les patients avec une maladie hépatique sévère ou un CHC par rapport aux patients sans CHC ou pathologie hépatique sévère. La protéine F serait ainsi présente à des concentrations plus élevées chez les patients infectés qui sont à un stade avancé de la maladie. La surproduction de protéine F semble aussi conduire à des interactions avec le cycle cellulaire qui pourrait favoriser la transformation et l'évolution vers le CHC, suggérant ainsi un rôle oncogène pour cette protéine [84, 85].

## 1.4. Cycle viral

### 1.4.1. Modèles d'étude

Le VHC n'est pas un virus facilement cultivable en culture cellulaire. Peu de lignées cellulaires sont permissives, les lignées primaires qui ont été utilisées initialement sont les hépatocytes primaires humains ou de chimpanzé, ensemble de cellules difficiles à obtenir et à cultiver [86, 87]. L'étude de la réplication a débuté avec l'utilisation d'ADNc génomique du VHC cloné à partir de sérum infectieux. Ces études ont été très difficiles et peu contributives, l'infection ne pouvait pas être entretenue dans ces systèmes.

L'analyse de la biologie cellulaire du VHC a pris un essor considérable avec la mise au point du modèle réplicon dans une lignée cellulaire continue, puis le développement de deux autres systèmes que sont les pseudoparticules HCV (HCVpp) et la culture cellulaire du clone infectieux JFH1 (HCVcc) [88].

En 1999, Lohman *et al.* ont construit le premier réplicon à partir d'une souche de VHC de génotype 1b [89]. En parallèle, Rice *et al.* ont développé la lignée cellulaire continue Huh7 issue d'hépatocarcinome humain permettant l'entretien du système réplicon [90]. Il s'agit d'ARN subgénomique de VHC au sein duquel la région codant les protéines structurales plus ou moins la région NS2, est remplacé par le gène codant la néomycine phosphotransférase, marqueur sélectif de résistance à la néomycine. Le gène Neo est placé sous le contrôle de l'IRES du VHC alors que les gènes des protéines non structurales sont sous le contrôle de l'IRES du virus de l'encéphalomyocardite murin (EMCV) (Figure 8). Ces constructions permettent en particulier des descriptions fines du mécanisme de réplication du VHC en lignée cellulaire Huh7. D'autres réplicons ont ensuite été développés et optimisés, contenant d'autres génotypes (1a, 1b et 2a) ; certaines mutations « d'adaptation » ont été identifiées comme permettant une augmentation du niveau de réplication ou étant indispensable à la réplication. Globalement, le modèle réplicon a permis la caractérisation du complexe de réplication au niveau biochimique et ultra structural, l'identification des partenaires impliqués et la localisation de ce complexe avec l'utilisation en particulier de la protéine NS5A taguée [91, 92]. Enfin l'utilisation de système réplicon-rapporteur a favorisé la recherche de nouvelles molécules antivirales. Cependant, l'inconvénient majeur du système réplicon est l'absence de production de particule virale, ce qui empêche une analyse complète du cycle viral.

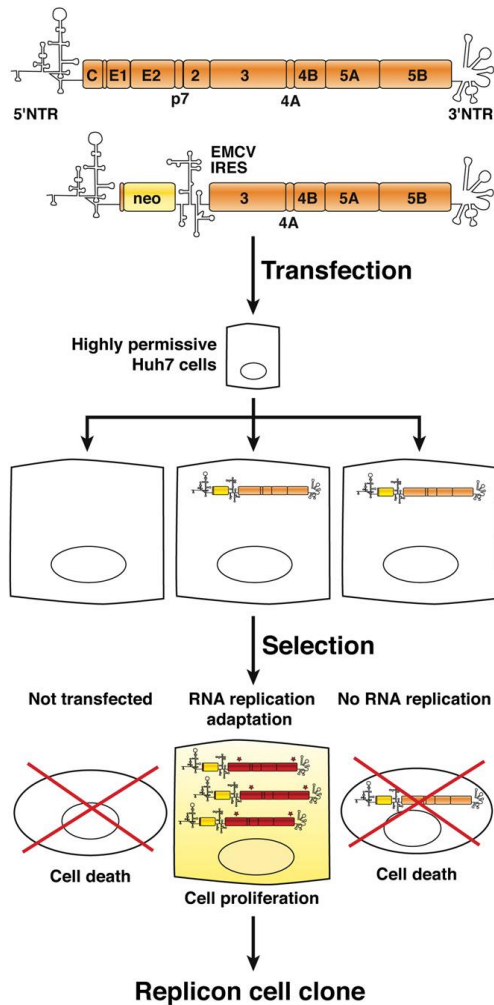


Figure 8. Schéma du modèle réplicon du VHC, d'après Lohman, 2013 [93].

Le développement des pseudoparticules a permis d'analyser un autre aspect du cycle viral : l'entrée du virus. Les pseudoparticules correspondent à l'assemblage *in vitro* des protéines d'enveloppe E1 et E2 du VHC dans des particules core du virus MLV (Murine Leukemia Virus) ou du VIH1 associées à la luciférase suite à la co-transfection de plusieurs constructions plasmidiques (Figure 9). Ces constructions, réalisées pour les six génotypes, ont permis de décrire le processus d'entrée du VHC ainsi que l'identification de nombreux partenaires cellulaires indispensables à cette étape. Cet outil se révèle précieux pour la recherche d'inhibiteurs d'entrée à visée thérapeutique.

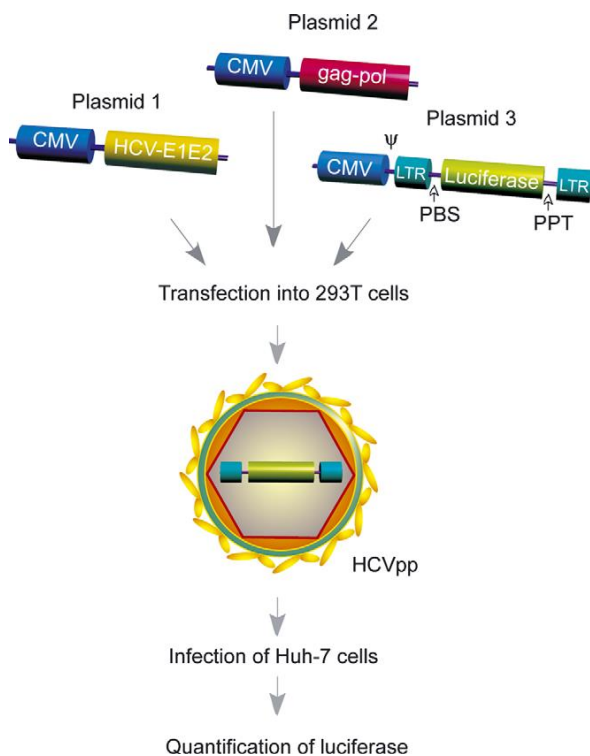
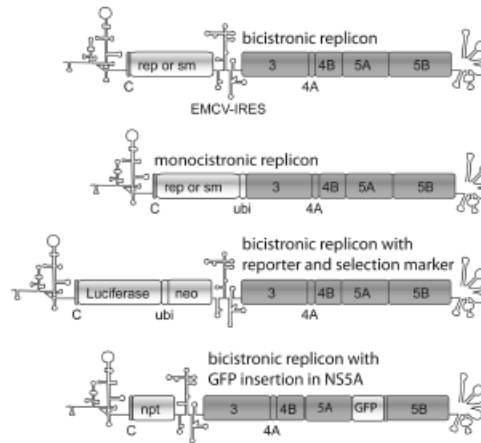


Figure 9. Schéma de production de pseudoparticules du VHC, d'après Voisset, 2004 [94].

Enfin en 2005, les travaux complémentaires et simultanés de trois équipes ont permis la mise au point d'un système complet de culture cellulaire d'un clone infectieux du VHC (Figure 10). Ainsi, l'équipe japonaise de Wakita a d'abord développé un réplicon subgénomique se répliquant en cellules Huh7 sans nécessité de mutations adaptatives [92]. Ce clone appelé JFH1 pour Japanese Fulminant Hepatitis, est issu d'un patient qui a présenté une hépatite fulminante lors de l'infection par une souche de VHC de génotype 2a ce qui est une présentation clinique rarissime. En collaborant avec le groupe de Bartenschlager, ils ont réussi à obtenir une production de particules virales infectieuses en transfectant le génome complet de JFH1 en cellules Huh7. La production était cependant limitée. Le système a été optimisé par les équipes de Chisari [95] d'une part et de Rice [96] d'autre part par l'utilisation d'une lignée cellulaire Huh7.5, lignée dérivée de Huh7 présentant une meilleure permissivité, un niveau de production virale élevé de l'ordre de  $10^4$  à  $10^6$  unités infectieuses/ml et supportant plusieurs passages sans diminution ou perte du pouvoir infectieux. Chisari *et al.* ont utilisé le clone JFH1 alors que Rice *et al.* ont développé un génome complet de génotype 2a, chimère des clones J6 (protéine C à NS2) et JFH1 (NS3 à NS5B). Depuis, d'autres clones chimériques ont été développés, constitués de la région 3' du JFH1 et de la région 5' des différents génotypes, permettant en particulier des études de sensibilités aux antiviraux en développement [97-100]. Les lignées cellulaires de type Huh7 restent les mieux adaptées, les

autres lignées cellulaires couramment utilisées ont une expression des interféron stimulés (ISGs) anti-VHC efficace limitant ou empêchant une réplication virale [101].

### A HCV replicons



### B HCVcc

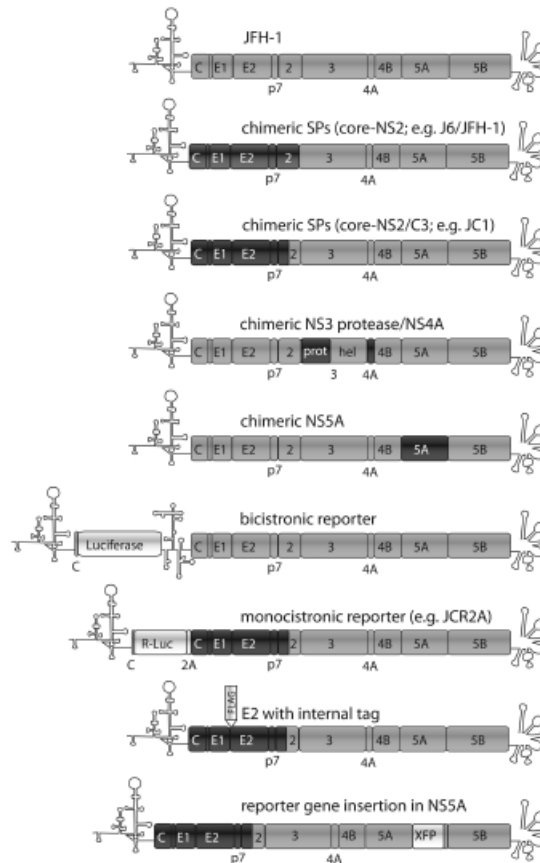


Figure 10. Modèles des constructions de culture cellulaire du VHC : (a) Réplicons mono et bi-cistronic avec gène rapporteur (luciférase) ou gène de sélection (néomycine phosphotransférase) ; (b) Modèles de HCVcc type sauvage JFH-1, type chimères, types mono- ou bicistronic de génomes rapporteurs. D'après Lohmann, 2013 [102].

Concernant les modèles animaux, les approches sont délicates [103]. Le chimpanzé est le seul animal qui peut être infecté par le VHC, il s'agit d'un modèle d'utilisation restreinte de par son coût et ses conditions d'utilisation complexes. Il est utilisé ponctuellement pour vérifier en particulier le potentiel infectieux de nouveaux modèles. Ainsi il a été démontré que le modèle HCVcc permettait l'obtention d'une infection chez le modèle primate. Le modèle chimpanzé est essentiellement réservé aux études vaccinales. Le second modèle animal est le modèle murin, il est aussi d'utilisation complexe et a été décliné en différentes approches. D'une part, des souris transgéniques ont été produites par microinjection embryonnaire de constructions génétiques permettant l'expression de différentes protéines du VHC chez la souris, ce modèle permet l'étude de la pathologie hépatique. D'autre part, les souris Trimeron ont été construites, ce sont des souris chimères qui sont constituées de trois environnements génétiques, elles sont sévèrement immunodéprimées, greffées avec un petit fragment de foie humain préalablement infecté par le VHC. Ce modèle est essentiellement utilisé pour l'évaluation des antiviraux spécifiques du VHC. Enfin, les souris chimériques uPA-SCID ont été développées, elles sont humanisées par transplantation d'hépatocytes primaires humains, et deviennent alors sensibles à l'infection par le VHC. Le phénotype uPA-SCID correspond à des souris avec des hépatocytes non fonctionnels (expression du gène de l'activateur du plasminogène de type urokinase qui entraîne un taux élevé de l'uPA dans le plasma et arrête la fonction hépatocytaire) et un système immunitaire déficient en lymphocytes T et B. Le modèle murin pourrait sembler plus accessible mais ce type de souris reste difficile à obtenir. En effet, ces souris sont immunodéprimées et donc fragiles, le succès de la xénogreffe hépatique est faible et l'obtention d'hépatocytes primaires difficile. Un fort taux de mortalité est observé au cours de la construction de ce modèle murin [104]. Le développement d'un modèle murin avec des hépatocytes permissifs au VHC est actuellement le centre des projets de recherche concernant ces modèles.

#### 1.4.2. Etapes du cycle viral

Bien qu'il ait été rapporté l'isolement de génomes viraux dans les cellules mononucléées circulantes (PBMC : peripheral blood mononuclear cell) [105] ainsi que dans le tissu cérébral de certains patients greffés [106], le VHC a un tropisme hépatique ; le foie et l'hépatocyte sont le lieu privilégié de sa répliation (Figure 11).



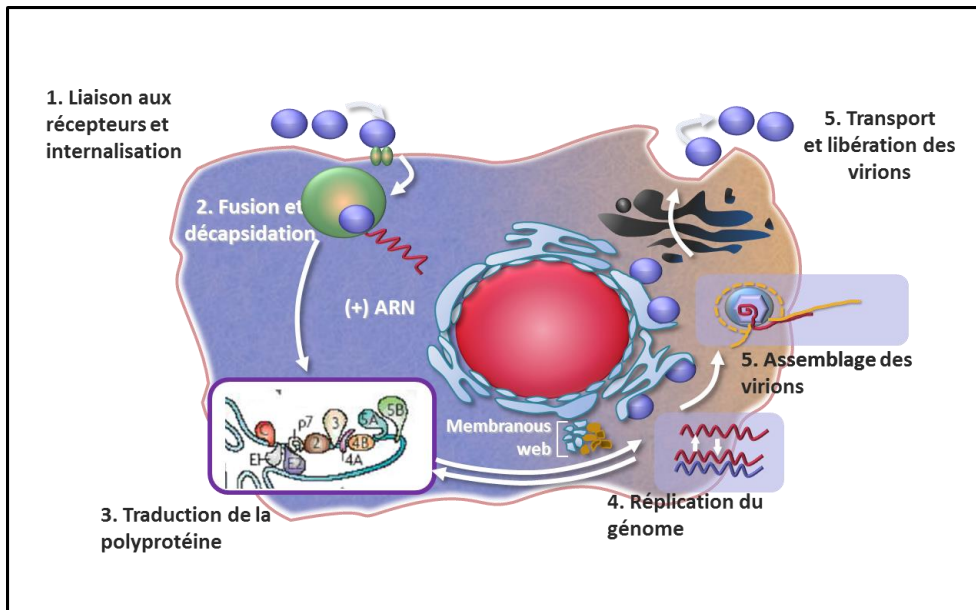


Figure 11. Schéma du cycle réplcatif du VHC adapté de Manns, 2007 [107].

L'infection débute par l'entrée du VHC dans sa cellule cible, principalement l'hépatocyte. La particule virale est constituée de la nucléocapside entourée de l'enveloppe qui se compose de la membrane cellulaire issue de l'hôte, riche en VLDL, avec notamment l'apolipoprotéine E (ApoE), et dans laquelle sont enchâssées les protéines virales E1 et E2. L'attachement à la surface de la cellule cible (Figure 12) est initiée par les VLDL et surtout l'ApoE qui se lie à des molécules d'attachement telles que les glycosaminoglycans (GAG) et les LDL-R (low density lipoprotein receptor) [108]. Ensuite, les protéines d'enveloppe interagissent avec les récepteurs de surface CD81 et SR-BI (Human Scavenger Receptor-class B type 1) avec une haute affinité, le site de liaison à CD81 a clairement été localisé dans une boucle extracellulaire de la protéine E2 [24]. Ces liaisons permettent la translocation du complexe d'attachement vers des sites riches en protéines claudin-1 (CLDN-1) et occludin (OCLN), protéines de type jonction serrées qui induisent l'endocytose au niveau de puits de clathrine. Les protéines de type claudin CLDN-6 et CLDN-9 semblent aussi participer à cette étape d'attachement/internalisation. Les lectines L-SIGN et DC-SIGN seraient impliquées dans l'attachement à d'autres cellules cibles comme les cellules dendritiques et endothéliales [109]. L'endosome formé est ensuite acidifié dans le cytoplasme, cela entraîne une fusion des membranes de l'enveloppe et de l'endosome suivie du relargage de la nucléocapside dans le cytoplasme [110, 111].

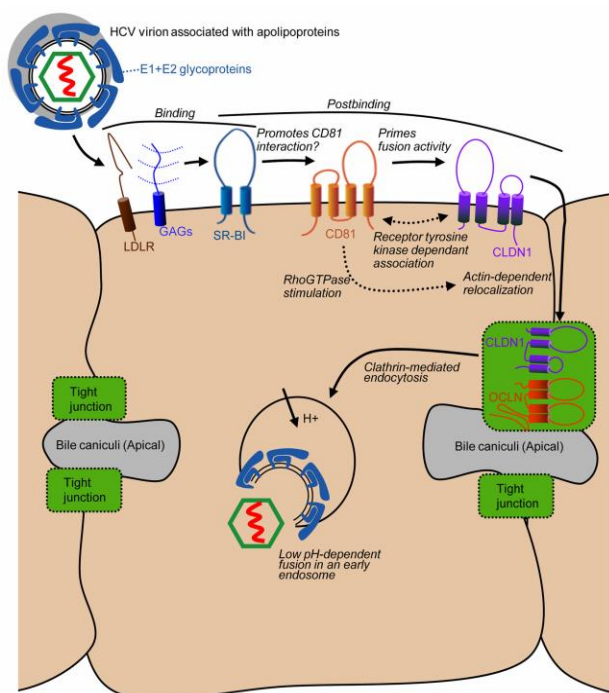


Figure 12. Schéma de l'étape d'entrée du VHC dans l'hépatocyte, d'après Ploss, 2012 [112].

Le virus est décapsidé et l'ARN viral est libéré dans le cytoplasme. Ce génome de polarité positive sert à la fois d'ARN messenger pour la synthèse protéique et de matrice pour la réplication virale. La traduction des protéines est sous le contrôle de l'IRES situé en 5'NC, la formation du complexe ribosomal 80S se réalise de façon cap-indépendante avec l'intervention de la région 3'NC [111]. Cette traduction, réalisée au niveau du réticulum endoplasmique permet la synthèse de la polyprotéine immature de 3010aa à 3033aa selon le génotype, elle est ensuite clivée par les protéases cellulaires et virales en protéines structurales et non structurales matures [89].

L'ensemble des protéines non structurales enchâssées dans les membranes du réticulum endoplasmique (RE) remaniées par l'intervention de NS4B forme le complexe de réplication viral, appelé le réseau membranaire (membranous web). L'ARN polymérase ARN dépendante synthétise un brin d'ARN de polarité négative qui sert ensuite de matrice à la synthèse de nombreux ARN de polarité positive, ils formeront les génomes des futurs virions [113, 114]. L'ARN viral nouvellement synthétisé est relargué du réseau membranaire par la protéine NS5A qui l'adresse vers les sites d'assemblage [115].

Les étapes ultimes d'assemblage des particules et de relargage sont encore mal connues car le modèle HCVcc permettant l'étude de ces phases est de mise au point récente.

Cependant plusieurs processus semblent émerger et deux phases, l'une précoce, la seconde tardive, peuvent être distinguées [116] .

Concernant la formation de la nucléocapside qui correspond à la phase précoce de l'assemblage, deux hypothèses sont proposées. La première suggère que la protéine core est d'abord transférée du RE à la surface des LD puis à nouveau recrutée par la membrane du RE au niveau des sites d'assemblage où elle interagit avec le complexe NS5A-ARN viral (Figure 13a). Les LD servent alors au transport de la protéine core des sites de traduction et de réplication de l'ARN vers les sites d'assemblage. La seconde hypothèse propose que la formation de la nucléocapside débute à la surface des LD. Dans ce cas, l'ARN viral est apporté à la protéine core par la protéine NS5A, présente également en surface des LD (Figure 13b). Des travaux récents ont confirmé ce lien étroit entre les LD et la protéine NS5A en rapportant la nécessité de recrutement par NS5A de la lipid-droplet-binding protein TIP47, indispensable à la réplication et à l'assemblage viral [117]. La phase tardive de l'assemblage correspond à l'enveloppement de la nucléocapside et l'incorporation des lipides, elle semble très fortement liée à la voie de synthèse des VLDL en « détournant » en quelque sorte cette voie à son profit. En effet, cette phase se produirait au niveau de microdomaines du RE spécialisés dans la synthèse des précurseurs des VLDL (appelés LuLD pour Luminal LD). Lors de cette synthèse, l'incorporation des apolipoprotéines apoB, apoC et apoE entraîne la formation de particules sphériques [118]. La nucléocapside serait insérée dans ces luLD avec une probable intervention des protéines virales NS2 et p7 [62, 63, 119]. Les glycoprotéines membranaires E1 et E2 enchâssées dans les membranes du RE glisseraient vers ces microdomaines pour se trouver dans l'enveloppe virale en formation [120]. Pour terminer, les particules virales bourgeonnent à partir de la lumière du RE dans le cytoplasme et sont ensuite excrétées par un mécanisme d'exocytose et circulent alors sous forme de lipoviroparticules (LVP) [121].

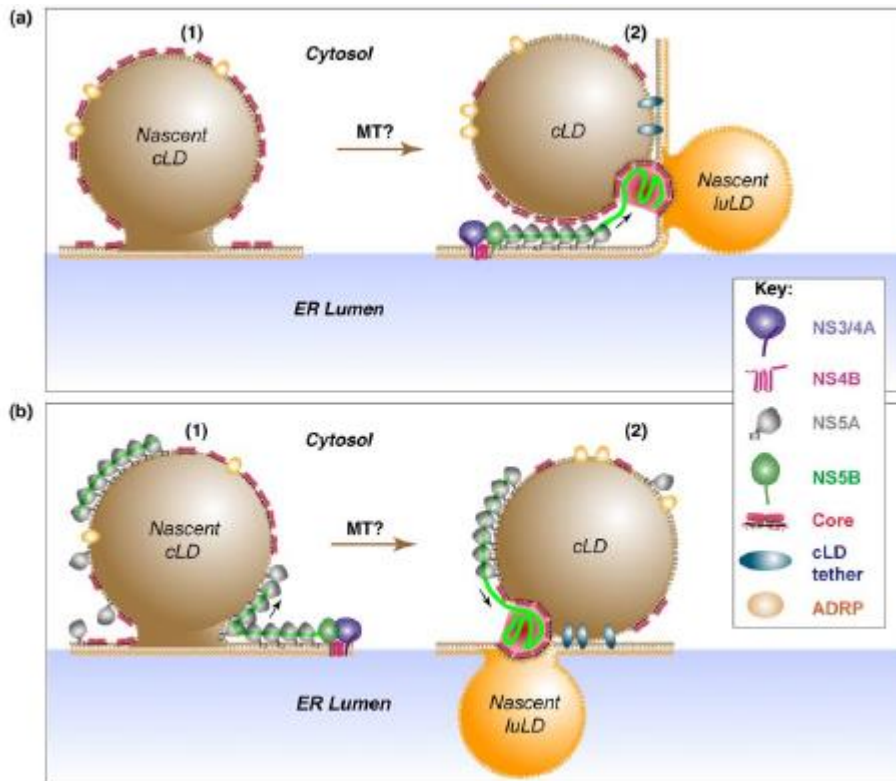


Figure 13. Modélisations de l'assemblage du VHC : (a) Initiation de l'assemblage par chargement de LD cytosoliques (cLD) avec de la protéine core et transport vers le « réseau membranaire » et le complexe de réplication ; (b) Initiation de l'assemblage sur les LD cytosoliques chargées des protéines core et NS5A, cette dernière recrutant le complexe de réplication (MT : microtubules transport). D'après Bartenschlager, 2011,[116].

## 2. Diversité génétique du VHC

### 2.1 Épidémiologie

Les dernières analyses publiées par l'OMS en 2004 estimaient que le nombre de personnes infectées par le VHC dans le monde était d'environ 120 à 130 millions de personnes [122, 123]. Depuis, les données mondiales ont été revues et les analyses combinées de séroprévalence publiées en 2013 montrent une augmentation de la séroprévalence de 2.3% à 2.8% et du nombre de patients infectés par le VHC de 122 millions à plus de 185 millions entre 1990 et 2005 (Figure 14).

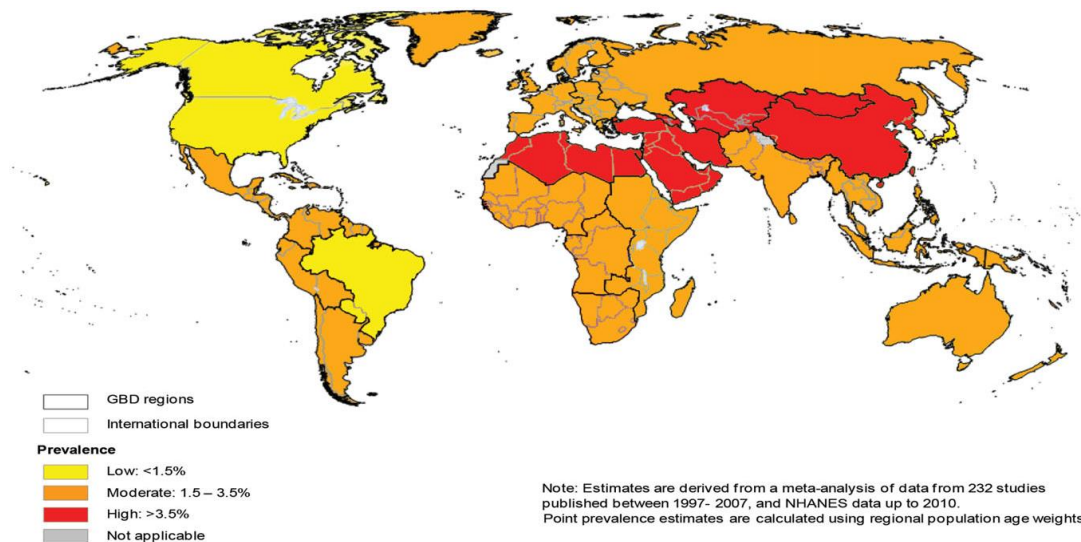


Figure 14. Carte représentant les estimations par régions GBD (Global Burden of Diseases, Injuries and Risk Factors) de la séroprévalence anti-VHC en 2005 (utilisation de données publiées jusqu'en 2007 et des données des USA jusqu'en 2010), d'après Modh Hanafiah, 2013 [1].

### 2.1.1. Modes de transmission

La voie principale de transmission du VHC est la voie parentérale en particulier par exposition percutanée large et /ou répétée à du sang contaminé. Ainsi, plusieurs facteurs de risques sont maintenant parfaitement identifiés [123]:

-La transfusion de sang ou de produits sanguins dérivés et la transplantation d'organe. Cette voie de transmission a été ubiquitaire jusqu'à la mise au point des premiers tests sérologiques en 1990 permettant la qualification des dons. La sécurisation des dons a été renforcée avec l'utilisation des tests de qualification génomique (DGV : diagnostic génomique viral). Le risque résiduel en transfusion est bien estimé pour les virus dépistés de façon systématique (VIH, VHB, VHC et HTLV), concernant le VHC le risque de transmission est très faible et a été évalué à un pour 12 500 000 dons [124]. Ce mode de transmission perdure néanmoins dans de nombreux pays où le don de sang est rémunéré et où le recours aux tests de qualification est inexistant. La transmission nosocomiale peut aussi être apparentée à la transmission par le sang précédemment décrite. L'usage de la voie injectable avec des procédures à l'hygiène défailante, dans le cadre de pratiques médicales (injections thérapeutiques avec des seringues souillées par un multi-usage, endoscopie, chirurgie

dentaire), paramédicale (acupuncture) ou non médicales (piercing, tatouage) conduirait à 2 millions de nouvelles infections par an et serait à l'origine d'environ 40% de l'ensemble des infections par le VHC dans le monde.

-La toxicomanie intraveineuse : la transmission survient par le partage d'aiguilles lors de l'injection mais aussi par l'utilisation commune du matériel d'injection (seringues et récipients de préparation de drogue, coton). L'injection de drogue par voie intraveineuse est le principal mode de transmission dans les pays développés, les études de séroprévalence montrent que la toxicomanie est le principal facteur de risque chez les patients séropositifs entre 30 et 50 ans (68% des cas des infections actives aux USA) ; de même la prévalence de l'infection à VHC au sein de la population toxicomane est très élevée et varie selon la durée de la toxicomanie (20 à 46% pour des toxicomanies inférieure à 5 ans, 64 à 94% pour des périodes de toxicomanie supérieures à 6 ans).

-Voies de transmission mineures : ces modes de transmission restent confidentiels par rapport à la contribution des voies de transmission précédemment décrites, l'efficacité de transmission est probablement faible.

Les accidents d'exposition au sang concernent quasi-exclusivement les professionnels de santé, les risques de transmission sont très faibles, estimé à 1,8% de séroconversion lors d'exposition à un patient source VHC positif lors de blessures profondes avec aiguilles contaminées. La prévalence chez les professionnels de santé est la même que dans la population générale, le risque de transmission de soignant à patient est très faible estimé à 0,5% lors d'exposition au cours de gestes chirurgicaux uniquement.

La transmission périnatale est de l'ordre de 4 à 7%, ne survenant que chez des femmes présentant une charge virale détectable lors de l'accouchement, le risque croît avec l'augmentation du niveau de charge virale. Un travail prolongé semble aussi favoriser la transmission alors que le mode d'accouchement ne semble pas influencer. L'allaitement n'est pas une voie de transmission. Enfin, il a été largement démontré que la co-infection par le VIH de la mère augmente de façon significative le risque de transmission du VHC à l'enfant [125, 126].

La transmission sexuelle est largement débattue. Plusieurs études ont été menées chez des couples hétérosexuels sérodiscordants ainsi que chez des patients hétérosexuels avec des partenaires multiples pour étayer ce mode de transmission [127, 128]. L'analyse phylogénétique des souches montraient souvent une origine commune mais restaient

insuffisantes pour démontrer la transmission, en outre il était souvent difficile d'exclure une autre voie de transmission telle que la toxicomanie en particulier. Cette transmission existerait mais serait associée à un faible risque chez les couples stables [129, 130] et à un risque plus élevé lorsque la personne source a des rapports sexuels non protégés multi-partenaires et est au stade d'infection aiguë [131]. Plus récemment et de façon ubiquitaire, de nombreux cas d'hépatite C aigue ont été recensés chez des hommes infectés par le VIH ayant des rapports sexuels avec les hommes. Les études épidémiologiques au sein de cette population comme l'étude Hepaig sont largement en faveur de ce mode de transmission émergent qui pourrait être facilité par une fragilité des muqueuses liée à l'infection par le VIH et d'autres infections sexuellement transmissibles [132, 133].

De façon très rare la transmission intrafamiliale a été suggérée et pourrait survenir par le partage d'objets de toilette tels que les rasoirs, brosses à dent ou ciseaux [134].

### 2.1.2. Prévalence de l'infection

Deux remarques doivent d'emblée accompagner les chiffres décrits dans ce paragraphe. D'une part, la majorité des données disponibles reposent sur la séroprévalence des anticorps anti-VHC et ne permet donc pas de distinguer les patients guéris et les patients avec une infection active, qu'elle soit en phase aiguë ou chronique. D'autre part, la plupart des études épidémiologiques sont menées sur des sous-groupes de populations tels que les donneurs de sang, les usagers de drogues, les personnes incarcérées, qui ne sont pas ou peu représentatifs de la population générale du pays où ils résident. Les études sur la population générale sont les plus utiles mais restent très rares car non réalisables pour beaucoup de pays [135, 136]. Plusieurs aires géographiques mondiales ont fait l'objet de revues récentes permettant une mise à jour des données épidémiologiques.

L'infection par le VHC est ubiquitaire mais deux aires géographiques peuvent être décrites : une aire de faible prévalence correspondant aux pays les plus industrialisés (Europe de l'Ouest, Amérique du Nord et Japon) ; subdivisée en une zone de très faible prévalence ( $\leq 1\%$ ) avec les pays scandinaves, le Royaume Unis, l'Allemagne, la France, l'Australie et le Canada ; et une zone de prévalence un peu plus élevée (aux alentours de 2%) comprenant par exemple les Etats-Unis, le Japon et l'Italie. L'Asie du Sud-est, l'Afrique et l'Europe de l'Est présentent des niveaux de prévalence élevés (supérieurs à 3.5%).

Ainsi en Amérique Latine, la prévalence est comprise entre 1,4% et 2,5% soit 6,8 à 8,9 millions d'adultes infectés par le VHC [137]. Le génotype 1 est majoritaire. Dans ces pays, le nombre de patients traités est faible ce qui risque de conduire à une augmentation des nouveaux cas d'une part et à la survenue des complications de l'infection à VHC, en particulier le CHC. La transmission par transfusion y apparaît comme la voie de transmission majeure, la transmission nosocomiale semble être le second mode de transmission via l'utilisation de matériel non stérilisé (aiguille, instruments pour soins dentaires, matériel pour tatouage), enfin la transmission par usage de drogue injectable semble minime.

Les données épidémiologiques en Europe, au Canada et en Israël sont issues de nombreuses données et registres, qui permettent de fournir une analyse solide [138]. Il apparaît ainsi de grandes différences entre les pays avec des prévalences variant de moins de 0,5% en Europe du Nord jusqu'à plus de 3% en Roumanie ou dans les zones rurales grecque et italienne. Il est estimé qu'entre 11,3 et 14,7 millions d'adultes ont une sérologie VHC positive dans cette zone géographique. Le facteur de risque de transmission majoritaire est l'utilisation de drogues par voie injectable et concerne près de la moitié des patients quel que soit le pays. Le génotype est aussi lié à la voie de transmission, ainsi les génotypes 1a et 3 sont associés à la toxicomanie intraveineuse alors que le génotype 1b est clairement associé à un antécédent de transfusion sanguine. La prévalence de l'infection par le VHC a augmenté depuis quelques années dans les anciens pays du bloc communiste. La cause identifiée est une explosion de la toxicomanie, en particulier chez les hommes jeunes. A l'inverse, la prévalence semble diminuer en Europe de l'Ouest. Ce déclin serait expliqué par la mise en place de mesures préventives, en particulier, une politique de dépistage massif, la mise à disposition de traitement de substitution et de kit d'injection à usage unique pour les toxicomanes, et enfin le traitement à large échelle de l'infection à VHC. Concernant certaines zones rurales avec de fortes prévalences, le risque identifié correspond à des transmissions nosocomiales. Le dernier point concernant l'épidémiologie de ces pays est une modification de la distribution des génotypes liée aux flux migratoires, l'augmentation de certains génotypes est ainsi associée à l'arrivée de migrants en provenance de pays qui entretiennent une histoire particulière avec le pays occidental (immigration des pays de l'Est vers l'Allemagne, du Pakistan et d'Inde vers la Grande Bretagne ou d'Afrique vers la Grèce).

La zone géographique comprenant l'Asie, l'Australie et l'Egypte est celle qui présente le plus grand nombre d'individus infectés avec une estimation variant de 49,3 à 64 millions de



patients anti-VHC positifs [139]. Cependant, les situations sont très variables selon les aires géographiques. Les niveaux de prévalence varient de 1 à 2 % mais montent jusqu'à 4 à 15% dans certains pays tels le Pakistan, Taiwan ou l'Égypte. Les voies de transmission sont très variables et surtout fortement liées à l'histoire des pays. En Égypte, la dissémination de l'infection a été clairement associée aux campagnes de lutte contre la schistosomiase, au Pakistan l'utilisation très importante de la voie injectable pour l'administration d'un très grand nombre de traitements a conduit aussi à cette dissémination. De même au Pakistan, l'habitude de se faire raser chez le barbier a été identifiée comme une autre voie majeure de transmission. Le recours traditionnel au tatouage en Asie est aussi un autre facteur de risque de transmission. La transmission par transfusion est la cause majeure dans beaucoup de pays, en effet, le don de sang est un acte rémunéré pour encore de nombreux pays, la qualification des dons par tests sérologiques n'existe pas, la sélection des donneurs se fait parfois sur questionnaire. Cette grande hétérogénéité épidémiologique se retrouve aussi au niveau moléculaire, le génotype 1 est majoritaire en Australie, en Chine, à Taiwan et en Asie du Nord ; le génotype 6 au Vietnam et en Asie du Sud-est et le génotype 2 est très prévalent en Corée et au Japon, en Inde et au Pakistan est principalement identifié le génotype 3, au Moyen Orient prédomine le génotype 4 (Égypte, Arabie Saoudite, Syrie) suivi du génotype 3 (Iraq, Iran). Il est important de considérer que les données de ces zones restent partielles, liées à des systèmes d'enregistrement et de surveillance encore en développement et à une densité de population très élevée.

La France compterait 500 à 600 000 individus infectés par le VHC et se situe en zone de prévalence intermédiaire. L'étude menée en 2003-2004 auprès des assurés sociaux rapportait une prévalence estimée de 0,86% (0,66-1,10), plus de la moitié des patients (53%) présentaient une hépatite C active avec un ARN détectable (Figure 15). Les facteurs indépendants associés à l'infection par le VHC étaient l'usage de drogue par voie IV, la transfusion sanguine avant 1992, le tatouage, des conditions de vie socio-économiques faibles, la naissance dans un pays présentant une forte prévalence de l'hépatite C et un âge supérieur à 29 ans [140]. Ces travaux montraient une stabilité de la prévalence par rapport à la précédente étude menée en 1994, et une amélioration du dépistage, près de 57,4% des patients séropositifs pour l'hépatite C connaissaient leur statut [141].

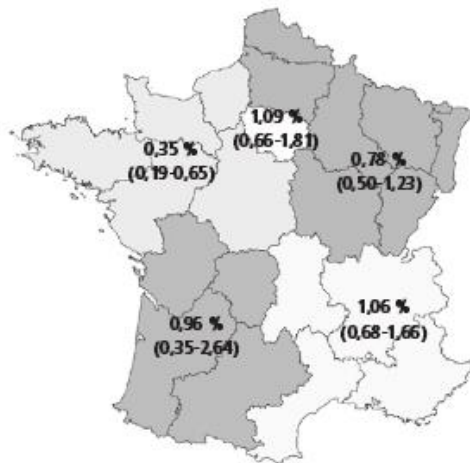


Figure 15. Estimations de la prévalence des anticorps anti-VHC par inter-région de résidence pour la population française en métropole âgée de 18 à 80 ans en 2003-2004, d'après Merre, 2005 [142].

### 2.1.3. Epidémiologie moléculaire

Concernant l'épidémiologie moléculaire, la répartition géographique des génotypes a été expliquée par de nombreuses analyses phylogénétiques. Ainsi, l'émergence du génotype 2 a probablement débuté il y a 90 à 150 ans alors que le génotype 1b est d'introduction plus récente (environ 60 à 70 ans) tout comme le génotype 3 (40 ans). Les génotypes 4 et 6 sont d'apparition beaucoup plus ancienne dans la population humaine, estimée à 350 et 700 ans respectivement [143]. Ainsi la présence d'une très grande diversité génétique dans une aire géographique limitée correspond à une implantation ancienne. Le génotype 2 est par exemple très fréquent en Afrique de l'Ouest (jusqu'à 100% des génotypes identifiés) avec une très grande diversité de sous-types, il en est de même en Afrique centrale où prédomine largement le génotype 4. A l'inverse un nombre limité de sous-types est identifié en Europe (les génotypes 1, 2 et 3 sont les plus souvent isolés) (Figure 16). Les études phylogénétiques montrent aussi une introduction du VHC dans les populations remontant à 500 à 2000 ans avec deux grandes origines géographiques : l'Afrique Central et de l'Ouest et l'Asie du Sud et du Sud-Est [22].

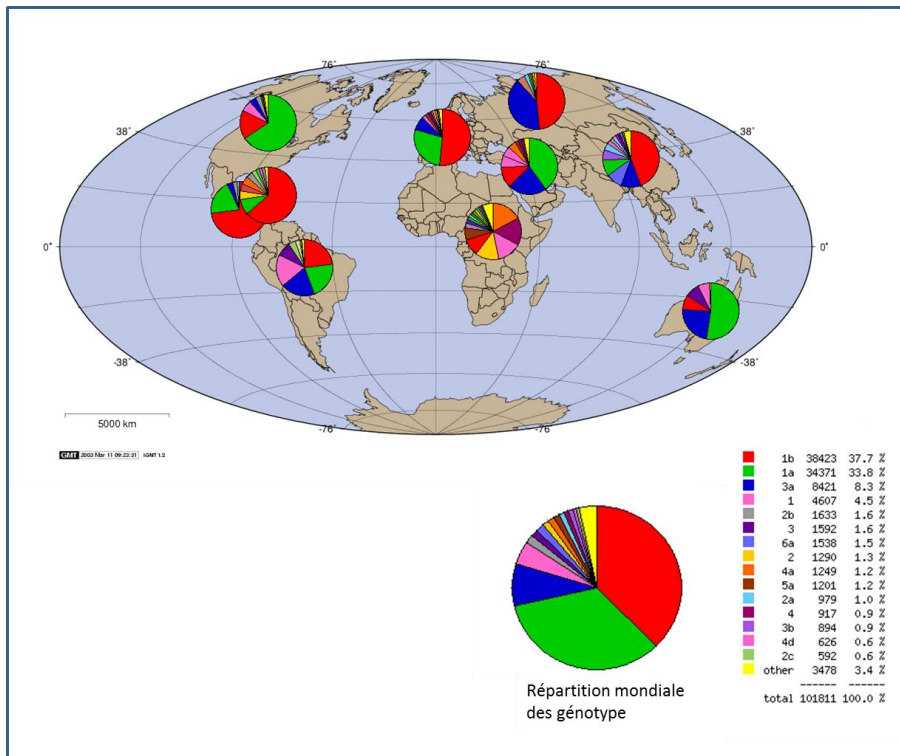


Figure 16. Distribution mondiale des génotypes et sous-types du VHC (D'après les données de la banque de séquences du VHC Los Alamos ; <http://hcv.lanl.gov/components/sequence/HCV/geo/geo.comp>).

Concernant la France, le génotype 1b prédomine représentant 40 à 50% des hépatites C avec pour origine principale la transfusion avant 1990, les génotypes 1a et 3a suivent avec 20% des cas chacun et concernent des transmissions par toxicomanie intraveineuse [144].

## 2.2. Variabilité génétique

Le VHC comme les autres virus à ARN tels que le VIH ou les enterovirus est caractérisé par une grande diversité génétique qui s'exprime au niveau inter-individuel (avec la notion de génotype et sous-type) et au niveau intra-individuel (avec l'existence de quasi-espèces). Les deux caractéristiques majeures du VHC à l'origine de cette diversité sont d'une part l'absence d'activité correctrice de l'ARN polymérase virale, elle est en effet dépourvue d'activité 3'exonucléasique et d'autre part un haut niveau de réplication avec une production virale quotidienne estimée à  $10^{12}$  virions [145, 146]. Le taux d'erreur de la polymérase est d'environ  $10^{-4}$  à  $10^{-5}$  par nucléotide incorporé ce qui correspond à une erreur en moyenne par nouveau génome synthétisé avec des taux de mutations variables selon les régions du génome [147, 148]. Les mutations surviennent au hasard et sont pour la plupart létales conduisant à un

virus défectif, les mutations non létales persistent et s'accumulent au cours des générations de virus, ces mutations peuvent moduler la capacité répliquative et peuvent fournir un avantage sélectif face aux différentes pressions de sélection susceptibles de s'exercer. Ce phénomène revêt une importance particulière dans la physiopathologie de l'infection mais a aussi un impact majeur sur le développement et la pertinence des méthodes de diagnostic et de suivi de l'infection ainsi que sur le choix thérapeutique et la possible résistance aux traitements antiviraux [8].

### 2.2.1. Génotypes et sous-types (échelle inter-individuelle)

La taxonomie du VHC a été plusieurs fois révisée et un consensus international avait été adopté en 2005 [149, 150]. Une nouvelle révision des règles de taxonomie du VHC a été publiée début 2014. Le nombre croissant de séquences disponibles et la co-existence de plusieurs bases de données avaient conduit à des discordances de nomenclature, rendant nécessaire ce travail mené par l'ICTV *Flaviviridae* Study Group. La nomenclature définit maintenant sept génotypes numérotés de 1 à 7 qui diffèrent au niveau de leur séquence nucléotidique de 31% à 33%, basée sur les analyses des régions core/E1, NS5B ou du génome complet [147]. L'ancienne terminologie « clade » est désormais abandonnée. Cette diversité s'exprime ensuite au niveau des sous-types qui ont été définis à l'intérieur de chaque phylum. Un sous-type doit présenter une divergence de sa séquence nucléotidique avec les autres sous-types d'au moins 15% sur la région core/E1 et la région NS5B, ce sous-type doit être identifié sur au moins deux isolats indépendants. La nomenclature révisée recense maintenant 67 sous-types assignés, 21 sous-types provisoires et 21 sous-types non assignés, en cours d'analyses (Figure 17). Le nombre de sous-type dans chaque phylum est très variable. Les analyses de phylo-géographie associées aux analyses d'horloge moléculaire ont permis de démontrer que les phylums comportant la plus grande diversité génétique correspondaient aux souches les plus anciennement implantées dans la population humaine, en particulier pour le génotype 2 [151]. Cette ancienneté a ainsi permis une plus longue évolution et « diversification » des souches dans la population [143]. A l'inverse, le génotype 1 implanté depuis le début du 20<sup>ème</sup> siècle seulement montre une plus faible évolution et un moindre nombre de sous-type [152]. Enfin, le troisième niveau de diversité génétique s'exprime aussi à travers la notion d'isolat, qui correspond à un groupe de souches du même sous-type partageant une homologie de séquence nucléotidique comprise entre 80 et 90%.

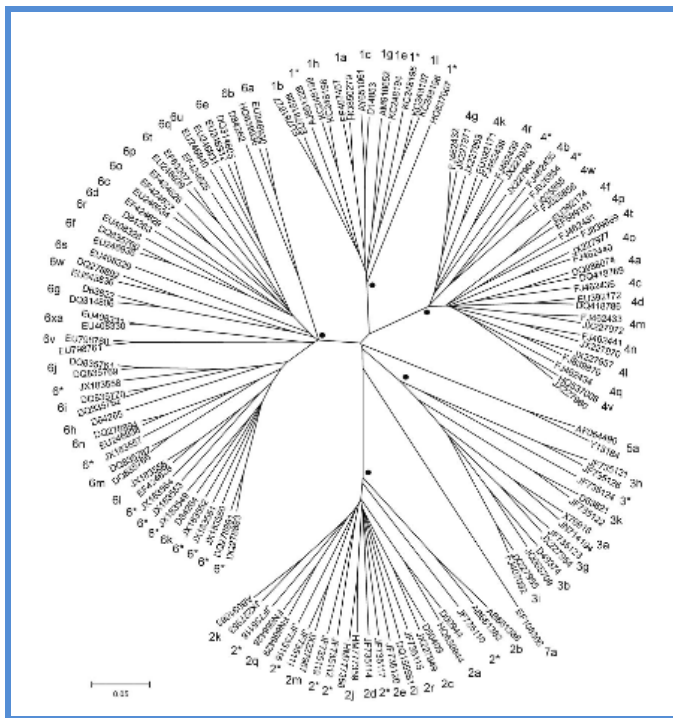


Figure 17. Arbre phylogénétique à partir de 129 séquences complètes de génome de VHC démontrant une répartition en 7 génotypes, d'après Smith, 2014, [3].

### 2.2.2. Les quasi-espèces (échelle intra-individuelle)

Le niveau ultime de l'expression de la diversité génétique du VHC se situe à l'échelle de l'individu infecté. En effet, le virus évolue à cause de l'absence d'activité 3'-5'exonucléotidase de l'ARN polymérase et du haut niveau de réplication virale. Ces deux éléments sont à l'origine d'un fort taux de mutations qui surviennent avec une fréquence plus élevée dans certaines régions telles que HVR1 dans E2 ou V3 dans NS5A [153]. Ainsi de nouveaux variants sont produits à chaque cycle de réplication, ils sont extrêmement proches génétiquement du virus « mère » mais toujours un peu différents [154]. La pression de sélection exercée par l'hôte et son environnement conduit à l'émergence de plusieurs sous-populations virales qui coexistent chez un individu (Figure 18) et qui définissent une quasispèce [155]. Le concept de quasi-espèce a été initialement défini par Eigen et Schuster [156, 157] qui ont modélisé la dynamique d'évolution des ARN réplicatifs afin d'expliquer l'évolution darwinienne et les origines de la vie. Ce concept a ensuite été repris en particulier par les biologistes travaillant sur les virus à ARN [158]. L'ensemble des variants de la quasispèce ne diffèrent entre eux que de quelques mutations et partagent au moins 90% d'identité de séquence nucléotidique. En outre, la quasispèce évolue au cours du temps dans

un processus darwinien et s'adapte aux conditions imposées par l'hôte. La sélection s'opère sur deux versants. La sélection négative correspond aux contraintes fonctionnelles du virus et à l'obligation de conserver une architecture tridimensionnelle des protéines virales compatible avec un cycle viral productif et la sélection compétitive permanente des variants possédant la meilleure capacité répliquative. La sélection positive est exercée par l'hôte et les conditions environnementales, elle correspond essentiellement à la réponse immunitaire de l'hôte et à la mise en œuvre de traitements antiviraux. Cette évolution permet une adaptation permanente à l'environnement où chaque variant participe au maintien du meilleur fitness viral. Une quasispèce est souvent composée de quelques variants majoritaires (les mieux adaptés) et de très nombreux variants minoritaires [159]. La composition pré-thérapeutique de la quasispèce a été largement étudiée dans les phénomènes de résistance aux thérapies basées sur l'interféron et pourrait être un facteur prédictif de réponse aux traitements. Comme nous l'avons mentionné précédemment, le degré de diversité génétique diffère selon les régions du génome. Ainsi, les régions 5'NC et 3'NC sont très conservées ainsi que les régions codant pour la capsidie et les enzymes NS3 et NS5B. A l'inverse, les régions codant pour les protéines d'enveloppe et plus précisément le domaine HVR1 présente une très grande diversité génétique. En effet, ce domaine est la cible des anticorps neutralisants, il est donc soumis à une forte pression de sélection. Les domaines ISDR et V3 de la région NS5A sont aussi porteurs d'une importante variabilité [160, 161]. Plusieurs travaux analysant les régions E2 et/ou NS5A ont montré qu'une forte diversité et complexité génétique pré-thérapeutique de la quasispèce étaient des facteurs prédictifs de réponse à la bithérapie anti-VHC. D'autres études rapportaient des résultats contradictoires et il semble que l'origine ethnique des patients, les clusters de souches impliqués, les posologies et durées de traitements ainsi qu'une absence de standardisation des méthodes d'analyse expliquaient ces résultats divergents [8]. Cette hétérogénéité génétique peut aussi participer à l'émergence de résistance aux nouvelles thérapies anti-protéase, anti-NS5A ou anti-polymérase à l'instar des phénomènes observés dans le traitement du VIH. Cette diversité intra-individuelle contribuerait aussi à la persistance virale et à l'évolution de l'infection, au tropisme cellulaire (notion de compartimentalisation), faciliterait la transmission (certains variants, parfois minoritaires, apparaissent très adaptés au nouvel hôte) et explique aussi l'absence d'un vaccin efficace actuellement [162, 163].

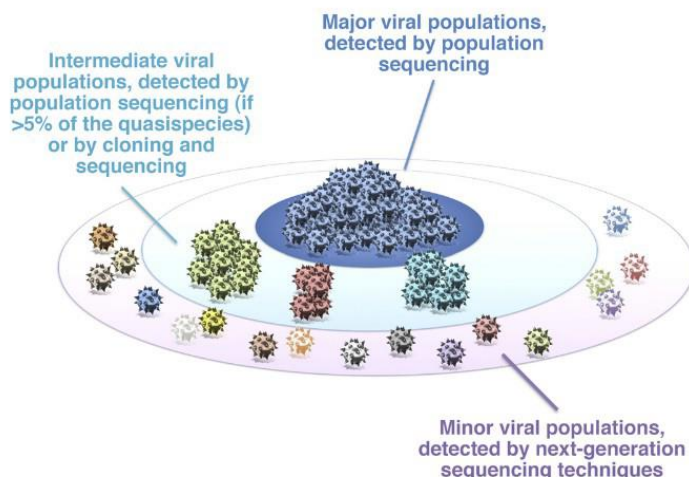


Figure 18. Distribution des quasi-espèces du VHC en populations virales majoritaires (plus de 10%), intermédiaires et minoritaires (moins de 1%), d'après Chevaliez, 2012 [164].

### 2.2.3. Les recombinaisons

La recombinaison est un mécanisme de diversification génétique et d'évolution qui semble rare chez le VHC [165]. Cependant ce phénomène existe et plusieurs souches recombinantes intergénotypes mais aussi intragénotypes ont été identifiées récemment, des recombinaisons intra sous-types ont aussi été rapportées dans deux études espagnoles [166]. La souche emblématique est la première souche recombinante intergénotypique RF1\_2k/1b décrite par Kalinina *et al.* à St Pétersbourg en 2002 [56]. Un cluster a ensuite été identifié et ce recombinant a été mis en évidence dans d'autres régions du monde mais chez des patients originaires d'Europe de l'Est, démontrant ainsi l'émergence, l'implantation et la circulation de cette souche. Le point de recombinaison des souches intergénotypiques a été identifié en NS2, à la jonction NS2/NS3 ou au début de NS3. Les recombinaisons intragénotypique et intra sous-type concernent essentiellement le génotype 1 et les points de recombinaisons ont été localisés dans différents gènes : core, E1-E2, NS3, NS5B et NS5A. Ces recombinaisons exigent qu'une même cellule, et a fortiori un même individu, soit infectée par deux souches génétiquement différentes.

Ces travaux montrent également la difficulté d'identifier les souches recombinantes. En effet, le génotypage pré-thérapeutique en pratique quotidienne est réalisé sur une seule cible (NS5B ou 5'NC), ce qui limite largement l'identification de telles souches. Il est ainsi probable que les événements de recombinaison soient aussi sous-estimés chez le VHC.

### 3. Prise en charge de la maladie à VHC

#### *3.1. Physiopathologie de l'infection par le VHC*

##### 3.1.1. L'hépatite aiguë

Après le contage, la phase d'incubation est en moyenne de 2 mois mais peut varier de 2 à 20 semaines. Il n'y a pas de critères diagnostiques précis pour définir l'hépatite C aiguë. C'est un faisceau d'indices qui amènent au diagnostic, avec par exemple la notion d'un facteur de risque dans les mois précédents, la présence de certains signes cliniques évocateurs, un bilan biologique perturbé avec notamment une augmentation du taux de transaminases de 2 à 20 fois la normale et la présence d'ARN viral du VHC chez un patient auparavant négatif. La recherche des anticorps anti-VHC n'est pas toujours contributive. En effet, la fenêtre sérologique peut être longue, la séroconversion survient entre 6 semaines et 6 mois après la contamination. Cliniquement, l'hépatite aiguë est difficile à identifier car cette phase est très souvent asymptomatique. Seulement 25 à 30% des patients vont présenter des signes cliniques parmi lesquels un ictère, un état fébrile, une asthénie, des troubles digestifs avec nausées, vomissements ou perte d'appétit, et des douleurs abdominales. Ces symptômes, s'ils sont présents, surviennent 6 à 8 semaines après le contage et persistent ensuite entre 3 et 12 semaines, ils disparaissent en général en même temps que la décroissance des transaminases et de l'ARN viral. L'hépatite aiguë peut être sévère et prolongée mais les formes fulminantes sont exceptionnelles [167].

##### 3.1.2. L'hépatite chronique

La résolution spontanée de l'infection aiguë par le VHC permettant une clairance virale survient chez 20 à 40% des patients, les chiffres sont variables selon les études car seulement 30% des infections aiguës ont une manifestation clinique, ce qui limite le suivi de l'ensemble des infections par le VHC (Figure 19). Bien qu'aucun facteur prédictif de la guérison spontanée au stade d'infection aiguë n'ait été clairement démontré, plusieurs éléments cliniques semblent émerger. Le sexe féminin, l'âge inférieur à 40 ans semblent favorables à la guérison, le second critère est aussi renforcé par l'observation d'un taux élevé de guérison chez l'enfant, de l'ordre de 80%. La présence d'un ictère pourrait aussi favoriser la clairance virale, suggérant une réponse immunitaire efficace. A l'inverse certaines maladies



infectieuses comme l'infection par le VIH ou par *Schistosoma mansoni* diminueraient les chances d'élimination du virus tout comme un état d'immunodépression causé par une greffe d'organe. Enfin, une diversité génétique élevée au sein de la quasiespèce semble péjorative alors qu'une homogénéité de la population virale paraît associée à une guérison spontanée. Le niveau de charge virale initiale et le génotype n'ont jamais été associés à une évolution particulière de l'infection aiguë [168, 169].

La persistance de l'ARN viral au-delà de 6 mois après la contamination est le critère retenu pour définir l'infection chronique par le VHC. Ainsi, près de 70 à 80% des adultes infectés n'éliminent pas le virus et développent une hépatite chronique [170]. C'est finalement à ce stade que le diagnostic est souvent fait, soit devant un bilan hépatique perturbé avec des transaminases élevées ou lors de bilans sérologiques systématiques. Peu d'études sont disponibles sur le suivi à long terme de patients infectés pour lesquels la date d'infection est connue. Cependant, les analyses menées sur des patients infectés par transfusion de sang ou injection d'immunoglobulines ont permis de caractériser l'infection chronique par le VHC. L'évolution lente et à bas bruit de la maladie est à l'origine d'une inflammation et d'une fibrose du tissu hépatique et conduit à une cirrhose dans 20% des cas après 20 ans d'évolution, puis peut évoluer vers le carcinome hépatocellulaire (CHC) qui a une incidence annuelle de 5% [171-173]. Cependant, plusieurs profils d'évolution vers la cirrhose peuvent être définis : évolution rapide avec survenue moins de 20 ans après la contamination, profil intermédiaire avec survenue 20 à 50 ans après, et profil d'évolution lente avec survenue plus de 50 après, et enfin l'existence de patients non progressifs. A partir du moment du diagnostic de cirrhose, les principales complications sont la décompensation, le CHC et le décès lié à l'insuffisance hépatique avec un risque à 5 ans de 18%, 7% et 9% respectivement [174]. Ces données globales doivent être confrontées au niveau individuel avec plusieurs facteurs qui influencent l'histoire naturelle de la maladie. Ainsi, la co-infection par les virus VIH et VHB accélère l'évolution de la maladie hépatique. La progression de la fibrose est aussi favorisée par l'âge auquel survient l'infection (en d'autres termes la durée de l'infection), le sexe masculin, l'origine ethnique (le taux de complications est ainsi plus élevé au Japon par rapport à l'Europe et l'Amérique du Nord), la consommation d'alcool. Enfin les co-morbidités que sont l'obésité, l'hémochromatose, les stéatoses non liées à une exogénose favorisent aussi une évolution péjorative de la maladie [172]. Au total, l'infection par le VHC serait à l'origine de 27% des cas de cirrhose et 25% des CHC au niveau mondial et entraînerait 8000 à 10000 décès par an [123, 175].

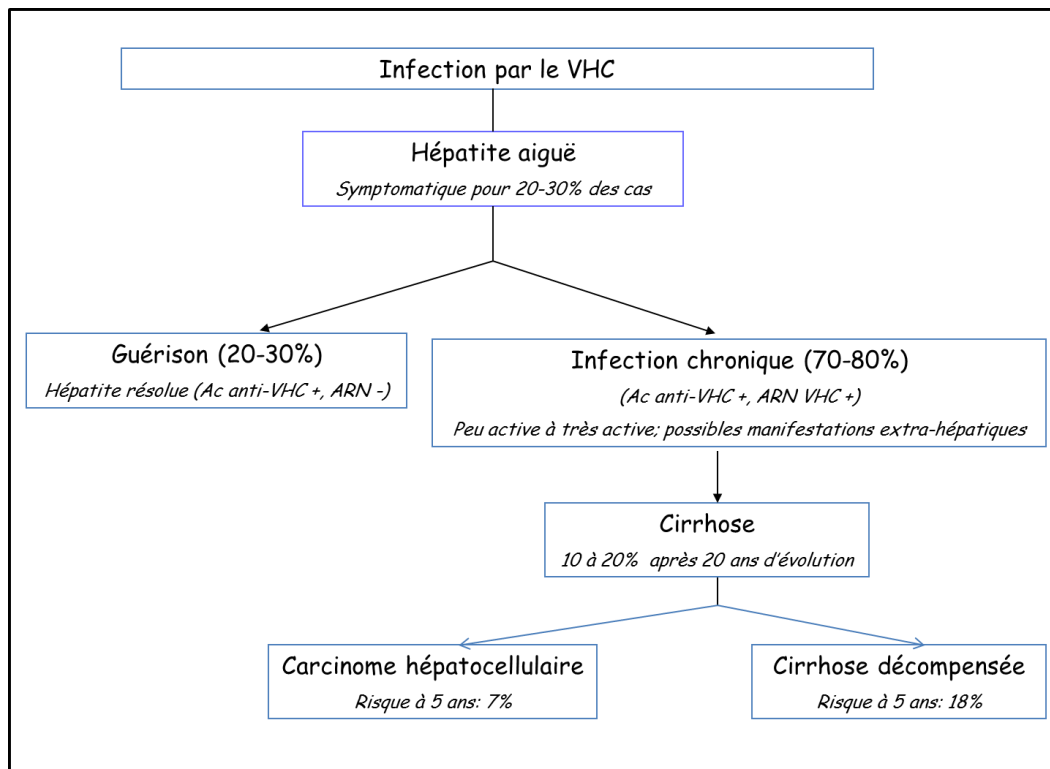


Figure 19. Histoire naturelle de l'infection par le VHC.

### 3.1.3. La réponse immunitaire au cours de l'infection par le VHC

La réponse immunitaire innée est mise en place en premier lieu lors de la primo-infection et orchestre ensuite la réponse immunitaire adaptative. Elle est essentiellement basée sur la production d'interférons de type I qui n'ont pas d'activité antiviral propre mais qui sont des médiateurs de l'expression d'effecteurs antiviraux ; ces interférons contribuent aussi à la cytotoxicité des cellules NK (Natural Killer), à la maturation des cellules dendritiques (CD) et à la différenciation des lymphocytes cytotoxiques. L'efficacité de ces réponses conditionne la clairance virale ou à l'inverse l'évolution vers l'infection chronique.

L'immunité innée anti-VHC débute au niveau des hépatocytes infectés et surtout des cellules dendritiques plasmacytoides (pDC). La reconnaissance des épitopes viraux par les récepteurs PRRPs (Pattern Recognition Receptors) active les voies de signalisation :

- des TLRs (Toll Like receptor) : les TLR 7/8 et 9 sont activés par la liaison aux ARN viraux alors que le TLR2 se lie aux protéines virales core et NS3 [176-178] (figure 20),
- de l'hélicase RIG-1 (retinoic acid-inducible gene-I), PRR intracellulaire activé par l'ARN simple brin du VHC [179, 180], (Figure 21).

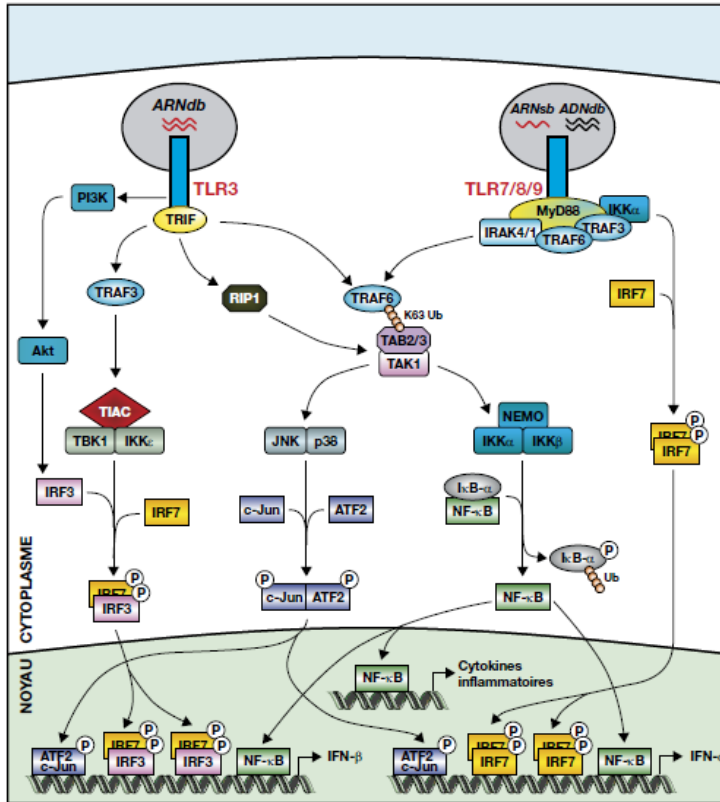


Figure 20. Voie d'induction de l'expression des *INF-I* par les TLRs, d'après Bouttier et al, 2008 [181].

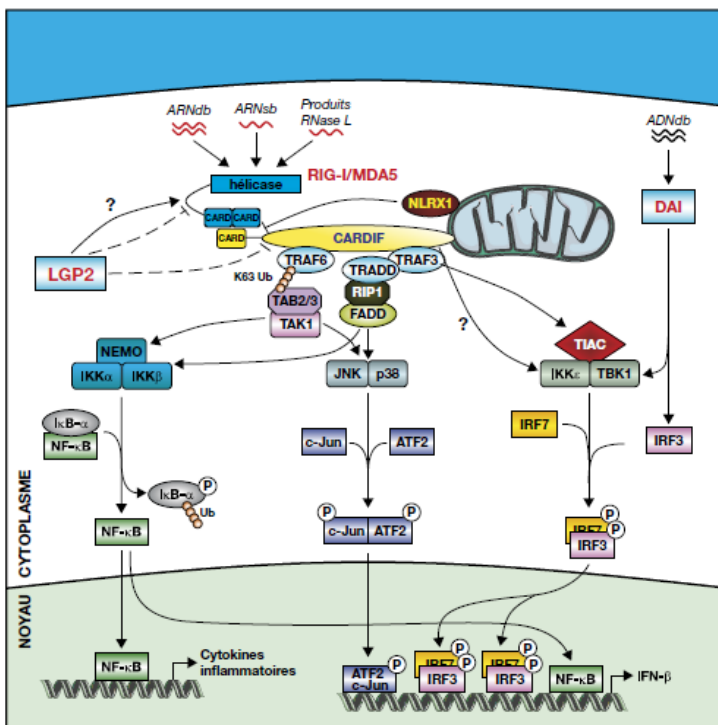


Figure 21. Voie intracytoplasmique d'induction de l'expression des *IFN-I* par l'ARN hélicases RIG-1, d'après Bouttier et al, 2008 [181].

Ces cascades de signalisation aboutissent à l'activation de facteurs de transcription, principalement le facteur NF- $\kappa$ B mais aussi les facteurs IRF3 et IRF7, et à la production de cytokines pro-inflammatoires (TNF $\alpha$ , IL-1, IL-6, interférons type I ( $\alpha/\beta$ )) [182, 183]. Les interférons produits vont ensuite activer l'expression des ISGs (interferon-stimulated genes) via la voie de signalisation cellulaire JAK/STAT, conduisant à l'expression de protéines effectrices antivirales telles que PKR, IFIT1, ISG20, ADAR1 ou OAS [184-186] (Figure 22). L'immunité innée anti-VHC est aussi médiée par les cellules NK (natural killer) ; elles ont une activité de cytotoxicité directe par exocytose de granules riches en perforine et granzymes et de cytotoxicité indirecte (induction de l'apoptose des cellules cibles par contact direct de ses protéines membranaires TRAIL, FasL ou TNF $\alpha$ ). Les cellules NK sont aussi capables de sécréter différentes cytokines à activité antivirale (Interféron $\gamma$ ) ou permettant une stimulation de la réponse immunitaire (activation des lymphocytes T et maturation des cellules dendritiques : cellules activatrices de la réponse adaptative) [187]. Il a été démontré que les protéines du VHC étaient capables d'interagir avec les acteurs de la réponse immunitaire innée et de les inhiber. La protéase NS3/4A est ainsi capable de cliver la protéine IPS-1 de la voie RIG-1 [186, 188, 189] et la protéine TRIF de la voie TLR3 [190] conduisant à une inhibition de la production d'interféron $\beta$  et de cytokines pro-inflammatoires. L'infection par le VHC conduit aussi à des perturbations de la voie JAK/STAT avec une probable implication de la protéine core, mais aussi de la protéine NS5A [191-193]. L'échappement viral au contrôle de l'immunité innée pourrait aussi être expliqué par une anomalie de répartition des différents phénotypes de cellules NK et une diminution des cellules NK à activité cytotoxique directe [194, 195]. Les protéines du VHC, notamment la protéine E2, entraîneraient également une diminution des capacités de dégranulation [196].

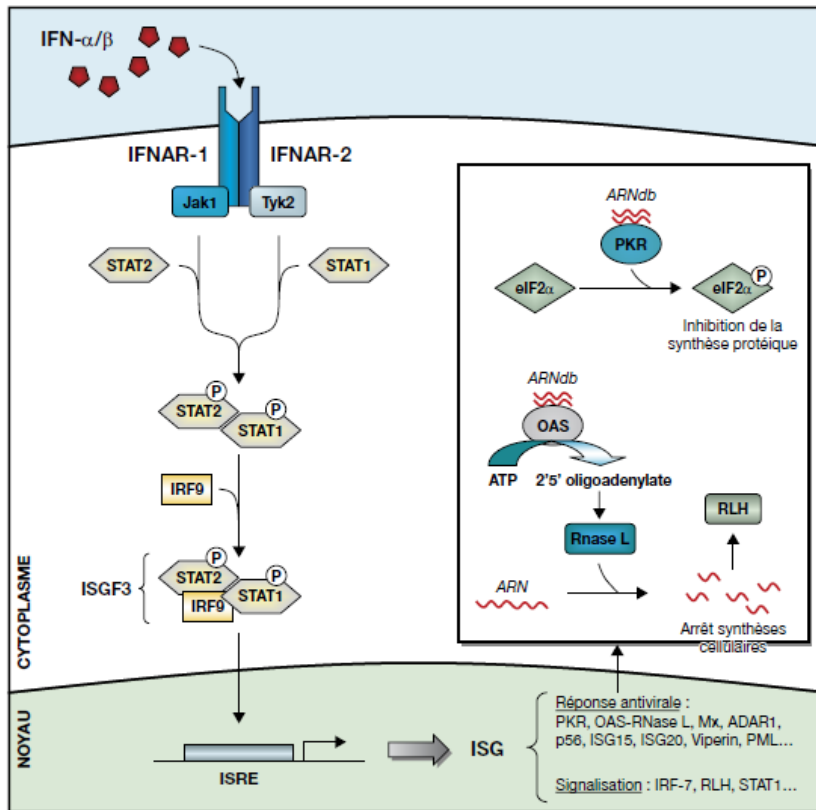


Figure 22. Voie de signalisation des interférons  $\alpha/\beta$ . D'après Bouttier et al, 2008 [181].

Concernant la réponse immunitaire adaptative, elle est médiée par les réponses cellulaire et humorale. Les lymphocytes T CD4+ sont stimulés par les cDC (cellules dendritiques conventionnelles) sécrétant de l'IL-12, et sont polarisés en lymphocytes TH1. Ils vont alors stimuler la prolifération des lymphocytes T CD8+, qui détruisent les hépatocytes infectés par le VHC par cytotoxicité directe. Ils activent aussi la prolifération et la différenciation des lymphocytes B qui donneront les plasmocytes sécrétant les anticorps anti-VHC. La réponse immunitaire adaptative à médiation cellulaire apparaît comme la réponse majeure dans le contrôle de la réplique virale et la clairance [197]. A l'inverse la réponse humorale apparaît de façon tardive après le contagé, il s'agit d'anticorps neutralisants mais qui sont bien souvent adaptés à la quasiespèce précédente et donc « en décalage » par rapport à la quasiespèce présente [26].

Ainsi, une réponse immunitaire initiale puissante semble associée à une éradication virale. A l'inverse, une réponse immunitaire de faible amplitude favoriserait la mise en place de mécanismes d'échappement par le VHC et l'évolution vers la chronicité. De probables facteurs génétiques de l'hôte sous-tendent cette variation d'efficacité de réponse immunitaire. Ainsi, certains allèles des molécules HLA-I (HLA-B57 et B27) et HLA-II (HLA-DQB1\*0301

et DBR1\*1101) seraient associés à une éradication spontanée du VHC [198]. De même, une éradication virale spontanée est plus fréquemment observée chez les patients présentant des cellules NK exprimant le récepteur inhibiteur KIR2DL3 avec HLA-C1 par rapport aux patients avec des cellules NK exprimant le récepteur KIR2DL2 [199].

#### 3.1.4. Les manifestations extra-hépatiques

L'infection chronique par le VHC est souvent accompagnée de manifestations extra-hépatiques. La persistance du virus entraîne une inflammation chronique « à bas bruit » et une stimulation permanente du système immunitaire [200]. Ainsi, l'échappement du virus au système immunitaire conduirait à l'accumulation d'immuns complexes et pourrait favoriser la survenue de maladies auto-immunes [201]. Des études sérologiques ont montré que 50 à 90% des patients présentant une cryoglobulinémie mixte (CM) de type II ou III avaient une sérologie VHC positive, d'autre part plus d'un tiers des patients infectés par le VHC avaient des immuns complexes circulants précipitant au froid, caractéristiques du syndrome cryoglobulinémique [202, 203]. Les CM sont à l'origine de vascularites, elles correspondent à la prolifération de clones de lymphocytes B produisant des IgM ayant une activité de facteur rhumatoïde (c'est-à-dire anti-IgG). Les manifestations cliniques engendrées sont diverses, triade classique des CM (purpura, arthralgie et asthénie) ou atteintes plus sévères au niveau neurologique (neuropathies périphériques) ou rénal (glomérulonéphrite membrano-proliférative). Des auto-anticorps sont aussi fréquemment trouvés (Ac anti-nucléaire, anti muscle lisse, anti-microsome foie ou rein (anti LKM1)) et expliquent les désordres immunitaires rapportés au cours de l'infection à VHC tels que le syndrome de Sjögren, les hépatites auto-immunes de type 1 ou 2 ou les thyroïdites auto-immunes. Au niveau physiopathologique, l'analyse des CM chez les patients infectés par le VHC montre une association d'IgM anti-IgG et d'anticorps IgG anti-VHC. Il semblerait que les antigènes E2 du VHC présentent une homologie structurale et antigénique avec le domaine variable des immunoglobulines humaines, et seraient ainsi reconnus par les auto-anticorps humains [204].

En outre, la survenue de lymphomes non-hodgkiniens (LNH) de type B semble aussi plus fréquente au cours de l'infection par le VHC avec des prévalences rapportées variant de 9 à 37% selon les études. Cette association pourrait être liée à une stimulation chronique des cellules B pour générer la réponse anticorps contre le VHC [205]. Il a ainsi été montré que le VHC peut infecter les lymphocytes B en utilisant le CD81 pour pénétrer dans ces cellules et stimule de façon chronique la réponse immunitaire. Cette stimulation directe via E2, et

indirecte via diverses interactions avec des cytokines et des facteurs cellulaires contrôlant le cycle cellulaire conduit à la sélection oligoclonale puis monoclonale de lymphocytes B, favorisant l'évolution maligne et la survenue de lymphomes B [206].

### *3.2. Prise en charge virologique de l'infection par le VHC*

#### 3.2.1. Circonstances de découverte

Comme nous l'avons expliqué, l'infection par le VHC est rarement symptomatique. Le diagnostic peut être porté au stade d'infection aiguë devant un tableau clinique évocateur chez un patient présentant un ictère, des nausées, une asthénie. Le plus souvent c'est le bilan biologique de première intention perturbé, avec en particulier une cytolysé d'importance variable (1,5 fois la normale jusqu'à 10 fois la normale), qui amènera à la recherche d'une infection par le VHC. Cependant, la découverte est fortuite pour une grande partie des patients lors d'un bilan systématique, suite à une consultation de médecine du travail par exemple. L'infection est aussi recherchée chez les patients décrivant un facteur de risque tel qu'une transfusion avant les années 1990 ou l'usage de drogue par voie intraveineuse. Cette recherche est aussi recommandée lors du diagnostic de l'infection par le VIH (et doit être répétée annuellement) et du diagnostic des cancers hématopoïétiques, en particulier les lymphomes. Enfin la recherche de l'infection est obligatoire et réglementaire lors d'un don de sang, d'organe, de tissus ou de cellules et lors d'un bilan d'accident d'exposition au sang (personnel professionnel victime ou patient source).

#### 3.2.2. Le diagnostic sérologique

La stratégie diagnostique recommandée par la Haute Autorité de Santé en France en 2011 repose sur la recherche des anticorps anti-VHC [207]. Ce diagnostic indirect recourt à l'utilisation de tests immuno-enzymatiques de type ELISA ultrasensibles de troisième génération qui présentent d'excellentes spécificité et sensibilité [208].

- En cas de dépistage négatif, on peut conclure à l'absence de contact avec le VHC sauf infection récente ou immunodépression sévère. Dans ces deux cas, une recherche de l'ARN du VHC doit être réalisée.

- La mise en évidence d'une sérologie positive ou douteuse doit être vérifiée par un deuxième test sur un second prélèvement de façon réglementaire [Nomenclature du J.O. du 12/08/1997]. Ce second test peut être un test ELISA (provenant d'une trousse différente que celle utilisée lors du dépistage) ou un immunoblot. Les tests de confirmation de type immunoblot / western blot permettent la mise en évidence de la spécificité anti-VHC des anticorps dépistés vis à vis des différentes protéines virales du VHC. Ces tests peuvent être utiles pour différencier une infection aiguë ou récente par le VHC et une réactivité faussement positive [209]. Un résultat positif confirmé conduit à recommander la recherche de l'ARN du VHC afin de distinguer une infection résolue d'une infection active.

### 3.2.3. Outils moléculaires de prise en charge et de suivi virologique

#### 3.2.3.1. Détection et quantification de l'ARN du VHC

La recherche d'une infection active, c'est-à-dire une réplication virale, doit être mise en œuvre devant toute sérologie VHC positive confirmée ou suspicion d'infection aiguë. Le génome est détectable entre 1 et 3 semaines après la contamination et sa persistance au-delà de six mois définit une hépatite chronique.

La recherche de l'ARN du VHC fait appel à des techniques de biologie moléculaire ultrasensibles qui permettent de déterminer la charge virale. Les techniques qui ont été successivement développées sont basées soit sur l'amplification de la cible (RT-PCR (Reverse Transcription-Polymerase Chain Reaction) et TMA (Transcription Mediated Amplification: amplification isotherme de l'ARN)), soit sur l'amplification du signal (technique des ADN branchés). Elles permettaient soit une détection soit une quantification de l'ARN viral [210]. Les techniques de dernière génération, aujourd'hui largement utilisées, font appel à la RT-PCR en temps réel et combinent les deux approches (détection et quantification). Ces techniques sont ultra-sensibles avec des seuils de détection variant de 10 à 15UI/ml, elles ont une grande linéarité (69 à 100<sup>6</sup> UI/ml), elles sont rapides et très largement automatisées [211].

Il faut insister sur le fait que le niveau de la charge virale n'est pas corrélé avec le degré d'atteinte hépatique et ne prédit pas la sévérité de la maladie hépatique. En revanche, la charge virale peut être un facteur prédictif de réponse au traitement. C'est pourquoi des mesures répétées de la charge virale sont utiles au cours du suivi des patients qui reçoivent un traitement anti-VHC. De même, la cinétique de la charge virale au cours du traitement anti-



VHC a une excellente valeur prédictive de réponse thérapeutique, cette donnée a été particulièrement bien étudiée chez les patients co-infectés par le VIH et le VHC lors de plusieurs études majeures [212-215]. La charge virale mesurée précocement, dès 4 semaines de traitement, a montré tout son intérêt chez les patients co-infectés participant à l'étude ANRS HC02 Ribavirine [216]. Cette mesure précoce est maintenant utilisée pour le suivi de toute thérapie anti-VHC.

### 3.3.2.2. Génotypage du VHC

Le bilan pré-thérapeutique comprend aussi une détermination du génotype viral. Il a été clairement démontré une forte implication de la diversité génétique dans la résistance aux traitements antiviraux [8]. La connaissance du génotype est indispensable car il permet à la fois de déterminer le schéma thérapeutique, en particulier le type de combinaison et la durée du traitement, et a aussi une valeur prédictive de succès thérapeutique. Le typage viral est réalisé par technique d'hybridation moléculaire de type Lipa ou par séquençage des régions 5'NC ou NS5B. Les deux techniques permettent une détermination du type viral, le séquençage de la région NS5B apporte une détermination du sous-type beaucoup plus fiable que l'hybridation moléculaire [217, 218]. La détermination précise du sous-type du génotype 1 est pour l'instant impérative en pré-thérapeutique.

## *3.3. Stratégies thérapeutiques*

### 3.3.1. Indications thérapeutiques

Le bilan préthérapeutique repose sur une évaluation clinico-biologique du patient. L'évaluation du stade de fibrose se place au premier plan de ce bilan ; elle est maintenant basée sur l'utilisation combinée de tests biologiques non invasifs (Fibromètre, Fibrotest) et de test d'élastométrie impulsionnelle (Fibroscan). Le recours à la ponction biopsie hépatique reste indiqué en cas de discordance des analyses précédentes. Les dernières recommandations de l'EASL (European Association for the Study of the Liver) ont été publiées en avril 2014 [219]. L'objectif du traitement anti-VHC est l'obtention de l'éradication virale permettant d'éviter la survenue d'une cirrhose, de sa décompensation le cas échéant, d'un CHC et d'un décès. Il est recommandé d'instaurer un traitement anti-VHC à partir d'une fibrose de stade F2 et de prioriser le traitement chez les patients présentant une fibrose plus sévère (F3 et F4).

Le traitement anti-viral des patients avec une cirrhose décompensée doit privilégier une combinaison thérapeutique sans interféron, voire sans ribavirine. Le traitement est par contre envisagé au cas par cas chez le patient avec une fibrose hépatique nulle à très modérée (F0-F1). Les autres indications thérapeutiques comprennent les hépatites aiguës, la tentative d'une éradication virale avant un projet de grossesse chez la femme, la co-infection par le VIH ou une tentative d'éradication virale avant une transplantation hépatique ou chez le patient transplanté hépatique présentant une récurrence de l'infection.

### 3.3.2. Arsenal thérapeutique

Les traitements reposaient encore récemment sur une bithérapie associant l'interféron alpha2 pégylé et la ribavirine. Le succès moyen était de 40 à 60%, mais variait selon le génotype viral impliqué (48% pour les génotypes 1, 4, 5 et 6 ; 88% pour les génotypes 2 et 3) [220-222]. Les échecs thérapeutiques de cette bithérapie étaient liés aux effets secondaires (asthénie, anémie sévère, troubles psychiatriques) mais aussi à l'émergence de résistance virale (non-réponse de certaines souches à l'interféron) [223]. De nouvelles stratégies thérapeutiques sont actuellement en plein essor, elles reposent sur l'utilisation d'antiviraux spécifiques du VHC (DAA : direct acting antiviral) associée ou pas à la bithérapie historique. Ces molécules représentent une véritable révolution dans la prise en charge de l'infection chronique par le VHC. En effet, des antiviraux ciblant la protéase, la polymérase, la protéine NS5A et la cellule hôte sont disponibles ou à des stades d'essais cliniques variables.

#### *3.3.2.1. La bithérapie conventionnelle*

L'association interféron pégylé  $\alpha 2a$  ou  $\alpha 2b$  et ribavirine conserve encore largement son indication en association avec les antiviraux spécifiques pour le traitement des infections chroniques par le VHC quel que soit le génotype. Cette bithérapie est aussi conservée en indication acceptable pour les génotypes non-1 par le groupe d'experts de l'EASL. Les durées de traitements sont définies selon le génotype et la réponse virologique précoce. Concernant les patients infectés par les génotypes 2 et 3, le traitement dure 24 semaines et peut être réduit à 16 semaines chez les patients présentant une charge virale indétectable à S4. Les patients infectés par les génotypes 4, 5 et 6 doivent recevoir la bithérapie pendant 48 semaines. Concernant le génotype 1, l'indication acceptable est retenue pour la bithérapie associée à un

inhibiteur de protéase de première génération. Pour terminer, la bithérapie conventionnelle reste le traitement de référence de l'hépatite C aiguë.

### 3.3.2.2. Les inhibiteurs de NS3/4A

#### 3.3.2.2.1. Les trithérapies de première génération

Les premières trithérapies combinant la bithérapie classique (interféron pégylé et ribavirine) et une anti-protéase (bocéprévir ou télaprévir) ont obtenu leur autorisation de mise sur le marché (AMM) et ont été proposées depuis 2011 aux patients infectés par le génotype 1 uniquement.

Le schéma thérapeutique standard du télaprévir comprenait une première séquence de 12 semaines associant l'interféron pégylé, la ribavirine et l'anti-protéase, suivie d'une séquence de 36 semaines de bithérapie sans télaprévir. Le schéma thérapeutique du bocéprévir était constitué d'une phase initiale (lead-in phase) de 4 semaines de bithérapie sans anti-protéase suivie d'une phase de trithérapie de 36 semaines. Les dernières recommandations américaines et européennes abandonnent quasiment ces molécules au profit des molécules sorties en 2014 [219, 224].

Bien que très encourageantes, ces combinaisons thérapeutiques avaient deux inconvénients. Les échecs thérapeutiques étaient fréquemment associés à la sélection de mutants résistants, avec une résistance croisée aux deux molécules ; la sélection est rapide et la barrière génétique est faible. Cependant la question de la résistance est très débattue. La préexistence de variants résistants minoritaires a été démontrée en particulier en recourant aux techniques d'ultra séquençage mais l'absence d'intégration du génome du VHC pose la question de l'archivage et de la ré-émergence de mutants lors d'une nouvelle exposition à un anti-viral déjà reçu et ayant montré un échec virologique. Au plan clinique, ces molécules ont montré une mauvaise tolérance, elles entraînent des effets indésirables importants, notamment des rashes cutanés et des anémies graves, qui ont pu conduire à l'arrêt du traitement et à une mortalité non négligeable notamment chez le patient cirrhotique [225].

### 3.3.2.2.2. Les anti-protéases de deuxième génération

Deux types de composés appartiennent à cette classe : les composés macrocycliques et les kétoamides linéaires. Beaucoup de ces molécules anti-NS3/4A de deuxième génération sont en essais cliniques de phase III et présentent des résultats très encourageants. Ces molécules ont un spectre d'activité plus large sur les différents génotypes et des profils de mutations associées à la résistance intéressants. Le siméprévir a obtenu son AMM aux USA et a une ATU en France en attendant son AMM européenne très prochainement. Cette molécule ne doit pas être prescrite chez un patient infecté par une souche de VHC de génotype 1a présentant la mutation Q80K (souche naturellement résistante au Siméprévir). D'autres inhibiteurs de protéase de seconde génération sont en cours d'étude de phase III ou de phase II (Tableau 1).

Dénomination Commune Internationale	Phase de développement
Siméprévir (Janssen)	ATU
Asunaprévir (BMS)	III
ABT-450/r (AbbVie)	III
Danoprévir (Roche)	II
Sovaprévir (Achillion)	II
GS-9451 (Gilead)	II
IDX320 (Idenix)	II
MK-5172 (2nde Génération Merck)	II
ACH-2684 (2nde Génération Achillion)	Ib

*Tableau 1. Liste non exhaustive des inhibiteurs de protéase NS3/4a en phase de développement.*

### 3.3.2.3. Les nouvelles classes thérapeutiques

Plusieurs cibles virales sont visées avec la même approche que celle utilisée dans le traitement du VIH (mêmes cibles et recours à des combinaisons thérapeutiques). Les objectifs de ces nouvelles lignes thérapeutiques sont d'obtenir un meilleur taux d'éradication virale, une meilleure tolérance clinique, un plus large spectre d'activité, le développement de combinaisons sans interféron avec une barrière génétique satisfaisante et des durées de traitement raccourcies.

### 3.3.2.3.1. Les inhibiteurs de la protéine NS5A

La méconnaissance du rôle de cette protéine dans la biologie virale du VHC explique l'intérêt tardif qui lui est maintenant porté en tant que cible thérapeutique. Néanmoins, plusieurs molécules anti-NS5A ont très récemment montré une activité anti-VHC très prometteuse (Tableau 2). Le daclatasvir a été la première molécule identifiée dans cette classe [226]. Le mécanisme d'action n'est pas totalement élucidé mais les premiers résultats montrent que ce sont des inhibiteurs directs de NS5A avec une liaison au domaine I aboutissant à une inhibition du cycle de réplication. Ainsi, le daclatasvir se lie au dimère de NS5A et entraîne une altération de la localisation subcellulaire de NS5A, ce qui inhibe la formation et l'activation du complexe de réplication, d'autre part, le daclatasvir bloque la synthèse de l'ARN viral, l'assemblage et la sécrétion des virions [227, 228].

Dénomination Commune Internationale	Phase de développement
Daclatasvir (BMS-790052 ; BMS)	III
Ledipasvir (GS-5885 Gilead)	III
ABT-267 (AbbVie)	III
PPI-668 (Presidio)	II
ACH-2928 (Achillion)	II
GSK2336805 (GSK)	II
BMS-824393 (BMS)	II
Samatasvir (IDX719 Idenix)	II
MK-8742 (2 <sup>nd</sup> e Génération Merck)	Ib
ACH-3102 (2 <sup>nd</sup> e Génération Achillion)	Ib
GS-5816 (2 <sup>nd</sup> e Génération Gilead)	Ib

Tableau 2. Liste non exhaustive des inhibiteurs de protéine NS5A en phase de développement.

Les molécules de première génération (daclatasvir, ledipasvir et ABT-267) sont caractérisées par une puissance anti-virale à faible dose (de l'ordre du pmol/L), une activité pan-génotypique et une faible barrière génétique. Les mutations ont été identifiées dans le domaine I (mutations primaires : M38T, Q30E/H/R, L31M/V, P32L et Y93H/N pour le génotype 1a et L31F/V, P32L et Y93H/N pour le génotype 1b) et une résistance croisée est observée entre les molécules de première génération. Les mutations associées à la résistance diffèrent légèrement selon les génotypes, la mutation en position Y93 est par contre identifiée quel que soit le génotype et confère un haut niveau de résistance y compris contre les inhibiteurs de seconde génération [229].

De fait, cette classe doit être utilisée en combinaison avec d'autres inhibiteurs spécifiques ou la bithérapie interféron pégylé-ribavirine. Cette dernière association n'est pas recommandée pour les patients infectés par le génotype 1a en l'absence de données pour l'instant. Le daclatasvir est en attente d'une AMM européenne.

#### 3.3.2.3.2. Les inhibiteurs de la polymérase NS5B

Plusieurs classes ont été développées: les inhibiteurs nucléosidiques d'une part et les inhibiteurs non nucléosidiques d'autre part (Tableau 3). Les premiers essais ont montré de bonnes efficacités avec un contrôle de la charge virale, cependant des effets secondaires graves allant jusqu'au décès ont conduit à des arrêts de développement (valopicitabine, balapiravir) et ont ralenti l'avancée des essais cliniques jusqu'à l'arrivée en 2014 du sofosbuvir avec une AMM européenne.

Les inhibiteurs nucléosidiques sont des analogues des nucléotides naturels, ils sont incorporés par la polymérase lors de la synthèse de l'ARN. Ces molécules sont des prodrogues et doivent être phosphorylées par les kinases cellulaires avant d'entrer en compétition avec les nucléotides naturels. Ils agissent comme des terminateurs de chaîne. La plupart de ces molécules ne sont pas modifiées en 3' mais au niveau du ribose en 2', ce qui crée un arrêt de synthèse par encombrement stérique [230]. La cible de ces molécules est très conservée parmi les génotypes permettant un large spectre d'activité, et une efficacité sur tous les génotypes. En outre, la barrière génétique apparaît élevée *in vitro*. Le sofosbuvir appartient à cette classe.

Les inhibiteurs non nucléosidiques (NNI) agissent de façon non compétitive en se liant à différents sites allostériques de la polymérase et inhibent les changements de conformation. Trois catégories de composés appartiennent à cette classe selon leur cible au niveau de la polymérase : le groupe des NNI1 (benzimidazoles) cible la poche située en haut du pouce ; le groupe des NNI2 (thiophènes, pyranoindoles, dihydroxypirones) se lie à la poche localisée à la base du pouce ; le groupe des NNI3 (benzothiazidines, acylpyrrolidines) cible la poche située à l'interface des domaines du pouce et des doigts de la polymérase [42]. Cette classe présente une barrière génétique basse qui conduit à une sélection rapide de souches résistantes et n'est active essentiellement que sur le génotype 1.

Dénomination Commune Internationale	Phase de développement	Classe
Sofosbuvir (Gilead)	AMM	Inhibiteur nucléosidique
Tégobuvir (Gilead)	II	Inhibiteur non nucléosidique (NNI3)
Setrobuvir (Roche)	II	Inhibiteur non nucléosidique (NNI2)
Deleobuvir (BI207127 BI)	II	Inhibiteur non nucléosidique (NNI1)
ABT-333 (AbbVie)	II	Inhibiteur non nucléosidique (NNI3)
ABT-072 (AbbVie)	II	Inhibiteur non nucléosidique (NNI3)
Lomibuvir (VX222 Vertex)	II	Inhibiteur non nucléosidique (NNI2)
GS-9669 (Gilead)	II	Inhibiteur non nucléosidique (NNI2)
BMS-791325 (BMS)	II	Inhibiteur non nucléosidique (NNI1)

*Tableau 3. Liste non exhaustive des inhibiteurs de polymérase NS5B en phase de développement ou commercialisés.*

#### 3.3.2.3.3. Stratégies d'utilisation des DAA

Les recommandations de l'EASL ont pris en compte les molécules disponibles ou qui seront approuvées par l'Agence Européenne du Médicament avant la fin de l'année 2014 [219]. En d'autres termes, seuls la bithérapie interféron pégylé-ribavirine, le sofosbuvir, le daclatasvir et le siméprévir sont pris en compte dans ces recommandations les plus récentes. Le tableau 4 résume les options thérapeutiques possibles en fonction du génotype, le niveau de recommandation est aussi rappelé.

Génotype	Combinaison	Durée	Contrôles de charge virale	Niveau de recommandation	Précautions particulières
1	IFN-Riba-Sof	12 S	J0, S4, S12; S12/S24 postT	A1	
	IFN-Riba-Sim	24 S ou 48S (12 S Tri + 12 ou 36 S Bi)	J0, S4, S12, S24/S48; S12/S24 postT	A1	Hors génotype 1a avec mutation Q80K (NS3)
	IFN-Riba-Dacl	24 S (12 S Tri + 12 S Bi ou Tri)	J0, S4, S10, S24; S12/S24 postT	B1	Uniquement génotype 1b
	Riba-Sof	24 S	J0, S2, S4, S12/24; S12/S24 postT	B2	
	Sof-Sim (+/- Riba)	12 S	J0, S2, S4, S12/24; S12/S24 postT	B1	
	Sof-Dacl (+/- Riba)	12 S ou 24 S	J0, S2, S4, S12/24; S12/S24 postT	B1	
2	Riba-Sof	12 S ou 16S ou 20S	J0, S2, S4, S12/24; S12/S24 postT	A1	
	IFN-Riba-Sof	12 S	J0, S4, S12; S12/S24 postT	B1	
3	IFN-Riba-Sof	12 S	J0, S4, S12; S12/S24 postT	A2	
	Riba-Sof	24 S	J0, S2, S4, S12/24; S12/S24 postT	A2	Hors rechuteurs avec cirrhose
	Sof-Dacl (+/- Riba)	12 S ou 24 S	J0, S2, S4, S12/24; S12/S24 postT	B1	
4	IFN-Riba-Sof	12 S	J0, S4, S12; S12/S24 postT	B1	
	IFN-Riba-Sim	24 S ou 48S (12 S Tri + 12 ou 36S Bi)	J0, S4, S12, S24/S48; S12/S24 postT	B1	
	IFN-Riba-Dacl	24 S (24S Tri ou 12 S Tri + 12S Bi)	J0, S4, S10, S24; S12/S24 postT	B1	
	Riba-Sof	24 S	J0, S2, S4, S12/24; S12/S24 postT	C2	
	Sof-Sim (+/- Riba)	12 S	J0, S2, S4, S12/24; S12/S24 postT	B2	
	Sof-Dacl (+/- Riba)	12 S ou 24 S	J0, S2, S4, S12/24; S12/S24 postT	B2	
5ou 6	IFN-Riba-Sof	12 S	J0, S4, S12; S12/S24 postT	B1	
	Riba-Sof	24 S	J0, S4, S12; S12/S24 postT	C2	

Tableau 4. Résumé des combinaisons thérapeutiques recommandées par l'EASL en fonction du génotype. IFN : interféron  $\alpha 2a/b$ , Riba : ribavirine, Sof : sofosbuvir ; Sim : siméprévir ; Dacl : daclatasvir ; S : semaine ; Tri : trithérapie ; Bi : bithérapie interféron peg-ribavirine ;



*J0 : charge virale préthérapeutique ; postT : charge virale post-traitement. L'option thérapeutique à privilégier est colorée en bleu et tient compte selon les génotypes du taux d'éradication observé dans les études, de la facilité d'utilisation de la combinaison, de la durée de traitement et du coût. Les autres options thérapeutiques comprennent en particulier des combinaisons sans interféron destinées notamment aux patients intolérants.*

#### 3.3.2.4. Les autres molécules en développement

##### 3.3.2.4.1. Les autres inhibiteurs spécifiques

Des développements de molécules ciblant NS4B et p7 sont en cours, mais à un stade moins avancé que les classes précédentes.

Trois classes d'anti-NS4B sont décrites : le clemizole hydrochloride, qui inhibe l'interaction avec l'ARN du VHC [231] ; les composés A2 et C4, décrits par Cho *et al.* [232], qui sembleraient inhiber l'interaction de NS4B avec les structures membranaires ou l'oligomérisation de NS4B elle-même; et enfin la silibinine qui perturberait l'interaction NS4B-NS3 et l'organisation des membranes cellulaires médiée par NS4B, mais dont le mécanisme d'action exact reste encore obscur [233].

Concernant la protéine p7, le BIT225, un inhibiteur dérivé amiloride, a été identifié par criblage et a montré une activité inhibitrice sur le virus de la diarrhée bovine (BVDV). L'association de cette molécule avec la combinaison interféron pégylé-ribavirine semblerait augmenter le taux de réponse virologique [234].

##### 3.3.2.4.2. Les inhibiteurs ciblant la cellule hôte (HTAs : Host Targeting Antiviral)

Cette approche originale cible la cellule hôte et les protéines qui sont stratégiques pour la réplication virale. Ces protéines étant conservées dans l'espèce humaine, cela permet de s'affranchir indirectement du problème du polymorphisme viral et de développer des molécules actives quel que soit le génotype du VHC.

La cyclophiline A (CYPA) est une peptidyl propyl isomérase qui catalyse l'isomérisation cis-trans des liaisons peptidiques au niveau de résidus proline et qui a un rôle essentiel dans le « folding » des protéines cellulaires. Cette protéine a fort niveau d'expression dans le cytosol

où elle joue un rôle crucial en permettant la réplication virale [235]. Le mécanisme reste mal compris ; cependant, il a été montré que CYP4 se lie aux domaines II et III de NS5A et participe à la liaison de l'ARN viral à la protéine NS5A. CYP4 semble aussi être nécessaire au recrutement de la polymérase NS5B dans le complexe de réplication membranaire [236]. Plusieurs inhibiteurs de CYP4 sont en développement (alisporivir, NIM811, SCY635) ; il s'agit de dérivés de la cyclosporine dépourvus d'activité immunosuppressive réalisant une inhibition de l'interaction CYP4-NS5A. L'alisporivir (anciennement Debio-025) est la molécule dont le développement est le plus avancé avec des essais cliniques de phase III débutés en 2011 et une activité antivirale sur les génotypes 1 à 4 [237].

Le microARN miR-122 est un facteur cellulaire fortement exprimé dans le foie qui stimule la réplication virale par trois mécanismes : (i) protection de la dégradation de l'ARN du VHC par la 5'-3' exoribonucléase 1, (ii) stimulation de la traduction du génome viral, (iii) augmentation de la réplication de l'ARN viral. Les sites de liaison du miR-122 sont conservés parmi les génotypes et les sous-types du VHC, rendant cette cible très attractive [238, 239]. Le miravirsén, inhibiteur de miR-122, est en phase d'étude clinique IIa sur le génotype 1, ces travaux montrent un contrôle de la charge virale VHC et l'obtention de l'indéteçtabilité, malheureusement un rebond viral a été observé chez la majorité des patients après l'arrêt du traitement ; d'autres études sont en cours [240].

D'autres cibles sont à des stades de développement très préliminaires. La stratégie basée sur l'inhibition de l'activité de phosphorylation de la PI4KIII $\alpha$ , indispensable à la réplication virale, s'est soldée par un échec dans le modèle souris [241]. L'inhibition de l'entrée du VHC est développée en ciblant le récepteur SR-B1, soit par utilisation d'un anticorps monoclonal avec une efficacité démontrée chez la souris [242], soit par utilisation de l'antagoniste ITX 5061 actuellement en évaluation de phase I [243]. Enfin, un inhibiteur de FASN (Fatty Acid synthase), le TVB-2640 est en étude pré-clinique [237]. En effet l'expression de FASN est augmentée en présence du VHC, en outre cet enzyme interagit avec le domaine N-terminal de la polymérase NS5B et augmente son expression [244], FASN apparait donc comme une cible antivirale intéressante.

La prise en charge de l'infection chronique par le VHC se trouve ainsi à un tournant majeur, avec un grand espoir pour l'ensemble des patients infectés. De très nombreuses études cliniques basées sur différentes associations thérapeutiques de DAA avec ou sans interféron

et/ou ribavirine sont actuellement présentées et publiées. Toutes rapportent des taux de succès thérapeutiques et de clairances virales exceptionnelles, proches des 100%. Les recommandations des sociétés savantes évoluent aussi très rapidement. Le frein majeur à l'utilisation de ces molécules est le coût très élevé de ces molécules dans un environnement financier de plus en plus contraint.

Dans ce contexte, l'étude du polymorphisme viral du VHC apparaît primordiale. En effet, le polymorphisme peut être à l'origine d'échec thérapeutique, qu'il soit primaire (souche naturellement résistante à une molécule) ou secondaire (émergence de souche résistante sous la pression thérapeutique).

## II. La protéine NS5A

La protéine NS5A est de taille variable selon les génotypes (entre 445 et 466 aa, 447 aa pour le génotype 1b). Il s'agit d'une phosphoprotéine qui existe sous deux formes : p56 à l'état basal et p58 à l'état hyper phosphorylé [245]. Elle est organisée en dimères ancrés dans les membranes intracellulaires par une hélice  $\alpha$ -amphiphile. Les données concernant sa structure tridimensionnelle sont encore incomplètes dans la mesure où seuls le domaine transmembranaire et le domaine I ont été cristallographiés, les domaines II et III semblent être des domaines très flexibles (Figure 23).

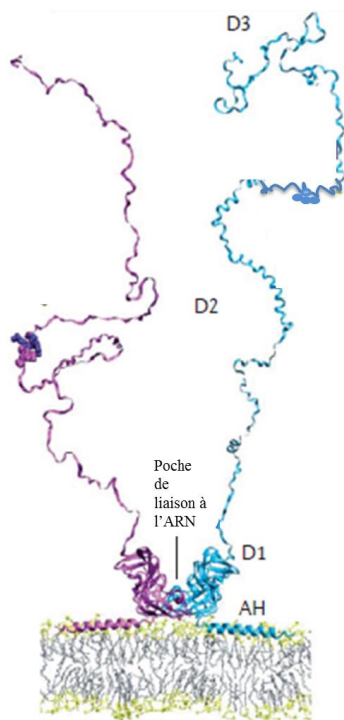


Figure 23. Représentation tridimensionnelle de la protéine NS5A organisée en dimère dans la membrane intracellulaire. AH : hélice  $\alpha$  amphiphile, D1 : domaine 1, formant la poche de liaison à l'ARN, D2 et D3 : domaines extra membranaires flexibles (unfolded domains). D'après Bartenschlager, 2013 [42].

Cependant plusieurs fonctions semblent se dégager des nombreuses études menées sur cette protéine. La protéine NS5A, bien que dépourvue d'une activité enzymatique, apparaît cependant comme une protéine pléiotropique. Elle joue ainsi des rôles clés à la fois lors de la réplication de l'ARN viral et de l'assemblage viral et dans plusieurs interactions avec les voies de signalisation de la cellule hôte.

La mise au point du système réplicon a initialement permis de montrer la présence indispensable de la protéine NS5A dans le complexe de réplication viral associée aux autres protéines virales non structurales (en particulier NS5B) [246, 247]. Les études avec les clones infectieux ont confirmé ces données et ont permis de préciser l'implication de NS5A lors du cycle viral dans sa globalité (phase de réplication et d'assemblage en particulier).

Les interactions entre la protéine NS5A et les voies de signalisation cellulaires sont aussi l'objet d'intenses investigations. Elles ont débuté avec la mise en évidence d'une résistance à l'interféron, expliquée par la capacité de la protéine NS5A à se lier et à inhiber la protéine kinase ARN double brin dépendante (PKR) cellulaire [248]. Depuis, il a été mis en évidence de très nombreuses liaisons entre la protéine NS5A et les protéines de la signalisation cellulaire. Le développement récent de molécules puissantes anti-VHC spécifique de la protéine NS5A, comme exposé précédemment, a permis de définir plus finement ses rôles dans la biologie du VHC.

## 1. Structure de la protéine NS5A et fonctions dans le cycle viral

La protéine NS5A comprend un domaine d'ancrage aux membranes cellulaires suivi de trois domaines I, II et III séparés de séquences de faible complexité (LCS : Low Complexity Sequences) (Figure 24) [249].

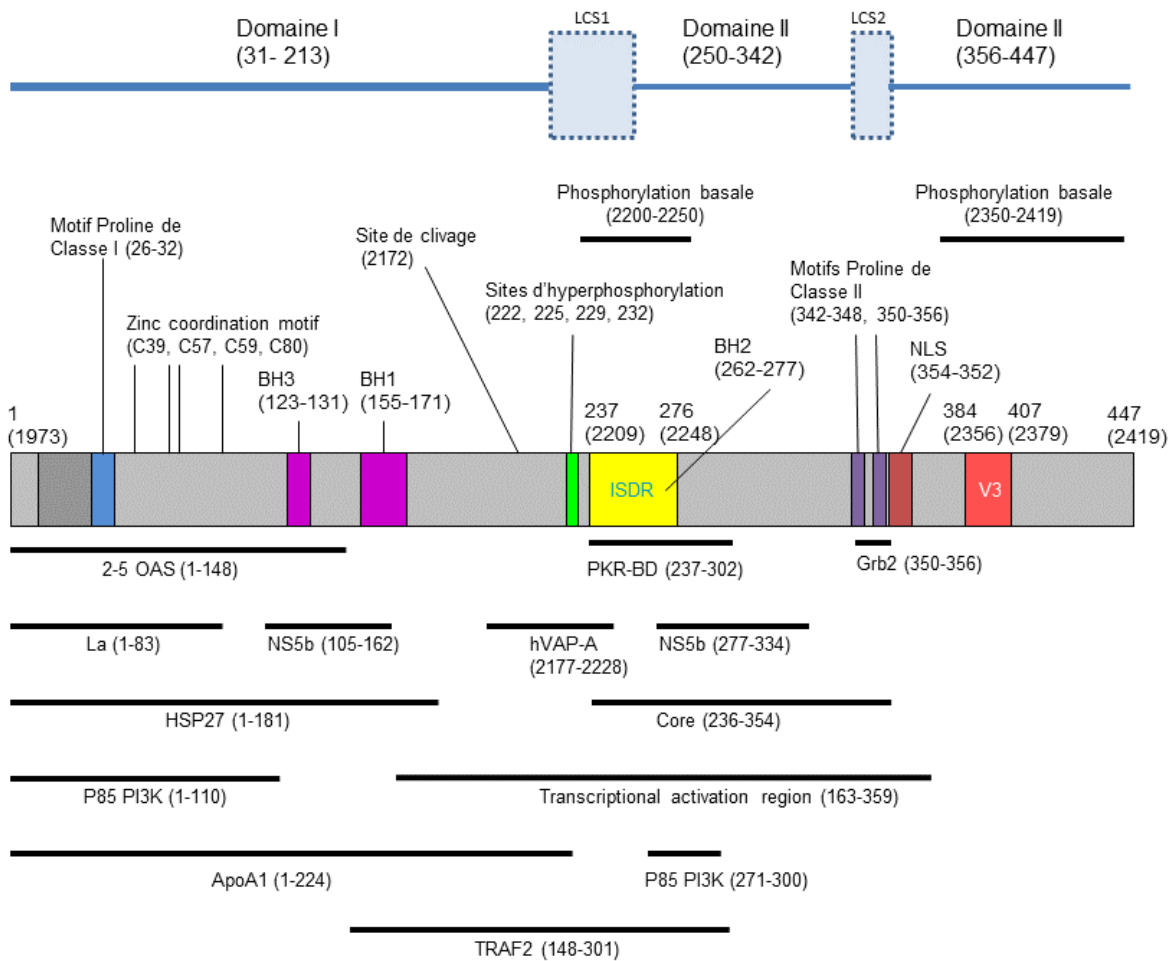


Figure 24. Représentation de l'organisation des domaines de NS5A et de ses principales caractéristiques, localisation des domaines d'interactions avec les protéines virales ou cellulaires, adapté de He, 2006 [250].

### 1.1. Le domaine d'ancrage

La protéine NS5A est une phosphoprotéine ancrée dans la bicouche lipidique des membranes intracellulaires dérivées du reticulum endoplasmique grâce à une hélice alpha amphiphile de 30 acides aminés située en N-terminal [251, 252]. Cette hélice permet la

localisation de la protéine NS5A dans les compartiments membranaires appropriés mais apparaît aussi indispensable au processus de réplication virale [252]. Cet ancrage, que seule la protéine NS5A possède parmi les protéines non structurales du VHC, est en quelque sorte le socle du complexe de réplication qui se forme au sein du réseau membranaire intracellulaire ou « membranous-web » présent dans la cellule infectée par le VHC.

## 1.2. Le domaine I

Le domaine I est constitué des 213 premiers acides aminés en N-terminal et est le domaine le plus conservé entre les génotypes du VHC. Il porte un site de liaison zinc-dépendant constitué de quatre résidus cystéine (C39, C57, C59 et C80) indispensables à la réplication virale et qui permet de définir NS5A comme une métalloprotéine [249]. Deux structures tridimensionnelles ont été résolues récemment (Figure 25). Les modèles proposés, bien que différents dans leur structure quaternaire, décrivent tous deux une organisation homodimérique du domaine I. Tellinghuisen *et al.* ont identifié un dimère en forme de pince avec un large sillon réalisant un site de liaison à l'ARN zinc-dépendant [253] tandis que Love *et al.* identifient une structure avec quatre sous-domaines dont la fonction dans la réplication ou l'assemblage viral reste à définir [254]. L'analyse biochimique comparative des ces structures par Hwang *et al.* a permis de définir NS5A comme une nouvelle classe de protéine de liaison à l'ARN en terme de structure, leurs travaux ont validé les données de cristallographie en démontrant que le domaine I et le domaine de liaison en C-terminal se lient à l'ARN viral [255].

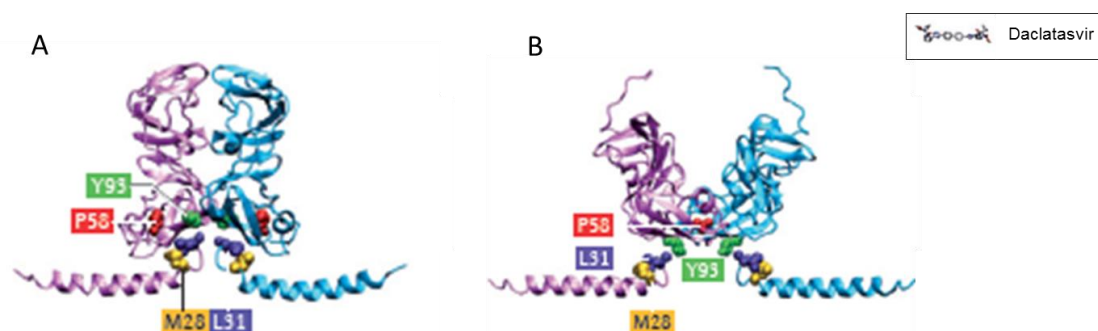


Figure 25. Représentations tridimensionnelles du domaine I de NS5A proposées par Tellinghuisen (A) [253] et par Love (B) [254]. Représentation de l'interaction avec le daclatasvir et des positions des mutations associées à la résistance à cette molécule. D'après Bartenschlager, 2013 [42].

D'autres travaux ont permis d'affiner la fonction de NS5A comme protéine de liaison à l'ARN. Foster *et al.* ont ainsi démontré que le domaine I est un site de haute affinité pour l'ARN viral mais que les domaines II et III sont aussi capables de se lier à l'ARN avec une plus faible affinité [256]. En outre, ils ont mis en évidence que la liaison à l'ARN viral se fait préférentiellement avec la région poly-uridylée en 3'NC de l'ARN, la polymérase NS5B se liant préférentiellement à la région X en 3'NC [255, 257]. Un des domaines d'interaction avec la polymérase NS5B est situé dans le domaine I. Plusieurs travaux très récents ont démontré que le domaine I de NS5A était le lieu de recrutement et d'induction de lipide-kinases, en particulier la phosphatidylinositol-4 kinase III alpha (PI4K $\alpha$ ), voie de signalisation indispensable et même critique pour la réplication virale car elle assure l'intégrité membranaire du complexe de réplication [258-260]. La PI4K $\alpha$  est majoritaire parmi les lipide-kinases du reticulum endoplasmique, elle a été identifiée comme modulant le trafic des vésicules cellulaires et elle est impliquée dans l'établissement de sites de réplication virale à partir des membranes du reticulum endoplasmique chez les enterovirus [261]. Ce recrutement de PI4K $\alpha$  permet la synthèse de phosphatidylinositol-phosphate, substrat probablement requis pour des protéines cellulaires impliquées dans le complexe de réplication. Ainsi, le domaine I de la protéine NS5A joue un rôle essentiel pour recruter et coordonner les protéines virales et cellulaires indispensables à la réplication du VHC.

### 1.3. Le domaine II

Alors que le domaine I est indispensable à la réplication virale, le domaine II (acides aminés 250-342) semble moins impératif. Les analyses par insertion ou délétion en système réplicon ont ainsi montré que la délétion de la moitié du domaine II en C-terminal et du domaine III [262] ou l'insertion de séquence hétérologue dans le domaine III telle que la gfp (green fluorescent protein) n'affecte pas la réplication virale [91]. Le domaine II est un des lieux d'interaction avec la polymérase virale NS5B [247], elle contient aussi la région déterminant la sensibilité à l'interféron (ISDR : Interferon sensitivity determining region, acides aminés 237-276), le domaine de liaison à la protéine kinase R (PKR-bd : protein Kinase R binding domain) et une région potentielle d'interaction avec bcl2. Le domaine II est aussi riche en résidus sérine identifiés comme des sites d'hyperphosphorylation (S2194, S2197, S2201, S2204). Il est important de noter que deux autres régions situées dans le domaine II (acides aminés 2200-2250) et en C-terminal du domaine III (acides aminés 2350-2419) interviennent dans la phosphorylation basale. Le domaine de liaison 2 (LCS2, situé



entre les domaines II et III) contient deux motifs polyproline, appelés PP2.1 et PP2.2, qui représente des sites de liaison aux domaines SH3 (Src homology 3) d'un grand nombre de protéines dont les src-tyrosine kinases [263]. Ces régions sembleraient impliquées dans la régulation de la réplication et de l'assemblage viral mais la démonstration reste encore vague [264].

#### 1.4. Le domaine III

Le domaine III (acides aminés 356-447) est le domaine le plus récemment étudié. En effet, son rôle restreint dans la réplication virale a de fait limité son exploration. Quelques études initiales ont identifié en C-terminal (acides aminés 354-352) une séquence potentielle de signal de localisation nucléaire (NLS : Nuclear Localization Signal) mais dont le rôle demeure incertain [250, 265]. Une étude récente a cependant démontré une localisation nucléaire du fragment C-terminal de la protéine NS5A indispensable à la réplication virale ; localisation médiée par une séquence NLS fonctionnelle associée à une activité caspase cellulaire. Ces travaux sont aussi en faveur d'une activité transactivatrice de gènes cellulaires de cette forme nucléaire de NS5A associée à une réplication virale efficace [266]. L'utilisation du modèle JFH-1 a permis une analyse complète du cycle viral et non plus de la réplication seule. Cette étape majeure a mis en lumière le rôle du domaine III crucial lors de l'assemblage viral [44, 267]. Il est maintenant établi que le domaine III est la région d'interaction avec la protéine core, cette interaction survient à la surface des gouttelettes lipidiques (LD : Lipid droplets) intracellulaires, il semblerait que le domaine I de NS5A assure la liaison aux LDs [44, 268]. Plusieurs résidus sérine et thréonine en C-terminal ont été initialement identifiés en système JFH1, sites phosphoaccepteurs probablement phosphorylés par des caséine kinases cellulaire de type II (CKII), les études ont rapporté leur rôle fondamental dans l'interaction core-NS5A et la formation de virions : le résidu S457 accompagné du motif SEED apparaissait comme favorable à la phosphorylation par les CKII [262] alors qu'une seconde équipe identifiaient un cluster (résidus S452, S454 et S457) [115]. Par la construction de clones chimériques, il a été récemment démontré que la production de virions est contrôlée de manière similaire en clone JFH-1 (génotype 2a) et H77S (génotype 1a). Ce travail a pu réconcilier les travaux précédents. Ainsi, Kim *et al.* confirment le rôle essentiel des trois positions conservés entre les génotypes dans le processus de production de virions [269]. Enfin le domaine III contient le domaine hypervariable V3, objet de notre travail. Ce domaine V3, localisé entre les positions 2356 à 2379 de la polyprotéine (soit 24

résidus entre les positions 384 et 407 sur NS5A), a été initialement décrit par Inshauspe *et al.* [270]. Plusieurs études ont ensuite mis en évidence une relation entre le polymorphisme de cette région au sein de la quasiespèce virale et la réponse au traitement par interféron [271, 272]. D'autres travaux ont ensuite montré des résultats similaires chez des patients traités par bithérapie combinant l'interféron et la ribavirine rapportant une diversité et une complexité de la quasiespèce statistiquement plus faible chez les patients répondeurs par rapport aux patients non répondeurs, en particulier sur le domaine V3 [273-276]. D'autres études n'identifient pas un domaine en particulier prédictif de la réponse au traitement, mais dans tous les travaux, le domaine V3 apparaît comme une région concentrant un fort polymorphisme [277, 278]. Une récente étude conduite au sein de notre laboratoire a montré que l'analyse du domaine V3 semblait un meilleur facteur prédictif de réponse au traitement par l'interféron chez les patients infectés par le génotype 1b, la région ISDR apparaissant peu contributive [10]. Le domaine V3 n'a pour l'instant pas été caractérisé au plan fonctionnel.

Pour terminer, les analyses structurales par spectroscopie à résonance magnétique nucléaire (RMN) du domaine II et du domaine III ont permis de montrer un état naturel peu structuré avec un haut degré de désordre et de flexibilité [279, 280]. Cette désorganisation apparente représente un avantage car elle permet à la protéine NS5A de s'adapter en changeant de conformation et en conséquence de se lier à de très nombreux partenaires cellulaires ou viraux.

Ainsi, la protéine NS5A est indispensable à la réplication virale et à l'assemblage, étapes sous le contrôle de phosphorylations. La protéine NS5A est trouvée à l'état basal de phosphorylation dans le réseau membranaire et représenterait la forme requise pour la réplication. A l'inverse l'état d'hyperphosphorylation, essentiellement médié par les caséines kinases I- $\alpha$ , nécessite la présence des protéines virales NS3, NS4A et NS4B. L'interaction de NS5A avec hVAP-33, protéine impliquée dans le trafic cellulaire, est inhibée par l'hyperphosphorylation et entraîne un arrêt de la formation du complexe de réplication et l'inhibition de la réplication virale [246, 281]. L'état de phosphorylation de NS5A conditionne aussi sa liaison à la polymérase virale NS5B, ce qui serait un mécanisme de contrôle indirect de la réplication virale. Enfin, il a été suggéré que la liaison de la polymérase virale NS5B à la matrice virale ARN serait modulée par la phosphorylation de NS5A, inhibant ainsi la réplication virale de façon directe. Grâce à la capacité de changement de conformation des domaines II et III et la présence de nombreux sites de phosphorylation, la

protéine NS5A interagit avec de nombreuses protéines cellulaires impliquées dans les voies de signalisation du système immunitaire, du contrôle du cycle cellulaire et de l'apoptose.

## 2. Protéine NS5A et système immunitaire de l'hôte

La réponse de l'hôte à l'infection par le VHC met en jeu l'immunité innée et l'immunité adaptative, elles contribuent à la clairance virale dans certains cas mais participent aussi à la persistance virale et à l'altération du tissu hépatique [282, 283]. En effet, les réponses immunitaires de l'hôte se trouvent perturbées par des interactions avec différentes protéines du VHC.

### 2.1. Interactions avec la réponse immunitaire innée

De façon générale, la réponse immunitaire innée met en jeu la production d'interférons qui sont des cytokines naturellement produites ; ces interférons stimulent en particulier la réponse immunitaire antivirale en activant plusieurs voies de signalisation cellulaire et induisent la synthèse de protéines à propriété antivirale. Concernant l'infection par le VHC, les interférons de type I (IFN $\beta$  et IFN $\alpha$ ), de type II (IFN $\gamma$ ) et de type III (IFN $\lambda$ ) participent à la réponse antivirale en utilisant des voies de signalisation communes et différentes [283]. Les études *in vitro* et chez la souris « humanisée » rapportent une altération par le VHC de la réponse anti-virale innée de façon globale [284, 285] ou impliquant une protéine en particulier comme NS5A [286]. Au cours de la phase d'incubation, il a été démontré une diminution de la production d'interféron  $\beta$  par l'interaction de la protéine NS3/4A avec les voies RIG-TLR3 avec pour conséquence une diminution de l'induction de la voie Jak/Stat et une baisse de l'expression des gènes ISGs (Interféron Stimulated Genes) [287, 288]; la protéine core interagit aussi avec la voie Jak/Stat en inhibant et dégradant StatI. Ces protéines NS3/4A et core sont aussi à l'origine d'une inhibition de la production d'IFN $\alpha$  par les cellules dendritiques circulantes (pDC) et d'une apoptose des pDC [289, 290]. Concernant la protéine NS5A, elle agit préférentiellement de façon directe en ciblant les protéines synthétisées par les ISGs, elle interagit moins sur la synthèse des interférons ou les voies de signalisation induite par ces cytokines.

D'une part, NS5A se lie à la PKR activée par les ARN double brin en formant des hétérodimères, la PKR intervient dans la régulation de la traduction cellulaire et notamment la phosphorylation du facteur d'initiation de la traduction eIF2 $\alpha$ . Des travaux récents semblent

montrer que la régulation de la PKR impliquerait l'IRES du VHC et le domaine III de NS5A, la concentration en protéine NS5A intracellulaire régulerait l'activation de la PKR participant ainsi au contrôle de la réplication virale [291]. L'interaction de la région NS5A avec la PKR met en jeu le domaine ISDR (interferon sensitivity determining region) [248, 292], qui couvre partiellement le domaine de liaison à la PKR. Des études contradictoires présentent le domaine ISDR comme une région pouvant être à l'origine de la résistance du VHC au traitement par l'interféron. Les premières études d'épidémiologie moléculaire rapportaient les isolats viraux avec une région ISDR conservée comme résistants à l'interféron et ceux avec des mutations dans la région ISDR comme sensibles [293, 294]. A l'heure actuelle, la question reste encore posée. Dans tous les cas, la liaison de la protéine NS5A à la PKR a pour conséquence la phosphorylation et l'inhibition du facteur eIF-2 $\alpha$ , et entraîne l'absence d'expression de nombreux ISGs. Les réponses interféron antivirales sont alors diminuées et facilitent la réplication virale [295].

D'autre part, la protéine NS5A peut aussi se lier et inhiber la 2'-5' oligoadénylate synthétase, enzyme dont la synthèse est induite par les interférons et qui est un effecteur de la réponse antivirale médiée par la RNaseL [193]. En outre, la protéine NS5A agit aussi positivement sur le promoteur du gène codant l'interleukine 8, cytokine capable d'atténuer les propriétés antivirales de l'interféron en inhibant de manière globale l'expression des ISGs ; ce qui favorise le développement de l'infection virale [296-298]. Enfin, il est démontré que la protéine NS5A est capable de se lier à plusieurs molécules des voies de transduction de l'interféron telles que la protéine STAT-1 [299] ou la sous-unité p85 de la PI3K [300].

## *2.2. Interaction avec la réponse immunitaire adaptative*

Les voies TH1 et TH2 de l'immunité adaptative sont aussi altérées au cours de l'infection à VHC. D'une part, il apparaît clairement que la réponse T CD4+, qui orchestre l'ensemble des réponses adaptatives, alliée à une production d'IL2 et d'IFN $\gamma$  et une réponse CD8+ efficace est corrélée à une clairance virale efficace du VHC et une guérison alors qu'une réponse T CD4+ absente est constatée chez les patients développant une infection chronique [197]. D'autre part, la réponse humorale montre une apparition tardive des anticorps neutralisants qui sont incapables d'éliminer le virus mais qui exercent cependant une pression de sélection, cette pression contribue largement à l'évolution de la quasiespèce, notamment sur la diversité génétique des protéines E1 et E2. Cette stimulation virale des lymphocytes B ferait le lit des manifestations extra-hépatiques de types cryoglobulines et

l'éventuelle évolution vers le lymphome non hodgkinien de type B. Enfin, la réplication virale à haut niveau constitue une activation permanente des cellules T VHC-spécifiques qui présentent alors des anomalies fonctionnelles ; il est ainsi mis en évidence une diminution puis une perte de la production d'IL-2, de l'activité cytotoxique, de TNF $\alpha$  puis d'IFN $\gamma$  [282]. L'infection virale semble aussi induire la production de cellules *Tregs* (sous-population de lymphocytes T CD4+) qui favoriseraient le passage à la chronicité [200]. Cependant l'implication précise de la protéine NS5A dans la modulation de la réponse adaptative reste mal connue.

### 3. Protéine NS5A et oncogénèse

Comme la protéine core et la protéine NS3, la protéine NS5A participe très probablement à l'oncogénèse en interagissant avec plusieurs voies de signalisation cellulaire qui régulent le cycle cellulaire de l'hépatocyte, l'apoptose et la réponse au stress oxydatif [250].

#### *3.1. Interactions avec la régulation du cycle cellulaire*

Trois grandes cascades de signalisation régulant la croissance cellulaire et médiées par les MAPK (mitogen-activated protein kinases) sont actuellement identifiées : la voie ERK (extracellular signal regulated kinase), la voie SAP/JNK (stress activated protein kinase / c-Jun N-terminal kinase) et la voie p38 MAPK [301]. Il a été montré que la protéine NS5A modulait ces trois voies notamment par l'interaction du motif polyproline C-terminal avec les domaines SH3 de la protéine cellulaire Grb2 (growth factor receptor-bound protein 2), protéine pivot dans le contrôle de la croissance cellulaire [300]. La séquestration de Grb2 par la protéine NS5A conduit à une baisse de l'activation des voies MAPK [302-304].

Concernant le cycle cellulaire, il est normalement régulé par l'activation séquentielle de complexes cyclines et cycline-dépendantes kinases (CDK), permettant le passage d'une phase du cycle à la phase suivante, chaque passage étant contrôlé par les voies p53 et Rb. En phase G1/S, l'activation par p53 du facteur de transcription p21 entraîne sa liaison à CDK2 et son inhibition. Cela permet l'arrêt temporaire du cycle laissant le temps de réparer l'ADN endommagé. Il a été montré que NS5A se lie à p53 et diminue ainsi le contrôle du cycle par p21, favorisant la prolifération. NS5A a aussi démontré sa capacité à diminuer l'expression de

la protéine ASPM (mitotic spindle protein M) qui conduit à des mitoses aberrantes et une instabilité chromosomique [305], l'hyperexpression de la protéine NS5A en culture cellulaire conduit aussi à une dérégulation des mitoses associée à une instabilité chromosomique et des aneuploidies [306]. Une autre étude a montré la présence d'une proportion d'hépatocytes en phase G1 significativement plus élevée dans des biopsies de foie de patients infectés par le VHC comparées à celles de sujets atteints d'hépatopathies d'origine alcoolique [307].

Une liaison de la protéine NS5A à la sous-unité p85 de la phosphatidylinositol 3-kinase (PI3K) a aussi été démontrée et augmente l'activation de la voie PI3K en stimulant l'activité catalytique de la sous unité p110 de l'enzyme. Cela conduit à la stabilisation du proto-oncogène  $\beta$ -caténine, à une stimulation de son activité de facteur de transcription et en conséquence à l'expression des gènes favorisant la survie cellulaire [308-310]. Il a été démontré récemment que la protéine NS5A activait aussi directement la protéine  $\beta$ -caténine indépendamment de la voie PI3K [311]. Pour terminer, la protéine NS5A est capable d'activer le facteur de transcription STAT 3 *in vitro* et dans les hépatocytes de souris transgéniques, STAT3 est stimulé par diverses cytokines et facteurs de croissance cellulaire et est habituellement impliqué dans le contrôle du cycle cellulaire et de l'apoptose [312, 313]. Ainsi, la protéine NS5A perturbe le cycle cellulaire et favoriserait la prolifération cellulaire.

### 3.2. Interactions avec les mécanismes cellulaires d'apoptose et de stress oxydatif

Outre la perturbation du cycle cellulaire qui facilite l'établissement d'une infection chronique, la mise en place de mécanismes d'échappement au phénomène de mort cellulaire programmée ou apoptose par la protéine NS5A est également suggérée. Cela favorise aussi la persistance virale. Il est d'emblée important de souligner que les patients infectés par le VHC présentent des niveaux plus élevés en protéines TRAIL (TNF-related apoptosis\_inducing ligand), TNF $\alpha$ , FAS et FASL qui sont les protéines contrôlant les différentes voies de l'apoptose [314].

La protéine NS5A est capable de se lier aux protéines TRADD et TRAF2 recrutées lors de la voie extrinsèque de l'apoptose médiée par le TNF $\alpha$  ; ces liaisons inhibent les caspases cellulaires à l'origine du déclenchement de l'apoptose, à l'inverse la protéine NS5A ne semble pas interagir et moduler la voie Fas [315]. La protéine NS5A est susceptible d'interagir avec les protéines de la famille Bcl-2 qui règlent la balance pro- et anti-apoptotique : NS5A se lie en particulier à Bad et Bax empêchant leur fonction pro-

apoptotique de s'exercer [316]. Plusieurs travaux ont aussi montré que la protéine NS5A inhibait l'apoptose médiée par p53. Ainsi, la protéine NS5A peut se lier et séquestrer la protéine anti-tumorale p53 dans le cytoplasme inhibant sa translocation nucléaire et l'induction de l'apoptose [317]. Elle forme aussi des complexes avec p53 et TBP (TATA binding-protein) empêchant la liaison à leur séquence d'ADN cible, suggérant un impact sur la régulation de l'activité transcriptionnelle cellulaire. Plus récemment, NS5A a été identifiée comme un puissant inhibiteur de la voie pro-apoptotique médiée par les LPS (Lipopolysaccharides) des bactéries à Gram négatif en diminuant l'expression des récepteurs de LPS de type TLR4 (Toll-like receptor 4) [318]. Enfin, Akkari *et al.* ont récemment montré que la protéine NS5A interagissait avec le TGF- $\beta$  et la protéine Ras avec pour conséquence une transformation des hépatocytes primaires vers des cellules tumorales [319].

L'infection par le VHC induit un stress oxydatif en partie en cause dans la genèse de la maladie hépatique. Plusieurs interactions des protéines virales avec les voies calciques et oxygène-dépendantes contrôlant le stress oxydatif ont été mises en évidence, la protéine core semblant la plus impliquée. Cependant quelques études rapportent aussi une intervention de NS5A dans ces processus. Etant établi que TRAF2 participe à la modulation de la réponse au stress du reticulum endoplasmique, l'interaction de NS5A avec TRAF2 précédemment décrite pourrait probablement affecter cette voie. Plusieurs études ont rapporté que la protéine NS5A semble aussi augmenter la production de ROS (reactive oxygen signalling) qui entraîne ensuite l'activation de la voie JNK et stimule la néoglucogenèse [320, 321]. Ainsi, les phosphorylations des protéines IRS1 (Insulin Receptor Substrat 1) se trouvent augmentées et entraînent une perturbation de la voie Akt/proteine kinase B, en particulier des facteurs de transcription FoxO1 et FoxO2. Or, il a été suggéré une association entre l'insulinorésistance et le risque de progression de la fibrose et la survenue du CHC chez les patients infectés par le VHC [320, 322]. En outre, cette activation de la production de ROS s'accompagne d'une altération de l'homéostasie calcique cellulaire et conduit à l'activation de plusieurs facteurs de transcription comme STAT3 et NF- $\kappa$ B [323]. Ce dernier facteur est primordial à la biologie cellulaire car il intervient dans la régulation de l'inflammation, l'immunité innée, la cancérogenèse et l'apoptose. L'activation de NF- $\kappa$ B par la protéine NS5A met, entre autre, en jeu des mécanismes calcium-dépendants [324].

Ces différentes liaisons aux protéines cellulaires convergent ensemble vers un rôle majeur de la protéine NS5A dans l'évolution de l'infection par le VHC vers une altération cellulaire et la transformation tumorale.

### III. Le polymorphisme génétique viral : origine et évolution

#### 1. Bases biologiques du polymorphisme viral: mutation, indels, recombinaison et duplication

La variabilité génétique du VHC comme nous l'avons précédemment développée s'exprime à différentes échelles, intra- et interindividuelles. Cette variabilité est la conséquence de la forte production de virus avec des génomes toujours porteurs de mutations ponctuelles nouvelles apparaissant au hasard, la sélection et l'implantation de chaque génome étant sélectionnées par l'interaction avec l'environnement et l'hôte.

Ainsi, le mode d'évolution principal des virus à ARN simple brin de polarité positive repose sur la mutation soit par transition (remplacement d'une base purique par une autre base purique ou d'une base pyrimidique par une autre base pyrimidique), soit par transversion (remplacement d'une base purique par une base pyrimidique ou l'inverse). Ces mutations peuvent être silencieuses (ou synonymes), elles n'entraînent pas de modification de la protéine, ou elles peuvent être non silencieuses (ou non-synonymes), conduisant à une modification de l'acide aminé codé et éventuellement des propriétés de la protéine qui sera traduite. Cette mutation peut aussi entraîner un codon stop et conduire à un génome défectif. Outre la modification ponctuelle de l'acide nucléique, la mutation peut aussi consister en des phénomènes d'insertion ou de délétion d'un ou plusieurs nucléotides (appelés indels). Lorsque l'indels est un multiple de 3 nucléotides, le cadre de lecture est conservé [325].

La recombinaison est un second mécanisme moléculaire d'évolution qui a été largement décrit chez les virus à ARN [326, 327], avec des fréquences variables selon les familles virales ; les taux de recombinaison sont élevés chez les retrovirus dont le VIH ( $10^{-4}$  à  $10^{-5}$ /site/génération) [328], les enterovirus [329] ou les *Corononaviridae* [330] ; alors que cette modalité d'évolution est rare chez les *Flaviviridae* dont le taux de recombinaison est estimé à  $10^{-8}$ /site/génération [331]. Ainsi, la recombinaison des génomes est bien identifiée chez le VIH1 avec l'existence de plus d'une soixantaine de formes circulantes recombinantes (appelées CRF) [328]. En revanche, la revue réalisée par Morel *et al.* [166] rapportent l'identification de seulement quelques recombinants du VHC intergénomiques dont le premier décrit par Kalinina *et al.* en 2002 [56], recombinant 2k/1b, mais aussi des recombinants intragénomiques et intra sous-types. Ces recombinants sont probablement



sous-estimés du fait du séquençage partiel et non du séquençage complet du génome en pratique clinique.

Le principal mécanisme régissant la recombinaison chez les virus à ARN est appelé le saut de matrice. L'ARN polymérase ARN dépendante associée à l'ARN matrice fait une pause au cours de la synthèse du nouveau génome et change de matrice sans se dissocier du brin en cours de synthèse puis reprend cette synthèse. La présence de structure en épingle à cheveux autour des points de recombinaison des souches de VHC semble aussi favoriser cette théorie. Un second mécanisme appelé non répliatif a été évoqué mais rarement décrit chez les virus ARN tels que les poliovirus ou le BVDV [166].

La recombinaison nécessite la co-infection de l'individu puis de la cellule cible par deux souches virales génétiquement distinctes. Selon le type de virus (Figure 26), elle consiste en une recombinaison simple (a) ou elle nécessite une production de virus « hétérozygote » suivie d'une recombinaison après infection de la cellule cible pour les retrovirus (b). Une alternative particulière à la recombinaison est le réassortiment qui consiste en un échange de segment chez les virus à ARN segmentés (c), la recombinaison intra-segment a néanmoins été démontrée chez ce type de virus.

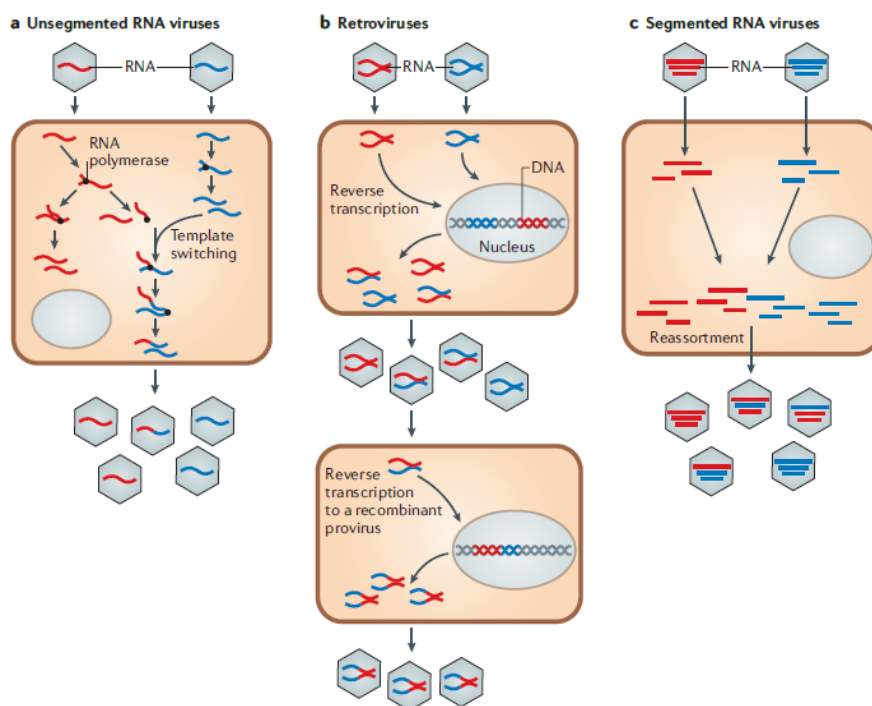


Figure 26. Représentation des modalités de recombinaison génétique selon les espèces virales. D'après Simon-Lorière et Holmes, 2011, [326].

Le principal mécanisme permettant ces recombinaisons repose sur un saut de polymérase autrement appelé « copy choice model » (Figure 27). Au cours de la synthèse du nouveau brin génomique, le complexe polymérase-brin néo-synthétisé peut se dissocier du brin matrice. La ré-association à une matrice de même nature à la même position ou à une position différente dans le même gène est qualifiée de recombinaison homologe, alors que la ré-association à une matrice de nature différente ou à des régions génomiques différentes conduit à des recombinaisons non-homologues. Enfin, la dissociation peut être abortive. Les changements de matrice semblent être favorisés par des taux de similarité de séquence élevés entre le brin « donneur » et le brin « accepteur », cela explique que les recombinaisons homologues sont les plus fréquentes. A l'inverse les recombinaisons non homologues sont issues d'échange entre régions génomiques distinctes ou entre ARN hétérologues. Ce type de recombinaison produit plus fréquemment des génomes délétères et est donc moins souvent observé.

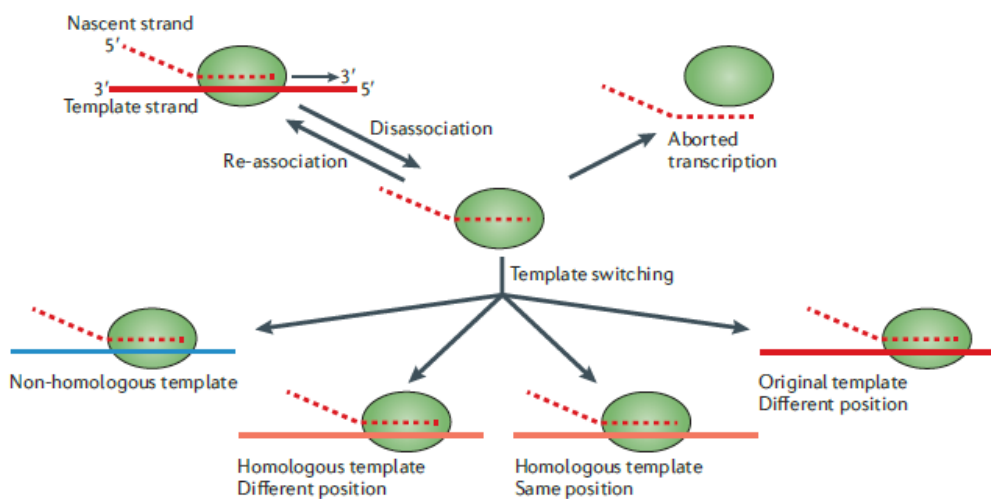


Figure 27. Schéma des différentes recombinaisons possible à partir du complexe ARN polymérase-brin d'ARN matrice-brin d'ARN néo-synthétisé. D'après Simon-Lorière et Holmes, 2011, [326].

Un second mécanisme par recombinaison non réplivative a aussi été décrit au cours de quelques travaux menés chez les poliovirus [332] et le bovine viral diarrhea virus (BDVD) [333]. Il consiste en un clivage suivi d'une ligation de l'ARN recombinant à la nouvelle matrice. Il s'agit alors d'un transfert latéral de gène, ce mécanisme nécessiterait la présence de structure secondaire de l'ARN favorable mais se dispense de l'intervention de la polymérase virale [334].

Enfin une dernière alternative d'évolution est la duplication de gène. Ce phénomène est largement décrit chez les vertébrés mais aussi chez les virus à ADN notamment. Il s'agit d'une recombinaison particulière, supportée soit par une double recombinaison soit par un transfert latéral de gène. Les duplications de gènes sont rares chez les virus à ARN, comme l'ont décrit récemment Simon-Lorrière et Holmes, qui ne relevaient que 9 événements de duplications de gènes parmi l'ensemble des virus à ARN dont les génomes complets étaient disponibles dans les bases de données (soit 1198 virus) [335]. De même les événements de duplications partielles restent très rares et essentiellement identifiés dans les régions non codantes ou inter géniques [336, 337]. L'évolution par duplication chez les virus à ARN conduit à une augmentation de la taille du génome avec pour conséquences une probabilité accrue d'apparition de mutations délétères et une baisse de la fitness virale. En outre, un génome de plus grande taille allonge le temps de répllication et ne favorise probablement pas la dissémination. L'hypothèse mécanistique majeure repose sur le saut de polymérase et une recombinaison avec une réassociation du brin au niveau de région avec des fortes homologies de séquences, hypothèse confortée par l'identification de recombinaisons homologues majoritairement. Il s'agit alors de la coexistence de deux gènes issus du même ancêtre, ces copies sont alors présentes sur la même branche d'un arbre lors d'une reconstruction phylogénétique.

Plusieurs travaux supportent aussi la possibilité que la duplication soit issue d'une recombinaison avec du matériel génétique d'un autre organisme, proche cependant, reposant sur un transfert latéral de gène plutôt qu'une duplication vraie. Cela est compatible avec l'observation que les similarités des séquences sont parfois plus élevées entre un des gènes dupliqués et le gène de l'organisme donneur qu'entre les deux copies de gènes. Cela a ainsi été montré avec la description de l'émergence d'un nouveau genre de DHV (Duck Hepatitis Virus) type 1 [338]. Cette souche possède une structure génomique typique de picornavirus à l'exception de la présence de 3 copies répétées consécutives du gène 2A, 2 copies sont plus proches des donneurs que de l'exemplaire ancestral. Ainsi, 2A1 est plus proche des Aphotavirus, le 2A3 se regroupe avec le parechovirus humain, le gène 2A2 n'a pas trouvé de lien identifié avec d'autres picornavirus. Les analyses menées sur le génome complet conduisent à l'identifier comme un nouveau clade parmi les picornavirus. Ainsi une duplication peut être la marque ou la cicatrice d'une recombinaison non homologue.

L'exploitation de ces informations génétiques est maintenant devenue le support indispensable à l'étude de l'évolution du monde vivant.

## 2. Evolution des génomes viraux à ARN

### 2.1. Principe

La phylogénie moléculaire est une discipline récente devenue maintenant incontournable pour reconstruire l'histoire évolutive des organismes vivants. L'augmentation considérable du nombre de séquences disponibles, la facilité croissante pour obtenir les séquences génomiques contemporaines et la puissance croissante des systèmes informatiques permettent le développement d'outils d'analyses statistiques complexes et robustes. Les virus à ARN, en particulier le VIH, les virus *Influenzae* ou le VHC, sont des modèles de référence qui ont largement contribué à l'essor de ces outils, ouvrant ainsi un nouveau champ d'étude de la virologie moléculaire.

En effet, l'originalité des virus à ARN réside dans leur dynamique d'évolution. Ils combinent un temps de génération court, une production de populations de grande taille et un taux de mutations élevé permettant des analyses informatives et robustes sur une échelle de temps courte « adapté au temps humain » (1 an par exemple) et de petites séquences génomiques (dès 300 nucléotides). L'échantillonnage est donc plus informatif et plus facile à obtenir qu'avec des organismes comme les eucaryotes dont le taux de substitution/site/an est un million de fois plus faible que celui des virus à ARN [339, 340]. Ces virus appartiennent avec les ADN mitochondriaux anciens à la catégorie des « measurably evolving populations » (MEPs), population d'intérêt majeur en termes d'analyses d'évolution [341]. Les séquences génétiques des MEPs peuvent être échantillonnées à différents temps (très rapprochés (ARN viraux) ou très éloignés (ADNmt anciens)) et présentent un nombre de mutations statistiquement significatifs. L'accumulation des mutations dans le temps et leur fixation dans la population peut être estimée et calculée, et permet des descriptions évolutives pertinentes. La notion de temps et d'horloge moléculaire est ainsi intégrée dans les modèles d'analyses phylogénétiques.

La large diffusion des outils de phylogénie moléculaire permet d'explorer de très nombreux aspects de la biologie virale, depuis l'épidémiologie des populations, l'investigation des transmissions, jusqu'à l'analyse des populations virales dans les compartiments chez un individu. Quel que soit la méthode utilisée, il est nécessaire de réaliser un alignement des séquences avec une grande minutie pour ne pas compromettre la fiabilité

des résultats. Ces jeux de données sont ensuite soumis à un codage informatique permettant leur utilisation dans des algorithmes statistiques. Ces algorithmes testent plusieurs hypothèses jusqu'à obtenir l'hypothèse la plus vraisemblable qui sera reconstruite sous la forme d'un arbre.

## 2.2. *Intérêt pratique*

L'ensemble des méthodes de phylogénie permettent d'aborder l'évolution d'une part à l'échelle de la population (inter-hôte) et de se concentrer alors sur les processus de transmission, la dynamique d'évolution est plus « neutre » dans ce cas, et d'autre part à l'échelle de l'individu (intra-hôte) au niveau duquel les processus de sélection sont alors plus finement étudiés.

Les outils les plus récents combinent la notion d'évolution et de processus écologique. En d'autres termes, les principes de phylogénie et de génétique des populations sont utilisés mais tiennent compte de l'évolution du pathogène au cours de l'infection et des processus de transmission (passage de barrière d'espèce ou transmission inter-hôte). Ces méthodes ont été développées pour répondre à différentes problématiques en utilisant une échelle de temps adaptée selon les modèles (notion d'horloge moléculaire) et en délimitant une zone géographique ou spatiale de travail.

Ainsi la notion de phylodynamie a ainsi émergée récemment [342]. Il peut s'agir de l'exploitation de ces méthodes à une échelle globale pour analyser l'influence des flux migratoires et des déplacements humains, pour reconstruire l'historique d'épidémies ou l'émergence de nouveaux virus. L'étude systématique de milliers de séquences d'isolats de virus grippaux à l'échelle mondiale et de façon annuelle a ainsi permis de montrer une complexité de la dynamique des épidémies composées d'épidémies hivernales en zones tempérées et de cas sporadiques tout au long de l'année en zone tropicale. La ré-émergence saisonnière est alimentée par une source persistante d'influenzae A virus en Asie du Sud-Est qui suit les réseaux de transports aériens et dissémine l'épidémie hivernale [343]. Les analyses de diversité ont montré aussi des pressions de sélection exercée par l'immunité de l'hôte à l'origine d'un taux d'évolution très élevé sur les acides aminés des sites antigéniques de l'hémagglutinine alors que les dynamiques d'évolution sont plus variables à l'échelle du génome entier [344, 345]. La dynamique d'évolution des virus grippaux est la résultante de la

combinaison de processus écologiques et de facteurs d'évolution intrinsèque. L'histoire d'autres épidémies a aussi pu être analysée et reconstruites. Nous pouvons citer par exemple, l'explosion de l'épidémie du VHC au 20<sup>ème</sup> siècle, virus implanté chez l'homme depuis plusieurs siècles, à travers l'avènement des pratiques transfusionnelles [346]. De même, les analyses de reconstruction phylogénétiques de l'histoire de la pandémie à HIV1 ont permis l'identification de multiples passages de barrière d'espèces, évènements fondateurs survenus en Afrique, de la dissémination à l'échelle mondiale et la datation du début de l'épidémie à la 1ère moitié du 20<sup>ème</sup> siècle [347, 348].

Une épidémie localisée à l'échelle d'une communauté (quelle que soit la taille de la communauté : une collectivité, une ville, un pays ou un continent) est aussi l'objet d'études de phylodynamie, il s'agit alors d'un cluster d'infection au niveau phylogénétique. Les objectifs peuvent être de déterminer la source de l'épidémie lors de l'émergence d'un nouveau virus comme cela a été le cas avec le SARS en 2003. Les analyses ont permis d'identifier une première source, la civette, puis le réservoir primaire, la chauve-souris et enfin le passage de barrière d'espèce a été daté à 4 ans grâce aux outils d'horloge moléculaire [349, 350]. Il est aussi possible de reconstruire le schéma d'invasion de la souche et d'établissement de la souche dans la population, ce qui est particulièrement intéressant en médecine vétérinaire.

Enfin, l'analyse des clusters de transmission et des chaînes de transmission, que l'histoire de transmission soit connue ou inconnue, est menée à l'aide de ces outils. Ces analyses permettent de mesurer des taux d'évolution, permettent aussi de valider les outils développés (validation des modèles d'horloge moléculaire) ou à l'inverse peuvent montrer la survenue d'incohérence dans ces modèles. La cohorte irlandaise « anti-D », qui correspond à un groupe de femmes transfusées accidentellement par le même lot de produit sanguin contaminé par du VHC à la même période, a permis ce type d'étude fournissant des conclusions robustes [351]. Quand l'histoire de transmission n'est pas connue, les objectifs peuvent être alors de rechercher la date de début d'épidémie ou la source épidémique.

L'évolution virale intra-hôte, c'est-à-dire à l'échelle de l'individu, représente le stade ultime d'expression de la diversité génétique et concerne particulièrement les virus à ARN à l'origine d'infection chronique de plusieurs mois à plusieurs années, permettant des échantillons itératifs dans le temps et des analyses de dynamique d'évolution virale pertinentes. Ces études concernent en premier lieu le VIH et le VHC. Ainsi, la dynamique

d'évolution du VIH chez l'hôte est principalement guidée par la sélection naturelle exercée par le système immunitaire. L'analyse des quasiespèces a montré que la diversité génétique est relativement basse quel que soit le moment de l'infection [352], cependant cette diversité génétique augmente progressivement au cours de l'infection chronique et est corrélée à l'évolution clinique [353] et au niveau d'échappement du VIH aux anticorps neutralisants [354]. Les analyses ont même montré des évolutions différentes des quasiespèces virales du VHC selon les compartiments cellulaires chez un même individu [355].

Ces analyses menées à l'échelle globale, locale ou intraindividuelle tendent vers des modélisations avec intégration des modèles mathématiques qui sont communes. En effet, en considérant la cellule ou le compartiment à l'échelle de l'hôte comme un individu à l'échelle populationnelle, les mêmes modèles intégrant les processus d'évolution et les processus écologiques peuvent être appliqués [356].

#### IV. Place de l'étude du domaine V3

La protéine NS5A est probablement impliquée dans l'évolution chronique de la maladie à VHC et la résistance au traitement antiviral. Une étude conduite au laboratoire de virologie du CHU d'Angers a montré que l'analyse du domaine V3 semblait être un meilleur facteur prédictif de réponse au traitement par l'interféron chez les patients infectés par le génotype 1b, la région ISDR apparaissant peu contributive [10]. Au cours de cette étude, l'amplification de la région V3 a permis de mettre en évidence chez 5 patients parmi 136, une duplication de ce domaine V3 [11, 357]. La prévalence de cet événement était de 3,67%. Quatre patients avaient une insertion de 31 acides aminés et le dernier une insertion de 27 acides aminés sans décalage du cadre de lecture. L'analyse de la séquence protéique de cette insertion a montré une augmentation de la polarité et de l'antigénicité ainsi qu'une diminution de l'hydrophobie, suggérant ainsi l'émergence d'un site antigénique. En outre, un patient présentait un lymphome et un second patient présentait une cryoglobulinémie associée à une insuffisance rénale chronique.

Dans ce contexte, nous avons réalisé le dépistage systématique de cette duplication du domaine V3 dans la population de patients infectés par le VHC de génotype 1b dépistés de façon prospective dans 19 laboratoires de virologie français, étude réalisée avec le soutien du groupe AC11 de l'ANRS (projet ANRS-Appel d'Offre 2006). Les objectifs du travail étaient :

- de définir la prévalence de ce polymorphisme et son implication dans la physiopathologie de l'infection par le VHC de génotype 1b à travers une étude épidémiologique.

Dans la mesure où le domaine V3 a été décrit comme un site hydrophile, cible possible de la réponse immunitaire à médiation humorale [272], la duplication de ce domaine pourrait être impliquée dans la réponse immunitaire antivirale qui joue déjà un rôle majeur dans l'évolution de la maladie. En effet, une réponse inflammatoire excessive accompagne l'infection par le VHC favorisant d'une part le développement du carcinome hépatocellulaire et d'autre part la survenue de pathologies dysimmunitaires. Il est ainsi rapporté une augmentation statistiquement significative des cryoglobulinémies mixtes et du risque de développer un lymphome non-hodgkinien chez les sujets infectés par le VHC [358]. Il apparaissait donc intéressant de rechercher une éventuelle implication de cette duplication en fonction de la présence de pathologies hépatiques, de cryoglobulinémies ou de lymphomes.



- de réaliser une étude d'épidémiologie moléculaire afin de déterminer l'éventuel mécanisme à l'origine de la duplication du domaine V3.

Nous avons séquencé les gènes NS5A complet, E1 et NS5B par séquençage direct et NS5A complet après clonage, afin de mener dans un second temps des analyses phylodynamiques à l'échelle interindividuelle et intraindividuelle. Nous avons aussi des prélèvements séquentiels de souches sauvages et de souches avec duplication que nous avons exploités en analyse de cinétique de quasiespèces. Nous avons pu appliquer une horloge moléculaire grâce à la disponibilité d'échantillons dès la contamination.

- de réaliser des analyses fonctionnelles au niveau protéiques et d'identifier de potentiels partenaires cellulaires préférentiels de cette protéine particulière.

Nous avons réalisé les productions en vecteurs d'expression des protéines NS5A mutées et sauvages, montrant leur expression en cultures de cellules de mammifères et en culture bactérienne. Nous avons réalisé un crible en système double-hybride levure contre une banque de foie. Nous avons initié la production des protéines NS5A mutées et sauvages pour permettre une validation des interactions en système pull-down.

## CHAPITRE 2 : EPIDEMIOLOGIE CLINIQUE DE LA DUPLICATION V3

### I. Introduction

Nous avons mis en évidence au sein de notre laboratoire un polymorphisme majeur du domaine V3 réalisant une duplication de 31 acides aminés de ce domaine sur des souches de VHC de génotype 1b. Cette insertion n'a encore jamais été décrite. Afin de préciser le rôle de ce polymorphisme dans l'histoire clinique de l'infection par le VHC de génotype 1b, nous avons mis en œuvre une étude d'épidémiologie clinique multicentrique impliquant 19 laboratoires de virologie hospitaliers français avec le soutien financier du groupe AC11 de l'ANRS obtenu en 2006.

Chaque laboratoire a inclus de façon consécutive tous les patients infectés chroniquement par le VHC de génotype 1b nouvellement identifiés par le virologue investigateur entre 2006 et 2009. Chaque inclusion était associée à une lettre d'information remise à chaque patient. La note d'information et la fiche de recueil anonyme des données clinico-biologiques ont été validées par le comité d'éthique du CHU de Brest (Avis CPP Ouest 6-15112006) et par la CNIL (Commission Nationale Informatique et Liberté). Aucun prélèvement supplémentaire n'était nécessaire et un aliquot de 500 $\mu$ L de sérum prélevé sur la sérothèque du laboratoire a été utilisé pour notre travail.

A l'issue de l'inclusion, les échantillons ont été transmis au laboratoire de virologie du CHU d'Angers à l'exception des centres de Brest, Rennes et Clermont-Ferrand qui ont réalisé les analyses de dépistage en interne. Afin de confirmer la présence d'une insertion et de définir précisément sa taille et sa composition, les échantillons identifiés en dépistage ont été séquencés.

Chaque patient inclus dans notre étude a fait l'objet d'un recueil rétrospectif de données cliniques et biologiques. Les analyses statistiques ont été menées à l'aide du logiciel SAS version 9.3 (SAS Institute, Inc., USA).

Nous avons initialement recruté 938 patients mais 117 échantillons n'ont pas pu être amplifiés. Parmi les 821 patients finalement inclus, nous avons identifié 25 patients infectés par une souche de VHC portant une insertion dans le domaine V3 soit une prévalence de 3.05% dans la population étudiée. Vingt et une souches avaient une insertion de 31 acides aminés, 3 souches avaient une insertion de 27 acides aminés et la dernière souche portait une

insertion de 12 acides aminés ; le cadre de lecture était conservé pour toutes les souches. Les analyses de séquence ont permis de confirmer qu'il s'agissait de duplication du domaine V3 avec des taux d'identité variant de 48 à 96% associés à des E-values supérieures à  $4^{-5}$ . La comparaison de prévalence a ensuite permis de montrer une corrélation entre la présence de la duplication V3 et la sévérité de la maladie hépatique: les patients avec une cirrhose avaient 2 fois plus de risque d'être porteur de la souche avec la duplication ( $p=0.04$ ). Le test de tendance montrait aussi que la prévalence de cette souche augmentait avec la présence d'une cirrhose ou d'un CHC en comparaison d'une absence de cirrhose. ( $p$  de tendance= $0.045$ ). Ce travail a identifié pour la première fois une duplication partielle d'un gène du VHC de génotype 1b dans la région NS5A. Nos analyses suggèrent aussi une possible implication de ce polymorphisme dans la sévérité de la maladie hépatique.

## II. Article 1

**Title:** Identification of a duplicated V3 domain in NS5A associated with cirrhosis and hepatocellular carcinoma in HCV-1b patients

**Short Title:** HCV NS5A-V3 duplication may favor liver diseases

**Authors:** H Le Guillou-Guillemette <sup>1,2</sup>, A Pivert <sup>1,2</sup>, S Bertrais <sup>2</sup>, P Veillon <sup>1,2</sup>, A Ducancelle <sup>1,2</sup>, E Bouthry <sup>1,2</sup>, S Alain <sup>4</sup>, V Thibault <sup>5</sup>, F Abravanel <sup>6</sup>, AR Rosenberg <sup>7</sup>, C Henquell <sup>8</sup>, E André-Garnier <sup>9</sup>, O Petsaris <sup>10</sup>, S Vallet <sup>10</sup>, JB Bour <sup>11</sup>, Y Baazia <sup>12</sup>, P Trimoulet <sup>13</sup>, P André <sup>14</sup>, C Gaudy-Graffin <sup>15</sup>, D Bettinger <sup>16</sup>, S Larrat <sup>17</sup>, A Signori-Schmuck <sup>17</sup>, H Saoudin <sup>18</sup>, B Pozzetto <sup>18</sup>, G Lagathu <sup>19</sup>, S Minjolle-Cha <sup>19</sup>, F Stoll-Keller <sup>20</sup>, JM Pawlotsky <sup>21</sup>, J Izopet <sup>6</sup>, C Payan <sup>10</sup> and F Lunel-Fabiani <sup>1,2,3</sup>.

1 Laboratoire de Virologie, CHU Angers, France

2 HIFIH Laboratory, UPRES 3859, SFR 4208, LUNAM University, Angers, France

3 Institut Vernes, Paris, France

4 Laboratoire de Virologie, CHU Limoges, France

5 Laboratoire de Virologie, CHU Pitié-Salpêtrière, Paris, France

6 Laboratoire de Virologie, CNR VHE, Inserm U1043, CHU Purpan, Toulouse, France

7 AP-HP, GHU Cochin, Laboratoire de Virologie, Université Paris Descartes, EA 4474 "Hepatitis C Virology", Paris, France

8 Laboratoire de Virologie, CHU Clermont-Ferrand, France

9 Laboratoire de Virologie, CHU Hôtel Dieu, Nantes, France

10 Département de Bactériologie-Virologie-Hygiène Hospitalière et Parasitologie-Mycologie, CHRU, LUBEM, EA3882, Brest, France

11 Laboratoire de Virologie, CHU, Dijon, France

12 Laboratoire de Virologie, CHU Avicenne, Bobigny, France

13 Laboratoire de Virologie, Hôpital Pellegrin Tripode, CHU Bordeaux, France

14 Laboratoire de Virologie, Centre de Biologie Nord, Hôpital de la Croix Rouse, Lyon, France

15 Université François Rabelais, Inserm U966, CHU Tours, France

16 Laboratoire de Virologie, CHU Besançon, France

17 Laboratoire de Virologie, UMI 3265 UJF-EMBL-CNRS, CHU, Unit of Virus Host Cell Interactions, Grenoble, France

18 Laboratoire de Bactériologie-Virologie, CHU Saint-Etienne, France

19 Laboratoire de Virologie, CHU Rennes, France

20 Institut de Virologie, CHU Strasbourg, Inserm U748, Strasbourg, France

21 Laboratoire de Virologie-Bactériologie, CHU Henri-Mondor, Créteil, France

**Corresponding author:** Dr Hélène Le Guillou-Guillemette, Laboratoire de Virologie, Département des agents infectieux, IBS, CHU Angers, 4 rue Larrey 49933 Angers cedex9, France. Phone: 33-241354709. Fax: 33-241354164. E-mail: heleguillou@chu-angers.fr

**Conflict of interest:** the authors involved in this study declare that they do not have anything to disclose regarding funding or conflicts of interest in respect to this manuscript.

**Abstract:**

The NS5A protein of the hepatitis C virus has been shown to be involved in the development of hepatocellular carcinoma. In a French multicenter study, we investigated the clinical and epidemiological features of a new HCV genotype 1b strain bearing a wide insertion into the V3 domain. We studied NS5A gene sequences in 821 French patients infected with genotype 1b HCV. We identified an uncharacterized V3 insertion without ORF disruption in 3.05% of the HCV sequences. The insertion comprised 31 amino acids (aa) for the majority of patients; 3 patients had 27 aa insertions and 1 had a 12 aa insertion. Sequence identity between the 31 aa insertions and the V3 domain ranged from 48 to 96% with E-values above  $4e^{-5}$ , thus illustrating sequence homology and a partial gene duplication event that to our knowledge has never been reported in HCV. No particular geographical localization was found for this strain. No association was found with extrahepatic diseases, i.e., cryoglobulinemia or lymphoma. Conversely, patients with cirrhosis were two times more likely to have HCV with this genetic characteristic ( $p=0.04$ ). Moreover, its prevalence increased with liver disease severity (from 3.0% in patients without cirrhosis to 9.4% in patients with both cirrhosis and HCC,  $p$  for trend=0.045). We identified a duplicated V3 domain in the HCV-1b NS5A protein for the first time. The duplication may be associated with unfavorable evolution of liver disease including a possible involvement in liver carcinogenesis.

**Key words:** hepatitis C virus, NS5A, hepatocellular carcinoma, cirrhosis, V3 domain

## Introduction

Infection by hepatitis C virus (HCV) becomes persistent in around 80% of cases and in turn a major cause of liver disease for the approximately 184 million HCV patients worldwide [1]. The infection is a major public health issue as these chronically infected patients are at risk of developing liver cirrhosis and hepatocellular carcinoma (HCC) [2].

HCV is a single-strand RNA virus belonging to the *Flaviviridae* family. Its genome encodes a polyprotein of about 3000 amino acids (aa) that is then cleaved by viral and host proteases into structural (Core, E1 and E2, p7) and non-structural (NS2, NS3, NS4A and NS4B, NS5A and NS5B) proteins. As for other RNA viruses, HCV is notable for its considerable genetic diversity: presently, 7 genotypes and more than 70 subtypes have been described [3].

NS5A is a phosphoprotein found in a basally phosphorylated 56 kDa form and a hyperphosphorylated 58 kDa form. The N-terminal domain contains an amphipathic alpha helix mediating membrane anchorage [4]. The cytoplasmic portion is divided into three domains separated by low complexity sequences. Domain I has been solved by protein crystallography; it is a zinc binding domain that may dimerize. Domains II and III are both natively unfolded [5]. NS5A is not yet fully characterized. However, we do know that it is a pleiotropic protein with several identified functions. Thus, studies strongly suggest that NS5A plays a role in resistance to interferon (IFN)-based therapy [6], in the establishment of chronic hepatitis and in liver carcinogenesis [7]. NS5A continues to be a major subject of research especially since it was recently identified as a target for new antivirals [8].

A variable region, named V3, was identified by Inchauspe *et al.* in the NS5A domain III (NS5A<sub>2356-2379</sub>) [9]. A correlation between the mutation level in V3 and the response to IFN therapy was firstly demonstrated by Duverlie *et al.* [10]. We and others confirmed that pre-therapeutic mutational analysis of the V3 domain in quasispecies appeared predictive of virological response to IFN [11].

In previous studies [11, 12] in HCV-1b-infected patients, we discovered a NS5A polymorphism that may consist in V3 domain duplication. One of those patients had cryoglobulinemia and developed a non-Hodgkin lymphoma. To better understand this unknown but intriguing HCV polymorphism, we performed a multicenter study to estimate its prevalence and identify its epidemiological features.



## **Patients and methods**

### *Study Patients*

Sera from consecutive patients newly diagnosed with chronic HCV-1b infection between 2006 and 2009 were selected at the routine genotyping step in each participating virology laboratory (19 French academic centers, Figure 1). No additional serum samples were required for the study.

### *Ethics Statement*

An information letter was sent to all patients and the study was approved by the Ethics Committee of the University Hospital of Brest (Avis CPP Ouest 6-15112006). The CPP do not required a written consent form as (i) the research was not interventional and any clinical or therapeutic consequences were involved (ii) no additional serum sample was done, research was performed in the HCV genotyping sample and focused on the HCV genome only. The study was also approved and registered by the National Commission for information technology and civil liberties (CNIL, 907057-09032007).

### *Demographic and clinical data*

Date of blood sample collection and biological parameters including HCV genotype, viral load and biochemical measures were obtained from the participating laboratories. Demographic (age, gender) and clinical (fibrosis evaluation and presence of cirrhosis (method, date and result), presence of HCC, cryoglobulinemia, lymphoma or co-infection (HIV or HBV), presumed date of infection and route of transmission if available) data at the time of inclusion were retrospectively collected from medical records. Using the collected data, we were able to determine fibrosis stages or the presence of cirrhosis in 542 patients. One hundred and eight patients were classified as cirrhotic, including 61 subjects (56.5%) with histologically confirmed cirrhosis (i.e., fibrosis stage evaluated as F4 according to the Metavir classification system). In 24 patients (22.2%) the diagnosis of cirrhosis was based on

both liver stiffness evaluation by transient elastography (Fibroscan) and a concordant diagnosis for at least one liver blood test (FibroMeter, Fibrotest, APRI) or corroborating clinical information. The 23 other patients were known to have cirrhosis according to the recorded clinical data, but the diagnostic details were not reported.

#### *NS5A amplification and direct nucleotide sequencing*

All samples were centralized and analyzed in the virology laboratory of the University Hospital of Angers, except for samples from Clermont-Ferrand and Rennes, which were screened on site. HCV RNA was extracted from 200 µL of each serum sample using the EasyMag automated extraction system (BioMérieux, Craponne, France). As previously described [13], the C-terminal portion of the NS5A gene (388 bp) was amplified by nested RT-PCR. Reverse transcription was combined with the first round of PCR using the outer primer set S2 and E2. A second-round primer set was used consisting of the inner primers S31 and I41b. The amplified products were resolved using 2% agarose gel electrophoresis, stained with ethidium bromide, and visualized with UV light. PCR products larger than expected were sequenced to assess the presence of an inserted domain. Samples bearing a confirmed insertion were then subjected to full length NS5A gene amplification and sequencing using outer primers E1 and E2 for the RT-PCR (1665 bp) and inner primers I3 and I41b (1344 bp) for the nested PCR.

NS5A PCR products were subjected to direct sequencing using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on the automated ABI3130xl, with forward and reverse primers used in the nested-PCR rounds, respectively. The aa were numbered according to the sequence of the HCV-1b HCV-J prototype (Genbank Accession number: D90208). If necessary, supplemental primers were also used for the NS5A sequencing (S2, AS2, S4, S31, AS32) [13].

### *Sequence analysis of the NS5A gene*

All sequences were aligned using SeqScape software (Applied BioSystems). When both were present, the aa sequences of the inserted domain and the V3 domain were compared pairwise using BLASTP [14]. Sequences were considered homologous when the expected (E)-value was above  $10^{-5}$ . This cut-off has already been used to assess duplication events in RNA viruses and was considered relatively stringent [15].

The sequence data reported in the present work were deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases: accession numbers KF420489 through KF420513.

### *Statistical analyses*

Statistical analyses were performed using SAS version 9.3 (SAS Institute, Inc., USA). The distribution of categorical variables was expressed as frequencies and percentages. Quantitative variables were described in mean values +/- standard deviation (SD) and median. Differences between patient groups were analyzed using the Fisher's exact test for nominal variables and the non-parametric Mann-Whitney-Wilcoxon test for continuous variables, unless otherwise indicated.

To assess the clinical relevance of the NS5A V3 insertion, we firstly compared the proportion of patients having this genetic characteristic as a function of several diseases. For this, 95% confidence intervals (CI) were calculated by the exact method assuming binomial distribution. We additionally performed an exact Cochran-Armitage trend test to assess whether the proportion of patients having the NS5A V3 insertion increased with liver disease severity (no cirrhosis, cirrhosis without HCC, cirrhosis with HCC). A multivariate ordinal logistic regression model was then used to assess the relationship of the NS5A insertion with liver disease after controlling for potential confounding factors (i.e., age and sex).

## Results

### *Identification of a duplicated V3 domain in the NS5A gene*

We initially enrolled 938 patients from 19 university virology laboratories. 117 patients were subsequently excluded due to failed NS5A amplification. Those excluded patients were younger on average ( $49.5 \pm 13.8$  vs  $54.8 \pm 14.6$  years, p-value from Student t-test=0.0002) and more frequently men (69.8% versus 50.3%,  $p < 0.0001$ ). However, no viral load differences were observed ( $5.80 \pm 0.86$  vs  $5.91 \pm 0.78$  log IU/mL, p-value from Student t-test=0.19) between excluded and included patients. Also, excluded patients did not significantly differ from included patients regarding disease duration, contamination date or alcohol abuse in the subsample of patients for whom the pertinent information was available (data not shown). Consequently, further statistical analyses were performed on 821 patients (Table 1).

Twenty-five of the 821 test subjects were infected by an HCV-1b strain bearing a NS5A gene longer than expected. Direct NS5A sequencing of these samples revealed V3 domain duplication. Insertion lengths were 31 aa for 21 patients, 27 aa for three, and 12 aa for one. The two V3 domains appeared in tandem without open reading frame (ORF) disruption as shown in Figure 2. The duplicated domains presented strong sequence similarities with a level of sequence identity ranging from 50 to 91%, E-values (expected values) ranged from  $4e^{-5}$  to  $2e^{-13}$ .

The prevalence of this novel duplication in our cohort was 3.05% (95% CI: 1.98-4.46). No particular geographical localization was seen for this mutant strain (Figure 1).

*Comparison between patients infected by the mutant HCV strain and those infected by wild-type HCV*

Patient characteristics according to V3 status are shown in Table 1. The only difference between the two groups was that patients infected by an HCV virus bearing a V3 duplicated domain were older ( $p=0.035$ ). No differences were found in terms of sex ratio, disease duration or HCV viral load.

In the subsample of patients for whom contamination information was available, time since contamination was highly related to age (Spearman rank correlation coefficient  $r=0.47$ ,  $p<0.0001$ ) but did not significantly differ according to V3 duplication status ( $p=0.91$ ).

*Clinical relevance of the V3 duplication*

To assess possible relationships between the V3 duplication and extrahepatic disease, we classified patients according to the presence of cryoglobulinemia or lymphoma. No significant difference in the prevalence of V3 duplication was observed between patients having cryoglobulinemia or lymphoma as compared to those free of these extrahepatic diseases (Figure 3A).

In contrast, a significant relationship was found between V3 duplication and liver disease. The proportion of patients having the mutant strain was two times higher in patients with hepatocellular carcinoma compared to those without. We also found a significant association between the presence of the V3 duplication and cirrhosis (Figure 3B). Furthermore, the prevalence of the mutant strain increased with liver disease severity ( $p$  for trend=0.0449) (Table 2). Unadjusted ordinal logistic regression analysis showed that the odds of having cirrhosis and/or HCC versus being non-cirrhotic was 2.4 times greater in patients with V3 duplication (OR [95% CI]: 2.41 [1.00-5.81],  $p=0.0502$ ). After adjustment for age and sex, the odds-ratio remained higher than 2.0, but it was non-significant at the conventional 0.05 level

(OR [95% CI]: 2.06 [0.84-5.02],  $p=0.1123$ ). Because of the proportional odds assumption of the ordinal logistic regression model, the presence of V3 duplication was associated with the same two-fold increase in odds for having HCC versus the combined category “cirrhotic but no HCC” or “non-cirrhotic”, after controlling for age and sex.

## Discussion

The pleiotropic nature of NS5A in the viral life cycle of HCV has sparked numerous analyses. In a previous work, we discovered an HCV-1b strain bearing a duplicated V3 domain in the NS5A gene [12]. Here, we performed a large multicenter study to characterize this polymorphism at the clinical level.

We found 25 cases with a high sequence identity between both V3 domains, illustrating partial gene duplication in NS5A. In our test population, the prevalence of the strain was 3.05% (95% CI: 1.98-4.46). To our knowledge, this polymorphism has never been reported before. Authors recently reported that gene duplication is infrequent in RNA viruses: they identified only nine cases of gene duplication among 1198 RNA virus species analyzed using a BLAST approach [15]. Specifically in HCV, insertion or deletion is quite rare. A large quasispecies study identified only infrequent, very small (from 1 to 4 aa) insertions in the E1-E2 or NS5A regions of HCV-1a and 1b. This work demonstrated that insertions were detected in all the specimens comprising sequential samples but not always in all the clones [16]. In our present work, we used direct sequencing to detect an insertion that was not only large but also a tandemly repeated domain. Moreover, an analysis performed at the quasispecies level showed that the V3 duplication was present in 100% of the clones of the 25 mutant HCVs (n=870); we ruled out technical PCR artifacts that could give rise to the duplication. We also identified the V3 duplication following kidney transplantation and its persistence in one patient (data not shown). Others studies have described small indels in the 5'UTR [17] or in the E1-E2 region, especially in the HVR1 subdomain [18]. In our work, the insertion contained 31 aa for almost all of the mutant HCVs (21/25). We thus verified that all the cases belonged to genotype 1b by sequencing the E1, NS5A and NS5B regions. Furthermore, Moradpour *et al.* demonstrated that domain III of NS5A tolerates wide heterologous insertion, for example by the GFP protein in the HCV replicon [19]. Our findings demonstrate that

insertions are also naturally tolerated as suggested *in vitro*. In contrast, past reports of genomic HCV polymorphisms have mainly described recombinant strains; none described duplicated domains [20]. Gene duplication in RNA viruses is possible, but it is a contradictory evolutionary pathway. Duplication leads to increased genome size, which may lead to new evolutionary opportunities, or conversely the enhancement of defects in production or replication [21]. Thus, a high mutation rate offers a better evolutionary path for RNA viruses. Therefore, a 3% prevalence for the mutated strain exhibiting duplication in the ORF should be considered as high.

In all cases, indels are one feature of the genetic variability characterizing RNA virus evolution, but the physiopathological consequences of these changes in HCV have not been well explored. The clinical relevance of indels in HCV has been mainly characterized by Gerotto *et al.* They showed that a 1-aa insertion or deletion in HVR1 was significantly associated with mixed cryoglobulinemia type 2 [22]. Other authors have correlated indels in the NS5A domain with genotype specific signatures and suggested that they participate in evolutionary dynamics, with a putative role in the resistance to antiviral therapy or host immune response evasion [23].

Gene duplications have been described as a major evolutionary tool for DNA viruses with large genomes encoding multigene families [24]. These phenomena represent a selective advantage, providing an opportunity for species diversification, first as genetic redundancy then co-evolution of the duplicated genes. This allows for the acquisition of new biological functions or the modification of virulence. However, biological constraints make gene duplication extremely rare in RNA viruses [25].

The few cases of partial gene duplication reported in RNA viruses were located in untranslated regions of *Flaviviridae* [26, 27] or in short intragenic regions. For example, in the bovine viral diarrhea virus, a duplicated NS3 gene was associated with increased disease



severity [28]. Similarly, an emerging subgroup (J) of the avian leukosis virus characterized by a 19bp insertion in the leader sequence has been associated with hemangioma, whereas other subgroups are associated with myeloid leucosis [29]. In human virology, Trento *et al.* described a new genotype of human respiratory syncytial virus (BA) bearing a 60-nucleotide duplication in the domain coding for the attachment G glycoprotein that emerged during successive epidemic seasons. This strain was initially identified at low circulation levels but then became the most prevalent circulating strain, possibly due to increased fitness and better adaptation to the host [30].

HCV infection severity before and after treatment has been linked to host parameters such as gender, age of infection, alcohol abuse and more recently genetic IL28B polymorphism. The virological characteristics associated with pejorative evolution are viral load and pre-therapeutic quasispecies diversity. A number of reports have associated aa variations in NS5A with issues in antiviral therapy [6, 11, 31] or the evolution of liver disease toward cancer [32]. Hung *et al.* showed that fewer aa substitutions in the NS5A PKRbd domain and the ISDR were correlated with HCC in HCV-1b patients [33]. El-Shamy *et al.* recently found this trend also: particular mutations in NS3 and NS5A were more closely associated with the development of HCC in their work [34]. The results of the present study suggest that the identified polymorphism in NS5A may also be associated with a pejorative progression of liver disease. We found a V3 domain duplication in around 3% of the studied patients, and observed that this genetic characteristic was much more prevalent in patients with cirrhosis and HCC. Moreover, this prevalence increased with liver disease severity, and the relationship with liver complications remained strong after adjustment for sex and age, although non-significant at the conventional 0.05 p-value. Unfortunately, we were unable to identify other factors that may potentially influence liver disease severity because the pertinent data were available only in a subsample of subjects. Date and route of transmission were not frequently

recorded, as mostly observed in prospective studies [35]. Time since contamination may be one of the potential confounding factors of the observed relationships with liver complications. However this factor was indirectly controlled by age adjustment in our study. Indeed, we found a high correlation between time since contamination and age among patients for whom this information was available. Moreover, in this subsample, time since contamination lost its significant relation to liver complications after age adjustment (data not shown).

Although we feel that the prevalence of the duplication in NS5A should be considered as elevated in the setting of RNA viruses and especially in that of HCV, the small number of patients with V3 insertions in our study population may lead to a lack of statistical power. This may explain the borderline or non-significant p-values and the wide confidence intervals, despite the strong observed differences. For these same reasons, the non-significant relationship with extrahepatic disease (cryoglobulinemia and lymphoma) should be interpreted cautiously. This kind of correlation was previously identified with single nucleotide insertions located in the HVR1-envelope [22]. Contrary to E1-E2, NS5A does not appear to be genuinely involved in immune system stimulation or clonal selection of lymphocytes. The V3 domain, located within NS5A domain III, has been shown to be dispensable for HCV replication *in vitro* [36], and our results suggest that its duplication does not influence viral load *in vivo*. Indeed, we observed no differences in viral loads as a function of V3 status. However, NS5A appeared to be indispensable for viral replication. As recently described, NS5A domain III seems to be involved in viral particle assembly [37] and the consequence of the V3 duplication in this step should be explored.

Here, we identified a wide duplication event in the NS5A protein of a HCV-1b strain and demonstrated that it may be associated with a higher risk of liver complications. The results of this cross-sectional study should be confirmed in a larger cohort of patients infected by the

HCV strain bearing the V3 duplication. Additional studies are now needed to fully characterize this emerging viral strain.

### **Acknowledgements**

The authors thank K. Erwin for proofreading the English manuscript.

The authors thank the AC11-HCV Study and AC33-Virology groups, and the following clinicians: Angers - I Fouchard-Hubert, P Calès; Avicenne - D Roulot; Besançon - V Di Martino; Bordeaux - V de Lédighen; Brest - JB Nousbaum, F Tanné; Clermont-Ferrand - A Abergel; Dijon - A Minello, P Hillon; Grenoble - JP Zarski, V Leroy, MN Hilleret; Limoges - V Loustaud-Ratti; Lyon - V Zoulim; Nantes - J Gournay; Paris-Cochin - P Podevin; Paris-Pitié Salpêtrière - P Lebray; Rennes - D Guyader; Saint-Etienne - M. Ecochard, X. Roblin; Strasbourg - M. Doffoël; Toulouse - S Métivier, K Barange; Tours - L D'Alteroche, Y Bacq.

The authors thank A Lefebvre, E Muel, Y Desille, I Renard, A Plages, A Rodallec and W Merrouche for their assistance in clinical data recording.

## References

1. Mohd Hanafiah K, Groeger J, Flaxman AD and Wiersma ST (2013) Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology* 57: 1333-1342.
2. Omland LH, Jepsen P, Krarup H, Schonning K, Lind B, et al. (2011) Increased mortality among persons infected with hepatitis C virus. *Clin Gastroenterol Hepatol* 9: 71-78.
3. Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, et al. (2005) Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* 42: 962-973.
4. Moradpour D, Brass V and Penin F (2005) Function follows form: the structure of the N-terminal domain of HCV NS5A. *Hepatology* 42: 732-735.
5. Hanouille X, Badillo A, Verdegem D, Penin F and Lippens G (2010) The domain 2 of the HCV NS5A protein is intrinsically unstructured. *Protein Pept Lett* 17: 1012-1018.
6. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, et al. (1996) Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* 334: 77-81.
7. Macdonald A and Harris M (2004) Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol* 85: 2485-2502.
8. Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, et al. (2010) Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. *Nature* 465: 96-100.
9. Inchauspe G, Zebedee S, Lee DH, Sugitani M, Nasoff M, et al. (1991) Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates. *Proc Natl Acad Sci U S A* 88: 10292-10296.
10. Duverlie G, Khorsi H, Castelain S, Jaillon O, Izopet J, et al. (1998) Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity. *J Gen Virol* 79 ( Pt 6): 1373-1381.
11. Veillon P, Payan C, Le Guillou-Guillemette H, Gaudy C and Lunel F (2007) Quasispecies evolution in NS5A region of hepatitis C virus genotype 1b during interferon or combined interferon-ribavirin therapy. *World J Gastroenterol* 13: 1195-1203.
12. Lunel-Fabiani F, Veillon P, Le Guillou-Guillemette H, Redureau J, Coeffic B, et al. (2005) Duplication of the V3 domain in Nonstructural 5A Protein related to Hepatitis C Virus extrahepatic manifestations. In: Abstracts, 45th Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC). Washington, DC, USA: American Society for Microbiology. pp. 485.
13. Veillon P, Payan C, Gaudy C, Goudeau A and Lunel F (2004) [Mutation analysis of ISDR and V3 domains of hepatitis C virus NS5A region before interferon therapy with or without ribavirin]. *Pathol Biol (Paris)* 52: 505-510.
14. Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, et al. (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J* 272: 5101-5109.
15. Simon-Loriere E and Holmes EC (2013) Gene Duplication Is Infrequent in the Recent Evolutionary History of RNA Viruses. *Mol Biol Evol*.
16. Torres-Puente M, Cuevas JM, Jimenez-Hernandez N, Bracho MA, Garcia-Robles I, et al. (2007) Contribution of insertions and deletions to the variability of hepatitis C virus populations. *J Gen Virol* 88: 2198-2203.

17. Revie D, Alberti MO, Braich RS, Bayles D, Prichard JG, et al. (2006) Discovery of significant variants containing large deletions in the 5'UTR of human hepatitis C virus (HCV). *Virology* 354: 217-223.
18. Hohne M, Schreier E and Roggendorf M (1994) Sequence variability in the encoding region of hepatitis C virus isolated from patients infected during a single source outbreak. *Arch Virol* 137: 25-34.
19. Moradpour D, Evans MJ, Gosert R, Yuan Z, Blum HE, et al. (2004) Insertion of green fluorescent protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus replication complexes. *J Virol* 78: 7400-7409.
20. Morel V, Fournier C, Francois C, Brochet E, Helle F, et al. (2011) Genetic recombination of the hepatitis C virus: clinical implications. *J Viral Hepat* 18: 77-83.
21. Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, et al. (2007) Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol Biol Evol* 24: 845-852.
22. Gerotto M, Dal Pero F, Loffreda S, Bianchi FB, Alberti A, et al. (2001) A 385 insertion in the hypervariable region 1 of hepatitis C virus E2 envelope protein is found in some patients with mixed cryoglobulinemia type 2. *Blood* 98: 2657-2663.
23. Cuevas JM, Gonzalez M, Torres-Puente M, Jimenez-Hernandez N, Bracho MA, et al. (2009) The role of positive selection in hepatitis C virus. *Infect Genet Evol* 9: 860-866.
24. Shackelton LA and Holmes EC (2004) The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol* 12: 458-465.
25. Holmes EC (2003) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol* 11: 543-546.
26. Gritsun TS and Gould EA (2006) The 3' untranslated region of tick-borne flaviviruses originated by the duplication of long repeat sequences within the open reading frame. *Virology* 354: 217-223.
27. Gritsun TS and Gould EA (2006) Direct repeats in the 3' untranslated regions of mosquito-borne flaviviruses: possible implications for virus transmission. *J Gen Virol* 87: 3297-3305.
28. Nagai M, Sakoda Y, Mori M, Hayashi M, Kida H, et al. (2003) Insertion of cellular sequence and RNA recombination in the structural protein coding region of cytopathogenic bovine viral diarrhoea virus. *J Gen Virol* 84: 447-452.
29. Lai H, Zhang H, Ning Z, Chen R, Zhang W, et al. (2011) Isolation and characterization of emerging subgroup J avian leukosis virus associated with hemangioma in egg-type chickens. *Vet Microbiol* 151: 275-283.
30. Trento A, Casas I, Calderon A, Garcia-Garcia ML, Calvo C, et al. (2010) Ten years of global evolution of the human respiratory syncytial virus BA genotype with a 60-nucleotide duplication in the G protein gene. *J Virol* 84: 7500-7512.
31. Puig-Basagoiti F, Fornis X, Furcic I, Ampurdanes S, Gimenez-Barcons M, et al. (2005) Dynamics of hepatitis C virus NS5A quasispecies during interferon and ribavirin therapy in responder and non-responder patients with genotype 1b chronic hepatitis C. *J Gen Virol* 86: 1067-1075.
32. De Mitri MS, Morsica G, Cassini R, Bagaglio S, Zoli M, et al. (2002) Prevalence of wild-type in NS5A-PKR protein kinase binding domain in HCV-related hepatocellular carcinoma. *J Hepatol* 36: 116-122.
33. Hung CH, Chen CH, Lee CM, Wu CM, Hu TH, et al. (2008) Association of amino acid variations in the NS5A and E2-PePHD region of hepatitis C virus 1b with hepatocellular carcinoma. *J Viral Hepat* 15: 58-65.

34. El-Shamy A, Shindo M, Shoji I, Deng L, Okuno T, et al. (2013) Polymorphisms of the core, NS3 and NS5A proteins of hepatitis C 1 virus genotype 1b associate with development of hepatocellular carcinoma. *Hepatology*.
35. Zeuzem S, Teuber G, Lee JH, Ruster B and Roth WK (1996) Risk factors for the transmission of hepatitis C. *J Hepatol* 24: 3-10.
36. Tellinghuisen TL, Foss KL, Treadaway JC and Rice CM (2008) Identification of residues required for RNA replication in domains II and III of the hepatitis C virus NS5A protein. *J Virol* 82: 1073-1083.
37. Appel N, Zayas M, Miller S, Krijnse-Locker J, Schaller T, et al. (2008) Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog* 4: e1000035.

## Figure legends

**Figure 1. Geographical recruitment and localization of the V3-duplicated HCV-1b strain.** The number of included patients is represented by proportional circles as indicated. X represents the number of patients with the V3-duplicated HCV strain in each center.

**Figure 2. Alignments of the 25 sequences showing a V3-NS5A duplicated domain with respect to the HCV-J 1b reference sequence.** The first V3 domain is located between positions 2353 and 2383 (light grey column), the second is tandemly located thereafter out to position 2414 (dark grey column). Dots (.) represent identical residues; dashes (-) indicate shorter insertions in the last four aligned sequences; X denotes a mixture.

**Figure 3. Comparisons of the proportion of patients having the V3-duplicated HCV strain according to clinical status for two extrahepatic diseases (A) and to the presence of HCC or cirrhosis (B).** Error bars represent 95% confidence intervals.

**Table 1. Characteristics of included patients according to HCV-1b NS5A V3 status**

	Total	Presence of V3 duplication		p
	n=821	No (n=796)	Yes (n=25)	
	n (%) or mean (sd) / median	n (%) or mean (sd) / median	n (%) or mean (sd) / median	
Male (%)	50.3	50.3	48	0.84 <sup>(a)</sup>
Age (years)	54.8 (14.6) / 54.7	54.6 (14.6) / 54.3	61.0 (14.2) / 59.7	0.035 <sup>(b)</sup>
Viral load (log UI/mL)	5.91 (0.78) / 5.98	5.91 (0.78) / 5.99	5.82 (0.62) / 5.88	0.43 <sup>(b)</sup>
Time since contamination (years)	n=191	n=176	n=15	
	23.8 (11.9) / 23.0	23.8 (12.2) / 23.0	24.7 (9.3) / 22.0	0.91 <sup>(b)</sup>
Contamination route known	n=279	n=265	n=14	
Transfusion	n=191 (68.5)	n=179 (67.6)	n=12 (85.7)	0.24 <sup>(b)</sup>
Injected drug use	n=43 (15.4)	n=42 (15.8)	n=1 (7.1)	0.70 <sup>(b)</sup>

<sup>(a)</sup> Fisher's exact test; <sup>(b)</sup> Non-parametric Mann–Whitney–Wilcoxon test



**Table 2. Patient characteristics and observed prevalence of V3 duplication according to combined status for cirrhosis and hepatocellular carcinoma.**

	No cirrhosis	Cirrhosis without HCC	Cirrhosis with HCC	
	n=432	n=87	n=32	
Variable	% or mean (sd) / median	% or mean (sd) / median	% or mean (sd) / median	p for linear trend
Male (%)	44.1	51.8	68.7	0.0056 <sup>(a)</sup>
Age (years)	54.1 (14.1) / 53.9	59.9 (12.2) / 62.0	65.2 (9.7) / 66.6	<0.0001 <sup>(b)</sup>
Viral load (log IU/ml)	5.91 (0.74) / 5.97	5.94 (0.79) / 6.00	5.67 (0.80) / 5.72	0.17 <sup>(b)</sup>
V3 duplication (%), 95% CI	3.01 (1.61-5.09)	5.75 (1.89-12.90)	9.38 (1.98-25.02)	<b>0.0449</b> <sup>(a)</sup>

<sup>(a)</sup> Cochran-Armitage test for trend; <sup>(b)</sup> General linear model including a linear contrast to test trend

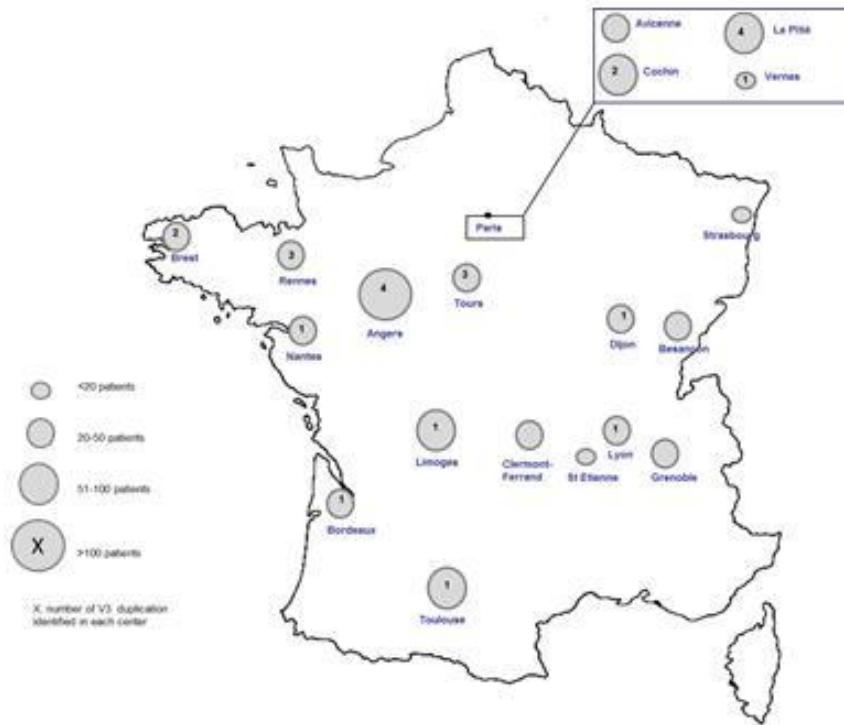


Figure 1

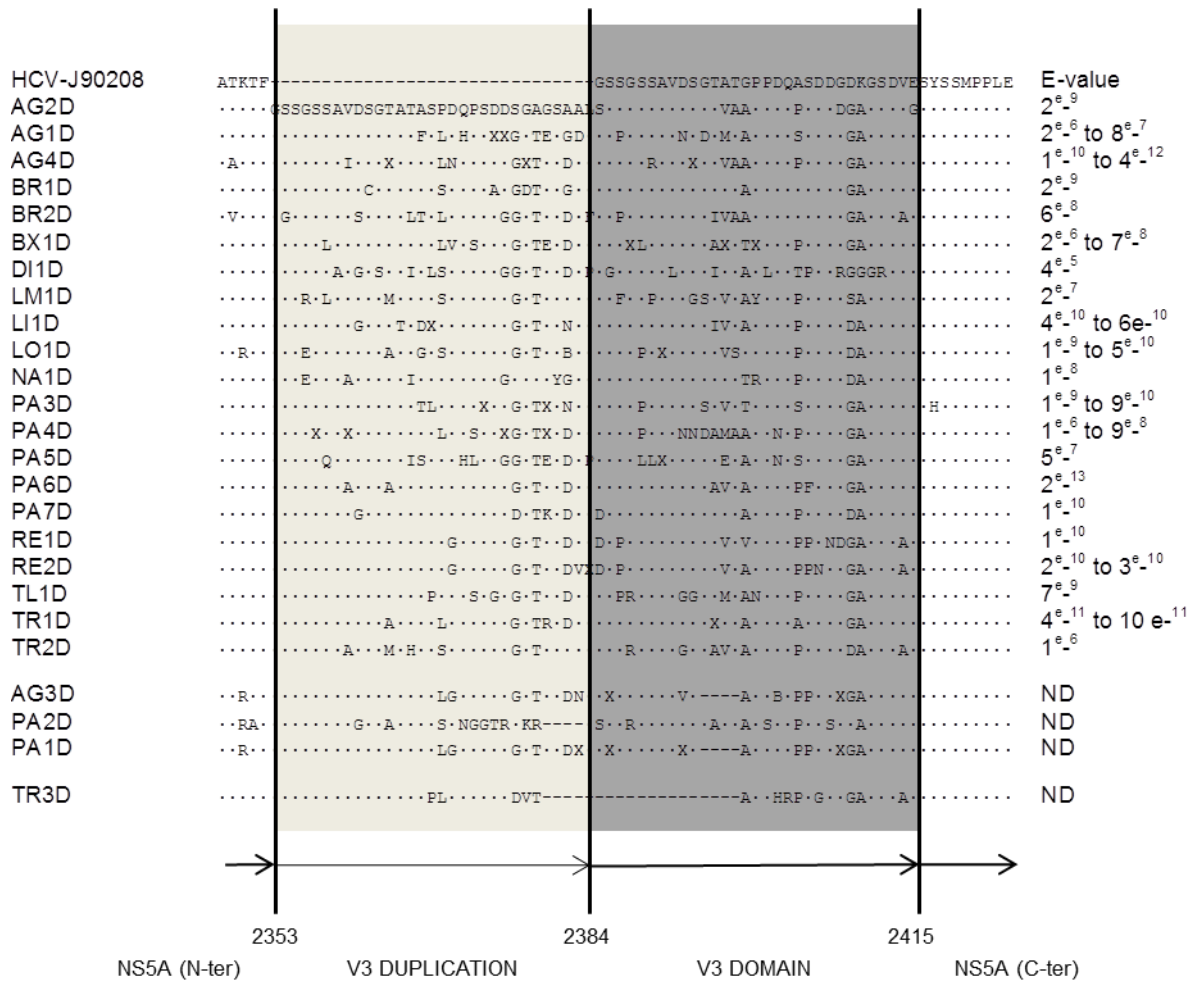
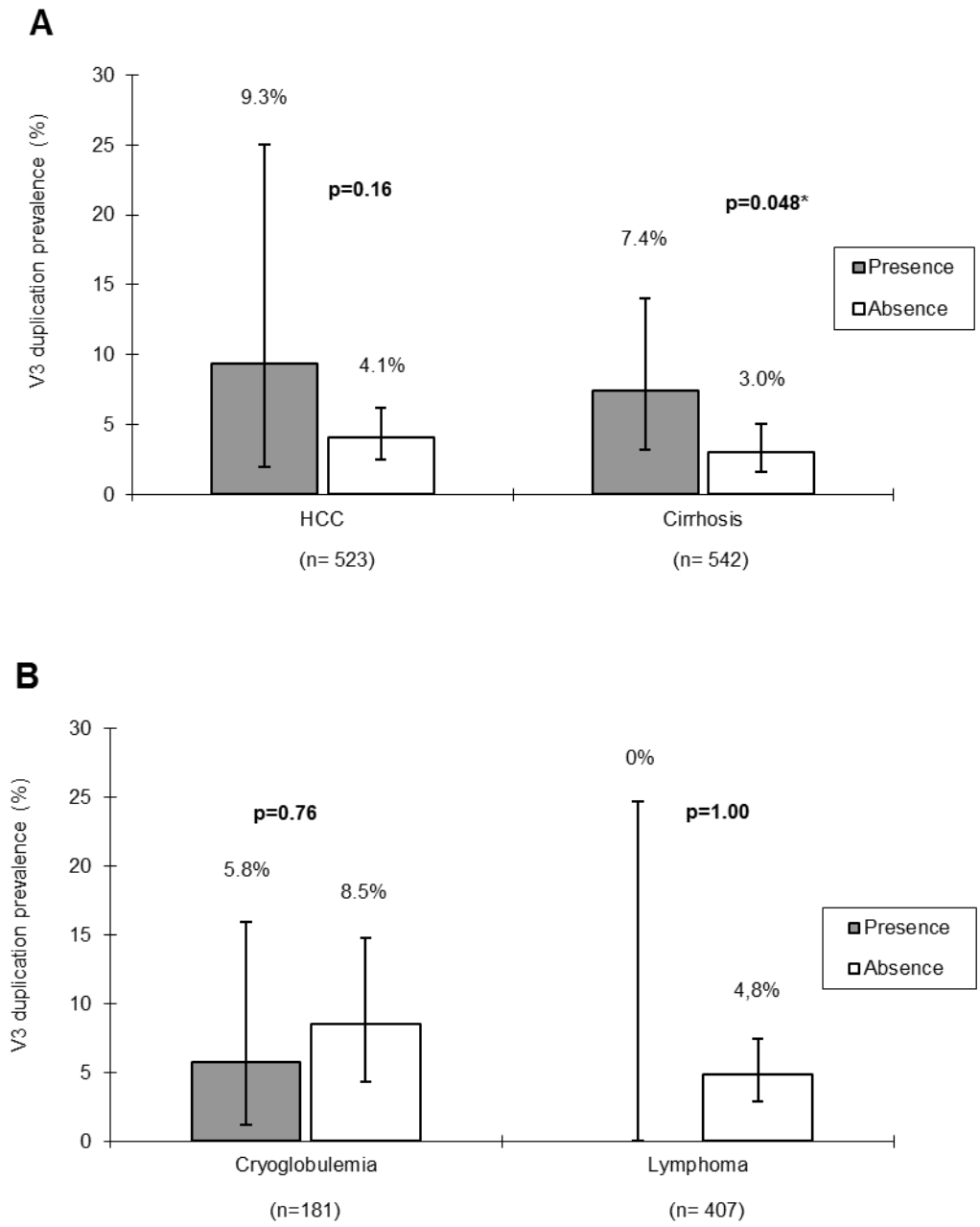


Figure 2



**Figure 3**

## CHAPITRE 3 : EPIDEMIOLOGIE MOLECULAIRE DES SOUCHES PORTANT UNE INSERTION DU DOMAINE V3

### I. Introduction

Chez les virus à ARN, les émergences de nouvelles souches proviennent principalement de mutations ou de recombinaisons homologues intra ou inter-géniques. Les évènements de recombinaisons non-homologues sont généralement délétères et sont par conséquent plus rarement observés.

Afin de caractériser l'origine des souches de VHC de génotype 1b présentant une insertion dans le domaine V3 et par là même de découvrir le mécanisme évolutif à l'origine de la duplication du domaine V3, nous avons mené des analyses moléculaires sur l'ensemble des souches précédemment identifiées au laboratoire. Nous avons réalisé d'une part un travail d'amplification, de clonage et de séquençage du gène NS5A entier, nous avons d'autre part analysé par séquençage direct les gènes NS5A et E1 de ces souches.

Ainsi, le séquençage direct des gènes NS5a et E1 a été effectué sur 27 souches mutantes. Une analyse à l'échelle de la quasi-espèce a aussi été menée par clonage et séquençage du gène NS5a chez l'ensemble des souches mutantes et d'autres souches sauvages sur des prélèvements itératifs. Nous avons analysé les polymorphismes de séquence sur ces souches mutantes ainsi que sur d'autres souches de type sauvage (1 domaine V3) à l'aide de méthodes bayésiennes basées sur la coalescence permettant l'inférence d'une histoire évolutive (temps de divergence, taux de mutation) et l'identification du mécanisme moléculaire à l'origine de la pseudo-duplication de V3.

Nous montrons que les souches mutantes représentent la totalité de la quasi-espèce infectant l'individu dès la contamination et persistent au cours de l'infection jusqu'à plus de 10 ans. L'étude de l'histoire évolutive révèle que les deux copies des domaines V3 des souches mutantes dérivent respectivement de deux populations parentales sauvages du génotype 1b et ayant divergé il y a environ cent ans. Enfin la pseudo-duplication du domaine V3 proviendrait probablement d'une recombinaison non-homologue entre souches de ces deux populations sauvages. Une plus grande sévérité de la maladie hépatique liée à la présence de cet haplotype mutant pourrait expliquer sa persistance temporelle.

Les recombinaisons non-homologues entre domaines ou gènes sont des évènements rarement observés chez les virus à ARN du fait de leur probable impact délétère sur le virus. L'observation *in vivo* de souches issues d'un tel évènement durant l'intégralité du processus infectieux constitue une première en la matière. Nous démontrons ainsi l'émergence réelle d'une nouvelle population de souches potentiellement plus virulente que la population sauvage. Ce résultat soulève la question de l'impact des recombinaisons non-homologues dans les émergences de nouvelles pathologies virales.

## II. Article 2

**Title:** Evolutionary history of V3 duplication in HCV 1b strains

**Authors:** H Le Guillou-Guillemette <sup>1,2</sup>, A Pivert <sup>1,2</sup>, E Bouthry <sup>1,2</sup>, C Henquell <sup>3</sup>, O Petsaris <sup>4</sup>, A Ducancelle <sup>1,2</sup>, P Veillon <sup>1,2</sup>, S Vallet <sup>4</sup>, C Payan <sup>4</sup>, F Lunel-Fabiani <sup>1,2</sup> and C Lemaire <sup>5</sup>, on behalf of the AC11 and AC33 ANRS study Groups.

### Author's affiliation

<sup>1</sup> Laboratoire de Virologie, CHU Angers, France

<sup>2</sup> HIFIH Laboratory, UPRES 3859, SFR 4208, LUNAM University, Angers, France

<sup>3</sup> Laboratoire de Virologie, CHU Clermont-Ferrand, France

<sup>4</sup> Département de Bactériologie-Virologie-Hygiène Hospitalière et Parasitologie-Mycologie, CHRU, LUBEM, EA3882, Brest, France

<sup>5</sup> Institut de Recherche en Horticulture et Semences, Université d'Angers, France

**Corresponding author:** Dr Hélène Le Guillou-Guillemette, Laboratoire de Virologie, Département des agents infectieux, IBS, CHU Angers, 4 rue Larrey 49933 Angers cedex9, France. Phone: 33-241354709. Fax: 33-241354164. E-mail: heleguillou@chu-angers.fr

### Key words

Hepatitis C virus, NS5A gene, V3 domain, non-homologous recombination, Bayesian methods

**Competing interest:** No

## Abstract

Emergences of new strain in RNA viruses are mainly due to mutations or intra and inter-genotype homologous recombination. Indeed, non-homologous recombinations may be deleterious and are rarely detected. In a previous French multicenter study, we identified 25 HCV-1b strains bearing two repeated tandemly V3 domains in the NS5A gene without ORF disruption. This polymorphism may be associated with unfavorable evolution of liver disease including a possible involvement in liver carcinogenesis. The present study aims to characterize the origin of these mutant strains and identified on which evolutionary mechanism the V3 duplication relies.

Direct sequencing of the entire NS5A and E1 genes were carried on 27 mutant strains. Quasispecies analyses in consecutive samples were also performed by cloning and sequencing the NS5A gene for all the mutant strains and other wild strains. We analyzed the mutants and wild type sequences polymorphisms using Bayesian methods to infer evolutionary history and molecular mechanism leading to the duplication-like.

We first showed that quasispecies are entirely composed by exclusively mutant's strains or wild type respectively. Moreover we demonstrated that the mutant quasispecies is present since the contamination and persists more than 10 years at least. The evolutionary history of V3 duplication reveals that the two V3 copies of the mutant strains derived from two parental HCV-1b wild type populations and diverged around 100 years ago. This V3 duplication-like is likely resulting from a non-homologous recombination that occurred between strains from these two wild-type populations. A higher liver disease severity due to the infection by this kind of HCV-1b mutants may explain its persistence.

We detected for the first time in chronically infected patients, persistently *in vivo* HCV strains that result from non-homologous recombination. Our study demonstrates the emergence of a new HCV population that may lead to more pejorative liver disease. This result emphasizes the possible consequences of the non-homologous recombination in the emergence of new viral diseases.



## INTRODUCTION

Evolution by genome recombination is widely described in numerous RNA viruses and has been proven to be relevant for epidemiology. Two models of RNA recombination are now described: the most frequently and best characterized model consists in a replicative mechanism with RNA template switching during viral RNA replication [1, 2], and the second one is a non-replicative joining of RNA fragments [3]. Evolution by recombination is associated with escape from immunity or allows a therapeutic escape with the emergence of drug resistant variants as observed in HIV infected patients receiving antiretroviral therapy [4]. Viral recombination could lead to an enhancement of the viral pathogenicity, or to a modified host range, it contributes to the emergence of new viral strain as Western Equine encephalitis virus [5]. Recombination between live attenuated vaccine strains and wild poliovirus had led to the abortion of the poliovirus eradication for instance [6]. However, the *Flaviviridae* family is characterized by a low recombination frequency. Thus, rare naturally HCV recombinant strains have been identified at the inter- and intra-genotypic levels [7, 8] without any change in gene length (i.e. no insertion); none of them described duplicated domains. However, several recent studies suggested that hepatitis C virus (HCV) recombination may be underestimated due to lack of accurate routine screening tools [9, 10]. Interestingly, gain of virulence may occurred by acquisition of new genes or portions of genes during a recombination event. These phenomena represent a selective advantage. They provide an opportunity for species diversification, first as genetic redundancy but then permitting co-evolution of the duplicated genes. This allows for the acquisition of new biological functions or the modification of virulence. Although, viral genomic duplications have been largely described in DNA viruses [11]; it appears infrequent in RNA viruses due to biological constraints [12] even if it remains possible [13]. Indeed, duplication leads to an increase in the genome size; this may lead to new evolutionary opportunities, or conversely the enhancement of defects in production or replication [14, 15]. The few cases of partial gene duplication reported in RNA viruses were located in untranslated regions of *Flaviviridae* [16, 17] or in short intragenic regions. For example, a duplication in the P1 gene of *Potyviridae*, a family of plant viruses, was reported to be the consequence of a former genomic recombination and to largely contribute to their adaptation to a wide range of host species [18]. Regarding the bovine viral diarrhea virus, a duplicated NS3 gene was associated with increased disease severity [19]. Thus, a high mutation rate offers a better way to evolution for

RNA viruses and large insertions or duplications have never been reported in HCV until now excepting in our previous work [20].

HCV infection leads to a persistent infection in around 80% of cases and is a major cause of liver disease with about 180 millions of people infected worldwide [21]. This infection besides a major problem of public health because chronically infected patients are at risk of developing liver cirrhosis and hepatocellular carcinoma [22, 23]. HCV is a single-strand RNA virus belonging to the *Flaviviridae* family. Its genome encodes a polyprotein of about 3000 amino acids which is cleaved by viral and host proteases into ten proteins: structural proteins (Core, E1 and E2, p7) and non-structural proteins (NS2, NS3, NS4A and NS4B, NS5A and NS5B). The pleiotropic nature of NS5A in the viral life cycle of HCV has sparked numerous analyses. Furthermore, clinical trials testing anti-NS5A drugs in combination with other antiviral therapies are bringing even more attention to the protein [24].

The HCV, as other RNA virus such as HIV, is characterized by a huge genetic diversity. Today, 7 genotypes and more than 67 subtypes are described [25]. The high replication level combined with a RNA-dependent RNA polymerase without a proof reading activity lead to the production of closely related but distinguishable viral variants in each infected patient named “quasispecies” [26]. The importance of the quasispecies heterogeneity has been largely suggested to be a way to evade the immune system and the therapeutic pressure. These mutations, insertions and deletions have also been demonstrated to enhance virulence. For example, different clinical outcomes in HIV1 subtype C were associated with this kind of polymorphism in Nef [27]. Variation of secondary structure of 5'-UTR due to indels and mutations might also be correlated to the changes of viral virulence in Enterovirus 71 [28]. Comparison with genome sequences of virulent and attenuated Duck Enteritis Virus (DEV) strains showed several polymorphisms correlated with pathogenic abilities [29]. Clinical relevance of indels in HCV has not been well explored until now. Physiopathological consequences in HCV were mainly characterized by Gerotto et al. They showed that a 1 aa insertion or deletion in HVR1 was significantly associated with mixed cryoglobulinemia type 2 [30]. Indels described in the NS5A domains have been correlated with genotype specific signature and suggested to participate to the evolutionary dynamics, with a putative role in the resistance to antiviral therapy, or to evade the immune host response [31].

In previous studies, our team discovered an HCV-1b strain bearing a duplicated V3 domain in the NS5A gene. NS5A is a phosphoprotein found in a basally phosphorylated form of 56 kDa and a hyperphosphorylated form of 58 kDa. The N-terminal domain contains an amphipathic alpha helix mediated the membrane anchorage [32, 33]. The cytoplasmic portion is divided into three domains separated by low complexity sequences (LCS). Domain I has been X-Ray solved and is a zinc binding domain that may dimerize. Domains II and III are both natively unfolded [34, 35]. Even though NS5A is not yet well-characterized, it is a pleiotropic protein whose several functions are now identified. Thus, studies strongly suggest a potential involvement of the NS5A protein in the resistance to the interferon based-therapy, in the establishment of a chronic hepatitis and in the liver carcinogenesis [36]. NS5A plays also critical roles during the viral life cycle. As recently described, NS5A domain III seems to be involved in the virus particle assembly [37]. The variable domain named V3, was identified by Inshaupe et al in the NS5A domain III (NS5A<sub>2356-2379</sub>) [38] and a correlation between the mutation level in V3 and the response to interferon therapy was demonstrated [39] and then confirmed by our team and others [40, 41]. These data underlined that quasispecies are one way used by RNA virus as a selective advantage.

Here we analyzed the evolutionary history and dynamics of a previously reported set of HCV 1b strains bearing two V3 domains tandemly repeated in the NS5A gene. We had identified a wide duplication event in the NS5A protein of a HCV-1b strain in a large multicenter study in France including more than 800 patients and demonstrated for the first time that this duplication may be associated with a higher risk of liver complications. To understand this polymorphism, we have sequenced the entire NS5A and E1 HCV genes taken from our previous studies [20, 40-42]. We then performed phylogenetic analyses in direct and clonal sequences sets. Our aim was to determine if the second V3 domain arose by duplication event or by another process like non-homologous recombination. Using molecular epidemiology tools, we compared mutant strains with wild type within and between hosts.

We showed that the extra-copy of the V3 domain in NS5A has not arose by simple duplication. More likely a double recombination event involving another HCV1-b may be the cause of the doubling of the V3 domain. Unfortunately, we do not identify any strain carrying a V3 domain similar to the one found in the new type of NS5A protein. At last relationships between higher mutation rate and severity in hepatitis are discussed.

## MATERIAL AND METHODS

### *Samples*

In the present study, we used previous set of samples:

-a set of NS5A quasispecies sequences taken from HCV 1b infected patients included in the NAIF study, with sequential samples from 8 patients (7 had a wild NS5A gene and 1 have a duplicated V3 gene (noted NS5A-dup in the following)) [40,41]

-a set of direct NS5A sequences taken from the 25 patients that have been identified to be infected by a V3 duplicated HCV 1b in our French multicenter study [20].

We have expanded our analyses using our French multicenter cohort: direct NS5A and E1 gene sequences was studied in 136 HCV strains (all the strains with the V3 duplication and a mean of ten strains taken in each center where a mutant strain was identified; Table 1); quasispecies study was also performed on the 25 newly mutant strains identified with a sequential analysis in three strains.

We also download all the complete NS5A genes from all HCV genotypes available from the Los Alamos HCV Sequence Database [42].

### *Experimental procedures*

#### *Direct NS5A sequences and NS5A quasispecies study*

RNA extraction, entire NS5A gene reverse transcription, amplification, cloning and sequencing, were explained in detail elsewhere [41]. Briefly, HCV RNA was extracted from 200 µL of each serum sample using the EasyMag automated extraction system (BioMérieux, Craponne, France). A full length NS5A gene amplification was performed using outer primers E1 and E2 for the RT-PCR (1665 bp) and inner primers I3 and I41b (1344 bp) for the nested PCR. NS5A PCR products were subjected to direct sequencing using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on the automated ABI3130xl, with primers previously used in the nested-PCR rounds.

We obtained 1120 direct NS5A sequences and we obtained a mean of 30 clones per samples (119 wild clones taken from 19 serums; 1001 mutant clones taken from 36 serums). HCV sequences obtained in this study have been deposited in Genbank with accession numbers XX to XX.

### *E1 gene amplification and direct nucleotide sequencing*

The E1 genomic segment (922 bp) was amplified with a two-step RT-PCR using an in-house protocol kindly transmitted by Dr C. Henquell. Briefly, a first mixture containing 10 $\mu$ L of HCV RNA previously extracted, 0,4 $\mu$ M of random primers (LifeTechnologies) and 0,5mM of each dNTP (Life Technologies) was activated for 5min at 65°C followed by 15sec at 25°C. A second mixture containing 200U of SuperScript®II Reverse Transcriptase (Invitrogen), 40U of RiboLock™ RNase Inhibitor (Fermentas), 1X of reverse transcriptase buffer and 8m M of DTT for a 10 $\mu$ L final volume was added to the previous reaction tube. Reverse transcription was carried out for 10min at 25°C and 50min at 50°C. For PCR amplification, a 10 $\mu$ L aliquot of cDNA was then added to a 50 $\mu$ L final reaction volume. The reaction mixture contained 0,4 $\mu$ M of Env-S (5'-TGG GYA ARG TCA TCG ATA CC-3') and Env-AS (5'-GGC 6GT 6CK RTT 6AT RTG CC-3', 6=inosin), forward and reverse primers respectively, 320 $\mu$ M of each dNTP, 1Xof polymerase buffer and 1,25U of *Taq* DNA polymerase (Qiagen). The PCR was then performed with a DNA denaturation for 2min at 95°C, followed by 45 cycles of denaturation (94°C for 15sec), annealing (49°C for 50sec) and elongation (72°C for 50sec) and a final elongation for 5min at 72°C. Twenty-five samples bearing a V3 insertion and ten samples originating from each laboratory, in which an insertion was identified, were studied (146 samples).

E1 PCR products were subjected to direct sequencing using the Big Dye Terminators v3.1 Cycle Sequencing kit (LifeTechnologies) on the automated ABI3130xl, with forward and reverse primers used in the previous two-PCR round. The E1 nucleotide sequences were assembled into a consensus sequence and then aligned to the HCVJ-1b prototype sequence (SeqScape software).

### *Sequences analysis*

1120 sequences of the quasispecies NS5A gene were aligned according to their amino-acid sequence using the Muscle [43] alignment algorithm implemented in the MEGA v6.0.5 software [44]. Prior to any inference the more likely substitution model was tested using jModelTest 2 [45]. The best model was found to be the HKY substitution model [46] and gamma heterogeneity and a proportion of invariant sites. This substitution model was used when useful in all subsequent analyses. All genealogical and phylogenetic relationships between strains on NS5A and E1 genes, were conducted using the models implemented in the BEAST v1.8.2 software [47]. For each kind of analysis, we used empirical base frequencies which permit to not overparametrize the models.

Each analysis using BEASTv1.8 was run using both a strict and uncorrelated lognormal relaxed clock [48]. The number of iterations and sampling intervals were tuned for each analysis in order to have good ESS values ( $>200$ ). Thus number of iterations varied from  $2 \times 10^7$  to  $4 \times 10^7$  with sampling intervals varying from 1,000 to 10,000. Estimated trees were analyzed using the TreeAnnotator software contained in the BEAST package using a burn in a step of 1000 and a posterior probability of 0.9. Several datasets were analyzed: intra-hosts analyses for NS5A carrying duplication or not, complete NS5A and E1 genes among wild and duplicated strains, NS5A V3 domains and NS5A without V3.

### *Intrahost dynamics*

Clones of NS5A were sequenced in 55 quasi-species. First, we looked for the occurrence of the duplication of V3 domain in order to evaluate the homogeneity of NS5A types within a quasi-species. In fact, quasi-species only composed of NS5A-dup strains would favor the hypothesis of a clearly differentiated population over a recurrent mutational effect in the wild-type NS5A locus. We inferred the demographic dynamics of each quasispecies using the Bayesian Skyline plot [49] model implemented in BEAST. This model permits to plot the variation of effective size against evolutionary times given here in coalescent units (2 times the effective size  $N_e$ ). Temporal samplings of a same quasispecies were grouped, decreasing the number of wild type from 16 to 9.

### *Differences in evolutionary rates between wild and NS5A-dup loci*

To test for differences in the rates of evolution between NS5A-dup and wild ones, we performed our analyses under the Hierarchical Phylogenetic Model [50] using three strains with duplicated V3 domain and seven wild. HPM incorporates fixed effects to test for differences in evolutionary dynamics across host populations in a formal statistical framework employing stochastic search variable selection and model averaging [51]. Compared to the classical strategy of independent analyses of sequences from each patient, the HPM is better to estimate evolutionary parameters such as nucleotide substitution rates. For the ten HCV strains a temporal sampling was available, that allows for estimation of mutation rate of the locus [52] at the quasispecies level. According to hosts the heterochronous sampling covers periods of several months (AG5sf) to 12 years (AG2) (see Table 1).

Statistical inferences of mutation rates as well as of demography were performed using the Extended Bayesian Skyline Plot (EBSP) model under the HPM framework. This model

assumes population fluctuation, a common situation in viruses; especially in humans where treatments can strongly influence the HCV's demography.

#### *Origins of NS5A-dup and the duplicated V3 domain*

The origins of the NS5A-dup were inferred from the analysis of their phylogenetic relationships and divergence time with wild strains. We used direct sequencing of NS5A locus on the 25 patients of the French multicenter study where the V3 duplication was found plus the 70 patients carrying the wild type NS5A strains. We used first the complete sequences of NS5A locus and then without the V3 domain(s). For comparison purposes, the same analysis was performed for the E1 locus.

A second set of phylogenetic analyses were performed only on the V3 domain. For the NS5A-dup locus, we compared the first copy of V3, R1, with the second R2 treating them as different loci. We then tested the hypothesis of a separate origin of the extra-copy of V3 in NS5A-dup rather than a duplication.

The hypothesis of an origin of the supplementary V3 domain from out of France was tested using 204 sequences of this locus in the Los Alamos database [53]. We performed all the analyses described in this section using the constant size coalescent model with random relaxed lognormal clock rate which was found to fit better our data.

## RESULTS

### *Intrahost dynamics*

Among the 55 quasispecies studied, we found no patient carrying a mix of NS5A-dup and wild type at any time. This result confirms the precedent recent clinical study [20] where patients were found to be infected by only one type of HCV strain, harbouring NS5A-dup or not. However certain strains do not have two complete domains (RE1, LM1 and TR3). They can have one complete domain and a piece of the second one, or in the case of TR3 have the 5-prime part of the first V3 domain and the 3-prime of the second domain.

Estimates of demographic dynamics using the Bayesian skyline model indicate that NS5A-dup strains have a significantly greater demographic growth than wild ones (Mann-Whitney U,  $p=0.001$ ) (Figure 1). Bayesian skyline plots presented different shapes among the 23 NS5a-dup and the 7 wild strains. For every strain, demographic dynamics seemed to reach a plateau. However if clear growth dynamics could have been evidenced for the NS5A-dup strains, wild types rather shown constant population sizes.

### *NS5A-dup loci evolve faster than wild ones*

The HPM procedure indicated that the NS5A-dup haplotypes showed a greater mean mutation rate ( $5.74E-4 \pm 7.2E-5$ ) than the wild ones ( $3.94E-4 \pm 8.2E-5$ ) (Mann-Whitney U,  $p < 0.01$ ) (Figure 2). In the same way, wild type NS5A quasispecies showed a greater tree height ( $42.71 \pm 5.4$ ) than NS5A-dup ones ( $33.05 \pm 2.4$ ) (Mann-Whitney U,  $p < 0.01$ ).

### *Origin of the V3 extra-copy*

First figure 3 indicates that the strains carrying NS5A-dup locus evolved separately from the wild ones. Because of the difference induced by the presence of an extra-domain V3 in NS5A-dup viruses, we performed the same analysis using sequences without the V3 domain. Again we found the same result (Figure 4), indicating that the clustering is really due to real divergence between strains. It appears that strains carrying NS5A-dup derived from the wild type about an hundred years ago. At last the phylogeny of the locus E1 is congruent with the NS5A ones (Supp. Info 1).

We analyzed the V3 domain alone. In NS5A-dup, we called R1 the first copy and R2 the second. The figure 5 shows that the copy R1 appears to be the most divergent one. R2 copies are the most related to the wild ones. This result is at odds with the hypothesis of a recent apparition of the R1 V3 domain by duplication of the R2 one. Conversely we claim that the first V3 domain in the NS5A-dup locus has another origin. Here again we show that the R1-cluster likely diverged from others about a century ago. Although R2 copies cluster with wild type domain, they form a distinct clade, confirming an independent evolution.

Finally, when performing the analysis with the world database of Los Alamos, we found that the V3 domain R1 likely came from a HCV1-b genotype (data not shown). However no V3 domain clearly clustered with the R1 copy (Figure 6). At this global scale, we confirm the common origin of R2 copies with the wild type of V3 domain.



## DISCUSSION

Emergence of new pathogens is always of high interest, especially when it concerns human health. The biological mechanisms of such emergences are poorly known. Especially the role of recombination in the creation of new viral populations is often underestimated as it needs co-infection events to occur. Less described is the non-homologous recombination mechanism that can occur within or between genotypes of viruses. We studied here the evolutionary dynamics of new HCV1-b strains carrying two V3 domains instead of one at the gene encoding the NS5A protein. Using clones sequences of quasispecies, direct sequencing as well as heterochronous sampling, we showed that the new V3 domain merely arose from non-homologous recombination events involving an unknown HCV1-b strain rather than by duplication. The population of NS5A-dup strains then probably derived from the wild ones a century ago. Finally we will discuss how higher evolutionary rates in the protein NS5A with two V3 domains may influence the severity of the hepatitis C in humans as recently proposed by us.

### *Origins of the NS5A-dup strains*

The phylogenetics analyses were all congruent whatever the dataset used : on NS5A including or not the V3 domain as well as on the E1 gene, all reached the same conclusion that the NS5A-dup strains diverged about a century ago from the wild type HCV1-b viruses. Higher tree heights estimated on direct sequences of wild type should confirm this result. Thus NS5A-dup HCV1-b strains should constitute a real emergence of a new family of HCV1-b. In fact the clear-cut clustering between wild and NS5A-dup strains would suggest a reproductive isolation between these two families. In fact, if mixing occurred we would not observe any clear clustering between NS5A-dup and wild ones. However we could question the ability of recombination between strains from different populations in HCV. Recombination represents the process by which genetic material is exchanged between strains. Thus the likelihood of such recombination events here suggests multiple infections from different donors, allowing the mixing of several populations of strains within a patient. In our dataset, we did not detect any mixing of populations within any patient. As a consequence there is little chance for recombination between the two families of HCV1-b to occur in a short period of several decades. Then, in the absence of recombination event between these strains, we could question the likelihood of reproductive isolation within HCV1-b. Is recombination lacking because of its rarity in such viruses [54, 55] or is it effective but deleterious because of

deleterious phenotypic effect such as incompatibilities [56, 57]? Thus, experimental recombination assays *in vitro* should permit to conclude to a real reproductive isolation.

More interesting are the phylogenetics of the V3 domain copies themselves. In fact, we firstly found that the first copy of V3 does not cluster with the second one. In the case of duplication one would expect a common origin for the two copies of a domain within a gene. Our results are at odds with this hypothesis. Indeed, the R1 copy of V3 comes from another population of HCV1-b as suggested by figures 5 and 6 when the R2 copy rather derived from the V3 domain observed in all the wild type NS5A genes. The comparison with world database of Los Alamos did not permit to clearly identify any potential ancestor for the R1 copy of V3. However comparisons with other genotypes of HCV, clearly shows that the R1 copy clusters with HCV1-b genotypes.

The question of the mechanism by which the R1 was added to the 5-prime side of the original V3 domain remains addressed. We claim here that this domain may have been introduced by non-homologous recombination between two divergent HCV1-b strains. Non-homologous recombination in HCV was recently described by Scheel et al. [55]. Although these authors demonstrated the potential for non-homologous recombination in RNA viruses, they described the mechanism in cell cultures. To our knowledge, we report here the first occurrence of a viable non-homologous recombination in a RNA virus that leads to the emergence of a new family of HCV1-b since a century. When homologous recombination, even rare, has been recognized to involve positive epistasis and produce recombinant with increased fitness [58], non-homologous recombination is often deleterious and likely to produce non-viable genotypes. In their experiment, Scheel et al. [55] observed a progressive deletion of extra-copies of genomes introduced by non-homologous. We also observed the occurrence of incomplete V3 domains in three direct sequences of NS5A-dup strains. Such a partial deletion could suggest that duplicated V3 domain should be slightly deleterious. However in their *in vitro* experiment Scheel et al. [55] showed that only a few generations are required to eliminate the extra-copies. We do not observe any quick elimination of extra V3 domain as they seemed to have persisted for about a century. The difference with Scheel et al. work [55] is that this duplication has experienced “real life” in human organisms, and more recently, faced clinical treatments. We could suggest than far from being deleterious, the supplementary V3 domain may provide an advantage in epidemiology of HCV1-b. This hypothesis is supported the observation of a better demographic dynamics (Figure 1) and intrahost mutation rates (Figure 2) for NS5A-dup. Our results showed here that the introduction of an extra-copy of a V3 domain did not alter the fitness of HCV strains.

Conversely, recent results suggest that such strains may have more negative effects on the patients than their wild counterparts [20].

#### *Implications for viral emergence*

NS5A-dup gene showed greater evolutionary rates than the wild locus. HPM framework permitted to get a good estimate of mutation rate for each quasispecies. This framework has proven its power in comparison of intrahost dynamics in HIV [51]. Such an analysis gives more accurate results than working on each quasispecies independently.

The difference in evolutionary rates between NS5A-dup and wild loci has no consequence on the demographic dynamics of the different families. Indeed, no differences were observed in the demographic curves estimated under the Bayesian Skyline Model. Clinical studies did not observe any difference in the viral load levels between patient respectively infected by the NS5A-dup and wild strains (data not shown).

It appears then that if carrying two V3 domains may not confer any net advantage in the intrahost population growth rate, it does not seem to have any deleterious effect. Moreover, NS5A-dup strains are found in several patients, and then should harbor all the transmission abilities, as shown by the identification of this HCV strain since the contamination (occurring by a kidney graft) and its persistence for at least 12 years. Their persistence in all the patients infected at different epochs claims again in favor of the absence of any deleterious effect of the extra-copy of V3 domain. In other words the acquisition of a supplementary V3 domain should be at least considered as selectively neutral if not positive.

#### CONCLUSION

Here we showed for the first time the impact of non-homologous recombination on the emergence of a new family of strains of HCV1-b. Recombination, although rare in viruses, could play an important role in the potential of emergence. However, underestimation of recombination rate in HCV should probably relate to methods of clinical diagnostics, which most of times focus on a single small region and not on large genomic portions. Here we have detected a non-homologous recombination event which is supposed to be rare than homologous one. This finding could have strong consequences on the epidemiology of such HCV. Indeed, as we evidenced, NS5a-dup strains may form a distinct clade, supposing a recent reproductive isolation with the wild strains. As this new population should have its own dynamics, considering of HCV1-b as a single population entity for epidemiology could therefore be source of bias. In other words, does NS5A-dup still belong to HCV1-b genotype

or is it a new emergent genotype (HCV1b cluster)? At last we could question the occurrence of such non-homologous recombination events in other viruses for which recombination is assumed to be more likely.

## REFERENCES

1. Simon-Loriere E and Holmes EC (2011) Why do RNA viruses recombine? *Nat Rev Microbiol* 9: 617-626.
2. Kirkegaard K and Baltimore D (1986) The mechanism of RNA recombination in poliovirus. *Cell* 47: 433-443.
3. Gmyl AP, Korshenko SA, Belousov EV, Khitrina EV and Agol VI (2003) Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces? *RNA* 9: 1221-1231.
4. Nora T, Charpentier C, Tenaillon O, Hoede C, Clavel F, et al. (2007) Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J Virol* 81: 7620-7628.
5. Weaver SC, Kang W, Shirako Y, Rumenapf T, Strauss EG, et al. (1997) Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses. *J Virol* 71: 613-623.
6. Kew O, Morris-Glasgow V, Landaverde M, Burns C, Shaw J, et al. (2002) Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. *Science* 296: 356-359.
7. Kalinina O, Norder H, Mukomolov S and Magnius LO (2002) A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol* 76: 4034-4043.
8. Gonzalez-Candelas F, Lopez-Labrador FX and Bracho MA (2011) Recombination in hepatitis C virus. *Viruses* 3: 2006-2024.
9. Morel V, Fournier C, Francois C, Brochot E, Helle F, et al. (2011) Genetic recombination of the hepatitis C virus: clinical implications. *J Viral Hepat* 18: 77-83.
10. Mes TH and van Doornum GJ (2011) Recombination in hepatitis C virus genotype 1 evaluated by phylogenetic and population-genetic methods. *J Gen Virol* 92: 279-286.
11. Shackelton LA and Holmes EC (2004) The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol* 12: 458-465.
12. Holmes EC (2003) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol* 11: 543-546.
13. Simon-Loriere E and Holmes EC (2013) Gene Duplication Is Infrequent in the Recent Evolutionary History of RNA Viruses. *Mol Biol Evol*.
14. Moya A, Holmes EC and Gonzalez-Candelas F (2004) The population genetics and evolutionary epidemiology of RNA viruses. *Nat Rev Microbiol* 2: 279-288.
15. Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, et al. (2007) Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol Biol Evol* 24: 845-852.
16. Gritsun TS and Gould EA (2006) The 3' untranslated region of tick-borne flaviviruses originated by the duplication of long repeat sequences within the open reading frame. *Virology* 354: 217-223.
17. Gritsun TS and Gould EA (2006) Direct repeats in the 3' untranslated regions of mosquito-borne flaviviruses: possible implications for virus transmission. *J Gen Virol* 87: 3297-3305.
18. Valli A, Lopez-Moya JJ and Garcia JA (2007) Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J Gen Virol* 88: 1016-1028.
19. Nagai M, Sakoda Y, Mori M, Hayashi M, Kida H, et al. (2003) Insertion of cellular sequence and RNA recombination in the structural protein coding region of cytopathogenic bovine viral diarrhoea virus. *J Gen Virol* 84: 447-452.
20. Le Guillou-Guillemette H, Pivert A, Bertrais S, Veillon P, Ducancelle A, et al. (2014) Identification of a duplicated V3 domain in NS5A associated with cirrhosis and hepatocellular carcinoma in HCV-1b patients. (under submission).
21. Mohd Hanafiah K, Groeger J, Flaxman AD and Wiersma ST (2013) Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology* 57: 1333-1342.
22. Omland LH, Farkas DK, Jepsen P, Obel N and Pedersen L (2010) Hepatitis C virus infection and risk of cancer: a population-based cohort study. *Clin Epidemiol* 2: 179-186.

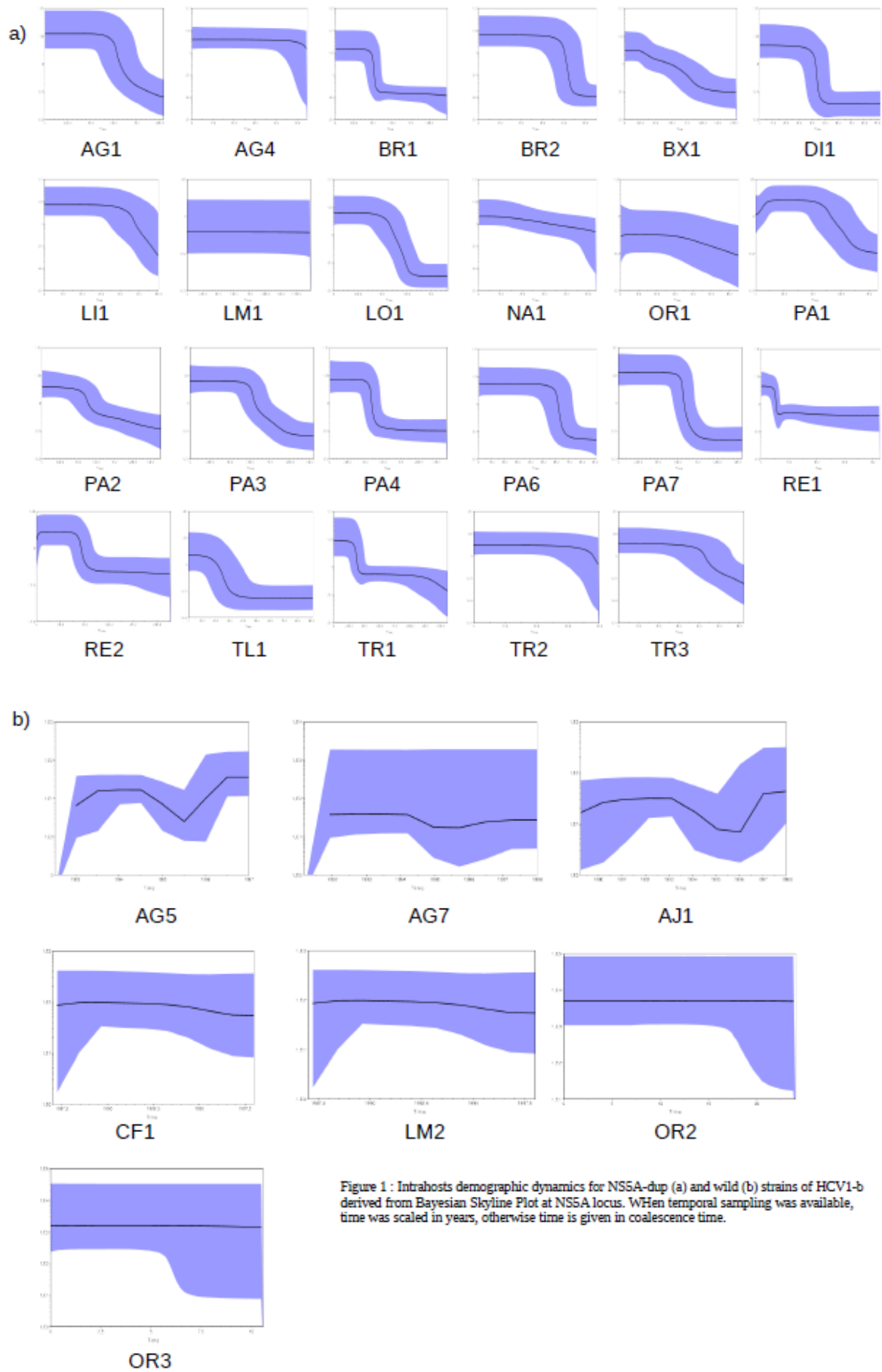
23. Perz JF, Armstrong GL, Farrington LA, Hutin YJ and Bell BP (2006) The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol* 45: 529-538.
24. Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, et al. (2010) Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. *Nature* 465: 96-100.
25. Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, et al. (2014) Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology* 59: 318-327.
26. Pawlotsky JM (2006) Hepatitis C virus population dynamics during infection. *Curr Top Microbiol Immunol* 299: 261-284.
27. Walker PR, Ketunuti M, Choge IA, Meyers T, Gray G, et al. (2007) Polymorphisms in Nef associated with different clinical outcomes in HIV type 1 subtype C-infected children. *AIDS Res Hum Retroviruses* 23: 204-215.
28. Wen HL, Si LY, Yuan XJ, Hao SB, Gao F, et al. (2013) Complete genome sequencing and analysis of six enterovirus 71 strains with different clinical phenotypes. *Virol J* 10: 115.
29. Wang J, Hoper D, Beer M and Osterrieder N (2011) Complete genome sequence of virulent duck enteritis virus (DEV) strain 2085 and comparison with genome sequences of virulent and attenuated DEV strains. *Virus Res* 160: 316-325.
30. Gerotto M, Dal Pero F, Loffreda S, Bianchi FB, Alberti A, et al. (2001) A 385 insertion in the hypervariable region 1 of hepatitis C virus E2 envelope protein is found in some patients with mixed cryoglobulinemia type 2. *Blood* 98: 2657-2663.
31. Cuevas JM, Gonzalez M, Torres-Puente M, Jimenez-Hernandez N, Bracho MA, et al. (2009) The role of positive selection in hepatitis C virus. *Infect Genet Evol* 9: 860-866.
32. Penin F, Brass V, Appel N, Ramboarina S, Montserret R, et al. (2004) Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J Biol Chem* 279: 40835-40843.
33. Moradpour D, Brass V and Penin F (2005) Function follows form: the structure of the N-terminal domain of HCV NS5A. *Hepatology* 42: 732-735.
34. Hanouille X, Badillo A, Verdegem D, Penin F and Lippens G (2010) The domain 2 of the HCV NS5A protein is intrinsically unstructured. *Protein Pept Lett* 17: 1012-1018.
35. Hanouille X, Badillo A, Wieruszkeski JM, Verdegem D, Landrieu I, et al. (2009) Hepatitis C virus NS5A protein is a substrate for the peptidyl-prolyl cis/trans isomerase activity of cyclophilins A and B. *J Biol Chem* 284: 13589-13601.
36. Macdonald A and Harris M (2004) Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol* 85: 2485-2502.
37. Appel N, Zayas M, Miller S, Krijnse-Locker J, Schaller T, et al. (2008) Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog* 4: e1000035.
38. Inchauspe G, Zebedee S, Lee DH, Sugitani M, Nasoff M, et al. (1991) Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates. *Proc Natl Acad Sci U S A* 88: 10292-10296.
39. Duverlie G, Khorsi H, Castelain S, Jaillon O, Izopet J, et al. (1998) Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity. *J Gen Virol* 79 ( Pt 6): 1373-1381.
40. Veillon P, Payan C, Gaudy C, Goudeau A and Lunel F (2004) [Mutation analysis of ISDR and V3 domains of hepatitis C virus NS5A region before interferon therapy with or without ribavirin]. *Pathol Biol (Paris)* 52: 505-510.
41. Veillon P, Payan C, Le Guillou-Guillemette H, Gaudy C and Lunel F (2007) Quasispecies evolution in NS5A region of hepatitis C virus genotype 1b during interferon or combined interferon-ribavirin therapy. *World J Gastroenterol* 13: 1195-1203.
42. Lunel-Fabiani F VP, Le Guillou-Guillemette H, Redureau J, Coeffic B, Subra JF, Causse X, Dubois P, Blanchi A, Gaudy C, Payan C (2005) Duplication of the V3 domain in Nonstructural 5A Protein related to Hepatitis C Virus extrahepatic manifestations. In: Abstracts, editor editors. 45th Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC). Washington, DC, USA: American Society for Microbiology. pp. 485.

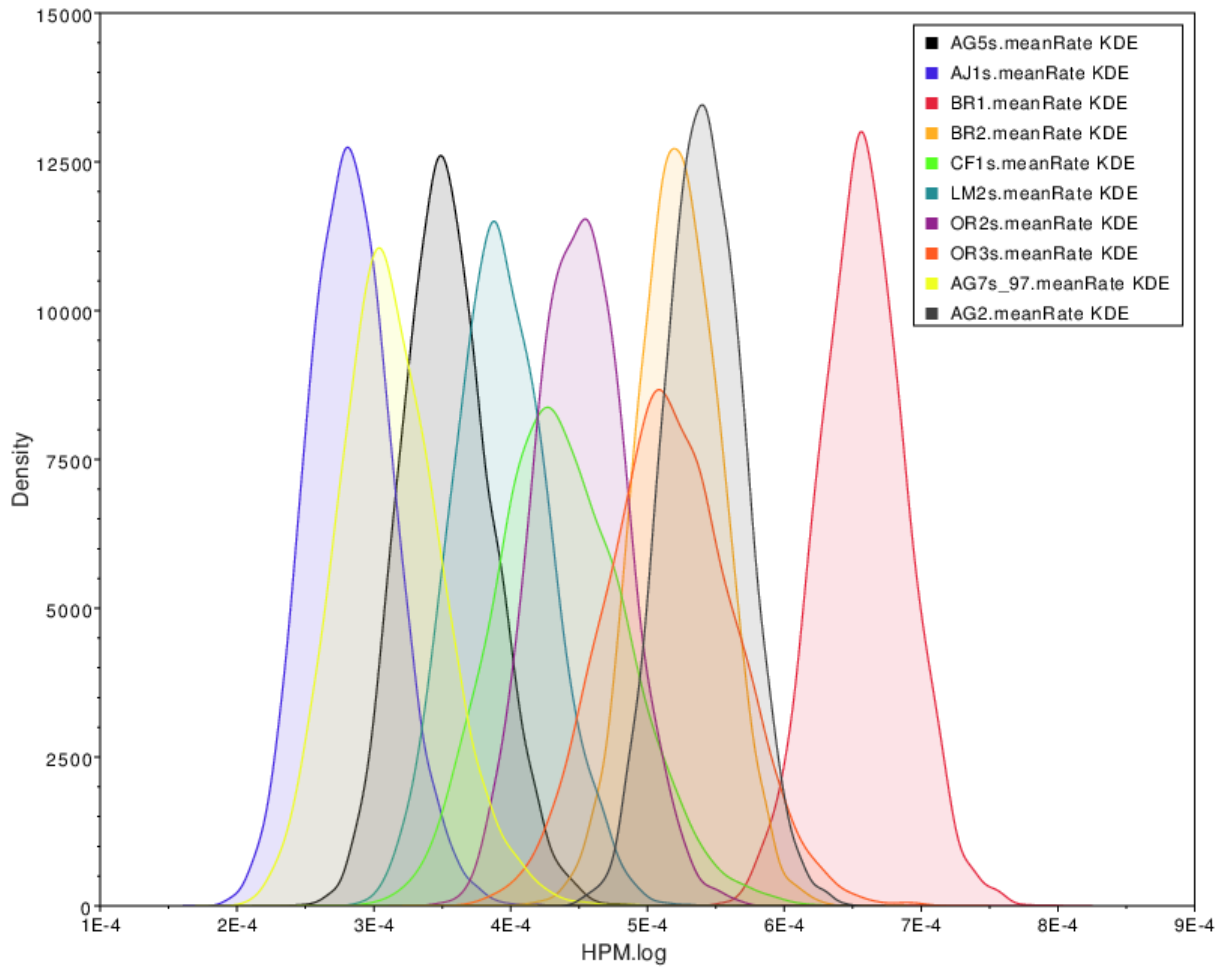
43. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
44. Tamura K, Stecher G, Peterson D, Filipinski A and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725-2729.
45. Darriba D, Taboada GL, Doallo R and Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9: 772.
46. Hasegawa M, Kishino H and Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22: 160-174.
47. Drummond AJ, Suchard MA, Xie D and Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29: 1969-1973.
48. Drummond AJ, Ho SY, Phillips MJ and Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4: e88.
49. Drummond AJ, Rambaut A, Shapiro B and Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22: 1185-1192.
50. Suchard MA, Kitchen CM, Sinsheimer JS and Weiss RE (2003) Hierarchical phylogenetic models for analyzing multipartite sequence data. *Syst Biol* 52: 649-664.
51. Edo-Matas D, Lemey P, Tom JA, Serna-Bolea C, van den Blink AE, et al. (2011) Impact of CCR5delta32 host genetic background and disease progression on HIV-1 intrahost evolutionary processes: efficient hypothesis testing through hierarchical phylogenetic models. *Mol Biol Evol* 28: 1605-1616.
52. Drummond AJ, Pybus OG, Rambaut A, Forsberg R and Rodrigo AG (2003) Measurably evolving populations. *TRENDS in Ecology and Evolution* 18: 481-488.
53. Kuiken C, Yusim K, Boykin L and Richardson R (2005) The Los Alamos hepatitis C sequence database. *Bioinformatics* 21: 379-384.
54. Simmonds P (2006) Recombination and selection in the evolution of picornaviruses and other Mammalian positive-stranded RNA viruses. *J Virol* 80: 11124-11140.
55. Scheel TK, Galli A, Li YP, Mikkelsen LS, Gottwein JM, et al. (2013) Productive homologous and non-homologous recombination of hepatitis C virus in cell culture. *PLoS Pathog* 9: e1003228.
56. Sanjuan R, Moya A and Elena SF (2004) The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A* 101: 8396-8401.
57. Carrasco P, de la Iglesia F and Elena SF (2007) Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco Etch virus. *J Virol* 81: 12979-12984.
58. Vuillaume F, Thebaud G, Urbino C, Forfert N, Granier M, et al. (2011) Distribution of the phenotypic effects of random homologous recombination between two virus species. *PLoS Pathog* 7: e1002028.

**Table 1:** Description of the strains of HCV1-b. For each strain the locality and date of sampling are given. Sequences of NS5A harbouring two V3 domains are indicated in bold and italic, the others belong to wild type. Direct sequences are given in a) and strains studied for quasispecies analysis are given in b).

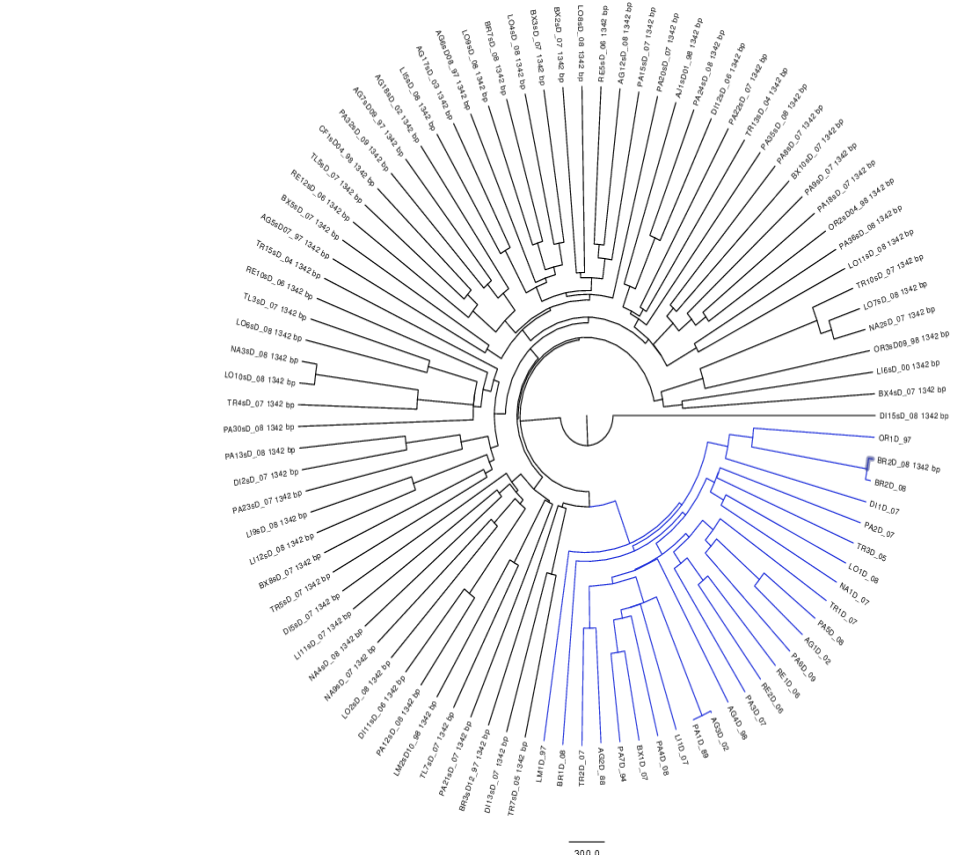
a)	Locality	Code	Date	b)	Locality	Code	Date	Quasispecies	n
	Angers	AG10s	04/04/2006		Ajaccio	AJ1s	24/09/1997	O	16
		AG12s	16/06/2008			AJ1s	25/01/1998	O	13
		AG13s	03/02/2009		<b><i>AG1</i></b>	<b><i>01/08/2002</i></b>	<b><i>O</i></b>	26	
		AG8s	23/01/2001		<b><i>AG2</i></b>	<b><i>18/02/2000</i></b>	<b><i>O</i></b>	32	
		AG9s	01/09/2004		<b><i>AG2</i></b>	<b><i>??/??/1998</i></b>	<b><i>O</i></b>	24	
		AG11s	01/02/1999		<b><i>AG2</i></b>	<b><i>??/??/1996</i></b>	<b><i>O</i></b>	26	
		<b><i>AG14s</i></b>	<b><i>01/01/2001</i></b>		<b><i>AG2</i></b>	<b><i>??/??/1988</i></b>	<b><i>O</i></b>	30	
		<b><i>AG15s</i></b>	<b><i>01/05/2004</i></b>		<b><i>AG3</i></b>	<b><i>01/01/2002</i></b>	<b><i>O</i></b>	32	
		<b><i>AG16s</i></b>	<b><i>12/03/2007</i></b>		<b><i>AG4</i></b>	<b><i>29/06/1998</i></b>	<b><i>O</i></b>	28	
		<b><i>AG17s</i></b>	<b><i>01/08/2003</i></b>		AG5s	09/05/1997	O	15	
		AG18s	01/11/2002		AG5s	29/07/1997	O	17	
		AG19s	08/11/2011		AG6s	29/09/1997	O	15	
	Bordeaux	BX2s	30/01/2007		AG7s	16/06/1997	O	17	
		BX3s	23/05/2007		AG7s	16/09/1997	O	6	
		BX4s	11/06/2007		<b><i>BX1</i></b>	<b><i>25/06/2007</i></b>	<b><i>O</i></b>	29	
		BX5s	19/09/2007		<b><i>BR1</i></b>	<b><i>11/07/2008</i></b>	<b><i>O</i></b>	35	
		BX6s	05/03/2007		<b><i>BR1</i></b>	<b><i>??/??/2007</i></b>	<b><i>O</i></b>	38	
		BX7s	09/08/2007		<b><i>BR1</i></b>	<b><i>??/??/2006</i></b>	<b><i>O</i></b>	37	
		BX8s	17/07/2007		<b><i>BR1</i></b>	<b><i>??/??/2007</i></b>	<b><i>O</i></b>	45	
		BX9s	23/05/2007		<b><i>BR1</i></b>	<b><i>??/??/2004</i></b>	<b><i>O</i></b>	37	
		BX10s	15/10/2007		<b><i>BR2</i></b>	<b><i>03/03/2008</i></b>	<b><i>O</i></b>	29	
		BX11s	05/03/2007		<b><i>BR2</i></b>	<b><i>??/??/2006</i></b>	<b><i>O</i></b>	20	
		BX12s	23/05/2007		<b><i>BR2</i></b>	<b><i>??/??/2004</i></b>	<b><i>O</i></b>	13	
		Brest	BR4s	30/06/2005		<b><i>BR2</i></b>	<b><i>??/??/2002</i></b>	<b><i>O</i></b>	27
	BR5s		26/01/2004		BR3s	28/01/1998	O	13	
	BR6s		05/01/2009		CF1s	17/09/1997	O	17	
	BR7s		24/09/2008		CF1s	23/01/1998	O	10	
	BR8s		18/01/2006		CF1s	24/04/1998	O	4	
	BR9s		15/02/2005		<b><i>DI1</i></b>	<b><i>17/01/2007</i></b>	<b><i>O</i></b>	30	
	BR10s		26/04/2007		<b><i>LM1</i></b>	<b><i>14/03/1997</i></b>	<b><i>O</i></b>	7	
	BR11s		31/01/2006		LM2s	13/03/1998	O	17	
	BR12s		13/03/2003		LM2s	24/07/1998	O	12	
	PA29s		13/06/2008		LM2s	14/10/1998	O	11	
	PA30s		04/04/2008		<b><i>LI1</i></b>	<b><i>17/01/2007</i></b>	<b><i>O</i></b>	30	
	Paris		PA31s	11/03/2009		<b><i>LO1</i></b>	<b><i>13/05/2008</i></b>	<b><i>O</i></b>	33
		PA32s	11/06/2009		<b><i>NA1</i></b>	<b><i>15/05/2007</i></b>	<b><i>O</i></b>	32	
		PA33s	20/12/2007		<b><i>OR1</i></b>	<b><i>10/12/1997</i></b>	<b><i>O</i></b>	8	
		PA34s	15/05/2008		OR2s	28/10/1997	O	14	
		PA35s	02/04/2008		OR2s	21/01/1998	O	11	
		PA36s	29/02/2008		OR2s	15/04/1998	O	12	
		PA37s	18/05/2009		OR3s	07/07/1998	O	15	
		PA38s	18/09/2008		OR3s	28/09/1998	O	15	
		DI2s	10/10/2007		<b><i>PA1</i></b>	<b><i>28/11/1989</i></b>	<b><i>O</i></b>	29	
		DI3s	04/07/2007		<b><i>PA2</i></b>	<b><i>02/04/2007</i></b>	<b><i>O</i></b>	31	
		DI4s	19/01/2007		<b><i>PA3</i></b>	<b><i>28/12/2007</i></b>	<b><i>O</i></b>	31	
		DI5s	21/05/2007		<b><i>PA4</i></b>	<b><i>07/03/2008</i></b>	<b><i>O</i></b>	29	
	DI6s	06/03/2008		<b><i>PA5</i></b>	<b><i>05/03/2008</i></b>	<b><i>O</i></b>	31		
	DI7s	08/01/2008		<b><i>PA6</i></b>	<b><i>25/05/2009</i></b>	<b><i>O</i></b>	27		
	DI8s	03/09/2007		<b><i>PA7</i></b>	<b><i>19/09/1994</i></b>	<b><i>O</i></b>	32		
	DI9s	26/09/2007		<b><i>RE1</i></b>	<b><i>14/10/2008</i></b>	<b><i>O</i></b>	39		
	DI10s	09/11/2006		<b><i>RE2</i></b>	<b><i>17/12/2009</i></b>	<b><i>O</i></b>	21		
	DI11s	21/11/2006		<b><i>TL1</i></b>	<b><i>12/03/2007</i></b>	<b><i>O</i></b>	31		
	DI12s	15/12/2006		<b><i>TR1</i></b>	<b><i>19/02/2007</i></b>	<b><i>O</i></b>	33		
	DI13s	04/07/2007		<b><i>TR2</i></b>	<b><i>30/10/2007</i></b>	<b><i>O</i></b>	29		
	DI14s	04/03/2008		<b><i>TR3</i></b>	<b><i>01/01/2005</i></b>	<b><i>O</i></b>	27		
	DI15s	17/01/2008							
	Limoges	LI2s	09/01/2008						
		LI4s	05/05/2008						
		LI5s	07/01/2008						
		LI9s	31/03/2008						
		LI10s	20/09/2006						
		LI11s	03/09/2007						
		LI12s	18/03/2008						
		LI3s	02/09/2003						
		LI6s	04/05/2000						
		LI7s	05/02/2001						
		LI8s	12/12/2003						
		LO2s	08/07/2008						
	Lyon	LO3s	06/06/2008						
		LO4s	19/08/2008						
		LO5s	04/04/2008						
		LO6s	26/08/2008						
		LO7s	03/06/2008						
		LO8s	08/07/2008						
		LO9s	02/07/2008						
		LO10s	01/09/2008						
		LO11s	14/04/2008						
		NA2s	06/12/2007						
		NA3s	07/01/2008						
		NA4s	11/02/2008						
	Nantes	NA5s	22/01/2008						
		NA6s	23/10/2007						
		NA7s	06/04/2007						
		NA8s	01/10/2007						
		NA9s	12/07/2007						
		NA10s	13/12/2007						
		NA11s	20/11/2007						
		PA8s	11/01/2007						
		PA9s	05/03/2007						
		PA10s	07/01/2008						
		PA11s	22/02/2008						
		Paris	PA12s	09/01/2008					
	PA13s		26/03/2008						
	PA14s		31/07/2007						
	PA15s		17/01/2007						
	PA16s		26/01/2007						
	PA17s		27/03/2007						
	PA18s		22/08/2007						
	PA19s		13/11/2007						
	PA20s		15/11/2007						
	PA21s		03/12/2007						
	PA22s		26/04/2007						
	PA23s		09/05/2007						
	PA24s	25/03/2008							
	Rennes	<b><i>RE3s</i></b>	<b><i>18/09/2006</i></b>						
		RE4s	03/11/2006						
		RE5s	23/02/2006						
		RE6s	16/08/2006						
		RE7s	23/10/2006						
		RE8s	26/09/2006						
		RE9s	31/05/2006						
		RE10s	27/06/2006						
		RE11s	28/02/2006						
		RE12s	06/12/2006						
		TL2s	11/02/2008						
		Toulouse	TL3s	02/02/2007					
	TL4s		06/07/2009						
	TL5s		23/04/2007						
	TL6s		28/05/2009						
	TL7s		27/11/2007						
	TL8s		12/06/2007						
	TL9s		22/01/2009						
	TL10s		11/01/2010						
	TL11s		06/09/2007						
	TL12s		25/06/2007						
	TL13s		17/01/2007						
	TL14s		05/01/2007						
	Tours	TR4s	11/04/2007						
		TR5s	21/02/2007						
		TR8s	22/10/2007						
		TR10s	12/01/2007						
		TR11s	26/02/2007						
		TR6s	23/05/2005						
		TR7s	04/11/2005						
		TR9s	11/05/2005						
		TR12s	01/12/2005						
		TR13s	31/03/2004						
		TR14s	21/01/2005						
		TR15s	18/05/2004						



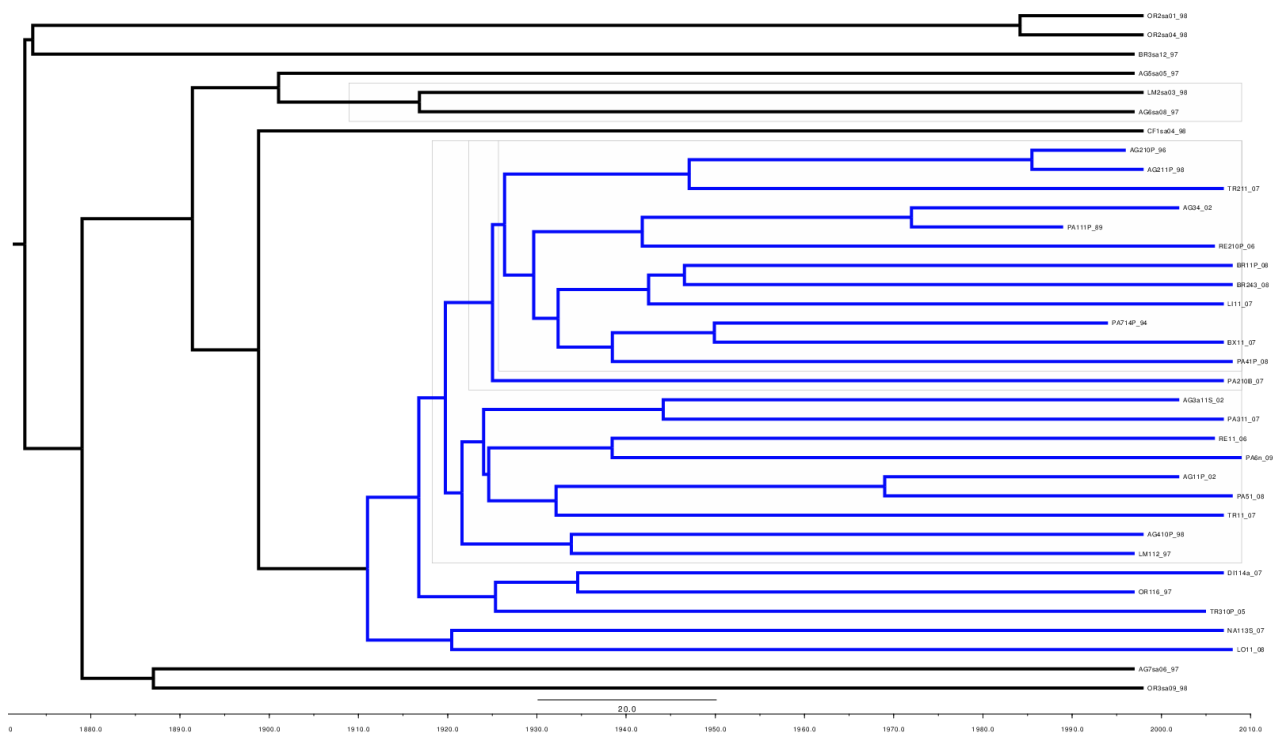




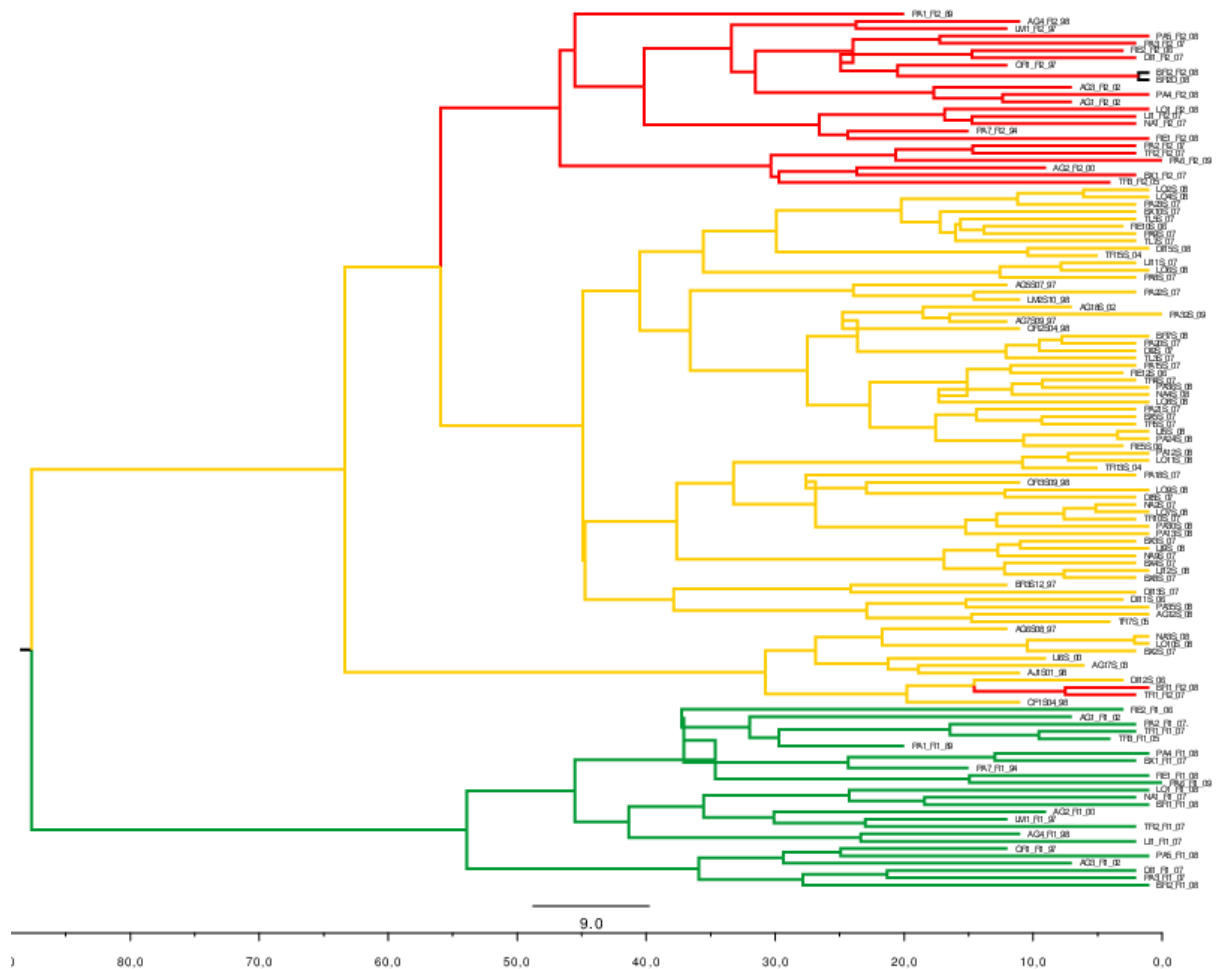
**Figure 2 :** Posterior distributions of mean substitution rates of NS5a for each quasispecies of HCV1-b. Each quasispecies has a different color.



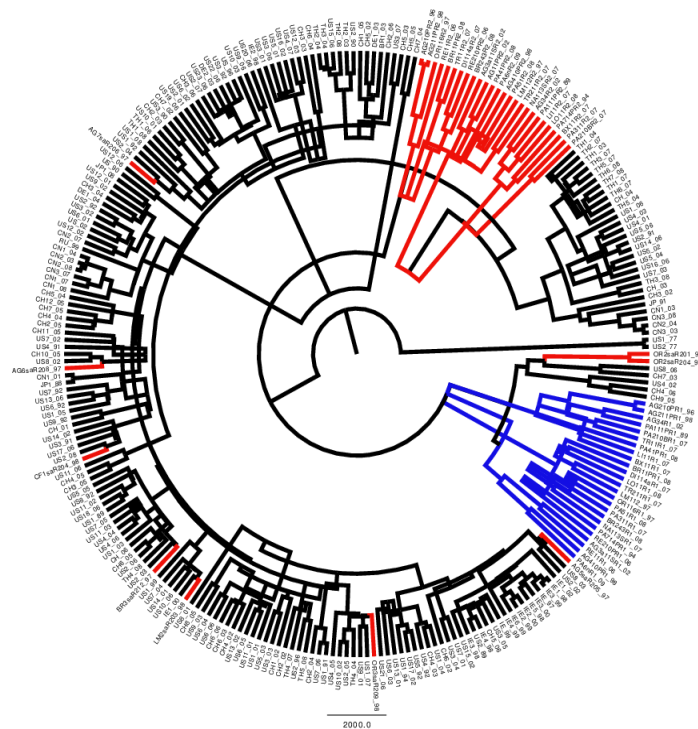
**Figure 3 : Divergence of the NS5A-dup from wild ones.** Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with random relaxed uncorrelated lognormal clock. NS5A-dup strains are indicated in blue.



**Figure 4 :** Divergence of NS5A-dup strains (in blue) from wild strains (in black). Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with random relaxed uncorrelated lognormal clock.

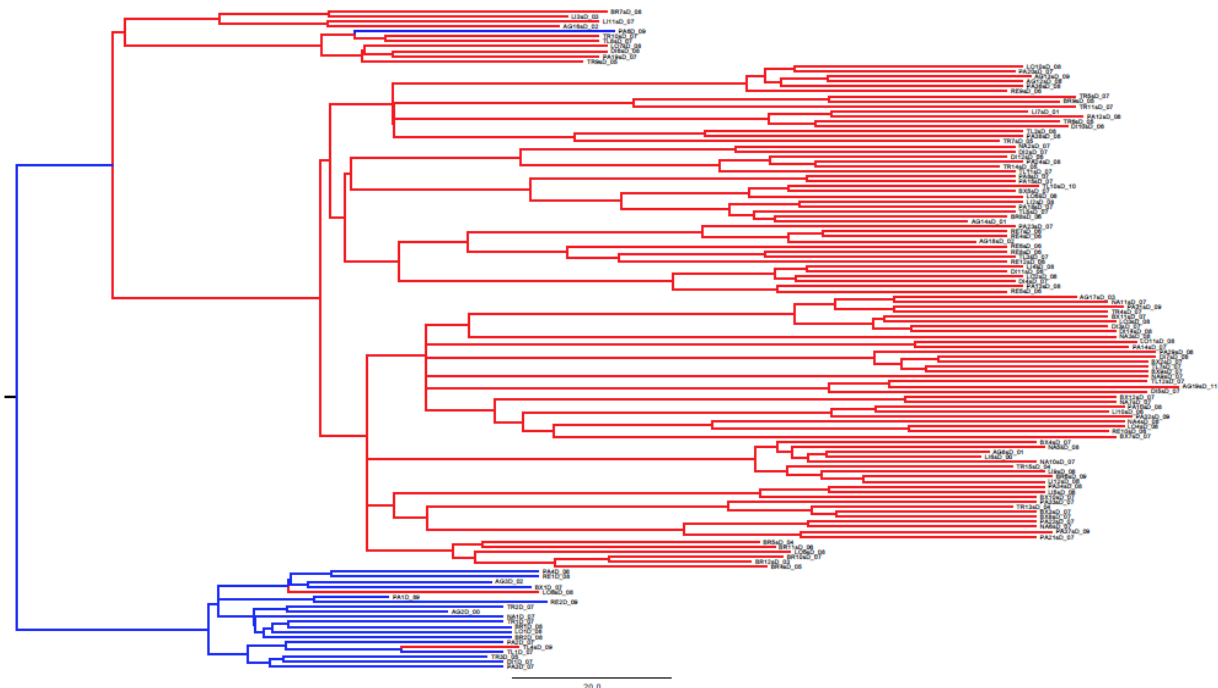


**Figure 5: First V3 domain (R1, in green) in NS5A-dup strains has a different origin than second V3 domain (R2, in red) which is related to wild-type domain (in yellow).** Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with random relaxed uncorrelated lognormal clock. Time is in x-axis and represents the divergence time between clades.



**Figure 6: R1 copies of V3 domain (in blue) have different origins than wild ones (in black) at the world scale. R2 copies (in red) have common origin with wild ones. Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with random relaxed uncorrelated lognormal clock.**

#### SUPPLEMENTARY INFORMATION



**SI1: Phylogeny of the E1 locus. NS5A-dup strains (red) are divergent from wild ones. Ninety percent posterior probability tree inferred from BEAST analysis under the coalescent constant size model with random relaxed uncorrelated lognormal clock.**

### I. Introduction

Le troisième volet de ce travail porte sur l'aspect fonctionnel de la duplication V3 dans le gène NS5A au niveau de la biologie du VHC. Nous avons souhaité caractériser les protéines partenaires des protéines NS5A portant une duplication. En effet, l'identification de protéines partenaires permet très souvent de caractériser le rôle de la protéine candidate au sein des voies de signalisation ou des voies métaboliques cellulaires. Les résultats de notre étude d'épidémiologie clinique suggéraient une éventuelle association entre l'infection par une souche portant la duplication V3 et une sévérité de la maladie hépatique, en particulier la présence d'une cirrhose et d'un CHC. L'identification de partenaires cellulaires originaux de cette protéine NS5A avec la duplication V3 impliqués dans les processus de fibrogenèse ou de carcinogenèse pourrait ainsi conforter les données observées au cours de ce premier travail.

De nombreuses approches ont été décrites et ont permis de cartographier les interactions de groupes de protéines, de protéomes complexes ou d'interactomes [359-362] et ont conduit à la constitution de banques de données spécifiques de domaines d'interaction [363, 364]. Les outils d'analyse développés sont basés sur des méthodes génétiques, biochimiques ou biophysiques [365, 366]. La caractérisation des interactions nécessite l'utilisation conjointe de plusieurs de ces méthodes pour valider les observations. Les techniques biochimiques sont les techniques les plus développées car elles sont facilement accessibles et elles sont les moins onéreuses. Nous avons choisi d'utiliser une combinaison de ces techniques.

Concernant les techniques biochimiques *in vivo*, il existe le phage display [367] et le double-hybride [368, 369] qui consistent toutes deux en un crible d'une banque d'ADN complémentaire (ADNc). Ces techniques *in vivo* permettent de mettre en évidence des interactions inconnues. Le phage display n'est quasiment plus utilisé. A l'inverse, la méthode de double-hybride est largement répandue avec différentes déclinaisons. Nous avons ainsi utilisé un crible en double-hybride levure pour notre analyse de première intention.

La co-immunoprécipitation et la précipitation du complexe protéique *in vitro* ou pull-down basées sur la mise en évidence du complexe protéique par précipitation sont parmi les techniques biochimiques *in vitro* les plus utilisées. Ces deux techniques ne dénaturent pas les protéines et utilisent respectivement un lysat cellulaire ou des protéines recombinantes. Ces techniques permettent uniquement de mettre en évidence des interactions entre des protéines

connues. Afin de valider en seconde intention les interactions protéine-protéine dépistées en crible double-hybride, nous avons fait appel à la technique de pull-down [370].

Pour caractériser nos variants NS5A au niveau fonctionnel et mettre en œuvre les techniques appropriées, nous avons développé plusieurs collaborations. Nous avons travaillé avec l'équipe du Dr E. Meurs qui dirige l'unité Hépacivirus et immunité innée de l'Institut Pasteur à Paris. Au cours du stage dans cette unité, nous avons réalisé plusieurs constructions en vecteur plasmidique permettant l'expression des protéines NS5A portant les deux domaines V3. Nous avons vérifié par immunofluorescence l'expression des protéines en culture cellulaire. Nous avons collaboré avec le Dr L. Tafforeau de l'équipe CIRI U1111/UMR5308 dirigée par le Dr V. Lotteau à Lyon. Cette équipe a développé une technique de crible d'interaction protéine-protéine en système double-hybride levure et a décrit l'interactome du VHC [371]. Le Dr L. Tafforeau a appliqué cette technologie à nos variants NS5A-V3 permettant ainsi d'identifier des protéines candidates. Enfin, nous avons souhaité valider les interactions précédemment observées en crible double-hybride. Nous avons initié et optimisé la production de ces protéines en système cellulaire eucaryote et en système bactérie avec l'objectif de valider les interactions par technique de pull-down. Nous avons réalisé ces expériences au sein de l'unité INSERM U892 du Dr Y. Delneste à Angers et avons pu bénéficier de leur expertise. Ce travail était l'objet du stage de Master II d'E. Bouthry que nous avons dirigé.



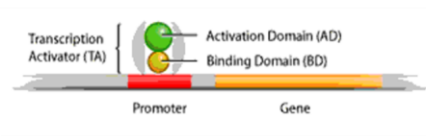
## II. Matériel et Méthodes

### 1. Crible en double-hybride levure, (Collaboration Dr L. Tafforeau, CIRI U1111/UMR5308, Lyon)

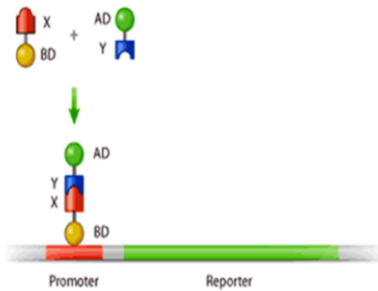
#### *1. 1. Principe du double-hybride en système levure*

La technique de double-hybride en système levure correspond à une analyse par criblage à haut débit semi-automatisée d'interactions protéine-protéine. Cette technique permet de confirmer et d'analyser les interactions entre deux protéines données et permet aussi d'identifier les partenaires d'une protéine donnée par le crible d'une banque d'ADNc. Le principe de cette technique est basé sur l'utilisation de facteurs de transcription de la levure qui sont constitués de deux domaines fonctionnels : un DNA-binding domain (DB) et un domaine de transactivation (AD) (Figure 28A). Chaque domaine est séparé et fusionné à une protéine d'intérêt dans un vecteur plasmidique, soit une protéine appât (X), soit une protéine proie (Y) conduisant à la production de protéines hybrides (Figure 28B). Deux souches de levures différentes et compatibles sont transformées respectivement avec la construction X ou Y et sélectionnées sur milieu de culture approprié. Les souches de levures sont ensuite croisées et les cellules diploïdes sont sélectionnées par culture. En effet, l'interaction entre ces deux protéines permet la reconstitution du facteur de transcription qui se lie alors à un gène rapporteur et active sa transcription. Le gène rapporteur restaure la capacité de la levure de croître en milieu sélectif déficient (en général dépourvu d'histidine, Figure 28C). Les colonies sont ensuite analysées par PCR et séquençage de l'ADNc afin d'identifier les vecteurs proies.

(A) Transcription chez la levure



(B) Principe du double-hybride chez la levure



(C) Détection de l'interaction



Figure 28. Principe du test rapporteur en double-hybride levure.

Cette technique de criblage à haut débit est très largement répandue, elle s'applique non seulement aux interactions protéines-protéines mais aussi à l'identification d'interactions protéines-ADN. Cette technique peut conduire à des faux positifs (protéines qui interagissent avec le promoteur des gènes rapporteurs, protéines « collantes (protéines heat shock), interactions coiled-coil (sous unité ribosomales, composants du cytosquelette), protéines non structurées). Ces faux positifs sont éliminés par l'analyse de l'expression des gènes rapporteurs en l'absence d'appât et les proies collantes sont connues. Les interactions permanentes et transitoires sont identifiées. A l'inverse les interactions de protéines présentant

des problèmes de structure (repliement), de stabilité ou nécessitant des modifications post-traductionnelles ne seront pas ou mal identifiées. Ainsi de nombreuses interactions déjà connues ne sont pas retrouvées avec cette technique. Néanmoins, il s'agit d'une technique robuste et rapide, réalisée *in vivo*, indépendante du niveau d'expression naturel des protéines, avec un rapport coût-efficacité avantageux qui a une place de choix en première intention [369, 372].

### 1.2. Constructions pour crible en double-hybride levure

La production de vecteur pour le crible en double-hybride a été réalisée selon le protocole développé au laboratoire de Lyon (Figure 29) [373] pour l'étude de l'interactome du VHC [371]. Les clones NS5A complets de 4 variants, précédemment produits en vecteur pGEMT-Easy au cours de l'étude phylogénétique, ont été repris (Ben, Dol, Gou, Rav) ainsi qu'une souche sauvage de génotype 1b (WT : wild type), Con1 fournie par l'équipe du Dr E. Meurs. Les souches Dol et Ben portent un domaine V3 inséré de 27aa et 12aa respectivement alors que les souches Rav et Gou ont une duplication complète de 31aa. Les inserts de NS5A ont été clonés dans les vecteurs gateway pDONR et pDEST-double-hybride (pGBKT7-gw, Clontech). Pour le crible double-hybride en levure, nous avons utilisé les souches AH109 et Y187 (Clontech). La première souche de levure a été transformée avec le plasmide "appât" (pGBKT7, c'est-à-dire le plasmide avec la partie Gal4-DNA binding fusionnée à notre protéine d'intérêt (NS5A)). La seconde a été transformée avec une librairie de plasmides "proies" (pACT2-gw, c'est-à-dire le plasmide avec la partie Gal4-Activation domain fusionnée à une librairie de cDNA de foie humain).

### 1.3. Crible

Après vérification que les protéines de fusion Gal4DB-NS5A\_variants n'étaient pas toxiques pour la levure et qu'il n'y avait pas d'auto-activation, nous avons croisé les 2 souches (ce sont 2 levures de type sexuel opposé permettant donc de générer un diploïde à partir des 2 souches) : après mélange à part égale des 2 souches, une culture sur milieu riche est réalisée pendant 4 à 5h (favorisant la reproduction), puis les cellules sont collectées et mises en culture sur milieu sélectif de cellules diploïdes. Nous avons sélectionné les levures où il y a interaction c'est-à-dire celles où le gène rapporteur encodant une enzyme impliquée dans la biosynthèse d'histidine est exprimé. En pratique, les levures ont été déposées sur un

milieu sans histidine et uniquement celles où il y a une interaction entre NS5A-variant et une protéine encodée par un cDNA humain peuvent pousser (environ 10 jours à 30°C).

Ces colonies ont été striées à nouveau sur milieu sélectif (2 fois 3 jours à 30°C). Une PCR a été réalisée sur ces colonies afin de séquencer les cDNA en question.

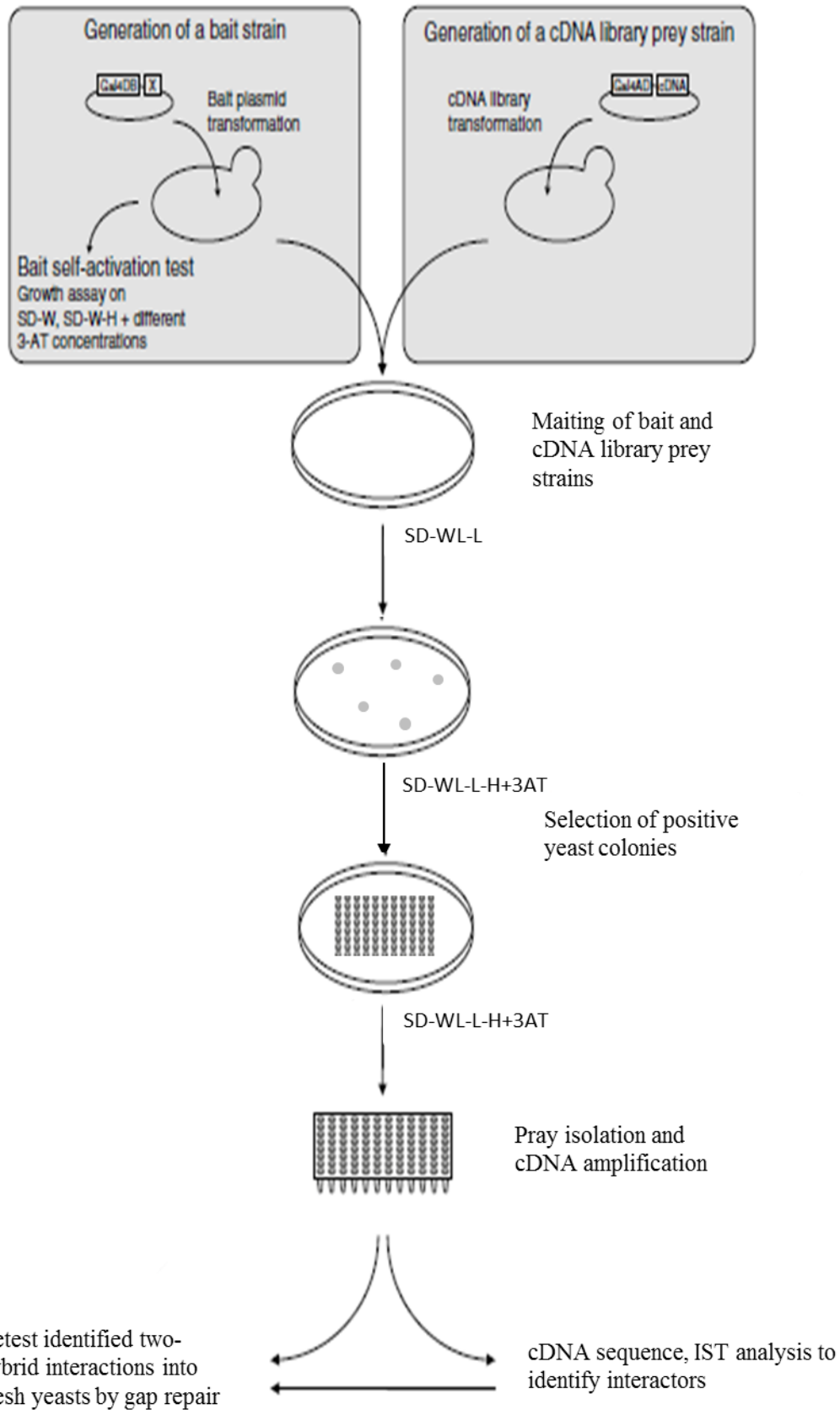


Figure 29. Schéma du crible en double-hybride basé sur la conjugaison de levures, d'après Tafforeau, 2012 [373].

## 1.4. Analyses

Un IST (Interaction Sequence Tag) est identifié pour chaque interaction. Un système bioinformatique développé au laboratoire de Lyon permet une annotation standardisée des protéines, une vérification des IST annotées et un export des données. La détermination du cadre de lecture et l'élimination de faux positifs sont réalisés au cours du traitement bioinformatique (la séquence Gal4 permet un alignement du cadre de lecture), la qualité de la séquence, le pourcentage d'identité, la E-value, l'annotation des positions et la database utilisée pour la protéine sont enregistrées.

## 2. Production de protéines et expériences de pull-down

### 2.1. Principe du pull-down

Il s'agit d'une technique de précipitation d'un complexe protéique qui permet de mettre en évidence une interaction directe entre deux protéines. La protéine d'intérêt est fusionnée à un « tag » qui permet sa liaison et sa purification sur une matrice de sépharose (chromatographie d'affinité sur colonne). La protéine candidate est ensuite incubée avec la protéine d'intérêt fixée à la matrice. Le complexe protéique qui s'est éventuellement formé est élué puis analysé par western blot (Figure 30). Ce test nécessite au préalable la production de protéines recombinantes, sous forme de protéines complètes ou d'un domaine particulier de la protéine. Il s'agit d'un test *in vitro* entre deux protéines, la mise en évidence de l'interaction est très dépendante de la concentration de chaque partenaire. L'intérêt est que cette technique ne dénature pas les protéines. Il s'agit d'une technique de validation d'interactions préalablement dépistées en criblage haut débit, cette technique est très spécifique mais nécessite la production et la purification préalable des protéines recombinantes. Ces méthodes sont de réalisation plus délicate et sont chronophages.

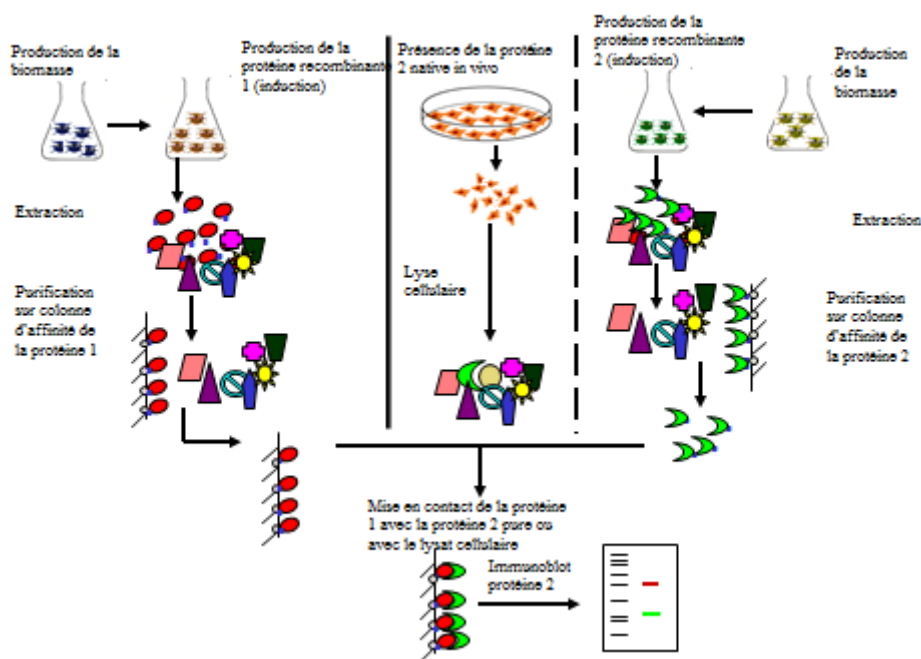


Figure 30. Principe du pull-down, d'après Gervais, 2010 [374].

## 2.2. Production en cellules de mammifères

### 2.2.1. Constructions des plasmides recombinants NS5A, (Collaboration Dr E. Meurs, Unité Hépacivirus et immunité innée, Institut Pasteur, Paris)

Les clones NS5A complets des mêmes 4 variants utilisés en crible double-hybride ainsi que le sauvage Con1 ont été repris et ligués en vecteur d'expression plasmidique pCI-Neo-3X-Flag, vecteur modifié disponible au laboratoire à Paris. Ce plasmide porte un tag triple Flag (3X-Flag, 3 répétitions successives du peptide Flag : DYDDDDK) et le gène de résistance à la néomycine. L'expression du peptide 3X-Flag est mise en évidence *in vitro* par un anticorps spécifique anti-Flag marqué. Les plasmides pGEMT-Easy sont digérés par l'enzyme de restriction NotI à raison de 1UI d'enzyme pour 10µg d'ADN, pendant 1h en Bain-Marie à 37°C. Les produits de digestion sont migrés en gel d'agarose 1% puis récupérés par découpage du gel et purifiés selon les recommandations du fournisseur (*Wizard*® SV Gel and PCR Clean-Up Start-Up kit, Promega). Une ligation des inserts en vecteur pCI-Neo-x3-Flag est réalisée avec la T4 DNA ligase (Promega Corp) ; 50µL de bactéries *E. coli* DH5- $\alpha$  disponibles à l'Institut Pasteur sont transformées avec 2.1 µL de produit de ligation, le mélange est incubé 30 min dans la glace, puis 45s en Bain-Marie à 42°C, puis 2 min dans la

glace, 450µL de milieu SOC (Invitrogen) sont ajoutés, le mélange est ensuite incubé 45 min à 37°C sous agitation douce. Les bactéries sont mises en culture une nuit à 37°C en milieu gélosé LB-ampicilline (100 µg/mL). Les constructions sont vérifiées par séquençage avant utilisation.

### 2.2.2. Transfections

Les lignées cellulaires 293T, cellules issues d'un cancer de rein de singe, et Huh7.5, lignée sélectionnée de la lignée Huh7 provenant d'un carcinome hépatocellulaire humain, disponibles au laboratoire à l'institut Pasteur, ont été utilisées pour l'expression *in vitro* des protéines NS5A. Ces cellules sont cultivées en milieu DMEM-10% SVF (Gibco) (DMEM : Dulbecco's Modified Eagle Medium, SVF : Serum de veau fœtal).

Deux autres lignées cellulaires ont été utilisées pour la production de protéine : les cellules Huh7, et les cellules CHO (Chinese Hamster Ovary) cultivées en milieu Ham F12. Ces lignées nous ont été fournies par l'Unité Inserm-U892 du Dr Y. Delneste d'Angers et par l'Unité Inserm-U966 du Pr P. Roingeard de Tours.

Les cultures s'effectuent à 37°C en atmosphère humide avec 5 % de CO<sub>2</sub>. Après trypsination, les cellules sontensemencées dans des plaques de 6 à 10 puits à raison de 10<sup>5</sup> à 1.5 x 10<sup>6</sup> cellules/mL selon la lignée, permettant d'obtenir une confluence des cellules comprise entre 90 et 95 % au moment de la transfection. Pour les expériences d'expression *in vivo*, des lamelles de verres sont préalablement déposées au fond des puits.

En immunofluorescence, 100 ou 800 ng d'ADN plasmidique sont mélangés à 50µL d'Opti-MEM (milieu de culture sans sérum Opti-MEM (Invitrogen)) pour un puit. Un mélange de 50µL d'Opti-MEM et 1.5µL de lipofectamine est ensuite ajouté et laissé en contact pendant 20 min à température ambiante. Cent microlitres du mélange précédent sont déposés dans chaque puits de Huh7.5 ou de 293T. Quatre heures après la transfection, le surnageant de culture est remplacé par 500µL de milieu MEM-SVF10%-1% acides aminés-1% pénicilline/streptomycine.

Pour la production de protéine, 4µg d'ADN plasmidique sont mélangés à 250 µL d'Opti-MEM chauffés à 37°C. En parallèle, 10 µL de lipofectamine sont mélangés à 250 µL d'Opti-MEM. Après incubation pendant 5 min des deux suspensions à température ambiante, les mélanges sont réunis, homogénéisés délicatement puis incubés à température ambiante pendant 20 min. Le mélange est déposé goutte à goutte dans chaque puits de culture cellulaire, réparti de manière homogène. Les cellules traitées sont incubées à 37°C en atmosphère



humide avec 5 % CO<sub>2</sub> pendant 24h. Les cellules transfectées sont alors réensemencées dans des flasques T75. La sélection se fait par ajout de néomycine (G418, Gibco) à la concentration de 500 µg/mL.

### 2.2.3. Immunofluorescence

Après 24h, les lamelles sont sorties des plaques de culture et transférées en plaques 24 puits adaptées à l'immunofluorescence. Les lamelles sont lavées deux fois avec un mélange PBS-Ca<sup>2+</sup>/Mg<sup>2+</sup>. La fixation est réalisée par ajout de PFA4% pendant 20 min. L'ensemble des lavages suivants est réalisé avec du PBS sans Ca<sup>2+</sup>/Mg<sup>2+</sup>. Les cellules sont perméabilisées par ajout de 500µL de [0.5% TritonX-100/CSK] pendant 15 min puis elles sont lavées deux fois pendant 5 min. Les lamelles sont mouillées avec du PBS sans Ca<sup>2+</sup>/Mg<sup>2+</sup>-Tween 0.05%. L'anticorps primaire (anticorps anti-Flag murin ou lapin, anticorps anti-GRP78 lapin (reticulum endoplasmique)) dilué au 1/200ème en PBS sans Ca<sup>2+</sup>/Mg<sup>2+</sup> est incubé pendant 1h. Après 3 lavages et un mouillage comme précédemment, l'anticorps secondaire marqué (anticorps anti-IgG1 souris-Cy3 ou Cy5, anticorps anti-IgG1 lapin-Cy3 ou Cy5) dilué au 1/200ème en PBS sans Ca<sup>2+</sup>/Mg<sup>2+</sup> est incubé pendant 45 min. Après lavage, le DAPI dilué au 1/1000ème en PBS sans Ca<sup>2+</sup>/Mg<sup>2+</sup> est ajouté pendant 5 min. Enfin, deux lavages pendant 5 min sont réalisés avant le montage des lamelles à l'aide du liquide de montage (Mowiol + 100µg de DABCO) et le lutage des lames. Pour certaines lamelles un marquage vital des lipides a été réalisé par ajout pendant 30 min de Bodipy dilué à 10 mg/ml en DMSO (Invitrogen).

### 2.2.4. Extraction des protéines totales

Le culot cellulaire de Huh7 ou de CHO, récupéré et lavé au PBS, est lysé dans 50 µl de tampon de lyse totale cellulaire (Tris pH 7,6 50 mM, Triton X®-100 1 %, NaCl 150 mM, EDTA 1 mM) auquel il faut ajouter extemporanément des inhibiteurs de protéases (PMSF : Phénylméthylsulfonyl Fluoride 1 mM, NaF 50mM et un mélange d'inhibiteurs de protéases (Roche)), incubé 30 min sur roue à 4°C et centrifugé 10 min à 4000g 4°C. Les surnageants, contenant les protéines totales, sont récupérés et conservés à -80°C.

La détermination de la concentration protéique se fait par technique colorimétrique par rapport à une gamme étalon d'albumine bovine sérique (BSA de 0 à 1,5 mg/mL, Thermo Scientific).

## 2.3. Production en bactéries

### 2.3.1. Construction du plasmide recombinant pFLAG-CTC<sup>TM</sup>

Le plasmide pFLAG-CTC<sup>TM</sup> (Sigma-Aldrich Corp) est un vecteur d'expression bactérien d'*E. coli*, utilisé pour l'expression cytoplasmique de protéines fusionnées en C-terminal à un peptide Flag (DYDDDDK). Ce vecteur contient le gène de résistance à l'ampicilline. La séquence codante de NS5A a été amplifiée par PCR à partir du plasmide pCI-neo-3X-Flag modifié à l'Institut Pasteur afin d'ajouter les sites nécessaires à la production en système bactérien. Les sites de restriction nécessaires au clonage dans pFLAG-CTC<sup>TM</sup> (Sigma-Aldrich), HindIII en 5' et KpnI en 3', sont insérés *via* les parties flanquantes des amorces utilisées lors de la PCR. Le codon d'initiation ATG est également ajouté en 5'.

La vérification de la présence de l'amplicon est réalisée par électrophorèse sur gel d'agarose à 2% (NuSieve) avec du BET (Bromure d'Ethidium) dans un tampon TBE (Tris Borate EDTA) 1X, révélée sous UV.

Le produit PCR obtenu et le plasmide pFLAG-CTC<sup>TM</sup> sont digérés par le couple d'enzymes de restriction HindIII/KpnI pendant 2h à 37°C, purifiés (Nucleospin Extract II®, Macherey-Nagel) et dosés par spectrométrie (ND-1000, Nanodrop). L'insertion du produit PCR (100 ng) dans le plasmide pFLAG-CTC<sup>TM</sup> est réalisée à l'aide de la T4 DNA ligase (Promega Corp) selon les recommandations du fournisseur.

Les plasmides sont transfectés dans des souches d'*E. coli* compétentes JM109 High Efficiency Competent Cells (Promega Corp) : 40 µL de bactéries compétentes sont mis en présence avec 4 µL produit de ligation et incubés 30 min dans la glace. Le choc thermique s'effectue par une incubation à 42°C pendant 50s puis dans la glace pendant 2 min. Les bactéries sont re-suspendues dans 360 µL de milieu LB (Luria-Broth) enrichi avec du glucose (concentration finale 10 mM) et incubées 1h30 à 37°C sous agitation 225 rpm. Les bactéries ayant intégré le plasmide sont sélectionnées en les étalant, à raison de 50 µL sur des boîtes de Pétri contenant du milieu LB complété par de l'ampicilline (100 µg/mL).

Après une nuit à 37°C, la présence de l'insert dans le plasmide est vérifiée chez les clones par PCR réalisée sur les colonies observées. Chaque colonie est diluée dans 10 µL d'eau stérile pour constituer la matrice PCR. L'amplification est réalisée à l'aide d'un couple d'amorces spécifiques du plasmide pFLAG-CTC<sup>TM</sup>. Une électrophorèse sur gel d'agarose à 2% avec du BET permet la révélation.

### 2.3.2. Purification des plasmides

Les colonies portant le plasmide recombinant sont mises en culture dans 2ml de LB contenant 100 µg/mL d'ampicilline et incubées une nuit à 37°C sous agitation à 250 rpm. Après 4 min de centrifugation à 3500 rpm à température ambiante, les plasmides sont purifiés à l'aide du kit Nucleospin plasmid® (Macherey-Nagel) élués dans 50 µL et dosés par spectrométrie (ND-1000, Nanodrop).

### 2.3.3. Culture et expression de la protéine NS5A

Les plasmides sont transfectés dans des souches d'*E. coli* compétentes BL21(DE3)pLysS (Promega Corp) suivant le protocole précédent. La souche bactérienne BL21 est inductible à l'IPTG (isopropyl-β-D-thiogalactopyranoside). Des volumes de 3 mL, 100mL et 1L de milieu LB-ampicilline (100 µg/mL) ont étéensemencés respectivement avec 100µL, 5 ml ou 25mL de pré-culture de bactéries transformées vérifiées de 16h. La culture est incubée à 37°C jusqu'à obtention d'une DO<sub>600</sub> comprise entre 0,5 et 0,8. La production de la protéine NS5A est induite par l'ajout d'IPTG à la concentration de 1mM et la culture est ré-incubée à 37°C pendant 4h sous agitation douce. Des aliquots (2x 1mL) sont prélevés avant et après induction pour l'analyse. La culture et les aliquots sont centrifugés à 4000g à 4°C pendant 20 min, les surnageants sont éliminés et les culots sont analysés immédiatement ou congelés à -80°C.

### 2.3.4. Extraction des protéines totales

Différents protocoles ont été testés à partir de 10 ml de culture précédente :

- Protocole TNG : Le culot bactérien est repris dans 5 mL de tampon TNG (Tris pH 7,6 20mM, NaCl 0,1 M, Glycérol 10 %). Les bactéries sont ensuite lysées par trois cycles de congélation à l'azote liquide et décongélation à 45°C en Bain-Marie, suivis de 3 cycles de 10s de sonication. Après ajout de PMSF 1mM, de 5µL de DNase et de MgCl<sub>2</sub> 25 mM, le lysat est incubé 20 min sur roue à 4°C. Un aliquot de fraction totale est prélevé puis le lysat est centrifugé 1h à 10000g à 4°C. Le surnageant récupéré correspond à la fraction soluble. Les aliquots de fraction totale et de fraction soluble sont conservés à - 80°C.

- Protocole Qiagen en conditions natives [375] : Le culot bactérien est repris en tampon de lyse (NaCl 300mM, NaH<sub>2</sub>PO<sub>4</sub> 50mM) à raison de 5ml/g de culot bactérien, auquel est ajouté du lysozyme (1ng/ml) et du PMSF (50µL/mL), le mélange est incubé 1h dans la glace. Une sonication (3 x 10s) est appliquée. Un aliquot de fraction totale est prélevé puis le lysat est centrifugé 1h à 10000g à 4°C. Le surnageant récupéré correspond à la fraction soluble. Les aliquots de fraction totale et de fraction soluble sont conservés à – 80°C.

- Protocole Qiagen en conditions dénaturantes en urée [375] : Le culot bactérien est repris en tampon de lyse (Tris pH7.6 10mM, NaH<sub>2</sub>PO<sub>4</sub> 100mM, urée 8M) à raison de 5ml/g de culot bactérien, auquel est ajouté du PMSF (50µL/mL), le mélange est incubé 1h dans la glace. Une sonication (3 x 10s) est appliquée. Un aliquot de fraction totale est prélevé puis le lysat est centrifugé 1h à 10000g à 4°C. Le surnageant récupéré correspond à la fraction soluble. Les aliquots de fraction totale et de fraction soluble sont conservés à – 80°C.

#### 2.4. Purification de la protéine NS5A

Le gel d'affinité anti-FLAG M2 (Sigma-Aldrich, TB A2220) contient des anticorps monoclonaux purifiés de souris IgG1 anti-Flag attachés de manière covalente à l'agarose par liaison hydrazide. Cette résine est utilisée pour la purification ou l'immunoprécipitation des protéines fusionnées Flag. Pour la purification des différentes protéines NS5A-Flag, la résine est préparée selon les recommandations du fournisseur.

Les surnageants issus des cultures bactériennes ou cellulaires sont incubés toute une nuit sur roue à 4°C avec la résine anti-FLAG. Après centrifugation 30s à 8000g à 4°C, le surnageant est récupéré (FN1 : fraction non retenue 1). La résine est récupérée et lavée trois fois avec 0.5mL de TBS. Chaque lavage est réalisé en tournant le tube sur roue pendant 1 à 2 min à température ambiante. Après chaque lavage, le tube est centrifugé 30s à 8000g à 4°C et le surnageant est conservé (FN2-4). La protéine NS5A est ensuite éluée de la résine : 100µL de solution d'éluion sont ajoutés à la résine, le mélange est incubé sur roue à température ambiante puis est centrifugé à 8000g pendant 30s à 4°C. L'éluat est déposé dans 10µL de tampon de lyse concentré (Tris pH 7.4 0.5M, NaCl 1.5M) et congelé à -80°C. Différents protocoles d'éluion ont été testés (incubation 10 ou 15 mn, à température ambiante ou à 30°C).

Les solutions d'élution suivantes ont été testées :

- solution de glycine 0.1 M pH 3.5
- solution de glycine 0.1M pH 2.8
- solution de glycine 0.2M pH 3.2
- solution de glycine 0.1 M pH 3.5 + NaCl 150Mm
- solution de glycine 0.1 M pH 2.8 + NaCl 150Mm
- solution de glycine 0.1 M pH 3.5 + MgCl<sub>2</sub> 2M

### *2.5. Pull-down*

L'interaction des partenaires de NS5A choisis à l'issue du crible double-hybride est vérifiée par la technique de pull-down. Les protéines cibles sont des protéines commercialisées (Tebu-bio) marquées par un tag GST. Cent nanogrammes de protéine purifiée NS5A-Flag liés à la résine M2 sont incubés avec 100 ng de protéine cible couplée à un tag GST avec 1 mL de tampon de lyse et du PMSF 1mM pendant 3h sur roue à température ambiante. Le mélange est centrifugé à 8000g pendant 2 mn à 4°C. Le complexe protéique est lavé trois fois avec du tampon de lyse totale (Tris p H 7.4 50mM, NaCl 150mM, EDTA 1mM, Triton X-100 1%). La protéine NS5A est éluée avec 100µL de solution de glycine 0.1M pH 2.8, le mélange est incubé sur roue à température ambiante puis est centrifugé à 8000g pendant 30s à 4°C. L'éluat est séparé en 2 aliquots destinés respectivement au western Blot avec l'anticorps primaire anti-GST et au western blot avec l'anticorps primaire anti-Flag.

## 3. Analyses des protéines produites

### *3.1. Gel SDS-PAGE coloré au bleu de Coomassie*

L'expression de la protéine NS5A est vérifiée sur un gel d'électrophorèse dénaturant SDS-PAGE (Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis), coloré au bleu de Coomassie. Une fraction de culture bactérienne ou cellulaire est mélangée à du tampon Laemmli puis dénaturée à 100°C pendant 5 min. Le tout est déposé sur un gel d'électrophorèse dénaturant 12% SDS-PAGE. La migration s'effectue à 70 V pendant 30 min puis à 120 V pendant 2h. Ce gel est ensuite coloré avec une solution d'Imperial Protein Stain (Thermo Scientific, 24615) puis décoloré à l'eau stérile afin de révéler les protéines. Une

semi-quantification de la protéine purifiée est réalisée par comparaison avec une gamme de BSA (bovine serum albumine).

### *3.2. Analyse par western blot*

Les protéines sont ensuite transférées sur une membrane de PolyVinylDiFluoride (Millipore) pendant 30 min à 25 V. Tous les lavages des étapes suivantes sont réalisés avec du TBS-Tween 0,05 %. Après transfert, la membrane est lavée puis saturée pendant 1 h à température ambiante dans une solution de TBS-Tween 0,05 %-lait écrémé 5 % sous agitation. La membrane est ensuite hybridée pendant une nuit à 4°C sous agitation lente avec une solution d'anticorps primaire de lapin anti-FLAG dilués au 1/5000<sup>ème</sup> dans du TBS-Tween 0,05 %-lait écrémé 3% (Sigma-Aldrich) ou d'anticorps anti-GST de souris dilués au 1/200<sup>ème</sup> dans de la BSA 1%. Les anticorps non liés à la membrane sont éliminés par des lavages successifs. La membrane est incubée 45 min à température ambiante sous agitation lente avec une solution d'anticorps secondaire de chèvre anti-anticorps de lapin ou anti-anticorps de souris conjugués à la peroxydase (Invitrogen) dilué au 1/10000<sup>ème</sup> dans du TBS-Tween 0,05 %-lait écrémé 3%. Après 3 lavages, la membrane est incubée pendant 1 min sous agitation lente dans un mélange volume/volume des réactifs du kit de détection de la chimioluminescence ECL<sup>TM</sup> (Amersham). La lecture se fait par le logiciel " Quantity One 1D " associé au lecteur " Molecular Imager ChemiDoc XRS System " (Biorad Laboratories).

### **III. Résultats**

#### 1. Crible en double-hybride levure

Les résultats du crible contre une banque d'ADNc de foie humain sont présentés dans le tableau 5. Chaque interaction entre une protéine NS5A (variant dupliqué V3 ou WT) et une protéine de la banque est identifiée par un chiffre dans la case correspondante. Ce chiffre est le nombre de « hits », c'est-à-dire le nombre de colonies identifiées avec cette interaction. Le nombre total de hits pour les variants dupliqués est aussi indiqué. La dernière colonne correspond aux interactions identifiées par l'équipe de Lyon entre les autres protéines du VHC et les protéines de la banque au cours de leurs travaux sur l'interactome du VHC [371].

Protéine Cible	Localisation	Noms	BEN	DOL	GOU	RAV	Nombre total de hits "variants NS5A"	Nombre de hits "NS5A_WT"	Interactions identifiées au cours de l'analyse de l'interactome du VHC	
ADFP	membrane plasmique	perilipin2 PLIN2		3		1	4		x	NS2
ALB	plasma	serum albumin	1				1		x	NS3
APOA1	plasma	Apolipoprotein A-I	4	74		171	249	85	x	NS2-3-4B
APOA2	plasma	Apolipoprotein A-II				2	2		x	NS3
ATF6	membrane, RE, noyau	Cyclic AMP-dependent transcription factor ATF-6 alpha	7		10		17		x	NS4B
BCLAF1	cytoplasme, noyau	Bcl-2-associated transcription factor 1 BTF			2		2			
BCLAF1			3	7		10				
BIN1	cytoplasme, noyau	Myc box-dependent-interacting protein 1	2				2	10	x	NS3
BIN1			2				2	8		
BIN1			1				1	3		
BIN1			1				1			
BRI3	membrane	pRGR2	1				1			
BZRAP1	cytoplasme, mitochondrie	Peripheral-type benzodiazepine receptor-associated protein 1	1				1			
C9orf23	noyau	Alba-like protein C9orf23	2				2	4		
CIDEB	cytosol	Cell death activator CIDE-B	2				2			
CMAS	noyau	N-acylneuraminate cytidyltransferase	1				1			
COMMD1	cytoplasme, noyau	COMMD1 copper metabolism (Murr1) domain containing 1			1		1			
DUS2L	cytoplasme, RE	tRNA-dihydrouridine synthase 2-like		1			1			
FGA	plasma	Fibrinogen alpha chain			1		1		x	NS3
GABARAPL2	golgi	Gamma-aminobutyric acid receptor-associated protein-like 2 GABA(A) receptor-associated protein-like 2	1		1		2			
HP	cytoplasme hépatocyte, plasma	Haptoglobin			1		1		x	NS3
HTRA1	espace extracellulaire	Human Serine protease HTRA1	1				1		x	NS3
JUNB	noyau	Transcription factor jun-B	1				1			
LGALS3	noyau	galectin-3, MAC2	1				1			
LRP5	membrane plasmique	Low-density lipoprotein receptor-related protein 5	1				1			
MPST	cytoplasme, mitochondrie	3-mercaptopyruvate sulfurtransferase tst2	1				1			
MYO1F	cytosol	Myosin-1f	1		1		2			
NRP1	membrane plasmique, espace extracellulaire	Neuropilin-1		1			1			
NUCB1	cytoplasme, golgi, membrane plasmique	nucleobindin-1	1				1	3		
OSM	espace extracellulaire	oncostatin M	7		6		13			
RASD1	membrane, cytoplasme, noyau	Dexamethasone-induced Ras-related protein 1	1				1			
RND3	membrane, golgi	Rho-related GTP-binding protein	2		1		3			
RPL23A	cytosol	60S ribosomal protein L23a	2		3		5			
RPL3	cytosol	60S ribosomal protein L3	4				4		x	NS3
RPL37	cytosol	60S ribosomal protein L37	1				1			
SORBS3	jonctions cellulaires, cytosquelette, noyau	vinexin Sorbin and SH3 domain-containing protein 3 SH3-containing adapter molecule 1 SCAM1	2		1		3	1	x	NS3
TP53BP2	cytoplasme, noyau	Apoptosis-stimulating of p53 protein 2 tumour suppressor 53 binding protein 2	1				1	4	x	Core
ZNF135	noyau	Zinc finger protein 135			1		1			
ZNF135			5	1		6	1			
HLA-C	membrane plasmique	HLA class I histocompatibility antigen	1				1			

Tableau 5. Résultats du crible des protéines variants dupliqués et sauvage (WT) de NS5A contre une banque d'ADNc de foie humain.



Les interactions les plus fréquentes des variants dupliqués NS5A sont colorées en rose. L'analyse des interactants identifiés a été réalisée à travers une revue de la littérature et une classification selon leur rôle dans la biologie cellulaire a été établie. Comme l'indique la figure 31, les protéines identifiées sont en premier lieu des protéines de structure (31%) puis se répartissent à un niveau comparable les protéines impliquées dans la signalisation ou la croissance cellulaire et la régulation de la transcription. Une part plus faible de partenaires identifiés sont des interactants connus avec le VHC ou sont des facteurs anti-apoptotiques.

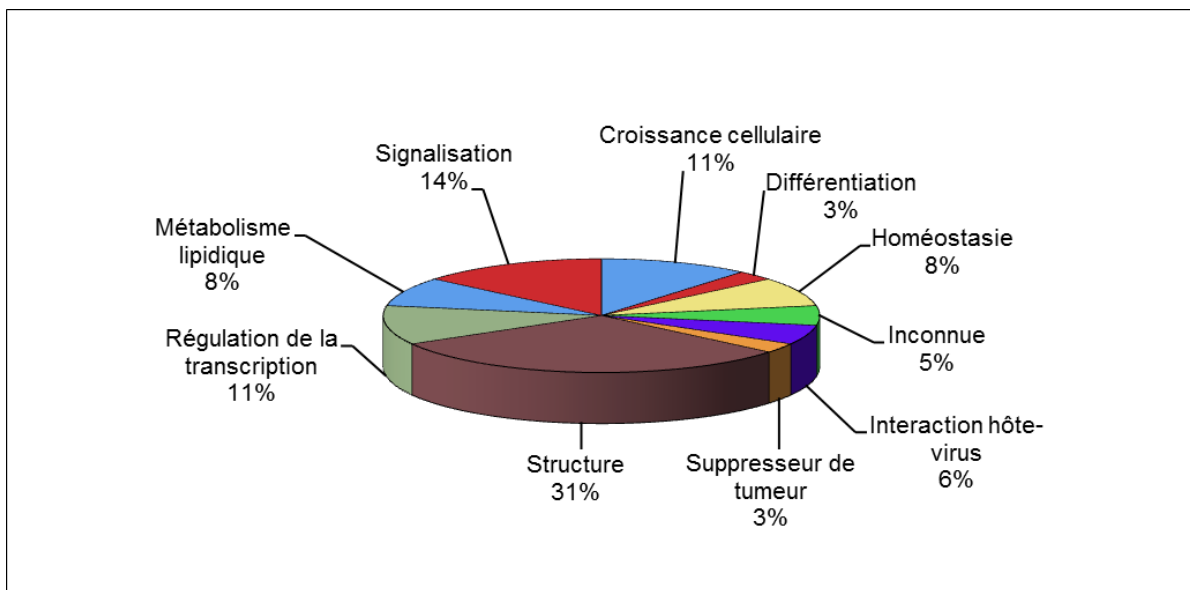


Figure 31. Analyse de la répartition des fonctions des partenaires identifiés au cours du criblage double-hybride en fonction du rôle dans la biologie cellulaire.

A partir de cette analyse, nous avons sélectionné les interactions à valider en première intention en pull-down. Le choix retenu est présenté dans le tableau 6. Ce choix a aussi été guidé par la disponibilité de ces interactants sous forme de protéine recombinante sur le marché ou auprès de nos collaborateurs. La protéine ApoA1 a été retenue comme témoin positif car son interaction avec NS5A a déjà été largement validée dans la littérature et est retrouvée dans notre crible avec le WT.

	Protéine Cible	Nombre total de hits "variants NS5A"	Nombre de hits "NS5A_WT"	Type de protéine	Interaction avec le VHC	Implication dans la biologie cellulaire
Interactant de NS5A	BIN1	6	21	Facteur suppresseur de tumeur à activité pro-apoptotique, (contrôle positif)	Interaction avec motifs SH3 des domaines I et II de NS5A	Carcinogénèse favorisée par la liaison de NS5A à Bin1, inhibition de l'activité pro-apoptotique de Bin1 [376-378]
	APOA1	249	85	Lipoprotéine (contrôle positif)	Interactions avec NS5A et Core au niveau des lipid droplets	Assemblage de la nucléocapside [379]
Effet sur la signalisation cellulaire et l'apoptose	ATF6	17		Facteur de transcription	Interaction avec NS4B dans les hépatocytes. Clivée et activée par NS4B et lors d'un stress du RE. Active le gène proapoptotique gadd153 dans les cellules infectées	Stress du RE (réponse UPR) [380-383]
	OSM	13		Cytokine	Inhibition de la replication du VHC en synergie avec l'interféron alpha	Prolifération cellulaire et réponse immunitaire [384-386]
	TP53BP2	1	4	coactivateur	Interaction avec la protéine core, inhibition de l'apoptose induite par le suppresseur de tumeur p53	Apoptose cycle cellulaire [387]
Effet sur l'assemblage du virus	ADFP	4		Lipoprotéine	Présente au niveau des lipid droplets où a lieu l'assemblage viral	Protéine présente dans les gouttelettes lipidiques intracellulaires des adipocytes, probable rôle dans le métabolisme lipidique de plusieurs types cellulaires dont les hépatocytes
Effet sur la réponse immunitaire	COMMD1	1		Protéine de régulation	-	Régulation de l'homéostasie du cuivre, du sodium, inhibe l'activation du facteur de transcription NF-kB par le TNF

Tableau 6. Sélection de protéines pour validation en pull-down. Les protéines retenues sont colorées en rose. Leur implication dans la biologie cellulaire et leur éventuel rôle dans le cycle du VHC sont explicités.

## 2. Expression de la protéine NS5A en cellules de mammifère

### *2.1. Immunofluorescence en lignées cellulaires*

L'expression de NS5A est positive pour les 2 lignées cellulaires 293T et Huh7.5 (Figure 32 A et B) validant ainsi les constructions. Les images sont de meilleure qualité avec les cellules Huh7.5 avec un nombre de cellules transfectées supérieur. Il n'est pas noté de différence de niveau d'expression ou du nombre de cellules transfectées entre le sauvage et les variants dupliqués dans les deux lignées. Le marquage anti-Flag NS5A est visible uniquement au niveau cytoplasmique pour les deux lignées. En Huh7.5, le marquage laisse apparaître des zones rondes non marquées (Figure 32 C1). Le double marquage anti-Flag NS5A/anti-LD (Lipid Droplets) montre que les zones rondes sont des lipid droplets (Figure 32 C2). Le triple marquage avec des anticorps anti-Flag/anti-LD/anti-RE (Reticulum Endoplasmique) montre une superposition des marquages anti-RE et anti-Flag NS5A (Figure 32 D1 à D3).

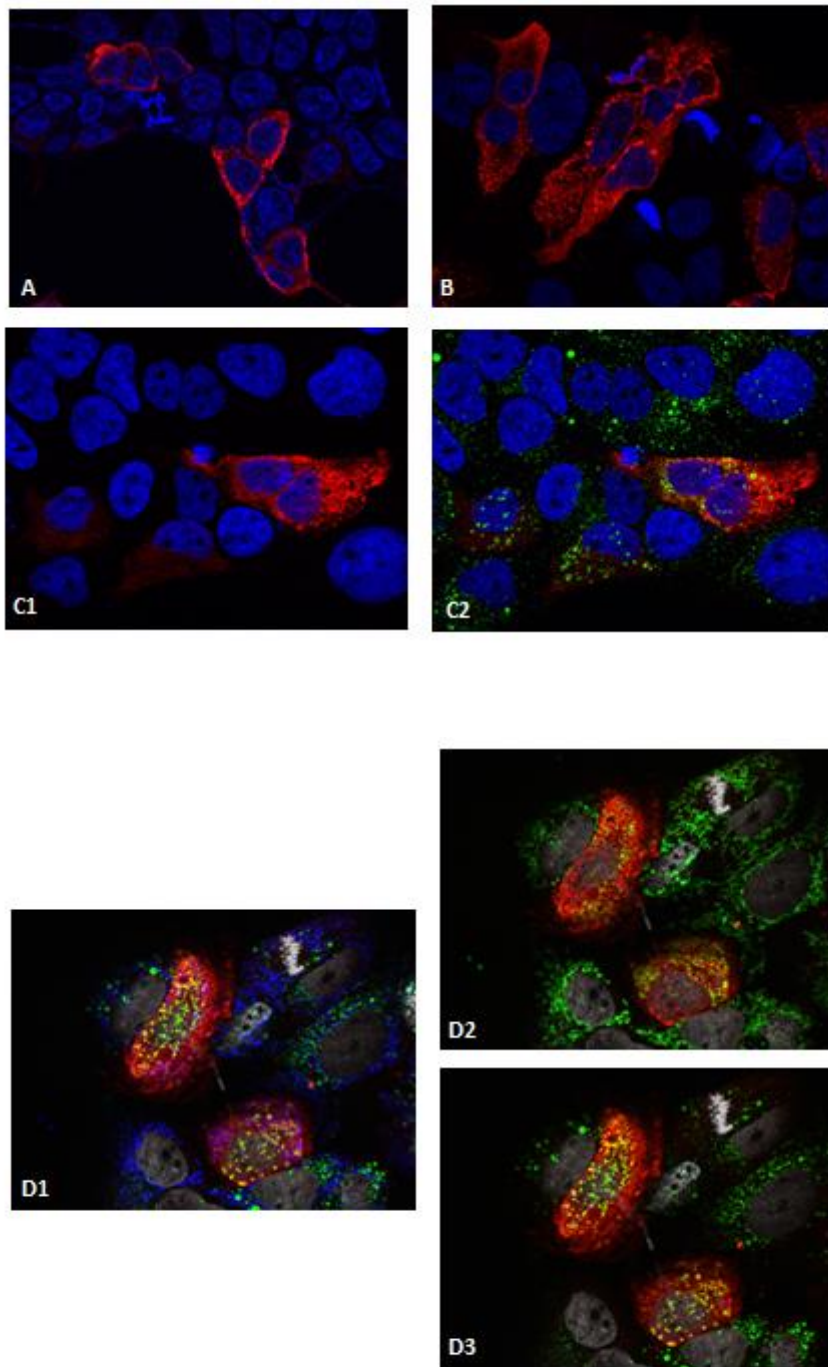


Figure 32. Analyses des cellules transfectées en microscopie à fluorescence. (A) Marquage anti-Flag NS5A en cellules 293T ; (B) marquage anti-Flag NS5A en cellules Huh7.5 ; (C) double marquage en cellules Huh7.5 (C1) anti-Flag NS5A seul (C2) anti-Flag NS5A et anti-LD ; (D) triple marquage en cellules Huh7.5 (D1) triple marquage anti-Flag NS5A (rouge) / anti-LD (vert) / anti-RE (bleu), (D2) double marquage anti-Flag NS5A / anti-RE, (D3) double marquage anti-Flag NS5A/anti-LD.

Les lysats cellulaires issus des cultures en Huh7.5 ont été contrôlés en western blot et montrent une production de la protéine NS5A pour chaque variant (Figure 33).

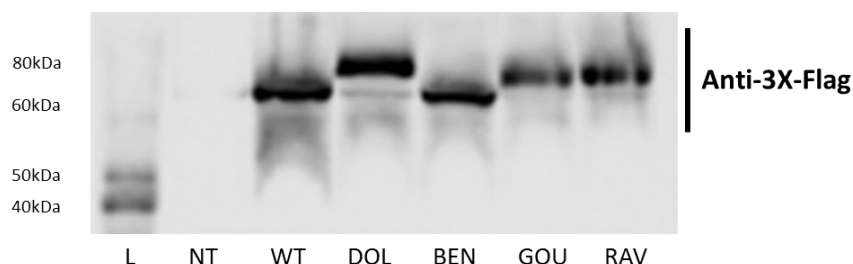


Figure 33. Analyse en western blot des lysats cellulaires Huh7.5 transfectés par les différents variants NS5A

## 2.2. Etablissement d'une lignée stable

L'expression de la protéine NS5A a été testée sur 2 types de cellules : cellules CHO et cellules Huh7. Après transfection du plasmide pCI-neo modifié recombinant ou non, la présence de la protéine NS5A a été vérifiée par une analyse en western blot (Figure 34).

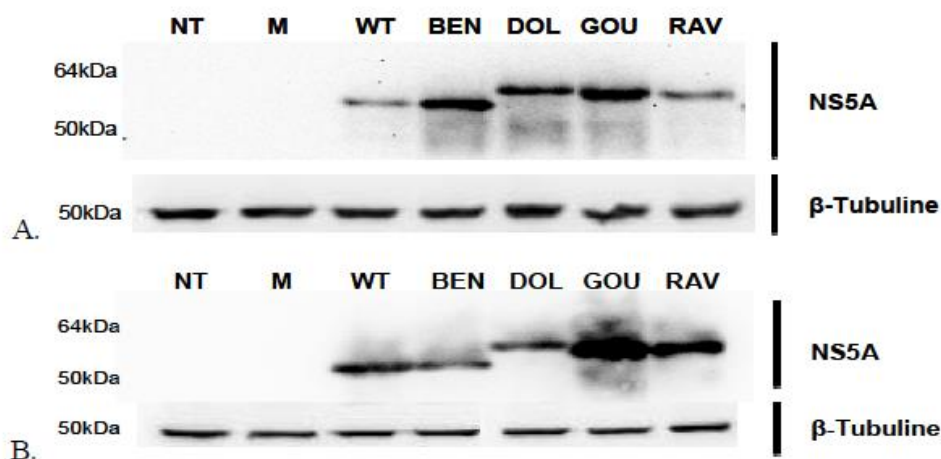


Figure 34. Western blot sur extraits de lysats cellulaires. (A) cellules CHO (B) cellules Huh7. *NT* : cellules non transfectées. *M* : « mock » plasmide sans insert NS5A.

La présence d'une bande de masse équivalente à celle de NS5A (située entre 50 et 64 kDa) sur le western blot montre l'expression d'une protéine uniquement dans les lignées transfectées, confirmant que les cellules transfectées avec un plasmide recombinant NS5A

expriment bien de manière continue la protéine cible. Une quantité équivalente de lysat de protéines totales a été déposée (20µg) pour chaque lignée transfectée. Le western blot présentant l'expression constante de la  $\beta$ -tubuline, utilisée comme contrôle de charge, montre que la quantité de protéines totales déposée est identique quelles que soient les conditions. Le rendement de production semble différent selon la forme de NS5A transfectée (sauvage ou dupliquée) : dans les cellules CHO, les mutants sont plus exprimés que le sauvage, avec un rendement plus élevé pour Ben et Gou par rapport à Dol et Rav. Cependant, dans les cellules Huh7, seuls Gou et Rav sont surexprimés par rapport au sauvage, avec un plus fort taux de Gou.

### 2.3. Phase de production

La figure 35 montre un profil protéique issu de la lyse des cellules Huh7 et CHO transfectées.

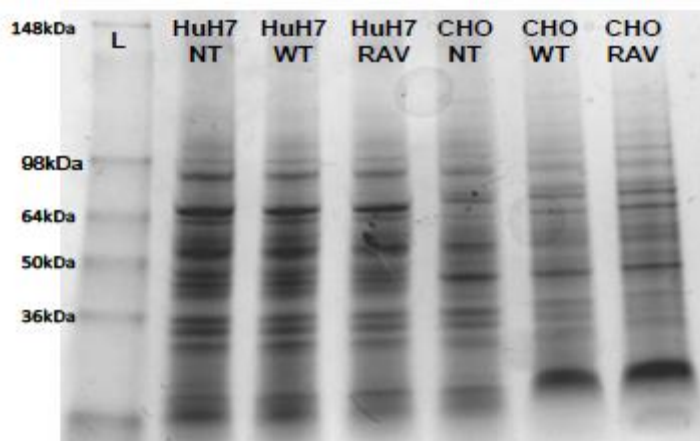


Figure 35. Analyse en gel SDS-PAGE des protéines cellulaires produites en cellules Huh7 et CHO. L : standard de poids moléculaire. NT : cellules non transfectées. WT : cellules transfectées avec la construction Wild type. RAV : cellules transfectées avec un variant de NS5A (variant RAV).

Afin de vérifier si la production de NS5A sur une flasque est suffisante, les fractions cellulaires obtenues après transfection des différents plasmides ont été analysées en gel SDS-PAGE. La même quantité de protéines totales issue de cellules non transfectées a été déposée en parallèle. Les profils protéiques ne semblent pas montrer de différence entre les diverses fractions. Il n'y a pas de détection de bande ayant une masse moléculaire voisine de celle de

la protéine NS5A (56 ou 58 kDa). Ces résultats montrent que si la protéine est exprimée, elle l'est en faible quantité. Il a donc été nécessaire d'augmenter le nombre de cellules utilisées afin d'améliorer le rendement.

#### 2.4. Optimisation de la quantité de protéine NS5A obtenue

Un essai de production en grand volume sur le sauvage WT a été réalisé. Après obtention d'une lignée stable, l'expression de NS5A a été amplifiée sur une surface équivalente à 1400 cm<sup>2</sup> de cellules Huh7. Les cellules récupérées ont été traitées comme précédemment décrit. Après purification sur résine M2 anti-FLAG, un aliquot a été analysé sur gel SDS-PAGE (Figure 36).

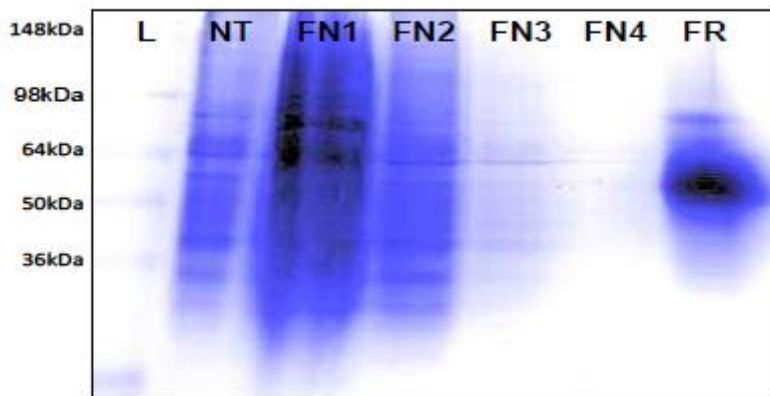


Figure 36. Analyse en gel SDS-PAGE des protéines cellulaires provenant d'une production sur 8 flasques de cellules Huh7 transfectées avec le recombinant WT après purification sur résine M2 anti-FLAG. **L** : standard de poids moléculaire. **NT** : cellules non transfectées. **FN1** : 1<sup>ère</sup> fraction non retenue. **FN2** : 2<sup>ème</sup> fraction non retenue. **FN3** : 3<sup>ème</sup> fraction non retenue. **FN4** : 4<sup>ème</sup> fraction non retenue. **FR** : fraction retenue.

Les fractions non retenues contiennent les protéines contaminantes qui ne se lient pas à la résine anti-FLAG. Ces protéines sont totalement éliminées dans la fraction FN1 et plus aucune n'est visible dans la fraction FN4. La présence d'une bande unique de masse équivalente à celle de NS5A (située entre 50 et 64 kDa) montre que la fraction retenue ne contient que la protéine NS5A sauvage.

La quantité de protéine NS5A produite a été estimée à l'aide d'une gamme BSA sur gel SDS-PAGE à environ 0.6µg/µL (Figure 37).

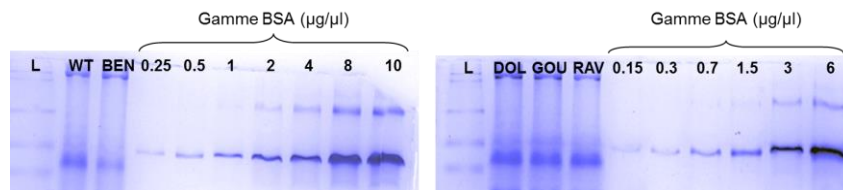


Figure 37. Semi-quantification de la protéine NS5A issue de la production sur 8 flasques de Huh7.

### 3. Expression de la protéine NS5A en système bactérien

#### *3.1. Construction du plasmide recombinant pFLAG-CTC<sup>TM</sup>*

La figure 38 présente l'amplification du gène NS5A obtenu à partir des 5 plasmides pCI-néo modifiés réalisée à l'aide d'amorces sur lesquelles ont été ajoutés les sites de restriction HindIII et KpnI ainsi que le codon d'initiation ATG. La taille des amplicons, située entre 1350 et 1450pb selon les formes de NS5A, correspond à la taille attendue.

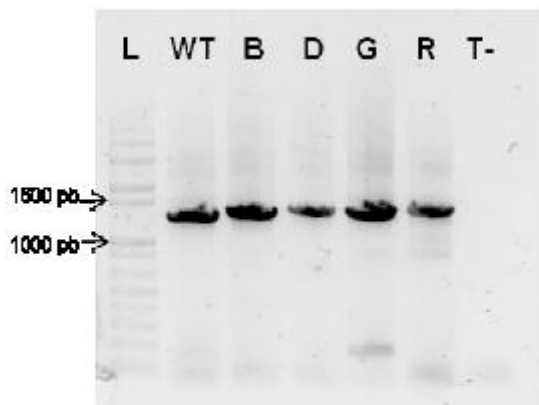


Figure 38. Vérification de la taille du gène NS5A par électrophorèse des produits de PCR. **L** : marqueur de taille. **WT** : wild-type. **B** : variant BEN. **D** : variant DOL. **G** : variant GOU. **R** : variant RAV.

Le plasmide pFLAG-CTC<sup>TM</sup> et les amplicons NS5A ont été digérés par les enzymes de restriction. Les séquences codant la protéine sauvage ou mutée de NS5A ont ensuite été



insérées dans le plasmide bactérien pFLAG-CTC<sup>TM</sup>. Après transformation dans les bactéries *E. coli* JM109, la présence de l'insert NS5A a été vérifiée sur plusieurs clones.

D'après la figure 39, les PCR sur tous les clones sauf sur Rav3 présentent une bande équivalente à NS5A. Les plasmides ayant intégré le gène NS5A sont purifiés puis introduits dans une souche de *E. coli* BL21(DE3)pLysS, capables ainsi de produire les différentes protéines NS5A après induction.

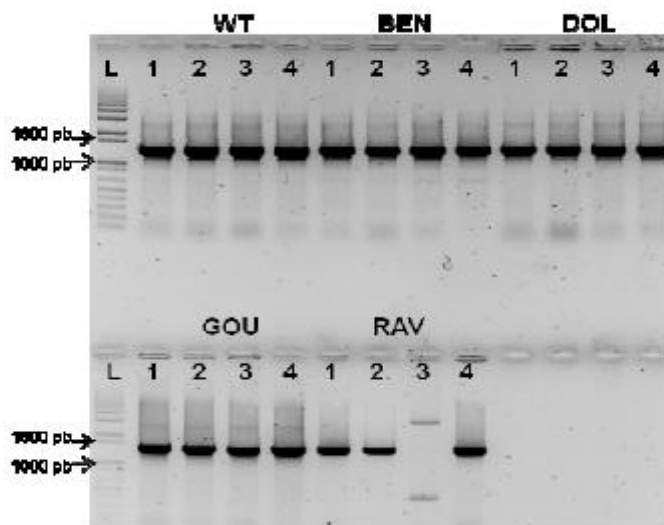


Figure 39. Vérification de la présence de l'insert dans le plasmide pFLAG-CTC<sup>TM</sup> par électrophorèse des produits de PCR réalisées sur les colonies pour les 5 variants (**WT**, **BEN**, **DOL**, **GOU**, **RAV**). **L** : marqueur de taille.

### 3.2. Production de la protéine NS5A

La production de la protéine NS5A a été analysée à partir de cultures réalisées sous un volume de 3mL de milieu LB sur l'ensemble des variants : WT, BEN, DOL, GOU, RAV. L'analyse a été réalisée en western blot (Figure 40).

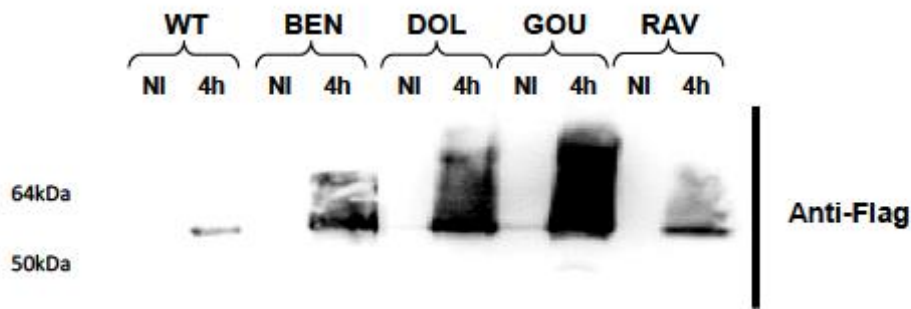


Figure 40. Analyse en western blot des protéines provenant des fractions solubles des 5 variants. NI : non induit.

L'ensemble des constructions permet une production de protéines attestée par la présence d'une bande spécifique. Cependant les quantités produites apparaissent faibles.

### 3.3. Optimisation de la quantité de protéine NS5A obtenue

Afin de répondre au problème de la quantité de protéine NS5A obtenue, un essai de production bactérienne en grand volume sur le sauvage WT a également été réalisé : 1L de culture a été testé. Le culot bactérien a été repris dans 30 mL de tampon de lyse (TNG) et 40  $\mu$ L de fraction soluble ont été déposés sur gel SDS-PAGE. La figure 41A ne montre pas de prédominance de la protéine NS5A, la lecture est difficile au vu de la concentration de l'ensemble des protéines produites par la bactérie. Cependant la protéine est produite en grande quantité, ce que montre l'analyse en western blot (Figure 41B). Les productions ultérieures ont été réalisées avec un volume intermédiaire de 100 mL de LB permettant une production de protéine suffisante (voir ci-après Figure 44).

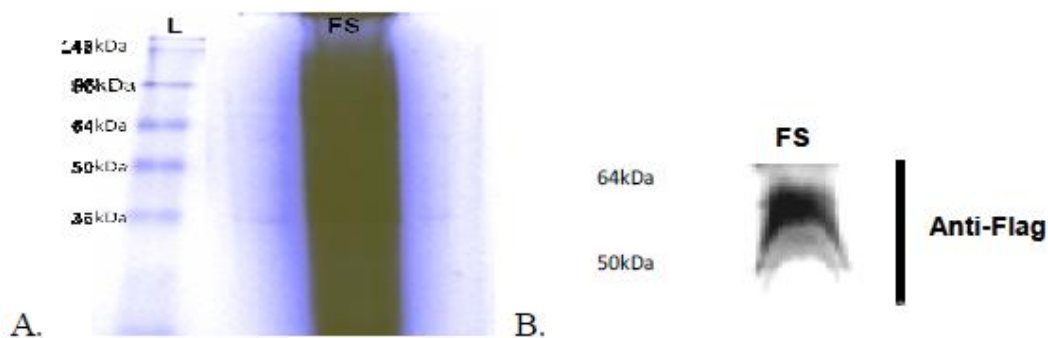


Figure 41. Analyse des protéines provenant de la fraction soluble (FS) obtenue à partir d'1L de culture bactérienne en SDS-PAGE (A) et western blot (B).

## 4. Interaction protéine-protéine : pull-down

### *4.1. Optimisation des conditions d'extraction de la protéine NS5A*

Trois protocoles de lyse (TNG, Qiagen en condition native, Qiagen en condition dénaturante-Urée) ont été testés afin d'améliorer les conditions d'extraction de nos protéines NS5A et d'augmenter la quantité de protéines obtenue. Les résultats sont présentés sur la figure 42. La lyse Qiagen en condition native n'est pas satisfaisante, la protéine NS5A n'est quasiment pas détectable en western blot. La lyse Qiagen en urée apparait drastique avec des profils en SDS-PAGE de lysats protéiques riches, la protéine NS5A est détectée en western blot mais la quantité reste insuffisante. Les images observées suggèrent une probable dégradation des protéines avec l'urée. La lyse en TNG montre un profil de protéines produites satisfaisant avec une large bande détectée en western blot. Ce protocole est retenu pour la production de protéines NS5A.

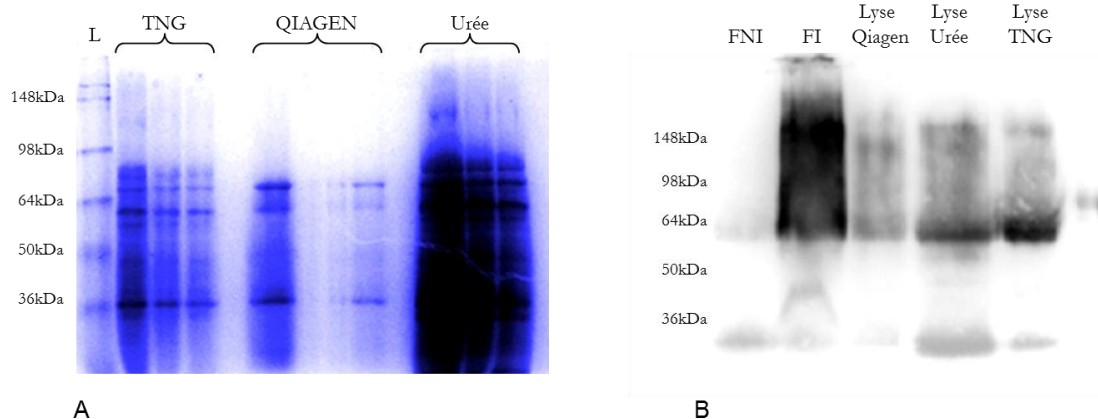


Figure 42. (A) Analyse des protéines en SDS-PAGE, (B) en western blot avec un anticorps anti-Flag (B) ; FNI : fraction non induite, FI : fraction induite.

### *4.2. Purification par affinité de la protéine NS5A : optimisation des conditions d'éluion*

Le test d'interaction pull-down s'effectue entre protéines purifiées : NS5A-Flag et partenaires cibles fusionnés à la GST. Le tag Flag de NS5A permet une détection sensible et une purification de la protéine NS5A en utilisant une résine anti-FLAG. Les éluats et les résines obtenus avec les solutions de glycine chauffées à 30°C ou en présence de NaCl ou de MgCl<sub>2</sub>. La présence de protéine NS5A dans l'éluat n'est obtenue qu'avec les solutions de

glycine seule chauffées (figure 43A). Ces solutions sont testées ensuite sous différentes conditions. La condition la plus favorable et qui est retenue est une purification en solution de glycine 0.1M pH 2,8 avec une incubation pendant 10 min à température ambiante (Figure 43B).

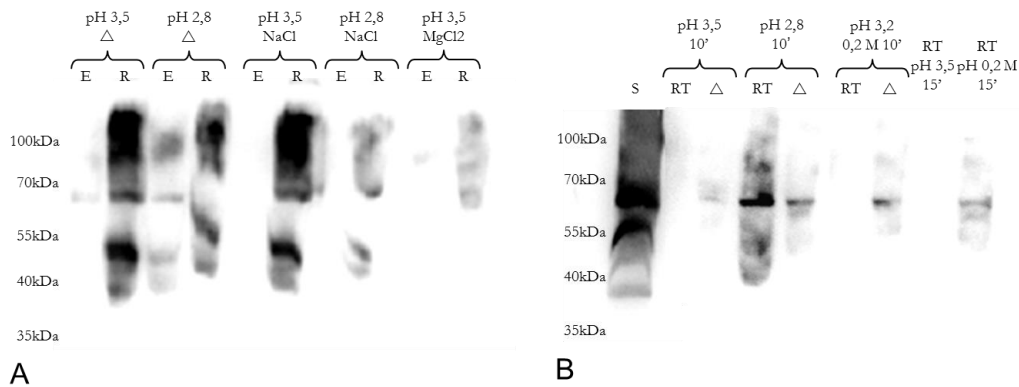


Figure 43. Analyse en western blot de la production de protéine NS5A dans l'éluat sous différentes conditions (RT : température ambiante,  $\Delta$  : 30 C; 10' : 10 minutes). A- 1<sup>ère</sup> série de test, dépôt des éluats (E) et de la résine (R) ; B- 2<sup>ème</sup> série de tests, dépôt des éluats uniquement, S : fraction soluble.

#### 4.3. Quantification de la protéine NS5A

La quantification de la protéine NS5A purifiée est nécessaire pour réaliser la technique de pull-down. En effet, la quantité de chaque interactant est le paramètre le plus sensible de cette technique. La quantité de protéine NS5A a été estimée après production bactérienne sous un volume de 100mL par comparaison avec une gamme BSA déposée sur un gel SDS-PAGE (Figure 44). L'éluat correspond à un étalon d'environ 300ng de protéine, soit une concentration de 7.5 ng/ $\mu$ L pour un volume d'éluat déposé de 40 $\mu$ L.

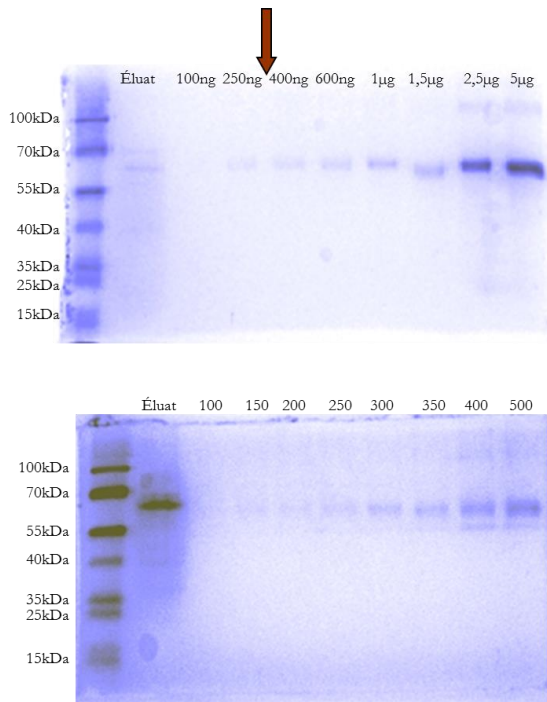


Figure 44. Estimation en SDS-PAGE de la quantité de protéine NS5A purifiée dans le lysat cellulaire du WT.

#### 4.4. Essais de pull-down

##### 4.3.1. Essai sur la production cellulaire en Huh7

L'interaction avec la protéine ADFP-GST a été testée sur la production cellulaire en Huh7 qui a permis d'obtenir 0,6 µg/µL de protéine NS5A. Afin de s'assurer que NS5A ne se fixe pas au tag GST, une interaction NS5A-GST a été recherchée en parallèle. L'analyse est effectuée par western blot avec un anticorps anti-GST. Le contrôle avec la GST seule ne montre pas d'interaction et le contrôle anticorps anti-GST est validé (Figure 45A). Aucune interaction entre la protéine NS5A et l'ADFP n'est identifiée (Figure 45B).



Figure 45. Analyse en western blot des interactions entre les 5 variants de NS5A et l'ADFP. (A) interaction GST-NS5A, Rés : résine seule contre ADFP ; (B) Interaction ADFP-NS5A, Rés : résine seule contre GST.

#### 4.3.2. Essai sur la production bactérienne

Les interactions avec les protéines recombinantes ApoA1-GST et ADFP-GST ont été testées sur un lysat bactérien issu d'une production réalisée sur un volume de 100mL de culture et purifié selon les conditions optimales définies auparavant. Les interactions ont été mises en œuvre en utilisant 100 ng environ de chaque partenaire. Les western blot obtenus sont « blanc » : aucune des protéines n'est détectée, que ce soit sur le dépôt de chaque protéine seule (western blot GST ou western Flag) ou celui des protéines.

Afin de vérifier les quantités de protéines et leur seuil de détection, les protéines commerciales ont été testées en western blot sans pull-down préalable. Seuls les dépôts de grande quantité sont détectés pour les trois protéines (Figure 46).

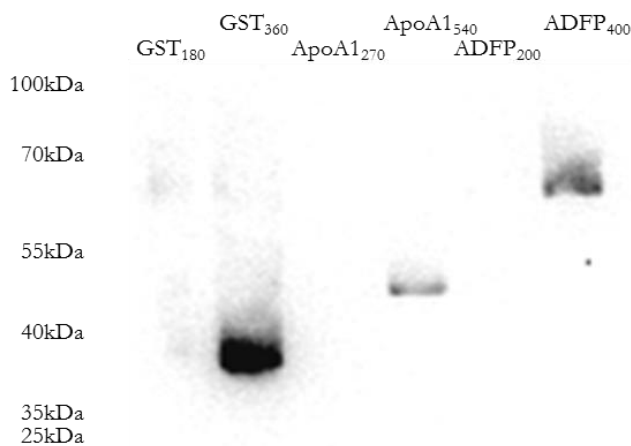


Figure 46. Analyse en western blot de différentes quantités (en ng) des protéines GST, ApoA1 et ADFP.

## IV. Discussion

La détermination du rôle potentiel de la duplication du domaine V3 de la protéine NS5A dans la physiopathologie de l'infection par le VHC est le troisième objectif de notre projet. Nous avons souhaité déterminer l'impact de cette duplication sur les interactions de la protéine NS5A avec les protéines cellulaires. La question était de savoir si la duplication entraînait des interactions avec des partenaires différents en comparaison de celles observées avec la protéine NS5A sauvage.

La majorité des duplications observées au cours de notre étude d'épidémiologie clinique étaient complètes c'est à dire constituées de deux domaines V3 de 31aa. Cependant quelques variants présentaient des duplications avec un des deux domaines V3 plus courts. Dans ce contexte, nous avons d'emblée fait le choix d'utiliser dans les études fonctionnelles plusieurs variants avec ces différentes tailles de domaine V3. Quatre variants de NS5A portant une duplication V3 partielle (souches Ben et Dol) ou complète (souches Gou et Rav), issus de souches de VHC de génotype 1b isolées de 4 patients atteints d'hépatite chronique et un variant sauvage (Con1) ont ainsi été analysés.

Nous avons mis en place une stratégie collaborative d'analyse d'interaction protéine-protéine impliquant 3 équipes et basée sur l'utilisation de techniques biochimiques *in vivo* et *in vitro*. L'approche fonctionnelle a d'abord nécessité la construction de vecteurs d'expression adaptés à la culture en lignées cellulaires de mammifères et en bactéries. Nos vecteurs étaient porteurs de tag de type Flag, marqueur peptidique très largement utilisé en biologie cellulaire, qui permet la production de protéines recombinantes. Ce peptide Flag fusionné à la protéine d'intérêt permet sa détection de façon sensible à l'aide d'anticorps anti-Flag marqués ou sa purification par anticorps anti-Flag fixés sur une résine. Nous avons ainsi obtenu une production des différents variants de la protéine NS5A dans 4 lignées cellulaires. La production en cellules Huh7.5 a été confirmée par immunofluorescence avec une localisation de NS5A au niveau du reticulum endoplasmique et une localisation au niveau des gouttelettes lipidiques des cellules. Ces observations sont en accord avec le caractère cytoplasmique essentiel de la protéine NS5A d'une part et ses interactions avec les lipoprotéines cellulaires d'autre part [42]. Nos constructions validées en lignées Huh7.5 et 293T ont été adressées au Dr L. Tafforeau qui a mis en œuvre une première approche fonctionnelle par des tests en double-hybride de levure contre une banque d'ADNc de foie. Cette stratégie permet

d'identifier notamment des interactions non connues. Elle autorise aussi un crible de plusieurs protéines candidates contre un très grand panel de protéines cibles [388]. Nous avons utilisé cette technologie comme un dépistage de première intention appliqué à nos 5 variants.

Nos résultats ont d'abord confirmé la nature de nos variants dupliqués par l'identification de très nombreux hits avec la protéine Bin1 pour un variant et avec l'apolipoprotéine A1 (ApoA1) pour 3 variants. Ces deux interactions ont aussi été identifiées avec le variant sauvage. En effet, ces deux protéines sont des interactants de la protéine NS5A connus et validés dans la littérature. Concernant Bin1, la protéine NS5A se lie au niveau de motifs SH3 de cette protéine pro-apoptotique suppresseur de tumeur [389]. L'interaction avec Bin1 a lieu au niveau des motifs proline PxxP du domaine II de NS5A qui ont une grande flexibilité et changent de conformation pour se lier à de nombreux partenaires [390, 391]. Cette interaction inhiberait l'activité anti-apoptotique de Bin1, favoriserait aussi la formation du réseau membranaire nécessaire à la réplication virale et modulerait l'hyperphosphorylation de la protéine NS5A [376-378]. Pour sa part, l'ApoA1 est associée à NS5A au niveau des gouttelettes lipidiques, les études suggèrent un rôle de NS5A dans la perturbation du métabolisme lipidique, elle contribuerait notamment à la stéatose hépatique observée dans les hépatites C chroniques [379, 392]. Dans les tests en double-hybride de levure, l'ApoA1 est la protéine qui a la plus forte affinité avec NS5A. Nous avons donc utilisé cette protéine comme témoin en pull-down. Testée en parallèle des autres partenaires cibles, l'interaction avec l'ApoA1 a plusieurs buts : valider l'interaction avec NS5A, confirmer la présence d'une quantité suffisante de NS5A afin de valider les résultats obtenus ou à l'inverse démontrer une quantité trop faible de protéine NS5A rendant l'expérience ininterprétable, et enfin valider une absence d'interaction entre ApoA1 et la protéine candidate.

Les autres interactions dépistées en double-hybride portaient sur des nombres de hits plus faibles, compris entre 1 et 10 hits pour un couple donné. De même, le nombre de hits cumulés pour les variants avec la duplication variait de 1 à 17 et de 1 à 4 pour le variant sauvage. Enfin, les interactions communes concernaient soit deux variants avec la duplication (ex : ADFP, ATF6, GABARAPL2...), soit un variant avec la duplication et le variant sauvage (ex : C9orf23, NUCB1, TP53BP2..). Seules les interactions avec SORBBS3 et ZNF135 concernaient 3 souches (2 avec duplication et le sauvage). Le nombre de hits et la communauté de l'interaction entre plusieurs variants ne suffisaient donc pas à choisir les interactions à valider ensuite en technique *in vitro*. L'implication des partenaires au niveau de la biologie cellulaire et leur éventuelle interaction avec d'autres protéines virales du VHC a



été le second critère retenu pour arrêter ce choix. Ainsi, la protéine NS5A semble impliquée dans le processus de carcinogenèse en interagissant notamment avec la protéine anti-tumorale p53 [317]. Notre étude d'épidémiologie clinique a en outre suggéré un impact négatif possible de la duplication V3 dans l'évolution de la pathologie hépatique. Il paraissait donc important de tester des interactions mettant en jeu des protéines impliquées dans l'oncogenèse. Nous avons ainsi retenu 3 partenaires impliqués dans les voies de signalisation cellulaire, l'ATF6 (17 hits), l'oncostatine M (13 hits) et le TP53BP2 (1 hit). La protéine ATF6 (Activating transcription factor 6) est activée lors d'une réponse UPR (unfolded protein response) traduite par un stress du reticulum endoplasmique avec l'accumulation de protéines mal repliées dans sa lumière [393]. Des études *in vitro*, ont montré que le VHC pouvait induire un stress du reticulum endoplasmique et altérer la réponse UPR [394]. L'oncostatine M (OSM) a d'abord été identifiée pour sa capacité à inhiber la prolifération de certaines lignées tumorales en particulier les mélanomes. Elle joue un rôle significatif dans la différenciation et la régénération hépatique [395]. D'après une étude récente, l'OSM pourrait inhiber la réplication du VHC en synergie avec l'IFN $\alpha$  [385]. La protéine TP53BP2 (Apoptosis-stimulating of p53 protein 2 ou ASPP2) joue un rôle crucial dans la régulation de l'apoptose et de la croissance cellulaire, notamment en agissant sur le gène suppresseur de tumeur p53 [396]. Il semblerait qu'elle puisse interagir avec les protéines de capsid du VHC, ce qui inhiberait l'apoptose induite par p53 [387]. De plus, NS5a empêche la translocation nucléaire de la protéine p53, suppresseur de tumeur en la séquestrant dans le cytoplasme et inhibant ainsi l'induction de l'apoptose [323].

Parmi les autres protéines candidates, nous avons retenu l'ADFP (4 hits) ou Adipose differentiation-related protein, qui appartient à la famille des lipid droplet protéin présentes au niveau des gouttelettes lipidiques des hépatocytes. Les lipoprotéines sont impliquées dans l'assemblage viral qui a lieu au niveau de gouttelettes lipidiques en présence de la protéine NS5A. Cette famille est impliquée notamment dans la physiopathologie de la stéatose hépatique à laquelle NS5A semble participer [397]. Concernant la réponse immunitaire, nous avons retenu la protéine COMMD1 (« COMM domain-containing protein 1 » ou protéine Murr1); elle appartient à la famille des protéines Murr1 impliquées dans la réponse immunitaire en inhibant l'activation du facteur de transcription NF- $\kappa$ B [398]. Or il a été démontré que la protéine NS5A interagissait avec les voies de signalisation de l'interféron et était capable d'activer NF- $\kappa$ B en induisant un stress oxydatif [308].

Suite à ces résultats, notre travail s'est attaché à la validation des interactions identifiées avec les candidats choisis par la technique de pull-down. Cette technique nécessite d'abord la production des protéines d'intérêt NS5A sous forme recombinantes. Ainsi, la production des 5 variants NS5A sauvage et mutés a été mise en œuvre sur deux systèmes : production en cellules de mammifères et production en bactéries. L'expression en cellules de mammifères de NS5A recombinante fusionnée à un tag triple Flag ayant été validée en lignées Huh7.5 et 293T, la construction plasmidique a été ré-utilisée pour la production en lignées Huh7 et CHO. A contrario, l'expression en système bactérien a nécessité la construction d'un plasmide recombinant adapté. Par souci de comparaison avec la production en lignée cellulaire, le tag utilisé a été le Flag. Les protéines ont été exprimées fusionnées à un peptide Flag ou 3xFlag en bactéries ou en cellules de mammifères, respectivement.

Les différents systèmes de production testés ont permis l'expression des protéines NS5A, vérifiée sur western blot. Concernant l'expression en cellules de mammifères, 2 lignées cellulaires ont été transfectées : la lignée hépatocytaire humaine Huh7 et la lignée ovarienne de Hamster CHO. Le rendement de production semblait différent selon la forme de NS5A transfectée (sauvage ou dupliquée) et selon le type cellulaire. En effet, les souches Ben et Gou étaient les variants les plus exprimés dans les cellules CHO tandis que dans les cellules Huh7, ce sont les variants Gou et Rav qui étaient les plus exprimés. Sur les deux lignées cellulaires, les protéines NS5A portant la duplication V3 étaient davantage exprimées que la protéine sauvage. Des tests de prolifération cellulaire pour évaluer l'avantage qu'induirait la duplication ont été initiés. Les résultats préliminaires semblent montrer que les cellules transfectées par certains mutants prolifèrent plus rapidement que celles transfectées par le sauvage, elles-mêmes proliférant plus rapidement que les cellules non transfectées. Ces résultats nécessitent néanmoins confirmation. Ces données seraient concordantes avec le rôle de la protéine NS5A dans la prolifération cellulaire et le contrôle du cycle cellulaire [250, 399].

Malgré une expression des protéines NS5A dans les deux systèmes de production, les rendements obtenus étaient faibles. Les tests d'interaction pull-down nécessitent une quantité suffisante (100 ng) de protéine NS5A pour pouvoir interagir avec les partenaires cibles. Des tests en grand volume ont donc été réalisés : une production sur une surface équivalente à 1400 cm<sup>2</sup> de cellules Huh7 et une production sur 1 litre de biomasse bactérienne. La production en cellules Huh7 sur une plus grande surface a conduit à un rendement médiocre.

Nous avons utilisé nos productions pour faire du pull-down sans obtenir de résultat satisfaisant (Figure 45). Nous avons privilégié l'optimisation de la production en système bactérien dans l'attente de l'arrivée d'un système de production de protéine en cellules en milieu liquide automatisé à haut rendement. En effet, la production en cellules de mammifères a l'avantage de respecter les modifications post traductionnelles.

Dans les différentes études publiées, les conditions d'expression de la protéine NS5A en bactéries sont très variables en fonction des constructions plasmidiques utilisées. [250, 400, 401]. Nous avons donc testé plusieurs conditions de production avec le sauvage WT afin de définir les conditions optimales, qui ont ensuite été vérifiées pour l'expression de l'ensemble des variants. Nous avons testé différentes conditions de lyse du culot bactérien et de purification de la protéine NS5A sous différents volumes de culture. Nous avons finalement défini une production optimale avec une culture bactérienne réalisée sous un volume de 100 mL suivie d'une lyse TNG du culot bactérien et d'une purification avec une solution de glycine 0.1M pH 2,8 incubée pendant 10 min à température ambiante. Avec ce protocole, nous avons obtenu une production estimée à 7.5 ng/μL de lysat cellulaire, ce qui est concordant avec les travaux de Huang *et al.* [400] qui montrent un rendement de 1 ng/μL à partir d'un lysat bactérien. Nous avons aussi obtenu l'élution de la protéine NS5A de la résine M2 anti-Flag qui était une des difficultés à laquelle nous nous sommes heurtées au début des productions.

L'obtention de protéines NS5A recombinantes purifiées en système bactérien nous a permis de débiter les analyses en pull-down. Nous avons testé un premier partenaire : l'ADFP en utilisant l'ApoA1 comme protéine témoin d'interaction. Aucune protéine n'a été détectée, pas même les témoins anticorps. Nous avons testé 100ng de chaque partenaire, cette quantité avait été retenue car déjà validée au laboratoire pour d'autres interactions. Cependant, la quantité de chaque partenaire est le paramètre le plus délicat du pull-down. Avant d'envisager d'autres quantités de partenaires dans de nouveaux pull-down, nous avons vérifié nos protéines partenaires et nous avons mis en évidence un seuil de détection plus élevé à environ 400ng et une absence de détection à la quantité de 100ng précédemment utilisée dans le pull-down. Les protéines étant conservées depuis quelques mois au laboratoire, la question d'une éventuelle dégradation s'est aussi posée.

Pour permettre la continuité de cette étude et la validation des résultats des tests en double-hybride, plusieurs axes vont être développés. De nouveaux tests en pull-down seront réalisés en faisant varier les quantités respectives de chaque partenaire avec un nouveau lot de protéines recombinantes GST. Les différents variants seront évalués d'emblée avec les autres partenaires retenus à la suite du crible en double-hybride levure. Nous avons d'abord commencé avec un seul partenaire et avons retenu l'ADFP. En effet, les résultats de l'étude épidémioclinique n'étaient pas encore obtenus et nous avons ciblé l'interaction de NS5A avec les lipoprotéines. Nous envisageons donc de focaliser cette analyse en privilégiant les partenaires impliqués dans l'oncogenèse et la fibrogenèse. En effet, nos travaux suggèrent que l'infection par une souche dupliquée pourrait être associée à une sévérité de la maladie hépatique, notamment la présence d'une cirrhose et d'un CHC. L'implication de la protéine NS5A dans la carcinogenèse hépatique est maintenant démontrée [402, 403], en outre cette protéine pourrait avoir des propriétés transactivatrices impliquant le domaine C-terminal où se situe notamment le domaine V3 [266].

D'autres approches pourraient être envisagées pour valider les résultats obtenus à la suite du test en double-hybride de levures. En effet, l'équipe de Lyon a développé une nouvelle stratégie de validation des interactions protéiques couplant au double-hybride et à la co-immunoprécipitation la spectrométrie de masse [404]. L'approche par une technique de coimmunoprécipitation pourrait également confirmer les interactions mises en évidence précédemment. Cette méthode, hautement spécifique, permet de détecter à la fois les interactions directes et indirectes. Elle nécessite l'utilisation de plusieurs anticorps très affins. De nombreuses interactions de la protéine NS5A ont ainsi été validées par cette méthode [405, 406]. L'utilisation couplée à des techniques biophysiques comme la spectrométrie de masse permet une approche structure-fonction originale, avec notamment la détermination des conformations adoptées selon l'interaction et la mise en évidence d'interactions transitoires [391, 407]. La très grande flexibilité des domaines II et III de NS5A est maintenant bien démontrée [280, 408] et explique la possibilité de liaison à de très nombreuses protéines. Une alternative est aussi envisagée à Angers avec la mise en place actuellement d'une plate-forme lentivirale.

L'identification d'un nouveau partenaire à l'issue de ces tests, nécessitera ensuite l'utilisation de réplicons sous-génomiques dérivés du clone infectieux JHF1 [409] et la mise en œuvre de constructions de réplicons modifiés portant les différents variants de NS5A ainsi

que les domaines III de NS5A seulement. Il s'agit en effet de définir si la duplication V3 est le domaine d'interaction avec la protéine partenaire identifiée. Cette stratégie combinée met aussi en œuvre des tests rapporteur luciférase ainsi que la quantification de l'expression des gènes codant les protéines partenaires *in vitro*. Ces stratégies ont permis la mise en évidence de l'inhibition de la protéine GADD45 $\alpha$  par NS5A soulignant le rôle de NS5A dans la survenue du CHC [403], ou encore l'interaction de la vipérine avec la protéine NS5A pour contrôler la réplication virale [410], ou enfin la suggestion d'une régulation de la réplication virale par activité transcriptionnelle du fragment C-terminal de la protéine NS5A [266].

L'identification d'un partenaire spécifique aux protéines NS5A avec une duplication V3 nous permettrait de mieux cerner l'implication de ce polymorphisme génétique dans la physiopathologie des infections par le virus de l'hépatite C.

## CHAPITRE 5 : CONCLUSION ET PERSPECTIVES

L'infection par le virus de l'hépatite C est un exemple rare en virologie « d'histoire résolue ». En effet, en moins de 30 ans, ce virus aura été découvert (1989), pris en charge tant au niveau diagnostique avec l'avènement simultané de la PCR, outil de sa découverte, qu'au niveau thérapeutique avec les premiers traitements immuno-modulateurs et sera très probablement éradiqué grâce à la révolution thérapeutique qui arrive. Néanmoins, cette histoire fulgurante à l'issue favorable ne doit pas conduire à un optimisme béat. Cette pathologie concerne actuellement plus de 180 millions de patients dont la majorité vit dans des pays à ressources limitées et pour lesquels l'accès au diagnostic d'une part et à ces thérapies d'autre part est probablement encore lointain. La particularité du VHC est sa diversité génétique, qui impacte sur la réponse au traitement, y compris sur les antiviraux spécifiques (DAA), et qui rend difficile la mise au point d'un vaccin. La compréhension des conséquences de cette diversité sur la physiopathologie reste alors d'un intérêt majeur.

Nos travaux se sont intéressés à une duplication du domaine V3 de la région NS5A découverte au laboratoire chez des souches de VHC de génotype 1b. Les virus sont des parasites intracellulaires obligatoires, et les virus à ARN portent en général des génomes de taille restreinte, avec une certaine notion d'économie et d'efficacité infectieuse. La présence de gènes dupliqués dans le génome viral pose alors la question de l'avantage éventuel procuré par cette portion génomique supplémentaire dans ce contexte d'économie. C'est pourquoi nous avons souhaité caractériser cette duplication du domaine V3 dans sa globalité : nous nous sommes intéressés à son éventuel impact clinique, à l'origine génétique de ce polymorphisme et à son possible rôle au niveau du cycle viral et de la biologie cellulaire du VHC.

Nous avons mis en œuvre une étude multicentrique associant 19 laboratoires de virologie français, qui a reçu le soutien du groupe hépatite C de l'AC11-virologie médicale en 2006. Nous avons inclus entre 2006 et 2009 les patients infectés par le VHC nouvellement génotypé 1b au sein de chaque laboratoire. Cette cohorte a constitué la base principale de notre étude. Nous avons utilisé en complément les données de séquences provenant de l'étude NAIF menée précédemment au laboratoire. Un total de 938 patients ont été identifiés par les laboratoires et ont été testés. Nous avons pu finalement inclure 821 patients. Ce premier travail nous a permis de confirmer l'existence de la duplication V3 au niveau d'une

importante cohorte française. Nous avons identifié une prévalence de 3,05% de la duplication V3 au sein de cette cohorte ce qui correspondait à 25 patients infectés par ce type de souche. Ce chiffre pourrait être considéré comme faible en terme de niveau de prévalence mais nous pensons qu'il n'est pas négligeable au regard de la biologie des virus à ARN. En effet, la duplication totale de gène conduisant à deux ORF en tandem a été rarement rapportée chez les virus à ARN ; à peine une dizaine de publications peuvent être recensées [335, 411-413]. De même quelques duplications partielles ont été identifiées dans les régions non codantes de flavivirus ou de tombusvirus [336, 414] ou dans les régions intergéniques du BVDV [415] ou du rotavirus du porc [337]. L'évolution des virus à ARN repose essentiellement sur les mutations ponctuelles associées à un niveau de réplication très élevé. En effet, un événement de duplication conduit à une augmentation de la taille du génome, une augmentation de la durée du cycle de réplication et une augmentation du risque d'erreur au cours de la réplication qui peut générer des virus défectifs [416]. Cet ensemble d'éléments apparaît défavorable à une transmission et une infectiosité efficaces.

Nos analyses moléculaires ont permis de montrer que cette duplication avait pour probable origine une recombinaison non-homologue issue de deux souches parentales de génotype 1b, nous estimons la survenue de cet événement il y environ 100 ans. Nos analyses réalisées sur les gènes NS5A et E1 étaient concordantes montrant un isolement de ces souches. La duplication de gène chez les virus à ARN est majoritairement issue de recombinaisons homologues [326, 327] ; les quelques souches recombinantes du VHC identifiés *in vivo* sont issus de ce type de mécanismes et confirment la possibilité de telle recombinaison chez le VHC [166]. Des travaux *in vitro* ont aussi validé ce mécanisme de recombinaison homologue avec un taux de recombinaison estimé à  $4.10^{-8}$  par nucléotide et  $4.10^{-4}$  par génome [331]. Sheel *et al.* ont démontré très récemment la possibilité de recombinaisons non-homologues chez le VHC à l'aide d'expériences *in vitro* réalisée avec des constructions JFH1 et J6 [417]. Notre étude permet de démontrer la réalité de ce type de recombinaison chez le VHC *in vivo* pour la première fois. En outre, nous montrons que nos souches recombinantes constituent la totalité de la quasiespèce, elles sont présentes dès l'infection et persistent jusqu'à plus de 10 ans chez les patients pour lesquels nous disposions de prélèvements consécutifs. Au regard de l'impact délétère des duplications et de la rareté des recombinaisons non-homologues [418] chez les virus à ARN, l'implantation de telles souches de VHC dans la population ne peut pas être considérée comme ayant une faible prévalence. De plus, cette implantation associée à la possibilité de transmission démontrée

chez un patient soulignent au moins l'absence d'impact négatif de la duplication sur la biologie du VHC, à défaut de souligner un quelconque avantage sélectif.

Au plan clinique, les recombinaisons peuvent conduire à un avantage sélectif avec l'émergence de mutants d'échappement ou peuvent permettre l'élimination de gènes délétères [327]. Les recombinaisons peuvent plus rarement conduire à l'émergence de nouvelles souches voire de nouveaux phylums ayant différentes pathogénèses. Par exemple, plusieurs recombinaisons entre souches du BDVD et gènes de l'hôte ont permis l'émergence d'une souche virulente cytopathogène létale chez l'hôte [419]. De même certains coronavirus du chat, qui recombinent entre eux, ont conduit à l'émergence de souches plus pathogènes à l'origine de péritonites graves [420]. Notre travail d'épidémiologie clinique avait pour but d'identifier un contexte particulier associé à ce type de souche. Or, le recueil de données cliniques a été le principal écueil de notre étude avec une part non négligeable de données manquantes, limitant les effectifs de certains groupes de patients. Néanmoins, les analyses statistiques ont montré une prévalence plus élevée des souches recombinantes chez les patients avec une pathologie hépatique sévère. Cette tendance observée dans notre étude se trouve aujourd'hui confortée par les résultats d'un autre travail mené à Brest, actuellement soumis à publication (Annexe 2). Cette équipe a réalisé le dépistage de la duplication V3 sur des échantillons provenant du centre de ressources biologiques de l'Hôpital Jean Verdier à Bondy et issus d'une cohorte de 72 patients infectés par le VHC de génotype 1b, bénéficiant d'un suivi longitudinal. Les patients étaient répartis en 3 groupes selon leur évolution clinique. Trois souches recombinantes ont été identifiées parmi les patients évoluant vers le CHC (n=28), deux souches recombinantes ont été dépistées chez les patients avec une cirrhose sans évolution vers le CHC (n=27) et aucune souche n'a été trouvée chez les patients avec une hépatite chronique avec une fibrose modérée (F1-F2). Les analyses menées sur les 2 domaines V3 issus des souches recombinantes du groupe de patients avec un CHC montraient une variabilité génétique supérieure à celle observée pour les souches recombinantes du groupe de patients avec cirrhose. La survenue d'un CHC semblait aussi associée à une évolution différente des caractéristiques génétiques et biochimiques du domaine V3 inséré. Cette étude a l'intérêt de faire appel à une cohorte de patients suivie dans le temps permettant une comparaison de l'évolution clinique des patients et une répartition en « groupe d'évolution » de la maladie hépatique. Malgré un nombre de patients plus faible par rapport à notre étude multicentrique, la même tendance est observée avec une prévalence de souches recombinantes plus élevée chez les patients présentant une plus grande sévérité de la



pathologie hépatique. Les analyses de variabilité génétique montrent un domaine V3-R1 différent du domaine V3-R2 et sont aussi en accord avec nos études moléculaires qui montraient un isolement reproductif du domaine V3 inséré.

La mise en évidence de partenaires protéiques spécifiques aux protéines NS5A dupliquées qui pourraient expliquer un éventuel rôle de la duplication dans la fibrose hépatique et le CHC, reste donc un challenge. Le crible en double-hybride levure nous a permis l'identification de partenaires différents entre les protéines NS5A dupliquées et la protéine NS5A sauvage. Certains partenaires retiennent tout particulièrement notre intérêt, comme l'ATF6, l'oncostatine M et le TP53BP2 car ces protéines participent aux voies de signalisation cellulaire impliquées dans l'apoptose et le contrôle de la carcinogenèse. Nous avons franchi un premier pas en réussissant la production des protéines NS5A marquées par des tags de type Flag à la fois en système cellulaire et en système bactérien. La production en cellule n'a pas été optimisée autant que souhaitée mais doit être retenue car elle conduit à des protéines ayant subies les modifications post-traductionnelles. Bien que les premiers pull-down n'aient pas été concluants, les analyses seront poursuivies avec les modifications techniques adaptées. Une approche fonctionnelle combinant plusieurs techniques s'avère impérative et nécessitera de nouvelles collaborations en focalisant probablement les analyses sur les voies de l'oncogenèse.

Nous avons découvert une duplication du domaine V3-NS5A chez des patients infectés par le VHC de génotype 1b. Nous montrons que ces souches sont issues d'une recombinaison non-homologue, mécanisme décrit pour la première fois *in vivo*. Ces souches se sont implantées dans la population, elles sont contemporaines de la contamination et persistent au cours de l'infection, ne montrant aucun impact délétère de ce type de recombinaison sur l'évolution du VHC. L'infection par ce type de souche pourrait être associée à une plus grande sévérité de la maladie hépatique avec une fréquence plus élevée de cirrhose et de CHC comme l'ont montré les études épidémiologiques réalisées à Angers et à Brest. Ce travail nécessite la recherche du rôle exact du domaine V3 dupliqué par la finalisation des études fonctionnelles.

## RÉFÉRENCES

1. Mohd Hanafiah K, Groeger J, Flaxman AD and Wiersma ST (2013) Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology* 57: 1333-1342.
2. Omland LH, Jepsen P, Krarup H, Schonning K, Lind B, et al. (2011) Increased mortality among persons infected with hepatitis C virus. *Clin Gastroenterol Hepatol* 9: 71-78.
3. Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, et al. (2014) Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology* 59: 318-327.
4. Beaumont E, Patient R, Hourieux C, Dimier-Poisson I and Roingeard P (2013) Chimeric hepatitis B virus/hepatitis C virus envelope proteins elicit broadly neutralizing antibodies and constitute a potential bivalent prophylactic vaccine. *Hepatology* 57: 1303-1313.
5. Law JL, Chen C, Wong J, Hockman D, Santer DM, et al. (2013) A hepatitis C virus (HCV) vaccine comprising envelope glycoproteins gpE1/gpE2 derived from a single isolate elicits broad cross-genotype neutralizing antibodies in humans. *PLoS One* 8: e59776.
6. Zein NN (2000) Clinical significance of hepatitis C virus genotypes. *Clin Microbiol Rev* 13: 223-235.
7. Di Marco V, Covolo L, Calvaruso V, Levrero M, Puoti M, et al. (2013) Who is more likely to respond to dual treatment with pegylated-interferon and ribavirin for chronic hepatitis C? A gender-oriented analysis. *J Viral Hepat* 20: 790-800.
8. Le Guillou-Guillemette H, Vallet S, Gaudy-Graffin C, Payan C, Pivert A, et al. (2007) Genetic diversity of the hepatitis C virus: impact and issues in the antiviral therapy. *World J Gastroenterol* 13: 2416-2426.
9. Pawlotsky JM (2013) Hepatitis C virus: standard-of-care treatment. *Adv Pharmacol* 67: 169-215.
10. Veillon P, Payan C, Le Guillou-Guillemette H, Gaudy C and Lunel F (2007) Quasispecies evolution in NS5A region of hepatitis C virus genotype 1b during interferon or combined interferon-ribavirin therapy. *World J Gastroenterol* 13: 1195-1203.
11. Lunel-Fabiani F VP, Le Guillou-Guillemette H, Redureau J, Coeffic B, Subra JF, Causse X, Dubois P, Blanchi A, Gaudy C, Payan C (2005) Duplication of the V3 domain in Nonstructural 5A Protein related to Hepatitis C Virus extrahepatic manifestations. In: Abstracts, editor editors. 45th Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC). Washington, DC, USA: American Society for Microbiology. pp. 485.
12. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, et al. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 244: 359-362.
13. Houghton M (2009) The long and winding road leading to the identification of the hepatitis C virus. *J Hepatol* 51: 939-948.
14. Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, et al. (1990) Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc Natl Acad Sci U S A* 87: 9524-9528.
15. Simmonds P (1999) Viral heterogeneity of the hepatitis C virus. *J Hepatol* 31 Suppl 1: 54-60.
16. King AM, Lefkowitz E, Adams Michael J and Carstens EB (2012) *Virus Taxonomy : Ninth Report of the International Committee on Taxonomy of Viruses*. London. 1338 p.
17. Simons JN, Leary TP, Dawson GJ, Pilot-Matias TJ, Muerhoff AS, et al. (1995) Isolation of novel virus-like sequences associated with human hepatitis. *Nat Med* 1: 564-569.

18. Simons JN, Pilot-Matias TJ, Leary TP, Dawson GJ, Desai SM, et al. (1995) Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci U S A* 92: 3401-3405.
19. Burbelo PD, Dubovi EJ, Simmonds P, Medina JL, Henriquez JA, et al. (2012) Serology-enabled discovery of genetically diverse hepaciviruses in a new host. *J Virol* 86: 6171-6178.
20. Stapleton JT, Fong S, Muerhoff AS, Bukh J and Simmonds P (2011) The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. *J Gen Virol* 92: 233-246.
21. Kapoor A, Simmonds P, Cullen JM, Scheel TK, Medina JL, et al. (2013) Identification of a pegivirus (GB virus-like virus) that infects horses. *J Virol* 87: 7185-7190.
22. Simmonds P (2013) The origin of hepatitis C virus. *Curr Top Microbiol Immunol* 369: 1-15.
23. Kaito M, Watanabe S, Tsukiyama-Kohara K, Yamaguchi K, Kobayashi Y, et al. (1994) Hepatitis C virus particle detected by immunoelectron microscopic study. *J Gen Virol* 75 ( Pt 7): 1755-1760.
24. Feneant L, Levy S and Cocquerel L (2014) CD81 and hepatitis C virus (HCV) infection. *Viruses* 6: 535-572.
25. Moradpour D, Cerny A, Heim MH and Blum HE (2001) Hepatitis C: an update. *Swiss Med Wkly* 131: 291-298.
26. Dustin LB and Rice CM (2007) Flying under the radar: the immunobiology of hepatitis C. *Annu Rev Immunol* 25: 71-99.
27. Wang L, Jeng KS and Lai MM Poly(C)-Binding Protein 2 Interacts with Sequences Required for Viral Replication in the Hepatitis C Virus (HCV) 5' Untranslated Region and Directs HCV RNA Replication through Circularizing the Viral Genome. *J Virol* 85: 7954-7964.
28. Weinlich S, Huttelmaier S, Schierhorn A, Behrens SE, Ostareck-Lederer A, et al. (2009) IGF2BP1 enhances HCV IRES-mediated translation initiation via the 3'UTR. *RNA* 15: 1528-1542.
29. Pestova TV, de Breyne S, Pisarev AV, Abaeva IS and Hellen CU (2008) eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. *EMBO J* 27: 1060-1072.
30. Terenin IM, Dmitriev SE, Andreev DE and Shatsky IN (2008) Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nat Struct Mol Biol* 15: 836-841.
31. Ali N, Pruijn GJ, Kenan DJ, Keene JD and Siddiqui A (2000) Human La antigen is required for the hepatitis C virus internal ribosome entry site-mediated translation. *J Biol Chem* 275: 27531-27540.
32. Ali N and Siddiqui A (1997) The La antigen binds 5' noncoding region of the hepatitis C virus RNA in the context of the initiator AUG codon and stimulates internal ribosome entry site-mediated translation. *Proc Natl Acad Sci U S A* 94: 2249-2254.
33. Cordes S, Kusov Y, Heise T and Gauss-Muller V (2008) La autoantigen suppresses IRES-dependent translation of the hepatitis A virus. *Biochem Biophys Res Commun* 368: 1014-1019.
34. Hahm B, Kim YK, Kim JH, Kim TY and Jang SK (1998) Heterogeneous nuclear ribonucleoprotein L interacts with the 3' border of the internal ribosomal entry site of hepatitis C virus. *J Virol* 72: 8782-8788.
35. Hwang B, Lim JH, Hahm B, Jang SK and Lee SW (2009) hnRNP L is required for the translation mediated by HCV IRES. *Biochem Biophys Res Commun* 378: 584-588.

36. Gontarek RR, Gutshall LL, Herold KM, Tsai J, Sathe GM, et al. (1999) hnRNP C and polypyrimidine tract-binding protein specifically interact with the pyrimidine-rich region within the 3'NTR of the HCV RNA genome. *Nucleic Acids Res* 27: 1457-1463.
37. Niepmann M (2009) Internal translation initiation of picornaviruses and hepatitis C virus. *Biochim Biophys Acta* 1789: 529-541.
38. Friebe P and Bartenschlager R (2002) Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. *J Virol* 76: 5326-5338.
39. Song Y, Friebe P, Tzima E, Junemann C, Bartenschlager R, et al. (2006) The hepatitis C virus RNA 3'-untranslated region strongly enhances translation directed by the internal ribosome entry site. *J Virol* 80: 11579-11588.
40. Berry KE, Waghray S and Doudna JA (2010) The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit. *RNA* 16: 1559-1569.
41. Bung C, Bochkayeva Z, Terenin I, Zinovkin R, Shatsky IN, et al. (2010) Influence of the hepatitis C virus 3'-untranslated region on IRES-dependent and cap-dependent translation initiation. *FEBS Lett* 584: 837-842.
42. Bartenschlager R, Lohmann V and Penin F (2013) The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. *Nat Rev Microbiol* 11: 482-496.
43. Dubuisson J (2007) Hepatitis C virus proteins. *World J Gastroenterol* 13: 2406-2415.
44. Appel N, Zayas M, Miller S, Krijnse-Locker J, Schaller T, et al. (2008) Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog* 4: e1000035.
45. Gawlik K, Baugh J, Chatterji U, Lim PJ, Bobardt MD, et al. (2014) HCV core residues critical for infectivity are also involved in core-NS5A complex formation. *PLoS One* 9: e88866.
46. McLauchlan J (2009) Lipid droplets and hepatitis C virus infection. *Biochim Biophys Acta* 1791: 552-559.
47. Banerjee A, Ray RB and Ray R (2010) Oncogenic potential of hepatitis C virus proteins. *Viruses* 2: 2108-2133.
48. Vieyres G, Thomas X, Descamps V, Duverlie G, Patel AH, et al. (2010) Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. *J Virol* 84: 10159-10168.
49. Roccasecca R, Ansuini H, Vitelli A, Meola A, Scarselli E, et al. (2003) Binding of the hepatitis C virus E2 glycoprotein to CD81 is strain specific and is modulated by a complex interplay between hypervariable regions 1 and 2. *J Virol* 77: 1856-1867.
50. Zeisel MB, Fofana I, Fafi-Kremer S and Baumert TF (2011) Hepatitis C virus entry into hepatocytes: molecular mechanisms and targets for antiviral therapies. *J Hepatol* 54: 566-576.
51. Khaliq S, Jahan S and Hassan S (2011) Hepatitis C virus p7: molecular function and importance in hepatitis C virus life cycle and potential antiviral target. *Liver Int* 31: 606-617.
52. Griffin SD, Beales LP, Clarke DS, Worsfold O, Evans SD, et al. (2003) The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. *FEBS Lett* 535: 34-38.
53. Griffin SD, Harvey R, Clarke DS, Barclay WS, Harris M, et al. (2004) A conserved basic loop in hepatitis C virus p7 protein is required for amantadine-sensitive ion channel activity in mammalian cells but is dispensable for localization to mitochondria. *J Gen Virol* 85: 451-461.
54. Griffin S (2010) Inhibition of HCV p7 as a therapeutic target. *Curr Opin Investig Drugs* 11: 175-181.

55. Welbourn S and Pause A (2007) The hepatitis C virus NS2/3 protease. *Curr Issues Mol Biol* 9: 63-69.
56. Kalinina O, Norder H, Mukomolov S and Magnius LO (2002) A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol* 76: 4034-4043.
57. Pietschmann T, Kaul A, Koutsoudakis G, Shavinskaya A, Kallis S, et al. (2006) Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *Proc Natl Acad Sci U S A* 103: 7408-7413.
58. Legrand-Abravanel F, Claudinon J, Nicot F, Dubois M, Chapuy-Regaud S, et al. (2007) New natural intergenotypic (2/5) recombinant of hepatitis C virus. *J Virol* 81: 4357-4362.
59. Yi M, Ma Y, Yates J and Lemon SM (2007) Compensatory mutations in E1, p7, NS2, and NS3 enhance yields of cell culture-infectious intergenotypic chimeric hepatitis C virus. *J Virol* 81: 629-638.
60. Dentzer TG, Lorenz IC, Evans MJ and Rice CM (2009) Determinants of the hepatitis C virus nonstructural protein 2 protease domain required for production of infectious virus. *J Virol* 83: 12702-12713.
61. Jones CT, Murray CL, Eastman DK, Tassello J and Rice CM (2007) Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus. *J Virol* 81: 8374-8383.
62. Popescu CI, Callens N, Trinel D, Roingeard P, Moradpour D, et al. (2011) NS2 protein of hepatitis C virus interacts with structural and non-structural proteins towards virus assembly. *PLoS Pathog* 7: e1001278.
63. Ma Y, Anantpadma M, Timpe JM, Shanmugam S, Singh SM, et al. (2011) Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins. *J Virol* 85: 86-97.
64. Erdtmann L, Franck N, Lerat H, Le Seyec J, Gilot D, et al. (2003) The hepatitis C virus NS2 protein is an inhibitor of CIDE-B-induced apoptosis. *J Biol Chem* 278: 18256-18264.
65. Lin C (2006) HCV NS3-4A Serine Protease.
66. Raney KD, Sharma SD, Moustafa IM and Cameron CE (2010) Hepatitis C virus non-structural protein 3 (HCV NS3): a multifunctional antiviral target. *J Biol Chem* 285: 22725-22731.
67. Meylan E, Curran J, Hofmann K, Moradpour D, Binder M, et al. (2005) Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* 437: 1167-1172.
68. Lin R, Lacoste J, Nakhaei P, Sun Q, Yang L, et al. (2006) Dissociation of a MAVS/IPS-1/VISA/Cardif-IKKepsilon molecular complex from the mitochondrial outer membrane by hepatitis C virus NS3-4A proteolytic cleavage. *J Virol* 80: 6072-6083.
69. Ogata S, Florese RH, Nagano-Fujii M, Hidajat R, Deng L, et al. (2003) Identification of hepatitis C virus (HCV) subtype 1b strains that are highly, or only weakly, associated with hepatocellular carcinoma on the basis of the secondary structure of an amino-terminal portion of the HCV NS3 protein. *J Clin Microbiol* 41: 2835-2841.
70. Ogata S, Nagano-Fujii M, Ku Y, Yoon S and Hotta H (2002) Comparative sequence analysis of the core protein and its frameshift product, the F protein, of hepatitis C virus subtype 1b strains obtained from patients with and without hepatocellular carcinoma. *J Clin Microbiol* 40: 3625-3630.
71. Kasprzak A, Adamek A, Przybyszewska W, Olejniczak K, Biczysko W, et al. (2009) p21/Waf1/Cipl cellular expression in chronic long-lasting hepatitis C: correlation with HCV proteins (C, NS3, NS5A), other cell-cycle related proteins and selected clinical data. *Folia Histochem Cytobiol* 47: 385-394.

72. Chatel-Chaix L, Baril M and Lamarre D (2010) Hepatitis C Virus NS3/4A Protease Inhibitors: A Light at the End of the Tunnel. *Viruses* 2: 1752-1765.
73. Poordad F, McCone J, Jr., Bacon BR, Bruno S, Manns MP, et al. (2011) Boceprevir for untreated chronic HCV genotype 1 infection. *N Engl J Med* 364: 1195-1206.
74. Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, et al. (2009) Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med* 360: 1839-1850.
75. Jacobson IM, McHutchison JG, Dusheiko G, Di Bisceglie AM, Reddy KR, et al. (2011) Telaprevir for previously untreated chronic hepatitis C virus infection. *N Engl J Med* 364: 2405-2416.
76. Jones DM, Patel AH, Targett-Adams P and McLauchlan J (2009) The hepatitis C virus NS4B protein can trans-complement viral RNA replication and modulates production of infectious virus. *J Virol* 83: 2163-2177.
77. Einav S, Elazar M, Danieli T and Glenn JS (2004) A nucleotide binding motif in hepatitis C virus (HCV) NS4B mediates HCV RNA replication. *J Virol* 78: 11288-11295.
78. Einav S, Gerber D, Bryson PD, Sklan EH, Elazar M, et al. (2008) Discovery of a hepatitis C target and its pharmacological inhibitors by microfluidic affinity analysis. *Nat Biotechnol* 26: 1019-1027.
79. Rai R and Deval J (2011) New opportunities in anti-hepatitis C virus drug discovery: targeting NS4B. *Antiviral Res* 90: 93-101.
80. Ranjith-Kumar CT, Sarisky RT, Gutshall L, Thomson M and Kao CC (2004) De novo initiation pocket mutations have multiple effects on hepatitis C virus RNA-dependent RNA polymerase activities. *J Virol* 78: 12207-12217.
81. Bressanelli S, Tomei L, Rey FA and De Francesco R (2002) Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *J Virol* 76: 3482-3492.
82. Vermehren J and Sarrazin C (2011) New HCV therapies on the horizon. *Clin Microbiol Infect* 17: 122-134.
83. Branch AD, Stump DD, Gutierrez JA, Eng F and Walewski JL (2005) The hepatitis C virus alternate reading frame (ARF) and its family of novel products: the alternate reading frame protein/F-protein, the double-frameshift protein, and others. *Semin Liver Dis* 25: 105-117.
84. Vassilaki N and Mavromara P (2009) The HCV ARFP/F/core+1 protein: production and functional analysis of an unconventional viral product. *IUBMB Life* 61: 739-752.
85. Dalagiorgou G, Vassilaki N, Foka P, Boumlic A, Kakkanas A, et al. (2011) High levels of HCV core+1 antibodies in HCV patients with hepatocellular carcinoma. *J Gen Virol* 92: 1343-1351.
86. Lanford RE, Sureau C, Jacob JR, White R and Fuerst TR (1994) Demonstration of in vitro infection of chimpanzee hepatocytes with hepatitis C virus using strand-specific RT/PCR. *Virology* 202: 606-614.
87. Ito T, Mukaigawa J, Zuo J, Hirabayashi Y, Mitamura K, et al. (1996) Cultivation of hepatitis C virus in primary hepatocyte culture from patients with chronic hepatitis C results in release of high titre infectious virus. *J Gen Virol* 77 ( Pt 5): 1043-1054.
88. Uprichard SL (2010) Hepatitis C virus experimental model systems and antiviral drug research. *Virol Sin* 25: 227-245.
89. Lohmann V, Korner F, Koch J, Herian U, Theilmann L, et al. (1999) Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* 285: 110-113.
90. Blight KJ, Kolykhalov AA and Rice CM (2000) Efficient initiation of HCV RNA replication in cell culture. *Science* 290: 1972-1974.

91. Moradpour D, Evans MJ, Gosert R, Yuan Z, Blum HE, et al. (2004) Insertion of green fluorescent protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus replication complexes. *J Virol* 78: 7400-7409.
92. Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, et al. (2005) Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* 11: 791-796.
93. Lohmann V and Bartenschlager R (2013) Hepatitis C virus replicons volume 3 and 4. *Gastroenterology* 144: 13-15.
94. Voisset C and Dubuisson J (2004) Functional hepatitis C virus envelope glycoproteins. *Biol Cell* 96: 413-420.
95. Zhong J, Gastaminza P, Cheng G, Kapadia S, Kato T, et al. (2005) Robust hepatitis C virus infection in vitro. *Proc Natl Acad Sci U S A* 102: 9294-9299.
96. Lindenbach BD, Evans MJ, Syder AJ, Wolk B, Tellinghuisen TL, et al. (2005) Complete replication of hepatitis C virus in cell culture. *Science* 309: 623-626.
97. Scheel TK, Gottwein JM, Carlsen TH, Li YP, Jensen TB, et al. (2011) Efficient culture adaptation of hepatitis C virus recombinants with genotype-specific core-NS2 by using previously identified mutations. *J Virol* 85: 2891-2906.
98. Scheel TK, Gottwein JM, Mikkelsen LS, Jensen TB and Bukh J (2011) Recombinant HCV variants with NS5A from genotypes 1-7 have different sensitivities to an NS5A inhibitor but not interferon-alpha. *Gastroenterology* 140: 1032-1042.
99. Gottwein JM, Jensen TB, Mathiesen CK, Meuleman P, Serre SB, et al. (2011) Development and application of hepatitis C reporter viruses with genotype 1 to 7 core-nonstructural protein 2 (NS2) expressing fluorescent proteins or luciferase in modified JFH1 NS5A. *J Virol* 85: 8913-8928.
100. Gottwein JM, Scheel TK, Jensen TB, Ghanem L and Bukh J (2011) Differential efficacy of protease inhibitors against HCV genotypes 2a, 3a, 5a, and 6a NS3/4A protease recombinant viruses. *Gastroenterology* 141: 1067-1079.
101. Sainz B, Jr., Barretto N, Yu X, Corcoran P and Uprichard SL (2012) Permissiveness of human hepatoma cell lines for HCV infection. *Virol J* 9: 30.
102. Lohmann V and Bartenschlager R (2013) On the History of Hepatitis C Virus Cell Culture Systems. *J Med Chem*.
103. Meuleman P and Leroux-Roels G (2009) HCV animal models: a journey of more than 30 years. *Viruses* 1: 222-240.
104. Houghton M (2011) Prospects for prophylactic and therapeutic vaccines against the hepatitis C viruses. *Immunol Rev* 239: 99-108.
105. Ducoulombier D, Roque-Afonso AM, Di Liberto G, Penin F, Kara R, et al. (2004) Frequent compartmentalization of hepatitis C virus variants in circulating B cells and monocytes. *Hepatology* 39: 817-825.
106. Vargas HE, Laskus T, Radkowski M, Wilkinson J, Balan V, et al. (2002) Detection of hepatitis C virus sequences in brain tissue obtained in recurrent hepatitis C after liver transplantation. *Liver Transpl* 8: 1014-1019.
107. Manns MP, Foster GR, Rockstroh JK, Zeuzem S, Zoulim F, et al. (2007) The way forward in HCV treatment--finding the right path. *Nat Rev Drug Discov* 6: 991-1000.
108. Moriishi K and Matsuura Y (2012) Exploitation of lipid components by viral and host proteins for hepatitis C virus infection. *Front Microbiol* 3: 54.
109. Tang H and Grise H (2009) Cellular and molecular biology of HCV infection and hepatitis. *Clin Sci (Lond)* 117: 49-65.
110. Blanchard E, Belouzard S, Goueslain L, Wakita T, Dubuisson J, et al. (2006) Hepatitis C virus entry depends on clathrin-mediated endocytosis. *J Virol* 80: 6964-6972.

111. Tsukiyama-Kohara K, Iizuka N, Kohara M and Nomoto A (1992) Internal ribosome entry site within hepatitis C virus RNA. *J Virol* 66: 1476-1483.
112. Ploss A and Evans MJ (2012) Hepatitis C virus host cell entry. *Curr Opin Virol* 2: 14-19.
113. Gosert R, Jendrszczok W, Berke JM, Brass V, Blum HE, et al. (2005) Characterization of nonstructural protein membrane anchor deletion mutants expressed in the context of the hepatitis C virus polyprotein. *J Virol* 79: 7911-7917.
114. Ferraris P, Blanchard E and Roingard P (2010) Ultrastructural and biochemical analyses of hepatitis C virus-associated host cell membranes. *J Gen Virol* 91: 2230-2237.
115. Masaki T, Suzuki R, Murakami K, Aizaki H, Ishii K, et al. (2008) Interaction of hepatitis C virus nonstructural protein 5A with core protein is critical for the production of infectious virus particles. *J Virol* 82: 7964-7976.
116. Bartenschlager R, Penin F, Lohmann V and Andre P (2011) Assembly of infectious hepatitis C virus particles. *Trends Microbiol* 19: 95-103.
117. Vogt DA, Camus G, Herker E, Webster BR, Tsou CL, et al. (2013) Lipid droplet-binding protein TIP47 regulates hepatitis C Virus RNA replication through interaction with the viral NS5A protein. *PLoS Pathog* 9: e1003302.
118. Benga WJ, Krieger SE, Dimitrova M, Zeisel MB, Parnot M, et al. (2010) Apolipoprotein E interacts with hepatitis C virus nonstructural protein 5A and determines assembly of infectious particles. *Hepatology* 51: 43-53.
119. Jirasko V, Montserret R, Lee JY, Gouttenoire J, Moradpour D, et al. (2010) Structural and functional studies of nonstructural protein 2 of the hepatitis C virus reveal its key role as organizer of virion assembly. *PLoS Pathog* 6: e1001233.
120. Phan T, Beran RK, Peters C, Lorenz IC and Lindenbach BD (2009) Hepatitis C virus NS2 protein contributes to virus particle assembly via opposing epistatic interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes. *J Virol* 83: 8379-8395.
121. Andre P, Komurian-Pradel F, Deforges S, Perret M, Berland JL, et al. (2002) Characterization of low- and very-low-density hepatitis C virus RNA-containing particles. *J Virol* 76: 6919-6928.
122. Perz JF FL, Pecoraro C, Hutin YJF, Armstrong GL (2004) Estimated global prevalence of hepatitis C virus infection. In: A. Infectious Diseases Society of America, VA editor editors. 42nd Annual Meeting of the Infectious Diseases Society of America. Boston, MA, USA.
123. Alter MJ (2007) Epidemiology of hepatitis C virus infection. *World J Gastroenterol* 13: 2436-2441.
124. Pillonel J, Brouard C, Laperche S, Barin F, Bernillon P, et al. (2009) [Quantitative estimate of the risk of blood donation contamination by infectious agents]. *Transfus Clin Biol* 16: 138-145.
125. Indolfi G and Resti M (2009) Perinatal transmission of hepatitis C virus infection. *J Med Virol* 81: 836-843.
126. Polis CB, Shah SN, Johnson KE and Gupta A (2007) Impact of maternal HIV coinfection on the vertical transmission of hepatitis C virus: a meta-analysis. *Clin Infect Dis* 44: 1123-1131.
127. Chayama K, Kobayashi M, Tsubota A, Koida I, Arase Y, et al. (1995) Molecular analysis of intraspousal transmission of hepatitis C virus. *J Hepatol* 22: 431-439.
128. Zylberberg H, Thiers V, Lagorce D, Squadrito G, Leone F, et al. (1999) Epidemiological and virological analysis of couples infected with hepatitis C virus. *Gut* 45: 112-116.
129. Tahan V, Karaca C, Yildirim B, Bozbas A, Ozaras R, et al. (2005) Sexual transmission of HCV between spouses. *Am J Gastroenterol* 100: 821-824.



130. Vandelli C, Renzo F, Romano L, Tisminetzky S, De Palma M, et al. (2004) Lack of evidence of sexual transmission of hepatitis C among monogamous couples: results of a 10-year prospective follow-up study. *Am J Gastroenterol* 99: 855-859.
131. Terrault NA (2002) Sexual activity as a risk factor for hepatitis C. *Hepatology* 36: S99-105.
132. van de Laar TJ, Matthews GV, Prins M and Danta M (2010) Acute hepatitis C in HIV-infected men who have sex with men: an emerging sexually transmitted infection. *AIDS* 24: 1799-1812.
133. Larsen C, Chaix ML, Le Strat Y, Velter A, Gervais A, et al. (2011) Gaining greater insight into HCV emergence in HIV-infected men who have sex with men: the HEPAIG Study. *PLoS One* 6: e29322.
134. Chang TT, Liou TC, Young KC, Lin XZ, Lin CY, et al. (1994) Intrafamilial transmission of hepatitis C virus: the important role of inapparent transmission. *J Med Virol* 42: 91-96.
135. Muhlberger N, Schwarzer R, Lettmeier B, Sroczynski G, Zeuzem S, et al. (2009) HCV-related burden of disease in Europe: a systematic assessment of incidence, prevalence, morbidity, and mortality. *BMC Public Health* 9: 34.
136. Omland LH, Christensen PB, Krarup H, Jepsen P, Weis N, et al. (2011) Mortality among Patients with Cleared Hepatitis C Virus Infection Compared to the General Population: A Danish Nationwide Cohort Study. *PLoS One* 6: e22476.
137. Kershenovich D, Razavi HA, Sanchez-Avila JF, Bessone F, Coelho HS, et al. (2011) Trends and projections of hepatitis C virus epidemiology in Latin America. *Liver Int* 31 Suppl 2: 18-29.
138. Cornberg M, Razavi HA, Alberti A, Bernasconi E, Buti M, et al. (2011) A systematic review of hepatitis C virus epidemiology in Europe, Canada and Israel. *Liver Int* 31 Suppl 2: 30-60.
139. Sievert W, Altraif I, Razavi HA, Abdo A, Ahmed EA, et al. (2011) A systematic review of hepatitis C virus epidemiology in Asia, Australia and Egypt. *Liver Int* 31 Suppl 2: 61-80.
140. Meffre C, Le Strat Y, Delarocque-Astagneau E, Dubois F, Antona D, et al. (2010) Prevalence of hepatitis B and hepatitis C virus infections in France in 2004: social factors are important predictors after adjusting for known risk factors. *J Med Virol* 82: 546-555.
141. Delarocque-Astagneau E, Meffre C, Dubois F, Pioche C, Le Strat Y, et al. (2010) The impact of the prevention programme of hepatitis C over more than a decade: the French experience. *J Viral Hepat* 17: 435-443.
142. Meffre C, Strat YL, Delarocque-Astagneau E, Antona D and Desenclos JC (2005) Estimation des taux de prévalence des anticorps anti-VHC et des marqueurs du virus de l'hépatite B chez les assurés sociaux du régime général de France métropolitaine, 2003-2004. Analyse descriptive, janvier 2005. INVS.
143. Simmonds P (2001) The origin and evolution of hepatitis viruses in humans. *J Gen Virol* 82: 693-712.
144. INVS (2008) Surveillance nationale de l'hépatite C à partir des pôles de référence volontaires / Données épidémiologiques 2001-2007.
145. Duarte EA, Novella IS, Weaver SC, Domingo E, Wain-Hobson S, et al. (1994) RNA virus quasispecies: significance for viral disease and epidemiology. *Infect Agents Dis* 3: 201-214.
146. Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, et al. (1998) Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science* 282: 103-107.

147. Smith DB, Pathirana S, Davidson F, Lawlor E, Power J, et al. (1997) The origin of hepatitis C virus genotypes. *J Gen Virol* 78 ( Pt 2): 321-328.
148. Domingo E, Escarmis C, Sevilla N and Baranowski E (1998) Population dynamics in the evolution of RNA viruses. *Adv Exp Med Biol* 440: 721-727.
149. Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, et al. (2005) Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* 42: 962-973.
150. Kuiken C and Simmonds P (2009) Nomenclature and numbering of the hepatitis C virus. *Methods Mol Biol* 510: 33-53.
151. Markov PV, Pepin J, Frost E, Deslandes S, Labbe AC, et al. (2009) Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *J Gen Virol* 90: 2086-2096.
152. Magiorkinis G, Magiorkinis E, Paraskevis D, Ho SY, Shapiro B, et al. (2009) The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. *PLoS Med* 6: e1000198.
153. Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, et al. The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC Evol Biol* 11: 131.
154. Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, et al. (2011) The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC Evol Biol* 11: 131.
155. Martell M, Esteban JI, Quer J, Genesca J, Weiner A, et al. (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol* 66: 3225-3229.
156. Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58: 465-523.
157. Eigen M and Schuster P (1977) The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* 64: 541-565.
158. Wilke CO (2005) Quasispecies theory in the context of population genetics. *BMC Evol Biol* 5: 44.
159. Fishman SL and Branch AD (2009) The quasispecies nature and biological implications of the hepatitis C virus. *Infect Genet Evol* 9: 1158-1167.
160. Bukh J, Purcell RH and Miller RH (1992) Sequence analysis of the 5' noncoding region of hepatitis C virus. *Proc Natl Acad Sci U S A* 89: 4942-4946.
161. Simmonds P, Smith DB, McOmish F, Yap PL, Kolberg J, et al. (1994) Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *J Gen Virol* 75 ( Pt 5): 1053-1061.
162. Holland JJ, De La Torre JC and Steinhauer DA (1992) RNA virus populations as quasispecies. *Curr Top Microbiol Immunol* 176: 1-20.
163. Gray RR, Salemi M, Klenerman P and Pybus OG (2012) A new evolutionary model for hepatitis C virus chronic infection. *PLoS Pathog* 8: e1002656.
164. Chevaliez S, Rodriguez C and Pawlotsky JM (2012) New virologic tools for management of chronic hepatitis B and C. *Gastroenterology* 142: 1303-1313 e1301.
165. Gonzalez-Candelas F, Lopez-Labrador FX and Bracho MA (2011) Recombination in hepatitis C virus. *Viruses* 3: 2006-2024.
166. Morel V, Fournier C, Francois C, Brochet E, Helle F, et al. (2011) Genetic recombination of the hepatitis C virus: clinical implications. *J Viral Hepat* 18: 77-83.
167. Maheshwari A and Thuluvath PJ (2010) Management of acute hepatitis C. *Clin Liver Dis* 14: 169-176; x.
168. Chen SL and Morgan TR (2006) The natural history of hepatitis C virus (HCV) infection. *Int J Med Sci* 3: 47-52.

169. Kamal SM (2008) Acute hepatitis C: a systematic review. *Am J Gastroenterol* 103: 1283-1297; quiz 1298.
170. Jang JY and Chung RT (2011) Chronic hepatitis C. *Gut Liver* 5: 117-132.
171. BEH (2003) *Epidémiologie de l'hépatite C: état des lieux*. N°16-17: 85-104.
172. Ascione A, Tartaglione T and Di Costanzo GG (2007) Natural history of chronic hepatitis C virus infection. *Dig Liver Dis* 39 Suppl 1: S4-7.
173. Thein HH, Walter SR, Gidding HF, Amin J, Law MG, et al. (2011) Trends in incidence of hepatocellular carcinoma after diagnosis of hepatitis B or C infection: a population-based cohort study, 1992-2007. *J Viral Hepat* 18: e232-241.
174. D'Amico G, Garcia-Tsao G and Pagliaro L (2006) Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol* 44: 217-231.
175. Perz JF, Armstrong GL, Farrington LA, Hutin YJ and Bell BP (2006) The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol* 45: 529-538.
176. Kawai T and Akira S (2007) Antiviral signaling through pattern recognition receptors. *J Biochem* 141: 137-145.
177. Wang N, Liang Y, Devaraj S, Wang J, Lemon SM, et al. (2009) Toll-like receptor 3 mediates establishment of an antiviral state against hepatitis C virus in hepatoma cells. *J Virol* 83: 9824-9834.
178. Eisenacher K, Steinberg C, Reindl W and Krug A (2007) The role of viral nucleic acid recognition in dendritic cells for innate and adaptive antiviral immunity. *Immunobiology* 212: 701-714.
179. Takahasi K, Yoneyama M, Nishihori T, Hirai R, Kumeta H, et al. (2008) Nonspecific RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol Cell* 29: 428-440.
180. Saito T, Owen DM, Jiang F, Marcotrigiano J and Gale M, Jr. (2008) Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature* 454: 523-527.
181. Bouttier M, Goncalves C, Journo C, Letienne J, Pina M, et al. (2008) Virus et interféron: mécanismes d'induction et stratégies d'échappement. *Virologie* 12: 159-173.
182. Ireton RC and Gale M, Jr. (2011) RIG-I like receptors in antiviral immunity and therapeutic applications. *Viruses* 3: 906-919.
183. Saito T and Gale M, Jr. (2007) Principles of intracellular viral recognition. *Curr Opin Immunol* 19: 17-23.
184. Samuel CE (2001) Antiviral actions of interferons. *Clin Microbiol Rev* 14: 778-809, table of contents.
185. Stetson DB and Medzhitov R (2006) Type I interferons in host defense. *Immunity* 25: 373-381.
186. Horner SM and Gale M, Jr. (2009) Intracellular innate immune cascades and interferon defenses that control hepatitis C virus. *J Interferon Cytokine Res* 29: 489-498.
187. Cheent K and Khakoo SI (2011) Natural killer cells and hepatitis C: action and reaction. *Gut* 60: 268-278.
188. Foy E, Li K, Sumpter R, Jr., Loo YM, Johnson CL, et al. (2005) Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-I signaling. *Proc Natl Acad Sci U S A* 102: 2986-2991.
189. Horner SM, Park HS and Gale M, Jr. (2012) Control of innate immune signaling and membrane targeting by the Hepatitis C virus NS3/4A protease are governed by the NS3 helix alpha0. *J Virol* 86: 3112-3120.

190. Loo YM, Owen DM, Li K, Erickson AK, Johnson CL, et al. (2006) Viral and therapeutic control of IFN-beta promoter stimulator 1 during hepatitis C virus infection. *Proc Natl Acad Sci U S A* 103: 6001-6006.
191. Larrea E, Aldabe R, Molano E, Fernandez-Rodriguez CM, Ametzazurra A, et al. (2006) Altered expression and activation of signal transducers and activators of transcription (STATs) in hepatitis C virus infection: in vivo and in vitro studies. *Gut* 55: 1188-1196.
192. de Lucas S, Bartolome J and Carreno V (2005) Hepatitis C virus core protein down-regulates transcription of interferon-induced antiviral genes. *J Infect Dis* 191: 93-99.
193. Taguchi T, Nagano-Fujii M, Akutsu M, Kadoya H, Ohgimoto S, et al. (2004) Hepatitis C virus NS5A protein interacts with 2',5'-oligoadenylate synthetase and inhibits antiviral activity of IFN in an IFN sensitivity-determining region-independent manner. *J Gen Virol* 85: 959-969.
194. Amadei B, Urbani S, Cazaly A, Fisticaro P, Zerbini A, et al. (2010) Activation of natural killer cells during acute infection with hepatitis C virus. *Gastroenterology* 138: 1536-1545.
195. Lin AW, Gonzalez SA, Cunningham-Rundles S, Dorante G, Marshall S, et al. (2004) CD56(+dim) and CD56(+bright) cell activation and apoptosis in hepatitis C virus infection. *Clin Exp Immunol* 137: 408-416.
196. Tseng CT and Klimpel GR (2002) Binding of the hepatitis C virus envelope protein E2 to CD81 inhibits natural killer cell functions. *J Exp Med* 195: 43-49.
197. Thimme R, Bukh J, Spangenberg HC, Wieland S, Pemberton J, et al. (2002) Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease. *Proc Natl Acad Sci U S A* 99: 15661-15668.
198. Amini M and Poustchi H (2012) Hepatitis C virus spontaneous clearance: immunology and genetic variance. *Viral Immunol* 25: 241-248.
199. Khakoo SI, Thio CL, Martin MP, Brooks CR, Gao X, et al. (2004) HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* 305: 872-874.
200. Barnaba V (2010) Hepatitis C virus infection: a "liaison a trois" amongst the virus, the host, and chronic low-level inflammation for human survival. *J Hepatol* 53: 752-761.
201. Jacobson IM, Cacoub P, Dal Maso L, Harrison SA and Younossi ZM (2010) Manifestations of chronic hepatitis C virus infection beyond the liver. *Clin Gastroenterol Hepatol* 8: 1017-1029.
202. Agnello V (1995) The aetiology of mixed cryoglobulinaemia associated with hepatitis C virus infection. *Scand J Immunol* 42: 179-184.
203. Vassilopoulos D and Calabrese LH (2002) Hepatitis C virus infection and vasculitis: implications of antiviral and immunosuppressive therapies. *Arthritis Rheum* 46: 585-597.
204. Landau DA, Rosenzweig M, Saadoun D, Klatzmann D and Cacoub P (2009) The B lymphocyte stimulator receptor-ligand system in hepatitis C virus-induced B cell clonal disorders. *Ann Rheum Dis* 68: 337-344.
205. Starkebaum G and Sasso EH (2004) Hepatitis C and B cells: induction of autoimmunity and lymphoproliferation may reflect chronic stimulation through cell-surface receptors. *J Rheumatol* 31: 416-418.
206. Machida K, Cheng KT, Pavio N, Sung VM and Lai MM (2005) Hepatitis C virus E2-CD81 interaction induces hypermutation of the immunoglobulin gene in B cells. *J Virol* 79: 8079-8089.
207. (2011) Stratégies de dépistage biologique des hépatites virales B et C. Synthèse – avis des groupes de travail et de lecture et recommandation de la HAS. In: HAS, editor editors.
208. Lunel-Fabiani F PC (2003) Outils virologiques dans le diagnostic et le suivi des hépatites C : utilisation et apport de nouveaux tests. *Gastroenterol Clin Biol* 27: 718-726.

209. Mohsen AH, Easterbrook P, Taylor CB and Norris S (2002) Hepatitis C and HIV-1 coinfection. *Gut* 51: 601-608.
210. Le Guillou-Guillemette H and Lunel-Fabiani F (2009) Detection and quantification of serum or plasma HCV RNA: mini review of commercially available assays. *Methods Mol Biol* 510: 3-14.
211. Chevaliez S (2011) Virological tools to diagnose and monitor hepatitis C virus infection. *Clin Microbiol Infect* 17: 116-121.
212. Torriani FJ, Rodriguez-Torres M, Rockstroh JK, Lissen E, Gonzalez-Garcia J, et al. (2004) Peginterferon Alfa-2a plus ribavirin for chronic hepatitis C virus infection in HIV-infected patients. *N Engl J Med* 351: 438-450.
213. Carrat F, Bani-Sadr F, Pol S, Rosenthal E, Lunel-Fabiani F, et al. (2004) Pegylated interferon alfa-2b vs standard interferon alfa-2b, plus ribavirin, for chronic hepatitis C in HIV-infected patients: a randomized controlled trial. *Jama* 292: 2839-2848.
214. Chung RT, Andersen J, Volberding P, Robbins GK, Liu T, et al. (2004) Peginterferon Alfa-2a plus ribavirin versus interferon alfa-2a plus ribavirin for chronic hepatitis C in HIV-coinfected persons. *N Engl J Med* 351: 451-459.
215. Nunez M, Miralles C, Berdun MA, Losada E, Aguirrebengoa K, et al. (2007) Role of weight-based ribavirin dosing and extended duration of therapy in chronic hepatitis C in HIV-infected patients: the PRESCO trial. *AIDS Res Hum Retroviruses* 23: 972-982.
216. Payan C, Pivert A, Morand P, Fafi-Kremer S, Carrat F, et al. (2007) Rapid and early virological response to chronic hepatitis C treatment with IFN alpha2b or PEG-IFN alpha2b plus ribavirin in HIV/HCV co-infected patients. *Gut* 56: 1111-1116.
217. Laperche S, Lunel F, Izopet J, Alain S, Deny P, et al. (2005) Comparison of hepatitis C virus NS5b and 5' noncoding gene sequencing methods in a multicenter study. *J Clin Microbiol* 43: 733-739.
218. Laperche S, Saune K, Deny P, Duverlie G, Alain S, et al. (2006) Unique NS5b hepatitis C virus gene sequence consensus database is essential for standardization of genotype determinations in multicenter epidemiological studies. *J Clin Microbiol* 44: 614-616.
219. Pawlotsky JM, Aghemo A, Dusheiko G, Forns X, Puoti M, et al. (2014) EASL Recommendations on treatment of hepatitis C, 2014. EASL Annual Meeting. London. pp. 1-11.
220. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, et al. (2001) Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 358: 958-965.
221. Hadziyannis SJ, Sette H, Jr., Morgan TR, Balan V, Diago M, et al. (2004) Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med* 140: 346-355.
222. Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, et al. (2002) Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 347: 975-982.
223. Poynard T, Yuen MF, Ratziu V and Lai CL (2003) Viral hepatitis C. *Lancet* 362: 2095-2100.
224. (2014) Recommendations for Testing, Managing, and Treating Hepatitis C. In: A. a. IDSA, editor editors.
225. Hezode C, Fontaine H, Dorival C, Larrey D, Zoulim F, et al. (2013) Triple therapy in treatment-experienced patients with HCV-cirrhosis in a multicentre cohort of the French Early Access Programme (ANRS CO20-CUPIC) - NCT01514890. *J Hepatol* 59: 434-441.
226. Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, et al. (2010) Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. *Nature* 465: 96-100.

227. Guedj J, Dahari H, Rong L, Sansone ND, Nettles RE, et al. (2013) Modeling shows that the NS5A inhibitor daclatasvir has two modes of action and yields a shorter estimate of the hepatitis C virus half-life. *Proc Natl Acad Sci U S A* 110: 3991-3996.
228. Lee C, Ma H, Hang JQ, Leveque V, Sklan EH, et al. (2011) The hepatitis C virus NS5A inhibitor (BMS-790052) alters the subcellular localization of the NS5A non-structural viral protein. *Virology* 414: 10-18.
229. Nakamoto S, Kanda T, Wu S, Shirasawa H and Yokosuka O (2014) Hepatitis C virus NS5A inhibitors and drug resistance mutations. *World J Gastroenterol* 20: 2902-2912.
230. Delang L, Neyts J, Vlieghe I, Abrignani S, Neddermann P, et al. (2013) Hepatitis C virus-specific directly acting antiviral drugs. *Curr Top Microbiol Immunol* 369: 289-320.
231. Einav S, Sklan EH, Moon HM, Gehrig E, Liu P, et al. (2008) The nucleotide binding motif of hepatitis C virus NS4B can mediate cellular transformation and tumor formation without Ha-ras co-transfection. *Hepatology* 47: 827-835.
232. Cho NJ, Dvory-Sobol H, Lee C, Cho SJ, Bryson P, et al. (2010) Identification of a class of HCV inhibitors directed against the nonstructural protein NS4B. *Sci Transl Med* 2: 15ra16.
233. Esser-Nobis K, Romero-Brey I, Ganten TM, Gouttenoire J, Harak C, et al. (2013) Analysis of hepatitis C virus resistance to silibinin in vitro and in vivo points to a novel mechanism involving nonstructural protein 4B. *Hepatology* 57: 953-963.
234. Chatel-Chaix L, Germain MA, Gotte M and Lamarre D (2012) Direct-acting and host-targeting HCV inhibitors: current and future directions. *Curr Opin Virol* 2: 588-598.
235. Peel M and Scribner A (2013) Cyclophilin inhibitors as antiviral agents. *Bioorg Med Chem Lett* 23: 4485-4492.
236. Liu Z, Yang F, Robotham JM and Tang H (2009) Critical role of cyclophilin A and its prolyl-peptidyl isomerase activity in the structure and function of the hepatitis C virus replication complex. *J Virol* 83: 6554-6565.
237. Baugh JM, Garcia-Rivera JA and Gallay PA (2013) Host-targeting agents in the treatment of hepatitis C: a beginning and an end? *Antiviral Res* 100: 555-561.
238. Li Y, Masaki T, Yamane D, McGivern DR and Lemon SM (2013) Competing and noncompeting activities of miR-122 and the 5' exonuclease Xrn1 in regulation of hepatitis C virus replication. *Proc Natl Acad Sci U S A* 110: 1881-1886.
239. Li YP, Gottwein JM, Scheel TK, Jensen TB and Bukh J (2011) MicroRNA-122 antagonism against hepatitis C virus genotypes 1-6 and reduced efficacy by host RNA insertion or mutations in the HCV 5' UTR. *Proc Natl Acad Sci U S A* 108: 4991-4996.
240. Janssen HL, Reesink HW, Lawitz EJ, Zeuzem S, Rodriguez-Torres M, et al. (2013) Treatment of HCV infection by targeting microRNA. *N Engl J Med* 368: 1685-1694.
241. Vaillancourt FH, Brault M, Pilote L, Uyttensprot N, Gaillard ET, et al. (2012) Evaluation of phosphatidylinositol-4-kinase IIIalpha as a hepatitis C virus drug target. *J Virol* 86: 11595-11607.
242. Lacey K, Vercauteren K, Grzyb K, Naddeo M, Verhoye L, et al. (2012) Novel human SR-BI antibodies prevent infection and dissemination of HCV in vitro and in humanized mice. *J Hepatol* 57: 17-23.
243. Zhu H, Wong-Staal F, Lee H, Syder A, McKelvy J, et al. (2012) Evaluation of ITX 5061, a scavenger receptor B1 antagonist: resistance selection and activity in combination with other hepatitis C virus antivirals. *J Infect Dis* 205: 656-662.
244. Huang JT, Tseng CP, Liao MH, Lu SC, Yeh WZ, et al. (2013) Hepatitis C virus replication is modulated by the interaction of nonstructural protein NS5B and fatty acid synthase. *J Virol* 87: 4994-5004.
245. Pawlotsky JM and Germanidis G (1999) The non-structural 5A protein of hepatitis C virus. *J Viral Hepat* 6: 343-356.

246. Gao L, Aizaki H, He JW and Lai MM (2004) Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. *J Virol* 78: 3480-3488.
247. Shirota Y, Luo H, Qin W, Kaneko S, Yamashita T, et al. (2002) Hepatitis C virus (HCV) NS5A binds RNA-dependent RNA polymerase (RdRP) NS5B and modulates RNA-dependent RNA polymerase activity. *J Biol Chem* 277: 11149-11155.
248. Gale MJ, Jr., Korth MJ, Tang NM, Tan SL, Hopkins DA, et al. (1997) Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology* 230: 217-227.
249. Tellinghuisen TL, Marcotrigiano J, Gorbalenya AE and Rice CM (2004) The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J Biol Chem* 279: 48576-48587.
250. He Y, Staschke KA and Tan SL (2006) HCV NS5A: A Multifunctional Regulator of Cellular Pathways and Virus Replication. *Hepatitis C Viruses: Genomes and Molecular Biology* Chapter 9. Norfolk (UK).
251. Brass V, Bieck E, Montserret R, Wolk B, Hellings JA, et al. (2002) An amino-terminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *J Biol Chem* 277: 8130-8139.
252. Penin F, Brass V, Appel N, Ramboarina S, Montserret R, et al. (2004) Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J Biol Chem* 279: 40835-40843.
253. Tellinghuisen TL, Marcotrigiano J and Rice CM (2005) Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature* 435: 374-379.
254. Love RA, Brodsky O, Hickey MJ, Wells PA and Cronin CN (2009) Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *J Virol* 83: 4395-4403.
255. Hwang J, Huang L, Cordek DG, Vaughan R, Reynolds SL, et al. (2010) Hepatitis C virus nonstructural protein 5A: biochemical characterization of a novel structural class of RNA-binding proteins. *J Virol* 84: 12480-12491.
256. Foster TL, Belyaeva T, Stonehouse NJ, Pearson AR and Harris M (2010) All three domains of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. *J Virol* 84: 9267-9277.
257. Huang L, Hwang J, Sharma SD, Hargittai MR, Chen Y, et al. (2005) Hepatitis C virus nonstructural protein 5A (NS5A) is an RNA-binding protein. *J Biol Chem* 280: 36417-36428.
258. Berger KL, Kelly SM, Jordan TX, Tartell MA and Randall G (2011) Hepatitis C Virus Stimulates the Phosphatidylinositol 4-Kinase III Alpha-Dependent Phosphatidylinositol 4-Phosphate Production That Is Essential for Its Replication. *J Virol* 85: 8870-8883.
259. Reiss S, Rebhan I, Backes P, Romero-Brey I, Erfle H, et al. (2011) Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment. *Cell Host Microbe* 9: 32-45.
260. Lim YS and Hwang SB (2011) Hepatitis C virus NS5A protein interacts with phosphatidylinositol 4-kinase type IIIalpha and regulates viral propagation. *J Biol Chem* 286: 11290-11298.
261. Hsu NY, Ilnytska O, Belov G, Santiana M, Chen YH, et al. (2010) Viral reorganization of the secretory pathway generates distinct organelles for RNA replication. *Cell* 141: 799-811.
262. Tellinghuisen TL, Foss KL, Treadaway JC and Rice CM (2008) Identification of residues required for RNA replication in domains II and III of the hepatitis C virus NS5A protein. *J Virol* 82: 1073-1083.

263. Macdonald A, Crowder K, Street A, McCormick C and Harris M (2004) The hepatitis C virus NS5A protein binds to members of the Src family of tyrosine kinases and regulates kinase activity. *J Gen Virol* 85: 721-729.
264. Hughes M, Gretton S, Shelton H, Brown DD, McCormick CJ, et al. (2009) A conserved proline between domains II and III of hepatitis C virus NS5A influences both RNA replication and virus assembly. *J Virol* 83: 10788-10796.
265. Ide Y, Zhang L, Chen M, Inchauspe G, Bahl C, et al. (1996) Characterization of the nuclear localization signal and subcellular distribution of hepatitis C virus nonstructural protein NS5A. *Gene* 182: 203-211.
266. Maqbool MA, Imache MR, Higgs MR, Carmouse S, Pawlotsky JM, et al. (2013) Regulation of hepatitis C virus replication by nuclear translocation of nonstructural 5A protein and transcriptional activation of host genes. *J Virol* 87: 5523-5539.
267. Hughes M, Griffin S and Harris M (2009) Domain III of NS5A contributes to both RNA replication and assembly of hepatitis C virus particles. *J Gen Virol* 90: 1329-1334.
268. Miyanari Y, Atsuzawa K, Usuda N, Watashi K, Hishiki T, et al. (2007) The lipid droplet is an important organelle for hepatitis C virus production. *Nat Cell Biol* 9: 1089-1097.
269. Kim S, Welsch C, Yi M and Lemon SM (2011) Regulation of the production of infectious genotype 1a hepatitis C virus by NS5A domain III. *J Virol* 85: 6645-6656.
270. Inchauspe G, Zebedee S, Lee DH, Sugitani M, Nasoff M, et al. (1991) Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates. *Proc Natl Acad Sci U S A* 88: 10292-10296.
271. Duverlie G, Khorsi H, Castelain S, Jaillon O, Izopet J, et al. (1998) Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity. *J Gen Virol* 79 ( Pt 6): 1373-1381.
272. Nousbaum J, Polyak SJ, Ray SC, Sullivan DG, Larson AM, et al. (2000) Prospective characterization of full-length hepatitis C virus NS5A quasispecies during induction and combination antiviral therapy. *J Virol* 74: 9028-9038.
273. Puig-Basagoiti F, Forn X, Furcic I, Ampurdanes S, Gimenez-Barcons M, et al. (2005) Dynamics of hepatitis C virus NS5A quasispecies during interferon and ribavirin therapy in responder and non-responder patients with genotype 1b chronic hepatitis C. *J Gen Virol* 86: 1067-1075.
274. El-Shamy A, Sasayama M, Nagano-Fujii M, Sasase N, Imoto S, et al. (2007) Prediction of efficient virological response to pegylated interferon/ribavirin combination therapy by NS5A sequences of hepatitis C virus and anti-NS5A antibodies in pre-treatment sera. *Microbiol Immunol* 51: 471-482.
275. Vuillermoz I, Khattab E, Sablon E, Ottevaere I, Durantel D, et al. (2004) Genetic variability of hepatitis C virus in chronically infected patients with viral breakthrough during interferon-ribavirin therapy. *J Med Virol* 74: 41-53.
276. Munoz de Rueda P, Casado J, Paton R, Quintero D, Palacios A, et al. (2008) Mutations in E2-PePHD, NS5A-PKRBD, NS5A-ISDR, and NS5A-V3 of hepatitis C virus genotype 1 and their relationships to pegylated interferon-ribavirin treatment responses. *J Virol* 82: 6644-6653.
277. Jardim AC, Yamasaki LH, de Queiroz AT, Bittar C, Pinho JR, et al. (2009) Quasispecies of hepatitis C virus genotype 1 and treatment outcome with peginterferon and ribavirin. *Infect Genet Evol* 9: 689-698.
278. Cuevas JM, Gonzalez M, Torres-Puente M, Jimenez-Hernandez N, Bracho MA, et al. (2009) The role of positive selection in hepatitis C virus. *Infect Genet Evol* 9: 860-866.
279. Liang Y, Ye H, Kang CB and Yoon HS (2007) Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. *Biochemistry* 46: 11550-11558.



280. Hanouille X, Badillo A, Verdegem D, Penin F and Lippens G (2010) The domain 2 of the HCV NS5A protein is intrinsically unstructured. *Protein Pept Lett* 17: 1012-1018.
281. Wang S, Wu X, Pan T, Song W, Wang Y, et al. (2011) Viperin inhibits hepatitis C virus replication by interfering with binding of NS5A to host protein hVAP-33. *J Gen Virol* 93: 83-92.
282. Rehmann B (2009) Hepatitis C virus versus innate and adaptive immune responses: a tale of coevolution and coexistence. *J Clin Invest* 119: 1745-1754.
283. Lemon SM (2010) Induction and evasion of innate antiviral responses by hepatitis C virus. *J Biol Chem* 285: 22741-22747.
284. Datta S, Hazari S, Chandra PK, Samara M, Poat B, et al. (2011) Mechanism of HCV's resistance to IFN-alpha in cell culture involves expression of functional IFN-alpha receptor 1. *Virol J* 8: 351.
285. Tsuge M, Fujimoto Y, Hiraga N, Zhang Y, Ohnishi M, et al. (2011) Hepatitis C virus infection suppresses the interferon response in the liver of the human hepatocyte chimeric mouse. *PLoS One* 6: e23856.
286. Kanda T, Steele R, Ray R and Ray RB (2009) Inhibition of intrahepatic gamma interferon production by hepatitis C virus nonstructural protein 5A in transgenic mice. *J Virol* 83: 8463-8469.
287. Arnaud N, Dabo S, Akazawa D, Fukasawa M, Shinkai-Ouchi F, et al. (2011) Hepatitis C virus reveals a novel early control in acute immune response. *PLoS Pathog* 7: e1002289.
288. Arnaud N, Dabo S, Maillard P, Budkowska A, Kalliampakou KI, et al. Hepatitis C virus controls interferon production through PKR activation. *PLoS One* 5: e10575.
289. Dolganiuc A, Chang S, Kodys K, Mandrekar P, Bakis G, et al. (2006) Hepatitis C virus (HCV) core protein-induced, monocyte-mediated mechanisms of reduced IFN-alpha and plasmacytoid dendritic cell loss in chronic HCV infection. *J Immunol* 177: 6758-6768.
290. Szabo G and Dolganiuc A (2006) HCV immunopathogenesis: virus-induced strategies against host immunity. *Clin Liver Dis* 10: 753-771.
291. Toroney R, Nallagatla SR, Boyer JA, Cameron CE and Bevilacqua PC (2010) Regulation of PKR by HCV IRES RNA: importance of domain II and NS5A. *J Mol Biol* 400: 393-412.
292. Pflugheber J, Fredericksen B, Sumpter R, Jr., Wang C, Ware F, et al. (2002) Regulation of PKR and IRF-1 during hepatitis C virus RNA replication. *Proc Natl Acad Sci U S A* 99: 4650-4655.
293. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, et al. (1995) Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest* 96: 224-230.
294. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, et al. (1996) Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* 334: 77-81.
295. Garaigorta U and Chisari FV (2009) Hepatitis C virus blocks interferon effector function by inducing protein kinase R phosphorylation. *Cell Host Microbe* 6: 513-522.
296. Girard S, Shalhoub P, Lescure P, Sabile A, Misek DE, et al. (2002) An altered cellular response to interferon and up-regulation of interleukin-8 induced by the hepatitis C viral protein NS5A uncovered by microarray analysis. *Virology* 295: 272-283.
297. Polyak SJ, Khabar KS, Paschal DM, Ezelle HJ, Duverlie G, et al. (2001) Hepatitis C virus nonstructural 5A protein induces interleukin-8, leading to partial inhibition of the interferon-induced antiviral response. *J Virol* 75: 6095-6106.

298. Polyak SJ, Khabar KS, Rezeiq M and Gretch DR (2001) Elevated levels of interleukin-8 in serum are associated with hepatitis C virus infection and resistance to interferon therapy. *J Virol* 75: 6209-6211.
299. Lan KH, Lan KL, Lee WP, Sheu ML, Chen MY, et al. (2007) HCV NS5A inhibits interferon-alpha signaling through suppression of STAT1 phosphorylation in hepatocyte-derived cell lines. *J Hepatol* 46: 759-767.
300. He Y, Nakao H, Tan SL, Polyak SJ, Neddermann P, et al. (2002) Subversion of cell signaling pathways by hepatitis C virus nonstructural 5A protein via interaction with Grb2 and P85 phosphatidylinositol 3-kinase. *J Virol* 76: 9207-9217.
301. Chang L and Karin M (2001) Mammalian MAP kinase signalling cascades. *Nature* 410: 37-40.
302. Macdonald A, Crowder K, Street A, McCormick C, Saksela K, et al. (2003) The hepatitis C virus non-structural NS5A protein inhibits activating protein-1 function by perturbing ras-ERK pathway signaling. *J Biol Chem* 278: 17775-17784.
303. Macdonald A, Chan JK and Harris M (2005) Perturbation of epidermal growth factor receptor complex formation and Ras signalling in cells harbouring the hepatitis C virus subgenomic replicon. *J Gen Virol* 86: 1027-1033.
304. Macdonald A and Harris M (2004) Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol* 85: 2485-2502.
305. Wu SC, Chang SC, Wu HY, Liao PJ and Chang MF (2008) Hepatitis C virus NS5A protein down-regulates the expression of spindle gene *Aspm* through PKR-p38 signaling pathway. *J Biol Chem* 283: 29396-29404.
306. Baek KH, Park HY, Kang CM, Kim SJ, Jeong SJ, et al. (2006) Overexpression of hepatitis C virus NS5A protein induces chromosome instability via mitotic cell cycle dysregulation. *J Mol Biol* 359: 22-34.
307. Freeman A, Hamid S, Morris L, Vowler S, Rushbrook S, et al. (2003) Improved detection of hepatocyte proliferation using antibody to the pre-replication complex: an association with hepatic fibrosis and viral replication in chronic hepatitis C virus infection. *J Viral Hepat* 10: 345-350.
308. Street A, Macdonald A, Crowder K and Harris M (2004) The Hepatitis C virus NS5A protein activates a phosphoinositide 3-kinase-dependent survival signaling cascade. *J Biol Chem* 279: 12232-12241.
309. Street A, Macdonald A, McCormick C and Harris M (2005) Hepatitis C virus NS5A-mediated activation of phosphoinositide 3-kinase results in stabilization of cellular beta-catenin and stimulation of beta-catenin-responsive transcription. *J Virol* 79: 5006-5016.
310. Park CY, Choi SH, Kang SM, Kang JI, Ahn BY, et al. (2009) Nonstructural 5A protein activates beta-catenin signaling cascades: implication of hepatitis C virus-induced liver pathogenesis. *J Hepatol* 51: 853-864.
311. Milward A, Mankouri J and Harris M (2010) Hepatitis C virus NS5A protein interacts with beta-catenin and stimulates its transcriptional activity in a phosphoinositide-3 kinase-dependent fashion. *J Gen Virol* 91: 373-381.
312. Waris G, Turkson J, Hassanein T and Siddiqui A (2005) Hepatitis C virus (HCV) constitutively activates STAT-3 via oxidative stress: role of STAT-3 in HCV replication. *J Virol* 79: 1569-1580.
313. Sarcar B, Ghosh AK, Steele R, Ray R and Ray RB (2004) Hepatitis C virus NS5A mediated STAT3 activation requires co-operation of Jak1 kinase. *Virology* 322: 51-60.
314. Mundt B, Wirth T, Zender L, Waltemathe M, Trautwein C, et al. (2005) Tumour necrosis factor related apoptosis inducing ligand (TRAIL) induces hepatic steatosis in viral hepatitis and after alcohol intake. *Gut* 54: 1590-1596.

315. Majumder M, Ghosh AK, Steele R, Zhou XY, Phillips NJ, et al. (2002) Hepatitis C virus NS5A protein impairs TNF-mediated hepatic apoptosis, but not by an anti-FAS antibody, in transgenic mice. *Virology* 294: 94-105.
316. Chung YL, Sheu ML and Yen SH (2003) Hepatitis C virus NS5A as a potential viral Bcl-2 homologue interacts with Bax and inhibits apoptosis in hepatocellular carcinoma. *Int J Cancer* 107: 65-73.
317. Lan KH, Sheu ML, Hwang SJ, Yen SH, Chen SY, et al. (2002) HCV NS5A interacts with p53 and inhibits p53-mediated apoptosis. *Oncogene* 21: 4801-4811.
318. Tamura R, Kanda T, Imazeki F, Wu S, Nakamoto S, et al. (2011) Hepatitis C Virus nonstructural 5A protein inhibits lipopolysaccharide-mediated apoptosis of hepatocytes by decreasing expression of Toll-like receptor 4. *J Infect Dis* 204: 793-801.
319. Akkari L, Gregoire D, Floc'h N, Moreau M, Hernandez C, et al. (2012) Hepatitis C viral protein NS5A induces EMT and participates in oncogenic transformation of primary hepatocyte precursors. *J Hepatol* 57: 1021-1028.
320. Deng L, Shoji I, Ogawa W, Kaneda S, Soga T, et al. (2011) Hepatitis C virus infection promotes hepatic gluconeogenesis through an NS5A-mediated, FoxO1-dependent pathway. *J Virol* 85: 8556-8568.
321. Dionisio N, Garcia-Mediavilla MV, Sanchez-Campos S, Majano PL, Benedicto I, et al. (2009) Hepatitis C virus NS5A and core proteins induce oxidative stress-mediated calcium signalling alterations in hepatocytes. *J Hepatol* 50: 872-882.
322. Hung CH, Wang JH, Hu TH, Chen CH, Chang KC, et al. (2010) Insulin resistance is associated with hepatocellular carcinoma in chronic hepatitis C infection. *World J Gastroenterol* 16: 2265-2271.
323. Gong G, Waris G, Tanveer R and Siddiqui A (2001) Human hepatitis C virus NS5A protein alters intracellular calcium levels, induces oxidative stress, and activates STAT-3 and NF-kappa B. *Proc Natl Acad Sci U S A* 98: 9599-9604.
324. Waris G, Livolsi A, Imbert V, Peyron JF and Siddiqui A (2003) Hepatitis C virus NS5A and subgenomic replicon activate NF-kappaB via tyrosine phosphorylation of Ikbpp and its degradation by calpain protease. *J Biol Chem* 278: 40778-40787.
325. Salemi M VA (2006) *The phylogenetic Handbook, a practical approach to DNA and protein phylogeny*.
326. Simon-Loriere E and Holmes EC (2011) Why do RNA viruses recombine? *Nat Rev Microbiol* 9: 617-626.
327. Worobey M and Holmes EC (1999) Evolutionary aspects of recombination in RNA viruses. *J Gen Virol* 80 ( Pt 10): 2535-2543.
328. Neher RA and Leitner T (2010) Recombination rate and selection strength in HIV intra-patient evolution. *PLoS Comput Biol* 6: e1000660.
329. Savolainen-Kopra C and Blomqvist S (2010) Mechanisms of genetic variation in polioviruses. *Rev Med Virol* 20: 358-371.
330. Woo PC, Lau SK, Huang Y and Yuen KY (2009) Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med (Maywood)* 234: 1117-1127.
331. Reiter J, Perez-Vilaro G, Scheller N, Mina LB, Diez J, et al. (2011) Hepatitis C virus RNA recombination in cell culture. *J Hepatol* 55: 777-783.
332. Gmyl AP, Belousov EV, Maslova SV, Khitrina EV, Chetverin AB, et al. (1999) Nonreplicative RNA recombination in poliovirus. *J Virol* 73: 8958-8965.
333. Gallei A, Pankraz A, Thiel HJ and Becher P (2004) RNA recombination in vivo in the absence of viral replication. *J Virol* 78: 6271-6281.
334. Chetverin AB, Chetverina HV, Demidenko AA and Ugarov VI (1997) Nonhomologous RNA recombination in a cell-free system: evidence for a transesterification mechanism guided by secondary structure. *Cell* 88: 503-513.

335. Simon-Loriere E and Holmes EC (2013) Gene Duplication Is Infrequent in the Recent Evolutionary History of RNA Viruses. *Mol Biol Evol*.
336. Gritsun TS and Gould EA (2006) The 3' untranslated region of tick-borne flaviviruses originated by the duplication of long repeat sequences within the open reading frame. *Virology* 354: 217-223.
337. Cao D, Barro M and Hoshino Y (2008) Porcine rotavirus bearing an aberrant gene stemming from an intergenic recombination of the NSP2 and NSP5 genes is defective and interfering. *J Virol* 82: 6073-6077.
338. Tseng CH, Knowles NJ and Tsai HJ (2007) Molecular analysis of duck hepatitis virus type 1 indicates that it should be assigned to a new genus. *Virus Res* 123: 190-203.
339. Jenkins GM, Rambaut A, Pybus OG and Holmes EC (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* 54: 156-165.
340. Drake JW and Holland JJ (1999) Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A* 96: 13910-13913.
341. Drummond AJ, Pybus OG, Rambaut A, Forsberg R and Rodrigo AG (2003) Measurably evolving populations. *TRENDS in Ecology and Evolution* 18: 481-488.
342. Pybus OG and Rambaut A (2009) Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* 10: 540-550.
343. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, et al. (2008) The global circulation of seasonal influenza A (H3N2) viruses. *Science* 320: 340-346.
344. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305: 371-376.
345. Fitch WM, Leiter JM, Li XQ and Palese P (1991) Positive Darwinian evolution in human influenza A viruses. *Proc Natl Acad Sci U S A* 88: 4270-4274.
346. Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, et al. (2001) The epidemic behavior of the hepatitis C virus. *Science* 292: 2323-2325.
347. Rambaut A, Robertson DL, Pybus OG, Peeters M and Holmes EC (2001) Human immunodeficiency virus. Phylogeny and the origin of HIV-1. *Nature* 410: 1047-1048.
348. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, et al. (1998) An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 391: 594-597.
349. Li W, Shi Z, Yu M, Ren W, Smith C, et al. (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310: 676-679.
350. Hon CC, Lam TY, Shi ZL, Drummond AJ, Yip CW, et al. (2008) Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J Virol* 82: 1819-1826.
351. Kenny-Walsh E (1999) Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin. Irish Hepatology Research Group. *N Engl J Med* 340: 1228-1233.
352. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, et al. (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 73: 10489-10502.
353. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, et al. (2007) Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Comput Biol* 3: e29.
354. Frost SD, Wrin T, Smith DM, Kosakovsky Pond SL, Liu Y, et al. (2005) Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proc Natl Acad Sci U S A* 102: 18514-18519.
355. Sobesky R, Feray C, Rimlinger F, Derian N, Dos Santos A, et al. (2007) Distinct hepatitis C virus core and F protein quasispecies in tumoral and nontumoral hepatocytes isolated via microdissection. *Hepatology* 46: 1704-1712.

356. Rodrigo AG, Shpaer EG, Delwart EL, Iversen AK, Gallo MV, et al. (1999) Coalescent estimates of HIV-1 generation time in vivo. *Proc Natl Acad Sci U S A* 96: 2187-2191.
357. Veillon P, Payan C, Gaudy C, Goudeau A and Lunel F (2004) [Mutation analysis of ISDR and V3 domains of hepatitis C virus NS5A region before interferon therapy with or without ribavirin]. *Pathol Biol (Paris)* 52: 505-510.
358. Agnello V and De Rosa FG (2004) Extrahepatic disease manifestations of HCV infection: some current issues. *J Hepatol* 40: 341-352.
359. Flores A, Briand JF, Gadat O, Andrau JC, Rubbi L, et al. (1999) A protein-protein interaction map of yeast RNA polymerase III. *Proc Natl Acad Sci U S A* 96: 7815-7820.
360. Flajolet M, Rotondo G, Daviet L, Bergametti F, Inchauspe G, et al. (2000) A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* 242: 369-379.
361. Ito T, Ota K, Kubota H, Yamaguchi Y, Chiba T, et al. (2002) Roles for the two-hybrid system in exploration of the yeast protein interactome. *Mol Cell Proteomics* 1: 561-566.
362. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403: 623-627.
363. Ng SK, Zhang Z, Tan SH and Lin K (2003) InterDom: a database of putative interacting protein domains for validating predicted protein interactions and complexes. *Nucleic Acids Res* 31: 251-254.
364. Ng KL, Huang CH and Liu HC (2008) Applications of domain-domain interactions in pathway study. *Comput Biol Chem* 32: 81-87.
365. Phizicky EM and Fields S (1995) Protein-protein interactions: methods for detection and analysis. *Microbiol Rev* 59: 94-123.
366. Suter B, Kittanakom S and Stagljar I (2008) Interactive proteomics: what lies ahead? *Biotechniques* 44: 681-691.
367. Scott JK and Smith GP (1990) Searching for peptide ligands with an epitope library. *Science* 249: 386-390.
368. Fields S and Song O (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340: 245-246.
369. Nickels BE (2009) Genetic assays to define and characterize protein-protein interactions involved in gene regulation. *Methods* 47: 53-62.
370. Bakker O, van Beeren HC, Emrich T, Holtke H and Wiersinga WM (1999) Interaction between nuclear hormone receptors and coactivators analyzed using a nonradioactive "Pull-Down" assay. *Anal Biochem* 276: 105-106.
371. de Chasse B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, et al. (2008) Hepatitis C virus infection protein network. *Mol Syst Biol* 4: 230.
372. Joung JK, Ramm EI and Pabo CO (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A* 97: 7382-7387.
373. Tafforeau L R-CC, and Lotteau V (2012) Virus–Human Cell Interactomes, Chapter 6; Two Hybrid Technologies: Methods and Protocols. Springer Science+Business Media.
374. Gervais M (2010) Etude des interactions protéine-protéine par double-hybride bactérien: applications en agro-alimentaire et en santé. UMR A191. Angers: Université d'Angers. pp. 300.
375. (2003) The QIAexpressionist™, A handbook for high-level expression and purification of 6xHis-tagged proteins. 128 p.
376. Livezey KW, Negorev D and Simon D (2000) Investigation of the expression of Bin1, a putative suppressor, in human hepatoma cells. *Cancer Genet Cytogenet* 116: 35-39.
377. Masumi A, Aizaki H, Suzuki T, DuHadaway JB, Prendergast GC, et al. (2005) Reduction of hepatitis C virus NS5A phosphorylation through its interaction with amphiphysin II. *Biochem Biophys Res Commun* 336: 572-578.

378. Nanda SK, Herion D and Liang TJ (2006) The SH3 binding motif of HCV [corrected] NS5A protein interacts with Bin1 and is important for apoptosis and infectivity. *Gastroenterology* 130: 794-809.
379. Shi ST, Polyak SJ, Tu H, Taylor DR, Gretch DR, et al. (2002) Hepatitis C virus NS5A colocalizes with the core protein on lipid droplets and interacts with apolipoproteins. *Virology* 292: 198-210.
380. Tong WY, Nagano-Fujii M, Hidajat R, Deng L, Takigawa Y, et al. (2002) Physical interaction between hepatitis C virus NS4B protein and CREB-RP/ATF6beta. *Biochem Biophys Res Commun* 299: 366-372.
381. Zheng Y, Gao B, Ye L, Kong L, Jing W, et al. (2005) Hepatitis C virus non-structural protein NS4B can modulate an unfolded protein response. *J Microbiol* 43: 529-536.
382. Ciccaglione AR, Marcantonio C, Tritarelli E, Equestre M, Vendittelli F, et al. (2007) Activation of the ER stress gene gadd153 by hepatitis C virus sensitizes cells to oxidant injury. *Virus Res* 126: 128-138.
383. Li S, Ye L, Yu X, Xu B, Li K, et al. (2009) Hepatitis C virus NS4B induces unfolded protein response and endoplasmic reticulum overload response-dependent NF-kappaB activation. *Virology* 391: 257-264.
384. Hosui A, Takehara T, Ohkawa K, Kanazawa Y, Tatsumi T, et al. (2006) Suppressive effect on hepatocyte differentiation of hepatitis C virus core protein. *Biochem Biophys Res Commun* 346: 1125-1130.
385. Ikeda M, Mori K, Ariumi Y, Dansako H and Kato N (2009) Oncostatin M synergistically inhibits HCV RNA replication in combination with interferon-alpha. *FEBS Lett* 583: 1434-1438.
386. Larrea E, Aldabe R, Gonzalez I, Segura V, Sarobe P, et al. (2009) Oncostatin M enhances the antiviral effects of type I interferon and activates immunostimulatory functions in liver epithelial cells. *J Virol* 83: 3298-3311.
387. Cao Y, Hamada T, Matsui T, Date T and Iwabuchi K (2004) Hepatitis C virus core protein interacts with p53-binding protein, 53BP2/Bbp/ASPP2, and inhibits p53-mediated apoptosis. *Biochem Biophys Res Commun* 315: 788-795.
388. Striebinger H, Koegl M and Bailer SM (2013) A high-throughput yeast two-hybrid protocol to determine virus-host protein interactions. *Methods Mol Biol* 1064: 1-15.
389. Zech B, Kurtenbach A, Krieger N, Strand D, Blencke S, et al. (2003) Identification and characterization of amphiphysin II as a novel cellular interaction partner of the hepatitis C virus NS5A protein. *J Gen Virol* 84: 555-560.
390. Schwarten M, Solyom Z, Feuerstein S, Aladag A, Hoffmann S, et al. (2013) Interaction of nonstructural protein 5A of the hepatitis C virus with Src homology 3 domains using noncanonical binding sites. *Biochemistry* 52: 6160-6168.
391. Feuerstein S, Solyom Z, Aladag A, Favier A, Schwarten M, et al. (2012) Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *J Mol Biol* 420: 310-323.
392. Wang AG, Lee DS, Moon HB, Kim JM, Cho KH, et al. (2009) Non-structural 5A protein of hepatitis C virus induces a range of liver pathology in transgenic mice. *J Pathol* 219: 253-262.
393. Wu J, Rutkowski DT, Dubois M, Swathirajan J, Saunders T, et al. (2007) ATF6alpha optimizes long-term endoplasmic reticulum function to protect cells from chronic stress. *Dev Cell* 13: 351-364.
394. Asselah T, Bieche I, Mansouri A, Laurendeau I, Cazals-Hatem D, et al. (2010) In vivo hepatic endoplasmic reticulum stress in patients with chronic hepatitis C. *J Pathol* 221: 264-274.

395. Hamada T, Sato A, Hirano T, Yamamoto T, Son G, et al. (2007) Oncostatin M gene therapy attenuates liver damage induced by dimethylnitrosamine in rats. *Am J Pathol* 171: 872-881.
396. Sullivan A and Lu X (2007) ASPP: a new family of oncogenes and tumour suppressor genes. *Br J Cancer* 96: 196-200.
397. Okumura T (2011) Role of lipid droplet proteins in liver steatosis. *J Physiol Biochem* 67: 629-636.
398. Burstein E, Hoberg JE, Wilkinson AS, Rumble JM, Csomos RA, et al. (2005) COMMD proteins, a novel family of structural and functional homologs of MURR1. *J Biol Chem* 280: 22222-22232.
399. Tan SL, Nakao H, He Y, Vijaysri S, Neddermann P, et al. (1999) NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand-dependent manner and perturbs mitogenic signaling. *Proc Natl Acad Sci U S A* 96: 5533-5538.
400. Huang L, Sineva EV, Hargittai MR, Sharma SD, Suthar M, et al. (2004) Purification and characterization of hepatitis C virus non-structural protein 5A expressed in *Escherichia coli*. *Protein Expr Purif* 37: 144-153.
401. Hanouille X, Badillo A, Wieruszeski JM, Verdegem D, Landrieu I, et al. (2009) Hepatitis C virus NS5A protein is a substrate for the peptidyl-prolyl cis/trans isomerase activity of cyclophilins A and B. *J Biol Chem* 284: 13589-13601.
402. Wang Q, Wang Y, Li Y, Gao X, Liu S, et al. (2013) NS5A<sup>ATP9</sup> Contributes to Inhibition of Cell Proliferation by Hepatitis C Virus (HCV) Nonstructural Protein 5A (NS5A) via MEK/Extracellular Signal Regulated Kinase (ERK) Pathway. *Int J Mol Sci* 14: 10539-10551.
403. Cheng D, Zhao L, Zhang L, Jiang Y, Tian Y, et al. (2013) p53 controls hepatitis C virus non-structural protein 5A-mediated downregulation of GADD45 $\alpha$  expression via the NF- $\kappa$ B and PI3K-Akt pathways. *J Gen Virol* 94: 326-335.
404. Germain MA, Chatel-Chaix L, Gagne B, Bonneil E, Thibault P, et al. (2014) Elucidating novel hepatitis C virus-host interactions using combined mass spectrometry and functional genomics approaches. *Mol Cell Proteomics* 13: 184-203.
405. Macdonald A, Mazaleyrat S, McCormick C, Street A, Burgoyne NJ, et al. (2005) Further studies on hepatitis C virus NS5A-SH3 domain interactions: identification of residues critical for binding and implications for viral RNA replication and modulation of cell signalling. *J Gen Virol* 86: 1035-1044.
406. Shelton H and Harris M (2008) Hepatitis C virus NS5A protein binds the SH3 domain of the Fyn tyrosine kinase with high affinity: mutagenic analysis of residues within the SH3 domain that contribute to the interaction. *Virol J* 5: 24.
407. Feuerstein S, Solyom Z, Aladag A, Hoffmann S, Willbold D, et al. (2011) <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N resonance assignment of a 179 residue fragment of hepatitis C virus non-structural protein 5A. *Biomol NMR Assign* 5: 241-243.
408. Hanouille X, Verdegem D, Badillo A, Wieruszeski JM, Penin F, et al. (2009) Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. *Biochem Biophys Res Commun* 381: 634-638.
409. Kato T, Date T, Miyamoto M, Furusaka A, Tokushige K, et al. (2003) Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology* 125: 1808-1817.
410. Helbig KJ, Eyre NS, Yip E, Narayana S, Li K, et al. (2011) The antiviral protein viperin inhibits hepatitis C virus replication via interaction with nonstructural protein 5A. *Hepatology* 54: 1506-1517.

411. Peng CW, Peremyslov VV, Mushegian AR, Dawson WO and Dolja VV (2001) Functional specialization and evolution of leader proteinases in the family Closteroviridae. *J Virol* 75: 12153-12160.
412. Valli A, Lopez-Moya JJ and Garcia JA (2007) Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J Gen Virol* 88: 1016-1028.
413. Walker PJ, Dietzgen RG, Joubert DA and Blasdell KR (2011) Rhabdovirus accessory genes. *Virus Res* 162: 110-125.
414. Panavas T, Panaviene Z, Pogany J and Nagy PD (2003) Enhancement of RNA synthesis by promoter duplication in tombusviruses. *Virology* 310: 118-129.
415. Nagai M, Sakoda Y, Mori M, Hayashi M, Kida H, et al. (2003) Insertion of cellular sequence and RNA recombination in the structural protein coding region of cytopathogenic bovine viral diarrhoea virus. *J Gen Virol* 84: 447-452.
416. Holmes EC (2003) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol* 11: 543-546.
417. Scheel TK, Galli A, Li YP, Mikkelsen LS, Gottwein JM, et al. (2013) Productive homologous and non-homologous recombination of hepatitis C virus in cell culture. *PLoS Pathog* 9: e1003228.
418. Lai MM (1992) RNA recombination in animal and plant viruses. *Microbiol Rev* 56: 61-79.
419. Becher P, Orlich M and Thiel HJ (2001) RNA recombination between persisting pestivirus and a vaccine strain: generation of cytopathogenic virus and induction of lethal disease. *J Virol* 75: 6256-6264.
420. Vennema H, Poland A, Foley J and Pedersen NC (1998) Feline infectious peritonitis viruses arise by mutation from endemic feline enteric coronaviruses. *Virology* 243: 150-157.



## ANNEXE

### **1. Genetic diversity of the Hepatitis C virus: impact and issues in the antiviral therapy. Highlight Topic 2007.**

Le Guillou-Guillemette H, Vallet S, Gaudy-Graffin C, Payan C, Pivert A, Goudeau A and Lunel-Fabiani F. *World Journal of Gastroenterology* 2007, Vol 13, 17: 2416-2426

Francoise Lunel Fabiani, Professor, Series Editor

## Genetic diversity of the hepatitis C virus: Impact and issues in the antiviral therapy

H Le Guillou-Guillemette, S Vallet, C Gaudy-Graffin, C Payan, A Pivert, A Goudeau, F Lunel-Fabiani

H Le Guillou-Guillemette, F Lunel-Fabiani, A Pivert, Laboratory of Virology, University Hospital of Angers, France and Laboratory HIFI, UPRES EA 3859, University of Angers, France  
S Vallet, C Payan, Department of Microbiology, EA 3882, CHU Morvan, University Hospital of Brest, France  
C Gaudy-Graffin, A Goudeau, Université François Rabelais, INSERM ERI 19 et CHRU de Tours, Tours, France  
Correspondence to: F Lunel-Fabiani, Professor, Laboratory of Virology, Department of infectious agents, University Hospital of Angers, 4 rue Larrey, 49933 Angers Cedex 9, France. [frlunel-fabiani@chu-angers.fr](mailto:frlunel-fabiani@chu-angers.fr)  
Telephone: +33-2-41354709 Fax: +33-2-41354164  
Received: 2006-12-23 Accepted: 2007-03-13

therapy. *World J Gastroenterol* 2007; 13(17): 2416-2426

<http://www.wjgnet.com/1007-9327/13/2416.asp>

### INTRODUCTION

The role of genetic variability in the natural history by hepatitis C Virus (HCV) infection and in the primary resistance to treatment remains unclear. In particular, the mechanisms underlying the resistance of HCV to Interferon treatment are very different from those by which the human immunodeficiency virus (HIV) becomes resistant to antiretroviral drugs. In this review, we will describe the impact of HCV genetic polymorphism in the treatment response.

### HCV VARIABILITY: ORIGIN AND CLASSIFICATION

HCV has a single-strand positive RNA genome and displays a high genetic diversity. This diversity results from defects in the repair activity of the RNA-dependent RNA polymerase (resulting in nucleotide substitutions) and from the absence of 5' to 3' exonuclease activity (lack of error correction)<sup>[1]</sup>. The mean frequency of nucleotide mutations varies from  $1.4 \times 10^3$  to  $1.9 \times 10^3$  substitutions per nucleotide and per year. This estimate is based on comparisons of the major sequences of complete genomes obtained after eight years of evolution in a chimpanzee and 13 years in a human<sup>[2-3]</sup>.

Some of the mutations accumulating during replication are silent or synonymous. These mutations have no impact on the amino-acid sequence of the viral protein, but may affect the secondary structure of the genomic RNA. Other so-called non-synonymous mutations lead to changes in protein sequence and the emergence of variants. Other mutations lead to the production of defective viral particles and are therefore lethal. The regions of the genome corresponding to essential viral functions (domains involved in translation or replication), or displaying major structural constraints (non-coding 5' and 3' ends) are the best conserved; indeed, the non coding 5' region is the most highly conserved region of the genome, with more than 90% identity between the sequences of distantly re-

### Abstract

The hepatitis C Virus (HCV) presents a high degree of genetic variability which is explained by the combination of a lack of proof reading by the RNA dependant RNA polymerase and a high level of viral replication. The resulting genetic polymorphism defines a classification in clades, genotypes, subtypes, isolates and quasispecies. This diversity is known to reflect the range of responses to Interferon therapy. The genotype is one of the predictive parameters currently used to define the antiviral treatment strategy and the chance of therapeutic success. Studies have also reported the potential impact of the viral genetic polymorphism in the outcome of antiviral therapy in patients infected by the same HCV genotype. Both structural and non structural genomic regions of HCV have been suggested to be involved in the Interferon pathway and the resistance to antiviral therapy. In this review, we first detail the viral basis of HCV diversity. Then, the HCV genetic regions that may be implicated in resistance to therapy are described, with a focus on the structural region encoded by the E2 gene and the non-structural genes NS3, NSSA and NSSB. Both mechanisms of the Interferon resistance and of the new antiviral drugs are described in this review.

© 2007 The WJG Press. All rights reserved.

**Key words:** Hepatitis C virus; Genetic diversity; Therapy resistance; E2; NS3; NSSA; NSSB

Le Guillou-Guillemette H, Vallet S, Gaudy-Graffin C, Payan C, Pivert A, Goudeau A, Lunel-Fabiani F. Genetic diversity of the hepatitis C virus: Impact and issues in the antiviral

lated strains<sup>61</sup>. The region encoding the capsid is also well conserved, with 81% to 88% sequence identity between isolates<sup>61</sup>. The most variable region of the genome is that encoding the envelope glycoproteins, E1 and E2. The sequences of hypervariable regions (HVR1 and HVR2) of E2 in strains isolated from different patients may differ by more than 50%. The polypeptides encoded by these hypervariable regions are therefore very tolerant to amino-acid substitutions<sup>6,7</sup>.

The classification of HCV, redefined and simplified in 1998, is based on the topology of the trees obtained by phylogenetic relationships between viral variants. HCV variants can be classified into six clades, and then into subtypes corresponding to subgroups of the most closely related viruses within a clade<sup>61</sup>. The virus genotype is indicated by an Arabic number (from 1 to 6), associated with a lower-case letter to indicate the subtype. This new nomenclature has led to the reclassification of virus genotypes 7, 8, 9 and 11 as genotype 6 viruses, and the reclassification of genotype 10 viruses as genotype 3 viruses. Recently, the status of HCV-genotype nomenclature has been re-examined in an attempt to resolve conflicting subtype and genotype designations. A complete list of currently recognized genotypes and subtypes was published<sup>62</sup>. The euH-CVDB (<http://hepatitis.ibcp.fr>), DDJB (<http://www.ddbj.nig.ac.jp>) and Los Alamos (<http://hcv.lanl.gov>) databases currently include a large number of HCV sequences, making it possible to compare an isolate with a large number of reference sequences<sup>63</sup>. Although currently restricted to specialist laboratories, the gold standard method for HCV genotyping is sequencing of the NS5B region followed by phylogenetic analysis including comparison to reference sequences<sup>11,12</sup>.

A chimeric virus generated by an intergenotypic homologous recombination event (genotype 2k/genotype 1b) in the NS2 gene was first described in the St Petersburg area: this demonstrated the occurrence of recombination phenomena in HCV<sup>13</sup>. Another recombinant form (genotype 2i/genotype 6p) was recently described in Vietnam, with a point of recombination between NS2 and NS3<sup>14</sup>. Such recombination events have already been described for GBV-C and Dengue virus<sup>15,16</sup>. In HCV, homologous recombination may be favoured by the nature of HCV risk behaviour in which there may be frequent exposures, for example among Intra Venous Drug Users. In practice, genotyping is based solely on polymorphism in a single genomic region (5'NC or NS5B), so it is currently impossible to estimate the frequency of recombination events.

Studies of the global distribution of HCV genotypes and analyses of their phylogenetic relationships have provided insight into their emergence and diversification over the centuries. The worldwide distribution of genotype 2 began some 90 to 150 years ago; that of subtype 1b began 60 to 70 years ago and that of genotype 3 began about 40 years ago<sup>17</sup>. Genotypes 4 and 6 emerged much earlier: 350 and 700 years ago, respectively<sup>18</sup>. The simultaneous presence of a large number of subtypes of a viral genotype in a limited geographic region is indicative of the long-standing endemic presence of the virus in the population studied. Genotype 2 is frequent in West Africa (59% to 100%,

depending on the region studied), and many subtypes have been identified (2c to 2l)<sup>19</sup>. The same is true of genotype 4 in Central Africa (4b, 4c, 4e and 4m)<sup>20</sup>. Conversely, the limited number of subtypes in Europe (where genotype 1, 2, 3 are the more frequently found), North America and Japan is consistent with the more recent introduction of HCV into these populations. Another level of complexity is found within a given infected patient termed as the quasispecies population. Quasispecies are populations of different but closely related genomes and data have suggested that Interferon alpha exerts a selective pressure on HCV quasispecies<sup>21,22</sup>. Thus, these molecular polymorphisms are clinically relevant and are one of the major factors in determining the outcome of the Interferon therapy.

Since 1986, before the discovery of HCV, the efficacy of Interferon alpha has been demonstrated in the treatment of chronic non A non B hepatitis<sup>23,24</sup>. Interferon alpha is a member of the cytokine family produced by cells in response to viral infections or other stimuli by binding to their specific receptors on the surface of target cells. Interferons stimulate a cascade of intracellular signalling pathways that result in the suppression of numerous Interferon-Stimulated Genes (ISGs) whose products (proteins) can mediate the effects of Interferon. Among these, are the INF induced, double stranded (ds) RNA dependant protein kinase (PKR), the 2' 5' oligoadenylate synthetase and the Mx proteins, of which the antiviral activities have been well demonstrated<sup>25</sup>. Interferon also acts as a stimulus of immunity and modulates cell growth, differentiation and apoptosis<sup>26</sup>.

In an effort to understand the role of different factors on the outcome of Interferon therapy, numerous studies have been performed to define predictors of response and both viral and host factors have been studied.

Among the viral factors examined, viral genotype seems to be the most important predicting factor; it appears to be a predictive parameter of the SVR as strong as well-established predictive parameters such as the pre-treatment viral load, the stage of fibrosis on liver, and the age and the sex of the patient. Trials clearly reported a high SVR rate of 76% to 84% in patients infected with HCV genotypes 2 or 3 whereas a weaker SVR of 42% to 52% was obtained in HCV genotype 1-infected patients<sup>27-29</sup>. The SVR is likely to be higher for genotype 2 than for genotype 3 where a shortened treatment period was experienced, but ongoing studies will determine the role of the initial viral load and a rapid virologic response<sup>30,31</sup>. Therapeutic outcomes are well known for these three genotypes because of their geographic distribution; genotype 1 represents more than 70% of the HCV infections in the Western world, genotypes 2 and 3 infect 10% to 20% of the other patients. Studies focused on the outcome of the antiviral therapy conducted in patients infected by HCV genotype 4 are mainly provided by Egyptian cohorts and results showed an intermediate SVR rate comprised between 55% to 69%<sup>32,33</sup>. Genotypes 5 and 6 are less studied because of their minor distribution, patients infected by these genotypes may achieve a SVR at a level between the SVRs of genotype 1 and genotypes 2-3<sup>34</sup>.

Differences in the SVR rate observed among differ-

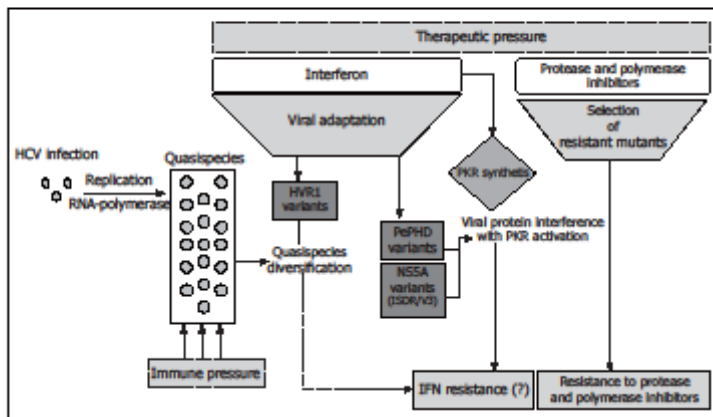


Figure 1 Mechanisms involved in HCV resistance to antiviral therapies.

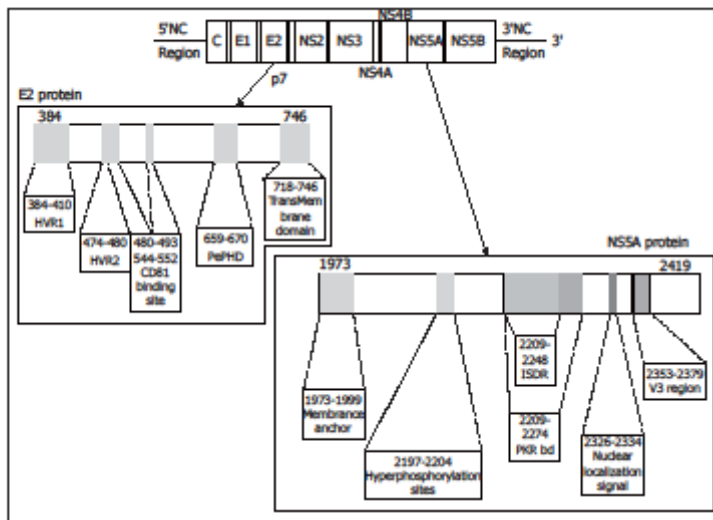


Figure 2 Linear representation of the E2 and NS5A proteins of the HCV, location of the main interesting domains of each protein. Each site of interest is represented by a grey zone on the protein and is linked to a box containing its name and amino acid positions.

ent HCV genotypes have highlighted a presumptive role of the viral genetic determinants and have suggested that some virus-encoded functions are involved in the response to Interferon. Although a very efficient *in vitro* or *in vivo* or animal models of HCV infection is not available, several studies have been undertaken to identify which viral gene may interfere with antiviral molecules. Both *in vivo* and *in vitro* studies demonstrated that the structural proteins E1 and E2, the non-structural proteins NS3, NS5A and NS5B may contribute to the resistance of the combined Interferon alpha and ribavirin therapy. Moreover, all these proteins share a genetic polymorphism involved in the resistance mechanisms as shown in Figure 1.

### GENOMIC DOMAINS CONTRIBUTING TO RESISTANCE

As previously described by Pawlotsky in this issue of the [www.wjgnet.com](http://www.wjgnet.com)

journal, HCV presents a high genetic diversity. The intra-genotype analysis of this diversity along the viral genome shows different degrees of variation; regions such as the 5' UTR and the core are highly conserved, the non-structural regions 2, 3, 5b and the 3'UTR are relatively variable whereas the envelope regions E1 and E2 and the NS4 and the NS5A genes exhibit the highest sequence diversity.

### Polymorphism and significance of amino acid substitutions within E2 regions in Interferon alpha resistance

E2 glycoprotein is a type I transmembrane protein of 70 kDa, with an N-terminal ectodomain and a C-terminal hydrophobic anchor. It assembles with E1 to form a heterodimer. It has been shown to interact with two potential HCV receptors, the human tetraspanin CD81 and the human scavenger receptor SR-BI, in experiments based on soluble E2 (sE2) binding (Figure 2). The role of E2 in the cell entry of HCV is extensively reviewed in this issue by

PKR		K	K	A	V	S	P	L	L	L	I	I	I	
		+		+										
PePHD domain		R	S	E	L	S	P	L	L	L	T	T	T	HCV-1 (1a)
		-	-	-	-	-	-	-	-	-	S	-	-	HCV-J (1b)
		-	Q	-	-	-	-	-	-	-	H	S	-	HCJ.06 (2a)
		-	G	Q	Q	-	-	-	-	-	H	S	-	HCJ.08 (2b)
		-	-	-	Q	H	-	-	-	-	H	S	-	HCV-NZL1 (3a)

Figure 3 Sequence homology between HCV E2-PePHD domain and PKR. PePHD domains from various HCV genotypes (1a, 1b, 2a, 2b, 3a) are aligned and compared with the PKR sequence. Identical and similar amino acids to PKR are shown by ( ) and (+), respectively. PKR auto-phosphorylation sites are underlined.

#### Dubuisson *et al.*

One of the major ways by which cellular Interferon alpha inhibits viral replication involves the Interferon-alpha-inducible double-stranded RNA-activated Protein Kinase R (PKR). Indeed the Interferon alpha binds to the PKR and leads to its autophosphorylation which in turn initiates the phosphorylation of eukaryotic initiation factor 2 alpha (eIF2 $\alpha$ ) by the PKR. This phosphorylation inhibits the RNA transcription. Indeed, eIF2 $\alpha$  is necessary to initiate the translation by forming a complex with GTP and met-tRNA and then allowing binding to the 40S ribosomal subunit. The E2 glycoprotein, or more precisely, a 12-amino acid domain of this protein located between residues 659 and 670 and known as PePHD (PKR-eIF2 $\alpha$  phosphorylation homology domain), is involved in PKR inhibition. Taylor *et al.* showed that PePHD was the main domain of E2 able to bind PKR *in vitro*<sup>31</sup>. The PePHD motif sequence is very similar to that of the auto-phosphorylation sites of PKR and its target, eIF2 $\alpha$ , and this similarity is more marked for genotype 1 viruses than for genotype 2 and 3 viruses (Figure 3). E2 glycoproteins of genotype 1 viruses (HCV-1 E2) behave *in vitro* as pseudosubstrates, inhibiting the kinase activity of PKR. In mammalian cells, the stimulation of translation by HCV-1 E2 is consistent with the hypothesis of PKR inhibition. This inhibitory activity has also been observed in a yeast model (*Saccharomyces cerevisiae* expressing HCV1 E2). The replacement of HCV-1 E2 by proteins identical to those of genotype 2 and 3 viruses abolishes this inhibitory effect. An interaction between genotype 1 strains and PKR, *via* PePHD, has therefore been proposed to account for the intrinsic resistance of the strains of this genotype to Interferon alpha<sup>31</sup>.

The clinical relationship between aminoacid sequence of PePHD and the outcome of Interferon therapy has been a matter of controversy. A few studies have addressed polymorphism of the PePHD region in patients carrying strains of genotypes 1, 2 and 3. Abid *et al.* studied a small number of patients infected with genotype 1b HCV. They initially showed that in some patients responding to treatment, the virus had a PePHD sequence identical to that of the HCV-J strain (RSELSPLLLSTT), calling Taylor's hypothesis into question<sup>34</sup>. For HCV 2a/b isolates, conflicting results about the association of PePHD mutations and treatment response have been published<sup>37,38</sup>. A number of studies focusing exclusively on the diversity of PePHD sequences in genotype 1b HCV before treatment have since shown strong conservation

of this motif, regardless of the response subsequently obtained<sup>39,40</sup>.

The genetic heterogeneity of the E2-HVR1 region can also be used to describe the composition of quasispecies precisely, provided a large number of clones are analysed. The preliminary studies in this area were based on analysis of single-strand polymorphisms or of a restricted number of clones (less than 10). A correlation between high levels of quasispecies complexity before Interferon alpha monotherapy and a lack of response to treatment was reported<sup>41,42</sup>. This led to the suggestion that genetic variability ensured a reservoir of potentially resistant strains. However, more recent studies with larger numbers of sequenced clones or based on SSCP enhanced sensitivity protocols have not confirmed the link between the genetic complexity of HCV and virological response<sup>43</sup>. These conflicting results illustrate the problems of methodological standardisation associated with studies of quasispecies.

#### Polymorphism of the HCV non-structural regions and impact on the treatment response.

The NS2, NS3 and NS4A/B proteins may not be implicated in the antiviral treatment resistance with the current combination pegylated Interferon alpha plus ribavirin. However, new therapies are in development targeting the protease encoded by NS3 and the polymerase encoded the NS5B. As already observed in the anti-HIV HAART treatment, the genetic polymorphism of NS3 and NS5B will be analysed in relation to therapy. HCV presents a high mutation rate and the antiviral pressure may favour the emergence of resistant strains. Mutation points have been already observed *in vitro* (replicon system).

#### Polymorphism of NSSA and Interferon resistance

The NS5A protein is the non structural HCV protein which is the protein most reported to be implicated in the Interferon resistance. Many interactions between NS5A and different molecules from several intracellular pathways have been also demonstrated in cellular *in vitro* systems. An overview of the key roles of the genetic heterogeneity of this protein will be detailed in this paragraph.

**Structure of the NS5A protein:** NS5A is a phosphoprotein found in a basally phosphorylated form of 56kDa and a hyperphosphorylated form of 58 kDa. It varies in length, from 445 amino acids in genotype 4a, 447 amino acids in genotype 1b, 448 in genotypes 1a and 1c, 450 in genotype 5a, 451 in genotype 6a, 452 in genotype 3 to 452 amino acids in genotype 2a and b. The amino acid sequence varies depending on the genotype. The NS5A protein is a pleiotropic protein involved both in the viral replication and in many interactions with cellular signalling pathways. Although its function remains unclear, many domains of interest have been described in this protein (Figure 2).

Many interactions between NS5A and cellular signalling pathways have been reported and the interacting sequence has been identified for some of them. Although studies have been conducted *in vitro*, results strongly suggest a potential involvement of the NS5A protein

in the establishment of a chronic hepatitis and in the carcinogenesis and outcome of a liver tumour (for review, see<sup>[44]</sup>). In this review, we will focus only on the protein interaction of NS5A with the Interferon pathway.

**Interaction with the cellular Interferon pathway:** Phosphorylation of PKR triggered a global translational repression comprising the viral replication. Gale *et al.*<sup>[45]</sup> identified in the C-terminal NS5A a Protein Kinase R binding domain. They demonstrated *in vitro* that the NS5A protein is able to bind and inhibit the Interferon- $\alpha$ -inducible double-stranded RNA-activated PKR, the interaction occurring *via* the NS5A PKR binding domain. In that way, the NS5A protein has been suggested to balance the Interferon cellular antiviral pathway and to be involved in the resistance to the Interferon based-therapy.

The NS5A protein is also able to interact with the Interferon pathway in a PKR-independent manner. It has been demonstrated *in vitro* that NS5A induces the expression of the pro-inflammatory chemokine interleukin 8 (IL-8) at both the mRNA and protein levels. This chemokine is known to inhibit directly the Interferon  $\alpha$  activity. The clinical relevance of these *in vitro* studies has been reported and it appears that the IL-8 serum levels were higher in infected patients than in healthy controls. A second study showed an increase of the pre-treatment IL-8 level in the non-responder than in the SVR<sup>[46,47]</sup>.

**Mutations in the NS5A protein and relationship with the Interferon resistance:** The 40 first amino acids of the PKR binding domain present a high level of variability. It was termed the Interferon sensitivity determining region (ISDR) reported to play a key role in the Interferon therapy response<sup>[48,49]</sup>. Molecular analysis of this 40 amino acid region showed the potential role of mutations in the resistance to Interferon therapy. They demonstrated a correlation between the success of Interferon therapy and the variability of the ISDR domain in Japanese patients with HCV genotype 1b or 2 infection. Patients infected with viral strains whose ISDR harbours more than four mutations different from the HCV-J sequence (the Japanese prototype strain defined as the wild-type) were responders to Interferon therapy whereas patients infected with a wild-type or strain harbouring less than four mutations in ISDR were non-responders. Numerous studies investigating the correlation between the mutations in ISDR in genotype 1b HCV and the outcome of the Interferon-based therapy have been undertaken. Japanese results were concordant with the initial study whereas European and north-American groups did not describe such a correlation and debate remains controversial. A recent meta-analysis has been conducted on 1230 ISDR sequences from HCV 1b-infected patients by Pascu *et al.*<sup>[50]</sup>. Sequences were provided from Japanese and European studies. Analyses were realised by logistic regression and clearly demonstrated an association between number of ISDR mutations and response to the treatment both in Japanese and European patients, irrespective from a geographical distribution. First, controversial results may be explained in part by the HCV European strains, most of the European HCV 1b strains present less than 3 mutations and it was difficult to identify a statistical

difference between the Interferon therapy response among European infected patients. It is also important to note that treatment schedules were not the same in Europe and in Japan. A correlation between the variability in the ISDR region and a SVR has been demonstrated in the genotype 2a-infected patients. Conversely, investigations in the patients infected by genotypes 2b or 3a did not find any relation between the successful therapy and genetic variability of the NS5A gene<sup>[51,52]</sup>. Additionally, *in vitro* studies did not report any interaction between NS5A genotype 3a and PKR in viral resistant strains to the Interferon therapy<sup>[53]</sup>.

The major implication of the variability in the resistance to antiviral agents has been recently pointed out, but few studies focused on the entire NS5A protein at the quasispecies level and its kinetics under therapy. Inshauspe *et al.*<sup>[54]</sup> identified another domain localised in the C-terminal region of NS5A and termed it the V3 domain. It is 27 amino acids in length and it harbours a high variability level. Duverlie *et al.*<sup>[55]</sup> demonstrated a relationship between the mutation level in V3 and the response to Interferon therapy. Resistant strains were highly conserved whereas sensitive strains were variable. Our group and others confirmed this correlation<sup>[56]</sup>. We followed HCV quasispecies diversity at baseline and under Interferon  $\alpha$ -ribavirin combined treatment in the entire NS5A gene and in each region of interest (PKR-bd, ISDR and V3). As reported by Puig-Basotti *et al.*<sup>[57]</sup>, the V3 domain showed a higher quasispecies diversity in responder patients than in non-responders in pre-treatment samples and these data confirmed the potential role of the genetic diversity in the success of the Interferon-based therapy.

#### NS3 protease as a target for specific antiviral therapy

Previous studies on the full-length genome sequence have led to the HCV NS3 domain as being classified as one of the less variable regions of the genome<sup>[58,59]</sup>. Nevertheless, it appears important to consider the diversity of HCV protease, because it is an attractive target for specific anti-HCV therapy.

Catalytic functions and the three-dimensional structure of HCV NS3 protease have been reviewed in this issue by Dubuisson *et al.* Some of the structural and functional constraints affecting the NS3 protease have been demonstrated by the definition of major domains, including the catalytic triad, the substrate binding pocket, the NS4A binding domain and the residues binding to the zinc ion<sup>[60]</sup>.

**Natural polymorphism of NS3 protease:** The diversity of the HCV NS3 protease gene in clinical samples has been studied<sup>[61,62]</sup>. Although the protease NS3 is considered to be one of the less variable genes in the HCV genome, variability of both nucleotide and amino acid sequences exists. The proportion of synonymous substitutions affecting the region is significantly higher than the proportion of non synonymous ones suggesting that the NS3 protease mutations are the products of random genetic drift rather than of positive selection. NS3 protease structural and chemical integrity is required for it to process the HCV polyprotein. The perpetuation of substitutions

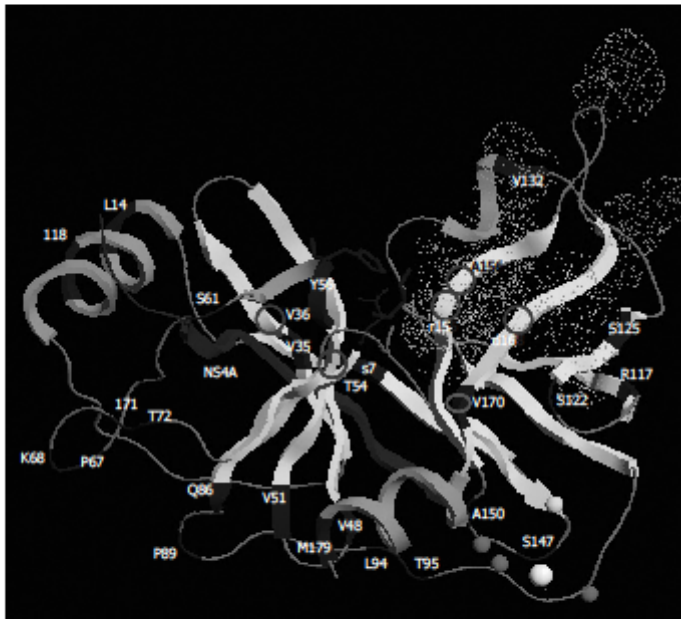


Figure 4 Three-dimensional structure of NS3 protease (PDB accession code 1NS3). Polymorphism and main residues implicated in resistance are shown. The protease is shown based on its secondary structure in light grey. Main polymorphic residues are shown in dark grey. The side chains of the residues forming the catalytic triad (H57, D81, and S139) are displayed in dark grey ball-and-stick representation. The NS4A cofactor is shown in dark grey and Zinc ion as a white ball. Zn<sup>2+</sup> ligands (C97, C99 and C149) and H145 residue are modeled in light grey spheres. Stars correspond to the side chains of the residues forming the S1 to S6 substrate binding pockets (grey dots cloud) of the enzyme. Residues implicated in resistance to protease NS3 inhibitors are shown with circles. This figure was prepared using the RasTop programme version 2.0.3-VF.

depends in large part on the extent to which the viral protease can tolerate them while remaining functional. The NS3 protease appears to be tolerant, particularly in its loop regions between helices and sheets (Figure 4). Sporadic mutations affecting the catalytic residues have already been observed from different patients. These catalytic-triad amino acid mutations probably lead to a loss of function, and therefore may be lethal. In that sense, they are not crucial for resistance to NS3 protease inhibitors<sup>[64]</sup>.

The NS3 region is known to contain a large number of T-cell epitopes. T-cells directed against the NS3 domain seem to be of particular importance both during the acute and chronic phase of the disease as the CD4<sup>+</sup> and CD8<sup>+</sup> T-cell response to the NS3 protease correlates with clearance and control of the infection<sup>[65]</sup>. As reviewed by Dubuisson *et al* in this issue, the NS3 serine protease influences the Interferon-dependent innate cellular host defence by inhibition of RIG-1 and TLR3 signalling<sup>[64,66]</sup>. As use of Interferon alpha is the standard of care for HCV, the potential of protease inhibition to enhance Interferon therapy has attracted much interest.

**NS3 Polymorphism and impact on response to antiviral specific treatment:** The NS3 protease three-dimensional structure description has provided the necessary detailed insight to permit rational inhibitor design. Therapies based on such inhibitors are still at the development and testing stage<sup>[64,67]</sup>. The generation of such therapies based on the inhibition of site-specific proteolysis has been clearly illustrated in the development of effective inhibitors of human immunodeficiency virus type 1 (HIV-1)<sup>[68]</sup>.

*In vitro* and *in vivo* phase I and II studies have shown that

evaluated NS3 protease inhibitors (BILN 2061, VX-950 and SCH 503034) developed drug resistant mutations<sup>[69,70]</sup>. These new drugs meet the same problems of viral resistance that exist in the specific anti-viral treatment of HIV and HBV infected individuals. In untreated patients, less than five percent of protease variants present resistant mutations to inhibitors SCH6, SCH 503034 and VX-950<sup>[61,71,72]</sup>. The reduced fitness of the most resistant variant may explain that such variants are rarely found in naïve patients<sup>[69,70,73]</sup>. The replicative fitness of resistant viruses is a critical parameter of viral resistance and is an important factor to consider in achieving sustained virological response. It is underestimated due to low replicative resistant strains not being detected.

Both HCV protease inhibitors evaluated in humans, VX-950 and SCH-503034, have demonstrated a very strong correlation between their antiviral effect, their serum trough levels, and the development of resistance<sup>[67,74]</sup>. Changing in viral quasiespecies with a selection of clones resistant to VX-950 has been observed in patients who experienced viral load breakthrough to VX-950 monotherapy<sup>[66]</sup>. Resistant variants may have poorer replicative fitness than wild-type viruses. However, VX-950-resistant clones isolated from patient plasma from the initial phase 1b study were thought to be associated with the lower doses of monotherapy, and selection for these clones was based on inadequate suppression of virus<sup>[75]</sup>. Four residues in the NS3 protease, when substituted, are known to be associated with resistance/reduced sensitivity to VX-950. R155K/T/S/M and A156T/V, located close to the VX-950 binding groove, a domain of NS3 protease, confer moderate to high level resistance to VX-950; V36A/M and

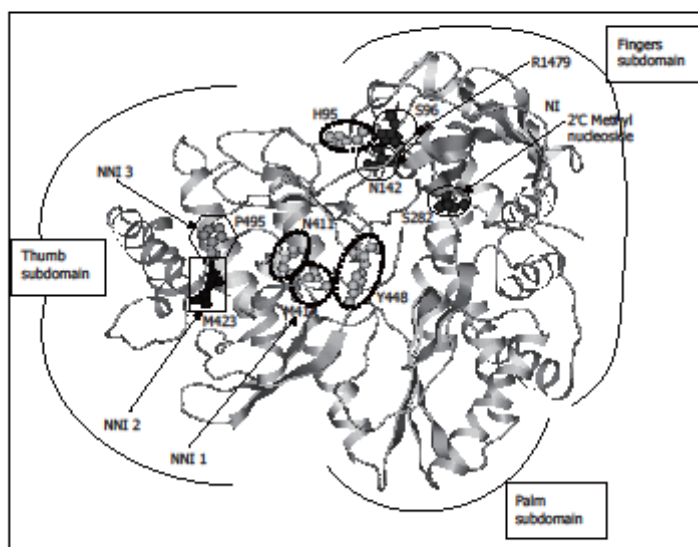


Figure 5 Three-dimensional structure of NS5B RNA-dependent RNA polymerase (POB accession code 1QUM). The three subdomains of the polymerase are shown (Fingers, Thumb and Palm) as well as the main residues targeted by each class of polymerase inhibitors (NNI 1: Non nucleoside class 1, NNI 2: Non nucleoside class 2, NNI 3: Non nucleoside class 3, NI: Nucleoside Inhibitor). This figure was prepared using the RasTop programme version 2.0.3-VF.

T54A located further from the binding groove confer low-level of resistance to VX-950 (Figure 4).

Three mutations, T54A, V170A and A156S mutations conferred low to moderate levels of resistance to SCH-503034 in a replicon model. Mutants with A156T substitution are highly resistant to this compound<sup>[73]</sup>. Combinations of VX-950 and SCH-503034 may be limited by the selection of cross-resistance mutations.

The polymorphic outside residues central to NS3 protease activity and substrate binding sites, may prove to be accessory substitutions of resistance mutations contributing to clinical resistance to HCV protease inhibitor therapy. This has been supposed for 3 separate second site mutations P89L, Q86R and G16R *in vitro* with the protease SCH6 inhibitor<sup>[74]</sup>. Moreover, resistance to future NS3 protease inhibitors could occur through mutations arising either in the protease gene itself or, as has been shown for HIV-1, in cleavage sites of the protein<sup>[9]</sup>.

Resistance to specific HCV inhibitors will probably emerge in case of insufficient antiviral pressure. Viral load monitoring and checking for emergence of NS3 protease resistance mutations could contribute to surveillance of sustained viral response.

#### NS5B polymorphism and impact of the anti-polymerase therapy

**Structure of the protein:** The last gene of the HCV is the NS5B which encodes the HCV polymerase, which is an RNA-dependent RNA polymerase (RdRP). This 68 kDa protein plays a central role in HCV replication and in viral diversity because of the lack of a proof-reading activity. The NS5B protein belongs to the tail-anchored proteins as previously detailed by Dubuisson *et al.* The crystal structure of this enzyme has been fully determined. The three dimensional structure is similar to other polymerases;

it shows a typical right-hand arrangement with fingers, palm and thumb domains. This organisation forms a cavity called a NTP tunnel in which the traffic of the nucleotides to the active site occurs<sup>[75,76]</sup> (Figure 5).

**Genetic polymorphism and enzymatic activity:** As all enzymatic proteins, the nucleic acid sequence is a fundamental determinant of the three-dimensional structure and it determines good enzymatic activity. It has been shown that the active site does not tolerate any mutations. Its amino acid sequence is highly conserved among all HCV genotypes and any mutation may inhibit viral replication. Several consensus sequence motifs have also been described along the NS5B polymerase; five of them are in the palm sub-domains (motifs A, B, C, D and E) and one is in the finger domain (motif F)<sup>[76,77]</sup>. Most of the single amino-acid substitutions introduced into the conserved motifs by directed mutagenesis resulted in decreased or abolished viral replication and enzymatic activity<sup>[80,81]</sup>.

**NS5B and viral genotype:** Although the NS5B protein remains a relatively well conserved protein because of its enzymatic function, the NS5B domain, along with other domains of the HCV, supports a slight variability which provides a basis for distinguishing viral type and subtype. Indeed, the nucleic acid sequence is less conserved than the 5'NC but sufficient to define type and viral subtype. The relation between the structure of the NS5B and the response to the Interferon therapy has rarely been studied and most studies were focused on genotype 1b. Results were contradictory. Vuilleumoz *et al.*<sup>[82]</sup> did not find any mutation differences (point and rate mutations) between responders and non-responders to the HCV NS5B strain; conversely Kumagai *et al.*<sup>[83]</sup> reported two polymorphisms in NS5B associated with an early viral clearance in the treatment of HCV genotype 1b-infected patients.



**New antiviral treatments targeting the HCV polymerase:** The HCV polymerase represents an important target for the new inhibitors of HCV replication and numerous assays are currently in development using the *in vitro* subgenomic HCV replicon system. Thanks to the knowledge of the three dimensional structure of NS5B, screening of many chemical molecules has been undertaken. Three classes of inhibitors are now reported and include nucleoside analogues, non nucleoside analogues and pyrophosphate mimics. Drugs belonging to the nucleoside inhibitors are in competition with the substrate of the RdRP and inhibit polymerase elongation. The binding site is localised in the fingers sub-domain of the polymerase. Molecules of this class have a 2' methylribose structure and some of them are now well characterised<sup>[64,65]</sup>. 2'-C-Methyl-Cytidine, 2'-C-Methyl-Adenosine and 2'-C-Methyl-Guanosine have been shown to be effective against genotype 1a and 1b in Con1 subgenomic replicons. A single mutation, S282T, confers a cross resistance to all the 2'-C-Methyl-Nucleoside, and to their prodrug as such the NM283 (prodrug of the 2'-C-Methyl-Cytidine known as the Valopicitabine)<sup>[66,67]</sup>. The R1479 (4'-Azido-Cytidine), which is another nucleoside inhibitor, is under development<sup>[68]</sup> and a recent study reported the mutation resistance points as S96T and S96T/N142T. This work showed no cross resistance between R1479 and the 2'C methyl nucleosides<sup>[69]</sup>. The NS5B genes of non genotype 1 (2a, 2b, 3, 4 and 6) have been studied in a genotype 1 replicon background; most of them reported a weak replication that enabled drug sensitivity assays. It is important to note that prototype sequences of genotype 4a and 6a encoded T282 and L495 which are implicated respectively in the *in vitro* resistance to the 2'-C-Methyl nucleoside inhibitors and to the benzimidazole non-nucleoside inhibitors<sup>[70]</sup>. NM283 (Idenix®) has demonstrated an anti-HCV activity in the trials of phases I and II a, it is now in the phase II b clinical development. Results of the phase I trial of R1626 (Roche®), another nucleoside inhibitor, were also encouraging. The first phase II development study is now starting<sup>[69]</sup>.

Compounds belonging to the non-nucleoside inhibitors (NNI) act as allosteric inhibitors that may affect the initiation step of the polymerisation (inhibition of the conformational transition prior to the formation of an efficient pre-elongation complex, inhibition of RNA binding to NS5B or non-competitive inhibitors of NTP incorporation). Different molecules belong to this class and have distinct targets. Indeed, three binding pockets have been localised and are used to define each class of NNI (1 to 3). Two of them are located in the thumb domain. The benzothiazidines group (NNI 1) binds to one pocket located at the interface between the thumb and palm domains and centralised around Met414. *In vitro* studies have been carried out mainly with the genotype 1b subgenomic replicon and substitutions have been shown to decrease susceptibility (M414T, C451R, G558R, H95R)<sup>[71]</sup> or to confer resistance (H95Q, N411S, M414L/T, Y448H)<sup>[72]</sup> to benzothiazine drugs from different pharmaceutical companies. The second pocket, located at the base of the thumb domain, binds to the large group of molecules comprising thiophene 2-carboxylic

acids, phenylalanine derivatives and substituted pyranones (NNI 2) and the binding central amino acid is Met423. *In vitro* assays reported differences in the inhibitory activity between the different genotypes tested. For example, Howe *et al.*<sup>[73]</sup> showed that the susceptibility level to the HCV-371 (NNI 2) was equal among the genotype 1b isolates but varied with genotypes 1a, 3a and 4 (5 to 59 fold less susceptible than genotype 1b). The last pocket is located at the top of the thumb around Pro495 and is targeted by the benzimidazole 5-carboxamide inhibitors (NNI 3)<sup>[64]</sup>. As with the other nucleoside inhibitors, sensitivity has been studied *in vitro* in the subgenomic replicon; Tomei *et al.*<sup>[74]</sup> have recently reported that HCV strains resistant to benzimidazole molecules harbored a substitution in Pro495. Interestingly, it has been shown that a chimaeric 1b:2b HCV replicon, harbouring a genotype 2b NS5B, was sensitive to the 2'-C-Methyl-Adenosine but completely insensitive to the NNI 1 (benzimidazole) and NNI 3 (thiophene)<sup>[64]</sup>. These results underlined that a genotype specificity may exist for each anti-polymerase inhibitor.

All of these results showed the importance of genetic diversity in the HCV therapeutic strategy. Preliminary results, both *in vitro* and *in vivo* trials, strongly suggest a relationship between genetic polymorphism and sensitivity to drug. Identification of the mutations set corresponding to each antiviral drug will be very helpful. It will represent a chance to anticipate treatment regimes at baseline based on the viral genotype, an approach currently only done for HIV therapy, many years after the beginning of HAART<sup>[75]</sup>.

Finally, the genetic diversity of the HCV largely impacts in the global approach of this viral disease as in the treatment management as well as in the development of new HCV antiviral therapies.

## ACKNOWLEDGMENTS

The authors thank the Fontevraud Group Study that largely contributes to the study of the HCV infection. Studies focused on the NS5A and the E2-PePHD diversity were realised with grants from the Agence Nationale de Recherche sur le SIDA et les Hépatites (ANRS). We thank Professor David Archer, School of Biology, University of Nottingham, UK, for careful reading of the manuscript.

## REFERENCES

- Steinhauser DA, Domingo E, Holland JJ. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* 1992; **122**: 281-288
- Ogata N, Alter HJ, Miller RH, Purcell RH. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc Natl Acad Sci USA* 1991; **88**: 3392-3396
- Okamoto H, Kojima M, Okada S, Yoshizawa H, Iizuka H, Tanaka T, Muchmore EE, Peterson DA, Ito Y, Mishiro S. Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. *Virology* 1992; **190**: 894-899
- Bukh J, Purcell RH, Miller RH. Sequence analysis of the 5' noncoding region of hepatitis C virus. *Proc Natl Acad Sci USA* 1992; **89**: 4942-4946
- Simmonds P, Smith DB, McOmish F, Yap PL, Kolberg J, Urdea MS, Holmes EC. Identification of genotypes of hepatitis

- C virus by sequence comparisons in the core, E1 and NS-5 regions. *J Gen Virol* 1994; **75**: 1053-1061
- 6 **Smith DB**. Evolution of the hypervariable region of hepatitis C virus. *J Viral Hepat* 1999; **6** Suppl 1: 41-46
  - 7 **Penin F**, Combet C, Germanidis G, Frainais PO, Deleage G, Pawlotsky JM. Conservation of the conformation and positive charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment. *J Virol* 2001; **75**: 5703-5710
  - 8 **Robertson B**, Myers G, Howard C, Brettin T, Bukh J, Gaschen B, Gojobori T, Maertens G, Mizokami M, Nainan O, Netesov S, Nishioka K, Shin I T, Simmonds P, Smith D, Stuyver L, Weiner A. Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy. *Arch Virol* 1998; **143**: 2493-2503
  - 9 **Simmonds P**, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, Halfon P, Inchauspe G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky JM, Penin F, Sablon E, Shin I T, Stuyver L, Thiel HJ, Viazov S, Weiner AJ, Widell A. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* 2005; **42**: 962-973
  - 10 **Kuiken C**, Mizokami M, Deleage G, Yusim K, Penin F, Shin IT, Charavay C, Tao N, Crisan D, Grando D, Dalwani A, Geourjon C, Agrawal A, Combet C. Hepatitis C databases, principles and utility to researchers. *Hepatology* 2006; **43**: 1157-1165
  - 11 **Laperche S**, Lunel F, Izopet J, Alain S, Deny P, Duverlie G, Gaudy C, Pawlotsky JM, Plantier JC, Pozzetto B, Thibault V, Tosi F, Lefrere JJ. Comparison of hepatitis C virus NS5b and 5' noncoding gene sequencing methods in a multicenter study. *J Clin Microbiol* 2005; **43**: 733-739
  - 12 **Laperche S**, Saune K, Deny P, Duverlie G, Alain S, Chaix ML, Gaudy C, Lunel F, Pawlotsky JM, Payan C, Pozzetto B, Tamalet C, Thibault V, Vallet S, Bouchardeau F, Izopet J, Lefrere JJ. Unique NS5b hepatitis C virus gene sequence consensus database is essential for standardization of genotype determinations in multicenter epidemiological studies. *J Clin Microbiol* 2006; **44**: 614-616
  - 13 **Kalinina O**, Norder H, Mukomolov S, Magnus LO. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol* 2002; **76**: 4034-4043
  - 14 **Noppompanth S**, Lien TX, Poovorawan Y, Smits SL, Osterhaus AD, Haagmans BL. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. *J Virol* 2006; **80**: 7569-7577
  - 15 **Holmes EC**, Worobey M, Rambaut A. Phylogenetic evidence for recombination in dengue virus. *Mol Biol Evol* 1999; **16**: 405-409
  - 16 **Worobey M**, Holmes EC. Homologous recombination in GB virus C/hepatitis G virus. *Mol Biol Evol* 2001; **18**: 254-261
  - 17 **Simmonds P**. The origin and evolution of hepatitis viruses in humans. *J Gen Virol* 2001; **82**: 693-712
  - 18 **Fybus OG**, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. *Science* 2001; **292**: 2323-2325
  - 19 **Jeannel D**, Fretz C, Traore Y, Kohdjo N, Bigot A, Pe Gamy E, Jourdan G, Kourouma K, Maertens G, Fumoux F, Fournel JJ, Stuyver L. Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa. *J Med Virol* 1998; **55**: 92-97
  - 20 **Fretz C**, Jeannel D, Stuyver L, Herve V, Lunel F, Boudifa A, Mathiot C, de The G, Fournel JJ. HCV infection in a rural population of the Central African Republic (CAR): evidence for three additional subtypes of genotype 4. *J Med Virol* 1995; **47**: 435-437
  - 21 **Okada S**, Akahane Y, Suzuki H, Okamoto H, Mishiro S. The degree of variability in the amino terminal region of the E2/NS1 protein of hepatitis C virus correlates with responsiveness to interferon therapy in viremic patients. *Hepatology* 1992; **16**: 619-624
  - 22 **Enomoto N**, Kurosaki M, Tanaka Y, Marumo F, Sato C. Fluctuation of hepatitis C virus quasiespecies in persistent infection and interferon treatment revealed by single-strand conformation polymorphism analysis. *J Gen Virol* 1994; **75**: 1361-1369
  - 23 **Kakumu S**, Arao M, Yoshioka K, Hayashi H, Kusakabe A, Hirofujii H, Kawabe M. Recombinant human alpha-interferon therapy for chronic non-A, non-B hepatitis: second report. *Am J Gastroenterol* 1990; **85**: 655-659
  - 24 **Hoofnagle JH**, Mullen KD, Jones DB, Rustgi V, Di Bisceglie A, Peters M, Waggoner JG, Park Y, Jones EA. Treatment of chronic non-A, non-B hepatitis with recombinant human alpha interferon. A preliminary report. *N Engl J Med* 1986; **315**: 1575-1578
  - 25 **Knapp S**, Yee LJ, Prodscham AJ, Hennig BJ, Hellier S, Zhang L, Wright M, Chiaramonte M, Graves M, Thomas HC, Hill AV, Thursz MR. Polymorphisms in interferon-induced genes and the outcome of hepatitis C virus infection: roles of MxA, OAS-1 and PKR. *Genes Immun* 2003; **4**: 411-419
  - 26 **Conrad B**. Potential mechanisms of interferon-alpha induced autoimmunity. *Autoimmunity* 2003; **36**: 519-523
  - 27 **Manns MP**, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomized trial. *Lancet* 2001; **358**: 958-965
  - 28 **Fried MW**, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncalves FL Jr, Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; **347**: 975-982
  - 29 **Hadziyannis SJ**, Sette H Jr, Morgan TR, Balan V, Diago M, Marcellin P, Ramadori G, Bodenheimer H Jr, Bernstein D, Rizzetto M, Zeuzem S, Pockros PJ, Lin A, Ackrill AM. Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med* 2004; **140**: 346-355
  - 30 **Dalgard O**, Mangia A. Short-term therapy for patients with hepatitis C virus genotype 2 or 3 infection. *Drugs* 2006; **66**: 1807-1815
  - 31 **von Wagner M**, Huber M, Berg T, Hinrichsen H, Rasenack J, Heintges T, Bergk A, Bernsmeier C, Haussinger D, Herrmann E, Zeuzem S. Peginterferon-alpha-2a (40KD) and ribavirin for 16 or 24 weeks in patients with genotype 2 or 3 chronic hepatitis C. *Gastroenterology* 2005; **129**: 522-527
  - 32 **El-Zayadi AR**, Attia M, Barakat EM, Badran HM, Hamdy H, El-Tawil A, El-Nakeeb A, Selim O, Saied A. Response of hepatitis C genotype-4 naive patients to 24 weeks of Peginterferon-alpha2b/ribavirin or induction-dose interferon-alpha2b/ribavirin/amantadine: a non-randomized controlled study. *Am J Gastroenterol* 2005; **100**: 2447-2452
  - 33 **Kamal SM**, El Tawil AA, Nakano T, He Q, Rasenack J, Hakam SA, Saleh WA, Ismail A, Aziz AA, Madwar MA. Peginterferon (alpha)-2b and ribavirin therapy in chronic hepatitis C genotype 4: impact of treatment duration and viral kinetics on sustained virological response. *Gut* 2005; **54**: 858-866
  - 34 **Nguyen MH**, Koeffe EB. Prevalence and treatment of hepatitis C virus genotypes 4, 5, and 6. *Clin Gastroenterol Hepatol* 2005; **3**: S97-S101
  - 35 **Taylor DR**, Shi ST, Romano PR, Barber GN, Lai MM. Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science* 1999; **285**: 107-110
  - 36 **Abid K**, Quadri R, Negro F. Hepatitis C virus, the E2 envelope protein, and alpha-interferon resistance. *Science* 2000; **287**: 1555
  - 37 **Saito T**, Ito T, Ishiko H, Yonaha M, Morikawa K, Miyokawa A, Mitamura K. Sequence analysis of PePHD within HCV E2 region and correlation with resistance of interferon therapy in Japanese patients infected with HCV genotypes 2a and 2b. *Am J Gastroenterol* 2003; **98**: 1377-1383
  - 38 **Watanabe H**, Nagayama K, Enomoto N, Itakura J, Tanabe Y, Sato C, Izumi N, Watanabe M. Amino acid substitutions in PKR-eIF2 phosphorylation homology domain (PePHD) of hepatitis C virus E2 protein in genotype 2a/2b and 1b in Japan and interferon efficacy. *Hepatol Res* 2003; **26**: 268-274
  - 39 **Gandy C**, Lambele M, Moreau A, Veillon P, Lunel F, Goudeau A. Mutations within the hepatitis C virus genotype 1b E2-PePHD domain do not correlate with treatment outcome. *J Clin Microbiol* 2005; **43**: 750-754

- 40 **Polyak SJ**, Nousbaum JB, Larson AM, Cotler S, Carithers RL Jr, Gretch DR. The protein kinase-interacting domain in the hepatitis C virus envelope glycoprotein-2 gene is highly conserved in genotype 1-infected patients treated with interferon. *J Infect Dis* 2000; **182**: 397-404
- 41 **Moribe T**, Hayashi N, Kanazawa Y, Mita E, Fusamoto H, Negi M, Kaneshige T, Igimi H, Kamada T, Uchida K. Hepatitis C viral complexity detected by single-strand conformation polymorphism and response to interferon therapy. *Gastroenterology* 1995; **108**: 789-795
- 42 **Toyoda H**, Kumada N, Nakano S, Takeda I, Sugiyama K, Osada T, Kiriya S, Sone Y, Kinoshita M, Hadama T. Quasispecies nature of hepatitis C virus and response to alpha interferon: significance as a predictor of direct response to interferon. *J Hepatol* 1997; **26**: 6-13
- 43 **Lopez-Labrador FX**, Ampurdanes S, Gimenez-Barcons M, Guisera M, Costa J, Jimenez de Anta MT, Sanchez-Tapias JM, Rodes J, Saiz JC. Relationship of the genomic complexity of hepatitis C virus with liver disease severity and response to interferon in patients with chronic HCV genotype 1b infection (correction of interferon). *Hepatology* 1999; **29**: 897-903
- 44 **Macdonald A**, Harris M. Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol* 2004; **85**: 2485-2502
- 45 **Gale MJ Jr**, Korth MJ, Tang NM, Tan SL, Hopkins DA, Dever TE, Polyak SJ, Gretch DR, Katze MG. Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology* 1997; **230**: 217-227
- 46 **Polyak SJ**, Khabar KS, Paschal DM, Ezelle HJ, Duverlie G, Barber GN, Levy DE, Mukaida N, Gretch DR. Hepatitis C virus nonstructural 5A protein induces interleukin-8, leading to partial inhibition of the interferon-induced antiviral response. *J Virol* 2001; **75**: 6095-6106
- 47 **Polyak SJ**, Khabar KS, Rezeiq M, Gretch DR. Elevated levels of interleukin-8 in serum are associated with hepatitis C virus infection and resistance to interferon therapy. *J Virol* 2001; **75**: 6209-6211
- 48 **Enomoto N**, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Izumi N, Marumo F, Sato C. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest* 1995; **96**: 224-230
- 49 **Enomoto N**, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Ogura Y, Izumi N, Marumo F, Sato C. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* 1996; **334**: 77-81
- 50 **Fascu M**, Martus P, Hohne M, Wiedenmann B, Hopf U, Schreiber E, Berg T. Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. *Gut* 2004; **53**: 1345-1351
- 51 **Murakami T**, Enomoto N, Kurosaki M, Izumi N, Marumo F, Sato C. Mutations in nonstructural protein 5A gene and response to interferon in hepatitis C virus genotype 2 infection. *Hepatology* 1999; **30**: 1045-1053
- 52 **Bagaglio S**, Bruno R, Lodrini S, De Mitri MS, Andreone P, Loggi E, Galli L, Lazzarin A, Morsica G. Genetic heterogeneity of hepatitis C virus (HCV) in clinical strains of HIV positive and HIV negative patients chronically infected with HCV genotype 3a. *J Biol Regul Homeost Agents* 2003; **17**: 153-161
- 53 **Castelain S**, Khorsi H, Roussel J, Francois C, Jaillon O, Capron D, Penin F, Wychowski C, Meurs E, Duverlie G. Variability of the nonstructural 5A protein of hepatitis C virus type 3a isolates and relation to interferon sensitivity. *J Infect Dis* 2002; **185**: 573-583
- 54 **Inchauspe G**, Zebedee S, Lee DH, Sugitani M, Nasoff M, Prince AM. Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates. *Proc Natl Acad Sci USA* 1991; **88**: 10292-10296
- 55 **Duverlie G**, Khorsi H, Castelain S, Jaillon O, Izopet J, Lunel F, Eb F, Penin F, Wychowski C. Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity. *J Gen Virol* 1998; **79**: 1373-1381
- 56 **Veillon P**, Payan C, Le Guillou-Guillemette H, Gaudy C, Lunel F, The Fontevraud Study Group. Quasispecies in NS5A region of HCV. *World J Gastroenterol* 2007; **13**: 1195-1203
- 57 **Puig-Basagoiti F**, Forns X, Furci I, Ampurdanes S, Gimenez-Barcons M, Franco S, Sanchez-Tapias JM, Saiz JC. Dynamics of hepatitis C virus NS5A quasispecies during interferon and ribavirin therapy in responder and non-responder patients with genotype 1b chronic hepatitis C. *J Gen Virol* 2005; **86**: 1067-1075
- 58 **Houghton M**. Hepatitis C virus. In: Howley DMKPM. *Fields Virology*. Philadelphia: Lippincott-Raven, 1996: 1035-1058
- 59 **Farci P**, Purcell RH. Clinical significance of hepatitis C virus genotypes and quasispecies. *Semin Liver Dis* 2000; **20**: 103-126
- 60 **Bartenschlager R**. The NS3/4A proteinase of the hepatitis C virus: unravelling structure and function of an unusual enzyme and a prime target for antiviral therapy. *J Viral Hepat* 1999; **6**: 165-181
- 61 **Vallet S**, Gouriou S, Nousbaum JB, Legrand-Quillien MC, Goudeau A, Picard B. Genetic heterogeneity of the NS3 protease gene in hepatitis C virus genotype 1 from untreated infected patients. *J Med Virol* 2005; **75**: 528-537
- 62 **Vallet S**, Gouriou S, Nkontchou G, Hotta H, Vilerio M, Legrand-Quillien MC, Beaugrand M, Trinchet JC, Nousbaum JB, Deny P, Gaudy C, Goudeau A, Picard B, Payan C. Is hepatitis C virus NS3 protease quasispecies heterogeneity predictive of progression from cirrhosis to hepatocellular carcinoma? *J Viral Hepat* 2007; **14**: 96-106
- 63 **Soderholm J**, Ahlen G, Kaul A, Frelin L, Alheim M, Barnfield C, Liljestrom P, Weiland O, Milich DR, Bartenschlager R, Sallberg M. Relation between viral fitness and immune escape within the hepatitis C virus protease. *Gut* 2006; **55**: 266-274
- 64 **Gale MJ Jr**. Effector genes of interferon action against hepatitis C virus. *Hepatology* 2003; **37**: 975-978
- 65 **Meylan E**, Curran J, Hofmann K, Moradpour D, Binder M, Bartenschlager R, Tschopp J. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* 2005; **437**: 1167-1172
- 66 **Kieffer T**, Sarrazin C, Bartels D, Hanzelka B, Mth U, Welker M, Wincheringer D, Lin C, Purdy S, Weegink C, Reesink H, Kwong A, Zeuzem S. Wild-type HCV NS3 protease re-emerges during follow-up after 14 days of dosing with VX-950 in patients with genotype 1 HCV. In: **Colombo M**. *Journal of Hepatology*. EASL 2006: Proceedings of the 41th Annual Meeting of the European Association for the Study of the Liver; 2006 Apr 26-30; Vienna, Austria. Shannon, Ireland: Elsevier, 2006: **44**: S7
- 67 **Zeuzem S**, Sarrazin C, Rouzier R, Tarral A, Brion N, Forestier N, Gupta S, Deckman D, Fellows K, Hussain M, Cutler D, Zhang J. Anti-viral activity of SCH503034, a HCV protease inhibitor, administered as monotherapy in hepatitis C genotype-1 (HCV-1) patients refractory to pegylated interferon (peg-IFN- $\alpha$ ). In: **Blei AT**. *Hepatology*. **AASLD 2005: Proceedings of the 56th Annual Meeting of the American Association for the Study of Liver Diseases**; 11-15 November 2005; San Francisco, USA. Hoboken: John Wiley & Sons, 2005: **42**: 233A
- 68 **Clavel F**, Hance AJ. HIV drug resistance. *N Engl J Med* 2004; **350**: 1023-1035
- 69 **Lin C**, Lin K, Luong YP, Rao BG, Wei YY, Brennan DL, Fulghum JR, Hsiao HM, Ma S, Maxwell JP, Cottrell KM, Perni RB, Gates CA, Kwong AD. In vitro resistance studies of hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061: structural analysis indicates different resistance mechanisms. *J Biol Chem* 2004; **279**: 17508-17514
- 70 **Sarrazin C**, Kieffer T, Bartels D, Hanzelka B, Mth U, Welker M, Wincheringer D, Lin C, Grossman T, Purdy S, Weegink C, Reesink H, Zeuzem S, Kwong A. Characterization of viral variants in the HCV NS3 protease domain of genotype 1 patients that are selected during 14 days of dosing with VX-950. In: **Blei AT**. *Hepatology*. **AASLD 2005: Proceedings of the 56th Annual Meeting of the American Association for the Study of Liver Diseases**; 2005 Nov 11-15; San Francisco, USA. Hoboken: John Wiley & Sons, 2005: **42**: 751A
- 71 **Franco S**, Clotet B, Martinez M. High genotypic and phe-

- notypic resolution of HCV NS3/4 protease quasispecies: mutations associated with drug resistance in treatment-naïve patients. In: **Lange J RD. Antiviral Therapy. HIV Drug Resistance Workshop 2006: Proceedings of the 15th International HIV Drug resistance Workshop; 2006 Jun 13-17; Sitges, Spain. London: International Medical Press, 2006: 11: S14**
- 72 **Winters MA, Welles SL, Holodniy M.** Hepatitis C virus protease gene diversity in patients coinfecting with human immunodeficiency virus. *J Virol* 2006; **80**: 4196-4199
- 73 **Tong X, Chase R, Skelton A, Chen T, Wright-Minogue J, Malcolm BA.** Identification and analysis of fitness of resistance mutations against the HCV protease inhibitor SCH 503034. *Antiviral Res* 2006; **70**: 28-38
- 74 **Reesink H, Zeuzem S, Weegink C, Forestier N, van Vliet A, van de Wetering de Rooij J, McNair L, Purdy S, Chu H, Jansen P.** Final results of a phase 1b, multiple-dose study of VX-950, a hepatitis C virus protease inhibitor. In: **Blei AT. Hepatology. AASLD 2005: Proceedings of the 56th Annual Meeting of the American Association for the Study of Liver Diseases; 2005 Nov 11-15; San Francisco, USA. Hoboken: John Wiley & Sons, 2005: 42: 234A**
- 75 **Yi M, Tong X, Skelton A, Chase R, Chen T, Prongay A, Bogen SL, Saksena AK, Njoroge PG, Veselenak RL, Pyles RB, Bourne N, Malcolm BA, Lemon SM.** Mutations conferring resistance to SCH6, a novel hepatitis C virus NS3/4A protease inhibitor. Reduced RNA replication fitness and partial rescue by second-site mutations. *J Biol Chem* 2006; **281**: 8205-8215
- 76 **Leeburg CA, Cable MB, Ferrari E, Hong Z, Mannarino AF, Weber PC.** Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct Biol* 1999; **6**: 937-943
- 77 **Bressanelli S, Tomei L, Roussel A, Incitti I, Vitale RL, Mathieu M, De Francesco R, Rey FA.** Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc Natl Acad Sci USA* 1999; **96**: 13034-13039
- 78 **Lohmann V, Korner F, Herian U, 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**: 8416-8428
- 79 **Lohmann V, Roos A, Korner F, Koch JO, Bartenschlager R.** Biochemical and kinetic analyses of NS5B RNA-dependent RNA polymerase of the hepatitis C virus. *Virology* 1998; **249**: 108-118
- 80 **Cheney IW, Naim S, Lai VC, Dempsey S, Bellows D, Walker MP, Shim JH, Horscroft N, Hong Z, Zhong W.** Mutations in NS5B polymerase of hepatitis C virus: impacts on in vitro enzymatic activity and viral RNA replication in the subgenomic replicon cell culture. *Virology* 2002; **297**: 298-306
- 81 **Labonte P, Axelrod V, Agarwal A, Aulabaugh A, Amin A, Mak P.** Modulation of hepatitis C virus RNA-dependent RNA polymerase activity by structure-based site-directed mutagenesis. *J Biol Chem* 2002; **277**: 38838-38846
- 82 **Vuillermoz I, Khattab E, Sablon E, Ottevaere I, Durantel D, Vieux C, Trepo C, Zoulim F.** Genetic variability of hepatitis C virus in chronically infected patients with viral breakthrough during interferon-ribavirin therapy. *J Med Virol* 2004; **74**: 41-53
- 83 **Kumagai N, Takahashi N, Kinoshita M, Tsunematsu S, Tsuchimoto K, Saito H, Ishii H.** Polymorphisms of NS5B protein relates to early clearance of hepatitis C virus by interferon plus ribavirin: a pilot study. *J Viral Hepat* 2004; **11**: 225-235
- 84 **Carroll SS, Tomassini JE, Bosserman M, Getty K, Stahlhut MW, Eldrup AB, Bhat B, Hall D, Simcoe AL, LaFemina R, Rutkowski CA, Wolanski B, Yang Z, Migliaccio G, De Francesco R, Kuo LC, MacCoss M, Olsen DB.** Inhibition of hepatitis C virus RNA replication by 2'-modified nucleoside analogs. *J Biol Chem* 2003; **278**: 11979-11984
- 85 **Carroll SS, Olsen DB.** Nucleoside analog inhibitors of hepatitis C virus replication. *Infect Disord Drug Targets* 2006; **6**: 17-29
- 86 **Migliaccio G, Tomassini JE, Carroll SS, Tomei L, Altamura S, Bhat B, Bartholomew L, Bosserman MR, Ceccacci A, Colwell LF, Cortese R, De Francesco R, Eldrup AB, Getty KL, Hou XS, LaFemina RL, Ludmerer SW, MacCoss M, McMasters DR, Stahlhut MW, Olsen DB, Hazuda DJ, Flores OA.** Characterization of resistance to non-obligate chain-terminating ribonucleoside analogs that inhibit hepatitis C virus replication in vitro. *J Biol Chem* 2003; **278**: 49164-49170
- 87 **Ludmerer SW, Graham DJ, Boots E, Murray EM, Simcoe A, Markel EJ, Grobler JA, Flores OA, Olsen DB, Hazuda DJ, LaFemina RL.** Replication fitness and NS5B drug sensitivity of diverse hepatitis C virus isolates characterized by using a transient replication assay. *Antimicrob Agents Chemother* 2005; **49**: 2059-2069
- 88 **Klumpp K, Leveque V, Le Pogam S, Ma H, Jiang WR, Kang H, Granycome C, Singer M, Laxton C, Hang JQ, Sarma K, Smith DB, Heindl D, Hobbs CJ, Merrett JH, Symons J, Cammack N, Martin JA, Devos R, 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**: 3793-3799
- 89 **Le Pogam S, Jiang WR, Leveque V, Rajyaguru S, Ma H, Kang H, Jiang S, Singer M, Ali S, Klumpp K, Smith D, Symons J, Cammack N, Najera I.** In vitro selected Con1 subgenomic replicons resistant to 2'-C-methyl-cytidine or to R1479 show lack of cross resistance. *Virology* 2006; **351**: 349-359
- 90 **Roberts S, Cooksley G, Shaw D, Berns H, Brandl M, Fettner S, Hill G, Ipe D, Klumpp K, Mannino M, O'Mara E, Tu Y, Washington C.** Interim results of a multiple ascending dose study of R1626, a novel nucleoside analog targeting HCV polymerase in chronic HCV patients. In: **Colombo M. Journal of Hepatology. EASL 2006: Proceedings of the 41th European Association for the Study of the Liver; 2006 Apr 26-30; Vienna, Austria. Shannon: Elsevier, 2006: 44: S269**
- 91 **Chan L, Pereira O, Reddy TJ, Das SK, Poisson C, Courchesne M, Proulx M, Siddiqui A, Yannopoulos CG, Nguyen-Ba N, Roy C, Nasturica D, Moinet C, Bethell R, Hamel M, L'Heureux L, David M, Nicolas O, Courtemanche-Asselin P, Brunette S, Bilimoria D, Bedard J.** Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 2: tertiary amides. *Bioorg Med Chem Lett* 2004; **14**: 797-800
- 92 **Mo H, Lu L, Pilot-Matias T, Pithawalla R, Mondal R, Masse S, Dekhtyar T, Ng T, Koev G, Stoll V, Stewart KD, Pratt J, Donner P, Rockway T, Maring C, Molla A.** Mutations conferring resistance to a hepatitis C virus (HCV) RNA-dependent RNA polymerase inhibitor alone or in combination with an HCV serine protease inhibitor in vitro. *Antimicrob Agents Chemother* 2005; **49**: 4305-4314
- 93 **Howe AY, Bloom J, Baldick CJ, Benetatos CA, Cheng H, Christensen JS, Chunduru SK, Coburn GA, Feld B, Gopalsamy A, Gorczyca WP, Herrmann S, Johann S, Jiang X, Kimberland ML, Krisnamurthy G, Olson M, Orłowski M, Swanberg S, Thompson I, Thorn M, Del Vecchio A, Young DC, van Zeijl M, Ellingboe JW, Upeslacijs J, Collett M, Mansour TS, O'Connell JF.** Novel nonnucleoside inhibitor of hepatitis C virus RNA-dependent RNA polymerase. *Antimicrob Agents Chemother* 2004; **48**: 4813-4821
- 94 **Kukolj G, McGibbon GA, McKercher G, Marquis M, Lefebvre S, Thauvette L, Gauthier J, Goulet S, Poupart MA, Beaulieu PL.** Binding site characterization and resistance to a class of non-nucleoside inhibitors of the hepatitis C virus NS5B polymerase. *J Biol Chem* 2005; **280**: 39260-39267
- 95 **Tomei L, Altamura S, Paonessa G, De Francesco R, Migliaccio G.** HCV antiviral resistance: the impact of in vitro studies on the development of antiviral agents targeting the viral NS5B polymerase. *Antiviral Chem Chemother* 2005; **16**: 225-245
- 96 **Graham DJ, Stahlhut M, Flores O, Olsen DB, Hazuda DJ, LaFemina RL, Ludmerer SW.** A genotype 2b NS5B polymerase with novel substitutions supports replication of a chimeric HCV 1b/2b replicon containing a genotype 1b NS3-5A background. *Antiviral Res* 2006; **69**: 24-30
- 97 **De Francesco R, Migliaccio G.** Challenges and successes in developing new therapies for hepatitis C. *Nature* 2005; **436**: 953-960

S- Editor Liu Y E- Editor Ma WH

**2. Duplication of the V3 domain in the genotype 1b hepatitis C virus NS5A protein: clonal analysis and physicochemical properties related to hepatocellular carcinoma occurrence.**

Petsaris O. Vallet S. Le Guillou-Guillemette H., Veillon P., S. Gouriou S., Barbier G., Nousbaum JB., Saliou P., NKontchou G., Trinchet JC., Lunel-Fabiani F., Payan C.

*Manuscrit soumis à J Viral Hepatitis.*

**Duplication of the V3 domain in the genotype 1b hepatitis C virus NS5A protein: clonal analysis and physicochemical properties related to hepatocellular carcinoma occurrence**

Petsaris O.<sup>1,2</sup>, Vallet S.<sup>1,2</sup>, Le Guillou-Guillemette H.<sup>3</sup>, Veillon P.<sup>3</sup>, S. Gouriou S.<sup>2</sup>, Barbier G.<sup>4</sup>, Nousbaum JB.<sup>5</sup>, Saliou P.<sup>6</sup>, NKontchou G.<sup>7</sup>, Trinchet JC.<sup>7,8</sup>, Lunel-Fabiani F.<sup>3</sup>, Payan C.<sup>1,2</sup>

<sup>1</sup> Centre Hospitalier Universitaire La Cavale Blanche, Laboratoire de Microbiologie, 29200 Brest cedex, France

<sup>2</sup> Université de Brest, Université Européenne de Bretagne, SFR ScInBioS, LUBEM (EA3882), UFR Médecine et Sciences de la Santé, 29200 Brest, France

<sup>3</sup> Centre Hospitalier Universitaire d'Angers, Laboratoire HIFI UPRES EA3859, 49000 Angers, France

<sup>4</sup> Université de Brest, Université Européenne de Bretagne, SFR ScInBioS, LUBEM (EA3882), ESIAB, 29280 Plouzané, France

<sup>5</sup> Centre Hospitalier Universitaire La Cavale Blanche, Service d'Hépatogastroentérologie, 29200 Brest, France

<sup>6</sup> Université de Brest, Université Européenne de Bretagne, Laboratoire de Santé Publique, Epidémiologie, UFR Médecine et Sciences de la Santé, 29200 Brest, France

<sup>7</sup> Hôpital Jean Verdier, Service d'Hépatogastroentérologie, Assistance Publique-Hôpitaux de Paris, UFR SMBH-Université Paris 13, 93143 Bondy cedex, France

<sup>8</sup> Centre de Ressources Biologiques, Hôpital Jean Verdier, Assistance Publique-Hôpitaux de Paris, 93143 Bondy cedex, France

Corresponding author: Prof. Christopher Payan, UFR Médecine et Sciences de la Santé, 22 avenue Camille Desmoulins, 29238 Brest cedex, France. Phone: 00-33-2-90-91-51-11; Fax: 00-33-2-90-91-51-01; E-mail: christopher.payan@univ-brest.fr

## Abstract

Hepatitis C virus non-structural protein 5A is known to play a role in development of hepatocellular carcinoma via interactions with host cell pathways. Hepatitis C virus genotype 1b strains presenting a wide insertion of 27 amino acids in the non-structural protein 5A V3 domain were studied in order to determine whether this V3 insertion, identified as a duplication, was associated with hepatocellular (HCC) carcinoma occurrence. Seventy-four sera were screened for the presence of this V3-like additional domain (V3-1). Three strains harboring the inserted domain were detected among patients who displayed progression to HCC (n=28), two strains among patients with liver cirrhosis (LC, n=27) and none among patients with chronic hepatitis (CH, n=19). Phylogenetic trees built from V3 and V3-1 sequences indicated that the latter clustered separately, and that these strains were probably transmitted after having already acquired the duplication. Between-sample analysis showed that HCC V3 and HCC V3-1 genetic variability was statistically higher than that of LC V3 and LC V3-1. Among HCC group sequences, the additional V3-1 domain and its four amino acid junctions with V3 domain exhibited significantly higher Shannon entropy values and positively selected amino acids. We further showed that, among HCC group sequences, these amino acid positions were more hydrophilic, and thus more strongly exposed, and were enriched in disorder-promoting highly flexible residues. Taken together, our results suggest that V3-1 evolution may differ in strains associated with HCC occurrence. The presence of an intrinsically “disordered” V3 duplicate may alter the NS5A protein network. Further investigations are necessary to elucidate the potential impact of V3 duplication in the context of carcinogenesis.

## 1 **Introduction**

2           According to WHO, global hepatitis C virus (HCV) infection prevalence is estimated  
3 at 3% of the world population, representing 170 million people. Ten to thirty percent of the  
4 latter will become cirrhotic, with a 3% to 8% per year risk of hepatocellular carcinoma (HCC)  
5 development [1]. In developed countries (the USA, Europe and Japan), about 60% of HCC  
6 cases are attributed to chronic hepatitis C infection [2]. Over the last two decades, the HCC  
7 incidence has been growing in these countries, and will probably continue due to a cohort  
8 effect of greater exposure to HCV from the 1950s to the 1980s via contact with contaminated  
9 blood [3].

10           Four HCV proteins are known to play an important role in hepatic oncogenesis: the  
11 core protein [14,15] and the non-structural proteins NS3 [16,17], NS5A [15] and NS5B [18].  
12 NS5A has recently attracted considerable interest because of its potential multiple functions.  
13 It may be a promising new target for anti-HCV therapy [19,20,21] and a host cell growth and  
14 apoptosis modulator via interactions with p21, p53 and cyclins [22,23]; it may be associated  
15 with the response to IFN- $\alpha$  therapy and possibly implicated in formation of a functional HCV  
16 replication complex. NS5A is a large (56-58 kDa, 447 amino acids) RNA binding hydrophilic  
17 phosphoprotein organized into three domains (Domains I, II and III) separated by low-  
18 complexity sequences [24,25,26,27]. The 30 amino-terminal amino acid residues of NS5A  
19 (Domain I) form an amphipathic  $\alpha$ -helix anchoring NS5A to the membrane-associated  
20 replication complex [28]. Domain II of NS5A (250-342) contains the interferon sensitivity-  
21 determining region (ISDR; amino acids 237-276) within the protein kinase R (PKR) binding  
22 domain (amino acids 237-309). Mutations in NS5A ISDR have been shown to correlate with  
23 IFN treatment responsiveness in genotype-1b-infected Japanese patients [29,30]. The protein  
24 kinase R binding domain is able to inhibit IFN-induced PKR involved in both the cellular  
25 antiviral response and in cell proliferation control and oncogenesis [31]. NS5A carboxy  
26 terminal domain III (356-447) is characterized by a high degree of genetic flexibility.  
27 Sequence variations in the hypervariable region V3 (383-406) or in the V3 and pre-region V3,  
28 called the interferon/ribavirin resistance-determining region (IRRDR), may correlate with an  
29 early and sustained virological response to IFN/RBV combination therapy [32,33,34]. The  
30 function of NS5A domain III in the viral replication cycle is still unknown. The NS5A C-  
31 terminus can tolerate large heterologous sequence insertions without impairing HCV  
32 replication [35]. On the other hand, domain III of NS5A may play a key role in hepatitis C  
33 virus infectious particle assembly, modulating virion production [24,27].



34 In a previous exploratory French study on HCV genotype 1b NS5A ISDR and V3  
35 domain mutations prior to IFN therapy, Veillon *et al.* detected strains harboring a never-  
36 before-described wide insertion in V3, mimicking a duplication of this domain, as two  
37 paralogs [32]. A multicenter study supported by the French National Agency for Research on  
38 AIDS and Viral Hepatitis (ANRS) was recently conducted in France (twenty laboratories) and  
39 showed a prevalence of 3.05% for HCV 1b strains, with a duplicated NS5A V3 domain that  
40 may correlate with severity of the liver disease, i.e. fibrosis level and HCC occurrence (Le  
41 Guillou *et al.*, ).

42 The aim of the present study was to describe V3 and V3-like (V3-l) domain genetic  
43 variability, to analyze selective pressure acting on the two paralogs and to estimate physico-  
44 chemical profiles of this duplication, while comparing strains associated with liver cirrhosis  
45 (LC) or with HCC.

46

## 47 **Materials and methods**

### 48 **Patients**

49 Patient sera were collected from the French Biological Resources Center of Jean  
50 Verdier Hospital (Bondy, France). Written consent was obtained for each patient. Seventy-  
51 four women with HCV-1b genotype-related liver disease were prospectively recruited  
52 according to the following criteria: Caucasian origin and absence of co-morbidities such as  
53 alcohol abuse, HIV or HBV co-infection. All were HCV-treatment-naive, and cirrhotic  
54 patients were followed up for eight years with periodic ultrasound examination and serum  $\alpha$ -  
55 fetoprotein monitoring, as previously reported [36]. Patients were classified into three groups  
56 according to the different degrees of liver disease severity at the time of serum collection:  
57 group CH included 19 patients with chronic hepatitis (F1-F2, METAVIR classification),  
58 group LC included 27 patients with liver cirrhosis (F4) without subsequent development of  
59 HCC, and group HCC included 28 patients with liver cirrhosis who subsequently developed  
60 HCC during follow-up. Patients were contaminated in Europe, in most cases via blood  
61 transfusion, and were matched so as to have the same presumed duration of hepatitis C  
62 infection. Viral evolution was also analyzed in two patients from whom sequential serum  
63 samples were available seven years apart: at the cirrhosis stage in one case (group LC), and  
64 before and after HCC in the other case (group HCC). Serum samples were stored at  $-70^{\circ}\text{C}$   
65 until use.

66

## 67 **HCV-RNA quantification and HCV genotyping**

68 Quantitative detection of HCV-RNA was performed using the RT-PCR-based method  
69 (Roche Diagnostics Cobas Amplicor HCV Monitor test v. 2.0, Meylan, France), with a cutoff  
70 level at 600 IU/mL. Genotyping of HCV was carried out by 5' non-coding region sequencing  
71 using the Trugene HCV method (TRUGENE 5'NC, Bayer HealthCare LLC, Berkeley, CA).  
72 In case of undetermined subtype, the NS5B region was sequenced to confirm genotype 1b  
73 according to French ANRS consensus protocol [37].

74

## 75 **HCV-RNA extraction and NS5A domain V3 amplification by RT-PCR**

76 Total HCV-RNA was extracted from 140 µL of serum using the QIAamp® RNA  
77 mini-kit (Qiagen, Courtaboeuf, France). The full-length NS5A gene (6246-7586, HCV-J  
78 strain numbering) was first amplified by reverse transcription, combined with an initial PCR  
79 using outer primers E1 and E2, previously described by Duverlie *et al.* [38]. Reverse  
80 transcription was performed using Superscript III reverse transcriptase included in the One-  
81 Step RT-PCR kit (Invitrogen, Life Technologies, Saint Aubin, France) and reverse primer E2  
82 (5'-GCTGCGAGATGTTGTGGCGTA-3', nucleotides 7968-7718, according to HCV-J  
83 numbering). Reverse transcription was combined with full-length NS5A amplification using  
84 the E2 primer described above and the E1 primer (5'-GAGGGGGCTGTGCAGTGGATG-3',  
85 nucleotides 6057-6077, HCV-J numbering), respectively, as non-sense and sense primers.  
86 Briefly, 25 µL of the reaction mix contained 2X reaction buffer, 200 µM dNTP mix, 1.6 mM  
87 magnesium sulfate, 0.5 µL of Superscript III reverse transcriptase/platinum Taq polymerase,  
88 200 nM of each sense and antisense primer and 5 µL RNA. Reverse transcription and the first  
89 PCR consisted of 1 cycle at 55°C for 30 min (for reverse transcription), followed by 1 cycle at  
90 94°C for 2 min, 40 cycles at 94°C for 30s, 55°C for 60s, 68°C for 3 min, with a final  
91 extension at 68°C for 5 min. To obtain NS5A V3 domain amplification, a nested PCR was  
92 performed with 5 µL of NS5A first PCR product as template, using internal sense primers  
93 S31 (5'-GGACTACGTCCCTCCGGTGG-3', nucleotides 7241-7260, HCV-J numbering) or  
94 S (5'-TTGCTACAAAGACCTTTGGCAG-3', nucleotides 7368-7568, HCV-J numbering), if  
95 failure to amplify with S31, and antisense I4 (5'-GCAGCAGACGACATCCTCAC-3',  
96 nucleotides 7567-7586, HCV-J numbering). Twenty-five µL of the reaction mix contained  
97 10X reaction buffer, 200 µM dNTP mix, 2 mM magnesium sulfate, 0.5µL AmpliTaq Gold  
98 DNA polymerase (Applied Biosystems, Life Technologies, Courtaboeuf, France), 200 nM of  
99 each primer and 5 µL of NS5A PCR product. The nested PCR consisted of 1 cycle at 94°C for  
100 5 min, 40 cycles at 94°C for 45s, 55°C for 45 s and 72°C for 3 min with a final extension at

101 72°C for 15 min. Fragment size was verified by ethidium bromide staining after  
102 electrophoresis in a 1% agarose gel (Eurobio, Les Ulis, France). The expected size of NS5A  
103 V3 PCR products was 345 nucleotides (using S31/I4 primers) and 218 nucleotides (with S3/I4  
104 primers). Fragments longer than expected were likely to carry the duplicated V3 domain and  
105 were studied.

106

### 107 **NS5A V3 cloning and DNA sequencing**

108         Amplified NS5A V3 domains likely to be duplicated were purified (High Pure PCR  
109 Product Purification kit; Roche Molecular Biochemicals, Meylan, France) and ligated into the  
110 pCR 2.1<sup>TM</sup>-TOPO vector (TOPO TA Cloning Kit; Invitrogen) following the manufacturer's  
111 instructions. Transformants were grown on LB agar plates containing 50 µg/mL of ampicillin  
112 and selected at random. Plasmids were extracted using the QIAprep Spin Miniprep kit  
113 (Qiagen, Courtaboeuf, France) and subjected to nested NS5A V3 PCR as described above.  
114 Amplicon size was verified by electrophoresis in 1% agarose gel. Plasmids containing an  
115 insert in the V3 domain were purified and sequenced with M13 universal and M13 reverse  
116 primers by the dideoxy chain termination method on a MegaBace DNA analysis system  
117 (Amersham Biosciences, Orsay, France). Amplified DNA products were also directly  
118 sequenced in both directions using S and I4 primers with the Big Dye Terminator sequencing  
119 kit (Applied Biosystems, Life Technologies, Courtaboeuf, France) on an ABI PRISM 310  
120 sequencer. Chromatograms were read by Chromas Lite software (<http://technelysium.com.au>)  
121 and analyzed by the Staden Package (Gap4) program (<http://staden.sourceforge.net>).

122

### 123 **NS5AV3 quasi-species analysis**

124         Major sequence alignment was manually edited using the GeneDoc 2.0.1 editor.  
125 Quasi-species sequences were aligned with the ClustalW program included in MEGA version  
126 4 software [39]. In order to clear evolutionary ties between the two genes, rooted phylogenetic  
127 trees were built from all V3 nucleotide sequences aligned with those of V3-l. Trees were  
128 constructed using the neighbor-joining method, a Kimura two-parameter model and the HCV-  
129 J strain as outgroup. Genetic variability was measured by within- and between-sample  
130 analyses. The within-sample approach consisted of examining each patient sequence  
131 independently and then the mean of the values obtained for each patient so as to obtain the  
132 value of the group. Between-sample analysis was based on pairwise comparison of all  
133 sequences of a clinical group except for genetic distances and dN/dS ratio calculations, where  
134 each sequence was directly compared to the HCV-J strain.

135 To compare V3 and V3-like diversity according to clinical outcome, genetic distances  
136 were estimated within these two regions using the Kimura two-parameter method  
137 implemented in MEGA4 software. We next analyzed sequence complexity, defined as  
138 Shannon entropy values, at nucleotide and amino acids level, using the Shannon Entropy  
139 Two-Tool available on the HCV LANL database (<http://hcv.lanl.gov>). This interface enables  
140 graphic representation of Shannon entropy, comparing site-by-site two opposite sets of  
141 sequences (positive and negative scale). Positional entropy variations were analyzed by  
142 comparing the cancer group (represented on a positive scale) with the cirrhosis group  
143 (negative scale).

144 The number of synonymous (dS) and non-synonymous (dN) substitutions per site was  
145 determined using the Nei-Gobojobori method with Jukes-Cantor correction in MEGA4  
146 software. The dN/dS ratio at each codon site was estimated using the SNAP tool  
147 (synonymous non-synonymous analysis program) available on the HCV LANL database.

148

#### 149 **Hydropathy profiles and physicochemical properties**

150 To delineate the hydrophobic profile of the protein in each clinical group, we used the  
151 Kyte-Doolittle hydropathy scale, implemented in DAMBE software [40], with a window size  
152 of 6, allowing detection of hydrophilic and possibly antigenic regions. To determine whether  
153 quasi-species variations between HCC and LC groups resulted in physicochemical changes  
154 (flexibility, surface exposure, presence of disorder- or order-promoting amino acids), a  
155 graphic comparison of positional amino acid properties was built using the Two-Sample  
156 Logos web-tool [41] available at <http://www.twosamplelogo.org/>. The Two-Sample Logo  
157 enables visual identification of physicochemical properties when comparing HCC (positive  
158 set) and LC group sequences (negative set) with the height of each amino acid symbol within  
159 the logo proportional to its frequency in the alignment. Only the HCC group was represented  
160 on the logo, with the upper section displaying overrepresented amino acids in the positive set  
161 and the lower section showing depleted amino acids (figure 5).

162

#### 163 **Pattern/motif search**

164 The number and nature of functional motifs potentially resulting from V3 duplication of HCC  
165 or LC group V3 major sequences were compared using the PROSITE program of the ExPASy  
166 Bioinformatics Resource Portal (<http://www.expasy.org/>).

167

168

## 169 **Statistical analysis**

170 Patient characteristics were compared between groups using analysis of variance or  
171 the Wilcoxon-Mann-Whitney test. The difference in V3 duplication prevalence was tested  
172 with Fischer's exact test. Differences between groups of sequences were assessed, when  
173 appropriate, by Wilcoxon-Mann-Whitney or Student's t-tests for quantitative variables.  $P <$   
174  $0.05$  was considered significant.

## 176 **Nucleotide sequence GenBank accession numbers**

177 A total of 126 different clonal sequences of the NS5A V3 gene have been deposited in  
178 the GenBank database (accession numbers KF648380-KF648505).

## 180 **Results**

### 181 **Characteristics of the patients**

182 Seventy-four patients with HCV genotype 1b chronic infection were divided into three  
183 clinical groups (HCC, LC and CH) and screened for the presence of V3 duplication. Patient  
184 general characteristics are presented in Table 1. The average age in each group was  
185 statistically different (respectively,  $56.07 \pm 13.0$  years,  $63.19 \pm 11.4$  y and  $71.09 \pm 8.6$  y in  
186 groups CH, LC and HCC, analysis of variance,  $p < 0.0001$ ). However, presumed durations of  
187 hepatitis C infection were similar and reached 46.5 years in the LC group and 46.75 in the  
188 HCC group. Group CH consisted of 26.3% of patients with an F1 METAVIR score and  
189 73.7% with an F2 score, while groups LC and HCC included only patients with an F4 score.  
190 No significant difference was observed in average viral load (respectively,  $2.98 \times 10^6$ ;  $2.78 \times$   
191  $10^6$  and  $4.10 \times 10^6$  UI/mL for HCC, LC and CH groups), nor in average serum alanine  
192 aminotransferase levels (respectively, 2.2, 2.5 and 2.2 times the upper limit of normal in the  
193 HCC, LC and CH groups). The average level of alpha-fetoprotein (AFP) was significantly  
194 higher ( $p = 0.0389$  in the Wilcoxon-Mann-Whitney test) among patients who developed HCC  
195 than in the other groups (mean AFP within normal limits).

### 197 **Presence and direct sequencing of V3 duplication**

198 Three insertions were detected in the HCC group ( $n=28$ ), two in the LC group ( $n=27$ )  
199 and none in the CH group ( $n=19$ ). Although the insertion prevalence was higher among  
200 patients with HCC (10.71%) than among those with LC (7.41%) or CH (0%), the difference  
201 was not statistically significant (Fischer's exact test,  $p=1$ ). For the two patients for whom

202 sequential serum samples were available, V3 duplication was also detected seven years later.  
203 In all cases of insertion in the V3 region, an additional domain of 31 amino acids was present,  
204 analyzed first by direct sequencing, as shown in Fig.1. The NS5A V3 region of HCV  
205 genotype 1b was located between amino acids 2353 and 2379. Here, 27 additional amino  
206 acids were present without changing the reading frame, joined to the initial V3 domain by a  
207 four-amino-acids junction. The percentage of nucleotide identity between these two domains  
208 (excluding the four amino acids forming junction) ranged from 0.71 to 0.79 (average 0.76) for  
209 the HCC group and from 0.79 to 0.85 (average 0.81) for the LC group (Table 2). Because of  
210 this high identity percentage, the additional domain was considered to be a duplication of V3.

211

### 212 **Quasispecies composition**

213 Seven HCV-1b isolates were studied in order to accurately characterize the  
214 quasispecies of NS5A V3 and V3-I domains. Four isolates were collected from 3 patients with  
215 HCC, 3 isolates from 2 patients with LC and, for one patient from each group, a second serum  
216 seven years later. The average number of clones per isolate was 28 in the HCC group and 22  
217 in the LC group, respectively, for a total clone number of 178. Finally, 167 sequences of 174  
218 nucleotides corresponding to V3 duplicated sequences were obtained. Eleven clones yielding  
219 non-duplicated V3 sequences (94 nucleotides in length) were excluded from further analysis.

220 When examining an alignment of the 167 clonal sequences of the HCV-1b V3 domain  
221 obtained in this study combined with the HCV-J 1b prototype and those previously published  
222 (accession numbers: AY808004-AY808042 in the GenBank database), we observed that non-  
223 duplicated sequences were aligned with the upstream V3 domain of our sequences  
224 corresponding to codons 2353 to 2379. Downstream V3 (codons 2383 to 2410) was the  
225 additional domain, so-called V3-like (V3-I) in our study.

226

### 227 **Phylogenetic trees**

228 Phylogenetic trees built from aligned V3 and V3-I clonal nucleotidic sequences  
229 showed that V3-I sequences clustered separately (Figure 2). Within these two clades,  
230 sequences were grouped according to the quasi-species of each patient. Clustering according  
231 to pathology was observed within the V3-I clade, where LC and HCC groups could be  
232 considered two sister clades despite a medium bootstrap resampling value (65). The V3 clade  
233 showed no specific distribution. The V3-I clade was genetically more distant from the  
234 outgroup (HCV-J) than the V3 clade. Furthermore, for each patient showing evolution to  
235 cancer, V3-I clusters exhibited longer branch lengths than V3, suggesting higher genetic

236 diversity for V3-1. No differences between the two paralogs in terms of viral evolution were  
237 found in the LC group. Viral evolution was analyzed in the two patients from whom  
238 sequential serum samples were available seven years apart: using the major clone  
239 evolutionary rate, the divergence time between V3 or V3-1 and the HCV-J prototype strain  
240 was estimated at fifty years, corresponding to the infection duration of our patients. Thus,  
241 these strains were transmitted after having already acquired the duplication.

242

### 243 **Clonal diversity of V3 domains as related to clinical outcome**

244 Via within-sample analysis, genetic distances calculated from V3 and V3-1 sequences  
245 were shown to be similar in each clinical group (HCC V3: 0.041 vs. V3-1: 0.0476,  $p=0.34$ ; LC  
246 V3: 0.0525 vs. V3-1: 0.0714, Wilcoxon-Mann-Whitney test,  $p=0.70$ ) (Table 2). Comparison  
247 of genetic distances in each domain (HCC V3:0.041 vs. LC V3: 0.0525,  $p=0.63$ ; HCC V3-1:  
248 0.0476 vs. LC V3-1: 0.0714, Wilcoxon-Mann-Whitney test,  $p=0.40$ ) did not show any  
249 significant differences. Between-sample analysis concerned all distances from HCV-J of the  
250 HCC group sequences compared to the LC group. As expected, between-sample genetic  
251 distances were clearly higher than within-sample distances. In the HCC group, V3-1  
252 sequences were found to be more distant from HCV-J than V3 sequences (average genetic  
253 distances were, respectively, 0.3414 and 0.2618, Student's t-test  $p<0.0001$ ). In the LC group,  
254 V3 and V3-1 sequences were statistically at the same distance from HCV-J (respectively  
255 0.2429 and 0.2353, Student's t-test,  $p=0.24$ ). Between-group comparisons in each domain  
256 showed statistically significant differences in genetic distances from HCV-J (HCC V3: 0.2618  
257 vs. LC V3: 0.2429,  $p=0.0007$ ; HCC V3-1: 0.3414 vs. LC V3-1: 0.2353, Student's t-test,  
258  $p<0.0001$ ). HCC V3 and HCC V3-1 were found to be statistically more distant from HCV-J  
259 than LC V3 and LC V3-1, respectively. Within the HCC group, V3-1 sequences were more  
260 mutated and thus more distant from HCV-J than those of V3.

261 Analyzing positional Shannon's entropy variations in one patient from whom sera  
262 before and after HCC development were available, we observed 10 positions exhibiting a  
263 significant increase in entropy after HCC occurrence. Three positions were located in the  
264 junction between V3 and V3-1, and 7 along V3-1. The same analysis conducted in a patient  
265 without evolution to HCC showed 6 positions of increased entropy distributed throughout the  
266 sequence, and two positions of depleted entropy (data not shown). In a comparison of the  
267 cancer group (represented on positive scale) to the cirrhosis group (negative scale), sequences  
268 corresponding to patients with evolution to cancer revealed higher variability at junction

269 amino acids 27, 28, 29 and 31, clustering particularly in the V3-I domain (positions 32, 35-36-  
270 37, 39, and 41-42-43), as shown in Figure 3.

271

### 272 **Analysis of dN/dS ratios in V3 and V3-I regions**

273 To assess selective pressure upon V3 and its duplicate, we computed, using within and  
274 between-sample analyses, the proportion of non-synonymous mutations (dN) and dN/dS  
275 ratios and compared them between V3 and V3-I according to their clinical group. In within-  
276 sample analysis of the HCC group, the V3-I domain revealed nearly 53% more non-  
277 synonymous mutations than V3 (ratio dN V3-I/dN V3 = 1.5327) vs. 38% for the LC group  
278 (dN V3-I/V3 = 1.3834) (Table 3). Along V3, the mean intra-patient dN /dS ratio was 0.363 in  
279 the HCC group vs. 0.342 (Wilcoxon-Mann-Whitney test,  $p=1$ ) in the LC group (Table 2).  
280 When examining V3-I, the mean intra-patient dN/dS ratio reached a slightly higher but  
281 statistically non-significant value in the HCC group (0.495 vs. 0.391, Wilcoxon-Mann-  
282 Whitney test,  $p=0.63$ ) compared to the LC group (Table 2). In between-sample analysis of the  
283 HCC group, the V3-I domain exhibited about 19% more non-synonymous mutations than V3  
284 (ratio dN V3-I/dN V3 = 1.1901). V3 and V3-I domains of the LC group maintained the same  
285 number of non-synonymous mutations (ratio dN V3-I/dN V3 = 0.9101) (Table 3). When  
286 comparing each sequence of the whole HCC or LC group to reference strain HCV-J, the mean  
287 values of dN/dS ratios were found to be statistically different in terms of the V3 or V3-I  
288 domains (average ratios dN/dS on HCC V3 = 0.2945 vs. HCC V3-I = 0.2442,  $p<0.0001$ ; LC  
289 V3 = 0.2290 vs. LC V3-I = 0.1896, Student's t-test,  $p<0.0001$ ) (Table 2). In each group, the  
290 ratio of V3-I to V3 kept the same values (Table 3). We next compared between- and within-  
291 sample variability of each group of sequences. In the HCC group, dN, calculated by between-  
292 sample analysis, was 6.67 times higher than dN obtained by the within-sample method for V3  
293 (vs. 4.37 in the LC group), and 5.18 times higher for V3-I (vs. 2.87 in the LC group) (Table  
294 3). Concerning dN/dS values, the ratios between/within had approximately the same values  
295 throughout V3 or V3-I in each group. Among sequences from the HCC group, non-  
296 synonymous nucleotide changes increased to the same extent as silent mutations. In the two  
297 clinical groups, the overall selection ratio between the two domains remained the same.

298 Despite the fact that V3 and V3-I seemed to be submitted to overall negative selection,  
299 we sought to identify amino acids that might be positively selected in relation to their  
300 hydrophobic or hydrophilic properties. When computing dN/dS ratios separately for each  
301 codon of all aligned HCC group sequences (between-sample analysis), a total of 21 amino  
302 acids were found to be positively selected, with eight and ten amino acids belonging,



303 respectively, to V3 and V3-1 domains. Three of the four amino acids that formed a junction  
304 between these two regions were positively selected. In the LC group, 13 amino acids were  
305 submitted to positive selection (4 in the V3 domain and 9 in the V3-1 domain). In contrast, the  
306 four amino acids forming a junction evolved under negative selection (Fig. 4a).

307 When comparing HCC and LC group sequences according to the Kyte-Doolittle scale, the  
308 distribution of average hydropathy appeared to be statistically different in most parts of the  
309 amino acid sequence ( $p$  ranged from 0.0241 to 0.00001, Wilcoxon-Mann-Whitney test),  
310 except for residues 1, 2 and 18 (Fig. 4b).

311

### 312 **Physicochemical properties**

313 Disordered regions in proteins lack a stable 3-D structure. Because their flexibility  
314 facilitates multiple partner interactions, such proteins often interfere with cell cycle regulation  
315 [42]. Polar or charged amino acids, presenting low mean hydropathy, are disorder-promoting.  
316 Statistically significant differences between HCC and LC group sequences in amino acid  
317 composition and their physicochemical properties are shown in Figure 5. Within the HCC  
318 group, sequences were found to be enriched in disorder-promoting residues along V3-1 and its  
319 junction with V3 (Fig. 4). At codons 29 and 31, HCC group sequences were depleted of  
320 neutral residues (D) and order-promoting residues (L), respectively, but were enriched in  
321 disorder-promoting residues, (E) and (P), respectively. Positions 27 and 31 were enriched in  
322 surface-exposed and high-flexibility residues (respectively, R and P) and were depleted of  
323 low-flexibility and buried residues (respectively G and L), while codon 27 was also found to  
324 be enriched in a positively charged residue (R) and depleted of neutral residue (G) (logos not  
325 shown). Along V3-1, six positions (39, 43, 46, 49, 55 and 56) were enriched in disorder-  
326 promoting residues.

327

### 328 **Pattern/motif search**

329 In control sequences containing a non-duplicated V3 domain, one myristoylation site  
330 and one caseine kinase II (CKII) phosphorylation site were recognized by PROSITE. In  
331 contrast, five or six myristoylation sites and from two to four caseine kinase II  
332 phosphorylation sites were present along sequences containing V3 and its duplicate V3-1.  
333 Three myristoylation sites and two CKII phosphorylation sites were identified at conserved  
334 positions. No clear difference was observed between HCC and LC groups in the distribution  
335 of functional motifs. However, HCC group sequences seemed to contain fewer CKII sites in  
336 the V3-1 domain.

## 337 **Discussion**

338           Compared to mutation changes, insertions or deletions are quite rare along the HCV  
339 genome. While deletions are likely to have damaging effects, insertions that are fixed by  
340 natural selection may be neutral or even beneficial to the virus [43]. Because NS5A is known  
341 to be potentially oncogenic, we studied NS5A V3 duplication quasi-species structure and its  
342 relation to HCC occurrence. Indeed, NS5A in HCV-1b variability has been well studied in  
343 correlation with responsiveness to IFN treatment, especially in the ISDR and V3 regions  
344 [44,45,46,47]. But only a few studies have investigated the relationship between NS5A amino  
345 acid variations in the PKR-bd region and HCC development in chronic HCV-1b patients  
346 [48,49,50]. To our knowledge, V3 quasi-species composition has not been studied in relation  
347 to development of HCC.

348           In this study, three strains carrying V3 duplications were detected in the HCC group,  
349 two in the LC group and none in the CH group. Although the insertion prevalence was higher  
350 among patients with HCC, the difference between groups was statistically non-significant,  
351 probably due to the rarity of these inserted strains. The NS5A V3 domain seemed to be  
352 duplicated, as it contained a downstream tandem repeat of equal length (81 nucleotides, 27  
353 amino acids) without interruption of the open reading frame. Direct sequencing of these  
354 strains showed the presence of a 27-amino-acids insertion joined to the V3 domain by a four-  
355 amino-acids junction.

356           The NS5A carboxy-terminal region is known to be flexible and to tolerate insertions  
357 and deletions at low frequency without impairing viral replication [51]. Insertions of a few  
358 amino acids have been described within the ISDR, and probably play a role in immune  
359 response evasion [43]. Within the V3 domain, two permissive sites enabling insertion of the  
360 green fluorescent protein (GFP) heterologous sequence have been identified, one after residue  
361 384 (corresponding to position 2356) and another after residue 418 [35]. Via transposon-  
362 mediated mutagenesis, other permissive sites for 57 pb (19 amino acids) insertions have been  
363 reported at residues 396, 410 and 417 [52]. Interestingly, the 31-amino-acids insertion we  
364 describe in this work is located at residues 410-411. Its length suggests the absence of  
365 selective restriction, maintaining the V3 region at an adequate size.

366           HCV genetic variability has been studied mainly in the complete NS5A and ISDR  
367 regions. A few reports have analyzed V3 domain mutations in relation to antiviral therapy  
368 outcome, but none investigated V3 variability according to liver disease stage or HCC  
369 occurrence [44,53,54,55]. By between-sample analysis and in pre-treatment, Veillon *et al.*

370 (2007), when studying European HCV-1b strains, found genetic distances on V3 that varied  
371 from 0.1836 (responder group) to 0.1372 (non-responders) [54]. Jardim *et al.* (2009),  
372 analyzing Brazilian strains, and El Hefnawi (2010), using strains of diverse geographic origin,  
373 obtained a genetic distance in V3 domain of, respectively, 0.2510 and 0.272 for the responder  
374 groups and 0.2430 and 0.166 for non-responders [44,55]. Noursbaum's study of American  
375 strains (2000) indicated shorter genetic distances in the V3 region in both patient response  
376 groups (respectively, 0.0521 for responders vs. 0.0291 for non-responders) [53]. Despite the  
377 fact that, in our study, between-sample analysis was performed by comparison with HCV-J,  
378 the values we obtained for V3 (0.2618 for the HCC group vs. 0.2429 for the LC group) and  
379 V3-1 (0.3414 for the HCC group vs. 0.2353 for the LC group) were in the range estimated by  
380 the studies mentioned previously [44,54,55]. Among patients with evolution to cancer,  
381 duplicate V3-1 showed statistically significant higher genetic diversity. Furthermore, the  
382 higher Shannon entropy values found in the HCC group were particularly concentrated along  
383 V3-1 and its junction with V3, suggesting more pronounced complexity of these regions.

384 Maintenance or loss of viral genetic variability depends on host immune selective  
385 pressure. HCV evasion strategy is based on overall increased variability. Although both dN  
386 and dS increase with the duration of chronic infection, a decrease in the dN/dS ratio is  
387 observed in the HCV genome over time, since selection pressure could act to preserve viral  
388 proteins [56]. Therefore, estimated duration of HCV infection is important for interpreting  
389 selection pressure ratios. The duration of HCV infection was very similar in our HCC and LC  
390 groups (respectively, 46.7 vs. 46.5 years). Within the HCC group, V3-1 was characterized by a  
391 higher dN/dS ratio than V3 (respectively, 0.4947 vs. 0.3634), especially in within-sample  
392 analysis, as linked to individual host balance. According to Li *et al.* (2008), the lowest dN/dS  
393 ratios, determined over the whole HCV genome, occurred at 21-38 years after infection,  
394 associated with advanced liver cirrhosis or severe liver disease (respectively, 0.096 vs. 0.211  
395 in case of mild liver disease) [56]. We observed higher ratios even at 40 years after infection,  
396 but our study focused on the hypervariable V3 zone only. Previously referenced studies  
397 focusing on V3 genetic distances during antiviral therapy also established dN/dS ratios,  
398 without indicating the stage of HCV-induced liver disease. According to Noursbaum's study  
399 (2000), the NS5A V3 region appeared to have a dN/dS ratio of around 0.5, higher in the  
400 responder group than in non-responders, suggesting that the region may be under positive  
401 selection pressure [53]. In contrast, Jardim's study (2009) found very low dN/dS ratios in this  
402 region (0.0060 in between-sample analysis) [44]. However, our results are consistent with

403 those of Veillon's study that reported a dN/dS value of 0.393 in the responder group in  
404 between-sample analysis [54].

405 Because duplication of a gene can lead to differentiation of the paralog, and thus to  
406 fixation of amino acid changes, we next identified the residues that were under positive  
407 selection pressure within each clinical group. In both groups, the V3-I domain exhibited a  
408 trend to fixing more mutations. In the HCC group, selection pressures were found to be  
409 positive at eight positions (29.6%) belonging to the V3 domain, and 10 positions (37.03%) of  
410 the V3-I domain. In the LC group, positive selection acted, respectively, on 4 (14.8%) and 9  
411 positions (33.3%). These results are consistent with the Cuevas study (2008), which reported  
412 10 of 24 V3 amino acids (41.7%) evolving under positive selection during therapy [46].

413 Amino acid changes were next analyzed for their physicochemical properties. A  
414 change in an amino acid group (hydrophobic or polar) suggests a modification in protein  
415 structure. However, the secondary structure of the NS5A carboxy-terminus remains unknown.  
416 Interestingly, sites undergoing positive selection were also those affected by hydrophathy  
417 changes. By comparing average hydrophathy profiles along V3 and V3-I, we observed that,  
418 overall, sequences from HCC group patients were more hydrophobic than those from the LC  
419 group, except for the region located between codons 25 and 35, that was clearly more  
420 hydrophilic and thus more strongly exposed. When combining logos predicting amino acid  
421 intrinsic properties, the sequences from HCC patients were found to be enriched in disorder-  
422 promoting residues between amino acids 27 and 31, corresponding to the junction between  
423 V3 and V3-I, and along the entire V3-I sequence. Among patients with HCC outcome, the  
424 short joining region was also enriched in surface-exposed and high flexibility residues. Thus,  
425 when comparing strains of the same infection duration, the additional V3-I domain seemed to  
426 be more divergent among patients with HCC outcome, notably because of some sites  
427 subjected to positive selective pressure located at the junction and along V3-I. Quasispecies  
428 analysis of these oncogenic strains showed that these regions were enriched in disorder-  
429 promoting amino acids and that the junction between V3 and V3-I was disordered, surface-  
430 exposed and of high flexibility.

431 Some authors used *in silico* methods to predict NS5A carboxy-terminus secondary  
432 structure. According to El Hefnawi [57], the region located at 374-410, overlapping V3, is  
433 intrinsically disordered, 100% exposed and of low complexity. Disorder-promoting amino  
434 acids are less exchangeable in disordered regions than in ordered ones, indicating a potential  
435 functional role, as possibly played by V3-I in HCC patients. Disordered proteins are also  
436 characterized by their preponderance toward contain tandem repeats [42].

437 Promiscuous proteins work as hubs in the cellular protein interaction network. They  
438 play a key role in cellular signaling pathways, the dysfunction of which may lead to  
439 carcinogenesis [58]. These proteins, engaged in multiple protein-protein interactions, can  
440 increase their fitness in the interaction network by incorporating intrinsic disorder and  
441 sequence repeats. Among HCV proteins, NS5A is one of the most strongly connected,  
442 interacting with a plethora of cellular proteins involved both in cell growth and apoptosis  
443 regulation [59]. Like other oncogenic proteins such as the SV40T antigen, papillomavirus E6,  
444 adenovirus E1B 55-kDa and HBx from HBV, NS5A is able to hold tumor suppressor p53 in  
445 the cell cytoplasm and to disrupt p53-p21/WAF1 cell cycle pathways. According to Lan [60],  
446 the NS5A C-terminus (amino acids 2135 to 2419) interacts with p53. We suggest that, among  
447 strains we have studied with known cancer outcome, V3 duplication, by addition of a  
448 disordered flexible V3-I domain, has as a consequence the enlargement of its available  
449 binding surface area, likely to more drastically interact with p53. Another argument  
450 suggesting a modification in NS5A V3 conformation is the increase in possible myristoylation  
451 sites due to adjunction of the V3-I domain. According to Yamasaki [61], covalent myristate  
452 binding is not reversible and could modify the protein's hydrophathy profile and its interaction  
453 network.

454 Gene duplication is a common phenomenon, known in all three domains of life,  
455 among which it represents a means of supplying new genetic material and contributing to  
456 organism diversification. After duplication, the additional gene or paralog, released from  
457 selective pressure, could have the possibility of divergent evolution, thus acquiring new  
458 functions. Evolution through duplications, deletions and mutations has been described for a  
459 few flavivirus genes. In the present work, we confirmed the relationship between V3  
460 duplication and HCC occurrence, as reported in a transversal epidemiologic study (Le Guillou  
461 *et al.*, unpublished data). The significance of this finding remains unclear and must be  
462 confirmed by other longitudinal studies. Whether or not V3 duplication has functional  
463 consequences might be ascertained by interactome methods [59].

464

465 Acknowledgments: This work was supported by a grant from the French research  
466 agency for AIDS and Hepatitis ANRS. We wish to thank Pr Ferec for the sequencing platform  
467 access and Jerry Bram, English medical translator; for reviewing this manuscript.

468

469 The authors declared no conflict of interest concerning this work.

470

## References

1. Lavanchy D (2009) The global burden of hepatitis C. *Liver Int* 29 Suppl 1: 74-81.
2. Bosch FX, Ribes J, Cleries R, Diaz M (2005) Epidemiology of hepatocellular carcinoma. *Clin Liver Dis* 9: 191-211, v.
3. Gomaa AI, Khan SA, Toledano MB, Waked I, Taylor-Robinson SD (2008) Hepatocellular carcinoma: epidemiology, risk factors and pathogenesis. *World J Gastroenterol* 14: 4300-4308.
4. Benvegnu L, Gios M, Boccato S, Alberti A (2004) Natural history of compensated viral cirrhosis: a prospective study on the incidence and hierarchy of major complications. *Gut* 53: 744-749.
5. Yoshida H, Shiratori Y, Moriyama M, Arakawa Y, Ide T, et al. (1999) Interferon therapy reduces the risk for hepatocellular carcinoma: national surveillance program of cirrhotic and noncirrhotic patients with chronic hepatitis C in Japan. IHIT Study Group. Inhibition of Hepatocarcinogenesis by Interferon Therapy. *Ann Intern Med* 131: 174-181.
6. Fattovich G, Stroffolini T, Zagni I, Donato F (2004) Hepatocellular carcinoma in cirrhosis: Incidence and risk factors. *Gastroenterology* 127: S35-S50.
7. El-Serag HB, Rudolph KL (2007) Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 132: 2557-2576.
8. Fattovich G, Ribero ML, Pantalena M, Diodati G, Almasio P, et al. (2001) Hepatitis C virus genotypes: distribution and clinical significance in patients with cirrhosis type C seen at tertiary referral centres in Europe. *J Viral Hepat* 8: 206-216.
9. Tagger A, Donato F, Ribero ML, Chiesa R, Portera G, et al. (1999) Case-control study on hepatitis C virus (HCV) as a risk factor for hepatocellular carcinoma: the role of HCV genotypes and the synergism with hepatitis B virus and alcohol. Brescia HCC Study. *Int J Cancer* 81: 695-699.
10. Tanaka Y, Kurbanov F, Mano S, Orito E, Vargas V, et al. (2006) Molecular tracing of the global hepatitis C virus epidemic predicts regional patterns of hepatocellular carcinoma mortality. *Gastroenterology* 130: 703-714.
11. Bruno S, Crosignani A, Maisonneuve P, Rossi S, Silini E, et al. (2007) Hepatitis C virus genotype 1b as a major risk factor associated with hepatocellular carcinoma in patients with cirrhosis: a seventeen-year prospective cohort study. *Hepatology* 46: 1350-1356.
12. Nakano T, Lau GM, Sugiyama M, Mizokami M (2012) An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region. *Liver Int* 32: 339-345.
13. Raimondi S, Bruno S, Mondelli MU, Maisonneuve P (2009) Hepatitis C virus genotype 1b as a risk factor for hepatocellular carcinoma development: a meta-analysis. *J Hepatol* 50: 1142-1154.
14. Anzola M (2004) Hepatocellular carcinoma: role of hepatitis B and hepatitis C viruses proteins in hepatocarcinogenesis. *J Viral Hepat* 11: 383-393.
15. Kasprzak A, Adamek A (2008) Role of hepatitis C virus proteins (C, NS3, NS5A) in hepatic oncogenesis. *Hepatol Res* 38: 1-26.
16. Vallet S, Gouriou S, Nkontchou G, Hotta H, Vilerio M, et al. (2007) Is hepatitis C virus NS3 protease quasi-species heterogeneity predictive of progression from cirrhosis to hepatocellular carcinoma? *J Viral Hepat* 14: 96-106.
17. Ogata S, Florese RH, Nagano-Fujii M, Hidajat R, Deng L, et al. (2003) Identification of hepatitis C virus (HCV) subtype 1b strains that are highly, or only weakly, associated with hepatocellular carcinoma on the basis of the secondary structure of an amino-terminal portion of the HCV NS3 protein. *J Clin Microbiol* 41: 2835-2841.
18. Banerjee A, Ray RB, Ray R (2010) Oncogenic potential of hepatitis C virus proteins. *Viruses* 2: 2108-2133.

19. Everson GT, Sims KD, Rodriguez-Torres M, Hezode C, Lawitz E, et al. (2013) Efficacy of an Interferon- and Ribavirin-free Regimen of Daclatasvir, Asunaprevir, and BMS-791325 in Treatment-naive Patients with HCV Genotype 1 Infection. *Gastroenterology*.
20. Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, et al. (2010) Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. *Nature* 465: 96-100.
21. Herbst DA, Reddy KR (2013) NS5A inhibitor, daclatasvir, for the treatment of chronic hepatitis C virus infection. *Expert Opin Investig Drugs* 22: 1337-1346.
22. Majumder M, Ghosh AK, Steele R, Ray R, Ray RB (2001) Hepatitis C virus NS5A physically associates with p53 and regulates p21/waf1 gene expression in a p53-dependent manner. *J Virol* 75: 1401-1407.
23. Ghosh AK, Steele R, Meyer K, Ray R, Ray RB (1999) Hepatitis C virus NS5A protein modulates cell cycle regulatory genes and promotes cell growth. *J Gen Virol* 80 ( Pt 5): 1179-1183.
24. Appel N, Zayas M, Miller S, Krijnse-Locker J, Schaller T, et al. (2008) Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog* 4: e1000035.
25. Tellinghuisen TL, Marcotrigiano J, Gorbalenya AE, Rice CM (2004) The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J Biol Chem* 279: 48576-48587.
26. Tellinghuisen TL, Foss KL, Treadaway JC, Rice CM (2008) Identification of residues required for RNA replication in domains II and III of the hepatitis C virus NS5A protein. *J Virol* 82: 1073-1083.
27. Tellinghuisen TL, Foss KL, Treadaway J (2008) Regulation of hepatitis C virion production via phosphorylation of the NS5A protein. *PLoS Pathog* 4: e1000032.
28. Brass V, Bieck E, Montserret R, Wolk B, Hellings JA, et al. (2002) An amino-terminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *J Biol Chem* 277: 8130-8139.
29. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, et al. (1995) Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest* 96: 224-230.
30. Pascu M, Martus P, Hohne M, Wiedenmann B, Hopf U, et al. (2004) Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. *Gut* 53: 1345-1351.
31. Cuddihy AR, Wong AH, Tam NW, Li S, Koromilas AE (1999) The double-stranded RNA activated protein kinase PKR physically associates with the tumor suppressor p53 protein and phosphorylates human p53 on serine 392 in vitro. *Oncogene* 18: 2690-2702.
32. Veillon P, Payan C, Gaudy C, Goudeau A, Lunel F (2004) [Mutation analysis of ISDR and V3 domains of hepatitis C virus NS5A region before interferon therapy with or without ribavirin]. *Pathol Biol (Paris)* 52: 505-510.
33. El-Shamy A, Sasayama M, Nagano-Fujii M, Sasase N, Imoto S, et al. (2007) Prediction of efficient virological response to pegylated interferon/ribavirin combination therapy by NS5A sequences of hepatitis C virus and anti-NS5A antibodies in pre-treatment sera. *Microbiol Immunol* 51: 471-482.
34. El-Shamy A, Nagano-Fujii M, Sasase N, Imoto S, Kim SR, et al. (2008) Sequence variation in hepatitis C virus nonstructural protein 5A predicts clinical outcome of pegylated interferon/ribavirin combination therapy. *Hepatology* 48: 38-47.
35. Moradpour D, Evans MJ, Gosert R, Yuan Z, Blum HE, et al. (2004) Insertion of green fluorescent protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus replication complexes. *J Virol* 78: 7400-7409.

36. Trinchet JC, Chaffaut C, Bourcier V, Degos F, Henrion J, et al. (2011) Ultrasonographic surveillance of hepatocellular carcinoma in cirrhosis: a randomized trial comparing 3- and 6-month periodicities. *Hepatology* 2011 Dec;54(6):1987-97 doi: 101002/hep24545.
37. Laperche S, Lunel F, Izopet J, Alain S, Deny P, et al. (2005) Comparison of hepatitis C virus NS5b and 5' noncoding gene sequencing methods in a multicenter study. *J Clin Microbiol* 43: 733-739.
38. Duverlie G, Khorsi H, Castelain S, Jaillon O, Izopet J, et al. (1998) Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity. *J Gen Virol* 79 ( Pt 6): 1373-1381.
39. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9: 299-306.
40. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* 92: 371-373.
41. Vacic V, Iakoucheva LM, Radivojac P (2006) Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. *Bioinformatics* 22: 1536-1537.
42. Szalkowski AM, Anisimova M (2011) Markov Models of Amino Acid Substitution to Study Proteins with Intrinsically Disordered Regions. *PLoS One* 6: e20488.
43. Torres-Puente M, Cuevas JM, Jimenez-Hernandez N, Bracho MA, Garcia-Robles I, et al. (2007) Contribution of insertions and deletions to the variability of hepatitis C virus populations. *J Gen Virol* 88: 2198-2203.
44. Jardim AC, Yamasaki LH, de Queiroz AT, Bittar C, Pinho JR, et al. (2009) pecies of hepatitis C virus genotype 1 and treatment outcome with peginterferon and ribavirin. *Infect Genet Evol* 9: 689-698.
45. Munoz de Rueda P, Casado J, Paton R, Quintero D, Palacios A, et al. (2008) Mutations in E2-PePHD, NS5A-PKRBD, NS5A-ISDR, and NS5A-V3 of hepatitis C virus genotype 1 and their relationships to pegylated interferon-ribavirin treatment responses. *J Virol* 82: 6644-6653.
46. Cuevas JM, Torres-Puente M, Jimenez-Hernandez N, Bracho MA, Garcia-Robles I, et al. (2008) Genetic variability of hepatitis C virus before and after combined therapy of interferon plus ribavirin. *PLoS One* 3: e3058.
47. Bouzgarrou N, Hassen E, Mahfoudh W, Gabbouj S, Schvoerer E, et al. (2009) NS5A(ISDR-V3) region genetic variability of Tunisian HCV-1b strains: Correlation with the response to the combined interferon/ribavirin therapy. *J Med Virol* 81: 2021-2028.
48. Hung CH, Chen CH, Lee CM, Wu CM, Hu TH, et al. (2008) Association of amino acid variations in the NS5A and E2-PePHD region of hepatitis C virus 1b with hepatocellular carcinoma. *J Viral Hepat* 15: 58-65.
49. Gimenez-Barcons M, Franco S, Suarez Y, Forns X, Ampurdanes S, et al. (2001) High amino acid variability within the NS5A of hepatitis C virus (HCV) is associated with hepatocellular carcinoma in patients with HCV-1b-related cirrhosis. *Hepatology* 34: 158-167.
50. De Mitri MS, Cassini R, Bagaglio S, Morsica G, Andreone P, et al. (2007) Evolution of hepatitis C virus non-structural 5A gene in the progression of liver disease to hepatocellular carcinoma. *Liver Int* 27: 1126-1133.
51. Appel N, Pietschmann T, Bartenschlager R (2005) Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *J Virol* 79: 3187-3194.
52. Liu S, Ansari IH, Das SC, Pattnaik AK (2006) Insertion and deletion analyses identify regions of non-structural protein 5A of Hepatitis C virus that are dispensable for viral genome replication. *J Gen Virol* 87: 323-327.



53. Nousbaum J, Polyak SJ, Ray SC, Sullivan DG, Larson AM, et al. (2000) Prospective characterization of full-length hepatitis C virus NS5A quasispecies during induction and combination antiviral therapy. *J Virol* 74: 9028-9038.
54. Veillon P, Payan C, Le Guillou-Guillemette H, Gaudy C, Lunel F (2007) Quasispecies evolution in NS5A region of hepatitis C virus genotype 1b during interferon or combined interferon-ribavirin therapy. *World J Gastroenterol* 13: 1195-1203.
55. ElHefnawi MM, Zada S, El-Azab IA (2010) Prediction of prognostic biomarkers for interferon-based therapy to hepatitis C virus patients: a meta-analysis of the NS5A protein in subtypes 1a, 1b, and 3a. *Virol J* 7: 130.
56. Li H, Hughes AL, Bano N, McArdle S, Livingston S, et al. (2008) Genetic Diversity of Near Genome-Wide Hepatitis C Virus Sequences during Chronic Infection: Evidence for Protein Structural Conservation Over Time. *PLoS One* 6: e19562.
57. El Hefnawi MM, El Behaidy WH, Youssif AA, Ghalwash AZ, El Housseiny LA, et al. (2009) Natural genetic engineering of hepatitis C virus NS5A for immune system counterattack. *Ann N Y Acad Sci* 1178: 173-185.
58. Fromer M, Shifman JM (2009) Tradeoff between stability and multispecificity in the design of promiscuous proteins. *PLoS Comput Biol* 5: e1000627.
59. de Chasse B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, et al. (2008) Hepatitis C virus infection protein network. *Mol Syst Biol* 4: 230.
60. Lan KH, Sheu ML, Hwang SJ, Yen SH, Chen SY, et al. (2002) HCV NS5A interacts with p53 and inhibits p53-mediated apoptosis. *Oncogene* 21: 4801-4811.
61. Yamasaki LH, Arcuri HA, Jardim AC, Bittar C, de Carvalho-Mello IM, et al. (2012) New insights regarding HCV-NS5A structure/function and indication of genotypic differences. *Virol J* 9: 14.

## Figure legends

**Fig. 1: Alignment of sequences obtained by direct sequencing of NS5A V3 PCR products.** The V3 domain is shaded in light gray, the additional domain V3-l in dark gray and the 4-amino-acids junction in black. Dots represent conserved positions in comparison to the HCV-J sequence (top, light gray zone), or to the same sequence artificially reported in the dark gray zone (HCV-J\*). K: patients with subsequent HCC occurrence; C: patients with liver cirrhosis. The alignment was edited using GeneDoc 2.0.1.

**Fig. 2: Phylogenetic tree drawn from 167 clonal sequences of NS5A V3 and its paralog V3-l.** Condensed taxa from each patient are represented by a proportional triangle; black for patients belonging to the HCC group and gray for LC patients. The HCV-J strain was included as an outgroup. The tree was built with neighbor-joining method, with a Kimura two-parameter model. The scale bar represents 2% of nucleotide sequence divergence.

**Fig. 3: Duplicated V3 relative Shannon entropy comparing cancer (positive scale) and cirrhosis groups (negative scale).** V3: amino acids 1-27; V3-l: amino acids 32-58. Entropy was calculated using the Shannon Entropy-Two tool available at the HCV LANL database. Positions exhibiting significant differences in variability for the cancer group are labeled in black and grouped in the dotted line square zone.

**Fig. 4: Graphic representation of variable selective pressures along V3 and V3-l sequences (A) and their hydropathy profiles (B).** V3: amino acids 1-27; V3-l: amino acids 32-58. A) dN/dS ratios calculated by between-sample analysis using the SNAP tool available on the HCV LANL database. dN/dS values are depicted at the y axis in HCC (HCC occurrence in black) or the LC group (LC without evolution to cancer, in gray). Codons are indicated at the x axis. B) Kyte and Doolittle plots of HCC (in black) and cirrhosis groups (in gray) were calculated using the DAMBE program, with a smoothing window equal to six residues. Each point of the curves is the clone set hydropathy average index. Amino acid positions are indicated at the x-axis; hydropathy values are indicated at the y-axis.

**Fig. 5: Two-sample logo of HCC sequence amino acid properties compared to those of LC.** Enriched amino acids in the HCC group are shown on the positive y-axis, depleted amino

acids on the negative y-axis. Disorder-promoting residues: A, R, S, Q, E, G, K, P; order-promoting residues: N, C, I, L, F, W, Y, V; disorder-order neutral residues: D, H, M, T.

Table 1. Patient characteristics and presence of V3 duplication according to the clinical stage of liver disease

	HCC	LC	CH	p-value
Number of patients	28	27	19	
Age (mean $\pm$ SD) (years)	71.09 $\pm$ 8.6	63.19 $\pm$ 11.4	56.07 $\pm$ 13.0	p < 0.0001
Gender	Female	Female	Female	
METAVIR score	F4 (100%)	F4 (100%)	F1 (26.3%) F2 (73.7%)	
HCV-RNA mean (range) (U/ml), <sup>*</sup>	2.98 [0.1 - 4.93] $\times 10^6$	2.78 [0.07 - 8.29] $\times 10^6$	4.10 [0.52 - 15] $\times 10^6$	NS
ALT mean (range) ( $\times$ ULN)	2.2 [1 - 6]	2.5 [1 - 5.83]	2.2 [1 - 5]	NS
AFP mean (range) (ng/ml)	64 [2 - 695]	10 [5 - 55]	5 [3 - 10]	p < 0.05
Number of patients with V3 duplication	3	2	0	p = 1

HCC: hepatocellular carcinoma; LC: liver cirrhosis; CH: chronic hepatitis.

ALT: alanine aminotransferase; ULN: upper level of normal range; AFP: alpha-fetoprotein; NS: not significant.

<sup>\*</sup>Ultr (Cobas AmpliCor Monitor; Roche Diagnostics, Meylan, France).

Table 2. HCV NSSA V3 and V3-I quasi-species analysis.

Patients and groups	No. of sequenced clones	No. of different sequences (nt, AA)	% identity between V3 and V3-like direct sequences (nt, AA)	Genetic distance <sup>™</sup>		Genetic distance from HCV-J <sup>†</sup>		Global dN/dS ratio <sup>™</sup>		Global dN/dS ratio compared to HCV-J <sup>†</sup>	
				V3	V3-I	V3	V3-I	V3	V3-I	V3	V3-I
<b>HCC (n = 4)</b>											
K <sub>A</sub>	27	12, 8	0.7160 (0.6296)	0.0090	0.0089			0.6996	0.5394		
K <sub>NS</sub>	24	21, 18	0.7160 (0.6296)	0.0669	0.0726			0.3374	0.3364		
K <sub>B</sub>	19	19, 16	0.8271 (0.8148)	0.0717	0.0773			0.2557	0.1430		
K <sub>C</sub>	32	26, 23	0.7901 (0.6296)	0.0292	0.067			0.1675	0.9161		
Mean	112		<b>0.7623 (0.6759)</b>	<b>0.041</b>	<b>0.0476</b>	<b>0.2618</b>	<b>0.3414</b>	<b>0.3634</b>	<b>0.4847</b>	<b>0.2946</b>	<b>0.2442</b>
<b>LC (n = 3)</b>											
C <sub>A</sub>	18	10, 8	0.9518 (0.6296)	0.0347	0.0298			0.6901	0.4699		
C <sub>B</sub>	30	28, 19	0.8025 (0.7407)	0.0655	0.0998			0.1792	0.2792		
C <sub>NS</sub>	18	10, 7	0.7901 (0.7037)	0.0572	0.0846			0.2572	0.2099		
Mean	66		<b>0.8148 (0.7407)</b>	<b>0.0626</b>	<b>0.0714</b>	<b>0.2429</b>	<b>0.2363</b>	<b>0.3422</b>	<b>0.3193</b>	<b>0.2290</b>	<b>0.1896</b>

HCC: hepatocellular carcinoma; LC: liver cirrhosis; V3: V3 domain; V3-I: V3-like domain.

K<sub>A</sub>, K<sub>B</sub>, K<sub>C</sub>: patients with HCC occurrence; C<sub>A</sub>, C<sub>B</sub>: patients with liver cirrhosis.

KAB and CBB serum samples were collected seven years after, respectively, KA and CB.

Mean genetic distance was estimated with the Kimura two-parameter method in MEGA 4 software; <sup>™</sup> within-sample; <sup>†</sup> between-samples.

The number of synonymous (dS) and non-synonymous (dN) substitutions per site was estimated using Nei and Gotoh's method with Jukes-Cantor correction implemented in

MEGA4 software.

Data discussed below are indicated in bold.

Table 3. Within- and between-sample comparisons of V3-I/V3 dN and dN/dS ratios

Mode of comparison	HCC		LC		
	dN	dN/dS	dN	dN/dS	
Within-sample, V3-I/V3 ratio	<b>1.5327</b>	1.3612	<b>1.3834</b>	0.9931	
Between-samples, V3-I/V3 ratio	1.1901	0.9293	<b>0.9101</b>	0.8277	
Between-samples/ within-sample ratio	V3	<b>6.67</b>	0.8104	<b>4.37</b>	0.8692
	V3-I	6.18	0.4936	<b>2.87</b>	0.5938

HCC: hepatocellular carcinoma; LC: liver cirrhosis; V3: V3 domain; V3-I: V3-like domain.  
Data discussed below are indicated in bold.

Figure 1.

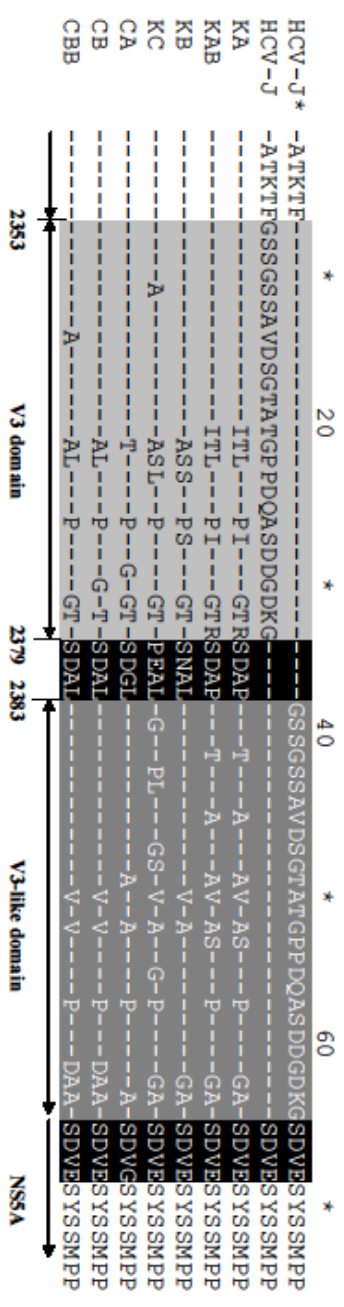


Figure 2.

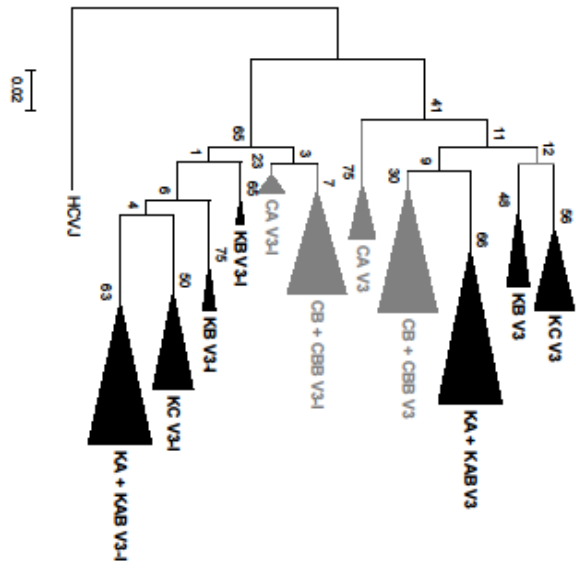




Figure 3.

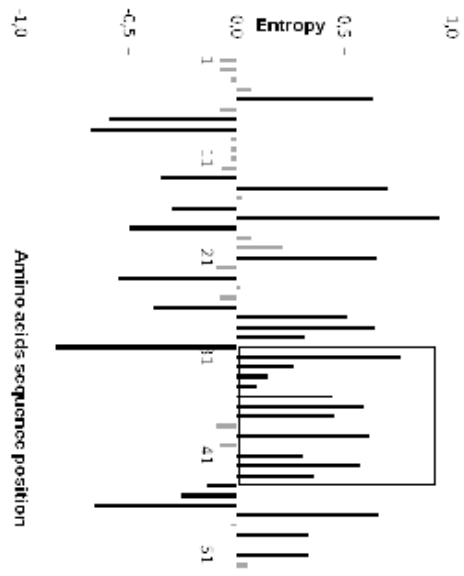


Figure 4.

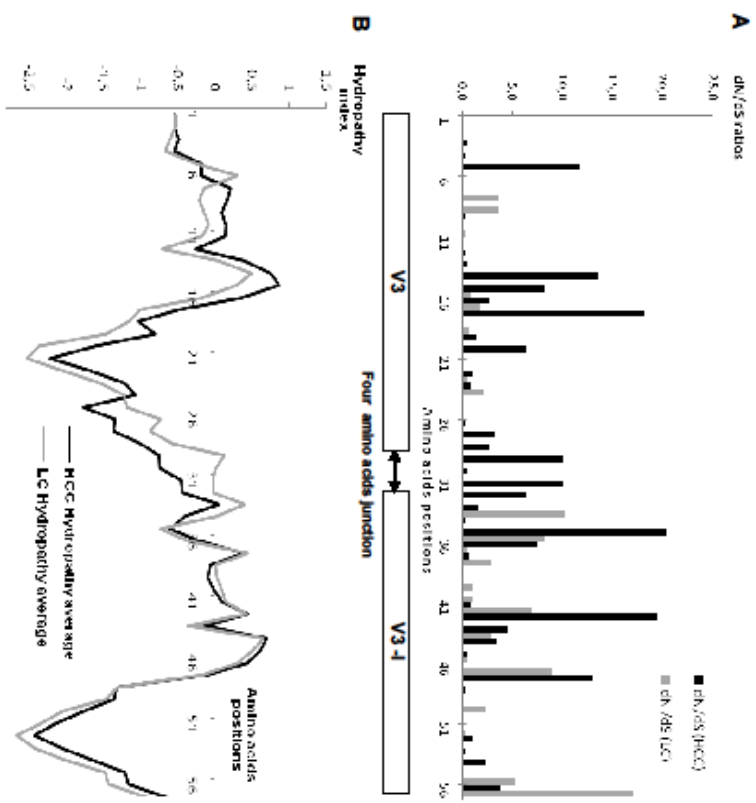
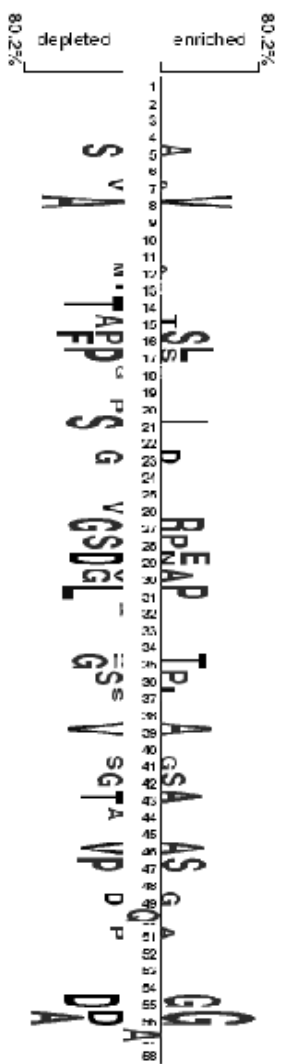


Figure 5.



# Thèse de Doctorat

Hélène LE GUILLOU-GUILLEMETTE

**Étude de la duplication du domaine V3 de la région NS5A du virus de l'hépatite C de génotype 1b : épidémiologie clinique et moléculaire et aspects fonctionnels**

**Study of the V3 duplicated domain of the NS5A region in the genotype 1b hepatitis C virus: clinical and molecular epidemiology and functional analyses.**

## Résumé

L'infection par le virus de l'hépatite C (VHC) est responsable d'hépatite chronique, de cirrhose et de carcinome hépatocellulaire. La protéine NS5A du VHC participe à la réplication et l'assemblage viral, et à l'oncogenèse. Notre étude a caractérisé un polymorphisme du domaine V3-NS5A du VHC de génotype 1b, découvert au laboratoire.

Nous avons identifié une duplication en tandem sans rupture d'ORF du domaine V3 avec une prévalence de 3,05%. La prévalence de cette souche augmentait avec la sévérité de la pathologie hépatique (de 3.0% groupe sans cirrhose jusqu'à 9.4% groupe cirrhose et CHC,  $p$  de tendance=0.045).

L'analyse des polymorphismes des séquences directes et clonales a montré que les souches dupliquées constituaient la totalité de la quasi-espèce et persistaient au moins 10 ans. Le mécanisme de duplication serait une recombinaison non-homologue entre souches de 2 populations sauvages.

Nos résultats fonctionnels préliminaires ont suggéré des protéines partenaires spécifiques des protéines NS5A dupliquées par un crible en système double-hybride levure. Nous avons réalisé la production des protéines NS5A dupliquées en systèmes cellulaire et bactérien, avec une optimisation en système bactérien. Les premiers tests de pull down n'ont pas encore permis de valider les interactions.

Nous avons identifié une duplication partielle V3-NS5A chez le VHC1b (i) potentiellement impliquée dans la carcinogenèse, (ii) présente durant l'intégralité de l'infection, (iii) issue d'une recombinaison non-homologue, événement encore jamais décrit chez le VHC *in vivo*. Nos travaux ont aussi permis l'identification de partenaires cellulaires et la production des protéines NS5A dupliquées.

## Mots clés

VHC, domaine V3, protéine NS5A, carcinome hépatocellulaire, recombinaison non-homologue, protéine recombinante

## Abstract

The hepatitis C virus (HCV) infection leads to chronic hepatitis, cirrhosis and hepatocellular carcinoma (HCC). The HCV NS5A protein has been shown to be involved in viral replication and assembly and in the liver carcinogenesis. We investigated a V3-NS5A polymorphism in HCV 1b identified in our laboratory.

We found two V3 domains tandemly duplicated without ORF disruption in 3.05% of the HCV sequences. The prevalence of this duplication increased with liver disease severity (from 3.0% in patients without cirrhosis to 9.4% in patients with both cirrhosis and HCC,  $p$  for trend=0.045).

Direct sequencing and clonal analyzes showed that quasispecies were entirely composed by mutants strains that persisted at least for 10 years. These V3 duplications were likely resulting from a non-homologous recombination that occurred between two wild-type strains.

Our preliminary functional results identified potential specific interactions between host-cell proteins and the V3-NS5A duplicated proteins using a yeast two-hybrid system. We expressed recombinant duplicated NS5A proteins in cell and bacterial cultures and developed an optimized protocol for bacterial cultures. A validation of the interactions has not been reached with the first pull-down assays.

We identified a V3-NS5A duplication in HCV 1b for the first time (i) associated with unfavorable evolution of liver disease including a possible involvement in liver carcinogenesis (ii) always present during the HCV infection (iii) resulting from a non-homologous recombination, mechanism that has never been described *in vivo* in HCV. We also suggested potential specific protein-protein interactions and performed recombinant NS5A protein production.

## Key Words

HCV, V3 domain, NS5A protein, hepatocellular carcinoma, non-homologous recombination, recombinant protein