Interactions homologues et hétérologues de la protéine centrale formant la capside du VHB

Mathilde Briday

THESE de DOCTORAT DE
L'UNIVERSITE CLAUDE BERNARD LYON 1

École Doctorale N° 205
École Doctorale Interdisciplinaire Sciences-Santé

Discipline : Biologie moléculaire et structurale, biochimie

Soutenue publiquement le 13/12/2022, par :
Mathilde BRIDAY

Homologous and heterologous interactions of the HBV capsid-forming core protein

---

Devant le jury composé de :

Pr. LOMBERGET Thierry	Université Lyon 1	Président
Dr. RINGKJOBING JENSEN Malene	CNRS, Grenoble	Rapporteure
Dr. DE ROCQUIGNY Hugues	INSERM, Tours	Rapporteur
Dr. BOUAZIZ Serge	CNRS, Paris	Examinateur
Dr. DURANTEL David	INSERM, Lyon	Examinateur
Dr. LAGURI Cédric	CNRS, Grenoble	Examinateur
Dr. BÖCKMANN Anja	CNRS, Lyon	Directrice de thèse
Dr. LECOQ Lauriane	CNRS, Lyon	Co-directrice de thèse

This thesis has been carried out under the supervision of Anja Böckmann and the co-supervision of Lauriane Lecoq, in the “Protein solid-state NMR” group which is part of the unit MMSB (Molecular Microbiology and Structural Biochemistry) – UMR CNRS 5086, at the IBCP:

Institut de Biologie et Chimie des Protéines

7, Passage du Vercors

69367 Lyon Cedex 07

France

Remerciements

Je souhaite commencer par adresser mes remerciements à ma directrice de thèse, Anja Böckmann, ainsi que ma co-directrice, Lauriane Lecoq. Elles ont su me guider pendant mes trois années de thèse, elles ont partagé avec moi leur expertise en RMN et en biologie structurale, elles m'ont soutenue pendant les moments d'hésitation et de doutes, mais elles ont aussi toujours été présentes pour les concrétisations, les accomplissements, et les moments de joie que nous avons partagés toutes ensemble. Je suis extrêmement reconnaissante envers elles, et je les remercie pour leur soutien infaillible pendant ces trois années.

Par la suite, je voudrais remercier toutes les personnes avec qui j'ai eu la chance de travailler, et en particulier les personnes de mon laboratoire : Marie Dujardin, ma première collègue de bureau, qui m'a aidé à maîtriser la Biorad; Marie-Laure Fogeron, avec qui j'ai appris la méthode du cell-free, que je voudrais particulièrement remercier pour son aide pour mon mémoire, pour les posters et pour les présentations orales; Roland Montserret, qui m'a tout appris sur l'ITC et le dichroïsme circulaire; Laura Cole, qui a produit plusieurs protéines pour moi; Louis Brigandat, mon collègue thésard, qui a analysé des grilles de microscopie pour moi, et qui a réalisé avec talent ma vidéo de départ; Yang Yang, mon autre collègue thésarde, avec qui j'ai pu échanger sur de nombreux sujets; Minh-Ha Nguyen, ma deuxième collègue de bureau, et Martí Ninot Pedrosa, les deux post-doctorants qui ont su répondre à mes interrogations tout au long de ma thèse; Louann Monnier, ma troisième collègue de bureau, qui a réalisé des expériences cell-free pour moi. Ensuite, je voudrais remercier mes collaborateurs à l'IBCP, qui m'ont apporté leur expertise : Virginie Gueguen-Chaignon en cristallographie et en DLS, et Éric Diesis pour la synthèse peptidique. Je voudrais également remercier Dorothée Bernard et Souad Boukoum, pour leur aide sur toutes mes questions administratives. J'ai une pensée particulière pour mes ami(e)s de l'IBCP, qui m'ont soutenue et avec qui j'ai pu partager de très bons moments : Amélie, Agathe, Fanny, Jorgo, Loïck, Charline, Camille, Romuald, et Jackson.

J'aimerais également remercier Thierry Lomberget et son équipe de chimistes, avec qui j'ai collaboré pendant ma thèse. François Hallé et Sylvie Radix m'ont fourni leur aide pour la production de petites molécules, et pour la rédaction de mon article scientifique en tant que première auteure.

Finalement, mes derniers remerciements sont adressés à mes ami(e)s à l'extérieur de l'IBCP, et également à ma famille. Toutes les personnes qui m'entourent m'ont aidé à passer trois belles années en dehors du laboratoire, ce qui m'a permis de décompresser après le travail. Je remercie chacune des personnes qui ont pu être présentes le jour de ma soutenance, pour leur soutien et pour leur écoute. J'ai une pensée particulière pour mes grands-parents, ma grand-mère, mon oncle et ma tante, mon parrain et sa femme, mes parents qui ont été à mes côtés depuis le début, et mes frères qui ont su me rassurer à chaque instant. Je finirais par remercier mon compagnon, Arnaud, qui a été mon pilier et sans qui je n'aurais sûrement jamais pu aller au bout de cette thèse.

Résumé substantiel

De nos jours, même avec l'existence d'un vaccin prophylactique, certaines régions du monde restent particulièrement touchées par les infections au virus de l'hépatite B (VHB). De plus, l'existence de traitements ne permet pas l'éradication complète de la maladie chez les patients déjà infectés. Ainsi, les études visant à mieux comprendre ce virus sont toujours d'actualité. Le cycle viral du VHB est complexe, et plusieurs étapes peuvent être bloquées pour pouvoir stopper la réplication du virus.

Nous nous sommes particulièrement intéressés à la protéine core (Cp) formant la capside du virus, et à ses interactions. Cette protéine est responsable de nombreuses fonctions, que ce soit en termes de base pour les structures des futurs virions, ou encore en tant que régulateur pour certaines étapes du cycle viral. Elle est stable en solution sous forme de dimère, et sa structure comporte plusieurs sites de liaisons potentiels : le spicule, la poche hydrophobe au cœur de la protéine, et la poche de liaison aux modulateurs d'assemblage de la capside (CAMs) à l'interface des dimères. Les deux premières régions sont formées par la dimérisation de deux monomères. Afin de mieux comprendre les mécanismes d'une des étapes les plus importantes du cycle viral, à savoir l'enveloppement de la capside par les protéines d'enveloppe, nous nous sommes particulièrement intéressés aux interactions impliquant la protéine core.

Tout d'abord, nous avons découvert qu'une petite molécule, le Triton X-100 (TX100), se lie dans la poche hydrophobe de la Cp. Le TX100 est un détergent utilisé lors l'étape de purification, et cette liaison, de l'ordre de la dizaine de micromolaire, induit un changement conformationnel détecté en résonance magnétique nucléaire (RMN) du solide, et plus récemment confirmé en cryo-microscopie électronique par une autre équipe. Le blocage de cette poche hydrophobe par un ligand pharmacologique pourrait potentiellement, à l'instar de certains mutants qui ont été testés en cellules, inhiber l'étape d'enveloppement de la capside et donc empêcher la formation de nouveaux virions infectieux. Cette découverte nous a mené à étudier la pharmacomodulation de ce ligand afin de comprendre l'importance de chaque fonction chimique présente sur la structure de cette molécule. Pour cela, nous avons réalisé un screening de molécules synthétisées par des collaborateurs chimistes, en utilisant la RMN en solution sur la Cp dimérique. Nous avons également évalué les paramètres thermodynamiques de ces interactions par titration calorimétrique isotherme (ITC). Finalement, nous avons docké le TX100 et de son homologue dans la poche hydrophobe. Ainsi, nous avons pu conclure sur la nécessité pour la molécule de comporter un cycle aromatique, une chaîne hydrophile de petite taille, et une chaîne aliphatique hydrophobe en position para. Des améliorations sur la structure mériteraient d'être plus approfondies pour

trouver un ligand qui pourraient notamment avoir des interactions multivalentes et une affinité plus grande.

Ensuite, nous nous sommes focalisés sur les interactions au niveau du spicule, et plus particulièrement entre les protéines de l'enveloppe et celle de la capside. Nous avons commencé par étudier des peptides qui inhibent cette interaction, appelés Oct1 et Oct2, et nous avons démontré qu'ils interagissaient spécifiquement au niveau du spicule. Par la suite, nous avons voulu démontrer que le domaine preS de la protéine d'enveloppe L interagissait avec la protéine core tronquée. Par RMN et ITC nous avons observé que preS se liait seulement avec la capside et pas avec le dimère, et que cette interaction était de l'ordre de quelques micromolaires. La technique de la co-expression en système acellulaire a permis de confirmer ces interactions dans des conditions proches des conditions physiologiques en termes de concentration plus proches de celles *in vivo*. Par RMN du solide, nous avons pu ensuite caractériser la zone de liaison entre ces deux protéines, et il semble que ce soit le spicule qui soit le plus impacté par l'interaction avec preS, même si ces résultats sont encore préliminaires à ce stade.

En conclusion, nos résultats apportent des informations nouvelles sur la poche hydrophobe du dimère de la protéine core, nous savons aujourd'hui quel type de structures sont capables d'interagir avec cette dernière, et comment envisager les prochains essais ; sur le spicule du dimère, qui semble interagir avec les peptides Oct, et peut-être partiellement avec les protéines d'enveloppe, et notamment le domaine preS de la protéine L.

Brief overview

Today, even with the existence of a prophylactic vaccine, some regions of the world remain particularly affected by hepatitis B virus infections. Moreover, the existence of treatments does not allow the complete eradication of the disease in patients already infected. Thus, studies to better understand this virus are still ongoing. The HBV viral cycle is complex, and several steps can be blocked to stop the replication of the virus. Several lines of research are currently under investigation.

We are particularly interested in the core protein (Cp) forming the virus capsid, and its interactions. This protein is responsible for many functions, either as a basis for the structures of future virions, or as a regulator for some steps of the viral cycle. It is stable in solution as a dimer, and its structure has several potential binding sites: the spike, the hydrophobic pocket at the core of the protein, and the binding pocket to capsid assembly modulators (CAMs) at the dimer interface. The first two regions are formed by the dimerization of two monomers. In order to better understand the mechanisms of one of the most important steps of the viral cycle, namely the envelopment of the capsid by the envelope proteins, we were particularly interested in the interactions involving the core protein.

First, we found that a small molecule, Triton X-100 (TX100), binds in the hydrophobic pocket of the Cp. TX100 is a detergent used during the purification step, and this binding, of the order of ten micromolar, induces a conformational change detected by solid-state nuclear magnetic resonance (NMR), and more recently confirmed by cryo-electron microscopy by another team. The blocking of this hydrophobic pocket by a pharmacological ligand could potentially, like some mutants that have been tested in cells, inhibit the capsid envelopment step, and thus prevent the formation of new infectious virions. This discovery led us to study the pharmacomodulation of this ligand in order to understand the importance of each chemical function present on the structure of this molecule. For this purpose, we performed a screening of molecules synthesized by chemical collaborators, using solution NMR on the dimeric Cp. We also evaluated the thermodynamic parameters of these interactions by isothermal calorimetric titration (ITC). Finally, we docked TX100 and its homolog in the hydrophobic pocket. Thus, we were able to conclude that the molecule must have an aromatic ring, a small hydrophilic chain, and a hydrophobic aliphatic chain in the para position. Further structural improvements would be required to find a ligand that could have multivalent interactions and a higher affinity.

Next, we focused on the interactions at the spike level, and more particularly between the envelope and capsid proteins. We started by studying peptides that inhibit this interaction, called Oct1 and Oct2, and demonstrated that they interact specifically at the spike. Then, we

wanted to demonstrate that the preS domain of the L envelope protein interacted with the truncated core protein. By NMR and ITC we observed that preS bound only to the capsid and not to the dimer, and that this interaction was of the order of a few micromolar. The technique of co-expression in cell-free system allowed us to confirm these interactions in conditions close to physiological conditions in terms of concentration closer to those *in vivo*. By solid-state NMR, we were then able to characterize the binding area between these two proteins, and it seems that it is the spike that is most impacted by the interaction with preS, even if these results are still preliminary at this stage.

In conclusion, our results bring new information on the hydrophobic pocket of the core protein dimer, we now know what kind of structures are able to interact with it, and how to envisage the next tests; on the spike of the dimer, which seems to interact with the Oct peptides, and maybe partially with the envelope proteins, and in particular the preS domain of the L protein.

Table of contents

List of abbreviations

1D	One dimension
2D	Two dimension
3D	Three dimension
-A-	
aa	Amino acid
ALT	Alanine aminotransferase
ASHV	Arctic squirrels hepatitis virus
AS	Ammonium sulfate
-B-	
BtHB	Bat hepatitis virus
-C-	
CAD	Cytosolic anchorage determinant
CAM	Capsid-assembly modulator
cccDNA	Covalently closed circular DNA
CD	Circular dichroism
CHB	Chronic hepatitis B
CHBV	Cranes hepatitis B virus
Cp	Core Protein
Cryo-EM	Electron cryo-microscopy
CSA	Chemical shift anisotropy
CSD	Chemical shift differences
CSP	Chemical shift perturbations
CTD	C-terminal domain
-D-	
DARR	Dipolar assisted rotational resonance
DDB1	Damage-specific DNA binding protein 1
DHBV	Ducks hepatitis B virus
DP	Differential power
DR	Direct repeat
-E-	
EGFR	Epidermal growth factor receptor
ER	Endoplasmic reticulum
ESCRT	Endosomal sorting complex required for transport
ERGIC	ER-Golgi intermediate compartment
E. coli	*Escherichia coli*

-F-

FID	Free induction decay
FT	Fourier transform

-G-

GSHV	Ground squirrels hepatitis virus

-H-

HBcAg	Hepatitis B ore antigen
HBeAg	Hepatitis B e antigen
HBsAg	Hepatitis B surface antigen
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HDV	Hepatitis D virus
HEV	Hepatitis E virus
HF	Hexylphenol
HHBV	Herons hepatitis B virus
HMQC	Heteronuclear multiple quantum coherence

-I-

ITC	Isothermal Titration Calorimetry

-K-

K_A	Affinity constant
K_D	Dissociation constant
kDa	Kilodaltons
kHz	KiloHertz

-M-

MAS	Magic-angle spinning
MBD	Matrix binding domain
MD	Matrix domain
mRNA	Messenger RNA
MVB	Multivesicular body

-N-

n	Stoichiometry
NA	Nucleos(t)ide analogue
NLS	Nuclear localization sequence
NMR	Nuclear Magnetic Resonance
NTCP	Sodium taurocholate co-transporter polypeptide
NTD	N-terminal domain

-O-

ORF	Open reading frame
OP	4-*tert*-octylphenol

-P-

PEG-IFNα	Pegylated-interferon-α
PEG	Polyethylene glycol
pgRNA	Pregenomic ARN
PHBV	Parrots hepatitis B virus
Pol	Polymerase
ppm	Parts per million
Pre-C/C	Pre-core/core

-R-

rcDNA	Relaxed circular DNA
RGHV	Ross geese hepatitis virus
RT	Reverse Transcriptase

-S-

SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel
SEC	Size exclusion chromatography
SOFAST	Band-selective optimized flip angle short transient
SN-beads	Strep-Tactin beads
ssNMR	Solid-state NMR
ssRNA	Single-stranded linear DNA
ST	Strep-Tag
STHBV	Storks hepatitis B virus
SVP	Subviral particles

-T-

TP	Terminal protein
TX100	Triton X-100

-V-

vRNA	Viral RNA

-W-

WB	Western Blot
WG-CFPS	Wheat germ cell-free protein synthesis
WHV	Woodchucks hepatitis virus
WHO	World health organization
WMHBV	Woolly monkey hepatitis B virus

List of publications

- **A pocket-factor–triggered conformational switch in the hepatitis B virus capsid**
  Lauriane Lecoq, Shishan Wang, Marie Dujardin, Peter Zimmermann, Leonard Schuster, Marie-Laure Fogeron, Mathilde Briday, Maarten Schledorn, Thomas Wiegand, Laura Cole, Roland Montserret, Stéphane Bressanelli, Beat H. Meier, Michael Nassal, and Anja Böckmann
  Proceedings of the National Academy of Sciences
  Volume 118, Issue 17, Pages 1-12
  Published April 20th, 2021

  Contribution: prepared samples, measured ITC, analyzed data.

- **Pharmacomodulation of a ligand targeting the HBV capsid hydrophobic pocket**
  Mathilde Briday, François Hallé, Lauriane Lecoq, Sylvie Radix, Juliette Martin, Roland Montserret, Marie Dujardin, Marie-Laure Fogeron, Michael Nassal, Beat H. Meier, Thierry Lomberget and Anja Böckmann
  Chemical Science
  Volume 13, Issue 30, Pages 8840-8847
  Accepted July 6th, 2022

  Contribution: prepared samples, measured ITC and NMR, analyzed data, wrote the initial manuscript.

- **Molecular elucidation of drug-induced abnormal assemblies of the Hepatitis B Virus capsid protein by solid-state NMR**
  Lauriane Lecoq, Louis Brigandat, Rebecca Huber, Marie-Laure Fogeron, Morgane Callon, Alexander Malär, Shishan Wang, Marie Dujardin, Mathilde Briday, Thomas Wiegand, David Durantel, Dara Burdette, Jan Martin Berke, Beat H. Meier, Michael Nassal, Anja Böckmann
  Nature Communications
  Volume 14, Issue 1, Pages 1-14
  Accepted January 18th, 2023

  Contribution: took EM micrographs, performed ITC, analyzed data.

- **An article on the interaction studies between Cp and preS is in preparation**

  Contribution: Performed sedimentation tests, CD, ITC, and NMR, analyzed data.

INTRODUCTION

CHAPTER 1

I. Introduction on hepatitis B virus

1. Discovery and classification

In 1965, a first serum antigen was discovered by Dr. Blumberg and his team in the blood of an Australian aborigine (Blumberg et al., 1965). At the beginning, the virus was called the "Australia Antigen", since the Australian aborigine's blood sample reacted with an antibody present in the serum of an American hemophiliac patient. For this discovery, in 1976, Dr Blumberg received the Nobel Prize in Medicine. Later, the link between hepatitis B and this serum antigen was confirmed by another study conducted by Dr. Prince's team (Prince, 1968). However, it was not until 1970 the morphology of hepatitis B virus (HBV) particles could be visualized thanks to electron microscopy (Dane & Cameron, 1970). Finally, the French team of Dr. Tiollais published the complete sequence of the virus (Galibert et al., 1979).

After this discovery, other viruses causing jaundice appeared, and the classification of the different human hepatitises has been implemented: the hepatitis C (HCV), D (HDV) and E (HEV) viruses.

Two different forms of hepatitis B can be distinguished (D. R. Milich et al., 1995). First, the acute hepatitis B, of which only 40% is accompanied by symptoms, and which has spontaneous cure in 90-95% of cases. In 5 to 10% of cases, it leads to chronic infection in adults. However, it appears more frequently in children infected early in life or in immunocompromised patients. Then, when acute hepatitis turns into fulminant hepatitis, it complicates about 1% of acute symptomatic hepatitis, leading to an overall mortality of about 90% in the absence of liver transplantation (Ichai & Samuel, 2019). Second, patients may go on to develop chronic hepatitis, which is defined by the persistence of hepatitis B surface antigen (HBsAg) more than six months after infection. Chronic HBV infection can be more or less active, ranging from simple inactive HBsAg carriage without lesions, to chronic active hepatitis. In this case, the virus is responsible for fibrosis that can lead to cirrhosis, the most severe stage of fibrosis. Moreover, numerous health problems can appear, including liver damage, liver failure, liver cancer, and more seriously even death.

2. Genotypes

HBV is a hepatotropic virus belonging to the *Hepadnaviridae* family (Rajoriya et al., 2017) and is classified in Orthohepadnavirus genus (**Figure 1**). This family is composed of enveloped viruses with icosahedral capsids or cores. Members of the Hepadnaviridae such as HBV are pararetroviruses, *i.e.* DNA viruses with an RNA intermediate, but HBV is also a hepatotropic DNA virus that replicates by reverse transcription.

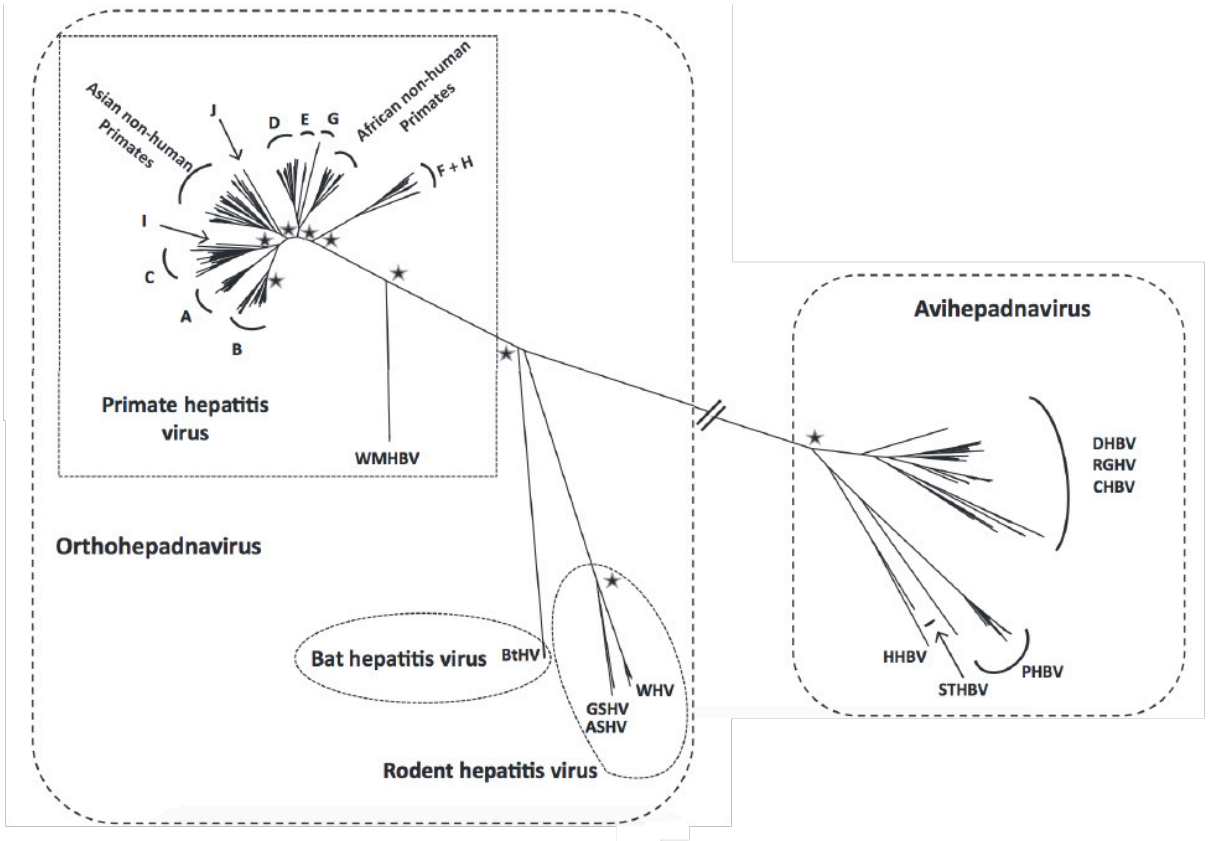

Figure 1. Phylogenetic tree of Hepadnavirus. *On the left side, the Orthohepadnavirus genus is represented, with primate hepatitis virus including A to J HBV genotypes, Asian non-human primates with gibbons and ourangutans, African non-human primates with chimpanzees and gorilla. On this genus there is also woolly monkey hepatitis B virus (WMHBV), bat hepatits virus (BtHB), ground squirrels hepatitis virus (GSHV), arctic squirrels hepatitis virus (ASHV) and woodchucks hepatitis virus (WHV). On the right side, the Avihepadnavirus genus is shown, with ducks hepatitis B virus (DHBV), ross geese hepatitis virus (RGHV), cranes hepatitis B virus (CHBV), herons hepatitis B virus (HHBV), storks hepatitis B virus (STHBV), and parrots hepatitis B virus (PHBV). Taken from (Locarnini et al., 2013)*

HBV can be classified into 10 main genotypes, from A to J, each of them having also sub-genotypes (**Figure 2**). All the genotypes differ by more than 8% at the nucleotide level (Ho et al., 2020), and 4 to 8% nucleotide differences for sub-genotypes. Their variability influences both the disease (effects) and also the treatments efficiency.

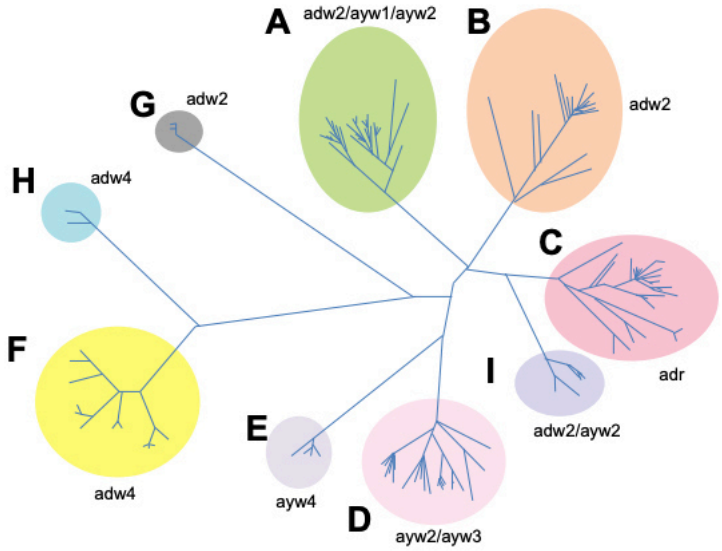

Figure 2. HBV genotypes tree. *The 9 major genotypes are represented by a specific color, with their specific sub-genotypes (ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw4, adwq, adrq+ and adrq-) based on HBsAg heterogeneity. Taken from (Rajoriya et al., 2017)*

3. Prevalence

Health services release updated hepatitis surveillance data showing a steady increase for HBV again in 2019. Indeed, HBV is an infectious disease that is widespread worldwide (**Figure 3**). Approximately 375 million chronic HBV carriers are estimated in the whole world. In 2016, only 10.5% of the estimated total population of people living with hepatitis B were aware of their infection, and only 16.7% of those diagnosed were on treatment.

A distinction is made between the different regions of infection. First, Central and South Africa and Western Asia, which are the regions where the population with chronic HBV is the highest (more than 8%). Second, the Mediterranean basin, North Africa, Russia, and some countries of South America are regions where the HBV carriers represent 2 to 5% of the total population. Finally, Europe, India, Australia and the North America are regions corresponding to low prevalence, with less than 2% of infected people.

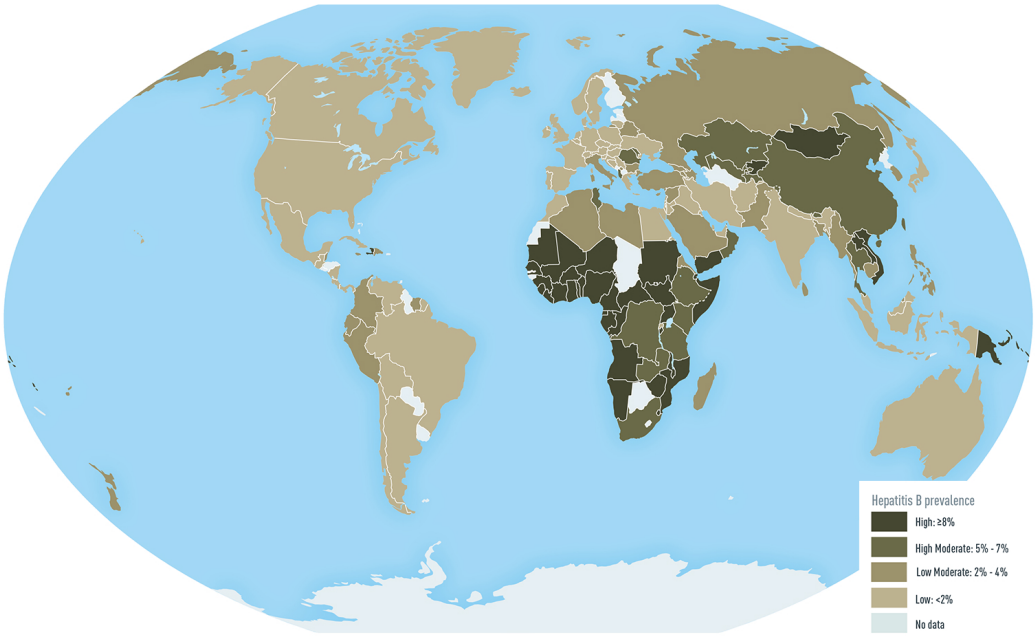

Figure 3. Prevalence map of HBV in the world. *Each regions are colored depending on their specific prevalence (from dark green for high prevalence, to light blue for low prevalence).Taken from (Schweitzer et al., 2015)*

Currently, the chronic hepatitis B (CHB) represents one of the deadliest diseases, causing nearly one million deaths per year, mainly in Africa (https://www.afro.who.int/health-topics/hepatitis).

4. Transmission and prevention

The transmission of HBV is possible through blood or other body fluids. Three main modes of transmission are distinguishable. The first mode is by blood transmission during transfusion of blood or blood products, by the use of contaminated medical equipment during care, injection drug use, and tattooing or piercing. The second mode is by sexual transmission through mucous membranes with an infected person, and finally by vertical transmission through transplacental passage from mother to child or at the time of delivery through blood contamination.

The first generation of HBV vaccine has been implemented in 1976 (Maupas et al., 1976), and recombinant vaccines were available from 1986, which are based on HBsAg gene inserted into yeast (**Table 1**). All hepatitis B vaccines that have been used since this date are made synthetically, therefore the hepatitis B vaccines do not contain any blood products or live virus, which is why you cannot get hepatitis B from the vaccine. The hepatitis B vaccine is recommended for all infants at birth and for children up to 18 years, as it is a safe and effective. The vaccine is also recommended for adults living with diabetes and those at high risk for infection. The hepatitis B vaccine is also known as the first “anti-cancer” vaccine because it prevents hepatitis B, the leading cause of liver cancer worldwide.

Table 1. List of recombinant HBV vaccines commercially available worldwide. *Taken from (Roldão et al., 2010)*

Trade name	Manufacturer	Country	Recombinant protein	Expression host
DTP-Hep B	P.T. Bio Farma	Indonesia	HBsAg S protein	Yeast (*Pichia pastoris*)
Engerix-B®	GlaxoSmithKline	Belgium	HBsAg S protein	Yeast (*Saccharomyces cerevisiae*)
Enivac HB	Panacea Biotec Ltd.	India	HBsAg S protein	Yeast (*P. pastoris*)
Euvax B	LG Life Sciences	South Korea	HBsAg S protein	Yeast (*S. cerevisiae*)
Gene Vac-B®	Serum Institute of India Ltd.	India	HBsAg S protein	Yeast (*Hansenula polymorpha*)
GenHevac B®	Pasteur-Mérieux Aventis	France	HBsAg S and M protein	Mammalian cells (CHO)
Heberbiovac HB	CIGB – Heber Biotec	Cuba	HBsAg S protein	Yeast (*P. pastoris*)
Hepavax-Gene®	Crucell	The Netherlands	HBsAg S protein	Yeast (*H. polymorpha*)
Recombivax HB®	Merck and Co., Inc.	USA	HBsAg S protein	Yeast (*S. cerevisiae*)
Revac-B+™	Bharat Biotech International Ltd.	India	HBsAg S protein	Yeast (*P. pastoris*)
Sci-B-Vac™	SciGen	Israel	HBsAg S, M and L protein	Mammalian cells (CHO)
Shanvac™-B	Shantha Biotechnics Ltd.	India	HBsAg S protein	Yeast (*P. pastoris*)

5. Treatments

On the World Health Organization (WHO) website, it is mentioned that there is currently no specific treatment for acute hepatitis B. However, there is care aimed at preserving the comfort of the patient.

In many cases, specific medication is indicated for HBV infection (Lok & McMahon, 2009). Currently, numerous molecules are commercially available or are under development (**Figure 4**). These drugs may modulate the immune system or target HBV more directly in order to stop the infection.

One of the earliest treatment options for HBV was the cytokine interferon. Today, the use of interferon is still common, as it has reduced renal clearance (Veronese & Harris, 2002). Current treatments for CHB include pegylated-interferon-α (Peg-IFNα) which is used as a first line treatment, and nucleos(t)ide analogues (NAs). However, neither of these treatments are sufficiently efficient for functional cure. Indeed, undetectable levels of HBsAg and HBV DNA in the serum, normality of alanine aminotransferase (ALT) and development of hepatitis B surface antibody (anti-HBs) (Bertoletti & Bert, 2018) are not achievable with these therapeutics.

Currently, the available antiviral therapies allow to control viral replication and to limit the progression to cirrhosis. However, these treatments require administration during the whole life of the patient. This is due to the frequent viral rebound upon treatment cessation, but also to the fact that immune modulation with interferon is only effective in a subgroup of patients who do not suffer from side-effects (Bertoletti & Bert, 2018).

HBV is often reactivated after stopping NAs because the exclusive use of antivirals do not directly target covalently closed circular DNA (cccDNA), which is the template for all viral

RNAs (H.-C. Yang & Kao, 2014). If the cccDNA, known as the persistent nuclear form of the viral genome, is not affected, it will resume to give rise to progeny virions.

Finally, the effectiveness of current antiviral treatments on the suppression of HBV replication has been observed, but even with time, they do not allow HBsAg loss and seroconversion (HBsAg loss of less than 10% in 5 years) (Lee et al., 2020).

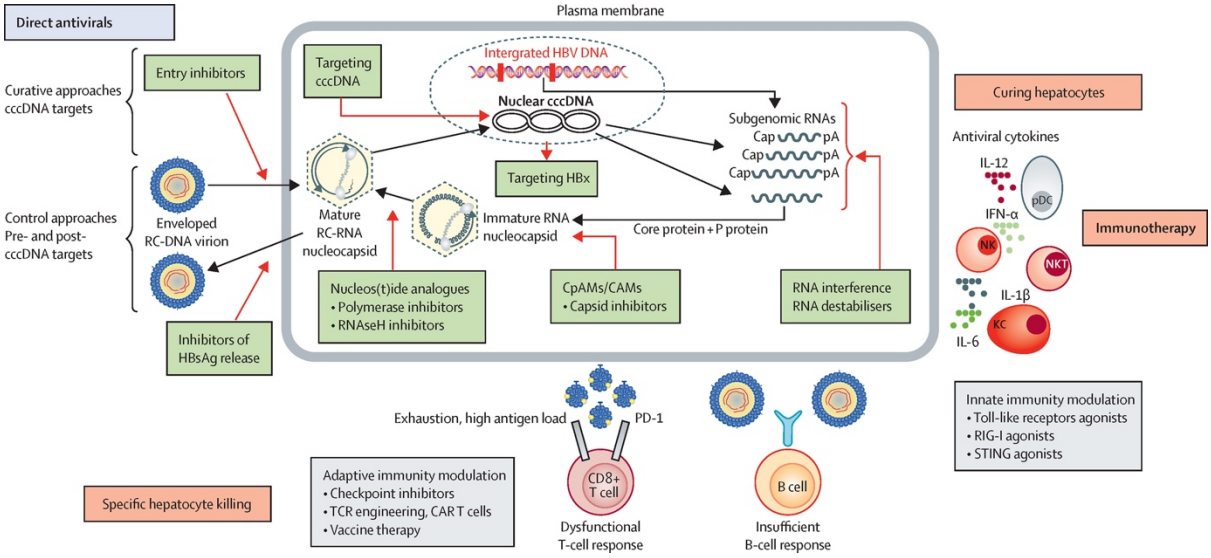

Figure 4. Summary of all treatments for HBV cure. *This figure shows the different perspectives of new treatments. Taken from (Revill et al., 2019)*

New efforts are directed on capsid assembly modulators (CAMs), as HBV capsid represents a more realistic target to achieve (Taverniti et al., 2022). As their name suggests, their aim is to prevent nucleocapsid assembly, by interacting at the interdimer interface (**Figure 5**). These small molecules have the ability to interfere with the kinetics of assembly and/or the correct interactions between core dimers. CAMs are different from NAs as they can have a blocking impact on HBV RNA encapsidation, viral replication, and production of HBV RNA/DNA-containing particles. CAMs mechanism of action differs from standard therapies, and by interfering in several steps of the viral life cycle of HBV (Lam et al., 2017), they may enhance viral suppression and can potentially be combined with other treatments. To date, several CAMs have entered clinical development (Taverniti et al., 2022).

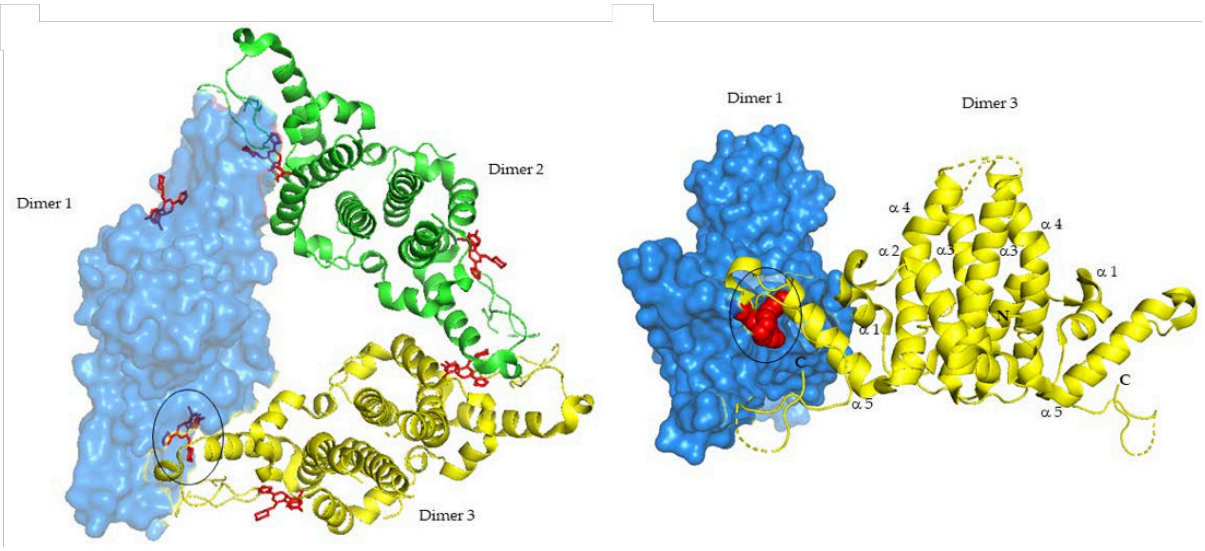

Figure 5. Interaction site of CAMs, with the example of NVR-010-001-E2. *Core dimers are colored in blue, green and tallow, and the CAM in red. Taken from (Taverniti et al., 2022)*

II. Viral life cycle

The HBV multiplication cycle occurs in the cytoplasm and in the nucleus of liver cells. The different stages will be detailed in this part (**Figure 6**).

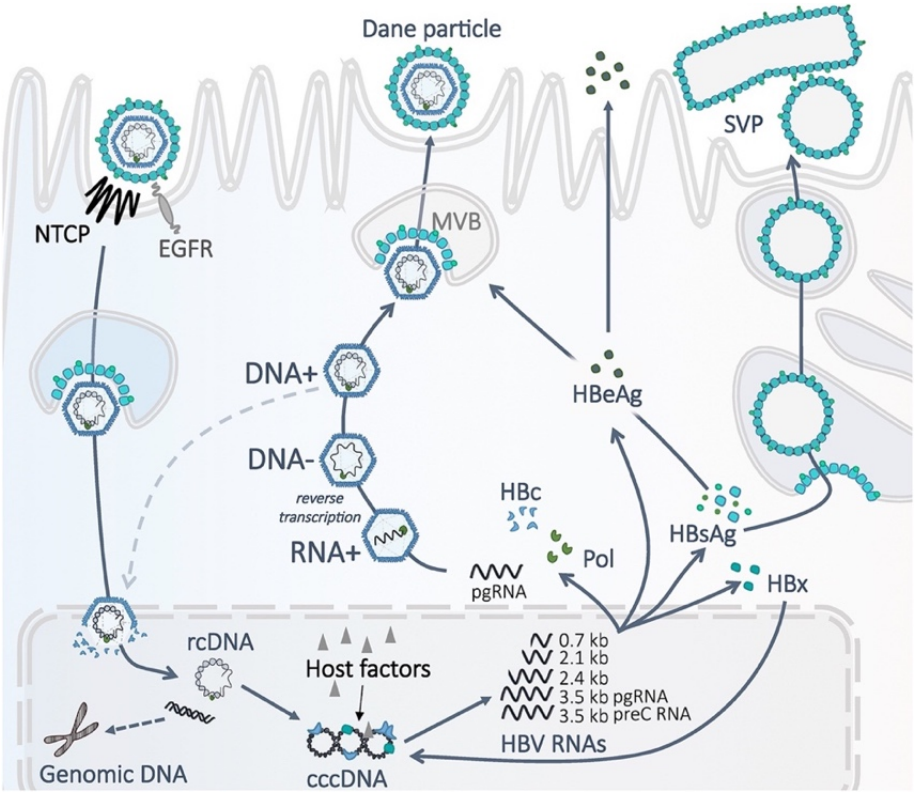

Figure 6. Schematic overview of HBV viral life cycle. *EGFR: epidermal growth factor receptor; NTCP: sodium taurocholate co-transporting polypeptide; rcDNA: relaxed circular DNA; cccDNA: covalently closed circular DNA; pgRNA: pregenomic RNA; ssDNA: single-stranded linear DNA; Cp: core protein; Pol: polymerase; ER: endoplasmic reticulum. Taken from (Tsukuda & Watashi, 2020)*

1. Life cycle steps

The entry of HBV into host hepatocytes is a regulated step. First, endocytic entry of the virion takes place during the non-specific interaction of the virus with the surface glycoproteins, and then by a specific binding with the cell receptor, NTCP (Leistner et al., 2007; Schulze et al., 2007; Verrier et al., 2016; Yan et al., 2012). A study recently supported that an EGFR, a receptor tyrosine kinase, allows the internalization of HBV by its direct interaction with NTCP (Iwamoto et al., 2020).

After the interaction of the virion with the surface of host cells, endocytose is taking place in order to release the nucleocapsid from the endosome to the cytoplasm. The virus is directed to the nucleus of hepatocytes along with the microtubules (Rabe et al., 2006). The virus core binds to the nuclear pore complex in an importin-dependent manner, and is imported to the nucleus (Kann et al., 1999; Rabe et al., 2006). Then, the relaxed-circular DNA (rcDNA) is released into the nucleus, it is repaired by host factors (Wei & Ploss, 2021) and finally forms the covalently closed circular DNA (cccDNA).

From this point on, the HBV viral RNAs (vRNAs) are transcribed from the cccDNA by host Pol2 and are exported from the nucleus to the cytoplasm. On one hand, the viral Cp and viral polymerase are subsequently translated from a full-length messenger RNA (mRNA), the pre-genomic RNA (pgRNA). The complex made by the polymerase and the pgRNA signals the RNA encapsidation by core proteins (Cps), and this allows the immature capsid to be formed. On the other hand, the envelope proteins are translated and transferred from the ER to the multivesicular bodies (MVBs). MVBs have the unique ability to generate intraluminal vesicles that bud away from the cytosol, a process topologically equivalent to that of enveloped virus budding. The formation of new virions is then possible when the nucleocapsid freshly formed after reverse transcription and digestion of linear pgRNA, is interacting with envelope proteins. The particles are finally secreted out of the cells (virus egress) or transported back to the nucleus for multiplication of the virus by maintaining a stable pool of cccDNA (Tsukuda & Watashi, 2020).

2. Viral particles secretion

Four different categories of viral particles can be identified in the serum of an infected patients. First, the whole virions which are infectious particles, also called Dane particles (**Figure 7**), circulate in the blood of infected subjects at a high concentration. Second, immature enveloped nucleocapsids can be secreted, they can either be empty or also contain immature DNA/RNA. Their presence is due to specific mutations on core protein, which accelerates the secretion of these particles even if they are not mature. Their role in HBV infection is still unknow and remains to be determined. Third, the subviral particles (SVPs) can also be secreted from host cells, that are only composed of envelope proteins and lack the capsid. They are therefore non-infectious, they can take the form of either spherical or filamentous particles (**Figure 7**) and are 100,000 times more concentrated than Dane particles (Bruss, 2007; Chai et al., 2008). The main function of these particles is to block the neutralizing effect of anti-HBV antibodies (Ho et al., 2020). Finally, HBV-replicating cells can also release

capsids which are not enveloped, called the ‘naked nucleocapsids’. This phenomenon can appear when the interaction between the capsid and the envelope proteins is defective. Except from Dane particles, all others are non-infectious.

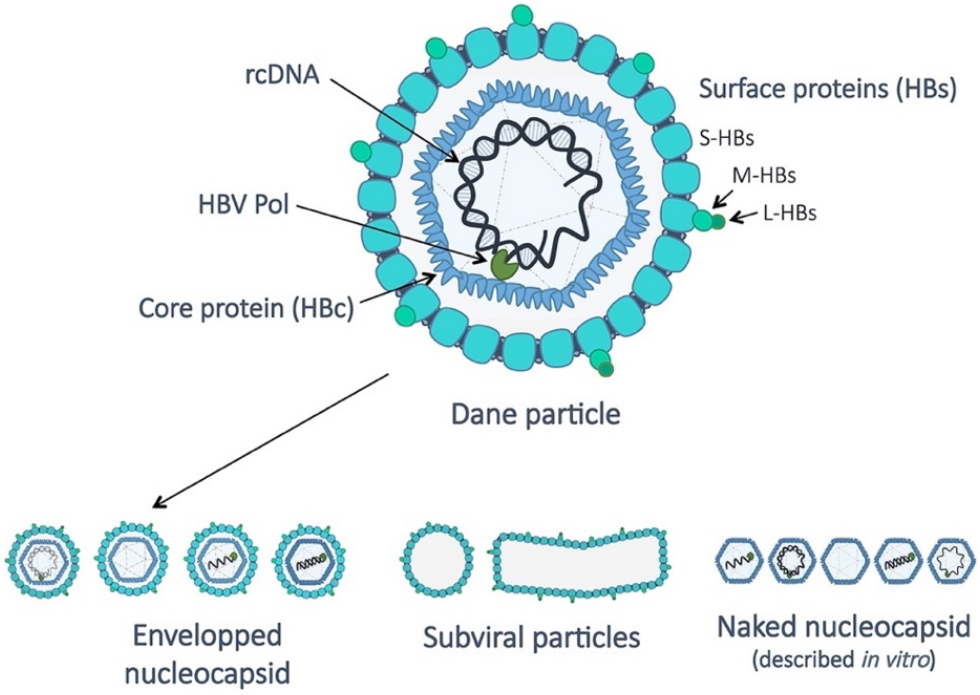

Figure 7. Composition of HBV secreted particles. *Dane particle corresponds to mature and enveloped virion, and the rest are abnormal particles. Taken from (Tsukuda & Watashi, 2020)*

III. Hepatitis B virus proteins

As introduced in the previous part, HBV infectious particles are composed of a capsid enclosing the polymerase and viral DNA, themself enveloped by the layer formed by lipids and surface proteins.

1. Genome

Hepatitis B virus is a small virus, with a genome made of only four overlapping open reading frames (ORFs) (**Figure 8**). Theses ORFs encode the polymerase (Pol), the core protein (HBc), the surface glycoproteins and X genes.

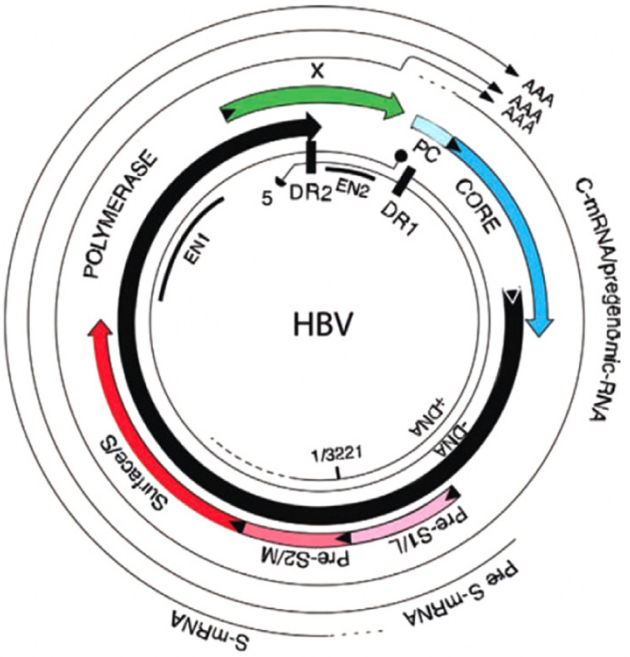

Figure 8. Representation of the HBV genome. *The longest ORF encodes the viral polymerase (Pol). The Pol ORF is fully overlapped by the envelope ORF (pre-S1, pre-S2 and S), whilst the precore/core (pre-C/C), and the X ORFs overlap partially with it. Taken from (Ho et al., 2020)*

The HBV genome, also known as rcDNA, consists of a partially double-stranded circular DNA molecule, with 3/4 of its circumference which is double-stranded. The complete strand has a negative (-) polarity, while the incomplete strand has a positive (+) polarity (Wong & Locarnini, 2018). This genome is made of 3182 to 3248 base-pairs, the number depending on the genotype (Locarnini et al., 2013). The minus-strand DNA encodes the complete viral genome and is covalently attached to the 5' ends of the viral polymerase (Wong & Locarnini, 2018).

2. Polymerase

The polymerase is the largest protein of HBV, and contains between 832 amino acids (aa) with a molecular weight of 90 kDa (Wong & Locarnini, 2018). This four-domain multifunctional enzyme is made of both a DNA-dependent DNA polymerase and an RNA-dependent DNA polymerase (reverse transcriptase) activities. HBV polymerase contains several important domains and motifs which define its functions: a terminal protein (TP) domain, a spacer domain, a reverse transcriptase (RT) domain and RNase H domain (**Figure 9**).

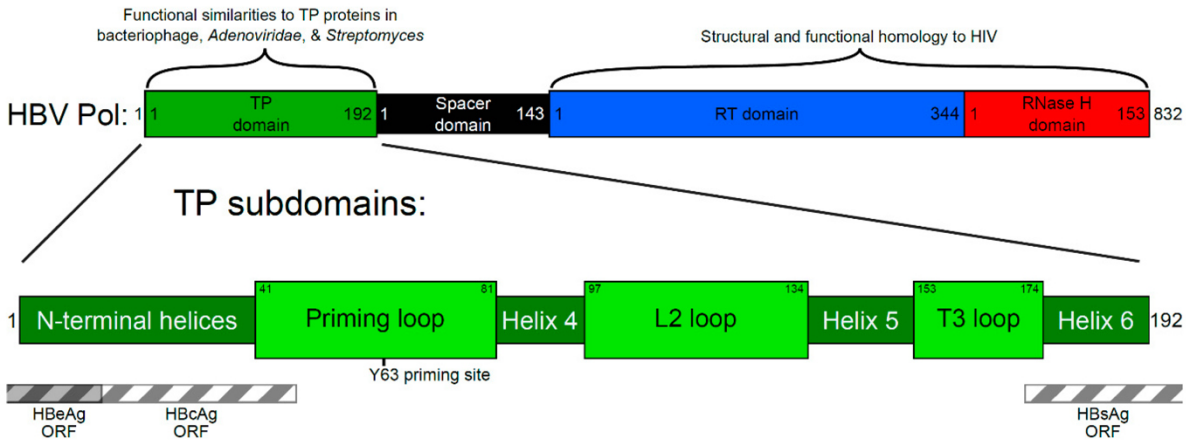

Figure 9. Schematic representation of HBV Pol domains and terminal protein subdomains. *Taken from (Buhlig et al., 2020)*

The main function of the polymerase is to replicate the HBV genome from an encapsidated pgRNA template, but is also responsible for viral RNA binding, RNA packaging, protein priming, template switching, DNA synthesis and RNA degradation (Clark & Hu, 2015).

It has to be known that HBV DNA Pol makes many errors, and these mistakes cannot be corrected because Pol lacks corrective 3'-5' exonuclease activity (proofreading activity). Therefore, nucleotide substitutions accumulate on the genome during successive replication cycles and are the source of HBV genetic variability.

3. HBV replication

To generate cccDNA, rcDNA is repaired by removal of Pol, completion of the (+) strand, removal of 5′ terminal structures (terminal protein in the (-) strand, oligoribonucleotide primer in the (+) strand), and ligation of the respective strands (Seeger & Mason, 2000). The full mechanism of rcDNA conversion into cccDNA is not completely understood but is still under investigation (Wei & Ploss, 2021). However, formation of cccDNA is one of the key steps for the establishment of infection. Indeed, negative strand nucleotide sequence of cccDNA serves as the template for transcription of genomic and subgenomic vRNAs, which will then be translated into viral proteins.

The viral RNAs transcribed from cccDNA include pgRNA, which serves as the template for reverse transcription (Seeger & Mason, 2000). When combined with the polymerase, also called viral reverse transcriptase, it forms the replication complex, packaged in the immature capsid (**Figure 10**). The Pol interacts with the encapsidation signal ε located next to the 5’ end of the pgRNA (Bartenschlager & Schaller, 1992), and then initiates their co-encapsidation. After that, protein-primed reverse transcription is underway. The 3’ nucleotide of the ε bulge and/or the first nucleotide of the upper stem serves as a template for the first DNA nucleotide covalent addition to a tyrosine residue in the terminal protein domain of the Pol. Then, the (-)-strand is reverse-transcribed from the 3’ end to the 5’ end of pgRNA (Wong & Locarnini, 2018). The template of pgRNA is degraded by the RNAse H domain of the Pol, while the (-)-strand is synthesized. However the 5’ capped region (including direct repeat 1 (DR1)) is kept and

translocated next to the DR2 of the newly (-)-strand (Tsukuda & Watashi, 2020). This allows to start the synthesis of the (+)-strand. The rcDNA is finally formed when the 50 to 70% of the (+)-strand is synthesized (Wong & Locarnini, 2018).

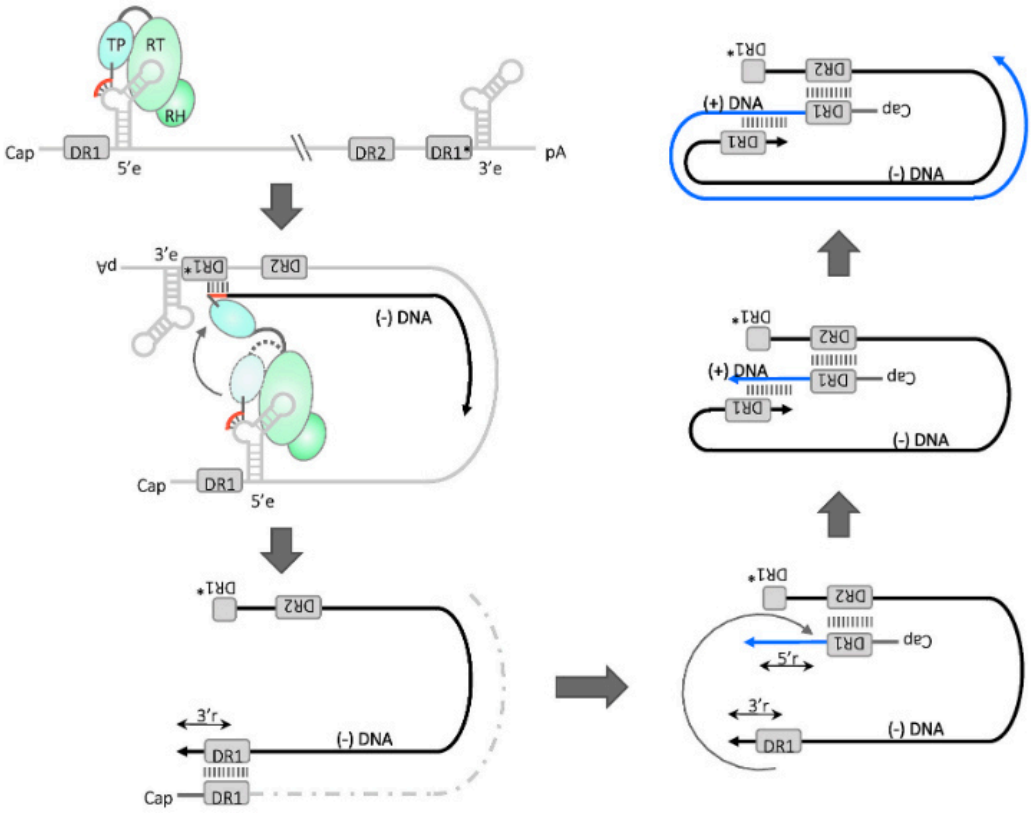

Figure 10. Scheme of the HBV replication. *The Pol is shown by TP, RT, RH complex. DR: direct repeat. Taken from (Tsukuda & Watashi, 2020)*

4. HBe protein

From the genome of HBV, the core gene encodes the core protein but also the secreted nonparticulate e antigen (HBeAg), meaning that HBeAg and core antigen (HBcAg) are made from the same open reading frame. The sequence of HBeAg is similar to the one of HBcAg; indeed they share the same core domain but differ in that HBeAg retains the N-terminal propeptide and lacks the RNA-binding arginine-rich C-terminal domain (CTD) (**Figure 11A**). This protein is expressed in the hepatocyte cytosol, but is also secreted in a soluble form in the serum (Revill et al., 2019). HBeAg has immunomodulatory properties, and particularly it decreases the immune response to HBcAg (Chen et al., 2004; D. Milich, 2003; D. R. Milich et al., 1990, 1998). In this way, immunomodulation mediated by HBeAg may predispose to chronicity and persistence after intrauterine or perinatal infection and prevent severe liver injury during infection in adults (D. Milich, 2003). Clinically, HBeAg is an indicator of active viral replication, infectivity, inflammation, severity of disease and response to antiviral therapy.

As HBeAg is produced from the core gene, its structure is similar to the one of HBcAg, with two monomers assembled into a dimer (**Figure 11B and C**) (DiMattia et al., 2013).

However, the HBeAg dimer interface is much smaller than that of HBcAg (Lawrence & Colman, 1993).

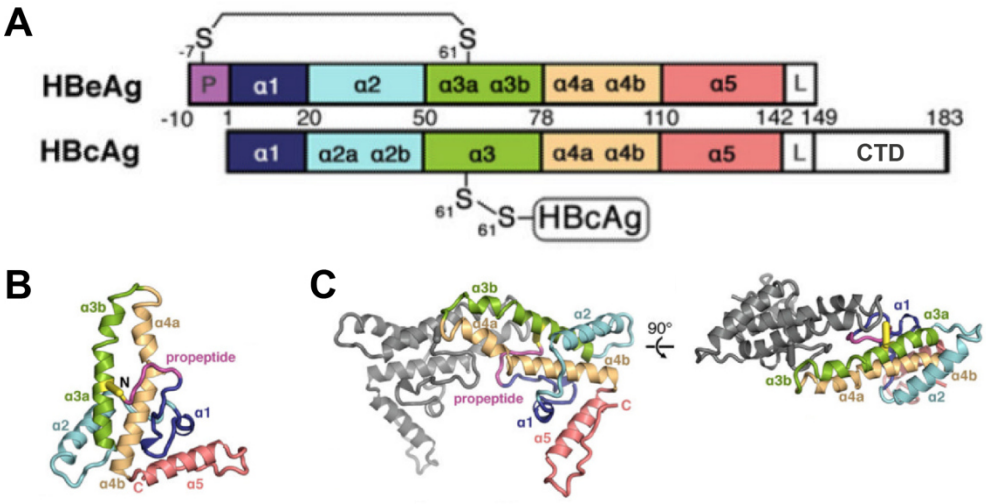

Figure 11. HBeAg sequence (A) and monomer/dimer structure (B and C). *All helices are specifically colored. P: propeptide; L: linker. Taken from (DiMattia et al., 2013)*

5. HBx protein

HBx protein is encoded by the shortest ORF and mRNA transcript of HBV genome (Bouchard & Schneider, 2004; Rawat & Bouchard, 2015). HBx is a 154-amino acid (aa) protein and is a transcriptional transactivator (Bouchard & Schneider, 2004). Indeed, HBx is required for efficient transcription of cccDNA, and so this protein plays a role in HBV replication (Seeger & Mason, 2015). It has been proved that HBx interacts with host proteins and facilitates HBV replication, but the mechanism of action of this protein has not yet been understood. (Keasler et al., 2007; Tan et al., 2018).

HBx has regulatory effects on cellular signaling pathways, DNA damage repair mechanisms, cell cycle and apoptosis, and so this protein is believed to be pontentially responsible for the development of HBV-associated hepatocellular carcinoma (HCC) (Clippinger et al., 2009; Leupin et al., 2005; Tang et al., 2005). Indeed, HBV transcription can be altered by the structural maintenance of chromosomes complex (Smc5/6). Smc5/6 binds to cccDNA and thus inhibits HBV replication, but HBx can interact with the cellular damage-specific DNA binding protein 1 (DDB1) to reset normal viral transcription (Decorsière et al., 2016; Murphy et al., 2016; Sekiba et al., 2022). Therefore, HBx is a potential target for anti-liver cancer molecules, but its structure is still unknown (Pathak et al., 2014).

6. Core protein

a. Sequence

The full-length core protein sequence is made of 183 aa and can be divided in three different parts: the α helix-rich assembly domain (aa 1 to 139), the linker (aa 140 to 149) and the arginine-rich CTD (aa 150 to 183) (**Figure 12**) (Venkatakrishnan & Zlotnick, 2016; Zlotnick et al., 2015).

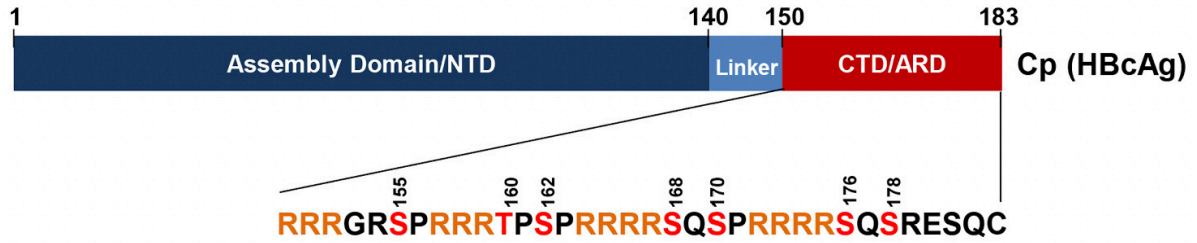

Figure 12. Sequence of the core protein. *The CTD sequence is detailed with the arginine residues in orange, and phosphorylation sites in red with their specific position on the sequence. NTD : N-terminal domain; CTD : C-terminal domain; ARD : arginine-rich domain. Taken from (Viswanathan et al., 2020)*

b. Dimerization

The core protein is a 20 kDa monomer, which is most likely unstable on its own, therefore it forms a dimer in solution (Wingfield et al., 1995). This dimerization occurs thanks to the interaction between the hydrophobic regions of two monomers, which corresponds to an interface formed by helices α3 and α4 of each monomer (Böttcher, 2021). The Cp dimerization can be further driven and stabilized by the formation of disulfide bonds. Each monomer contains a conserved cysteine at position 61 that can form an intradimer disulfide bridge (**Figure 13**), but it has been shown that it was not essential for dimer formation (Nassal et al., 1992). Another intermolecular disulfide bridge can be formed by the two cysteines 48 of each monomer, but only in high concentration of oxidant, and this bridge is also dispensable for dimerization (Nassal et al., 1992). Indeed, cysteine 48 is not located at the dimer interface and the distance between the two residues of each monomer is too important to allow the disulfide bond formation in the native shell (**Figure 13**) (Wynne et al., 1999).

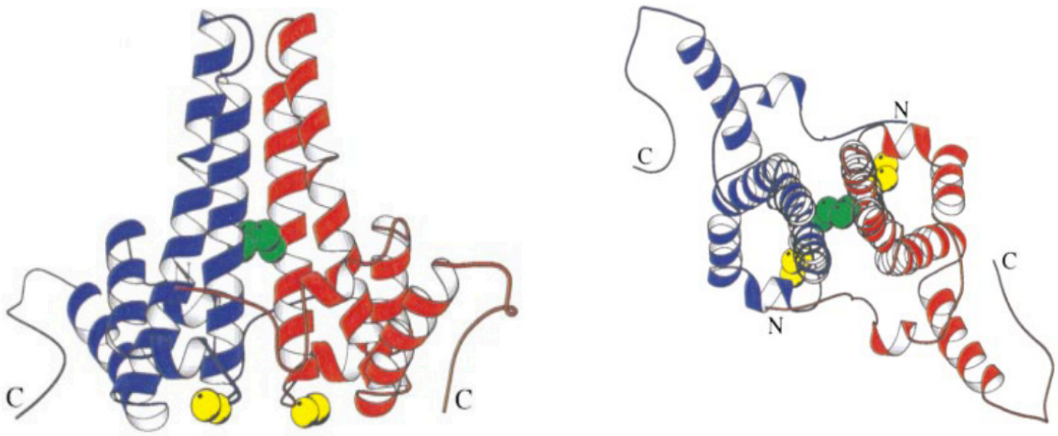

Figure 13. HBV core protein dimer (PDB 1QGT). *Two orthogonal views of the Cp dimer are represented, on the left with a front view, and on the right from the top of the dimer. Each monomer is colored, as well as Cys48 in yellow and Cys61 in green, which forms a disulfide bridge between the two monomers. Taken from (Wynne et al., 1999)*

c. Structure

The structure of HBc assembly domain has been studied by numerous studies using electron cryo-microscopy (cryo-EM) (Böttcher et al., 1997; Böttcher & Nassal, 2018; Conway et al., 1997; Crowther, 1994; Schlicksup et al., 2018) and X-ray crystallography (Alexander et al., 2013; Venkatakrishnan & Zlotnick, 2016; Wynne et al., 1999). Many structures only resolve the assembly domain up to aa 144 and in some exceptional cases up to aa 152 (Böttcher & Nassal, 2018). However, the density upstream of residue 144 is much weaker than other residues, indicating more flexibility or occupancy, and some of the C-terminal different conformations. This small sequence between the assembly domain and the CTD is the linker-region, and is involved in several steps of HBV replication, including virion secretion (Liu et al., 2018).

The HBV assembly domain is mainly α helical, which differs from the other coat proteins of icosahedral viruses that normally have a jellyroll or β-barrel structure (Wynne et al., 1999). HBV Cp is made of five α helices interconnected by loops (**Figure 14**): α1 helix (aa 13 to 17) which is the shortest one, α2 (aa 27 to 43), α5 (aa 112 to 127), while α3 (aa 50 to 73) and α4 helices (aa 79 to 110) which are the longest ones and form together the spike of the capsid.

The basic structural and functional unit of Cp is the dimeric form of the protein. Indeed, the stability of the dimer is possible by the hydrophobic interaction between α3 and α4 helices of two Cp monomers, which allows the formation of a stable four-helix bundle at their interface (Böttcher et al., 1997; Conway et al., 1997; Wynne et al., 1999).

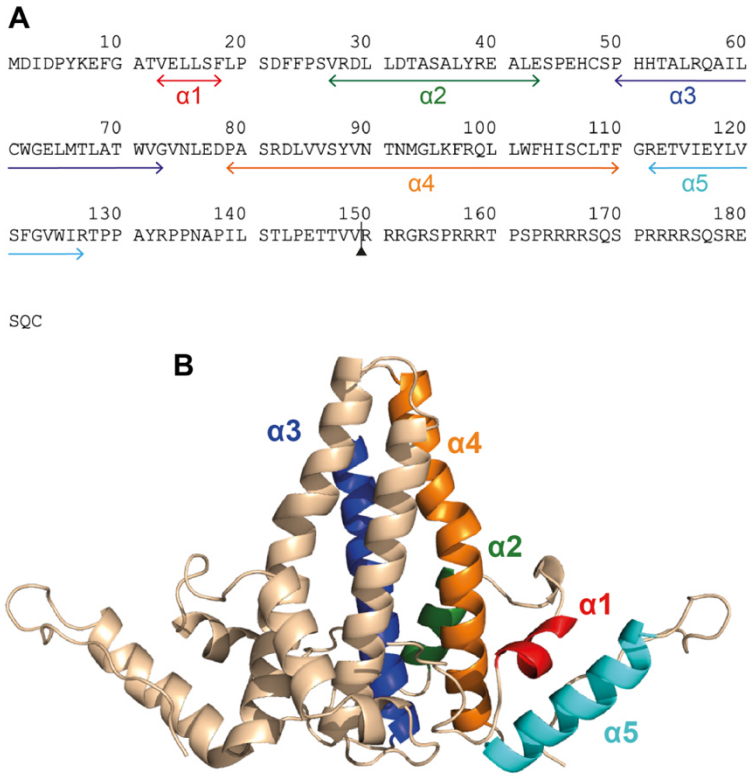

Figure 14. HBV core protein sequence and dimer α helical structure (PDB 1QGT). *A) Amino acid sequence. B) Dimer structure in cartoon representation. Each helix is positioned on the sequence, specifically colored on the structure.*

d. Functions

Cp has multiple many roles and is involved in many steps in the HBV life cycle. Its most important role is capsid assembly as it is the building block; but it also modulates virtually all other steps of the viral life cycle (Zlotnick et al., 2015). First, Cp binds to the newly formed cccDNA to participate in the formation of the viral minichromosome, and so Cp affects the epigenetics of nuclear viral DNA (Bock et al., 2001). Then, Cp is implicated in the regulation of some cellular functions. It binds to the promoters of many cellular genes and could inhibit their expression (Du et al., 2009; Guo et al., 2012; Kwon & Rho, 2003), and interacts with RNA export factors (C.-C. Yang et al., 2014). Moreover, Cp participates in regulation of reverse transcription. In fact, the nucleic acid chaperone activity of Cp helps maintain proper folding of pgRNA and its reverse transcription into rcDNA (Diab et al., 2018). Finally, Cp carries both nuclear localization signals (NLS) and HBsAg binding sites, respectively necessary for transfer between cytoplasm and nucleus, and for infectivity. In addition to all the structural functions of the core protein, it also has nucleic acid chaperone activity, through the basic residues of its CTD. Cp has a regulatory power on the replication of the viral genome, and more particularly on its structural rearrangement (Chu et al., 2014). For all these reasons, Cp is central in many studies and is one of the targets of HBV treatments.

7. Envelope proteins

a. *The three surface proteins*

The surface gene of HBV produce the three different envelope proteins: the small (S), middle (M) and large (L) HBs (**Figure 15**). They are all translated from 2.4/2.1 kb mRNAs, transcribed from the initial cccDNA. These proteins have a common C-terminus, but they differ from the N-terminal domain. The respective length of HBs S, HBs M and HBs L are 226, 281 and 400 aa. HBs M is longer from the HBs S as it has a supplementary preS2 domain, and HBs L has two supplementary preS1 and preS2 domains (Inoue et al., 2021). Moreover, HBs L is myristoylated at the N-terminus, a condition essential for infection (Gripon et al., 1995).

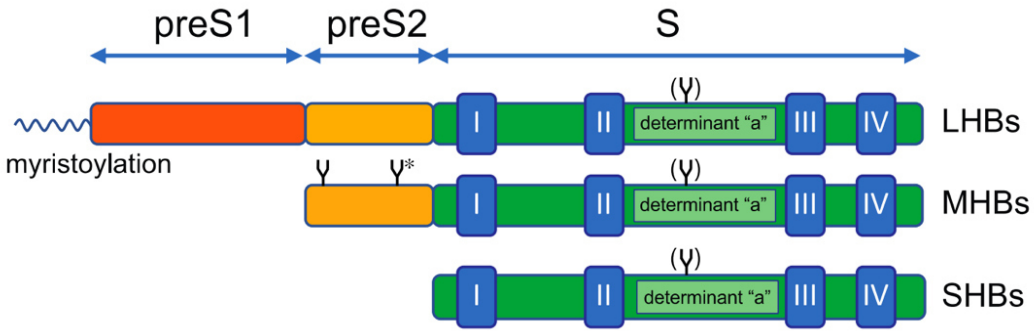

Figure 15. Structure of the three HBV envelope proteins. *Each domain is colored: blue for transmembrane domains, green for each inter transmembrane domains, orange for preS2 and red for preS1. Taken from (Inoue et al., 2021)*

For all envelope proteins, 4 transmembrane domains are present on the HBs S sequence, and an additional determinant "a". This determinant "a" is an antigenic domain that is a major target of neutralizing antibodies, as it is exposed on the outside of the envelope. Moreover, the common S-domain harbors an N-glycosylation site, which is partially used in all three surface proteins. When fully synthesized, all of them sharing the same S-domain, and thanks to that they can be anchored to the ER membrane. Finally, at the end of the viral cycle, the three envelope proteins are transported from ER membranes to the MVBs to form the envelope. It has been demonstrated that the ER retention reaction is not required. Indeed, mutations on HBs L, leading to proteins that cannot retain HBsAg particles in the ER anymore, still allow virion formation. MVBs are considered to participate in the formation of the HBV envelope. Indeed, they interact with endosomal sorting complex required for transport (ESCRT) which is necessary for virion release (Böttcher, 2021).

Importantly, M proteins are not necessary for virion production, whereas S and L proteins are required (Bruss & Ganem, 1991).

b. *HBs L and preS domain*

The preS domain is formed by both hydrophilic preS1 and preS2 domains, and it has many functions along its sequence (**Figure 16**). Amino acids 2–48 of the pre-S1 sequence specifically interact with the high-affinity entry receptor and NTCP. Amino acids 21-47 of the preS1 domain likely bear the major epitope for cell attachment (Cooper et al., 2003). Residues 49–75 are also required for infection, but the precise function of this region is not known (Urban

et al., 2014). Residues 96-116, located on both preS1 and preS2 domains, are involved in the interaction with nucleocapsids for their envelopment (Bruss, 1997; Poisson et al., 1997).

```
         10          20          30          40          50          60
MGQNLSTSNP LGFFPDHQLD PAFRANTANP DWDFNPNKDT WPDANKVGAG AFGLGFTPPH

         70          80          90         100         110         120
GGLLGWSPQA QGILQTLPAN PPPASTNRQS GRQPTPLSPP LRNTHPQAMQ WNSTTFHQTL

        130         140         150         160
QDPRVRGLYF PAGGSSSGTV NPVLTTASPL SSIFSRIGDP ALN
```

Figure 16. preS amino acid sequence.

The myristic acid of HBs L is essential to enable the virus particle to interact with the cellular plasma membrane (Gripon et al., 1995; Urban et al., 2014), but also allows the preS1 and preS2 domains to transit between the cytoplasmic side and the luminal side (Prange & Streeck, 1995). This property allows preS1 to switch from the cytoplasmic side when it is required for the envelopment of the capsid, to the luminal side when it is exposed on the viral surface of mature viral particles and so essential for NTCP interaction for the infection (**Figure 17**) (Yan et al., 2012). Indeed, HBs L is necessary for infection and viral morphogenesis, as the pre-S1 region encodes the entry receptor-binding domain (Urban et al., 2014). L protein is also involved in the interaction with Hsc70 chaperone. Hsc70 selectively binds to cytosolic anchorage determinant (CAD) and may contribute to virion formation by stabilizing the cytosolic configuration of preS and facilitating the contacts between the L protein and the nucleocapsid (Prange et al., 1999).

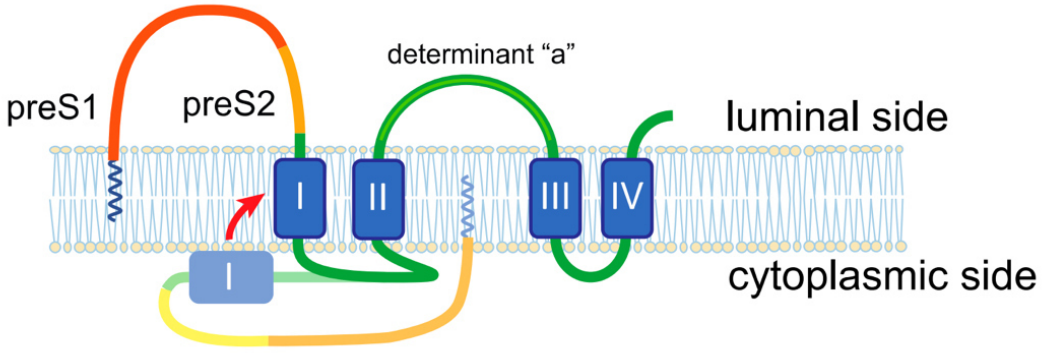

Figure 17. Conformation of HBs L through membrane. *This scheme shows the position switch of preS1 (red arrow): when it is on the luminal side and on the cytoplasmic side (opaquer). Taken from (Inoue et al., 2021)*

In the mature virion and in HBs particles, the preS1 domain is exposed on the surface, where it covers both the S domain and parts of the preS2 domain (Wong & Locarnini, 2018). The preS domain contains several immunogenic T- and B-cell epitopes (Toita, 2015), therefore it is a good candidate for an effective vaccine.

IV. HBV capsid

1. Morphology and assembly

The icosahedral viral capsid is formed by the assembly of numerous copies of the capsid protein. This assembly is initiated by the binding of the Pol to pgRNA (Ponsel & Bruss, 2003) and is driven by the hydrophobic interaction between the interdimer interfaces. The aa at the interdimer region were identified on the crystal structure of the dimer (**Figure 18**): 14E, 18F, 120V, 124V, 127R, 129P, 132Y, 134P and 135P from one monomer and 23F, 25P, 29D, 37L, 39R, 122F and 139I from the other monomer (Zheng et al., 2018).

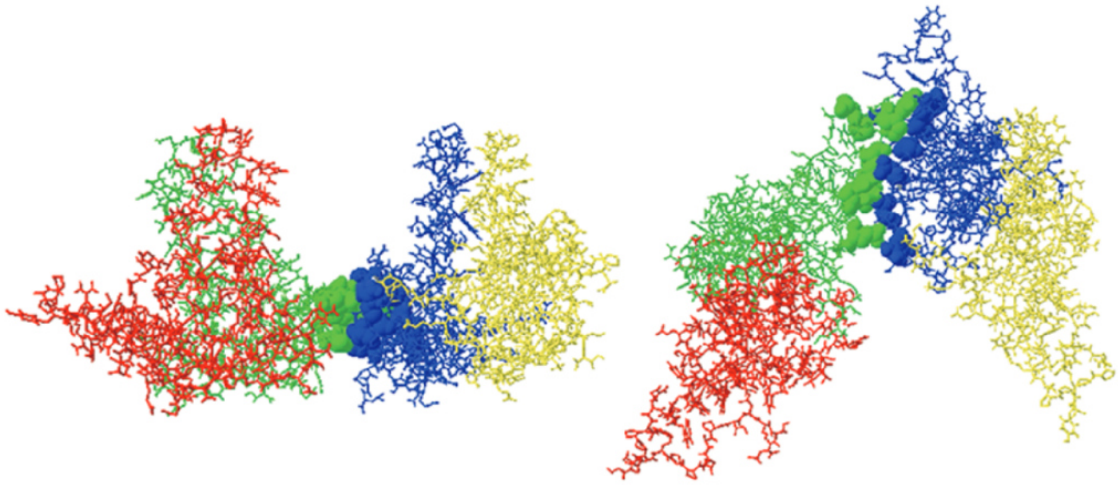

Figure 18. Structure view of the interdimer interface. *Each monomer has its own color. On the left, the interface is viewed from the side, and on the right from above. Taken from (Zheng et al., 2018)*

Two different types of capsids can be distinguished: the T=4 which is dominant (approximately 95%) and the T=3 which is minor (around 5%). The HBV virion is approximatively 42 nm in diameter with core diameter of around 36 nm for T=4 and 32 nm for T=3 (Crowther, 1994), and so T=3 capsids represent 70% of the volume of the T=4 ones (Böttcher, 2021).

The T=4 is composed of 240 copies of Cp dimers, whereas the T=3 is only composed of 180 copies. All the core proteins forming the capsids have the same amino acid sequence; however, the Cp copies can take four quasi-equivalent conformations, respectively named A, B, C or D (**Figure 19**). The major difference between these conformations is located in the region involved in interdimer interaction (especially residues 128 to 136) (Wynne et al., 1999).

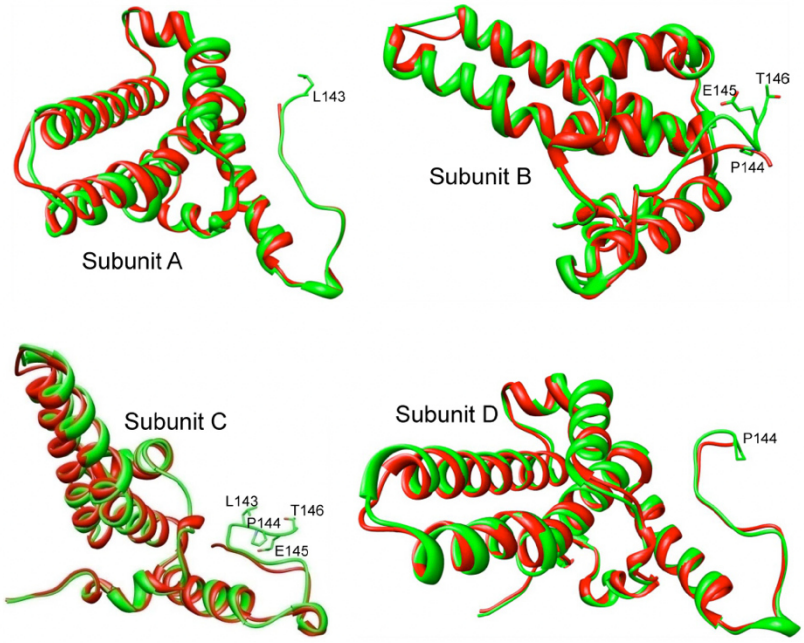

Figure 19. The different conformations of Cp in the capsid. *The crystal structures of the first 142/3 residues of Cp149 are colored in red. The cryo-EM structure of the full-length Cp, with only 143 residues for subunit A, 146 residues for B and C, 144 residues for D, is colored in green. Taken from (Yu et al., 2013)*

In the T=4 capsids, among the 240 Cp copies, there are 60 asymmetrical units, each consisting of two dimeric entities formed by AB and CD conformations, which have a nearly two-fold symmetry (**Figure 20A**). The A conformation interfaces other A conformations, and it forms the five-fold axis. AB dimers extend from the five-fold to the quasi-six-fold axes, and each three-fold vertices are composed of 3 CD dimers which connect the vertices of the quasi-hexagon. Finally, each quasi-sixfold is surrounded by B-C-D-B-C-D subunits (**Figure 20A**). There are holes around the capsid's five-fold, quasi-sixfold, three-fold, and local three-fold axes (**Figure 20A and B**). The five-fold and quasi-sixfold axes primarily involve residues in the proline-rich loop (amino acids 128 to 136) and the C-terminal part of the sequence (amino acids 137 to 142/3). In particular, N137 forms a ring with P134 and P135 and provides a structural clamp to hold the ring in place. The three-fold and the local three-fold channels involve the short α1 helix (amino acids 13 to 17) and the long α2 helix (amino acids 27 to 43).

Contrary to the T=4 capsids, the T=3 capsids only have three quasi-equivalent conformations, A B and C, and the quasi-sixfold vertices is made by the arrangement B-C-B-C-B-C (as the D conformation does not exist for the T=3 capsids). Dimeric contacts surround the axis and are established when the loop near the C-terminus of the assembly domain fits into a helical loop extension structure at the C-terminus of the assembly domain of the adjacent subunit.

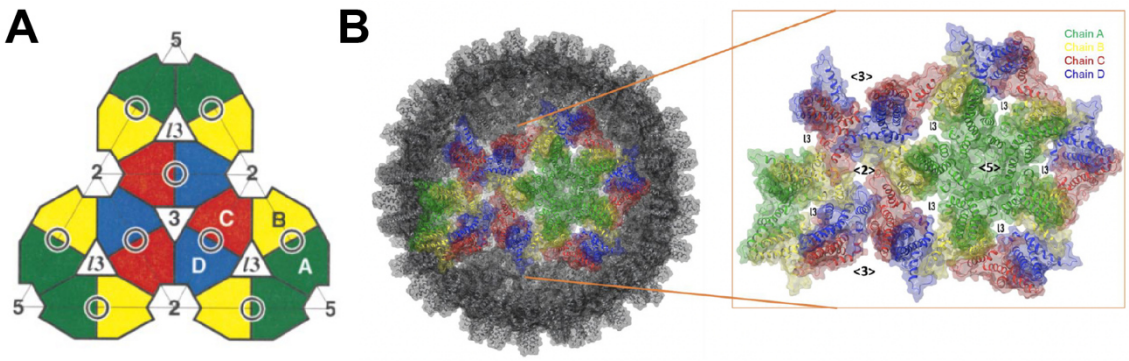

Figure 20. Structure of T=4 HBV capsid. *A) Schematic representation of the assembly of different Cp conformations and their symmetrical arrangement, with a specific color for each conformation. Taken from (Wynne et al., 1999) B) Cartoon representation of the whole capsids, with a zoom on the conformation arrangement. Taken from (Viswanathan et al., 2020)*

HBV capsids can be assembled *in vitro* with the CTD-deleted Cp, Cp149. For this purpose, necessary conditions need to be in place: Cp149 dimer concentration (which has to sufficiently high), temperature and ionic strength (assembly is facilitate in presence of salts in the buffer). At the end, the resulting Cp149 capsids are morphologically similar to capsids purified from HBV expressing cells (Wingfield et al., 1995; Zlotnick et al., 1997). An important finding suggests that cellular proteins may interact with the linker region of the core protein to regulate capsid assembly. Indeed, it has been demonstrated that capsids can be assembled in hepatocytes expressing Cp142, but not Cp149, two forms of Cp with C-terminal truncation after residues 142 and 149, respectively (Ludgate et al., 2016).

It was shown that certain point mutation on the Cp could affect the assembly of the capsid, and notably the mutation of the Tyrosine 132 into Alanine (Böttcher, 2021) which is the best-known. Moreover, two other well-known mutation can also abolish the capsid formation, the phenylalanine 23 into alanine and the leucine 42 into alanine. All these mutations lead to dimers only, which are stable under a wide range of conditions (Alexander et al., 2013).

2. C-terminal domain

The 34-amino acid CTD, containing four arginine-rich clusters, is essential for nucleic acid binding activity, and especially packaging of the pgRNA (Birnbaum & Nassal, 1990; Nassal, 1992). The CTD is subject to phosphorylation, which is required for RNA packaging and plays a role in reorganizing packaged nucleic acid and the solution behavior of the core (de Rocquigny et al., 2020; Lan et al., 1999; J. C.-Y. Wang et al., 2012). Indeed, the CTD contains eight serine/threonine residues that can potentially be phosphorylated or dephosphorylated during the viral replication cycle (**Figure 21A**) (Basagoudanavar et al., 2007; Ning et al., 2017; Pugh et al., 1989). Among them, the major phosphorylation sites are S155, S162 and S170. Thus, the CTD accommodates the main functional sites of HBc that control ordered progression of viral maturation and its localization within the host cell (Böttcher, 2021).

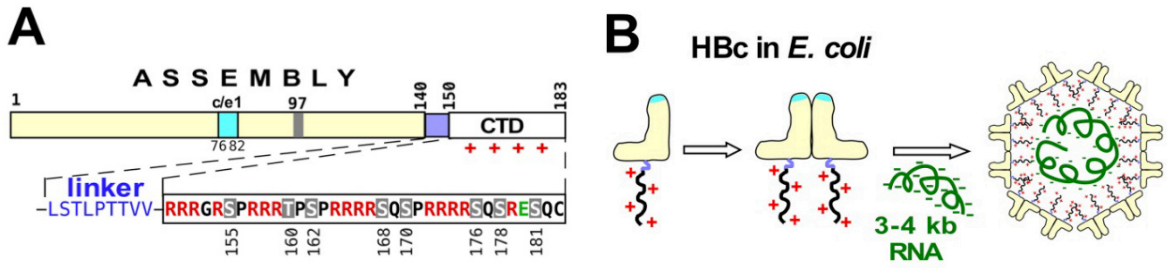

Figure 21. CTD characterization. *A) Organization of each HBc domain, with a zoom on linker and CTD sequences. B) Formation of Cp dimers followed by non-specific pgRNA packaging. Taken from (Heger-Stevic et al., 2018)*

To mimic hyperphosphorylated Cp, it is possible to substitute the seven serine residues in the CTD with aspartic acid, and this results in assembled capsids without detectable packaged RNA. On the contrary, to mimic the dephosphorylated Cp, the seven phosphor-acceptor residues are substituted with alanine, resulting in assembled capsids that non-specifically package viral or cellular RNA (Heger-Stevic et al., 2018; Hu & Liu, 2017; Ludgate et al., 2016; Ning et al., 2011). This is because the arginine-rich clusters in the CTD can non-specifically bind RNA through charge interaction. The RNA binding activity of the CTD can be neutralized by the introduction of negatively charged moieties through phosphorylation (**Figure 21B**) (Chua et al., 2010).

Under expression conditions close to those of physiological conditions, the CTD of Cp was shown to be important for capsid formation (Ludgate et al., 2016). Indeed, Cp149 is only able to assemble when its concentration is high, whereas Cp183 can assemble at concentrations as low as 5 nM (S. Wang et al., 2019).

3. Hydrophobic pocket

When two monomers assemble into a dimer, a hydrophobic pocket is formed in the center of the spikes, which can be associated with a tunnel that crosses the two monomers, containing specifically important residues as M1, I3, P5, Y6, E8, F9, L15, L16, F18, L19, F23, F24, V27, L42, L55, A58, I59, L60, C61, W62, F/I97, R98, Q99, L100, W102, F103, I105, F110, V115, Y118, L119, F122, W125, and L140 (**Figure 22**). These residues are located on α1, the loop connecting α1 and α2, and α4, and α5 helices. They play an important role in maintaining the stability of the monomer fold as these amino acids are highly conserved in human variants (Wynne et al., 1999). The anterior pivot point of α3 is also surface exposed and is the most positively charged region of the spike and leads to the entrance of the hydrophobic pocket (Böttcher, 2021).

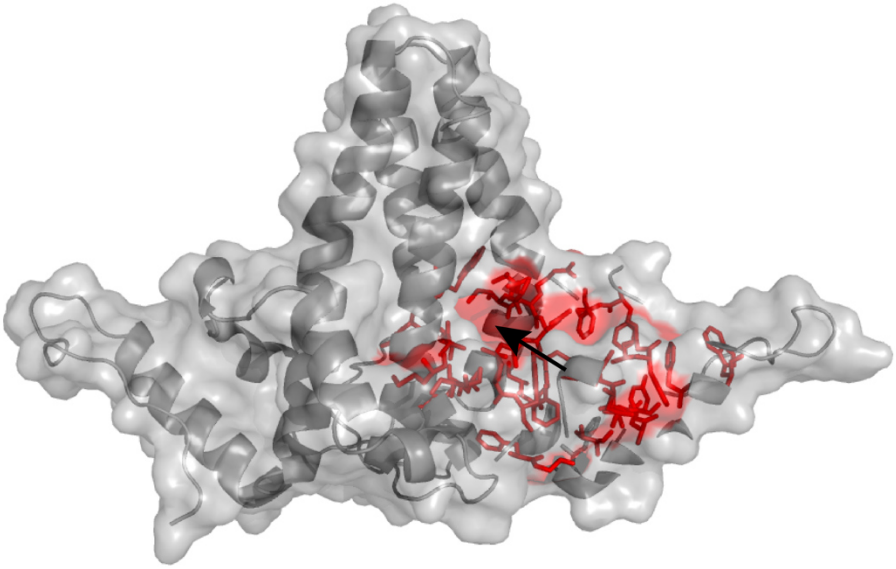

Figure 22. Hydrophobic pocket of HBV dimer (PDB 1QGT). *The entrance to the pocket is indicated by an arrow. The residues belonging to the pocket are labelled in red, on the right monomer.*

Recently, we found a ligand, Triton X-100 (TX100), that interact in the hydrophobic pocket and lead to a conformational switch which may be critical for specific interaction with envelop proteins and virion morphogenesis (Lecoq et al., 2021). The significant structural impact has been characterized by both nuclear magnetic resonance (NMR) (Lecoq et al., 2021) and cryo-EM (Makbul, Kraft, et al., 2021). These studies are important as the hydrophobic pocket has been shown to be central for capsid assembly and envelopment. Indeed, the naturally occurring mutation F97L has an impact on assembly kinetics and so it increases the speed of assembly of dimers into capsids (Ceres et al., 2004). Moreover, two naturally occurring mutations, P5T and L60V, are causing low-level secretion of HBV virions (Le Pogam & Shih, 2002). Finally, the mutation F122V/S/Y is leading to capsid without packaged pgRNA (Pairan & Bruss, 2009). All these residues are located in the hydrophobic pocket, and their mutations impact on the particle composition and secretion.

4. Spike

As first seen in cryo-EM (Crowther, 1994), the structure of the HBV capsid shows characteristic spikes radiating from the surface of the shell, which are formed by the four-helix bundle (Wynne et al., 1999).

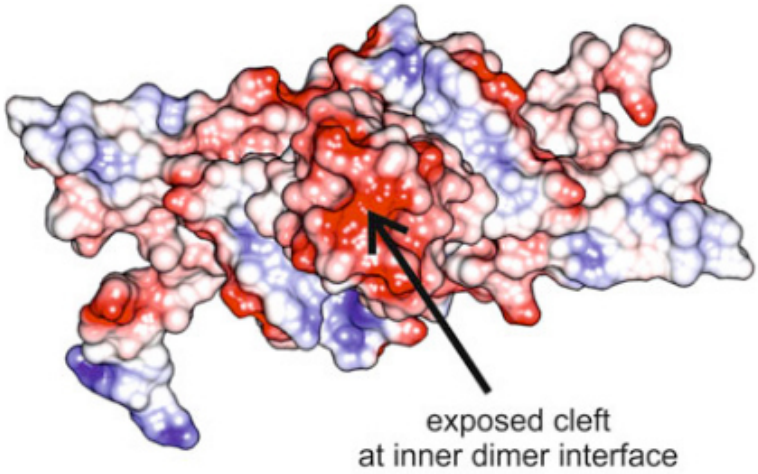

Figure 23. View of the dimer spike. *Dimer is in surface representation, and colored with the relative surface charges (blue is for positive charges and red for negative ones) Taken from (Böttcher, 2021)*

The tips of the spikes are the most negatively charged exposed sites of the capsids (**Figure 23**). They are the specific interface sites between the capsid and the envelope in virions (Dryden et al., 2006; Seitz et al., 2007) and bind to fragments of preS and HBs S (Muhamad et al., 2015). Indeed, it has been observed that the residues making the surface of the dimer spike are well conserved. This information is consistent with the involvement of the spike in the assembly and stability of viral particles (Wynne et al., 1999).

Small peptides are able to bound specifically to the capsid, and particularly to the tip of the spike (Böttcher et al., 1998; Freund et al., 2008). They have been selected by phage display, and they have a high affinity and a strong antiviral activity (Dyson & Murray, 1995). What is interesting about these peptides is that they are able to block the interaction between Cp and HBs L, which may suggest that they are able of blocking the viral cycle. The two peptides are named Oct1 and Oct2 and have the respective sequence: GSLLGRMKGA and MHRSLLGRMKGA (Freund et al., 2008). They were studied by isothermal titration calorimetry, cryo-EM and solution NMR. Smaller peptides are able to bind to Cp, as ALLGRMKG (Dyson & Murray, 1995) or only SLLGRM (Makbul, Khayenko, et al., 2021) which was studied by cryo-EM.

5. Maturation

a. Maturation signal

The maturation signal, that enables the envelopment of rcDNA-filled capsids by the envelope proteins to form infectious virions, is still not fully understood. Freshly synthesized capsids are held in the cytoplasm until maturation (Gerelsaikhan et al., 1996). The three major hypotheses will be described, including i) a structural change in the capsid, ii) a specific phosphorylation state of the CTD and iii) a difference in proline cis-trans isomerization state.

The capsid matures after pgRNA reverse-transcription into rcDNA, this is the first signal during maturation (Roseman et al., 2005). A structural change in the nucleocapsid is associated with the completion of the minus-strand DNA or the initiation of plus-strand synthesis, a step that is required for DNA to be packaged into the viral capsid. During that process, nucleotides can diffuse freely into and out of the capsid shell because of the fenestration of the capsid. Indeed, X-ray studies have shown that capsids have large pores around the icosahedral quasi six-fold, three-fold and quasi three-fold axes. These pores are necessary for the nucleotides access to the capsid interior where occurs the DNA synthesis (Wynne et al., 1999). The structural change occurs on core protein C-terminus (**Figure 24**), which is intrinsically disordered and placed on the borders of holes in the nucleocapsid. On immature particles, the core C-terminus would be oriented toward the interior of the capsid. On mature particles, the core C-terminus is exposed to the exterior for half of the chains, exposing the nuclear localization sequence (NLS) signals. For the other half of capsid proteins, the C-terminus would be exposed to the interior and bound to the viral genome.

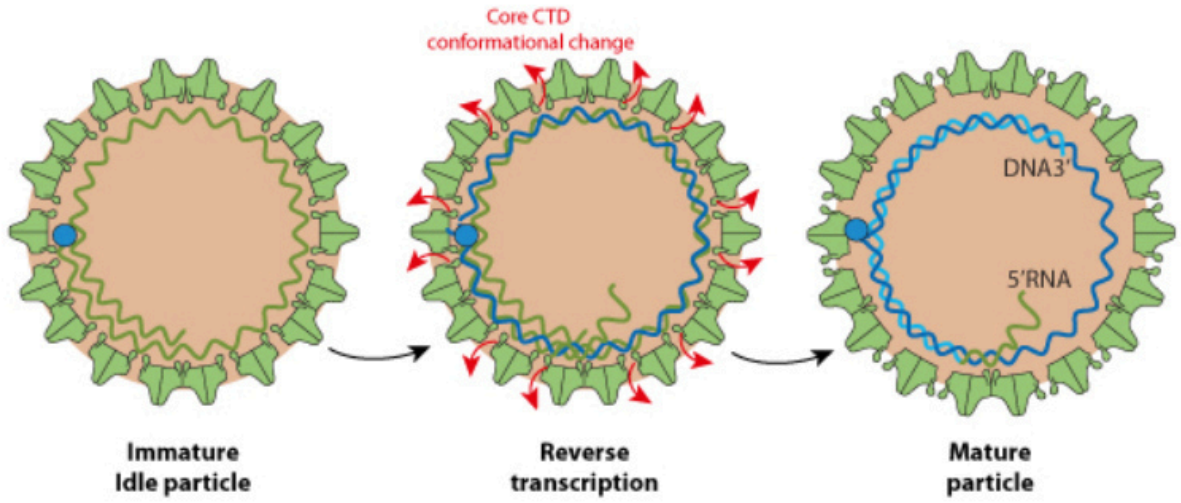

Figure 24. HBV maturation process. *The maturation of the capsids is possible after reverse transcription, allowing CTD conformational change to obtain mature particle packing cccDNA. Taken from https://viralzone.expasy.org/1436*

The second possible signal of maturation is located on the CTD, and more specifically on its phosphorylation state. Indeed, recent research has shown that phosphorylation and dephosphorylation events in the nucleocapsid are related to the cccDNA formation (Luo et al., 2020) and the regulation of the reverse transcription and the capsids maturation (**Figure 25**) (Basagoudanavar et al., 2007; Ning et al., 2017). The immature capsids are phosphorylated, and on the contrary all of the phosphorylation sites were found to be dephosphorylated in the mature capsids before envelopment and secretion (Perlman et al., 2005).

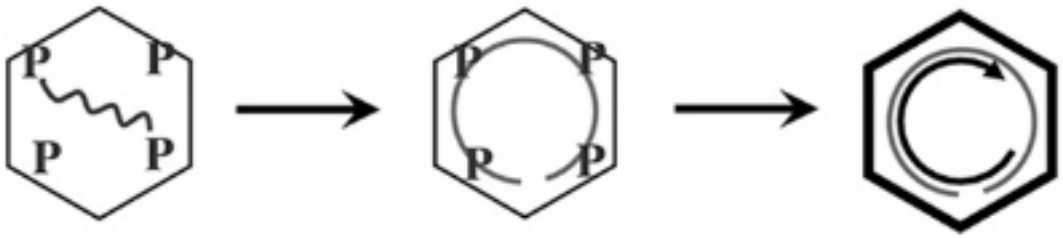

Figure 25. Schematic representation of the maturation of HBV capsids through phosphorylation and dephosphorylation steps. *The immature capsid is represented by a hexagon with thin lines. The thick-lined hexagons symbolize the rcDNA-containing capsid. Taken from (Ning et al., 2017)*

b. New insights

We discovered recently that TX100 interacts with the hydrophobic pocket of the capsid (Lecoq et al., 2021), changing its conformation and particularly residue F97 (Makbul, Kraft, et al., 2021), and we hypothesized that this binding event can mimic the structural maturation signal previously explained. This discovery leads to two new maturation scenarios.

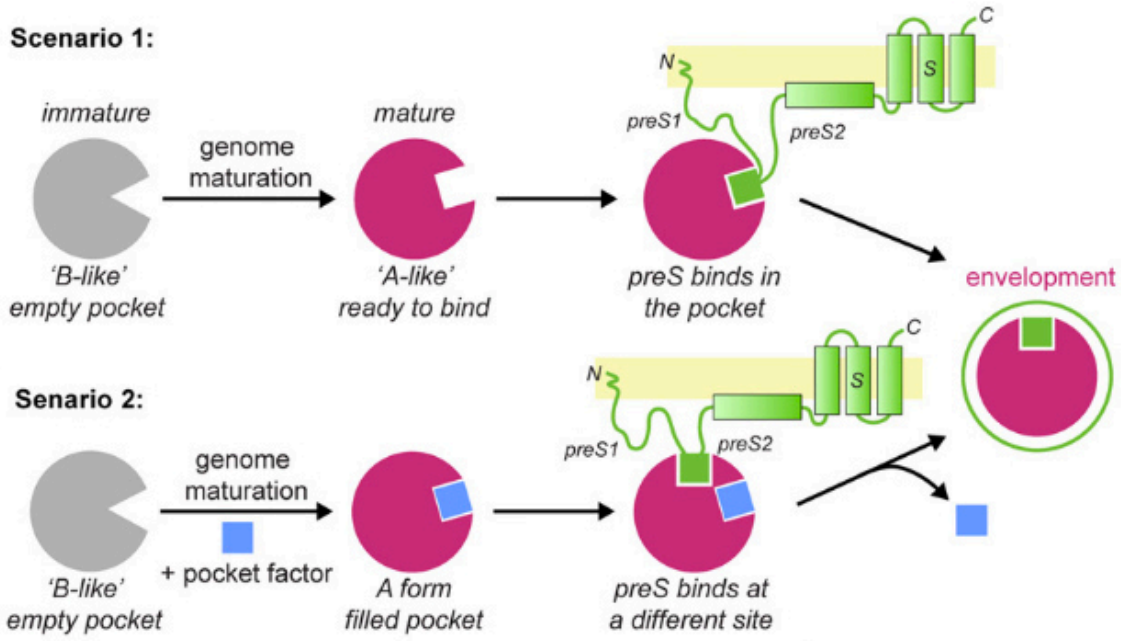

Figure 26. Possible scenarios for the formation of virion particles. *Taken from (Lecoq et al., 2021)*

In a first scenario (**Figure 26A**), the TX100 ligand mimics a natural virus-derived pocket factor. The preS1/2 domain would be a prime candidate since it appears to be required for virion formation (Bruss, 2004). In this sense, the F97L mutation would make pocket accessibility easier, allowing the factor to interact more readily, which could explain premature envelopment. In a second scenario (**Figure 26B**), TX100 mimics a cellular molecule that induces a conformational change that allows the capsid to be ready for envelopment. This structural change would allow preS to bind to a distinct interaction site.

6. Envelopment

All stages of the viral cycle are successive, including maturation and envelopment of the nucleocapsid which are also linked by several mechanisms. It is hypothesized that envelopment is regulated by a structural signal that signals the genome maturation state. Unfortunately, it is difficult to observe the envelopment step *in vitro* (Bruss, 2007).

a. Interaction with envelope proteins

After maturation, capsid envelopment strictly depends on the large envelope protein. Encapsidation of viral nucleocapsids and secretion of mature viral particles require HBs L proteins with cytoplasmic preS region (Bruss, 1997; Bruss & Ganem, 1991; Poisson et al., 1997; Ponsel & Bruss, 2003). Indeed, the virion release depends on HBs L as preS1 is necessary. Direct interaction between the nucleocapsid and envelope proteins was demonstrated in cell-free binding assays and cell-based assays, and several host factors have been identified to work in the process of envelope formation (Inoue et al., 2021). However, it is poorly understood where HBV capsids recruit the envelope to the host cell HBV capsids recruit the envelope during budding (Lambert et al., 2007). The virion formation starts in a specific compartment named the ER-Golgi intermediate compartment (ERGIC) (Huovila et al., 1992). After this step, the mechanisms in place are obscure.

The secretion process for infectious virions depends on MVB-ESCRT machinery, which involves gamma 2-adaptin, CHMP3/4, Vps4, Nedd4, and α-taxilin (Hoffmann et al., 2013; Lambert et al., 2007; Watanabe et al., 2007).

On the capsid side, the region named matrix binding domain (MBD) seems to be important for envelopment (Le Pogam et al., 2000; Ponsel & Bruss, 2003; Roseman et al., 2005), as mutations on MDB residues affect the virions packaging (Ning et al., 2018). On the envelope side, a small sequence of 22 amino acids on preS, named matrix domain (MD), is necessary for HBV assembly (Bruss, 1997, 2007; Schittl & Bruss, 2014). Moreover, NMR-studies suggest that the HBs S fragment forms a short helical segment and the preS fragment a tight loop, which both bind with one face to the cleft at the tips of the spikes (Muhamad et al., 2015).

b. Mutagenesis studies

Mutagenesis analyses identified amino acid residues of Cp on the shell of nucleocapsids that are essential for the envelopment of DNA-containing nucleocapsids (Koschel et al., 2000; Orabi et al., 2015; Pairan & Bruss, 2009; Ponsel & Bruss, 2003). While the tip of the spike of capsid appeared not to be directly involved in the morphogenesis of virions, alanine substitution of amino acid residues L60, L95, K96 or I126 abolish virion secretion (Ponsel & Bruss, 2003). *In vivo* experiments show that mutating any residue of the pocket prevent the interaction of HBs L with HBc (Pastor et al., 2019). All these observations suggest that the pocket could be involved in the interaction of preS but is generally too small to accommodate parts of the preS-sequence directly. Hydrophobic pocket residues are

located on the outer surface of capsid in the depressions between the spikes and possibly bind to envelope proteins and initiate budding of virion particles.

One of the most interesting mutants is the naturally occurring mutation of phenylalanine 97 to leucine. This mutation causes premature secretion of immature nucleocapsids containing less mature forms of viral DNA (Ceres et al., 2004; Yuan et al., 1999). The side chain of the residue 97 is located in the inner space of the spike and close to a hydrophobic pocket showing structural differences in mature versus immature capsids. Possibly, the F97L mutation induces a conformational change causing a constitutive or early expression of the envelopment signal. According to (Shih et al., 2020), only the F97L mutation would lead to a premature secretion. It is important to know that substitution or deletion mutations of the Cp linker region examined thus far completely blocks the production of empty virions, suggesting a critical role of the linker region in empty virion morphogenesis (Liu & Hu, 2019).

Additional point mutation (P5T and P130T) in a quite distant area of the core protein can both restore the wild-type phenotype (Yuan & Shih, 2000), demonstrating the complex nature of the maturation signal for envelopment.

c. Secretion of other particles

First, the egress of naked capsids does not require the ESCRT machinery contrary to mature virions, but rather involves other factors such as Alix and HGS (Bardens et al., 2011; Chou et al., 2015).

Then, empty capsids can be efficiently enveloped and secreted as mature virion particles (Blondot et al., 2016; Hu & Liu, 2017). To date, no studies about viral maturation are able to explain why massive numbers of empty virions are secreted (Ning et al., 2011). Spherical subviral particles predominantly contain HBs S and are released from the endoplasmic reticulum through the general secretory pathway (Patient et al., 2009). However, filamentous subviral particles that contain HBs L share the ESCRT-dependent egress pathway (Jiang et al., 2016). Thus, the virus egress pathway is closely associated with the cellular membrane sorting machinery.

Finally, a model has been proposed to explain how immature capsids can be enveloped (**Figure 27**). This hypothetic process is composed of two secretion signals and a blocking signal generated by the single strand pgRNA (Ning et al., 2018). The first secretion signal (I) is responsible for the basal secretion of empty virions and is inhibited by the blocking signal on immature capsid. The second secretion signal (II) is supposed to be located at the MBD region of Cp and interacts with the MD sequence of L protein as described above. This second signal is necessary for complete virions secretion, and for empty virion, it acts as an enhancer. The blocking signal is present only on the surface of immature nucleocapsids to

prohibit their envelopment, whereas the empty capsid would not possess this signal as it induced by the packaging of pgRNA or ssDNA.

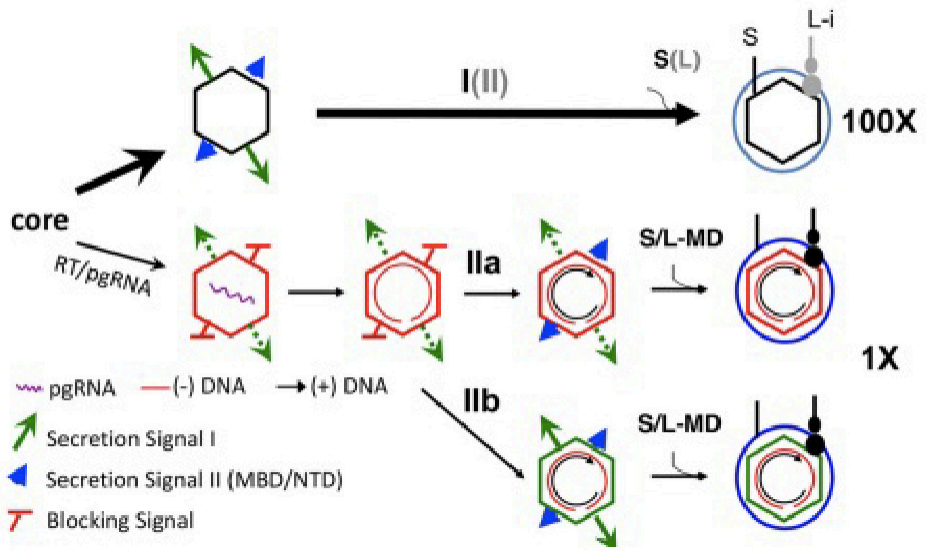

Figure 27. Model for two distinct signals in the secretion of empty versus complete HBV virions. *Taken from (Ning et al., 2018)*

CHAPTER 2

I. Protein expression in bacteria

The use of bacteria for heterologous expression of proteins was established in the 1980s when molecular cloning techniques became widely available. Recombinant expression of proteins of interest in *Escherichia coli* (*E. coli*) is an important tool for many purposes, including structural biology, and is still one of the dominant bacterial expression systems since it is a well-understood host with broad application range.

E. coli is the most popular organism for the expression of recombinant proteins since it displays numerous advantages (**Table 2**): first, it has incomparable fast growth kinetics, allowing protein production in short times. Indeed, in glucose-salts medium and in the optimal environmental conditions, a culture can reach the stationary phase in few hours (Sezonov et al., 2007). However, it is important to know that a considerable decrease in generation time can be caused by the expression of a recombinant protein due to a metabolic burden on the microorganism (Bentley et al., 1990). Second, it is possible to achieve high cell density cultures in a lot of different media (Shiloach & Fass, 2005). However, exponential growth in complex media leads to lower densities (Sezonov et al., 2007). Third, rich complex media can be made from readily available and inexpensive components, and allows for isotope labelling required for NMR.

Disadvantages of *E. coli* expression system include the difficulty in expressing some mammalian proteins, the accumulation of proteins as inclusion bodies, the protease contamination from host proteins leading to degradation of expressed protein, and also the endotoxin accumulation. Also, this bacterial system lacks some enzyme necessary for the phosphorylation and/or glycosylation of the proteins (Heger-Stevic et al., 2018). However, new methods have been implemented in order to get fully-functional proteins, such as using yeast enzyme in *E.coli* systems (Acharya et al., 2005).

Table 2. Pros and cons for different expression systems. *(El-Baky & Redwan, 2015)*

Expression system	Advantages	Disadvantages
E. coli	• Very fast growth rate • Very high protein yield • Very low production cost	• Eukaryotic co- and posttranslational modifications not possible • Codon usage is different from eukaryotes • Producing insoluble proteins
Yeast	• Fast growth rate • High protein yield • Low production cost • Good protein folding • Perform simple glycosylation	• Lack the complex posttranslational modification pathways
Mammalian cell lines	• Very good protein folding • Perform simple glycosylation • Perform complex glycosylation	• Very slow growth rate • Very low protein yield • Very high production cost
Insect cell lines	• High expression yield • Very good protein folding • Perform simple glycosylation • Perform complex glycosylation	• Slow growth rate • High production cost
Cell-free systems	• Rapid optimization of expression constructs • High expression yield • Reaction conditions can be modified to solve insolubility problems	• High production cost

The protein expression process starts with the administration of a vector that inserts target DNA fragments into the *E. coli* cells (**Figure 28**). After that, bacteria cells are grown in specific medium in order to produce the initial preculture, then transferred into a bigger volume of medium to start the main culture. Protein expression is initiated by induction and followed by a centrifugation to remove the supernatant and recover cells. The pellet is then dissolved in lysis buffer and cells are lysed. Another centrifugation allows to remove the cellular debris and to keep the protein of interest in the supernatant. Final step is a purification process (by chromatography for example).

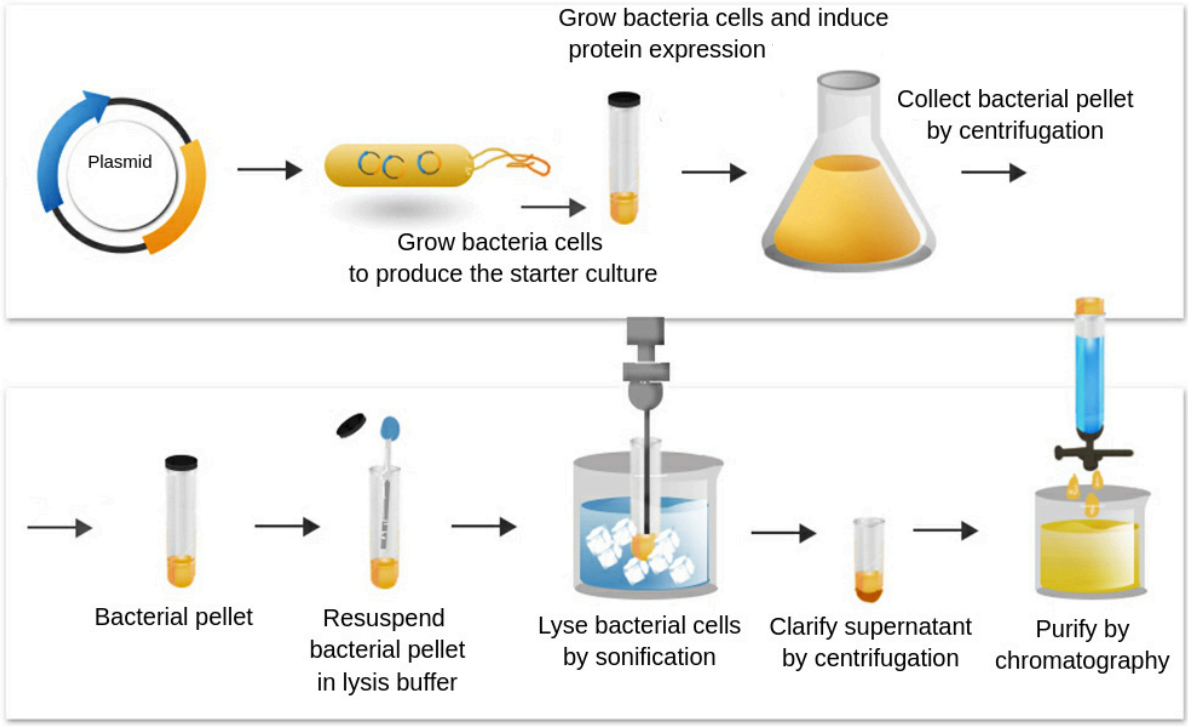

Figure 28. Protein expression and production by bacterial system. *Adapted from https://theory.labster.com/bacteria-expression-systems/*

The *E. coli* cells show good tolerance to deuterium (Paliy et al., 2003), allowing for cost effective labeling with ^{13}C, ^{15}N and ^{2}H.

II. Isotopic labeling

Several strategies have been developed for efficient isotopic labeling of recombinant proteins in bacteria (Cai et al., 1997; Marley et al., 2001; Ross et al., 2004). The aim of these methods consists of obtaining high cell densities by the use of inexpensive unlabeled medium for a small culture, and its subsequent transfer in labeled medium before expression.

Isotopic labeling is defined as a technique used to replace all nuclei of a certain element with its respective isotope through a reaction, metabolic pathway, or cell. For the production of ^{13}C and ^{15}N labeled proteins, the labeled precursors typically used are uniformly ^{13}C-labeled glucose or glycerol, and ^{15}N-labeled ammonium chloride or ammonium sulfate, respectively the source of carbone and nitrogen (Hong, 2010). These compounds are directly incorporated into the growth medium for protein expression.

After the transfer of the preculture into the labeled medium, the metabolic pathway of *E. coli* cells allows the amino acids of the protein to be labeled from the conversion of glucose and ammonium chloride. The last step is the incorporation of the resultant isotopes into the protein of interest by biosynthesis (Lacabanne et al., 2018; Verardi et al., 2012).

In addition of ^{13}C and ^{15}N labelling, it is possible to deuterate the proteins. For this purpose, the water and the glucose used in the growth medium must be replaced by deuterium oxide (D_2O) and deuterated glucose, in order to convert all ^{1}H in ^{2}H.

Most labeled proteins are produced in *E. coli* bacterial system, and it requires a specific growth medium, also known as M9 minimal medium (Marley et al., 2001; Studier, 2005). This medium is economic and highly efficient; however, it has a specific composition, and its preparation has to follow a strict order (**Table 3**). The addition of traces ensures that the large number of metal ion containing enzymes in *E. coli* can function optimally (Cai et al., 1997). The vitamins are supplemented in order to improve cell growth and protein expression (Atlas, 2010).

Table 3. Composition of M9 medium. *Taken from (Lacabanne et al., 2018)*

M9 medium components	**Concentration (final)**	**Trace elements**	**Concentration (final)**
$MgSO_4$	2 mM	EDTA	0.17 mM
$CaCl_2$	100 µM	$CaCl_2$	0.54 mM
Na_2HPO_4	38 mM	$CuSO_4$	0.027 mM
KH_2PO_4	22 mM	$MnCl_2$	0.095 mM
NaCl	8.6 mM	NaCl	0.086 mM
Antibiotic	25-50 μg/ml	H_3BO_3	0.003 mM
Glucose	2 g/L	$ZnSO_4$	0.024 mM
NH_4Cl	2 g/L	$FeSO_4$	0.216 mM
Trace elements (100x)	1 x	ascorbic acid	0.011 mM
Vitamins solution (100x)	1 x		

New methods for uniform isotopic labeling have been set up in order to improve labeled protein production, as increasing buffering capacity, changing the composition of the minimal medium, or improving the expression protocol (Azatian et al., 2019; Cai et al., 2016; Studier, 2005).

Uniform ^{13}C and ^{15}N-labeling plays an indispensable role in structure determination of proteins, as it is the simplest and most cost-effective biosynthetic labeling method. Isotopic labelling is used to enhance the signal-to-noise ratios of the NMR spectra (Hong, 1999) but also allows for site-specific interrogation of structures and intermolecular contacts. Indeed, a single protein sample can in principle provide all the structural constraints (dihedral angles and distances).

III. Wheat germ cell-free protein synthesis (WG-CFPS)

Cell-free protein synthesis is a powerful tool for protein expression. Indeed, cell-free systems allows the production of proteins that would be toxic to cells, undergo proteolysis in cells or aggregate in inclusion bodies. They have been used for the production of various kinds of proteins, including membrane proteins (Fogeron et al., 2017; Klammt et al., 2004; Makino et al., 2014) and proteins that are toxic to cells (Verardi et al., 2012).

While a large number of cell-free systems exist today, the main three systems are based on *E. coli* lysates (prokaryotic), wheat germ extracts and rabbit reticulocytes lysates (both eukaryotic). Here we will focus on the system using wheat-germ extracts which combines advantages of the eukaryotic translational machinery with yields compatible with structural studies (Takai et al., 2010). This system is particularly suited for the expression of difficult-to-express proteins such as integral membrane proteins. Indeed, it is an opened system which allows the addition of detergents and/or lipids which promote the synthesis of membrane proteins in solubilized form as soon as they leave the ribosome. As opposed to basic cellular mechanisms, WG-CFPS creates an optimal environment for protein synthesis due to all metabolic resources being assigned to the protein. It contains the cellular components necessary for protein synthesis (transfer RNA, ribosomes and initiation, elongation, and termination factors). The extract is supplemented with an energy-generating system (phosphocreatine/phosphocreatine kinase), and with spermidine to stimulate the efficiency of chain elongation. Only exogenous amino acids and mRNA are needed to initiate translation (**Figure 29**).

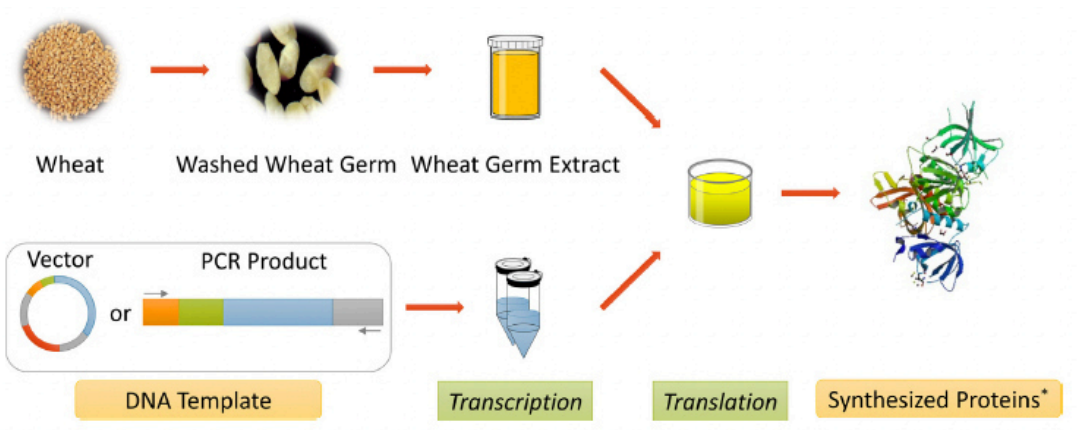

Figure 29. Overview of wheat germ cell free expression. *Taken from (Harbers, 2014)*

Once optimal translation conditions have been established, WG-CFPS is a simple process, protein expression and purification can be done in a short period of time (1-2 days). Labeling of proteins for structural studies like NMR can be done in a very specific and efficient manner (Fogeron et al., 2021; S. Wang et al., 2019).

The main drawbacks of wheat germ cell free system include the commercial extracts prices which are expensive and also a reduced capacity for post-translational modifications (David et al., 2019). In our lab, we sort the germs, and we produce our own wheat-germ extract from non-treated durum wheat seeds in order to decrease the price of protein production.

WG-CFPS allows the co-expression of two proteins, which is particularly interesting in the context of the present work to study the potential interaction between Cp183 and the preS1 and preS domains, HBs S and HBs L proteins, directly upon synthesis. To the end, the mRNAs of both proteins are *in vitro* synthesized and added to the translation reaction. While the Cp183 protein is synthesized without tag, a Strep-Tactin is fused to the envelope proteins (preS1, preS, HBs S or HBs L), which allow to capture them on Strep-Tactin magnetic beads. Further analysis by SDS-PAGE followed by western blotting allows then to determine whether the proteins interact or not. All proteins of interest here are soluble proteins. If Cp183 is detected on the blot (using an anti-Cp183 antibody kindly provided by Michael Nassal) in the same fraction than envelope proteins this means that both of them interact upon synthesis.

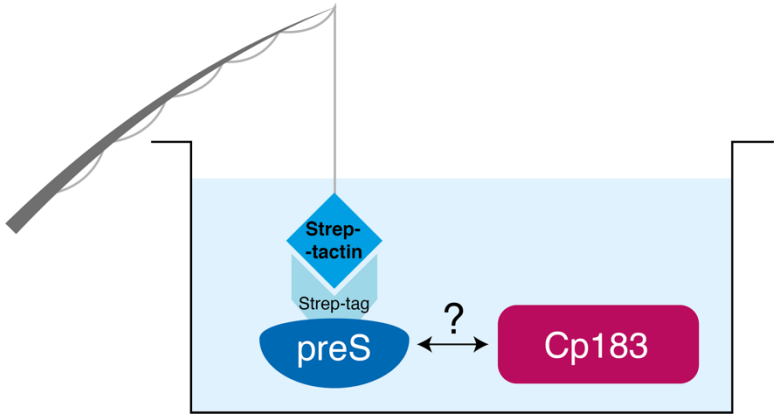

Figure 30. WG-CFS co-expression test model. *The fishing rod models the magnetic beads, which capture the envelope proteins (in this example it is preS) and if an interaction takes place with the capsid protein (here Cp183) upon synthesis, then they will be fished together and detected on the western blot (WB) of Strep-Tactin beads.*

IV. Nuclear Magnetic Resonance

The scientific community has granted the discovery of NMR to Isidor Isaac Rabi, who received the Nobel Prize in Physics in 1944. In 1946, NMR in condensed matter (water and paraffin) was discovered simultaneously by Edward Purcell at Harvard and Felix Bloch at Stanford using different instrumentation and techniques.

1. Principle

The principle of NMR is based on the fact that all nuclei have a spin and are electrically charged. If an external magnetic field is applied, an energy transfer is possible between the base energy to a higher energy level. The energy transfer takes place at a wavelength that corresponds to radio frequencies and when the spin returns to its base level, energy is emitted

at the same frequency. The signal that matches this transfer is measured and is called the Free Induction Decay (FID). This signal depends on the chemical environment of the atom, thus giving access to details of the electronic structure of a molecule and its individual functional groups. After its acquisition, the signal is processed by Fourier transform (FT), phase adjustment, chemical shift calibration, integration, in order to yield an NMR spectrum of the concerned nucleus (**Figure 31**).

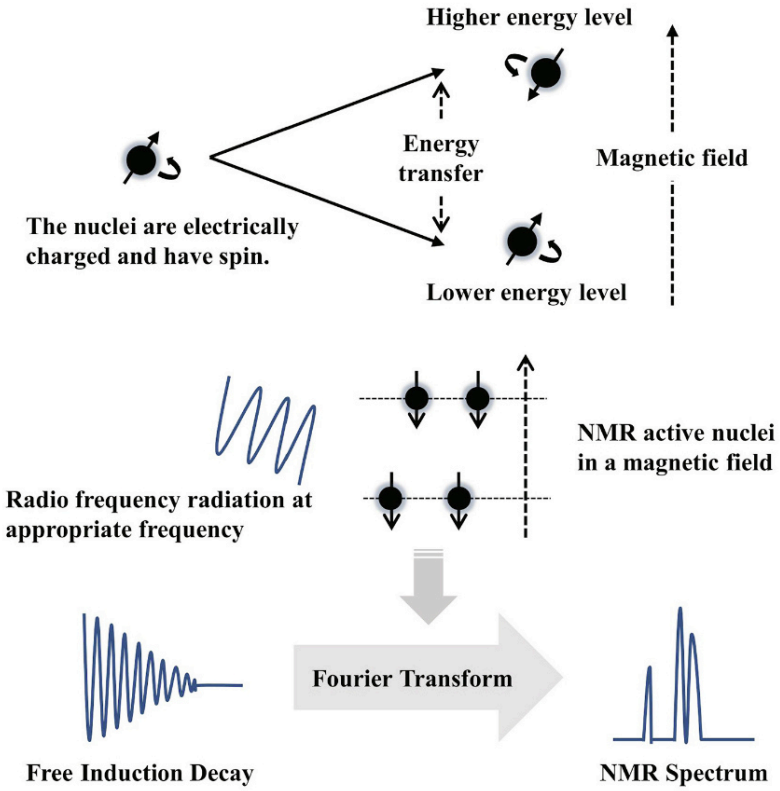

Figure 31. Scheme about NMR principle. *Taken from (Cao et al., 2021)*

To be able to detect nuclei, they should have a spin different from zero. Indeed, each nucleus have a spin quantum, which results from the sum of its protons and neutrons. If an atom has an even number of protons and neutrons, the spin quantum will be null. On the contrary, if this is an odd number, the quantum spin will have a non-zero value. Nuclei from isotopes of many elements (*e.g.* ^{1}H, ^{13}C, ^{2}H, ^{6}Li, ^{10}B, ^{11}B, ^{14}N, ^{15}N, ^{17}O, ^{19}F, ^{23}Na, ^{29}Si, ^{31}P, ^{35}Cl, ^{113}Cd, ^{129}Xe, ^{195}Pt) have been studied and the most commonly used nuclei are ^{1}H, ^{15}N and ^{13}C by high-field NMR spectroscopy since they are the most abundant in macromolecules. Most of the isotopes used for NMR experiments are naturally abundant, but in small quantities (**Table 4**). In this way, proteins which are mostly composed of proton, carbon and oxygen must be labeled.

Table 4. Different isotopes properties. *Adaptated from (Gerothanassis et al., 2002)*

Nucleus	Spin	Natural Abundance %	Gyromagnetic ratio (MHz/T)
^{1}H	1/2	99.985	42.6
^{2}H	1	0.015	6.5
^{7}Li	3/2	92.58	16.5
^{13}C	1/2	1.108	10.7
^{14}N	1	99.63	3.1
^{15}N	1/2	0.37	4.3
^{17}O	5/2	0.037	5.8
^{19}F	1/2	100	40.1
^{23}Na	3/2	100	11.2
^{31}P	1/2	100	17.2

When the sample of interest is labeled, it is placed in the probe inside an NMR spectrometer, which generates the strong magnetic field necessary for nuclei detection. For most NMR experiments, a homogeneous magnetic field strength is needed. The intramolecular magnetic field around an atom in a molecule changes the resonance frequency, and in this way NMR spectroscopy can provide detailed information about the structure, dynamics, reaction state, and chemical environment of molecules such as proteins.

2. Chemical shifts

Due to their different electron density, all nuclei do not resonate at the same frequency. In fact, electrons are charged, rotate and generate a magnetic field that is the reverse of the applied magnetic field. The electronic shielding reduces the magnetic field at the nucleus, therefore the frequency needed to obtain resonance is decreased. This is known as chemical shift and provides an explanation for the ability of NMR to ascertain chemical structures, which rely on electron density. Nuclei with higher degrees of shielding due to a higher electron density have their NMR frequency shifted up, which corresponds to a low chemical shift. On the contrary, less shielding shifts NMR frequency down, which gives a high chemical shift. The chemical shift of each atom is expressed in parts per million (ppm), a value that is independent from the magnetic field.

The most prominent feature of NMR spectroscopy is its ability to differentiate chemically identical atoms, experiencing small variations in their magnetic environment influenced by their neighboring atoms. This property, substantiated by several decades of scientific and technological developments, allows for the assignment of each peak in the spectrum to a specific atom within a molecule of interest. While a one-dimensional (1D) spectrum can be sufficient to assign atoms of a small molecule (**Figure 32**), for complex molecules like proteins, the number of nuclei is more important than a small molecule, and the assignment becomes too difficult with a unidimensional spectrum. Therefore, multi-dimensional spectra can be recorded.

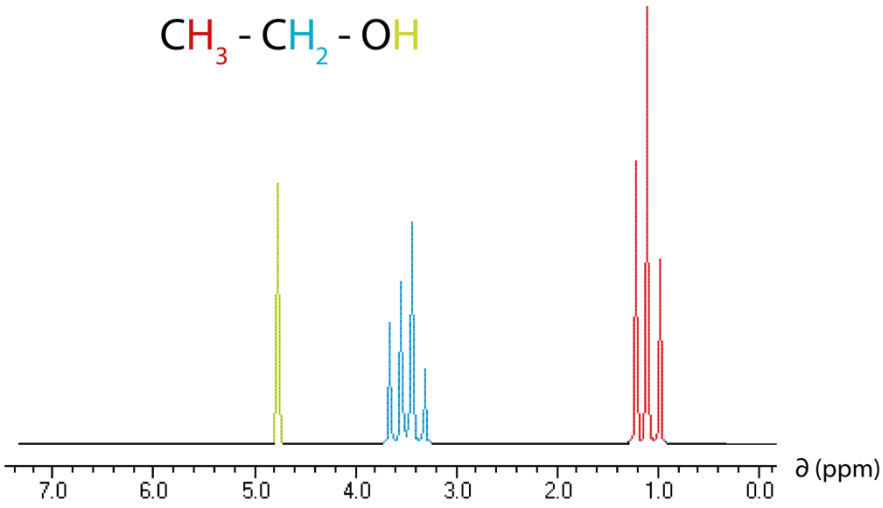

Figure 32. 1D ^{1}H spectrum of ethanol. *Depending on the chemical environment of each proton, the chemical shifts are different.*

3. Multi-dimensional detection

Multidimensional NMR spectroscopy forms the basis of the determination of the three-dimensional structure of biomolecules like proteins, by providing the resolution necessary to analyze their complex spectra.

1D NMR spectra result from a FT of the FID signal acquired just after the pulse sequence (excitation step with variation of magnetization). A two-dimensional (2D) spectrum has its detection step separated from its preparation step by evolution and mixing time intervals (**Figure 33**). During the evolution time, the chemical shift of the first nucleus labels the magnetization. Then, during the mixing time, the magnetization is transferred from the first nucleus to the second. The transfer of magnetization results in correlations between signals from the two different nuclei. There are two mechanisms to transfer the magnetization: scalar coupling or dipolar interaction. Depending on the type of interaction used to accomplish this transfer, the evaluation of the spectra allows for the extraction of either a resonance assignment or structurally relevant parameters (Schmieder, 2006). During the detection, the magnetization is marked with the chemical shift of the second nucleus.

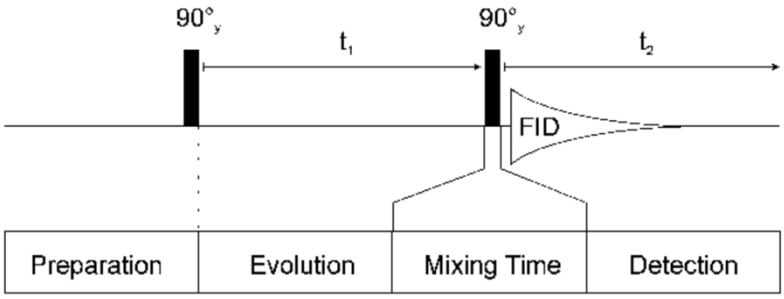

Figure 33. Time intervals for a 2D NMR experiment. *Taken from https://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/2dnmr.htm*

In this way, two-dimensional spectra can be recorded, based on correlations such as ^{1}H-^{15}N, ^{13}C-^{15}N but also ^{13}C-^{13}C. In a homonuclear experiment, chemical shifts of the same type of nucleus are recorded in both dimensions (δ_H). If no transfer takes place during the mixing time, the same chemical shift is recorded in both dimensions, and this results in a diagonal peak (**Figure 34**). If there is transfer of magnetization between two nuclei, a cross peak appears at their respective chemical shifts (δ_1 and δ_2 in **Figure 34**). This indicates a connection between the nuclei corresponding to the chemical shifts. In a heteronuclear experiment, two different chemical shifts (δ_H and δ_X in **Figure 34**) are recorded and in the case of a transfer of magnetization during the mixing time. In this way, a cross peak appears, indicating a connection between the nuclei involved.

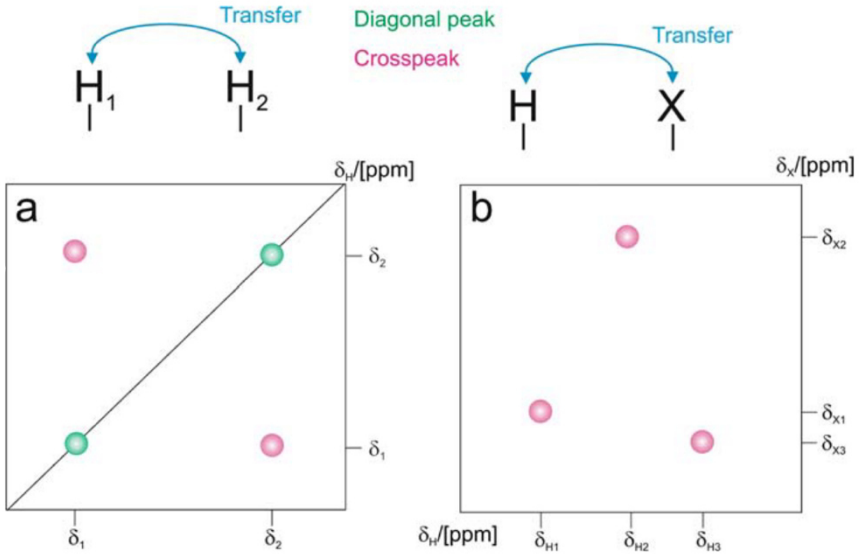

Figure 34. Homonuclear and heteronuclear experiments. *Homonuclear is cross-peaks (in pink) are represented on the left and heteronuclear on the right.*

The two-dimensional experiment can even be extended to three or more nuclei. With the use of 2D and 3D spectra, it is then possible to perform the assignment of a peak to its specific nucleus in the protein sequence (Fricke et al., 2017). For example, it is possible to record a HNCA spectrum (**Figure 35**). In this case, the magnetization passes from the hydrogen to the nitrogen, and then to the Cα via scalar coupling. At the end, the magnetization goes back to nitrogen and hydrogen for signal detection.

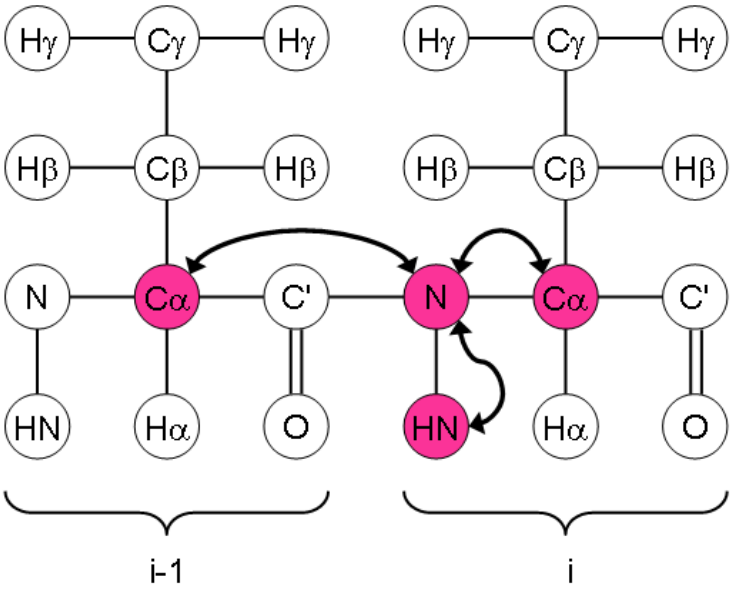

Figure 35. Schematic view of a 3D spectra: HNCA. *This scheme shows the magnetisation transfer from one nucleus to the next ones. Taken from: https://www.protein-nmr.org.uk/solution-nmr/spectrum-descriptions/hnca/)*

4. Liquid-state NMR

a. Principle

Liquid-state NMR, also called solution NMR, is widely used for numerous applications, including the determination of molecular structure, conformations and associated dynamics, the study of complex mixtures or intermolecular interactions, and this technique is applicable for small organic molecules to proteins or nano-objects. Yet, this technique has some limitations, notably it requires a soluble sample, a condition which can be difficult to reach for some molecules, like membrane proteins. Moreover, solution NMR is usually limited to the study of macromolecules with a maximum size of approximately 40 kilodaltons (kDa) or less, since the slowing of the overall tumbling for large proteins causes rapid transverse relaxation and line broadening beyond detection.

b. Binding modes

Currently, NMR is widely used for drug discovery, as this is a non-destructive technique. It allows to get information about the localization of ligand binding site, but also ligand binding modes (Q. Li & Kang, 2020). When a ligand interacts with a protein, three different exchange modes can occur: the fast, intermediate, or slow exchange (**Figure 36**). In slow exchange, only the signal of the free and the bound forms are observable. This is because there is not significant interconversion between the two forms during the detection period of the NMR experiment ($k_{ex} << \Delta\omega$). In this case, the intensity of each peak directly corresponds to the population of that species, and it is usually the case for high affinity interactions. On the contrary, in fast exchange, the signal of the free-protein peak is moving

towards the fully ligand-saturated protein peak. This is because there is rapid interconversion between the two forms and hence an average during the detection time of the NMR experiment (k_{ex} >> $\Delta\omega$). In the intermediate exchange regime, the same observation is possible, with peak broadening along the titration, leading to intensity differences of the successive peaks. Indeed, the interconversion between the two forms is stable during the NMR experiment (k_{ex} ~ $\Delta\omega$) (Becker et al., 2018; Kleckner & Foster, 2011; Oh et al., 2020).

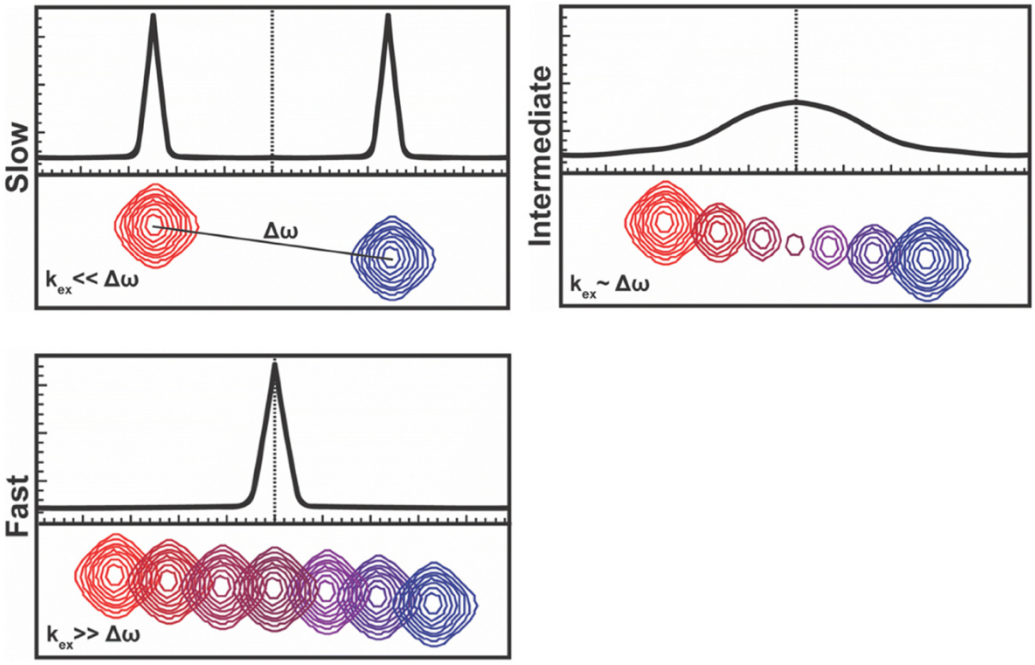

Figure 36. Representation of the three different exchange regimes. *k_{ex} is the total exchange rate; $\Delta\omega$ is the chemical shift difference between the two peaks.*

c. Solution NMR spectra

During my thesis, the main spectra I recorded are band-selective optimized flip angle short transient (SOFAST) - heteronuclear multiple quantum coherence (HMQC) which correlates the chemical shift of the protons to a heteronucleus (here ^{15}N) via the J coupling interaction between the nuclei. These experiments have been used especially in biomolecular NMR as they allow high-throughput protein screening, by the fact that they the recording of 2D spectra in few seconds (Schanda et al., 2005). An example of $^{2}H^{13}C^{15}N$-Cp149 dimer spectrum is given in **Figure 37**.

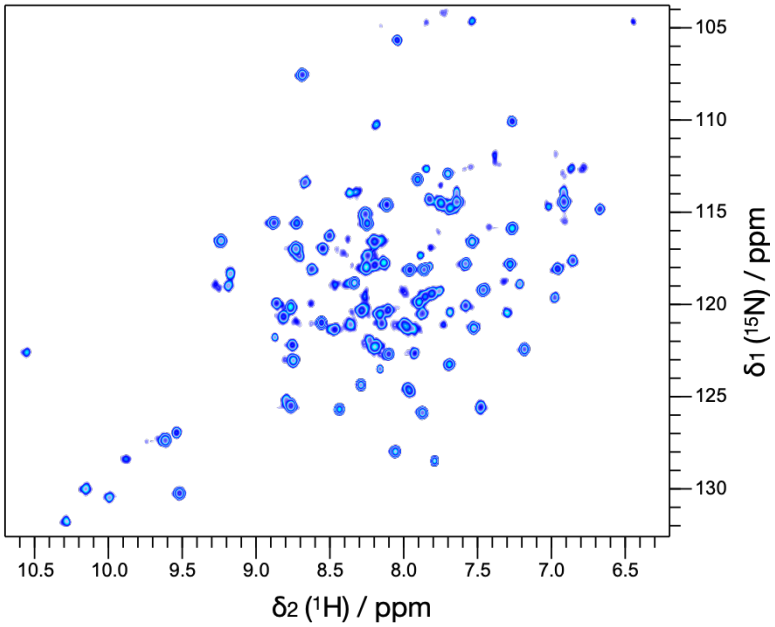

Figure 37. ^{1}H-^{15}N-SOFAST spectrum of $^{2}H^{13}C^{15}N$-Cp149 dimer in 50 mM HEPES, 5 mM DTT, pH 7.5.

5. Solid-state NMR

a. Magic-Angle spinning

Solid-state NMR (ssNMR) spectra are characterized by larger linewidths than in solution NMR, which can be utilized to give quantitative information on the molecular structure, conformation, and dynamics of the material.

In ssNMR, magic-angle spinning (MAS) is an approach used to improve the resolution as well as the sensitivity of the recorded spectra (Schledorn et al., 2020). The sample is spun at a frequency of 1 to 130 kiloHertz (kHz) at an angle of 54.74°, in regard to the magnetic field.

There are three different interactions between the nuclei in NMR: dipolar, chemical shift anisotropy (CSA), and quadrupolar, generating spectral broad linewidths difficult to analyze. However, these interactions are dependent on position and MAS is used to average them in ssNMR. Dipole interactions between nuclei are approximately zero at 54.74° (**Figure 38**) (Alia & Ganapathy, 2009). MAS can partially average the quadrupolar interactions. However, nucleus-electron interactions, known as CSA, are non-zero. In solution, these interactions are averaged out due to molecular movement, molecules can adopt any orientation depending on the magnetic field. In solid-state, MAS can overcome this problem.

Indeed, this technique allows to transform wide spectral lines into narrow lines, increasing the resolution of the spectra, and is able to suppress the CSA.

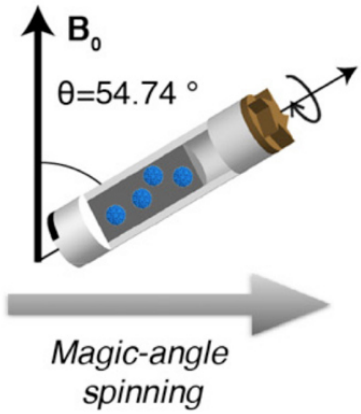

Figure 38. Representation of the MAS technique. *Taken from (Lecoq et al., 2020)*

In order to improve the signal and the resolution of the spectra, MAS can be combined with faster rotor spinning (Böckmann et al., 2015). Indeed, the combination of these two parameters have enabled research into systems previously not amenable to atomic-resolution analysis (Polenova et al., 2015). In fact, when proteins spin faster, the direct detection of proton spectra are possible as the ^{1}H dipolar couplings are weakened. These high rotation frequencies allow to determine the structures of a wide range proteins, and notably viral nucleocapsids (Andreas et al., 2016) The increased sensitivity and resolution attained at high magnetic fields enables analysis of very large macromolecular systems that would be intractable at lower fields and lower MAS frequencies.

b. Rotor filling

Contrary to solution NMR, the sample used for solid-state NMR should be in a solid form, as a sediment, a powder or crystals etc. In this state, the sample must be transferred in a specific container, to be then inserted in the solid-state NMR probe, called an NMR rotor. After years of development, numerous commercial Zirconia cylindrical rotors exist now, with smaller and smaller size, going down to 0.5 mm diameter (**Figure 39**). Indeed, the faster the rotor can turn, the higher the rotation frequencies can be. The advantage of these smaller rotors is the improvement of the signal to noise ratio, however due to less quantities of sample, the signal could be reduced, therefore a balance between these two parameters must be found.

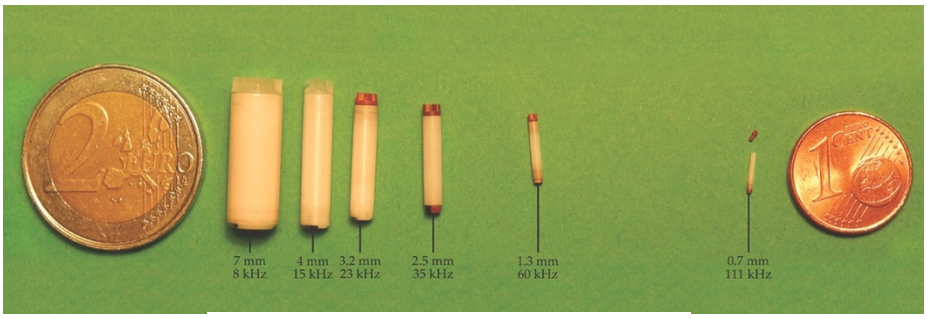

Figure 39. Different size of rotors.

The rotors could be open on one end (for 3.2 mm rotors and higher diameters) or open on both ends (for 1.3, 0.7 mm rotors and smaller). If both ends are open, the bottom cap has to be firstly closed. Then, the rotor is inserted in the filling tool, with a funnel allowing to insert the viscous sample into the rotor (**Figure 40**).

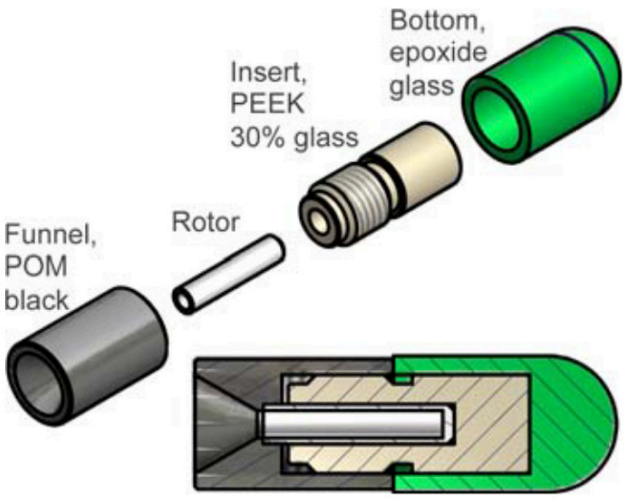

Figure 40. Filling tool representation. *Taken from (Bockmann et al., 2009)*

The protein sample is sedimented into the rotors by ultracentrifugation during several hours, depending on the sample density, until the packing of the whole sample (**Figure 41**). The rotor must be verified to be intact before filling. At the end of the process, the drive cap is pushed on the rotor to close it. The drive cap contains the wings that will allow the rotation of the sample.

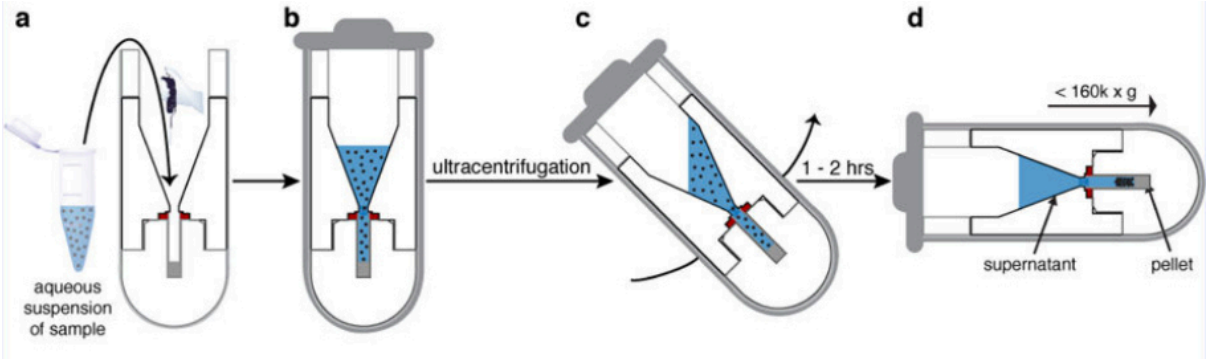

Figure 41. Sedimentation of the sample into the rotor. *Taken from (Mandal et al., 2018)*

c. Solid-state NMR spectra

For the study of labeled Cp149 capsids in 3.2 mm rotors, I mainly recorded 2D ^{13}C-^{13}C spectra, more specifically called dipolar assisted rotational resonance (DARR) spectra. These

spectra are homonuclear, explaining the presence of a diagonal. The magnetization is transferred from the proton to the ^{13}C, and then to the other closest ^{13}C. An example of $^{2}H^{13}C^{15}N$-Cp149 dimer spectrum is given in **Figure 42**.

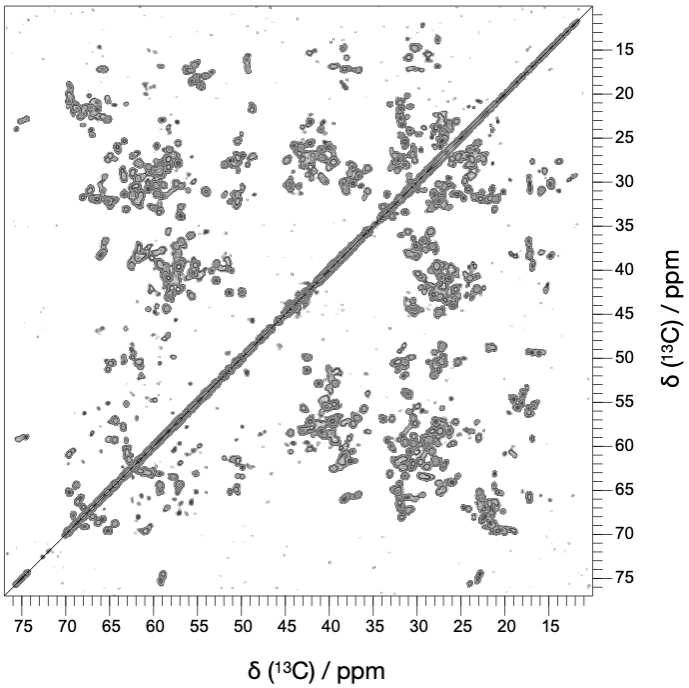

Figure 42. DARR spectrum of $^{13}C^{15}N$-Cp149 reassembled capsids in 50 mM HEPES, 5 mM DTT, pH 7.5.

When the labeled capsids have been sedimented in 1.3 mm rotors, I recorded 2D hNH spectra, which allow to transfer the magnetization from hydrogen to ^{15}N and finally back to hydrogen. An example of $^{2}H^{13}C^{15}N$-Cp149 dimer spectrum is given in **Figure 43**.

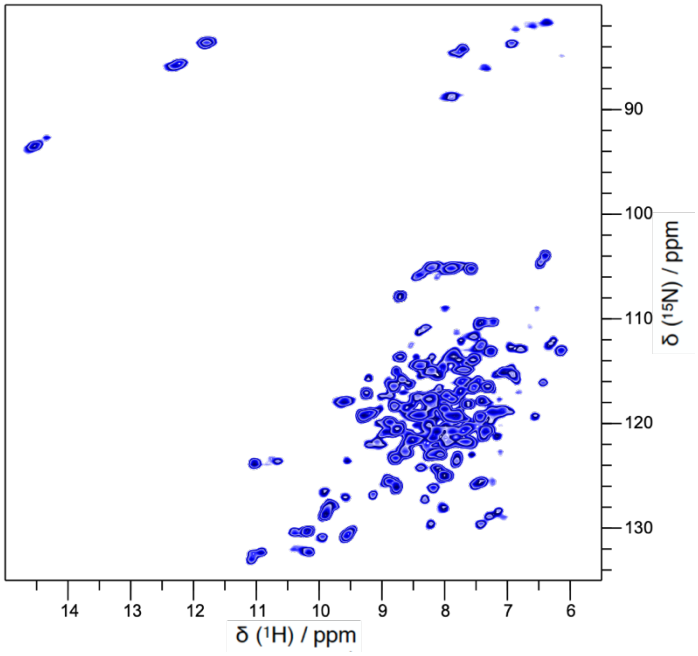

Figure 43. hNH spectrum of $^{13}C^{15}N$-Cp149 reassembled capsids in 50 mM HEPES, 5 mM DTT, pH 7.5.

V. Isothermal Titration Calorimetry

1. Principle

Isothermal titration calorimetry (ITC) is a physical technique used to determine the thermodynamic properties of an interaction in solution. It is most often used to study the binding of small molecules to larger macromolecules (Pierce et al., 1999). While the most abundant applications lie in the field of biochemical characterization, its use has increased in other fields because of minimal sample and buffer restrictions.

The principle of this technique is based on the simple measurement of heat. Because of the relationship between the change in enthalpy and the change in internal energy of a system, the heat absorbed or released is equal to the enthalpy when the system is at constant pressure.

The ITC system consists of two compartments, the reference cell and the sample cell, which are enclosed in an adiabatic jacket, meaning that there is no temperature variation (**Figure 44**). The two cells are connected to a thermoelectric device, which is sensitive to temperature variation inside the cells. This device is directly connected to a feedback power supply.

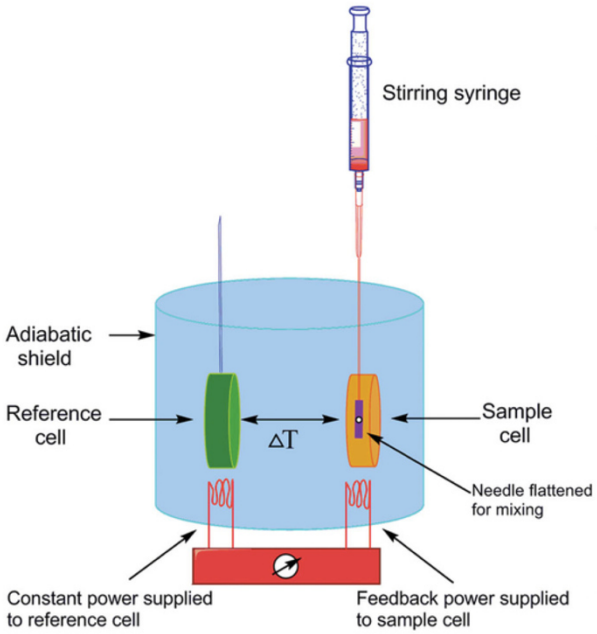

Figure 44. ITC instrument and its composition.

The protein of interest is inserted in the sample cell, while the reference cell is used as a control; it contains distillated water. The syringe is filled with the ligand which will be gradually added inside the sample cell. At the beginning, the power level between the two cells is calibrated to a specific value. Then, when a constant power is supplied to the sample cell heater, and the thermocouples detect a difference in temperature (ΔT) between the two cells, power is reduced or increased to the sample cell feedback heater proportionally to ΔT, thereby returning to the equilibrium state. In fact, during interaction event between the protein and the ligand, heat is given off in an exothermic reaction, and so the heater decreases its power to

the cell to maintain a constant temperature, resulting in a positive plotted event (the plot is following the event in the cell with heat released). A positive value indicates that this amount of power is reduced to maintain steady state. This feedback power used to maintain constant temperature is referred to as the differential power (DP). The opposite situation is true for an endothermic event.

In comparison with other biophysical techniques, ITC has several key advantages: full thermodynamic characterization in a single experiment, no need for reporter labels as heat is a universal signal, native technique as compounds can be retrieved after, interaction in solution, quite fast acquisition time, and finally possibility of performing experiment with optically dense solutions or unusual systems.

However, this technique has also major drawbacks: signal is proportional to the binding enthalpy so non-covalent complexes may exhibit rather small binding enthalpies, heat is a universal signal, and each process contributes to the global measured heat, large amount of sample is needed, kinetically slow processes may be overlooked, limitation of binding affinities, and finally no kinetic information for binding interactions could be accessed.

2. Interaction results

During the first injection, the microcalorimeter measures all the heat released until the bonding reaction reaches an equilibrium (**Figure 45**). The amount of heat measured is directly proportional to the number of bonds. Then, the molar ratio between the ligand and the protein is gradually increased by a series of ligand injections. The protein is progressively saturated, the number of bonds with the ligand decreases and the heat variation starts to decrease. At the end, the reaction has reached saturation and the sample cell contains an excess of ligand compared to the protein. After all the successive injections, the results are visible as a curve with a succession of positive or negative peaks depending on the type of interaction.

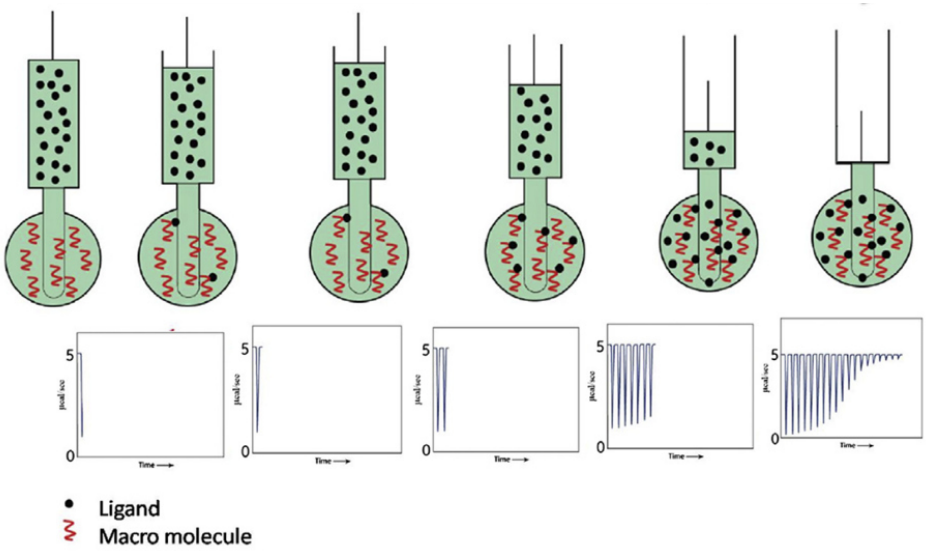

Figure 45. ITC experiment. *The reaction in the cell is directly correlated to the information collected by the computer system just below. Taken from (Srivastava, 2019)*

Afterwards, the data processing software integrate the area of each peak, and plot them against the ligand/protein ratio (**Figure 46**). The resulting isotherm can be fitted to a binding model to extract the thermodynamic data. The values extracted under isothermal include the affinity constant (K_A), binding stoichiometry (n), changes in free energy (ΔG˚), enthalpy (ΔH˚) and entropy (ΔS˚). The molar ratio taken at the center of the binding isotherm gives the stoichiometry of the reaction. The affinity constant and the enthalpy are deduced directly from the isotherm and are respectively the slope value of the line that intercepts the exponential phase of the isothermal curve and the amount of heat released per mole of bound ligand.

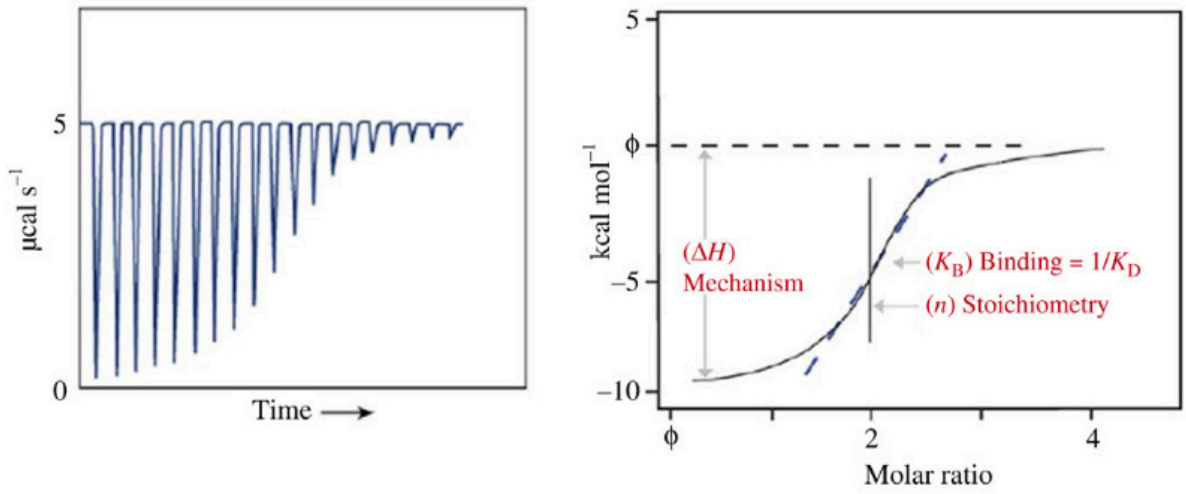

Figure 46. Thermodynamic data obtained from ITC curve. *Taken from (Srivastava, 2019)*

RESULTS

RESULTS I

I. Introduction

As described int the introduction, the capsid-forming core protein is the building block of nucleocapsids and is involved in many crucial steps of the HBV life cycle, it is therefore emerging as a new antiviral target (Diab et al., 2018; Viswanathan et al., 2020). A new and unexplored option to interfere with nucleocapsid dynamics in general, and nucleocapsid envelopment in particular, is offered by the hydrophobic pockets formed in the Cp dimer at the base of the spike (**Figure 47A**), which we have recently shown by solid-state NMR to bind the detergent TX100 (Lecoq et al., 2021), followed by the resolution of cryo-EM structure of the complex (Makbul, Kraft, et al., 2021). Binding causes well-detectable conformational alterations in the capsid structure which essentially localize to a region implicated in HBV nucleocapsid envelopment, *i.e.* virion formation.

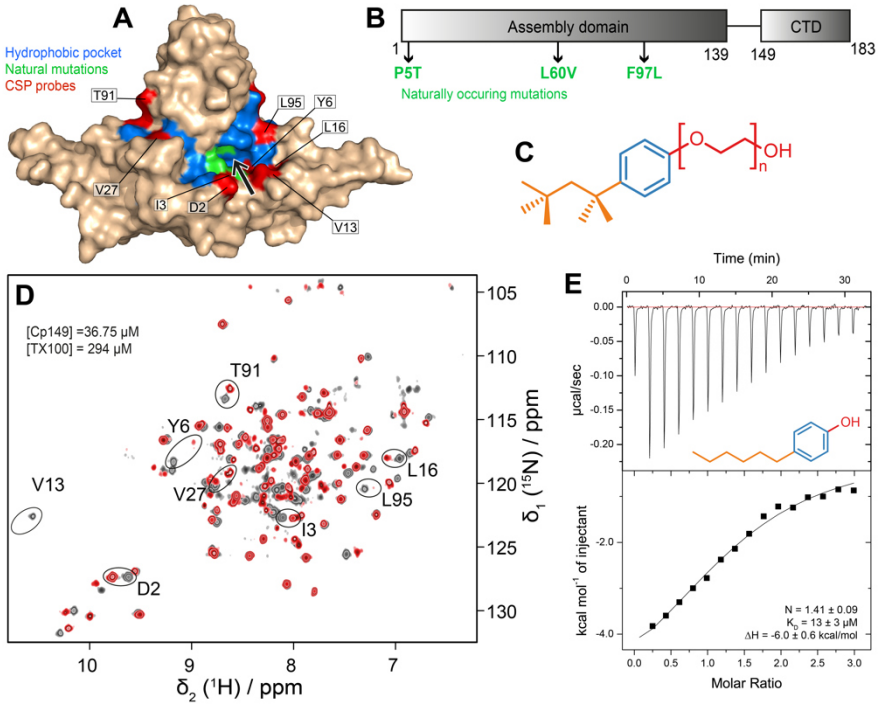

Figure 47. Workflow to assess binding to the hydrophobic pocket. *A) Structure of the Cp149 dimer in surface representation. The residues of the hydrophobic pocket impacted by the interaction of TX100 (Lecoq et al., 2021) are highlighted in blue. The eight residues whose NMR signals were used to quantify CSPs are highlighted in red. The pocket entrance is indicated by an arrow. Naturally occurring mutations are plotted in green. B) Sequence of the full-length core protein (Cp183) with mutations highlighted. C) Structure of the TX100 molecule (compound **1a**), with the hydrophilic tail in red, the aromatic moiety in blue, and the alkane part in para-position in orange. D) Solution-state NMR spectra of the $^{2}H^{13}C^{15}N$-Cp149 apo dimer in grey, and with eight equivalent TX100 in red, with peaks used to assess CSPs circled and labelled with their amino-acid type and position in the sequence. The assigned apo spectrum is shown in **Figure 54**, and the full set of CSPs between the two spectra are shown in **Figure 55**. E) ITC raw data and binding isotherms displaying the titration of 4-hexylphenol (compound **2a**) against Cp149 reassembled capsids at pH 7.5 and 298 K (K_D = 13 ± 3 µM). For an overview of ITC runs see **Figure 63**.*

Envelopment occurs when the initially encapsidated viral pregenomic pgRNA is reverse transcribed, inside the capsid, into mature, double-stranded DNA, signalling readiness of the capsid for interaction with the envelope proteins. The mechanism of this regulation is still undetermined, although several hypotheses have been proposed (Lecoq et al., 2021; Ning et al., 2011; Summers & Mason, 1982; Xi et al., 2022). Indeed, envelopment is modulated by several naturally-occurring Cp mutations (K. Li et al., 2007; Yuan et al., 1999), many of them being located in and around Cp's hydrophobic pocket, such as F/I97L, P5T and L60V (**Figure 47A-C**). Mutations in other envelopment-disabling positions also localize to the hydrophobic pocket, namely at L95 and K96 (Ponsel & Bruss, 2003). Interestingly, uncoupling of envelopment from genome maturation by the F/I97L mutation can be reverted by the A119F mutation in the preS envelope protein domain (Le Pogam et al., 2000). In all cases, the most plausible interpretation for the mutationally altered envelopment phenotypes is a modification of the core-envelope interaction. In addition to the hydrophobic pocket in the Cp spike base, the spike tip has also been identified as interacting domain, as indicated by inhibition of envelopment by spike-binding peptides (Böttcher et al., 1998; Makbul, Khayenko, et al., 2021). The detailed binding site(s) of the envelope proteins on the core protein, as well as a molecular view of the interaction, remain to be determined.

We have recently found that TX100 binds to the hydrophobic pocket in Cp (Lecoq et al., 2021). Indeed, a conformational change of the hydrophobic pocket of the Cp has been observed between two forms of capsids: the form A, also called *E.coli* capsids as they are the ones obtained after production and purification; the form B, also called reassembled capsids, which are capsids obtained after disassembly and reassembly. The chemical shift differences (CSDs) between these two forms are located all around the hydrophobic pocket (**Figure 48**).

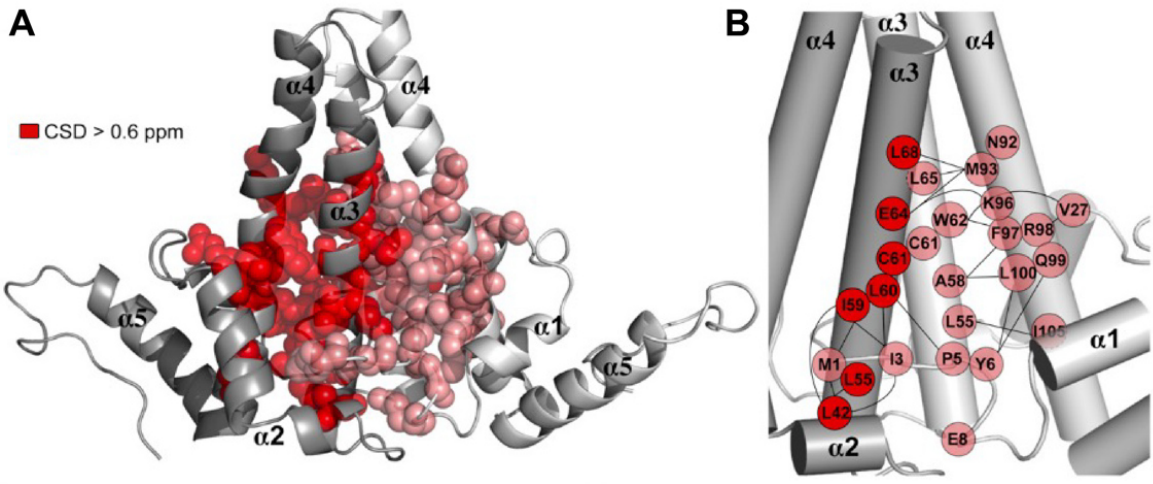

Figure 48. Residues impacted by TX100 binding. *The different chains of the dimer are represented by dark gray and red for chain C, and light gray and light red for chain D. A) High CSDs (> 0.6 ppm) are plotted on the dimer, represented in cartoon, and residues in spheres. B) Distances (≤ 4 Å) between residues with high CSDs are shown by black lines. Taken from (Lecoq et al., 2021)*

In order to find what was the reason of this conformational switch, an extraction was performed using chloroform/methanol (Lecoq et al., 2021). The chloroform phase was analyzed by NMR, and revealed that the molecule correspond to the structure of TX100, a detergent used during the purification step of Cp production in *E.coli* expression system. The structure was confirmed by the comparison with the spectra of commercial TX100 which fully overlay with the molecule in the chloroform solution.

II. Results

After having identified that TX100 binds inside the hydrophobic pocket of Cp, we were interested in characterizing the interaction by ITC (**Figure 49**). These experiments represent my contribution to the work described in (Lecoq et al., 2021). We found that TX100 was interacting with Cp at the micromolar range, with a dissociation constant (K_D) of 8.3 ± 0.5 µM, and with a stoichiometry of one molecule of TX100 per Cp monomer (*i.e.* two molecules per Cp dimer) and so two TX100 with the dimer (the stoichiometry (n) is about 0.89, and represents the number of molecule per Cp monomer).

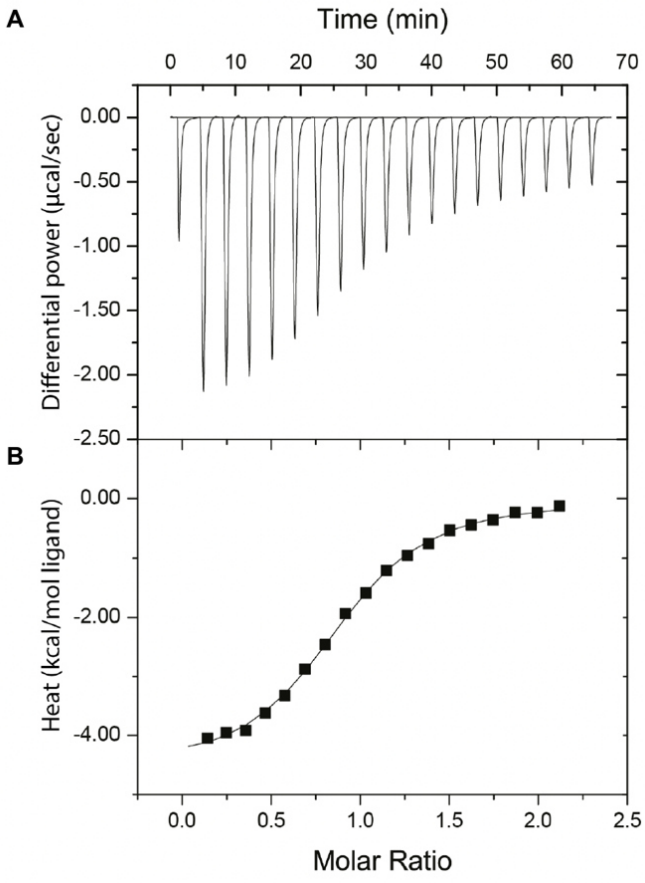

Figure 49. ITC thermograph of Cp149 reassembled capsids titrated by TX100. *A) Raw ITC isotherms. B) Binding curve at pH 7.5 and 298K. the best fit give the following results: n=0.89 and K_A=1.20 ± 0.07 *10^5 M^{-1}. Taken from (Lecoq et al., 2021)*

Our finding thus opens a new avenue to target the capsid-envelope interaction. A first step towards by establishing this molecule as a lead compound for the development of new antivirals being capable of blocking capsid envelopment was to investigate which moieties were central for binding, and which variations thereof were tolerated. We thus investigated the pharmacophore of Cp hydrophobic pocket binders, in order to establish how the three chemical moieties of the molecule contribute to capsid binding. This work was one of the major topics of my thesis work.

To test a large number of compounds, we here used for this solution NMR on the Cp dimers instead of solid-state NMR of the capsids for the sake of higher throughput. The residues impacted by TX100 binding (Lecoq et al., 2021), and with isolated signals in the spectra are highlighted in blue in **Figure 47A**. The chemical variations we explored are shown in **Figure 47C**, and concern firstly the TX100 highly hydrophobic branched aliphatic chain carrying five methyl groups (in orange); secondly, the aromatic ring (in blue); and thirdly the $(CH_2\text{-}O)_n$ hydrophilic tail (in red), connected to the aromatic ring in the *para* position to the aliphatic chain.

We tested in this work the capsid's hydrophobic pocket binding capacity for compounds with variations in the three TX100 moieties. To explore the chemical space of the pocket, we used 28 *de novo* synthesized compounds and 12 commercially available molecules (for a complete overview see **Figure 50**).

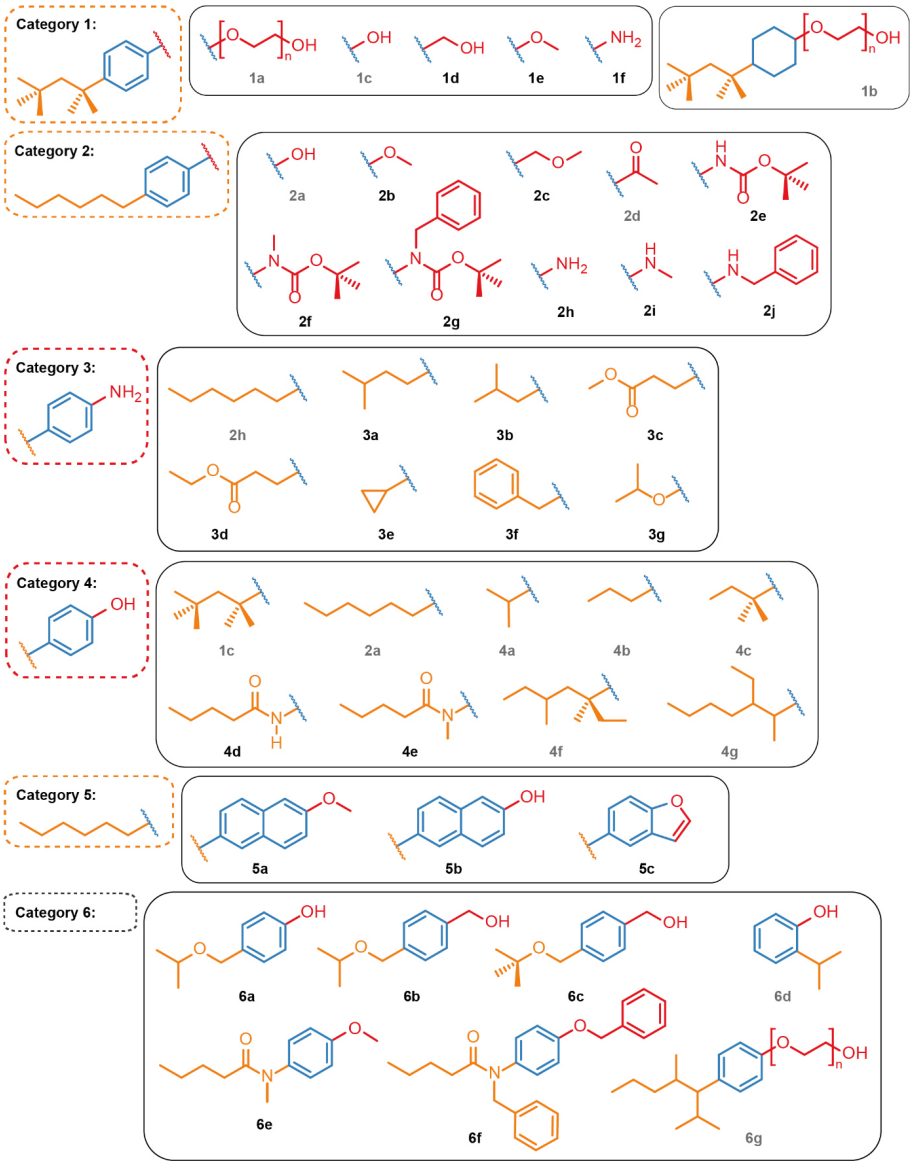

Figure 50. All molecules used in this study. *They are categorized in function of the moiety changed compared to Triton X-100, the compound of reference (**1a**).*
*Commercial compounds are indicated with grey labels, and are: **1a**, Triton X-100 (TX100); **1b**, Reduced Triton X-100; **1c**, 4-tert-octylphenol (OP); **2a**, 4-hexylphenol (4HF); **2d**, 4'-heptylacetophenone; **2e**, 4-hexylaniline; **4a**, 4-isopropylphenol; **4b**, 4-propylphenol; **4c**, 4-tert-amylphenol; **4f**, 4-(3,5-dimethyl-3-heptyl)phenol; **4g**, 4-(2-ethyl-1-methylhexyl)phenol; **6d**, 2-isopropylphenol; **6g**, NP40.*

To assess binding, we first used solution-state NMR on the truncated Cp dimer (Cp149), which allowed to speed up analysis of the large number of compounds compared to solid-state NMR of assembled capsids (Lecoq et al., 2018, 2020). As shown in the example spectra in **Figure 47D**, chemical-shifts perturbations (CSPs) are a clear indicator for TX100 binding, and represent, when the exchange rate is in slow regime, the chemical shifts of the bound form. This is for example the case for TX100 and 4-*tert*-octylphenol (OP) (compounds **1a** and **1c**) as can be seen in the spectra in **Figure 51 and Figure 52** where, upon titration on Cp149 dimer, only peaks of the free and the bound forms can be observed.

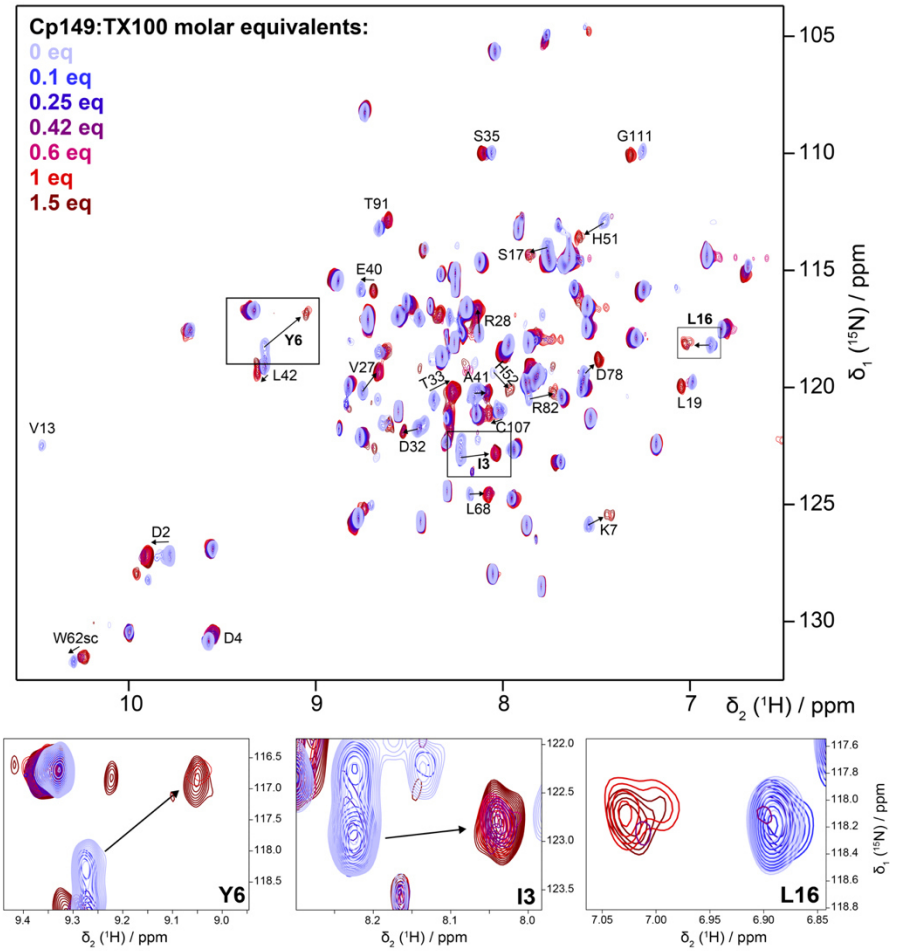

Figure 51. Titration of TX100 (1a) on Cp149 dimer ([Cp149$_{monomer}$] = 140 µM, from 0 to 1.5 molar equivalents monomer:TX100). *The residues which move the most are indicated by an arrow. A zoom is shown for residues Y6, I3 and L16. At the beginning of the titration (0, 0.2 and 0.5 equivalents, blue gradient spectra), the chemical shift observed correspond to the free form and intensities start to decrease. Then, at 0.85 equivalents (purple spectrum), the signals broaden and in some cases are not visible (Y6) or the two forms are visible simultaneously (L16). Finally, above 1.2 equivalents (pink to brown spectra), the intensity of the signal corresponding to the bound form increases. The NMR signal broadening is an indication that the interaction occurs in the slow exchange regime, which usually occurs when the K_D becomes stronger than 10 µM (Williamson 2013).*

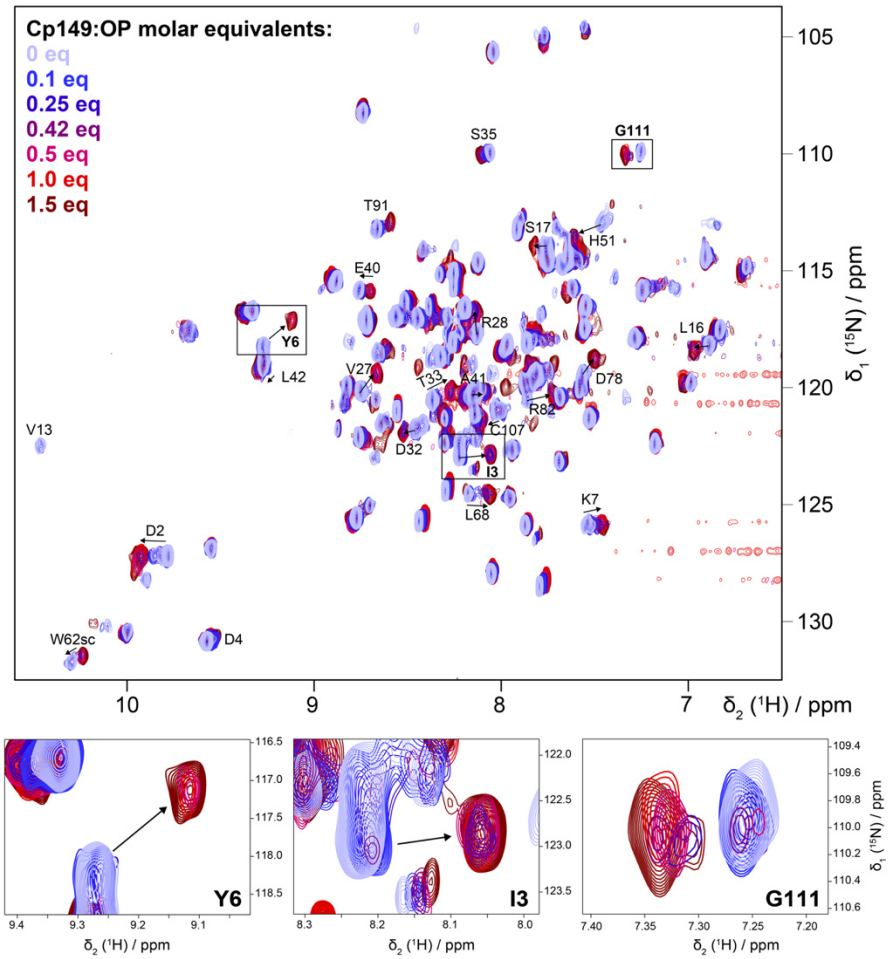

Figure 52. Titration of the Cp149 dimer with OP (1c) ([Cp149monomer] = 140 µM, from 0 to 1.5 molar equivalents monomer:OP). *The residues that move the most are indicated by an arrow. A zoom is shown for residues Y6, I3 and G111. The behaviour of the NMR signals is substantially similar to what was observed upon titration with TX100 (**Figure 51**), with an apparent slow exchange regime.*

For compounds with lower affinity (higher K_D) resulting in a fast exchange regime, peaks can be found at intermediate positions. This is the case for 4-hexylphenol (4HF, compound **2a**) for which a titration on Cp149 is shown in **Figure 53**. In this case, if one considers that the compounds have similar binding modes and contacts, the extent of the CSPs induced by the different compounds can be correlated with the strength of binding.

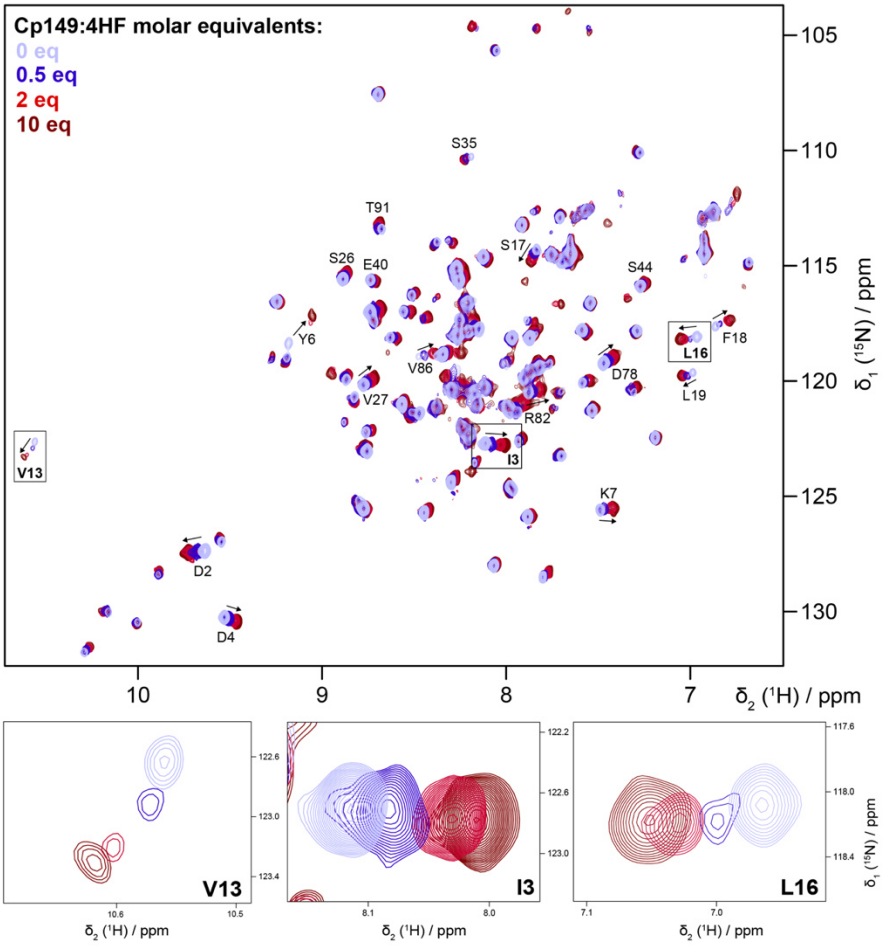

Figure 53. Titration of Cp149 dimer with 4HF (2a) ([Cp149] ≈ 80 µM, from 0 to 10 molar equivalents monomer:4HF). *The residues that move the most are indicated by an arrow. A zoom is shown for residues V13, I3 and L16. Contrary to TX100 and OP, here the titration indicates intermediate to fast exchange, where the observed chemical shift change is the weighted average of the shifts in the free and bound states (Williamson, 2013).*

Where binding constants were in the adequate range, we complemented the NMR data by isothermal titration calorimetry (ITC) measurements to quantify binding affinities with Cp149 reassembled capsids, as shown in the example of 4HF in **Figure 47E**. ITC reveals that only compounds **1a** and **1c** have a K_D below 10 µM, while all other compounds tested have lower affinities (see **Table 5**) and are therefore, as shown for 4HF (**Figure 53**), in fast exchange. Thus, one can consider that CSPs are representative of the binding strength at same molar ratios (4 molar equivalents) (Williamson, 2013).

Chemical-shift assignments of the Cp149 dimer at pH 7.5, used as reference for CSP calculations, are shown in **Figure 54** and were deposited at the BMRB under accession number 51294. The assigned reference spectrum is shown in **Figure 54A**.

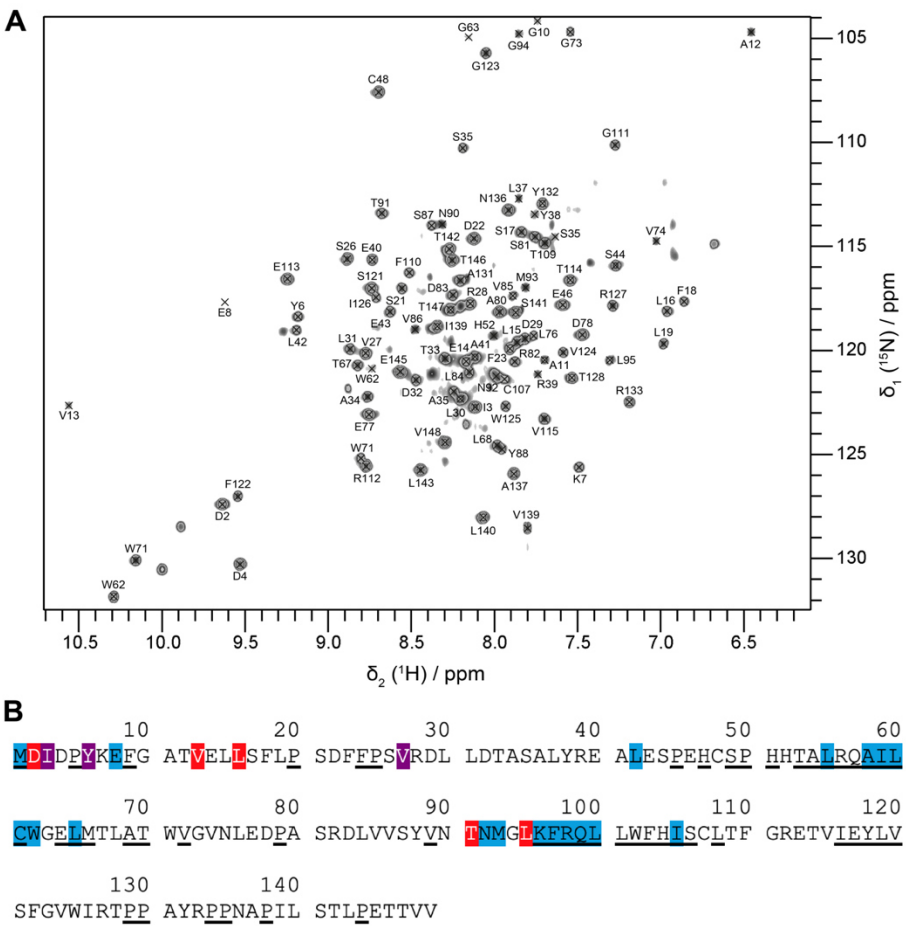

Figure 54. Cp149 assignments in solution NMR. *A) Solution-state NMR spectrum of the Cp149 dimer. B) Assignment reported on the sequence of Cp149. The residues of the hydrophobic pocket impacted by the interaction of TX100 are highlighted in blue. The eight residues whose NMR signals were used to quantify CSPs are highlighted in red, or purple if they are also part of the blue category, namely D2, I3, Y6, V13, L16, V27, T91 and L95. Unassigned residues are underlined. Resonance backbone assignments were deposited under BMRB accession number 51294.*

Unlike in solid-state NMR analysis, where we mainly used ^{13}C detection on protonated samples (Lecoq et al., 2021), in solution NMR not all amino acids of the core protein are visible due to incomplete back exchange of several NH in the hydrophobic core of the Cp dimer after expression in deuterated medium. While this prevented the use of some NH localized directly in the hydrophobic pocket, we selected eight amino acids located near the entry of the pocket (in red in **Figure 47A**), which can still be detected to measure CSPs (as shown on the sequence in **Figure 54B**). As exemplified in **Figure 55** for TX100, CSPs can then be derived upon addition of each ligand.

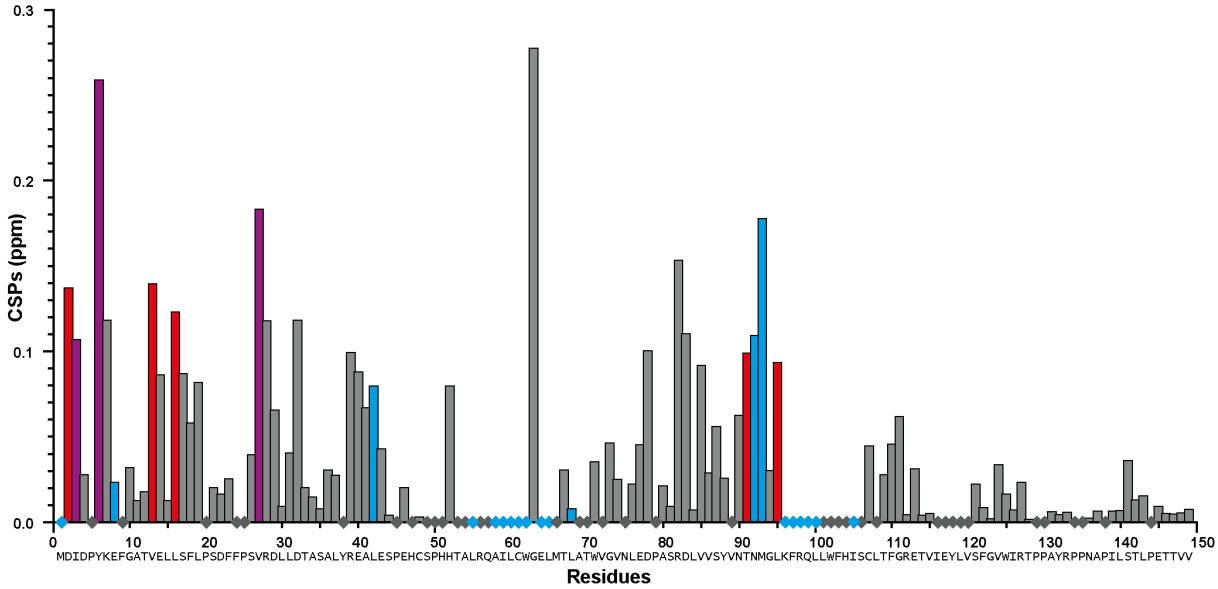

Figure 55. Full ^{1}H-^{15}N-CSP graph induced by TX100 (1a) on Cp149 dimer in solution NMR. *CSPs were calculated by comparing ^{1}H-^{15}N-BTROSY spectra of $^{2}H^{13}C^{15}N$-Cp149 dimer alone and with 4 molar equivalents of TX100 ($[Cp149_{monomer}]$ = 80 µM and [TX100] = 320 µM). The unassigned residues are shown with a small diamond and are mainly due to incomplete amide back-exchange in the solvent-protected regions. The residues of the hydrophobic pocket the most impacted by the interaction of TX100 according to solid-state NMR data are highlighted in blue (Lecoq et al., 2021). The eight residues used to establish binding efficiency are highlighted in red, or purple if they are also part of the blue category.*

1. The aromatic moiety is central for pocket binding

First, to investigate the importance of the aromatic moiety for binding, we tested the interaction of reduced TX100 (compound **1b**), shown in **Figure 56A**, where the aromatic ring is replaced by a cyclohexane moiety. The solution NMR spectrum recorded on Cp149 with compound **1b** only showed very small CSPs (**Figure 56B**) (for full spectra, see **Figure 64-Figure 69**), and ITC did not yield meaningful information. This suggests the central importance of the aromatic moiety to yield a productive interaction with the hydrophobic pocket. We thus considered in the following only compounds conserving the aromatic ring of TX100.

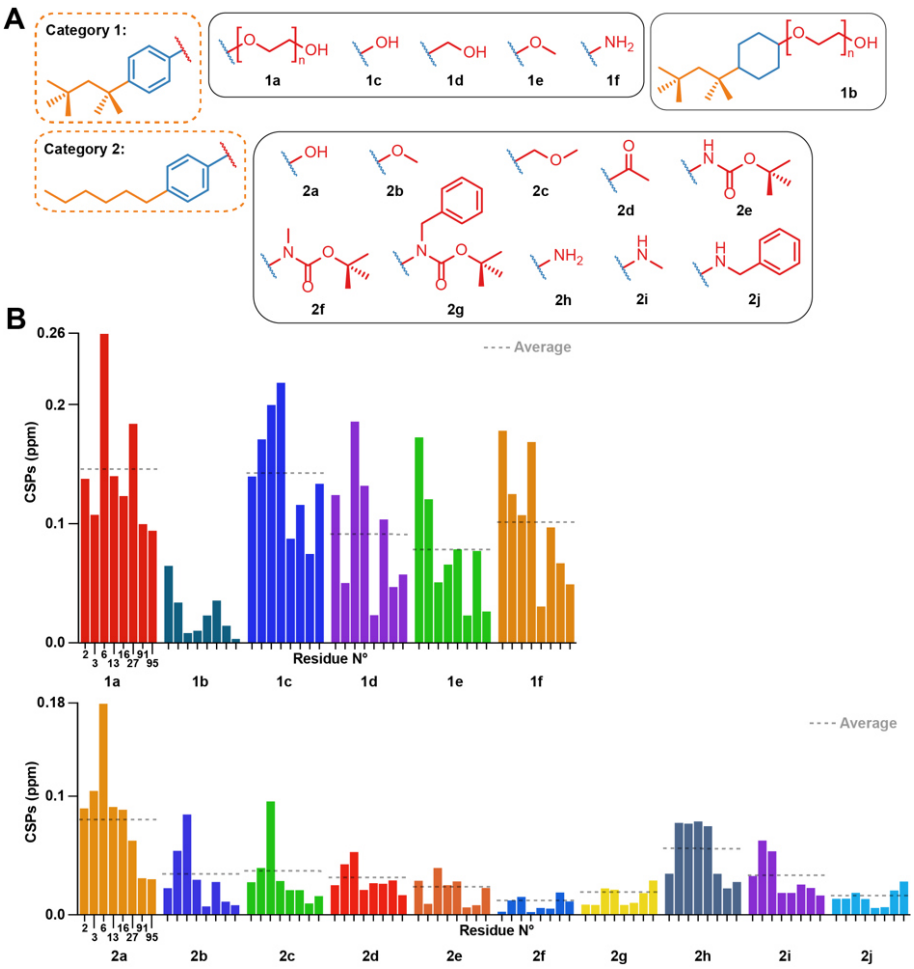

Figure 56. Effects of modulations of the TX100 aromatic ring and hydrophilic tail*. A) Chemical structures of compounds tested from categories 1 and 2. B) NMR CSPs induced by the compounds for the eight selected capsid resonances (for full spectra see* ***Figure 64 and Figure 65****). CSP averages are indicated by dotted lines. Residue numbers of the representative peaks are given below the CSPs; all CSPs throughout the manuscript are shown at the same scale.*

2. The hydrophilic tail is dispensable for pocket binding, and can be replaced by -OH, and to a lesser extent, by CH_2-OH or -NH_2

We next determined whether the hydrophilic chain of TX100 is important for binding to the hydrophobic pocket. **Figure 56A** shows the compounds used grouped in two categories, one carrying a *tert*-octyl group as hydrophobic moiety (category 1), the other a linear hexyl chain (category 2), which resulted in slightly higher water solubility. The NMR spectra of the Cp149 dimer in the presence of all listed compounds (**Figure 64 and Figure 65**) resulted in the CSPs shown in **Figure 56B**, with those from categories 1 and 2 shown in separate plots. Category 1 compounds (except **1b**) showed high CSPs values, with averages (indicated by dotted lines) higher than 0.05 ppm when compared to the apo state. In particular, compound **1c**, namely OP, gave CSPs as high as those derived for TX100. Category 2 compounds induced consistently smaller CSPs, indicating that the hexyl moiety binds less strongly than

the *tert*-octyl moiety. Intra-category comparisons revealed that replacement of the polyethylene glycol (PEG) chain by -OH and -NH_2 impacts binding, but only to a small extent, a shared result for both categories. However, replacing the PEG chain by bulkier groups (compounds **2e**, **2f**, **2g**, **2j**) weakened the interaction. The presence of a possible hydrogen-bonding partner could thus reveal another element of the pharmacophore. The binding parameters of the different compounds with the capsid determined by ITC are reported in **Table 5**. Individual curves are shown in **Figure 63**.

Table 5. Thermodynamic parameters derived from ITC measurements. *Data are shown for Cp149 reassembled capsids with 7 different compounds, including the K_D and the enthalpy (ΔH). For compounds **1b**, **2d**, **2e**, **4f**, **4g**, and **6d**, ITC did not allow to determine binding parameters (see **Figure 63**). The TX100 value is taken from previous data published in (Lecoq et al., 2021).*

Ligand	K_D (μM)	ΔH (kcal mol^{-1})
1a	8.3 ± 0.5	-4.59 ± 0.05
1c	7.5 ± 0.5	-4.79 ± 0.14
2a	13 ± 3	-6.0 ± 0.6
4a	18 ± 7	-5 ± 3
4b	14 ± 7	-7 ± 4
4c	30 ± 4	-12.7 ± 1.2
6g	68 ± 7	-3.41 ± 0.09

In general, the data showed that the enthalpies for all compounds are negative and in the same range, indicating that the interaction is exothermic. The results confirmed compounds **1c** and **2a**, both bearing a hydroxyl group instead of the PEG tail, as best alternative binders to TX100. Hence, while the hydrophilic moiety as a whole plays a minor role in binding, the position of the first oxygen in the chain is important. H-bond donor is required as methyl ether (**2b**), methoxymethyl (**2c**) and ketone (**2d**) resulted in rather small CSPs, and no measurable interaction by ITC. The positioning of bulky groups, like those in compounds **2e**, **2f**, **2g** and **2j**, resulted in the smallest CSPs, *i.e.* less efficient binding.

3. Fully hydrophobic aliphatic, aromatic and functionalized chains can successfully replace the tert-octyl group

We next assessed variations in the hydrophobic element. We investigated compounds featuring different chain lengths and branching, and insertions of oxygen/nitrogen atoms, assigned to categories 3 and 4 (**Figure 57A**). We started from both -OH and -NH_2 groups instead of the hydrophilic tail, based on their previously assessed roughly similar binding affinities. We investigated different branched hydrocarbon chains, but also amide, ester, and ether groups to establish whether their polarity would favour binding to the accessible main chain amide or carboxyl groups in the capsid's hydrophobic pocket. **Figure 57B** displays the CSPs measured for categories 3 and 4 (for full spectra, see **Figure 66 and Figure 67**). The CSPs indicate an optimal atom number of four to eight carbons, as indicated by the similar

CSPs induced by compounds **1c**, **2a**, **2h**, **3a**, **3b** and **4c** (average CSP between 0.046 and 0.142 ppm) (see **Table 6** for the list of all averaged CSPs and **Table 7** for solution NMR parameters). Three-carbon chains reduced CSPs to below 0.050 ppm, as observed for compounds **4a** and **4b**. When long alkyl chains were introduced *para* to the phenolic OH group, the interaction also weakened, as seen for compounds **4f** and **4g**. The CSPs also revealed that investigated groups with heteroatoms did not establish new favourable interactions, as demonstrated by the small CSPs of compounds **3c**, **3d**, **3e**, **4d**, and **4e** (average CSP < 0.029 ppm). Surprisingly, while a cyclopropyl moiety (**3e**) did not bind to the pocket, a benzyl moiety (**3f**) displayed substantial binding, as indicated by a CSP of 0.071 ppm. ITC measurements showed comparable affinities (between 14 and 30 µM) for compounds **4a**, **4b**, and **4c**, and no measurable interactions for **4f** and **4g,** consistent with the NMR data.

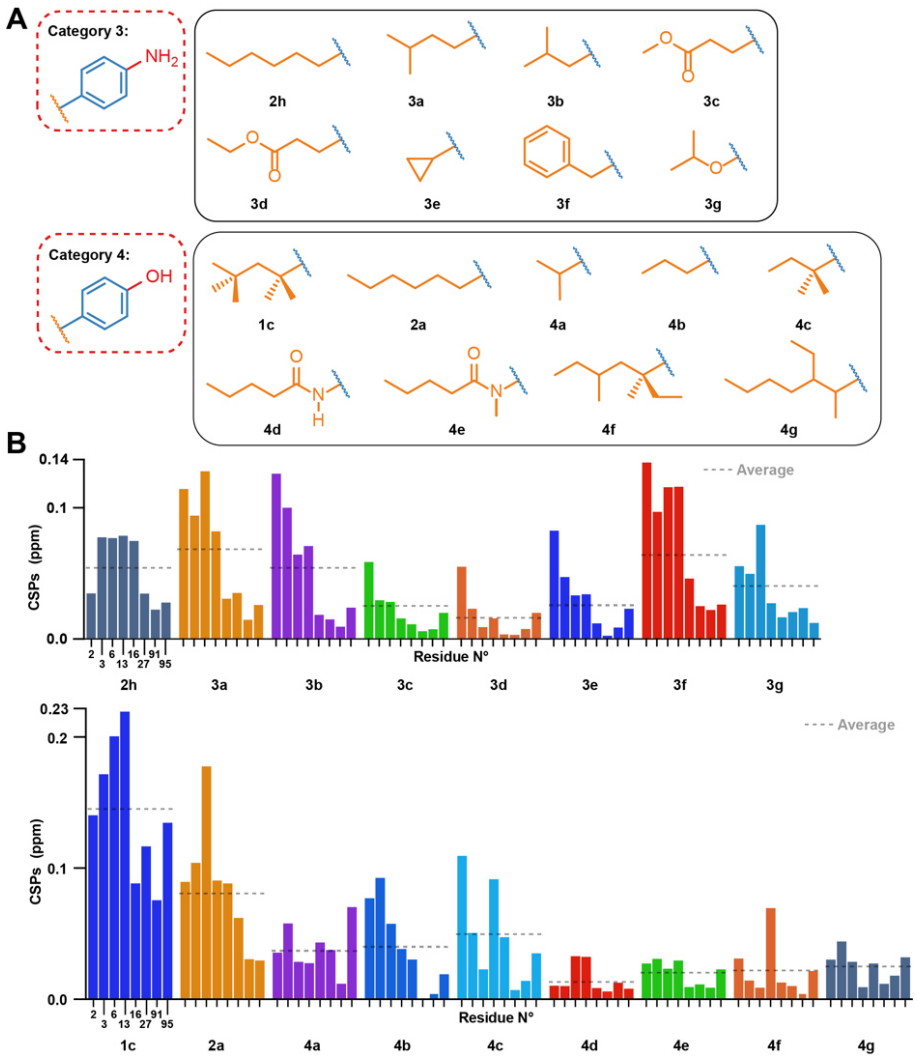

Figure 57. Effects of modulations of the TX100 hydrophobic moiety*. A) Compounds tested from categories 3 and 4. B) NMR chemical-shift perturbations induced by compounds on the eight selected capsid resonances (for full spectra see* ***Figure 66 and Figure 67****). CSPs averages are indicated by dotted lines.*

4. A single aromatic ring is essential for binding, and hydrophilic/hydrophobic groups need to be in para position

We further investigated the impact of increasing the size of the aromatic moiety. In the respective category 5 compounds (**Figure 58A**), the aromatic part is replaced by a naphthalene (**5a-b**) or a benzofuran (**5c**) ring. The CSPs for these molecules are very small (**Figure 58B** upper panel; for full spectra see **Figure 68)**, showing that fused aromatic ring systems block the interaction, disabling binding of compounds **5a**, **5b** and **5c** into the hydrophobic pocket of Cp.

Subsequently, we also established a variety of simultaneous modifications of the hydrophobic and hydrophilic groups (category 6). The only compound of this category displaying a significant interaction with the capsid was **6c**, with an average CSP of 0.048 ppm (for full spectra, see **Figure 69**). Analysis of CSPs induced by **6b** and **6c** bearing both a benzyl alcohol on the hydrophilic part shows that a terminal *tert*-butyl group on the hydrophobic chain seems to favor the interaction, since the CSPs for **6c** are more than twice as high as for **6b**. Yet, CSPs for **6c** still remain lower than for **1d**, confirming the even higher impact of the 4-*tert*-butyl moiety on binding. Both compounds **6d** and **6g** were tested by ITC (**Table 5**), the latter displaying weak affinity, and the former no measurable data. The lack of affinity of **6d**, compared to its *para* isomer **4a** is revealed by ITC (**Figure 49**), which highlights that the relative position of the hydrophilic and hydrophobic groups is central for the interaction. NMR however report small, but, within the error bars, same CSPs for **4a** and **6d**.

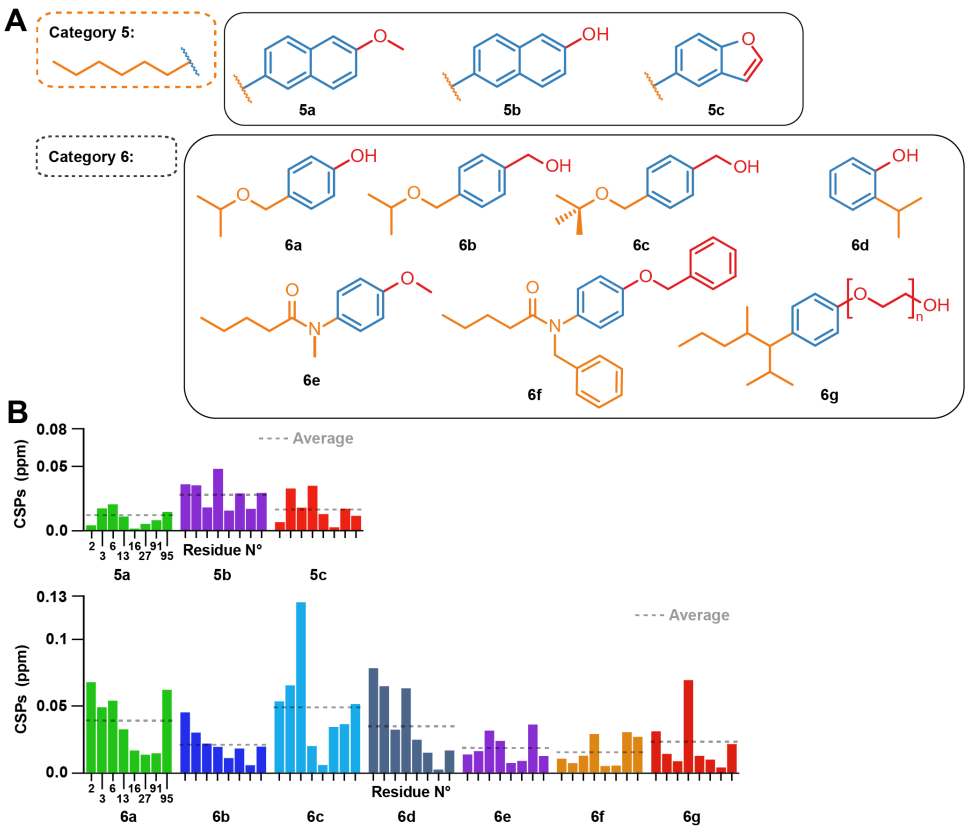

Figure 58. Effect of modulations of the TX100 aromatic ring, and concurrently on the hydrophilic and hydrophobic groups. *A) Compounds tested from categories 5 and 6. B) NMR chemical shift perturbations induced by the compounds on the eight selected capsid resonances (for full spectra see* ***Figure 68 and Figure 69****). CSPs averages are indicated by dotted lines.*

5. Docking of TX100 and OP on the core protein dimer

We used the information derived from the binding study to dock the 4-*tert*-octylphenol (OP) and TX100 molecules to the Cp hydrophobic pocket. For this purpose, we defined a flexible box comprising residues which showed significant CSPs on binding of TX100 (Lecoq et al., 2021). We computed multiple binding models and, based on our NMR and ITC results, selected those where the hydrophilic moiety is positioned outside the hydrophobic pocket. **Figure 59A and Figure 59B** show the best models obtained by docking respectively molecule **1c** (OP) and **1a** (TX100). In both models, the *tert*-octyl chain share a similar position inside the pocket. However, the aromatic rings are oriented differently, with the PEG tail from TX100 protruding outside the pocket near I59-L60 and the N-terminus of the protein (M1 to Y6) (**Figure 59A**), while the phenol of OP is rather directed towards Q99, V13, L16 and S17 (**Figure 59B**). Comparison of both docked models with the recent cryo-EM structure of the HBV capsid in presence of TX100 (PDB 7PZK) (Makbul, Kraft, et al., 2021) is shown in **Figure 59C and Figure 59D**. In the cryo-EM structure, the hydrophobic moiety of TX100 can insert deeper in the pocket as permitted by the rotation of F97 aromatic ring identified in (Makbul, Kraft, et al., 2021). The docking with OP (**Figure 59D**) likely represents a more realistic orientation since it is found closer to the cryo-EM structure than docking with TX100 (**Figure 59C**). Yet, even if the docking did not predict the F97 sidechain rotation, it correctly modelled the reorientation of K96 sidechain in the case of TX100, where it aligns well with the cryo-EM structure (**Figure 59C**).

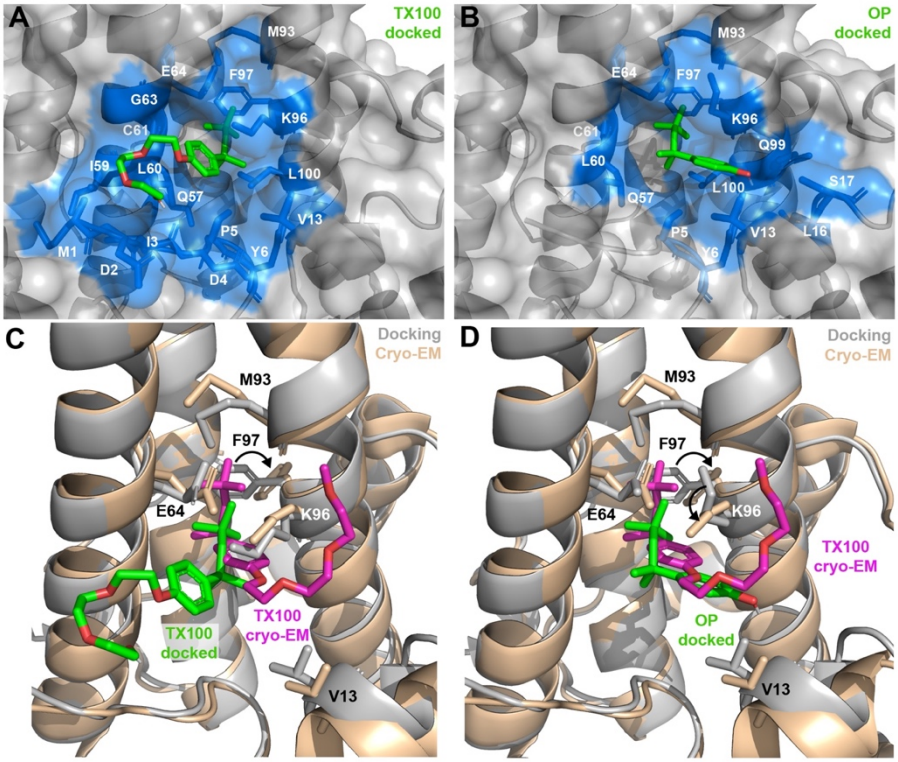

Figure 59. Computed models of OP and TX100 bound in the hydrophobic pocket of the Cp149 dimer. *Docked model of A) OP (compound **1c**) and B) TX100 (**1a**) in the hydrophobic pocket of Cp149 dimer. Residues located in the direct proximity of ligands (< 5 Å) are represented in blue sticks. C-D) Comparison of bound TX100 in the cryo-EM structure (in magenta, protein in wheat, PDB 7PZK) (Makbul, Kraft, et al., 2021) with the docked models (in green, protein in grey) for C) TX100 and D) OP. Neighbouring residues V13, E64, M93, K96 and F97 are shown as sticks. In both cryo-EM structure and docked models, TX100 contains n = 3 ethylene oxide units in the PEG tail.*

6. Solid-state NMR CSPs can distinguish the conformational changes caused by the different chemical entities

The CSPs of the bound state reveal the impact of the binder on the protein structure, especially when ^{13}C CSPs are considered, which mostly reflect changes in backbone dihedral angles. In a first approximation, small carbon CSPs reflect small changes, and large carbon CSPs substantial impact on the structure. Amide proton and nitrogen chemical shifts, in contrast, are more sensitive to differences in hydrogen bonding and chemical environment variations. We thus established the structural changes using ^{13}C CSPs for compounds **1c** and **2a**, and compared them to those induced by TX100. Extracts from ^{13}C-^{13}C-dipolar assisted rotational resonance (DARR) spectra for P5, A58 and R98 are shown in **Figure 60A,** and ^{13}C-CSPs for these and other residues of interest are shown in **Figure 60B** for the three compounds. Full DARR spectra are shown in **Figure 61A** and more extracts in **Figure 61B**.

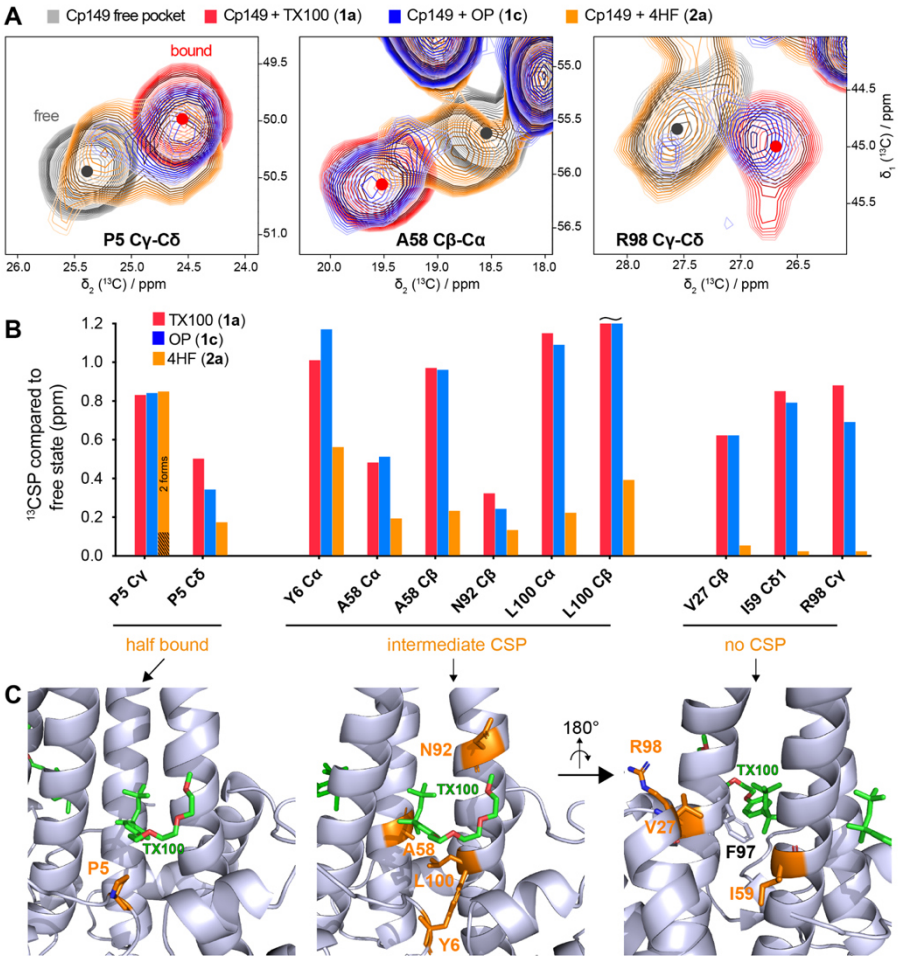

Figure 60. Solid-state NMR can probe the impact of the ligand's hydrophilic tail on the hydrophobic pocket*. A) Comparison of 2D extracts from ^{13}C-^{13}C-DARR solid-state NMR spectra of Cp149 capsids reassembled without compounds (empty pocket, grey), and in presence of TX100 (bound pocket, red), OP (blue), or 4HF (orange). Full aliphatic regions of DARR spectra are shown in* ***Figure 61****. B) Comparison of ^{13}C-CSPs induced by the three compounds. Three groups could be identified based on their behaviour when bound to 4HF: P5 at the entrance of the pocket showing a mixture of free/bound pocket; the residues showing intermediate CSPs; and the residues showing no CSPs with 4HF. C) Structure of Cp183-TX100 (PDB 7PZK) (Makbul, Kraft, et al., 2021) with residues involved in the three CSP classes shown in orange. The right panel is rotated by 180° to display the bottom of the hydrophobic pocket and the proximity with F97 sidechain.*

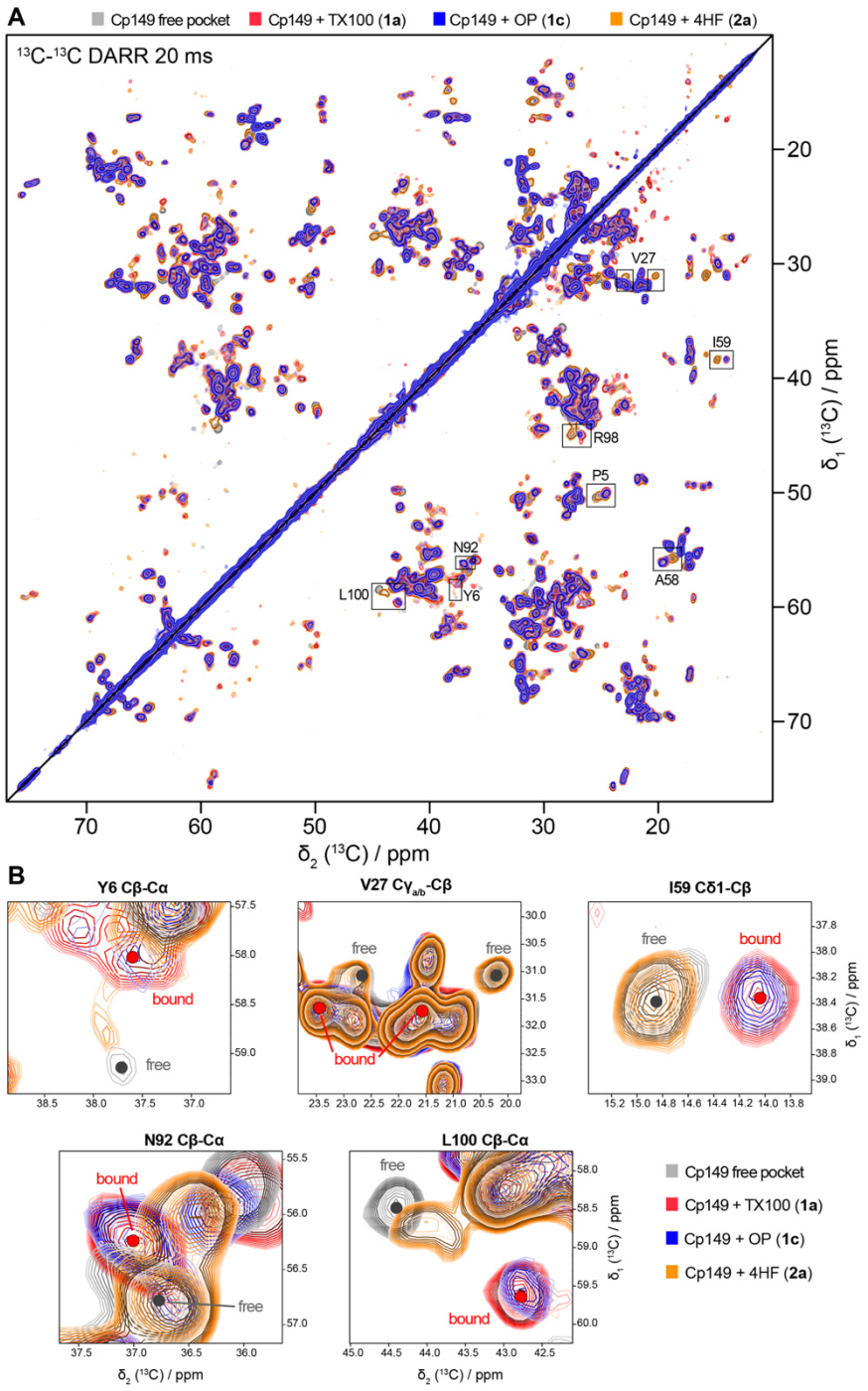

Figure 61. Solid-state NMR spectra comparison of free and bound pocket. *A) Comparison of the aliphatic region of 2D ^{13}C-^{13}C-DARR solid-state NMR spectra of Cp149 capsids reassembled in absence of compounds (empty pocket, in grey), Cp149 capsids reassembled with TX100 (bound pocket, in red), with 4-tert-octylphenol (OP, in blue) and with 4-hexylphenol (4HF, in orange). B) Extracts of 5 correlation peaks from the 2D DARR spectra. Extracts from P5, A58 and R98 are shown in the main text in Figure 6, together with the CSPs.*

Interestingly, OP induces CSPs highly similar to binding of TX100, with spectra almost perfectly overlapping for most residues. This confirms that the PEG tail is not involved in the interaction, and that replacement by a hydroxyl group has almost no impact, even on residues surrounding the pocket. On the contrary, the binding with 4HF induces smaller carbon CSPs for residues within the hydrophobic pocket. In particular, the P5 Cγ-Cδ cross peak in presence of 4HF reveals the presence of two species, respectively corresponding to the bound and free hydrophobic pockets with each about 50 % occupancy, as based on peak intensities, suggesting that only half of the pockets are occupied. P5 Cγ is located in the direct vicinity of the ligand, at around 4 Å from the aromatic ring according to the cryo-EM structure ù and docked models (**Figure 59**). As such, it is a good indicator of binding and can be used to estimate the percentage of occupancy. However, for the other residues in the vicinity of the ligand such as Y6, A58, N92 and L100, we detect a broad peak, at an intermediate chemical shift between the free and the bound state, as illustrated for A58 (**Figure 60A-B**). Finally, a third group could be identified, represented by residues V27, I59 and R98, which show almost no CSPs upon 4HF binding, but high CSPs upon binding of TX100 and OP (**Figure 60B**). Interestingly, these residues are located at the bottom of the hydrophobic pocket, behind F97 (**Figure 60C**). Hence, the lack of significant CSPs for these residues, as illustrated for R98, is likely due to an absence of a larger conformational rearrangement in presence of 4HF, as the F97 side-chain rotation observed in presence of TX100 (Makbul, Kraft, et al., 2021). We therefore conclude that, compared to the branched *tert*-octyl moiety, the less sterically hindered hexyl group might not be able to induce the rotation of the F97 side chain (Makbul, Kraft, et al., 2021).

III. Discussion

The exploration of 40 different TX100 analogues allowed us to identify the structural requirements for micromolar binding of TX100 to the capsid-forming HBV Cp. Results from NMR and ITC were complementary, with NMR allowing to characterize also weaker binding events, and ITC able to provide a K_D for notably the strongest binders. To summarize (**Figure 62**), we established that, with respect to binding, i) the TX100 hydrophilic tail can be reduced to a single hydroxyl function without significant loss; ii) its replacement by -NH_2, -OCH_3 and -CH_2OH is possible with minor losses; iii) in contrast, replacement of the aromatic ring by a cyclohexane ring has a deleterious effect, either due to steric hindrance or the different relative orientation of the hydrophobic/hydrophilic moiety; vi) a hydrophobic moiety of 4-8 atoms is needed to establish significant binding. The *para* positioning on the aromatic ring proved to be essential, and exceedingly bulky groups on either side hindered binding. Further studies will now attempt to improve the potency of TX100 derivatives by combining hydrophobic, aromatic and hydrophilic moieties. Perspectives include (i) increase lipophilicity of the hydrophobic chain using fluorine bearing carbons (CF_2/CF_3), thereby decreasing solvation and strengthening affinity to the targeted hydrophobic pocket, and (ii) link the hydrophilic moiety to a functionalized molecule, to induce multivalent interactions with the Cp dimer, for instance

targeting the capsid-assembly modulator (CAM) binding pocket (Bourne et al., 2006; Viswanathan et al., 2020).

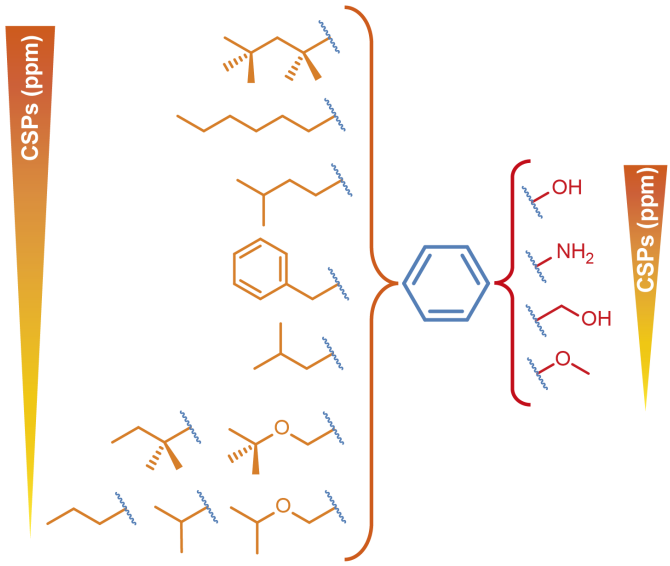

Figure 62. Summary of interacting molecules in the hydrophobic pocket of the HBV core protein. *Overview of the different moieties attached to either side of the central aromatic ring, classified as function of their CSPs. Only compounds with CSPs above 0.038 ppm are shown.*

A recent cryo-EM structure (Makbul, Kraft, et al., 2021) allowed to determine the position of the previously identified TX100 molecule (Lecoq et al., 2021) in the hydrophobic pocket; while electron density was observed in the pocket already in previous EM studies (Böttcher, 2021; Böttcher & Nassal, 2018), the nature of the molecule could not be determined, likely due to the lack of resolution. Indeed, while the EM density allows positioning of TX100 mainly based on the aromatic ring and the terminal -$C(CH_3)_3$ group, chemical composition and hydrophilic tail were not accessible from the electron densities.

A docking study based on NMR chemical shifts, coupled with the reasonable assumption that the hydrophobic part, rather than the hydrophilic part, fits into the hydrophobic pocket, results in a bundle of structures, many of which show an orientation very similar to that obtained by cryo-EM. Both structural models position the oxygen of the phenolic ring in a position where it can interact with residue K96 at the pocket entry.

In contrast to NMR, cryo-EM reported only limited structural changes in the presence of TX100, predominantly repositioning of the F97 aromatic ring. NMR indeed detects numerous and beyond-the-pocket-extending CSPs between the apo and bound states. This is likely due to the higher sensitivity of NMR, which can identify, with atomic resolution, changes in local dihedral angle orientations, but also differences in the local magnetic field as induced by different positioning of aromatic side chains. When considering that the largest change described in the cryo-EM study is the movement of the aromatic F97 ring, one can deduce that cryo-EM can detect changes which give rise to ^{13}C CSPs > 1.5 ppm, which is 5 times higher than the limit under which we consider the CSP insignificant (Figure 3A in (Block

et al., 2021)). Thus, while cryo-EM has the clear advantage to position the binders inside the pocket, solid-state NMR can identify the residues impacted by binding, including those which react in an allosteric manner.

By presenting a detailed analysis of the binding properties of the hydrophobic pocket that plays a central role in envelopment, our work thus opens a new avenue to target both the capsid-envelope interaction, and probably the dynamics of the HBV capsid in the viral life cycle in general (Niklasch et al., 2021). The pharmacomodulation through the experimental identification of the important parts of the ligands binding to the hydrophobic pocket is a first step towards the development of a potential binder that would not exhibit endocrine disrupting activity. It opens possibilities to design amino substituted compounds, which could lead to an antiviral treatment.

One should mention that TX100 is known to cause endocrine disruption (Nimrod & Benson, 1996) and, besides a variety of other phenols, particularly 4-tert-octylphenol is considered as a major environmental pollutant (Olaniyan et al., 2018), as it is toxic to many organisms. In order to consider TX100 as a starting point for further compound development, this toxicity must be addressed. Alternatives to TX100 are currently under development, as for instance the new detergent Nereid, an environmentally-friendly replacement candidate for TX100 based on substitution of the phenol with a benzylic alcohol (Farcet et al., 2021). Interestingly, we here show that the benzylic alcohol version (compound **1d**) retains strong binding capacity to Cp, which points out a direction towards less endocrine-disrupting Cp pocket binders.

IV. Conclusion

By presenting a detailed analysis of the binding properties of the hydrophobic pocket that plays a central role in envelopment, our work opens a new avenue to target both the capsid envelope interaction, and probably the dynamics of the HBV capsid in the viral life cycle in general.2 The pharmacomodulation through the experimental identification of the important parts of the ligands binding to the hydrophobic pocket is a first step towards the development of a potential binder that would not exhibit endocrine disrupting activity. It opens possibilities to design amino substituted compounds, which could lead to an antiviral treatment.

V. Supplementary figures

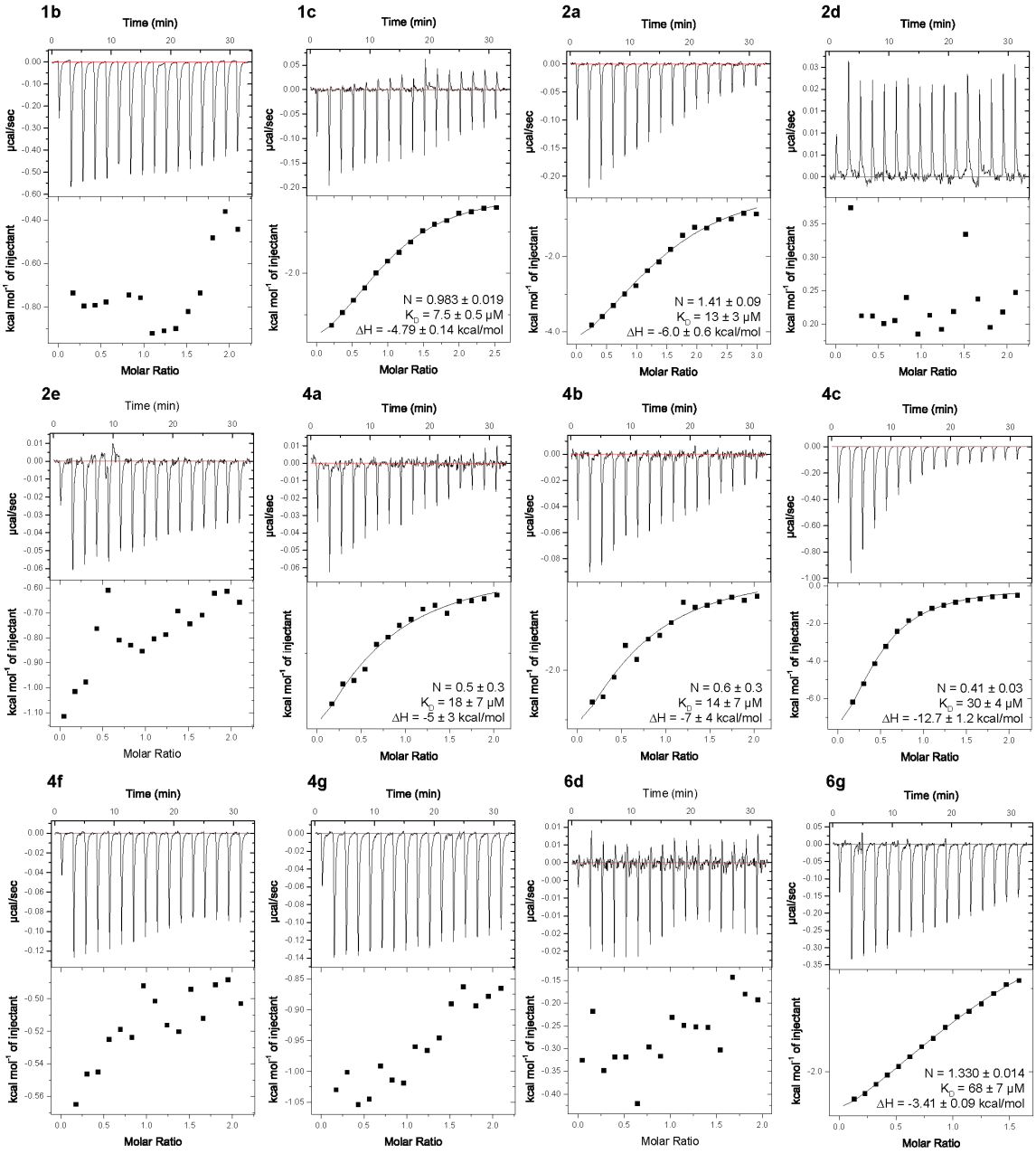

Figure 63. ITC raw data and binding isotherms obtained for Cp149 reassembled capsid upon additions of the indicated compounds. *Experiments were done in 50 mM HEPES buffer at pH 7.5 and 298 K, with typical concentrations of capsid at 100 µM ([Cp149$_{monomer}$]) and ligand concentrations of 1000 µM.*

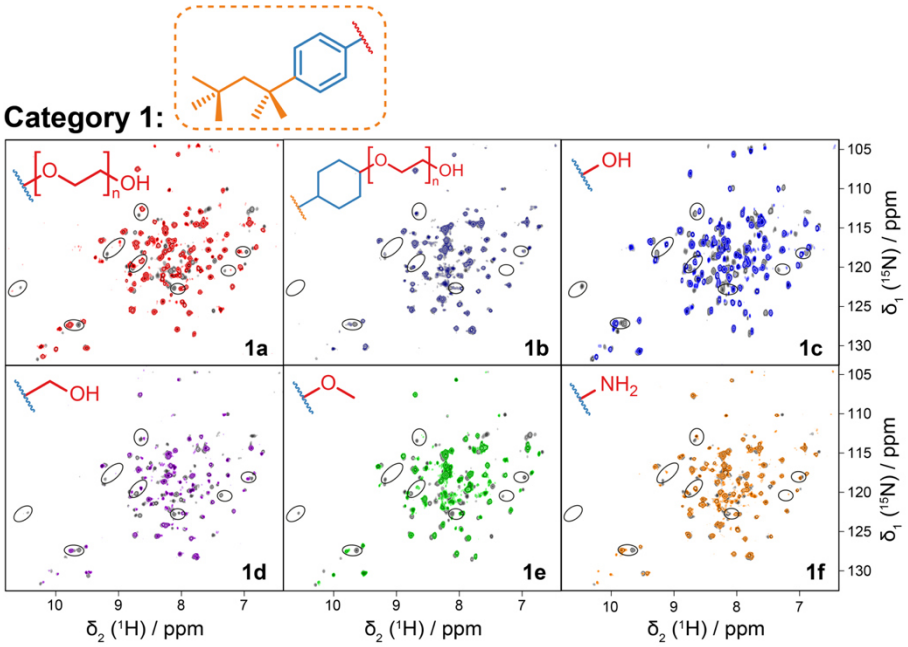

Figure 64. *^{1}H-^{15}N-SOFAST spectra of Cp149 dimer control (in grey) and in presence of 4 molar equivalents of compounds from the category 1 as displayed in* ***Figure 50****. The residues used for the determination of the interaction are circled. Note that control spectra may be different due to slight temperature changes coming from variable calibration from one spectrometer to another. Each spectrum recorded in presence of ligand is therefore overlayed with its corresponding control spectrum in light grey. Concentrations are for compound* ***1a****, [Cp149$_{monomer}$] = 80 µM; compound* ***1d****, [Cp149$_{monomer}$] = 90 µM; compounds* ***1b****,* ***1f****, [Cp149$_{monomer}$] = 120 µM; compound* ***1c****, [Cp149$_{monomer}$] = 120 µM.*

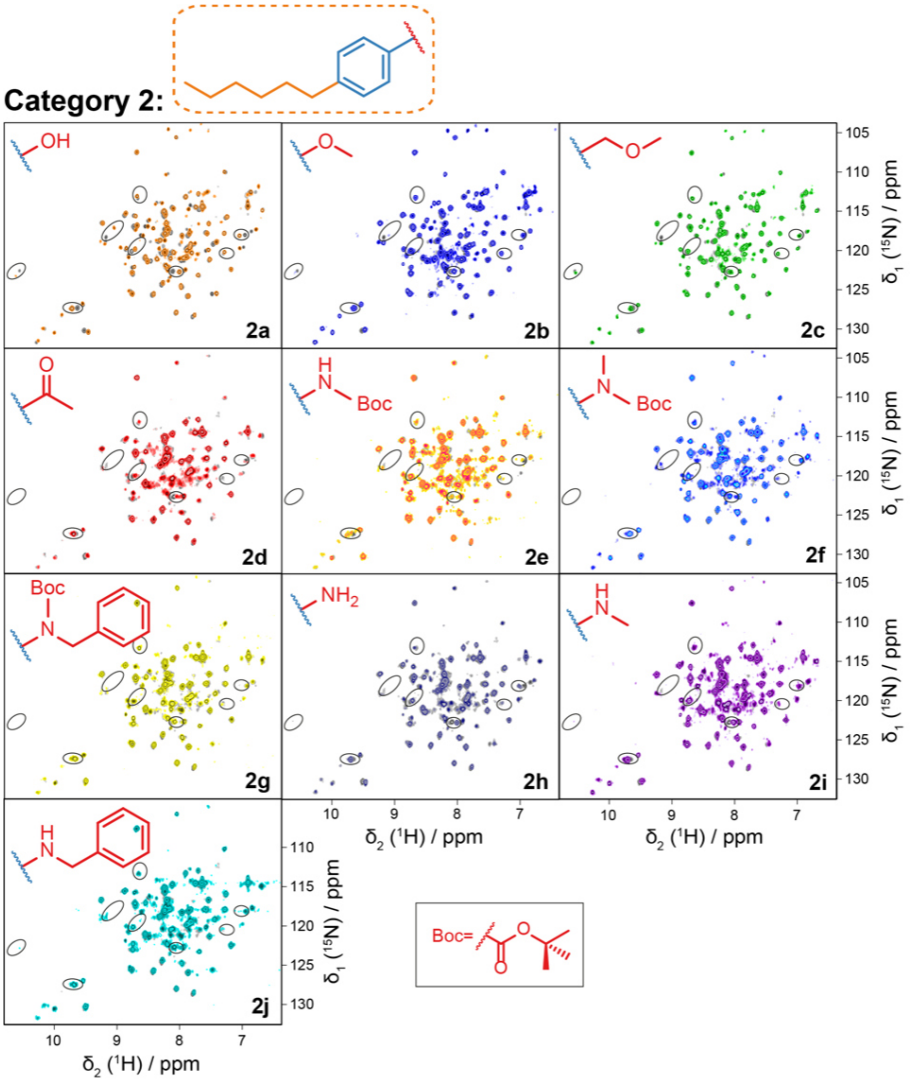

Figure 65. *^{1}H-^{15}N-SOFAST spectra of Cp149 dimer control (in grey) and in presence of 4 molar equivalents of compounds from the category 2 as displayed in* ***Figure 50****. For compounds* ***2a****,* ***2b****,* ***2c****,* ***2e****,* ***2f, 2g****,* ***2h****,* ***2i****,* ***2j****, [Cp149$_{monomer}$] = 80 µM, and for compound* ***2d*** *[Cp149$_{monomer}$] = 100 µM.*

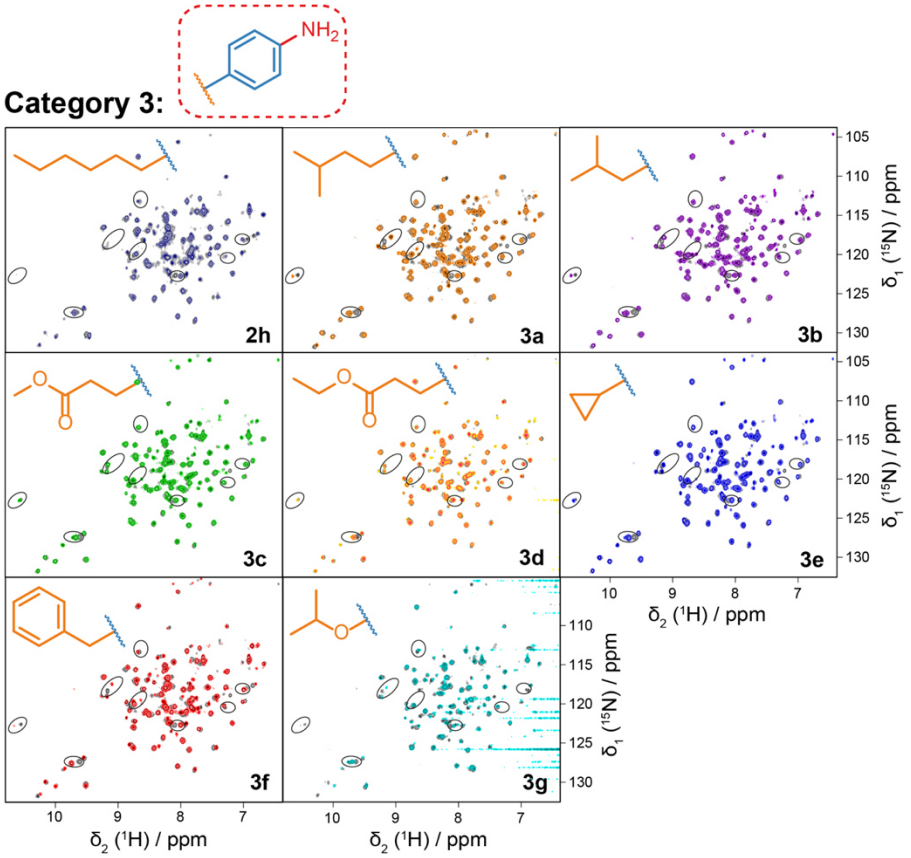

Figure 66. *^{1}H-^{15}N-SOFAST spectra of Cp149 dimer control (in grey) and in presence of 4 molar equivalents of compounds from the category 3 as displayed in* ***Figure 50****. For compounds* ***3a****,* ***3b****,* ***3c****,* ***3d****,* ***3e****,* ***3f****,* ***3g****, [Cp149$_{monomer}$] = 90 µM.*

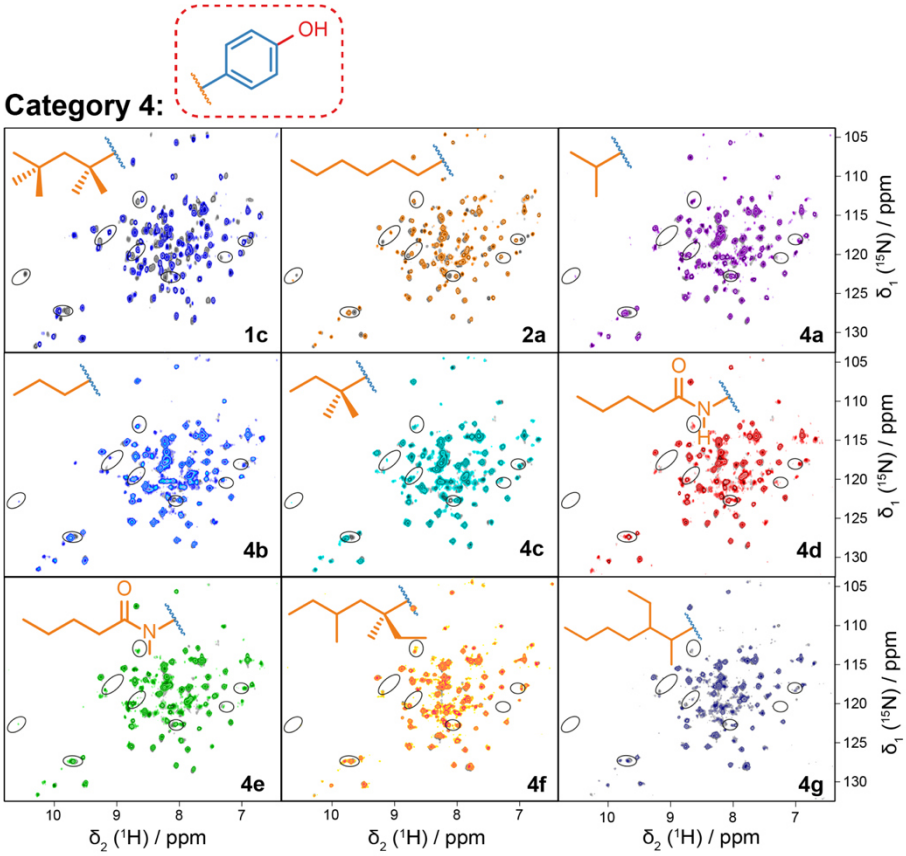

Figure 67. *^{1}H-^{15}N-SOFAST spectra of Cp149 dimer control (in grey) and in presence of 4 molar equivalents of compounds from the category 4 as displayed in* ***Figure 50****. For compounds* ***4a***, ***4b***, ***4c***, *[Cp149$_{monomer}$] = 100 µM; for compounds* ***4d***, ***4e***, *[Cp149$_{monomer}$] = 80 µM; for compounds* ***4f***, ***4g***, *[Cp149$_{monomer}$] = 120 µM.*

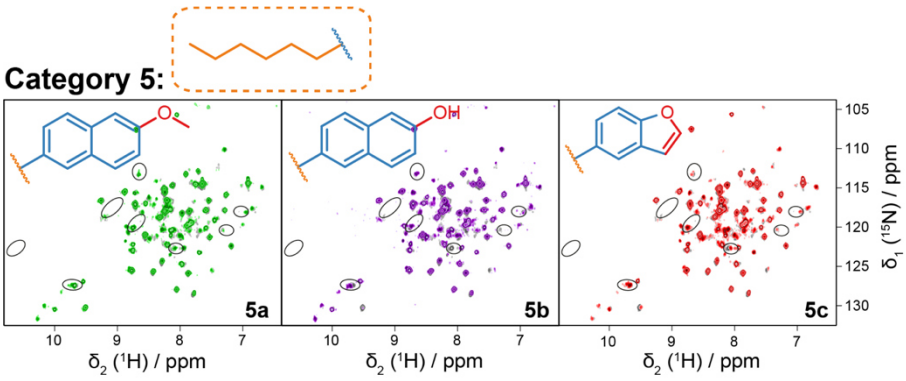

Figure 68. *^{1}H-^{15}N-SOFAST spectra of Cp149 dimer control (in grey) and in presence of 4 molar equivalents of compounds from the category 5 as displayed in* ***Figure 50****. For compounds* ***5a***, ***5b***, ***5c***, *[Cp149$_{monomer}$] = 80 µM.*

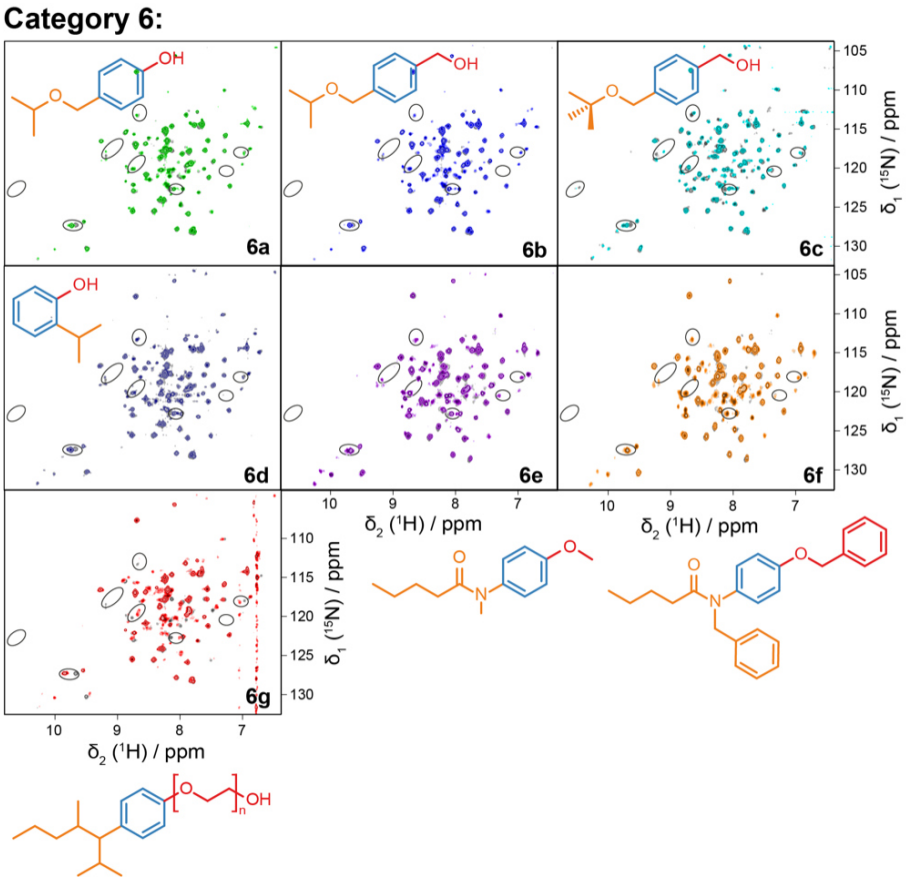

Figure 69. *^{1}H-^{15}N-SOFAST spectra of Cp149 dimer control (in grey) and in presence of 4 molar equivalents of compounds from the category 6 as displayed in* ***Figure 50****. For compounds* ***6a****,* ***6b****,* ***6d****,* ***6g****, [Cp149$_{monomer}$] = 120 µM; for compound* ***6c****,* ***6e****,* ***6f****, [Cp149$_{monomer}$] = 80 µM.*

Table 6. *Average CSPs calculated on the eight residues (D2, I3, Y6, V13, L16, V27, T91 and T95) for all compounds.*

Category 1:

Compound	1a	1b	1c	1d	1e	1f
Average CSPs (ppm)	0.143	0.024	0.142	0.090	0.077	0.103
Standard deviation (Stdev, ppm)	0.018	0.007	0.017	0.018	0.017	0.018

Category 2:

Compound	2a	2b	2c	2d	2e	2f	2g	2h	2i	2j
Average CSPs (ppm)	0.083	0.029	0.031	0.029	0.020	0.008	0.014	0.052	0.030	0.014
Stdev (ppm)	0.016	0.009	0.009	0.004	0.004	0.002	0.003	0.009	0.006	0.002

Category 3:

Compound	2h	3a	3b	3c	3d	3e	3f	3g
Average CSPs (ppm)	0.052	0.064	0.052	0.021	0.016	0.029	0.072	0.035
Stdev (ppm)	0.009	0.015	0.015	0.006	0.006	0.009	0.016	0.009

Category 4:

Compound	1c	2a	4a	4b	4c	4d	4e	4f	4g
Average CSPs (ppm)	0.142	0.083	0.038	0.039	0.046	0.014	0.019	0.021	0.024
Stdev (ppm)	0.017	0.016	0.006	0.011	0.012	0.004	0.003	0.007	0.004

Category 5:

Compound	5a	5b	5c
Average CSPs (ppm)	0.009	0.027	0.016
Standard deviation (ppm)	0.002	0.004	0.004

Category 6:

Compound	6a	6b	6c	6d	6e	6f	6g
Average CSPs (ppm)	0.038	0.021	0.048	0.036	0.018	0.015	0.021
Standard deviation (ppm)	0.007	0.004	0.012	0.009	0.004	0.004	0.007

VI. Materials and Methods

1. Molecular binders

Commercial compounds (**1a, b, c**; **2a, d**; **4a, b, c, f, g**; **6d, g**) were purchased from Sigma Aldrich. The synthesis of the other compounds used is described in the chemical syntheses annex of the Supplementary Information. Stock solutions of commercial and *de novo* compounds were prepared by diluting commercial solution or powder in DMSO, and then in the same buffer as the protein sample (except compounds **1a**, **1b** and **6g** which were directly diluted in protein's buffer). The typical concentration of the stock solutions was at 10 mM, and then the final solutions were calculated according to the experiments made, *i.e.* for ITC, the ligand should be at least 10 times more concentrated than the protein. For the final concentrations, it was established that the DMSO concentration should not exceed 5 % for either NMR or ITC experiments.

2. Expression and purification of Cp149

Plasmids of pET-28a2-HBc149opt were transformed into *E. coli* BL21*CP cells and grown at 37 °C overnight in 5 mL LB minimal medium culture, in presence of chloramphenicol and kanamycin. Precultures were diluted to yield a 1 L culture, and cells were grown at 37 °C until the optical density at 600 nm (OD_{600}) reached 0.6. Expression of Cp149 was then induced with 1 mM IPTG overnight at 20 °C. Cells were collected by centrifugation at 6,000 × g for 20 min and resuspended in 12 mL TN300 buffer (50 mM TRIS at pH 7.5, 300 mM NaCl, 2.5 mM EDTA, 5 mM DTT). During 1 h, the cell suspensions were incubated on ice with 1 mg/mL chicken lysozyme, 1X protease inhibitor mixture solution, and 0.5% Triton X-100 (TX100). Then, a total of 6 µL benzonase nuclease (0.5 µL/mL of protein solution) was added to digest nucleic acids for 45 min at room temperature. Cells were broken by sonication using a minimum of 10 cycles of 15 sec of sonication and 50 sec of cooling down on ice. Cell lysates were centrifuged at 30,000 g for 30 min to remove cell debris. The supernatant was loaded onto a 10 to 60 % sucrose gradient buffered with TN buffer and centrifuged in SW-32Ti Beckman Coulter swinging bucket rotor at 141,800 g for 3 h at 4 °C. Capsids in gradient fractions were identified by 15 % sodium dodecyl sulfate–polyacrylamide gel (SDS-PAGE). Then, proteins were precipitated by 35 % saturated ammonium sulfate (AS) during 30 min incubation and were centrifuged at 20,000 g for 40 min. Pellets were resuspended in 10 mL purification buffer (50 mM TRIS pH 7.5, 5 % sucrose, 5 mM DTT, and 1 mM EDTA). The proteins were centrifuged again for 20 min at 14,000 g to remove insoluble impurities. The capsids were dialyzed in purification buffer to remove remaining AS. Labelled proteins were expressed following the same protocols as unlabeled ones, except that those bacteria were grown in M9 minimal medium (with D_2O instead of H_2O) containing ^{13}C-deuterated glucose and $^{15}NH_4Cl$, and the induction was made at 25 °C.

3. Capsid disassembly and reassembly *in vitro*

Capsids were dialyzed overnight against disassembly buffer (50 mM $NaHCO_3$ pH 9.6, 5 mM DTT) at 4 °C. Capsid disassembly into dimers was achieved by the addition of urea powder at a final concentration of 3 M incubated overnight at 4 °C. To separate the dimers from remaining aggregates, size exclusion chromatography (SEC) was performed using a HiPrep 16/60 Sephacryl S-200 HR column (120 mL dead volume), previously equilibrated in disassembly buffer. The protein sample was filtered with a 0.45 µm PVDF membrane, before its injection on the column. The SEC also allows to remove TX100 from the hydrophobic pocket of the core protein (Lecoq et al., 2021). Fractions containing dimers were pooled and concentrated with an Amicon (10,000 MWCO). For solution NMR experiments, Cp149 dimer solution was dialyzed in 50 mM HEPES buffer at pH 7.5, 5 mM DTT and directly used. Dimer samples are always controlled using negative staining electron microscopy to ensure the absence of capsids before running solution NMR experiments. For ITC and solid-state NMR, the dimer solution was dialyzed in reassembly buffer (50 mM TRIS pH 7.5, 500 mM NaCl, 5 mM DTT) overnight at 4 °C. The resulting reassembled capsids were checked by negative straining electron microscopy and finally dialyzed in 50 mM HEPES buffer at pH 7.5.

4. Solution NMR

Freshly prepared ^{2}H-^{13}C-^{15}N Cp149 dimer in 50 mM HEPES buffer (pH 7.5), 5 mM DTT was quantified by absorbance measurements, and concentrated to approximately 100 µM (in monomer concentration). The different compounds were added at a Cp monomer:compound ratio of 1:4. Solutions were transferred into 3 mm solution NMR tubes, and ^{1}H-^{15}N-SOFAST spectra were recorded on all samples on a 600 MHz spectrometer (Bruker Avance II) or 950 MHz spectrometer (Bruker Advance III) equipped with a triple-resonance cryo-probe. NMR acquisition parameters are detailed in **Table 7**. The assignment from Cp149 dimer was transferred from BMRB deposition number 15969 (Freund et al., 2008) and confirmed by 3D HNCO, HNCA and HNcoCA spectra recorded on the 950 MHz spectrometer on the control sample, since differences in the protein sequence and in pH did not allow a complete assignment transfer. New assignments were deposited in the BMRB under deposition number 51294. All spectra were recorded at 295 K and processed using TopSpin 4.0.8 (Bruker Biospin) and analyzed with the CcpNmr Analysis package, Version 2.4.2 (Vranken et al., 2005). Chemical-shift perturbations (CSPs) between Cp149 dimer without and with ligand were calculated for each atom according to: $\Delta\delta_{HN} = \sqrt{(\Delta\delta_H)^2 + \left(\frac{\gamma_N}{\gamma_H}\Delta\delta_N\right)^2}$. A full CSP graph is shown for TX100 in **Figure 55**. For the other compounds, eight residues showing significant CSPs upon TX100 binding and isolated on the 2D spectrum were taken as representative probes: D2, I3, Y6, V13, L16, V27, T91 and T95. Average CSPs calculated on these 8 residues are detailed in **Table 6** for all compounds.

For the NMR titrations with TX100, OP and 4HF, ^{2}H-^{13}C-^{15}N Cp149 dimer in 50 mM HEPES buffer at pH 7.5 and containing 5 mM DTT was quantified by absorbance measurements. ^{1}H-^{15}N-SOFAST spectra were recorded for all ratios on the 600 MHz spectrometer (Bruker Avance II). NMR acquisition parameters are detailed in **Table 8**. For the titrations with TX100 and OP, the protein was concentrated to approximately 140 µM (in monomer concentration) and 14 titration points were done from ratio 0 to 1.5 monomer:ligand. Titrations are shown in **Figure 51 and Figure 52**, on which only 7 points are shown for clarity. For the titration with 4HF, ^{2}H-^{13}C-^{15}N Cp149 dimer was concentrated to approximately 70 µM (in monomer concentration) and 4 titration points were done at ratio 0, 0.5, 2 and 10 molar equivalents of monomer:ligand. Titration is shown in **Figure 53**. Stock solutions were done in HEPES buffer for TX100 and in 100 % DMSO for OP and 4HF.

Table 7. *NMR experimental details of solution NMR measurements recorded on $^{2}H^{13}C^{15}N$-Cp149 dimer. sw stands for spectral width. 3 mm tubes were used for all recordings. Final percentage of DMSO varied from 0 to 5 %. The concentration is given for the monomer unit.*

Ligand	[Cp149] (µM)	Field (MHz)	sw {^{15}N} (ppm)	Increments {^{15}N}	Number of scans	Exp. time
Control	73	600	29	200	64	1h08
1a, 1c, 1e, 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h, 2i, 2j, 4a, 4b, 4c, 4d, 4e, 5a, 5b, 5c, 6e, 6f	73	600	29	200	64	1h08

Control	100	600	29	200	48	51 min
3a, 3b, 3c, 3d, 3e, 3f	100	600	29	200	48	51 min
Control	102	600	29	200	64	1h08
1d, 6c, 3g	102	600	29	200	64	1h08
Control	117	950	29	316	64	1h38
1b, 1f, 4f, 4g, 6a, 6b, 6d, 6g	117	950	29	316	64	1h38

Table 8. *NMR experimental details of solution NMR ^{1}H-^{15}N-SOFAST spectra recorded on $^{2}H^{13}C^{15}N$-Cp149 dimer for titration with TX100, OP and 4HF. sw stands for spectral width. 3 mm tubes were used for all recordings. Final percentage of DMSO varied from 0 to 5 %. The concentration is given for the monomer unit.*

Ligand	**[Cp149] (µM)**	**Field (MHz)**	**sw {^{15}N} (ppm)**	**Increments {^{15}N}**	**Number of scans**	**Exp. time**
TX100 and OP	140	600	29	100	24	12 min
4HF	70	600	29	100	64	33 min

5. Solid-state NMR

For solid-state NMR, freshly prepared Cp149 reassembled capsids in 50 mM HEPES buffer (pH 7.5) were incubated with 4 to 8 equivalents of compounds to be investigated (TX100, OP and 4HF) for 2 hours at room temperature. They were then concentrated using Amicon Ultra centrifugal filter units (Merck, 50 kDa cut-off) to about 20 mg/ml in 1 mL and sedimented into 3.2 mm zirconium rotors by ultracentrifugation (200,000 g, 14 h, 4 °C) using a home-made filling tool. Rotors were immediately closed after the addition of 1 µL of saturated DSS solution for chemical-shift referencing.

2D ^{13}C-^{13}C-DARR (Takegoshi et al., 2001) spectra were recorded using a 3.2 mm triple-resonance (^{1}H, ^{13}C, ^{15}N) wide-bore probe head at a static magnetic field of 18.8 T corresponding to 800 MHz proton resonance frequency (Bruker Avance II). All spectra were referenced to DSS and recorded at a sample temperature of 4 °C according to the resonance frequency of the supernatant water. Assignments were derived from those presented in (Lecoq et al., 2018) (bound-state) and (Lecoq et al., 2021) (unbound state). NMR acquisition parameters are detailed in **Table 9**. All spectra were processed using TopSpin 4.0.3 (Bruker Biospin) and analyzed with the CcpNmr Analysis package, Version 2.4.2.

Chemical-shift differences between Cp149 reassembled capsid without and with ligand were calculated for each carbon atoms according to: $\Delta\delta_C = \delta_C[bound] - \delta_C[unbound]$.

Table 9. *NMR experimental details of solid-state NMR measurements. sw stands for spectral width. Experiments were recorded on an 800 MHz wide-bore spectrometer at a 17.5 kHz MAS frequency and at a sample temperature estimated at 5 °C. The ^{15}N-insert was removed from the probe for the DARR experiments of samples with OP and 4HF to increase the signal-to-noise ratio. Experimental time for samples with Triton X-100 and with OP was longer to respectively compensate the fact that it was recorded with the ^{15}N-insert (TX100), and that the rotor was not full (OP).*

Sample	**Cp149+TX100 (1a)**	**Cp149+OP (1c)**	**Cp149+4HF (2a)**
Experiment	2D DARR	2D DARR	2D DARR
Transfer 1	HC-CP	HC-CP	HC-CP
Field [kHz]	67.1 (^{1}H) 50 (^{13}C)	65.8 (^{1}H) 50 (^{13}C)	65 (^{1}H) 50 (^{13}C)
Shape	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H
^{13}C carrier [ppm]	58.6	58.6	58.6
time [ms]	0.8	0.7	0.9
Transfer 2	DARR	DARR	DARR
Field [kHz]	17.5(^{1}H)	17.5(^{1}H)	17.5(^{1}H)
^{13}C carrier [ppm]	100	100	100
time [ms]	20	20	20
t_1 increments	2560	2560	2560
sw (t_1) [kHz]	93.75	93.75	93.75
Acq. time (t_1) [ms]	13.7	13.7	13.7
t_2 increments	3072	3072	3072
sw (t_2) [kHz]	93.8	93.8	93.8
Acq. time (t_2) [ms]	16.4	16.4	16.4
^{1}H decoupling	SPINAL64	SPINAL64	SPINAL64
Field [kHz]	90	90	90
Interscan delay d1 [s]	2.6	2.6	2.6
Number of scans	20	20	8
Measurement time	38 h 40	38 h 40	15 h

6. Isothermal titration calorimetry

Samples of freshly prepared Cp149 reassembled capsid were dialysed into 50 mM HEPES buffer at pH 7.5 and then quantified by absorbance measurements. ITC experiments were performed using a MicroCal iTC200 instrument (Malvern Panalytical, Malvern, Worcestershire, UK) at a temperature of 278 K. Briefly, binding enthalpies were obtained by injecting small volumes of compounds into the microcalorimeter reaction cell containing Cp149 reassembled capsid sample under stirring conditions (1,000 rpm) and a reference power of 5 µcal/s. In a typical ITC experiment, after a temperature equilibration delay, the measurement was started by a 1 µL injection of ligand (injection duration 2 s), followed by 15 injections of 2.5 µL (injection duration 5 s) in 2 min intervals. The heat evolution after each injection was obtained from the time integral of the calorimetric signal and the data were analysed assuming a single site binding model (one set of sites) using the MicroCal Origin software package. The thermodynamic parameters were calculated with the software provided by the instrument supplier. Control experiments for heats generated by mixing and dilution were performed under identical conditions and used for data correction in subsequent analysis.

7. Docking

Structure files of Cp149 dimer and OP ligand were prepared using MGLTools 1.5.6 (Morris et al., 2009). The initial PDB file of the ligand was obtained by extracting the TX100 coordinates from PDB structure 1UEH (Chang et al., 2003) and manually removing the hydrophilic tail. The ligand was set entirely flexible in the docking calculation, with four rotatable bonds. The coordinates of Cp149 were taken from PDB structure 1QGT (Wynne et al., 1999), and mutation of residue 97 to Phenylalanine was introduced with USCF Chimera (Pettersen et al., 2004). The receptor protein was prepared by selecting a search box encompassing the following residues showing large CSPs (Lecoq et al., 2021), that were also set flexibles: residues 1, 3, 6, 8, 27, 55, 58, 59 to 62, 64, 65, 68, 92, 93, 96 to 100 and 105 in one monomer and residues 42, 55, 58, 59, 60 to 62, 64, 65 and 68 in the other monomer, resulting in a total number of 69 rotatable bonds. The docking calculation was then performed using AutoDock Vina 1.1.2 (Trott & Olson, 2009). The resulting docked models were visualized with PyMoL (Delano W. L. http://www.pymol.org/).

VII. Contributions

I prepared Cp samples, together with Marie Dujardin and Marie-Laure Fogeron, measured ITC, together with Roland Montserret, and NMR spectra, together with Lauriane Lecoq; Thierry Lomberget, Francois Hallé and Sylvie Radix designed compounds, François Hallé and Sylvie Radix synthesized compounds; Juliette Martin made docking experiments; Mathilde Briday and Lauriane Lecoq analyzed data; Anja Böckmann, Thierry Lomberget, and Lauriane Lecoq designed the study, together with Michael Nassal and Beat H. Meier. Mathilde Briday, Lauriane Lecoq and Anja Böckmann prepared the initial manuscript, together with Michael Nassal and Beat H. Meier, and input from all authors.

RESULTS II

I. Introduction

The interaction between the capsid and the envelope remains a topic of importance in understanding the mechanism of infection of hepatitis B virus. Indeed, if the capsid is not enveloped, then the resulting virus particle will not be infectious. Thus, research is still underway to unravel the mystery of this peculiar interaction, including small peptides (Oct) that could inhibit the binding between the envelope proteins and the core protein. The first to discuss Oct peptides were Dyson and Murray, who selected these peptides from a random hexapeptide library displayed on filamentous phage. These peptides have the particularity to block the interaction between the L envelope protein and the capsid (Dyson & Murray, 1995). Based on what has been investigated previously, Oct peptides bind to the capsid protein through interactions involving the spike region (**Figure 70**) (Freund et al., 2008). They found that Oct2 binding on the core protein dimer induces structural changes in the tertiary structure, but not the secondary.

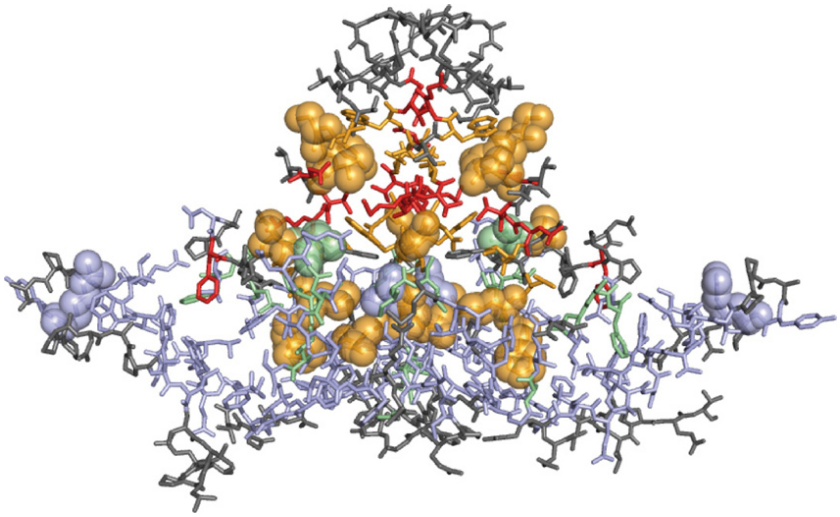

Figure 70. Residues impacted by the binding of Oct2 peptide on Cp dimer. *The residues are classified by four colors: light blue for residues which are negligibly affected; green for residues which have small CSPs (< 0.04 ppm in the ^{1}H dimension); orange for residues which have large CSPs (> 0.1 ppm); red for non-assigned residues. Taken from (Freund et al., 2008)*

Moreover, cryo-EM measurements showed that these peptides increased the electron density at the binding cleft at the very tip of the spike (Böttcher et al., 1998). More precisely, it has been shown that residues 77 and 78, respectively a glutamic acid and an aspartic acid, are important for binding as their mutation in alanine drastically reduce the affinity of the core protein with the peptides. Finally, Makbul studied the “SLLGRM” motif which has been identified as the core binding motif by peptide array screening (Makbul, Khayenko, et al., 2021). This specific small peptide binds at the interface between the two monomers on the

spike, leading to a more open conformation at the tips of the spike (**Figure 71**) (Makbul, Khayenko, et al., 2021).

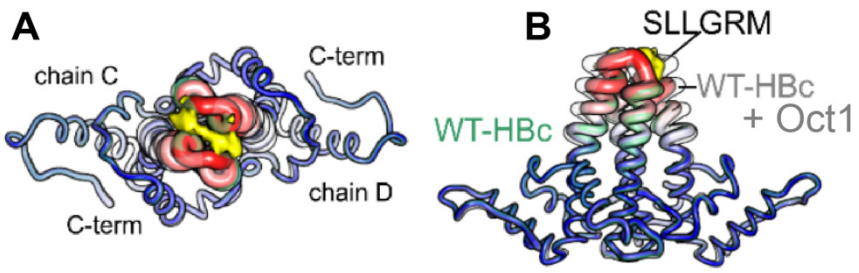

Figure 71. *Model of the WT-HBc + Oct1 with mapping of « SLLGRM » motif. Adapted from (Makbul, Khayenko, et al., 2021)*

All these studies suggest that it is the interaction of peptides on the spike tip that is responsible for blocking the binding of the envelope protein L to the core protein. We are therefore interested in studying these envelope-capsid interactions in order to obtain a more complete view on how this interaction can be abolished by the binding of peptides. For the moment, this envelope-capsid interaction has been mainly investigated in using in-cell virology approaches, but there are several mutational studies which rather localize the binding site rather to the spike base than to the spike tip. We decided to first focus on characterizing Oct1 and Oct2 peptides by NMR in order to derive approaches that can reveal molecular interfaces. This shall allow to investigate the more complex case of preS binding (larger protein, possibly lower affinity), which we mainly investigated by ITC, but also chemical shift perturbations on the capsid side. Ultimately, we aim at determining how the core protein is interacting with the envelope proteins.

II. Results

1. Interactions between Cp and Oct peptides

We used ITC and solid-state NMR to be able to characterize the interaction between Cp149 and Oct peptides. We used Oct1 and Oct2, which sequences are respectively GSLLGRMKGA and MHRSLLGRMKGA. First, we performed ITC measurements to be able to evaluate the binding parameters. While Freud and coworkers tested the binding of Oct peptides with Cp149 dimer at different temperatures and at pH 9.5, we tested the interaction of each peptide with Cp149 capsids at 298 K and pH 7.5, to be closer to physiological conditions. Since the initial curves recorded did not show the first plateau typical of ITC experiments. Thus, we modified the protein and ligand concentrations to obtain interpretable curves. After optimization, we obtained ITC curves for each of the peptides (**Figure 72**). The results show that both peptides interact with Cp149 capsid in the micromolar range, with a dissociation constant of 5.5 ± 0.7 µM and 5.8 ± 1.1 µM for Oct1 and Oct2 respectively.

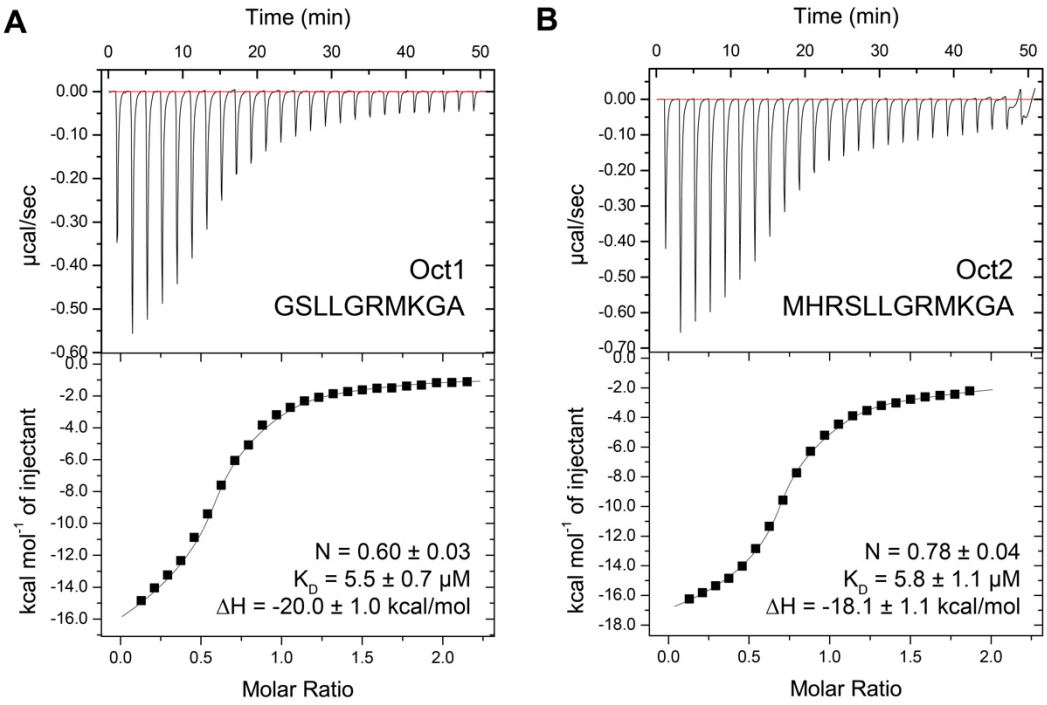

Figure 72. ITC curves for both Oct peptides. *A) ITC raw data and binding isotherms displaying the titration of Oct1 against Cp149 reassembled capsids at pH 7.5 and 298 K. B) ITC raw data and binding isotherms displaying the titration of Oct2 against Cp149 reassembled capsids at pH 7.5 and 298 K. Experiments were done in 50 mM HEPES buffer, with [Cp149$_{monomer}$] = 40 µM and [Oct] = 400 µM. Sequences of both peptides are annotated, as thermodynamic results.*

It can be seen that the peptides show a strong interaction with Cp149 capsid, and these results agree with those obtained by Freund and his colleagues for the Cp149 dimer (K_D = 3.6 and 2.0 for Oct1 and Oct2 respectively). Moreover, when testing the interaction of Oct2 with Cp149 capsid, they found a stoichiometry slightly smaller than with the dimer. In our case, the stoichiometry of Oct2 is similar to the one they measured (n = 0. 78 ± 0.04 compared to 0.85 ± 0.09 in their paper), however the stoichiometry of Oct1 is slightly smaller than for Oct2 (condition not tested in (Freund et al., 2008)).

After demonstrating that the interaction between the two peptides and the core protein was measurable and quantifiable, we decided to analyze the Cp+Oct complexe by solid-state NMR. Indeed, this technique allows us to obtain more information about the binding of these peptides specifically to the capsid, which cannot be analyzed in solution-state NMR because of its size. For this purpose, we mixed a labeled $^{13}C^{15}N$-Cp149 solution with each peptide solution to achieve a Cp monomer:Oct = 1:4 ratio, condition necessary to ensure that all Cp binding sites are occupied with Oct peptides. Afterwards, the solution containing about 20 mg of protein was concentrated and then centrifuged to fill a 3.2mm rotor. A control sample of Cp149 capsids reassembled in absence of Oct was also prepared in order to study the CSPs between the reference spectrum and those with the Oct peptides. 2D DARR spectra were recorded (**Figure 73**). We can see that for both peptides, differences are observed in the

spectra compared to the control, confirming the interaction on the capsid and allowing to localize the binding. For the control sample, the assignments that have been made previously were used (Lecoq et al., 2018), and a set of 3D spectra was recorded on both Cp-Oct samples in order to assign the bound forms and compare the chemical shifts with the control.

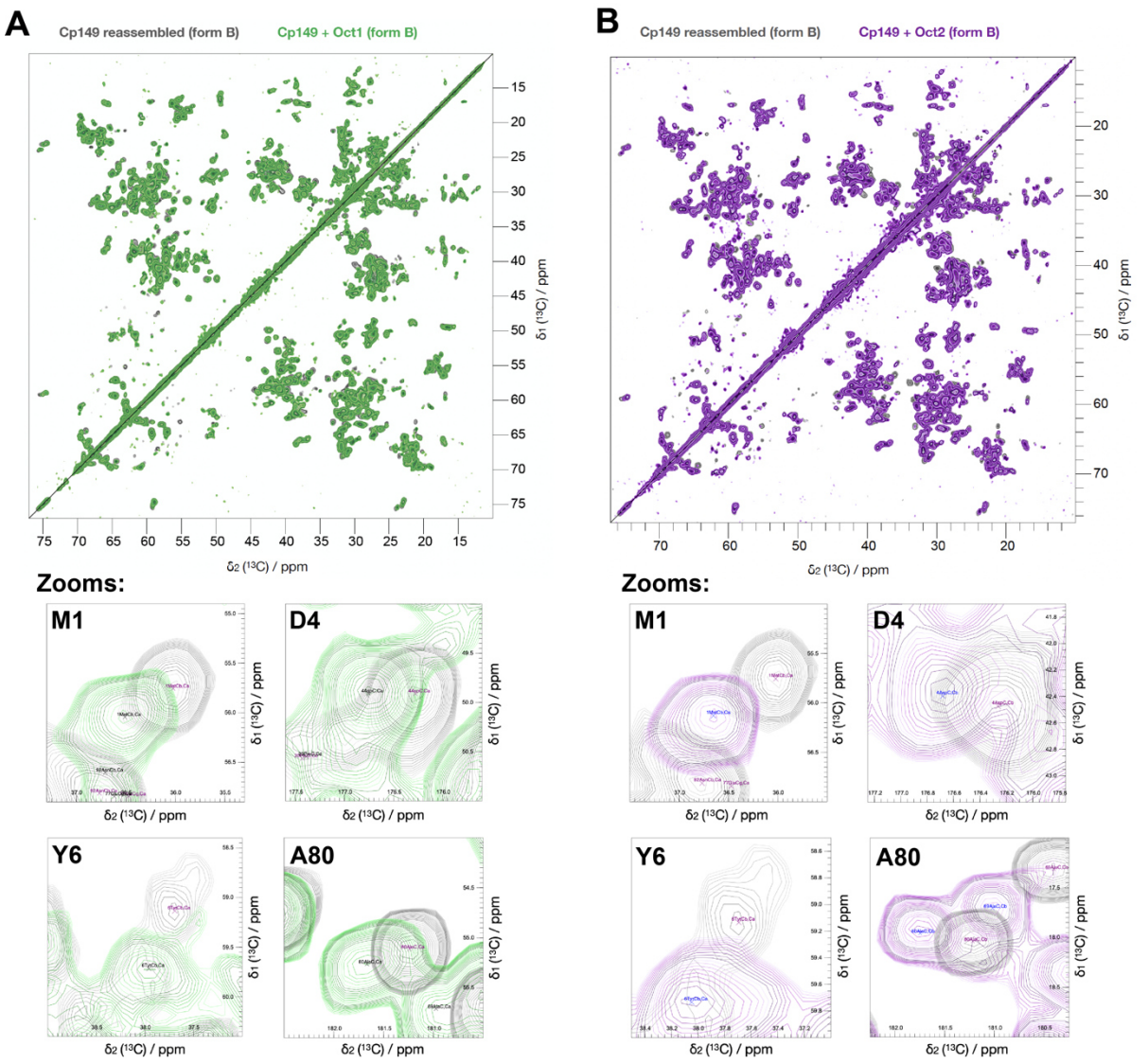

Figure 73. DARR spectra of Cp149 and Oct peptides. *A) Comparison of the aliphatic region of 2D DARR spectra of labelled $^{13}C^{15}N$-Cp149 reassembled capsids in absence (grey spectrum) or in presence Oct1 (green spectrum). B) Comparison of the aliphatic region of 2D DARR spectra of labelled $^{13}C^{15}N$-Cp149 reassembled capsids in absence (grey spectrum) or in presence Oct2 (purple spectrum). Experimental parameters are detailed in* ***Table 12****.*

Freund and coworkers identified a list of amino acids highly impacted by the binding of Oct2 on Cp149 dimer: S26, A35, Q57, C61, W62, G63, D64, L68, A69, W71, V86, S87, Y88, V89, K96, Q99, L100, and F103. Almost all these residues showed high CSPs in our solid-state NMR analyses, and we further identified additional peaks that have chemical shifts above 0.3 ppm (**Figure 74A and C**). The identified residues are highlighted on the dimer structure of the core protein (**Figure 74B and D**) for each peptide.

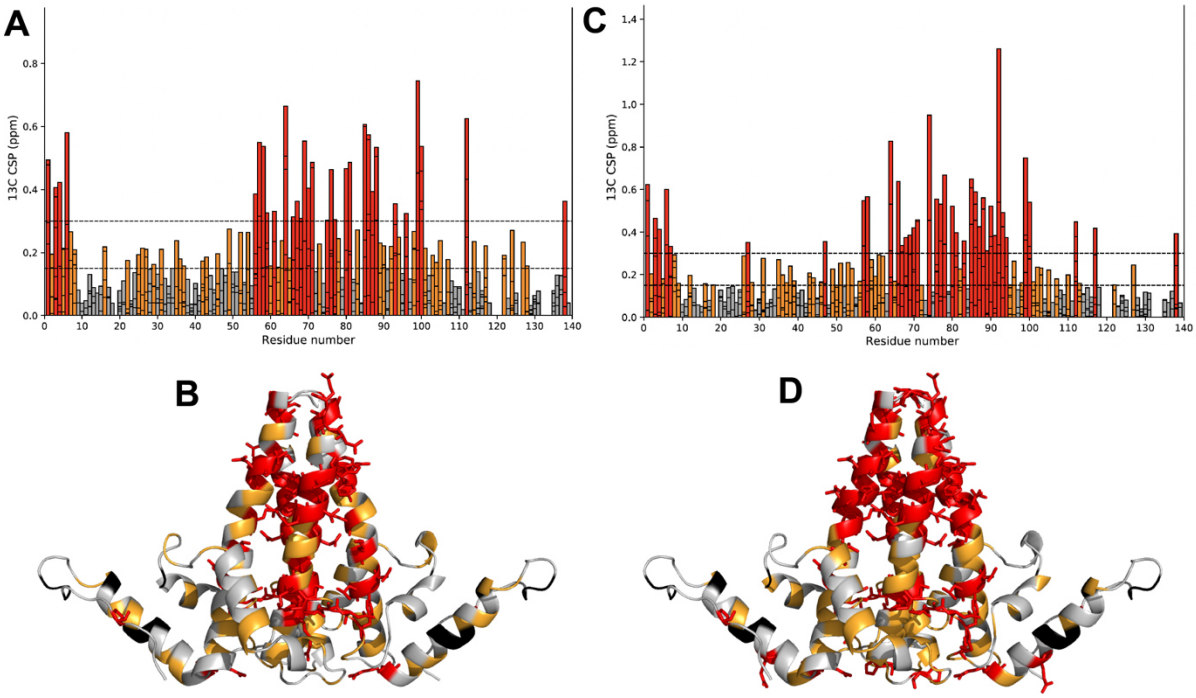

Figure 74. Residues of Cp149 dimer impacted by the Oct binding. *A) and C) CSPs graphs showing, for each residue, the difference of chemical shift between the unbound and the bound form. B) and D) Structure of the dimer (PDB 1QGT) in presence of respectively Oct1 and Oct2 peptide, mapping amino acids with high (colored in red) and medium (colored in orange) CSPs.*

By comparing the two CSPs graphs, we can see that both peptides induce globally the same perturbations on the structure of Cp149 dimer. Moreover, we can see that the residues mainly impacted by the binding of the peptides are those of the spike, but also some located more in the center of the dimer, which could be due to allosteric effects. Finally, it is possible to see that the Oct2 peptide induces some more CSPs than Oct1 (**Figure 75**). It is quite possible to think that because Oct2 is larger, it has more binding sites than Oct1, and therefore induces more structural changes on the core protein structure.

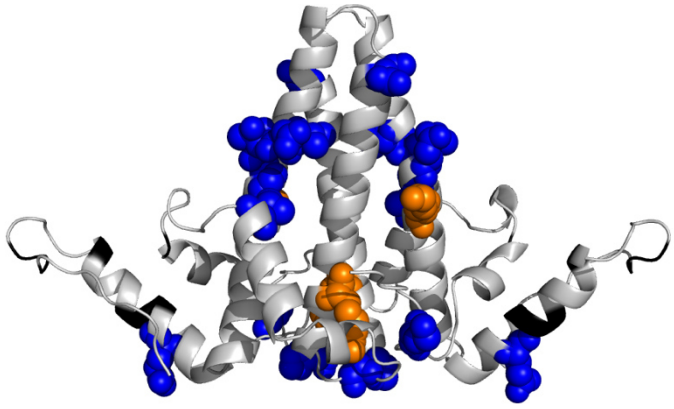

Figure 75. Residues of Cp149 dimer (PDB 1QGT) impacted by Oct2 but not by Oct1 interaction and vice versa. *The supplementary residues impacted by Oct1 are colored in orange, and the ones by Oct2 in blue.*

In particular, it is possible to notice the presence of high CSPs localized at the base of the dimer, which could be due to allosteric effects.

Oct is now further used to derive NMR approaches for polarization transfer experiments between a ligand and protein, in collaboration with Beat H. Meier's group at ETH Zurich. The here present work, including the assignments of the Oct-bound form, presents the basis for these investigations.

2. Envelope interaction

It has already been shown that the L protein of the envelope interacts with the core protein, especially the preS part, and even more precisely the amino acids at the junction between the domains of preS1 and preS2 (Bruss, 1997, 2004; Poisson et al., 1997). However, on the core protein side, the binding site is not known. Still, a variety of mutations, natural and non-natural ones, that impact envelopment, have been identified (Koschel et al., 2000; Pairan & Bruss, 2009; Pastor et al., 2019; Ponsel & Bruss, 2003). We here aim at deeper characterizing the structural details of this interaction between the capsid and different parts of the envelope proteins.

An important thing to know is that the preS domain is an intrinsically disordered protein (IDP) and is a highly unstable protein that needs to be kept under special buffer conditions which stabilize it. First of all, this protein cannot be concentrated to more than 50 µM otherwise it precipitates. Moreover, it requires a buffer solution containing protease inhibitor to avoid its degradation. Finally, thawing and refreezing cycles must be avoided as much as possible to prevent degradation with different temperature cycles. At the beginning, we were only able to produce preS in a wheat-germ cell free system, but after design of expression vectors by Michael Nassal, and optimization of expression in our lab, this protein can now be produced in bacteria.

a. Interaction between Cp149 dimer and preS

First, we studied the interaction between preS and the Cp149 dimer. We evaluated this interaction by ITC and solution-state NMR. The results obtained are unambiguous and prove that there is no interaction between these two proteins. By ITC, we recorded an experiment with a freshly disassembled Cp149 dimer solution, and a freshly produced preS solution. Normally, we fill the sample cell with Cp149 solution, but as preS could not be concentrated, we inversed the solution in the cell and in the syringe. The isotherm corresponding to this experiment shows identical peaks heights to those recorded in the blank experiment (dilution of Cp149 from the syringe into the buffer solution that is in the cell). The fact that the two isotherms are similar proved that the signals detected during the experiment between Cp149 dimer and preS correspond only to the heat released by the dilution of Cp149 dimer in the sample cell.

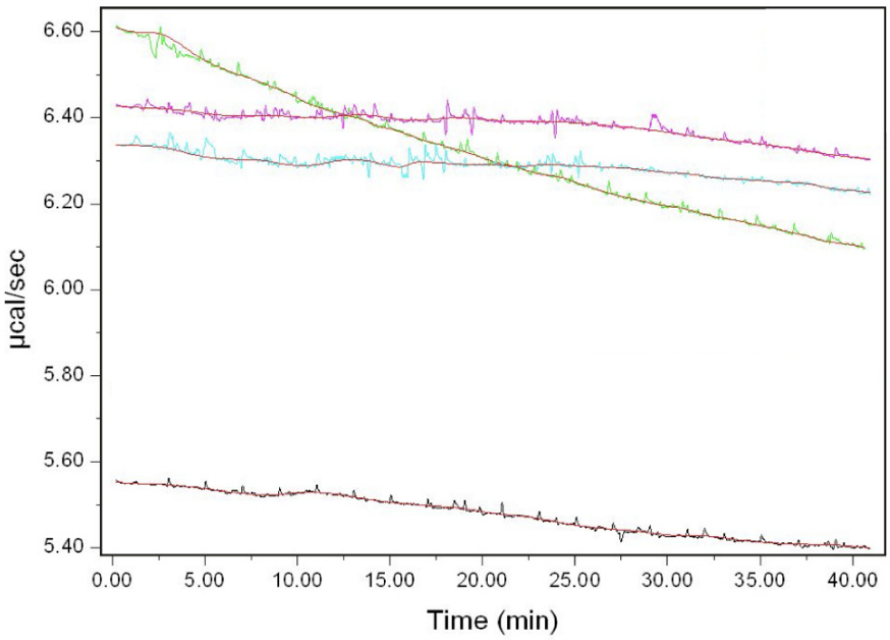

Figure 76. Raw ITC curves of preS titrated by Cp149 dimer. *The black curve corresponds to the blank experiment, and the cyan, green and pink curves to the triplicate titration experiments. Experiments were done with typical concentration of [preS] = 35 µM and [Cp149$_{monomer}$] = 350 µM.*

We then checked the interaction between Cp149 dimer and preS using solution-state NMR. We acquired a reference ^{1}H-^{15}N-SOFAST spectrum of freshly prepared labeled ^{13}C^{15}N-preS, and then we titrated preS with successive additions of 0, 0.5, 1, 2 and 5 equivalents of Cp149 dimer (**Figure 77**). We can see along the titration and even up to 5 equivalents, we do not detect any change on the spectra, either in terms of intensities or chemical shift. This corroborates the first results we obtained by ITC, namely that preS does not interact with the HBV core protein in its dimeric form.

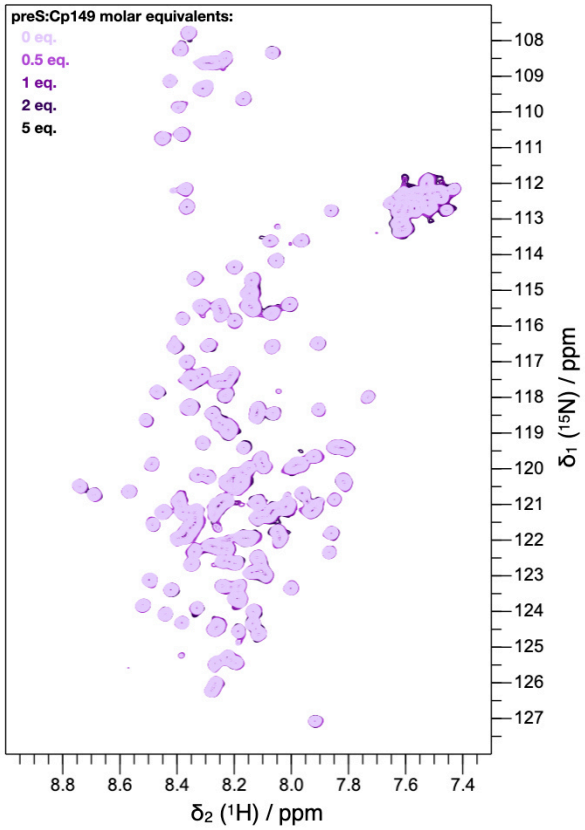

Figure 77. ^{1}H-^{15}N solution-state NMR spectra of the $^{13}C^{15}N$-preS titrated with Cp149 dimer. *The addition of Cp149 is indicated by a color gradient. preS concentration was 13.5 µM. All experimental parameters are detailed in* ***Table 10****.*

To get the confirmation on both proteins' sides, we also recorded a reference ^{1}H-^{15}N-SOFAST spectrum with freshly prepared labeled $^{2}H^{13}C^{15}N$-Cp149 dimer, with 1 equivalent of preS. We have reiterated the measure after 2.5 hours to verify that the interaction does not occur even with time (**Figure 78**). We can see that there are no observable CSPs and therefore no structural changes detected on the spectrum of labeled Cp149 dimer.

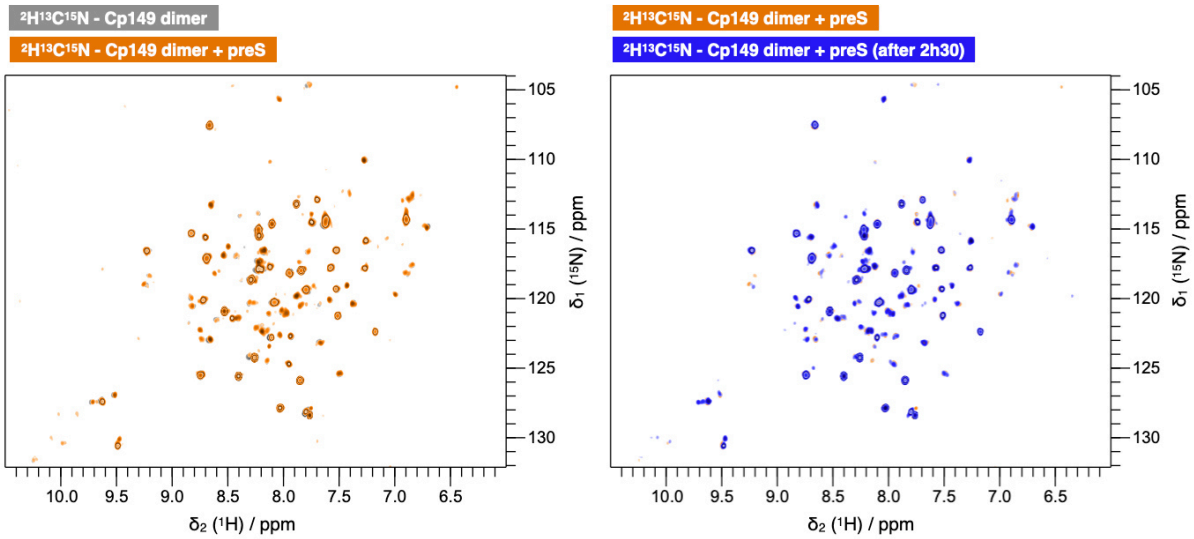

Figure 78. Solution-state NMR spectra of $^{2}H^{13}C^{15}N$-Cp149 dimer with preS. *The left panel shows the comparison between the control spectrum in grey and the spectrum with 1 equivalent of preS in orange. The right panel shows the difference between the spectrum just after the addition of preS in orange, and after 2.5 hours of incubation in blue. Experimental parameters are detailed in* ***Table 11****.*

In summary, both ITC and solution NMR titrations on preS and Cp149 dimer sides converge to the conclusion that preS does not seem to interact with Cp149 dimer.

b. Interaction between Cp149 capsid (form B) and preS

We were interested in the interaction between the Cp149 capsid and the preS protein. For this, since the capsid cannot be directly analyzed in solution NMR because of its size (~ 40 kDa), we looked at the interaction from preS side. For this, we filled NMR tube with labeled $^{13}C^{15}N$-preS, and we recorded spectra before and after the successive addition of Cp149 capsids. The backbone assignment of preS was already performed in the lab (BMRB number 51186) and reported in (Fogeron et al., 2022), and could be used for the reference spectrum of preS. We added small volumes of concentrated Cp149 capsids in order to perform a complete titration (**Figure 79**), which led to no chemical shift perturbations, but instead to a progressive decrease in the peaks intensities along the titration. This phenomenon is likely due to the fact that preS interacts with the Cp149 capsid, forming a complex that becomes too large to be observed in solution NMR, which explains the progressive disappearance of the preS peaks. No precipitation was observed in the NMR tube at the end of the titration, suggesting that the peaks disappearance is not due to protein aggregation. This was confirmed by an SDS-PAGE showing that both proteins are not degraded after the titration (data not shown).

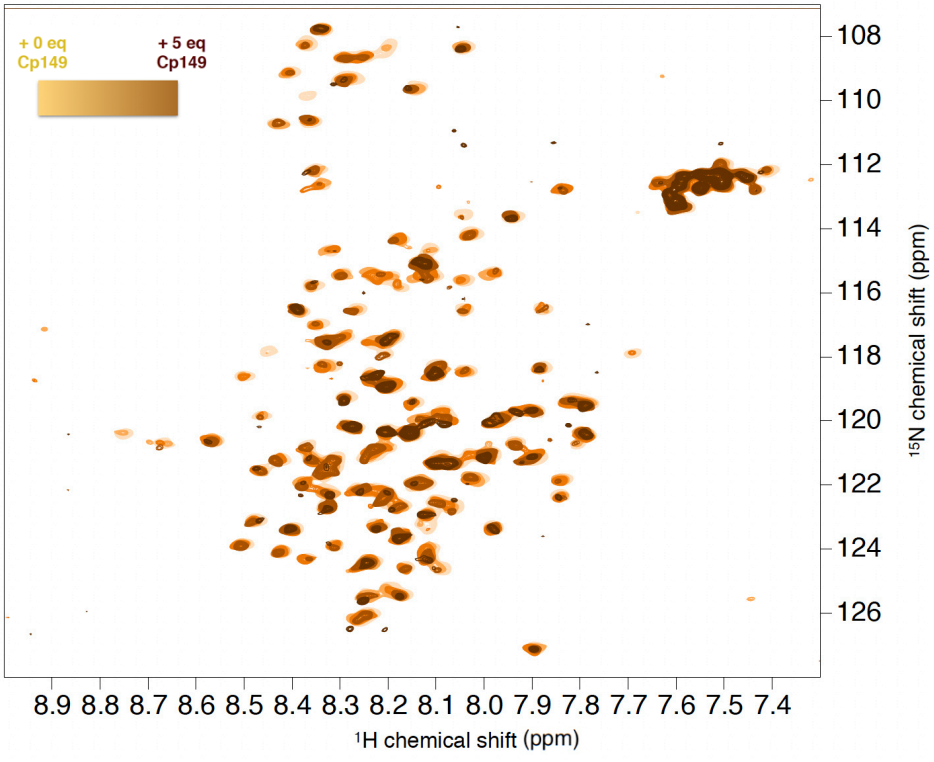

Figure 79. Solution NMR spectra of $^{13}C^{15}N$-preS titration by Cp149 capsids. *The addition of Cp149 is indicated by a color gradient. Spectra were recorded at 0, 0.5, 1, 2 and 5 equivalents. Experimental parameters are detailed in* ***Table 10****.*

The analysis of peak intensities reveal that they all decrease mainly at 1 equivalent, and then the intensities decrease in a less drastic way. After the addition of 2 equivalents of Cp149 capsids, most of the peaks intensities have decreased by about 50% (**Figure 80**), which means that all the preS protein is interacting with the capsids. There is no region of preS that remains independently flexible compared to the rest of the complex, as it can happen for some IDPs (Hosek et al., 2020). Finally, since no CSPs were observed and since the intensity decreases seem to affect preS in a global manner, it is difficult to use these data to get information on the localization of the binding site on preS.

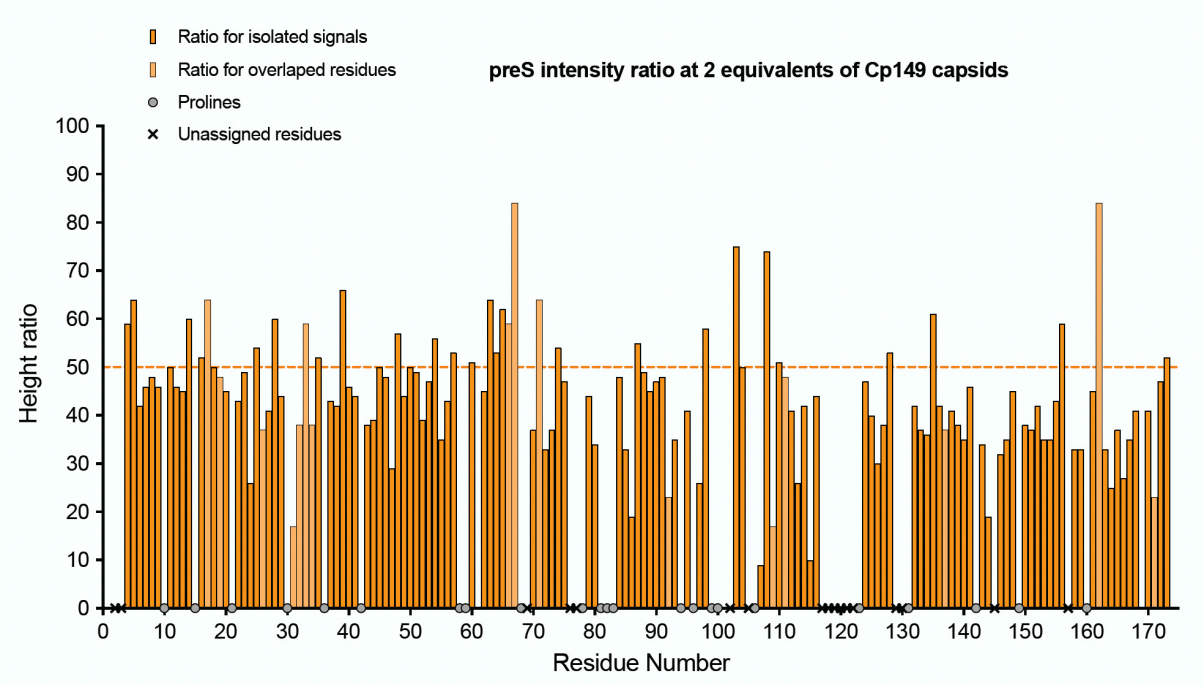

Figure 80. Intensity decrease of preS residues titrated by Cp149. *PreS intensities after addition of 2 molar equivalents of Cp149 capsids are compared with the reference intensities without Cp149.*

After studying the interaction by solution-state NMR on preS side, we wanted to explore the same kind of analysis on the Cp149 capsids side, in order to localize the binding site on the core protein. For this purpose, we first tested the co-sedimentation of these two proteins in order to check that we could sediment the two proteins to fill an NMR rotor. We sedimented the Cp+preS mixture following a strict preparation protocol: preS was added drop by drop in a small volume of highly concentrated Cp149 solution (200 µM) to allow the whole envelope protein domain to interact with the capsid, and work on ice to avoid any degradation. Subsequently, this mixture was centrifuged into small plastic tubes used specifically for small scale testing (max volume ~ 300 µL) under the same conditions that we apply to fill the rotors

(200,000g, 17 h, 4 °C), to see if the proteins were able to co-sediment. At the end of the centrifugation cycle, the supernatant was collected, and the sediment was resuspended in the same buffer to be loaded on SDS-PAGE (**Figure 81**). We can see that most of the proteins have co-sedimented (lane “Sediment”), even if some small quantities remain in the supernatant. This experiment allows us to validate that during a rotor filling rotor filling, most of the proteins would be inside the sediment in the NMR rotors.

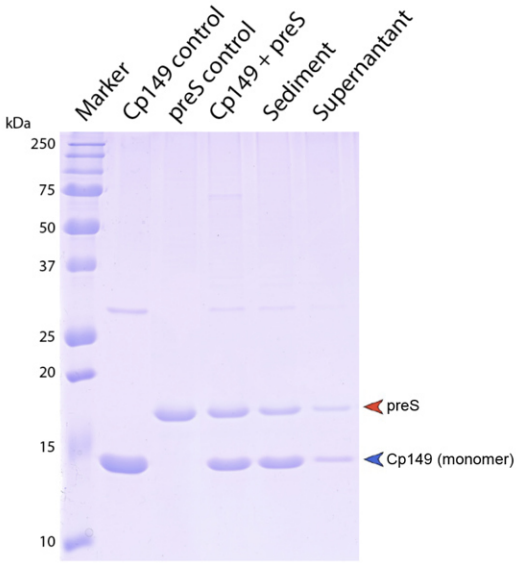

Figure 81. SDS-PAGE analysis of co-sedimentation of core and envelope proteins.

We first prepared 3.2 mm rotors, with $^{13}C^{15}N$-Cp149 capsids mixed with 1.5 equivalent of preS. The first sample we recorded showed large CSPs on the 2D DARR spectra (**Figure 82A**), but upon observation of the NCA spectrum (**Figure 82B**), there was visible oxidation at cysteines 48 and 61 (observable on the zoom on (**Figure 82B**). After analyzing the CSPs, we concluded that the changes observed were likely caused by oxidation of cysteine 48 (**Figure 82C**).

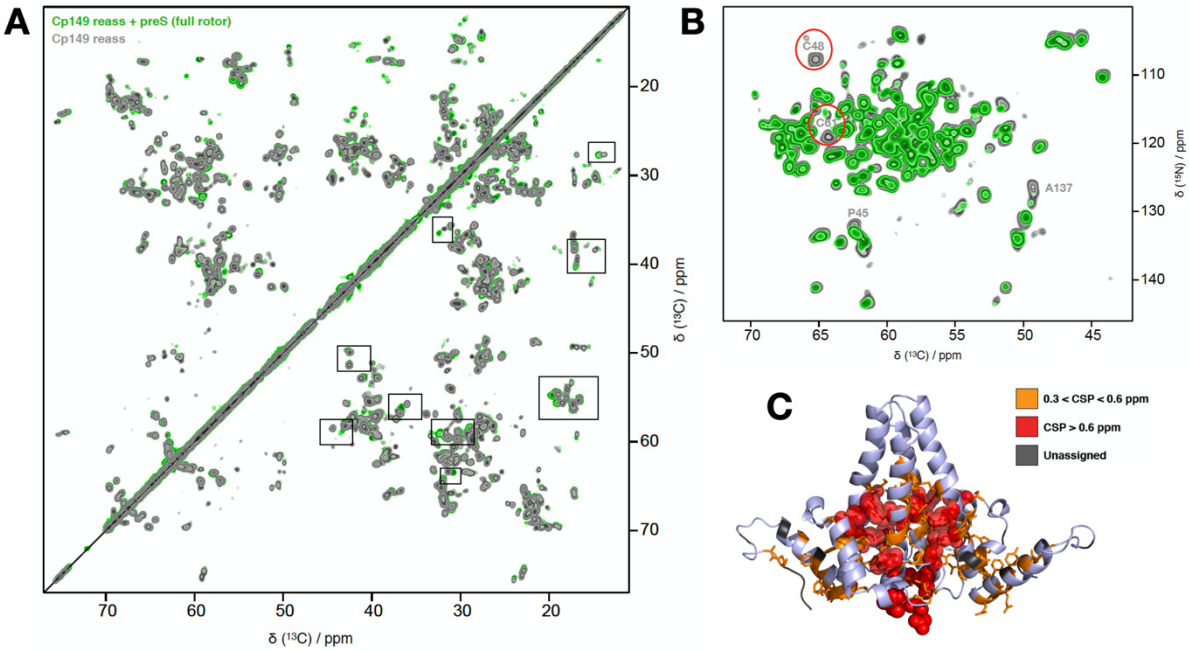

Figure 82. Solid-state NMR spectra for 3.2 mm rotor filled with $^{13}C^{15}N$-Cp149 capsids and 1.5 eq. preS. *A) DARR aliphatic region overview. B) NCA spectrum with oxidation of the cysteines. C) Cp149 dimer structure (PDB 1QGT) with high CSPs plotted in red. Experimental parameters are detailed in* ***Table 12****.*

We looked for the reason for this oxidation, but no clear answer could be obtained, it might be important to mention that the proteins we had mixed were stored in Cp149 buffer (50 mM HEPES pH 7.5, 5 mM DTT). However, with time and optimization, we realized that preS is particularly sensitive and needs to be kept in specific conditions. Thus, this sample was prepared again, but with both proteins in the preS buffer (20 mM $NaPO_4$ pH 6, 50 mM NaCl, 1X protease inhibitor). 2D DARR and NCA spectra were recorded, revealing small differences between the spectra in presence and in absence of preS (**Figure 83A**). In this buffer, there was no visible oxidation of cysteines (**Figure 83B**). Interestingly, the changes observed occur for residues in the hydrophobic pocket, *i.e.* residues which are affected by TX100 binding, as illustrated for alanine 58, where both forms A and B are visible (**Figure 83C**). In this case, we cannot know whether the changes are due to the presence of the two capsid forms in the sediment, or to the interaction with preS.

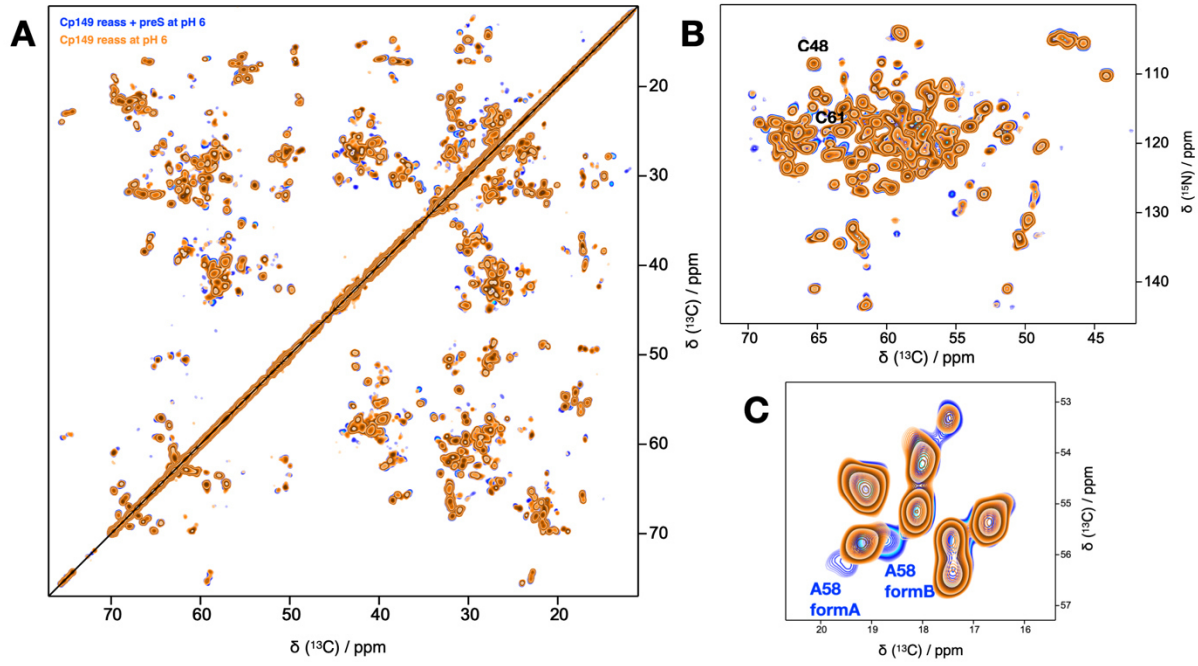

Figure 83. Solid-state NMR spectra for the second 3.2 mm rotor filled with $^{13}C^{15}N$-Cp149 capsids and 1.5 eq. preS. *A) DARR aliphatic region overview. B) NCA spectrum with no oxidation of the cysteines. C) Zoom of the alanine region. Experimental parameters are detailed in* ***Table 12****.*

Since preS has a low yield and stability, and in order to add the information on the protons (which are usually more sensitive to interactions), we prepared a new sample to fill a 1.3 mm rotor, for which only ~ 3 mg of protein are required compared to ~ 15 mg). This rotor diameter allows to spin at 60 kHz MAS frequency, enabling proton-detection spectra recording

of deuterated proteins. 2D hNH spectra of $^{2}H^{13}C^{15}N$-Cp149 reassembled capsids in preS buffer (20 mM PO_4 pH 6, 50 mM NaCl, 1X protease inhibitor) in presence and in absence of preS were recorded (**Figure 84**). Some small CSPs are observed, revealing an interaction between the two proteins.

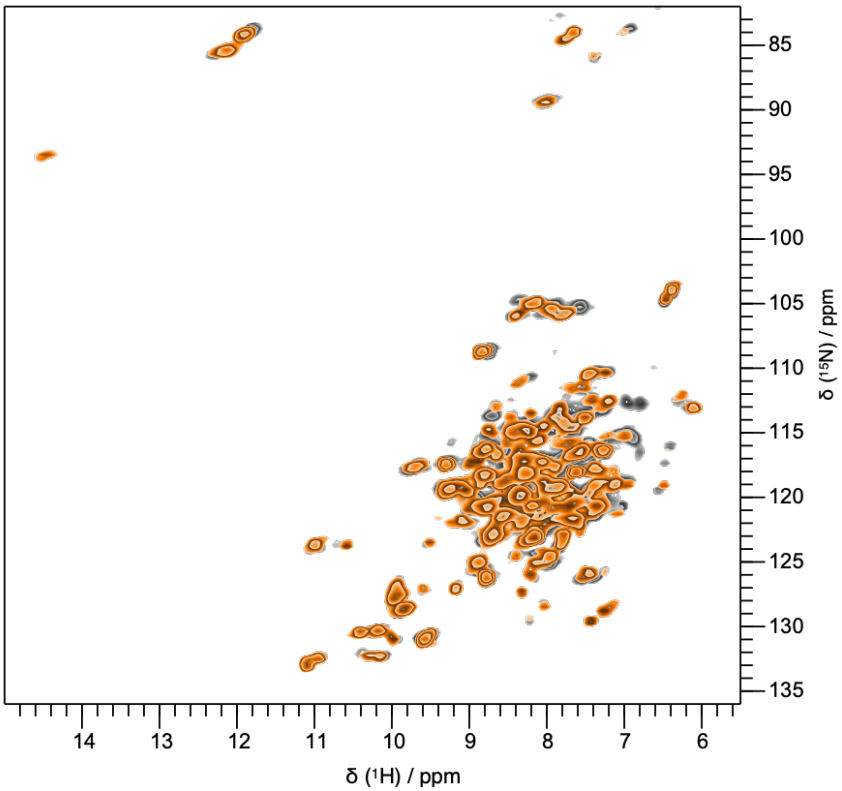

Figure 84. Solid-state NMR hNH spectrum acquired from 1.3mm rotor filled with $^{2}H^{13}C^{15}N$-Cp149 capsids and 1.5 eq. preS. *The reference spectrum is colored in grey, and the sample of labeled Cp149 with preS is shown in orange. Experimental parameters are detailed in* ***Table 13****.*

In order to increase resolution and to get detailed information on the binding, we recorded 3D spectra to allow precise assignment of the Cp+preS complex, and so to be able to accurately distinguish residues when peaks are overlapped on hNH spectra. 3D hCANH also allowed to check the oxidation state of C48, which is in reduced form in both samples, excluding that CSPs could be due to that. ^{1}H-^{15}N CSPs were calculated for each residue and are shown in (**Figure 85A**). Residues that are predominantly impacted (CSPs > 0.075 ppm) are mapped on the structure of the Cp149 dimer (**Figure 85B**), revealing that some of them are located on the spike, and that other impacted residues are located more at the base of the dimer or even close to the inter-dimer area. From these results, we can see that not only one part of the protein is impacted, which could mean that preS could interact at different sites on the Cp dimer. In addition, it can be seen that the hydrophobic pocket residues are not necessarily impacted, which would tend to suggest that the shifts observed on the previously recorded DARRs could come from a TX100 problem in the pocket of some capsids.

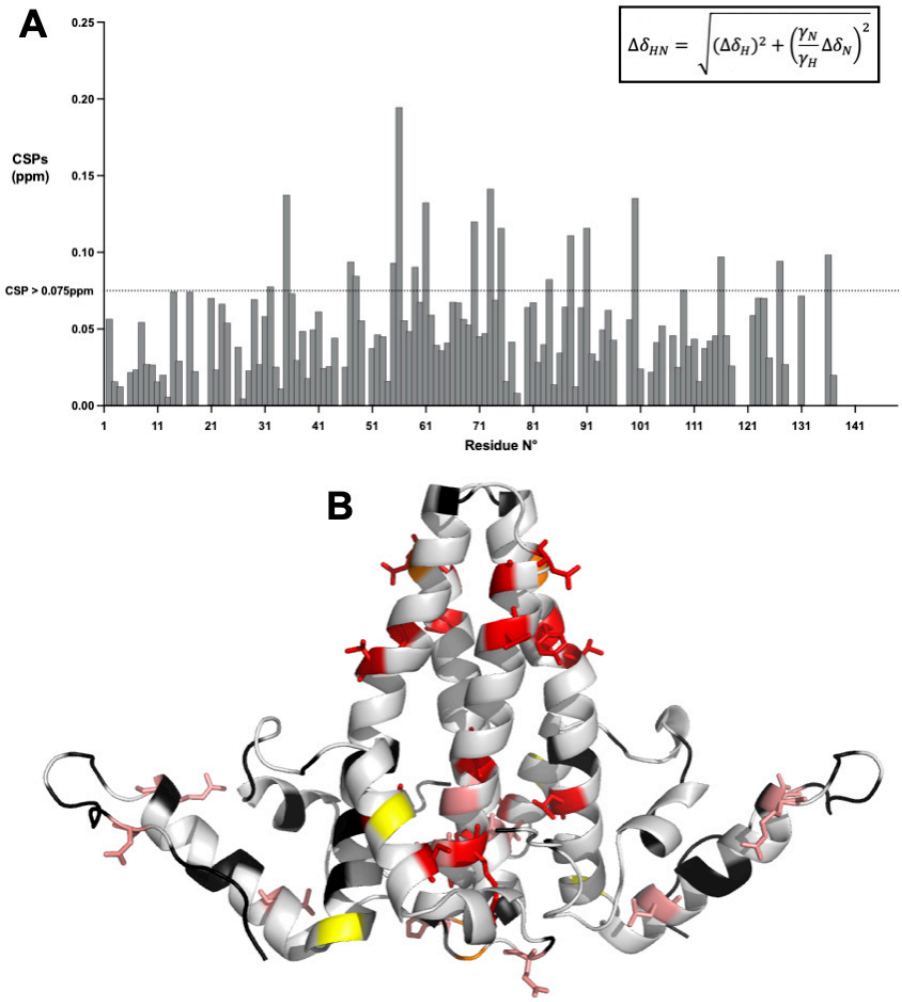

Figure 85. CSPs observed for the comparison between 2D hNH spectra. *A) CSPs graph plotted from hNH spectrum of $^{2}H^{13}C^{15}N$-Cp149 reassembled capsids (comparison between the reference the sample with 1.5 equivalent preS). B) Cp149 dimer structure with residues impacted by the binding of preS from hNH spectra analysis. Red residues correspond to CSPs > 0.1 ppm, pink to 0.1 > CSP > 0.09 ppm, orange to 0.09 > CSP > 0.08 ppm, and yellow to 0.08 > CSP > 0.075 ppm. Unassigned residues are colored in black.*

CSPs were calculated for ^{1}H, ^{15}N and ^{13}Ca nuclei separately using hCANH spectra (**Figure 86A**), and then their respective CSPs were added together to see the overall result for each residue. Some residues display high CSPs (> 1.2 ppm) including Y6, S17, T33, T53, R56, C61, L65, T70, V86, N90, N92, N136 which are mapped on the dimer structure (**Figure 86B**). These large CSPs are mainly due to displacements at the carbon level which were not considered on the CSPs plot of the hNH spectrum. It is possible to notice that the residues at the top of the spike do not seem to be affected, but rather the area just below. There are also residues at the interface between the two dimers which have high CSPs. We can assume that some residues are affected because of allosteric effects.

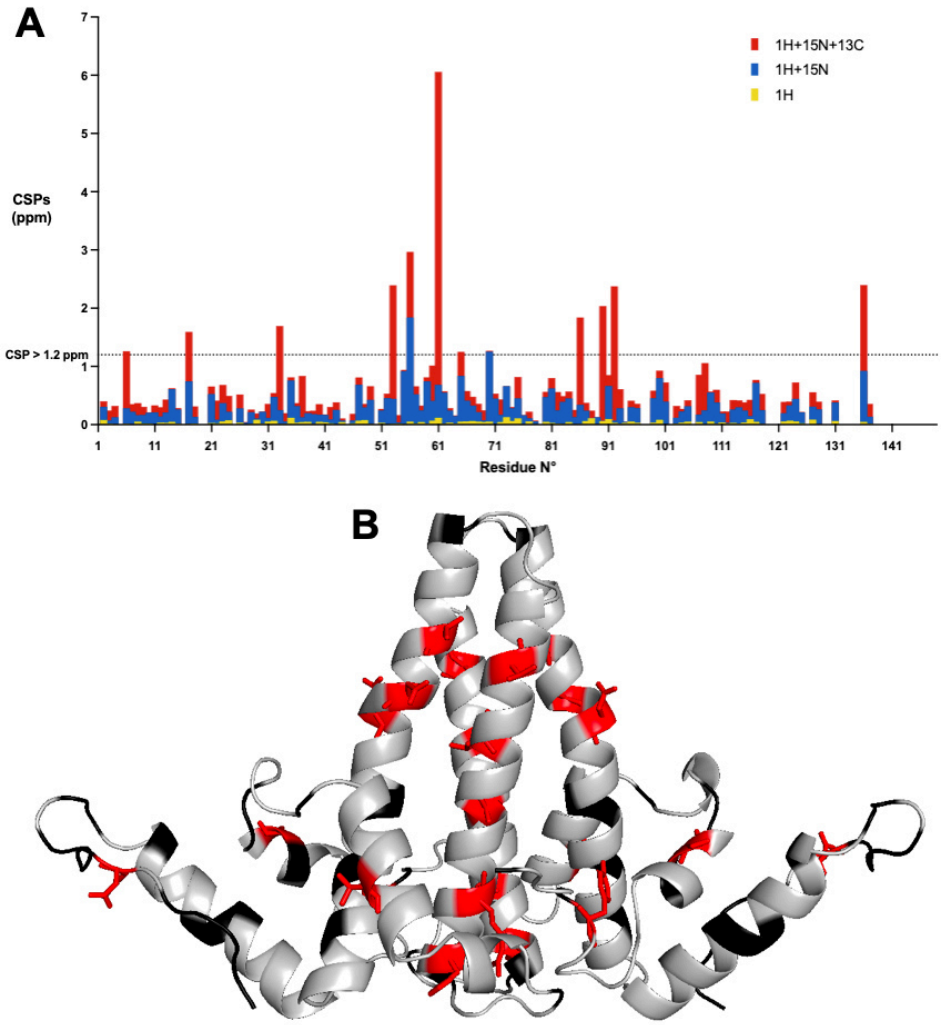

Figure 86. CSPs observed for the comparison between 3D hCANH spectra. *A) CSPs graph plotted from hCANH spectrum of $^{2}H^{13}C^{15}N$-Cp149 reassembled capsids (comparison between the reference the sample with 1.5 equivalent preS). The results are given for each nuclei: yellow for ^{1}H, blue for $^{1}H+^{15}N$ and red for $^{1}H+^{15}N+^{13}C$. B) Cp149 dimer structure with residues impacted by the binding of preS from hCANH spectra analysis. Red residues correspond to CSP > 0.075 ppm. Unassigned residues are colored in black.*

These results are interesting because they are partially in agreement with the results highlighted in (Ponsel & Bruss, 2003). Indeed, it was shown that 11 mutations on Cp would allow the formation of nucleocapsids but would block the envelopment and thus the formation of virion. Among these mutations, we were able to detect high CSPs on the hNH spectra (> 0.075 ppm) for 3 of them, S17, R127 and N136, which may allow us to predict that our results could be relevant.

We then used ITC to obtain information about thermodynamic parameters of the binding event. First, we started by performing an experiment under similar conditions as the ones used for the first 3.2 mm rotor sample, *i.e.* with both proteins in HEPES buffer at pH 7.5. The ITC experiment showed that in this case, Cp149 reassembled capsids and preS do not interact (**Figure 87**). This confirms that the CSPs observed in **Figure 82C** were indeed due to

the different oxidation state of the cysteine residues of Cp149, and not to an interaction with preS.

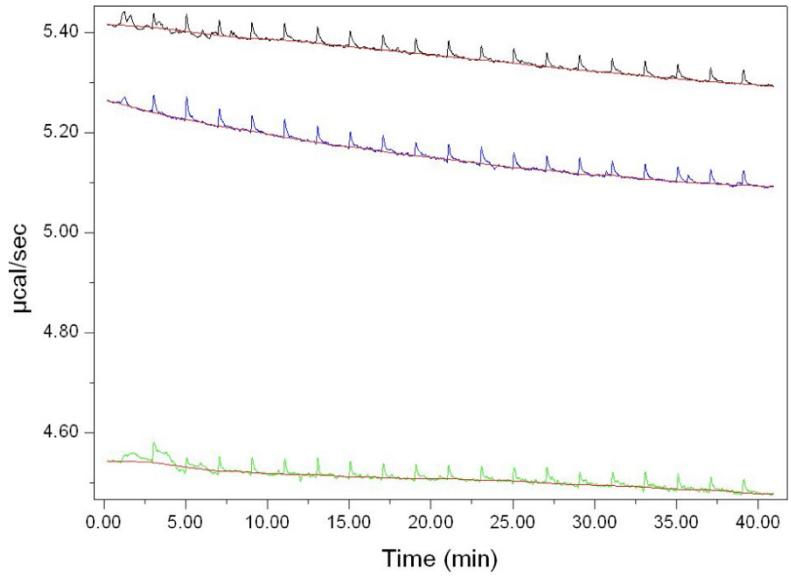

Figure 87. Raw ITC curves of preS titrated by Cp149 dimer in HEPES buffer at pH 7.5. *The green curve corresponds to the blank experiment, and the black and blue curves to the duplicate titration experiment. Experiments were done with typical concentration of [preS] = 26 µM and [Cp149$_{monomer}$] = 200 µM.*

In order to understand this result, we performed circular dichroism (CD) analysis of preS at the two different pH (**Figure 88**). These graphs showed that in the buffer at pH 7.5 (corresponding to capsids HEPES buffer), the protein does not have the same curve as when it is at pH 6 (the one corresponding to preS phosphate buffer).

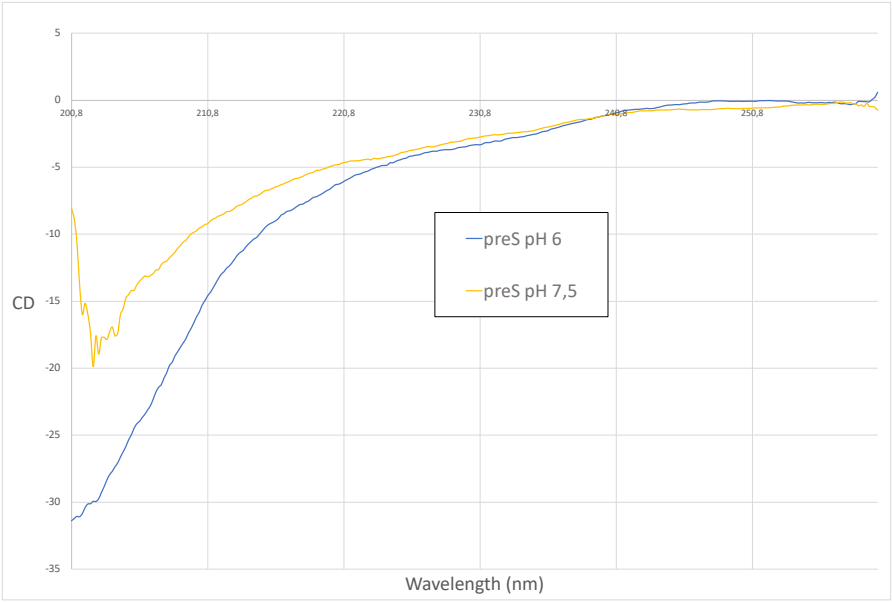

Figure 88. CD curves of preS in 20 mM PO_4 50 mM NaCl, 1X protease inhibitor (preS buffer) at two different pH.

We can thus see that preS is affected by its storage in phosphate buffer at pH 7.5. However, the CD curve does not allow us to give an exact reason for the interaction problems between the two proteins in 50 mM HEPES, 5 mM DTT at pH 7.5, as HEPES cannot be used for CD because it creates interferences.

After confirming that the two proteins did not interact in HEPES buffer at pH 7.5, we dialyzed them both in phosphate buffer at pH 6 to see if the interaction could be characterized under these conditions. This experiment allowed us to determine the thermodynamic information of the binding (**Figure 89**). First, the peaks measured by ITC are positive, whereas they were always negative in the measurements between Cp149 and the hydrophobic pocket ligands. This information tells us that the reaction is endothermic, meaning that it absorbs heat from outside. This indicates that the interaction is most likely hydrophobic . Moreover, these measurements showed that the interaction was strong, in the order of a few µM. This seems surprising compared to the results obtained with the NMR spectra which do not show large CSPs (except for hCANH spectra). Finally, the ITC allowed us to determine the stoichiometry (n = 0.6 for the preS:monomer ratio) which indicates that one preS protein binds to one dimer of Cp149 in the capsid (*i.e.* just to remember that an isolated dimer does not interact with preS).

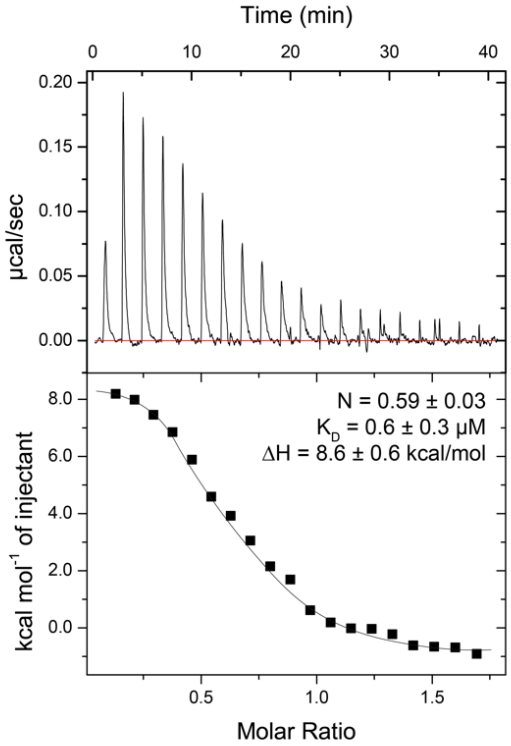

Figure 89. ITC raw data and binding isotherms displaying the titration of Cp149 reassembled capsids against preS. *Experiments were done with typical concentration of [preS] = 26 µM and [$Cp149_{monomer}$] = 200 µM.*

Finally, according to (Sahu et al., 2016), it is possible to see changes from endothermic to exothermic reaction in the ITC curves for interactions involving IDPs when measurements are made at different temperatures. We have therefore measured the system at different temperature, by making successive jumps of 5 °C to see the impact on the curves (**Figure 90**). Thus, we can see that the peaks became more and more negative when the temperature increases. Unfortunately, we could not complete the process because the preS protein does not tolerate being subjected to high temperatures, making the measurements unusable.

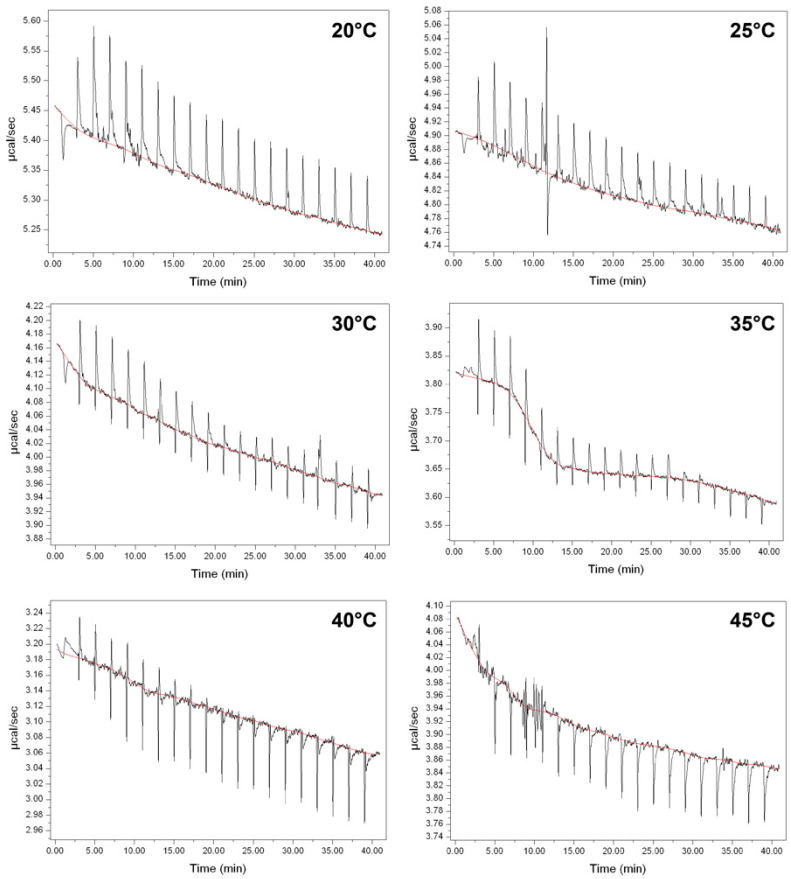

Figure 90. ITC curves of preS titrated by Cp149 at different temperatures. *Raw curves with temperature indicated on the top right corner of each run. Experiments were done with typical concentration of [preS] = 18 µM and [Cp149$_{monomer}$] = 100 µM.*

c. Interaction between Cp149 capsid (form A) and preS

After proving and characterizing the interaction between the Cp149 reassembled capsids (form B) and preS, we wanted to see if the binding between the two proteins was still possible with the Cp149 capsids in form A, *i.e.* with the hydrophobic pocket obstructed by TX100. In order to be sure that all hydrophobic pockets were full of TX100, we added an excess of ligand and then performed a dialysis to remove unbound TX100. We then titrated preS with Cp149 capsids (form A) using solution-state NMR. The recorded spectra showed that preS peaks decrease (**Figure 91**), as already observed for reassembled capsids (no

TX100), however the decrease is less drastic as more peaks are still visible even after adding 5 equivalents of capsids (comparison on **Figure 91**). These observations therefore suggests that the Cp149 capsid form A interacts less with preS, which could be due to conformational changes induced by TX100 on the structure of the core protein.

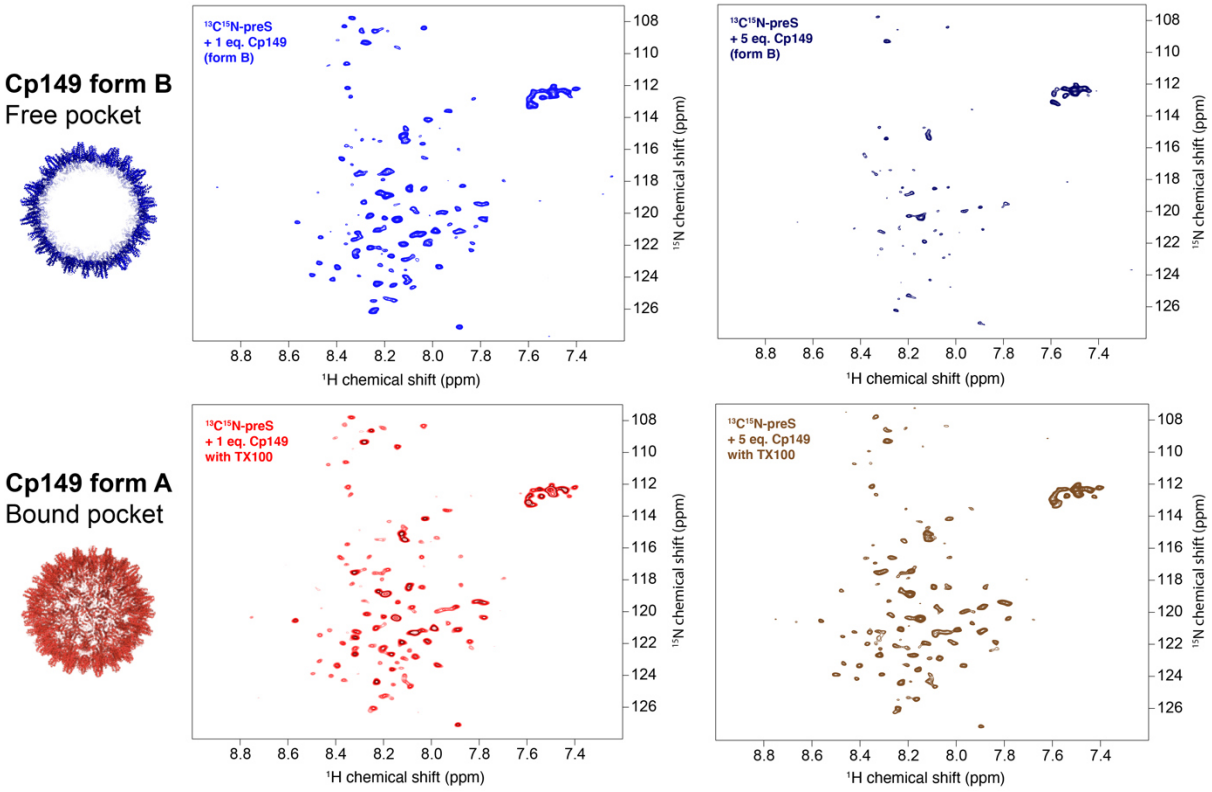

Figure 91. Titration of $^{13}C^{15}N$-preS by Cp149 capsids (form A) compared with titration of preS by Cp149 capsids (form B). *The upper panel shows the spectra recorded with addition of Cp149 form B and the lower panel the spectra for addition of Cp149 form A. On the left spectra, 1 equivalent of Cp149 was added, and on the right side, 5 equivalents. Experimental parameters are detailed in* ***Table 10****.*

To confirm our observation, we prepared a new 1.3 mm rotor with $^{13}C^{15}N$-Cp149 mixed with 1.5 eq. preS, and we recorded an hNH spectrum to see if we could observe any changes (**Figure 92**). We see indeed less or even non at all CSPs on this spectrum, confirming that TX100 may interfere in the interaction between both proteins. In order to evaluate the CSPs for the Cp form A interaction with preS, additional spectra will have to be recorded, including 3D, in order to evaluate whether the large CSPs observed for form B in these spectra are present also for form A. Indeed, the study of an hCANH spectrum could allow to observe changes even if they are less important.

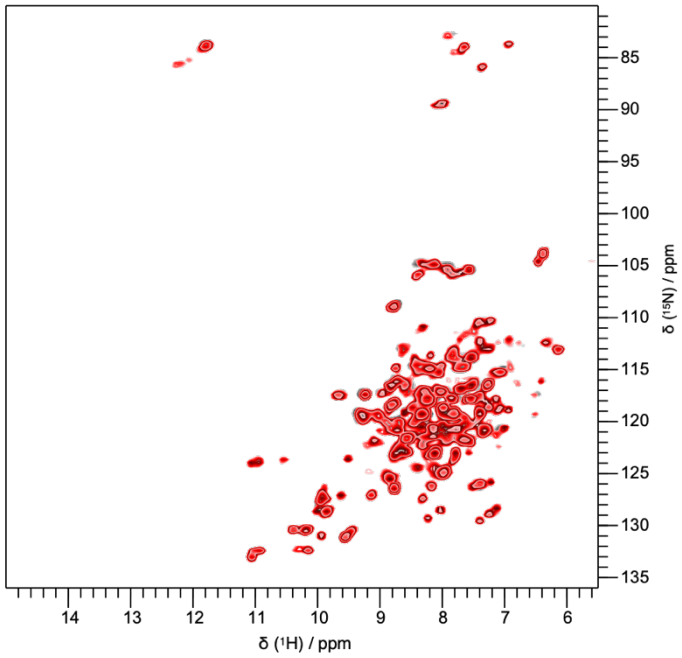

Figure 92. Solid-state NMR hNH spectrum acquired from 1.3 mm rotor filled with $^{2}H^{13}C^{15}N$-Cp149 capsids form A and 1.5 eq. preS. *The reference spectrum is colored in grey, and the sample of labeled Cp149 form A with preS is shown in red. Experimental parameters are detailed in* ***Table 13****.*

d. Cell-free co-expression tests

After using ITC, solution-state NMR and solid-state NMR to prove the interaction between the preS domain of HBs L and the Cp149 protein, we wanted to perform co-expression tests in cell-free system for the full-length Cp183. For this purpose, we performed these tests using wheat germ cell-free protein synthesis, a system that is used in our laboratory. Thus, we co-expressed the proteins in bilayer mode in order to identify if there is an interaction using western blot analysis.

First, we tested the interaction between preS1 and preS domains of HBs L with Cp183. For the interaction with preS1 domain, a band corresponding to Cp183 can be observed the SN-Beads fraction when it is co-expressed with preS1, but also when it is produced alone (**Figure 93**). In the latter case the band is however weaker, which could lead us to believe that Cp183 could interact with preS1. However, this result should be taken with caution, and must be confirmed by reproducing the experiment.

Since we already proved the interaction between preS and Cp149, we wanted to evaluate the binding of preS with Cp183. We tested two preS constructs with a Strep-Tag in N-ter or in C-ter position (ST-preS and preS-ST, when ST means Strep-Tag). The bands corresponding to ST-preS and preS-ST are quite weak, meaning that their expression level is low. Both proteins are nevertheless detected in the CFS and SN-Beads fractions (**Figure 93**). Interestingly, Cp183 is found in the SN-Beads fraction together with preS-ST, but only very weakly with ST-preS, although the expression level of preS-ST is even lower than the one of ST-preS. This means that the interaction site could be in the N-terminal region of preS, and

that the N-terminal tag could hinder the interaction. As the results with preS1, these observations should be taken with caution, as they are not always reproducible.

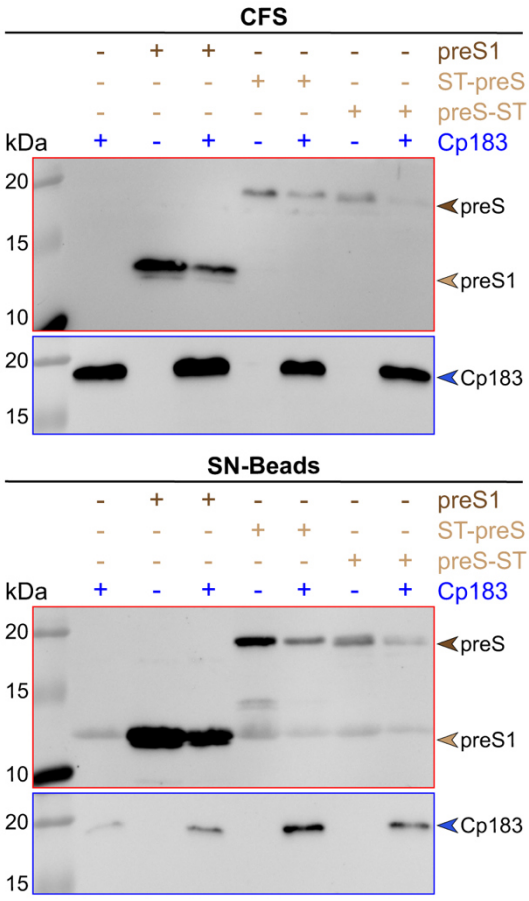

Figure 93. Western blot of preS1 or preS/Cp183 co-expression test. *Two different antibodies were used, the StrepMAB Classic in red frame, and the anti-Cp183 in the blue frame.*

Knowing that the preS fragment interacts with Cp183, we next tested the interaction between Cp183 and full-length HBs S and HBs L (**Figure 94**). A band corresponding to Cp183 is clearly detected on the SN-beads together with HBs L, while only a very weak aspecific band is observed in this fraction when Cp183 is expressed alone (**Figure 94 a)**).

Interestingly, we also tested the interaction between Cp149 and HBs L. This preliminary result has to be confirmed but it seems that these two proteins do not interact upon synthesis. No band corresponding in the SN-beads fraction when Cp149 was co-expressed with HBs L (**Figure 94 b)**). This result is in agreement with our previous observations showing that Cp149 dimer did not interact with preS. Indeed, in a cell-free system, Cp149 protein does not form capsids but only dimers (S. Wang et al., 2019). This would imply that Cp149 dimer might not interact with HBs L either. We thus show here that Cp183 and HBs L interact upon cell-free synthesis. In addition, in order to better characterize the interactions between the capsid and the envelope, we also tested the potential interaction of Cp183 with HBs S. A band corresponding to Cp183 is clearly detected on the SN-beads together with HBs S, clearly

meaning that they interact upon synthesis (**Figure 94 c)**). These data suggest that at least two binding sites are involved in the interaction between the capsid and the envelope proteins, one is located in the preS, and the second one in the S domain. These results in WG-CFPS are interesting because according to (Poisson et al., 1997), each of the preS1 and S proteins interact with the core protein, and we could see the same interpretations.

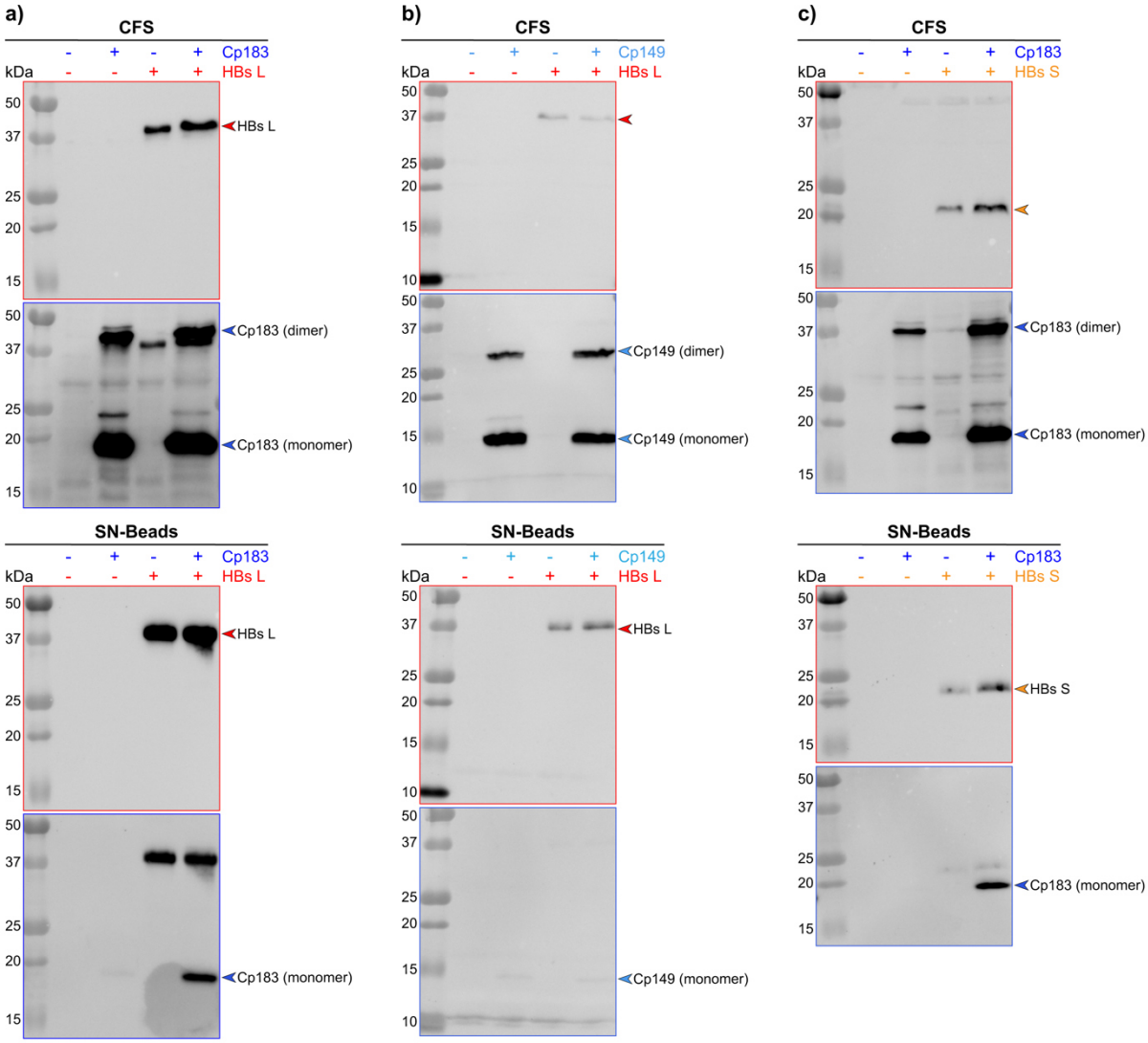

Figure 94. Western blot analysis of full-length envelope proteins/ core proteins co-expression. a) Cp183/HBs L. b) Cp149/HBs L. c) Cp183/HBs S. *Two different antibodies were used, the StrepMAB Classic in red frame, and the anti-Cp149/183 in the blue frame. In panel a, the band corresponding to HBs L is observed also in the blue frame because the same membrane was incubated with the two antibodies successively without stripping.*

The results we have obtained and described are preliminary, and it would be important to be able to reproduce these co-expression tests to confirm them. Furthermore, it would be interesting to test the interaction between envelope proteins and the Cp149/Cp183-F97L mutant, or to see the impact of preS mutations on the interaction with Cp. Finally, it should be considered that preS is phosphorylated when produced in wheat-germ cell-free system (Fogeron et al., 2021), and this may also impact the interaction.

III. Discussion

Using ITC and solid-state NMR, we were able to characterize the interaction between Cp149 and the Oct peptides, and we could see that these peptides interact mainly at the spike of the core protein, and that they bound with micromolar affinity. These results are comparable with those obtained and discussed in (Freund et al., 2008) and (Makbul, Khayenko, et al., 2021), both showing that Oct peptides bind to the spike of the core protein. However, our study provides additional data with previously unobserved CSPs, located notably on amino acids distant from the spike. Our results also validate the postulate in (Böttcher et al., 1998) indicating that residues 77 and 78 of Cp149 are important for peptide binding, and which in our case present high CSPs on solid-state NMR analysis. Spectra involving polarization transfer between Oct2 peptide and the core protein may provide further information on the binding site and on the atoms of the peptide that interact with those of the protein. It could also be interesting to test the structural changes induced by Oct2 binding on a Cp mutant, *e.g.* F97L or L60V and P5T, which respectively causes premature (Ceres et al., 2004) or low-secretion (Le Pogam et al., 2000). Indeed, in their recent study, Makbul and Khayenko have observed that the peptide binding seem to be independent of the secretion phenotype (Makbul, Khayenko, et al., 2021).

Then, we demonstrated using ITC and NMR that there is a very specific interaction between the HBV capsid and the preS domain *in vitro* (**Figure 95**), which is in line with the results obtained in (Bruss, 1997; Pastor et al., 2019). Importantly, we showed that preS is not able to interact with Cp dimer, but exclusively with the Cp capsids, and in specific conditions (20 mM PO_4 pH 6, 50 mM NaCl). We show here using ITC that this interaction takes place in the micromolar affinity and is endothermic. This likely indicates hydrophobic interaction between the two proteins. Whether this affinity will be higher with envelopment-modifying mutants that interact with immature capsids remains to be established. Interestingly, our data give a first indication that TX100, a compound we have shown to insert into the Cp hydrophobic pocket, interferes at least partially with the binding of preS to Cp capsids. Indeed, in solution NMR we could see that preS is less affected by the binding with the free-pocket capsid (form B) than with the TX100-bound-pocket capsid (form A) since the intensity decrease was weaker. In addition, in solid-state NMR, no CSPs were observed on the 'bound' capsid upon addition of preS (**Figure 92**), while significant CSPs indicate binding between the capsid with the free pocket and preS. Yet, the precise localization of the binding still needs to be further addressed since the CSPs observed are rather dispersed on the Cp structure and are not fully understood yet. In addition, it would be interesting to assess whether, as Dyson and Murray suggest (Dyson & Murray, 1995), Oct peptides, once bound to the Cp spike, could inhibit the interaction between Cp and preS *in vitro*.

Finally, co-expression tests in wheat-germ cell free system allowed us to confirm our first hypotheses on the interaction between preS and Cp, but they also allowed us to consider the binding between preS1/HBs S/HBs L and Cp (**Figure 95**). These results are in agreement with those described in (Poisson et al., 1997). Furthermore, the phosphorylation of preS in the

cell-free system suggests that preS is still able to interact with Cp when it is phosphorylated, but this is another research lead to find out if this is important.

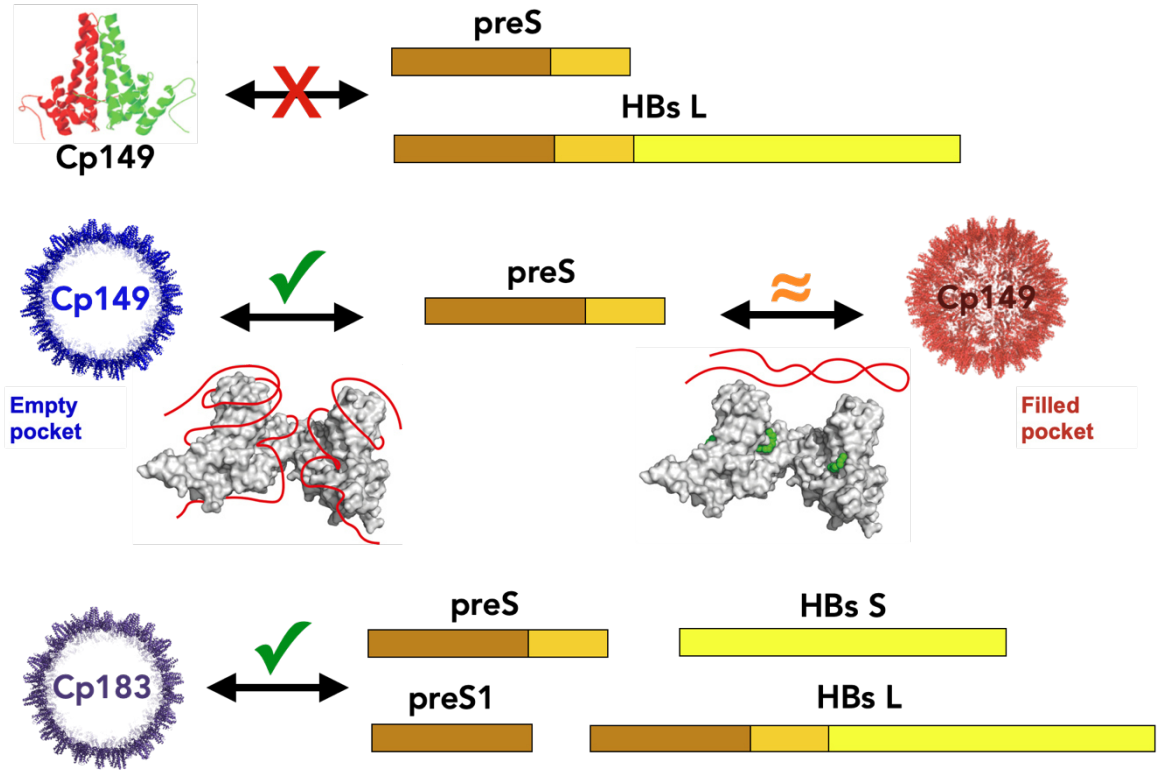

Figure 95. Scheme of the different interactions between core and envelope proteins. *For Cp149 and preS interaction, the binding of preS (represented by a red line) is modelized on the Cp dimer structure (PDB 1QGT).*

The work here, if still preliminary, sets the stage to investigate the Cp and envelope protein interactions in detail. Subsequently, if we manage to get information on the precise localization of the binding on the preS side, it will be interesting to build a docking model of the Cp-preS complex. For this, we can consider looking at the NMR spectra of labeled preS in presence and absence of unlabeled capsid by solid-state NMR. A highly interesting venue is to assess also the interaction of mutant forms of the Cp, especially those which impact envelopment, with preS. Next steps shall include to complete the CSP analysis of the Cp form A/preS interaction and compare it to form B. The same approach shall be used to investigate preS binding to a preformed Cp/Oct complex. Finally, the investigation, also by NMR, of the more complex S/Cp interaction shall be pursued. Also, polarization-transfer methods between the Cp and its ligands will be established using Cp/Oct as a model, with the goal of adapting them to study contacts in the Cp/preS complex.

IV. Conclusion

We have investigated here the interactions of the Cp with different interaction partners. Starting from the observation that a small-molecule binder can insert in the hydrophobic pocket (Briday et al., 2022; Lecoq et al., 2021), we further investigated its binding mode, and established all biochemical tools to investigate its impact on the interaction between the capsid and the viral envelope proteins. We could establish that a binding event to the hydrophobic pocket indeed modifies the preS/Cp interaction, which presents a first step towards the establishment of molecules which could interfere with this step. Our work sets the stage for further work of Cp/L interactions, in order to establish how preS interacts with the capsid.

V. Materiel and methods

1. Bacterial expression and purification of Cp149

The protein is produced using the same protocol described in the paragraph “Expression and purification of Cp149” on page 91.

2. Capsid disassembly and reassembly *in vitro*

The protein preparation follows the same steps as described in the paragraph “Capsid disassembly and reassembly *in vitro*” on page 92. For ITC and solution NMR experiments, Cp149 dimer solution was dialyzed in 20 mM $NaPO_4$ pH 6, 50 mM NaCl, with addition of 1X EDTAfree protease inhibitor and directly used. Dimer samples are always controlled using negative staining electron microscopy to ensure the absence of capsids before running ITC and solution NMR experiments. For ITC and solid-state NMR, the capsids were dialyzed in 50 mM HEPES pH 7.5 for the tests with Oct peptides, and 20 mM PO_4 50 mM NaCl supplemented with 1X EDTAfree protease inhibitor overnight at 4 °C for the tests with preS. The resulting reassembled capsids were checked by negative straining electron microscopy.

3. Bacterial expression and purification of preS

E. coli BL21 (DE3) cells were transformed with the plasmids pRSF-H6GB1-HpreS1/2 and grown at 37 °C in 2 tubes filled with 5 mL LB minimum medium culture in presence of kanamycin. Precultures were diluted to yield a 2 L culture, and cells were grown at 37 °C until T7 promoter-controlled protein expression reached an OD_{600} of 1.2 using 1 mM of IPTG overnight at 30 °C. Then cells were harvested by centrifugation (6,000 g, 20 min, 4 °C), and finally resuspended in 4 ml/g of cell pellet lysis buffer (20 mM $NaPO_4$ pH 7.5, 500 mM NaCl, 40 mM imidazole) supplemented with EDTAfree protease inhibitor. Lysis was performed by incubating cells with 1 mg/ml of lysozyme for 50 min at 4 °C under stirring, and nucleic acids digestion was performed by the addition of benzonase nuclease 250 U/µL (6 µL/L of culture) supplemented with 2 mM $MgCl_2$ during 30 min at room temperature. Cellular membranes were broken using a microfluidizer at 15,000 psi (three successive runs). Soluble proteins in the supernatant were isolated by centrifugation (25,000 g, 30 min, 4 °C) and filtrated using a 0.45 µm PVDF membrane. Then, the solution was loaded into a 5 mL HisTrap affinity column, connected to an NGC chromatography system. The His-tagged proteins were eluted with elution buffer (20 mM NaPO4 pH 7.5, 500 mM NaCl, 1 M imidazole) after having washed the

HisTrap column with lysis buffer. The eluted proteins were dialyzed in TEV protease reaction buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 5 mM DTT) using a 3.5 kDa cut-off membrane, at 4 °C under slow stirring. GB1 fusion proteins were cleaved using homemade TEV protease 10 U/µl (0.2 mL or 2,000 U per milligram of preS protein) by incubating overnight at 4 °C under stirring. Cleaved preS protein was recovered using Strep-Tactin resin in batch mode. 1X EDTA-free protease inhibitor was added to the purified protein solution. PreS was dialyzed using a 3.5 kDa cut-off membrane overnight in final buffer (20 mM $NaPO_4$ pH 6, 50 mM NaCl). Protein concentration was determined by Nanodrop measurement and the absorbance at 280 nm. Subsequently, the protein solution was concentrated by immerging the dialysis bag containing the protein and protease inhibitor into Sephadex G-25 powder at 4 °C. Concentration was followed by weighing the dialysis bag before adding the powder and every 12–20 h thereafter. This approach was used to avoid preS sticking to cellulose membranes used in concentrators as Amicon which are commonly used for other proteins. The protein concentration was measured by Nanodrop, and proteins are frozen in liquid nitrogen before their storage at -80 °C to ensure their stability. Protein solutions in all purification steps were analyzed using 15% polyacrylamide SDS-PAGE gels. Labelled proteins were expressed following the same protocols as unlabeled ones, except that those bacteria were grown in M9 minimal medium containing ^{13}C-deuterated glucose and $^{15}NH_4Cl$.

4. Solution NMR

a. preS protein

Freshly prepared ^{13}C-^{15}N-preS in 20 mM $NaPO_4$ pH 6, 50 mM NaCl, 1X EDTAfree protease inhibitor (pH 6) was quantified by absorbance measurements and concentrated to approximately 10 µM. Cp149 dimer and capsids (form B) were added at preS:Cp monomer ratio of 0.5, 1, 2 and 5, and ratio 1 and 5 for Cp149 capsids (form A). Solutions were transferred into 3 mm solution NMR tubes (for the titration with Cp149 dimer) and in 5 mm tubes (for titration with Cp149 capsids), and ^{1}H-^{15}N-SOFAST spectra were recorded on all samples on a 600 MHz spectrometer (Bruker Avance II) or 950 MHz spectrometer (Bruker Advance III) equipped with a triple-resonance cryo-probe. NMR acquisition parameters are detailed in **Table 10**. The assignment of preS was previously done in (Fogeron et al., 2022) and is accessible in the BioMagResBank (http://www.bmrb.wisc.edu/) under accession code 51186. All spectra were recorded at 295 K and processed using TopSpin 4.0.8 (Bruker Biospin) and analyzed with the CcpNmr Analysis package, Version 2.4.2 (Vranken et al., 2005). The intensities of each peak for each amino acid were calculated by CcpNmr software between the control sample and the sample preS + 2 eq. Cp149 (form B).

Table 10. *NMR experimental details of solution NMR ^{1}H-^{15}N-SOFAST spectra recorded at 298 K on ^{13}C-^{15}N-preS for titration with Cp149 dimer, Cp149 capsids (form B) and Cp149 capsids (form A). 100 points were recorded in ^{15}N dimension, with a spectral width of 24 ppm. The Cp149 concentration is given for the monomer unit.*

Ligand	[preS] (µM)	Field (MHz)	Number of scans	Exp. time
Cp149 dimer	13.5	950	24 (0 eq.)	40 min

			26 (0.5 eq.)	38 min
			28 (1 eq.)	41 min
			34 (2 eq.)	50 min
			52 (5 eq.)	1h16
Cp149 capsids (form B)	10	600	32 (0 and 0.5 eq)	35 min
			48 (1 and 2 eq.)	52 min
			64 (5 eq.)	1h10
Cp149 capsids (form A)	10	600	16 (0 and 1 eq.)	17 min
			32 (5 eq.)	35 min

b. Cp149 dimer protein

Freshly prepared ^{2}H-^{13}C-^{15}N-Cp149 dimer in 20 mM $NaPO_4$ pH 6, 50 mM NaCl, 1X protease inhibitor was quantified by absorbance measurements, and concentrated to approximately 45 µM (in monomer concentration). preS was added at a Cp monomer:preS ratio of 1:1. Solutions were transferred into 5 mm Shigemi solution NMR tubes, and ^{1}H-^{15}N-SOFAST spectra were recorded on a 600 MHz spectrometer (Bruker Avance II) equipped with a triple-resonance cryo-probe. NMR acquisition parameters are detailed in **Table 11**. All spectra were recorded at 295 K and processed using TopSpin 4.0.8 (Bruker Biospin) and analyzed with the CcpNmr Analysis package, Version 2.4.2 (Vranken et al., 2005).

Table 11. *NMR experimental details of solution NMR measurements recorded on $^{2}H^{13}C^{15}N$-Cp149 dimer. sw stands for spectral width. 3 mm tubes were used for all recordings. The concentration is given for the monomer unit.*

Ligand	[Cp149] (µM)	Field (MHz)	sw {^{15}N} (ppm)	Increments {^{15}N}	Number of scans	Exp. time
preS	45	600	29	200	64 (0 eq.)	1h08
					128 (1 eq.)	2h17

5. Solid-state NMR

a. Carbon detection using 3.2mm rotors

Freshly prepared ^{3}C-^{15}N-Cp149 reassembled capsids in 50 mM HEPES, 5 mM DTT (pH 7.5) were incubated with 4 equivalents of Oct1 and Oct2 peptides for 1 hour on ice. Freshly prepared ^{13}C-^{15}N-Cp149 reassembled capsids in 50 mM HEPES, 5 mM DTT (pH 7.5) and in 20 mM PO4, 50 mM NaCl and 1x protease inhibitor (pH 6) were incubated with 1 equivalent of preS for 1 hour on ice. They were then concentrated using Amicon Ultra centrifugal filter units (Merck, 50 kDa cut-off) to about 20 mg/ml in 1 mL and sedimented into 3.2 mm zirconium rotors by ultracentrifugation (200,000 g, 14 h, 4 °C) using a home-made filling tool. Rotors were immediately closed after the addition of 1 µL of saturated DSS solution for chemical-shift referencing.

2D ^{13}C-^{13}C-DARR (Takegoshi et al., 2001) (for Oct1, Oct2 and preS) and 2D NCA (for preS) spectra were recorded using a 3.2 mm triple-resonance (^{1}H, ^{13}C, ^{15}N) wide-bore probe head at a static magnetic field of 18.8 T corresponding to 800 MHz proton resonance frequency (Bruker Avance II). All spectra were referenced to DSS and recorded at a sample temperature of 4 °C according to the resonance frequency of the supernatant water. Assignments were derived from those presented in (Lecoq et al., 2018) (bound-state, BMRB number 27317) and (Lecoq et al., 2021) (unbound state, BMRB number 28122). NMR acquisition parameters are detailed in **Table 12**. All spectra were processed using TopSpin 4.0.3 (Bruker Biospin) and analyzed with the CcpNmr Analysis package, Version 2.4.2.

For Oct1 and Oct2 peptides, chemical-shift differences between Cp149 reassembled capsid without and with ligand were calculated for each carbon atoms according to: $\Delta\delta_C = \delta_C[bound] - \delta_C[unbound]$.

Table 12. *NMR experimental details of solid-state NMR measurements of 3.2 mm rotors. sw stands for spectral width. Experiments were recorded on an 800 MHz wide-bore spectrometer at a 17.5 kHz MAS frequency and at a sample temperature estimated at 4 °C. The ^{15}N-insert was removed from the probe to increase the signal-to-noise ratio.*

Sample	**Cp149+Oct1**	**Cp149+Oct2**	**Cp149+preS (pH 7.5)**	**Cp149+preS (pH 6)**
Experiment	2D DARR	2D DARR	2D DARR	2D DARR
Transfer 1	HC-CP	HC-CP	HC-CP	HC-CP
Field [kHz]	67.0 (^{1}H) 50 (^{13}C)	65.7 (^{1}H) 50 (^{13}C)	69.0 (^{1}H) 50 (^{13}C)	66.7 (^{1}H) 50 (^{13}C)
Shape	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H
^{13}C carrier [ppm]	58.6	58.6	58.6	58.6
time [ms]	1.0	0.8	0.6	1.0
Transfer 2	DARR	DARR	DARR	DARR
Field [kHz]	17.5 (^{1}H)	17.5 (^{1}H)	17.5 (^{1}H)	17.5 (^{1}H)
^{13}C carrier [ppm]	100	100	100	100
time [ms]	20	20	20	20
t_1 increments	2560	2560	1450	1450
sw (t_1) [kHz]	93.75	93.75	52.5	52.5
Acq. time (t_1) [ms]	13.7	13.7	13.8	13.8
t_2 increments	3072	3072	1300	1300
sw (t_2) [kHz]	93.75	93.75	39.7	39.7
Acq. time (t_2) [ms]	16.4	16.4	16.4	16.4

^{1}H decoupling	SPINAL64	SPINAL64	SPINAL64	SPINAL64
Field [kHz]	90	90	90	90
Interscan delay d1 [s]	2.6	2.6	2.6	2.6
Number of scans	12	16	16	28
Measurement time	22 h 36	1 d 6 h	17 h 03	1 d 6 h

Sample	**Cp149+preS (pH 7.5)**	**Cp149+preS (pH 6)**
Experiment	2D NCA	2D NCA
Transfer 1	HN-CP	HN-CP
Field [kHz]	62.9 (^{1}H) 45.5 (^{15}N)	62.9 (^{1}H) 45.5 (^{15}N)
Shape	Tangent ^{1}H	Tangent ^{1}H
^{13}C carrier [ppm]	58.6	58.6
time [ms]	0.9	1.0
Transfer 2	NCa-CP	NCa-CP
Field [kHz]	6 (^{13}C)	6 (^{13}C)
Field [kHz]	10.2 (^{15}N)	10.6 (^{15}N)
Shape	Tangent ^{13}C	Tangent ^{13}C
^{13}C carrier [ppm]	58.6	58.6
time [ms]	7	10
t_1 increments	400 (^{15}N)	400 (^{15}N)
sw (t_1) [kHz]	16.2	16.2
Acq. time (t_1) [ms]	12.3	12.3
t_2 increments	2304 (^{13}C)	2304 (^{13}C)
sw (t_2) [kHz]	100	100
Acq. time (t_2) [ms]	11.5	11.5
^{1}H decoupling	SPINAL64	SPINAL64
Field [kHz]	90	90
Interscan delay d1 [s]	2.6	2.6
Number of scans	32	32
Measurement time	9 h 20	9 h 20

b. Proton detection using 1.3mm rotors

Freshly prepared $^{2}H^{13}C$-^{15}N-Cp149 reassembled capsids in 50 mM HEPES, 5 mM DTT (pH 7.5) and in 20 mM $NaPO_4$, 50 mM NaCl, 1x protease inhibitor (pH 6) were incubated with 1 equivalent of preS for 1 hour on ice. They were then concentrated using Amicon Ultra centrifugal filter units (Merck, 50 kDa cut-off) to about 20 mg/ml in 1 mL and sedimented into 1.3 mm zirconium rotors by ultracentrifugation (200,000 g, 14 h, 4 °C) using a home-made filling tool. Rotors were immediately closed after the addition of 1 µL of saturated DSS solution for chemical-shift referencing.

2D hNH spectra were recorded using a 1.3 mm triple-resonance (^{1}H, ^{13}C, ^{15}N) wide-bore probe head at a static magnetic field of 18.8 T corresponding to 800 MHz proton resonance frequency (Bruker Avance II). All spectra were referenced to DSS and recorded at 60 kHz MAS frequency and at a sample temperature of 23 °C according to the resonance frequency of the supernatant water. Assignments were transferred from (Lecoq et al., 2019) (BMRB number 27845). NMR acquisition parameters are detailed in **Table 13**. All spectra were processed using TopSpin 4.0.3 (Bruker Biospin) and analyzed with the CcpNmr Analysis package, Version 2.4.2.

Table 13. *NMR experimental details of solid-state NMR measurements of 1.3 mm rotors. sw stands for spectral width. Experiments were recorded on an 800 MHz wide-bore spectrometer at a 60 kHz MAS frequency and at a sample temperature estimated at 23 °C. Decoupling was set at 10 kHz for the 3 channels, and ^{1}H MISSISSIPI water suppression at 20 kHz for 100 ms.*

Sample	**Cp149 (pH 6) (form B)**		
Experiment	hNH	hCANH	hCONH
Transfer 1	HN-CP	HCa-CP	HCO-CP
Field [kHz]	40.8 (^{1}H) 18.8 (^{15}N)	13.8 (^{1}H) 32.1 (^{13}C)	14.9 (^{1}H) 32.1 (^{13}C)
Shape	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H
^{13}C carrier [ppm]		56	56
time [ms]	1.0	3	3.5
Transfer 2	NH-CP	CN-CP	CN-CP
Field [kHz]	43.1 (^{1}H)		
Field [kHz]		29 (^{13}C)	31.3 (^{13}C)
Field [kHz]	18.8 (^{15}N)	16.7 (^{15}N)	16.7 (^{15}N)
Shape	Tangent ^{1}H	Tangent ^{13}C	Tangent ^{13}C
^{13}C carrier [ppm]		56	178
time [ms]	0.5	10	12

Transfer 3		NH-CP	NH-CP
Field [kHz]		42.5 (^{1}H)	37.8 (^{1}H)
Field [kHz]		18.8 (^{15}N)	18.8 (^{15}N)
Shape		Tangent ^{1}H	Tangent ^{1}H
time [ms]		0.5	0.5
t_1 increments	320 (^{15}N)	48 (^{13}C)	40 (^{13}C)
sw (t_1) [ppm]	80	22	15
Acq. time (t_1) [ms]	24.7	5.4	6.6
t_2 increments	4096 (^{1}H)	32 (^{15}N)	32 (^{15}N)
sw (t_2) [ppm]	97.7	30	30
Acq. time (t_2) [ms]	26.2	6.6	6.6
t_3 increments		4096 (^{1}H)	4096 (^{1}H)
sw (t_3) [ppm]		59	59
Acq. time (t_3) [ms]		43.4	43.4
Interscan delay d1 [s]	1.5	1.5	1.5
Number of scans	64	80	144
Measurement time	9 h 17	2 d 9 h	3 d 14 h

Sample	**Cp149+preS (pH 6) (form B)**		
Experiment	hNH	hCANH	hCONH
Transfer 1	HN-CP	HC-CP	HC-CP
Field [kHz]	38.2 (^{1}H) 18.8 (^{15}N)	13.8 (^{1}H) 32.1 (^{13}C)	14.9 (^{1}H) 32.1 (^{13}C)
Shape	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H
^{13}C carrier [ppm]		56	56
time [ms]	1.0	3	3.5
Transfer 2	NH-CP	CN-CP	CN-CP
Field [kHz]	39.1 (^{1}H)		
Field [kHz]		30.5 (^{13}C)	31.3 (^{13}C)
Field [kHz]	18.8 (^{15}N)	16.7 (^{15}N)	16.7 (^{15}N)
Shape	Tangent ^{1}H	Tangent ^{13}C	Tangent ^{13}C
^{13}C carrier [ppm]		56	178

time [ms]	0.5	12	10
Transfer 3		NH-CP	NH-CP
Field [kHz]		42.5 (^{1}H)	42.5 (^{1}H)
Field [kHz]		18.8 (^{15}N)	18.8 (^{15}N)
Shape		Tangent ^{1}H	Tangent ^{1}H
time [ms]		0.5	0.5
t_1 increments	320 (^{15}N)	54 (^{13}C)	40 (^{13}C)
sw (t_1) [ppm]	80	25	15
Acq. time (t_1) [ms]	24.7	5.3	6.6
t_2 increments	4096 (^{1}H)	32 (^{15}N)	32 (^{15}N)
sw (t_2) [ppm]	97.7	30	30
Acq. time (t_2) [ms]	26.2	6.6	6.6
t_3 increments		4096 (^{1}H)	4096 (^{1}H)
sw (t_3) [ppm]		59	59
Acq. time (t_3) [ms]		43.4	43.4
Interscan delay d1 [s]	1.5	1.5	1.5
Number of scans	64	80	80
Measurement time	9 h 16	2 d 16 h	2 d

Sample	**Cp149 (pH 6) (form A)**	**Cp149+preS (pH 6) (form A)**
Experiment	hNH	hNH
Transfer 1	HN-CP	HN-CP
Field [kHz]	40.8 (^{1}H) 18.8 (^{15}N)	39.9 (^{1}H) 18.8 (^{15}N)
Shape	Tangent ^{1}H	Tangent ^{1}H
^{13}C carrier [ppm]		
time [ms]	1.0	1.0
Transfer 2	NH-CP	NH-CP
Field [kHz]	43.1 (^{1}H)	42.1 (^{1}H)
Field [kHz]	18.8 (^{15}N)	18.8 (^{15}N)
Shape	Tangent ^{1}H	Tangent ^{1}H

time [ms]	0.5	0.5
t_1 increments	320 (^{15}N)	320 (^{15}N)
sw (t_1) [ppm]	80	80
Acq. time (t_1) [ms]	24.7	24.7
t_2 increments	4096 (^{1}H)	4096 (^{1}H)
sw (t_2) [ppm]	97.7	97.7
Acq. time (t_2) [ms]	26.2	26.2
Interscan delay d1 [s]	1.5	1.5
Number of scans	96	144
Measurement time	13 h 55	20 h 53

6. Isothermal titration calorimetry

The ITC experiments were set up in the same way as previously described in the paragraph “Isothermal titration calorimetry”, on page 96.

a. Cp149 and Oct peptides

Samples of freshly prepared Cp149 reassembled capsid were dialyzed into 50 mM HEPES buffer at pH 7.5 and then quantified by absorbance measurements. The Oct peptides are dissolved in HEPES buffer to prepare stock solutions at 1000 µM, then the final solutions were calculated in order to respect the typical 10 times ratio between the sample concentration and the ligand concentration. For these ITC experiments, after a temperature equilibration delay, the measurement was started by a 1 µL injection of ligand (injection duration 2 s), followed by 24 injections of 1.6 µL (injection duration 3.2 s) in 2 min intervals.

b. preS and Cp149

Samples of freshly prepared preS, Cp149 dimer and capsids (forms A and B) were dialyzed into 20 mM $NaPO_4$, 50 mM NaCl, 1x protease inhibitor and then quantified by absorbance measurements. For this specific case, as preS cannot be concentrated more than 50 µM, this protein was introduced in the sample cell, and the syringe was filled with Cp149 samples approximately 10 times more concentrated than preS. For these ITC experiments, after a temperature equilibration delay, the measurement was started by a 1 µL injection of ligand (injection duration 2 s), followed by 19 injections of 1.6 µL (injection duration 3.2 s) in 2 min intervals.

7. Circular Dichroism

Far UV circular dichroism (CD) spectra were recorded on a Chirascan spectrometer (Applied Photophysics, Leatherhead, United Kingdom) calibrated with 1S-(+)-10-

camphorsulfonic acid. Measurements were carried out at room temperature in a 0.1 cm path length quartz cuvette. Spectra were measured in a 180–260 nm wavelength range with an increment of 0.2 nm, band pass of 0.5 nm, and integration time of 1 s. Spectra were processed, baseline-corrected, and smoothed with the Chirascan software. PreS was in 20 mM phosphate buffer, 50 mM NaCl (pH 6) at a concentration of 11.8 µM.

8. Cell-free co-expression tests

Co-expression tests have been performed in the WG-CFPS system with homemade wheat germ extract using the bilayer method as described in (Fogeron et al., 2017). Transcription was performed for 6 hours at 37 °C in Transcription buffer containing RNAsin (1U/µL), SP6 polymerase (1U/µL) (CellFree Sciences), 10 mM rNTP mix (Promega) and 0.1 µg/µL of plasmid DNA. Translation was performed in 96-well plates for 16 hours at 22 °C without shaking. The translation reaction consisted of 20 µL of translation mix in the bottom layer, overlaid by 200 µL of feeding buffer (top layer) in one well. The translation mix contained ½ volume of transcription reaction, ½ volume of wheat germ extract, 40 ng/mL creatine kinase, 6 mM amino acid mixture (average concentration of each amino acid 0.3 mM). For co-expression, mRNAs corresponding to the two proteins of interest were simply mixed in equal volumes. The feeding buffer contained 1x SUB-AMIX buffer (CellFree Sciences Japan) supplemented with 6 mM amino acid mixture. In the case of the membrane proteins HBs S and HBs L, 0.05% (w/v) Brij-58 was added both to the translation mix and the feeding buffer to ensure protein synthesis in a detergent-solubilized form.

After protein synthesis, benzonase was added to the total cell-free sample (CFS) and the mixture was incubated for 30 minutes at RT on the wheel. After a centrifugation step (20,000 g, 30 min, 4 °C), the supernatant was separated from the pellet. The pellet was centrifuged a second time (20,000 g, 5 min, 4 °C) on order to remove residual supernatant, and resuspended in 1X Laemmli for SDS-PAGE and Western-blot analysis. The supernatant was then incubated with magnetic beads coated with Strep-Tactin (IBA Lifesciences) to capture the envelope proteins or their subdomains through the Strep-tag fused at their N- and/or C-terminus. After washing in the presence of 0.1% DDM to limit potential aspecific binding, the magnetic beads were resuspended in 1X Laemmli. This sample is called SN-Beads.

For SDS-PAGE analysis, the samples were heated at 95 °C for 5 min. Then, these gels were transferred (iBLOT) onto Western-blot membranes. They were incubated in milk for 1h30 at RT, then with a primary antibody for several hours at RT. The membranes were rinsed with PBS-T buffer and incubated with the secondary antibody for 2h at RT. A final washing step with PBS-T buffer was necessary before spraying the membranes with reagent to reveal the bands, and observe them with a chemiluminescence fusion camera. For each assay with the preS1 and preS domains, we prepared two membranes, one incubated with a mouse anti-STREP primary antibody and a HRP-coupled anti-mouse secondary antibody in order to detect envelope proteins with strep-tag, and the second membrane with a rabbit anti-CORE primary antibody (kindly provided by Michael Nassal), and a HRP-coupled anti- rabbit

secondary antibody in order to detect core proteins. For assays with the HBs S and L proteins, the same membrane was incubated successively with the anti-Strep and anti-core antibodies.

VI. Contributions

I prepared Cp samples, performed co-sedimentation tests, measured CD, ITC, and NMR spectra, together with Lauriane Lecoq; Eric Diesis performed the Oct peptides synthesis Laura Cole prepared the preS samples; Marie-Laure Fogeron performed the co-expression tests using WG-CFPS; Mathilde Briday and Lauriane Lecoq analyzed data; Anja Böckmann, Lauriane Lecoq and Marie-Laure Fogeron designed the study, together with Michael Nassal and Beat H. Meier.

REFERENCES

Acharya, A., Xu, X.-J., Husain-Ponnampalam, R. D., Hoffmann-Benning, S., & Kuo, M.-H. (2005). Production of constitutively acetylated recombinant p53 from yeast and Escherichia coli by tethered catalysis. *Protein Expression and Purification*, *41*(2), 417–425. https://doi.org/10.1016/j.pep.2005.01.015

Alexander, C. G., Jürgens, M. C., Shepherd, D. A., Freund, S. M. V., Ashcroft, A. E., & Ferguson, N. (2013). Thermodynamic origins of protein folding, allostery, and capsid formation in the human hepatitis B virus core protein. *Proceedings of the National Academy of Sciences*, *110*(30). https://doi.org/10.1073/pnas.1308846110

Alia, A., & Ganapathy, S. (2009). Magic angle spinning (MAS) NMR: a new tool to study the spatial and electronic structure of photosynthetic complexes. *Photosynth Res*, 11.

Andreas, L. B., Jaudzems, K., Stanek, J., Lalli, D., Bertarello, A., Le Marchand, T., Cala-De Paepe, D., Kotelovica, S., Akopjana, I., Knott, B., Wegner, S., Engelke, F., Lesage, A., Emsley, L., Tars, K., Herrmann, T., & Pintacuda, G. (2016). Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proceedings of the National Academy of Sciences*, *113*(33), 9187–9192. https://doi.org/10.1073/pnas.1602248113

Atlas, R. M. (2010). *Handbook of microbiological media* (4th ed). ASM Press ; CRC Press/Taylor & Francis.

Azatian, S. B., Kaur, N., & Latham, M. P. (2019). Increasing the buffering capacity of minimal media leads to higher protein yield. *Journal of Biomolecular NMR*, *73*(1–2), 11–17. https://doi.org/10.1007/s10858-018-00222-4

Bardens, A., Döring, T., Stieler, J., & Prange, R. (2011). Alix regulates egress of hepatitis B virus naked capsid particles in an ESCRT-independent manner: Alix mediates HBV capsid release. *Cellular Microbiology*, *13*(4), 602–619. https://doi.org/10.1111/j.1462-5822.2010.01557.x

Bartenschlager, R., & Schaller, H. (1992). Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *The EMBO Journal*, *11*(9), 3413–3420. https://doi.org/10.1002/j.1460-2075.1992.tb05420.x

Basagoudanavar, S. H., Perlman, D. H., & Hu, J. (2007). Regulation of Hepadnavirus Reverse Transcription by Dynamic Nucleocapsid Phosphorylation. *Journal of Virology*, *81*(4), 1641–1649. https://doi.org/10.1128/JVI.01671-06

Becker, W., Bhattiprolu, K. C., Gubensäk, N., & Zangger, K. (2018). *Investigating Protein–Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy*. 13.

Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., & Kompala, D. S. (1990). Plasmid-encoded protein: The principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnology and Bioengineering*, *35*(7), 668–681. https://doi.org/10.1002/bit.260350704

Bertoletti, A., & Bert, N. L. (2018). Immunotherapy for Chronic Hepatitis B Virus Infection. *Gut and Liver*, *12*(5), 497–507. https://doi.org/10.5009/gnl17233

Birnbaum, F., & Nassal, M. (1990). Hepatitis B virus nucleocapsid assembly: Primary structure requirements in the core protein. *Journal of Virology*, *64*(7), 3319–3330. https://doi.org/10.1128/jvi.64.7.3319-3330.1990

Biswas, T., & Tsodikov, O. V. (2010). An easy-to-use tool for planning and modeling a calorimetric titration. *Analytical Biochemistry*, *406*(1), 91–93. https://doi.org/10.1016/j.ab.2010.06.050

Block, T. M., Chang, K.-M., & Guo, J.-T. (2021). Prospects for the Global Elimination of Hepatitis B. *Annual Review of Virology*, *8*(1), 437–458. https://doi.org/10.1146/annurev-virology-091919-062728

Blondot, M.-L., Bruss, V., & Kann, M. (2016). Intracellular transport and egress of hepatitis B virus. *Journal of Hepatology*, *64*(1), S49–S59. https://doi.org/10.1016/j.jhep.2016.02.008

Blumberg, B. S., Alter, H. J., & Visnich, S. (1965). A “New” Antigen in Leukemia Sera. *Journal of the American Medical Association*, *191*(7), 101–106.

Bock, C. T., Schwinn, S., Locarnini, S., Fyfe, J., Manns, M. P., Trautwein, C., & Zentgraf, H. (2001). Structural organization of the hepatitis B virus minichromosome. *Journal of Molecular Biology*, *307*(1), 183–196. https://doi.org/10.1006/jmbi.2000.4481

Böckmann, A., Ernst, M., & Meier, B. H. (2015). Spinning proteins, the faster, the better? *Journal of Magnetic Resonance*, *253*, 71–79. https://doi.org/10.1016/j.jmr.2015.01.012

Bockmann, A., Gardiennet, C., Verel, R., Hunkeler, A., Loquet, A., Pintacuda, G., Emsley, L., Meier, B. H., & Lesage, A. (2009). Characterization of different water pools in solid-state NMR protein samples. *J Biomol NMR*, 9.

Böttcher, B. (2021). Hepatitis B Core Protein Capsids. In J. R. Harris & J. Marles-Wright (Eds.), *Macromolecular Protein Complexes III: Structure and Function* (Vol. 96, pp. 451–470). Springer International Publishing. https://doi.org/10.1007/978-3-030-58971-4_14

Böttcher, B., & Nassal, M. (2018). Structure of Mutant Hepatitis B Core Protein Capsids with Premature Secretion Phenotype. *Journal of Molecular Biology*, *430*(24), 4941–4954. https://doi.org/10.1016/j.jmb.2018.10.018

Böttcher, B., Tsuji, N., Takahashi, H., Dyson, M. R., Zhao, S., Crowther, R. A., & Murray, K. (1998). Peptides that block hepatitis B virus assembly: Analysis by cryomicroscopy, mutagenesis and transfection. *The EMBO Journal*, *17*(23), 6839–6845. https://doi.org/10.1093/emboj/17.23.6839

Böttcher, B., Wynne, S. A., & Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron microscopy. *Nature*, *386*, 88–91.

Bouchard, M. J., & Schneider, R. J. (2004). The Enigmatic X Gene of Hepatitis B Virus. *Journal of Virology*, *78*(23), 12725–12734. https://doi.org/10.1128/JVI.78.23.12725-12734.2004

Bourne, C. R., Finn, M. G., & Zlotnick, A. (2006). Global Structural Changes in Hepatitis B Virus Capsids Induced by the Assembly Effector HAP1. *Journal of Virology*, *80*(22), 11055–11061. https://doi.org/10.1128/JVI.00933-06

Briday, M., Hallé, F., Lecoq, L., Radix, S., Martin, J., Montserret, R., Dujardin, M., Fogeron, M.-L., Nassal, M., Meier, B. H., Lomberget, T., & Böckmann, A. (2022). Pharmacomodulation of a ligand targeting the HBV capsid hydrophobic pocket. *Chemical Science*, *13*(30), 8840–8847. https://doi.org/10.1039/D2SC02420A

Bruss, V. (1997). A short linear sequence in the pre-S domain of the large hepatitis B virus envelope protein required for virion formation. *Journal of Virology*, *71*(12), 9350–9357. https://doi.org/10.1128/jvi.71.12.9350-9357.1997

Bruss, V. (2004). Envelopment of the hepatitis B virus nucleocapsid. *Virus Research*, *106*(2), 199–209. https://doi.org/10.1016/j.virusres.2004.08.016

Bruss, V. (2007). *Hepatitis B virus morphogenesis*. 1–9.

Bruss, V., & Ganem, D. (1991). The role of envelope proteins in hepatitis B virus assembly. *Proceedings of the National Academy of Sciences*, *88*(3), 1059–1063. https://doi.org/10.1073/pnas.88.3.1059

Buhlig, T. S., Bowersox, A. F., Braun, D. L., Owsley, D. N., James, K. D., Aranda, A. J., Kendrick, C. D., Skalka, N. A., & Clark, D. N. (2020). Molecular, Evolutionary, and Structural Analysis of the Terminal Protein Domain of Hepatitis B Virus Polymerase, a Potential Drug Target. *Viruses*, *12*(5), 570. https://doi.org/10.3390/v12050570

Cai, M., Huang, Y., Sakaguchi, K., Clore, G. M., Gronenborn, A. M., & Craigie, R. (1997). *An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli*. *11*, 97–102.

Cai, M., Huang, Y., Yang, R., Craigie, R., & Clore, G. M. (2016). A simple and robust protocol for high-yield expression of perdeuterated proteins in Escherichia coli grown in shaker flasks. *Journal of Biomolecular NMR*, *66*(2), 85–91. https://doi.org/10.1007/s10858-016-0052-y

Campbell, I. D. (2013). The evolution of protein NMR. *Biomedical Spectroscopy and Imaging*, *2*(4), 245–264. https://doi.org/10.3233/BSI-130055

Cao, R., Liu, X., Liu, Y., Zhai, X., Cao, T., Wang, A., & Qiu, J. (2021). Applications of nuclear magnetic resonance spectroscopy to the evaluation of complex food constituents. *Food Chemistry*, *342*, 128258. https://doi.org/10.1016/j.foodchem.2020.128258

Carlomagno, T. (2005). Ligand-Target Interactions: What Can We Learn from NMR? *Annual Review of Biophysics and Biomolecular Structure*, *34*(1), 245–266. https://doi.org/10.1146/annurev.biophys.34.040204.144419

Ceres, P., Stray, S. J., & Zlotnick, A. (2004). Hepatitis B Virus Capsid Assembly Is Enhanced by Naturally Occurring Mutation F97L. *Journal of Virology*, *78*(17), 9538–9543. https://doi.org/10.1128/JVI.78.17.9538-9543.2004

Chai, N., Chang, H. E., Nicolas, E., Han, Z., Jarnik, M., & Taylor, J. (2008). Properties of Subviral Particles of Hepatitis B Virus. *Journal of Virology*, *82*(16), 7812–7817. https://doi.org/10.1128/JVI.00561-08

Chang, S.-Y., Ko, T.-P., Liang, P.-H., & Wang, A. H.-J. (2003). Catalytic Mechanism Revealed by the Crystal Structure of Undecaprenyl Pyrophosphate Synthase in Complex with Sulfate, Magnesium, and Triton. *Journal of Biological Chemistry*, *278*(31), 29298–29307. https://doi.org/10.1074/jbc.M302687200

Chen, M. T., Billaud, J.-N., Sällberg, M., Guidotti, L. G., Chisari, F. V., Jones, J., Hughes, J., & Milich, D. R. (2004). A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. *Proceedings of the National Academy of Sciences*, *101*(41), 14913–14918. https://doi.org/10.1073/pnas.0406282101

Chen, R. (2012). Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology Advances*, *30*(5), 1102–1107. https://doi.org/10.1016/j.biotechadv.2011.09.013

Chi, S.-W., Kim, D.-H., Lee, S.-H., Chang, I., & Han, K.-H. (2007). Pre-structured motifs in the natively unstructured preS1 surface antigen of hepatitis B virus. *Protein Science*, *16*(10), 2108–2117. https://doi.org/10.1110/ps.072983507

Chou, S.-F., Tsai, M.-L., Huang, J.-Y., Chang, Y.-S., & Shih, C. (2015). The Dual Role of an ESCRT-0 Component HGS in HBV Transcription and Naked Capsid Secretion. *PLOS Pathogens*, *11*(10), e1005123. https://doi.org/10.1371/journal.ppat.1005123

Chu, T.-H., Liou, A.-T., Su, P.-Y., Wu, H.-N., & Shih, C. (2014). Nucleic Acid Chaperone Activity Associated with the Arginine-Rich Domain of Human Hepatitis B Virus Core Protein. *Journal of Virology*, *88*(5), 2530–2543. https://doi.org/10.1128/JVI.03235-13

Chua, P. K., Tang, F.-M., Huang, J.-Y., Suen, C.-S., & Shih, C. (2010). Testing the Balanced Electrostatic Interaction Hypothesis of Hepatitis B Virus DNA Synthesis by Using an *In Vivo* Charge Rebalance Approach. *Journal of Virology*, *84*(5), 2340–2351. https://doi.org/10.1128/JVI.01666-09

Ciulli, A. (2013). Biophysical Screening for the Discovery of Small-Molecule Ligands. In M. A. Williams & T. Daviter (Eds.), *Protein-Ligand Interactions* (Vol. 1008, pp. 357–388). Humana Press. https://doi.org/10.1007/978-1-62703-398-5_13

Clark, D. N., & Hu, J. (2015). Unveiling the roles of HBV polymerase for new antiviral strategies. *Future Virology*, *10*(3), 283–295. https://doi.org/10.2217/fvl.14.113

Clippinger, A. J., Gearhart, T. L., & Bouchard, M. J. (2009). Hepatitis B Virus X Protein Modulates Apoptosis in Primary Rat Hepatocytes by Regulating both NF-κB and the Mitochondrial Permeability Transition Pore. *Journal of Virology*, *83*(10), 4718–4731. https://doi.org/10.1128/JVI.02590-08

Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J., & Steven, A. C. (1997). Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature*, *386*, 91–94.

Cooper, A., Paran, N., & Shaul, Y. (2003). The earliest steps in hepatitis B virus infection. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1614*(1), 89–96. https://doi.org/10.1016/S0005-2736(03)00166-4

Crowther, R. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell*, *77*(6), 943–950. https://doi.org/10.1016/0092-8674(94)90142-2

Dane, D. S., & Cameron, C. H. (1970). Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *The Lancet*, 695–698.

David, G., Fogeron, M.-L., Montserret, R., Lecoq, L., Page, A., Delolme, F., Nassal, M., & Böckmann, A. (2019). Phosphorylation and Alternative Translation on Wheat Germ Cell-Free Protein Synthesis of the DHBV Large Envelope Protein. *Frontiers in Molecular Biosciences*, *6*, 138. https://doi.org/10.3389/fmolb.2019.00138

de Rocquigny, H., Rat, V., Pastor, F., Darlix, J. L., Hourioux, C., & Roingeard, P. (2020). Phosphorylation of the Arginine-Rich C-Terminal Domains of the Hepatitis B Virus (HBV) Core Protein as a Fine Regulator of the Interaction between HBc and Nucleic Acid. *Viruses*, *12*(7), 738. https://doi.org/10.3390/v12070738

Decorsière, A., Mueller, H., van Breugel, P. C., Abdul, F., Gerossier, L., Beran, R. K., Livingston, C. M., Niu, C., Fletcher, S. P., Hantz, O., & Strubin, M. (2016). Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature*, *531*(7594), 386–389. https://doi.org/10.1038/nature17170

Diab, A., Foca, A., Zoulim, F., Durantel, D., & Andrisani, O. (2018). The diverse functions of the hepatitis B core/capsid protein (HBc) in the viral life cycle: Implications for the development of HBc-targeting antivirals. *Antiviral Research*, *149*, 211–220. https://doi.org/10.1016/j.antiviral.2017.11.015

DiMattia, M. A., Watts, N. R., Stahl, S. J., Grimes, J. M., Steven, A. C., Stuart, D. I., & Wingfield, P. T. (2013). Antigenic Switching of Hepatitis B Virus by Alternative

Dimerization of the Capsid Protein. *Structure*, *21*(1), 133–142. https://doi.org/10.1016/j.str.2012.10.017

Dryden, K. A., Wieland, S. F., Whitten-Bauer, C., Gerin, J. L., Chisari, F. V., & Yeager, M. (2006). Native Hepatitis B Virions and Capsids Visualized by Electron Cryomicroscopy. *Molecular Cell*, *22*(6), 843–850. https://doi.org/10.1016/j.molcel.2006.04.025

Du, J., Liang, X., Liu, Y., Qu, Z., Gao, L., Han, L., Liu, S., Cui, M., Shi, Y., Zhang, Z., Yu, L., Cao, L., Ma, C., Zhang, L., Chen, Y., & Sun, W. (2009). Hepatitis B virus core protein inhibits TRAIL-induced apoptosis of hepatocytes by blocking DR5 expression. *Cell Death & Differentiation*, *16*(2), 219–229. https://doi.org/10.1038/cdd.2008.144

Dyson, M. R., & Murray, K. (1995). Selection of peptide inhibitors of interactions involved in complex protein assemblies: Association of the core and surface antigens of hepatitis B virus. *Proceedings of the National Academy of Sciences*, *92*(6), 2194–2198. https://doi.org/10.1073/pnas.92.6.2194

El-Baky, N. A., & Redwan, E. M. (2015). Therapeutic Alpha-Interferons Protein: Structure, Production, and Biosimilar. *Preparative Biochemistry and Biotechnology*, *45*(2), 109–127. https://doi.org/10.1080/10826068.2014.907175

Fanning, G. C. (n.d.). Therapeutic strategies for hepatitis B virus infection: Towards a cure. *Drug Discovery*, 18.

Farcet, J., Kindermann, J., Karbiener, M., Scheinecker, R., Kostner, O., & Kreil, T. R. (2021). Synthesis of “Nereid,” a new phenol-free detergent to replace Triton X-100 in virus inactivation. *Journal of Medical Virology*, *93*(6), 3880–3889. https://doi.org/10.1002/jmv.26708

Feng, H., Chen, P., Zhao, F., Nassal, M., & Hu, K. (2013). Evidence for Multiple Distinct Interactions between Hepatitis B Virus P Protein and Its Cognate RNA Encapsidation Signal during Initiation of Reverse Transcription. *PLoS ONE*, *8*(8), e72798. https://doi.org/10.1371/journal.pone.0072798

Fogeron, M.-L., Badillo, A., Penin, F., & Böckmann, A. (2017). Wheat Germ Cell-Free Overexpression for the Production of Membrane Proteins. In J.-J. Lacapere (Ed.), *Membrane Protein Structure and Function Characterization* (Vol. 1635, pp. 91–108). Springer New York. https://doi.org/10.1007/978-1-4939-7151-0_5

Fogeron, M.-L., Lecoq, L., Cole, L., Harbers, M., & Böckmann, A. (2021). Easy Synthesis of Complex Biomolecular Assemblies: Wheat Germ Cell-Free Protein Expression in Structural Biology. *Frontiers in Molecular Biosciences*, *8*, 639587. https://doi.org/10.3389/fmolb.2021.639587

Fogeron, M.-L., Lecoq, L., Cole, L., Montserret, R., David, G., Page, A., Delolme, F., Nassal, M., & Böckmann, A. (2022). Phosphorylation of the Hepatitis B Virus Large Envelope Protein. *Frontiers in Molecular Biosciences*, *8*, 821755. https://doi.org/10.3389/fmolb.2021.821755

Freund, S. M. V., Johnson, C. M., Jaulent, A. M., & Ferguson, N. (2008). Moving towards High-Resolution Descriptions of the Molecular Interactions and Structural Rearrangements of the Human Hepatitis B Core Protein. *Journal of Molecular Biology*, *384*(5), 1301–1313. https://doi.org/10.1016/j.jmb.2008.10.020

Fricke, P., Chevelkov, V., Zinke, M., Giller, K., Becker, S., & Lange, A. (2017). Backbone assignment of perdeuterated proteins by solid-state NMR using proton detection and ultrafast magic-angle spinning. *Nature Protocols*, *12*(4), 764–782. https://doi.org/10.1038/nprot.2016.190

Fung, S., Choi, H. S. J., Gehring, A., & Janssen, H. L. A. (2022). Getting to HBV cure: The promising paths forward. *Hepatology*, *76*(1), 233–250.

Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P., & Charnay, P. (1979). Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in E. coli. *Nature*, *281*(5733), 646–650. https://doi.org/10.1038/281646a0

Gazina, E. V., Fielding, J. E., Lin, B., & Anderson, D. A. (2000). Core Protein Phosphorylation Modulates Pregenomic RNA Encapsidation to Different Extents in

Human and Duck Hepatitis B Viruses. *Journal of Virology*, *74*(10), 4721–4728. https://doi.org/10.1128/jvi.74.10.4721-4728.2000

Gerelsaikhan, T., Tavis, J. E., & Bruss, V. (1996). Hepatitis B virus nucleocapsid envelopment does not occur without genomic DNA synthesis. *Journal of Virology*, *70*(7), 4269–4274. https://doi.org/10.1128/jvi.70.7.4269-4274.1996

Gerothanassis, I. P., Troganis, A., Exarchou, V., & Barbarossou, K. (2002). NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY: BASIC PRINCIPLES AND PHENOMENA, AND THEIR APPLICATIONS TO CHEMISTRY, BIOLOGY AND MEDICINE. *CHEMISTRY EDUCATION: RESEARCH AND PRACTICE IN EUROPE*, *3*(2), 229–252.

Gibbs, E. B., & Showalter, S. A. (2015). Quantitative Biophysical Characterization of Intrinsically Disordered Proteins. *Biochemistry*, *54*(6), 1314–1326. https://doi.org/10.1021/bi501460a

Gripon, P., Le Seyec, J., Rumin, S., & Guguen-Guillouzo, C. (1995). Myristylation of the Hepatitis B Virus Large Surface Protein Is Essential for Viral Infectivity. *Virology*, *213*(2), 292–299. https://doi.org/10.1006/viro.1995.0002

Guo, Y., Kang, W., Lei, X., Li, Y., Xiang, A., Liu, Y., Zhao, J., Zhang, J., & Yan, Z. (2012). Hepatitis B viral core protein disrupts human host gene expression by binding to promoter regions. *BMC Genomics*, *13*(1), 563. https://doi.org/10.1186/1471-2164-13-563

Harbers, M. (2014). Wheat germ systems for cell-free protein expression. *FEBS Letters*, *588*(17), 2762–2773. https://doi.org/10.1016/j.febslet.2014.05.061

Heger-Stevic, J., Zimmermann, P., Lecoq, L., Böttcher, B., & Nassal, M. (2018). Hepatitis B virus core protein phosphorylation: Identification of the SRPK1 target sites and impact of their occupancy on RNA binding and capsid structure. *PLOS Pathogens*, *14*(12), e1007488. https://doi.org/10.1371/journal.ppat.1007488

Ho, J. K.-T., Jeevan-Raj, B., & Netter, H.-J. (2020). Hepatitis B Virus (HBV) Subviral Particles as Protective Vaccines and Vaccine Platforms. *Viruses*, *12*(2), 126. https://doi.org/10.3390/v12020126

Hoffmann, J., Boehm, C., Himmelsbach, K., Donnerhak, C., Roettger, H., Weiss, T. S., Ploen, D., & Hildt, E. (2013). Identification of α-taxilin as an essential factor for the life cycle of hepatitis B virus. *Journal of Hepatology*, *59*(5), 934–941. https://doi.org/10.1016/j.jhep.2013.06.020

Hong, M. (1999). *Resonance assignment of 13C/15N labeled solid proteins by two- and three-dimensional magic-angle-spinning NMR*. 14.

Hong, M. (2010). Isotopic Labeling for NMR Spectroscopy of Biological Solids. *ISOTEC® Stable Isotopes: Products for Solid State NMR*, 6–11.

Hosek, T., Bougault, C. M., Lavergne, J.-P., Martinez, D., Ayala, I., Fenel, D., Restelli, M., Morlot, C., Habenstein, B., Grangeasse, C., & Simorre, J.-P. (2020). Structural features of the interaction of MapZ with FtsZ and membranes in Streptococcus pneumoniae. *Scientific Reports*, *10*(4051), 1–13. https://doi.org/10.1038/s41598-020-61036-9

Hu, J., & Liu, K. (2017). Complete and Incomplete Hepatitis B Virus Particles: Formation, Function, and Application. *Viruses*, *9*(3), 56. https://doi.org/10.3390/v9030056

Huovila, A. P., Eder, A. M., & Fuller, S. D. (1992). Hepatitis B surface antigen assembles in a post-ER, pre-Golgi compartment. *Journal of Cell Biology*, *118*(6), 1305–1320. https://doi.org/10.1083/jcb.118.6.1305

Ichai, P., & Samuel, D. (2019). Management of Fulminant Hepatitis B. *Current Infectious Disease Reports*, *21*(7), 25. https://doi.org/10.1007/s11908-019-0682-9

Inoue, J., Sato, K., Ninomiya, M., & Masamune, A. (2021). Envelope Proteins of Hepatitis B Virus: Molecular Biology and Involvement in Carcinogenesis. *Viruses*, *13*(6), 1124. https://doi.org/10.3390/v13061124

Iwamoto, M., Saso, W., Nishioka, K., Ohashi, H., Sugiyama, R., Ryo, A., Ohki, M., Yun, J.-H., Park, S.-Y., Ohshima, T., Suzuki, R., Aizaki, H., Muramatsu, M., Matano, T., Iwami, S., Sureau, C., Wakita, T., & Watashi, K. (2020). The machinery for endocytosis of epidermal growth factor receptor coordinates the transport of incoming hepatitis B virus to the endosomal network. *Journal of Biological Chemistry*, *295*(3), 800–807. https://doi.org/10.1016/S0021-9258(17)49936-4

Jiang, B., Himmelsbach, K., Ren, H., Boller, K., & Hildt, E. (2016). Subviral Hepatitis B Virus Filaments, like Infectious Viral Particles, Are Released via Multivesicular Bodies. *Journal of Virology*, *90*(7), 3330–3341. https://doi.org/10.1128/JVI.03109-15

Jürgens, M. C., Vörös, J., Rautureau, G. J. P., Shepherd, D. A., Pye, V. E., Muldoon, J., Johnson, C. M., Ashcroft, A. E., Freund, S. M. V., & Ferguson, N. (2013). The hepatitis B virus preS1 domain hijacks host trafficking proteins by motif mimicry. *Nature Chemical Biology*, *9*(9), 540–547. https://doi.org/10.1038/nchembio.1294

Kann, M., Sodeik, B., Vlachou, A., Gerlich, W. H., & Helenius, A. (1999). Phosphorylation-dependent Binding of Hepatitis B Virus Core Particles to the Nuclear Pore Complex. *Journal of Cell Biology*, *145*(1), 45–55. https://doi.org/10.1083/jcb.145.1.45

Kastritis, P. L., & Bonvin, A. M. J. J. (2013). On the binding affinity of macromolecular interactions: Daring to ask why proteins interact. *Journal of The Royal Society Interface*, *10*(79), 20120835. https://doi.org/10.1098/rsif.2012.0835

Keasler, V. V., Hodgson, A. J., Madden, C. R., & Slagle, B. L. (2007). Enhancement of Hepatitis B Virus Replication by the Regulatory X Protein In Vitro and In Vivo. *Journal of Virology*, *81*(6), 2656–2662. https://doi.org/10.1128/JVI.02020-06

Klammt, C., Lohr, F., Schafer, B., Haase, W., Dotsch, V., Ruterjans, H., Glaubitz, C., & Bernhard, F. (2004). High level cell-free expression and specific labeling of integral membrane proteins. *European Journal of Biochemistry*, *271*(3), 568–580. https://doi.org/10.1111/j.1432-1033.2003.03959.x

Kleckner, I. R., & Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1814*(8), 942–968. https://doi.org/10.1016/j.bbapap.2010.10.012

Koschel, M., Oed, D., Gerelsaikhan, T., Thomssen, R., & Bruss, V. (2000). Hepatitis B Virus Core Gene Mutations Which Block Nucleocapsid Envelopment. *Journal of Virology*, *74*(1), 1–7. https://doi.org/10.1128/JVI.74.1.1-7.2000

Kwon, J. A., & Rho, H. M. (2003). Transcriptional Repression of the Human p53 Gene by Hepatitis B Viral Core Protein (HBc) in Human Liver Cells. *Biological Chemistry*, *384*(2). https://doi.org/10.1515/BC.2003.022

Lacabanne, D., Meier, B. H., & Böckmann, A. (2018). Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy. *Journal of Biomolecular NMR*, *71*(3), 141–150. https://doi.org/10.1007/s10858-017-0156-z

Lam, A. M., Ren, S., Espiritu, C., Kelly, M., Lau, V., Zheng, L., Hartman, G. D., Flores, O. A., & Klumpp, K. (2017). Hepatitis B Virus Capsid Assembly Modulators, but Not Nucleoside Analogs, Inhibit the Production of Extracellular Pregenomic RNA and Spliced RNA Variants. *Antimicrobial Agents and Chemotherapy*, *61*(8), 1–14. https://doi.org/10.1128/AAC.00680-17

Lambert, C., Döring, T., & Prange, R. (2007). Hepatitis B Virus Maturation Is Sensitive to Functional Inhibition of ESCRT-III, Vps4, and γ2-Adaptin. *Journal of Virology*, *81*(17), 9050–9060. https://doi.org/10.1128/JVI.00479-07

Lan, Y. T., Li, J., Liao, W., & Ou, J. (1999). Roles of the Three Major Phosphorylation Sites of Hepatitis B Virus Core Protein in Viral Replication. *Virology*, *259*(2), 342–348. https://doi.org/10.1006/viro.1999.9798

Lawrence, M. C., & Colman, P. M. (1993). Shape Complementarity at Protein/Protein Interfaces. *Journal of Molecular Biology*, *234*, 946–950.

Le Pogam, S., & Shih, C. (2002). Influence of a Putative Intermolecular Interaction between Core and the Pre-S1 Domain of the Large Envelope Protein on Hepatitis B Virus

Secretion. *Journal of Virology*, *76*(13), 6510–6517. https://doi.org/10.1128/JVI.76.13.6510-6517.2002

Le Pogam, S., Yuan, T. T.-T., Sahu, G. K., Chatterjee, S., & Shih, C. (2000). Low-Level Secretion of Human Hepatitis B Virus Virions Caused by Two Independent, Naturally Occurring Mutations (P5T and L60V) in the Capsid Protein. *Journal of Virology*, *74*(19), 9099–9105. https://doi.org/10.1128/JVI.74.19.9099-9105.2000

Lecoq, L., Fogeron, M.-L., Meier, B. H., Nassal, M., & Böckmann, A. (2020). Solid-State NMR for Studying the Structure and Dynamics of Viral Assemblies. *Viruses*, *12*(10), 1–27. https://doi.org/10.3390/v12101069

Lecoq, L., Schledorn, M., Wang, S., Smith-Penzel, S., Malär, A. A., Callon, M., Nassal, M., Meier, B. H., & Böckmann, A. (2019). 100 kHz MAS Proton-Detected NMR Spectroscopy of Hepatitis B Virus Capsids. *Frontiers in Molecular Biosciences*, *6*, 58. https://doi.org/10.3389/fmolb.2019.00058

Lecoq, L., Wang, S., Dujardin, M., Zimmermann, P., Schuster, L., Fogeron, M.-L., Briday, M., Schledorn, M., Wiegand, T., Cole, L., Montserret, R., Bressanelli, S., Meier, B. H., Nassal, M., & Böckmann, A. (2021). A pocket-factor–triggered conformational switch in the hepatitis B virus capsid. *Proceedings of the National Academy of Sciences*, *118*(17), 1–12. https://doi.org/10.1073/pnas.2022464118

Lecoq, L., Wang, S., Wiegand, T., Bressanelli, S., Nassal, M., Meier, B. H., & Böckmann, A. (2018a). Localizing Conformational Hinges by NMR: Where Do Hepatitis B Virus Core Proteins Adapt for Capsid Assembly? *ChemPhysChem*, *19*(11), 1336–1340. https://doi.org/10.1002/cphc.201800211

Lecoq, L., Wang, S., Wiegand, T., Bressanelli, S., Nassal, M., Meier, B. H., & Böckmann, A. (2018b). Solid-state [13C–15N] NMR resonance assignment of hepatitis B virus core protein. *Biomolecular NMR Assignments*, *12*(1), 205–214. https://doi.org/10.1007/s12104-018-9810-y

Lee, H. W., Lee, J. S., & Ahn, S. H. (2020). Hepatitis B Virus Cure: Targets and Future Therapies. *International Journal of Molecular Sciences*, *22*(1), 213. https://doi.org/10.3390/ijms22010213

Lee, P.-C., Chao, Y., Chen, M.-H., Lan, K.-H., Lee, I.-C., Hou, M.-C., & Huang, Y.-H. (2020). Risk of HBV reactivation in patients with immune checkpoint inhibitor-treated unresectable hepatocellular carcinoma. *Journal for ImmunoTherapy of Cancer*, *8*(2), 1–9. https://doi.org/10.1136/jitc-2020-001072

Leistner, C. M., Gruen-Bernhard, S., & Glebe, D. (2007). Role of glycosaminoglycans for binding and infection of hepatitis B virus. *Cellular Microbiology*, *10*(1), 122–133. https://doi.org/10.1111/j.1462-5822.2007.01023.x

Leupin, O., Bontron, S., Schaeffer, C., & Strubin, M. (2005). Hepatitis B Virus X Protein Stimulates Viral Genome Replication via a DDB1-Dependent Pathway Distinct from That Leading to Cell Death. *Journal of Virology*, *79*(7), 4238–4245. https://doi.org/10.1128/JVI.79.7.4238-4245.2005

Li, K., Zoulim, F., Pichoud, C., Kwei, K., Villet, S., Wands, J., Li, J., & Tong, S. (2007). Critical Role of the 36-Nucleotide Insertion in Hepatitis B Virus Genotype G in Core Protein Expression, Genome Replication, and Virion Secretion. *Journal of Virology*, *81*(17), 9202–9215. https://doi.org/10.1128/JVI.00390-07

Li, Q., & Kang, C. (2020). A Practical Perspective on the Roles of Solution NMR Spectroscopy in Drug Discovery. *Molecules*, *25*(13), 2974. https://doi.org/10.3390/molecules25132974

Liu, K., & Hu, J. (2019). Secretion of empty or complete hepatitis B virions: Envelopment of empty capsids versus mature nucleocapsids. *Future Virology*, *14*(2), 95–105. https://doi.org/10.2217/fvl-2018-0128

Liu, K., Luckenbaugh, L., Ning, X., Xi, J., & Hu, J. (2018). Multiple roles of core protein linker in hepatitis B virus replication. *PLOS Pathogens*, *14*(5), e1007085. https://doi.org/10.1371/journal.ppat.1007085

Locarnini, S., Littlejohn, M., Aziz, M. N., & Yuen, L. (2013). Possible origins and evolution of the hepatitis B virus (HBV). *Seminars in Cancer Biology*, *23*(6), 561–575. https://doi.org/10.1016/j.semcancer.2013.08.006

Lok, A. S. F., & McMahon, B. J. (2009). Chronic hepatitis B: Update 2009. *Hepatology*, *50*(3), 661–662. https://doi.org/10.1002/hep.23190

Ludgate, L., Liu, K., Luckenbaugh, L., Streck, N., Eng, S., Voitenleitner, C., Delaney, W. E., & Hu, J. (2016). Cell-Free Hepatitis B Virus Capsid Assembly Dependent on the Core Protein C-Terminal Domain and Regulated by Phosphorylation. *Journal of Virology*, *90*(12), 5830–5844. https://doi.org/10.1128/JVI.00394-16

Luo, J., Xi, J., Gao, L., & Hu, J. (2020). Role of Hepatitis B virus capsid phosphorylation in nucleocapsid disassembly and covalently closed circular DNA formation. *PLOS Pathogens*, *16*(3), e1008459. https://doi.org/10.1371/journal.ppat.1008459

Maity, S., Gundampati, R. K., & Suresh Kumar, T. K. (2019). NMR Methods to Characterize Protein-Ligand Interactions. *Natural Product Communications*, *14*(5), 1–17. https://doi.org/10.1177/1934578X19849296

Makbul, C., Khayenko, V., Maric, H. M., & Böttcher, B. (2021). Conformational Plasticity of Hepatitis B Core Protein Spikes Promotes Peptide Binding Independent of the Secretion Phenotype. *Microorganisms*, *9*(5), 956. https://doi.org/10.3390/microorganisms9050956

Makbul, C., Kraft, C., Grießmann, M., Rasmussen, T., Katzenberger, K., Lappe, M., Pfarr, P., Stoffer, C., Stöhr, M., Wandinger, A.-M., & Böttcher, B. (2021). Binding of a Pocket Factor to Hepatitis B Virus Capsids Changes the Rotamer Conformation of Phenylalanine 97. *Viruses*, *13*(11), 1–15. https://doi.org/10.3390/v13112115

Makino, S., Beebe, E. T., Markley, J. L., & Fox, B. G. (2014). Cell-Free Protein Synthesis for Functional and Structural Studies. In Y. W. Chen (Ed.), *Structural Genomics* (Vol. 1091, pp. 161–178). Humana Press. https://doi.org/10.1007/978-1-62703-691-7_11

Mandal, A., Boatz, J. C., & Wheeler, T. (2018). *On the use of ultracentrifugal devices for routine sample preparation in biomolecular magic-angle-spinning NMR*. *67*(3), 165–178.

Marley, J., Lu, M., & Bracken, C. (2001). *A method for efficient isotopic labeling of recombinant proteins*. *20*, 71–75.

Maupas, P., Coursaget, P., Goudeau, A., & Bagros, P. (1976). *IMMUNISATION AGAINST HEPATITIS B IN*. 1367–1370.

Milich, D. (2003). Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection. *Hepatology*, *38*(5), 1075–1086. https://doi.org/10.1053/jhep.2003.50453

Milich, D. R., Chen, M. K., Hughes, J. L., & Jones, J. E. (1998). *The Secreted Hepatitis B Precore Antigen Can Modulate the Immune Response to the Nucleocapsid: A Mechanism for Persistence*. 5–10.

Milich, D. R., Jones, J. E., Hughes, J. L., Price, J., Raney, A. K., & McLachlan, A. (1990). Is a function of the secreted hepatitis B e antigen to induce immunologic tolerance in utero? *Proceedings of the National Academy of Sciences*, *87*(17), 6599–6603. https://doi.org/10.1073/pnas.87.17.6599

Milich, D. R., Sallberg, M., & Maruyama, T. (1995). The humoral immune response in acute and chronic hepatitis B virus infection. *Springer Seminars in Immunopathology*, *17*, 149–166. https://doi.org/10.1007/BF00196163

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. https://doi.org/10.1002/jcc.21256

Muhamad, A., Ho, K. L., Abdul Rahman, Mohd. B., Tejo, B. A., Uhrín, D., & Tan, W. S. (2015). Hepatitis B virus peptide inhibitors: Solution structures and interactions with the viral capsid. *Org. Biomol. Chem.*, *13*(28), 7780–7789. https://doi.org/10.1039/C5OB00449G

Murphy, C. M., Xu, Y., Li, F., Nio, K., Reszka-Blanco, N., Li, X., Wu, Y., Yu, Y., Xiong, Y., & Su, L. (2016). Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication. *Cell Reports*, *16*(11), 2846–2854. https://doi.org/10.1016/j.celrep.2016.08.026

Nassal, M. (1992). The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *Journal of Virology*, *66*(7), 4107–4116. https://doi.org/10.1128/jvi.66.7.4107-4116.1992

Nassal, M. (2008). Hepatitis B viruses: Reverse transcription a different way. *Virus Research*, *134*(1–2), 235–249. https://doi.org/10.1016/j.virusres.2007.12.024

Nassal, M. (2015). HBV cccDNA: Viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut*, *64*(12), 1972–1984. https://doi.org/10.1136/gutjnl-2015-309809

Nassal, M., Rieger, A., & Steinau, O. (1992). Topological analysis of the hepatitis B virus core particle by cysteine-cysteine cross-linking. *Journal of Molecular Biology*, *225*(4), 1013–1025. https://doi.org/10.1016/0022-2836(92)90101-O

Niklasch, M., Zimmermann, P., & Nassal, M. (2021). The Hepatitis B Virus Nucleocapsid—Dynamic Compartment for Infectious Virus Production and New Antiviral Target. *Biomedicines*, *9*(11), 1577. https://doi.org/10.3390/biomedicines9111577

Nimrod, A. C., & Benson, W. H. (1996). Environmental Estrogenic Effects of Alkylphenol Ethoxylates. *Critical Reviews in Toxicology*, *26*(3), 335–364. https://doi.org/10.3109/10408449609012527

Ning, X., Basagoudanavar, S. H., Liu, K., Luckenbaugh, L., Wei, D., Wang, C., Wei, B., Zhao, Y., Yan, T., Delaney, W., & Hu, J. (2017). Capsid Phosphorylation State and Hepadnavirus Virion Secretion. *Journal of Virology*, *91*(9), 1–16. https://doi.org/10.1128/JVI.00092-17

Ning, X., Luckenbaugh, L., Liu, K., Bruss, V., Sureau, C., & Hu, J. (2018). Common and Distinct Capsid and Surface Protein Requirements for Secretion of Complete and Genome-Free Hepatitis B Virions. *Journal of Virology*, *92*(14), 1–19. https://doi.org/10.1128/JVI.00272-18

Ning, X., Nguyen, D., Mentzer, L., Adams, C., Lee, H., Ashley, R., Hafenstein, S., & Hu, J. (2011). Secretion of Genome-Free Hepatitis B Virus – Single Strand Blocking Model for Virion Morphogenesis of Pararetrovirus. *PLOS Pathogens*, *7*(9), 1–14.

Oh, K.-I., Kim, J., Park, C.-J., & Lee, J.-H. (2020). Dynamics Studies of DNA with Non-canonical Structure Using NMR Spectroscopy. *Int. J. Mol. Sci.*, *21*, 1–22.

Olaniyan, L. W. B., Okoh, O. O., Mkwetshana, N. T., & Okoh, A. I. (2018). Environmental Water Pollution, Endocrine Interference and Ecotoxicity of 4-tert-Octylphenol: A Review. In P. de Voogt (Ed.), *Reviews of Environmental Contamination and Toxicology Volume 248* (Vol. 248, pp. 81–109). Springer International Publishing. https://doi.org/10.1007/398_2018_20

Orabi, A., Bieringer, M., Geerlof, A., & Bruss, V. (2015). An Aptamer against the Matrix Binding Domain on the Hepatitis B Virus Capsid Impairs Virion Formation. *Journal of Virology*, *89*(18), 9281–9287. https://doi.org/10.1128/JVI.00466-15

Pairan, A., & Bruss, V. (2009). Functional Surfaces of the Hepatitis B Virus Capsid. *Journal of Virology*, *83*(22), 11616–11623. https://doi.org/10.1128/JVI.01178-09

Paliy, O., Bloor, D., Brockwell, D., Gilbert, P., & Barber, J. (2003). Improved methods of cultivation and production of deuteriated proteins from E. coli strains grown on fully deuteriated minimal medium. *Journal of Applied Microbiology*, *94*(4), 580–586. https://doi.org/10.1046/j.1365-2672.2003.01866.x

Pastor, F., Herrscher, C., Patient, R., Eymieux, S., Moreau, A., Burlaud-Gaillard, J., Seigneuret, F., de Rocquigny, H., Roingeard, P., & Hourioux, C. (2019). Direct interaction between the hepatitis B virus core and envelope proteins analyzed in a

cellular context. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-52824-z

Pathak, R. K., Baunthiyal, M., Taj, G., & Kumar, A. (2014). Virtual screening of natural inhibitors to the predicted HBx protein structure of Hepatitis B Virus using molecular docking for identification of potential lead molecules for liver cancer. *Bioinformation*, *10*(7), 428–435. https://doi.org/10.6026/97320630010428

Patient, R., Hourioux, C., & Roingeard, P. (2009). Morphogenesis of hepatitis B virus and its subviral envelope particles. *Cellular Microbiology*, *11*(11), 1561–1570. https://doi.org/10.1111/j.1462-5822.2009.01363.x

Payne, S. (2017). Family Hepadnaviridae. In *Viruses* (pp. 321–327). Elsevier. https://doi.org/10.1016/B978-0-12-803109-4.00038-6

Pellecchia, M. (2005). Solution Nuclear Magnetic Resonance Spectroscopy Techniques for Probing Intermolecular Interactions. *Chemistry & Biology*, *12*(9), 961–971. https://doi.org/10.1016/j.chembiol.2005.08.013

Perlman, D. H., Berg, E. A., O'Connor, P. B., Costello, C. E., & Hu, J. (2005). Reverse transcription-associated dephosphorylation of hepadnavirus nucleocapsids. *Proceedings of the National Academy of Sciences*, *102*(25), 9020–9025. https://doi.org/10.1073/pnas.0502138102

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pierce, M. M., Raman, C. S., & Nall, B. T. (1999). Isothermal Titration Calorimetry of Protein–Protein Interactions. *Methods*, *19*(2), 213–221. https://doi.org/10.1006/meth.1999.0852

Poisson, F., Severac, A., Hourioux, C., Goudeau, A., & Roingeard, P. (1997). Both Pre-S1 and S Domains of Hepatitis B Virus Envelope Proteins Interact with the Core Particle. *Virology*, *228*(1), 115–120. https://doi.org/10.1006/viro.1996.8367

Polenova, T., Gupta, R., & Goldbourt, A. (2015). Magic Angle Spinning NMR Spectroscopy: A Versatile Technique for Structural and Dynamic Analysis of Solid-Phase Systems. *Anal. Biochem.*, *87*, 5458−5469.

Ponsel, D., & Bruss, V. (2003). Mapping of Amino Acid Side Chains on the Surface of Hepatitis B Virus Capsids Required for Envelopment and Virion Formation. *Journal of Virology*, *77*(1), 416–422. https://doi.org/10.1128/JVI.77.1.416-422.2003

Pope, B., & Kent, H. M. (1996). High Efficiency 5 Min Transformation of Escherichia Coli. *Nucleic Acids Research*, *24*(3), 536–537. https://doi.org/10.1093/nar/24.3.536

Prange, R., & Streeck, R. E. (1995). Novel transmembrane topology of the hepatitis B virus envelope proteins. *The EMBO Journal*, *14*(2), 247–256. https://doi.org/10.1002/j.1460-2075.1995.tb06998.x

Prange, R., Werr, M., & Löffler-Mary, H. (1999). Chaperones Involved in Hepatitis B Virus Morphogenesis. *Biological Chemistry*, *380*(3). https://doi.org/10.1515/BC.1999.042

Prince, A. M. (1968). AN ANTIGEN DETECTED IN THE BLOOD DURING THE INCUBATION PERIOD OF SERUM HEPATITIS. *Proceedings of the National Academy of Sciences*, *60*, 814–821.

Prozeller, D., Morsbach, S., & Landfester, K. (2019). Isothermal titration calorimetry as a complementary method for investigating nanoparticle–protein interactions. *Nanoscale*, *11*(41), 19265–19273. https://doi.org/10.1039/C9NR05790K

Pugh, J., Zweidler, A., & Summers, J. (1989). Characterization of the major duck hepatitis B virus core particle protein. *Journal of Virology*, *63*(3), 1371–1376. https://doi.org/10.1128/jvi.63.3.1371-1376.1989

Puthenveetil, R., & Vinogradova, O. (2019). Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments. *Journal of*

Biological Chemistry, *294*(44), 15914–15931. https://doi.org/10.1074/jbc.REV119.009178

Rabe, B., Glebe, D., & Kann, M. (2006). Lipid-Mediated Introduction of Hepatitis B Virus Capsids into Nonsusceptible Cells Allows Highly Efficient Replication and Facilitates the Study of Early Infection Events. *Journal of Virology*, *80*(11), 5465–5473. https://doi.org/10.1128/JVI.02303-05

Rajoriya, N., Combet, C., Zoulim, F., & Janssen, H. L. A. (2017). How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B. Time for an individualised approach? *Journal of Hepatology*, *67*(6), 1281–1297. https://doi.org/10.1016/j.jhep.2017.07.011

Rawat, S., & Bouchard, M. J. (2015). The Hepatitis B Virus (HBV) HBx Protein Activates AKT To Simultaneously Regulate HBV Replication and Hepatocyte Survival. *Journal of Virology*, *89*(2), 999–1012. https://doi.org/10.1128/JVI.02440-14

Revill, P. A., Chisari, F. V., Block, J. M., Dandri, M., Gehring, A. J., Guo, H., Hu, J., Kramvis, A., Lampertico, P., Janssen, H. L. A., Levrero, M., Li, W., Liang, T. J., Lim, S.-G., Lu, F., Penicaud, M. C., Tavis, J. E., Thimme, R., Zoulim, F., … Schinazi, R. F. (2019). A global scientific strategy to cure hepatitis B. *The Lancet Gastroenterology & Hepatology*, *4*(7), 545–558. https://doi.org/10.1016/S2468-1253(19)30119-0

Revill, P., Yuen, L., Walsh, R., Perrault, M., Locarnini, S., & Kramvis, A. (2010). Bioinformatic analysis of the hepadnavirus e-antigen and its precursor identifies remarkable sequence conservation in all orthohepadnaviruses. *Journal of Medical Virology*, *82*(1), 104–115. https://doi.org/10.1002/jmv.21645

Roldão, A., Mellado, M. C. M., Castilho, L. R., Carrondo, M. J., & Alves, P. M. (2010). Virus-like particles in vaccine development. *Expert Review of Vaccines*, *9*(10), 1149–1176. https://doi.org/10.1586/erv.10.115

Roseman, A. M., Berriman, J. A., Wynne, S. A., Butler, P. J. G., & Crowther, R. A. (2005). A structural model for maturation of the hepatitis B virus core. *Proceedings of the*

National Academy of Sciences, *102*(44), 15821–15826. https://doi.org/10.1073/pnas.0504874102

Ross, A., Kessler, W., Krumme, D., Menge, U., Wissing, J., van den Heuvel, J., & Flohé, L. (2004). Optimised fermentation strategy for 13C/15N recombinant protein labelling in Escherichia coli for NMR-structure analysis. *Journal of Biotechnology*, *108*(1), 31–39. https://doi.org/10.1016/j.jbiotec.2003.10.014

Sahu, D., Bastidas, M., Lawrence, C. W., Noid, W. G., & Showalter, S. A. (2016). Assessing Coupled Protein Folding and Binding Through Temperature-Dependent Isothermal Titration Calorimetry. In *Methods in Enzymology* (Vol. 567, pp. 23–45). Elsevier. https://doi.org/10.1016/bs.mie.2015.07.032

Schanda, P., Kupče, Ē., & Brutscher, B. (2005). SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds. *Journal of Biomolecular NMR*, *33*(4), 199–211. https://doi.org/10.1007/s10858-005-4425-x

Schittl, B., & Bruss, V. (2014). Mutational profiling of the variability of individual amino acid positions in the hepatitis B virus matrix domain. *Virology*, *458–459*, 183–189. https://doi.org/10.1016/j.virol.2014.04.030

Schledorn, M., Malär, A. A., Torosyan, A., Penzel, S., Klose, D., Oss, A., Org, M., Wang, S., Lecoq, L., Cadalbert, R., Samoson, A., Böckmann, A., & Meier, B. H. (2020). Protein NMR Spectroscopy at 150 kHz Magic-Angle Spinning Continues To Improve Resolution and Mass Sensitivity. *ChemBioChem*, *21*(17), 2540–2548. https://doi.org/10.1002/cbic.202000341

Schlicksup, C. J., Wang, J. C.-Y., Francis, S., Venkatakrishnan, B., Turner, W. W., VanNieuwenhze, M., & Zlotnick, A. (2018). Hepatitis B virus core protein allosteric modulators can distort and disrupt intact capsids. *ELife*, *7*, 1–23. https://doi.org/10.7554/eLife.31473

Schlicksup, C. J., & Zlotnick, A. (2020). Viral structural proteins as targets for antivirals. *Current Opinion in Virology*, *45*, 43–50. https://doi.org/10.1016/j.coviro.2020.07.001

Schmieder, P. (2006). Multidimensional NMR Spectroscopy. *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, 1204–1208.

Schulze, A., Gripon, P., & Urban, S. (2007). Hepatitis B virus infection initiates with a large surface protein-dependent binding to heparan sulfate proteoglycans. *Hepatology*, *46*(6), 1759–1768. https://doi.org/10.1002/hep.21896

Schweitzer, A., Horn, J., Mikolajczyk, R. T., Krause, G., & Ott, J. J. (2015). Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. *The Lancet*, *386*(10003), 1546–1555. https://doi.org/10.1016/S0140-6736(15)61412-X

Seeger, C., & Mason, W. S. (2000). Hepatitis B Virus Biology. *Microbiology and Molecular Biology Reviews*, *64*(1), 51–68. https://doi.org/10.1128/MMBR.64.1.51-68.2000

Seeger, C., & Mason, W. S. (2015). Molecular biology of hepatitis B virus infection. *Virology*, *479–480*, 672–686. https://doi.org/10.1016/j.virol.2015.02.031

Seitz, S., Urban, S., Antoni, C., & Böttcher, B. (2007). Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions. *The EMBO Journal*, *26*(18), 4160–4167. https://doi.org/10.1038/sj.emboj.7601841

Sekiba, K., Otsuka, M., Funato, K., Miyakawa, Y., Tanaka, E., Seimiya, T., Yamagami, M., Tsutsumi, T., Okushin, K., Miyakawa, K., Ryo, A., & Koike, K. (2022). HBx-induced degradation of Smc5/6 complex impairs homologous recombination-mediated repair of damaged DNA. *Journal of Hepatology*, *76*(1), 53–62. https://doi.org/10.1016/j.jhep.2021.08.010

Selzer, L., Katen, S. P., & Zlotnick, A. (2014). The Hepatitis B Virus Core Protein Intradimer Interface Modulates Capsid Assembly and Stability. *Biochemistry*, *53*(34), 5496–5504. https://doi.org/10.1021/bi500732b

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). *Escherichia coli* Physiology in Luria-Bertani Broth. *Journal of Bacteriology*, *189*(23), 8746–8749. https://doi.org/10.1128/JB.01368-07

Shih, C., Wu, S.-Y., Chou, S.-F., & Yuan, T.-T. T. (2020). Virion Secretion of Hepatitis B Virus Naturally Occurring Core Antigen Variants. *Cells*, *10*(1), 43. https://doi.org/10.3390/cells10010043

Shiloach, J., & Fass, R. (2005). Growing E. coli to high cell density—A historical perspective on method development. *Biotechnology Advances*, *23*(5), 345–357. https://doi.org/10.1016/j.biotechadv.2005.04.004

Srivastava, V. K. (2019). Isothermal titration calorimetry. *Data Processing Handbook for Complex Biological Data Sources*, 125–138.

Starr, C. A., Barnes, L. F., Jarrold, M. F., & Zlotnick, A. (2022). Hysteresis in Hepatitis B Virus (HBV) Requires Assembly of Near-Perfect Capsids. *Biochemistry*, *61*(7), 505–513. https://doi.org/10.1021/acs.biochem.1c00810

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234. https://doi.org/10.1016/j.pep.2005.01.016

Summers, J., & Mason, W. S. (1982). Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell*, *29*(2), 403–415. https://doi.org/10.1016/0092-8674(82)90157-X

Takai, K., Sawasaki, T., & Endo, Y. (2010a). Practical cell-free protein synthesis system using purified wheat embryos. *Nature Protocols*, *5*(2), 227–238. https://doi.org/10.1038/nprot.2009.207

Takai, K., Sawasaki, T., & Endo, Y. (2010b). *The Wheat-Germ Cell-Free Expression System*. 7.

Takegoshi, K., Nakamura, S., & Terao, T. (2001). 13C±1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chemical Physics Letters*, 7.

Tan, G., Xu, F., Song, H., Yuan, Y., Xiao, Q., Ma, F., Qin, F. X.-F., & Cheng, G. (2018). Identification of TRIM14 as a Type I IFN-Stimulated Gene Controlling Hepatitis B Virus Replication by Targeting HBx. *Frontiers in Immunology*, *9*, 1–14. https://doi.org/10.3389/fimmu.2018.01872

Tang, H., Delgermaa, L., Huang, F., Oishi, N., Liu, L., He, F., Zhao, L., & Murakami, S. (2005). The Transcriptional Transactivation Function of HBx Protein Is Important for Its Augmentation Role in Hepatitis B Virus Replication. *Journal of Virology*, *79*(9), 5548–5556. https://doi.org/10.1128/JVI.79.9.5548-5556.2005

Taverniti, V., Ligat, G., Debing, Y., Kum, D. B., Baumert, T. F., & Verrier, E. R. (2022). Capsid Assembly Modulators as Antiviral Agents against HBV: Molecular Mechanisms and Clinical Perspectives. *Journal of Clinical Medicine*, *11*(5), 1–12. https://doi.org/10.3390/jcm11051349

Toita, R. (2015). Applications of human hepatitis B virus preS domain in bio- and nanotechnology. *World Journal of Gastroenterology*, *21*(24), 7400. https://doi.org/10.3748/wjg.v21.i24.7400

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, NA-NA. https://doi.org/10.1002/jcc.21334

Tsukuda, S., & Watashi, K. (2020). Hepatitis B virus biology and life cycle. *Antiviral Research*, *182*, 1–10. https://doi.org/10.1016/j.antiviral.2020.104925

Urban, S., Bartenschlager, R., Kubitz, R., & Zoulim, F. (2014). Strategies to Inhibit Entry of HBV and HDV Into Hepatocytes. *Gastroenterology*, *147*(1), 48–64. https://doi.org/10.1053/j.gastro.2014.04.030

Venkatakrishnan, B., & Zlotnick, A. (2016). The Structural Biology of Hepatitis B Virus: Form and Function. *Annual Review of Virology*, *3*(1), 429–451. https://doi.org/10.1146/annurev-virology-110615-042238

Verardi, R., Traaseth, N. J., Masterson, L. R., Vostrikov, V. V., & Veglia, G. (2012). Isotope Labeling for Solution and Solid-State NMR Spectroscopy of Membrane Proteins. In H. S. Atreya (Ed.), *Isotope labeling in Biomolecular NMR* (Vol. 992, pp. 35–62). Springer Netherlands. https://doi.org/10.1007/978-94-007-4954-2_3

Veronese, F. M., & Harris, J. M. (2002). Introduction and overview of peptide and protein pegylation. *Advanced Drug Delivery Reviews*, *54*(4), 453–456. https://doi.org/10.1016/S0169-409X(02)00020-0

Verrier, E. R., Colpitts, C. C., Bach, C., Heydmann, L., Weiss, A., Renaud, M., Durand, S. C., Habersetzer, F., Durantel, D., Abou-Jaoudé, G., López Ledesma, M. M., Felmlee, D. J., Soumillon, M., Croonenborghs, T., Pochet, N., Nassal, M., Schuster, C., Brino, L., Sureau, C., … Baumert, T. F. (2016). A targeted functional RNA interference screen uncovers glypican 5 as an entry factor for hepatitis B and D viruses: VIRAL HEPATITIS. *Hepatology*, *63*(1), 35–48. https://doi.org/10.1002/hep.28013

Viswanathan, U., Mani, N., Hu, Z., Ban, H., Du, Y., Hu, J., Chang, J., & Guo, J.-T. (2020). Targeting the multifunctional HBV core protein as a potential cure for chronic hepatitis B. *Antiviral Research*, *182*, 104917. https://doi.org/10.1016/j.antiviral.2020.104917

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics*, *59*(4), 687–696. https://doi.org/10.1002/prot.20449

Wang, J. C.-Y., Dhason, M. S., & Zlotnick, A. (2012). Structural Organization of Pregenomic RNA and the Carboxy-Terminal Domain of the Capsid Protein of Hepatitis B Virus. *PLoS Pathogens*, *8*(9), 1–10. https://doi.org/10.1371/journal.ppat.1002919

Wang, S., Fogeron, M.-L., Schledorn, M., Dujardin, M., Penzel, S., Burdette, D., Berke, J. M., Nassal, M., Lecoq, L., Meier, B. H., & Böckmann, A. (2019). Combining Cell-Free Protein Synthesis and NMR Into a Tool to Study Capsid Assembly Modulation.

Frontiers in Molecular Biosciences, *6*, 1–11. https://doi.org/10.3389/fmolb.2019.00067

Watanabe, T., Sorensen, E. M., Naito, A., Schott, M., Kim, S., & Ahlquist, P. (2007). Involvement of host cellular multivesicular body functions in hepatitis B virus budding. *Proceedings of the National Academy of Sciences*, *104*(24), 10205–10210. https://doi.org/10.1073/pnas.0704000104

Wei, L., & Ploss, A. (2021a). Hepatitis B virus cccDNA is formed through distinct repair processes of each strand. *Nature Communications*, *12*(1), 1–13. https://doi.org/10.1038/s41467-021-21850-9

Wei, L., & Ploss, A. (2021b). Mechanism of Hepatitis B Virus cccDNA Formation. *Viruses*, *13*(8), 1–19. https://doi.org/10.3390/v13081463

Wider, G. (2000). Structure Determination of Biological Macromolecules in Solution Using Nuclear Magnetic Resonance Spectroscopy. *BioTechniques*, *29*(6), 1278–1294. https://doi.org/10.2144/00296ra01

Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *73*, 1–16. https://doi.org/10.1016/j.pnmrs.2013.02.001

Wingfield, P. T., Stahl, S. J., Williams, R. W., & Steven, A. C. (1995). *Hepatitis Core Antigen Produced in Escherichia coli: Subunit Composition, Conformational Analysis, and In Vitro Capsid Assemblyt*. 14.

Wong, D. J., & Locarnini, S. A. (2018). Molecular Virology and Life Cycle. In J.-H. Kao & D.-S. Chen (Eds.), *Hepatitis B Virus and Liver Disease* (pp. 1–23). Springer Singapore. https://doi.org/10.1007/978-981-10-4843-2_1

Wynne, S. A., Crowther, R. A., & Leslie, A. G. W. (1999). The Crystal Structure of the Human Hepatitis B Virus Capsid. *Molecular Cell*, *3*(6), 771–780. https://doi.org/10.1016/S1097-2765(01)80009-5

Xi, J., Liu, H., & Hu, J. (2022). Regulation of Hepatitis B Virus Virion Release and Envelopment Timing by Nucleocapsid and Envelope Interactions. *Journal of Virology*, *96*(1), e01305-21. https://doi.org/10.1128/JVI.01305-21

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., Huang, Y., Qi, Y., Peng, B., Wang, H., Fu, L., Song, M., Chen, P., Gao, W., Ren, B., Sun, Y., Cai, T., Feng, X., Sui, J., & Li, W. (2012). Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *ELife*, *1*, 1–28. https://doi.org/10.7554/eLife.00049

Yang, C.-C., Huang, E.-Y., Li, H.-C., Su, P.-Y., & Shih, C. (2014). Nuclear Export of Human Hepatitis B Virus Core Protein and Pregenomic RNA Depends on the Cellular NXF1-p15 Machinery. *PLoS ONE*, *9*(10), e106683. https://doi.org/10.1371/journal.pone.0106683

Yang, H.-C., & Kao, J.-H. (2014). Persistence of hepatitis B virus covalently closed circular DNA in hepatocytes: Molecular mechanisms and clinical significance. *Emerging Microbes & Infections*, *3*(1), 1–7. https://doi.org/10.1038/emi.2014.64

Yu, X., Jin, L., Jih, J., Shih, C., & Hong Zhou, Z. (2013). 3.5Å cryoEM Structure of Hepatitis B Virus Core Assembled from Full-Length Core Protein. *PLoS ONE*, *8*(9), e69729. https://doi.org/10.1371/journal.pone.0069729

Yuan, T. T.-T., Sahu, G. K., Whitehead, W. E., Greenberg, R., & Shih, C. (1999). The Mechanism of an Immature Secretion Phenotype of a Highly Frequent Naturally Occurring Missense Mutation at Codon 97 of Human Hepatitis B Virus Core Antigen. *Journal of Virology*, *73*(7), 5731–5740. https://doi.org/10.1128/JVI.73.7.5731-5740.1999

Yuan, T. T.-T., & Shih, C. (2000). A Frequent, Naturally Occurring Mutation (P130T) of Human Hepatitis B Virus Core Antigen Is Compensatory for Immature Secretion Phenotype of Another Frequent Variant (I97L). *Journal of Virology*, *74*(10), 4929–4932. https://doi.org/10.1128/jvi.74.10.4929-4932.2000

Zheng, C., Fu, Y., Xu, Z., Zou, Y., & Deng, K. (2018). Hepatitis B virus core protein dimer-dimer interface is critical for viral replication. *Molecular Medicine Reports*. https://doi.org/10.3892/mmr.2018.9620

Zlotnick, A., Cheng, N., Stahl, S. J., Conway, J. F., Steven, A. C., & Wingfield, P. T. (1997). Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein: Implications for morphogenesis and organization of encapsidated RNA. *Proceedings of the National Academy of Sciences*, *94*(18), 9556–9561. https://doi.org/10.1073/pnas.94.18.9556

Zlotnick, A., Venkatakrishnan, B., Tan, Z., Lewellyn, E., Turner, W., & Francis, S. (2015). Core protein: A pleiotropic keystone in the HBV lifecycle. *Antiviral Research*, *121*, 82–93. https://doi.org/10.1016/j.antiviral.2015.06.020

CONCLUSION AND PERSPECTIVES

In the current context, it seems important to continue research on the hepatitis B virus (HBV) in order to be able, one day, to eradicate this disease which still affects hundreds of thousands of people every year. The treatments available today are restrictive and expensive, resulting in blocked access for some patients, especially those in poor countries.

Thus, we have focused on one part of the HBV viral cycle, the envelopment, to try to better understand this mechanism and in this sense consider how to block the replication of the virus. We performed *in vitro* studies, producing more specifically the truncated core protein (Cp149) by bacterial culture using the *Escherichia coli* production system. When we disassembled the protein and then reassembled it in order to purify it, we saw a difference in conformation when analyzing both forms in solid-state nuclear magnetic resonance (NMR). This switch is caused by Triton X-100 (TX100) that we added during the purification step. This detergent binds directly in the hydrophobic pocket formed by the Cp dimer, which is an important binding site for envelopment. From here we distinguished the form A, bound to TX100, and the form B, with its free pocket. This discovery allowed us to foresee the possibility of modulating this ligand to design one with better affinity and which is less toxic. For this purpose, solution NMR is a powerful technique that allowed us to screen several molecules, and to define the best structure. These results were also supported by the use of isothermal titration calorimetry (ITC) to characterize the thermodynamic parameters of binding. Through our research, we were able to conclude that the structure of the ligand that binds in the hydrophobic pocket must respect some rules: an aromatic ring, a short hydrophilic chain and a hydrophobic chain in para position. These results allow us to foresee further tests that we could perform in the future, in particular with a molecule with several interaction sites. Indeed, it would be interesting to design a compound that would bind to the hydrophobic pocket and also to the capsid-assembly modulators (CAMs) binding site.

Subsequently, we focused on the interactions between the capsid and the envelope of HBV, in order to better understand how the envelopment step appears. Before studying these interactions, we focused on the study of Oct peptides, which are thought to bind to the Cp dimer and inhibit capsid-envelope interactions. Solution NMR studies showed that these peptides bind specifically to the spike of the Cp dimer, and we were able to confirm these results by providing additional results through solid-state NMR analysis. Our results also show possible allosteric effects leading to the detection of chemical shift perturbations (CSPs) at the dimer base. Spectra with polarization transfer are currently being acquired and analyzed.

After the study of Oct peptides, we focused on capsid-envelope interactions. Indeed, it was shown that capsid and envelope proteins interact, in particular the S and L proteins of the envelope (HBs S and HBs L respectively), whereas the M protein does not bind to Cp. Thus, these results also opened up the question of the preS domain, which is found in Nter in

addition to the HBs S sequence to give HBs L. This domain, which thus appears to be important for these capsid-envelope interactions, was the focus of our initial analyses. First, we studied the interaction between the Cp149 dimer and preS, which led us to notice that these two proteins could not interact, results that were obtained by ITC, and in solution NMR looking from the point of view of both preS and Cp149. This information was crucial for our further research. Next, we wanted to try to determine the binding sites on each protein, first on preS side and then on Cp149 capsids side. Through solution NMR titration, we observed decreases in intensity on the spectrum of preS after successive additions of Cp149. After a study of the intensity decreases of each peak, we were able to conclude that the whole preS protein was interacting with Cp capsids, not a particular part of its sequence. Next, using solid state NMR, we sought to determine the binding site on the Cp149 capsids. A study of CSPs led us to assume that preS interacts partly at the spike, but also at other sites. More spectra should be recorded to get more information about the binding site on Cp149. The ITC allowed us to calculate the stoichiometry and dissociation constant associated with this binding, showing that a preS protein interacts with a Cp149 dimer in the capsid at the micromolar range. Finally, we wanted to evaluate the ability of TX100 to inhibit the interaction between Cp and preS. For this purpose, we again performed solution NMR experiments on the preS side, and solid-state NMR experiments on the capsid side. Both techniques led to the same conclusion, one showing that the decrease in peaks was no longer as obvious, and the other showing almost no CSPs. Thus, we could assume that TX100 was able to interfere in the interaction between the capsid and the preS domain, but not entirely. In terms of perspectives, it would be interesting to study more deeply the chemical shifts between the capsid (form A) and preS, and then to consider performing these analyses with the whole core protein (Cp183).

In the laboratory, we have the capacity to produce our proteins of interest using the cell-free system, and more specifically the wheat-germ expression system, which allows us to have all the metabolic resources assigned to the protein. Using this technique, we evaluated the different possibilities of capsid-envelope interactions. Thus, we observed that Cp149 dimer did not bind to the preS domain. Subsequently, we noticed that the Cp183 protein seems to interact with preS1, preS, HBs S, HBs L, which agrees with previous *in vivo* studies that showed the same results. This production technique is interesting because it allows expression conditions close to those of physiology, and also to produce phosphorylated preS.

All the results we obtained allow us to foresee the possibility of blocking the interaction between the capsid and the envelope. For the continuation of this project, it seems most interesting to consider the study of mutants of the core protein or of preS, to see their impact on the binding.