Production, assembly and solid-state NMR analysis of various hepatitis B virus capsids

Shishan Wang

N°d'ordre NNT : xxx

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 205
Ecole Doctorale Interdisciplinaire Sciences-Santé

Discipline : Biochimie

Soutenue publiquement le 26/09/2019, par :
Shishan WANG

Production, assembly and solid-state NMR analysis of various hepatitis B virus capsids

---

Devant le jury composé de :

Xavier Hanoulle, CR CNRS, Université de Lille, Rapporteur

Malene Rinkjobing Jensen, DR2 CNRS, Université Grenoble Alpes, Rapporteur

Lucie Etienne, CR CNRS, École normale supérieure de Lyon, Examinatrice

Olivier Walker, Maître de conférence, Université Claude Bernard Lyon 1, Examinateur

Beat Meier, Professeur, École polytechnique fédérale de Zurich, Examinateur

Anja Böckmann, DR1 CNRS, Université Claude Bernard Lyon 1, Directrice de thèse

This thesis has been carried out under the supervision of Anja Böckmann, in the team « solid-state NMR of proteins » that is part of the unit UMR 5086 – MMSB (Molecular Microbiology and Structural Biochemistry) at IBCP:

Institut de Biologie et Chimie des Protéines

7, passage du Vercors

F-69367 Lyon cedex 07

Résumé

L'hépatite B est une maladie du foie qui pose un problème majeur de santé publique, avec 257 millions de personnes atteints d'une infection chronique et environ 780 000 morts par an. Il n'existe à ce jour aucun traitement permettant de guérir complètement de l'infection, malgré l'utilisation notamment d'analogues de nucléos(t)ides et d'interférons. Afin d'atteindre l'objectif de l'organisation mondiale de la santé, qui vise à éliminer l'hépatite B d'ici 2030, de nouvelles thérapies ont besoin d'être développées. Étant donné son rôle clé dans le cycle de vie du virus de l'hépatite B (VHB), la protéine core qui forme la capside virale est aujourd'hui l'une des cibles avec le plus grand potentiel thérapeutique.

Nos recherches sont focalisées sur la caractérisation des capsides du VHB dans différents états conformationnels en utilisant des techniques de biochimie et de RMN du solide, afin de révéler leur conformation précise sous différentes conditions, incluant l'interaction des capsides avec des antiviraux, et la relation entre la conformation de la capside et la maturation du virus. Un système d'expression bactérienne ainsi qu'un système acellulaire de synthèse de protéine à base de germes de blé ont été établis au laboratoire pour produire les capsides, et des protocoles pour désassembler puis réassembler les capsides en présence de différents types d'ARN ont été implémentés. Des échantillons de capsides formées dans *E. coli* et réassemblées *in vitro* ont été analysés par RMN. Les différentes formes de capsides observées incluent les protéines tronquées Cp140 et Cp149, la protéine entière Cp183, phosphorylée P-Cp183, et enfin des mutants.

Dans un premier temps, nous avons préparé des échantillons pour l'attribution séquentielle de la protéine core par RMN du solide. L'utilisation de la détection carbone en RMN requiert plusieurs dizaines de milligrammes d'échantillon, qui

ont pu être produits en utilisant l'expression bactérienne en milieu minimum contenant des isotopes marqués. Les attributions séquentielles ont été réalisées sur la protéine tronquée Cp149, qui donne des spectres très similaires à Cp183. Cet échantillon a également été utilisé pour identifier les différences conformationnelles entre les 4 monomères de la capside, qui sont provoquées par la symétrie icosaédrale T=4.

Ensuite, l'objet principal de cette thèse a été l'investigation et la comparaison d'une large variété de capsides, incluant Cp183, P-Cp183, Cp149, Cp140 qui forme principalement des capsides de symétrie icosaédrale T=3, et les mutants Cp140 C61A et Cp183 F97L. Nous avons étudié tous ces échantillons dans leur forme autoassemblée dans les bactéries *E. coli*, ainsi que dans leur forme réassemblée. Pour le réassemblage de la protéine entière, qui requiert la présence d'acides nucléiques, nous avons testé différents types d'ARN, y compris l'ARN viral prégénomique. Nous avons étudié différentes symétries (T=3 et T=4), ainsi que les états d'oxydation de la capside, et comparé les différences de conformation grâce aux perturbations de déplacements chimiques observées dans les spectres RMN. Nous avons pu identifier les acides aminés impliqués dans les changements conformationnels majeurs entre les différentes préparations.

La RMN du solide en détection proton a récemment émergé comme un outil important pour l'analyse de protéines produites en quantités moindres. Nous avons appliqué cette stratégie à l'analyse des capsides de Cp149 afin d'obtenir l'attribution des protons amides. Pour cela, de la protéine deutérée a été produite en bactéries, ce qui a nécessité l'adaptation des protocoles d'expression.

La détection proton par RMN du solide peut être combinée avec succès à la synthèse des protéines en système acellulaire, qui donne de faibles rendements par rapport aux cultures en bactéries. Cette approche est particulièrement intéressante pour analyser la modulation de l'assemblage des capsides induite par la présence de drogues. Bien que nous ayons commencé

à étudier l'impact de modulateurs d'assemblage par RMN en détection carbone sur des capsides formées dans *E. coli* et réassemblées (données préliminaires non montrées dans ce manuscrit), la détection proton ouvre la voie vers l'analyse de l'impact de ces modulateurs sur l'assemblage des protéines core directement à la sortie du ribosome. Nous montrons ici que l'expression en système acellulaire combinée à la RMN du solide en détection proton peut être utilisée pour analyser les déplacements chimiques de la capside, et dans le futur leurs modulations conformationnelles.

Objectifs de la thèse

L'objectif de la thèse était de caractériser structuralement la capside du virus de l'hépatite B (VHB) en utilisant la RMN du solide. Tout comme en cristallographie par rayons X et en microscopie électronique, la préparation d'échantillons n'est pas seulement l'étape initiale d'un travail structural, mais est également centrale pour concevoir les objets adéquats qui vont permettre d'illuminer la compréhension mécanistique du virus. La capside du VHB passe par une grande variété d'états durant son cycle de vie, que nous cherchons à imiter *in vitro* afin d'obtenir une vision détaillée sur sa conformation, ce qui est difficile à accomplir par d'autres techniques. En effet, les déplacements chimiques RMN ont l'avantage d'être extrêmement sensibles aux changements conformationnels des protéines. Ce projet a pu être réalisé grâce à notre collaborateur Michael Nassal, qui non seulement a contribué par son immense connaissance sur le VHB et en particulier sur la capside, mais également par la confection de tous les vecteurs d'expression et protocoles que nous avons utilisé au cours de ce travail. Nous avons ainsi pu établir les protocoles d'expression et purification de la capside au laboratoire, et les optimiser pour les besoins de la RMN du solide.

Cette thèse décrit nos efforts dans la préparation d'un grand nombre d'échantillons pour la caractérisation des capsides par RMN et pour la conception de stratégies visant à comprendre la diversité conformationnelle des capsides.

Le Chapitre 1 contient un résumé des connaissances générales sur l'hépatite B et le VHB, incluant : les traitements actuels, la structure du virus, le cycle de vie du virus, le génome et les protéines virales ; la maturation et l'enveloppement de la capside ; les modulateurs de l'assemblage de la capside actuellement en développement.

Le Chapitre 2 est la section de méthodologie, qui introduit les technologies expérimentales utilisées dans cette thèse. Les échantillons de protéine ont été produits soit par système d'expression en bactérie, soit par synthèse en système acellulaire à base de germes de blé, et ont ensuite été caractérisés par RMN du solide.

Le Chapitre 3 contient les résultats de nos recherches, publiés ou non publiés. La première section détaille la production de tous les échantillons de la protéine core utilisés dans cette thèse. Les seconde et troisième parties montre le travail publié sur la caractérisation par RMN des capsides Cp149, incluant l'attribution RMN en détection carbone et l'identification des 4 conformations correspondant aux différentes sous-unités de la capside. La quatrième section montre la découverte d'une nouvelle conformation de la capside qui pourrait être liée au signal de maturation dans le virus (papier en préparation). La cinquième partie décrit l'application de la spectroscopie RMN à 100 kHz MAS en détection proton pour la caractérisation des capsides (papier en révision). Enfin, la dernière partie des résultats montre l'étude par RMN de la modulation de l'assemblage des capsides en utilisant le système de synthèse des protéines en système acellulaire à base de germes de blé (papier soumis).

Le dernier chapitre contient nos conclusions et perspectives pour le travail présenté dans cette thèse.

Summary

Hepatitis B is a widely spread liver disease which causes a heavy burden for human health, with 257 millions of people affected by chronic infection and about 780,000 deaths per year. Yet, infected patients cannot be completely cured by current treatments using notably nucleos(t)ide analogues and interferons. In order to achieve the goal of the World Health Assembly (WHA), who wishes to eliminate hepatitis B by 2030, new therapies need to be developed. Given its critical role for the Hepatitis B virus (HBV) life cycle, the core protein (Cp) which forms the viral capsid is today one of the antiviral targets with the highest potential.

Our research focuses on the characterization of HBV capsids in different conformational states using biochemistry and solid-state NMR, aiming at revealing their precise conformation under different conditions, including the interaction of capsids with antivirals, and the correlation between capsid conformation and viral maturation. For sample preparation, both a bacterial expression system and a wheat germ cell-free protein synthesis system have been established in the laboratory to produce HBV capsids, and protocols to disassemble and reassemble capsids with different nucleic acids have been implemented. Both capsids preformed in *E. coli* and capsids reassembled *in vitro* were addressed to NMR studies. Different capsids forms include the truncated versions Cp140 and Cp149, the full-length protein Cp183, the phosphorylated P-Cp183 and mutant forms.

First, we have prepared samples for the sequential assignment of the protein using solid-state NMR. The use of carbon-13 detection asks for several tens of milligrams of sample, which were produced using labeled isotopes and bacterial expression in minimal media. Sequential assignments were performed using the truncated capsid Cp149, which showed highly similar spectra to Cp183. This sample was also used to identify conformational differences

between the four different monomers in the capsid, which are due to the T=4 icosahedral symmetry.

Then, the main body of the thesis is the investigation and comparison of a variety of different capsid forms, including Cp183, P-Cp183, Cp149, Cp140, another truncated form resulting in mainly T=3 icosahedral assemblies, and Cp140 C61A and Cp183 F97L mutants. We investigated all samples in both the *E. coli*-produced and reassembled forms, which needs for the full-length protein the presence of nucleic acids, of which we tested several, including the viral pregenomic RNA. We investigated different symmetries, as well as oxidation states of the capsid, and compared the differences via chemical shift perturbations observed in NMR spectra. We reported in a site-specific manner the major conformational changes observed between the different preparations.

Proton-detected solid-state NMR has recently emerged as a tool for analyzing proteins with the need of less sample amount. We have applied this strategy to the analysis of the Cp149 capsids, in order to obtain sequential assignments of the amide proton resonances. For this, deuteration of the protein in bacteria was used as well, needing adaptation of sample preparation protocols.

Proton detection can be successfully combined with cell-free protein synthesis, which gives low yields compared to bacterial expression. This approach is of potential interest to analyze capsid assembly modulation induced by the presence of drugs. While we have started in the framework of this thesis to analyze the capsid in presence of different capsid assembly modulators by carbon-13 detected NMR on *E. coli* and reassembled capsids (preliminary results not reported here), proton detection opens the way to an analysis of the impact of capsid modulation directly on the exit of the core proteins from the ribosome, on assembly. We showed that cell-free expression combined with proton-detection solid-state NMR can be used to analyze capsid chemical shifts, and thus in future work the conformational modulations.

Objectives of the thesis work

The objective of the thesis was to introduce the hepatitis B virus (HBV) capsid into structural characterization by solid-state NMR. As in X-ray crystallography and electron microscopy, sample preparation is not only the initial step of structural work, but is central in devising objects that can illuminate the mechanistic understanding. The HBV capsid goes through a variety of states during its life cycle, which we aim to mimic *in vitro* in order to analyze its conformation in a detailed view that is difficult to obtain by other techniques. Indeed, NMR chemical shifts are a highly sensitive probe for the conformational changes of proteins. The basis to the realization of the project was established by an important body of work from our collaborator Michael Nassal, who not only contributed with his immense knowledge of HBV in general and especially the capsid to this work, but most importantly also with all optimized expression vectors he contributed. This allowed us to establish the protocols in the lab, and modify and optimize them to the needs of solid-state NMR sample preparation.

This thesis describes our efforts in providing a large number of samples for NMR characterization of HBV capsids, as well as devising strategies to understand their conformational diversity. The NMR part was carried out by Lauriane Lecoq from the group, as well as Maarten Schledorn from ETH.

Chapter 1 provides a summary of general knowledge about hepatitis B and the hepatitis B virus, including: epidemic and current treatments of hepatitis B; the viral structure, life cycle, genome and proteins; viral capsid maturation and envelopment; the developed assembly modulators for HBV capsid.

Chapter 2 is a methods section which introduces the experimental technologies used in this thesis. Protein samples were produced either in a bacterial expression system or by wheat-germ cell-free synthesis, and characterized by solid-state NMR.

Chapter 3 contains the results of our research, both published or unpublished. The first part details the production of all the HBV core protein samples used in this thesis. The second and third parts show the published work on the NMR characterization of Cp149 capsids, including the NMR assignment using carbon-detection and the identification of the four different capsid subunits conformations. The fourth section is about the discovery of a new capsid conformation which could be linked to the maturation of the viral capsid (to be submitted). The fifth part displays the application of proton-detection with 100 kHz MAS NMR spectroscopy in HBV capsid characterization (paper under review). Finally, the last results section introduces the combination of cell-free protein synthesis and NMR in capsid assembly modulation study (paper submitted).

The last chapter reports our conclusions, as well as a perspective, for the work presented in this thesis.

Related publications and contributions

Publications:

1. **Solid-state [13C-15N] NMR resonance assignment of hepatitis B virus core protein.** Lauriane Lecoq#, Shishan Wang#, Thomas Wiegand, Stéphane Bressanelli, Michael Nassal*, Beat H. Meier*, Anja Böckmann*. *Biomolecular NMR Assignments*. 2018. 12, 205–214.

Contribution: as a co-first author, I executed biochemical experiments part of this paper, contributed to writing the paper.

Lauriane Lecoq executed the experiments of solid-state NMR.

2. **Localizing conformational hinges by NMR: where do HBV core proteins adapt for capsid assembly?** Lauriane Lecoq, Shishan Wang, Thomas Wiegand, Stéphane Bressanelli, Michael Nassal*, Beat H. Meier*, Anja Böckmann*. *ChemPhysChem*. 2018. 19, 1336–1340.

Contribution: Prepared samples for NMR characterization, contributed to writing the paper.

Manuscript in preparation:

3. **NMR reveals a second conformation of the hepatitis B virus capsid.** Lauriane Lecoq#, Shishan Wang#, Marie-Laure Fogeron, Marie Dujardin, Michael Nassal, Beat H. Meier, Anja Böckmann

Contribution: executed and analyzed biochemical experiments and electron microscopy, prepared all different samples for NMR characterization, including the transcription of pgRNA, contributed substantially to discussion and analysis of results; contributed to writing the paper.

Manuscript under review:

4. **100 kHz MAS proton-detected NMR spectroscopy of hepatitis B virus capsids**. Lauriane Lecoq#, Maarten Schledorn#, Shishan Wang, Susanne

Smith-Penzel, Alexander A. Malär, Morgane Callon, Michael Nassal, Beat H. Meier*, Anja Böckmann*.

Contribution: Prepared the NMR samples, optimized the conditions for protein deuteration and contributed to writing the paper.

Manuscript submitted:

5. **Combining cell-free protein synthesis and NMR into a tool to study capsid assembly modulation.** Shishan Wang, Marie-Laure Fogeron, Maarten Schledorn, Marie Dujardin, Susanne Penzel, Dara Burdette, Jan Martin Berke, Michael Nassal, Lauriane Lecoq, Beat H. Meier & Anja Böckmann.

Contribution: carried out protein bacterial and cell-free syntheses and analyses, generated NMR samples, recorded and analyzed electron micrographs, analyzed results, wrote the paper.

Acknowledgements

Time is a tireless warrior always keeping on going, along with whom I have spent almost 4 years in Lyon, my favorite city. During my stay in France, I have gained a lot in both research and life. It is a really special experience for me to live abroad, beginning with expectancy, full of surprise and harvest. Here, I express my heartfelt gratitude to all the people who gave me concern and helps.

First, many thanks to my supervisor, Dr. Anja Böckmann, who gave me the opportunity to study in such an excellent group, with exciting research topics, harmonious work atmosphere and enthusiastic group members. I am also really grateful for everything she has done for me, which made things easier to adapt myself to the study and life in France. Meanwhile, I should give my appreciation to Dr. Lauriane Lecoq for the excellent work in NMR characterization on this project, and for the selfless help I received from her, which made my study going smoothly.

Thanks to the collaborators of this project: Prof. Michael Nassal from University of Freiburg, who shared the plasmids and bacterial needed for this project; Prof. Beat Meier from ETH Zurich, who guided this research with Dr. Anja Böckmann; and Maarten Schledorn from ETH Zurich, who took part in NMR data collection analysis.

I would like to give my thanks to my laboratory colleague for the kind helps to my study and living in France. They are: Roland Montserret, a respectable humorous man, who picked me up for the airport and helped me deal with all kinds of formalities when I first arrived France; Marie-Laure Fogeron, a patient and courteous person, who gave many helps and guidance in cell-free production; Guillaume David, a young and promising man, from whom I received a lot of useful advices for matters of the university; Marie Dujardin, a diligent researcher and nice teammate, as well as a good cooperator of this project.

Thanks to my former groupmates, from whom I also benefited a lot, including Aurélie Badillo, Denis Lacabanne, Anaïs Papin, Lauriane Varrel, Claire Chuilon, François Penin, and Karem Cobas from Cuba.

Thanks to the referees Dr. Malene R. Jensen and Dr. Xavier Hanouille who accepted to read and evaluate my work, to all the members of my PhD jury: Dr. Lucie Etienne, Pr. Olivier Walker and Pr. Beat Meier, and also to the members of my PhD committee for their time and advices along my PhD: Dr. Stéphane Bressanelli and Dr. Christophe Grangeasse.

Thanks to my friends in Lyon, along with whom I spent a very nice time abroad, and my family in China, whose support is very important for me, encouraging me keeping on going.

Last but not least, I would like to express my sincere gratitude to the China Scholarship Council (CSC). It is this financial support of CSC for my PhD study that made this wonderful experience happen.

CONTENTS

Chapter II

Chapter III

Abbreviations

CAMs	capsid assembly modulators
cccDNA	covalently closed circular DNA
CFS	Cell-free systems
CHB	chronic hepatitis B
Cp	core protein
dsL DNA	double-stranded linear DNA
HBeAg	hepatitis B e antigen
HBsAg	hepatitis B surface antigen
HBV	Hepatitis B virus
HCC	hepatocellular carcinoma
HSPG	hepatocyte-associated heparin sulfate proteoglycans
MBD	matrix binding domain
MD	matrix domain (MD)
MVBs	multivesicular bodies
NAs	nucleos(t)ide analogues
NPCs	nuclear pore complexes
NTCP	sodium taurocholate co-transporting polypeptide
ORFs	open reading frames
PDB	Protein Data Bank
pgRNA	pregenomic RNA
Pol	polymerase
rcDNA	relaxed circular DNA
ssDNA	single-stranded linear DNA
WGE-CF	wheat germ cell-free
WGE	wheat germ extracts

HEPATITIS B BACKGROUND

Contents

● Hepatitis B

➢ A major burden for public health

Hepatitis B infection is a major health problem in developing and also developed countries (Figure 1). According to the report of World Health Organization (WHO) in 2015, about 6% of the adult population in Western Pacific and African regions is infected, about 3.3%, 2.0% and 1.6% of the general population is infected in the Eastern Mediterranean, South-East Asia and European regions respectively, and 0.7% of the population is infected in the American continent. Nearly half of the human population lives in high hepatitis B epidemic areas, and one third of the global population was or is in contact with the virus of hepatitis B (Bertoletti et al., 2018). No less than 2 billion people are estimated being infected by Hepatitis B in the past or present (WHO, 2017; 2015).

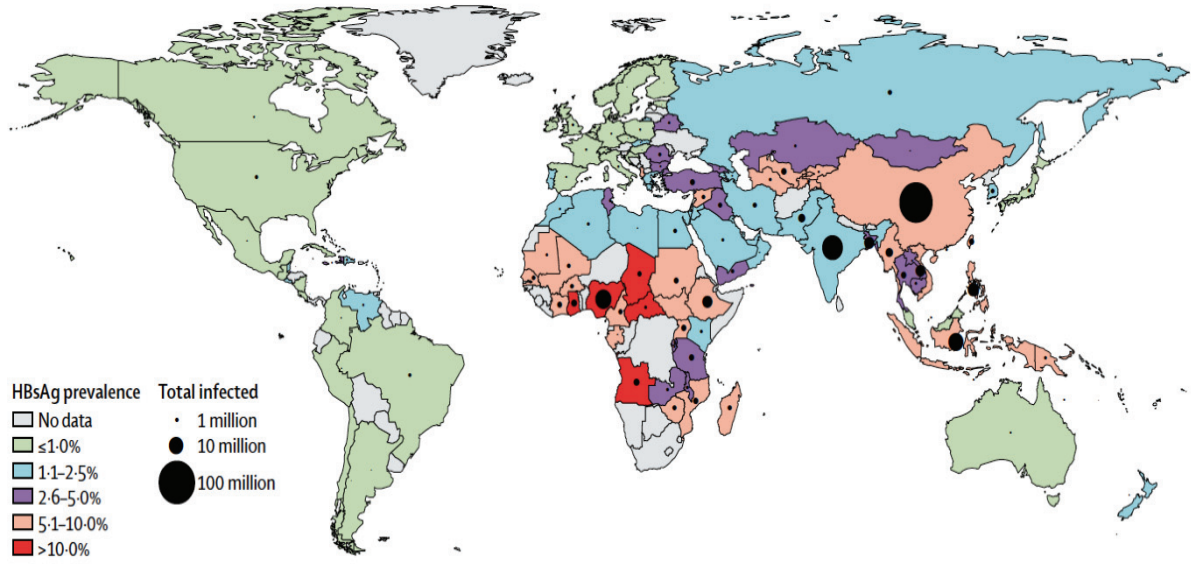


Figure 1. HBsAg prevalence estimates for 2016. As a global health problem, hepatitis B infection occurs in almost all countries with different prevalence. More than 250 million people are infected by hepatitis B virus and remain unable to be completely cured. Figure source: (Collaborators, 2018).

The hepatitis B virus (HBV) can be detected in the body fluids of a person who is infected, such as blood, saliva, semen, vaginal secretions. Meanwhile, the virions in body fluids can spread by getting into the blood stream of new patient and finally leading to liver infection. The main transmission of HBV occurs via blood, sexual contact and mother-to-child (WHO, 2016). Acute hepatitis B infection is self-limiting, and in most infection events, adults could recover within 6 months without treatment (Bell and T. Nguyen, 2009; Hollinger and Lau, 2006; Hyams, 1995; Lavanchy, 2004). In about 5–10% of infections, subjects are incapable of clearing the HBV virions and evolve into chronic infection (Liang, 2009). Chronic hepatitis B (CHB) infection is defined as persistent infection, and is characterized by the presence of hepatitis B surface antigen (HBsAg) in blood or serum for more than 6 months (Hyams, 1995; Lavanchy, 2004). The transition from acute infection to chronic infection is closely related to the age of the infected patient: it happens less than 12% in adult and children above 5 years old among different ethnic groups; 20-60% in infants under 5 years old, and about 90% in neonates with infected mothers (Bell and T. Nguyen, 2009; Goldstein et al., 2005; Hoofnagle et al., 2007; Hyams, 1995; McMahon, 2009). Approximately 15-40% of CHB patients have the risk of deterioration to cirrhosis, liver failure or hepatocellular carcinoma (HCC) (Lok, 2002). It is estimated that 30% of cirrhosis and more than half of diagnosed hepatocellular carcinoma can be contributed by hepatitis B infection (Fako and X. W. Wang, 2017; Perz et al., 2006), leading to 45% of male patients' and 15% of female patients' life-span significantly shortened (Bell and T. Nguyen, 2009).

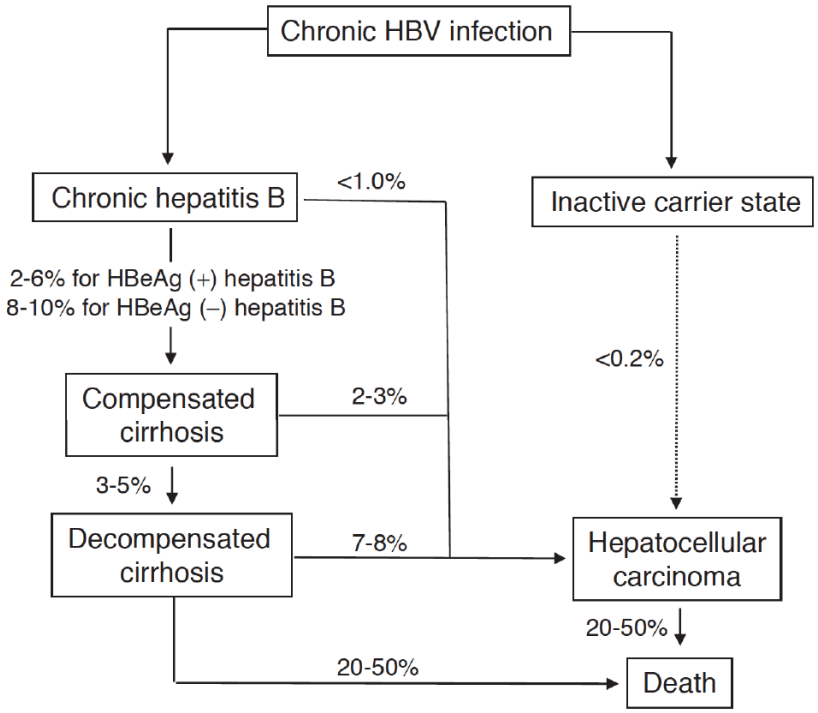
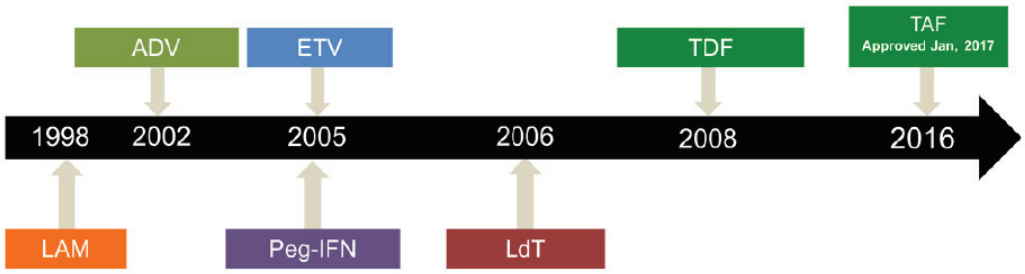


Figure 2. Progression of chronic HBV infection. The risk of deterioration from different clinical states is displayed with estimated annual rates (%).According to: (Christoph Seeger et al., 2013).

There is currently no cure to clear hepatitis B infection (Levrero et al., 2016; Lin and Kao, 2016; Yuen et al., 2018). The treatment of acute hepatitis B consists mainly in supplying fluids lost from vomiting and diarrhea, because no treatment has been shown to be really helpful (Mantzoukis et al., 2017; WHO, 2015). Current medicines used in CHB treatments are nucleos(t)ide analogues (NAs) and interferon (PEG-IFN-α) (Figure 3), but none of them can really cure the infection (Mak et al., 2017; Nijampatnam and Liotta, 2019). The NAs, as well as oral antiviral agents recommended by WHO such as tenofovir and entecavir, can suppress viral replication to retard the progression of cirrhosis, and reduce incidence of hepatic carcinoma (Bartenschlager et al., 2019; Hung et al., 2015; Pei et al., 2017; Yuen et al., 2018), but they are incompetent to clear the cccDNA, which causes persistent infection, and the need of long-term treatment suffers from emergence of resistance (Lam et al., 2017; Mak et al., 2017; Nijampatnam and Liotta, 2019). Meanwhile, the PEG-IFN-α can be tolerated only by a small part of patients, with modest efficacy and possible severe side effects (Mak et al., 2017; Nijampatnam and Liotta, 2019).

a



b

Drug and patient population	Resistance at year of therapy expressed as percentage of patients					
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Lamivudine	23	46	55	71	80	–
Telbivudine, HBeAg-Pos	4.4	21	–	–	–	–
Telbivudine, HBeAg-Neg	2.7	8.6	–	–	–	–
Adefovir, HBeAg-Neg	0	3	6	18	29	–
Adefovir (LAM-resistant)	Up to 20	–	–	–	–	–
Tenofovir	0	0	0	0	–	–
Entecavir (naïve)	0.2	0.5	1.2	1.2	1.2	1.2
Entecavir (LAM resistant)	6	15	36	46	51	57

Figure 3. Medicines for HBV infection. a): timeline for the approval of hepatitis B medicines by U.S. Food and Drug Administration (FDA). Six nucleos(t)ide analogues: lamivudine (LAM), adefovir (ADV), entecavir (ETV), telbivudine (LdT), tenofovir (TDF), tenofovir alafenamide (TAF). A interferon: PEG-IFN-α. Taken from: (Schinazi et al., 2018) b): incidence of drug resistance with nucleos(t)ide analogues treatements. Taken from: (Christoph Seeger et al., 2013).

Therefore, vaccination of hepatitis B vaccine is the most effective way to prevent the spread of hepatitis B since the 1980s. The 3-shot series vaccine recommend by WHO has covered 84% of new born child with 95% efficacy (Figure 4) (WHO, 2017; 2015). A mathematical model reveals that it prevented about 65 million people from infection of hepatitis B, and more than 1 million people were saved from disease caused by hepatitis B infection in the year 2000 (Goldstein et al., 2005). Despite of the great contribution made by the vaccine in the battle against hepatitis B, there are still an estimated 257 million people living with hepatitis B infection, resulting in over 700,000 deaths every year (Collaborators, 2018; GBD 2016 Causes of Death Collaborators, 2017; Gish et al., 2015; Schlicksup et al., 2018a; WHO, 2017; 2015). This makes hepatitis B infection one of the top death cause in humans, and largely contributes to

hepatocellular carcinoma, which is the 5th frequent cancer in the world (Lavanchy, 2004; Lozano et al., 2012; Parkin et al., 2001).

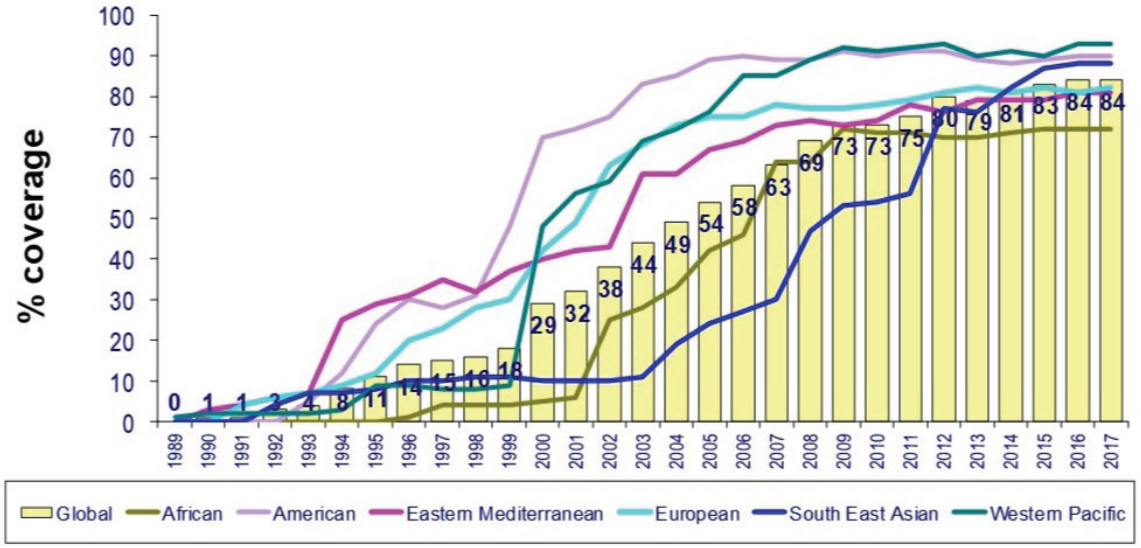


Figure 4. Coverage of 3rd dose of hepatitis B vaccine in infants global from 1989 to 2017. Figure source: World Health Organization (WHO) website.

The first global health sector strategy on viral hepatitis was adopted by the World Health Assembly (WHA) in 2016. Its goal is to eliminate all five hepatic viruses (hepatitis A, B, C, D and E), and especially hepatitis B and C, as a major public health burden, by 2030 (WHO, 2016). In order to achieve this goal, the member countries and secretariat of WHO have to take actions to reduce the incidence of viral hepatitis from 6-10 million cases to 0.9 million cases, and the annual deaths by hepatitis from 1.4 million to 0.5 million or less by 2030 (WHO, 2016). The actions are mainly focused on raising public awareness and preventing transmission; develop advanced medicines, diagnostics and treatment services; and promoting the health equity in society (WHO, 2016).

➢ The hepatitis B virus

Hepatitis B is caused by the hepatitis B virus, an enveloped DNA virus belonging to the Hepadnavirus family, which attacks liver tissue specifically. The Hepadnavirus family

was divided into two genera: the avihepadnaviruses, which infect birds, such as duck, heron, stork, ross goose, parrot, etc.; and the orthohepadnaviruses, that infect mammals and include the human HBV, chimpanzees HBV, woodchuck hepatitis virus, orangutans hepatitis virus, bat hepatitis virus, etc. (Locarnini et al., 2013). However, three new genera, including metahepadnaviruses, parahepadnaviruses and herpetohepadnaviruses, have been discovered and proposed recently (Dill et al., 2016; Hahn et al., 2015; Lauber et al., 2017) (Figure 5). Meanwhile, a sister taxon of hepadnavirus family, which is formed by non-enveloped fish viruses and thus referred as nackednaviruses, has also been revealed (Lauber et al., 2017).

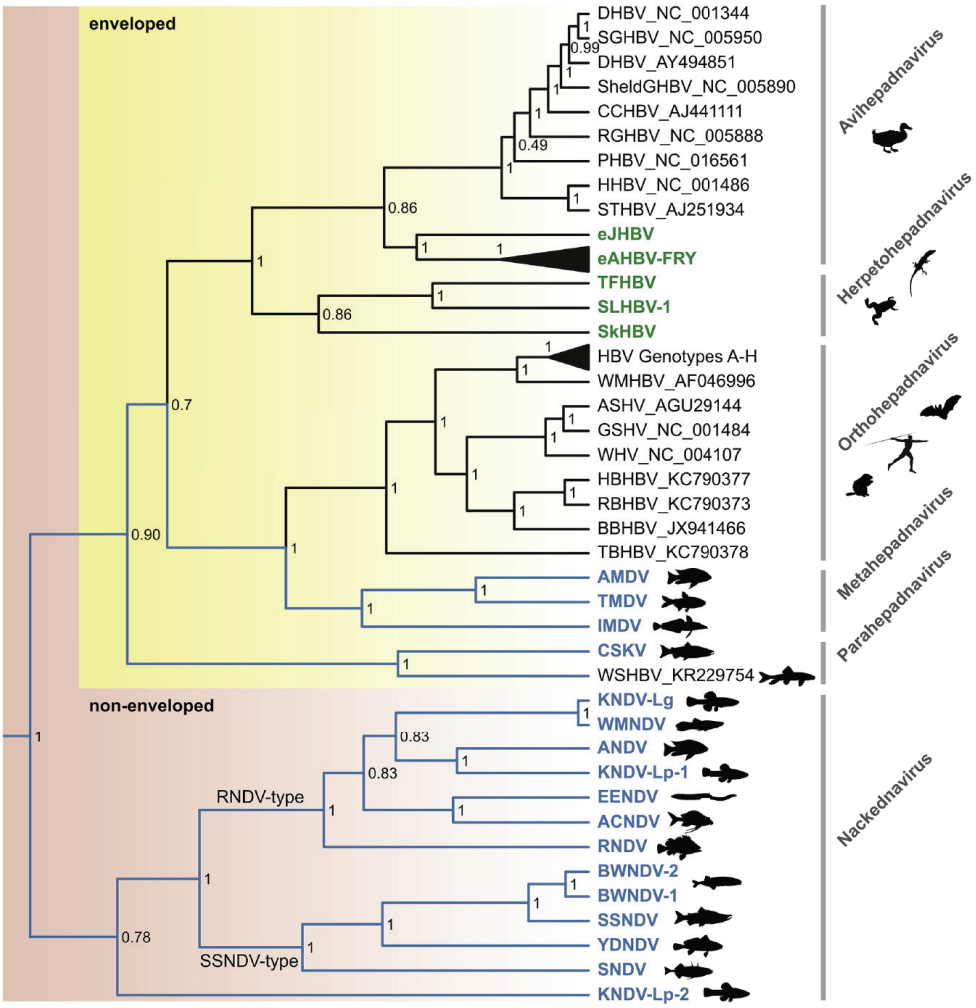


Figure 5. Phylogenetic Relationship of Hepadnavirus family and Nackednaviruses. Taken from: (Lauber et al., 2017).

The origin of HBV has always been controversial (Jazayeri et al., 2010; Littlejohn et al., 2016; Paraskevis et al., 2013; Simmonds, 2001), while recent research based on the

evidence of archeology and genetics reveals that the history of HBV infection to human can date back to at least 4500 years ago (Mühlemann et al., 2018). But it is only in 1946 that the virus was first suspected to exist (MacCallum, 1947) and the surface antigens were first discovered in the blood of aboriginal patient in Australia by Baruch Blumberg in 1965 (Blumberg and Alter, 1965). The HBV particle was then discovered by David Dane and his collaborator using electron microscopy in 1970 (Dane et al., 1970), and the HBV particle is therefore also called Dane particle.

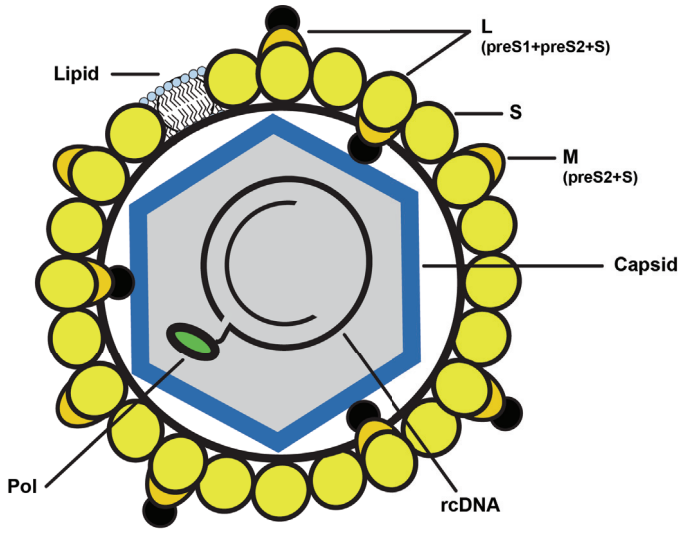


Figure 6. The schematic structure of hepatitis B virus.

Mature virions of HBV are spherical particles with a diameter of 42-47 nm. The particle is constituted by an outer envelope composed of host cell lipids and surface proteins, and an inner icosahedral nucleocapsid consisting of a core protein shell and a copy of viral DNA associating with a DNA polymerase (Figure 6) (Warner and Locarnini, 2012). There are 3 kinds of surface proteins, which commonly are referred to as hepatitis B surface antigens (HBsAg), constituting the envelope of the virus. They are translated from the same gene sequence with different start codons and same stop codons, therefore all of the 3 surface proteins share the same peptide sequence at the C-terminal. The small HBV surface protein (abbr. SHBs or S) is the smallest surface protein; the middle HBV surface protein (abbr. MHBs or M) contains S and an additional domain called preS2 ; the large HBV surface protein (abbr. LHBs or L) contains S, preS2 and an additional domain called preS1 (Bertoletti et al., 2018). The 3 surface proteins envelope the capsid of HBV, an icosahedral shell formed by 120 or 90 copies

of core protein (Cp) dimers, which are called T=4 or T=3 capsids, respectively (Zlotnick et al., 1996). Cp comprises a 140 amino acid-long assembly domain, a linker of 9 amino acids and a 34-36 residues C-terminal arginine-rich domain (Heger-Stevic et al., 2018; Zlotnick et al., 2015). The genome of HBV is a partially double-stranded relaxed circular DNA (rcDNA), which is packaged inside the icosahedral capsid together with the viral polymerase protein (Pol) (Warner and Locarnini, 2012; Wong and Locarnini, 2017).

➢ The genotypes of HBV

Since the complete sequence of the HBV genome was determined in 1979 (Galibert et al., 1979), nine genotypes of HBV have been identified and classified based on the genetic differences of the complete genome greater than 7.5% and named from genotype A to genotype I (Kramvis, 2014; Pourkarim et al., 2014). Five of the nine genotypes (A-E) contribute to around 96% of chronic HBV infections with 17% of genotype A, 14% of genotype B, 26% of genotype C, 22% of genotype D and 18% of genotype E, meanwhile, genotypes from F to I only contribute to no more than 2% of human CHB (Velkov et al., 2018). Furthermore, nearly 40 subgenotypes are subdivided from genotypes A–D, F, H, I, with the divergence of intergroup nucleotide between 4% to 8%, named with an additional digit following the genotype letter, such as A1, A2, B1, B2, etc (Kramvis, 2014; Pourkarim et al., 2014; Rajoriya et al., 2017; Tong and Revill, 2016). In addition to the nine genotypes, a novel variant strain, which is called genotype J in most publications, was isolated from an 88-year-old male Japanese who was potentially infected during the Second World War. According to phylogenetic analysis, the genome of the strain is closest to gibbon HBV instead of human HBV and it might be a recombinant of gibbon HBV and human HBV genotype C (Locarnini et al., 2013; Tatematsu et al., 2009). But with no further reports on its infection to human, whether this variant should be considered as a new genotype is still dissented by some researchers (Velkov et al., 2018).

Distribution of HBV genotypes is different depending on the geographical and ethnic population (**Table 1**). Genotype A is a frequent variant in North America, northern

Europe and Africa; genotype B and genotype C are the dominant variants in East Asia; genotype D is the most common variant in southern Europe, Middle East and India; genotype E mainly exists in Sub-Saharan Africa; genotype F is the predominant variant in Latin America; genotype G and genotype H are found mostly in the USA and Latin America and rarely in other places; genotype I and J occurred mostly in East Asia (Locarnini et al., 2013; Rajoriya et al., 2017; Tong and Revill, 2016; Velkov et al., 2018). Recent statistical research based on global available HBV genotyping Data reveals that most of genotype A patients live in Sub-Saharan Africa (72.2%) and Asia (17.2%); while the overwhelming majority of genotype B (98.9%) and genotype C (98.6%) patients live in Asia; patients infected by genotype D mainly live in Asia (61.9%), Africa (22.0%) and Europe (13.5%); and finally about 97.0% of genotype E patients live in Sub-Saharan Africa (Velkov et al., 2018).

Genotype	Subtype	Genome length (bp)	Core protein (AA)	Global distribution
A	A1	3221	185	Africa, Asia
	A2			Northern Europe, North America
B	B1	3215	183	Japan
	B2			Rest of Asia
	B3			Indonesia, China
	B4			Vietnam, Cambodia
C	C1	3215	183	Far East
	C2			Far East
	C3			Polynesia
	C4			Australian indigenous population
	C5			Philippines
D	D1	3182	183	Europe, Middle East, Egypt, India, Asia
	D2			Europe, Japan
	D3			Europe, Asia, South Africa, USA
	D4			Australia, Japan, Papua New Guinea

E		3212	183	Sub-Saharan Africa, UK, France
F	F1	3215	183	Central and South America, Bolivia
	F2			Brazil, Venezuela, Nicaragua
	F3			Venezuela, Panama, Columbia
	F4			Argentina, Bolivia, France
G		3248	195	USA, Germany, Japan, France
H		3215	183	USA, Japan, Nicaragua
I	I1	3215	183	Vietnam, Cambodia
	I2			Laos
J		3182	183	Japan

Table 1. Major genotypes of HBV and their distributions.

The geographical distribution, as well as the length of genome and core protein, is shown in the table. Modified from (Locarnini et al., 2013; Wong and Locarnini, 2017).

- **The serotypes of HBV**

HBV can also be classified into four major serotypes: *adw, adr, ayw and ayr*, according to the S protein variants of amino acids 122 and 160, with subtype *d* corresponding to K122, subtype *y* R122, subtype *w* K160, and subtype *r* corresponding to R160 (Norder et al., 1992; Okamoto et al., 1988; Warner and Locarnini, 2012). The "*a*" indicates a common region of HBsAg to all HBV wild-type strains containing an unknown number of epitopes (Echevarría and Avellón, 2006). Moreover, at least ten HBsAg subtypes have been distinguished by serological analysis and designated *adw2*, *adw3*, *adw4q*$^{-}$, *adrq*$^{+}$, *adrq*$^{-}$, *ayw1*, *ayw2*, *ayw3*, *ayw4* and *ayr* (Echevarría and Avellón, 2006; Kay and Zoulim, 2007). Sequencing on codons 101-180 of S genes of HBV strains suggest that the mutations of residue 127 are important for the *w1*-*w4* subtypes. The P127, T127 and L127 are encoded by *w1*/*w2*, *w3* and *w4* respectively, and the subtype *w1* also encodes for R122, F134 , and/or A159 in addition to P127 (Norder et al., 2004). The *q* is a determinant which is found to be absent from *adw4* and some strains of *adr* (Couroucé-Pauty et al., 1978; Magnius et al., 1975).

Serotyping was important for detecting HBsAg for the safety of blood supply and therefore widely used in epidemiological studies (Kay and Zoulim, 2007). Despite that a correlation has been found between HBsAg subtypes and genotypes of HBV (Figure 7) (Norder et al., 1992; Rajoriya et al., 2017), the variations of serotypes are limited by only several amino acids residues of HBsAg and can't reflect HBV distribution in geographic and epidemiologic as well as genotypes classification (Moriya et al., 2002). As a result, the HBV genotypes, rather than serotypes, are more and more widely used in HBV epidemiology, transmission, and geographic distribution study (Warner and Locarnini, 2012).

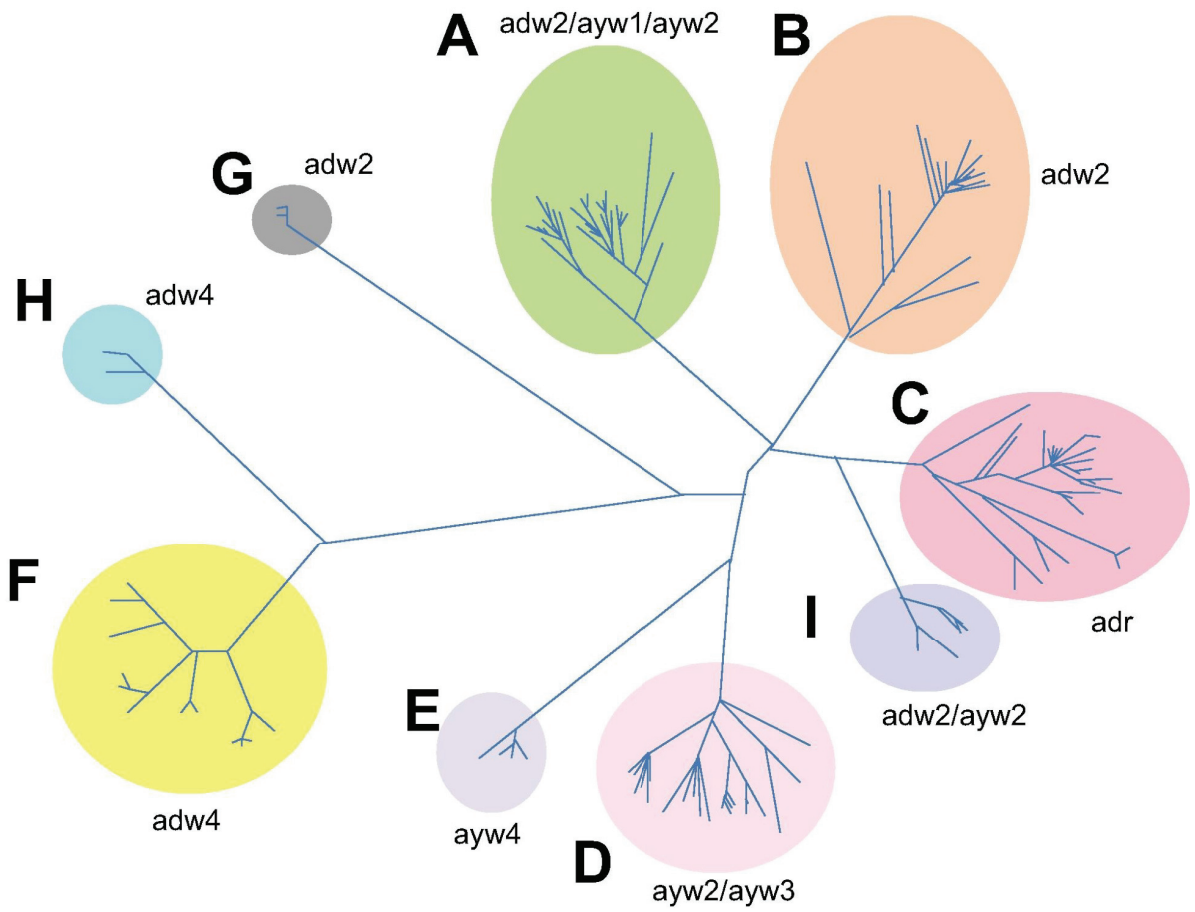


Figure 7. Serotypes phylogenetic tree. Serotypes are shown in the unrooted phylogenetic tree basing on HBV genotypes. The *ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw4, adwq, adrq+* and *adrq-* are serological subtypes identified based on HBsAg heterogeneity. Source: (Rajoriya et al., 2017)

➢ The HBV life cycle

The life cycle of the HBV can be divided into 4 main events (Figure 8)**:** a) attachment, penetration and uncoating of HBV virions; b) cccDNA conversion and transcription activation; c) capsid assembly and replication of the viral genome; d) capsid envelopment and secretion of viral particles (Wong and Locarnini, 2017).

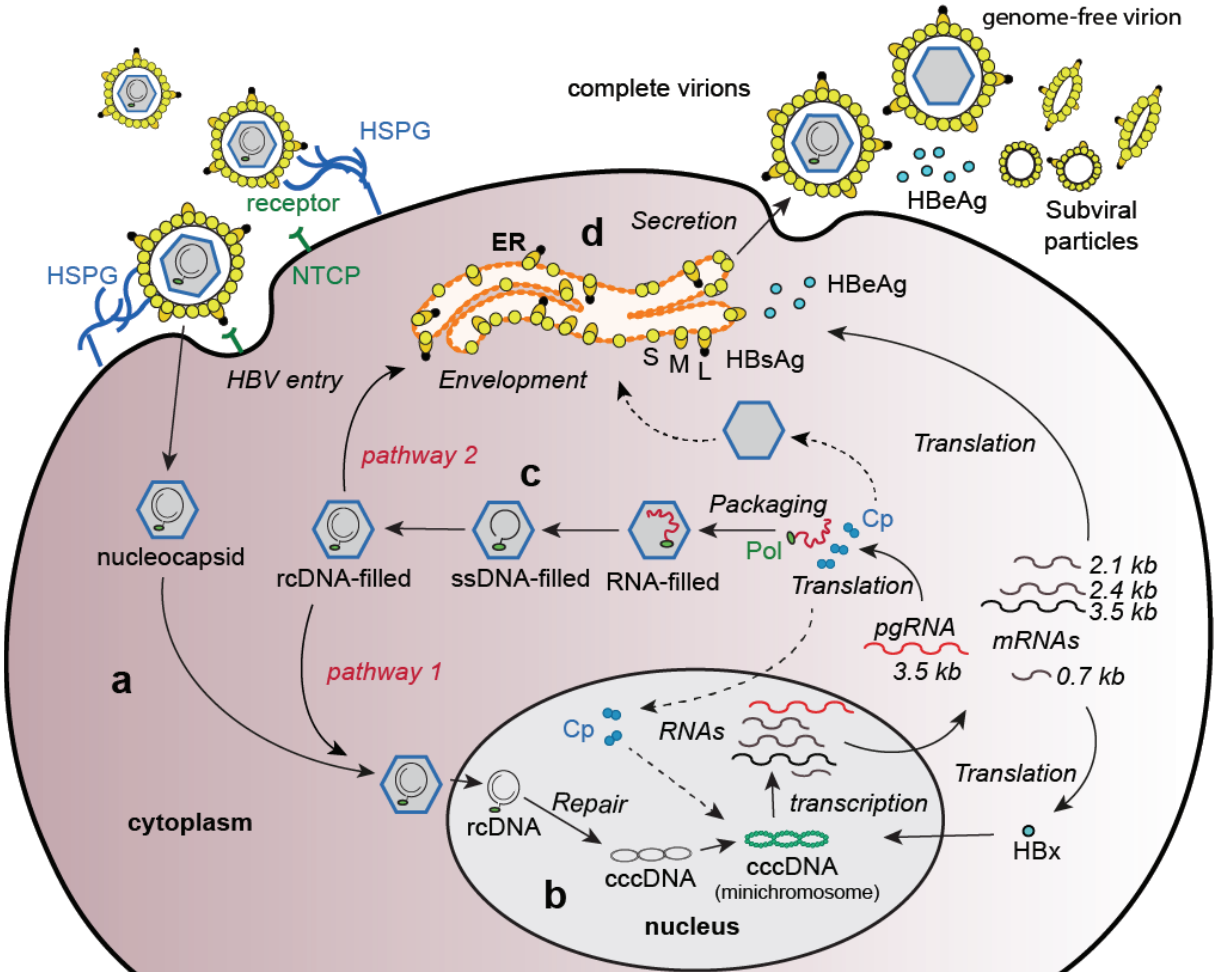


Figure 8. The life cycle of HBV. HSPG: heparin sulfate proteoglycans; NTCP: sodium taurocholate co-transporting polypeptide; rcDNA: relaxed circular DNA; cccDNA: covalently closed circular DNA; pgRNA: pregenomic RNA; ssDNA: single-stranded linear DNA; Cp: core protein; Pol: polymerase; ER: endoplasmic reticulum.

Attachment, penetration and uncoating of HBV virions. HBV virions first adhere to hepatocyte-associated heparin sulfate proteoglycans (HSPG) (Schulze et al., 2007), a low-affinity receptor for viruses, and then bind to sodium taurocholate co-transporting

polypeptide (NTCP) (Bertoletti et al., 2018). NTCP is a high-affinity receptor on hepatocytes surface for hepatitis B and D virus (H. Yan et al., 2012). The binding of HBV to NTCP leads to the transport of the HBV nucleocapsid (NC) into hepatocytes by endocytosis (Ni et al., 2014; H. Yan et al., 2014). The envelope of the virion, including lipids and surface proteins, is removed by a yet unknown mechanism, while the capsids enter the hepatocyte cytoplasm. Capsids supposedly undergo structural changes in the C-terminal of the core proteins due to phosphorylation in cytoplasm, which leads to the exposure of the presumed nuclear localization signals (Kann et al., 1999; Rabe et al., 2003). Meanwhile, they are directed toward the karyotheca via an active microtubule-dependent mechanism, and are then uncoated around or in the nuclear pore complexes (NPCs) to release viral genome into nucleus (Gallucci and Kann, 2017; Rabe et al., 2003; 2006).

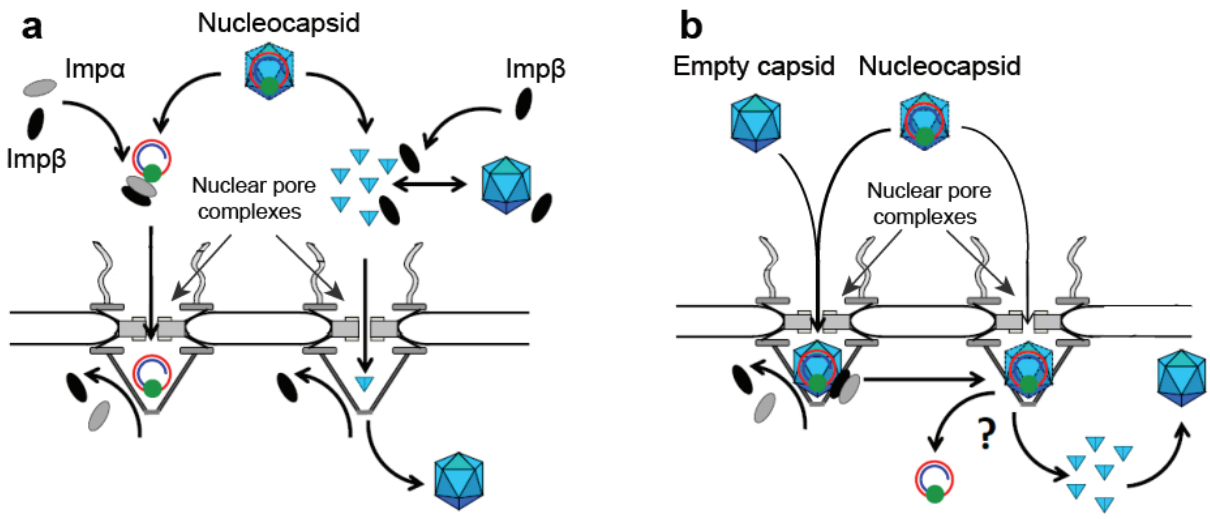


Figure 9. Nuclear import of HBV genome. Two hypothetical models are proposed for the nuclear entry of HBV genome, as reviewed in (Gallucci and Kann, 2017). The capsids and Cp dimers are shown by blue icosahedra or triangles. The rcDNA is shown as partly double-strand with red and blue color. The nuclear pore complexes (NPCs) consist of about 30 kinds of nucleoporins (Nups), of which the Nup153 and Tpr Protein form a nuclear basket on the karyoplasm side (Alber et al., 2007; Appen and M. Beck, 2016; Krull et al., 2004). The importin α and β (impα, impβ), which are nuclear transport receptors responsible for active transport of macromolecules, are shown with gray and black ellipse respectively. a): the polymerase-mediated import model. Capsids are considered to be disassembled in cytoplasm or at the NPCs, releasing the Pol-DNA complexes and Cp dimers in the cytoplasm side. The exposed rcDNA is transported to the karyoplasm by the help of importin α and β, which bind to the viral polymerase. The released Cp dimers in the

cytoplasm could eventually reassemble to empty capsids. The importin β could disassemble the empty capsids and transport the released dimers into nucleus. The binding of importins is dissolved at the nuclear basket after the Pol-DNA complexes or Cp dimers entry. b): the intact capsid import model. Nucleocapsids enter nucleus through NPCs via the interaction with importin α and β. The importin α and β are dissolved in the nuclear basket, and the capsids are disassembled by the interaction with Nup153. Then the Pol-DNA complexes and Cp dimers are released to karyoplasm, where rcDNA is repaired to cccDNA and the empty capsids could be reassembled from released Cp dimers. Adapted from: (Gallucci and Kann, 2017).

cccDNA conversion and transcription activation. The genome of HBV is then converted from the partially double-stranded relaxed circular DNA (rcDNA) to a completely double-stranded covalently closed circular DNA (cccDNA), which is responsible for the transcription of pgRNA, as well as other viral messenger RNAs (mRNAs), as the major template (Warner and Locarnini, 2012; Wong and Locarnini, 2017). In the nucleus of host cells, cccDNA persists in the form of viral minichromosome by integrating into the nucleosome together with histone and non-histone proteins (Allweiss and Dandri, 2017; Bock et al., 1994; Newbold et al., 1995). The chromatin-like structure of cccDNA minichromosomes is probably involved in the regulation of viral genomic transcription in the infected cells (Q. Gong et al., 2011; Shi et al., 2012).

Five major RNA species are transcribed from the viral cccDNA template by the host RNA polymerase II (Ganem and Schneider, 2001; Wong and Locarnini, 2017) (Figure 8), and are subsequently capped at the 5' terminal and polyadenylated at the 3' terminal (Warner and Locarnini, 2012). These RNAs are then transported to the cytoplasm to start translation of the viral protein products. The longest transcribed viral RNA, with 3.5 kilo-bases (kb), is called precore mRNA, and is the messenger RNA for the precore protein (HBeAg); another 3.5 kb-long transcribed RNA is the pregenomic RNA (pgRNA), which is 33 nucleotides (nt) shorter at the 5' end than precore mRNA (Warner and Locarnini, 2012). The pgRNA is responsible for the translation of the core and polymerase proteins (Pol), and after its packaging into the capsid it serves as a template for the reverse transcription of the viral genome (Locarnini and Zoulim, 2010; Warner and Locarnini, 2012; Wong and Locarnini, 2017). The three surface proteins (HBsAgs) are encoded in two subgenomic mRNAs, in which the 2.4 kb mRNA is responsible for the translation of L protein, while the 2.1 kb mRNA is responsible for

the translation of both M and S proteins (Warner and Locarnini, 2012). There is also a small 0.7 kb long mRNA which encodes the gene of hepatitis B X protein (Warner and Locarnini, 2012).

Capsid assembly and replication of the viral genome. Core protein dimers are translated from transcribed pgRNA and spontaneously assemble into capsids in the cytosol. Analysis of the serum of HBV-infected chimpanzees and patients have interestingly revealed that the vast majority of secreted particles did not contain any viral genome, and that only a small proportion of nucleocapsids could package the viral genome to form complete and infectious virions (Luckenbaugh et al., 2015; Ning et al., 2011). Though the details of capsid assembly remain unclear *in vivo*, it is speculated to be achieved through multiple pathways (Lutomski et al., 2018; 2017). Supposedly, assembly of the HBV capsid originates from a trimer of Cp dimers, of which the structure is still unknown, probably a closed ring or open structure (Figure 10) (Uetrecht et al., 2011; Zlotnick et al., 1999). Other oligomers, such as dimers, tetrameric and pentameric dimers, might also be involved in the assembly process (Holmes et al., 2015; Uetrecht et al., 2011). For nucleocapsids, the assembly is suggested to initiate with the binding of translated polymerase to the ε signal on pgRNA, and then followed by core protein dimers or oligomer of dimers continuously binding to the Pol-RNA complex until the capsid is completely formed (Bartenschlager and Schaller, 1992; Patel et al., 2017; Zlotnick et al., 1999). In this process, the minus-sense DNA, a single-stranded linear DNA (ssDNA), is reverse transcribed from pgRNA, which is degraded afterward by the RNase H activity of the viral polymerase, and finally the ssDNA is repaired to form the partially double-stranded relaxed circular DNA (rcDNA) (Bertoletti et al., 2018). Meanwhile, phosphorylation of the nucleocapsids, which is necessary for the envelopment of virus particles, appears to be coupled with single-stranded linear DNA synthesis (Warner and Locarnini, 2012). Mature nucleocapsids (NCs) can then be transported to the karyotheca to release their rcDNA into the nucleus and amplify the cccDNA pool (Wong and Locarnini, 2017).

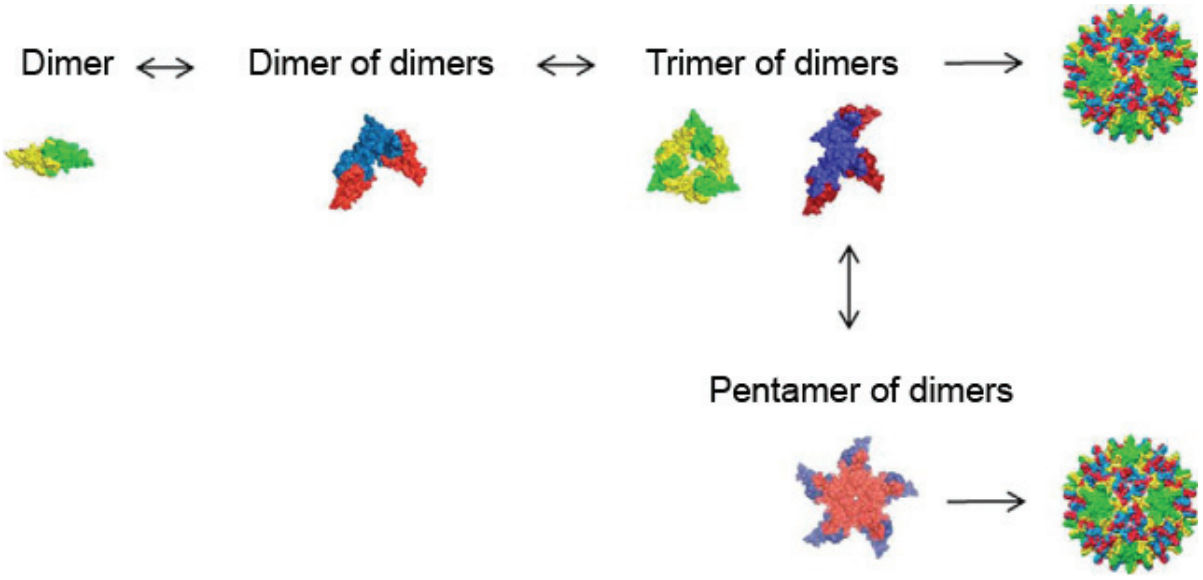


Figure 10. Possible routes of HBV capsid assembly. Capsid assembly is thought to originate from an intermediate of trimeric or pentameric dimers. The dimers or oligomeric dimers might be integrated on the assembly intermediates directly to complete the capsid, while another possibility is the oligomeric dimers should be dissolved into dimers, and then are conferred with the competence to form capsids (Holmes et al., 2015). Taken from: (Holmes et al., 2015).

Capsid envelopment and secretion of viral particles. Mature nucleocapsids can also be enveloped with surface proteins and ER-Golgi derived lipids and subsequently secreted out of the host cells (Bertoletti et al., 2018; Warner and Locarnini, 2012; Wong and Locarnini, 2017). The maturation of NCs is coupled with the replication of the viral genome, and mature, but also empty phosphorylated, NCs are enveloped and secreted, with the former representing complete virions (K. Liu and Hu, 2019; Ning et al., 2018). The release of HBV particles is suggested to progress by budding into multivesicular bodies (MVBs) of late endosomes and followed by transport through the exosomal pathway (Figure 11) (Lambert et al., 2007; Prange, 2012; Selzer and Zlotnick, 2015). Evidence for HBV particles located in the late endosome, as well as in large intracellular compartments, support this speculation (Falcón et al., 2008; Roingeard and Sureau, 1998). The empty, phosphorylated capsids are also enveloped and secreted via the same pathway (Prange, 2012; Selzer and Zlotnick, 2015), resulting in non-infectious genome-free particles, in almost 100-fold excess compared to complete virions (Hu and K. Liu, 2017).

Besides mature and incomplete HBV virions, two types of subviral particles (SVPs), constituted by the envelope proteins and forming small spherical or filamentous empty particles of a diameter from 17 to 25 nm, are secreted in the blood, in amounts up to 100,000 fold compared to HBV complete virions (Bertoletti et al., 2018; Hu and K. Liu, 2017; Warner and Locarnini, 2012; Wong and Locarnini, 2017). HBeAg, a soluble nonstructural protein dimer of HBV, can also be detected in patient's blood at high levels, and its secretion is independent of viral assembly and maturation (DiMattia et al., 2013; Warner and Locarnini, 2012; Wong and Locarnini, 2017).

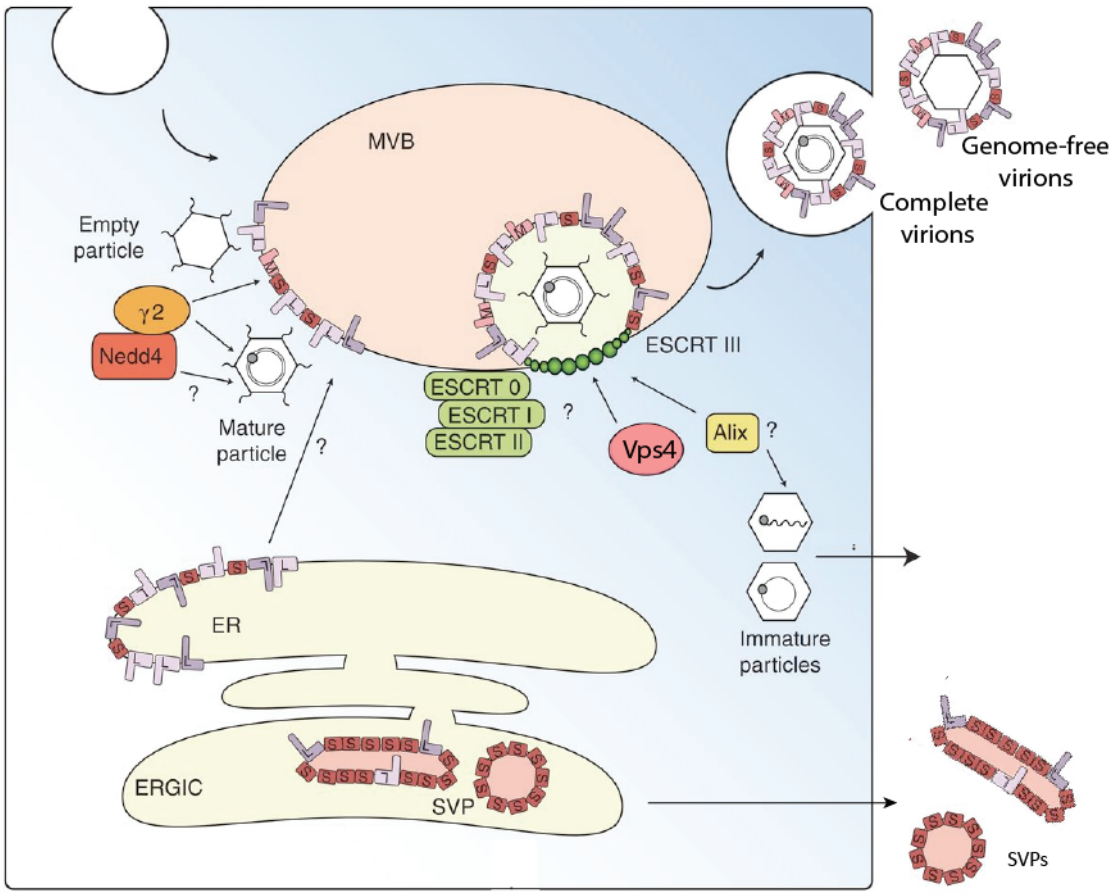


Figure 11. The secretion of HBV particles and subviral particles. MVB: multivesicular bodies. ESCRT: endosomal sorting complex required for transport. Alix: apoptosis-linked gene 2 -interacting protein X. ERGIC: ER-Golgi intermediate compartment. The synthesized surface proteins are transported from ER membrane to MVB membrane with unclear mechanism. Mature nucleocapsids and empty capsids are enveloped with surface proteins by budding into MVBs, which is mediated by the protein complexes of ESCRT-0, -I, -II, -III, and Vps4 ATPase, and the cellular Nedd4 ubiquitin ligase and c2-adaptin might be involved in this process (Prange, 2012; Selzer and Zlotnick, 2015). The enveloped particles are released via the fusion of MVB-derived

exosomes and plasma membrane (Prange, 2012). The Alix shows the ability of facilitating immature NCs to be released as nude capsids *in vitro* (Bardens et al., 2011). Differently, the SVPs are released via constitutive secretory pathway by budding into the lumen of ER or ERGIC (Prange, 2012). Adapted from: (Selzer and Zlotnick, 2015).

● Genome and proteins of HBV

➢ rcDNA

The genome of HBV, known as rcDNA, is a small relaxed-circular DNA molecule containing a complete minus-strand and an incomplete plus-strand (Figure 12) (Wong and Locarnini, 2017). The rcDNA contains 3182 to 3248 base-pairs depending on the genotype of HBV (Locarnini et al., 2013). Two direct repeats (DR) sequences of 11 nucleotides, referred as DR1 and DR2, are located at the 5' ends of both strands of rcDNA (Ganem and Schneider, 2001). The 226 base-pairs which overlap between the two DR sequences hold the rcDNA in circular construction, with a gap between the 3' and 5' ends of the minus-strand, which is bridged by the shorter plus-strand (Wong and Locarnini, 2017). The minus-strand DNA encodes the complete viral genome, and is covalently bound to the viral polymerase at its 5' ends (Wong and Locarnini, 2017). While the 5'-end of the plus-strand is capped by an 18-bases oligoribonucleotide derived from the 5' end of the pgRNA (Will et al., 1987), the 3' end of the plus-strand is at a variable position, which leads the plus-strand to cover 20-80% of the whole length genome (Wong and Locarnini, 2017). As a result, the circle of rcDNA is just partly double-stranded. Besides encoding information and regulating expression of viral genes, the rcDNA is involved in two key processes of HBV life cycle: the conversion from rcDNA to cccDNA, which generates production of viral RNAs using the host enzymes; and the reverse transcription of rcDNA from pgRNA in the viral nucleocapsid, which is important for viral maturation (Locarnini and Zoulim, 2010).

1. Open reading frames (ORFs)

The recognized coding information, including viral protein genes, promoters, enhancers and other regulatory genes, is located on the minus-strand within four overlapping ORFs (Figure 12): the longest ORF (Pol-ORF) is the gene encoding the viral polymerase. It overlaps with the other three ORFs; the envelope ORF encodes

the surface proteins (HBsAg) with the three regions preS1, preS2 and S; the precore/core ORF encodes the core protein and the N-terminally extended precore protein (HBeAg); the X ORF is the gene of the transactivating protein X (HBx) (Wen, 2004; Wong and Locarnini, 2017). The four ORFs of the HBV genome are not only controlled by their respective promoters, but also by two enhancer elements which show strong transcription activity in hepatic cells. While the enhancer I (Enhl) is located at the upstream of the X ORF and within the region of the Pol-ORF, it regulates the transcription of all four HBV ORFs (Huan and Siddiqui, 1993; H. Su and Yee, 1992; Warner and Locarnini, 2012; Yen, 1993). The enhancer II (EnhII), which is active only in hepatic cells, is located within the region of the X ORF, and was found to be involved in the regulation of basal core promoter and preS2/S promoters transcription (Warner and Locarnini, 2012; Yee, 1989; Yuh and Ting, 1993; 1990).

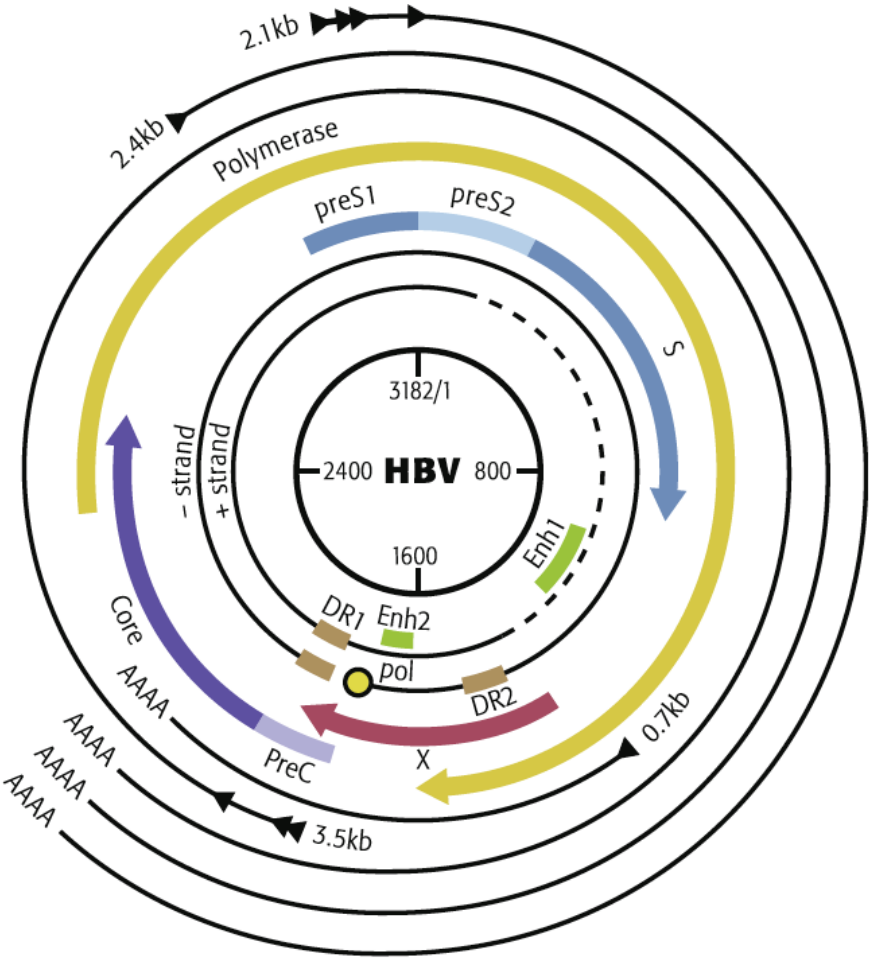


Figure 12. The genome of hepatitis B virus. The map is made based on the genome of HBV genotype D. The partially completed plus-strand DNA is shown as a solid line, supplemented by a

dashed line for the missing part. The minus-strand DNA is shown as a circular solid line with a gap next to the DR1 locus. All ORFs and regulatory elements including the direct repeat sequences and enhancers are also shown on the map. All transcribed RNAs use the same termination signal located in the core ORF, so that they are overlapped at the 3' end. The viral polymerase (shown as yellow circle) is covalently linked to the minus strand at the 5' terminal of the minus-strand DNA. Figure source: (Gish et al., 2015).

2. The generation of cccDNA

The exact mechanism of cccDNA formation has not been fully understood yet. It probably takes place in successive four events: the single-stranded gap region is repaired by the viral polymerase and a cellular polymerase (Seeger and Mason, 2015; Warner and Locarnini, 2012); the 5' terminal structures of the minus-strand and the plus-strand, including the Pol protein and the capped oligoribonucleotide, are removed by cellular endonucleases such as the tyrosyl-DNA phosphodiesterase 2 (Lucifora et al., 2017; Nassal, 2015); the short terminal redundancy on the minus-strand, which derives from viral DNA synthesis, is removed by the cellular endonucleases (Seeger and Mason, 2015; Warner and Locarnini, 2012); covalent ligation of the gaps in both strands closes the DNA molecular circle to form cccDNA (Locarnini and Zoulim, 2010; Warner and Locarnini, 2012). Once formed, the cccDNA is integrated into the nucleosome with histone and non-histone proteins and finally forms the viral minichromosome (Bock et al., 1994; Newbold et al., 1995). The core protein of HBV takes part in the formation of the minichromosome as a structural component, and can reduce the space between nucleosomes of the nucleoprotein complexes (Bock et al., 2001). The cccDNA accumulates and persists stably in the nucleus of infected cells with 30–50 copies for each cell (Newbold et al., 1995; Tuttleman et al., 1986), as a template for transcription of the viral RNAs using the cellular RNA polymerase II (Bertoletti et al., 2018). The HBx protein and the acetylation status of nucleic H3/H4 histones have been shown to play a role in the replication of HBV and the transcription of the viral genome through binding to the cccDNA minichromosome (Belloni et al., 2009; Lucifora et al., 2011; L. Luo et al., 2013; Rivière et al., 2015).

3. The reverse transcription of rcDNA

The replication of the HBV genome (Figure 13) is initiated by the formation of the replication complex, which is composed of pgRNA and the viral polymerase. Both are packaged in the immature capsid, coupling replication and the maturation process of the viral nucleocapsid (Warner and Locarnini, 2012). The transcribed pgRNA contains two stem loop structures, known as the epsilon loop (ε), at its 5′ end and 3′ end respectively. These two epsilon loops are critical for the initiation of reverse transcription (Warner and Locarnini, 2012; Wong and Locarnini, 2017). The N-terminal region of Pol binds to the epsilon loop at the 5' end of the pgRNA, which leads to a conformational change that will stimulate its enzymatic activity and initiate the viral genome replication (Bartenschlager and Schaller, 1992; J. Beck and Nassal, 2007). Then the core protein dimers bind to the pgRNA-Pol complex continuously, which is probably the main factor for selective packaging of pgRNA, to assemble immature nucleocapsids (Bartenschlager and Schaller, 1992). Coupled with the formation of viral nucleocapsids, the synthesis of the minus-strand starts with the priming of a 3 nt oligomer which is complementary to the at 5' end epsilon loop (Warner and Locarnini, 2012; Zoulim and Seeger, 1994). The Pol-oligomer primer is then translocated to the 3' end epsilon loop and the minus-strand is reverse-transcribed from the 3' end to the 5' end of pgRNA (Will et al., 1987; Wong and Locarnini, 2017). With the extension of the minus-strand, the template pgRNA is degraded by the polymerase except its 5' capped region which includes the DR1 repeat sequence (Ganem and Schneider, 2001). The residual DR1 RNA fragment is transferred to the complementary DR2 element at the 5' end of the newly synthesized minus-strand (ssDNA), and start the synthesis of the plus-strand as a primer (Will et al., 1987). The 3' end of the synthesized plus-strand fragment goes a little further than 5' end of the minus-strand, and is subsequently translocated to 3' end of the template minus-strand to continue the synthesis (Nassal, 2015). The plus-strand synthesis terminates when it extends to about 50 to 70% of the viral genome-length, with a yet unclear mechanism, and the partially double-stranded rcDNA is subsequently formed (Wong and Locarnini, 2017).

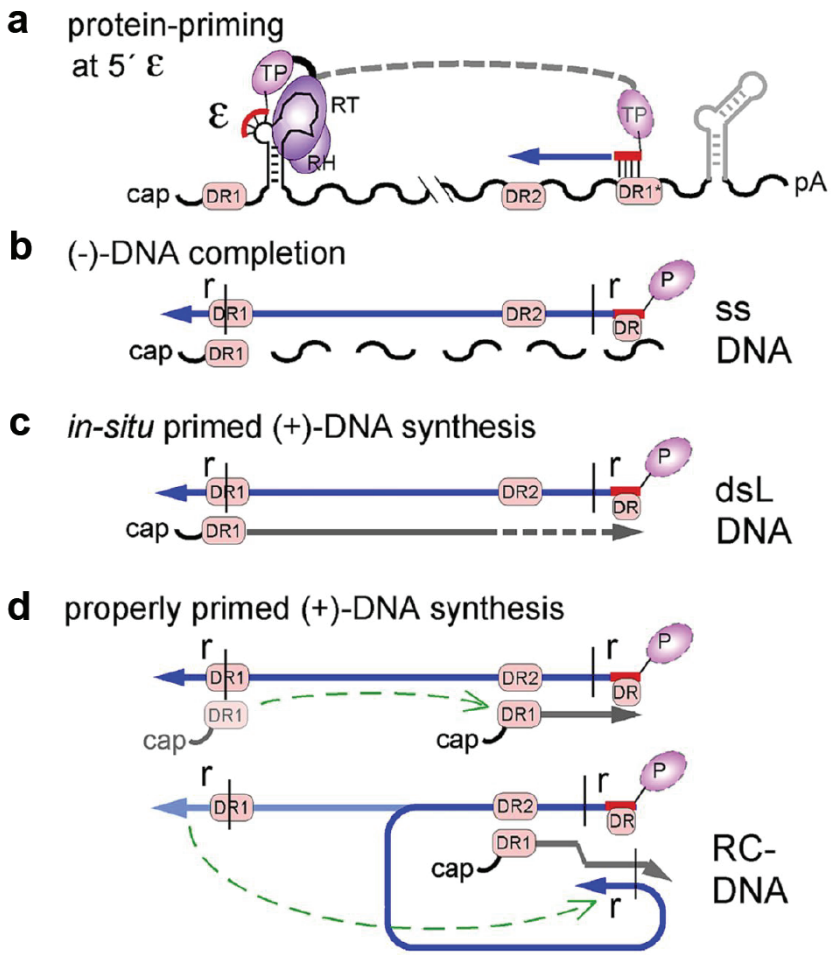


Figure 13. The replication of rcDNA. The black wavy line represents the pgRNA template for reverse transcription, which is longer than the viral genome length of HBV with a cap and poly A (pA) tail. The epsilon loops (ε) of pgRNA are shown with hairpin structures. The newly synthesized viral DNA strands are shown with blue (minus-strand) or grey (plus-strand) arrow lines. a): the initiation of reverse transcription. b): the elongation of minus-strand DNA (ssDNA). c): the synthesis of double-stranded linear DNA (dsL DNA). d): the synthesis of short plus-strand DNA and the formation of partly double-stranded circles of rcDNA. Figure source：(Nassal, 2015).

4. dsL DNA

Besides rcDNA, another replication product, double-stranded linear DNA (dsL DNA), is also found in infected cells with a frequency of about 5–20% compared with rcDNA (Christoph Seeger et al., 2013; W. Yang and Summers, 1999). It is derived from an aberration in the rcDNA synthesis process, when the residual DR1 RNA fragment primes the plus-strand synthesis directly from the 3' end of the newly synthesized minus-strand, instead of being translocated to DR2 element at the 5' end (Figure 13-c)

(Nassal, 2015). It can be inaccurately repaired to a cccDNA-like circular molecule due to a host nonhomologous end-joining pathway, so that the circular molecule is usually functionally defective (Nassal, 2015). The dsL DNA can also be integrated into the host chromosomal genome (W. Yang and Summers, 1999). The integration neither is managed by viral genes nor has any role in HBV replication, but it might take part in the formation of HBV-associated hepatocellular carcinoma (Sung et al., 2012).

- **HBsAg**

The HBV surface antigen (HBsAg) was first isolated by Baruch Blumberg, who won a Nobel prize for this discovery, in the 1960s (Blumberg and Alter, 1965). HBsAg can be detected in the serum of newly infected patients in the first several weeks after infection. It is the main component of the HBV envelop and its secretion correlates closely with viral replication, so that it not only is the first viral marker discovered, but also remains the major clinical hallmark for HBV infection so far (Krugman et al., 1979; Maurizia Rossana Brunetto, 2010). HBsAg includes 3 forms of HBV surface proteins, called S (small), M (medium), and L (large) protein, respectively. They are translated by host ribosomes at the rough-surfaced endoplasmic reticulum (ER), and span the ER membrane with four transmembrane domains resulting in their complex topology structure (Figure 14) (Bruss et al., 1994; Ostapchuk et al., 1994; Prange and Streeck, 1995). Only a minor proportion of HBsAg proteins can be assembled into the envelop of viral particles and secreted as HBV virions, and correspondingly, the majority of HBsAg proteins are secreted from the host cell in the form of self-assembled spherical or filamentous subviral particles (SVPs) (Heermann et al., 1984). SVPs are the most abundant particles in the serum of HBV patients, with a ratio up to 100,000 to 1 compared to mature virions. Though the SVPs are non-infectious, they are suspected to adsorb virus-neutralizing antibodies with such a huge quantity, which might promote the spread of HBV virions and contribute to persistent infection of HBV patients (Bertoletti et al., 2018; Warner and Locarnini, 2012). It is worth mentioning that the maturation of hepatitis D virus (HDV), which co-infects about 5 % of HBV patients as a satellite virus (WHO, 2017), also requires the participation of HBV surface proteins (Abbas and Afzal, 2013; Negro, 2014).

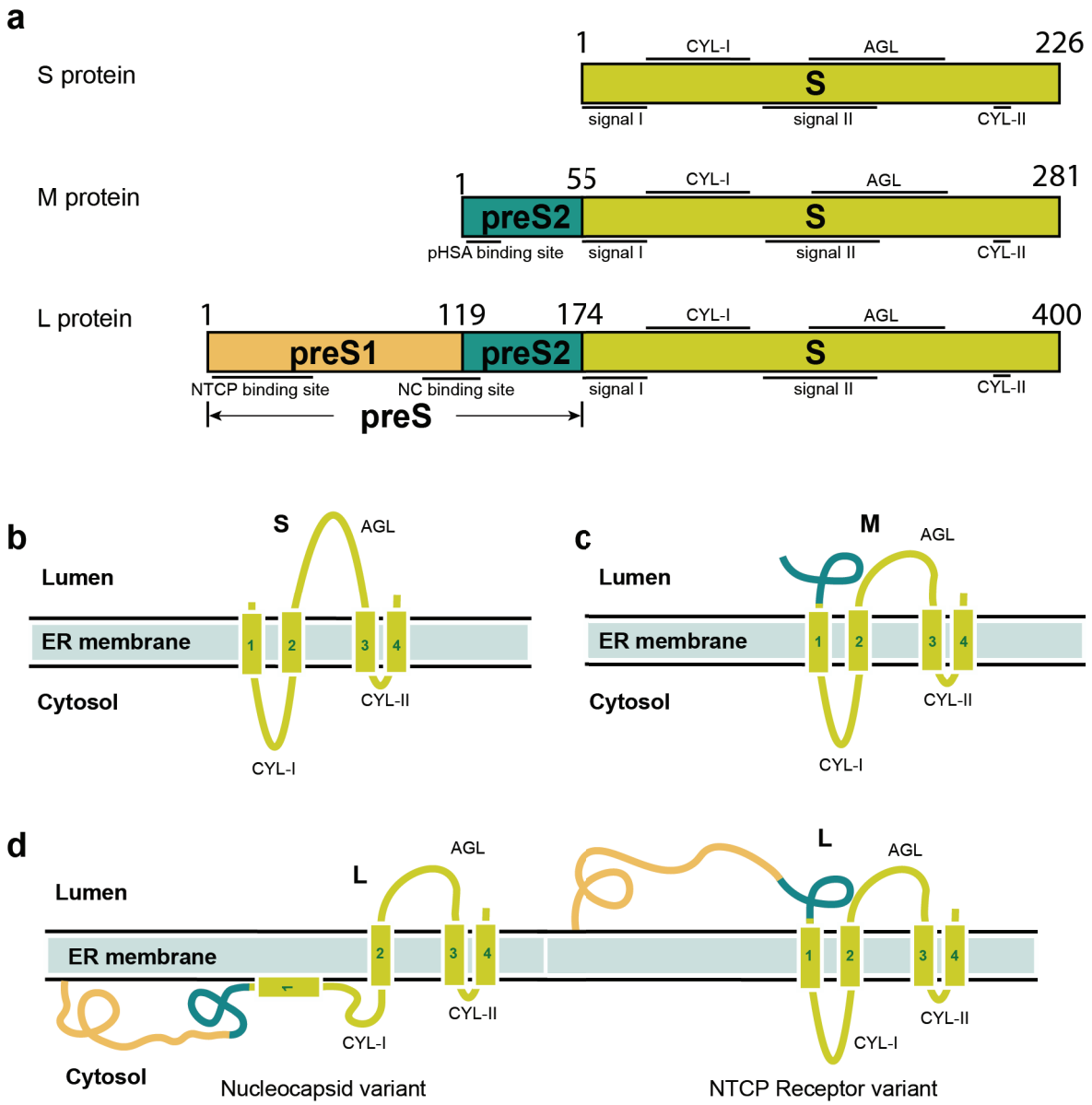


Figure 14. Topology structure of HBsAg proteins. a): structural domains and signal sequences of S, M and L proteins. The S, preS2, preS1 are shown with *chartreuse*, *dark green*, *orange* respectively. Black lines are used to indicate specific regions of HBsAgs, including: transmembrane signal I and II; binding sites of polymerized human serum albumin (pHSA), sodium taurocholate co-transporting polypeptide (NTCP), Nucleocapsids (NC); conformational antigenic loop (AGL); cytosolic loops (CYL-I and II). b-d): the topology models of the HBsAg proteins on the endoplasmic reticulum (ER) membrane. HBsAg proteins are fixed on the ER membrane by four embedded transmembrane domains (TM1-4), which are connected by the antigenic loop (AGL) and two cytosolic loops (CYL). Two variants of L protein exist, important in the envelopment of mature capsid (left) and in the context of cell entry (right) respectively.

1. S protein

The S protein, which consists of 226 amino acid residues (Figure 14), is the shortest but most plentiful surface protein. It has four putative transmembrane domains (TM1-4) spanning the ER membrane, and can assemble spontaneously with ER lipids in the host cell (Simon et al., 1988). There is a N-terminal transmembrane signal (signal I) located within the first 30 amino acids of S protein, and an internal transmembrane signal (signal II) located inside the region from residue 84 to residue 136 (Eble et al., 1987). Signals I and II serve as signal sequences to translocate the HBsAg proteins to the ER membrane, but they will not be cleaved by host signal peptidases (Eble et al., 1987). While acting as the predominant component of all viral and subviral particles (Maurizia Rossana Brunetto, 2010), the S protein is responsible for the major antigenic determinants of HBsAg (Carman et al., 1990). The antigenic determinant is a conformational antigenic loop (AGL) with several cross-linked cysteine residues on the domain from the residues 99 to 169, and responsible for all serotypes of HBV (Wong and Locarnini, 2017). The antigenic determinant is highly conserved probably because it has a low affinity binding site for HSPG on hepatocytes surface, which is important for the HBV virions entry (Sureau and Salisse, 2013).

2. M protein

The M protein can be found in all HBV particles as a minor component, but it is not essential for the replication, secretion, and infection of HBV (Bruss and Ganem, 1991; Bruss and Vieluf, 1995; Fernholz et al., 1991). The M protein shares the whole sequence of the S protein at the C-terminal, and has an additional domain with 55 amino acids residues at its N-terminal (Figure 14), which is known as preS2 (Ganem and Schneider, 2001). Moreover, the topological structure of M protein is predicted to be similar to the S protein, except the additional preS2 domain, which locates in the ER lumen (Eble et al., 1990). The preS2 domain has a major antigenic epitope at its central region (Wong and Locarnini, 2017). The part of preS2 from residues 3 to 16 is the region which binds to host polymerized human serum albumin (pHSA) (Kann and Gerlich, 2005), but the effect of this interaction is still not clear.

3. L protein

The L protein is essential for the formation of filamentous SVPs and for the envelopment of HBV capsids, as it constitutes 10-20 % of all surface proteins in both of these particles, while it only accounts for 2 % of the surface proteins in the spherical particles (Maurizia Rossana Brunetto, 2010). Compared to the M protein, it has an additional preS1 domain at its N-terminus, which consists of 108 or 119 additional amino acids depending on the genotype (Figure 14) (Wong and Locarnini, 2017). Together, preS1 and preS2 are called the preS domain. Due to lack of transmembrane signal, the synthesized preS domain, as well as the TM1 domain, initially stays on the cytosolic side, while the hydrophobic C-terminus of the L protein spans the ER membrane with help of the signal II (Bruss and Vieluf, 1995; Urban et al., 2014). The preS domain on the cytosolic side takes part in the envelopment of HBV by interacting with viral capsids and is subsequently enclosed within viral particles (Urban et al., 2014). The region involved in the interaction with the nucleocapsids (NC) has been located to residue 103 - 127, which consists of the last 17 amino acids of preS1 and the first 8 amino acids of preS2 (Bruss, 1997; Le Seyec et al., 1998; Poisson et al., 1997). About half of the L molecules undergo complex post-translational changes in the topology, leading to the translocation of preS domains from the cytosolic side to the luminal space of ER, which exposes the perS domains to the surface of enveloped viral particles (Bruss, 2007; Bruss et al., 1994; Lambert and Prange, 2001; Persing et al., 1987; Seitz et al., 2016). Myristoylation is not necessary for viral replication and secretion (Bruss and Ganem, 1991), but is important for the infectivity of HBV virions (Gripon et al., 1995). The N-terminal sequence of preS1 from myristoylated residue 2 to residue 48 represents the specific binding site for NTCP receptor on hepatocyte surface and is therefore involved in the attachment and entry of HBV (Wong and Locarnini, 2017). The sequence from residues 49 to 75 is also important for HBV infectivity, but it is not yet understood why (Urban et al., 2014). Moreover, there are some antigenic sites in the preS region, which are responsible for the immune responses in the HBV patient, can be recognized by host B cells and T cells (Wong and Locarnini, 2017).

➢ Core protein

1. The sequence and function of core protein

The core protein (Cp) of HBV, which is referred to as HBV core-antigen (HBcAg) in serology, is one of the two translational products of pgRNA (Wong and Locarnini, 2017). It is usually constituted of 183 amino acids (aa) with a molecular weight of around 21 kDa, except for genotype A (185 aa) and for genotype G (195 aa) (Locarnini et al., 2013; Warner and Locarnini, 2012). Within the HBV particles, 90 or 120 Cp dimers form a T=3 or T=4 icosahedral protein shell. Cp183 contains a N-terminal domain (NTD) of 140 aa, which is responsible for the formation of the capsid structure; a linker of 9 aa; and a C-terminal domain (CTD) of 34 aa, which is required for the packaging of pgRNA and the release of the viral genome in the nucleus (Figure 15) (S. Liu et al., 2010; Petit and Pillot, 1985; Rabe et al., 2003). The NTD is competent to form the capsid-shell structure alone when it is over-expressed in *E. coli* or under high ionic strength conditions *in vitro* (Wingfield et al., 1995; Zlotnick et al., 1996), while a recent study has shown that the CTD is required for capsid assembly in a mammalian cell-free system, which is closer to physiological conditions (Ludgate et al., 2016). The CTD is rich in arginines, which confers it abundant positive charges. It contains seven conserved potential phosphorylation sites on serines or threonines (Heger-Stevic et al., 2018; Kann and Gerlich, 2005; J. C. Y. Wang et al., 2012; Yeh and Ou, 1991). Phosphorylation and dephosphorylation events can regulate the charges of the CTD, contributing to viral genome packaging and release (Heger-Stevic et al., 2018; Ludgate et al., 2016; Ning et al., 2017; Zhao et al., 2018). The 9 aa linker was thought to have no role for Cp function, but recent study indicates that it might participate in multiple events of HBV life cycle, including pgRNA packaging, viral DNA synthesis, phosphorylation of the CTD and virions secretion (K. Liu et al., 2018). Besides being the structural component of the viral capsid, Cp is also involved in the generation of the cccDNA minichromosome, and probably takes part in the modulation of genome transcription (Allweiss and Dandri, 2017; Bock et al., 2001).

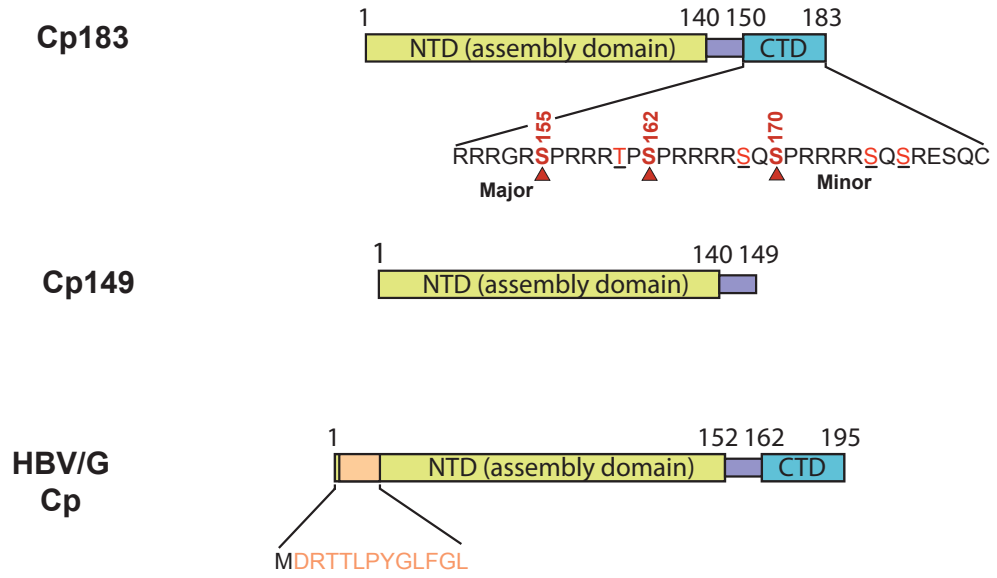


Figure 15. Core proteins of HBV. The N-terminal domain of Cp is shown in yellow, the linker in purple, and the CTD in blue with its detailed sequence. The potential phosphorylation sites of the CTD are emphasized in red, and marked with triangles for the three major sites which have been confirmed to be crucial for viral genome packaging (Machida et al., 1991; Yeh and Ou, 1991). Cp149 is an artificially truncated variant of the full-length protein (Cp183), in which the CTD is absent. Cp of genotype G, referred to as HBV/G Cp, contains an additional N-terminal sequence of 12 residues (orange) replacing the initial methionine (K. Li et al., 2007).

2. The structure of Cp

The Cp shows dimeric behavior when it is analyzed by gel filtration or assembled into capsids as the basic unit (Crowther et al., 1994; Nassal et al., 1992; Wingfield et al., 1995; S. Zhou and Standring, 1992). The NTD is an α-helix-rich region (Figure 16-a and b) (Bottcher et al., 1997; Conway et al., 1997; Wynne et al., 1999), and the dimers form spikes composed of two α3 and two α4 helices of the NTD (Figure 16-c) (Wynne et al., 1999; Yu et al., 2013). While residue C48 is located at the spike bottom and found in the reduced state, an intradimer disulfide crosslink can be formed by the C61-C61 within the helices spike (Wynne et al., 1999), which is necessary neither for formation of the dimer, nor assembly and envelopment of the capsid (Nassal, 1992; Nassal et al., 1992; Yu et al., 2013). However, this disulfide bond can reduce the stability of the capsid, and might thus be important in the uncoating and release of the

packaged genome (Selzer et al., 2014). In addition, the last residue (C183) can also cross-link different Cp dimers, which is considered to be helpful to capsid stability (Ning et al., 2018). The structure of the CTD is missing from the cryo-EM maps due to its high flexibility and disordered nature (Böttcher and Nassal, 2018; Yu et al., 2013).

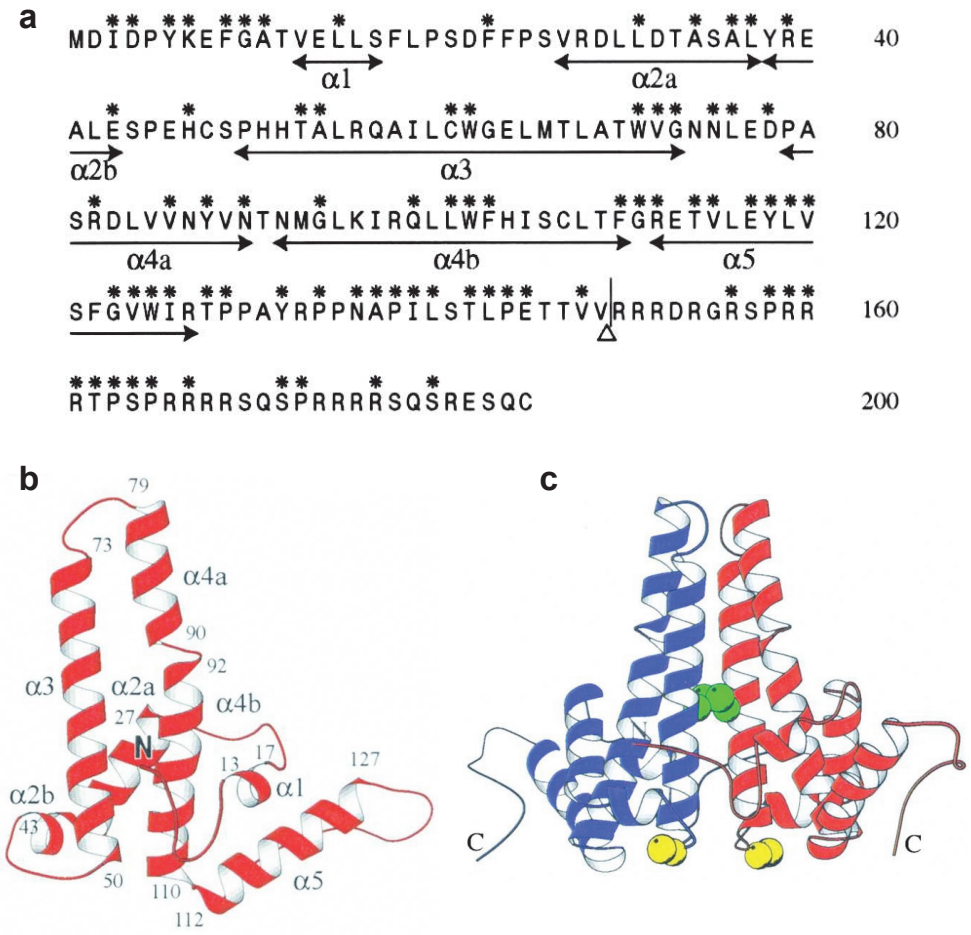


Figure 16. The sequence and crystal structure of Cp. a): Sequence of Cp with the localization of the 5 α helices. The Δ indicates the truncated position of Cp149. Fully conserved residues are labelled with stars. Cp is mainly constituted by helices: α1 helix from residue 13 to 17; α2 helix from residue 27 to 43; α3 helix from residue 50 to 72; α4 helix from residue 79 to 110; and α5 helix from residue 112 to 127. The α2 and α4 helices are slightly bent between residues of 37-38 or 90-92, resulting in 2 pairs of sub-helices, α2a/b or α4a/b, respectively. b): Representation of the monomer of Cp149, with helices indicated. The first 12 residues at the N-terminal and residues 128-142/3 are irregular, and the last several residues (143/4-149) are missing in the structure. c): Crystal structure of the Cp149 dimer. The C48 and C61 are identified in yellow and green respectively. Adapted from: (Wynne et al., 1999)

- **HBeAg**

The HBeAg is encoded in the same ORF (precore/core ORF) and shares most of the amino acid sequence with the core protein. Still, processing and physiological function are quite different (Warner and Locarnini, 2012). In contrast to Cp, the HBeAg protein is translated from the 3.5 kb precore mRNA instead of the pgRNA, and undergoes a series of post-translational modification in the ER (Figure 17) (Wong and Locarnini, 2017). Moreover, it does not take part in viral replication and secretion, acting as an accessory protein for HBV infection rather than a structural protein for viral particles (Warner and Locarnini, 2012). The function of HBeAg in HBV infection is still not completely clear, but it is believed that it can affect the immune response system of patient and can contribute to the persistent infection of HBV (M. Chen et al., 2005; M. T. Chen et al., 2004; Hadziyannis, 2001; Milich and Liang, 2003; Milich et al., 1998; Visvanathan et al., 2007). Despite the HBeAg-negative infections caused by the HBV precore mutant (Alexopoulou and Karayiannis, 2014; Brunetto et al., 1989), the secretion of HBeAg generally correlates with the high levels of HBV virions replication (P. Chen et al., 2017; Liang and Ghany, 2002). Therefore, the HBeAg is widely used to monitor the viral replication in clinical diagnosis and treatment together with other serological markers (HBsAg, HBV DNA, anti-HBc, etc) (Bonino et al., 2010; Krugman et al., 1979; Schmilovitz-Weiss et al., 1993).

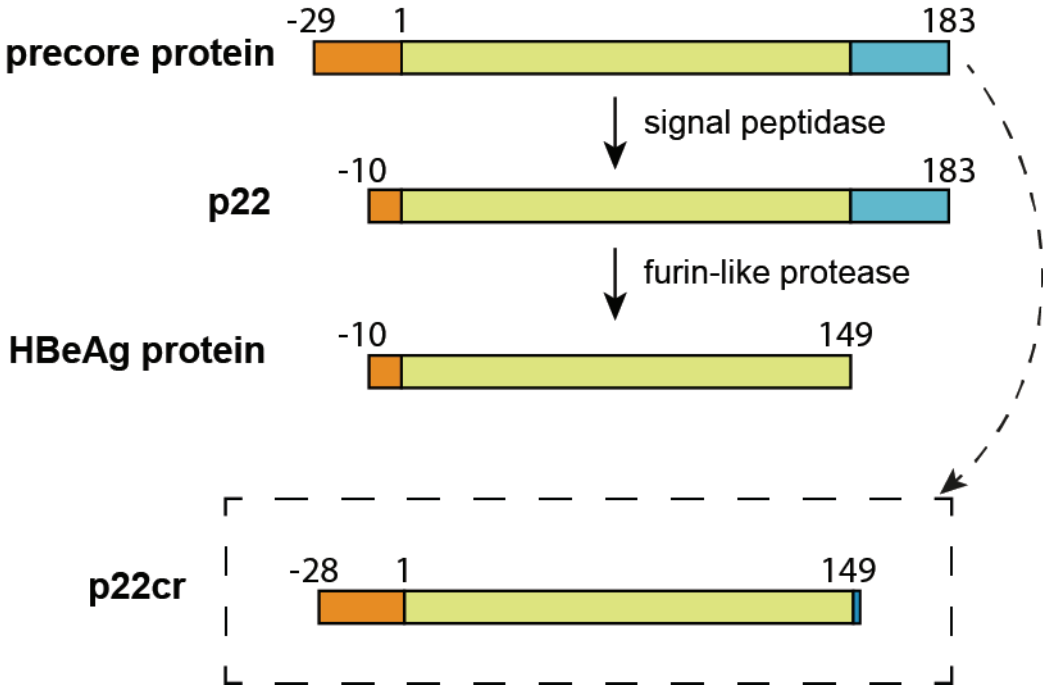


Figure 17. Generation of HBeAg. Compared with Cp, initially synthesized HBeAg protein (precore protein) has a 29 amino acid residues surplus at its N-terminus, of which the first 19 residues act as a signal peptide to translocate the precore protein into the ER lumen, where it is subsequently cleaved by the host signal peptidase (Dienes et al., 1995; Warner and Locarnini, 2012). The generation of mature HBeAg protein needs a second cleavage by a furin-like protease to remove the CTD region of the intermediate type (p22) (Messageot et al., 2003; Standring et al., 1988; Takahashi et al., 1983; Walsh and Locarnini, 2012), with a variable cleavage site at the core protein region from residues 149 to 154 (Eren et al., 2018; Messageot et al., 2003; Takahashi et al., 2000; Walsh and Locarnini, 2012; Watts et al., 2011). The p22cr is derived from the precore protein, with most of the N-terminal additional sequence and several residues of CTD remaining.

The HBeAg secreted in patient sera is in dimeric form with an intramolecular C(-7)-C61 disulfide bond, whose structure is significantly different from the one of Cp149 dimer (Figure 18) (DiMattia et al., 2013; Zlotnick et al., 2015). Although the tertiary structures of HBeAg and Cp are similar, the monomer subunit of HBeAg dimer rotates by 140°, compared with quaternary structure Cp dimer (DiMattia et al., 2013). HBeAg in its reduced state has been shown to form capsids which look similar to T=3 capsid of Cp149 but less regular, and which are facilitated by high ionic strength *in vitro* (DiMattia et al., 2013; Watts et al., 2011). Moreover, p22 was reported to form genome-free hybrid capsids collaborating with Cps (Scaglioni et al., 1997), yet these hybrid capsids were recently found to have the ability to package DNA (Z. Yan et al., 2019).

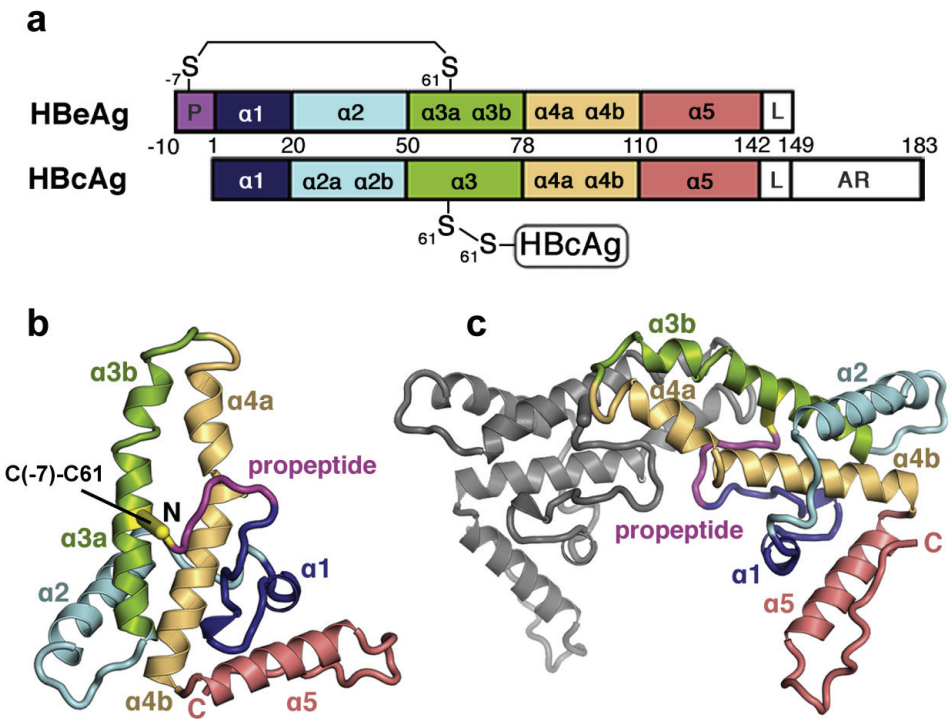


Figure 18. Sequence and structure of HBeAg. a): HBeAg contains similar helices as Cp but contains an additional N-terminal domain and no CTD. An intramolecular disulfide between Cysteine (-7) and 61 can be formed in HBeAg monomer (b), which results in the assembly-inactive structure of HBeAg dimer (c). Taken from: (DiMattia et al., 2013).

A derivative of the precore protein, which consists of the N-terminal signal peptide, assembly domain and maybe a few amino acids of CTD of core protein, referred as p22cr in some publications, is also reported to be detected in the sera of HBV patient (Figure 17) (E.-Q. Chen et al., 2017; Gou et al., 2017; Kimura et al., 2005; Maasoumy et al., 2015). It was shown that p22cr, rather than the core protein, could be the component of empty capsid in genome-free virions (Kimura et al., 2005). However, no further studies were published to support this statement. A recent study confirmed that the entire CTD can be detected in empty HBV virions from patient sera, questioning the previous report about p22cr genome-free capsid which might have been caused by detection mistakes (Hu and K. Liu, 2017; Ning et al., 2017). Nevertheless, as a new member of core-related antigens (HBcrAgs, including HBcAg and HBeAg), the p22cr has been regarded as a new potential marker for HBV diagnosis and treatment (Yuen et al., 2018).

- **Polymerase protein**

The polymerase protein (Pol) is the biggest protein encoded by HBV with a molecular weight of 90 kDa (Wong and Locarnini, 2017). This enzyme comprises four major domains with distinct functions (Figure 19-a): a terminal protein (TP) domain, a spacer domain, a reverse transcriptase domain and RNase H domain. The N-terminal TP domain is a functional domain containing a conserved tyrosine residue (Y63) which binds to the first nucleotide of the minus-strand DNA to initiate reverse transcription together with the 5' end ε loop of pgRNA (Figure 19-b) (Jones and Hu, 2013a; Lanford et al., 1997), and it is also involved in triggering nucleocapsid assembly (Bartenschlager and Schaller, 1992; Jones et al., 2014). The spacer domain serves as a linker, but no other exact function has been found except that 3 conserved Cys residues at its C-terminal might be involved in pgRNA packaging (S. Kim et al., 2009; Lamontagne et al., 2016; Wong and Locarnini, 2017). The third part of Pol, a typical reverse transcriptase (RT) domain, which shares a conserved YMDD sequence with the RT of retroviruses, is responsible for the synthesis of viral genome using the templates of pgRNA and new minus-strand DNA (Figure 19-b) (Menéndez-Arias et al., 2014). The C-terminal domain, a RNase H activity (RH) region, is involved in the degradation of pgRNA during reverse transcription (Figure 19-b), which is important for the synthesis of the plus-strand DNA (Champoux and Schultz, 2009; Ko et al., 2014; Wong and Locarnini, 2017).

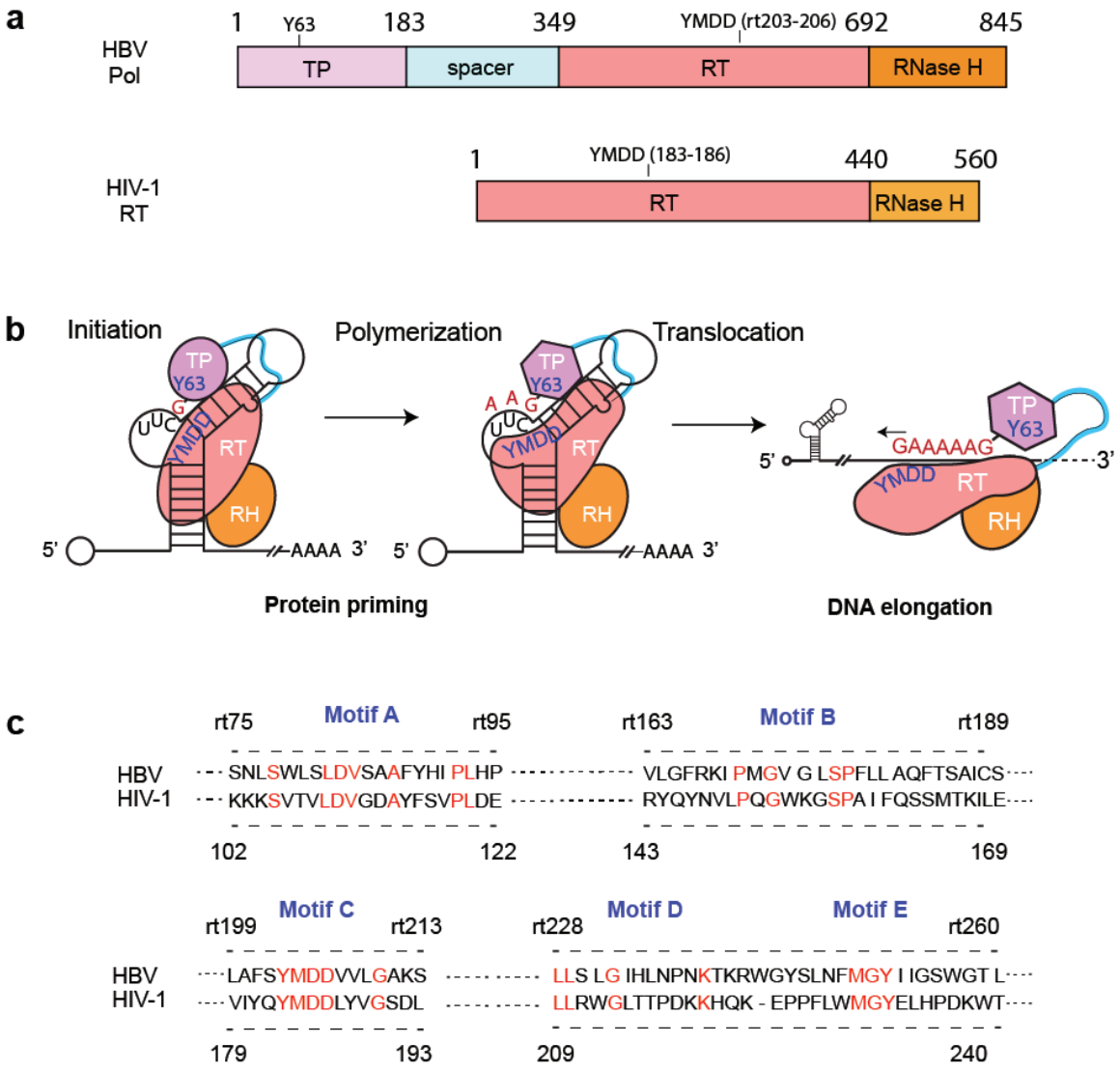


Figure 19. The HBV polymerase and HIV-1 reverse transcriptase. a): structural domains of the HBV polymerase (Pol) and HIV-1 reverse transcriptase (RT). b): priming and elongation of the minus-strand DNA. Pol primes reverse transcription by repairing the nucleotide (G) linked by residue Y63 to the internal bulge of 5' end ε loop of pgRNA (left). Two following nucleotides (A) are linked to the initiating nucleotide (G) by the activity of the YMDD sequence (mid). The complex of Pol and the GAA primer is translocated to 3' of pgRNA with a topological structure change of Pol, and the minus-strand DNA elongates, while pgRNA is degraded (right). c): main conserved motifs in the RT domain of HBV Pol and HIV RT. Adapted from: (Jones and Hu, 2013b; Menéndez-Arias et al., 2014)

Despite the vital role played by Pol in HBV infection, no high-resolution structure has been obtained yet. However, as Pol contains high sequence homology with the HIV reverse transcriptase, with up to 20 % similarity for the RT domain and more than 30%

for the RNase H domain (Figure 19-c) (Das et al., 2001; Tavis et al., 2013), the well-studied HIV reverse transcriptase is usually taken as an ideal model for HBV Pol protein study (Clark and Hu, 2015; Menéndez-Arias et al., 2014). Moreover, most of the nucleos(t)ide analogues (NAs), which are widely used in HBV therapy, can also be used as antiviral agents targeting the HIV reverse transcriptase (Menéndez-Arias et al., 2014; Warner and Locarnini, 2012).

- **HBx protein**

HBx is composed of 154 amino acids and is a nonstructural protein conserved for all orthohepadnaviruses members (Wong and Locarnini, 2017). While it is dispensable for duck HBV (Meier et al., 2003) or absent in other avian hepatitis viruses (Funk et al., 2007), the HBx protein has been confirmed to be vital for WHV replication (H. S. Chen et al., 1993; Zoulim et al., 1994). While initially, the role of HBx in viral replication was controversial (Blum et al., 1992; Melegari et al., 1998; Yaginuma et al., 1987), a more recent research demonstrated that cell types can interfere with the relation between HBx and HBV replication (D.-Y. Gong et al., 2013). The studies on animal models have revealed that HBx is required by and can distinctly heighten HBV replication as a regulatory protein (Keasler et al., 2009; Slagle and Bouchard, 2016; Tsuge et al., 2010; Xu et al., 2002; Zhang et al., 2018). Though the multiple functions of HBx have not been totally understood, a lot of studies show that it is not only directly involved in viral replication, but also modulates the cellular metabolism to benefit viral replication and persistence (Slagle and Bouchard, 2018; Wong and Locarnini, 2017). For example, HBx was shown to take part in the following events: methylation of cccDNA (Lee et al., 2019); exportation or degradation of host restriction factors (R. Chen et al., 2017; Decorsière et al., 2016; Gao et al., 2017; X. Luo et al., 2016; van de Klundert et al., 2016); interaction with transcription promoting factor (Kouwaki et al., 2016); activation of specific cellular signal pathway (Rawat and Bouchard, 2015); modulation of the intracellular calcium level (Casciano et al., 2017; Geng et al., 2012) and host gene expression (Sengupta et al., 2017).

● Maturation and envelopment of viral Capsids

➢ The capsid arrangement

The capsid of HBV is an icosahedral superstructure formed by large numbers of Cp dimers (Wynne et al., 1999; Yu et al., 2013; S. Zhou and Standring, 1992; Zlotnick et al., 1996). Two types of capsids are identified: the dominant type T=4 and the minor type T=3, which contribute to about 95% and 5% of the HBV particles in patient serum respectively (Crowther et al., 1994; Zlotnick et al., 2015). Each dimer in the capsid can be distinguished into two adjoining subunits according to their quasi-equivalent environments (Figure 20) (Conway et al., 2003; Zlotnick et al., 1996). The larger T=4 capsid consists of 120 copies of Cp dimers with 4 types of subunits (A, B, C, D), whereas the smaller T=3 capsid has 3 only types of subunits (A, B, C) and is formed with 90 copies of Cp dimers (Crowther et al., 1994; Zlotnick et al., 1996).

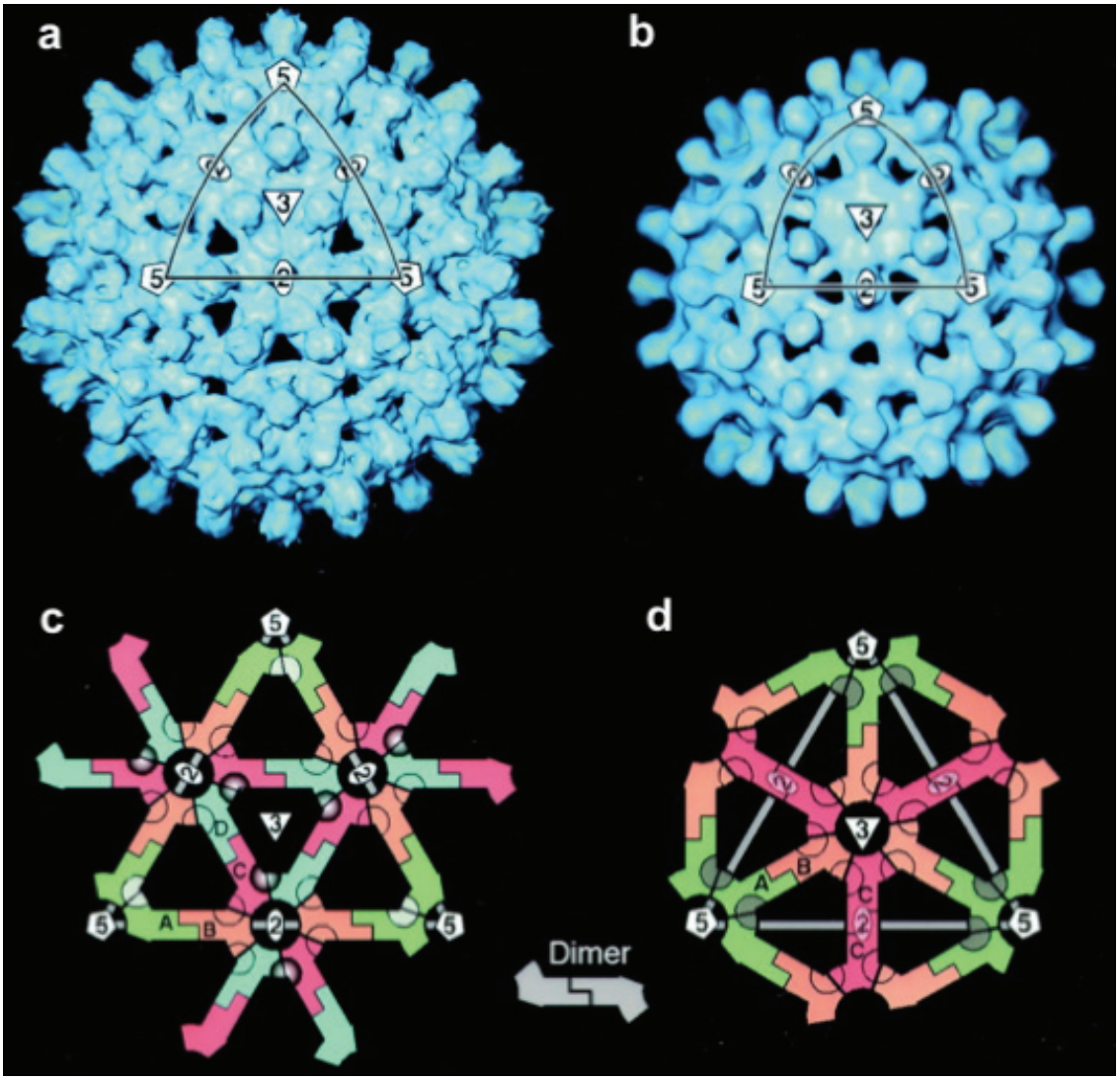


Figure 20. HBV capsids construction and the arrangement of subunits. The T=4 capsid contains 60 AB dimers and 60 CD dimers, while the T=3 capsid is formed by 60 AB dimers and 30 CC dimers. Both types of capsid are porous (as shown in panels *a* and *b*), with the interspace between dimers accounting for about 16 % of the surface. The subunit arrangements of T=4 and T=3 capsids are shown in panels *c* and *d*, and the subunits of A, B, C, D are marked by *green*, *orange*, *pink*, *indigo* color respectively. The symmetry axes of 2, 3, 5-fold on capsids are marked with 2, 3, 5. Taken from: (Conway et al., 2003)

The fold structures of the seven subunits in T=4/3 capsid are very similar, especially the spike structure of the four types of dimers (AB and CD of T=4 capsid, AB and CC of T=3 capsid) which looks almost identical according to cryo-electron microscopy, X-ray, and solid-state NMR studies (Lecoq et al., 2018; Wynne et al., 1999; Yu et al., 2013; Zlotnick et al., 1996). The main differences between subunits locate in the N-

terminal and C-terminal of the NTD and residues of the linker, which are the regions involving adjacent dimers (Lecoq et al., 2018; Wynne et al., 1999; Yu et al., 2013).

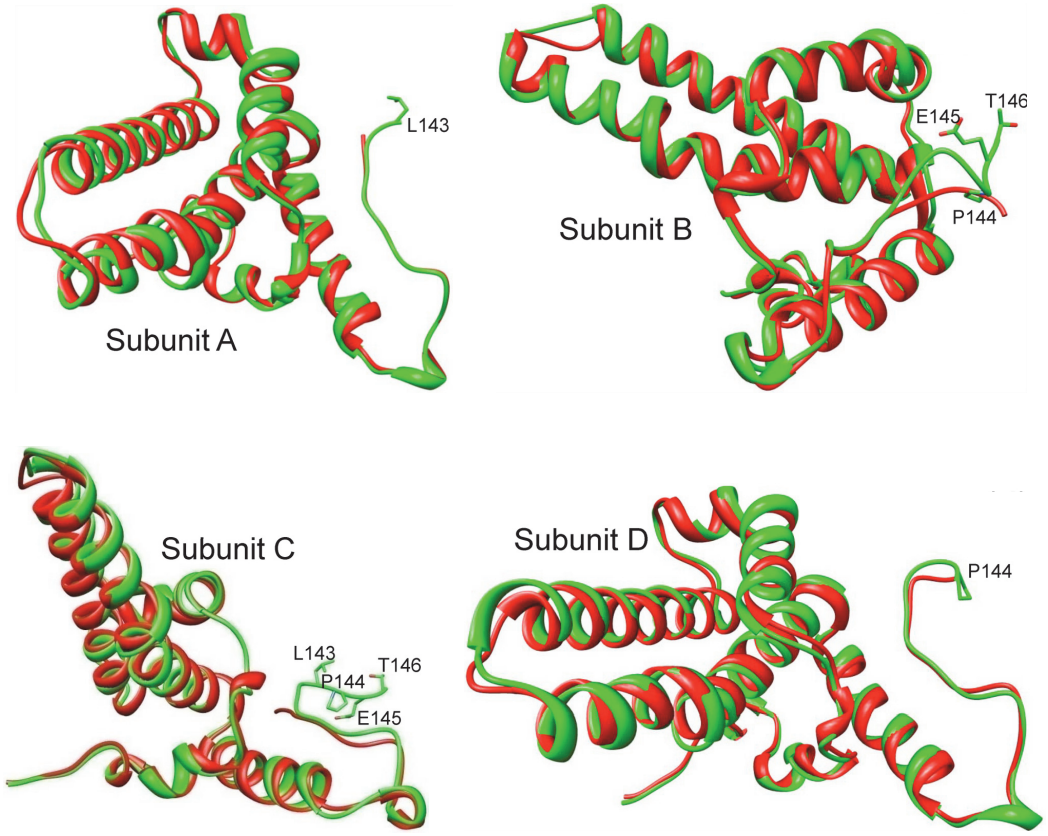


Figure 21. The asymmetric subunits structures of T4 capsid. In red is the crystal structures of the first 142/3 residues of Cp149. In green is the cryo-EM structure of the full-length Cp, with only 143 residues for subunit A, 146 residues for B and C, 144 residues for D. Adapted from : (Yu et al., 2013).

The cryo-EM structure of HBV capsid was solved at 3.5 Å resolution on the full-length Cp183 and revealed a double-layer model for HBV capsid: besides the ordered outer layer formed by the NTD and the linker, an inner layer is constituted by the CTD and packaged RNAs (Figure 22-a) (Roseman et al., 2005a; Yu et al., 2013). Around the 2-fold and 5-fold axes holes, the residues of the linker are found to be responsible for the connection between the two layers (Roseman et al., 2005a; Yu et al., 2013). Moreover, even if the CTD remains mostly invisible in structural studies, it was shown to be partially exposed through the 3-fold axes related holes (Figure 22-b) (Yu et al.,

2013), which is different with the previous speculation that the 2-fold axe hole is responsible for CTD tails exposure (C. Chen et al., 2011; Wynne et al., 1999). This visual evidence is supported by biochemical studies about the exposure of CTD (Gallina et al., 1989; J. Kim and J. Wu, 2015; 2014; Meng et al., 2011; Seifer and Standring, 1994).

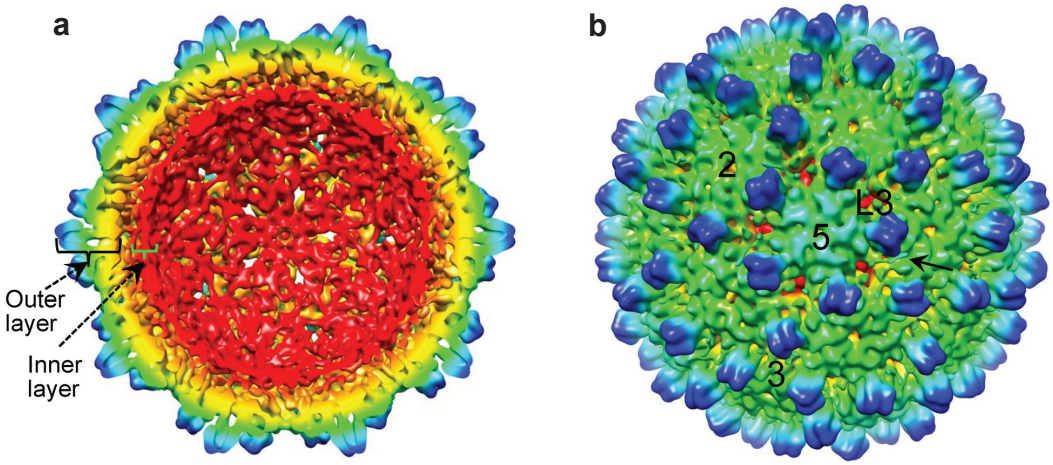


Figure 22. The double-layer capsid model reconstructed by cryo-EM. The inside view (a) and outside view (b) of radially colored capsid structure. The spikes and outer layer of capsid is in *blue* and *green*. The inner layer formed by CTD and RNA is *red*. The transition area between the two layers is in *yellow*. The 2, 3/L3, 5 refer to the holes of 2-fold, 3-fold related and 5-fole axe respectively. Adapted from : (Yu et al., 2013)

The exposure of the CTD is certainly transient due to its highly dynamic characteristics (C. Chen et al., 2011). Recent studies have shown that, compared to empty capsid, the transient exposure of CTD is restricted by the packaging nucleic acids (J. Kim and J. Wu, 2015; Meng et al., 2011). In RNA-filled capsids, the 10 % of residues which do not interact with RNA are part of the exposed tails of CTD (J. Kim and J. Wu, 2015). These tails are mainly composed of residues 170 to 175, and could contain the binding motif for serine-arginine protein kinase (SRPK) (J. Kim and J. Wu, 2014), which might contribute to the phosphorylation of CTD *in vivo* (Daub et al., 2002; Zheng et al., 2005). While a significant exposure of the CTD was observed in RNA-filled capsid, the mature capsid filled with rcDNA seems to restrict almost the whole CTD inside the capsid (Figure 23) (Meng et al., 2011).

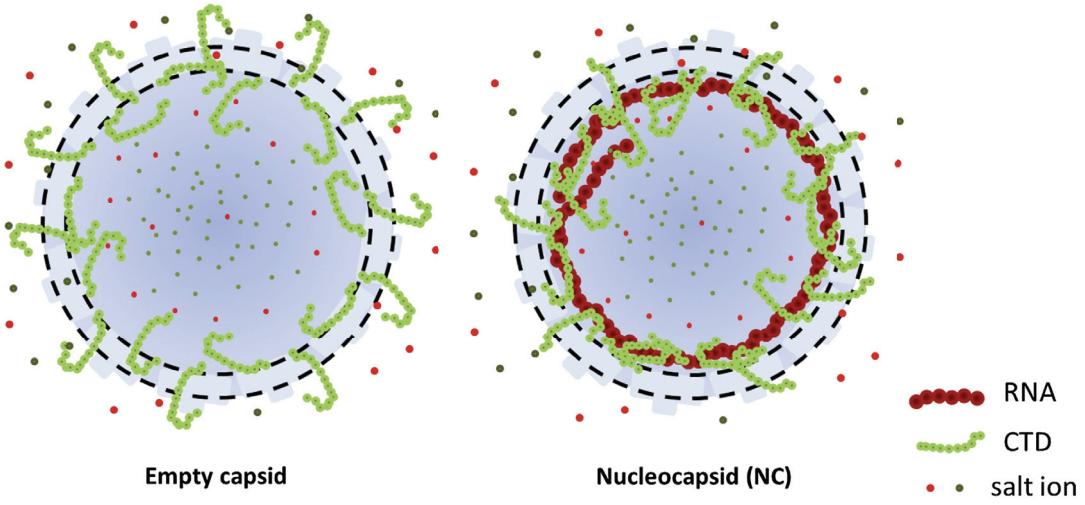


Figure 23. The CTD exposure diagram of empty capsid and RNA-filled capsid. Compared with the empty casid (left), the nucleocapsid packaging RNAs (right) exposes less CTD residues. Adapted from (J. Kim and J. Wu, 2014)

➢ The formation of capsid

The assembly of capsid *in vitro* is widely established as a three-phase event: first the initiation with a trimer complex of Cp dimers; then the elongation by adding subsequent dimers to the initiate complex nucleus; and the final completion with the insertion of last subunit (Endres and Zlotnick, 2002; Hagan, 2014; Hagan and Elrad, 2010; Lutomski et al., 2017; H. D. Nguyen et al., 2007; Zandi et al., 2006; Zlotnick et al., 1999). The weak interaction between dimers would be responsible for this process (Ceres et al., 2004; Zlotnick, 2007; 2003), and the assembly intermediates could be trapped by enhancing the association energy with high concentration of salt (Pierson et al., 2014; Zlotnick, 2007). Using charge detection mass spectrometry (CDMS), a series of intermediates, mainly composed of 104-117 dimers, have been detected and identified (Figure 24) (Pierson et al., 2014). These intermediates are considered to be in low free energy states, which are harder to gain or loss dimers (Pierson et al., 2014).

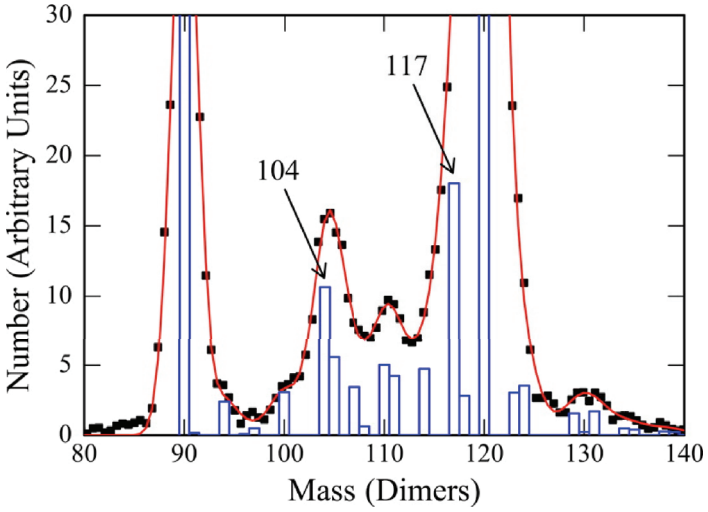


Figure 24. Distribution of assembly intermediates. Analysis based on the CDMS data. The mass is shown as the number of dimers. The black squares, which indicate the quantity of measured intermediates, are fitted by a red curve. The blue histograms indicate the intensities of intermediates. It is worth noting that no intermediates were detected for T=3 capsids. Adopted from: (Pierson et al., 2014).

Additional cryo-EM investigations confirm that these intermediates are not aberrant structures, but incomplete capsids, which represent different assembly states or pathways (Lutomski et al., 2018; Pierson et al., 2014) (Figure 25).

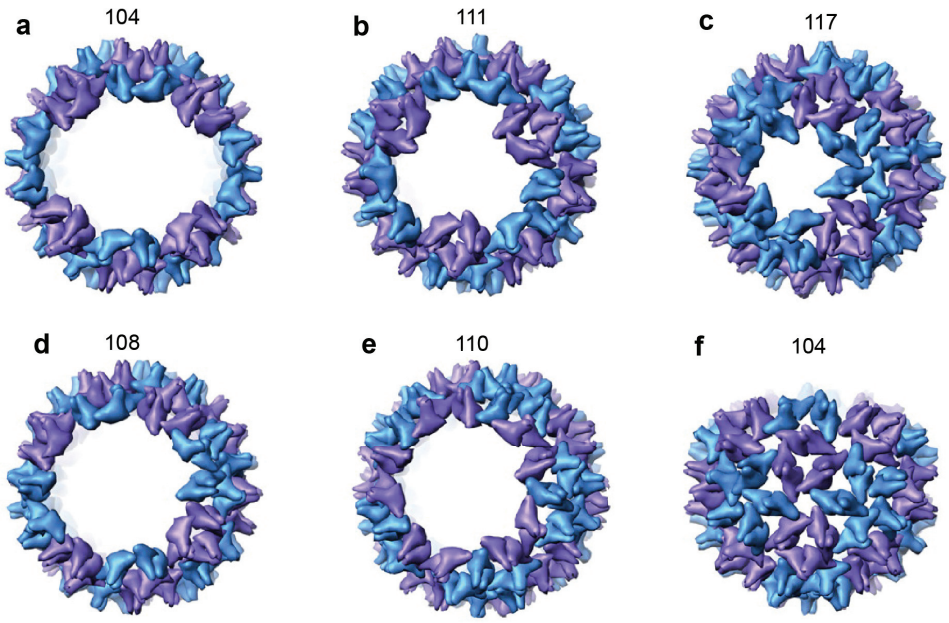


Figure 25. Structure models for assembly intermediates of HBV capsid. Dimer numbers of incomplete capsids are shown above the intermediate structural models. AB and CD dimers are

shown with blue and purple blocks respectively. The intermediate types of 104-, 111-, 117- dimers (a-c), of 108-dimers (d), and of 110-dimers (e) appear to follow different assembly pathways. The side view of 104-dimers is shown in (f).

However, more recent findings reveal that the final completed capsid might not result from the last subunit insertion, but correction of imperfect capsids (Lutomski et al., 2017). These imperfect capsids, which have been reported as aberrant intermediates in the previous study (Pierson et al., 2014), are overgrown with more than 120 dimers, and then slowly corrected (Figure 26) (Lutomski et al., 2017; Uetrecht et al., 2008). This finding is consistent with the opinion that imperfect capsids are less stable than perfect capsid (Singh and Zlotnick, 2003; J. C.-Y. Wang et al., 2015), and suggests a probable proof-reading mechanism of capsid assembly (Lutomski et al., 2017).

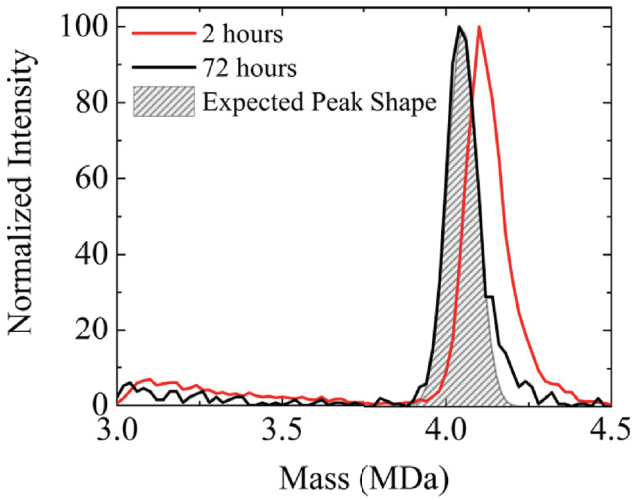


Figure 26. Time dependent correction of overgrown capsids. The assembly reaction measured within 2 h (red) and 72 h (black) are shown in overlaid spectra. As time goes on, the peak of overgrown capsid turns into the peak located in the expected position of normal capsids. Figure taken from (Lutomski et al., 2017).

However, the conditions inside hepatocytes are highly complex, and many of the host factors, such as Hsp90 (Shim et al., 2011), nucleophosmin B23 (Jeong et al., 2014), filamin B (Y. Li et al., 2018), protein kinase A and C (H. Kang et al., 2008), are probably involved in the capsids formation, leading to a poor understanding of this process *in vivo* (Nassal, 2008; Selzer and Zlotnick, 2015).

➢ The CTD phosphorylation balance

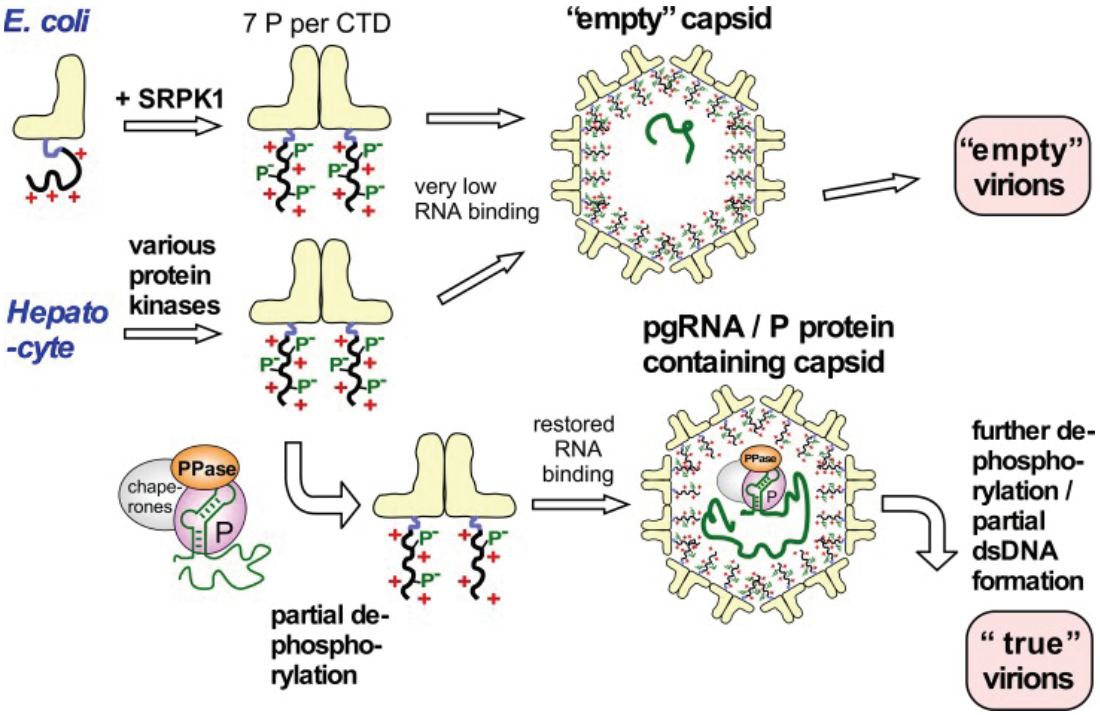


Figure 27. Schematic representation of Cp assembly in *E. coli* and hepatocytes. The red pluses indicate the positive charges on CTD, while the green P^- indicates the negative charge of phosphorylated residues. The RNAs are shown with green curves. Figure taken from (Heger-Stevic et al., 2018)

Despite the fact that capsids filled with nonspecific RNA can be harvested from bacteria or produced *in vitro* (C. Chen et al., 2011; Porterfield et al., 2010; Wingfield et al., 1995), the capsids harvested from human cells are assembled without any non-specific RNA (Ning et al., 2011). It is speculated that there might be a molecular chaperone which interacts with the CTD to block the nonspecific packaging *in vivo* (C. Chen et al., 2011). It has been shown that binding of unphosphorylated CTD to SRPK can suppress capsid assembly and prohibit RNA binding, and the follow-up phosphorylation of CTD looses binding gradually (C. Chen et al., 2011; Heger-Stevic et al., 2018; Zheng et al., 2005). The full phosphorylation of the CTD was also recently reported to block specific and nonspecific RNA binding, which indicates that partial dephosphorylation could occur prior to pgRNA packaging (Heger-Stevic et al., 2018; Zhao et al., 2018). The Pol-pgRNA complex, which interacts with phosphorylated CTD,

is speculated to replace the bound SRPK to activate he capsid assembly, which leads to the specific packaging of pgRNA *in vivo* (C. Chen et al., 2011; Kann and Gerlich, 1994; Ludgate et al., 2016; Patel et al., 2017). The replication of rcDNA is thought to correlate with dephosphorylation of the CTD to balance the charges on the nucleocapsid (Perlman et al., 2005; P.-Y. Su et al., 2016). The partial synthesis of viral rcDNA might be a result of the timely envelopment of nucleocapsid, due to the termination of dNTPs supply (Heger-Stevic et al., 2018). Similarly, the uncoating of naked viral core in the infected cells might be caused by the restart of DNA synthesis and the re-phosphorylation of CTD, which disturbed the electrostatic homeostasis of nucleocapsids (Heger-Stevic et al., 2018; P.-Y. Su et al., 2016). Given that hundreds of protein kinases and phosphatases are encoded by human genome (Manning et al., 2002), it is likely that more than one of them take part in the CTD phosphorylation (Selzer and Zlotnick, 2015). However, how does the CTD phosphorylation and dephosphorylation happen *in vivo* remains to be further elucidated (Heger-Stevic et al., 2018).

- **Capsid envelopment**

Envelopment of HBV mature capsids to form virions depends on the interaction of the core protein with surface proteins, especially the L protein (Ning et al., 2018). On the Cp side, a region known as the matrix binding domain (MBD), including the spike tip area mentioned above and the residues around a hydrophobic pocket of Cp dimers, are predicted as important for envelopment (Le Pogam et al., 2000; Ponsel and Bruss, 2003; Roseman et al., 2005a). The residues of the MBD are discontinuous but exposed on the capsid surface, and mutating or blocking MBD could affect the virions packaging distinctly (Ning et al., 2018; Orabi et al., 2015). On the envelope side, a region of the L protein, which spans the preS1 and preS2 regions with 22 aa, and referred to as the matrix domain (MD), has been identified to be important for HBV assembly (Bruss, 2007; 1997; Schittl and Bruss, 2014). In addition, residues on the cytosolic loop I of S region are also considered to be related with capsid envelopment (Biswas et al., 2013; Löffler-Mary et al., 2000; Poisson, 1997).

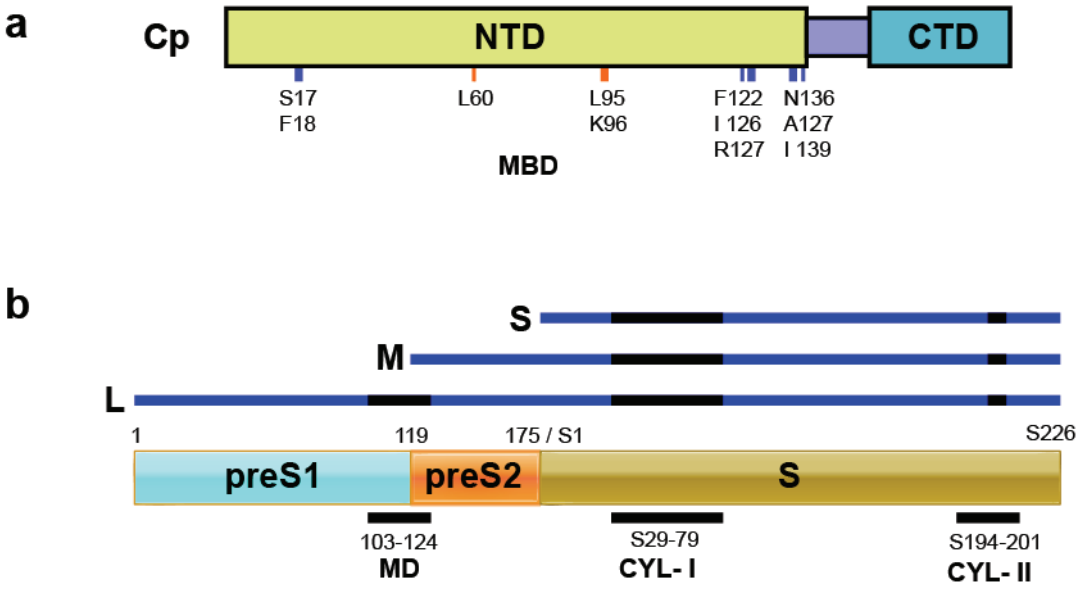


Figure 28. The matrix domain and the matrix binding domain. a) The matrix binding domain (MBD) of Cp is located on the NTD with discrete segments. L60, L95 and K96 from two Cp monomers form a ring-like groove surrounding the base of spike. S17, F18, F122, I126, R127, N136, A137 and I139 lie close to the pores of capsid. b) The matrix domain (MD) is located on the preS domain of L, from R103 to S124. The MBD of Cp and MD of L are considered to be responsible for envelopment of HBV capsid. The cytosolic loop I (CYL-I), which locates at the S domain from residues 29 to 79, is probably involved in HBV capsid envelopment. Besides, the cytosolic loop II (CYL-II) of S protein from residues 194 to 201 has been revealed to be critical for HDV maturation (Blanchet and Sureau, 2006; Komla-Soukha and Sureau, 2006). Figure is modified from: (Ning et al., 2018)

A cryo-EM study has demonstrated that the types of HBV capsids (T=3 and T=4) are not related to envelopment and secretion of complete virions and empty virions (Dryden et al., 2006). Furthermore, cryo-EM structures of viral particles purified from a serum of a chronically infected patient revealed that the envelop is in contact with the charged spike of the capsid *via* electrostatic interactions and with a high degree of plasticity (Seitz et al., 2007). Two types of compact particles, referred as A/B-map, and a sort of gapped particle, referred as C-map, are distinguished by cryo-EM (Figure 29), but they all share a similar packing mechanism, with one asymmetric unit corresponding to up to 3 surface proteins (Seitz et al., 2007).

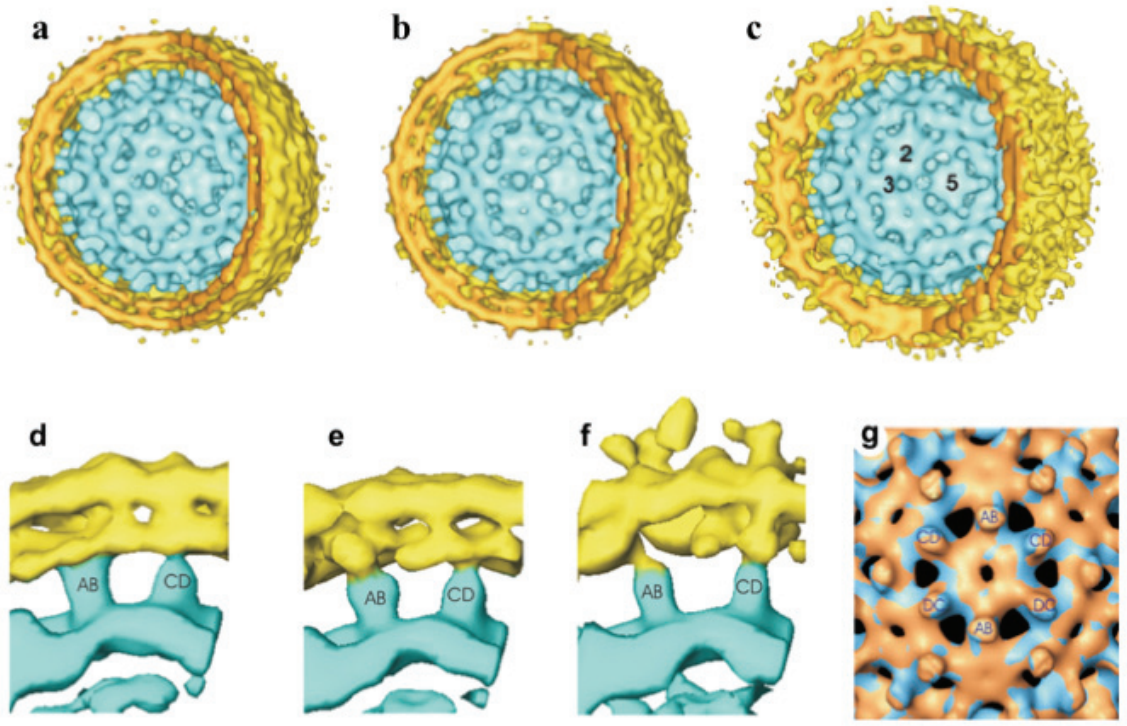


Figure 29. The morphological phenotypes of HBV particles. A-map (a), B-map (b) and C-map (c) phenotypes of HBV particles are shown in the sectional view maps, in which the envelopes and the inside T=4 capsids are distinguished with yellow or blue respectively. The contact of AB and CD spike with envelope is shown in side view, with d, e, f corresponding to A-, B- and C-map. The distance between the inner surface of envelope and the surface of the core is variable: A-map < B-map < C-map. The superposition of capsids (g) displays a displacement between CD-pikes of the A-map (blue) and B-map (orange). Taken from: (Seitz et al., 2007).

The difference between the 3 types of particles is not only in position and size of contact area, but also in the inside of the capsid structure (Figure 29-g) (Seitz et al., 2007). For the B-map and C-map particles, the position of capsid spikes is consistent with the previous obtained crystal structure (Wynne et al., 1999), but the CD-spikes shift about 4 to 6 Å in the A-map particles (Seitz et al., 2007). This is probably due to the space between core and envelope which is crowded in the A-map particle and therefore gives more conformational stress to the capsid compared to the B-map and C-map particles (Seitz et al., 2007). In all types of particles, the surface proteins bind to the tips of AB-spikes intimately, as well as the tips of CD spikes with relatively small area, around the charged residues D78 of capsid subunits (Seitz et al., 2007).

➢ The maturation hypotheses of HBV capsid

In secreted HBV virions, both mature nucleocapsids and empty capsids can be found with significantly different level (Hu and K. Liu, 2017). The density of complete HBV virions, containing rcDNA within the capsid, is about 10^9/ml in HBV patient serum, while the genome-free virions, containing phosphorylated capsids devoid of RNA or DNA, can be detected up to 10^{11}/ml (Luckenbaugh et al., 2015; Ning et al., 2011) (Figure 30). In addition, the secretion of RNA virions, containing pgRNA within the capsid, was confirmed to be relatively rare (10^6/ml) (Jansen et al., 2016; J. Wang et al., 2016). Finally, a frequent natural mutation I97L (or F97L) in the *adr* subtype leads to another type of immature virions containing ssDNA (Yuan et al., 1999a; 1999b). Therefore it is conjectured that there are probably some signals exposed on the capsid surface responsible for the selective envelopment and secretion (Hu and K. Liu, 2017; Ning et al., 2018; 2017).

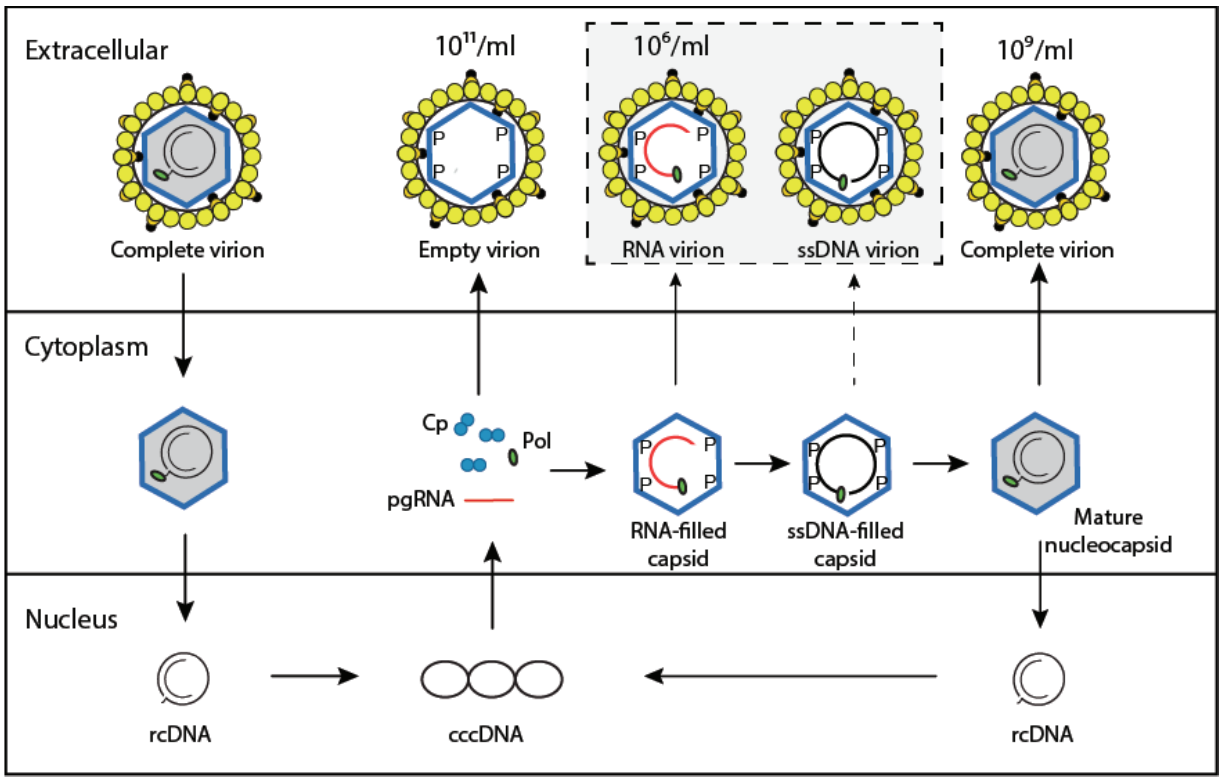


Figure 30. Selective secretion of HBV virions. The secretion virions of HBV are mainly composed of empty and complete particles. While the RNA-filled virions only account for 0.01‰ of the total secreted virions, and the ssDNA-filled virions are just detected in the specific HBV types with Cp mutation of I97L or F97L.

1. The classic maturation signal hypothesis

Early studies surmised that the elongation of the plus-strand DNA induces a structural change which indicates the maturation of nucleocapsid, resulting to the envelopment and secretion (Seeger and Hu, 1997; Summers and Mason, 1982). Later, structural differences have been detected between the spikes of empty Cp149 and RNA-containing Cp183 capsids produced in *E. coli* cells and anlayzed by cryo-EM (Roseman et al., 2005b). The differences are located in the hydrophobic pocket of capsid spike (Roseman et al., 2005b), of which several residues are considered to be important for virion maturation (Chua et al., 2003; Le Pogam et al., 2000; Ponsel and Bruss, 2003; Yuan et al., 1999b). They suggested a model, consistent with the previous hypothesis, that these structural changes are related to the rcDNA synthesis and might be a positive signal for nucleocapsid maturation (Figure 31-a) (Roseman et al., 2005b). Meanwhile, another report, based on the close association of phosphorylation with pgRNA packaging and the nonphosphorylation state of complete virions, postulated that the dephosphorylation of nucleocapsids might trigger envelopment and secretion (Perlman et al., 2005). Yet, all conjectures cannot accurately explain the massive secretion of empty virions, independent from viral genome replication or dephosphorylation of the capsid (Ning et al., 2017). Moreover, recent research has shown that, despite the fact that phosphorylation and dephosphorylation events in the nucleocapsid are related to its maturation, they are not correlated with virion secretion (Ning et al., 2017).

2. The single strand blocking signal hypothesis.

Whereas the maturation model presuming a positive signal is not able to explain the secretion of empty virions, a single strand (ss) blocking model is proposed for the negative secretion of pgRNA virions and ssDNA virions (Ning et al., 2011). Two key events occur in this model: the packaging of any ss nucleic acid, including both pgRNA and ssDNA, would trigger a blocking signal in the immature nucleocapsids to prohibit their envelopment, whereas the empty capsid would not possess this signal; then the removal of pgRNA or ssDNA, rather than the rcDNA synthesis, would induce the structural change required for the capsid envelopment concomitantly with the

elimination of the negative signal (Figure 31-b) (Ning et al., 2011). In addition, it is predicted that the empty capsid likely shares the same structural characteristic than the mature nucleocapsid (Ning et al., 2011). This model is able to justify the rare secretion or absence of RNA-filled and ssDNA-filled virions, as well as the large number secretion of empty and complete virions. However, the detailed mechanism, including the ss blocking signal and the uncertain structural change for envelopment, is still unclear (Ning et al., 2018).

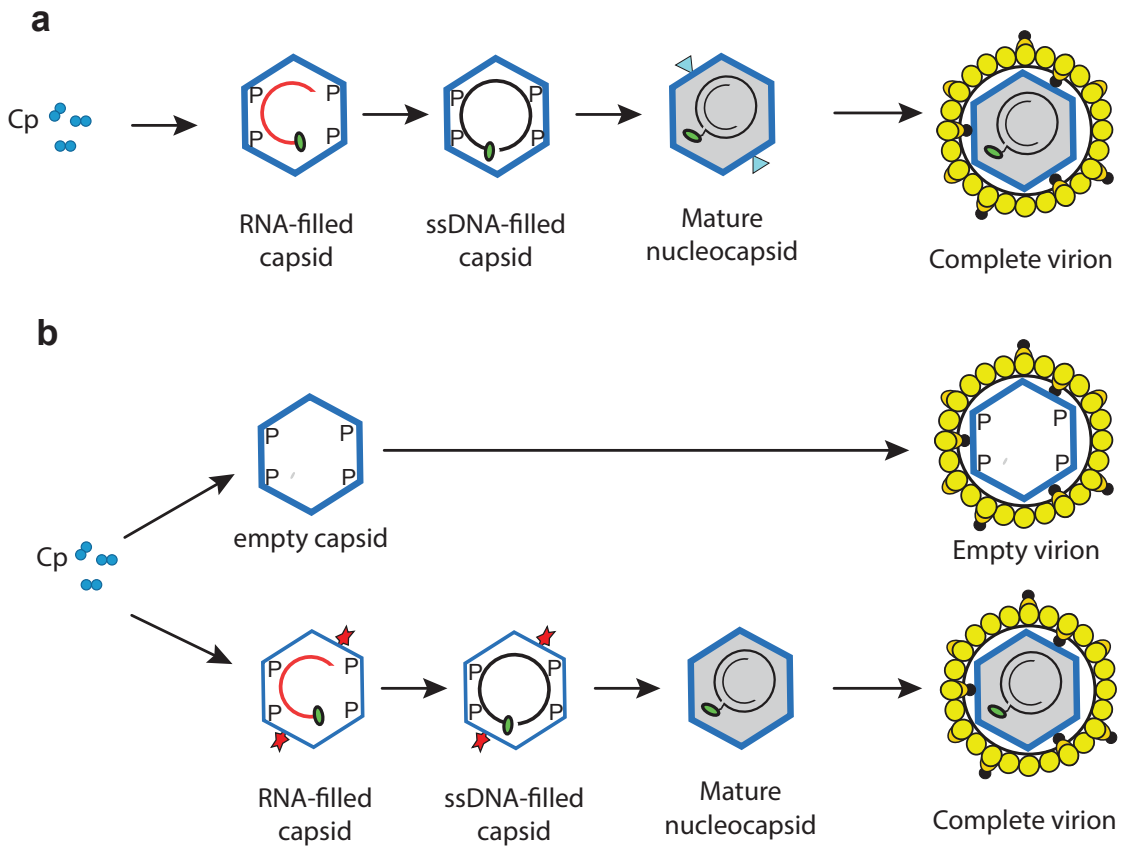


Figure 31. Maturation models for HBV secretion. a) The classic maturation signal model: the complete rcDNA replication or the dephosphorylation of the CTD induces a positive signal (cyan arrows) for the maturation of nucleocapsid. b) The single strand blocking model: a negative signal is displayed on the immature nucleocapsids surface and subsequently removed along with the double strand rcDNA synthesis.

3. The “two-signal” regulation hypothesis

A “two-signal” model, supported by a series of indirect biological evidences, was proposed recently, giving more details on the HBV particle secretion (K. Liu and Hu, 2019; Ning et al., 2018). In this hypothesis, there might be two positive secretion signals on both empty capsids and nucleocapsids, and a negative blocking signal on the immature nucleocapsids (Figure 32). The secretion signal I, which probably interacts with the S protein, is responsible for the basal secretion of empty virions, and is inhibited by the blocking signal on immature capsid. The secretion signal II, which is vital for complete virions secretion and just acts as an enhancer for empty virions secretion, is surmised to locate at the MBD region of Cp and interact with the MD sequence of L protein. The blocking signal is only displayed on the surface of immature nucleocapsids, induced by the packaging of pgRNA or ssDNA.

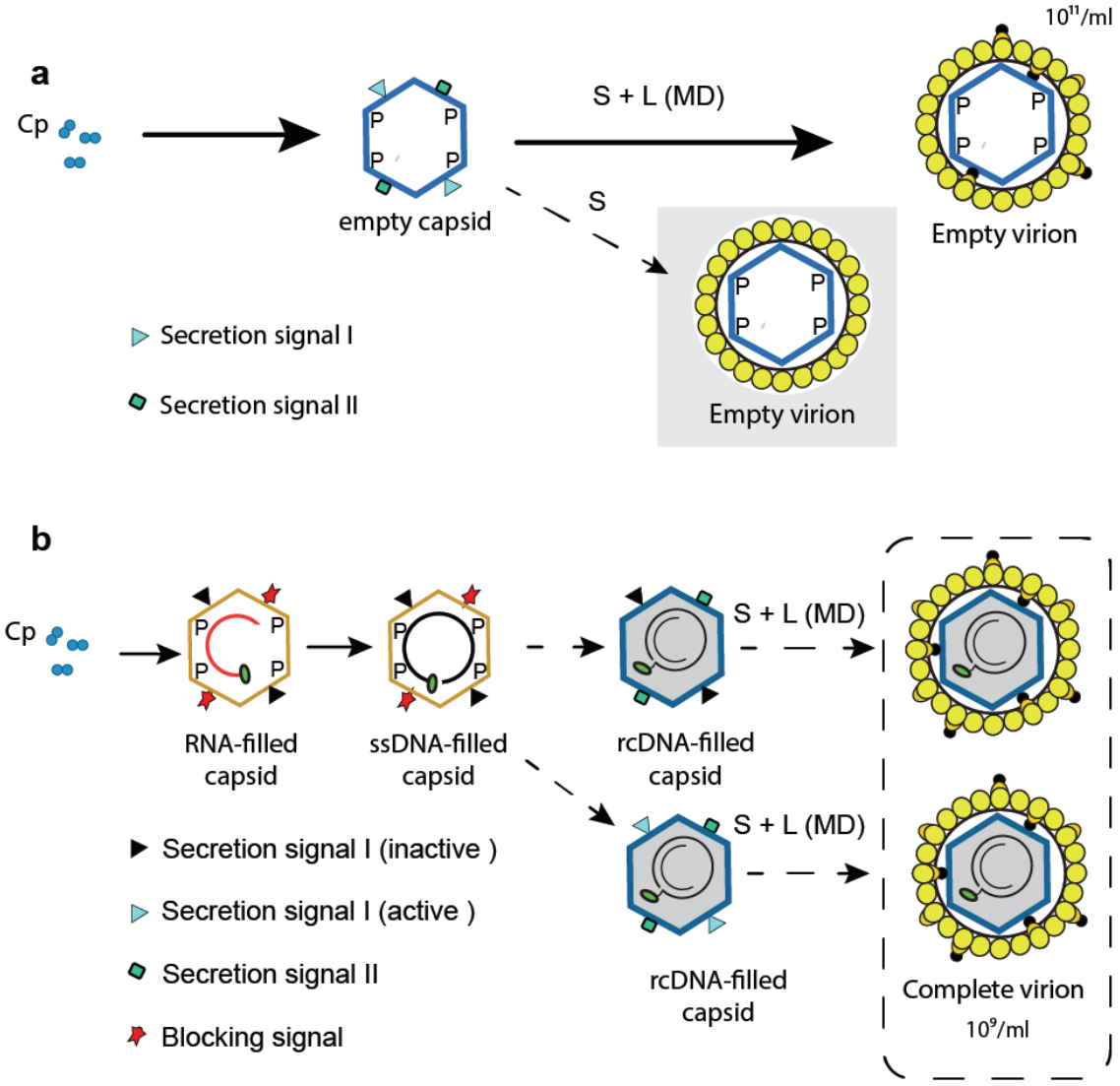


Figure 32. The "two-signal" secretion model. a) Secretion model for empty virions. The S and L envelope proteins form empty virions as the empty capsids contain the 2 signals, and the S protein alone is also capable of basal level secretion for empty virions but is not competent to secret complete virions. b) Secretion model for complete virions. The presence of pgRNA and ssDNA induce the blocking signal, which suppresses the secretion signal I (black triangles) to inhibit immature nucleocapsid secretion and can be eliminated by the emergence of secretion signal II. There are two possibilities for nucleocapsid maturation: either the emergence of signal II is sufficient for nucleocapsid envelopment (brown capsid), or the clearance of blocking signal releases the secretion signal I and together with the secretion signal II, they induce the nucleocapsid envelopment (blue capsid). The secretion of mature nucleocapsid requires both the S protein and the MD sequence of L protein. Adapted from: (Ning et al., 2018).

● Capsid assembly modulators (CAMs)

The CAMs are small molecule compounds which impair viral replication by disturbing capsid assembly (Nijampatnam and Liotta, 2019; L. Yang et al., 2019). Their mechanism of action is to prevent the encapsidation of the Pol-pgRNA complex by accelerating the kinetics of capsid assembly, and therefore blocking the reverse transcription of the pgRNA and the formation of rcDNA. Moreover, the interaction of CAMs and Cps is demonstrated to restrain the formation of cccDNA and restore innate signaling, which cannot be achieved by current nucleos(t)ide analogues (NAs) treatment (Belloni et al., 2014; Berke et al., 2017a; Gruffaz et al., 2013). Therefore, CAMs are now considered as one of the most potent candidates for hepatitis B treatment, and several compounds are in clinical trials (Feng et al., 2018; Schinazi et al., 2018).

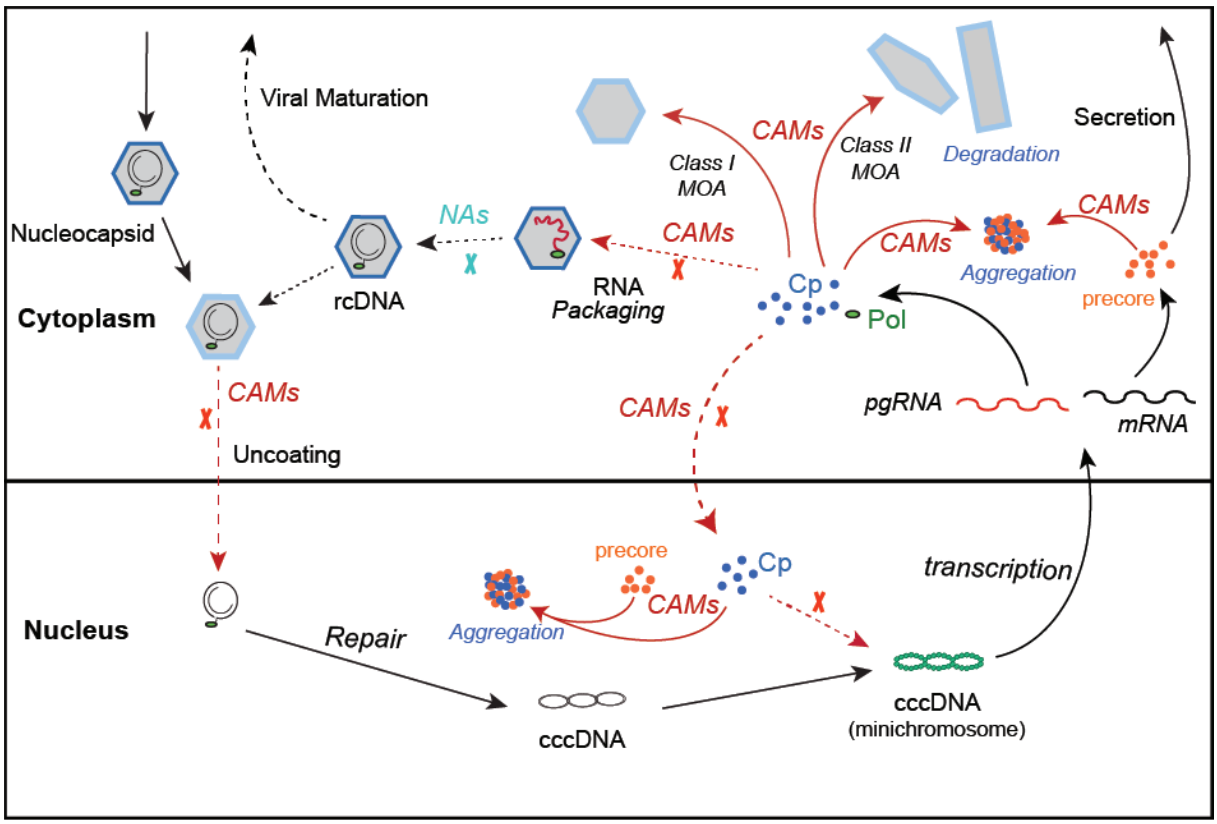


Figure 33. Mechanisms of action (MOA) of CAMs. The impact of CAMs on HBV replication is mainly the inhibition of pgRNA packaging, with two class MOAs, which result in apparently normal capsids devoid of viral genome or aberrant capsids respectively. Since Cp takes multiple roles in

the HBV lifecycle, the influence of CAMs can also take place during the uncoating of nucleocapsid, or also during minichromosome formation, while the NAs inhibit viral replication by focusing on the obstruction of reverse transcription only. Moreover, some CAMs are also found to reduce HBeAg secretion by inducing precore protein aggregation.

In recent years, several new scaffolds of CAMs have been discovered and developed (Corcuera et al., 2018; Nijampatnam and Liotta, 2019). Based on different mechanisms of action (MOA) between small molecules and capsids, three well studied scaffolds of CAMs have been classified into two functional classes: the first class compounds, including phenylpropenamides (PPAs) and sulfamoylbenzamides (SBAs), promote empty capsid assembly with apparently normal shape; the second class compounds, composed of heteroaryldihydropyrimidines (HAPs), leads to aberrant capsids (Berke et al., 2017a; Corcuera et al., 2018; Lahlali et al., 2018; Z. Zhou et al., 2017). We refer to the two classes compounds here as CAM-N and CAM-A respectively, to avoid the nomenclature confusion in previous publications (Schinazi et al., 2018; L. Yang et al., 2019).

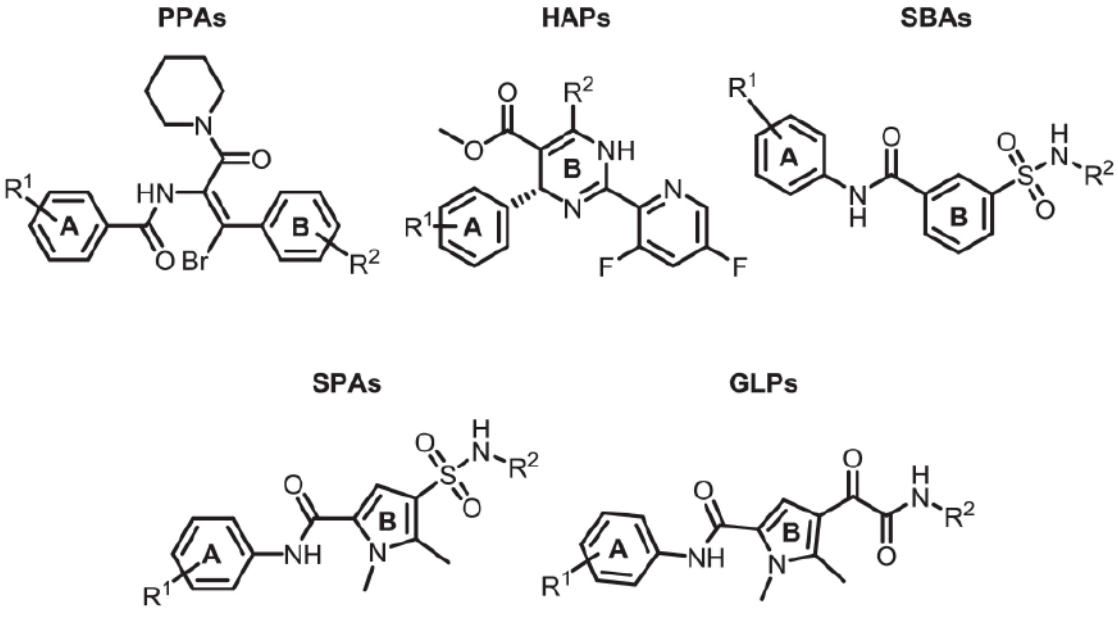


Figure 34. Five common chemical scaffolds of CAMs. The PPAs and SBAs, as well as most of the new developed CAMs including SPAs and GLPs, are CAM-N MOA compounds. The CAM-A MOA compounds are mainly made up by HAP and its derivatives. Taken from: (Nijampatnam and Liotta, 2019)

- **CAM-N compounds**

More than ten chemical scaffolds of CAMs have been discovered so far (Huber et al., 2018a; Nijampatnam and Liotta, 2019; L. Yang et al., 2019). Besides PPAs and SBAs, most of the novel non-HAP CAMs exhibit CAM-N MOA compounds, which accelerate capsid assembly without encapsidation of pgRNA, and empty capsids are formed as a result (Corcuera et al., 2018; Nijampatnam and Liotta, 2019), such as sulfamoylpyrroloamides (SPAs), glyoxamoylpyrroloxamides (GLPs), benzamides (BAs), dibenzo-thiazepin-2-one (DBT), pyrazolyl-thiazole (PT) (Corcuera et al., 2018; Lahlali et al., 2018; S. Wu et al., 2017).

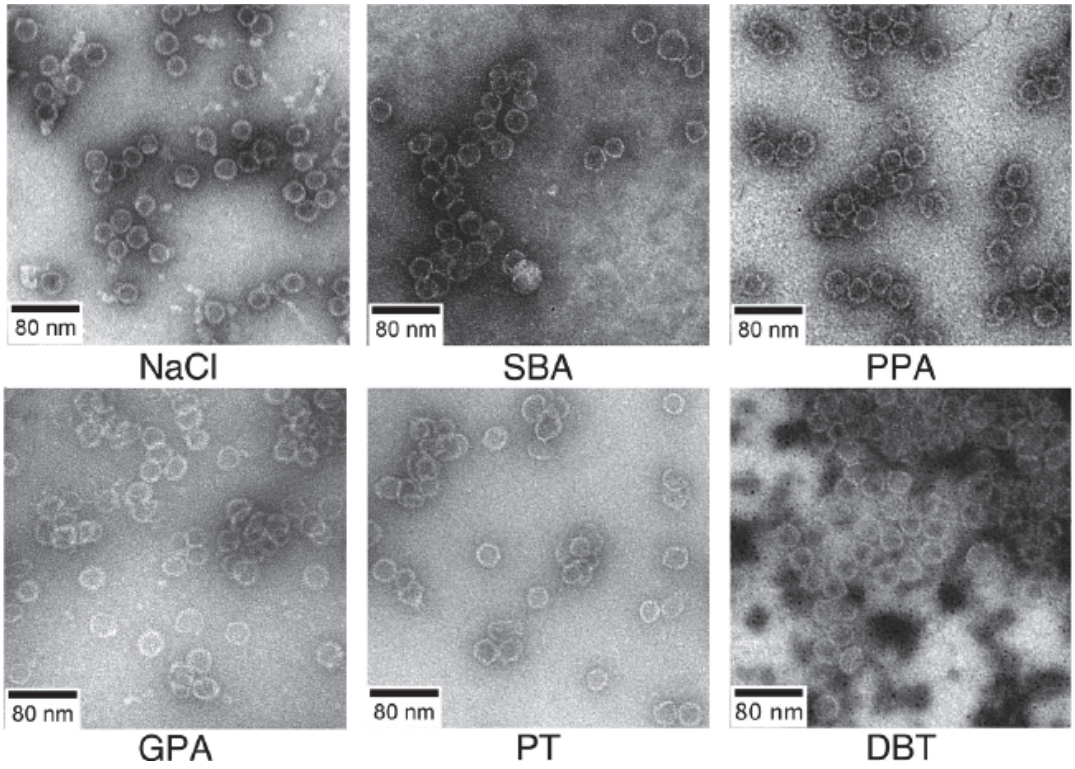


Figure 35. Electron micrographs of capsids assembled with CAM-N compounds. The CAM-N compounds lead to ostensibly intact capsids, which are indistinguishable from normal capsids assembled with NaCl. PPA: phenylpropenamide; SBA: sulfamoylbenzamide; GPA: glyoxamide-pyrrolamide; DBT: dibenzo-thiazepin-2-one; PT: pyrazolyl-thiazole. Adopted from: (Corcuera et al., 2018)

1. The phenylpropenamides (PPAs)

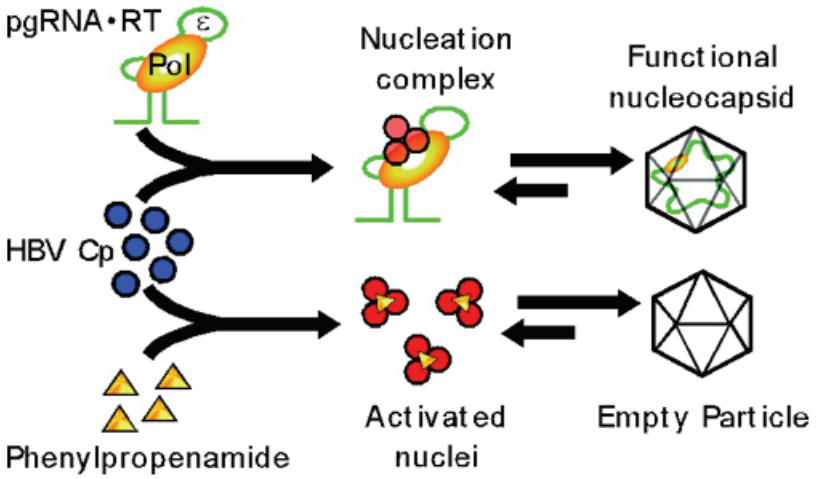


Figure 36. Mechanism model of PPAs promoting capsid assembly. The PPAs molecules stimulate the assembly of empty capsids which causes the free Cp dimers are consumed rapidly, and thus limit pgRNA packaging to inhibit nucleocapsid formation. Adapted from (Katen et al., 2010)

The first member of PPAs, AT-61, was reported to have antiviral activity about 20 years ago (King et al., 1998; Perni et al., 2000). While the AT-61 and its derivative AT-130 inhibit HBV replication in cell cultures, it was shown that their activity is not related to the cells resistance conferred by NAs, which indicates a different mechanism (Delaney et al., 2002; King et al., 1998). Unlike the NAs, AT-61 and AT-130 do not affect the activity of Pol, as well as the transcription and translation of viral genome, but can evidently reduce pgRNA packaging (Figure 36) (Delaney et al., 2002; Feld et al., 2007; Perni et al., 2000). Crystallization study shows that the AT-130 binds to the hydrophobic pockets of subunit B and C of capsids, while the B-pocket is its favorite binding site (Figure 37) (Katen et al., 2013). The binding, occurring between the piperidine ring of PPAs and the hydrophobic pocket (Katen et al., 2013; P. Wang et al., 2011), causes the Cp dimers to keep an assembly-active state, which accelerates the initiate nucleation step of assembly (Katen et al., 2010; Kondylis et al., 2019b). In the presence of AT-130 or another PPA derivative B-21, the assembled products also contain large numbers of incomplete particles (Kondylis et al., 2019b). This is probably due to the depletion of free Cp dimers and to the increased association energy of

dimers due to the binding of PPAs compounds, which prevents intermediates from dissociation (Kondylis et al., 2019b).

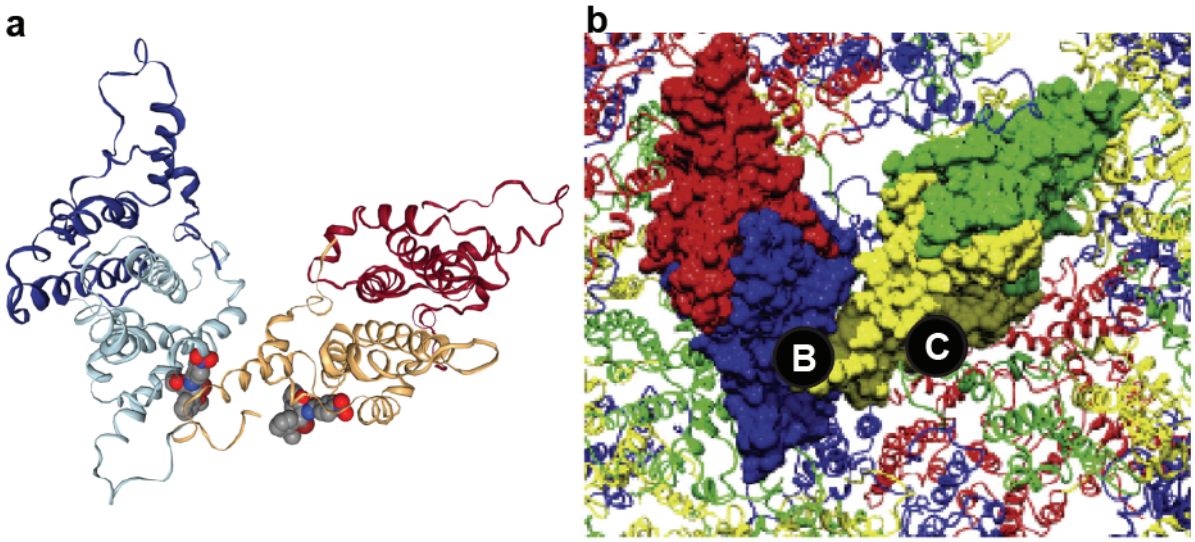


Figure 37. Binding mode of AT-130 to HBV capsid. a): crystal structures of Cp dimers binding with AT-130 at the hydrophobic pockets between Cp dimers (PDB number: 4G93). b): AT-130 molecules bind to the hydrophobic pockets of subunit B, as well as bind to subunit C, but favor B-pocket. Taken from: (Katen et al., 2013).

2. The sulfamoylbenzamides (SBAs)

The SBAs were discovered later than PPAs and HAPs, but show potent antiviral activity, and several member compounds are in the process of preclinical or clinical study (Feng et al., 2018; Nijampatnam and Liotta, 2019). DVR-01 and DVR-23, which are early representatives of SBA compounds discovered in 2013, display good ability in suppression of HBV replication with the same mechanism as AT-61, but do not affect the replication of other hepadnaviruses (e.g. DHBV, WHV) (Campagna et al., 2013). The crystal structure obtained from SBA-R01-bound to the assembly incompetent Cp149-Y132A mutant (forming hexamers instead of capsids), indicates that SBA compounds bind in the same hydrophobic pockets as PPAs and HAPs (Z. Zhou et al., 2017). NVR 3-778, whose structure is just partially disclosed, is the most advanced compound of SBA series and its Phase 1 clinical trial is nearly finished (Feng et al., 2018; Lam et al., 2019). The treatments with 600 mg BID (abbreviation for "bis in die",

twice a day) of NVR 3-778 alone or combined with PegIFN show positive antiviral efficacy (Feng et al., 2018). JNJ-379 and JNJ-399, which are promising antivirals but with no disclosed structures, are speculated to be derivatives of SBA compounds (Feng et al., 2018). JNJ-379 can either cause empty capsid assembly or interrupt the disassembly of capsid, resulting in the inhibition of both pgRNA encapsidation and cccDNA formation (Dusheiko, 2018). Mutations of residues at the hydrophobic pockets may reduce JNJ-379 efficacy (Berke et al., 2017b). Similar dual MOA was also observed with JNJ-632 and AB-432, which not only causes genome-free capsids with apparently intact shape, but also prevents the establishment of cccDNA in early infection steps (Berke et al., 2017a; Mani et al., 2018; Vandyck et al., 2018). Both JNJ-379 and AB-432 are undergoing Phase I clinical trials (Feng et al., 2018; Mani et al., 2018).

3. The sulfamoylpyrroloamides (SPAs)

The SPAs compounds were recently established as HBV antivirals (Vandyck et al., 2014; Vendeville et al., 2017). They are described to reduce viral mRNA and pgRNA quantity and to be involved in multiple steps of HBV replication (Nijampatnam and Liotta, 2019). For example, JNJ-827 was demonstrated to suppress pgRNA packaging as well as cccDNA formation in cells, with an additional effect on the secretion level of HBeAg (Lahlali et al., 2018).

4. The glyoxamoylpyrroloxamides (GLPs)

The GLPs are developed from SPA series, and are the newest scaffolds of CAMs (Nijampatnam and Liotta, 2019). Though only limited information of GLPs is disclosed, they are suggested to have the abilities of affecting capsid assembly, impeding cccDNA establishment, and disrupting preformed capsids (Bassit et al., 2018; Nijampatnam and Liotta, 2019; Schinazi et al., 2017; Vandyck et al., 2015). The GLP-26 exhibits highly potent activity in Hep AD38 cells and *in vivo*, and the combination treatment using GLP-26 with entecavir, an NAs drugs which has been used to treat hepatitis B, displays excellent efficacy (Bassit et al., 2018).

➢ CAM-A compounds

Most of CAM-A compounds are developed based on HAP series (Feng et al., 2018). Different from most other CAM scaffolds, the CAM-A compounds inhibit viral replication by prompting the Cp dimers to assembly into lager deviant capsids (Kondylis et al., 2019a; Schlicksup et al., 2018b; Stray and Zlotnick, 2006; Venkatakrishnan et al., 2016; Weber et al., 2002). Besides, they can also lead to Cp depletion via proteasomal degradation (Deres et al., 2003), and aggregation related with intracellular promyelocytic leukemia nuclear bodies (Huber et al., 2018b).

1. The heteroaryldihydropyrimidines (HAPs)

It has been shown that HAPs can increase the nucleation efficiency of Cps and misdirect the assembly to the incorrect path *in vitro* (Bourne et al., 2008; C. Liu et al., 2017; Stray et al., 2005), and also reduce the level of both Cp and viral DNA *in vivo* (Deres et al., 2003; Feng et al., 2018; Katen et al., 2013). Under negative stain electron microscopy, the assembly products misdirected by HAPs show distinct structures: for example, HAP-1 and HAP-12 lead to irregular particles (Bourne et al., 2008; Stray et al., 2005), while HAP-7 causes broken capsids (Bourne et al., 2008) and BAY 41-4109 and HAP-18 induce tubular structures (Bourne et al., 2008; C. Liu et al., 2017).

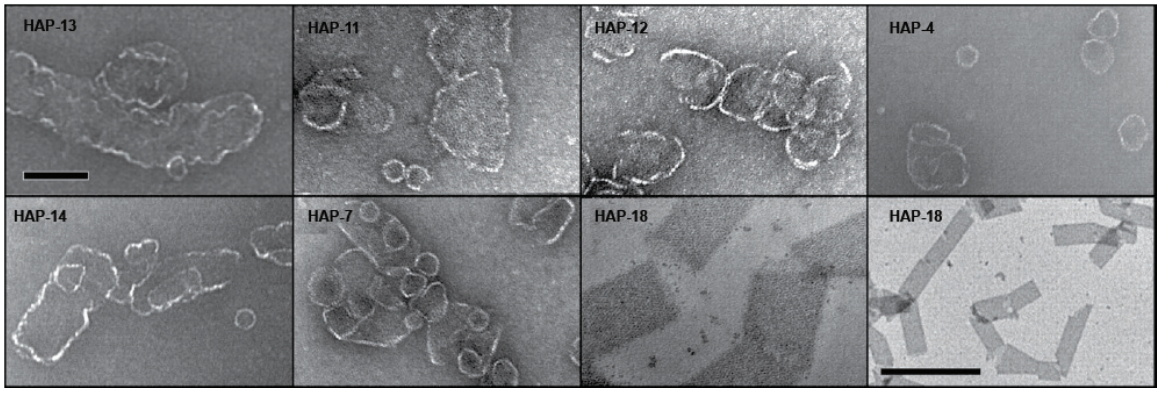


Figure 38. Electron micrographs of assembly products affected by HAPs *in vitro*. Electron micrographs display Cp149 dimers incubated with different HAP compounds at 37 °C for 24 h, with molar ratio of Cp149 dimers: HAP molecular = 1: 3. HAPs drive Cp149 dimers to assemble into various polymers, such as irregular particles, broken capsids, or tubular structures. The scale bar

(100 nm) in panel HAP-13 is shared by all micrographs, except the last one with a scale bar of 500 nm. Adopted from: (Bourne et al., 2008).

HAPs not only mislead the assembly of Cps in varied structures, but also distort preformed capsids in heterogeneous ways. In order to obtain structures at a good resolution, the HAP-binding structures are mainly resolved from mutants of truncated Cp (Berke et al., 2017a; Klumpp et al., 2015; Lahlali et al., 2018; Schlicksup et al., 2018b). A HAP-1-binding crystal structure bound with resolution of 5 Å shows that the hydrophobic pockets of capsid subunits are responsible for the interaction with HAPs, and contrary to AT-130, HAP-1 prefers subunit C to subunit B (Bourne et al., 2006; Katen et al., 2013). The presence of HAP-1 induces specific subunit movements, such as the opening of 3-fold vertices, the flattening of quasi-six-fold vertices, the bulge of 5-fold vertices, and finally results in changes of the capsid structure (Bourne et al., 2006). A cryo-EM study on HAP-TAMRA also shows the ability of HAP to distort preformed capsids, including conferring sharp angles on capsids, flatting the quasi-six-fold regions, and breaking capsid shapes (Schlicksup et al., 2018b). HAP-TAMRA compounds are also found to bind to hydrophobic pockets of subunits B and C in T=4 capsids, but without preference, and only binding to subunit B in T=3 capsids (Figure 39) (Schlicksup et al., 2018b). HAP-18, which interacts with almost the same Cp atoms as HAP-1, equally binds to subunits B and C, and brings very modest quaternary changes to preformed capsids (Venkatakrishnan et al., 2016). Distinct from the expanding effect of HAP-1 on preformed capsids, the diameter of aberrant capsids induced by HAP-18 is decreased by 3 Å (Venkatakrishnan et al., 2016).

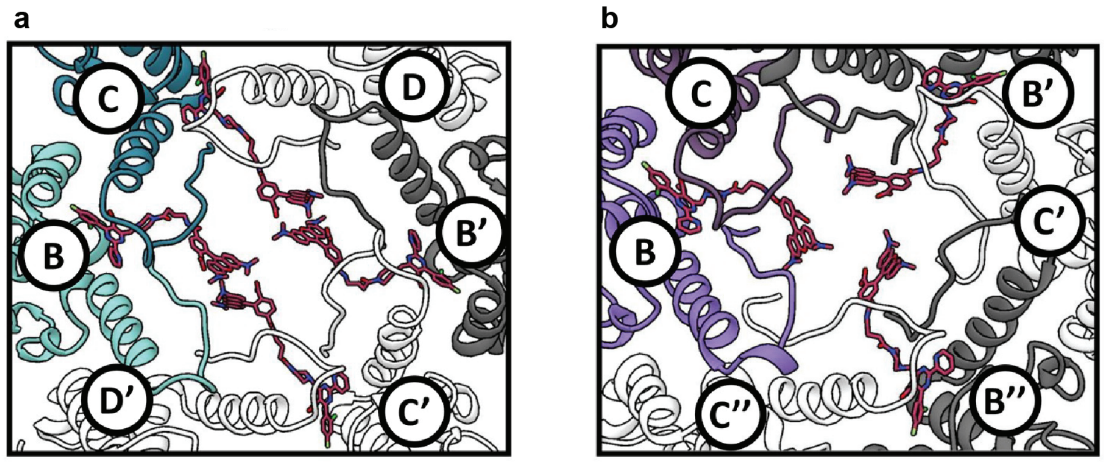


Figure 39. The binding of HAP-TAMRA on hydrophobic pocket. In T=4 capsids, four HAP-TAMRA molecules (red) locate in quasi-sixfold cluster with binding to subunit B and C, while the subunit D is not bound with the small molecules due to block of adjacent subunit B (a). The subunit C is blocked by B in T=3 capsids and therefore only subunit B is bound with HAP-TAMRA (b). Similar block also happens in five-fold cluster to prevent subunit A from binding. Taken from: (Schlicksup et al., 2018b).

The V124W mutant of Cp, which was designed to bury the HAP binding site by filling the hydrophobic pocket with a bulky tryptophan residue, shows resistance to HAP-1 and its derivative HAP-12 (Tan et al., 2013). In addition, the V124W dimers have a more compact conformation than WT dimers, and show high assembly activity of empty capsid (Tan et al., 2013). This is consistent with the reports showing that the subunits of capsid are more compact than free dimers (Bourne et al., 2009; Packianathan et al., 2010), and indicates that the allosteric change caused by V124W mutation or HAPs binding might related to the assembly-active state of Cp (Tan et al., 2013; 2015). Filling of the pocket around quasi-six-fold hole activates the assembly-active state of Cp dimers and strengthens the interdimer interactions (Bourne et al., 2006; Klumpp et al., 2015; Venkatakrishnan and Zlotnick, 2016). Given that HAPs cause capsid dissociation and finally result in broken capsids (Berke et al., 2017a; Lahlali et al., 2018; Stray and Zlotnick, 2006), the stabilization of HAPs to non-capsid polymers seems to be a contradictory effect (Schlicksup et al., 2018b; Stray et al., 2005).

Interestingly, an HAP derivative, HAP-R01, was found to not only misassemble HBeAg into irregular capsids *in vitro*, but also to cause the precore protein to be consumed in the cytoplasm and deposited in the nucleus in cell cultures (Z. Yan et al., 2019). Consistently, the level of HBeAg could be significantly lowered by HAP-R01 treatment in serum of infected FVB/N mice (Z. Yan et al., 2019).This effect was not observed in control tests with AT-130 and entecavir (ETV), whether other HAPs, especially the similar compounds under clinical trial, have this direct suppression effect on HBeAg which needs further investigation (Z. Yan et al., 2019).

2. Ciclopirox (CPX)

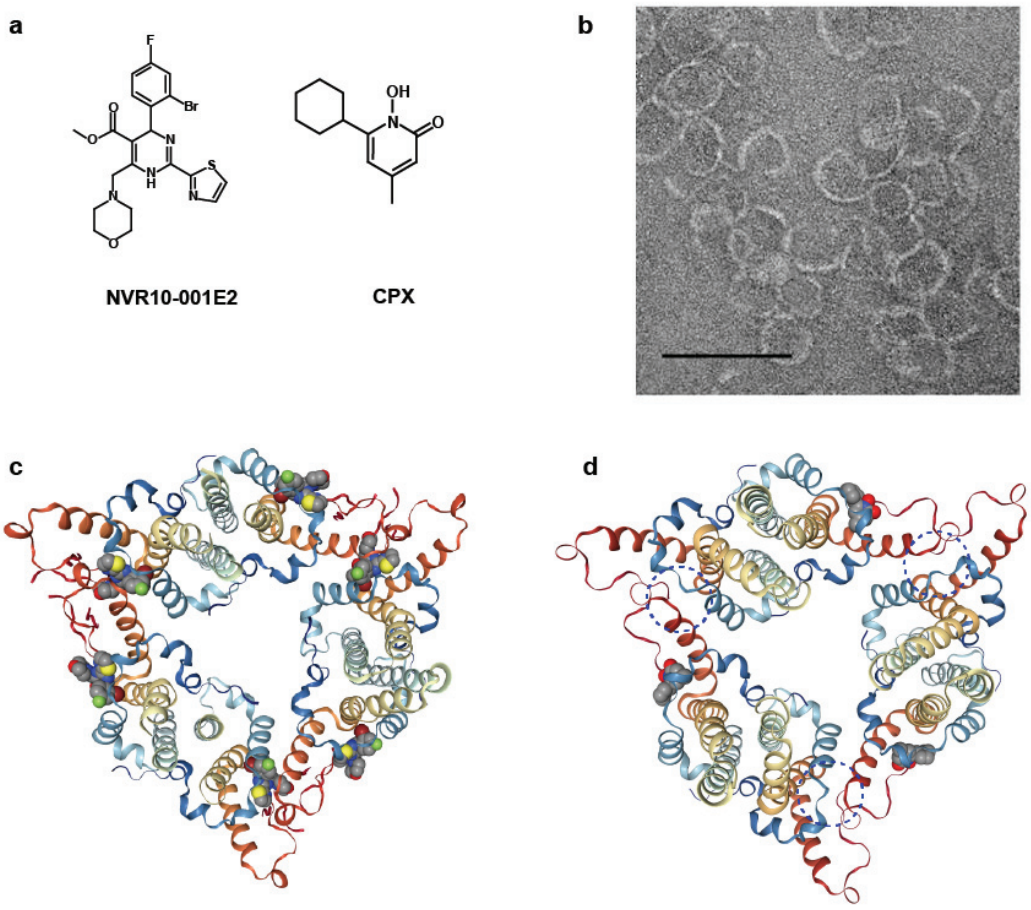


Figure 40. Ciclopirox (CPX) interacts with Cp by CAM-A MOA. CPX has a distinct chemical structure from HAP series (a) but exhibits similar effects on capsid assembly (b). Scale bar = 100 nm. (c) The crystal structure of the Cp149-Y132A hexamer binding with HAP derivative (NVR10-001E2, PDB code 5E0I) and (d) with CPX (PDB code 6J10). CPX also binds to the hydrophobic

pockets, whereas unlike the HAP derivatives, it is absent in pockets of the dimer-dimer interfaces (blue circles).

Recent research demonstrates that ciclopirox can act as a novel CAM compound with the CAM-A MOA (J.-A. Kang et al., 2019). CPX has been applied to treatment for fungal infection of skin and nails over 20 years (H. Zhou et al., 2010), and was reported to inhibit transcription of HIV-1 (Cáceres et al., 2016), but it does not seem to affect the transcription of HBV (J.-A. Kang et al., 2019). Similarly to HAP derivatives, CPX also misdirects Cp dimers to assemble into broken capsids and decreases the level of intracellular Cp probably due to the misassembled polymers which are degraded via the proteasome pathway (Deres et al., 2003; J.-A. Kang et al., 2019; Stray and Zlotnick, 2006; G. Wu et al., 2013). Meanwhile, the combination of CPX with NAs gives good synergistic result, indicating that it might be a promising small-molecular antiviral for hepatitis B treatment (J.-A. Kang et al., 2019).

- **New class MOA compounds**

Some of the novel CAMs show different MOA than CAM-N and CAM-A compounds. For example, the isothiafludine (NZ-4) inhibits pgRNA encapsidation by interacting with the CTD instead of the hydrophobic pocket of Cps (L. Yang et al., 2014; 2016); BCM-599 and its related compounds suppress viral replication by preventing Cp dimers from assembly (Cho et al., 2014); and HF9C6 interacts with Cp through the hydrophobic pocket but leads to multicapsid aggregates *in vitro* and prevents Cp or capsid from entering nuclear in infected cells (Huber et al., 2018a). Those reports suggest that CAMs have various MOA in inhibition of viral replication, and might efficiently suppress HBV infection through a combination of treatments with current antivirals.

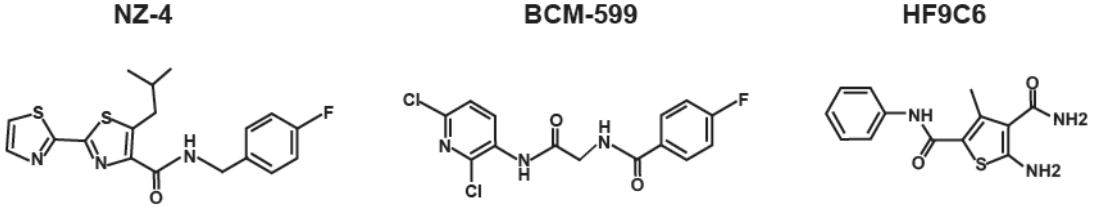


Figure 41. Chemical Structure of novel MOA compounds.

● References

Abbas, Z., Afzal, R., 2013. Life cycle and pathogenesis of hepatitis D virus: A review. WJH 5, 666–675. doi:10.4254/wjh.v5.i12.666

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., Sali, A., Rout, M.P., 2007. The molecular architecture of the nuclear pore complex. Nature 450, 695–701. doi:10.1038/nature06405

Alexopoulou, A., Karayiannis, P., 2014. HBeAg negative variants and their role in the natural history of chronic hepatitis B virus infection. World J. Gastroenterol. 20, 7644–7652. doi:10.3748/wjg.v20.i24.7644

Allweiss, L., Dandri, M., 2017. The Role of cccDNA in HBV Maintenance. Viruses 9, 156. doi:10.3390/v9060156

Appen, von, A., Beck, M., 2016. Structure Determination of the Nuclear Pore Complex with Three-Dimensional Cryo electron Microscopy. J. Mol. Biol. 428, 2001–2010. doi:10.1016/j.jmb.2016.01.004

Bardens, A., Döring, T., Stieler, J., Prange, R., 2011. Alix regulates egress of hepatitis B virus naked capsid particles in an ESCRT-independent manner. Cell. Microbiol. 13, 602–619. doi: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, 3413–3420.

Bartenschlager, R., Urban, S., Protzer, U., 2019. Towards curative therapy of chronic viral hepatitis. Z Gastroenterol 57, 61–73. doi:10.1055/a-0824-1576

Bassit, L., Cox, B., Ono, S.K., Verma, K., Yoon, J., Amblard, F., Schinazi, R.F., 2018. Novel and Potent HBV Capsid Modulator Reduces HBeAg and cccDNA in Core-site Directed T109I Mutant in HepNTCP Cells, in:. Presented at the The International Liver Congress EASL - European Association for the Study of the Liver Paris France.

Beck, J., Nassal, M., 2007. Hepatitis B virus replication. World J. Gastroenterol. 13, 48–64.

Bell, S.J., Nguyen, T., 2009. The management of hepatitis B. Australian Prescriber 32, 99–104.

Belloni, L., Li, L., Palumbo, G.A., Chirapu, S.R., Calvo, L., Finn, M.G., Lopatin, U., Zlotnick, A., Levrero, M., 2014. HAPs hepatitis B virus (HBV) capsid inhibitors prevent HBc interaction with the viral minichromosome and selected host cell genes to inhibits transcription and affect cccDNA stability. Digestive and Liver Disease 46, e9. doi:10.1016/j.dld.2014.01.024

Belloni, L., Pollicino, T., De Nicola, F., Guerrieri, F., Raffa, G., Fanciulli, M., Raimondo, G., Levrero, M., 2009. Nuclear HBx binds the HBV minichromosome and modifies the epigenetic regulation of cccDNA function. PNAS 106, 19975–19979. doi:10.1073/pnas.0908365106

Berke, J.M., Dehertogh, P., Vergauwen, K., Van Damme, E., Mostmans, W., Vandyck, K., Pauwels, F., 2017a. Capsid Assembly Modulators Have a Dual Mechanism of Action in Primary Human Hepatocytes Infected with Hepatitis B Virus. Antimicrob. Agents Chemother. 61, e00560–17–14. doi:10.1128/AAC.00560-17

Berke, J.M., Verbinnen, T., Tan, Y., Dehertogh, P., Vergauwen, K., Vandyck, K., Lenz, O., 2017b. The HBV capsid assembly modulator JNJ-379 is a potent inhibitor of viral

replication across full length genotype A-H clinical isolates in vitro, in:. Presented at the The Liver Meeting Washington DC October 2017.

Bertoletti, A., Lucifora, J., Zoulim, F., 2018. Virology and Pathogenesis of Hepatitis B, in: Zakim and Boyer's Hepatology. Elsevier Inc., pp. 464–473.e5. doi:10.1016/B978-0-323-37591-7.00031-8

Biswas, S., Candotti, D., Allain, J.-P., 2013. Specific amino acid substitutions in the S protein prevent its excretion in vitro and may contribute to occult hepatitis B virus infection. J. Virol. 87, 7882–7892. doi:10.1128/JVI.00710-13

Blanchet, M., Sureau, C., 2006. Analysis of the cytosolic domains of the hepatitis B virus envelope proteins for their function in viral particle assembly and infectivity. J. Virol. 80, 11935–11945. doi:10.1128/JVI.00621-06

Blum, H.E., Zhang, Z.S., Galun, E., Weizsäcker, von, F., Garner, B., Liang, T.J., Wands, J.R., 1992. Hepatitis B virus X protein is not central to the viral life cycle in vitro. J. Virol. 66, 1223–1227.

Blumberg, B., Alter, H., 1965. A “New” Antigen in Leukemia Sera

. JAMA 191, 541–546.

Bock, C.T., Schranz, P., Schröder, C.H., Zentgraf, H., 1994. Hepatitis B virus genome is organized into nucleosomes in the nucleus of the infected cell. Virus Genes 8, 215–229.

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. J. Mol. Biol. 307, 183–196. doi:10.1006/jmbi.2000.4481

Bonino, F., Piratvisuth, T., Brunetto, M.R., Liaw, Y.-F., 2010. Diagnostic markers of chronic hepatitis B infection and disease. Antiviral Therapy 15 Suppl 3, 35–44. doi:10.3851/IMP1622)

Bottcher, B., Wynne, S.A., Crowther, R.A., 1997. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature 386, 88–91. doi:10.1038/386088a0

Bourne, C., Lee, S., Venkataiah, B., Lee, A., Korba, B., Finn, M.G., Zlotnick, A., 2008. Small-molecule effectors of hepatitis B virus capsid assembly give insight into virus life cycle. J. Virol. 82, 10262–10270. doi:10.1128/JVI.01360-08

Bourne, C.R., Finn, M.G., Zlotnick, A., 2006. Global structural changes in hepatitis B virus capsids induced by the assembly effector HAP1. J. Virol. 80, 11055–11061. doi:10.1128/JVI.00933-06

Bourne, C.R., Katen, S.P., Fulz, M.R., Packianathan, C., Zlotnick, A., 2009. A mutant hepatitis B virus core protein mimics inhibitors of icosahedral capsid self-assembly. Biochemistry 48, 1736–1742. doi:10.1021/bi801814y

Böttcher, B., Nassal, M., 2018. Structure of Mutant Hepatitis B Core Protein Capsids with Premature Secretion Phenotype. J. Mol. Biol. 430, 4941–4954. doi:10.1016/j.jmb.2018.10.018

Brunetto, M.R., Stenlert, M., Rizzetto, M., Verme, G., Will, H., Bonino, F., 1989. A HBV variant is responsible for anti-HBe positive hepatitis B. Journal of Hepatology 9, S11–1. doi:10.1016/0168-8278(89)90199-2

Bruss, V., 2007. Hepatitis B virus morphogenesis. World J. Gastroenterol. 13, 65–73. doi:10.3748/wjg.v13.i1.65

Bruss, V., 1997. A short linear sequence in the pre-S domain of the large hepatitis B virus envelope protein required for virion formation. J. Virol. 71, 9350–9357.

Bruss, V., Ganem, D., 1991. The role of envelope proteins in hepatitis B virus assembly. Proc. Natl. Acad. Sci. U.S.A. 88, 1059–1063. doi:10.1073/pnas.88.3.1059

Bruss, V., Lu, X., Thomssen, R., Gerlich, W.H., 1994. Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein. The EMBO Journal 13, 2273–2279.

Bruss, V., Vieluf, K., 1995. Functions of the internal pre-S domain of the large surface protein in hepatitis B virus particle morphogenesis. J. Virol. 69, 6652–6657.

Campagna, M.R., Liu, F., Mao, R., Mills, C., Cai, D., Guo, F., Zhao, X., Ye, H., Cuconati, A., Guo, H., Chang, J., Xu, X., Block, T.M., Guo, J.-T., 2013. Sulfamoylbenzamide derivatives inhibit the assembly of hepatitis B virus nucleocapsids. J. Virol. 87, 6931–6942. doi:10.1128/JVI.00582-13

Carman, W.F., Zanetti, A.R., Karayiannis, P., Waters, J., Manzillo, G., Tanzi, E., Zuckerman, A.J., Thomas, H.C., 1990. Vaccine-induced escape mutant of hepatitis B virus. The Lancet 336, 325–329.

Casciano, J.C., Duchemin, N.J., Lamontagne, R.J., Steel, L.F., Bouchard, M.J., 2017. Hepatitis B virus modulates store-operated calcium entry to enhance viral replication in primary hepatocytes. PLoS ONE 12, e0168328. doi:10.1371/journal.pone.0168328

Cáceres, C.J., Angulo, J., Contreras, N., Pino, K., Vera-Otarola, J., López-Lastra, M., 2016. Targeting deoxyhypusine hydroxylase activity impairs cap-independent translation initiation driven by the 5'untranslated region of the HIV-1, HTLV-1, and MMTV mRNAs. Antiviral Research 134, 192–206. doi:10.1016/j.antiviral.2016.09.006

Ceres, P., Stray, S.J., Zlotnick, A., 2004. Hepatitis B virus capsid assembly is enhanced by naturally occurring mutation F97L. J. Virol. 78, 9538–9543. doi:10.1128/JVI.78.17.9538-9543.2004

Champoux, J.J., Schultz, S.J., 2009. Ribonuclease H: properties, substrate specificity and roles in retroviral reverse transcription. FEBS J. 276, 1506–1516. doi:10.1111/j.1742-4658.2009.06909.x

Chen, C., Wang, J.C.-Y., Zlotnick, A., 2011. A kinase chaperones hepatitis B virus capsid assembly and captures capsid dynamics in vitro. PLoS Pathog 7, e1002388. doi:10.1371/journal.ppat.1002388

Chen, E.-Q., Feng, S., Wang, M.-L., Liang, L.-B., Zhou, L.-Y., Du, L.-Y., Yan, L.-B., Tao, C.-M., Tang, H., 2017. Serum hepatitis B core-related antigen is a satisfactory surrogate marker of intrahepatic covalently closed circular DNA in chronic hepatitis B. Scientific RepoRts 7, 173. doi:10.1038/s41598-017-00111-0

Chen, H.S., Kaneko, S., Girones, R., Anderson, R.W., Hornbuckle, W.E., Tennant, B.C., Cote, P.J., Gerin, J.L., Purcell, R.H., Miller, R.H., 1993. The woodchuck hepatitis virus X gene is important for establishment of virus infection in woodchucks. J. Virol. 67, 1218–1226.

Chen, M., Sällberg, M., Hughes, J., Jones, J., Guidotti, L.G., Chisari, F.V., Billaud, J.-N., Milich, D.R., 2005. Immune tolerance split between hepatitis B virus precore and core proteins. J. Virol. 79, 3016–3027. doi:10.1128/JVI.79.5.3016-3027.2005

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. Proc. Natl. Acad. Sci. U.S.A. 101, 14913–14918. doi:10.1073/pnas.0406282101

Chen, P., Xie, Q., Lu, X., Yu, C., Xu, K., Ruan, B., Cao, H., Gao, H., Li, L., 2017. Serum HBeAg and HBV DNA levels are not always proportional and only high levels of HBeAg most likely

correlate with high levels of HBV DNA: A community-based study. Medicine (Baltimore) 96, e7766. doi:10.1097/MD.0000000000007766

Chen, R., Zhao, X., Wang, Y., Xie, Y., Liu, J., 2017. Hepatitis B virus X protein is capable of down-regulating protein level of host antiviral protein APOBEC3G. Scientific RepoRts 7, 40783. doi:10.1038/srep40783

Cho, M.H., Jeong, H., Kim, Y.S., Kim, J.-W., Jung, G., 2014. 2-amino-N-(2,6-dichloropyridin-3-yl)acetamide derivatives as a novel class of HBV capsid assembly inhibitor. J. Viral Hepat. 21, 843–852. doi:10.1111/jvh.12214

Christoph Seeger, Fabien Zoulim, William S. Mason, 2013. HEPADNAVIRUSES, in: Fields Virology Sixth Edition. pp. 2185–2221.

Chua, P.K., Wen, Y.-M., Shih, C., 2003. Coexistence of two distinct secretion mutations (P5T and I97L) in hepatitis B virus core produces a wild-type pattern of secretion. J. Virol. 77, 7673–7676. doi:10.1128/JVI.77.13.7673-7676.2003

Clark, D.N., Hu, J., 2015. Hepatitis B virus reverse transcriptase - Target of current antiviral therapy and future drug development. Antiviral Research 123, 132–137. doi:10.1016/j.antiviral.2015.09.011

Collaborators, T.P.O., 2018. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. The Lancet Gastroenterology & Hepatology 1–21. doi:10.1016/S2468-1253(18)30056-6

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. doi:10.1038/386091a0

Conway, J.F., Watts, N.R., Belnap, D.M., CHENG, N., Stahl, S.J., Wingfield, P.T., Steven, A.C., 2003. Characterization of a Conformational Epitope on Hepatitis B Virus Core Antigen and Quasiequivalent Variations in Antibody Binding. J. Virol. 77, 6466–6473. doi:10.1128/JVI.77.11.6466-6473.2003

Corcuera, A., Stolle, K., Hillmer, S., Seitz, S., Lee, J.-Y., Bartenschlager, R., Birkmann, A., Urban, A., 2018. Novel non-heteroarylpyrimidine (HAP) capsid assembly modifiers have a different mode of action from HAPs in vitro. Antiviral Research 158, 135–142. doi:10.1016/j.antiviral.2018.07.011

Couroucé-Pauty, A.M., Lemaire, J.M., Roux, J.F., 1978. New hepatitis B surface antigen subtypes inside the ad category. Vox Sang. 35, 304–308.

Crowther, R.A., Kiselev, N.A., Bottcher, B., Berriman, J.A., Borisova, G.P., Ose, V., Pumpens, P., 1994. Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell 77, 943–950.

Dane, D.S., Cameron, C.H., Briggs, M., 1970. Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. The Lancet 1, 695–698.

Das, K., Xiong, X., Yang, H., Westland, C.E., Gibbs, C.S., Sarafianos, S.G., Arnold, E., 2001. Molecular modeling and biochemical characterization reveal the mechanism of hepatitis B virus polymerase resistance to lamivudine (3TC) and emtricitabine (FTC). J. Virol. 75, 4771–4779. doi:10.1128/JVI.75.10.4771-4779.2001

Daub, H., Blencke, S., Habenberger, P., Kurtenbach, A., Dennenmoser, J., Wissing, J., Ullrich, A., Cotten, M., 2002. Identification of SRPK1 and SRPK2 as the major cellular protein kinases phosphorylating hepatitis B virus core protein. J. Virol. 76, 8124–8137. doi:10.1128/JVI.76.16.8124-8137.2002

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, 386–389. doi:10.1038/nature17170

Delaney, W.E., Edwards, R., Colledge, D., Shaw, T., Furman, P., Painter, G., Locarnini, S., 2002. Phenylpropenamide Derivatives AT-61 and AT-130 Inhibit Replication of Wild-Type and Lamivudine-Resistant Strains of Hepatitis B Virus In Vitro. Antimicrob. Agents Chemother. 46, 3057–3060. doi:10.1128/AAC.46.9.3057-3060.2002

Deres, K., Schröder, C.H., Paessens, A., Goldmann, S., Hacker, H.J., Weber, O., Krämer, T., Niewöhner, U., Pleiss, U., Stoltefuss, J., Graef, E., Koletzki, D., Masantschek, R.N.A., Reimann, A., Jaeger, R., Gross, R., Beckermann, B., Schlemmer, K.-H., Haebich, D., Rübsamen-Waigmann, H., 2003. Inhibition of hepatitis B virus replication by drug-induced depletion of nucleocapsids. Science 299, 893–896. doi:10.1126/science.1077215

Dienes, H.P., Gerken, G., Goergen, B., Heermann, K., Gerlich, W., Meyer zum Büschenfelde, K.H., 1995. Analysis of the precore DNA sequence and detection of precore antigen in liver specimens from patients with anti-hepatitis B e-positive chronic hepatitis. Hepatology 21, 1–7.

Dill, J.A., Camus, A.C., Leary, J.H., Di Giallonardo, F., Holmes, E.C., Ng, T.F.F., 2016. Distinct Viral Lineages from Fish and Amphibians Reveal the Complex Evolutionary History of Hepadnaviruses. J. Virol. 90, 7920–7933. doi:10.1128/JVI.00832-16

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, 133–142. doi: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, 843–850. doi:10.1016/j.molcel.2006.04.025

Dusheiko, G., 2018. Current and future directions for the management of hepatitis B. S. Afr. Med. J. 108, 22–30. doi:10.7196/SAMJ.2018.v108i8b.13497

Eble, B.E., Lingappa, V.R., Ganem, D., 1990. The N-terminal (pre-S2) domain of a hepatitis B virus surface glycoprotein is translocated across membranes by downstream signal sequences. J. Virol. 64, 1414–1419.

Eble, B.E., MacRae, D.R., Lingappa, V.R., Ganem, D., 1987. Multiple topogenic sequences determine the transmembrane orientation of the hepatitis B surface antigen. Mol. Cell. Biol. 7, 3591–3601.

Echevarría, J.M., Avellón, A., 2006. Hepatitis B virus genetic diversity. J. Med. Virol. 78, S36–S42. doi:10.1002/jmv.20605

Endres, D., Zlotnick, A., 2002. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. Biophysical Journal 83, 1217–1230. doi:10.1016/S0006-3495(02)75245-4

Eren, E., Watts, N.R., Dearborn, A.D., Palmer, I.W., Kaufman, J.D., Steven, A.C., Wingfield, P.T., 2018. Structures of Hepatitis B Virus Core- and e-Antigen Immune Complexes Suggest Multi-point Inhibition. Structure 26, 1314–1326.e4. doi:10.1016/j.str.2018.06.012

Fako, V., Wang, X.W., 2017. Molecular Carcinogenesis of HBV-Related HCC, in: Hepatitis B Virus and Liver Disease. Springer Singapore, Singapore, pp. 143–162. doi:10.1007/978-981-10-4843-2_8

Falcón, V., Menéndez, I., Acosta-Rivero, N., Shibayama, M., la Rosa, de, M.C., Muñoz, J.L., Miranda-Sánchez, M., Gavilondo, J.V., Lopez, D., Dueñas-Carrera, S., Grá, B., Chinea, G., Garcia, L.T., García, W.P., Vidal, E.R., Arús-Soler, E.R., da Silva, J.O., Álvarez, F.S., Acosta, E.J.O., Seoane, J.D., Morales-grillo, J., Pentón, E., Kourí, J.B., Tsutsumi, V., 2008. Ultrastructural Evidences of Hepatitis B Infection in Human Liver Biopsies Disclose Complex Assembly and Morphogenesis Pathways for Hepatitis B Virus. American J. of Infectious Diseases 4, 1–8.

Feld, J.J., Colledge, D., Sozzi, V., Edwards, R., Littlejohn, M., Locarnini, S.A., 2007. The phenylpropenamide derivative AT-130 blocks HBV replication at the level of viral RNA packaging. Antiviral Research 76, 168–177. doi:10.1016/j.antiviral.2007.06.014

Feng, S., Gao, L., Han, X., Hu, T., Hu, Y., Liu, H., Thomas, A.W., Yan, Z., Yang, S., Young, J.A.T., Yun, H., Zhu, W., Shen, H.C., 2018. Discovery of Small Molecule Therapeutics for Treatment of Chronic HBV Infection. ACS Infect Dis 4, 257–277. doi:10.1021/acsinfecdis.7b00144

Fernholz, D., Stemler, M., Brunetto, M., Bonino, F., Will, H., 1991. Replicating and virion secreting hepatitis B mutant virus unable to produce preS2 protein. Journal of Hepatology 13 Suppl 4, S102–4.

Funk, A., Mhamdi, M., Will, H., Sirma, H., 2007. Avian hepatitis B viruses: molecular and cellular biology, phylogenesis, and host tropism. World J. Gastroenterol. 13, 91–103. doi:10.3748/wjg.v13.i1.91

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, 646–650.

Gallina, A., Bonelli, F., Zentilin, L., Rindi, G., Muttini, M., Milanesi, G., 1989. A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. J. Virol. 63, 4645–4652.

Gallucci, L., Kann, M., 2017. Nuclear Import of Hepatitis B Virus Capsids and Genome. Viruses 9, 21. doi:10.3390/v9010021

Ganem, D., Schneider, R., 2001. Hepadnaviridae: The Viruses and Their Replication, in: Fields, B.N., Peter M, M.H., Diane E, P.D.G., Robert A, P.D.L., Malcolm A, M.M., Roizman, B., Stephen E, M.S., David M, P.D.K. (Eds.), Fields - Virology 4th Ed. pp. 2923–2969.

Gao, Y., Feng, J., Yang, G., Zhang, S., Liu, Y., Bu, Y., Sun, M., Zhao, M., Chen, F., Zhang, W., Ye, L., Zhang, X., 2017. Hepatitis B virus X protein-elevated MSL2 modulates hepatitis B virus covalently closed circular DNA by inducing degradation of APOBEC3B to enhance hepatocarcinogenesis. Hepatology 66, 1413–1429. doi:10.1002/hep.29316

GBD 2016 Causes of Death Collaborators, 2017. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 390, 1151–1210. doi:10.1016/S0140-6736(17)32152-9

Geng, X., Huang, C., Qin, Y., McCombs, J.E., Yuan, Q., Harry, B.L., Palmer, A.E., Xia, N.-S., Xue, D., 2012. Hepatitis B virus X protein targets Bcl-2 proteins to increase intracellular calcium, required for virus replication and cell death induction. Proc. Natl. Acad. Sci. U.S.A. 109, 18471–18476. doi:10.1073/pnas.1204668109

Gish, R.G., Given, B.D., Lai, C.-L., Locarnini, S.A., Lau, J.Y.N., Lewis, D.L., Schluep, T., 2015. Chronic hepatitis B: Virology, natural history, current management and a glimpse at future opportunities. Antiviral Research 1–12. doi:10.1016/j.antiviral.2015.06.008

Goldstein, S.T., Zhou, F., Hadler, S.C., Bell, B.P., Mast, E.E., Margolis, H.S., 2005. A mathematical model to estimate global hepatitis B disease burden and vaccination impact. International Journal of Epidemiology 34, 1329–1339. doi:10.1093/ije/dyi206

Gong, D.-Y., Chen, E.-Q., Huang, F.-J., Leng, X.-H., Cheng, X., Tang, H., 2013. Role and functional domain of hepatitis B virus X protein in regulating HBV transcription and replication in vitro and in vivo. Viruses 5, 1261–1271. doi:10.3390/v5051261

Gong, Q., Chen, S., Guo, J., Sun, H., Zheng, G., Liu, Q., Ren, H., He, S., 2011. Chromosome remodeling related to hepatitis B virus replication in HepG2 cells. DNA Cell Biol. 30, 347–354. doi:10.1089/dna.2010.1172

Gou, Y., Zhao, Y., Rao, C., Feng, S., Wang, T., Li, D., Tao, C., 2017. Predictive Value of Hepatitis B Core-Related Antigen (HBcrAg) During the Natural History of Hepatitis B Virus Infection. Clin. Lab. 63, 1063–1070. doi:10.7754/Clin.Lab.2017.161034

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, 292–299. doi:https://doi.org/10.1006/viro.1995.0002

Gruffaz, M., Testoni, B., Luangsay, S., Ait-Goughoulte, M., Petit, M.A., Ma, H., Klumpp, K., Javanbakht, H., Durantel, D., Zoulim, F., 2013. 378 HEPATITIS B CORE (HBC) PROTEIN IS A KEY AND VERY EARLY NEGATIVE REGULATOR OF THE INTERFERON RESPONSE. Journal of Hepatology 58, S155–S156. doi:10.1016/S0168-8278(13)60380-3

Hadziyannis, S., 2001. Hepatitis B e antigen–negative chronic hepatitis B. Hepatology 34, 617–624. doi:10.1053/jhep.2001.27834

Hagan, M.F., 2014. Modeling Viral Capsid Assembly. Adv Chem Phys 155, 1–68. doi:10.1002/9781118755815.ch01

Hagan, M.F., Elrad, O.M., 2010. Understanding the concentration dependence of viral capsid assembly kinetics--the origin of the lag time and identifying the critical nucleus size. Biophysical Journal 98, 1065–1074. doi:10.1016/j.bpj.2009.11.023

Hahn, C.M., Iwanowicz, L.R., Cornman, R.S., Conway, C.M., Winton, J.R., Blazer, V.S., 2015. Characterization of a Novel Hepadnavirus in the White Sucker (Catostomus commersonii) from the Great Lakes Region of the United States. J. Virol. 89, 11801–11811. doi:10.1128/JVI.01278-15

Heermann, K.H., Goldmann, U., Schwartz, W., Seyffarth, T., Baumgarten, H., Gerlich, W.H., 1984. Large surface proteins of hepatitis B virus containing the pre-s sequence. J. Virol. 52, 396–402.

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 Pathog 14, e1007488. doi:10.1371/journal.ppat.1007488

Hollinger, F.B., Lau, D.T.Y., 2006. Hepatitis B: The Pathway to Recovery Through Treatment. Gastroenterology Clinics of North America 35, 895–931. doi:10.1016/j.gtc.2006.10.002

Holmes, K., Shepherd, D.A., Ashcroft, A.E., Whelan, M., Rowlands, D.J., Stonehouse, N.J., 2015. Assembly Pathway of Hepatitis B Core Virus-like Particles from Genetically Fused Dimers. J. Biol. Chem. 290, 16238–16245. doi:10.1074/jbc.M114.622035

Hoofnagle, J.H., Doo, E., Liang, T.J., Fleischer, R., Lok, A.S.F., 2007. Management of hepatitis B: Summary of a clinical research workshop. Hepatology 45, 1056–1075. doi:10.1002/hep.21627

Hu, J., Liu, K., 2017. Complete and Incomplete Hepatitis B Virus Particles: Formation, Function, and Application. Viruses 9, 56–17. doi:10.3390/v9030056

Huan, B., Siddiqui, A., 1993. Regulation of hepatitis B virus gene expression. Journal of Hepatology 17, S20–S23.

Huber, A.D., Pineda, D.L., Liu, D., Boschert, K.N., Gres, A.T., Wolf, J.J., Coonrod, E.M., Tang, J., Laughlin, T.G., Yang, Q., Puray-Chavez, M., Ji, J., Singh, K., Kirby, K.A., Wang, Z., Sarafianos, S.G., 2018a. Novel Hepatitis B Virus Capsid-Targeting Antiviral that Aggregates Core Particles and Inhibits Nuclear Entry of Viral Cores. ACS Infect Dis acsinfecdis.8b00235. doi:10.1021/acsinfecdis.8b00235

Huber, A.D., Wolf, J.J., Liu, D., Gres, A.T., Tang, J., Boschert, K.N., Puray-Chavez, M.N., Pineda, D.L., Laughlin, T.G., Coonrod, E.M., Yang, Q., Ji, J., Kirby, K.A., Wang, Z., Sarafianos, S.G., 2018b. The Heteroaryldihydropyrimidine Bay 38-7690 Induces Hepatitis B Virus Core Protein Aggregates Associated with Promyelocytic Leukemia Nuclear Bodies in Infected Cells. mSphere 3, 254. doi:10.1128/mSphereDirect.00131-18

Hung, C.-H., Hu, T.-H., Lu, S.-N., Lee, C.-M., Chen, C.-H., Kee, K.-M., Wang, J.-H., Tsai, M.-C., Kuo, Y.-H., Chang, K.-C., Chiu, Y.-C., Chen, C.-H., 2015. Tenofovir versus entecavir in treatment of chronic hepatitis B virus with severe acute exacerbation. Antimicrob. Agents Chemother. 59, 3168–3173. doi:10.1128/AAC.00261-15

Hyams, K.C., 1995. Risks of Chronicity Following Acute Hepatitis B Virus Infection: A Review. Clinical Infection Diseases 20, 992–1000.

Jansen, L., Kootstra, N.A., van Dort, K.A., Takkenberg, R.B., Reesink, H.W., Zaaijer, H.L., 2016. Hepatitis B Virus Pregenomic RNA Is Present in Virions in Plasma and Is Associated With a Response to Pegylated Interferon Alfa-2a and Nucleos(t)ide Analogues. J. Infect. Dis. 213, 224–232. doi:10.1093/infdis/jiv397

Jazayeri, S.M., Alavian, S.M., Carman, W.F., 2010. Hepatitis B virus: origin and evolution. J. Viral Hepat. 17, 229–235. doi:10.1111/j.1365-2893.2009.01193.x

Jeong, H., Cho, M.-H., Park, S.-G., Jung, G., 2014. Interaction between nucleophosmin and HBV core protein increases HBV capsid assembly. FEBS Letters 588, 851–858. doi:10.1016/j.febslet.2014.01.020

Jones, S.A., Clark, D.N., Cao, F., Tavis, J.E., Hu, J., 2014. Comparative analysis of hepatitis B virus polymerase sequences required for viral RNA binding, RNA packaging, and protein priming. J. Virol. 88, 1564–1572. doi:10.1128/JVI.02852-13

Jones, S.A., Hu, J., 2013a. Protein-primed terminal transferase activity of hepatitis B virus polymerase. J. Virol. 87, 2563–2576. doi:10.1128/JVI.02786-12

Jones, S.A., Hu, J., 2013b. Hepatitis B virus reverse transcriptase: diverse functions as classical and emerging targets for antiviral intervention. Emerg Microbes Infect 2, e56–11. doi:10.1038/emi.2013.56

Kang, H., Yu, J., Jung, G., 2008. Phosphorylation of hepatitis B virus core C-terminally truncated protein (Cp149) by PKC increases capsid assembly and stability. Biochem. J. 416, 47–54. doi:10.1042/BJ20080724

Kang, J.-A., Kim, S., Park, M., Park, H.-J., Kim, J.-H., Park, S., Hwang, J.-R., Kim, Y.-C., Jun Kim, Y., Cho, Y., Sun Jin, M., Park, S.-G., 2019. Ciclopirox inhibits Hepatitis B Virus secretion by blocking capsid assembly. Nat Commun 10, 2184. doi:10.1038/s41467-019-10200-5

Kann, M., Gerlich, W.H., 2005. Structure and Molecular Virology, in: Viral Hepatitis, Third Edition. John Wiley & Sons, Ltd, Oxford, UK, pp. 147–180. doi:10.1002/9780470987131.ch10

Kann, M., Gerlich, W.H., 1994. Effect of core protein phosphorylation by protein kinase C on encapsidation of RNA within core particles of hepatitis B virus. J. Virol. 68, 7993–8000.

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. The Journal of Cell Biology 145, 45–55.

Katen, S.P., Chirapu, S.R., Finn, M.G., Zlotnick, A., 2010. Trapping of Hepatitis B Virus Capsid Assembly Intermediates by Phenylpropenamide Assembly Accelerators. ACS Chem. Biol. 5, 1125–1136. doi:10.1021/cb100275b

Katen, S.P., Tan, Z., Chirapu, S.R., Finn, M.G., Zlotnick, A., 2013. Assembly-directed antivirals differentially bind quasiequivalent pockets to modify hepatitis B virus capsid tertiary and quaternary structure. Structure 21, 1406–1416. doi:10.1016/j.str.2013.06.013

Kay, A., Zoulim, F., 2007. Hepatitis B virus genetic variability and evolution. Virus Res. 127, 164–176. doi:10.1016/j.virusres.2007.02.021

Keasler, V.V., Hodgson, A.J., Madden, C.R., Slagle, B.L., 2009. Hepatitis B virus HBx protein localized to the nucleus restores HBx-deficient virus replication in HepG2 cells and in vivo in hydrodynamically-injected mice. Virology 390, 122–129. doi:10.1016/j.virol.2009.05.001

Kim, J., Wu, J., 2015. A Thermodynamic Model for Genome Packaging in Hepatitis B Virus. Biophysical Journal 109, 1689–1697. doi:10.1016/j.bpj.2015.08.021

Kim, J., Wu, J., 2014. A theoretical study of SRPK interaction with the flexible domains of hepatitis B capsids. Biophysical Journal 107, 1453–1461. doi:10.1016/j.bpj.2014.07.032

Kim, S., Lee, J., Ryu, W.-S., 2009. Four conserved cysteine residues of the hepatitis B virus polymerase are critical for RNA pregenome encapsidation. J. Virol. 83, 8032–8040. doi:10.1128/JVI.00332-09

Kimura, T., Ohno, N., Terada, N., Rokuhara, A., Matsumoto, A., Yagi, S., Tanaka, E., Kiyosawa, K., Ohno, S., Maki, N., 2005. Hepatitis B virus DNA-negative dane particles lack core protein but contain a 22-kDa precore protein without C-terminal arginine-rich domain. J. Biol. Chem. 280, 21713–21719. doi:10.1074/jbc.M501564200

King, R.W., Ladner, S.K., Miller, T.J., Zaifert, K., Perni, R.B., Conway, S.C., Otto, M.J., 1998. Inhibition of human hepatitis B virus replication by AT-61, a phenylpropenamide derivative, alone and in combination with (-)beta-L-2',3'-dideoxy-3'-thiacytidine. Antimicrob. Agents Chemother. 42, 3179–3186.

Klumpp, K., Lam, A.M., Lukacs, C., Vogel, R., Ren, S., Espiritu, C., Baydo, R., Atkins, K., Abendroth, J., Liao, G., Efimov, A., Hartman, G., Flores, O.A., 2015. High-resolution crystal structure of a hepatitis B virus replication inhibitor bound to the viral core protein. Proc. Natl. Acad. Sci. U.S.A. 112, 15196–15201. doi:10.1073/pnas.1513803112

Ko, C., Shin, Y.-C., Park, W.-J., Kim, S., Kim, J., Ryu, W.-S., 2014. Residues Arg703, Asp777, and Arg781 of the RNase H domain of hepatitis B virus polymerase are critical for viral DNA synthesis. J. Virol. 88, 154–163. doi:10.1128/JVI.01916-13

Komla-Soukha, I., Sureau, C., 2006. A tryptophan-rich motif in the carboxyl terminus of the small envelope protein of hepatitis B virus is central to the assembly of hepatitis delta virus particles. J. Virol. 80, 4648–4655. doi:10.1128/JVI.80.10.4648-4655.2006

Kondylis, P., Schlicksup, C.J., Brunk, N.E., Zhou, J., Zlotnick, A., Jacobson, S.C., 2019a. Competition between Normative and Drug-Induced Virus Self-Assembly Observed with Single-Particle Methods. J. Am. Chem. Soc. 141, 1251–1260. doi:10.1021/jacs.8b10131

Kondylis, P., Schlicksup, C.J., Katen, S.P., Lee, L.S., Zlotnick, A., Jacobson, S.C., 2019b. Evolution of Intermediates during Capsid Assembly of Hepatitis B Virus with Phenylpropenamide-Based Antivirals. ACS Infect Dis 5, 769–777. doi:10.1021/acsinfecdis.8b00290

Kouwaki, T., Okamoto, T., Ito, A., Sugiyama, Y., Yamashita, K., Suzuki, T., Kusakabe, S., Hirano, J., Fukuhara, T., Yamashita, A., Saito, K., Okuzaki, D., Watashi, K., Sugiyama, M., Yoshio, S., Standley, D.M., Kanto, T., Mizokami, M., Moriishi, K., Matsuura, Y., 2016. Hepatocyte Factor JMJD5 Regulates Hepatitis B Virus Replication through Interaction with HBx. J. Virol. 90, 3530–3542. doi:10.1128/JVI.02776-15

Kramvis, A., 2014. Genotypes and genetic variability of hepatitis B virus. Intervirology 57, 141–150. doi:10.1159/000360947

Krugman, S., Overby, L.R., Mushahwar, I.K., Ling, C.M., Frösner, G.G., Deinhardt, F., 1979. Viral hepatitis, type B. Studies on natural history and prevention re-examined. The New England Journal of Medicine 300, 101–106. doi:10.1056/NEJM197901183000301

Krull, S., Thyberg, J., Björkroth, B., Rackwitz, H.-R., Cordes, V.C., 2004. Nucleoporins as components of the nuclear pore complex core structure and Tpr as the architectural element of the nuclear basket. Mol. Biol. Cell 15, 4261–4277. doi:10.1091/mbc.e04-03-0165

Lahlali, T., Berke, J.M., Vergauwen, K., Foca, A., Vandyck, K., Pauwels, F., Zoulim, F., Durantel, D., 2018. Novel Potent Capsid Assembly Modulators Regulate Multiple Steps of the Hepatitis B Virus Life Cycle. Antimicrob. Agents Chemother. 62, 672. doi:10.1128/AAC.00835-18

Lam, A.M., Espiritu, C., Vogel, R., Ren, S., Lau, V., Kelly, M., Kuduk, S.D., Hartman, G.D., Flores, O.A., Klumpp, K., 2019. Preclinical Characterization of NVR 3-778, a First-in-Class Capsid Assembly Modulator against Hepatitis B Virus. Antimicrob. Agents Chemother. 63, e01734–18. doi:10.1128/AAC.01734-18

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. Antimicrob. Agents Chemother. 61, 2212. doi: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 gamma 2-adaptin. J. Virol. 81, 9050–9060. doi:10.1128/JVI.00479-07

Lambert, C., Prange, R., 2001. Dual topology of the hepatitis B virus large envelope protein: determinants influencing post-translational pre-S translocation. J. Biol. Chem. 276, 22265–22272. doi:10.1074/jbc.M100956200

Lamontagne, R.J., Bagga, S., Bouchard, M.J., 2016. Hepatitis B virus molecular biology and pathogenesis. HR 2, 163–48. doi:10.20517/2394-5079.2016.05

Lanford, R.E., Notvall, L., Lee, H., Beames, B., 1997. Transcomplementation of nucleotide priming and reverse transcription between independently expressed TP and RT domains of the hepatitis B virus reverse transcriptase. J. Virol. 71, 2996–3004.

Lauber, C., Seitz, S., Mattei, S., Suh, A., Beck, J., Herstein, J., Börold, J., Salzburger, W., Kaderali, L., Briggs, J.A.G., Bartenschlager, R., 2017. Deciphering the Origin and Evolution

of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses. Cell Host Microbe 22, 387–399.e6. doi:10.1016/j.chom.2017.07.019

Lavanchy, D., 2004. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. J. Viral Hepat. 1–11.

Le Pogam, S., Yuan, 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. J. Virol. 74, 9099–9105.

Le Seyec, J., Chouteau, P., Cannie, I., Guguen-Guillouzo, C., Gripon, P., 1998. Role of the pre-S2 domain of the large envelope protein in hepatitis B virus assembly and infectivity. J. Virol. 72, 5573–5578.

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

Lee, H., Jeong, H., Lee, S.Y., Kim, S.S., Jang, K.L., 2019. Hepatitis B Virus X Protein Stimulates Virus Replication Via DNA Methylation of the C-1619 in Covalently Closed Circular DNA. Mol. Cells 42, 67–78. doi:10.14348/molcells.2018.0255

Levrero, M., Testoni, B., Zoulim, F., 2016. HBV cure: why, how, when? Curr Opin Virol 18, 135–143. doi:10.1016/j.coviro.2016.06.003

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. J. Virol. 81, 9202–9215. doi:10.1128/JVI.00390-07

Li, Y., Sun, Y., Sun, F., Hua, R., Li, C., Chen, L., Guo, D., Mu, J., 2018. Mechanisms and Effects on HBV Replication of the Interaction between HBV Core Protein and Cellular Filamin B. Virol Sin 33, 162–172. doi:10.1007/s12250-018-0023-4

Liang, T.J., 2009. Hepatitis B: The virus and disease. Hepatology 49, S13–S21. doi:10.1002/hep.22881

Liang, T.J., Ghany, M., 2002. Hepatitis B e Antigen--the dangerous endgame of hepatitis B. The New England Journal of Medicine 347, 208–210. doi:10.1056/NEJMe020060

Lin, C.L., Kao, J.H., 2016. Review article: novel therapies for hepatitis B virus cure - advances and perspectives. Aliment Pharmacol Ther 44, 213–222. doi:10.1111/apt.13694

Littlejohn, M., Locarnini, S., Yuen, L., 2016. Origins and Evolution of Hepatitis B Virus and Hepatitis D Virus. Cold Spring Harbor Perspectives in Medicine 6, a021360. doi:10.1101/cshperspect.a021360

Liu, C., Fan, G., Wang, Z., Chen, H.-S., Yin, C.-C., 2017. Allosteric conformational changes of human HBV core protein transform its assembly. Scientific RepoRts 1–9. doi:10.1038/s41598-017-01568-9

Liu, K., Hu, J., 2019. Secretion of empty or complete hepatitis B virions: envelopment of empty capsids versus mature nucleocapsids. Future Virology 14, 95–105. doi:10.2217/fvl-2018-0128◯C

Liu, K., Luckenbaugh, L., Ning, X., Xi, J., Hu, J., 2018. Multiple roles of core protein linker in hepatitis B virus replication. PLoS Pathog 14, e1007085–25. doi:10.1371/journal.ppat.1007085

Liu, S., He, J., Shih, C., Li, K., Dai, A., Zhou, Z.H., Zhang, J., 2010. Structural comparisons of hepatitis B core antigen particles with different C-terminal lengths. Virus Res. 149, 241–244. doi:10.1016/j.virusres.2010.01.020

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, 561–575. doi:10.1016/j.semcancer.2013.08.006

Locarnini, S., Zoulim, F., 2010. Review Molecular genetics of HBV infection. Antiviral Therapy 1–12. doi:10.3851/IMP1619)

Lok, A.S.-F., 2002. Chronic Hepatitis B. The New England Journal of Medicine 346, 1682–1683.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., Alvarado, M., Anderson, H.R., Anderson, L.M., Andrews, K.G., Atkinson, C., Baddour, L.M., Barker-Collo, S., Bartels, D.H., Bell, M.L., Benjamin, E.J., Bennett, D., Bhalla, K., Bikbov, B., Bin Abdulhak, A., Birbeck, G., Blyth, F., Bolliger, I., Boufous, S., Bucello, C., Burch, M., Burney, P., Carapetis, J., Chen, H., Chou, D., Chugh, S.S., Coffeng, L.E., Colan, S.D., Colquhoun, S., Colson, K.E., Condon, J., Connor, M.D., Cooper, L.T., Corriere, M., Cortinovis, M., de Vaccaro, K.C., Couser, W., Cowie, B.C., Criqui, M.H., Cross, M., Dabhadkar, K.C., Dahodwala, N., De Leo, D., Degenhardt, L., Delossantos, A., Denenberg, J., Jarlais, Des, D.C., Dharmaratne, S.D., Dorsey, E.R., Driscoll, T., Duber, H., Ebel, B., Erwin, P.J., Espindola, P., Ezzati, M., Feigin, V., Flaxman, A.D., Forouzanfar, M.H., Fowkes, F.G.R., Franklin, R., Fransen, M., Freeman, M.K., Gabriel, S.E., Gakidou, E., Gaspari, F., Gillum, R.F., Gonzalez-Medina, D., Halasa, Y.A., Haring, D., Harrison, J.E., Havmoeller, R., Hay, R.J., Hoen, B., Hotez, P.J., Hoy, D., Jacobsen, K.H., James, S.L., Jasrasaria, R., Jayaraman, S., Johns, N., Karthikeyan, G., Kassebaum, N., Keren, A., Khoo, J.-P., Knowlton, L.M., Kobusingye, O., Koranteng, A., Krishnamurthi, R., Lipnick, M., Lipshultz, S.E., Ohno, S.L., Mabweijano, J., MacIntyre, M.F., Mallinger, L., March, L., Marks, G.B., Marks, R., Matsumori, A., Matzopoulos, R., Mayosi, B.M., McAnulty, J.H., McDermott, M.M., McGrath, J., Mensah, G.A., Merriman, T.R., Michaud, C., Miller, M., Miller, T.R., Mock, C., Mocumbi, A.O., Mokdad, A.A., Moran, A., Mulholland, K., Nair, M.N., Naldi, L., Narayan, K.M.V., Nasseri, K., Norman, P., O'Donnell, M., Omer, S.B., Ortblad, K., Osborne, R., Ozgediz, D., Pahari, B., Pandian, J.D., Rivero, A.P., Padilla, R.P., Perez-Ruiz, F., Perico, N., Phillips, D., Pierce, K., Pope, C.A., Porrini, E., Pourmalek, F., Raju, M., Ranganathan, D., Rehm, J.T., Rein, D.B., Remuzzi, G., Rivara, F.P., Roberts, T., De León, F.R., Rosenfeld, L.C., Rushton, L., Sacco, R.L., Salomon, J.A., Sampson, U., Sanman, E., Schwebel, D.C., Segui-Gomez, M., Shepard, D.S., Singh, D., Singleton, J., Sliwa, K., Smith, E., Steer, A., Taylor, J.A., Thomas, B., Tleyjeh, I.M., Towbin, J.A., Truelsen, T., Undurraga, E.A., Venketasubramanian, N., Vijayakumar, L., Vos, T., Wagner, G.R., Wang, M., Wang, W., Watt, K., Weinstock, M.A., Weintraub, R., Wilkinson, J.D., Woolf, A.D., Wulf, S., Yeh, P.-H., Yip, P., Zabetian, A., Zheng, Z.-J., Lopez, A.D., Murray, C.J.L., AlMazroa, M.A., Memish, Z.A., 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380, 2095–2128. doi:10.1016/S0140-6736(12)61728-0

Löffler-Mary, H., Dumortier, J., Klentsch-Zimmer, C., Prange, R., 2000. Hepatitis B virus assembly is sensitive to changes in the cytosolic S loop of the envelope proteins. Virology 270, 358–367. doi:10.1006/viro.2000.0268

Lucifora, J., Arzberger, S., Durantel, D., Belloni, L., Strubin, M., Levrero, M., Zoulim, F., Hantz, O., Protzer, U., 2011. Hepatitis B virus X protein is essential to initiate and maintain virus

replication after infection. Journal of Hepatology 55, 996–1003. doi:10.1016/j.jhep.2011.02.015

Lucifora, J., Salvetti, A., Marniquet, X., Mailly, L., Testoni, B., Fusil, F., Inchauspé, A., Michelet, M., Michel, M.-L., Levrero, M., Cortez, P., Baumert, T.F., Cosset, F.-L., Challier, C., Zoulim, F., Durantel, D., 2017. Detection of the hepatitis B virus (HBV) covalently-closed-circular DNA (cccDNA) in mice transduced with a recombinant AAV-HBV vector. Antiviral Research 145, 14–19. doi:10.1016/j.antiviral.2017.07.006

Luckenbaugh, L., Kitrinos, K.M., Delaney, W.E., IV, Hu, J., 2015. Genome-free hepatitis B virion levels in patient sera as a potential marker to monitor response to antiviral therapy. J. Viral Hepat. 22, 561–570. doi:10.1111/jvh.12361

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. J. Virol. 90, 5830–5844. doi:10.1128/JVI.00394-16

Luo, L., Chen, S., Gong, Q., Luo, N., Lei, Y., Guo, J., He, S., 2013. Hepatitis B virus X protein modulates remodelling of minichromosomes related to hepatitis B virus replication in HepG2 cells. Int. J. Mol. Med. 31, 197–204. doi:10.3892/ijmm.2012.1165

Luo, X., Huang, Y., Chen, Y., Tu, Z., Hu, J., Tavis, J.E., Huang, A., Hu, Y., 2016. Association of Hepatitis B Virus Covalently Closed Circular DNA and Human APOBEC3B in Hepatitis B Virus-Related Hepatocellular Carcinoma. PLoS ONE 11, e0157708. doi:10.1371/journal.pone.0157708

Lutomski, C.A., Lyktey, N.A., Pierson, E.E., Zhao, Z., Zlotnick, A., Jarrold, M.F., 2018. Multiple Pathways in Capsid Assembly. J. Am. Chem. Soc. 140, 5784–5790. doi:10.1021/jacs.8b01804

Lutomski, C.A., Lyktey, N.A., Zhao, Z., Pierson, E.E., Zlotnick, A., Jarrold, M.F., 2017. Hepatitis B Virus Capsid Completion Occurs through Error Correction. J. Am. Chem. Soc. 139, 16932–16938. doi:10.1021/jacs.7b09932

Maasoumy, B., Wiegand, S.B., Jaroszewicz, J., Bremer, B., Lehmann, P., Deterding, K., Taranta, A., Manns, M.P., Wedemeyer, H., Glebe, D., Cornberg, M., 2015. Hepatitis B core-related antigen (HBcrAg) levels in the natural history of hepatitis B virus infection in a large European cohort predominantly infected with genotypes A and D. Clin. Microbiol. Infect. 21, 606.e1–10. doi:10.1016/j.cmi.2015.02.010

MacCallum, F.O., 1947. Homologous Serum Hepatitis

. Proceedings of the Royal Society of Medicine 39(10), 655–657.

Machida, A., Ohnuma, H., Tsuda, F., Yoshikawa, A., Hoshi, Y., Tanaka, T., Kishimoto, S., Akahane, Y., Miyakawa, Y., Mayumi, M., 1991. Phosphorylation in the carboxyl-terminal domain of the capsid protein of hepatitis B virus: evaluation with a monoclonal antibody. J. Virol. 65, 6024–6030.

Magnius, L., Kaplan, L., Vyas, G.N., Perkins, H.A., 1975. A new virus specified determinant of hepatitis B surface antigen. Acta Pathol Microbiol Scand B 83, 295–297.

Mak, L.-Y., Wong, D.K.-H., Seto, W.-K., Lai, C.-L., Yuen, M.-F., 2017. Hepatitis B core protein as a therapeutic target. Expert Opin. Ther. Targets 21, 1153–1159. doi:10.1080/14728222.2017.1397134

Mani, N., Cole, A.G., Phelps, J.R., Ardzinski, A., Cobarrubias, K.D., Cuconati, A., Dorsey, B.D., Evangelista, E., Fan, K., Guo, F., Guo, H., Guo, J.-T., Harasym, T.O., Kadhim, S., Kultgen, S.G., Lee, A.C.H., Li, A.H.L., Long, Q., Majeski, S.A., Mao, R., McClintock, K.D., Reid,

S.P., Rijnbrand, R., Snead, N.M., Micolochick Steuer, H.M., Stever, K., Tang, S., Wang, X., Zhao, Q., Sofia, M.J., 2018. Preclinical Profile of AB-423, an Inhibitor of Hepatitis B Virus Pregenomic RNA Encapsidation. Antimicrob. Agents Chemother. 62, 1893. doi:10.1128/AAC.00082-18

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S., 2002. The Protein Kinase Complement of the Human Genome. Science 298, 1912–1934.

Mantzoukis, K., Rodríguez-Perálvarez, M., Buzzetti, E., Thorburn, D., Davidson, B.R., Tsochatzis, E., Gurusamy, K.S., 2017. Pharmacological interventions for acute hepatitis B infection: an attempted network meta-analysis. Cochrane Database Syst Rev 3, CD011645. doi:10.1002/14651858.CD011645.pub2

Maurizia Rossana Brunetto, 2010. A new role for an old marker, HBsAg. Journal of Hepatology 52, 475–477. doi:10.1016/j.jhep.2009.12.020

McMahon, B.J., 2009. The natural history of chronic hepatitis B virus infection. Hepatology 49, S45–S55. doi:10.1002/hep.22898

Meier, P., Scougall, C.A., Will, H., Burrell, C.J., Jilbert, A.R., 2003. A duck hepatitis B virus strain with a knockout mutation in the putative X ORF shows similar infectivity and in vivo growth characteristics to wild-type virus. Virology 317, 291–298.

Melegari, M., Scaglioni, P.P., Wands, J.R., 1998. Cloning and characterization of a novel hepatitis B virus x binding protein that inhibits viral replication. J. Virol. 72, 1737–1743.

Menéndez-Arias, L., Álvarez, M., Pacheco, B., 2014. Nucleoside/nucleotide analog inhibitors of hepatitis B virus polymerase: mechanism of action and resistance. Curr Opin Virol 8, 1–9. doi:10.1016/j.coviro.2014.04.005

Meng, D., Hjelm, R.P., Hu, J., Wu, J., 2011. A theoretical model for the dynamic structure of hepatitis B nucleocapsid. Biophysical Journal 101, 2476–2484. doi:10.1016/j.bpj.2011.10.002

Messageot, F., Salhi, S., Eon, P., Rossignol, J.-M., 2003. Proteolytic processing of the hepatitis B virus e antigen precursor. Cleavage at two furin consensus sequences. J. Biol. Chem. 278, 891–895. doi:10.1074/jbc.M207634200

Milich, D., Liang, T.J., 2003. Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection. Hepatology 38, 1075–1086. doi: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. J. Immunol. 160, 2013–2021.

Moriya, T., Kuramoto, I.K., Yoshizawa, H., Holland, P.V., 2002. Distribution of hepatitis B virus genotypes among American blood donors determined with a PreS2 epitope enzyme-linked immunosorbent assay kit. J. Clin. Microbiol. 40, 877–880. doi:10.1128/JCM.40.3.877-880.2002

Mühlemann, B., Jones, T.C., Damgaard, P. de B., Allentoft, M.E., Shevnina, I., Logvin, A., Usmanova, E., Panyushkina, I.P., Boldgiv, B., Bazartseren, T., Tashbaeva, K., Merz, V., Lau, N., Smrčka, V., Voyakin, D., Kitov, E., Epimakhov, A., Pokutta, D., Vicze, M., Price, T.D., Moiseyev, V., Hansen, A.J., Orlando, L., Rasmussen, S., Sikora, M., Vinner, L., Osterhaus, A.D.M.E., Smith, D.J., Glebe, D., Fouchier, R.A.M., Drosten, C., Sjögren, K.-G., Kristiansen, K., Willerslev, E., 2018. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature 557, 418–423. doi:10.1038/s41586-018-0097-z

Nassal, M., 2015. HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. Gut 64, 1972–1984. doi:10.1136/gutjnl-2015-309809

Nassal, M., 2008. Hepatitis B viruses: reverse transcription a different way. Virus Res. 134, 235–249. doi:10.1016/j.virusres.2007.12.024

Nassal, M., 1992. Conserved cysteines of the hepatitis B virus core protein are not required for assembly of replication-competent core particles nor for their envelopment. Virology 190, 499–505.

Nassal, M., Rieger, A., Steinau, O., 1992. Topological analysis of the hepatitis B virus core particle by cysteine-cysteine cross-linking. J. Mol. Biol. 225, 1013–1025.

Negro, F., 2014. Hepatitis D virus coinfection and superinfection. Cold Spring Harbor Perspectives in Medicine 4, a021550–a021550. doi:10.1101/cshperspect.a021550

Newbold, J.E., Xin, H., Tencza, M., Sherman, G., Dean, J., Bowden, S., Locarnini, S., 1995. The covalently closed duplex form of the hepadnavirus genome exists in situ as a heterogeneous population of viral minichromosomes. J. Virol. 69, 3350–3357.

Nguyen, H.D., Reddy, V.S., Brooks, C.L., 2007. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. Nano Lett. 7, 338–344. doi:10.1021/nl062449h

Ni, Y., Lempp, F.A., Mehrle, S., Nkongolo, S., Kaufman, C., Fälth, M., Stindt, J., Königer, C., Nassal, M., Kubitz, R., Sültmann, H., Urban, S., 2014. Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. Gastroenterology 146, 1070–1083. doi:10.1053/j.gastro.2013.12.024

Nijampatnam, B., Liotta, D.C., 2019. Recent advances in the development of HBV capsid assembly modulators. Curr Opin Chem Biol 50, 73–79. doi:10.1016/j.cbpa.2019.02.009

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. J. Virol. 91, 1. doi: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. J. Virol. 92, JVI.00272–18. doi: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 Para-retrovirus. PLoS Pathog 7, e1002255–14. doi:10.1371/journal.ppat.1002255

Norder, H., Couroucé, A.-M., Coursaget, P., Echevarria, J.M., Lee, S.-D., Mushahwar, I.K., Robertson, B.H., Locarnini, S., Magnius, L.O., 2004. Genetic Diversity of Hepatitis B Virus Strains Derived Worldwide: Genotypes, Subgenotypes, and HB_sAg Subtypes. Intervirology 47, 289–309. doi:10.1159/000080872

Norder, H., Couroucé, A.M., Magnius, L.O., 1992. Molecular basis of hepatitis B virus serotype variations within the four major subtypes. J. Gen. Virol. 73 (Pt 12), 3141–3145. doi:10.1099/0022-1317-73-12-3141

Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R.I., Imai, M., Miyakawa, Y., Mayumi, M., 1988. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J. Gen. Virol. 69 (Pt 10), 2575–2583. doi:10.1099/0022-1317-69-10-2575

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. J. Virol. 89, 9281–9287. doi:10.1128/JVI.00466-15

Ostapchuk, P., Hearing, P., Ganem, D., 1994. A dramatic shift in the transmembrane topology of a viral envelope glycoprotein accompanies hepatitis B viral morphogenesis. The EMBO Journal 13, 1048–1057.

Packianathan, C., Katen, S.P., Dann, C.E., Zlotnick, A., 2010. Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly. J. Virol. 84, 1607–1615. doi:10.1128/JVI.02033-09

Paraskevis, D., Magiorkinis, G., Magiorkinis, E., Ho, S.Y.W., Belshaw, R., Allain, J.-P., Hatzakis, A., 2013. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology 57, 908–916. doi:10.1002/hep.26079

Parkin, D.M., Bray, F., Ferlay, J., Pisani, P., 2001. Estimating the world cancer burden: Globocan 2000. Int. J. Cancer 94, 153–156.

Patel, N., White, S.J., Thompson, R.F., Bingham, R., Weiß, E.U., Maskell, D.P., Zlotnick, A., Dykeman, E.C., Tuma, R., Twarock, R., Ranson, N.A., Stockley, P.G., 2017. HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. Nature Microbiology 1–10. doi:10.1038/nmicrobiol.2017.98

Pei, Y., Wang, C., Yan, S.F., Liu, G., 2017. Past, Current, and Future Developments of Therapeutic Agents for Treatment of Chronic Hepatitis B Virus Infection. J. Med. Chem. 60, 6461–6479. doi:10.1021/acs.jmedchem.6b01442

Perlman, D.H., Berg, E.A., O'connor, P.B., Costello, C.E., Hu, J., 2005. Reverse transcription-associated dephosphorylation of hepadnavirus nucleocapsids. Proc. Natl. Acad. Sci. U.S.A. 102, 9020–9025. doi:10.1073/pnas.0502138102

Perni, R.B., Conway, S.C., Ladner, S.K., Zaifert, K., Otto, M.J., King, R.W., 2000. Phenylpropenamide derivatives as inhibitors of hepatitis B virus replication. Bioorg. Med. Chem. Lett. 10, 2687–2690.

Persing, D.H., Varmus, H.E., Ganem, D., 1987. The preS1 protein of hepatitis B virus is acylated at its amino terminus with myristic acid. J. Virol. 61, 1672–1677.

Perz, J.F., Armstrong, G.L., Farrington, L.A., Hutin, Y.J.F., Bell, B.P., 2006. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. Journal of Hepatology 45, 529–538. doi:10.1016/j.jhep.2006.05.013

Petit, M.A., Pillot, J., 1985. HBc and HBe antigenicity and DNA-binding activity of major core protein P22 in hepatitis B virus core particles isolated from the cytoplasm of human liver cells. J. Virol. 53, 543–551.

Pierson, E.E., Keifer, D.Z., Selzer, L., Lee, L.S., Contino, N.C., Wang, J.C.Y., Zlotnick, A., Jarrold, M.F., 2014. Detection of late intermediates in virus capsid assembly by charge detection mass spectrometry. J. Am. Chem. Soc. 136, 3536–3541. doi:10.1021/ja411460w

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, 115–120. doi:10.1006/viro.1996.8367

Poisson, F.E.A., 1997. Both Pre-S1 and S Domains of Hepatitis B Virus Envelope Proteins Interact with the Core Particle 1–6.

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. J. Virol. 77, 416–422. doi:10.1128/JVI.77.1.416-422.2003

Porterfield, J.Z., Dhason, M.S., Loeb, D.D., Nassal, M., Stray, S.J., Zlotnick, A., 2010. Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. J. Virol. 84, 7174–7184. doi:10.1128/JVI.00586-10

Pourkarim, M.R., Amini-Bavil-Olyaee, S., Kurbanov, F., Van Ranst, M., Tacke, F., 2014. Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. World J. Gastroenterol. 20, 7152–7168. doi:10.3748/wjg.v20.i23.7152

Prange, R., 2012. Host factors involved in hepatitis B virus maturation, assembly, and egress. Med Microbiol Immunol 201, 449–461. doi:10.1007/s00430-012-0267-9

Prange, R., Streeck, R.E., 1995. Novel transmembrane topology of the hepatitis B virus envelope proteins. The EMBO Journal 14, 247–256.

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. J. Virol. 80, 5465–5473. doi:10.1128/JVI.02303-05

Rabe, B., Vlachou, A., Pante, N., Helenius, A., Kann, M., 2003. Nuclear import of hepatitis B virus capsids and release of the viral genome. Proceedings of the National Academy of Sciences 100, 9849–9854.

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, 1281–1297. doi: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. J. Virol. 89, 999–1012. doi:10.1128/JVI.02440-14

Rivière, L., Gerossier, L., Ducroux, A., Dion, S., Deng, Q., Michel, M.-L., Buendia, M.-A., Hantz, O., Neuveut, C., 2015. HBx relieves chromatin-mediated transcriptional repression of hepatitis B viral cccDNA involving SETDB1 histone methyltransferase. Journal of Hepatology 63, 1093–1102. doi:10.1016/j.jhep.2015.06.023

Roingeard, P., Sureau, C., 1998. Ultrastructural analysis of hepatitis B virus in HepG2-transfected cells with special emphasis on subviral filament morphogenesis. Hepatology 28, 1128–1133. doi:10.1002/hep.510280431

Roseman, A.M., Berriman, J.A., Wynne, S.A., Butler, P.J.G., Crowther, R.A., 2005a. A structural model for maturation of the hepatitis B virus core. Proceedings of the National Academy of Sciences 102, 15821–15826.

Roseman, A.M., Berriman, J.A., Wynne, S.A., Butler, P.J.G., Crowther, R.A., 2005b. A structural model for maturation of the hepatitis B virus core. Proc. Natl. Acad. Sci. U.S.A. 102, 15821–15826. doi:10.1073/pnas.0504874102

Scaglioni, P.P., Melegari, M., Wands, J.R., 1997. Posttranscriptional regulation of hepatitis B virus replication by the precore protein. J. Virol. 71, 345–353.

Schinazi, R.F., BOUCLE, S., Amblard, F., SARI, O., Bassit, L., Emory University, 2017. Elimination of hepatitis B virus with antiviral agents. PCT Application. WO2017156255A1.

Schinazi, R.F., Ehteshami, M., Bassit, L., Asselah, T., 2018. Towards HBV curative therapies. Liver Int 38 Suppl 1, 102–114. doi:10.1111/liv.13656

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. doi:10.1016/j.virol.2014.04.030

Schlicksup, C.J., Wang, J.C.-Y., Francis, S., Venkatakrishnan, B., Turner, W.W., VanNieuwenhze, M., Zlotnick, A., 2018a. Hepatitis B virus core protein allosteric modulators can distort and disrupt intact capsids. Elife 7, 1–23. doi:10.7554/eLife.31473.001

Schlicksup, C.J., Wang, J.C.-Y., Francis, S., Venkatakrishnan, B., Turner, W.W., VanNieuwenhze, M., Zlotnick, A., 2018b. Hepatitis B virus core protein allosteric modulators can distort and disrupt intact capsids. Elife 7, 13046. doi:10.7554/eLife.31473

Schmilovitz-Weiss, H., Levy, M., Thompson, N., Dusheiko, G., 1993. Viral markers in the treatment of hepatitis B and C. Gut 34, S26–35. doi:10.1136/gut.34.2_Suppl.S26

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, 1759–1768. doi:10.1002/hep.21896

Seeger, C., Hu, J., 1997. Why are hepadnaviruses DNA and not RNA viruses? Trends in Microbiology 5, 447–450.

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

Seifer, M., Standring, D.N., 1994. A protease-sensitive hinge linking the two domains of the hepatitis B virus core protein is exposed on the viral capsid surface. J. Virol. 68, 5548–5555.

Seitz, S., Iancu, C., Volz, T., Mier, W., Dandri, M., Urban, S., Bartenschlager, R., 2016. A Slow Maturation Process Renders Hepatitis B Virus Infectious. Cell Host Microbe 20, 25–35. doi:10.1016/j.chom.2016.05.013

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, 4160–4167. doi:10.1038/sj.emboj.7601841

Selzer, L., Katen, S.P., Zlotnick, A., 2014. The hepatitis B virus core protein intradimer interface modulates capsid assembly and stability. Journal of Biological Chemistry 53, 5496–5504. doi:10.1021/bi500732b

Selzer, L., Zlotnick, A., 2015. Assembly and Release of Hepatitis B Virus. Cold Spring Harbor Perspectives in Medicine 5, a021394–17. doi:10.1101/cshperspect.a021394

Sengupta, I., Das, D., Singh, S.P., Chakravarty, R., Das, C., 2017. Host transcription factor Speckled 110 kDa (Sp110), a nuclear body protein, is hijacked by hepatitis B virus protein X for viral persistence. J. Biol. Chem. 292, 20379–20393. doi:10.1074/jbc.M117.796839

Shi, L., Li, S., Shen, F., Li, H., Qian, S., Lee, D.H.S., Wu, J.Z., Yang, W., 2012. Characterization of nucleosome positioning in hepadnaviral covalently closed circular DNA minichromosomes. J. Virol. 86, 10059–10069. doi:10.1128/JVI.00535-12

Shim, H.Y., Quan, X., Yi, Y.-S., Jung, G., 2011. Heat shock protein 90 facilitates formation of the HBV capsid via interacting with the HBV core protein dimers. Virology 410, 161–169. doi:10.1016/j.virol.2010.11.005

Simmonds, P., 2001. Reconstructing the origins of human hepatitis viruses. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 356, 1013–1026. doi:10.1098/rstb.2001.0890

Simon, K., Lingappa, V.R., Ganem, D., 1988. Secreted hepatitis B surface antigen polypeptides are derived from a transmembrane precursor. The Journal of Cell Biology 107, 2163–2168.

Singh, S., Zlotnick, A., 2003. Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. J. Biol. Chem. 278, 18249–18255. doi:10.1074/jbc.M211408200

Slagle, B.L., Bouchard, M.J., 2018. Role of HBx in hepatitis B virus persistence and its therapeutic implications. Curr Opin Virol 30, 32–38. doi:10.1016/j.coviro.2018.01.007

Slagle, B.L., Bouchard, M.J., 2016. Hepatitis B Virus X and Regulation of Viral Gene Expression. Cold Spring Harbor Perspectives in Medicine 6, a021402–20. doi:10.1101/cshperspect.a021402

Standring, D.N., Ou, J.H., Masiarz, F.R., Rutter, W.J., 1988. A signal peptide encoded within the precore region of hepatitis B virus directs the secretion of a heterogeneous population of e antigens in Xenopus oocytes. Proc. Natl. Acad. Sci. U.S.A. 85, 8405–8409.

Stray, S.J., Bourne, C.R., Punna, S., Lewis, W.G., Finn, M.G., Zlotnick, A., 2005. A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. Proc. Natl. Acad. Sci. U.S.A. 102, 8138–8143. doi:10.1073/pnas.0409732102

Stray, S.J., Zlotnick, A., 2006. BAY 41-4109 has multiple effects on Hepatitis B virus capsid assembly. J. Mol. Recognit. 19, 542–548. doi:10.1002/jmr.801

Su, H., Yee, J.-K., 1992. Regulation of hepatitis B virus gene expression by its two enhancers. Proceedings of the National Academy of Sciences 89, 2708–2712.

Su, P.-Y., Yang, C.-J., Chu, T.-H., Chang, C.-H., Chiang, C., Tang, F.-M., Lee, C.-Y., Shih, C., 2016. HBV maintains electrostatic homeostasis by modulating negative charges from phosphoserine and encapsidated nucleic acids. Nature Publishing Group 1–19. doi:10.1038/srep38959

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, 403–415.

Sung, W.-K., Zheng, H., Li, S., Chen, R., Liu, X., Li, Y., Lee, N.P., Lee, W.H., Ariyaratne, P.N., Tennakoon, C., Mulawadi, F.H., Wong, K.F., Liu, A.M., Poon, R.T., Fan, S.T., Chan, K.L., Gong, Z., Hu, Y., Lin, Z., Wang, G., Zhang, Q., Barber, T.D., Chou, W.-C., Aggarwal, A., Hao, K., Zhou, W., Zhang, C., Hardwick, J., Buser, C., Xu, J., Kan, Z., Dai, H., Mao, M., Reinhard, C., Wang, J., Luk, J.M., 2012. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat Genet 44, 765–769. doi:10.1038/ng.2295

Sureau, C., Salisse, J., 2013. A conformational heparan sulfate binding site essential to infectivity overlaps with the conserved hepatitis B virus A-determinant. Hepatology 57, 985–994. doi:10.1002/hep.26125

Takahashi, K., Brotman, B., Usuda, S., Mishiro, S., Prince, A.M., 2000. Full-Genome Sequence Analyses of Hepatitis B Virus (HBV) Strains Recovered from Chimpanzees Infected in the Wild: Implications for an Origin of HBV. Virology 267, 58–64. doi:10.1006/viro.1999.0102

Takahashi, K., Machida, A., Funatsu, G., Nomura, M., Usuda, S., Aoyagi, S., Tachibana, K., Miyamoto, H., Imai, M., Nakamura, T., Miyakawa, Y., Mayumi, M., 1983. Immunochemical structure of hepatitis B e antigen in the serum. J. Immunol. 130, 2903–2907.

Tan, Z., Maguire, M.L., Loeb, D.D., Zlotnick, A., 2013. Genetically altering the thermodynamics and kinetics of hepatitis B virus capsid assembly has profound effects on virus replication in cell culture. J. Virol. 87, 3208–3216. doi:10.1128/JVI.03014-12

Tan, Z., Pionek, K., Unchwaniwala, N., Maguire, M.L., Loeb, D.D., Zlotnick, A., 2015. The interface between hepatitis B virus capsid proteins affects self-assembly, pregenomic RNA packaging, and reverse transcription. J. Virol. 89, 3275–3284. doi:10.1128/JVI.03545-14

Tatematsu, K., Tanaka, Y., Kurbanov, F., Sugauchi, F., Mano, S., Maeshiro, T., Nakayoshi, T., Wakuta, M., Miyakawa, Y., Mizokami, M., 2009. A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. J. Virol. 83, 10538–10547. doi:10.1128/JVI.00462-09

Tavis, J.E., Cheng, X., Hu, Y., Totten, M., Cao, F., Michailidis, E., Aurora, R., Meyers, M.J., Jacobsen, E.J., Parniak, M.A., Sarafianos, S.G., 2013. The hepatitis B virus ribonuclease H is sensitive to inhibitors of the human immunodeficiency virus ribonuclease H and integrase enzymes. PLoS Pathog 9, e1003125. doi:10.1371/journal.ppat.1003125

Tong, S., Revill, P., 2016. Overview of hepatitis B viral replication and genetic variability. Journal of Hepatology 64, S4–S16. doi:10.1016/j.jhep.2016.01.027

Tsuge, M., Hiraga, N., Akiyama, R., Tanaka, S., Matsushita, M., Mitsui, F., Abe, H., Kitamura, S., Hatakeyama, T., Kimura, T., Miki, D., Mori, N., Imamura, M., Takahashi, S., Hayes, C.N., Chayama, K., 2010. HBx protein is indispensable for development of viraemia in human hepatocyte chimeric mice. J. Gen. Virol. 91, 1854–1864. doi:10.1099/vir.0.019224-0

Tuttleman, J.S., Pourcel, C., Summers, J., 1986. Frmation of the Pool of Covalently Close ular Viral DNA in Hepadnavirus-Infected. Cell 47, 451–460. doi:https://doi.org/10.1016/0092-8674(86)90602-1

Uetrecht, C., Barbu, I.M., Shoemaker, G.K., van Duijn, E., Heck, A.J.R., 2011. Interrogating viral capsid assembly with ion mobility-mass spectrometry. Nat Chem 3, 126–132. doi:10.1038/nchem.947

Uetrecht, C., Versluis, C., Watts, N.R., Roos, W.H., Wuite, G.J.L., Wingfield, P.T., Steven, A.C., Heck, A.J.R., 2008. High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. Proc. Natl. Acad. Sci. U.S.A. 105, 9216–9220. doi:10.1073/pnas.0800406105

Urban, S., Bartenschlager, R., Kubitz, R., Zoulim, F., 2014. Strategies to inhibit entry of HBV and HDV into hepatocytes. Gastroenterology 147, 48–64. doi:10.1053/j.gastro.2014.04.030

van de Klundert, M.A.A., van den Biggelaar, M., Kootstra, N.A., Zaaijer, H.L., 2016. Hepatitis B Virus Protein X Induces Degradation of Talin-1. Viruses 8, 281. doi:10.3390/v8100281

Vandyck, K., HACHÉ, G.Y.P., LAST, S.J., MC GOWAN, D.C., Rombouts, G., VERSCHUEREN, W.G., RABOISSON, P.J.-M.B., 2014. Sulphamoylpyrrolamide derivatives and the use thereof as medicaments for the treatment of hepatitis B. US Patent Application. US20160115125A1.

Vandyck, K., KESTELEYN, B.R.R., Pieters, S.M.A., Rombouts, G., VERSCHUEREN, W.G., RABOISSON, P.J.-M.B., 2015. Glyoxamide substituted pyrrolamide derivatives and the use thereof as medicaments for the treatment of hepatitis B. PCT Application. WO2015011281.

Vandyck, K., Rombouts, G., Stoops, B., Tahri, A., Vos, A., Verschueren, W., Wu, Y., Yang, J., Hou, F., Huang, B., Vergauwen, K., Dehertogh, P., Berke, J.M., Raboisson, P., 2018. Synthesis and Evaluation of N-Phenyl-3-sulfamoyl-benzamide Derivatives as Capsid

Assembly Modulators Inhibiting Hepatitis B Virus (HBV). J. Med. Chem. 61, 6247–6260. doi:10.1021/acs.jmedchem.8b00654

Velkov, S., Ott, J., Protzer, U., Michler, T., 2018. The Global Hepatitis B Virus Genotype Distribution Approximated from Available Genotyping Data. Genes 9, 495–14. doi:10.3390/genes9100495

Vendeville, S.M.H., LAST, S.J., Demin, S.D., GROSSE, S.C., HACHÉ, G.Y.P., Hu, L., Pieters, S.M.A., Rombouts, G., Vandyck, K., VERSCHUEREN, W.G., RABOISSON, P.J.-M.B., UC, J.S.I., 2017. Cyclized sulfamoylarylamide derivatives and the use thereof as medicaments for the treatment of hepatitis B. US Patent. US20170002025A1.

Venkatakrishnan, B., Katen, S.P., Francis, S., Chirapu, S., Finn, M.G., Zlotnick, A., Sundquist, W.I., 2016. Hepatitis B Virus Capsids Have Diverse Structural Responses to Small-Molecule Ligands Bound to the Heteroaryldihydropyrimidine Pocket. J. Virol. 90, 3994–4004. doi:10.1128/JVI.03058-15

Venkatakrishnan, B., Zlotnick, A., 2016. The Structural Biology of Hepatitis B Virus: Form and Function. Annu. Rev. Virol. 3, 429–451. doi:10.1146/annurev-virology-110615-042238

Visvanathan, K., Skinner, N.A., Thompson, A.J.V., Riordan, S.M., Sozzi, V., Edwards, R., Rodgers, S., Kurtovic, J., Chang, J., Lewin, S., Desmond, P., Locarnini, S., 2007. Regulation of Toll-like receptor-2 expression in chronic hepatitis B by the precore protein. Hepatology 45, 102–110. doi:10.1002/hep.21482

Walsh, R., Locarnini, S., 2012. Hepatitis B precore protein: pathogenic potential and therapeutic promise. Yonsei Med. J. 53, 875–885. doi:10.3349/ymj.2012.53.5.875

Wang, J., Shen, T., Huang, X., Kumar, G.R., Chen, X., Zeng, Z., Zhang, R., Chen, R., Li, T., Zhang, T., Yuan, Q., Li, P.-C., Huang, Q., Colonno, R., Jia, J., Hou, J., McCrae, M.A., Gao, Z., Ren, H., Xia, N., Zhuang, H., Lu, F., 2016. Serum hepatitis B virus RNA is encapsidated pregenome RNA that may be associated with persistence of viral infection and rebound. Journal of Hepatology 65, 700–710. doi:10.1016/j.jhep.2016.05.029

Wang, J.C.-Y., Chen, C., Rayaprolu, V., Mukhopadhyay, S., Zlotnick, A., 2015. Self-Assembly of an Alphavirus Core-like Particle Is Distinguished by Strong Intersubunit Association Energy and Structural Defects. ACS Nano 9, 8898–8906. doi:10.1021/acsnano.5b02632

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 Pathog 8, e1002919. doi:10.1371/journal.ppat.1002919

Wang, P., Naduthambi, D., Mosley, R.T., Niu, C., Furman, P.A., Otto, M.J., Sofia, M.J., 2011. Phenylpropenamide derivatives: anti-hepatitis B virus activity of the Z isomer, SAR and the search for novel analogs. Bioorg. Med. Chem. Lett. 21, 4642–4647. doi:10.1016/j.bmcl.2011.05.077

Warner, N., Locarnini, S., 2012. Replication of Hepatitis B Virus, in: Zakim and Boyers Hepatology, Sixth Edition. Elsevier Inc., pp. 86–96. doi:10.1016/B978-1-4377-0881-3.00006-1

Watts, N.R., Conway, J.F., Cheng, N., Stahl, S.J., Steven, A.C., Wingfield, P.T., 2011. Role of the Propeptide in Controlling Conformation and Assembly State of Hepatitis B Virus e-Antigen. J. Mol. Biol. 409, 202–213. doi:10.1016/j.jmb.2011.03.049

Weber, O., Schlemmer, K.-H., Hartmann, E., Hagelschuer, I., Paessens, A., Graef, E., Deres, K., Goldmann, S., Niewoehner, U., Stoltefuss, J., Haebich, D., Ruebsamen-Waigmann, H., Wohlfeil, S., 2002. Inhibition of human hepatitis B virus (HBV) by a novel non-nucleosidic compound in a transgenic mouse model. Antiviral Research 54, 69–78.

Wen, Y.-M., 2004. Structural and functional analysis of full-length hepatitis B virus genomes in patients: implications in pathogenesis. J. Gastroenterol. Hepatol. 19, 485–489. doi:10.1111/j.1440-1746.2003.03158.x

WHO, 2017. Global Hepatitis Report 2017. World Health Organization 1–83.

WHO, 2016. Global Health Sector Strategies on Viral Hepatitis, 2016-2021 1–44.

WHO, 2015. Guidelines for the prevention, care and treatment of persons with chronic hepatitis B infection. World Health Organization 1–166.

Will, H., Reiser, W., Weimer, T., Pfaff, E., Büscher, M., Sprengel, R., Cattaneo, R., Schaller, H., 1987. Replication strategy of human hepatitis B virus. J. Virol. 61, 904–911.

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 assembly. Biochemistry 34, 4919–4932.

Wong, D.J., Locarnini, S.A., 2017. Molecular Virology and Life Cycle, in: Hepatitis B Virus and Liver Disease. Springer Singapore, Singapore, pp. 1–23. doi:10.1007/978-981-10-4843-2_1

Wu, G., Liu, B., Zhang, Y., Li, J., Arzumanyan, A., Clayton, M.M., Schinazi, R.F., Wang, Z., Goldmann, S., Ren, Q., Zhang, F., Feitelson, M.A., 2013. Preclinical characterization of GLS4, an inhibitor of hepatitis B virus core particle assembly. Antimicrob. Agents Chemother. 57, 5344–5354. doi:10.1128/AAC.01091-13

Wu, S., Zhao, Q., Zhang, P., Kulp, J., Hu, L., Hwang, N., Zhang, J., Block, T.M., Xu, X., Du, Y., Chang, J., Guo, J.-T., 2017. Discovery and Mechanistic Study of Benzamide Derivatives That Modulate Hepatitis B Virus Capsid Assembly. J. Virol. 91, e00519–17–18. doi:10.1128/JVI.00519-17

Wynne, S.A., Crowther, R.A., Leslie, A.G., 1999. The Crystal Structure of the Human Hepatitis B Virus Capsid. Molecular Cell 90, 5830–5844.

Xu, Z., Yen, T.S.B., Wu, L., Madden, C.R., Tan, W., Slagle, B.L., Ou, J.-H., 2002. Enhancement of hepatitis B virus replication by its X protein in transgenic mice. J. Virol. 76, 2579–2584.

Yaginuma, K., Shirakata, Y., Kobayashi, M., Koike, K., 1987. Hepatitis B virus (HBV) particles are produced in a cell culture system by transient expression of transfected HBV DNA. Proc. Natl. Acad. Sci. U.S.A. 84, 2678–2682.

Yan, H., Peng, B., Liu, Y., Xu, G., He, W., Ren, B., Jing, Z., Sui, J., Li, W., 2014. Viral Entry of Hepatitis B and D Viruses and Bile Salts Transportation Share Common Molecular Determinants on Sodium Taurocholate Cotransporting Polypeptide. J. Virol. 88, 3273–3284. doi:10.1128/JVI.03478-13

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, e00049. doi:10.7554/eLife.00049

Yan, Z., Wu, D., Hu, H., Zeng, J., Yu, X., Xu, Z., Zhou, Z., Zhou, X., Yang, G., Young, J.A.T., Gao, L., 2019. Direct Inhibition of Hepatitis B e Antigen by Core Protein Allosteric Modulator. Hepatology 386, 1546. doi:10.1002/hep.30514

Yang, L., Liu, F., Tong, X., Hoffmann, D., Zuo, J., Lu, M., 2019. Treatment of Chronic Hepatitis B Virus Infection Using Small Molecule Modulators of Nucleocapsid Assembly: Recent Advances and Perspectives. ACS Infect Dis acsinfecdis.8b00337. doi:10.1021/acsinfecdis.8b00337

Yang, L., Shi, L.-P., Chen, H.-J., Tong, X.-K., Wang, G.-F., Zhang, Y.-M., Wang, W.-L., Feng, C.-L., He, P.-L., Zhu, F.-H., Hao, Y.-H., Wang, B.-J., Yang, D.-L., Tang, W., Nan, F.-J., Zuo, J.-P., 2014. Isothiafludine, a novel non-nucleoside compound, inhibits hepatitis B virus replication through blocking pregenomic RNA encapsidation. Acta Pharmacol. Sin. 35, 410–418. doi:10.1038/aps.2013.175

Yang, L., Wang, Y.-J., Chen, H.-J., Shi, L.-P., Tong, X.-K., Zhang, Y.-M., Wang, G.-F., Wang, W.-L., Feng, C.-L., He, P.-L., Xu, Y.-B., Lu, M.-J., Tang, W., Nan, F.-J., Zuo, J.-P., 2016. Effect of a hepatitis B virus inhibitor, NZ-4, on capsid formation. Antiviral Research 125, 25–33. doi:10.1016/j.antiviral.2015.11.004

Yang, W., Summers, J., 1999. Integration of Hepadnavirus DNA in Infected Liver: Evidence for a Linear Precursor. J. Virol. 73, 9710–9717.

Yee, J.K., 1989. A liver-specific enhancer in the core promoter region of human hepatitis B virus. Science 246, 658–661.

Yeh, C.T., Ou, J.H., 1991. Phosphorylation of hepatitis B virus precore and core proteins. J. Virol. 65, 2327–2331.

Yen, T.S.B., 1993. Regulation of hepatitis B virus gene expression. Virology 4, 33–42. doi:https://doi.org/10.1016/1044-5773(93)80006-A

Yu, X., Jin, L., Jih, J., Shih, C., Zhou, Z.H., 2013. 3.5Å cryoEM structure of hepatitis B virus core assembled from full-length core protein. PLoS ONE 8, e69729. doi:10.1371/journal.pone.0069729

Yuan, T.T., Sahu, G.K., Whitehead, W.E., Greenberg, R., Shih, C., 1999a. 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. J. Virol. 73, 5731–5740.

Yuan, T.T., Tai, P.C., Shih, C., 1999b. Subtype-independent immature secretion and subtype-dependent replication deficiency of a highly frequent, naturally occurring mutation of human hepatitis B virus core antigen. J. Virol. 73, 10122–10128.

Yuen, M.-F., Chen, D.-S., Dusheiko, G.M., Janssen, H.L.A., Lau, D.T.Y., Locarnini, S.A., Peters, M.G., Lai, C.-L., 2018. Hepatitis B virus infection. Nature Reviews Disease Primers 4, 1–21. doi:10.1038/nrdp.2018.35

Yuh, C.H., Ting, L.P., 1993. Differentiated liver cell specificity of the second enhancer of hepatitis B virus. J. Virol. 67, 142–149.

Yuh, C.H., Ting, L.P., 1990. The genome of hepatitis B virus contains a second enhancer: cooperation of two elements within this enhancer is required for its function. J. Virol. 64, 4281–4287.

Zandi, R., van der Schoot, P., Reguera, D., Kegel, W., Reiss, H., 2006. Classical nucleation theory of virus capsids. Biophysical Journal 90, 1939–1948. doi:10.1529/biophysj.105.072975

Zhang, J.-F., Xiong, H.-L., Cao, J.-L., Wang, S.-J., Guo, X.-R., Lin, B.-Y., Zhang, Y., Zhao, J.-H., Wang, Y.-B., Zhang, T.-Y., Yuan, Q., Zhang, J., Xia, N.-S., 2018. A cell-penetrating whole molecule antibody targeting intracellular HBx suppresses hepatitis B virus via TRIM21-dependent pathway. Theranostics 8, 549–562. doi:10.7150/thno.20047

Zhao, Q., Hu, Z., Cheng, J., Wu, S., Luo, Y., Chang, J., Hu, J., Guo, J.-T., 2018. Hepatitis B Virus Core Protein Dephosphorylation Occurs during Pregenomic RNA Encapsidation. J. Virol. 92, e02139–17. doi:10.1128/JVI.02139-17

Zheng, Y., Fu, X.-D., Ou, J.H.J., 2005. Suppression of hepatitis B virus replication by SRPK1 and SRPK2 via a pathway independent of the phosphorylation of the viral core protein. Virology 342, 150–158. doi:10.1016/j.virol.2005.07.030

Zhou, H., Shen, T., Luo, Y., Liu, L., Chen, W., Xu, B., Han, X., Pang, J., Rivera, C.A., Huang, S., 2010. The antitumor activity of the fungicide ciclopirox. Int. J. Cancer 127, 2467–2477. doi:10.1002/ijc.25255

Zhou, S., Standring, D.N., 1992. Hepatitis B virus capsid particles are assembled from core-protein dimer precursors. Proc. Natl. Acad. Sci. U.S.A. 89, 10046–10050.

Zhou, Z., Hu, T., Zhou, X., Wildum, S., Garcia-Alcalde, F., Xu, Z., Wu, D., Mao, Y., Tian, X., Zhou, Y., Shen, F., Zhang, Z., Tang, G., Najera, I., Yang, G., Shen, H.C., Young, J.A.T., Qin, N., 2017. Heteroaryldihydropyrimidine (HAP) and Sulfamoylbenzamide (SBA) Inhibit Hepatitis B Virus Replication by Different Molecular Mechanisms. Scientific RepoRts 1–12. doi:10.1038/srep42374

Zlotnick, A., 2007. Distinguishing reversible from irreversible virus capsid assembly. J. Mol. Biol. 366, 14–18. doi:10.1016/j.jmb.2006.11.034

Zlotnick, A., 2003. Are weak protein-protein interactions the general rule in capsid assembly? Virology 315, 269–274.

Zlotnick, A., CHENG, N., Conway, J.F., Booy, F.P., Steven, A.C., Stahl, S.J., Wingfield, P.T., 1996. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. Journal of Biological Chemistry 35, 7412–7421. doi:10.1021/bi9604800

Zlotnick, A., Johnson, J.M., Wingfield, P.W., Stahl, S.J., Endres, D., 1999. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. Biochemistry 38, 14644–14652. doi:10.1021/bi991611a

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. doi:10.1016/j.antiviral.2015.06.020

Zoulim, F., Saputelli, J., Seeger, C., 1994. Woodchuck hepatitis virus X protein is required for viral infection in vivo. J. Virol. 68, 2026–2030.

Zoulim, F., Seeger, C., 1994. Reverse transcription in hepatitis B viruses is primed by a tyrosine residue of the polymerase. J. Virol. 68, 6–13.

Methods to study HBV core protein

Contents

● Bacterial expression of recombinant proteins

➢ Recombinant expression in *E. coli*

The *Escherichia coli* (*E. coli*) expression system is a well-developed approach, and the most widely used means to heterologous protein production for functional and structural studies (Gomes et al., 2016; Kaur et al., 2018). According to the data of the Worldwide Protein Data Bank (wwPDB), about 90 % of publications used *E. coli* as the host cell for heterologous protein expression up to August 2018 (Broadway, 2019) (Figure 42).

Compared with other cell-based expression systems developed from yeast, plant, insect or mammalian cells, the *E. coli* expression system has advantages in many factors, including low cost, good yield, fast production, convenient management, well-studied genetics, and efficient isotope labeling (Joseph et al., 2015; Kaur et al., 2018; Lacabanne et al., 2019a; 2018) (Table 2). Meanwhile, due to the cellular difference between bacteria and eukaryotic cells, the negative factors for recombinant protein expression in *E. coli* cannot be neglected. For example, the reduced condition of the *E. coli* cytoplasm, provided by thioredoxin reductase and glutaredoxin, could affect the formation of disulfide bond which is required for some protein (Sørensen and Mortensen, 2005); some enzymes responsible for posttranslational modifications are absent in *E. coli* cell, leading to a lack in the expressed protein of phosphorylation or glycosylation (Gopal and Kumar, 2013; Heger-Stevic et al., 2018; Kaur et al., 2018). But these disadvantages can be overcome to some extent by using engineered strains or by co-expressing specific enzyme with the target protein (Gopal and Kumar, 2013; Heger-Stevic et al., 2018). As a result of wide use and continuous developing for long years, the recombinant expression in *E. coli* systems can benefit from abundant commercialized vectors, cell strains, fusion tags, and well-introduced strategies (reviewed in

(Correa and Oppezzo, 2014; Gileadi, 2017; Gopal and Kumar, 2013; Joseph et al., 2015; Kaur et al., 2018)).

Host system	Merits	Demerits
Escherichia coli	Easy Quick Economical Rapid growth rate Capacity for continuous fermentation	Does not possess necessary machinery for removing introns from transcripts Foreign gene might contain sequences that act as termination signals resulting in premature termination and loss of gene expression Codon bias Lack of post translational modifications Glycosylation is extremely uncommon in bacteria Production of proteins in the insoluble form or in the form of inclusion bodies Degradation of proteins Accumulation of endotoxins
Bacillus subtilis	Does not produce LPS/endotoxins Can be transformed readily with many bacteriophages and plasmids Capable of secreting functional extracellular proteins directly into the culture medium	Production of extracellular proteases which can recognize and degrade heterologous proteins Instability of plasmids Reduced or non expression of the protein of interest
Yeast system	Rapid growth in low cost medium Appropriate post-translational modifications Safety of the system is guaranteed No endotoxins production	Hyperglycosylation of proteins Codon bias Inefficient in secreting the proteins into growth medium leading to intracellular retention
Filamentous fungus	High-level of expression	Complex Lack of knowledge on physiology
Baculovirus / Insect system	High level of expression Appropriate posttranslational modifications Safe for vertebrates Excellent tool for recombinant glycoprotein production	Continuous expression not possible More demanding culture conditions
Mammalian cells / system	Proper protein folding Appropriate post-translational modifications and product assembly Proper glycosylation	High cost Complicated technology Potential contamination with animal viruses
Transgenic plants	Easy scaling up at low cost Proteins can be localized to different organs at different growth stages High yield	Expression levels are target dependent Functional assays yet to be developed
Transgenic animals	Proper protein folding Appropriate post-translational modifications and product assembly Proper glycosylation	Relatively longer production period Low yield Higher costs

Table 2. Comparison of different expression systems for recombinant proteins. Taken from: (Gomes et al., 2016).

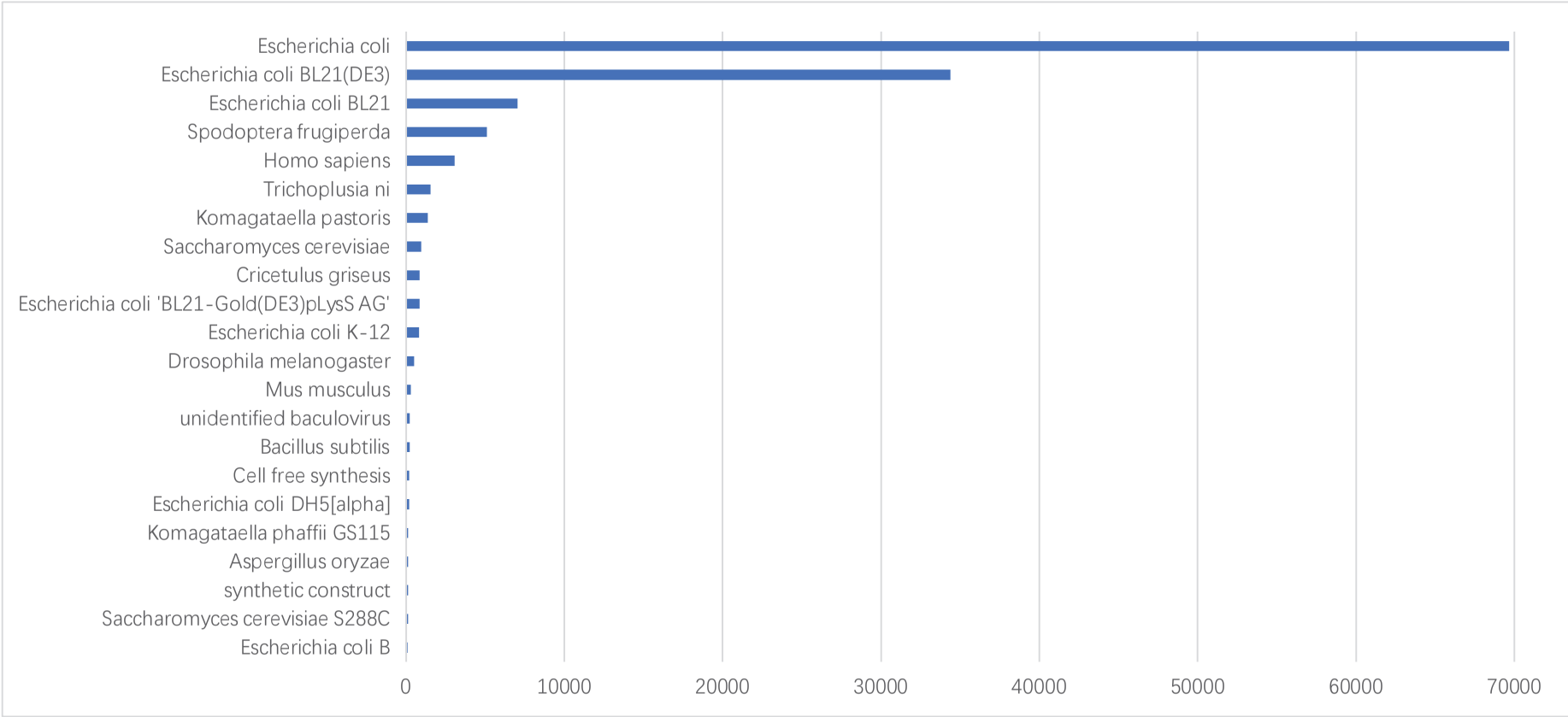


Figure 42. Protein Data Bank (PDB) data distribution by expression organism. The quantity distribution of protein structures from different expression Organism is shown above. Data source: RCSB PDB website, up to June, 2019.

➢ Isotopic labeling

Contrary to X-ray crystallography and cryo-electron microscopy (cryo-EM) studies, the protein sample used for nuclear magnetic resonance (NMR) research needs to be isotopically labeled, because homonuclear ^{1}H spectroscopy is of limited use in solid-state NMR due to the still broad proton resonances. Especially for studies of proteins and other biomacromolecules by solid-state NMR, isotopic labeling can not only improve detection sensitivity but also allows the survey on site-specific structures and intermolecular contacts (Hong, 2010). The main isotopes used in NMR studies are ^{1}H, ^{13}C and ^{15}N. While the natural abundance of ^{1}H is higher than 99.9 %, the labeling of other isotopes must be executed with artificially extracted high abundance materials (Verardi et al., 2012).

1. Culture media used for isotopic labeling

The M9 minimal medium is one of the most used media for *E. coli* growth, and is generally used for isotopic labeling of recombinant proteins with economic cost and high efficiency (Pardi, 1992). The supplement of trace elements and vitamins to M9 minimal medium is widely used to enhance host cell growth and target protein expression (Atlas, 2010). In order to reduce the cost and to gain higher yield, the improved “rich medium” which contains algal lysate is introduced to be used for expression of isotopic labeling proteins (Fiaux et al., 2004; Suzuki et al., 2005). Increasing buffering capacity and modifying composition of M9 minimal medium, as well as the improvement on expression protocol, are also reported to be able to improve isotopic labeling protein yield (Azatian et al., 2019; Cai et al., 2016; Klopp et al., 2018; O'Brien 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		

Table 3. Components of M9 minimal medium (Lacabanne, 2017). The trace elements are listed on the right. The vitamin solution used (BME Vitamins 100x solution) is from Sigma-Aldrich.

2. Uniform isotopic labeling

Uniform isotopic labeling is the simplest and most cost-effective way for solid-state NMR protein preparation, and is widespread applied in spectroscopic characterization of target proteins (Baldus, 2010; Hong, 2010). The protein sample is usually overexpressed in *E. coli* with the minimal media combined with the use of ^{13}C-labeled glucose and ^{15}N-labeled ammonium chloride, which serve as the sole nitrogen source and carbon source (Hong, 2010; Lacabanne et al., 2019a). The labeled amino acids are converted from glucose and ammonium chloride, following the metabolic pathways of *E. coli*, and the isotopes are finally introduced into the target protein by biosynthesis (Hoogstraten and Johnson, 2008; Lacabanne et al., 2018; Verardi et al., 2012). For the expression of deuterated protein, the H_2O and glucose used in the minimal media should be replaced respectively by D_2O and deuterated glucose.

3. Selective or specific isotopic labeling

Based on the knowledge of metabolic pathways of *E. coli*, recombinant proteins can be selectively labeled or unlabeled by certain strategies, such as usage of mixed carbon sources or specific labeled precursors (Hong and Jakes, 1999; Lacabanne et al., 2019a; 2018; Verardi et al., 2012). Residue-selective isotope labeling and unlabeling can be achieved by adding certain amino acids in the minimal medium, so that the protein synthesis is processed with the nitrogen source and carbon source of substrate, as well as with the specific supplied amino acids directly (Lacabanne et al., 2018; Verardi et al., 2012; Vuister et al., 1994). The summary of amino-acid biosynthesis for the *E. coli* BL21 strain is shown in Figure 43.

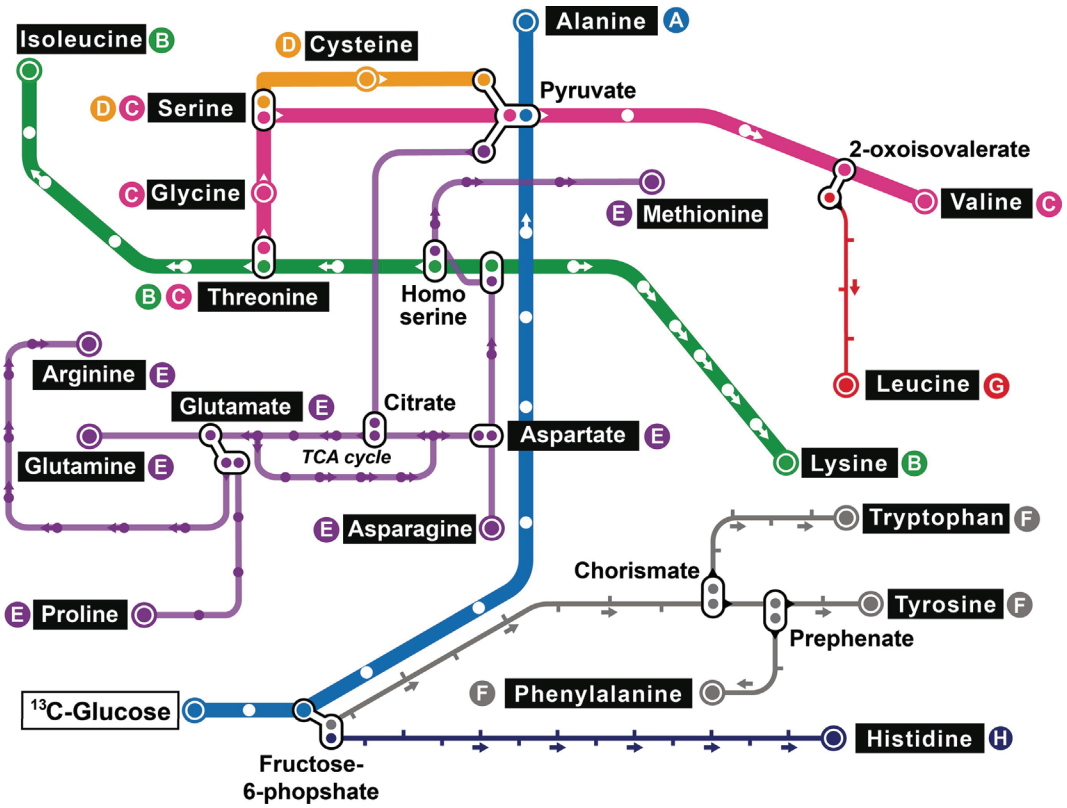


Figure 43. The metabolic pathways of *E. coli* for amino acids biosynthesis. This "subway map" represents a summary of amino-acid biosynthesis for the BL21 strain. Some slight differences may exist in other *E. coli* strains. Each station on the map represents a metabolic step for biosynthesis of amino acids, and the arrows are used to denote the non-

reversible steps. The alanine is derived from pyruvate, which is originated from glucose through glycolysis pathway (line A, blue). The citrate, produced from pyruvate, generates aspartate and glutamate via tricarboxylic acid (TCA) cycle, and finally result in products of arginine, glutamine, proline and asparagine (line E, purple). The threonine, isoleucine and lysine are generated from aspartate (line B, green). The catabolism of threonine leads to generation of serine and glycine (line C, pink), and the serine subsequently generates cysteine (line D, orange), as well as 2-oxoisovalerate through the pyruvate station. The 2-oxoisovalerate is precursor for both valine (line C, pink) and leucine (line G, red). The fructose-6-phosphate can be converted into histidine (line H, dark blue) and aromatic amino acids, tyrosine, phenylalanine, tryptophan (line F, gray). Taken from: (Lacabanne et al., 2018).

The amino acid for which selective labeling or unlabeling shall be achieved should not lie on the main biosynthesis pathway of other amino acids, in which case the conversion between amino acids shall disturb selective labeling or unlabeling efficiency (Lacabanne et al., 2019a; 2018). For example, additional Glu and Asp can take part in biosynthesis of other amino acids, leading to diffusion of selective labeling, which is called scrambling (Lacabanne et al., 2019a).

In addition, the specifically labeled carbon sources have also been used in recombinant protein expression for solid-state NMR study, including [1-^{13}C]- or [2-^{13}C]-glucose, [1,3-^{13}C]- or [2-^{13}C]-glycerol, [1,2-^{13}C]- or [1-^{13}C]-pyruvate and isotopic labeling bicarbonate, $^{13}CH_3$-methyl specifically labeled precursors, (Kurauskas et al., 2017; Verardi et al., 2012).

● Wheat Germ Cell-Free Overexpression

➢ Cell-free systems and wheat germ cell-free system

Cell-free systems (CFS) were firstly established about 50 years ago based on lysates of *E. coli* (DeVries and Zubay, 1967; Lederman and Zubay, 1967; NIRENBERG and MATTHAEI, 1961; Zubay et al., 1967). Subsequently, the material sources of CFS have been widely expanded to other organisms, such as HeLa cells (Reichman and Penman, 1973), yeast (Wang et al., 2008), wheat germ (Roberts and Paterson, 1973), insects (Ezure et al., 2006), rabbit reticulocyte (Craig et al., 1992), and these systems still undergo large improvements in both preparation and applications (reviewed in (Chong, 2014; Gregorio et al., 2019; Lu, 2017; Perez et al., 2016; Villarreal and Tan, 2017)).

The wheat germ extracts cell-free (WGE-CF) system was first used to produce RNAs in 1973 (Roberts and Paterson, 1973) and improved to be an efficient and powerful protein expression system in recent 20 years (Madin et al., 2000; Sawasaki et al., 2002a; Tsuge et al., 2010). Compared with classical cell-based and other cell-free protein expression systems, the WGE-CF system displays several advantages (Table 4). As a cell-free protein expression system, it not only can produce proteins whose expression is toxic for the host cell, but also is open for the addition of required supplements, such as chaperones, detergents, lipids or interesting drugs (Fogeron et al., 2017). The WGE-CF system has no bias in codon usage for eukaryotic genes, and the shortage of endogenous proteases reduces degradation of recombinant expressed proteins (Fogeron et al., 2017; Harbers, 2014; Takahashi et al., 2009). Meanwhile, the low translation rate and the coupling of translation and protein folding are helpful to produce proteins which keep their activity (Fogeron et al., 2017). Moreover, the advantages of relatively high yield compared to other CFS, medium economical cost, and efficient isotopic labeling strategies without the problem of scrambling, allows the WGE-CF system to be a suitable and useful

tool for NMR studies (David et al., 2018; Fogeron et al., 2016; Kohno and Endo, 2007; Noirot et al., 2011; Novikova et al., 2018; Vinarov et al., 2004).

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

Table 4. Comparison between main cell-free and cell-based protein expression systems. Taken from: (Fogeron, 2015).

➢ Wheat germ extracts and mRNA preparation

Protocols for wheat germ extracts (WGE) preparation have been introduced by (Takai et al., 2010a). They can be carried out in a laboratory setting, despite a time-consuming process as it takes 4 or 5 days and needs eye selection for choosing good embryo particles (Fogeron et al., 2017; Harbers, 2014). Commercial WGE, as for example produced by CellFree Sciences Co., Ltd., is available but with relative high cost (Fogeron et al., 2017) (700 € per ml).

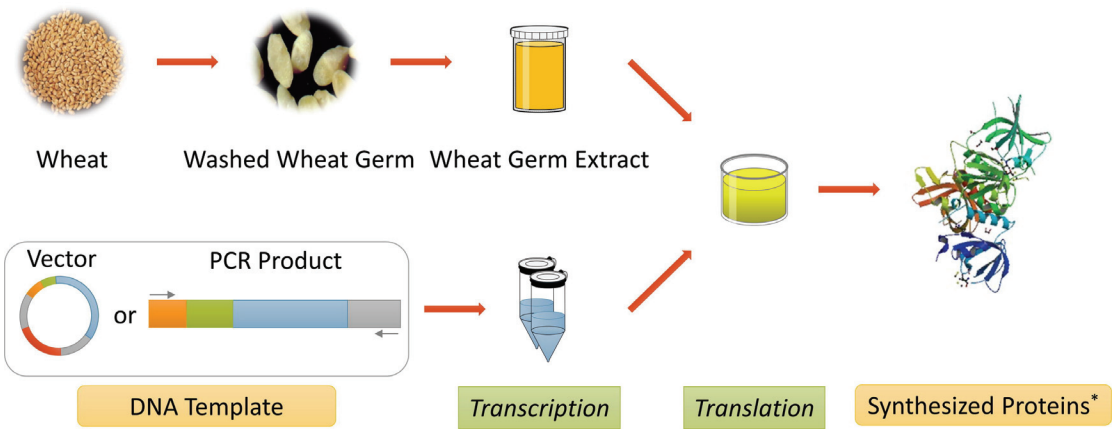


Figure 44. Outline of protein expression with WGE-CF system. Wheat germ extracts can be produced from commercial wheat seeds, with a series of steps including seed crushing, embryo selection, washing to remove endosperm, and extracting from washed embryo particles. The mRNA template used for recombinant protein expression can be transcribed from either circular vectors or linear DNA segment with required functional elements. The elements of DNA template are highlighted in the figure, including the promoter (orange) and enhancer (green) for transcription, the target gene (blue), the vector derived sequence (grey), and the antibiotic resistance gene (red) of vector. The expression is executed by mixing transcribed mRNA solution with wheat germ extracts, supplemented with amino acids and creatine kinase. Taken from: (Harbers, 2014)

Both circular or linear DNA sequence can be used to produce mRNA templates for WGE-CF protein expression (Sawasaki et al., 2000; Takai et al., 2010a). Besides the transcriptional elements such as promotor and enhancer, about 1,500 nucleotides length noncoding sequence are recommended to exist in the

3'-end of mRNA transcripts to prevent mRNA from degradation (Harbers, 2014; Sawasaki et al., 2002b). Although the transcription can be coupled with the translation reaction, the expression of recombinant protein is usually performed by adding pre-transcribed mRNA into WGE translation reaction (Harbers, 2014).

➢ Reaction modes of WGE-CF translation

Various reaction modes have been developed for translation of recombinant protein in WGE-CF system (Harbers, 2014; Takai et al., 2010a; 2010b), such as the batch mode (Kawasaki et al., 2003), the repeat-batch mode (Sawasaki et al., 2005), the bilayer mode (Sawasaki et al., 2002a), the dialysis mode (Kohno and Endo, 2007).

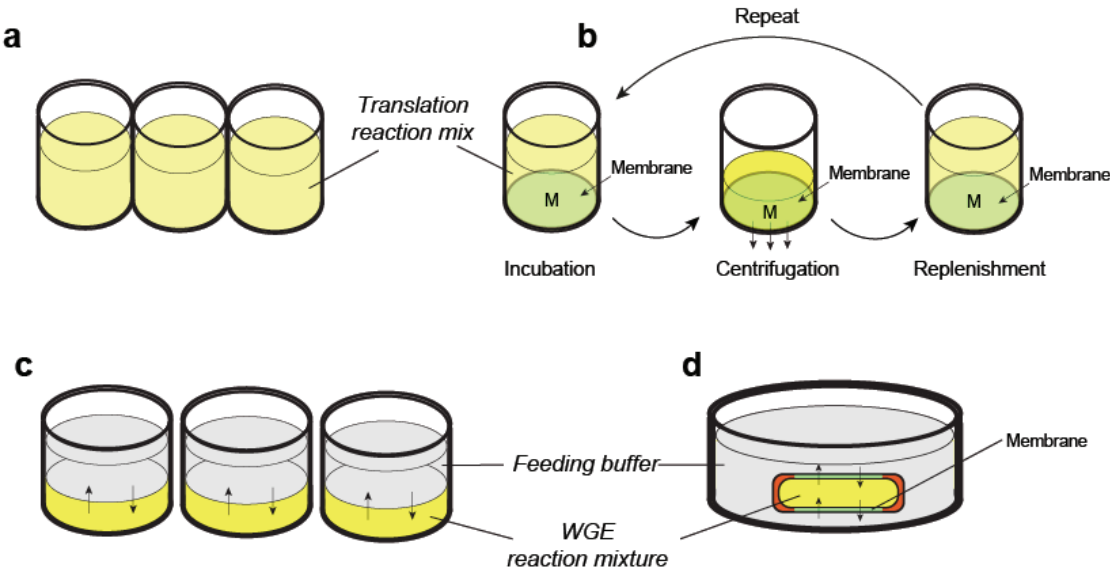


Figure 45. WGE-CF translation modes. a): batch mode. All reagents are mixed together for translation incubation. b): repeat-batch mode. The membrane (light green) at bottom of the reaction vial is used to filter small molecular weight inhibitors by centrifugation. New substrates can be replenished into the vial to maintain translation reaction. c): bilayer mode. The reaction mix (yellow) is overlaid by feeding buffer (gray) layer, and these two layers will gradually merge with each other over time to keep protein translation. d): dialysis mode. The translation reaction mix is contained in a dialysis cassette which is soaked in the feeding buffer. Through the membrane of the dialysis cassette, fresh substrates in feeding buffer continuously spread into the reaction mix, while the small molecular weight inhibitors are seeped out.

In the batch mode, the translation reaction is performed by mixing all reagents together. Though this method is useful for high-throughput expression screening (Sawasaki et al., 2002b; Schwarz et al., 2008), the yield of protein is commonly insufficient for biochemical analysis due to the limitation of incubation time, which ordinarily lasts a few hours (Harbers, 2014; Sawasaki et al., 2002a).

The repeat-batch mode is carried out similarly to the batch mode but with a filter membrane at the bottom of reaction vial (Beebe et al., 2011; Harbers, 2014). Small molecular weight inhibitors, produced during the translation process, can be removed by centrifugation, and then new reaction buffer is replenished to the reaction vial (Harbers, 2014). Translation reaction is maintained with refreshed substrates by repeated centrifugation and replenishment (Harbers, 2014). Continuous translation reaction can be also managed automatically with the ‘‘Filter-Feed’’ method, which drives the filtering by pressure and replenishes new substrates and mRNA template from time to time (Beebe et al., 2011; 2014; Harbers, 2014).

In the bilayer mode, the translation reaction mix is injected into well of flat-bottom plate as the bottom layer, overlaid by the feeding buffer as supplement (Fogeron et al., 2017; Takai et al., 2010a). The diffusion between reaction mix layer and feeding buffer layer allows new substrates to be blended into reaction mix to maintain protein translation (Harbers, 2014). This mode is more efficient for protein expression than the batch mode, and suitable for high-throughput screening as well, therefore it has been widely used for WGE-CF protein expression (Fogeron et al., 2017; Harbers, 2014; Sawasaki et al., 2005; Takai et al., 2010b).

The dialysis mode is more efficient than the batch mode and the bilayer mode, and is hence usually used in large-scale protein production (David et al., 2018; Takai et al., 2010b). The translation reaction is maintained for longer times by substance exchange between reaction mix and feeding buffer (Harbers, 2014; Takai et al., 2010b). If the mRNA is supplemented into the reaction mix and the

feeding buffer is changed in time, the reaction time can be extended to as long as two weeks (Takai et al., 2010b).

- **Isotopic labeling in WGE-CF system**

Like in other cell-free systems, isotopic labeling for translated protein is specific and efficient in the WGE-CF system (Abdine et al., 2011; Fogeron, 2015; Lacabanne et al., 2019a). While in *E. coli* expression system, other metabolic products, such as mRNA and host proteins are also labeled along with the recombinant protein during the cells proliferation and growth, in cell-free synthesis the translated protein is isotopic labeled specifically with ^{13}C and ^{15}N labelled amino acids. Therefore, the impact of other biomolecules is reduced during NMR sample characterization as they are not labeled, as for example the packaged RNA inside viral capsids. For selective isotopic labeling, the problem of scrambling, which commonly occurs in cell-based protein expression, is severely limited and can be further eliminated by adding inhibitors to the reaction mix (Kohno, 2010; Takai et al., 2010b; Vinarov et al., 2004).

Solid-state NMR

Solid-state NMR is a powerful tool for structural and dynamic studies of biomacromolecules (Müller et al., 2013). Compared to X-ray crystallography and solution NMR, solid-state NMR is not limited by sample crystallinity, solubility, molecular size, or even purity, and has been applied for example to the characterization of membrane proteins (Judge et al., 2014; Lacabanne et al., 2019b; Ladizhansky, 2017; Opella, 2015), fibrillar proteins (Niu et al., 2014; van der Wel, 2017; Wälti et al., 2016), viral capsids (Bayro and Tycko, 2016; Han et al., 2010; Lecoq et al., 2018) and subviral particles (David et al., 2018), but also to protein-protein complexes (Gardiennet et al., 2016; Sun et al., 2011; Tang et al., 2011) or protein-DNA/RNA (Abramov and Goldbourt, 2014; Carlomagno, 2014; Sergeyev et al., 2011; Wiegand et al., 2019). While structure determination has been achieved mainly for several smaller proteins (Loquet et al., 2012; Schütz et al., 2015), structural studies by solid-state NMR have contributed to a variety of insight in biology (Struppe et al., 2017).

Though solid-state NMR can be used for large molecular weight proteins in theory, it is still needs further technique developements, in order to overcome spectral crowding and to enhance sensitivity (Demers et al., 2018; Fricke et al., 2015). For example, the structure determination of proteins with more than 200 amino acids, which lead to crowded spectra, is usually impeded by overlapped peaks which are unanalyzable (Demers et al., 2018). However, the spectra of large proteins can be simplified to some extent by specialized isotope labeling as described above, but also by high-dimensional spectroscopy, which is helpful for chemical shift assignment and distance restraints collection (Franks et al., 2007; Huber et al., 2011; Linser et al., 2014; Sperling et al., 2010; Zinke et al., 2017). Moreover, several advances have been made in magic-angle spinning (MAS) solid-state NMR (Demers et al., 2018; Fricke et al., 2015), such as the proton-detection techniques, thanks to the development of smaller and smaller rotors which allow to reach extremely high spinning frequencies (> 110

kHz) and to enhance the sensitivity per mass unit (Agarwal et al., 2014; Demers et al., 2011). Fast-MAS is usually combined with the technique of proton (^{1}H) detection (Fricke et al., 2017; Ishii and Tycko, 2000; Zhang et al., 2017). Another technique used to increase the sensitivity is the dynamic nuclear polarization (DNP) (Akbey and Oschkinat, 2016; Ni et al., 2013), which transfers the large polarization of electrons to nuclei at low temperatures.

- **Chemical shifts**

In an NMR spectrum, each peak represents a nucleus or a set of nuclei, which appear at a specific chemical shift. Due to their sensitivity to bond lengths, bond and torsion angles, the chemical shifts are one of the most fundamental informative parameters for protein studies by NMR (Berjanskii and Wishart, 2017; Szilágyi, 1995) as they can give detailed information on the protein structure, dynamics and ligand binding, which can then be used to analyze protein secondary and super-secondary structure, protein folding, protein-protein/ligand interactions, protein flexibility, residue-specific accessible surface area, as reviewed in (Williamson, 2013).

- **Magic angle spinning (MAS)**

Magic angle spinning is a widely used technique, which gives high resolution spectra for solid-state NMR studies. In MAS solid-state NMR, the rotor containing the protein sample is rotated around an axis at an angle of 54.74° with respect to the direction of the spectrometer magnetic field, which can largely reduce signal broadening through averaging of homogenous contributions to the line width (Alia et al., 2009).

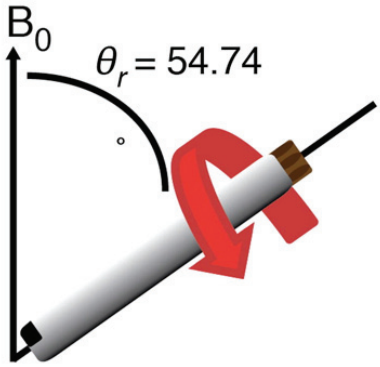


Figure 46. Magic angle spinning schematic diagram. The arrow of B_0 indicates the direction of the constant, homogeneous magnetic field in the NMR spectrometer. θ_r refers the magic angle of 54.74°. The black semi-circle marker at rotor bottom is used to monitor MAS frequency. Taken from: (Penzel et al., 2018)

➢ Filling protein sample into NMR rotors

The protein samples labeled with NMR active nuclei, such as ^{13}C and ^{15}N, are filled into NMR rotors in the state of crystalline samples (Ashbrook and McKay, 2016; Martin and Zilm, 2003), supramolecular assemblies sediments (Quinn and Polenova, 2017), liposomes (Yao et al., 2014), etc. A series of commercial rotors for solid-state NMR are available, with outer diameter ranging from 0.7 mm to 9.5 mm (Figure 47-a). The MAS frequencies are limited by the rotors dimensions: the smaller the diameter is, the higher MAS frequencies can be reached (Demers et al., 2018). The sensitivity of NMR is also affected by rotors dimensions. Although smaller rotors give better signal/noise (S/N) ratio per unit sample mass, larger rotors containing more sample quantities result in higher overall S/N ratio than smaller ones (Demers et al., 2011; Samoson et al., 2010). In the case of viral capsids, the protein sample is sedimented into the NMR rotor by ultracentrifugation using a special filling tool that can sediment up to 1.5 ml of protein solution (**Figure 47**-b). The rotor must be verified to be intact before filling, and the cap should not be damaged when closing the rotor. Before

closing the rotor, 1 μl of a saturated solution of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) is added to further calibrate chemical shifts.

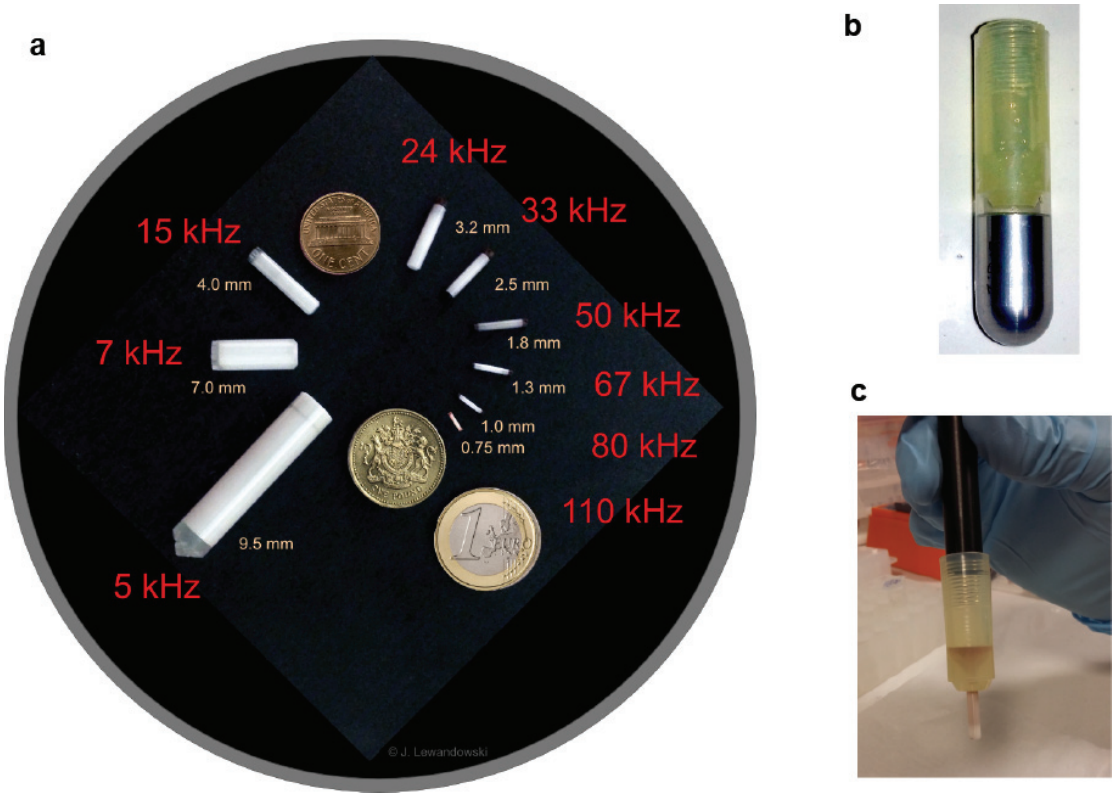


Figure 47. Rotors and filling tool for solid-state NMR studies. a): rotors with different diameters used for solid-state NMR studies and their corresponding MAS frequencies. Picture source: http://www.warwick.ac.uk. b): filling tool for 3.2 mm rotors. c): filled 3.2 mm rotor with HBV Cp sediment resulting from ultracentrifugation at 200,000 g, 4 °C, for 14h.

References

Abdine, A., Verhoeven, M.A., Warschawski, D.E., 2011. Cell-free expression and labeling strategies for a new decade in solid-state NMR. N Biotechnol 28, 272–276. doi:10.1016/j.nbt.2010.07.014

Abramov, G., Goldbourt, A., 2014. Nucleotide-type chemical shift assignment of the encapsulated 40 kbp dsDNA in intact bacteriophage T7 by MAS solid-state NMR. Journal of Biomolecular NMR 59, 219–230. doi:10.1007/s10858-014-9840-4

Agarwal, V., Penzel, S., Szekely, K., Cadalbert, R., Testori, E., Oss, A., Past, J., Samoson, A., Ernst, M., Böckmann, A., Meier, B.H., 2014. De novo 3D structure determination from sub-milligram protein samples by solid-state 100 kHz MAS NMR spectroscopy. Angew. Chem. Int. Ed. Engl. 53, 12253–12256. doi:10.1002/anie.201405730

Akbey, Ü., Oschkinat, H., 2016. Structural biology applications of solid state MAS DNP NMR. J. Magn. Reson. 269, 213–224. doi:10.1016/j.jmr.2016.04.003

Alia, A., Ganapathy, S., de Groot, H.J.M., 2009. Magic Angle Spinning (MAS) NMR: a new tool to study the spatial and electronic structure of photosynthetic complexes. Photosyn. Res. 102, 415–425. doi:10.1007/s11120-009-9478-3

Ashbrook, S.E., McKay, D., 2016. Combining solid-state NMR spectroscopy with first-principles calculations - a guide to NMR crystallography. Chem. Commun. (Camb.) 52, 7186–7204. doi:10.1039/c6cc02542k

Atlas, R.M., 2010. Handbook of Microbiological Media - 4th Edition. CRC Press.

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, 11–17. doi:10.1007/s10858-018-00222-4

Baldus, M., 2010. Solid-state NMR on Larger Biomolecules. ISOTEC Stable Isotopes Products for Solid State NMR 2–4.

Bayro, M.J., Tycko, R., 2016. Structure of the Dimerization Interface in the Mature HIV-1 Capsid Protein Lattice from Solid State NMR of Tubular Assemblies. J. Am. Chem. Soc. 138, 8538–8546. doi:10.1021/jacs.6b03983

Beebe, E.T., Makino, S.-I., Markley, J.L., Fox, B.G., 2014. Automated cell-free protein production methods for structural studies. Methods Mol. Biol. 1140, 117–135. doi:10.1007/978-1-4939-0354-2_9

Beebe, E.T., Makino, S.-I., Nozawa, A., Matsubara, Y., Frederick, R.O., Primm, J.G., Goren, M.A., Fox, B.G., 2011. Robotic large-scale application of wheat cell-free translation to structural studies including membrane proteins. N Biotechnol 28, 239–249. doi:10.1016/j.nbt.2010.07.003

Berjanskii, M.V., Wishart, D.S., 2017. Unraveling the meaning of chemical shifts in protein NMR. Biochim Biophys Acta Proteins Proteom 1865, 1564–1576. doi:10.1016/j.bbapap.2017.07.005

Broadway, N., 2019. Protein Expression. Materials and Methods 2, 123. doi:dx.doi.org/10.13070/mm.en.2.123

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, 85–91. doi:10.1007/s10858-016-0052-y

Carlomagno, T., 2014. Present and future of NMR for RNA-protein complexes: a perspective of integrated structural biology. J. Magn. Reson. 241, 126–136. doi:10.1016/j.jmr.2013.10.007

Chong, S., 2014. Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. Curr Protoc Mol Biol 108, 16.30.1–11. doi:10.1002/0471142727.mb1630s108

Correa, A., Oppezzo, P., 2014. Overcoming the Solubility Problem in E. coli: Available Approaches for Recombinant Protein Production, in: Chen, Y.W. (Ed.), Structural Genomics, Methods and Protocols. Springer New York, New York, NY, pp. 27–44. doi:10.1007/978-1-4939-2205-5_2

Craig, D., Howell, M.T., Gibbs, C.L., Hunt, T., Jackson, R.J., 1992. Plasmid cDNA-directed protein synthesis in a coupled eukaryotic in vitro transcription-translation system. Nucleic Acids Res. 20, 4987–4995. doi:10.1093/nar/20.19.4987

David, G., Fogeron, M.-L., Schledorn, M., Montserret, R., Haselmann, U., Penzel, S., Badillo, A., Lecoq, L., André, P., Nassal, M., Bartenschlager, R., Meier, B.H., Böckmann, A., 2018. Structural Studies of Self-Assembled Subviral Particles: Combining Cell-Free Expression with 110 kHz MAS NMR Spectroscopy. Angew. Chem. Int. Ed. Engl. 57, 4787–4791. doi:10.1002/anie.201712091

Demers, J.-P., Chevelkov, V., Lange, A., 2011. Progress in correlation spectroscopy at ultra-fast magic-angle spinning: basic building blocks and complex experiments for the study of protein structure and dynamics. Solid State Nucl Magn Reson 40, 101–113. doi:10.1016/j.ssnmr.2011.07.002

Demers, J.-P., Fricke, P., Shi, C., Chevelkov, V., Lange, A., 2018. Structure determination of supramolecular assemblies by solid-state NMR: Practical considerations. Progress in Nuclear Magnetic Resonance Spectroscopy 109, 51–78. doi:10.1016/j.pnmrs.2018.06.002

DeVries, J.K., Zubay, G., 1967. DNA-directed peptide synthesis. II. The synthesis of the alpha-fragment of the enzyme beta-galactosidase. Proc. Natl. Acad. Sci. U.S.A. 57, 1010–1012. doi:10.1073/pnas.57.4.1010

Ezure, T., Suzuki, T., Higashide, S., Shintani, E., Endo, K., Kobayashi, S.-I., Shikata, M., Ito, M., Tanimizu, K., Nishimura, O., 2006. Cell-free protein synthesis system prepared from insect cells by freeze-thawing. Biotechnol. Prog. 22, 1570–1577. doi:10.1021/bp060110v

Fiaux, J., Bertelsen, E.B., Horwich, A.L., Wüthrich, K., 2004. Uniform and residue-specific 15N-labeling of proteins on a highly deuterated background. Journal of Biomolecular NMR 29, 289–297. doi:10.1023/B:JNMR.0000032523.00554.38

Fogeron, M.-L., 2015. Development of a wheat germ cell-free expression system for the production, the purification and the structural and functional characterization of eukaryotic membrane proteins : application to the preparation of hepatitis C viral proteins. lyon.

Fogeron, M.-L., Badillo, A., Penin, F., Böckmann, A., 2017. Wheat Germ Cell-Free Overexpression for the Production of Membrane Proteins, in: Chen, Y.W. (Ed.), Structural Genomics, Methods in Molecular Biology. Springer New York, New York, NY, pp. 91–108. doi:10.1007/978-1-4939-7151-0_5

Fogeron, M.-L., Jirasko, V., Penzel, S., Paul, D., Montserret, R., Danis, C., Lacabanne, D., Badillo, A., Gouttenoire, J., Moradpour, D., Bartenschlager, R., Penin, F., Meier, B.H., Böckmann, A., 2016. Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus. Journal of Biomolecular NMR 1–12. doi:10.1007/s10858-016-0040-2

Franks, W.T., Kloepper, K.D., Wylie, B.J., Rienstra, C.M., 2007. Four-dimensional heteronuclear correlation experiments for chemical shift assignment of solid proteins. Journal of Biomolecular NMR 39, 107–131. doi:10.1007/s10858-007-9179-1

Fricke, P., Chevelkov, V., Shi, C., Lange, A., 2015. Strategies for solid-state NMR investigations of supramolecular assemblies with large subunit sizes. J. Magn. Reson. 253, 2–9. doi:10.1016/j.jmr.2014.10.018

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. Nat Protoc 12, 764–782. doi:10.1038/nprot.2016.190

Gardiennet, C., Wiegand, T., Bazin, A., Cadalbert, R., Kunert, B., Lacabanne, D., Gutsche, I., Terradot, L., Meier, B.H., Böckmann, A., 2016. Solid-state NMR chemical-shift perturbations indicate domain reorientation of the DnaG primase in the primosome of Helicobacter pylori. Journal of Biomolecular NMR 64, 189–195. doi:10.1007/s10858-016-0018-0

Gileadi, O., 2017. Recombinant Protein Expression in E. coli : A Historical Perspective, in: Chen, Y.W. (Ed.), Structural Genomics, Methods in Molecular Biology. Springer New York, New York, NY, pp. 3–10. doi:10.1007/978-1-4939-6887-9_1

Gomes, A.R., Byregowda, S.M., Veeregowda, B.M., Balamurugan, V., 2016. An Overview of Heterologous Expression Host Systems for the Production of Recombinant Proteins. Adv. Anim. Vet. Sci. 4, 346–356. doi:10.14737/journal.aavs/2016/4.7.346.356

Gopal, G.J., Kumar, A., 2013. Strategies for the Production of Recombinant Protein in Escherichia coli. Protein J 32, 419–425. doi:10.1007/s10930-013-9502-5

Gregorio, N.E., Levine, M.Z., Oza, J.P., 2019. A User's Guide to Cell-Free Protein Synthesis. Methods Protoc 2, 24. doi:10.3390/mps2010024

Han, Y., Ahn, J., Concel, J., Byeon, I.-J.L., Gronenborn, A.M., Yang, J., Polenova, T., 2010. Solid-state NMR studies of HIV-1 capsid protein assemblies. J. Am. Chem. Soc. 132, 1976–1987. doi:10.1021/ja908687k

Harbers, M., 2014. Wheat germ systems for cell-free protein expression. FEBS Letters 588, 2762–2773. doi: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 Pathog 14, e1007488. doi:10.1371/journal.ppat.1007488

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

Hong, M., Jakes, K., 1999. Selective and extensive 13C labeling of a membrane protein for solid-state NMR investigations. Journal of Biomolecular NMR 14, 71–74.

Hoogstraten, C.G., Johnson, J.E., 2008. Metabolic labeling: Taking advantage of bacterial pathways to prepare spectroscopically useful isotope patterns in proteins and nucleic acids. Concepts Magn. Reson. 32A, 34–55. doi:10.1002/cmr.a.20103

Huber, M., Hiller, S., Schanda, P., Ernst, M., Böckmann, A., Verel, R., Meier, B.H., 2011. A proton-detected 4D solid-state NMR experiment for protein structure determination. ChemPhysChem 12, 915–918. doi:10.1002/cphc.201100062

Ishii, Y., Tycko, R., 2000. Sensitivity enhancement in solid state (15)N NMR by indirect detection with high-speed magic angle spinning. J. Magn. Reson. 142, 199–204. doi:10.1006/jmre.1999.1976

Joseph, B.C., Pichaimuthu, S., Srimeenakshi, S., 2015. An Overview of the Parameters for Recombinant Protein Expression in Escherichia coli. J Cell Sci Ther 06, 1–7. doi:10.4172/2157-7013.1000221

Judge, P.J., Taylor, G.F., Dannatt, H.R.W., Watts, A., 2014. Solid-State Nuclear Magnetic Resonance Spectroscopy for Membrane Protein Structure Determination, in: Chen, Y.W. (Ed.), Structural Genomics, Methods in Molecular Biology. Springer New York, New York, NY, pp. 331–347. doi:10.1007/978-1-4939-2230-7_17

Kaur, J., Kumar, A., Kaur, J., 2018. Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements. Int. J. Biol. Macromol. 106, 803–822. doi:10.1016/j.ijbiomac.2017.08.080

Kawasaki, T., Gouda, M.D., Sawasaki, T., Takai, K., Endo, Y., 2003. Efficient synthesis of a disulfide-containing protein through a batch cell-free system from wheat germ. Eur. J. Biochem. 270, 4780–4786. doi:10.1046/j.1432-1033.2003.03880.x

Klopp, J., Winterhalter, A., Gébleux, R., Scherer-Becker, D., Ostermeier, C., Gossert, A.D., 2018. Cost-effective large-scale expression of proteins for NMR studies. Journal of Biomolecular NMR 71, 247–262. doi:10.1007/s10858-018-0179-0

Kohno, T., 2010. NMR assignment method for amide signals with cell-free protein synthesis system. Methods Mol. Biol. 607, 113–126. doi:10.1007/978-1-60327-331-2_11

Kohno, T., Endo, Y., 2007. Production of protein for nuclear magnetic resonance study using the wheat germ cell-free system. Methods Mol. Biol. 375, 257–272. doi:10.1007/978-1-59745-388-2_13

Kurauskas, V., Schanda, P., Sounier, R., 2017. Methyl-Specific Isotope Labeling Strategies for NMR Studies of Membrane Proteins. Methods Mol. Biol. 1635, 109–123. doi:10.1007/978-1-4939-7151-0_6

Lacabanne, D., 2017. Solid-state NMR studies of the ABC transporter BmrA in its lipid environment. Lyon.

Lacabanne, D., Fogeron, M.-L., Wiegand, T., Cadalbert, R., Meier, B.H., Böckmann, A., 2019a. Protein sample preparation for solid-state NMR investigations. Progress in Nuclear Magnetic Resonance Spectroscopy 110, 20–33. doi:10.1016/j.pnmrs.2019.01.001

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

Lacabanne, D., Orelle, C., Lecoq, L., Kunert, B., Chuilon, C., Wiegand, T., Ravaud, S., Jault, J.-M., Meier, B.H., Böckmann, A., 2019b. Flexible-to-rigid transition is central for substrate transport in the ABC transporter BmrA from Bacillus subtilis. Commun Biol 2, 149. doi:10.1038/s42003-019-0390-x

Ladizhansky, V., 2017. Applications of solid-state NMR to membrane proteins. Biochim Biophys Acta Proteins Proteom 1865, 1577–1586. doi:10.1016/j.bbapap.2017.07.004

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

Lederman, M., Zubay, G., 1967. DNA-directed peptide synthesis. 1. A comparison of T2 and Escherichi coli DNA-directed peptide synthesis in two cell-free systems. Biochim. Biophys. Acta 149, 253–258. doi:10.1016/0005-2787(67)90706-x

Linser, R., Bardiaux, B., Andreas, L.B., Hyberts, S.G., Morris, V.K., Pintacuda, G., Sunde, M., Kwan, A.H., Wagner, G., 2014. Solid-state NMR structure determination from diagonal-compensated, sparsely nonuniform-sampled 4D proton-proton restraints. J. Am. Chem. Soc. 136, 11002–11010. doi:10.1021/ja504603g

Loquet, A., Sgourakis, N.G., Gupta, R., Giller, K., Riedel, D., Goosmann, C., Griesinger, C., Kolbe, M., Baker, D., Becker, S., Lange, A., 2012. Atomic model of the type III secretion system needle. Nature 486, 276–279. doi:10.1038/nature11079

Lu, Y., 2017. Cell-free synthetic biology: Engineering in an open world. Synth Syst Biotechnol 2, 23–27. doi:10.1016/j.synbio.2017.02.003

Madin, K., Sawasaki, T., Ogasawara, T., Endo, Y., 2000. A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. Proc. Natl. Acad. Sci. U.S.A. 97, 559–564. doi:10.1073/pnas.97.2.559

Martin, R.W., Zilm, K.W., 2003. Preparation of protein nanocrystals and their characterization by solid state NMR. J. Magn. Reson. 165, 162–174.

Ni, Q.Z., Daviso, E., Can, T.V., Markhasin, E., Jawla, S.K., Swager, T.M., Temkin, R.J., Herzfeld, J., Griffin, R.G., 2013. High Frequency Dynamic Nuclear Polarization. Acc. Chem. Res. 46, 1933–1941. doi:10.1021/ar300348n

NIRENBERG, M.W., MATTHAEI, J.H., 1961. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc. Natl. Acad. Sci. U.S.A. 47, 1588–1602. doi:10.1073/pnas.47.10.1588

Niu, Z., Zhao, W., Zhang, Z., Xiao, F., Tang, X., Yang, J., 2014. The molecular structure of Alzheimer β-amyloid fibrils formed in the presence of phospholipid vesicles. Angew. Chem. Int. Ed. Engl. 53, 9294–9297. doi:10.1002/anie.201311106

Noirot, C., Habenstein, B., Bousset, L., Melki, R., Meier, B.H., Endo, Y., Penin, F., Böckmann, A., 2011. Wheat-germ cell-free production of prion proteins for solid-state NMR structural studies. N Biotechnol 28, 232–238. doi:10.1016/j.nbt.2010.06.016

Novikova, I.V., Sharma, N., Moser, T., Sontag, R., Liu, Y., Collazo, M.J., Cascio, D., Shokuhfar, T., Hellmann, H., Knoblauch, M., Evans, J.E., 2018. Protein structural biology using cell-free platform from wheat germ. Adv Struct Chem Imaging 4, 13. doi:10.1186/s40679-018-0062-9

O'Brien, E.S., Lin, D.W., Fuglestad, B., Stetz, M.A., Gosse, T., Tommos, C., Wand, A.J., 2018. Improving yields of deuterated, methyl labeled protein by growing in H2O. Journal of Biomolecular NMR 71, 263–273. doi:10.1007/s10858-018-0200-7

Opella, S.J., 2015. Solid-state NMR and membrane proteins. Journal of Magnetic Resonance 253, 129–137. doi:10.1016/j.jmr.2014.11.015

Pardi, A., 1992. Isotope labelling for NMR studies of biomolecules. Curr. Opin. Struct. Biol. 2, 832–835. doi:10.1016/0959-440X(92)90107-I

Penzel, S., Smith, A.A., Ernst, M., Meier, B.H., 2018. Setting the magic angle for fast magic-angle spinning probes. J. Magn. Reson. 293, 115–122. doi:10.1016/j.jmr.2018.06.002

Perez, J.G., Stark, J.C., Jewett, M.C., 2016. Cell-Free Synthetic Biology: Engineering Beyond the Cell. Cold Spring Harb Perspect Biol 8, a023853. doi:10.1101/cshperspect.a023853

Quinn, C.M., Polenova, T., 2017. Structural biology of supramolecular assemblies by magic-angle spinning NMR spectroscopy. Q. Rev. Biophys. 50, e1. doi:10.1017/S0033583516000159

Reichman, M., Penman, S., 1973. Stimulation of polypeptide initiation in vitro after protein synthesis inhibition in vivo in HeLa cells. Proc. Natl. Acad. Sci. U.S.A. 70, 2678–2682. doi:10.1073/pnas.70.9.2678

Roberts, B.E., Paterson, B.M., 1973. Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. Proc. Natl. Acad. Sci. U.S.A. 70, 2330–2334. doi:10.1073/pnas.70.8.2330

Samoson, A., Tuherm, T., Past, J., Reinhold, A., Heinmaa, I., Anupõld, T., Smith, M.E., Pike, K.J., 2010. Fast Magic-Angle Spinning: Implications. John Wiley & Sons, Ltd, Chichester, UK. doi:10.1002/9780470034590.emrstm1017

Sawasaki, T., Gouda, M.D., Kawasaki, T., Tsuboi, T., Toscana, Y., Takai, K., Endo, Y., 2005. The Wheat Germ Cell-Free Expression System, in: Chemical Genomics, Methods for High-Throughput Materialization of Genetic Information. Humana Press, Totowa, NJ, pp. 131–144. doi:10.1007/978-1-59259-948-6_10

Sawasaki, T., Hasegawa, Y., Tsuchimochi, M., Kamura, N., Ogasawara, T., Kuroita, T., Endo, Y., 2002a. A bilayer cell-free protein synthesis system for high-throughput screening of gene products. FEBS Letters 514, 102–105.

Sawasaki, T., Hasegawa, Y., Tsuchimochi, M., Kasahara, Y., Endo, Y., 2000. Construction of an efficient expression vector for coupled transcription/translation in a wheat germ cell-free system. Nucleic Acids Symp. Ser. 9–10.

Sawasaki, T., Ogasawara, T., Morishita, R., Endo, Y., 2002b. A cell-free protein synthesis system for high-throughput proteomics. Proc. Natl. Acad. Sci. U.S.A. 99, 14652–14657. doi:10.1073/pnas.232580399

Schütz, A.K., Vagt, T., Huber, M., Ovchinnikova, O.Y., Cadalbert, R., Wall, J., Güntert, P., Böckmann, A., Glockshuber, R., Meier, B.H., 2015. Atomic-resolution three-dimensional structure of amyloid β fibrils bearing the Osaka mutation. Angew. Chem. Int. Ed. Engl. 54, 331–335. doi:10.1002/anie.201408598

Schwarz, D., Dötsch, V., Bernhard, F., 2008. Production of membrane proteins using cell-free expression systems. Proteomics 8, 3933–3946. doi:10.1002/pmic.200800171

Sergeyev, I.V., Day, L.A., Goldbourt, A., McDermott, A.E., 2011. Chemical shifts for the unusual DNA structure in Pf1 bacteriophage from dynamic-nuclear-polarization-enhanced solid-state NMR spectroscopy. J. Am. Chem. Soc. 133, 20208–20217. doi:10.1021/ja2043062

Sperling, L.J., Berthold, D.A., Sasser, T.L., Jeisy-Scott, V., Rienstra, C.M., 2010. Assignment strategies for large proteins by magic-angle spinning NMR: the 21-kDa disulfide-bond-forming enzyme DsbA. J. Mol. Biol. 399, 268–282. doi:10.1016/j.jmb.2010.04.012

Struppe, J., Quinn, C.M., Lu, M., Wang, M., Hou, G., Lu, X., Kraus, J., Andreas, L.B., Stanek, J., Lalli, D., Lesage, A., Pintacuda, G., Maas, W., Gronenborn, A.M., Polenova, T., 2017. Expanding the horizons for structural analysis of fully protonated protein assemblies by NMR spectroscopy at MAS frequencies above 100 kHz. Solid State Nucl Magn Reson 87, 117–125. doi:10.1016/j.ssnmr.2017.07.001

Sun, S., Han, Y., Paramasivam, S., Yan, S., Siglin, A.E., Williams, J.C., Byeon, I.-J.L., Ahn, J., Gronenborn, A.M., Polenova, T., 2011. Solid-State NMR Spectroscopy of Protein Complexes, in: Chen, Y.W. (Ed.), Structural Genomics, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 303–331. doi:10.1007/978-1-61779-480-3_17

Suzuki, H., Shimada, Y., Kobayashi, M., Kudo, M., Nozawa, T., Wang, Z.-Y., 2005. Isotopic labeling of proteins by utilizing photosynthetic bacteria. Anal. Biochem. 347, 324–326. doi:10.1016/j.ab.2005.05.012

Szilágyi, L., 1995. Chemical shifts in proteins come of age. Progress in Nuclear Magnetic Resonance Spectroscopy 27, 325–442. doi:10.1016/0079-6565(95)01011-2

Sørensen, H., Mortensen, K., 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*

. Microb Cell Fact 4, 1–8. doi:10.1186/1475-2859-4-1

Takahashi, H., Nozawa, A., Seki, M., Shinozaki, K., Endo, Y., Sawasaki, T., 2009. A simple and high-sensitivity method for analysis of ubiquitination and polyubiquitination based on wheat cell-free protein synthesis. BMC Plant Biol. 9, 39. doi:10.1186/1471-2229-9-39

Takai, K., Sawasaki, T., Endo, Y., 2010a. Practical cell-free protein synthesis system using purified wheat embryos. Nat Protoc 5, 227–238. doi:10.1038/nprot.2009.207

Takai, K., Sawasaki, T., Endo, Y., 2010b. The Wheat-Germ Cell-Free Expression System. CPB 11, 272–278. doi:10.2174/138920110791111933

Tang, M., Sperling, L.J., Berthold, D.A., Schwieters, C.D., Nesbitt, A.E., Nieuwkoop, A.J., Gennis, R.B., Rienstra, C.M., 2011. High-resolution membrane protein structure by joint calculations with solid-state NMR and X-ray experimental data. Journal of Biomolecular NMR 51, 227–233. doi:10.1007/s10858-011-9565-6

Tsuge, M., Hiraga, N., Akiyama, R., Tanaka, S., Matsushita, M., Mitsui, F., Abe, H., Kitamura, S., Hatakeyama, T., Kimura, T., Miki, D., Mori, N., Imamura, M., Takahashi, S., Hayes, C.N., Chayama, K., 2010. HBx protein is indispensable for development of viraemia in human hepatocyte chimeric mice. J. Gen. Virol. 91, 1854–1864. doi:10.1099/vir.0.019224-0

van der Wel, P.C.A., 2017. Insights into protein misfolding and aggregation enabled by solid-state NMR spectroscopy. Solid State Nucl Magn Reson 88, 1–14. doi:10.1016/j.ssnmr.2017.10.001

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. Adv. Exp. Med. Biol. 992, 35–62. doi:10.1007/978-94-007-4954-2_3

Villarreal, F., Tan, C., 2017. Cell-free systems in the new age of synthetic biology. Front. Chem. Sci. Eng. 11, 58–65. doi:10.1007/s11705-017-1610-x

Vinarov, D.A., Lytle, B.L., Peterson, F.C., Tyler, E.M., Volkman, B.F., Markley, J.L., 2004. Cell-free protein production and labeling protocol for NMR-based structural proteomics. Nat. Methods 1, 149–153. doi:10.1038/nmeth716

Vuister, G.W., Kim, S.-J., Wu, C., Bax, A., 1994. 2D and 3D NMR Study of Phenylalanine Residues in Proteins by Reverse Isotopic Labeling. J. Am. Chem. Soc. 116, 9206–9210. doi:10.1021/ja00099a041

Wang, X., Liu, J., Zheng, Y., Li, J., Wang, H., Zhou, Y., Qi, M., Yu, H., Tang, W., Zhao, W.M., 2008. An optimized yeast cell-free system: sufficient for translation of human papillomavirus 58 L1 mRNA and assembly of virus-like particles. J. Biosci. Bioeng. 106, 8–15. doi:10.1263/jbb.106.8

Wälti, M.A., Ravotti, F., Arai, H., Glabe, C.G., Wall, J.S., Böckmann, A., Güntert, P., Meier, B.H., Riek, R., 2016. Atomic-resolution structure of a disease-relevant Aβ(1-42) amyloid fibril. Proc. Natl. Acad. Sci. U.S.A. 113, E4976–84. doi:10.1073/pnas.1600749113

Wiegand, T., Cadalbert, R., Lacabanne, D., Timmins, J., Terradot, L., Böckmann, A., Meier, B.H., 2019. The conformational changes coupling ATP hydrolysis and translocation in a bacterial DnaB helicase. Nat Commun 10, 31. doi:10.1038/s41467-018-07968-3

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

Müller, H., Etzkorn, M., Heise, H., 2013. Solid-State NMR Spectroscopy of Proteins, in: Modern NMR Methodology, Topics in Current Chemistry. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 121–156. doi:10.1007/128_2012_417

Yao, X., Dürr, U.H.N., Gattin, Z., Laukat, Y., Narayanan, R.L., Brückner, A.-K., Meisinger, C., Lange, A., Becker, S., Zweckstetter, M., 2014. NMR-based detection of hydrogen/deuterium exchange in liposome-embedded membrane proteins. PLoS ONE 9, e112374. doi:10.1371/journal.pone.0112374

Zhang, R., Mroue, K.H., Ramamoorthy, A., 2017. Proton-Based Ultrafast Magic Angle Spinning Solid-State NMR Spectroscopy. Acc. Chem. Res. 50, 1105–1113. doi:10.1021/acs.accounts.7b00082

Zinke, M., Fricke, P., Samson, C., Hwang, S., Wall, J.S., Lange, S., Zinn-Justin, S., Lange, A., 2017. Bacteriophage Tail-Tube Assembly Studied by Proton-Detected 4D Solid-State NMR. Angew. Chem. Int. Ed. Engl. 56, 9497–9501. doi:10.1002/anie.201706060

Zubay, G., Lederman, M., DeVries, J.K., 1967. DNA-directed peptide synthesis. 3. Repression of beta-galactosidase synthesis and inhibition of repressor by inducer in a cell-free system. Proc. Natl. Acad. Sci. U.S.A. 58, 1669–1675. doi:10.1073/pnas.58.4.1669

RESULTS

Contents

● Overexpression, Purification and Reconstruction of HBV Capsids

This part is about the sample preparation of preformed capsids and reassembled capsids, which is a crucial step for protein structural studies.

➢ Abstract

The Cp of HBV was first reported to be synthesized in *E. coli* in 1979 (Burrell et al., 1979; Pasek et al., 1979), and was subsequently confirmed to assemble into capsids in *E. coli* cells, with diameter of about 30 nm (Cohen and Richmond, 1982). Cps were reported to be purified from *E. coli* using a series of purification methods, such as ammonium sulfate precipitation, sucrose gradient, gel filtration, and the proteins obtained from *E. coli* can be disassembled and reassembled *in vitro* (Heger-Stevic et al., 2018a; Porterfield et al., 2010; Wynne et al., 1999; Zlotnick et al., 1996). Based on the protocols from the literature and from our collaborator Michael Nassal (Freiburg University), we overexpressed and purified Cps of truncated types, full-length types, phosphorylated type and mutant types. The capsids formed in *E. coli* and reassembled *in vitro*, as well as capsids produced in WGE-CF system were characterized using solid-state NMR. In this chapter I will detail all the plasmids of Cp used for *E. coli* expression and their characteristics.

➢ Expression of HBV core protein

1. Plasmids and *E. coli* strains

The plasmids detailed in Table 5 and *E. coli* strains for Cps expression were generously provided by Michael Nassal (Freiburg University), except the plasmid for HBV geonoty G Cp was kindly given by Timothy Booth (University of Manitoba). The pRST-T7 plasmids encode codon-optimized Cp under control of a T7 RNA polymerase promoter. In addition, for the production of full-length protein phosphorylated at the seven phosphorylation sites (Heger-Stevic et al., 2018c), the plasmid encodes also for the soluble fragment dNS1 of SRPK1 kinase under a Tet promoter. Plasmids were all transformed into BL21 Star CodonPlus (DE3) competent cells, also called BL21*CP(DE3). The ‘Star’ provides a genomic defect in RNase E that reduces mRNA degradation. In addition, the CodonPlus cells carry an extra plasmid that provides rare

E. coli tRNAs and promotes expression of genes with codons rarely used in *E. coli* (Gopal and Kumar, 2013; Heger-Stevic et al., 2018b). This extra plasmid confers resistance to chloromycetin (also called chloramphenicol). After transformation, monoclones were incubated on LB-agar plate at 37°C overnight and inoculated in 5 ml LB media grown overnight to prepare glycerol stocks which were then stored at -80 °C. All the culture media we used contained 34 µg/ml chloromycetin for the CodonPlus plasmid, as well as 50 µg/ml kanamycin or 100 µg/ml ampicillin according to the expression plasmids resistance.

Cps	Plasmids	Resistance
Cp140	pRSF-T7-HBc140opt	Kanamycin
Cp140C61A	pRSF-T7-HBc140opt-C61A	Kanamycin
Cp149	pET28a2-HBc149opt	Ampicillin
Cp183	pRSF-T7-HBc183opt	Kanamycin
P-Cp183	pRSF-Tet-SRPK1-T7-HBc183opt	Kanamycin
Cp183 F97L	pRSF-T7-HBc183opt-F97L	Kanamycin
Cp183 EEE	pRSF-T7-HBc183opt-EEE	Kanamycin
HBV/G Cp	pET28b-HBV/G core	Kanamycin

Table 5. Plasmids of Cps in the thesis. P-Cp183 refers phosphorylated full-length Cp protein (Cp183), which was co-expressed with serine-arginine-phosphokinase 1 (SRPK1). The Cp183 EEE is a phosphorylation mimic protein with mutation of S155E, S162E, and S170E. HBV/G Cp refers to core protein of HBV from the genotype G, which has an additional N-terminal domain and contains a total of 195 amino acids. Full protein sequences are given in Annexe 1.

2. Expression condition screening

For each construct, 10 µl of glycerol stock was inoculated into 100 ml LB medium and grew overnight at 37 °C, 160 rpm/min. Cells were pelleted from preculture by centrifugation at 20 °C, 4000 g for 10 min, and resuspended into large volume culture with ratio of preculture:culture = 1:20. While cells grew to OD_{600} in range of 2.0-3.2 at 37 °C, 160 rpm/min, Cps expression was induced by 1 mM IPTG overnight. The ^{13}C and ^{15}N labeled proteins were produced in M9 medium with ^{13}C-labeled glucose and ^{15}N-labeled ammonium chloride. For deuterated protein production, all components of M9 medium were lyophilized for 48 h and resuspended in D_2O.

The expression of Cps shows distinct behaviors, which are significantly related to the cell density at the induction time. While the cells growth is slightly affected by the OD_{600} of induction for Cp149 expression, it is significantly inhibited when the induction is done at low cell density (OD_{600} < 2.0) in cultures of Cp140, Cp180, P-Cp180, which leads to less cell pellets and subsequently to lower yield of proteins (Figure 48).

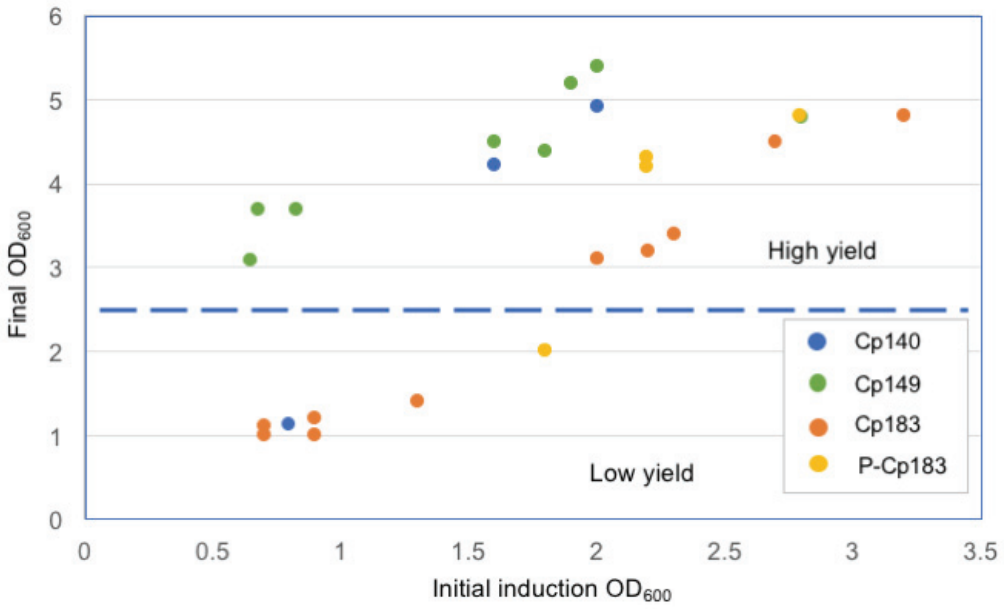


Figure 48. The influence of initial induction density of cultures for Cps expression. The expression of Cp149 (green) is slightly affected by initial density of induced culture, while the low initial density for the induction of Cp140 (blue), Cp183 (orange) and P-Cp183 (yellow) cultures leads to smaller cells quantities. Though the protein expression is not linearly related to final cell

density, the yield of Cps falls sharply if the final OD_{600} of the induced cultures is below 3.0. High yields can be obtained when the cultures are induced at cell density within a range of 2.0-3.2.

The inhibition effect on cell growth is probably not caused by IPTG but the accumulation of Cp in *E. coli* cells. Parallel tests show that 0.2, 0.5 and 1.0 mM IPTG have almost equal impact on Cp183 cultures, on both cell growth and protein expression (Figure 49). The initial OD_{600} for the induction seems not to affect the expression quantity of Cps in *E. coli* cells, however, good yield (about 25 mg/l) could be obtained if cultures were induced at high cell density ($2.0 < OD_{600} < 3.2$).

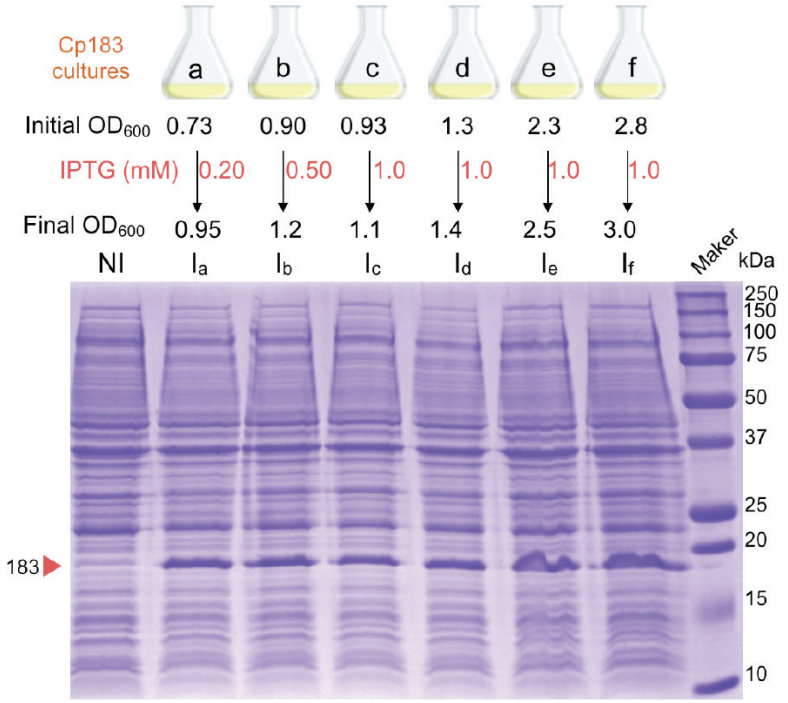


Figure 49. Expression tests on Cp183 expression. The induction tests were done on 50 ml M9 medium at 37 °C for 3 h. Cell pellets from 1 ml of non-induced (NI) and induced (I_a-I_f) cultures were resuspended in 1* SDS-PAGE loading buffer with same density (60 µl buffer for each unit of OD_{600}), and 10 µl of each sample were loaded onto SDS-PAGE. All the samples induced with different concentration (a-c) and initial cell density (c-f) give a protein band on SDS-PAGE with similar intensity, which means the accumulation of Cp183 is limited at certain quantity in *E. coli* cells.

Low temperature is usually helpful for the expression of soluble proteins (Correa and Oppezzo, 2014; Vera et al., 2007). For full-length core proteins, Cp183 is indeed expressed better at 20 °C than 37°C, and the expression of P-Cp183 can only be

detected by SDS-PAGE in the culture induced at 16 °C. On the contrary, the expression of Cp149 is good at or above 25 °C, and the 20 °C is the optimal induction temperature when the purity is taken into account (Figure 50).

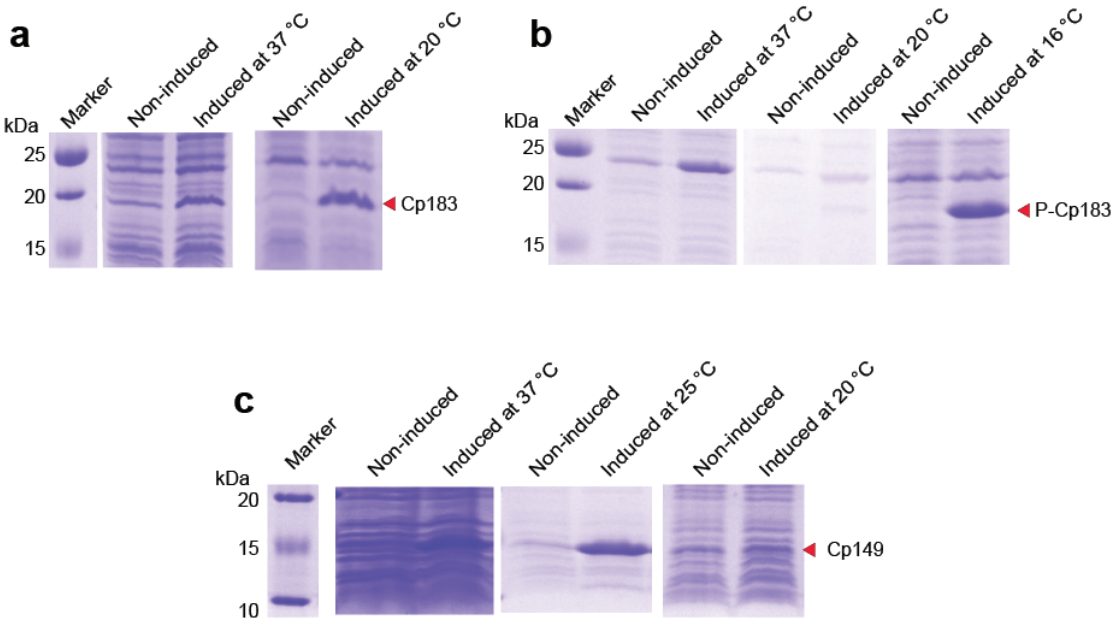


Figure 50. The expression of Cps influenced by temperature. The target protein bands are indicated by red triangle. The optimal condition for Cp183 (a), P-Cp183 (b), Cp149 (c) is an overnight induction respectively at 20 °C, 16 °C and 25 °C.

➢ Purification of of HBV core protein

Buffers	Components
TN300	50 mM Tris pH 7.5, 300 mM NaCl, 5 mM DTT, 1 mM EDTA
Sucrose gradient	50 mM Tris pH 7.5, 300 mM NaCl, 5 mM DTT, 10-60% sucrose, 1 mM EDTA
Disassembly buffer A	50 mM $NaHCO_3$ pH 9.6, 5 mM DTT

Disassembly buffer B	50 mM Tris pH 7.5, 2 M Gua-HCl, 500 mM LiCl, 5 mM DTT
Disassembly buffer C	50 mM Tris pH 7.5, 2.5 M Gua-HCl, 5 mM DTT
Reassembly buffer A	50 mM Tris pH 7.5, 500 mM NaCl, 5 mM DTT
Reassembly buffer B	50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT
Dilution buffer	50 mM HEPES pH 7.5, 500 mM LiCl, 5 mM DTT
Purification buffer	50 mM Tris pH 7.5, 5% sucrose, 5 mM DTT, 1 mM EDTA

Table 6. Buffers used for the Cps sample preparation.

The overnight induced cultures were centrifuged at 6,000 g for 20 min to pellet the cells. Cell pellets from 1 liter culture were resuspended in 15 ml TN300 buffer (Table 6) and lysed with chicken lysozyme (1 mg/ml) (except for Cp140, which migrates to the same position with chicken lysozyme on SDS-PAGE), 50× protease inhibitor cocktail solution (300 µl) and Triton-X-100 (5%) for 45 min. 6 µl of benzonase nuclease were added to digest cellular nucleic acids for 30 min at room temperature. The suspension then underwent intermittent sonication (10 short bursts per 10s, 40 % cycle duty, 10 cycles) to thoroughly lyse remaining cells.

Lysate solution was centrifuged at 8,000g for 1 h. Capsids in the supernatant were separated by 10-60 % (m/v) sucrose gradient (Figure 51). Truncated Cps usually migrated in the 20-30 % sucrose layers, corresponding to a density around 1.10 g/ml, while the full-length Cps migrated in the 30-50 % sucrose layers, corresponding to a density around 1.18 g/ml, and the Cp from genotype G was the most dense capsid with a density around 1.23 g/ml as it was mainly found in the 40-60 % fractions.

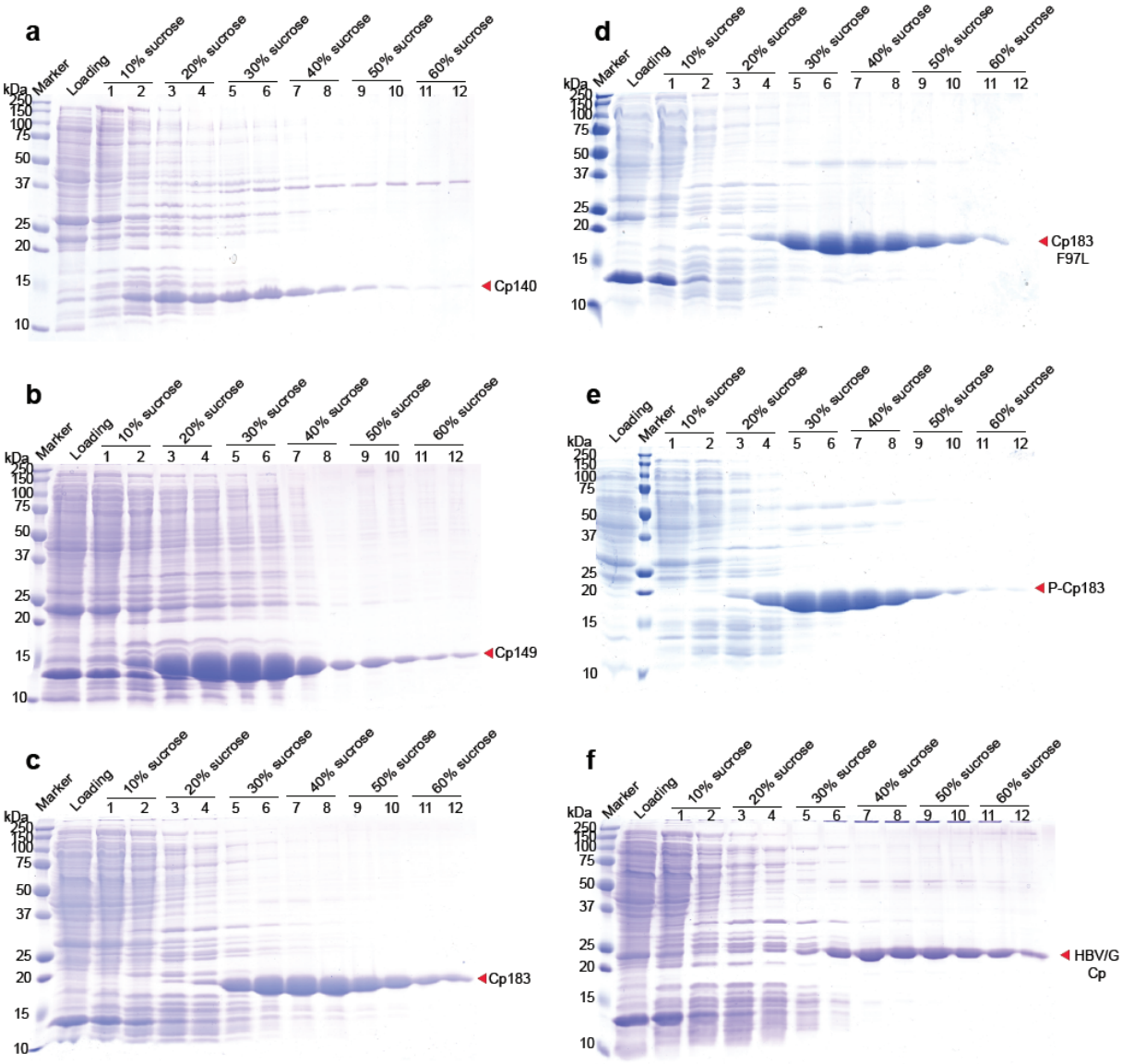


Figure 51. Sucrose gradient separation of HBV Cps. Sucrose gradients were prepared in Beckman tubes used for SW 32 Ti Rotor, with 5 ml each layer, while 7.5 ml cell lysate were loaded on the top of sucrose layers. Capsids were separated by centrifugation at 141,800 g for 3 h at 4 °C. Based on the difference of density, capsids and protein contaminants migrated to different sucrose layers. In addition, while the truncated Cp140 and Cp149 mainly migrated to 20-40 % layers (a-b), most capsids of full-length proteins, such as Cp183, P-C183 and Cp183 mutants, migrated to 30-50 % layers (c-e). For HBV/G Cp, capsids were generally found in 40-60 % sucrose fractions (f).

The fractions containing the protein were then precipitated by ammonium sulfate (AS) of 20 % saturation (for truncated Cps) or 40 % saturation (for full-length Cps). Protein pellets were precipitated by centrifugation at 20,000g for 1 h, and resuspended in 15 ml purification buffer (Table 6). Insoluble pellets were removed by centrifugation at

14,000 g for 15 min and the supernatant was dialyzed in purification buffer to remove remaining AS. Final purity was >90 % as shown by SDS-PAGE (Figure 52). The typical yields of Cps purification are displayed in Table 7.

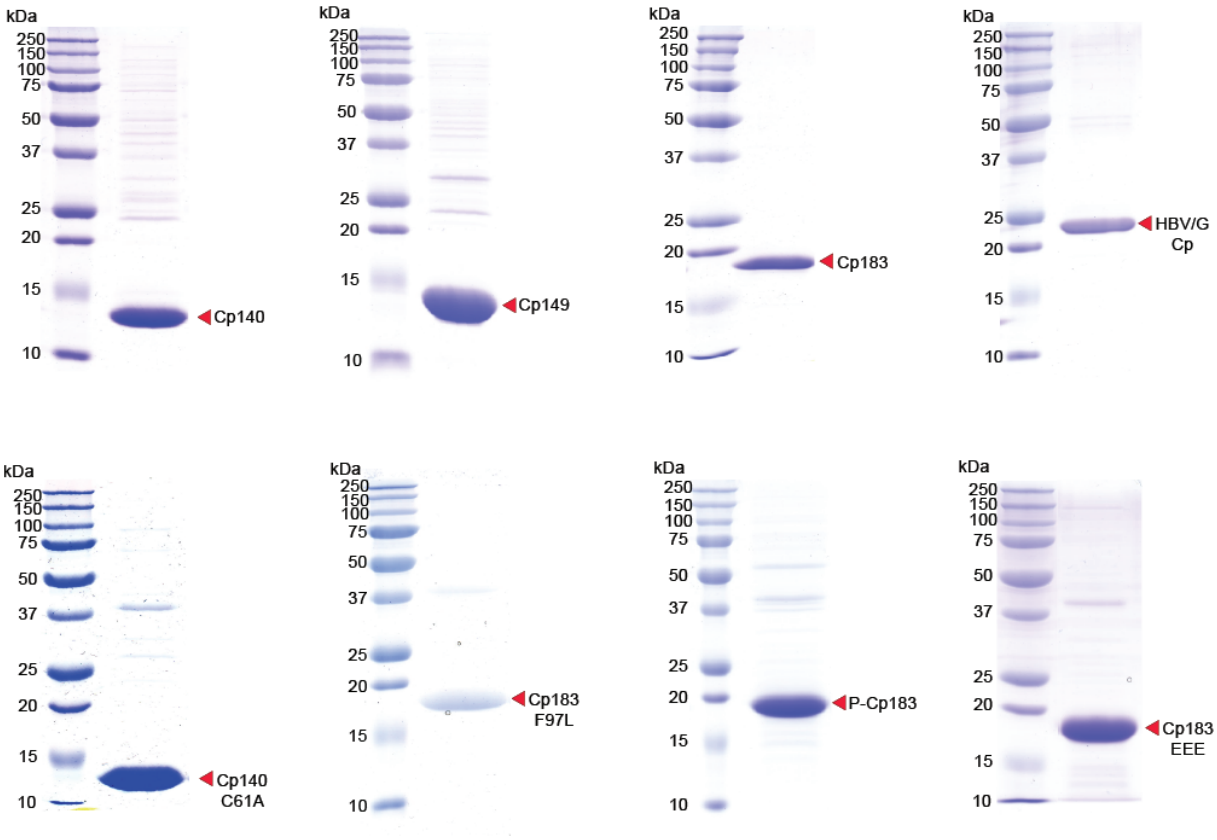


Figure 52. Purity of HBV Cp samples. Proteins were further purified from sucrose fractions by a step of ammonium sulfate precipitation, and finally dialyzed against purification buffer.

Cps	Induction OD and temperature	Purification methods	Typical yield
Cp140	OD_{600} > 1.5 25 °C	10-60% sucrose gradient 20% saturated AS precipitation	80 mg/L
Cp149	OD_{600} > 0.6 25 °C	10-60% sucrose gradient 20% saturated AS precipitation	100 mg/L

Cp183	OD_{600} > 2.0 20 °C	10-60% sucrose gradient 40% saturated AS precipitation	25 mg/L
P-Cp183	OD_{600} > 2.2 16 °C	10-60% sucrose gradient 40% saturated AS precipitation	50 mg/L
HBV/G Cp	OD_{600} > 1.5 20 °C	10-60% sucrose gradient 40% saturated AS precipitation	20 mg/L

Table 7. Summary of HBV Cps expression and purification.

- **Reconstruction of HBV Capsids**

1. Production of viral pgRNA *in vitro*

Template plasmid, pCHT-3101pgRNA-SNAP (6.3 kbp), was transformed into and amplified in DH5α cells. Culture was incubated in 300 ml LB containing 100 µg/ml ampicillin at 37 °C, overnight. Cells were pelleted to extract circular DNA, using a commercial QIAprep Spin Miniprep Kit, which was subsequently digested into linear DNA by *Xma1* enzyme (New England Biolabs, Inc.). Linearized pCHT-3101pgRNA-SNAP was isolated by organic solvents extraction (Green and Sambrook, 2017; 2016). In brief, contaminants, such as enzymes, were removed first with a volume mixture of phenol:chloroform:isoamyl alcohol (25 :24 :1, pH7.9, Sigma-Aldrich, Inc.) and then with a volume of chloroform. DNA in the aqueous phase was precipitated by 2.5 volume of ethanol, supplemented by 1/10 volume of sodium acetate solution (3 M, pH 5.2). Pellet was washed with 800 µl of 70% ethanol once, and finally dissolved in RNase-free H_2O. Absorbance at 260 nm was measured by nanodrop and the DNA was diluted to a final concentration of 1 µg/µl in RNase-free H_2O.

Reagents	Stock solution	Final concentration	Volume (µl)
rNTP mix	25 mM	5 mM	200
TRIS	500 mM	50 mM	100
$MgCl_2$	500 mM	25 mM	50
Triton	0.1 mM	0.01 mM	100
DTT	100 mM	10 mM	100
Spermidine	25 mM	2.5 mM	100
RNAsin	80 U/µl	1.6 U/µl	20
DNA template	1 µg/µl	0.1 µg/µl	100
T7 polymerase	0.8 µg/µl	0.06 µg/µl	75
ddH_2O	**-**	**-**	155
Total	**-**	-	**1000**

Table 8. Reaction mix for 1 ml of pgRNA transcription. The rNTP mix contains a mix of equal volumes of ATP, UTP, CTP and GTP. The T7 polymerase was prepared in the laboratory.

The pgRNA was transcribed *in vitro* using T7 polymerase (Beckert and Masquida, 2010), with the linearized plasmid serving as template (Table 8). The transcription mix was incubated at 37 °C for 2 h, and then 25 mM EDTA were added into the mix to stop the reaction. The reaction product was analyzed on a 0.8 % agarose gel (Figure 53), and the transcribed pgRNA was purified following the same protocol as the one used for the linear DNA. About 3 mg of pgRNA could be obtained from 1 ml transcription reaction mix (100 µg of DNA).

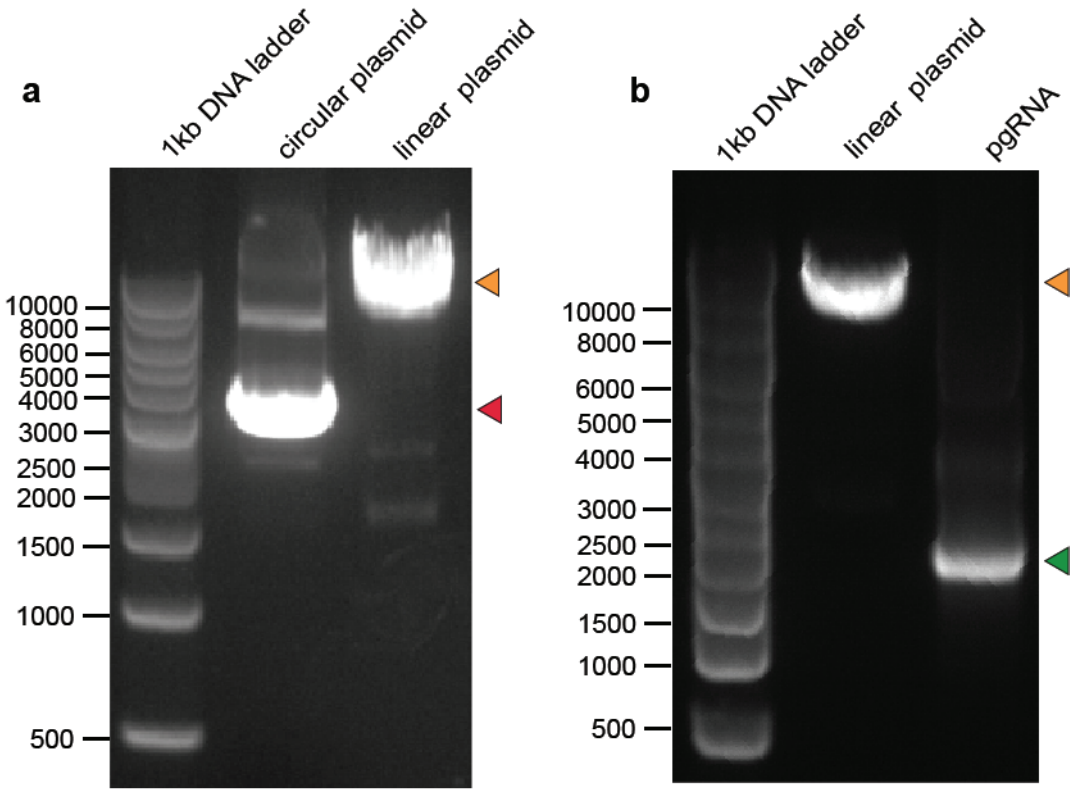


Figure 53. DNA template preparation and pgRNA transcription. a): circular pCHT-3101pgRNA-SNAP plasmid was cut at the unique restriction site of *Xma1* enzyme, resulting in a linear plasmid which will serve as DNA template in the transcription reaction. b): characterization of transcribed pgRNA. The gels were made with 0.8 % agarose, and the 1 kb DNA ladder was bought from Sigma-Aldrich, Inc. (Product No.D 3937). The positon of circular and linear plasmid bands, as well as the pgRNA band, are indicated with red, orange and green triangles respectively.

2. Disassembly and reassembly of HBV capsids

For truncated Cps, preformed capsids from *E. coli* were dialyzed against disassembly buffer A (Table 6) overnight at 4 °C. Solid urea was added to the protein solution to a final concentration of 3 M and incubated on ice for 2 h in order to dissociate capsids. Cp dimers were then separated by gel filtration using a *HiPrepTM 16/60 Sephacryl® S-200 HR* column of 120 ml (Figure 54-a). Dimers were dialyzed against reassembly buffer A (Table 6) containing 500 mM NaCl to reform capsids. Remaining dimers were separated again by gel filtration (Figure 55-a). Reassembled empty capsids were dialyzed against purification buffer overnight at 4 °C.

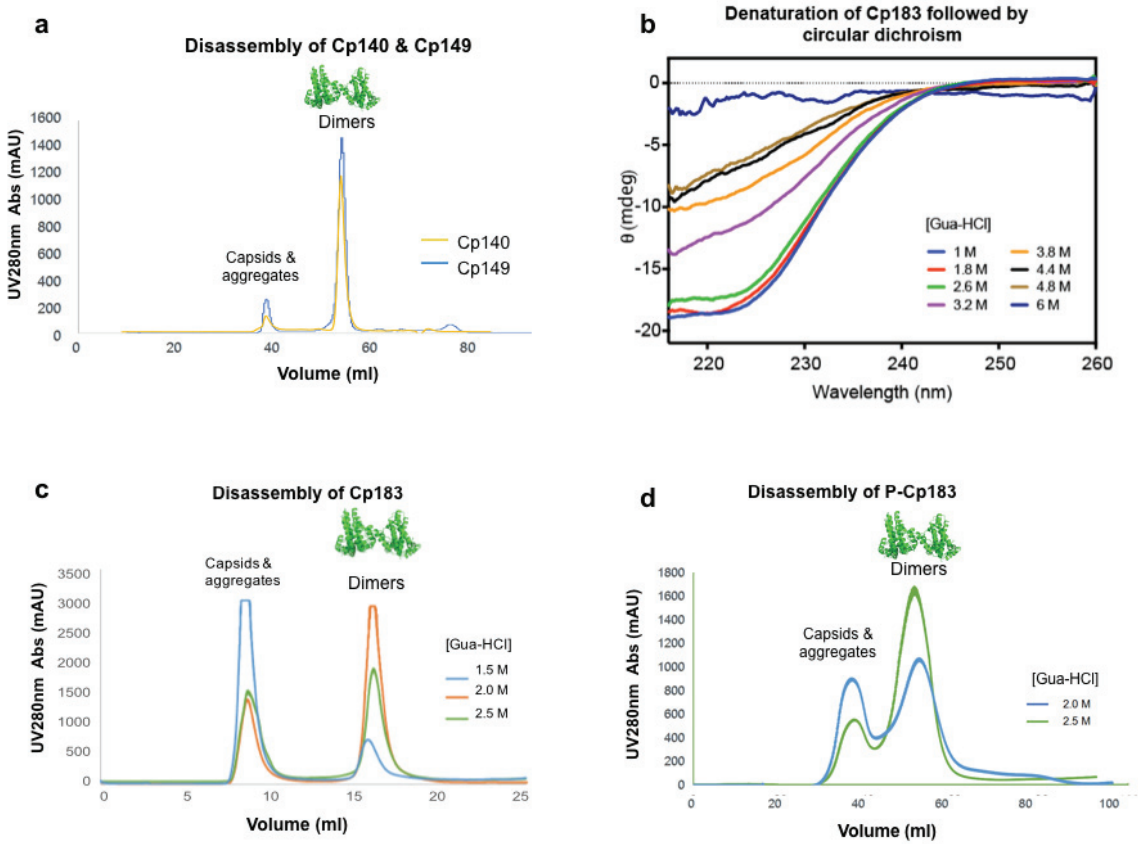


Figure 54. Disassembly of HBV Cps. a): disassembly of Cp140 and Cp149 using 3 M urea. b): the denaturation tendency of Cp183 caused by Gua-HCl from in 1 M to 6 M, followed by circular dichroism at wavelength of 222 nm. c): disassembly test of Cp183. The disassembly shows best efficiency with a concentration of Gua-HCl of 2 M. d): disassembly of P-Cp183. The best condition for P-Cp183 disassembly is 2.5 M Gua-HCl instead of 2M Gua-HCl. The columns used for gel filtration are HiPrep 16/60 Sephacryl S-200 HR (a, d) and Superdex 200 10/300 column (c), produced by GE Healthcare.

For full-length Cps, preformed capsids from *E. coli* were dialysis against disassembly buffer containing Guanidine instead of Urea as published in (Porterfield et al., 2010). The impact of Guanidine concentration on Cp183 was evaluated using circular dichroism (Figure 54-b) and showed that Gua-HCl below 2.6 M does not denaturate Cp183 secondary structure. Then, the disassembly efficiency was checked for Cp183 (Figure 54-c) and P-Cp183 (Figure 54-d) as a function of Gua-HCl concentration in order to find optimal disassembly conditions. Thereby, Cp183 and Cp183 mutants capsids were disassembled by dialysis against disassembly buffer B (Table 6) containing 2 M Gua-HCl, while P-Cp183 capsids were disassembled by dialysis

against disassembly buffer C (Table 6) containing 2.5 M Gua-HCl, both overnight at 4 °C. Cp dimers were separated similarly as for truncated Cps. Disassembly was more efficient at low temperature, by keeping buffers at 4 °C prior to column equilibration and protein injection. Encapsidated RNA precipitated by LiCl was removed by centrifugation (20,000 g, 15 min at 4 °C). As P-Cp183 capsids do not contain RNA, the presence of LiCl and this centrifugation step are not necessary.

For the reassembly of phosphorylated or phospho-mimetic empty capsids, P-Cp183 or Cp183 EEE dimers were reassembled by dialysis against reassembly buffer A (Table 6) overnight at room temperature. For the other full-length proteins, the reassembly step required the addition of RNA in order to equilibrated the charges and avoid protein precipitation. For the reassembly of RNA-packaged capsids, Cp183 WT and mutant dimers in 2 M Gua-HCl at ~0.8 mg/ml were diluted 4 times in dilution buffer (Table 6) to reach a final concentration of 0.5 M Gua-HCl, and dialyzed in presence of RNA against reassembly buffer B (Table 6) overnight at room temperature. Two types of RNA were used for the reassembly: a mix of RNAs from torula yeast bought from Sigma-Aldrich with a Cp dimer:RNA molar ratio of 20:1, and the pgRNA prepared as described above with a Cp dimer:RNA molar ratio of 120:1.

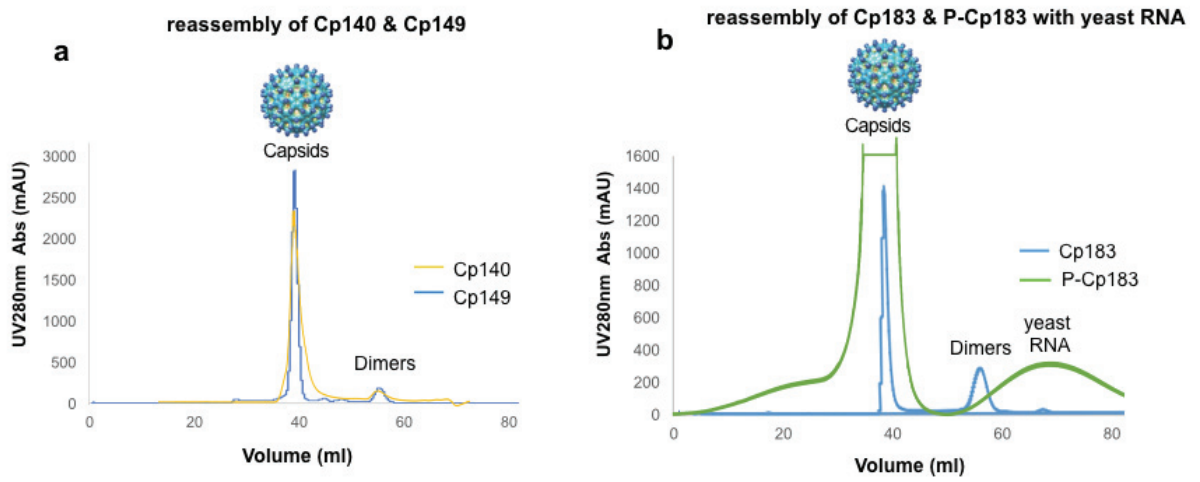


Figure 55. Reassembly of HBV Cps. The reassembled capsids of truncated (a) and full-length (b) Cps were analyzed by gel filtration with column of HiPrep™ 16/60 Sephacryl S-200 HR. Contrary to Cp183, P-Cp183 did not package yeast RNA, due to the phosphorylation which neutralized the positive charges of the CTD.

➢ **Negative staining electron microscopy (EM)**

Capsids autoassembled in *E. coli* or reassembled *in vitro* were characterized by negative staining EM. The protein solutions were diluted to 1 mg/ml and 5 µl were deposited on the carbon-coated grids. The samples were incubated on the grids for 2 min and then the excess of solution was removed using Whatman paper. The negative staining was carried out by layering the grid on 50 µl of 2 % phosphotungstic acid (PTA) at pH 7 for 2 min at room temperature. Excess PTA solution remaining on the grids was removed using Whatman paper. Sample grids were analyzed using a JEM-1400 transmission electron microscope operating at 100 kV. The capsids observed in electron micrographs display diameters of about 30 nm or 26 nm (Figure 56, Figure 57), which are consistent with the size of T=4 and T=3 capsids found in the literature (Selzer et al., 2014).

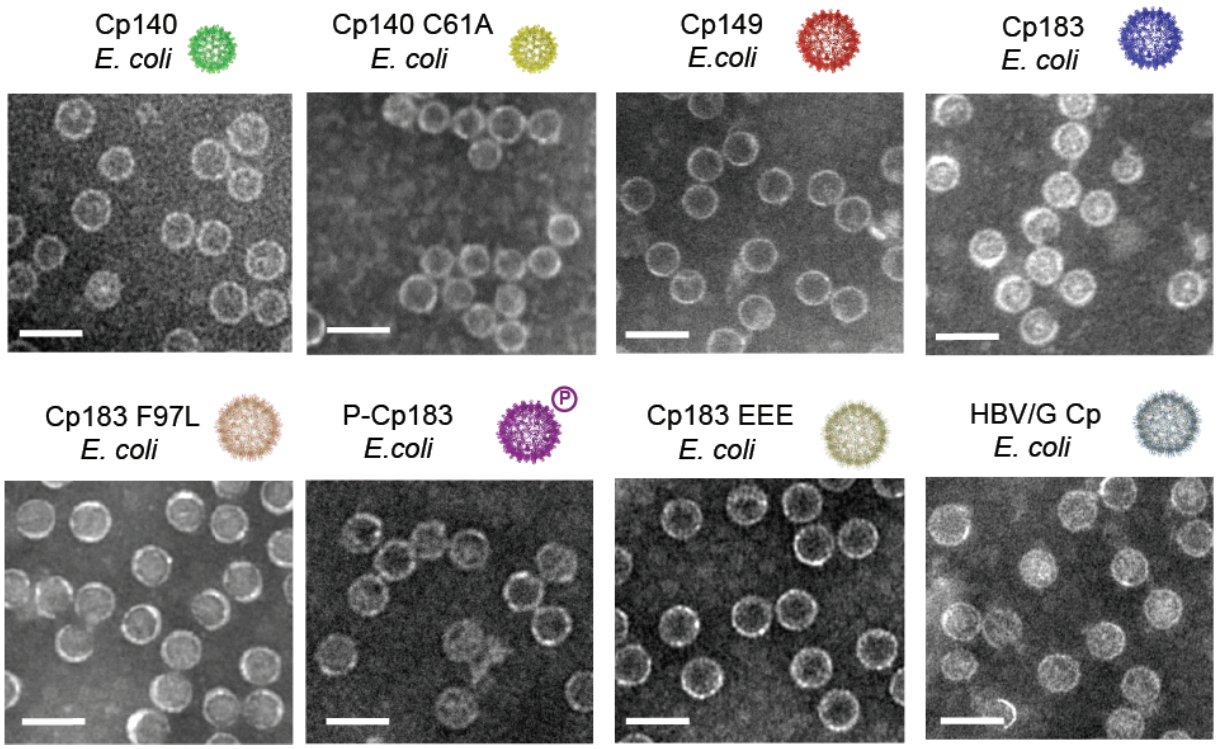


Figure 56. Electron micrographs of autoassembled capsids in *E. coli*. Scale bar is for 50 nm.

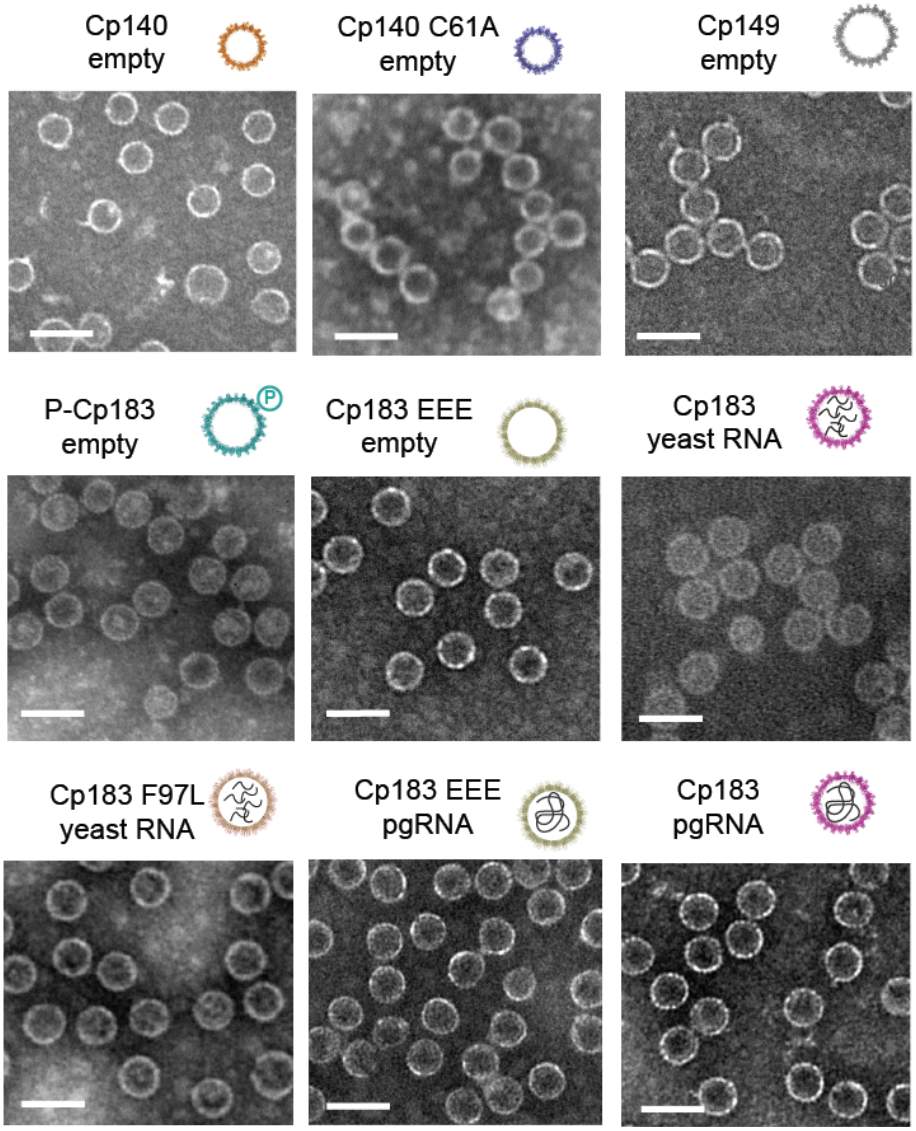


Figure 57. Electron micrographs of reassembled capsids. Scale bar is for 50 nm.

➢ Interaction of HBV capsids with CAMs

The interactions between Cps and capsid assembly modulators (CAMs) were carried out *in vitro* using preformed capsids from *E. coli* or Cp149 dimers which were subsequently reassembled into capsids in presence of CAMs. The CAMs were provided by Gilead and Johnson-Johnson companies and include three CAM-N compounds: AT-130, JNJ-632 and JNJ827, and one CAM-A compound (HAP): JNJ-890. The three CAMs were dissolved in DMSO to prepare 10 mM stock solutions. Preformed ^{13}C-^{15}N-Cp149 and ^{13}C-^{15}N-Cp183 capsids were incubated at 37 °C for 2 h

with a molar ratio of Cp monomer:CAMs of 1:4. For Cp149 reassembled with AT-130, ^{13}C-^{15}N-Cp149 dimers were incubated at 4 °C for 24 h with 150 mM NaCl together with a molar ratio of Cp monomer:AT-130 of 1:4, and subsequently dialyzed against purification buffer (Table 6) overnight at 4 °C. The interaction of Cps and CAMs was analyzed by negative staining EM (Figure 58), and will be characterized by solid-state NMR.

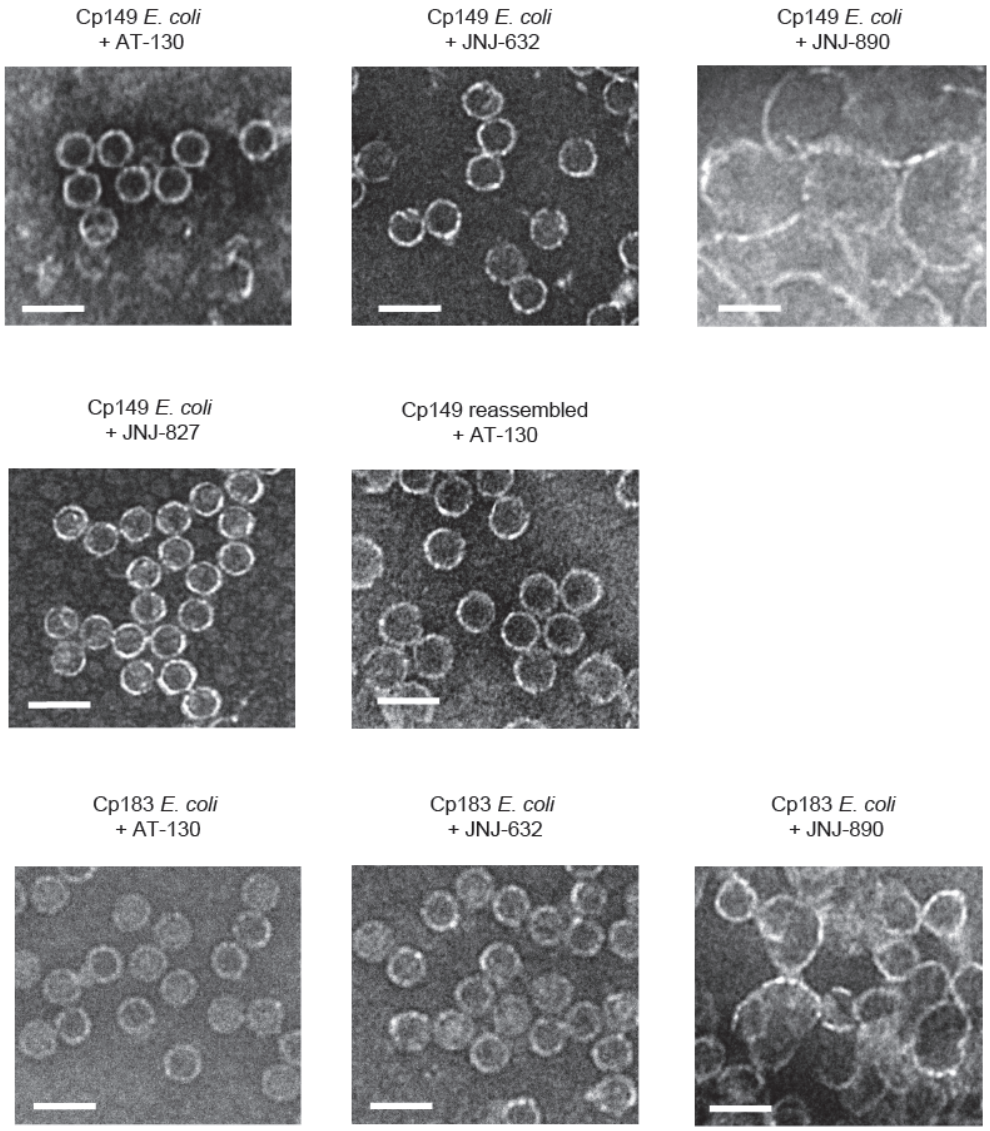


Figure 58. Electron micrographs of HBV capsids in interaction with CAMs. The AT-130, JNJ-632, JNJ-827, which belong to CAM-N class results in apparently normal capsids; JNJ-890, which belongs to CAM-A class leads to aberrant capsids. Scale bars indicate 50 nm.

➢ References

Beckert, B., Masquida, B., 2010. Synthesis of RNA by In Vitro Transcription, in: Chen, Y.W. (Ed.), Structural Genomics, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 29–41. doi:10.1007/978-1-59745-248-9_3

Burrell, C.J., MacKay, P., Greenaway, P.J., Hofschneider, P.H., Murray, K., 1979. Expression in Escherichia coli of hepatitis B virus DNA sequences cloned in plasmid pBR322. Nature 279, 43–47.

Cohen, B.J., Richmond, J.E., 1982. Electron microscopy of hepatitis B core antigen synthesized in E. coli. Nature 296, 677–679.

Correa, A., Oppezzo, P., 2014. Overcoming the Solubility Problem in E. coli: Available Approaches for Recombinant Protein Production, in: Chen, Y.W. (Ed.), Structural Genomics, Methods and Protocols. Springer New York, New York, NY, pp. 27–44. doi:10.1007/978-1-4939-2205-5_2

Gopal, G.J., Kumar, A., 2013. Strategies for the Production of Recombinant Protein in Escherichia coli. Protein J 32, 419–425. doi:10.1007/s10930-013-9502-5

Green, M.R., Sambrook, J., 2017. Isolation of High-Molecular-Weight DNA Using Organic Solvents. Cold Spring Harb Protoc 2017, pdb.prot093450. doi:10.1101/pdb.prot093450

Green, M.R., Sambrook, J., 2016. Precipitation of DNA with Ethanol. Cold Spring Harb Protoc 2016, pdb.prot093377. doi:10.1101/pdb.prot093377

Heger-Stevic, J., Kolb, P., Walker, A., Nassal, M., 2018a. Displaying Whole-Chain Proteins on Hepatitis B Virus Capsid-Like Particles, in: Chen, Y.W. (Ed.), Structural Genomics, Methods in Molecular Biology. Springer New York, New York, NY, pp. 503–531. doi:10.1007/978-1-4939-7808-3_33

Heger-Stevic, J., Kolb, P., Walker, A., Nassal, M., 2018b. Displaying Whole-Chain Proteins on Hepatitis B Virus Capsid-Like Particles, in: Chen, Y.W. (Ed.), Structural Genomics, Methods in Molecular Biology. Springer New York, New York, NY, pp. 503–531. doi:10.1007/978-1-4939-7808-3_33

Heger-Stevic, J., Zimmermann, P., Lecoq, L., Böttcher, B., Nassal, M., 2018c. Hepatitis B virus core protein phosphorylation: Identification of the SRPK1 target sites and impact of their occupancy on RNA binding and capsid structure. PLoS Pathog 14, e1007488. doi:10.1371/journal.ppat.1007488

Pasek, M., Goto, T., Gilbert, W., Zink, B., Schaller, H., Mackay, P., Leadbetter, G., Murray, K., 1979. Hepatitis B virus genes and their expression in E. coli. Nature 282, 575–5. doi:10.1038/282575a0

Porterfield, J.Z., Dhason, M.S., Loeb, D.D., Nassal, M., Stray, S.J., Zlotnick, A., 2010. Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. J. Virol. 84, 7174–7184. doi:10.1128/JVI.00586-10

Selzer, L., Katen, S.P., Zlotnick, A., 2014. The hepatitis B virus core protein intradimer interface modulates capsid assembly and stability. Biochemistry 53, 5496–5504. doi:10.1021/bi500732b

Vera, A., González-Montalbán, N., Arís, A., Villaverde, A., 2007. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol. Bioeng. 96, 1101–1106. doi:10.1002/bit.21218

Wynne, S.A., Crowther, R.A., Leslie, A.G., 1999. The Crystal Structure of the Human Hepatitis B Virus Capsid. Molecular Cell 90, 5830–5844.

Zlotnick, A., CHENG, N., Conway, J.F., Booy, F.P., Steven, A.C., Stahl, S.J., Wingfield, P.T., 1996. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. Journal of Biological Chemistry 35, 7412–7421. doi:10.1021/bi9604800

● NMR characterization of HBV core proteins

The work in this part was done with Lauriane Lecoq, in collaboration with Beat H. Meier, and has been published in following paper:

Solid-state [^{13}C-^{15}N] NMR resonance assignment of Hepatitis B Virus core protein

Lauriane Lecoq#, Shishan Wang#, Thomas Wiegand, Stéphane Bressanelli, Michael Nassal*, Beat H. Meier*, Anja Böckmann*.

Biomolecular NMR Assignments. 2018. 12, 205–214.

Contribution: as a co-first author, I executed biochemical experiments part of this paper, contributed to writing the paper.

Lauriane Lecoq executed the experiments of solid-state NMR.

➢ Abstract

Each year, nearly 900,000 deaths are due to serious liver diseases caused by chronic hepatitis B virus (HBV) infection. The viral particle is composed of an outer envelope and an inner icosahedral nucleocapsid formed by multiple dimers of a ~20 kDa self-assembling core protein. Here we report the solid-state ^{13}C and ^{15}N resonance assignments of the HBV core protein in its capsid form. A secondary chemical shift analysis of the 140 visible residues suggests an overall alpha-helical three-dimensional fold similar to that derived from X-ray crystallography of the capsid and from solution-state NMR of the core protein dimer. Interestingly, however, at three distinct regions the chemical shifts in solution differ significantly between core proteins in the capsid state versus in the dimer state, strongly suggesting the respective residues to be involved in capsid assembly.

Keywords:

Hepatitis B virus • Core protein • Nucleocapsid • Solid-state NMR • Assignments

➢ Biological Context

More than 2 billion people have been infected with hepatitis B virus (HBV), and more than 250 million thereof have become chronic carriers of the virus. Each year, close to 900,000 people die from hepatitis B as they develop chronic liver disease, including cancer (World Health Organization 2017). While an effective prophylactic vaccine is available to prevent infection, current treatments for chronic HBV infection are rarely curative and subject to the emergence of resistant virus variants (Zoulim and Durantel 2015). Hence new therapeutic options are urgently sought (Revill et al. 2016).

The HBV virus particle (virion) is composed of an outer envelope bearing three related surface antigen (HBsAg) proteins and an inner nucleocapsid formed by multiple copies of a single about 20 kDa core protein (Cp) species. Cp forms stable dimers capable of

self-assembling the capsid shell. The predominant form consists of 120 Cp dimers arranged inT=4 icosahedral symmetry. In infection, newly forming nucleocapsids initially co-package a complex of one copy of the viral pregenomic (pg) RNA plus one copy of the viral polymerase (Pol); maturation involves capsid-assisted reverse transcription of the pgRNA into a relaxed circular DNA by Pol (Nassal 2015). The N terminal ~140 amino-acids of Cp, modeled by C terminally truncated variants like Cp149, are required and sufficient for capsid shell formation, even in bacteria (Gallina et al. 1989; Birnbaum and Nassal 1990). The downstream arginine-rich C-terminal domain (CTD) generally binds nucleic acid (Birnbaum and Nassal 1990) but is also required for specific viral pgRNA packaging (Nassal 1992). 3D structure first by cryo-electron microscopy (EM) reconstruction revealed mostly an α-helical fold (Conway et al. 1997; Böttcher et al. 1997). The crystallographic structure of Cp149 solved in 1999 at a resolution of 3.3 Å (Wynne et al. 1999) confirmed the mainly α-helical structure. Since then, further structures of the HBV capsid have been solved either by cryo-EM (Crowther et al. 1994; Dryden et al. 2006; Chen et al. 2011; Yu et al. 2013; Selzer et al. 2015; Patel et al. 2017) or crystallography (Bourne et al. 2006; Katen et al. 2013). Yet, despite these structural studies, many steps along the viral life cycle involving the core protein are not yet fully understood; namely, as the conformation of the capsids is believed to be a function of the different maturation states (Ning et al. 2017), it is central to structurally analyze the capsids under a larger variety of conditions. In this context, solid-state NMR is an attractive technique since it allows the study of large molecular assemblies as viral capsids at physiological temperature, including those not forming crystalline states, which enables the study of different constructs, formed in various environments. First applications of solid-state NMR to virus capsids have been described in the literature and made significant advances in the understanding of different viral cycles. For example, Polenova and coworkers have determined a structural model of the human immunodeficency virus 1 (HIV-1) capsids and have shown that the internal flexibility of the capsid protein is essential for its assembly and interaction with host partners (Han et al. 2010; Quinn et al. 2015; Suiter et al. 2015). Solid-state NMR also allowed to characterize the dynamics of the filamentous and icosahedral fd bacteriophage as well as to probe the capsid regions involved in the interaction with the DNA genome of the virus (Morag et al. 2014; Abramov et al. 2015).

More recently, the *de novo* structure of the *Acinetobacter* phage 205 has been solved (Andreas et al. 2016) using the advances in NMR technology with faster spinning frequencies (110 kHz) and ^{1}H detection techniques (Penzel et al. 2015; Stanek et al. 2016). All these studies have paved the way towards the analysis of a large variety of viral capsid assemblies using magic-angle spinning (MAS)-NMR, in order to understand the viral life cycle.

Here we report the ^{13}C and ^{15}N resonance assignments of the HBV core protein Cp149 in the capsid form expressed in *E. coli* using 2D and 3D solid-state NMR experiments. The spectra obtained show narrow lines (between 70 and 100 Hz in ^{13}C). We compare the obtained chemical shifts and secondary structures to previous structural studies by X-ray and liquid-state NMR, and highlight that the differences are mainly located at the dimer interfaces. Preliminary data indicate this is extendable to full-length Cp183 particles, either filled with RNA, or mostly devoid of RNA upon extensive phosphorylation of the C-terminal domain.

➢ Materials and methods

Protein expression and purification; sample preparation

Cp149 was expressed and purified as described previously (Heger-Stevic et al. 2017). Briefly, the pET-28a2-Cp149opt plasmid was transformed into *E. coli* BL21* Codon + (DE3) cells and grown at 37 °C overnight on a LB-agar plate containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. Then, a single colony was inoculated into 3 ml LB medium and grown at 37 °C for 5 h. It was then transferred into 50 ml of standard M9 minimal medium containing $^{15}NH_4Cl$ and ^{13}C-glucose as the sole nitrogen and carbon sources. The culture was incubated for 4 hours at 37 °C and transferred into 100 ml fresh ^{13}C-^{15}N M9 medium for an overnight incubation at 25 °C. 850 ml of ^{13}C-^{15}N M9 medium were inoculated with the 150 ml overnight ^{13}C-^{15}N M9 culture. When OD_{600} reached 0.7, the bacterial cells were grown for another 6 h at 25 °C after induction with 1 mM IPTG.

Bacterial cells were pelleted by centrifugation (6,000 *g*, 20 min) and re-suspended in 15 ml of lysis buffer (50 mM Tris pH 7.4, 300 mM NaCl, 5 mM DTT). 15 mg of chicken lysozyme, 300 µl of protease inhibitor cocktail solution 50X and 750 *µ*l of 10% Triton-X-100 were added to the cell suspension and incubated on ice for 45 min. Benzonase nuclease was added to the suspension and mixed for 30 min at room temperature. The cells were lysed by sonication and then centrifuged at 8000 *g* for 1 h to remove cell debris. The supernatant was loaded onto a step gradient from 10% to 70% (m/v) sucrose (buffered in 50 mM Tris pH 7.5, 300 mM NaCl, 5 mM DTT) and centrifuged in SW-32Ti Beckman Coulter rotor at 28,800 *g* for 3 h at 4 °C. The gradients were divided into 14 fractions. The fractions containing Cp149 capsids were identified by SDS-PAGE and precipitated by slow addition of $(NH_4)_2SO_4$ to 20% saturation. After incubation on ice for 1 h and centrifugation at 25,000 *g* for 1 h, pellets were resuspended in purification buffer (50 mM Tris pH 7.5, 5 mM EDTA, 5 mM DTT, 5% sucrose). The protein solution was centrifuged again for 15 min to remove insoluble pellets and the supernatant containing soluble capsids was dialyzed overnight in purification buffer at 4 °C.

Electron microscopy

5 µL of sample were loaded on a carbon-coated grid and incubated for 2 minutes. The grid was then layered on top of 50 µL of 2 % phosphotungstic acid (m/v) pH 7 and incubated at room temperature for 2 minutes for negative staining. The grids were examined with a JEM-1400 transmission electron microscope operating at 100 kV.

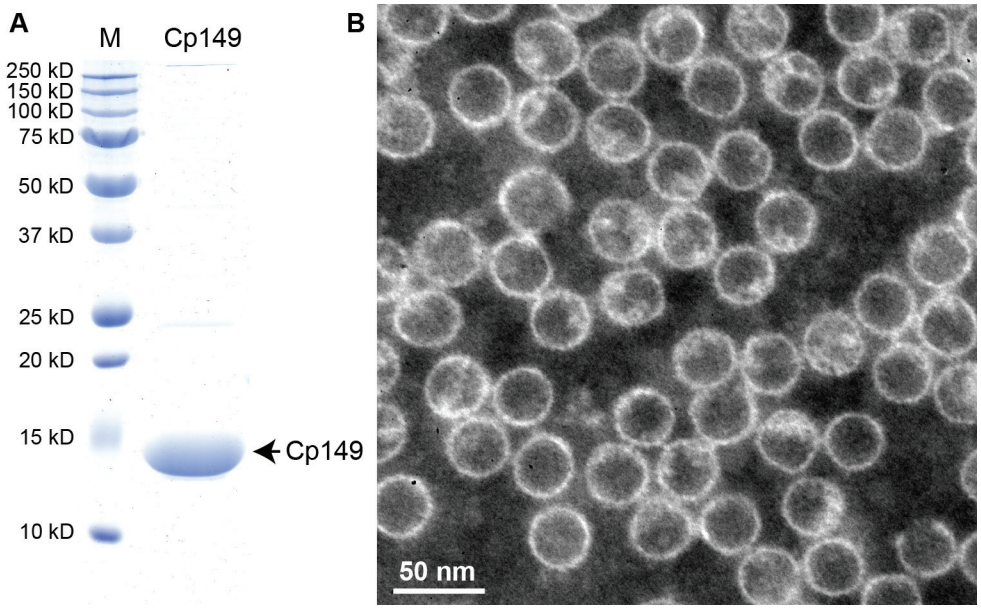


Fig. 1 Sample characterization. **A.** SDS-PAGE 15% of a molecular weight marker (Precision Plus Protein™ Unstained Standards from BioRad) and 20 μg of purified Cp149 expressed in *E. coli*. **B.** Negative-stain EM picture of Cp149 concentrated at 1 mg/ml.

Solid-state NMR spectroscopy

For NMR sample preparation, we started from 40 mg of soluble capsids which were concentrated to a final volume of l ml and sedimented into a thin-wall 3.2 mm zirconium rotor by centrifugation (180,000 *g*, 14 h, 4 °C) using a home-made filling tool (Böckmann et al. 2009). All NMR experiments were conducted using a 3.2 mm triple-resonance (^{1}H, ^{13}C, ^{15}N) probe head at static magnetic field of 18.8 T corresponding to 800 MHz proton resonance frequency (Bruker Avance II). Assignments were obtained at a MAS frequency of 17.5 kHz using a combination of 2D and 3D correlation experiments, including: 2D DARR, 3D NCACB, NCACX, NCOCX, CANCO and CCC (Schuetz et al. 2010; Habenstein et al. 2011). The 3D CCC was performed with the probe in double-resonance (^{1}H, ^{13}C) mode to increase the signal to noise ratio by a factor of 1.4. Experimental details are given in Table 1. All spectra were referenced using DSS and recorded at a sample temperature of 4 °C, as determined by the resonance frequency of the supernatant water (Böckmann et al. 2009). Two-dimensional DARR, NCA and NCO spectra were recorded on several other preparations of the same protein, with always reproducible spectra. Three-dimensional

spectra were subsequently recorded for 2–6 days each on a single sample of Cp149. Side-chain assignments were achieved using NCACB, NCACX, NCOCX and CCC spectra (Schuetz et al. 2010; Habenstein et al. 2011). All spectra were processed using TopSpin 3.2 (Bruker Biospin) by zero filling to no more than double the number of acquired points, the time domain signals were apodized with a squared cosine function (SSB 2.5–3 depending on spectrum and dimension). Spectra analysis and assignment was done with the CcpNmr Analysis package (Vranken et al. 2005; Stevens et al. 2011).

➢ Assignments and data deposition

1. Sequential assignments and secondary chemical shifts

^{13}C and ^{15}N backbone and side-chain chemical shifts of the Cp149 capsid protein have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number 27317. Capsids expressed with high yield of minimum 50 mg per liter culture, and final preparations showed a high purity level as shown in Figure 1A. Purified proteins were analyzed under the electron microscope. As described previously (Birnbaum and Nassal 1990), capsids were assembled directly during bacterial expression, as numerous HBV core protein capsids were observed under the electron microscope (see Figure 1B for negative-stain EM pictures). Capsid preparations were homogenous and showed mainly T=4 capsids with a typical diameter of 30 ± 3 nm. The 2D ^{13}C-^{15}N NCA spectrum of assigned Cp149 capsids is displayed in Figure 2A and shows narrow peaks with full widths at half maximum (FWHM) between 70 and 100 Hz in ^{13}C, which is comparable to line widths obtained previously in viral capsid proteins such as the HIV-1 virion (Suiter et al. 2015) and is indicative for homogeneous preparations of uniformly labeled proteins. Representative planes of the 3D NCACB, NCACO, CANCO and NCOCA experiments are shown in Figure 2B. Assignments of Cp149 were completed at 87 % for all backbone residues, with missing assignments predominantly found at the protein C-terminal end including residues 140 to 149 for which no cross peaks were detected in 3D spectra. In detail, 95.6 % of N. 97.1 % of Cα. 96.2 % of Cβ. 97.8 % of C' and 75.3 % of side chain non-H were unambiguously assigned for residues 1–139 of Cp149.

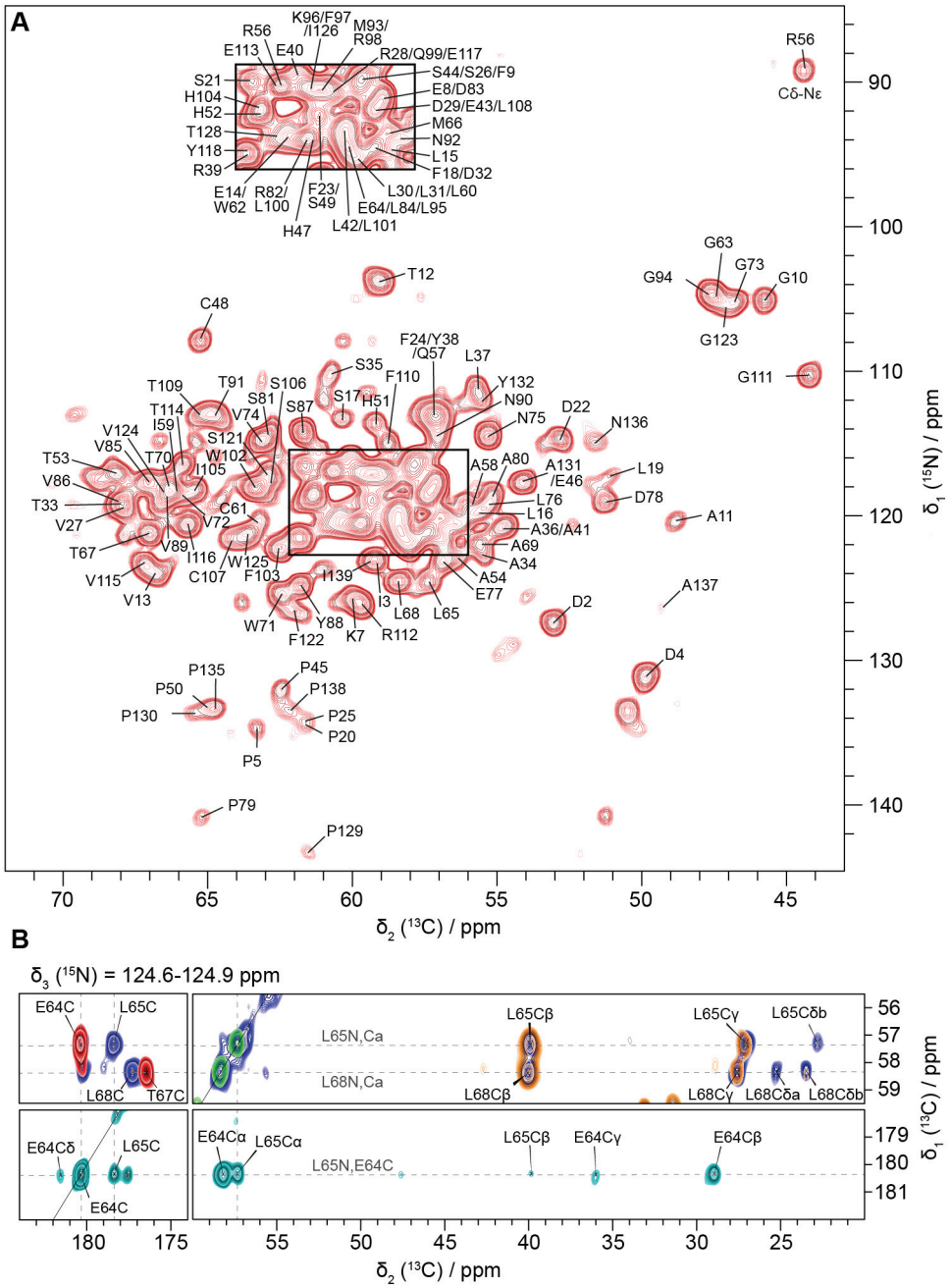


Fig. 2 Well-resolved spectra allow for unambiguous assignment: A) 2D ^{13}C-^{15}N NCA spectrum of Cp149 capsids. All assigned N-Cα intra-residue correlations are labeled with the residue number as well the Cα-Cδ intra-residue correlations for proline residues. The middle of the spectrum is copied inside the black rectangle for more clarity. **B)** Representative plane for residue Leu65 of 3D CANCO (red), NCACX (dark blue), NCACB (green for positive signals and orange for negative signals) and NCOCX (cyan) spectra. For experimental details see Table 1.

A secondary chemical-shift analysis (Figure 3) reveals a backbone fold similar to that of the X-ray capsid structure as described in (Wynne et al. 1999). Helices α1 (residues 13-17) and α2 (residues 27-43) appear to be the same length in both capsids, furthermore the smaller values of secondary chemical shifts obtained for residues 37 and 38 are representative of the 90°-kink observed in the X-ray structure. NMR identifies two additional α-helical turns of four residues each, between residues 5-8 and 21-24, and possibly a short β-strand or extended conformation from residues 1 to 4. Helix α3 (residues 50-76 in solid-state NMR) is predicted to be longer by three residues, while helix α5 is predicted to be two-residues shorter (residues 112-125). Another difference is observed in helix α4 that is not divided into two smaller helices in the NMR predictions but instead would form a continuous 30-amino acids helix, shorter by one residue compared to the X-ray structure.

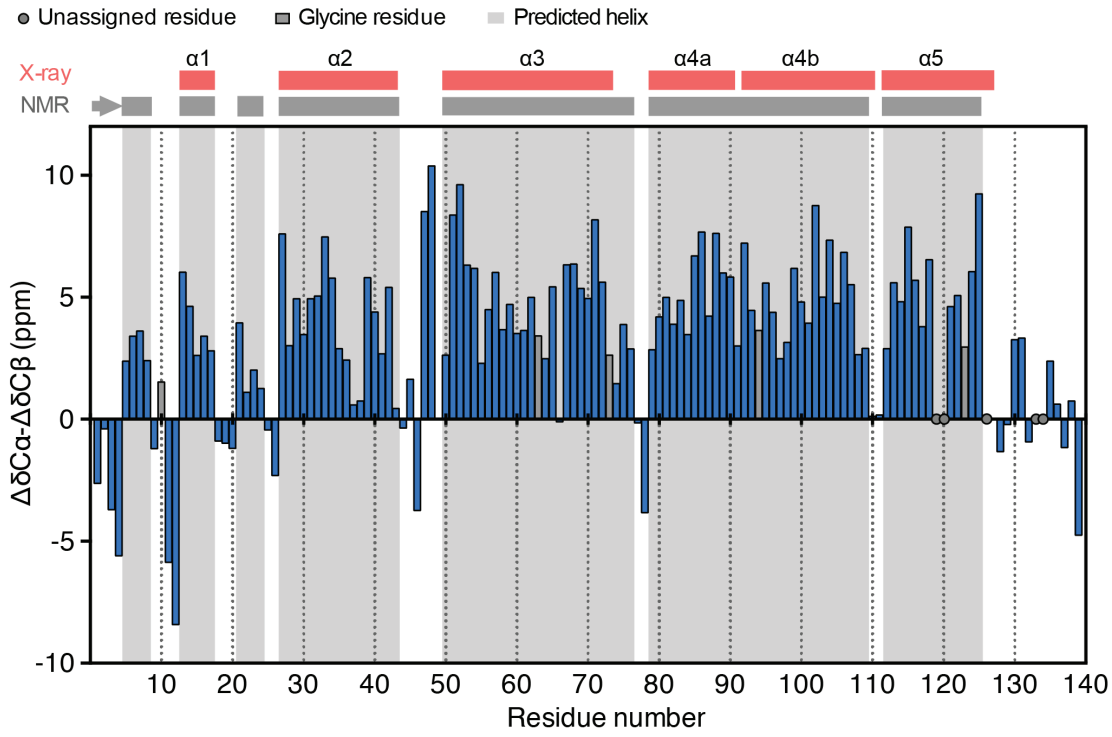


Fig. 3 Differences between secondary chemical shifts of Cα and Cβ resonances relative to their random coil shift (Wishart et al. 1995) as a function of Cp149 residue number. Missing assignments are represented as grey circles and include residues L119, V120, I126, R133, P134, as well as the C-terminal end from 140 to 149. For glycine residues (shown as grey bar), the deviation of the Cα shift from the random coil value is plotted. The α-helices found in the X-ray structure of Cp149 capsid as described in (Wynne et al. 1999) are shown as pink boxes at the top of the graph and the ones predicted from solid-state NMR chemical shifts are shown as grey boxes.

Four residues in a row with a positive value are indicative of an α-helix and three residues in a row with a negative value indicate a *β-sheet.* For the three Cys residues, the random coil chemical shifts corresponding to the reduced form were taken.

2. Comparison to Cp149 dimer chemical shifts from solution-state NMR

Chemical shifts of the dimeric core protein have previously been determined using solution-state NMR on a protein comprising a slightly different amino-acid sequence (Freund et al. 2008); furthermore, to prevent the dimers from assembling, solution spectra were taken at pH 9.7. The backbone assignments here presented for the capsid form of Cp149 are slightly more complete than the ones determined for the Cp149 dimer, as conformational exchange induced the loss of several signals in the solution-state study. Also, the high pH used in that study might have been responsible for solvent-saturation transfer effects. A comparison between the dimer and capsid chemical shifts is shown in Figure 4A for Cα and Cβ resonances. Chemical shifts of the six residues which were different between the two Cp149 constructs used were disregarded (A35, D40, D64, T74, V93 and L116 in the solution study), and also, chemical shift differences, which we interpret as chemical-shift perturbations (CSPs) upon assembly, observed for neighboring residues in the sequence, as residues N75 and L76. While overall chemical shifts are rather similar between the dimer and capsid form, interesting exceptions are observed in three distinct regions identified in Figure 4B. The first region includes residues 18 to 23, which are involved in contacts with two different surrounding dimers in the capsid structure. The residues with which they are in interaction on the neighboring dimers comprise amino acids 136-138, which however have not been assigned for the dimer in solution. Other significant CSPs are observed in a second region from residues 55 to 62, which could be due to the different oxidation state of Cys61, or alternatively to allosteric changes induced by the interaction between the dimers in the capsid form. These changes could be transmitted via a third region showing CSPs comprising residues 96 and 97, which is located in between regions 1 and 2. Met1 also shows an isolated CSP, probably due to the close proximity with the second region affected by capsid assembly. These results reveal the three major regions undergoing a conformational change between the core protein in solution and the core protein assembled as a capsid.

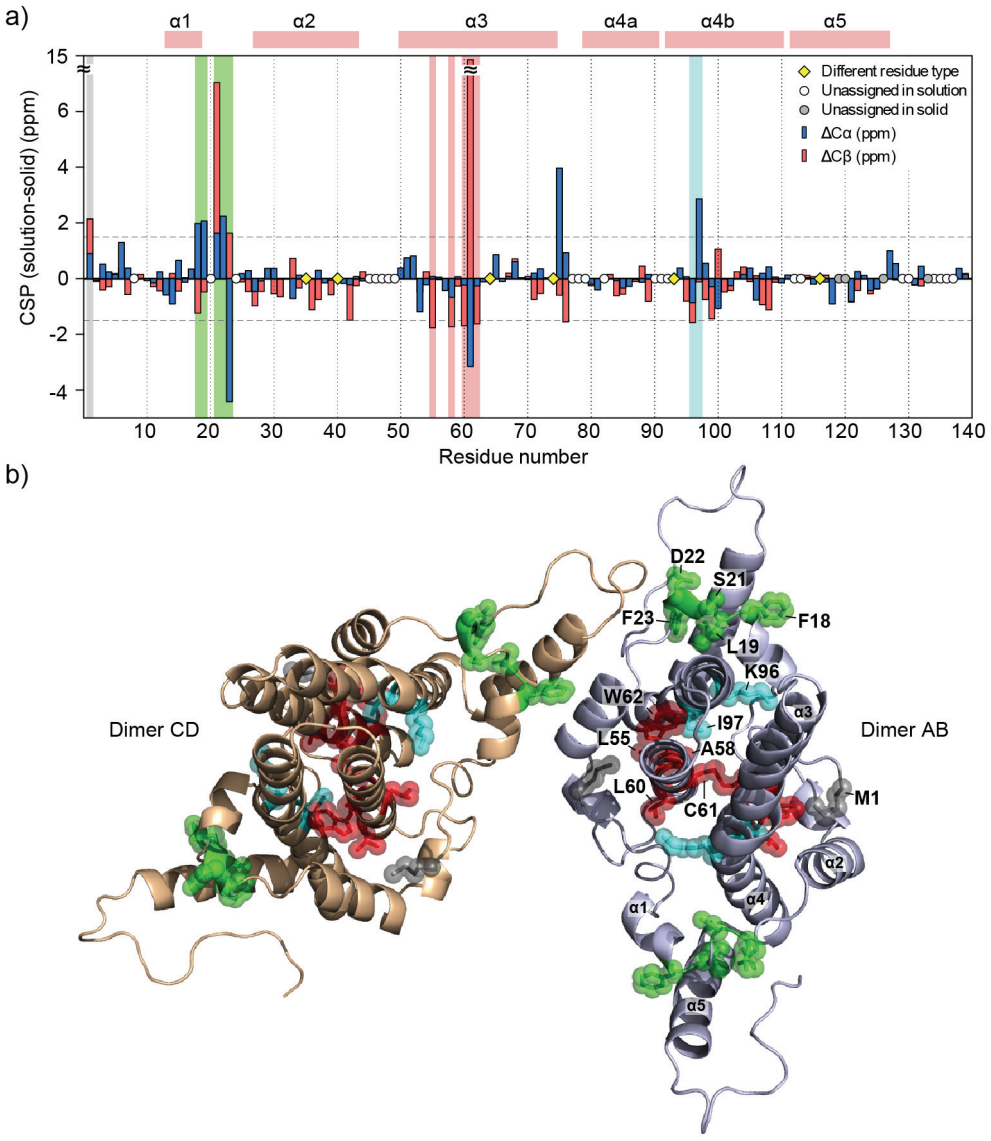


Fig. 4 A) Differences between Cα (in blue) and Cβ (in red) resonances of dimeric Cp149 (from solution-state NMR) and of the Cp149 capsids (from solid-state NMR). Yellow diamonds stand for residues which were different between the two Cp149 constructs. Missing data points are due to missing assignments in either the solution-state (white circles) or the solid-state NMR study (grey circles). Secondary structure elements extracted from the X-ray structure are shown above the graph as pink boxes. Chemical shift differences below -1.5 ppm or above 1.5 ppm (grey dotted lines) are considered as significant. Three regions were identified and are colored in green, red, and blue. Isolated Met1 is highlighted in grey. **B)** Sphere representation of residues with significant chemical-shift differences between Cp149 dimer and Cp149 capsid. It includes residues M1, F18, L19, S21, D22, F23, L55, A58, L60, C61, W62, K96 and I97, which are labeled on subunit B and colored as in panel A. Residues N75 and L76 are not included, as the V74 to T74 substitution in the dimer in solution is likely to be at the origin of the observed CSPs. Helices are labeled on

subunit A. PDB: 1QGT. The figures involving structures were all prepared using the PyMol software (Warren Delano, http://www.pymol.org).

➢ Conclusion

The contribution of MAS-NMR spectroscopy to understanding structural and dynamical features of virus-like particles has become significant overt the past years. Here we have shown that assembled hepatitis B virus capsids can be studied by solid-state NMR. This is of particular importance to further analyze the conformational dynamics the core protein is thought to undergo upon its multiple interactions with viral and cellular factors that eventually yield a mature infectious virion. Labeled Cp149 protein can be produced in pure form and in large quantity in *E. coli* and the protein yields high-quality spectra with narrow lines and a dispersion typical for an α-helical protein. Sequential assignments are to 87 % complete for Cp149 backbone, and yield secondary chemical shifts which are compatible with available X-ray structures of the capsid. Signals for residues downstream of residue 140 are not detectable by either technique, owing to the dynamic behavior of the C-terminus. However, a comparison with the chemical shifts reported for the Cp149 dimer (Freund et al. 2008) highlights conformational differences in three distinct regions of the protein, suggesting they are structurally affected by the transition from the dimer state to the capsid state. In summary, the work presented here sets the stage for further, high resolution NMR studies of the HBV core protein which is central in HBV infection and thus is also becoming a highly attractive target for novel anti-HBV therapeutics (Zlotnick et al. 2015; Durantel and Zoulim 2016; Pei et al. 2017).

➢ Acknowledgements

This work was supported by the French ANR (ANR ANR-14-CE09-0024B), the Swiss National Science Foundation (Grant 200020_159707), and by the Marie Skłodowska-Curie program (H2020-MSCA-IF-2016 748516). We thank the Centre d'Imagerie Quantitative Lyon-Est (CIQLE) for the access and training on the electron microscopy platform.

Experiment	DARR	NCA	NCO	NCACB	NCACX	NCOCX	CANCO	CCC
Transfer 1	HC-CP	HN-CP	HN-CP	HN-CP	HN-CP	HN-CP	HC-CP	HC-CP
Field [kHz]	65.7(H)/50(C)	56.1(H)/43.1(N)	56.1(H)/43.1(N)	57.7(H)/43.1(N)	61.7(H)/43.1(N)	60.2(H)/43.1(N)	65.1(H)/50(C)	65.2(H)/50(C)
Shape	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H
^{13}C carrier [ppm]	56.5							55
time [ms]	1	0.9	0.9	0.9	0.8	0.8	0.6	0.7
Transfer 2	DARR	NCa-CP	NCO-CP	NCa-CP	NCa-CP	NCO-CP	CaN-CP	DREAM
Field [kHz]	17.5(H)	6(C)/10.9(N)	6(C)/11.5(N)	6(C)/11.5(N)	6(C)/10.8(N)	6(C)/11.6(N)	6(C)/11.5(N)	8(C)
Shape	-	Tangent ^{13}C	Tangent ^{13}C	Tangent ^{13}C	Tangent ^{13}C	Tangent ^{13}C	Tangent ^{13}C	Tangent ^{13}C
^{13}C carrier [ppm]	97.8	54	174.9	59.9	58.6	176.5	59.3	56
time [ms]	20	6	6	6	6.5	3	6.5	4
Transfer 3				DREAM	DARR	DARR	NCO-CP	DARR

Field [kHz]				8(C)	17.5(H)	17.5(H)	6(C)/10.9(N)	17.5(H)
Shape				Tangent ^{13}C	-	-	Tangent ^{13}C	-
Carrier [ppm]				56	58.6	176.5	176.2	43.8
time [ms]				3	70	30	7	80
t_1 increments	2560	1344	1344	74	64	96	135	200
sw (t_1) [kHz]	93.75	62.5	62.5	5.7	5.7	5.7	8.0	15.1
Acq. time (t_1) [ms]	13.7	10.8	10.8	6.5	5.6	8.4	8.4	6.6
t_2 increments	3072	2304	2304	120	85	108	100	200
sw (t_2) [kHz]	93.8	100	100	8.0	8.0	6.0	5.7	15.1
Acq. time (t_2) [ms]	16.4	11.5	11.5	7.5	5.3	9.0	8.8	6.6
t_3 increments				2304	2304	2304	2304	2816
sw (t_3) [kHz]				100	100	100	100	93.75

Acq. time (t_3) [ms]				11.5	11.5	11.5	11.5	15.0
^{1}H decoupling	SPINAL64	SPINAL64	SPINAL64	SPINAL64	SPINAL64	SPINAL64	SPINAL64	SPINAL64
Field [kHz]	90	90	90	90	90	90	90	90
Interscan delay d1 [s]	3	2.6	2.6	3	2.9	3	3	2.75
Number of scans	16	24	24	16	16	16	8	4
Measurement time	34 h	23 h	23 h	5 days	3 days	6 days	4 days	5.2 days

Table 1. NMR experimental details. sw stands for spectral width. Experiments were recorded on an 800 MHz spectrometer at a 17.5 kHz MAS frequency and at a sample temperature of 5 °C. The ^{15}N-insert was removed from the probe for the CCC experiment to increase the signal to noise ratio.

➢ References

Abramov G, Morag O, Goldbourt A (2015) Magic-angle spinning NMR of intact bacteriophages: Insights into the capsid, DNA and their interface. Journal of Magnetic Resonance 253:80–90. doi: 10.1016/j.jmr.2015.01.011

Andreas LB, Jaudzems K, Stanek J, et al (2016) Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. Proc Natl Acad Sci USA 113:9187–9192. doi: 10.1073/pnas.1602248113

Birnbaum F, Nassal M (1990) Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. Journal of Virology 64:3319–3330.

Bourne CR, Finn MG, Zlotnick A (2006) Global Structural Changes in Hepatitis B Virus Capsids Induced by the Assembly Effector HAP1. Journal of Virology 80:11055–11061. doi: 10.1128/JVI.00933-06

Böckmann A, Gardiennet C, Verel R, et al (2009) Characterization of different water pools in solid-state NMR protein samples. J Biomol NMR 45:319–327. doi: 10.1007/s10858-009-9374-3

Böttcher B, Wynne SA, Crowther RA (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature 386:88–91. doi: 10.1038/386088a0

Chen C, Wang JC-Y, Zlotnick A (2011) A Kinase Chaperones Hepatitis B Virus Capsid Assembly and Captures Capsid Dynamics in vitro. PLoS Pathog 7:e1002388–10. doi: 10.1371/journal.ppat.1002388

Conway JF, Cheng N, Zlotnick A, et al (1997) Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. Nature 386:91–94. doi: 10.1038/386091a0

Crowther RA, Kiselev NA, Böttcher B, et al (1994) Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell 77:943–950.

Dryden KA, Wieland SF, Whitten-Bauer C, et al (2006) Native Hepatitis B Virions and Capsids Visualized by Electron Cryomicroscopy. Molecular Cell 22:843–850. doi: 10.1016/j.molcel.2006.04.025

Durantel D, Zoulim F (2016) New antiviral targets for innovative treatment concepts for hepatitis B virus and hepatitis delta virus. Journal of Hepatology 64:S117–S131. doi: 10.1016/j.jhep.2016.02.016

Freund SMV, Johnson CM, Jaulent AM, 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 384:1301–1313. doi: 10.1016/j.jmb.2008.10.020

Gallina A, Bonelli F, Zentilin L, et al (1989) A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. Journal of Virology 63:4645–4652.

Habenstein B, Wasmer C, Bousset L, et al (2011) Extensive de novo solid-state NMR assignments of the 33 kDa C-terminal domain of the Ure2 prion. J Biomol NMR 51:235–243. doi: 10.1007/s10858-011-9530-4

Han Y, Ahn J, Concel J, et al (2010) Solid-State NMR Studies of HIV-1 Capsid Protein Assemblies. J Am Chem Soc 132:1976–1987. doi: 10.1021/ja908687k

Heger-Stevic J, Kolb P, Walker A, Nassal M (2017) Displaying whole-chain proteins on hepatitis B virus capsid-like particles. Virus-Derived Nanoparticles for Advanced Technologies, in press

Katen SP, Tan Z, Chirapu SR, et al (2013) Assembly-Directed Antivirals Differentially Bind Quasiequivalent Pockets to Modify Hepatitis B Virus Capsid Tertiary and Quaternary Structure. Structure/Folding and Design 21:1406–1416. doi: 10.1016/j.str.2013.06.013

Morag O, Abramov G, Goldbourt A (2014) Complete Chemical Shift Assignment of the ssDNA in the Filamentous Bacteriophage fd Reports on Its Conformation and on Its Interface with the Capsid Shell. J Am Chem Soc 136:2292–2301. doi: 10.1021/ja412178n

Nassal M (2015) HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. Gut 64:1972. doi: 10.1136/gutjnl-2015-309809

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 66:4107–4116.

Ning X, Basagoudanavar SH, Liu K, et al (2017) Capsid Phosphorylation State and Hepadnavirus Virion Secretion. Journal of Virology 91:e00092–17–16. doi: 10.1128/JVI.00092-17

Patel N, White SJ, Thompson RF, et al (2017) HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. Nature Microbiology 1–10. doi: 10.1038/nmicrobiol.2017.98

Pei Y, Wang C, Yan SF, Liu G (2017) Past, Current, and Future Developments of Therapeutic Agents for Treatment of Chronic Hepatitis B Virus Infection. J Med Chem acs.jmedchem.6b01442–19. doi: 10.1021/acs.jmedchem.6b01442

Penzel S, Smith AA, Agarwal V, et al (2015) Protein resonance assignment at MAS frequencies approaching 100 kHz: a quantitative comparison of J-coupling and dipolar- coupling-based transfer methods. J Biomol NMR 1–22. doi: 10.1007/s10858-015-9975-y

Quinn CM, Lu M, Suiter CL, et al (2015) Magic angle spinning NMR of viruses. Progress in Nuclear Magnetic Resonance Spectroscopy 86-87:21–40. doi: 10.1016/j.pnmrs.2015.02.003

Revill P, Testoni B, Locarnini S, Zoulim F (2016) Global strategies are required to cure and eliminate HBV infection. Nature Reviews Gastroenterology And Hepatology 13:239–248. doi: 10.1038/nrgastro.2016.7

Schuetz A, Wasmer C, Habenstein B, et al (2010) Protocols for the Sequential Solid-State NMR Spectroscopic Assignment of a Uniformly Labeled 25 kDa Protein: HET-s(1-227). Chem Eur J of Chem Bio 11:1543–1551. doi: 10.1002/cbic.201000124

Selzer L, Kant R, Wang JCY, et al (2015) Hepatitis B Virus Core Protein Phosphorylation Sites Affect Capsid Stability and Transient Exposure of the C-terminal Domain. J Biol Chem 290:28584–28593. doi: 10.1074/jbc.M115.678441

Stanek J, Andreas LB, Jaudzems K, et al (2016) NMR Spectroscopic Assignment of Backbone and Side-Chain Protons in Fully Protonated Proteins: Microcrystals, Sedimented Assemblies, and Amyloid Fibrils. Angew Chem 128:15730–15735. doi: 10.1002/ange.201607084

Stevens TJ, Fogh RH, Boucher W, et al (2011) A software framework for analysing solid-state MAS NMR data. J Biomol NMR 51:437–447. doi: 10.1007/s10858-011-9569-2

Suiter CL, Quinn CM, Lu M, et al (2015) MAS NMR of HIV-1 protein assemblies. Journal of Magnetic Resonance 253:10–22. doi: 10.1016/j.jmr.2014.12.009

Vranken WF, Boucher W, Stevens TJ, et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696. doi: 10.1002/prot.20449

Wishart DS, Bigam CG, Holm A, et al (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J Biomol NMR 5:67–81.

World Health Organization (2017) Global hepatitis report. In: www.who.int. Apr 2017

Wynne SA, Crowther RA, Leslie AG (1999) The crystal structure of the human hepatitis B virus capsid. Molecular Cell 3:771–780.

Yu X, Jin L, Jih J, et al (2013) 3.5 Å cryo-EM Structure of Hepatitis B Virus Core Assembled from Full-Length Core Protein. PLoS ONE 8:e69729–11. doi: 10.1371/journal.pone.0069729

Zlotnick A, Venkatakrishnan B, Tan Z, et al (2015) Core protein: A pleiotropic keystone in the HBV lifecycle. Antiviral Research 121:82–93. doi: 10.1016/j.antiviral.2015.06.020

Zoulim F, Durantel D (2015) Antiviral Therapies and Prospects for a Cure of Chronic Hepatitis B. Cold Spring Harbor Perspectives in Medicine 5:a021501–a021501. doi: 10.1101/cshperspect.a021501

● Localizing conformational hinges by NMR

The work in this part was done with Lauriane Lecoq, in collaboration with Beat H. Meier, and has been published in following paper:

Localizing conformational hinges by NMR: where do HBV core proteins adapt for capsid assembly?

Lauriane Lecoq, Shishan Wang, Thomas Wiegand, Stéphane Bressanelli, Michael Nassal*, Beat H. Meier*, Anja Böckmann*

ChemPhysChem. 2018. 19, 1336–1340.

Contribution: Prepared samples for NMR characterization, contributed to writing the paper.

➢ Abstract

The hepatitis B virus (HBV) icosahedral nucleocapsid is assembled from 240 chemically identical core protein molecules and, structurally, comprises four groups of symmetrically nonequivalent subunits. We show here that this asymmetry is reflected in solid-state NMR spectra of the capsids in which peak splitting is observed for a subset of residues. We compare this information to dihedral angle variations from available 3D structures, and also to computational predictions of dynamic domains and molecular hinges. We find that while, at the given resolution, dihedral angles variations directly obtained from the X-ray structures are not precise enough to be interpreted, the chemical-shift information from NMR correlates, and interestingly goes beyond, information from bioinformatics approaches. Our study reveals the high sensitivity with which NMR can detect the residues allowing the subtle conformational adaptations needed in lattice formation. Our findings are important for understanding the formation and modulation of protein assemblies in general.

Keywords: Hepatitis B virus • Nucleocapsid • Core protein • Solid-state NMR • Asymmetric unit

➢ Introduction

When proteins assemble into larger superstructures, the constituting molecules are in many cases no longer symmetry equivalent, as they need to conform to a minimum energy configuration. This is observed in protein crystals, but also in protein fibrils or viral capsids and envelopes. In crystals, a subunit from which the entire structure can be generated by symmetry operations is called the asymmetric unit, and this concept is also used in other symmetric assemblies like icosahedra with a triangulation number T>1, where the monomers need to distort in nonequivalent classes of conformations upon assembly.(Caspar and Klug, 1962) These distortions can either be distributed over the protein chain, or be localized at a few key residues, i.e. hinges, interconnected

by predominantly rigid bodies. The identification of hinge residues is important in a variety of areas, and computational software, as for example DynDom,(Taylor et al., 2014) has been developed to determine dynamic domains, hinge axes and hinge bending residues from two protein conformers, and was successfully applied to compare structures with a resolution higher than 3 Å.(Koike et al., 2017)

The HBV particle is composed of an outer envelope and an inner nucleocapsid formed by a self-assembling core protein (Cp) of 20 kDa. The first ~140 amino acids of Cp are sufficient for capsid assembly.(Birnbaum and Nassal, 1990; Gallina et al., 1989) First cryo-electron microscopy (cryo-EM),(Böttcher et al., 1997; Conway et al., 1997) and later on crystallography,(Wynne et al., 1999) revealed the structure of Cp149, as a mainly α-helical protein which dimerizes via formation of a four-helix bundle. In the assembled capsid, these bundles appear as perpendicular dimeric spikes punctuating the capsid shell that is formed by lateral dimer-dimer contacts. The T=4 icosahedral symmetry accommodates four distinct (A, B, C and D) environments with 60 molecules each. Two types of dimers, 60 of type AB and 60 of type CD,(Zlotnick et al., 1996) assemble in 12 fivefold vertices composed solely by subunit A, and 30 quasi-sixfold vertices composed by subunits B, C and D. Because the HBV capsid follows classical quasi-equivalence,(Caspar and Klug, 1962) the contacts between dimers around the five- and quasi-sixfold vertices are very similar and involve the same residues. More recently, cryo-EM revealed structural details of HBV capsids down to 3.5 Å resolution, resulting in a similar structure to X-ray for the four molecules.(Yu et al., 2013) EM studies of the capsid have further shown that the local structural modifications caused by the insertion of additional amino acids leads to global structural changes where hinge residues play an important role.(Böttcher et al., 2006) Hinge-mediated motions may also enable the dynamic structural changes suspected to accompany viral genome maturation during HBV replication.(Nassal, 2008)

Virus capsids can also be studied by solid-state NMR.(Abramov et al., 2015; Andreas et al., 2016; Quinn et al., 2015) NMR chemical shifts, and especially those of carbon-13, are highly sensitive to backbone dihedral angles, and can detect conformational differences by chemical-shift perturbations, typically when an interactant is added, or in protein complexes (for a review see (Williamson, 2013)). When different

conformations (or more general crystallographically inequivalent species) are present in a single entity, this results in the observation of peak splitting in the spectra.(Andreas et al., 2010; 2015; Böckmann et al., 2003; Frericks Schmidt et al., 2007; Gath et al., 2014; Ma et al., 2015) We show here, at the example of the HBV capsid, that peak splitting is indeed observed in the spectra for a specific subset of residues, and that these sites correlate with molecular hinges, but also the dynamic domains, identified by the bioinformatics approach in DynDom.

➢ Results and disscussion

We here investigated *E. coli* self-assembled Cp149 HBV capsids forming homogeneous T=4 icosahedra with a typical diameter of 30 ± 3 nm (Figure 1a). The ^{13}C and ^{15}N resonances of the Cp149 protein were assigned using 3D spectroscopy,(Lecoq et al., 2018) and the annotated 2D ^{13}C-^{13}C Dipolar Assisted Rotational Resonance(Takegoshi et al., 2001) (DARR) spectrum of the capsids is displayed in Figure 1b. Sequential assignments reveal that a subset of residues experiences resonance splitting, with up to 2.5 ppm of chemical-shift differences amongst themselves in the ^{13}C and/or ^{15}N dimensions. Indeed, peak splitting is clearly distinguishable for residues A11, T12, L16-D22, T33-S35, L108-F110, S121-W125, R127, T128, P130-Y132 and N136-P138 (Figure S1). Chemical shifts of all forms are given in Table S1.

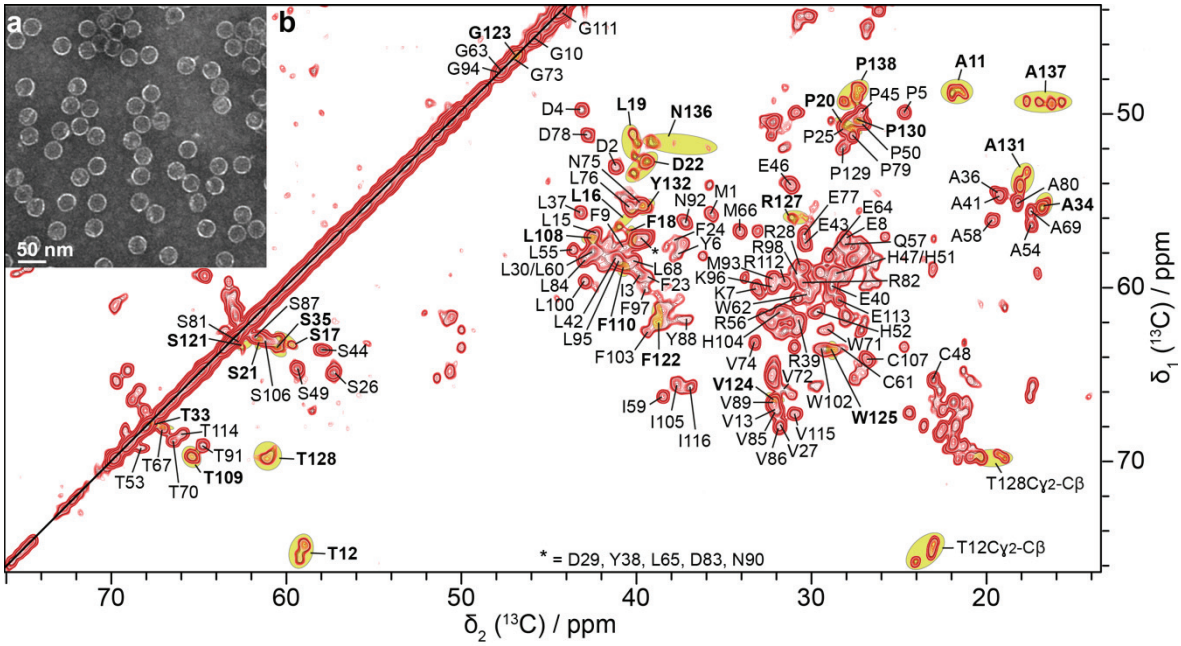


Figure 1. a) Negative-stain EM micrograph of Cp149 capsids. **b)** 2D ^{13}C-^{13}C DARR solid-state NMR spectrum of Cp149 capsids recorded with a mixing time of 20 ms. All assigned Cα-Cβ intra-residue correlations are labeled with the residue number except for proline residues where the Cγ-Cδ peaks are labeled. Residues with peak splitting are indicated in bold letters and corresponding resonances are highlighted in yellow. The spectrum was processed using a squared cosine apodization (SSB 3).

Where four distinct signals are observed (as for example A137 shown in Figure 2a), they can be assigned to the four crystallographic distinct molecules in the asymmetric unit represented in Figure 2b. When two sets of peaks are observed, it suggests that at least one monomer is significantly distinct from the others. In total, detectable peak splitting was observed for 28 residues, representing almost 20 % of the protein.

Nearly all residues showing peak splitting are located close to the fivefold (Figure 2c) and quasi-sixfold (Figure 2d) vertices, while none of the residues within the spikes of the capsids is concerned. About half of them are located in loops, where the backbone is less constrained by hydrogen bonds than in regular secondary structure elements, and the remaining are mainly located towards the beginning or end of the base plate helices (Figure S2). For A137, the four Cα-Cβ correlation peaks can be easily distinguished in the 2D spectra (Figure 2a). An analysis of the multiple peaks from a same residue in the 2D DARR spectrum (Figure S3) and 3D spectra (Figure S4)

reveals similar intensities, which suggests that the peaks indeed correspond respectively to the 60 A, B, C and D subunits per capsid. The same behavior is observed for the other residues displaying peak quadrupling, namely T12, S35 and T128. On the contrary, when only two or three peaks are observed in the NMR spectra, one peak is usually more intense, representing 120 or 180 subunits with a rather similar environment.

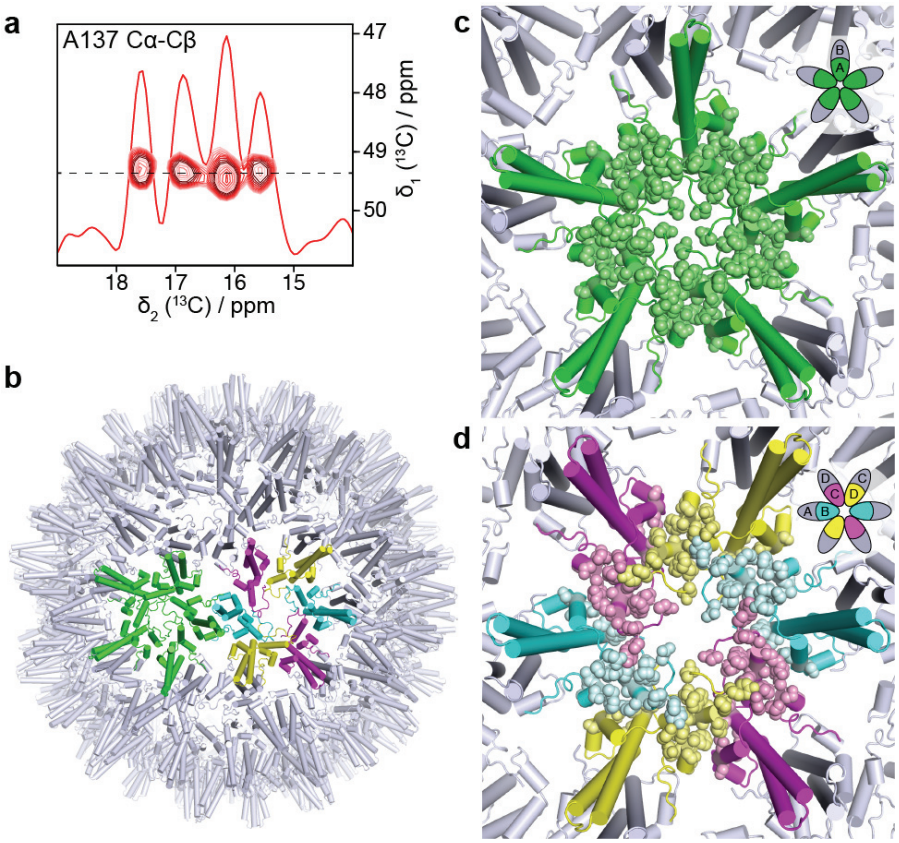


Figure 2. a) Extract from a 2D DARR spectrum, as well as a 1D trace, displaying the Cα-Cβ correlations of Ala137 showing a peak quadrupling, corresponding to the four molecules in the unit cell. **b)** Capsid structure (PDB: 1QGT(Wynne et al., 1999)), with localization of the monomers on the five- and quasi-sixfold vertices. **c, d)** Sphere representation of the side chains of the 28 residues showing NMR peak splitting, located near the fivefold **(c)** and quasi-sixfold **(d)** interfaces, shown on the capsid structure. On the top right are given the schematic representations of the subunit arrangement in the T=4 capsid, described in(Wynne et al., 1999). A, B, C and D subunits are shown in green, cyan, magenta and yellow, respectively. The A subunits form pentamers while B, C and D subunits form the hexamers.

The nonequivalence of the four subunits is also visible in the X-ray structure,(Wynne et al., 1999) as seen in Figure 3a where the four monomers are aligned using PyMOL.(The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC, 2015) Some residues display significant differences between the four monomers such as Y132, N136 or A137. As a consequence, the intermolecular contacts in the five- and quasi-sixfold vertices are indeed different, as shown in Figures 3b and c at the example of A137. Naively, one would assume that the deviation between the dihedral angles of the molecules in the unit cell would correlate with the residues showing peak splitting. An analysis of the local variations in backbone torsion angles φ and ψ between the four monomers in the X-ray structure (as standard deviations), which are shown in black in Figure 3d, however reveals that this is not the case. Instead, the variations are rather uniform along the sequence, with few N- and C-terminal residues, plus some residues in the spike loop displaying higher variations. It has however already been realized that the sum of torsion-angle rotations needed to superimpose two structures to the experimental accuracy can be 10 times larger than when just adding φ and ψ differences along the protein backbone,(Chapman et al., 2015) even for structures showing higher precision than Cp149 (1.7 vs 3.5 Å). Using computer programs which localize differences between two given (X-ray) structures(Chapman et al., 2015; Echols et al., 2003; Hayward and Berendsen, 1998), one obtains for example for two nominally identical structures of dihydrofolate reductase at 2.2 Å (PDB codes 1RG7(Dunbar et al., 1997) and 1DDS(Sawaya and Kraut, 1997)) a difference in dihedral angles of 9°±11°.(Chapman et al., 2015) Even between the dihedral angles obtained from two high resolution and nominally identical structures, significant dihedral angle variations (4.4°) are found (Figure S5).

Backbone torsion angles can also be derived from the NMR secondary chemical shifts. Using a databank approach, the program TALOS+(Shen et al., 2009) delivers φ and ψ from the chemical shifts of the residue to be investigated, plus for one neighbor on each side. The standard deviation between the dihedral angles predicted from the split resonances (Table S1) is shown in red in Figure 3d. Evidently, the torsion-angle variations differ between the two methods: while NMR shows an average deviation of 1.5° and only 38 residues involved in changes, the X-ray analysis lacks the clear pattern observed by NMR, with an average deviation of 14°, and no link between the

two dihedral-angle variations is apparent. Given the observations described above, this is no great surprise and the variations in the X-ray structure are probably background differences.(Chapman et al., 2015) (For similar results from the comparison between the X-ray and EM capsid structures see Figure S6 and S7).

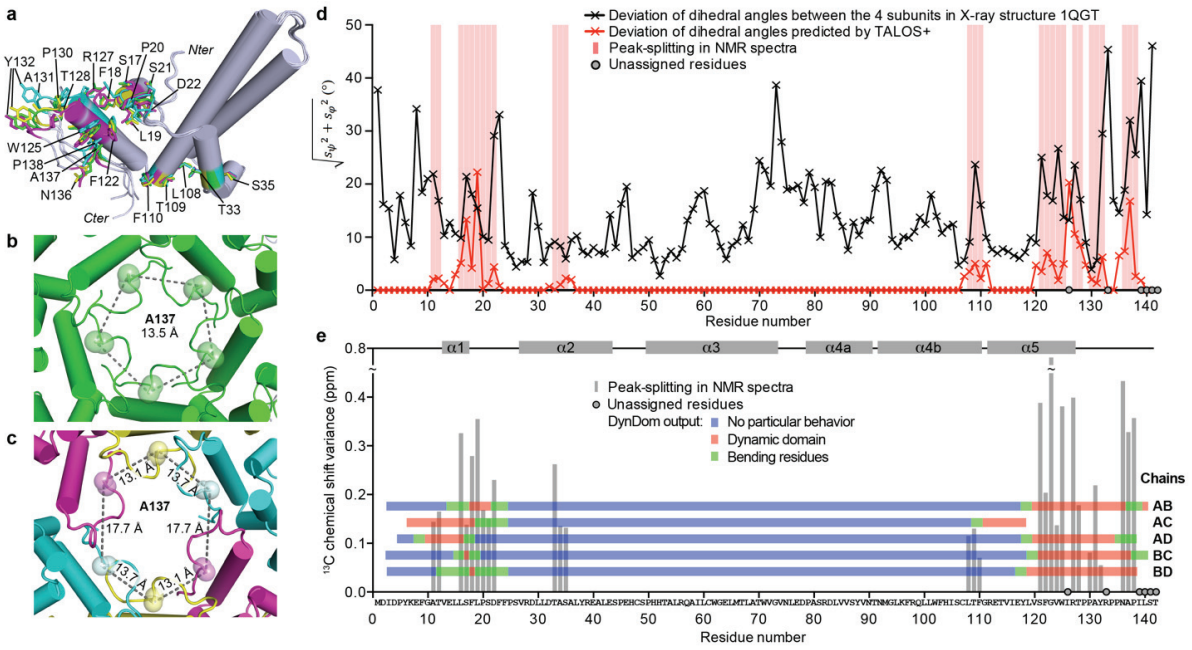


Figure 3. **a)** Overlay of the four capsid subunits to visualize on one hand the differences in the X-ray structures between the molecules in the asymmetric unit (PDB: 1QGT(Wynne et al., 1999)), and on the other hand the residues for which peak splitting is detected by solid-state NMR (shown as sticks in the color of the corresponding subunit: green, cyan, magenta and yellow for A, B, C and D subunit respectively). **b)** Zoom into the five-fold vertices with equivalent distances shown for the Cβs of A137 between the five A subunits. **c)** Zoom into the quasi-sixfold vertices with non-equivalent distances shown for the Cβs of A137 between each B, C and D subunit. **d)** Plots of the rms of the standard deviations between the backbone dihedral angles of the four subunits of the X-ray structure of Cp149 (PDB 1QGT(Wynne et al., 1999), black line), and the corresponding value predicted from the chemical shifts of the different forms observed by NMR using TALOS+(Shen et al., 2009) (red line). Residues displaying peak splitting in the NMR spectra are highlighted by pink bars. Unassigned residues are marked by grey circles. **e)** ^{13}C chemical shift variance of the different forms observed by NMR (grey bars) and identification of residues involved in interdomain motions using the DynDom software(Taylor et al., 2014) (horizontal bars). The “dynamic” domains (in red) and the bending residues (in green) are determined by comparing pairwise the A, B, C and D chains of the X-ray structure (PDB 1QGT(Wynne et al., 1999)).

Confronting the NMR data to predictions from the program DynDom (Taylor et al., 2014) designed to identify hinges from the comparison of two X-ray structures was more successful. DynDom defines “dynamic” domains (with dynamics referring there to differences between two static structures, and not actual flexibility), as well as residues involved in interdomain bending (http://fizz.cmp.uea.ac.uk/dyndom/dyndomMain.do). Applied to the capsid monomers in a pairwise analysis using the DynDom webserver (Figure 3e), “dynamic” domains (in red) and bending residues (in green) were identified for all pairs (with exception of the C and D monomers which do not reveal any “dynamic” domains or hinges). The individual comparisons yield similar regions involved in interdomain bending, one near the N-terminal, and another in the C-terminal domain. The overlap of the pairwise analyses is, with the exception of AC, rather consistent. DynDom classifies the residues in three classes: rigid domains, hinges and, in addition, the “dynamic” domains which are not necessarily rigid, but show an intradomain variation smaller than the interdomain motion. For the C-terminal, DynDom clearly differentiates between “dynamic” domains (residues ~121-135) and bending residues (~117-120), while in the N-terminal region this distinction is less clear-cut, and globally comprises residues 8-24. Residues T109 and F110 were only detected as bending residues between subunits A and C. The differences observed between the pairwise comparisons of monomers A-D illustrate the limit of the approach, but overall, good agreement is found. It is interesting that both the “dynamic” domain and hinge residues identified using DynDom strongly correlate with the residues for which NMR peak splitting is observed (observed chemical shift variance shown as grey bars in Figure 3e), which reveals that actually the “dynamic” domains show a variability comparable to the hinges as response to constraints from capsid assembly. In particular, splitting in the C-terminal region for residues S121 to Y132, including helix α5, coincide with the “dynamic” domain from the DynDom analysis. The large chemical-shift variations observed for residues 136-138 fall into the range of the mechanical hinge predicted by DynDom. Interestingly, no peak splitting is observed for the predicted hinge residues 118-120, but only from residue 121 onwards. These residues are part of helix α5 spanning residues R112-R127, and even if the hinge seems shifted compared to the DynDom prediction, it still comprises parts of the helix. Residues A11, T12 and L16-D22 displaying peak splitting coincide with the N-terminal region identified as “dynamic”

hinges by DynDom, which however shows a rather large distribution over the different pairs, including regions not highlighted by NMR. T109 and F110, exclusively detected as bending residues between A and C monomers by DynDom, display, together with L108, peak splitting in the NMR spectra. Interestingly, residue G111 in this same region has been described as hinge in an early EM study.(Böttcher et al., 2006) Residues T33 to S35 are only identified by NMR peak splittings, and are located in helix α2 which is facing helices α1 and α5, part of the N-terminal and C-terminal hinge regions respectively. Overall, these results show that the peak splitting observed in the NMR spectra correlates to a large extent with both the hinges *and* the "dynamic" domains identified by the DynDom method.

It has been suggested early on that the accurate recognition of sites involved in interdomain motions, here linked to the creation of adequate interfaces on mutimerization, may identify possible target sites for drugs.(Hayward and Berendsen, 1998) And indeed, in the available structures of the core protein in complex with inhibitors,(Klumpp et al., 2015; Zhou et al., 2017) T33, T109, F110, F122, V124, W125, T128 and Y132 residues are in close contact with small molecule assembly modulators (Figure S8), which are currently intensely investigated as potential new anti-HBV drugs.(Mak et al., 2017) Interference with the possibility to create adequate conformers via hinge motion, or inversely the stabilization of certain conformations promoting assembly, is likely involved in assembly modulation by these drugs. Alternatively, the hinges might play a role in accommodating assembly modulators at critical locations at protein interfaces.

➢ Conclusion

In summary, we have demonstrated that structural differences between quasi-equivalent HBV core proteins occur in strategic positions near the interfaces between the dimers, and no distortions are found in the helices that form the spikes of the capsid. This immediate information from NMR correlates well with information on molecular hinges extracted via bioinformatics methods, here DynDom, from high-resolution 3D structures but is more explicit because it is a direct residue-specific observation not

requesting extensive postprocessing of the primary data. The site-selective information from NMR is of particular importance to further analyze in detail the different conformations which the HBV capsid is thought to access during genome and particle maturation,(Nassal, 2008) as well as its interactions with assembly modulators.

➢ Acknowledgements

This work was supported by the Swiss National Science Foundation (Grant 200020_159707), the French National Research Agency (ANR) (ANR-14-CE09-0024B), by the LABEX ECOFECT (ANR-11-LABX-0048) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the ANR, the DFG grant NA154/9-4 and the Federal Ministry for Economic Affairs and Energy grant ZIM ZF4010301 CR5 (MN), the TGIR-RMN-THC Fr3050 CNRS, the Chinese Scientific Council (SW) and the European Marie Skłodowska-Curie program (H2020-MSCA-IF-2016 748516). BM acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 741863, FASTER). We thank the Centre d'Imagerie Quantitative Lyon-Est (CIQLE) for access to the EM platform, and Bettina Böttcher for interesting and helpful discussions.

➢ References

Abramov, G., Morag, O., Goldbourt, A., 2015. Magic-angle spinning NMR of intact bacteriophages: Insights into the capsid, DNA and their interface. J. Magn. Reson. 253, 80–90. doi:10.1016/j.jmr.2015.01.011

Andreas, L.B., Eddy, M.T., Pielak, R.M., Chou, J., Griffin, R.G., 2010. Magic Angle Spinning NMR Investigation of Influenza A M2 18–60: Support for an Allosteric Mechanism of Inhibition. J. Am. Chem. Soc. 132, 10958–10960. doi:10.1021/ja101537p

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. Proc. Natl. Acad. Sci. U.S.A. 113, 9187–9192. doi:10.1073/pnas.1602248113

Andreas, L.B., Reese, M., Eddy, M.T., Gelev, V., Ni, Q.Z., Miller, E.A., Emsley, L., Pintacuda, G., Chou, J.J., Griffin, R.G., 2015. Structure and Mechanism of the Influenza A M2 18–60Dimer of Dimers. J. Am. Chem. Soc. 137, 14877–14886. doi:10.1021/jacs.5b04802

Birnbaum, F., Nassal, M., 1990. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. J. Virol. 64, 3319–3330.

Böckmann, A., Lange, A., Galinier, A., Luca, S., Giraud, N., Juy, M., Heise, H., Montserret, R., Penin, F., Baldus, M., 2003. Solid state NMR sequential resonance assignments and conformational analysis of the 2x10.4 kDa dimeric form of the Bacillus subtilis protein Crh. J. Biomol. NMR 27, 323–339.

Böttcher, B., Vogel, M., Ploss, M., Nassal, M., 2006. High Plasticity of the Hepatitis B Virus Capsid Revealed by Conformational Stress. J. Mol. Biol. 356, 812–822. doi:10.1016/j.jmb.2005.11.053

Böttcher, B., Wynne, S.A., Crowther, R.A., 1997. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature 386, 88–91. doi:10.1038/386088a0

Caspar, D.L.D., Klug, A., 1962. Physical Principles in the Construction of Regular Viruses. Cold Spring Harb. Sym. 27, 1–24. doi:10.1101/SQB.1962.027.001.005

Chapman, B.K., Davulcu, O., Skalicky, J.J., Brüschweiler, R.P., Chapman, M.S., 2015. Parsimony in Protein Conformational Change. Structure 23, 1190–1198. doi:10.1016/j.str.2015.05.011

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. doi:10.1038/386091a0

Dunbar, J., Yennawar, H.P., Banerjee, S., Luo, J., Farber, G.K., 1997. The effect of denaturants on protein structure. Protein Sci. 6, 1727–1733. doi:10.1002/pro.5560060813

Echols, N., Milburn, D., Gerstein, M., 2003. MolMovDB: analysis and visualization of conformational change and structural flexibility. Nucleic Acids Res. 31, 478–482. doi:10.1093/nar/gkg104

Frericks Schmidt, H.L., Sperling, L.J., Gao, Y.G., Wylie, B.J., Boettcher, J.M., Wilson, S.R., Rienstra, C.M., 2007. Crystal Polymorphism of Protein GB1 Examined by Solid-State NMR Spectroscopy and X-ray Diffraction. J. Phys. Chem. B 111, 14362–14369. doi:10.1021/jp075531p

Gallina, A., Bonelli, F., Zentilin, L., Rindi, G., Muttini, M., Milanesi, G., 1989. A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. J. Virol. 63, 4645–4652.

Gath, J., Bousset, L., Habenstein, B., Melki, R., Böckmann, A., Meier, B.H., 2014. Unlike Twins: An NMR Comparison of Two α-Synuclein Polymorphs Featuring Different Toxicity. PLoS ONE 9, e90659–11. doi:10.1371/journal.pone.0090659

Hayward, S., Berendsen, H.J., 1998. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins 30, 144–154.

Klumpp, K., Lam, A.M., Lukacs, C., Vogel, R., Ren, S., Espiritu, C., Baydo, R., Atkins, K., Abendroth, J., Liao, G., Efimov, A., Hartman, G., Flores, O.A., 2015. High-resolution crystal structure of a hepatitis B virus replication inhibitor bound to the viral core protein. Proc. Natl. Acad. Sci. U.S.A. 112, 15196–15201. doi:10.1073/pnas.1513803112

Koike, R., Amemiya, T., Horii, T., Ota, M., 2017. Structural changes of homodimers in the PDB. J. Struct. Biol. doi:10.1016/j.jsb.2017.12.004

Lecoq, L., Wang, S., Wiegand, T., Bressanelli, S., Nassal, M., Meier, B.H., Böckmann, A., 2018. Solid-state [13C-15N] NMR resonance assignment of hepatitis B virus core protein. Biomol. NMR Assign. 12, 205–214. doi:10.1007/s12104-018-9810-y

Ma, P., Xue, Y., Coquelle, N., Haller, J.D., Yuwen, T., Ayala, I., Mikhailovskii, O., Willbold, D., Colletier, J.-P., Skrynnikov, N.R., Schanda, P., 2015. Observing the overall rocking motion of a protein in a crystal. Nat. Commun. 6, 8361. doi:10.1038/ncomms9361

Mak, L.-Y., Wong, D.K.-H., Seto, W.-K., Lai, C.-L., Yuen, M.-F., 2017. Hepatitis B core protein as a therapeutic target. Expert Opin. Ther. Tar. doi:10.1080/14728222.2017.1397134

Nassal, M., 2008. Hepatitis B viruses: Reverse transcription a different way. Virus Res. 134, 235–249. doi:10.1016/j.virusres.2007.12.024

Quinn, C.M., Lu, M., Suiter, C.L., Hou, G., Zhang, H., Polenova, T., 2015. Magic angle spinning NMR of viruses. Prog. Nucl. Mag. Res. Sp. 86-87, 21–40. doi:10.1016/j.pnmrs.2015.02.003

Sawaya, M.R., Kraut, J., 1997. Loop and Subdomain Movements in the Mechanism of Escherichia coliDihydrofolate Reductase: Crystallographic Evidence †,‡. Biochemistry 36, 586–603. doi:10.1021/bi962337c

Shen, Y., Delaglio, F., Cornilescu, G., Bax, A., 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR 44, 213–223. doi:10.1007/s10858-009-9333-z

Takegoshi, K., Nakamura, S., Terao, T., 2001. 13C–1H dipolar-assisted rotational resonance in magic-angle spinning NMR. Chem. Phys. Lett. 344, 631–637. doi:10.1016/s0009-2614(01)00791-6

Taylor, D., Cawley, G., Hayward, S., 2014. Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. Bioinformatics 30, 3189–3196. doi:10.1093/bioinformatics/btu506

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC, 2015. The PyMOL Molecular Graphics System, Version 2.0.

Williamson, M.P., 2013. Using chemical shift perturbation to characterise ligand binding. Prog. Nucl. Mag. Res. Sp. 73, 1–16. doi:10.1016/j.pnmrs.2013.02.001

Wynne, S.A., Crowther, R.A., Leslie, A.G., 1999. The crystal structure of the human hepatitis B virus capsid. Mol. Cell 3, 771–780.

Yu, X., Jin, L., Jih, J., Shih, C., Hong Zhou, Z., 2013. 3.5 Å cryo-EM Structure of Hepatitis B Virus Core Assembled from Full-Length Core Protein. PLoS ONE 8, e69729–11. doi:10.1371/journal.pone.0069729

Zhou, Z., Hu, T., Zhou, X., Wildum, S., Garcia-Alcalde, F., Xu, Z., Wu, D., Mao, Y., Tian, X., Zhou, Y., Shen, F., Zhang, Z., Tang, G., Najera, I., Yang, G., Shen, H.C., Young, J.A.T., Qin, N., 2017. Heteroaryldihydropyrimidine (HAP) and Sulfamoylbenzamide (SBA) Inhibit Hepatitis B Virus Replication by Different Molecular Mechanisms. Sci. Rep. 7, 42374. doi:10.1038/srep42374

Zlotnick, A., Cheng, N., Conway, J.F., Booy, F.P., Steven, A.C., Stahl, S.J., Wingfield, P.T., 1996. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. Biochemistry 35, 7412–7421. doi:10.1021/bi9604800

➢ Supplementary Material

Experimental Section

Protein expression and purification. ^{13}C-^{15}N-labeled Cp149 was expressed in *E. coli* and purified as described previously.(Lecoq et al., 2018)

Electron microscopy. 5 µL of Cp149 capsids concentrated at 1 mg/ml were loaded on a carbon-coated grid and incubated for 2 minutes. The grid was then layered on top of 50 µL of 2 % phosphotungstic acid (m/v) pH 7 and incubated at room temperature for 2 minutes for negative staining. The grids were examined with a JEM-1400 transmission electron microscope operating at 100 kV.

Solid-state NMR spectroscopy. ^{13}C-^{15}N-labeled Cp149 capsids were sedimented into a thin-wall 3.2 mm zirconium rotor (180,000 *g*, 14 h) using a home-made filling tool.(Böckmann et al., 2009) The DARR spectrum was recorded on an 800 MHz Bruker Avance II spectrometer using a 3.2 mm triple-resonance probe, at 17.5 kHz spinning frequency and 4 °C according to the resonance frequency of the supernatant water(Böckmann et al., 2009). 17.5 kHz proton irradiation during 20 ms for the DARR transfer, and 90 kHz decoupling SPINAL-64(Fung et al., 2000) decoupling, were used. An interscan delay of 3 s and 16 scans resulted in a total acquisition time of 1d 8h. Resonance assignment of the split peaks was performed using the 3D spectra described previously.(Lecoq et al., 2018) Spectra analysis was done with the CcpNmr Analysis package.(Stevens et al., 2011; Vranken et al., 2005)

Dihedral angles deviation between the X-ray monomer structures. The dihedral angles of the four monomers in the X-ray structure of Cp149 (PDB 1QGT(Wynne et al., 1999)) were extracted using VADAR online software http://vadar.wishartlab.com/.(Willard et al., 2003) Then, the deviation of φ and ψ angles between the four monomers was calculated for each residue according to formula:

Deviation of dihedrals = $\sqrt{s_\psi{}^2 + s_\varphi{}^2}$ (°)

where $s_x = \sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2/(N-1)})$ with $\bar{x}$ being the mean value.

NMR chemical shift and TALOS dihedral angle deviations. The variance of ^{13}C-chemical shifts was calculated for residues showing peak splitting in the NMR spectra and divided by the number of assigned carbon atoms (minimum 3 for C', Cα and Cβ): $\sqrt{\sum_{i=1}^{N} s_i{}^2}/N$ where s_i is the standard deviation of ^{13}C chemical shifts between the different forms observed in the NMR spectra, and *N* is the number of ^{13}C chemical shifts available. The variance was calculated between two, three or four sets of chemical shifts depending on the number of assigned forms (see Table S1). Then, dihedral angles were calculated using TALOS+ based on the chemical shifts of each form. TALOS+ was used rather than TALOSN as it gives more weight to the actual chemical shift and less to the neighboring residues. The deviation of dihedral angles between the different forms was calculated using the same formula as for the X-ray structure analysis.

Protein domain motion analysis. The program DynDom(Hayward and Berendsen, 1998; Taylor et al., 2014) was used to pairwise compare subunits of the X-ray structure 1QGT.

```
         10          20          30          40          50          60
MDIDPYKEFG ATVELLSFLP SDFFPSVRDL LDTASALYRE ALESPEHCSP HHTALRQAIL

         70          80          90         100         110         120
CWGELMTLAT WVGVNLEDPA SRDLVVSYVN TNMGLKFRQL LWFHISCLTF GRETVIEYLV

        130         140
SFGVWIRTPP AYRPPNAPIL STLPETTVV
```

Peak splitting in NMR spectra: 2 forms, 3 forms, 4 forms

Figure S1. Amino-acid sequence of Cp149 showing residues with NMR peak doubling, tripling and quadrupling in blue, purple and red, respectively**.**

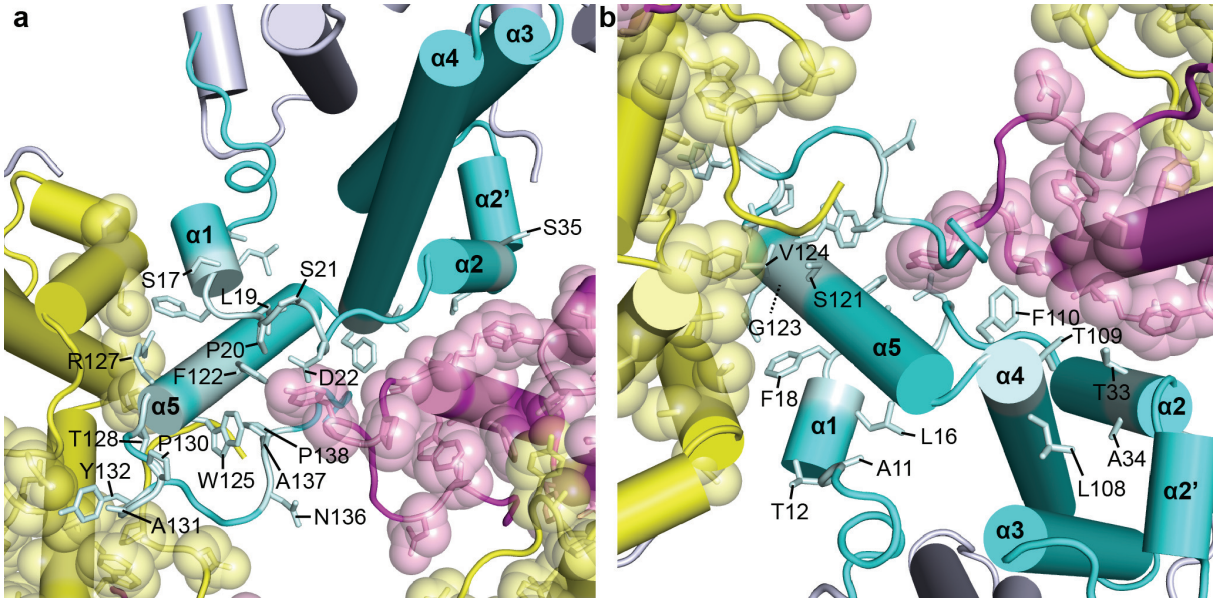


Figure S2. Zoom on the 28 residues displaying multiple peaks in the NMR spectra displayed as sticks on the capsid structure (PDB: 1QGT(Wynne et al., 1999)) in B subunit (colored in cyan) and as spheres and sticks in C and D subunits (colored in magenta and yellow, respectively). Residues are labeled on B subunit in **a)** view from the top of the capsid structure (16 residues) and **b)** view from the inside of the capsid structure (12 residues). Secondary structure elements are labeled in bold and helices are shown as cylindrical cartoon.

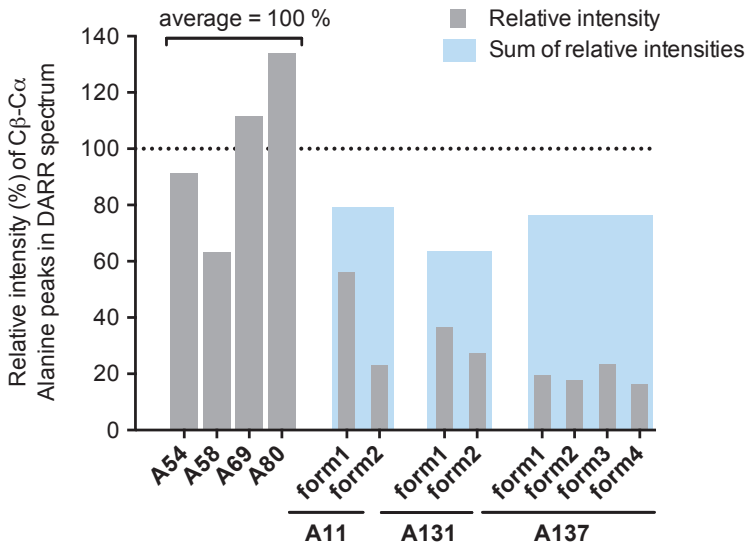


Figure S3. Intensity comparison of isolated Cβ-Cα correlation peaks in the DARR spectrum for alanine residues. The average intensity taken as a reference was calculated on the four single resonance and isolated alanine residues displayed in grey (A54, A58, A69 and A80). Relative intensities of alanine residues with several forms in the NMR spectrum (A11, A131 and A137) are displayed in grey and their sums are displayed in blue. A34, A38 and A41 are not shown here as A38 and A41 are overlapped and as the two forms of A34 are overlapped on DARR spectrum, which prevents from an accurate intensity value determination. The summed intensities of the split signals of A11, A131 and A137 are slightly weaker than the individual intensities of the alanines not showing peak splitting, which can be explained by an increased flexibility due to their location in loops.

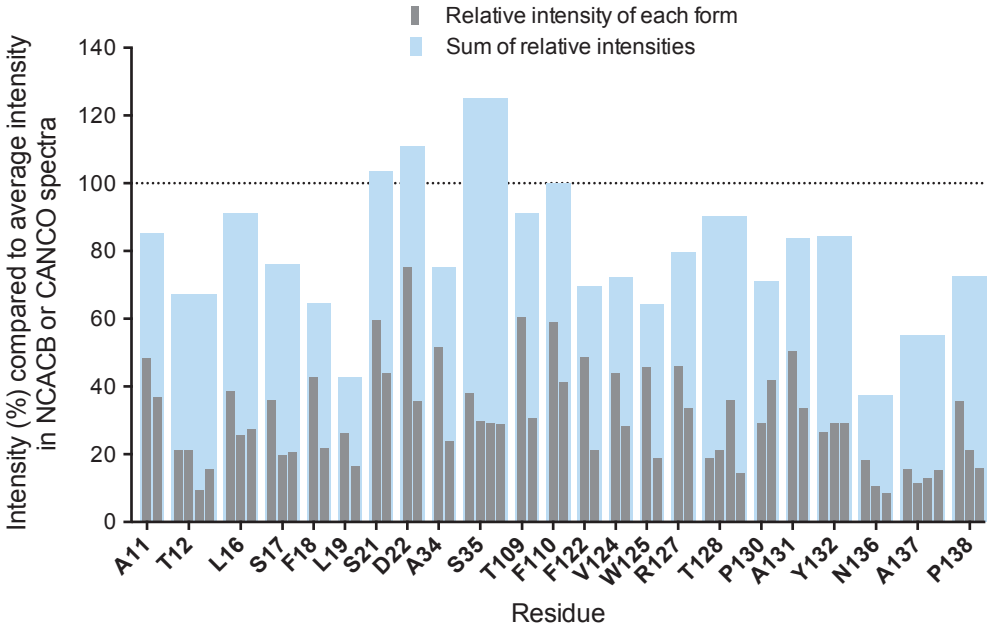


Figure S4. Relative peaks intensity in 3D NCACB or CANCO of residues with peak splitting compared to the average intensity of residues with a single resonance. In grey is the relative intensity of each form while in blue is the sum of the two, three or four peaks relative intensities. Only isolated peaks were

included, therefore P20, T33, L108, S121 and G123 are not present in the graph due to overlaps in all 3D spectra. NCACB peaks were used for most of residues, except for S21, A34, S35, V124, R127, A131 and Y132 where CANCO peaks were used as they were more isolated.

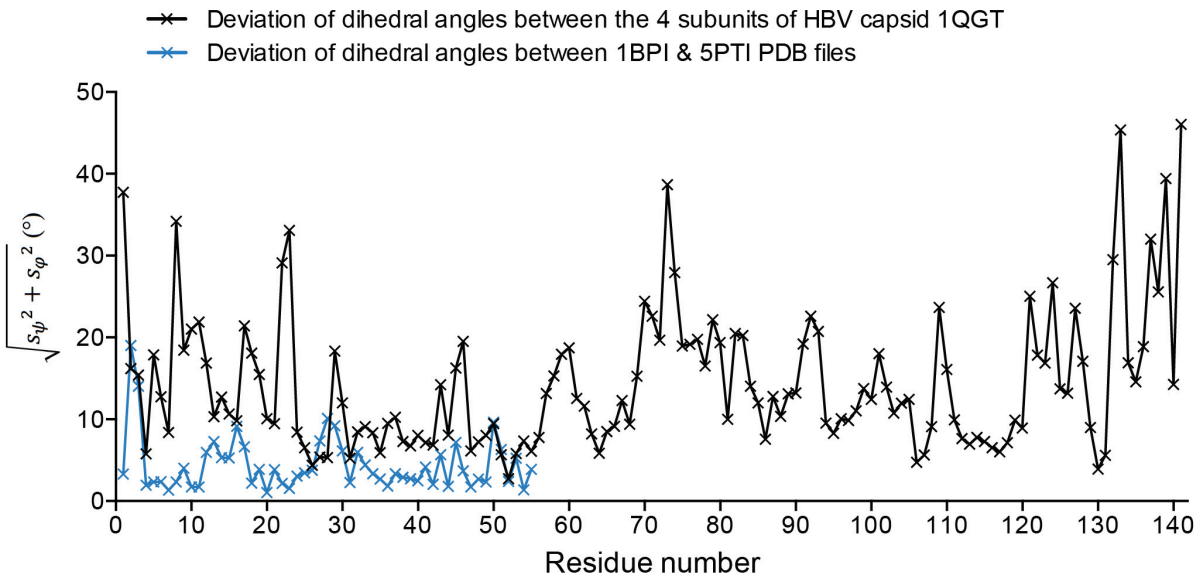


Figure S5. We wanted to estimate whether the variations in the dihedral angles between the different monomers in the X-ray structure pinpoint real differences or whether they are within the variations given by the limited resolution of 3.3 Å of the X-ray structure. We thus compared the dihedral angle variations observed between the four molecules in the asymmetric unit of HBV capsid (PDB: 1QGT(Wynne et al., 1999), black line) and the ones observed between two different high resolution structures of bovine pancreatic trypsin inhibitor (BPTI, blue line) at 1.0 Å (PDB: 5PTI(Wlodawer et al., 1984)) and 1.1 Å (PDB: 1BPI(Parkin et al., 1996)) to highlight the variations as a function of resolution. The three C-ter residues of BPTI structures are omitted for clarity. The variance $\sqrt{(s_\psi{}^2 + s_\varphi{}^2)}$ between the BPTI structures is 4.4°, versus 14.0° between the four subunits of the HBV capsid. Furthermore, for ubiquitin, an rmsd of 5.7° was found for the differences between φ angles between a 1.8 Å X-ray structure and NMR-determined high-precision values.(Wang and Bax, 1996) Thus, even very well defined X-ray structures show considerable variations of the dihedral angles. Therefore the variations observed in the HBV capsid could indeed be caused by the precision effects alone.

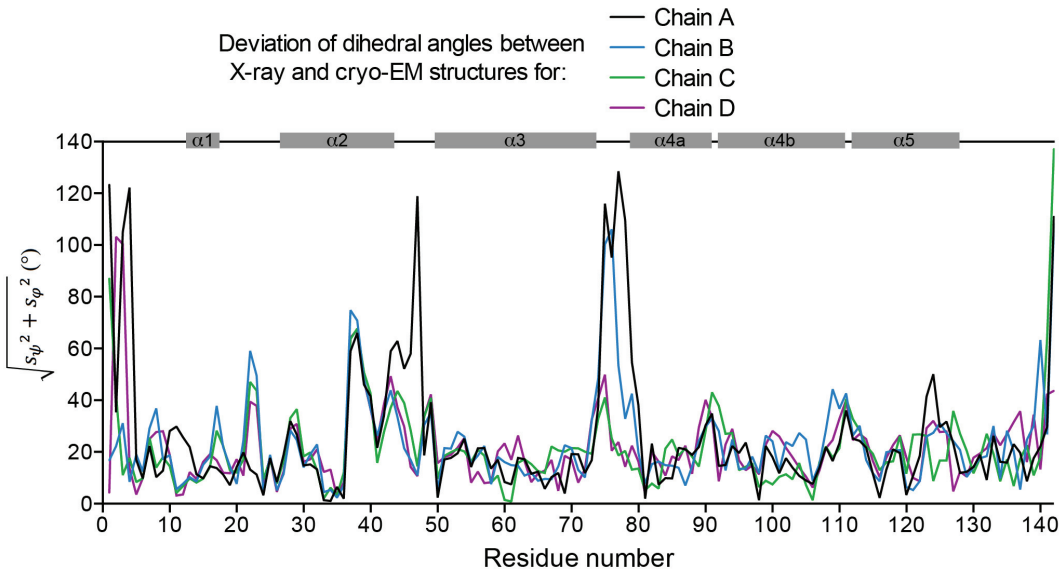


Figure S6. Comparison of the deviation of dihedral angles for the 4 chains of HBV core protein between the 3.3 Å X-ray structure (PDB: 1QGT(Wynne et al., 1999)) and the 3.5 Å cryo-EM structure (PDB: 3J2V(Yu et al., 2013)). Variations are significantly more important than when comparing two X-ray structures with a good resolution as in Figure S5, with an average difference of 25.3 ° for chain A, 22.8 ° for chain B, 20.3 ° for chain C and 21.7 ° for chain D for residues 1 to 140. This significant difference is partly due to mobility in the spike turn region comprising residues 75-80, and also to genuine differences in the C-terminal part.

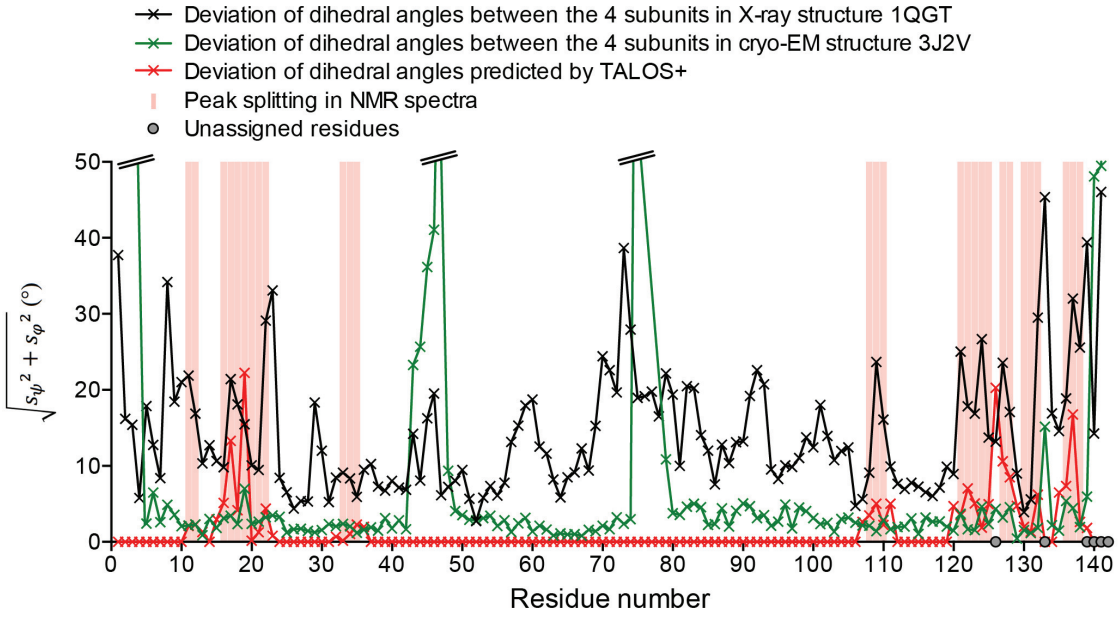


Figure S7. Comparison of the dihedral angle variations observed between the four molecules in the asymmetric unit of the 3.3 Å X-ray structure (PDB: 1QGT(Wynne et al., 1999) in black), the 3.5 Å cryo-EM structure (PDB: 3J2V(Yu et al., 2013), in green) and predicted from NMR chemical shifts (using

TALOS+, in red). The variations observed between the four monomers in the EM structure is for most residues surprisingly small with an average deviation of 2.9° if we exclude the four regions with exceptionally high deviations due to structural flexibility (the N and C-termini residues as well as residues 43 to 47 and 75 to 80, resulting in an average deviation of 10.2°). Apart from residue 19, the EM differences in dihedral angles show little correlation with the data from NMR.

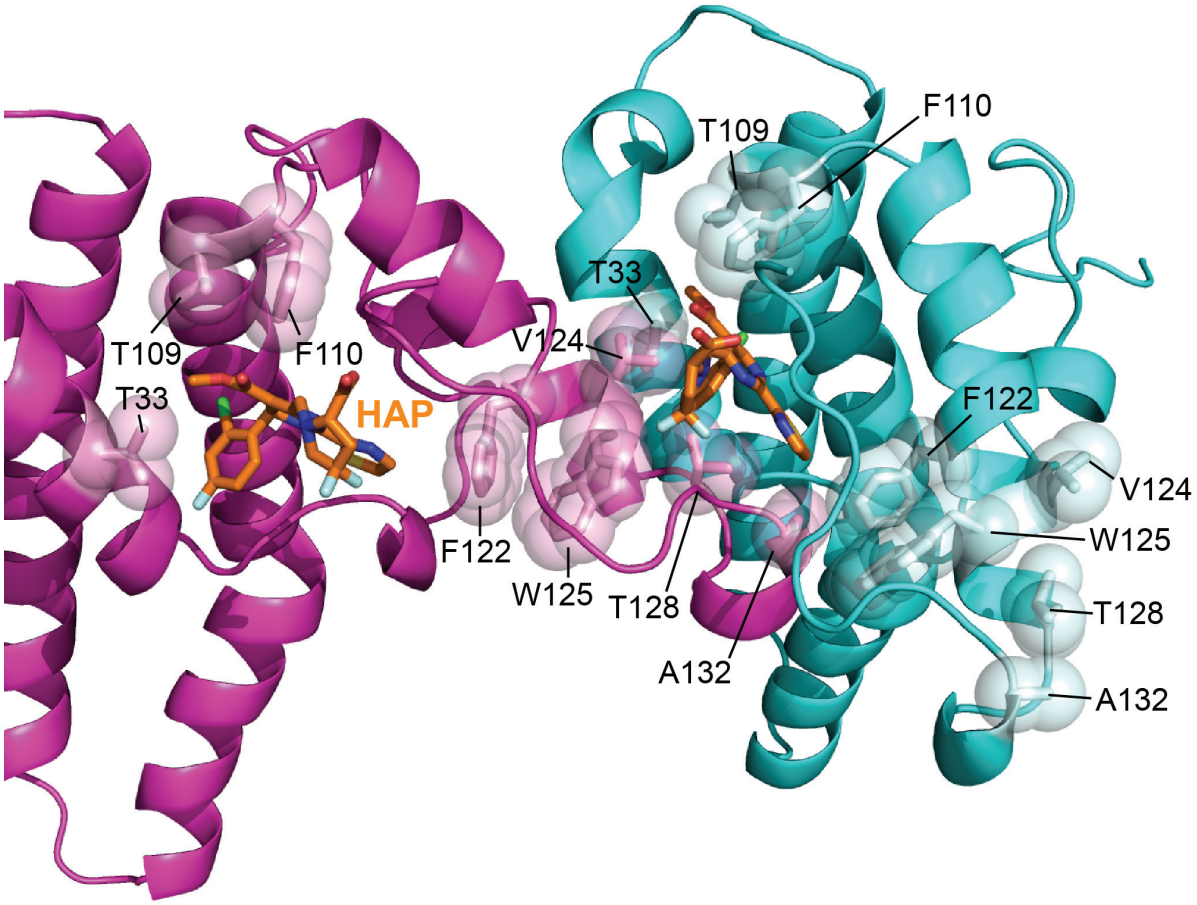


Figure S8. Hexamer structure of the HBV core protein Y132A mutant in complex with heteroaryldihydropyrimidine (HAP_R01, in orange sticks) (PDB: 5WRE(Zhou et al., 2017)) with residues identified as hinges by both NMR peak splitting and DynDom(Hayward and Berendsen, 1998; Taylor et al., 2014) program and less than 5 Å from the inhibitor shown as sticks and spheres. One HAP molecule is close to T33, T109, F110, F122, W125 from one subunit (here subunit B in cyan) and to V124, T128 and A132 from the neighboring subunit (here subunit C in magenta).

Residue	Chemical shifts (ppm)					
Ala11	N	C	Cα	Cβ		
Form1	120.5	172.9	48.8	21.8		
Form2	120.4	172.9	48.9	21.2		
Thr12	N	C	Cα	Cβ	Cγ2	
Form1	103.7	175.6	58.9	74.6	22.9	
Form2	103.7	175.7	58.9	74.9	23.0	
Form3	103.9	175.8	59.2	75.0	23.0	
Form4	103.7	176.0	59.1	75.6	23.9	
Leu16	N	C	Cα	Cβ	Cγ	
Form1	119.9	179.9	55.6	40.1	28.6	
Form2	118.6	179.4	56.4	40.9	29.8	
Form3	121.1	179.6	55.2	39.7	28.0	
Ser17	N	C	Cα	Cβ		
Form1	113.4	173.8	60.4	62.8		
Form2	115.6	173.3	61.1	62.7		
Form3	111.6	172.5	59.5	63.3		
Phe18	N	C	Cα	Cβ		
Form1	121.3	177.3	57.2	39.9		
Form2	122.9	176.3	56.8	40.4		
Leu19	N	C	Cα	Cβ	Cγ	
Form1	117.4	174.1	51.1	40.1	26.1	
Form2	118.0	173.8	51.7	39.8	28.0	
Pro20	N	C	Cα	Cβ	Cγ	Cδ
Form1	134.5	177.4	61.6	32.7	27.8	50.4
Form2	135.5	177.1	62.0	32.1	26.9	50.6
Ser21	N	C	Cα	Cβ		
Form1	116.5	174.4	61.6	62.9		
Form2	114.5	175.0	61.6	62.8		
Asp22	N	C	Cα	Cβ	Cγ	
Form1	114.7	176.4	52.8	39.3	181.0	
Form2	115.2	175.9	53.4	40.1	180.3	
Thr33	N	C	Cα	Cβ	Cγ2	
Form1	119.1	175.5	67.7	67.6	21.6	
Form2	119.1	174.6	66.7	68.2	21.7	
Ala34	N	C	Cα	Cβ		
Form1	122.7	178.9	55.4	16.7		
Form2	122.7	178.4	55.4	16.6		
Ser35	N	C	Cα	Cβ		
Form1	110.3	176.6	60.7	63.1		
Form2	110.8	176.5	60.9	63.1		
Form3	108.1	176.6	60.4	63.4		
Form4	112.1	176.1	61.0	63.1		
Leu108	N	C	Cα	Cβ		
Form1	118.3	177.9	57.1	42.4		
Form2	119.1	178.2	57.2	42.1		
Thr109	N	C	Cα	Cβ	Cγ2	
Form1	113.1	174.4	65.2	69.7	20.8	
Form2	114.9	174.9	65.4	69.5	20.3	
Phe110	N	C	Cα	Cβ		
Form1	115.0	175.5	58.7	40.4		
Form2	116.3	175.7	58.8	40.5		
Ser121	N	C	Cα	Cβ		
Form1	117.6	176.8	62.6	63.3		
Form2	114.2	178.1	61.6	62.9		

Phe122	N	C	Cα	Cβ		
Form1	126.6	177.2	61.9	38.7		
Form2	123.5	176.8	61.3	38.3		
Gly123	N	C	Cα			
Form1	105.7	174.2	47.0			
Form2	105.2	176.4	47.2			
Val124	N	C	Cα	Cβ	Cγα	Cγβ
Form1	118.5	178.1	66.5	31.9	21.2	22.8
Form2	120.3	178.9	66.3	31.9	20.9	23.2
Trp125	N	C	Cα	Cβ		
Form1	121.4	177.4	63.5	29.0		
Form2	126.0	179.0	63.8	29.2		
Arg127	N	C	Cα	Cβ		
Form1	119.2	174.9	56.0	31.0		
Form2	118.4	176.5	55.8	30.4		
Thr128	N	C	Cα	Cβ	Cγ2	
Form1	119.2	171.6	60.9	69.8	20.2	
Form2	121.0	171.6	61.1	69.7	18.8	
Form3	119.8	171.6	60.8	69.4	19.2	
Form4	120.0	171.6	60.9	69.7	20.1	
Pro130	N	C	Cα	Cβ	Cγ	Cδ
Form1	133.7	177.7	65.4	32.0	27.2	50.4
Form2	133.4	177.2	65.0	32.0	27.2	50.4
Ala131	N	C	Cα	Cβ		
Form1	117.6	178.2	54.2	18.0		
Form2	116.4	177.9	53.4	17.6		
Tyr132	N	C	Cα	Cβ		
Form1	112.9	172.8	55.4	39.3		
Form2	113.8	172.8	55.4	39.5		
Form3	112.2	173.3	55.7	39.7		
Asn136	N	C	Cα	Cβ	Cγ	
Form1	114.9	172.7	51.5	39.0	176.7	
Form2	114.8	172.5	51.9	35.8	177.5	
Form3	113.7	171.8	51.8	37.5	177.0	
Ala137	N	C	Cα	Cβ		
Form1	127.4	174.9	49.3	17.6		
Form2	126.3	174.8	49.3	16.9		
Form3	129.0	173.9	49.5	16.1		
Form4	128.6	174.2	49.4	15.6		
Pro138	N	C	Cα	Cβ	Cγ	Cδ
Form1	133.4	176.3	62.4	31.6	27.0	49.2
Form2	132.8	176.1	62.4	31.4	27.2	48.6
Form3	133.4	173.5	62.0	32.1	28.0	49.7

Table S1. Chemical shift values of the different forms observed for residues showing peak splitting. All values are given in ppm.

References :

Böckmann, 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 45, 319–327. doi:10.1007/s10858-009-9374-3

Fung, B.M., Khitrin, A.K., Ermolaev, K., 2000. An improved broadband decoupling sequence for liquid crystals and solids. J. Magn. Reson. 142, 97–101. doi:10.1006/jmre.1999.1896

Hayward, S., Berendsen, H.J., 1998. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins 30, 144–154.

Lecoq, L., Wang, S., Wiegand, T., Bressanelli, S., Nassal, M., Meier, B.H., Böckmann, A., 2018. Solid-state [13C-15N] NMR resonance assignment of hepatitis B virus core protein. Biomol. NMR Assign. 12, 205–214. doi:10.1007/s12104-018-9810-y

Parkin, S., Rupp, B., Hope, H., 1996. Structure of bovine pancreatic trypsin inhibitor at 125 K definition of carboxyl-terminal residues Gly57 and Ala58. Acta Cryst D52, 18–29. doi:10.1107/S0907444995008675

Stevens, T.J., Fogh, R.H., Boucher, W., Higman, V.A., Eisenmenger, F., Bardiaux, B., van Rossum, B.-J., Oschkinat, H., Laue, E.D., 2011. A software framework for analysing solid-state MAS NMR data. J. Biomol. NMR 51, 437–447. doi:10.1007/s10858-011-9569-2

Taylor, D., Cawley, G., Hayward, S., 2014. Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. Bioinformatics 30, 3189–3196. doi:10.1093/bioinformatics/btu506

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 59, 687–696. doi:10.1002/prot.20449

Wang, A.C., Bax, A., 1996. Determination of the Backbone Dihedral Angles φ in Human Ubiquitin from Reparametrized Empirical Karplus Equations. J. Am. Chem. Soc. 118, 2483–2494. doi:10.1021/ja9535524

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D., Wishart, D.S., 2003. VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res. 31, 3316–3319. doi:10.1093/nar/gkg565

Wlodawer, A., Walter, J., Huber, R., Sjölin, L., 1984. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. J. Mol. Biol. 180, 301–329.

Wynne, S.A., Crowther, R.A., Leslie, A.G., 1999. The crystal structure of the human hepatitis B virus capsid. Mol. Cell 3, 771–780.

Yu, X., Jin, L., Jih, J., Shih, C., Hong Zhou, Z., 2013. 3.5 Å cryo-EM Structure of Hepatitis B Virus Core Assembled from Full-Length Core Protein. PLoS ONE 8, e69729–11. doi:10.1371/journal.pone.0069729

Zhou, Z., Hu, T., Zhou, X., Wildum, S., Garcia-Alcalde, F., Xu, Z., Wu, D., Mao, Y., Tian, X., Zhou, Y., Shen, F., Zhang, Z., Tang, G., Najera, I., Yang, G., Shen, H.C., Young, J.A.T., Qin, N., 2017. Heteroaryldihydropyrimidine (HAP) and Sulfamoylbenzamide (SBA) Inhibit Hepatitis B Virus Replication by Different Molecular Mechanisms. Sci. Rep. 7, 42374. doi:10.1038/srep42374

● NMR reveals a second conformation of the hepatitis B virus capsid

The work in this part was done with Lauriane Lecoq, in collaboration with Michael Nassal and Beat H. Meier:

Lauriane Lecoq#, Shishan Wang#, Marie-Laure Fogeron, Marie Dujardin, Michael Nassal, Beat H. Meier, Anja Böckmann

Contribution: executed and analyzed biochemical experiments and electron microscopy, prepared all different samples for NMR characterization, including the transcription of pgRNA, contributed substantially to discussion and analysis of results; contributed to writing the paper.

➢ Abstract

The viral capsid plays a central role in hepatitis B virus (HBV) particle maturation, and the maturation signal hypothesis formulated early-on predicts that a structural change signals for envelopment in mature capsids. Still, the structures obtained by both X-ray and cryo-EM were consistently described as highly similar. We here show using solid-state NMR that the HBV capsid can take two different, well-defined but substantially different conformations. These two forms exist for a variety of capsid states and constructs, including from empty truncated, RNA-filled wild-type, phosphorylated capsids, as well as capsids from mutant core proteins. Both forms exist independently from the C61 interdimer disulfide bond and T=3/T=4 icosahedral symmetry. The slow spontaneous transition between the two conformations is catalyzed in disassembly/reassembly reactions. The largest conformational differences between the two forms cluster around the spike base, between helices 3 and 4 and the dimer interface. These changes are much larger than those induced by the presence of the C-terminal domain, nucleic acids, phosphorylation or different symmetry. We hypothesize that these two states of the nucleocapsid play a role in viral particle maturation.

➢ Introduction

Chronic hepatitis B infection affects more than 250 million people today(World Health Organization, 2017) and is the leading cause for terminal liver disease. The hepatitis B virus (reviewed in (Seeger et al., 2007)) is a small enveloped virus, which genomic information codes only for a few genes: the envelope proteins HBs, the core protein Cp, the polymerase P and HBx with yet undetermined function. Cp is a 183-residue protein with two domains: the assembly domain that forms the contiguous capsid shell, and the C-terminal domain (CTD, residues 150-183) that is amongst other responsible for RNA packaging. The two domains are connected by a linker (residues 141-149). The hepatitis B virus T4 capsid assembles from 120 copies of the dimeric core protein.

It packs the pregenomic (pg) RNA, as well as a copy of the viral polymerase, which transcribes the pgRNA to double-stranded relaxed circular (rc) DNA inside the capsid on maturation before envelopment of the particle. It is unclear how readiness for envelopment is conveyed, and Summers and Mason (Summers and Mason, 1982) postulated early-on that a conformational change during maturation might give the signal. Alternatively, the interplay of two events has been put forward to trigger this conformational change: achievement of genome maturation inside, sensed by the phosphorylation status of the CTD and transmitted by the domain to the outside.

As it has been recently reported that not only mature, rcDNA containing capsids with a defined C-terminal phosphorylation state are enveloped, but that a variety of subviral particles are secreted, including enveloped empty capsids (Luckenbaugh et al., 2015), and also phosphorylated ones (X. Ning et al., 2017), (reviewed in (Hu and Liu, 2017)), it becomes clear that neither C-terminal phosphorylation/dephosphorylation nor nucleic acid maturation are triggering envelopment. This strengthens today the hypothesis of an untriggered, spontaneous conformational change, as also described in HBV envelope maturation(Seitz et al., 2016).

The capsid structure has been investigated by a range of structural biology techniques, and besides a 3.3 Å X-ray structure (Wynne et al., 1999a; 1999b) of the N-terminal assembly domain, structures of the full-length capsid have been determined by cryo-electron-microscopy (cryo-EM) (Böttcher et al., 1997; Crowther et al., 1994), the latest to date at 3.5 Å resolution (Yu et al., 2013). The assembly domain forms T=4, and to a lesser extent, T=3 icosahedral capsids, made respectively from 120 and 90 Cp dimers. It was shown that by decreasing the C-terminal tail length, the number of T=3 capsids increased at the expense of T=4 capsids (Zlotnick et al., 1996). In both types of capsids, four monomers of Cp assemble in the asymmetric unit, which arrange around five-fold and quasi-sixfold axes. The core protein structure displays five α-helices, with helices 3 and 4 forming a hairpin. In the dimer, these hairpins assemble into the four-helical bundle which tips form the capsid spikes when assembled. Helix 5 is involved, together with the downstream loop, in interdimer interactions at the five- and quasi-sixfold capsid vertices. The highly positively charged CTD largely escapes as of today structural investigations, even if it could be localized to regions near the

inner surface of the capsid (Patel et al., 2017; Zlotnick et al., 1997). The different structures have mostly been described as similar to the X-ray structure, besides differences due to small differences attributed to the absence/presence of the CTD (Yu et al., 2013), the presence of RNA as opposed to DNA (Roseman et al., 2005), or due to drug binding (Schlicksup et al., 2018).

We here used NMR to investigate structural features of the HBV capsid. It is recognized that NMR chemical shifts detect conformational differences with very high sensitivity(Williamson, 2013), which has made the success of NMR for example in fibril studies, where one of its strong points resided in its ability to identify polymorphism (Meier et al., 2017). Notably, ^{13}C chemical shifts are directly related to protein backbone dihedral angles, and are for example used as restraints in structure calculation. Changes therein, called chemical-shift perturbations (CSPs), thus directly inform on changes in the dihedral angles and thus the structure. This is in contrast to ^{1}H CSPs, which are more sensitive to surface effects (Williamson, 2013). We have recently sequentially assigned the NMR signals of the HBV capsid, and could point out the residues which conformationally adapt on dimer-to-capsid transition (Lecoq et al., 2018a). Also, we identified the residues which form the hinges which accommodate in the capsid the formation of the five-fold and quasi-six-fold vertices (Lecoq et al., 2018b). Relying once more on the high sensitivity of the spectral fingerprints, we here identified a well-defined second conformation of the viral capsid, which surprisingly escaped detection during the last 20 years of intense structural scrutiny of this large and complex protein assembly.

➢ Material and Methods

Expression and Purification Cp140, Cp140 C61A, Cp149, Cp183, P-Cp183.

Proteins were expressed in *E. coli* and purified as described previously (Heger-Stevic et al., 2017; Lecoq et al., 2018a) with minor modifications depending on the constructs, as detailed below. Briefly, plasmids of pRSF_T7-HBc140opt, pRSF_T7-HBc140C61Aopt pET-28a2-HBc149opt, pRSF_T7-HBc183opt and pRSF_Tet-SRPK1dNS1_T7-HBc183opt were transformed into *E. coli* BL21* CodonPlus (DE3)

cells and grown at 37 °C in M9 minimal medium culture containing ^{13}C-glucose and $^{15}NH_4Cl$ (for more details on labeled culture see (Lecoq et al., 2018a)). When OD600 reached between 0.7 and 2.0, expression of Cp140, Cp140 C61A and Cp149 was induced with 1 mM IPTG overnight at 25 °C. For full-length proteins Cp183 and P-Cp183, expression was induced when OD600 reached 2.0 at a temperature of 20 °C and 16 °C, respectively. Cells were collected by centrifugation at 6,000 g for 20 min and resuspended in 15 ml (for one liter of culture) of TN300 buffer (50 mM Tris, 300 mM NaCl, 5 mM DTT, pH 7.5). The cells suspension were incubated on ice with 15 mg of chicken lysozyme, 300 µl of protease inhibitor cocktail solution 50× and 750 µl of 10 % Triton-X-100 for 45 min. 6 µl of Benzonase nuclease was added to digest nuclear acids for 30 min at room temperature. Cells were broken by sonication using a minimum of 10 cycles of 10 s of sonication and 50 s of cooling down on ice. Cell lysates were centrifuged at 8,000 g for 1 h to remove cell debris. The supernatant was loaded onto 10 to 60 % sucrose gradient buffered with 50 mM Tris pH 7.5, 300 mM NaCl, 5 mM DTT and centrifuged in SW-32Ti Beckman Coulter swinging bucket rotor at 140,000 g for 3 h at 4°C. Capsids in gradient fractions were identified by 15 % SDS-Polyacrylamide gel and precipitated by 20 % (for Cp140 and Cp149) or 40 % (for Cp183 and P-Cp183) saturated $(NH_4)_2SO_4$. After incubation on ice for 1 h and centrifugation at 20,000 g for 1 h, pellets were resuspended in 10 ml of purification buffer (50 mM Tris pH 7.5, 5 mM DTT). The protein solution was centrifuged again for 15 min to remove insoluble pellet. The supernatant containing soluble capsids was dialyzed overnight against purification buffer at 4 °C.

In-vitro transcription of pgRNA.

The pCHT-3101pgRNA-SNAP plasmid (6.3 kbp) was transformed in DH5α cells and grow overnight at 37 °C in presence of ampicillin. The circular plasmid was extracted with a commercial kit (QIAprep Spin Miniprep Kit), and then digested for 2 h at 37 °C by *Xma1* enzyme (New England Biolabs, Inc) to recover linear DNA. DNA was purified by extraction with one volume of phenol:chloroform, again with one volume of chloroform, then precipitated with 2.5 volume of 100 % ethanol and 3 M sodium acetate at pH 5.2. The pellet was resuspended in RNase-free H_2O and the concentration was checked by nanodrop based on the absorbance value at 260 nm. For the *in vitro*

transcription (Beckert and Masquida, 2011), linearized plasmid at 0.1 μg/μl was used as a template and incubated with 0.06 μg/μl of home-made T7 RNA polymerase in presence of 50 mM TRIS-HCl at pH 7.5, 5 mM of ribonucleotide triphosphates (rNTPs) mix (Promega), 25 mM $MgCl_2$, 0.01 mM Triton, 10 mM DTT, 2.5 mM spermidine, and 1.6 U/μl of RNAsin. After 2 h incubation at 37 °C, 25 mM EDTA were added to stop the transcription. The freshly-transcribed 3.2 kb pgRNA was purified with the same protocol as the linear DNA, resuspended in RNase-free H_2O and loaded on a 0.8 % agarose gel to check RNA quality prior to reassembly with Cp183 dimer. Typically, 1 μg of DNA was able to transcribe 30 μg of pgRNA.

Capsid disassembly and reassembly.

For Cp140, Cp140 C61A and Cp149, capsids were dialyzed against disassembly buffer A (50 mM NaHCO3, 5 mM DTT, pH 9.6) at 4 °C overnight and disassembled by adding solid urea to a final concentration of 3 M. Dimers were separated by size exclusion chromatography (SEC) using HiPrep™ 16/60 Sephacryl® S-200 HR column (120 ml) equilibrated in disassembly buffer A. Dimers (~2 mg/ml) were reassembled into capsids by dialysis against reassembly buffer A (50 mM Tris, 250 mM NaCl, 5 mM DTT, pH 7.5).

Capsids of Cp183 purified from *E. coli* were disassembled by dialysis against disassembly buffer B (2 M GuHCl, 50 mM Tris, 0.5 M LiCl, 5 mM DTT, pH 7.5) at 4 °C overnight. Disassembly was about 60 % complete for full-length capsids, while for truncated capsids disassembly was virtually complete. Encapsidated RNA precipitated by LiCl was removed by centrifugation (20,000 g, 15 min at 4 °C). Dimers were separated by size exclusion chromatography (SEC) using HiPrep™ 16/60 Sephacryl® S-200 HR column (120 ml) equilibrated in disassembly buffer B. For reassembly of Cp183 capsids, dimers of Cp183 (~0.8 mg/ml) in 2 M GuHCl were diluted 4 times with dilution buffer (50 mM HEPES, 0.5 M LiCl, 5 mM DTT, pH 7.5) to a final concentration of 0.5 M GuHCl, and dialyzed in presence of pgRNA against reassembly buffer B (50 mM HEPES, 150 mM NaCl, 5 mM DTT, pH 7.5) overnight at room temperature. The molar ratio of Cp183 dimer:pgRNA for reassembly was 60:1. For P-Cp183, the same protocol was applied with the exception that 2.5 M GuHCl was used instead of 2 M, and no RNA was added on reassembly. For the five samples, the reassembled capsids

were loaded on HiPrep™ 16/60 Sephacryl® S-200 HR column to separate pure capsids from residual dimers or RNA, and were dialyzed overnight against purification buffer at 4 °C to be in the same conditions as capsids purified from *E. coli*.

SDS-PAGE under non-reducing conditions.

20 µl protein solutions were incubated with 250 mM iodoacetamide at room temperature for 20 min to block free thiols. Samples were mixed with 5 µl 5x non-reducing loading buffer (250 mM Tris·HCl, 10% SDS, 30% (v/v) Glycerol, 0.05% (w/v) Bromophenol Blue, pH 7.5) and 10 µl were loaded onto 15% SDS-PAGE.

Separation of Cp140 T=3/T=4 capsid.

T=3 capsids of Cp140 were separated from T=4 capsid by sucrose density gradient centrifugation as described previously (Zlotnick et al., 1996). Sucrose gradients (5%, 10%, 15%, 20%, 25%, 30%) were prepared with TN300 buffer. Each layer of sucrose solution was 6 ml. Capsid sample (1 ml) was loaded onto the top of the gradient and separated at 140,000 g for 3 h at 4 °C in SW-32Ti Beckman Coulter swinging bucket rotor. Sucrose gradient fractions containing Cp140 capsid were analyzed by electron microscopy, and fractions containing mainly T=3 capsids were pooled and dialyzed against purification buffer at 4 °C.

NMR rotor filling.

After dialysis in purification buffer, soluble capsids purified from *E. coli* and capsids reassembled *in vitro* were concentrated using Amicon Ultra centrifugal filter units (Merck, 50 kDa cut-off) to about 30 mg/ml in 1 ml and sedimented into 3.2 mm zirconium rotor by centrifugation (200,000 g, 14 h, 4 °C) using a home-made filling tool (Böckmann et al., 2009). Rotors were immediately closed with watertight caps.

NMR spectroscopy.

All NMR spectra were recorded using a 3.2 mm triple-resonance (^{1}H, ^{13}C, ^{15}N) probe head at a static magnetic field of 18.8 T corresponding to 800 MHz proton resonance frequency (Bruker Avance II). Assignment of the form A was described previously and was performed on Cp149 capsids purified from *E. coli* (Lecoq et al., 2018a). Similarly,

assignment of the conformation B was obtained at a MAS frequency of 17.5 kHz using a combination of 2D and 3D correlation experiments, including: 2D DARR, NCA, NCO, 3D NCACB, NCACX, NCOCX, CANCO and CCC which were performed on the sample of P-Cp183 capsids reassembled *in vitro*. On all the other samples, only 2D spectra were recorded including 2D DARR, NCA and NCO, except on the sample of Cp149 capsids purified from *E. coli* and stored for one month at room temperature for which a 3D NCACX was recorded in order to unambiguously identify the resonances of oxidized C61. Experimental details are given in Table S1. All spectra were referenced using DSS and recorded at a sample temperature of 4 °C according to the resonance frequency of the supernatant water (Böckmann et al., 2009). Spectra were processed using TopSpin 3.2 (Bruker Biospin) and analyzed with the CcpNmr Analysis package (Stevens et al., 2011; Vranken et al., 2005).

Data deposition.

The ^{13}C and ^{15}N backbone and side-chain chemical shifts of the core protein of HBV capsids in conformation B have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number XXXXX (to be deposited).

Calculation of A and B percentage.

The percentage of each conformation was calculated based on the intensity of six isolated peaks corresponding to P5 Cγ-Cδ, P5 Cγ-Cα and I59 Cδ1-Cβ resonances for which the distinction between the two conformations A and B was clearly visible (leading to two peaks per resonance). If the signal was nonexistent, a symbolic peak was added at the virtual position of the missing conformation. Peaks heights were fitted using the Gaussian fit function of CcpNmr software (Stevens et al., 2011; Vranken et al., 2005) and the percentage of conformation B was calculated according to:

$$Conformation\ B\ (\%) = \frac{1}{n}\sum_{n}^{1}\frac{I_B}{[I_A + I_B]} * 100$$

where n is 3 here, I_A is the intensity of conformation A and I_B the intensity of conformation B for each of the 3 resonances. Error was calculated as the standard deviation of conformation B percentage between the three resonances.

Calculation of CSPs.

Chemical-shift perturbations between the conformation A and B were calculated for each assigned carbon atoms according to: $\Delta\delta_C = |\delta_C[A] - \delta_C[B]|$.

➢ Results

1. NMR spectra reveal two different, but well-defined capsid conformations

HBV capsids can be produced using expression in *E. Coli*, where they self-assemble in the bacteria during expression, and can further be purified as such (see material and methods). 2D ^{13}C DARR spectra can be directly recorded on sediments of the capsid preparations (as previously shown for CTD-truncated Cp (Cp149) (Lecoq et al., 2018b; 2018a)). Figure 1 shows the spectrum obtained for the Cp183 capsids isolated from *E. coli* in blue. The Cp183 spectrum could mainly be assigned by transferring the sequential assignments of Cp149 (Lecoq et al., 2018a). The 34 additional residues from the C-terminal domain are not observed in the DARR spectra, likely due to their dynamic behavior. We also recorded the corresponding spectrum on Cp183 capsids reassembled in the presence of *in-vitro* transcribed pgRNA (based on protocols described in (Porterfield et al., 2010)), and which is shown in pink in Figure 1.

We clearly observe important chemical-shift perturbations (CSPs) between the ^{13}C spectral fingerprints for the two preparations, which reveal two distinct conformations we refer to in the following as A and B. Indeed, the observation of CSPs in NMR spectra is directly related to conformational change, as the chemical shifts are highly sensitive to the electronic environment and thus the dihedral angles of the protein backbone and side chains. These differences reveal that the capsid takes on two different conformations between the *E. coli* RNA-filled capsids, and the reassembled pgRNA containing capsids.

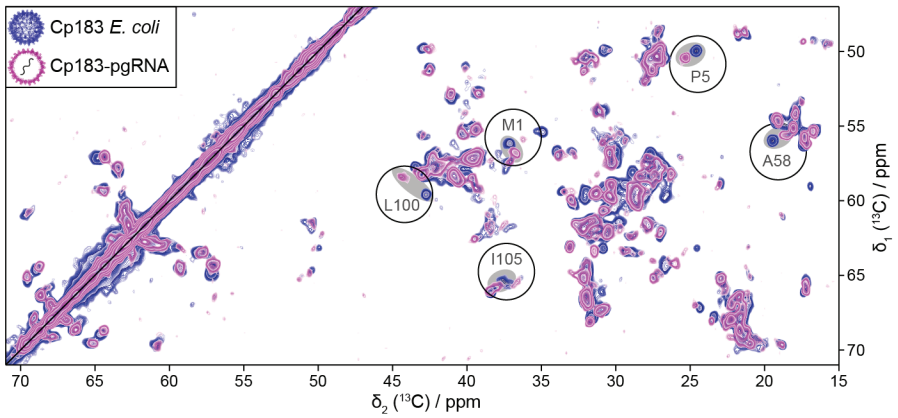


Figure 1. Two sets of chemical shifts are observed for Cp183 and Cp183-pgRNA. Overlay of the Cα-Cβ region of 2D ^{13}C DARR spectra of E. coli Cp183 and Cp183-pgRNA. Spectra were recorded on samples prepared from capsids isolated from *E. Coli* (in blue), and on capsids disassembled/reassembled in the presence of *in-vitro* transcribed pgRNA (in pink). Examples of chemical shift perturbations are highlighted by grey background and are labelled with the corresponding residue number. Signal assignments for Cp183-pgRNA have been transferred from sequential assignments done using a set of 3D experiments (see Material and Methods) on reassembled P-Cp183 capsids (see below) as representative sample (Figure S1).

2. The A to B transition progresses over time and can be catalyzed by reassembly

To identify the origin and specificity of the two conformations observed for Cp183, we produced a variety of capsids including truncated capsids lacking the CTD and the linker (Cp140 and Cp140 C61A mutant) or only the CTD (Cp149), as well as capsids made of full-length protein with 7 phosphorylated sites, P-Cp183, which is obtained by co-expression of protein kinase SRPK1 in *E. Coli* (Heger-Stevic et al., 2017; Rabe et al., 2016). The latter is, as also the truncated constructs, devoid of nucleic acids, as shown by the absence of signals stemming from nucleic acids in the spectra (Figure S2) when compared to Cp183 *E. Coli* capsids which contain ^{13}C labeled RNA. Spectra were recorded on both *E. Coli* capsids and *in-vitro* reassembled capsids, which are reassembled with unlabeled pgRNA for Cp183, and without nucleic acids for the truncated capsids and P-Cp183. Figure 2a compares the proline 5 Cγ-Cδ signal, as a distinct signature and representative signal of the conformational change operating between forms A and B, for all capsids. It can be seen that all investigated constructs

can exist in either conformation A or B. Cp149, Cp183 and P-Cp183 capsids isolated from *E. Coli* consistently displayed highest percentages of A, while the corresponding reassembled capsids were mainly found in conformation B (>85 %) (Figure 2b). One should note that consistent preparation of pure A states proved delicate for most constructs, and often small bymixtures of B were observed in the samples. Statistics on 14 Cp149/Cp183 samples gave an average of 80±14 % of A for E. coli assembled capsids, and for 6 reassembled capsids 93±5 % of B. For the constructs of Cp showing a complete assembly domain including the linker, the major determinant for B forms was thus the disassembly and reassembly reaction. In comparison, Cp140 showed always a mixture of A and B forms, both in the *E. Coli* and reassembled capsids, with however B being dominant in the latter. We also investigated the Cp140 C61A mutant, which fingerprints resulted in A/B mixtures, both in *E. Coli* and reconstituted forms, but with different ratios. Thus, in contrast to Cp149 and Cp183, Cp140 is conformationally more variable, and its behavior is remnant of amyloid polymorphism, as pure forms were difficult to obtain, and external conditions, as well as mutations, changed ratios with contrasting results. For example, *E. coli* and reassembled Cp140 showed an A/B mixture, while the *E. coli* Cp140 C61A mutant resulted in rather pure A. Reassembly of both only increased B, but did not convert all A to B.

In order to determine whether disassembly/reassembly was needed to transform A into B, we analyzed if A could spontaneously convert to B simply with time. For this, two samples from the same batch of Cp149 from *E. Coli* were left for 1 month either at room temperature (RT) or at 4 °C. The recorded spectra showed that the percentage of B clearly increased in both, indicating that the conformation indeed slowly changes from A to B. However, the transition was not or only weakly temperature dependent, since the A/B ratios were similar for samples left at RT and 4 °C. It is interesting to note that once the capsids are sedimented in the NMR rotor, which represents a confined environment, form A is stable over months.

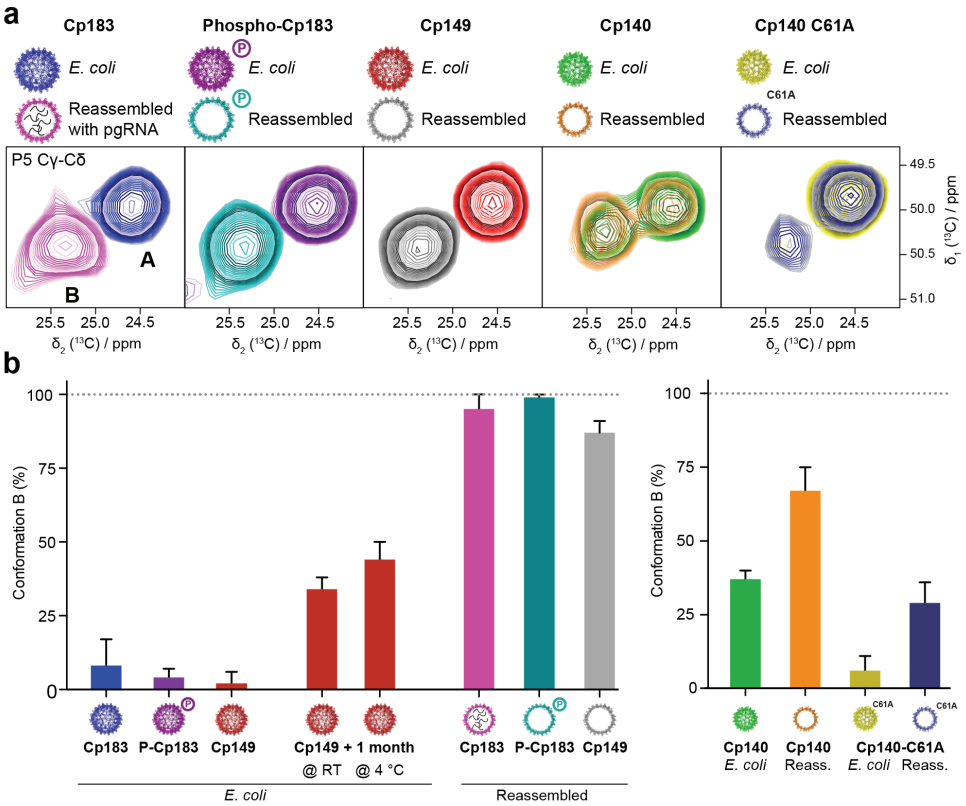


Figure 2. Two different forms are observed for all capsid constructs. a) Region of 2D DARR spectra corresponding to the Cγ-Cδ correlation peak of Pro5 as representative signature for A and B forms. For each construct of core protein, spectral extracts both from capsids isolated from *E. coli* and *in vitro* reassembled capsids are shown. For Cp149, Cp183 and P-Cp183, the capsids from *E. coli* show clear signals of conformation A, while the *in vitro* reassembled capsids show signals typical for B. For Cp140, a mixture of signals corresponding to conformations A and B is visible. **b**) Percentage of conformation B in the different capsids shown in panel a) for capsids from *E. coli* and reassembled capsids. In addition, two samples from Cp149 left one month at different temperatures are shown. The percentage of B was calculated based on the intensity of six representative and isolated peaks in the DARR spectrum (for P5 Cγ-Cδ, P5 Cγ-Cα and I59 Cδ1-Cβ resonances, see Material and Methods). Error bars represent the standard deviation between the percentages obtained for the different resonance signals analyzed.

3. A and B forms are compatible with both T=3 and T=4 icosahedral capsids

In contrast to Cp149 and Cp183, which form a majority of T=4 icosahedral capsids, the truncation of residues from the linker has been shown to result mainly in T=3 icosahedral capsids (Zlotnick et al., 1996). In order to investigate if the observation of two states A and B is related to the presence of T=3 and T=4 capsids, we analyzed all samples by electron microscopy (EM) using negative staining. Figure 3 summarizes

the icosahedral symmetries observed for the samples for which spectra are shown in Figure 2. As expected, for Cp183, P-Cp183 and Cp149, both types of capsids yielded a large majority of T=4 symmetries, independent of the two conformations A and B observed by NMR. In addition, we prepared a sample containing a majority of T=3 capsids, which were isolated by fractioning Cp140 *E. Coli* and reassembled capsids using an additional sucrose gradient. As virtually pure T=3 capsids were observed by EM, the mixtures of A and B conformations in the NMR spectra of Cp140 cannot be explained by the capsid symmetry. Cp140 C61A yielded mainly T=3, yet both *E. Coli* and reassembled capsids were rather found in form A. As all T=4 Cp149 and Cp183 capsids can access both A and B, and T=3 Cp140 a mixture of A and B, we can thus conclude that both forms A and B are compatible with the two icosahedral arrangements T=3 and T=4, and that icosahedral symmetry is no a determinant for A or B states.

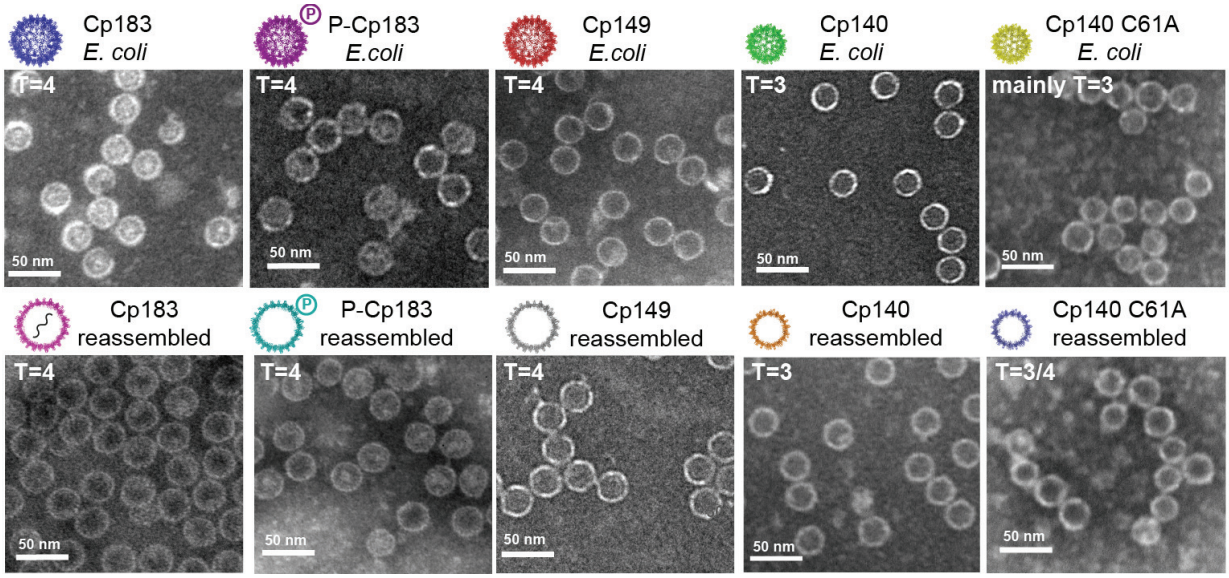


Figure 3. The two forms are independent on icosahedral symmetry. Negative stain EM of the different capsid samples concentrated at 1 mg/ml. Cp183, P-Cp183 and Cp149 form quasi exclusively T=4 capsids in both *E. Coli* and reassembled capsids, while Cp140 forms both T=3 and T=4 capsids, and mainly T=3 capsid can be isolated by an additional sucrose gradient to exclude T=4 capsids. Cp140 C61A yields mainly T=3 *E. Coli* capsids but rather a mixture of T=3 and T=4 after reassembly.

4. A and B forms are observed independently from the oxidation state of C61

C61 is located in the dimer interface between the two core protein monomers, and can form a covalent disulfide bond between them, resulting in a stable dimer when C61 is in its oxidized form (Selzer et al., 2014). C61 disulfide bond formation has been shown not to be essential in capsid formation, and the capsid exists in reduced (Yu et al., 2013) and oxidized (Wynne et al., 1999a) forms in the different available structures. To understand whether the C61 oxidation state has an impact on the formation of the conformation A or B, we analyzed C61 resonance signals in the ^{13}C NMR spectra. Cysteine Cα, and mainly Cβ chemical shifts, clearly identify the oxidation states (Sharma and Rajarathnam, 2000). As the protein is purified in the presence of DTT, most samples show reduced Cys chemical shifts, as reflected by the red and grey bars for forms A and B in Figure 4a and b. The blue bars corresponding to the sample kept one month at RT show that a substantial amount of oxidized C61 is present, as indicated by the observation of a signal corresponding to the oxidized form (Figure S3) which was sequentially assigned using 3D spectroscopy. Also, a SDS-PAGE run under non-reducing conditions (Figure 4c) indicates that the sample kept at RT reveals the presence of oxidized dimer. Still, the sample kept at RT which is to a large extent oxidized show mixed A/B forms, clearly indicating that the C61 disulfide bridge is compatible with both forms, and that its formation does not induce either A or B forms. We also assessed the oxidation state of C61 for all other samples, which were all found to be reduced, independently of the A/B ratio (Figure S4).

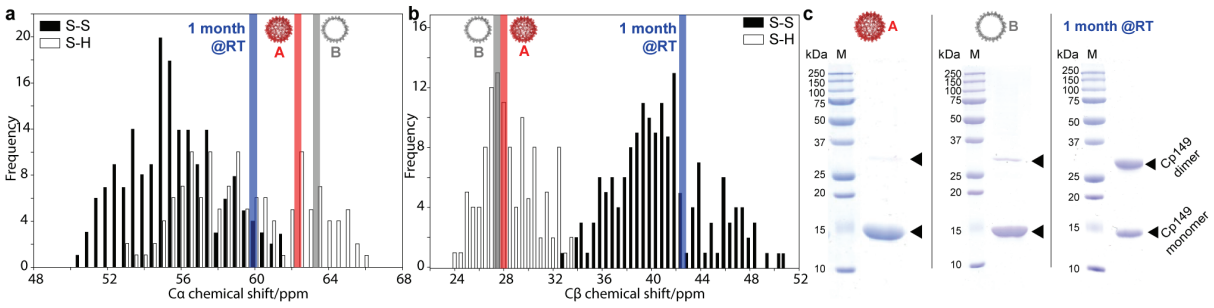


Figure 4. The two forms are independent on the oxidation state of C61. The chemical-shift distribution of cysteine Cα (**a**) and Cβ (**b**) as a function of redox state (from data bank and adapted from literature as compiled by (Sharma and Rajarathnam, 2000)) is given for reference in black and white, and the position of the Cα and Cβ resonances in different Cp149 preparations as determined here is indicated by colored bars. While for Cp149 in forms A (red bar) and B (grey bar)

Cα and notably Cβ clearly show a reduced state, Cp149 stored one month at RT (blue bar), displays C61 chemical shifts corresponding to an oxidized form. This is equally reflected in the SDS-PAGE gel run under non-reducing conditions using 250 mM iodoacetamide (**c**), where the protein stored at RT displays more S-S bound dimer.

5. Structural differences between forms A and B concentrate in the spike base

The two different states are distinguished by their chemical shift. Figure 5a shows a comparison between the carbon chemical shifts of *E. Coli* Cp183 and pgRNA-reassembled Cp183 capsids (for comparisons between the five constructs, see Figure S5). CSPs up to 1.8 ppm in ^{13}C can be observed, which indicates substantial conformational differences (Williamson, 2013). We consider CSPs smaller than 0.3 ppm not significant (in grey), and larger than 0.6 ppm substantial (on red). Intermediate CSPs are shown in orange. The larger CSPs between forms A and B concentrate in three major regions: the N-terminus, and the residues clustering around positions 60 and 100. In addition, residues V27 and L42 show isolated differences. The three major regions are distant in the primary structure (Figure 5a), but actually colocalize in the 3D fold. This can be seen in Figure 5b, where CSPs > 0.6 ppm are plotted as red spheres on the 3D Cp149 crystal structure (Wynne et al., 1999a). The residues clearly concentrate in the spike base. The interactions between the residues are detailed in Figure 5c, as lines connecting neighboring residues (< 4 Å, but long-range in the sense that they are more than 5 residues apart in the sequence) in the structure. In this network, extensive *intra*monomer contacts exist between helices 3 and 4 (L55/I105, A58/F97/L100, W62/F97, L65/M93), supplemented by contacts between in helices 2 and 3 (V27/W62, L42/L55), and 2 and 4 (V27/R98), as well as between the N-terminus and helix 4 (Y6/Q99/L100). *Inter*monomer contacts exist between the N-terminus and helices 2 and 3 (M1/I59, I3/L42/I59/L60 and P5/L60), between the salt-bridge forming residues E64 and K96, and the C61-C61 disulfide. Changes are observed up to the middle of the spike, where residues and L68 faces M93. Only E8 remains isolated from this network. CSPs are virtually absent from the *inter*dimer interfaces, which form the fivefold and quasi-sixfold vertices in the capsid; and also the spike tip is totally unaffected.

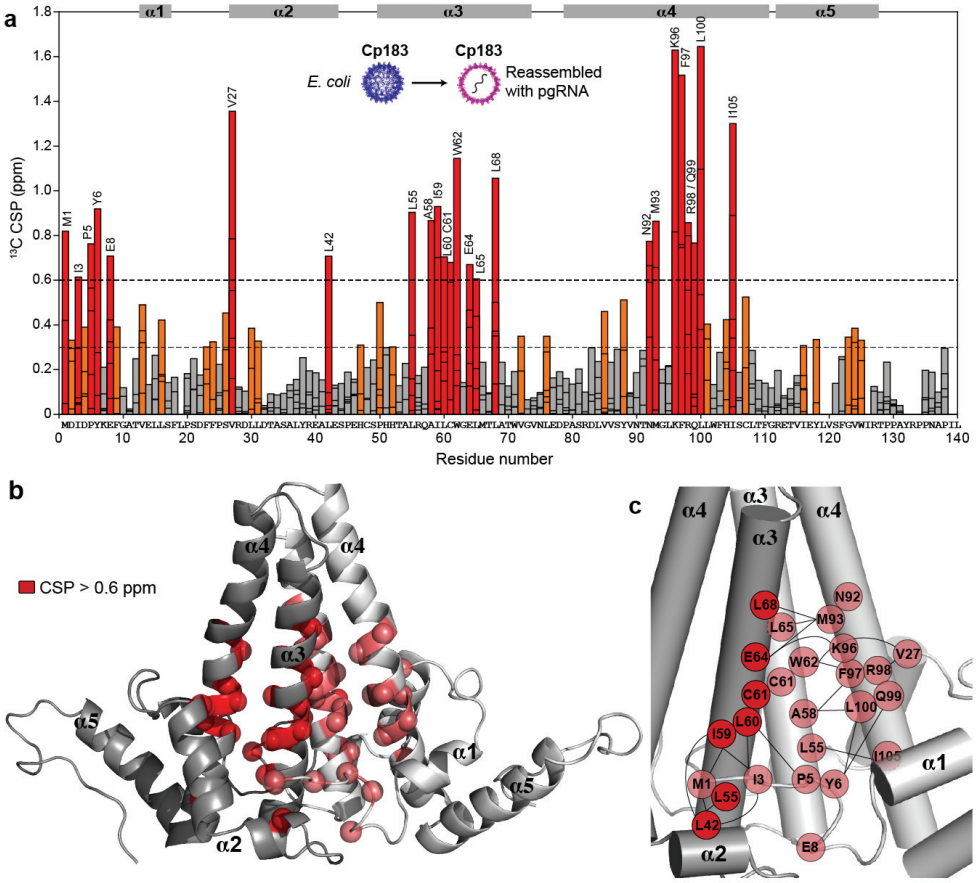


Figure 5. **Localization of structural differences between conformations A and B. a**) Histogram showing chemical-shift perturbations (CSPs) induced between forms A and B, calculated as $CSP = |\delta_C[A] - \delta_C[B]|$. A single bar is plotted for each assigned carbon atom (see for example V27, where the five transparently superimposed bars correspond to the CSPs of the five assigned carbon resonances). The protein sequence up to residue 140 is given below the graph. CSPs larger than 0.6 ppm are colored red, those smaller than 0.3 ppm grey, and the values in between orange. Names and numbers of amino acids showing high CSPs are given on the graph. **b**) Cα of residues with CSPs > 0.6 ppm are mapped as colored spheres on the X-ray structure (subunits B and C) of Cp149 dimer (PDB 1QGT(Wynne et al., 1999b)). The structural rearrangement between forms A and B occurs in the base of the capsid spike. Color code: dark grey and red for subunit C, light grey and salmon for subunit D. **c**) Proximities in the 3D structure (≤ 4 Å) between residues with CSPs > 0.6 ppm are indicated as solid lines on the structure with cylindrical helices.

6. Truncation, phosphorylation, nucleic acid content and different icosahedral symmetry induce only minor conformational differences

The CSPs observed in the spectra of A and B forms represent the largest conformational differences observed between any forms of capsid we investigated. M1 shows a particular behavior in that its chemical shift is different between most preparations and forms, even between nominally or virtually identical capsids. It is however always in a well-defined state, and is never missing from any spectrum, which means that it is neither flexible nor statically disordered. When comparing the other constructs between them, minor changes can however be observed as well (Figure 2), which allow to evaluate the impact of the presence of the C-terminal domain, different nucleic acids, phosphorylation, as well as of icosahedral symmetry.

In order to compare capsids with and without CTD, spectra recorded on Cp149 and P-Cp183 (Figures 6a), both devoid of nucleic acids, and between Cp149 and Cp183 (Figure 6b, where Cp183 capsids contain nucleic acids), were compared and their CSP analyzed, , for both *E. coli* and reassembled capsids. The most striking observation is that the CTD is not observed in any of the full-length protein spectra, presumably due to intermediate dynamics, which interference with NMR parameters as magic-angle spinning and proton decoupling, which can broaden signals beyond detection or compromise polarization transfers. The CTD residues are neither observed in INEPT spectra which would have displayed them if they showed very fast local rotational correlation times as observed in small proteins in solution. One can see that the CSPs are mostly below 0.3 ppm, meaning that the presence of the CTD does not induce a significant structural reorganization in the rest of the protein. Interestingly however, signals for several glutamates, in particular E113, but to a lesser extent also E43, are attenuated in the P-Cp183 spectra (see Figure S6b, c), indicating that interactions, likely with the CTD residues, heterogenously broaden and thus attenuate the resonances of this residue. This is the case for both A and B forms.

To assess the impact of the presence of nucleic acids, Figure 6c shows the CSPs between Cp183 and P-Cp183, again for forms A and B. The major difference between the two constructs are the nucleic acids signals which are visible in the *E. coli* Cp183

spectrum, in the region around 63-95 ppm (Figure S2). Nucleic acid resonances match the chemical shifts typically for the sugar moieties from ribose in RNA, revealing that Cp183 packs mainly RNA from *E. coli*, and not DNA (Figure S7). Despite the fact that one capsid is filled with nucleic acid and the other is devoid of material, only very limited chemical shift differences are observed between the Cp183 and P-Cp183 capsids, neither in A nor B forms. A small effect is seen, when comparing Cp183-pgRNA and reassembled P-Cp183, for E46, which signals shows less intensity in the former, suggesting that the RNA interacts, likely via the CTD, with E46, which sticks out towards the interior of the capsid, underneath the spikes. This is not observed for *E. coli* Cp183 vs P-Cp183, hinting at slightly different interactions for E. coli and pgRNA.

To compare the T=3 and T=4 capsids, Figure 6d shows the CSPs detected between spectra from Cp140 and Cp149 capsids, also for both forms A and B. One can see that, while mainly small CSPs are present, some residues show substantial CSPs, which localize mainly to the interdimer interfaces, to peaks which show peak splitting, caused by the asymmetry of A, B, C and D subunits near the five-fold and quasi-sixfold vertices (Lecoq et al., 2018b). One can observe also the four distinct peaks observed for the four different monomers in the asymmetric unit cell show slightly different resonances, which differences however remain small, as shown at the example of A137 in Figure S8.

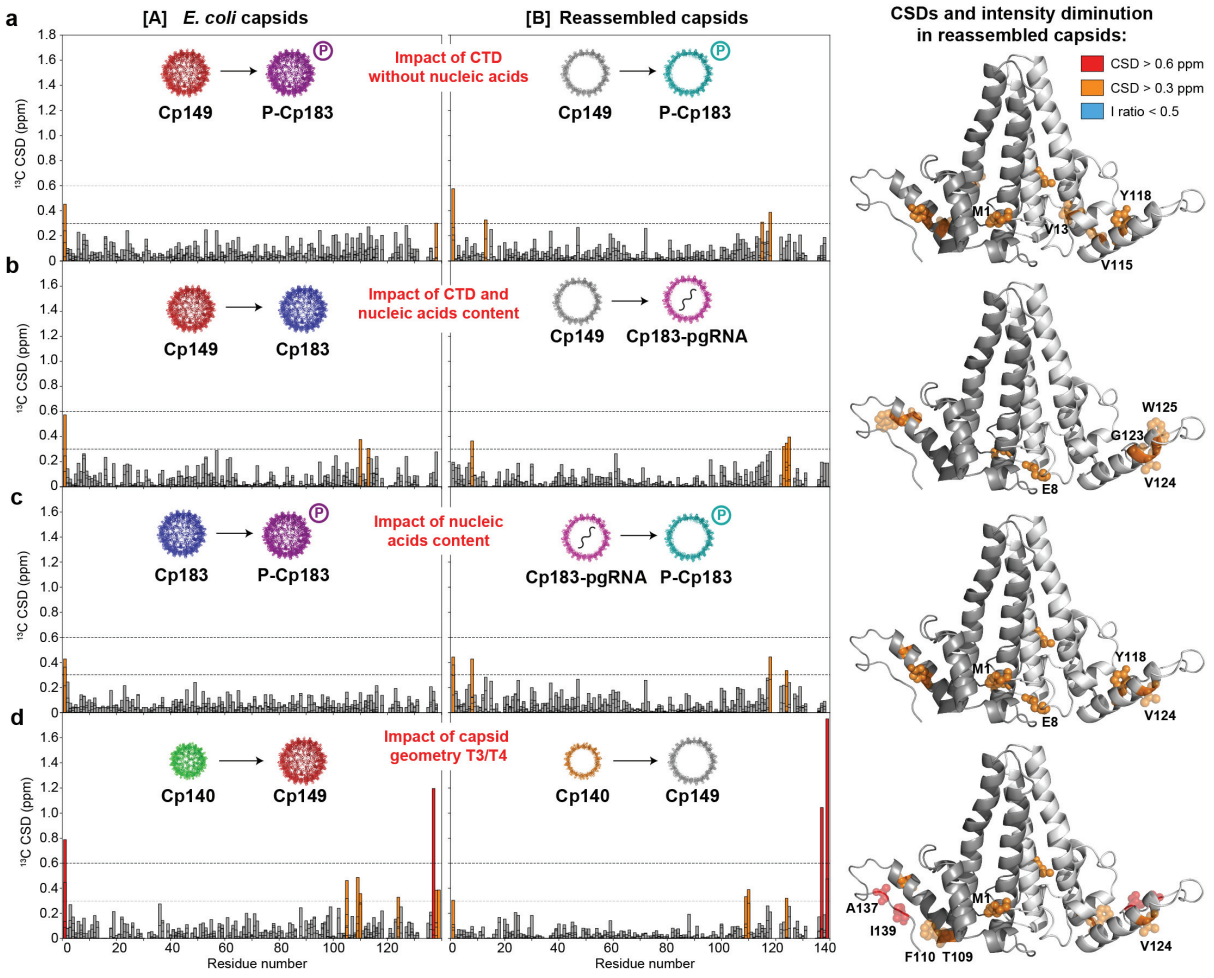


Figure 6. Impact of different constructs on the capsid structure. CSPs induced by **a**) the truncation of the C-terminal domain in absence of nucleic acids (Cp149 *vs*. P-Cp183) **b**) the truncation of the CTD and nucleic acid content (Cp149 *vs*. Cp183), **c**) the CTD phosphorylation and nucleic acid content (Cp183 *vs*. P-Cp183), and **d**) the T=3 *vs* T=4 capsid geometry (Cp140 *vs*. Cp149). For each comparison, the chemical shift differences are shown for both *E. coli* (form A) and reassembled (form B) capsids. Residues with CSPs above 0.3 ppm and 0.6 ppm in reassembled capsids are shown on the structure (PDB 1QGT(Wynne et al., 1999b)) in orange and red, respectively.

➢ Discussion

1. The switch between A and B forms is the dominant difference between capsid forms

We have investigated ten different forms of the HBV capsid. Samples cover the range of truncated, full-length, phosphorylated, empty and RNA-filled capsids. We have

measured for all forms chemical shifts, and compared them in order to identify the impact of the presence of the C-terminal, the icosahedral symmetry, and the nucleic acid-content. Overall, the conformational changes associated to these variations were small, but we could identify an interaction of the C-terminal domain with the around amino acid E113, as well as a slight impact of the presence of pgRNA on residue E46, pointing towards the capsid center. Also, conformational changes could be identified at the interdimer interfaces between the two different capsid symmetries.

The single most important conformational change is however surprisingly detected independently of the type of construct, in the form of the two different conformations A and B which we identified for all five different core protein constructs tested. Indeed, we could show that the two forms are neither function of the nucleic-acid content, the phosphorylation state, the oxidation state, nor the icosahedral symmetry of the capsid. We showed that the transition between the two states takes places spontaneously, but can be catalyzed by disassembly/reassembly. The chemical shift differences observed between the two conformations revealed significant conformational differences, which localize at the base of the spikes, at the intradimer interface. The key residues in this conformational transition from A to B belong to the hydrophobic core of Cp, but also comprise surface residues, as the intermonomer salt bridge forming E64 and K96. Importantly, no significant CSPs are observed for the residues located at the interdimer interfaces, which we previously reported to show NMR peak-splitting (Lecoq et al., 2018a). This clearly shows that, even if NMR chemical shifts cannot directly measure the spatial organization of the core proteins in the shell, the capsid quaternary structure remains largely unaffected by the A/B transition.

2. Capsid structures in the literature represent form B

We have shown here that capsids isolated from *E. coli* slowly transit from form A to B over a period of months. Amongst the X-ray structures of HBV capsids available to date, four were solved from capsids isolated from *E. coli* (Bourne et al., 2006; Tan et al., 2007; Wynne et al., 1999a). Where crystallization conditions were described, diffraction-quality crystals were only obtained after several months (Bourne et al., 2006; Wynne et al., 1999b). More recently, it has been reported that long-term storage

of capsids induced higher order in cryo-EM micrographs (Rabe et al., 2016). There, “fresh” capsids (vitrified 2 weeks post-preparation) were shown to present heterogeneities, notably in the N-terminus, while “aged” capsids (vitrified after several months) showed higher resolution and lower B factors, described as in line with increasing ordering with time. Interpretation of these observations in the light of our results suggests that for x-ray diffraction, form B is likely more prone to form high-quality crystals. In the EM study, the “fresh” sample probably contained to a certain extent both forms A and B, and the “aged” capsids pure form B. The differences between the two structures obtained from “fresh” and “aged” capsids have been described as a stretching of the spike helices and the inter-dimer contact region; resolution (8 Å) is however not sufficient to pinpoint structural details on a per-residue basis. Finally, the highest-resolution cryo-EM structure (3.5 Å) to date has been determined using *E. coli* capsids(Yu et al., 2013). Still, experimental methods given do not allow to assess how aged the preparations were, besides that they were described as fresh. The structure was described to be highly similar to the x-ray structure from(Wynne et al., 1999b), and indeed a comparison of the two structures does not highlight specific differences which could be related to A/B forms. It seems thus that until now mainly capsid form B has been determined at high resolution. Form A might have appeared in the literature in EM work from Roseman and coworkers (Roseman et al., 2005), which describes two different capsid forms at 7.6 Å resolution, which however have been interpreted as stemming from different nucleic acid content, on one hand RNA and on the other DNA (for further discussion see below).

3. A possible role for A and B conformers in capsid maturation

Is the second conformation of the capsids simply an artefact produced by capsid assembly in *E. coli*, or is the existence of forms A and B and their interconversion important in capsid maturation? During HBV replication, only the mature capsids containing rcDNA are selected for envelopment, but not immature ones containing single-stranded DNA or pgRNA, in order to create complete and infectious virions. In addition, phosphorylated capsids devoid of DNA can also be enveloped (Hu and Liu, 2017; X. Ning et al., 2017) and are secreted as empty virions in even larger amounts than rcDNA-containing virions. It was hypothesized early-on that the envelopment of

virions would require a maturation-associated structural change of the capsid (Hu and Liu, 2017; Summers and Mason, 1982).

In this context, mutations in amino-acid residue F97 (or I97) have been associated with an immature secretion phenotype (Yuan et al., 1999). The mutation of F97L impacts, as shown in Figure S9, several residues in the protein, but large differences are only observed for residues L55 and E64 in form B. The impact of the mutation slightly differs between forms A and B, and shows a larger impact compared to for example the C61A mutation (Figure S9b). Residue F97 is located at the bottom of a hydrophobic pocket which has been previously identified using computer-aided structure analysis of the capsid (B. Ning and Shih, 2004), and which was defined as including V27, L31, A58, I59, C61, W62, M93, K96, I97, L100 and L101 residues (B. Ning and Shih, 2004). All these residues, with exception of L31, have been identified here to take part in the conformational transition between forms A and B, as can be seen in Figure 7a, where the hydrophobic pocket as identified by Ning et al. (B. Ning and Shih, 2004) is shown on the structure in surface mode in a, and the residues identified here to change conformation between A and B forms in b. Additional residues in vicinity of the pocket show CSPs, which form a sort of groove around the spike base, and include in addition the charged E64 which forms a saltbridge with K96, which actually lines the top of the entrance to the hydrophobic pocket. Ning and Shi proposed that a conformational change is transmitted through the residues in this hydrophobic pocket via the C terminus which can sense the growing genome maturity via phosphorylation or dephosphorylation (B. Ning and Shih, 2004). Our data does not support this, as both forms are observed as well in the phosphorylated capsids. Our data neither complies with the hypothesis that phosphorylation events can trigger this difference, at least not with the here used hyper-phosphorylated form, which can access both A and B types. As mentioned above, structural modifications in this hydrophobic pocket have been observed between recombinant RNA-containing capsids from *E. coli*, and DNA-filled cores isolated from viral particles using electron microscopy (Roseman et al., 2005). It was hypothesized that these differences are due to the nature of the nucleic acid packed inside the capsids, and that they might be related to the maturation signal. While we did not evaluate DNA-filled capsids, there are no major conformational differences between the behavior of C-terminally truncated, full-length *E. coli* RNA-

filled capsids, and pgRNA filled capsids, and the conclusion lies near that the two forms observed might indeed correspond to the A and B forms identified here. One should note that we do not detect deviation from symmetry for the tip residues of the spike, as described in EM micrographs (Roseman et al., 2005), which would result in NMR spectra in peak splittings as previously observed for the resonances at the interdimer interfaces (Lecoq et al., 2018b).

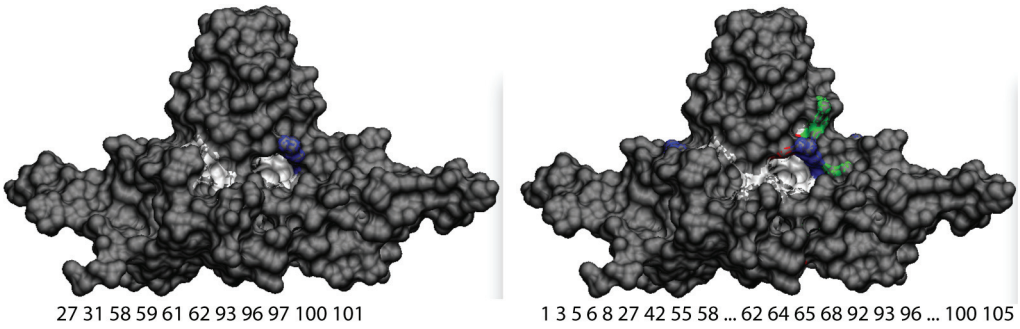


Figure 7. The residues changing conformation between forms A and B form part of the hydrophobic pocket around residue F97. a) residues identified to line the hydrophobic pocket as determined by (B. Ning and Shih, 2004). **b**) amino acids sensing CSPs induced between forms A and B capsids (in red in Figure 5) shown on the structure (PDB 1QGT (Wynne et al., 1999b)) color coded according to residue type (white, hydrophobic; red, acidic; blue, basic; and green, polar) .

It is tempting to hypothesize that one of the two conformations revealed here corresponds to the “single-strand blocking” state (X. Ning et al., 2011) while the other one would represent the “envelopment-signaling” state. Transition from A to B in vivo could either just be a function of time, or maturation of nucleic acid content could play the role of a trigger; we have so far however not detected any preference of certain preparations, be it with or without nucleic acids, phosphorylated capsid or not, for states A or B. It is however clear that B is the thermodynamically more stable form, as transition from A to B occurs, spontaneously as a function of time, or triggered, by for example disassembly/reassembly, but no transition from B to A has been observed under any conditions. If the two forms play indeed a role in maturation, and how exactly one of them would signal readiness for envelopment, remains to be determined.

➢ Conclusions

We have identified, using the high sensitivity of NMR chemical shifts, two different, but well-defined structural forms for the HBV capsid. They are independently observed of any modification of the viral capsid, and can be accessed universally by all forms investigated. Form A is metastable, and can transit over time to form B. This transformation is the single most important change observed, while truncation, presence of nucleic acids, phosphorylation and icosahedral symmetry only have minor impact on the capsid structure. We hypothesize that the identified forms might be related to capsid maturation and the signal for envelopment, as the structural differences between them colocalize with residues in a hydrophobic pocket around residue F97, which mutant F97L has been shown to relate to the premature envelopment phenotype.

➢ Acknowledgements

This work was supported by the French Agence Nationale de Recherche (ANR-14-CE09-0024B), the LABEX ECOFECT (ANR-11-LABX-0048) within the Université de Lyon program Investissements d'Avenir (ANR-11-IDEX-0007), by the Swiss National Science Foundation (Grant 200020_159707), the Chinese Scientific Council (SW) and the European Marie Skłodowska-Curie program (H2020-MSCA-IF-2016 748516). We thank the Centre d'Imagerie Quantitative Lyon-Est (CIQLE) for access to the EM platform.

➢ References

Beckert, B., Masquida, B., 2011. Synthesis of RNA by in vitro transcription. Methods Mol. Biol. 703, 29–41. doi:10.1007/978-1-59745-248-9_3

Bourne, C.R., Finn, M.G., Zlotnick, A., 2006. Global Structural Changes in Hepatitis B Virus Capsids Induced by the Assembly Effector HAP1. J. Virol. 80, 11055–11061. doi:10.1128/JVI.00933-06

Böckmann, 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 45, 319–327. doi:10.1007/s10858-009-9374-3

Böttcher, B., Wynne, S.A., Crowther, R.A., 1997. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature 386, 88–91. doi:10.1038/386088a0

Crowther, R.A., Kiselev, N.A., Böttcher, B., Berriman, J.A., Borisova, G.P., Ose, V., Pumpens, P., 1994. Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell 77, 943–950.

Heger-Stevic, J., Kolb, P., Walker, A., Nassal, M., 2017. Displaying whole-chain proteins on hepatitis B virus capsid-like particles. Virus-Derived Nanoparticles for Advanced Technologies.

Hu, J., Liu, K., 2017. Complete and Incomplete Hepatitis B Virus Particles: Formation, Function, and Application. Viruses 9, 56–17. doi:10.3390/v9030056

Lecoq, L., Wang, S., Wiegand, T., Bressanelli, S., Nassal, M., Meier, B.H., Böckmann, A., 2018a. Solid-state [13C-15N] NMR resonance assignment of hepatitis B virus core protein. Biomol. NMR Assign. 12, 205–214. doi:10.1007/s12104-018-9810-y

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

Luckenbaugh, L., Kitrinos, K.M., Delaney, W.E., Hu, J., 2015. Genome-free hepatitis B virion levels in patient sera as a potential marker to monitor response to antiviral therapy. Journal of Viral Hepatitis 22, 561–570. doi:10.1111/jvh.12361

Meier, B.H., Riek, R., Böckmann, A., 2017. Emerging Structural Understanding of Amyloid Fibrils by Solid-State NMR. Trends in Biochemical Sciences 42, 777–787. doi:10.1016/j.tibs.2017.08.001

Ning, B., Shih, C., 2004. Nucleolar localization of human hepatitis B virus capsid protein. J. Virol. 78, 13653–13668. doi:10.1128/JVI.78.24.13653-13668.2004

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. J. Virol. 91, e00092–17–16. doi:10.1128/JVI.00092-17

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 Para-retrovirus. PLoS Pathog 7, e1002255–14. doi:10.1371/journal.ppat.1002255

Patel, N., White, S.J., Thompson, R.F., Bingham, R., Weiß, E.U., Maskell, D.P., Zlotnick, A., Dykeman, E.C., Tuma, R., Twarock, R., Ranson, N.A., Stockley, P.G., 2017. HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. Nature Microbiology 1–10. doi:10.1038/nmicrobiol.2017.98

Porterfield, J.Z., Dhason, M.S., Loeb, D.D., Nassal, M., Stray, S.J., Zlotnick, A., 2010. Full-Length Hepatitis B Virus Core Protein Packages Viral and Heterologous RNA with Similarly High Levels of Cooperativity. J. Virol. 84, 7174–7184. doi:10.1128/JVI.00586-10

Rabe, M., Aisenbrey, C., Pluhackova, K., de Wert, V., Boyle, A.L., Bruggeman, D.F., Kirsch, S.A., Böckmann, R.A., Kros, A., Raap, J., Bechinger, B., 2016. A Coiled-Coil Peptide Shaping Lipid Bilayers upon Fusion. Biophys. J. 111, 2162–2175. doi:10.1016/j.bpj.2016.10.010

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. Proc. Natl. Acad. Sci. U.S.A. 102, 15821–15826. doi:10.1073/pnas.0504874102

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, e31473–23. doi:10.7554/eLife.31473

Seeger, C., Zoulim, F., Mason, W.S., 2007. Hepadnaviruses. Book chapter 1–38.

Seitz, S., Iancu, C., Volz, T., Mier, W., Dandri, M., Urban, S., Bartenschlager, R., 2016. A Slow Maturation Process Renders Hepatitis B Virus Infectious. Cell Host and Microbe 20, 25–35. doi:10.1016/j.chom.2016.05.013

Selzer, L., Katen, S.P., Zlotnick, A., 2014. The hepatitis B virus core protein intradimer interface modulates capsid assembly and stability. J. Biol. Chem. 53, 5496–5504. doi:10.1021/bi500732b

Sharma, D., Rajarathnam, K., 2000. 13C NMR chemical shifts can predict disulfide bond formation. J. Biomol. NMR 18, 165–171.

Stevens, T.J., Fogh, R.H., Boucher, W., Higman, V.A., Eisenmenger, F., Bardiaux, B., van Rossum, B.-J., Oschkinat, H., Laue, E.D., 2011. A software framework for analysing solid-state MAS NMR data. J. Biomol. NMR 51, 437–447. doi:10.1007/s10858-011-9569-2

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, 403–415.

Tan, W.S., McNae, I.W., Ho, K.L., Walkinshaw, M.D., 2007. Crystallization and X-ray analysis of the T = 4 particle of hepatitis B capsid protein with an N-terminal extension. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 63, 642–647. doi:10.1107/S1744309107033726

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 59, 687–696. doi:10.1002/prot.20449

Williamson, M.P., 2013. Using chemical shift perturbation to characterise ligand binding. Prog. Nucl. Mag. Res. Sp. 73, 1–16. doi:10.1016/j.pnmrs.2013.02.001

World Health Organization, 2017. Global hepatitis report [WWW Document]. www.who.int. URL http://www.who.int/hepatitis/publications/global-hepatitis-report2017/en/ (accessed 11.23.17).

Wynne, S.A., Crowther, R.A., Leslie, A.G., 1999a. The crystal structure of the human hepatitis B virus capsid. Mol. Cell 3, 771–780.

Wynne, S.A., Leslie, A.G.W., Butler, P.J.G., Crowther, R.A., 1999b. Crystallization of hepatitis B virus core protein shells: determination of cryoprotectant conditions and preliminary X-ray characterization. Acta Crystallogr. D Biol. Crystallogr. 55, 557–560. doi:10.1107/s0907444998012621

Yu, X., Jin, L., Jih, J., Shih, C., Hong Zhou, Z., 2013. 3.5 Å cryo-EM Structure of Hepatitis B Virus Core Assembled from Full-Length Core Protein. PLoS ONE 8, e69729–11. doi:10.1371/journal.pone.0069729

Yuan, 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. J. Virol. 73, 5731–5740.

Zlotnick, A., Cheng, N., Conway, J.F., Booy, F.P., Steven, A.C., Stahl, S.J., Wingfield, P.T., 1996. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. Biochemistry 35, 7412–7421. doi:10.1021/bi9604800

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. Proc. Natl. Acad. Sci. U.S.A. 94, 9556–9561.

➢ Supplementary Material

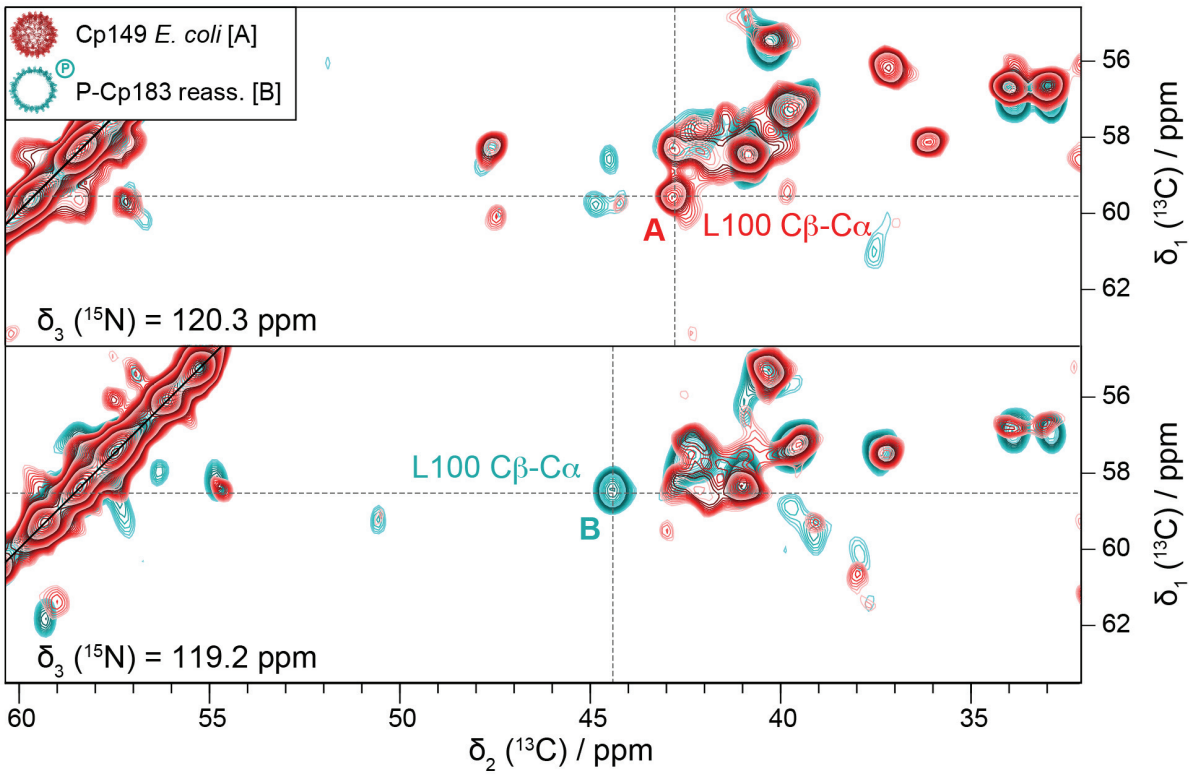


Figure S1. 3D extracts of NCACX spectra recorded on Cp149 *E. coli* capsids (form A in red) and P-Cp183 reassembled capsids (form B in cyan), and **b)** Cp183 *E. coli* capsids (form A in blue) and Cp183 reassembled capsids with pgRNA (form B in magenta). Assignment is shown for residue L100 which has a different chemical shift in the three dimensions (^{15}N, ^{13}C and ^{13}C) between the two conformations A and B (at the intersection of the dotted lines).

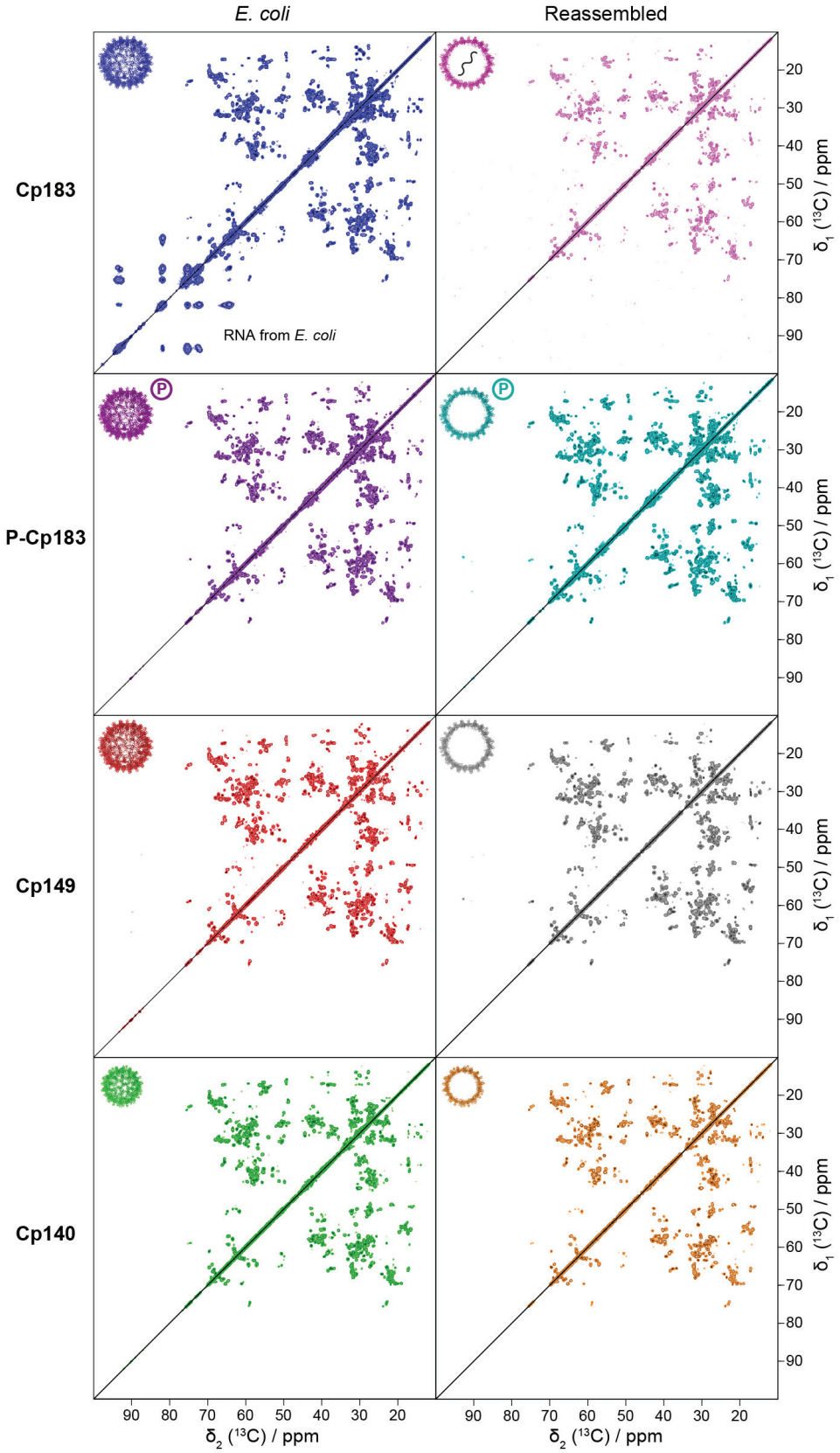


Figure S2. Aliphatic region of DARR spectra with a 20 ms mixing time for 8 capsid samples obtained either from *E. coli* (left panels) or after *in-vitro* reassembly (right panels). Capsids from

Cp183 are the only ones able to package RNA, which is isotopically labeled in the case of the *E. coli* sample.

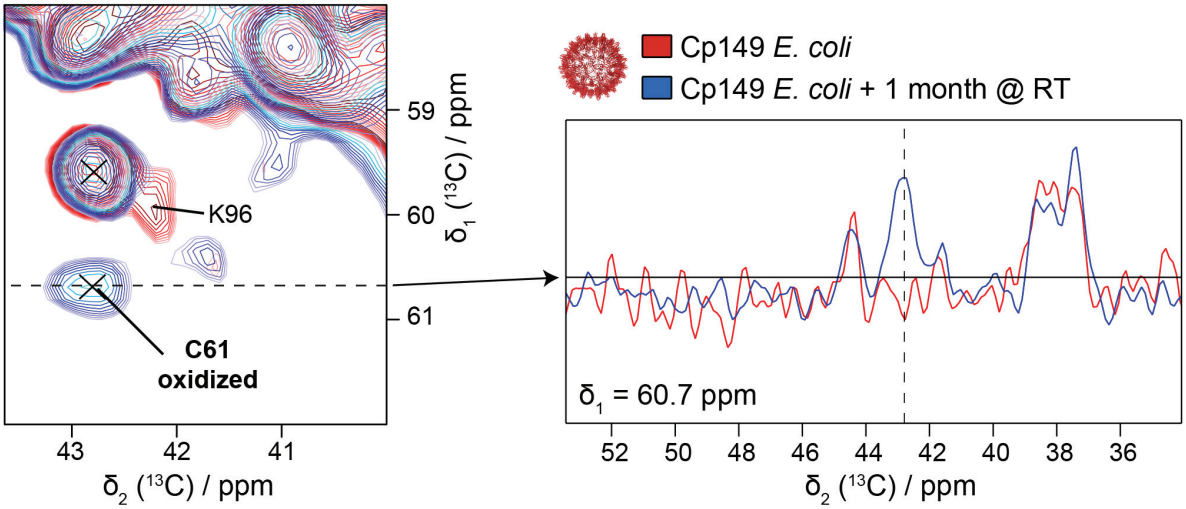


Figure S3. 2D region of oxidized Cys61 C -C (left panel) and corresponding 1D trace at 60.7 ppm (right panel) for Cp149 from *E. coli* (in red) and equivalent sample left for 1 month at room temperature (RT, in blue). Only the sample left at RT shows the Cys61 at a chemical shift typical for cysteines involved in disulfide bonds, with C at 60.7 ppm and C at 42.8 ppm.

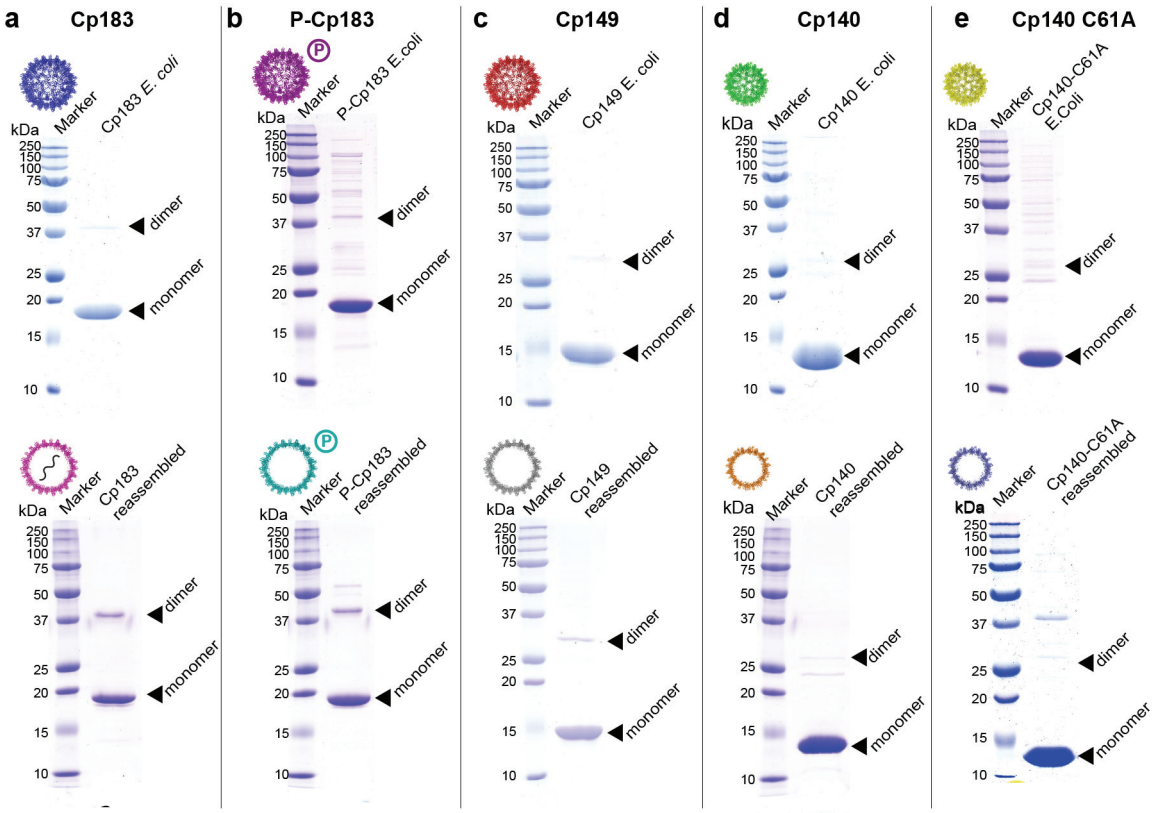


Figure S4. SDS-PAGE in non-reducing conditions for the 10 samples.

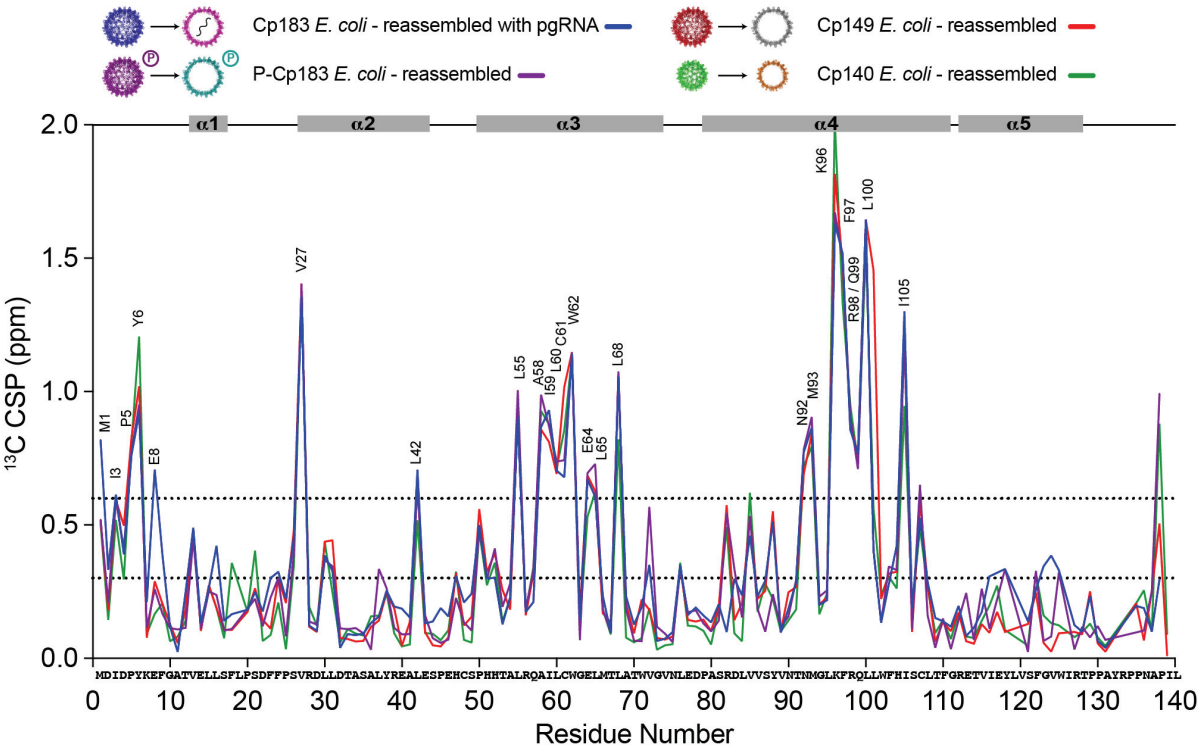


Figure S5. Chemical shift perturbations (CSPs) induced between forms A and B, calculated as

$\mathrm{CSP} = |\delta_C[A] - \delta_C[B]|$, for the five different constructs used in this study. For clarity, only the maximum CSP (in carbon) is shown for each residue. The protein sequence up to residue 140 is given below the graph. Names of amino acids showing high CSPs are given on the graph. The four different constructs (Cp183, P-Cp183, Cp149 and Cp140) give similar CSPs as detailed for *E. coli* vs reassembled Cp183 in the main text.

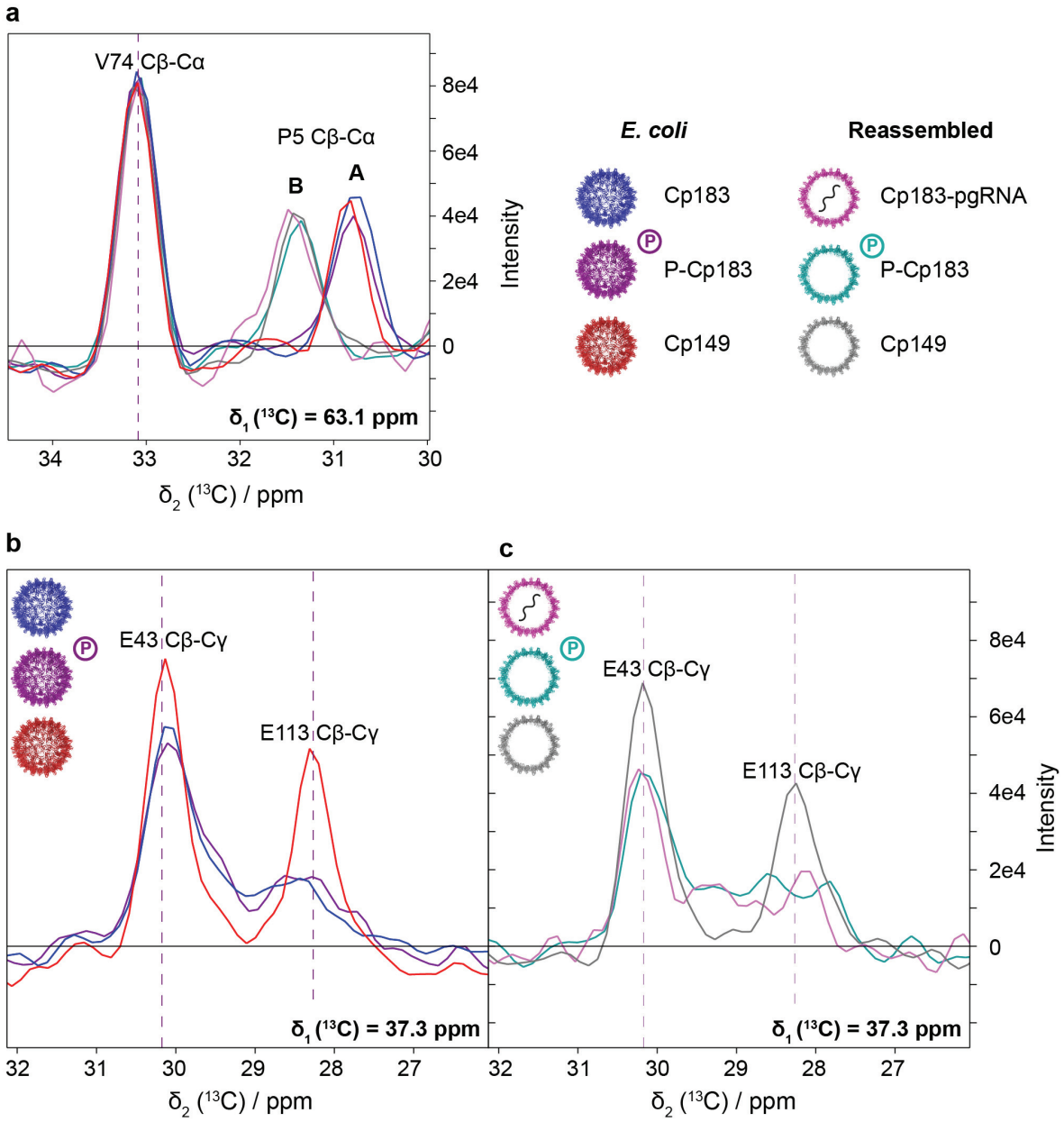


Figure S6. 1D traces from DARR spectra (20 ms) for full-length proteins (Cp183 and P-Cp183) and truncated protein (Cp149) both from *E. coli* and *in vitro* reassembled capsids. **a**) Intensities were scaled on the Cβ-Cα correlation peak of V74, which is located in a structured region and has the same chemical shift in forms A and B, and can thus be used to reference intensity between different spectra. As an example, P5 Cβ-Cα traces are shown where the difference between A and

B forms is clear. *E. coli* capsids from Cp183 (blue), P-Cp183 (purple) and Cp149 (red) are all in form A while reassembled capsids of Cp183 with pgRNA (pink), P-Cp183 (cyan) and Cp149 (grey) are all in form B. **b**) Intensities of E43 and E113 Cβ-Cγ correlation peak are shown for the *E. coli* capsids, and **c**) the reassembled capsids, showing that the presence of the CTD, phosphorylated or not, has an effect on the peak intensity of both glutamate residues, especially E113.

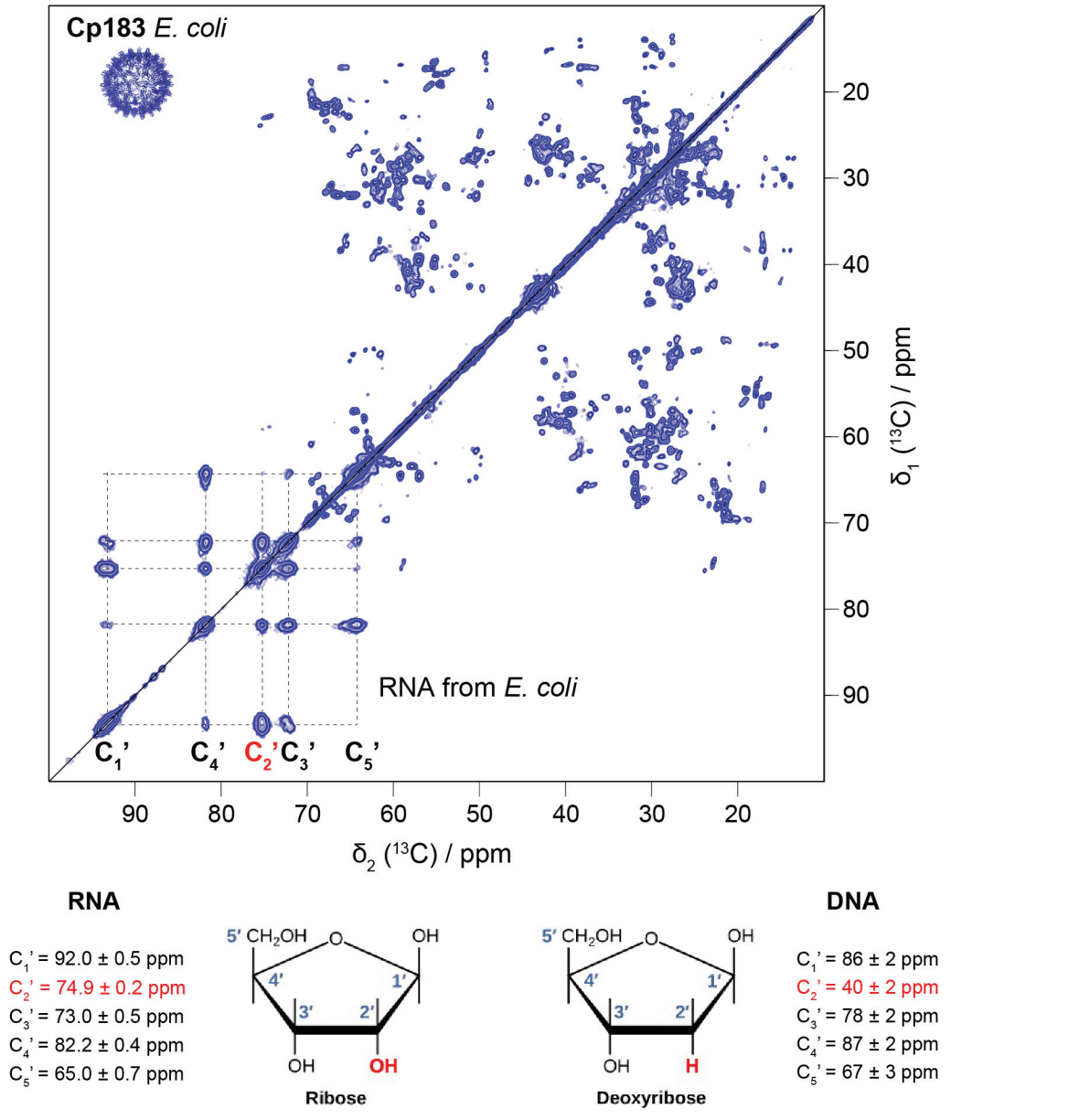


Figure S7. DARR 20 ms of Cp183 from *E. coli*. and typical chemical shifts of sugar ribose and deoxyribose found in the BMRB. The chemical shift around 75 ppm matches the C_2' chemical shift of ribose, which means that Cp183 packs RNA from *E. coli*. and not DNA, as no correlations are observed around 40 ppm.

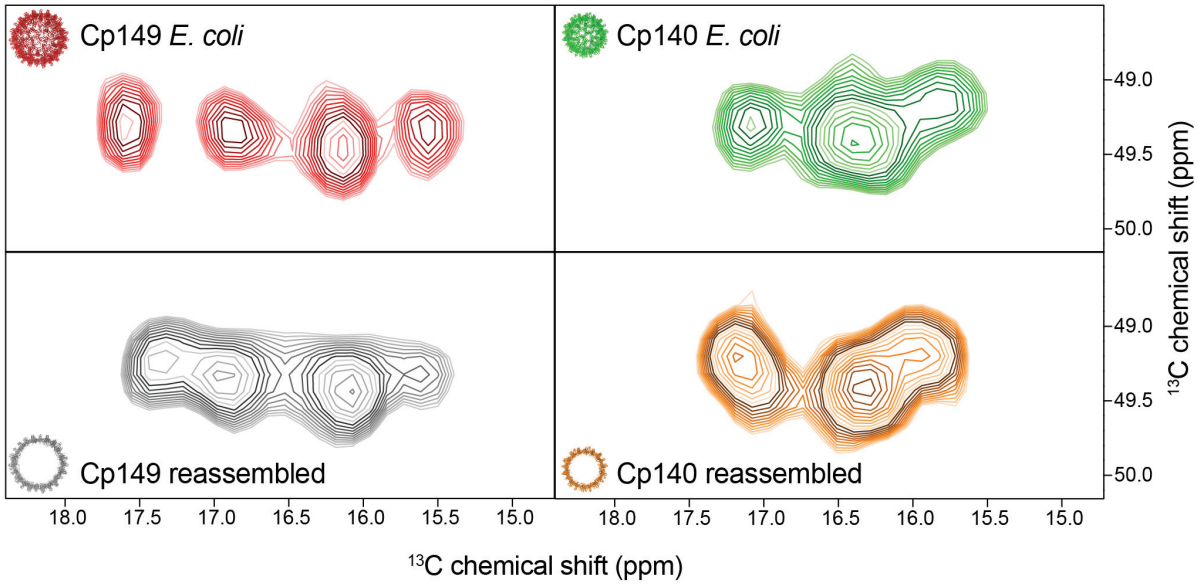


Figure S8. Zoom on Ala137 Cb-Ca correlation peak in T=3 and T=4 capsids. Cp149 *E. coli* capsids (in red) and reassembled capsids (in grey) display 4 distinct peaks corresponding to the A, B, C and D subunits, while Cp140 *E. coli* capsids (in green) and reassembled capsids (in orange) display 3 distinct peaks with one more intense than the two others. It means that two subunits have the same environment in the T=3 capsids.

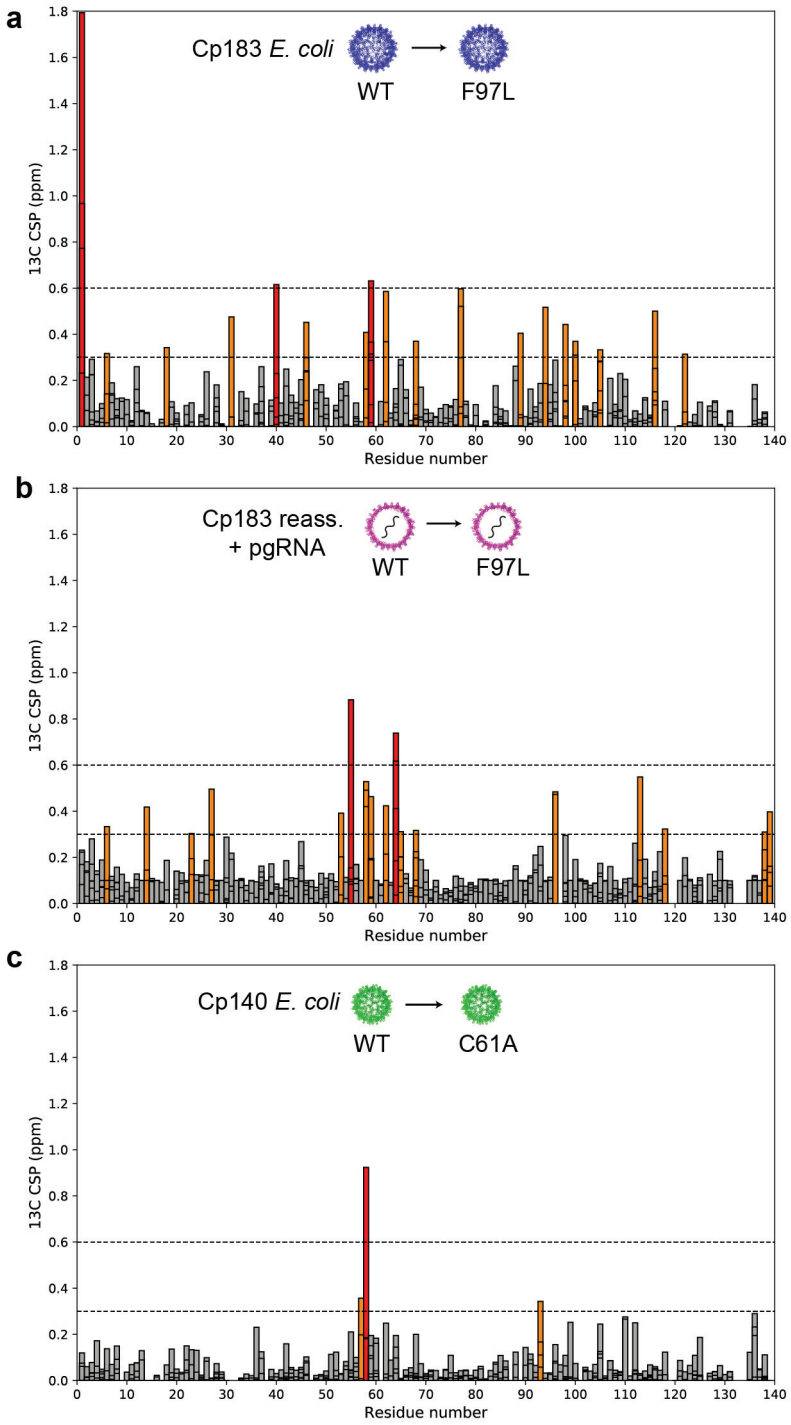


Figure S9. CSPs induced by the F97L mutation. CSPs are shown between Cp183 and the Cp183-F97L mutant for *E. coli* (**a**) and reassembled capsids (**b**). The impact of the mutation Cp140-C61A is shown in (**c**) to highlight how other mutations impact chemical shifts.

● 100 kHz MAS proton-detected NMR spectroscopy of hepatitis B virus capsids

The work in this part was done with:

Lauriane Lecoq[#], Maarten Schledorn[#], Shishan Wang, Susanne Smith-Penzel, Alexander A. Malär, Morgane Callon, Michael Nassal, Beat H. Meier[*], Anja Böckmann[*]

Contribution: Prepared the NMR samples, optimized the conditions for protein deuteration and contributed to writing the paper.

➢ Abstract

We sequentially assigned the fully-protonated capsids made from core proteins of the Hepatitis B virus using proton detection at 100 kHz magic-angle spinning (MAS) in 0.7 mm rotors and compare sensitivity and assignment completeness to previously obtained assignments using carbon-detection techniques in 3.2 mm rotors and 17.5 kHz MAS. We show that proton detection shows a global gain of a factor ~50 in mass sensitivity, but that signal-to-noise ratios and completeness of the assignment was somewhat higher for carbon-detected experiments for comparable experimental times. We also show that deuteration and H^N back protonation improves the proton linewidth at 100 kHz MAS by a factor of 1.5, from an average of 170 to 110 Hz, and by a factor of 1.3 compared to deuterated capsids at 60 kHz MAS in a 1.3 mm rotor. Yet, several H^N protons cannot be back-exchanged due to solvent inaccessibility, which results in a total of 15 % of the amides missing in the spectra.

➢ Introduction

Proton detection at 60 kHz magic-angle spinning (MAS) in 1.3 mm rotors in most cases requires protein deuteration (Andreas et al., 2015). This often sacrifices expression yields in bacteria, and also reduces the number of proteins to be studied to those for which exchangeable protons, most importantly amide protons, can to a large extent be back exchanged from ^2H to ^1H. Compared to 60 kHz, MAS frequencies of 100 kHz further average ^1H dipolar interactions by a factor of ~0.6 (Penzel et al., 2019), improving resolution in fully-protonated systems and allowing to resolve the resonances of small proteins (Cala-De Paepe et al., 2017; Lakomek et al., 2017; Schubeis et al., 2018). Still, comparison of ^1H linewidths on protonated and deuterated proteins has shown that deuteration improves resolution even at this frequency (Böckmann et al., 2015; Penzel et al., 2015; Xue et al., 2017).

The sequential assignment of protein resonances is a prerequisite in nuclear magnetic resonance (NMR) to further structural studies, and, as in ^{13}C-detected spectroscopy, extended sets of pulse sequences have been developed to enable backbone and sidechain assignments of protonated proteins around 100 kHz MAS (Higman, 2018; Stanek et al., 2016). Nevertheless, today, only a handful of fully-protonated proteins were assigned at these spinning frequencies, including the crystalline model protein GB1 (Andreas et al., 2016), a phage capsid assembly (Andreas et al., 2016), the prion domain of HET-s(218-289) (Smith et al., 2017; Stanek et al., 2016), a heptahelical transmembrane protein (Lalli et al., 2017), the HIV-1 capsid (Struppe et al., 2017), the extracellular domain of neonatal Fc receptor (Stöppler et al., 2018) and a carbonic anhydrase of 29 kDa (Vasa et al., 2019). A 26-mer model RNA was also assigned, demonstrating the potential applications for nucleic acids (Marchanka et al., 2018). Except for the phage capsid assembly (Andreas et al., 2016), the ^{13}C and ^{15}N chemical shifts were available beforehand from solution NMR or ^{13}C-detected solid-state NMR for these proteins, as also in the present study.

Here we investigate HBV nucleocapsids formed by multiple copies of the core protein (Cp). Cp forms stable dimers which self-assemble to form the capsid shell. The predominant form consists of 120 Cp dimers arranged in T = 4 icosahedral symmetry. The core protein has a 149-residue assembly domain (Birnbaum and Nassal, 1990; Gallina et al., 1989) named Cp149, and a 34-residue disordered nucleic acid binding C-terminal domain (Nassal, 1992). Residues 140 to 149 form a flexible linker between the N-terminal and C-terminal domains (Watts et al., 2002). For a T = 4 capsid, there are four unique subunit environments, A, B, C, and D, that are occupied by chemically identical but structurally distinct AB and CD dimers. We have recently sequentially assigned the ^{13}C and ^{15}N resonances of Cp149 capsids, which revealed narrow lines comparable to microcrystalline proteins (Lecoq et al., 2018b) or other viral capsids studied by NMR (Abramov et al., 2015; Han et al., 2010). In this work, we describe the proton-detected sequential assignments of the amide protons of the fully-protonated protein at 100 kHz, using a set of experiments which do in principle not rely on previous knowledge of ^{13}C and ^{15}N resonances. We compare the completeness of assignments to the ^{13}C equivalent spins, discuss the use of protonated pCp149 *versus* deuterated dCp149 samples for this protein, and compare spectra recorded on the deuterated

protein at 60 and 100 kHz, to conclude on the assets of each rotor sizes, namely 3.2 mm, 1.3 mm and 0.7 mm.

➢ Materials and methods

Sample preparation

Uniformly ^{13}C-^{15}N labeled fully-protonated pCp149 capsid samples were prepared as previously described in (Lecoq et al., 2018b). Preparation of the uniformly ^{2}H-^{13}C-^{15}N labeled dCp149 resulted in a slightly modified expression protocol. Cells from 10 ml inoculated LB medium were collected by centrifugation (4,000 *g*, 10 min, 20 °C) and resuspended in 100 ml of D_2O-based M9 minimal medium containing 1 g/l of $^{15}NH_4Cl$ and 2 g/l of deuterated ^{13}C-glucose as sole nitrogen and carbon sources, respectively. The deuterated culture was incubated overnight at 37 °C, and then transferred into 900 ml of fresh D_2O-based M9 medium. When the culture reached an OD_{600} of 2.0, protein expression was induced by adding 1 mM IPTG and cells were grown overnight at 25 °C. Protonated and deuterated capsids were purified as previously described in (Lecoq et al., 2018b), allowing the accessible labile deuterons in the deuterated sample to back-exchange to protons. For fast-MAS NMR measurements, capsids were dialyzed overnight at 4 °C in the solid-state-NMR buffer (50 mM TRIS pH 7.5, 5 mM DTT) and < 0.5 mg (Supplementary Table 3) of capsids were filled into 0.7 mm rotors using home-made filling tools (Böckmann et al., 2009) by centrifugation (200,000 *g*, 14h, 4°C). A minute amount of saturated to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) solution was added to the protein sediment before closing the rotor for chemical shift referencing.

NMR spectroscopy and data processing

The carbon-detected spectra used for comparison are described in (Lecoq et al., 2018b). Shortly, they were acquired on a wide-bore 800 MHz Bruker Avance II spectrometer equipped with a 3.2 mm triple-resonance MAS probe at 17.5 kHz MAS and a sample temperature of 4 °C, as determined from the relationship T (°C) = 455 –

90 * δ_{H2O} described in (Gottlieb et al., 1997), where the water chemical shift (δ_{H2O}) corresponds to the supernatant water signal (Böckmann et al., 2009).

The ^{1}H-detected spectra at 100-110 kHz were acquired on a wide-bore 850 MHz Bruker Avance III spectrometer equipped with a 0.7 mm triple-resonance MAS probe and referenced DSS. The magic angle was set using a 0.7 mm rotor with glycine ethyl ester, optimizing intensity and *J*-coupling-based splitting of the CO resonance (Penzel et al., 2018). The MAS frequency was set to 100 kHz and the VT gas temperature to 273 K using a nitrogen gas flow of 400 l/h, corresponding to a sample temperature of 22 °C, extrapolated from the water chemical shift in a ^{1}H 1D (Böckmann et al., 2009; Gottlieb et al., 1997) as detailed above. On the uniformly ^{13}C-^{15}N labeled fully-protonated sample, a set of four three-dimensional (3D) spectra (hCANH, hCONH, hCAcoNH, hncaCBcaNH) (Penzel et al., 2015) and one two-dimensional (2D) fingerprint spectrum (hNH) were recorded. For comparison, 2D hNH and hCANH spectra with virtually identical acquisition parameters (except for the MAS frequency which was set to 110 kHz in the hCANH) were recorded on dCp149 capsids. Acquisition parameters are given in Supplementary Table 1.

The 2D hNH and 3D hCANH spectra (Barbet-Massin et al., 2014; Penzel et al., 2015) ^{1}H-detected spectra at 60 kHz on dCp149 were acquired on a narrow-bore 700 MHz Bruker Avance II spectrometer equipped with a 1.3 mm triple-resonance MAS probe and referenced to DSS. The MAS frequency was set to 60 kHz and the VT gas temperature to 248 K using a nitrogen gas flow of 1400 l/h, corresponding to a sample temperature of 16 °C, extrapolated from the water chemical shift in a ^{1}H 1D (Böckmann et al., 2009; Gottlieb et al., 1997) as detailed above. Acquisition parameters are given in Supplementary Table 2.

All spectra were processed using TopSpin 4.0.3 (Bruker Biospin) by zero filling to no more than double the number of acquired points. Spectra were apodized in the direct and indirect dimensions with a shifted sine-bell window function (SSB = 3), except for the determination of linewidths, where no apodization was applied. Spectral analyses and resonance assignments were performed using CcpNmr Analysis 2.4.2 (Stevens et al., 2011; Vranken et al., 2005). Peak positions, linewidths and peak intensities were fitted using the parabolic fit function integrated in CcpNmr. The standard deviations of

the average linewidths were calculated by the square root of the difference between the individual linewidths minus their mean value, divided by the number of lines minus one.

For comparison of sensitivites of spectra from 3.2 mm (thin-wall), 1.3 mm and 0.7 mm rotors, we calculated the total mass by weighting the empty and full rotors (Supplementary Table 3).

The T_2' bulk relaxation times were measured using a Hahn-echo inserted after an hNH dipolar-coupling based polarization transfer sequence. The resulting proton-detected bulk amide signal has been recorded for different variable delay points. The peak area of each signal was extracted using TopSpin and then exported to MATLAB (MATLAB 2016R, The MathWorks Inc., Natick, MA 2016), where the relaxation series was fitted with a mono-exponential decay function. The fit error was derived using a bootstrapping approach.

Assignment deposition

$^{1}H^{N}$ chemical shifts of the fully-protonated ^{13}C-^{15}N-labeled Cp149 capsids at 100 kHz MAS were deposited in the Biological Magnetic Resonance Data Bank (BMRB) under accession number 27845. ^{15}N, $^{13}C'$, $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts, for which some resonances slightly differ from ^{13}C-detection based assignments (BMRB 27317 (Lecoq et al., 2018b)), were deposited as well.

➢ Results and discussion

1. Sequential assignments of the amide proton resonances

For amide proton assignments, the pCp149 capsid sample was used. Backbone atoms were assigned using four 3D spectra: hCANH showing intra-residue connections; hCONH and hCAcoNH showing inter-residue connections, and hncaCBcaNH connecting the Cβ (and Cα) to the NH. Representative extracts of the spectra are shown in Figure 1A. The four spectra allowed to assign 90 % of the H^{N} spins of the protein for residues 1–139 (the ten residues from the linker are excluded from the

statistic analysis as they are not visible in the NMR spectra). Using only the proton-detected spectra shown in Figure 1, 83 % of N (including prolines), 91 % of Cα, 23 % of Cβ and 80 % of C' could be assigned. Peak assignments are summarized as the mean chemical shift from the combined 3D experiments, and back-predicted on the 2D hNH spectrum in Figure 1B.

2. Comparison to ^{13}C-detection-based resonance assignments and signal-to-noise

We have recently assigned the ^{15}N and ^{13}C backbone and sidechain resonances of pCp149 capsids using ^{13}C-detected solid-state NMR at 17.5 kHz MAS in a 3.2 mm (thin-wall) rotor (Lecoq et al., 2018b). There, the assignment for residues 1-139 was complete at 97 % for N, C', Cα and Cβ atoms, and 76 % of side-chain heteronuclei. C-terminal residues 140 to 149 are excluded from the statistics as no cross peaks were detected in either ^{13}C- or ^{1}H-detected experiments. Except this region, only 4 residues are missing from the ^{13}C-detection-based assignment of HBV capsids. An additional 11 residues compared to the ^{13}C-detection based procedure could not be assigned due to low intensity or missing signals in the ^{1}H-detected spectra (L16, L19, L30, Q99, V120, S121, R127, P129, Y132, P138, I139, see Supplementary Figure 1). Thus, the ^{13}C-^{15}N backbone assignment completeness using ^{1}H-detection is slightly lower when compared to ^{13}C-detection NMR. However, the latter obviously completely lacks information on ^{1}H shifts, due to the strong proton line broadening at 17.5 kHz.

Experimental parameters used for proton and carbon-detection spectra are compared in Figure 2A. Measurement times are roughly the same between the two approaches for this protein (despite the difference in sample amount of about a factor of one hundred), with about 8-12 h used to record a good 2D fingerprint spectrum (C-C DARR or NCA for carbon-detection, and hNH for proton-detection), and about 2 weeks to record the set of 3D experiments used for the assignment. The reason for the lower assignment completeness in ^{1}H-detection lies mainly in the by a factor of about 2 lower signal-to-noise ratio (SNR) per square root time of the proton-detected spectra (Figure 2). This brings the divided intensity of some residues with peak doubling or quadrupling (*vide infra*) below the detection limit, which remained therefore unassigned. In detail, the SNR per square root time calculated on 2D experiments shows a gain by a factor

of 1.8 between hNH and DARR-CC spectra, and of 2.4 between hNH and NCA spectra, based on the intensities of 14 representative residues (Figure 2F). Corresponding 3D spectra show an SNR per square root time increase by a factor of 2.3 between the hCONH and the CANCO spectra, and almost a factor of 3 between the hCAcoNH and the NCOCX spectra for the $[C\alpha]_{i-1}$ correlation signals (Figure 2F). With respect to the latter, one should note that alternative schemes exist for the hCAcoNH experiment, where J-coupling based transfers (Barbet-Massin et al., 2013) could improve sensitivity under certain conditions. 3D NCACX spectra yield almost twice the SNR per square root time compared to hCANH spectra on the $[C\alpha]_i$ correlation signal, and it additionally provides the carbon sidechain shifts. Finally, the main difference is observed for the experiment leading to the Cβ assignment, where the NCACB shows better sensitivity than the hncaCBcaNH. Globally, the SNR per square root time is therefore still higher within similar experimental time when using 3.2 mm and carbon detection versus 0.7 mm and proton detection by a factor of about 2 for both 2D and 3D spectra. Still, especially when available sample amounts prove limiting, the gain in mass sensitivity clearly outweighs the loss in SNR for this protein. Indeed, with a total mass here of 0.56 ± 0.04 mg in a 0.7 mm rotor *versus* 55 ± 3 mg in a thin-wall 3.2 mm rotor (mass ratio ~100, Supplementary Table 3), and even when considering a loss of roughly a factor of two in SNR, the global gain in sensitivity remains ~50 in favor of the ^{1}H-detection approach using 0.7 mm rotors. In addition, we compared the SNR for dCp149 capsids at 60 kHz in a 1.3 mm rotor (Figure 2F). We found a similar sensitivity for the 3D hCANH, even if the 2D hNH reveals a better SNR compared to pCp149 capsids at 100 kHz. With a total mass of 4.3 ± 0.1 mg in the 1.3 mm rotor (mass ratio ~8 compared to 0.7 mm), there is therefore no clear advantage of using 1.3 mm rotors rather than 0.7 mm in terms of sensitivity for this protein.

One can ask whether the expected sensitivity effects could be predicted. Data, as well as predictions, related to sensitivity have been described in a recent publication (Mandala and Hong, 2019) where a gain using proton detection of a factor 1.5 was predicted for a 3D spectrum of microcrystalline SH3. As also detailed there, some parameters can be easily compared between the two approaches, as for example the rotor volume (roughly a factor 1/100, *vide supra*), the increased efficiency of the smaller coil which we estimated to scale like the inverse rotor diameters, namely 3.2/0.7 ≈ 4.6

(and 1.3/0.7) ≈ 1.8 (Webb, 1997; Samoson et al., 2010). A factor of 8 is gained by detecting protons instead of carbons. A comparison of these parameters results indeed in a factor around 2.5 in favor of carbon detection (for a simple 1D spectrum detected starting with proton equilibrium polarization). In addition, other factors need in principle be taken into account: for instance, the ratio of the product of the linewidth in all dimensions of the experiment (in Hz), which, as the proton linewidth is often comparable with the carbon linewidth, is on the order of one. Next, the efficiency of the probe circuits, the noise figure of the preamplifier, and efficiencies of the different polarization transfers used during the 2D and 3D experiments, also play a major role. As these values depend highly on the protein, temperature, spectrometer and/or probe parameters, it is therefore difficult to make good predictions for the relative SNR in proton-detected (0.7 mm) and carbon-detected (3.2 mm) experiments; but for full rotors and to a reasonable approximation they are roughly equal within a factor of two or three.

Finally, ^{1}H-detection has the crucial advantage to provide the amide-proton resonances, which are not only sensitive probes for conformational differences, but also for non-covalent interactions such as hydrogen bonds or ring-current effects. We conclude that in cases where enough sample is available, the carbon- and proton-detected approaches are truly complementary. Obviously, when sample amount is limiting, proton detection above 100 kHz is a must, and clearly shows competitive sensitivity.

3. Proton chemical shifts are more sensitive to detect capsid subunits

We have shown that the presence of the A, B, C, D subunits in the icosahedral HBV capsid causes peak splitting in carbon-detected spectra (Lecoq et al., 2018a). This behavior was observed for residues A11-T12, L16-D22, T33-S35, L108-F110, V115, S121-W125, R127-T128, P130-Y132, N136-P140, representing a total of 28 residues over 139 visible ones, *i.e.* about 20 % of the protein. In proton-detected spectra, NMR peak splitting was detected for globally the same residues, including A11-T12, D22, A34-S35, T109, V115, F122-V124, T128, A131, N136-A137, representing a total of 14 residues over 128 possibly visible NH correlations (prolines were removed), *i.e.* about

11 % of the protein. The congruency between the two approaches once more shows that the high sensitivity of NMR to detect asymmetric features in molecular assemblies applies for proton chemical shifts as well. The remaining 9 % for which one would also expect NMR peak splitting in the proton spectra are too close to the noise and could therefore not be detected.

For residues which chemical shifts are different for the four subunits, the proton's high sensitivity led to larger chemical shift differences between the 4 protein subunits. This is illustrated for residue A137 in Figure 2B, which shows extracts of the hNH and NCA spectra of pCp149 capsids, with the signals assigned to the 4 core protein subunits. The respective intensities of the four correlation peaks are displayed in Figure 2C, and show roughly the same distribution in both spectra. The four A137 signals equally show a higher SNR in the NCA than in the hNH when comparing two experiments with a similar experimental time, here about 8 hours (Figure 2D). The chemical shift perturbations (CSPs) between the peaks of the four different subunits reveal that, while Cα remains almost unaffected, H^N, N and Cβ nuclei are the most sensitive for A137, with ^1H showing CSPs up to nearly 1 ppm (Figure 2E). This emphasizes the complementarity of the three types of nuclei to measure chemical-shift perturbations as indicator in interaction studies.

4. Deuteration *versus* protonation: incomplete back-exchange *versus* increased proton linewidths

The choice between deuterated and protonated proteins for NMR studies using proton detection has been lately discussed in the literature (Cala-De Paepe et al., 2017; Linser, 2017; Schubeis et al., 2018; Xue et al., 2017). It has been suggested that MAS frequencies above 100 kHz present an important opening towards the use of fully-protonated proteins, since resolution starts to become high enough to resolve most resonances in 100-200 amino-acid proteins. This allows to bypass the more complex deuterated sample preparation and, even more importantly, back-exchange of amide protons, which can be difficult without a denaturation-renaturation step if they are not solvent accessible. Full protonation also represents an advantage to ease access to side chain resonances. A systematic linewidth comparison for protonated *vs.*

deuterated ubiquitin at 126 kHz MAS is given in (Penzel et al., 2019), and predicts significant differences in linewidths (> 80 Hz) even for this highest yet described spinning frequency, between protonated and deuterated proteins.

While here we have used pCp149 capsids for assignments, we wanted to assess which advantages, if any, can be obtained through the use of dCp149. The hNH spectra of the pCp149 and dCp149 capsids are shown in Figures 3A and B. A 1D extract of C48 H^N is shown as inset to illustrate the observed difference in proton linewidth on a single resonance. Figures 3C and D show the FIDs of the 2D spectra in the direct acquisition dimension, where the signal decay in the proton dimension can be seen. The signal decay is clearly shorter in the protonated sample than in the deuterated one. In order to quantify, we measured bulk T_2' relaxation times, which resulted in 2.5 ± 0.1 ms for the protonated sample and 11.6 ± 0.2 ms for the deuterated sample (Supplementary Figure 2). This shows indeed that the homogeneous contribution to the linewidth is by a factor of 4.6 smaller in dCp149 than in pCp149, as is expected by dilution of the strongly coupled proton network with deuterium. One can note that while the proton linewidths are smaller for dCp149, the nitrogen lines appear broader compared to pCp149 by an average factor of 1.2. The same value was found with and without ^{13}C decoupling during the indirect acquisition dimension.

We also measured a 2D hNH of dCp149 at 60 kHz in a 1.3 mm rotor for comparison, which is shown in Supplementary Figure 3A. The proton linewidths of the three samples were measured for the assigned residues in the 3D hCANH spectra at 60 kHz and 100 kHz for dCp149, and 110 kHz for pCp149. The resulting linewidths are shown in Figures 3E-G and Supplementary Figure 3B-C. The mean proton linewidth was determined to 170 ± 80 Hz for pCp149 at 100 kHz, 110 ± 50 Hz for dCp149 at 110 kHz, and 130 ± 60 Hz for dCp149 at 60 kHz. This results in an average factor of 1.5 between the linewidths of pCp149 and dCp149 at 100-110 kHz, and 1.3 between dCp149 in 0.7 mm rotor at 100 kHz and dCp149 in 1.3 mm rotor at 60 kHz. The improvement of 20 Hz between proton linewidths measured on dCp149 at 110 kHz compared to 60 kHz is close to values obtained for deuterated ubiquitin (improvement of 13 Hz in (Penzel et al., 2019)) and for deuterated GB1 (improvement of only 6 Hz in (Cala-De Paepe et al., 2017)).

Despite the better resolution of dCp149 vs pCp149 at MAS ≥ 100 kHz, it should be mentioned that 15 % of the assigned residues are completely missing from the 2D and 3D spectra of the deuterated sample dCp149 due to incomplete back-exchange (Figure 4A and Supplementary Figure 3C). In particular, mainly two regions are affected, namely A54 to T62 and L101 to L108, both located at the base of the spike where the bottoms of helices α3 and α4 interact in the dimer. These residues are buried and thus protected from the solvent, as shown in Figure 4B. Only the sidechains of I59, W62 and I116 are accessible, while the H^N remain inaccessible. In HBV capsids, the advantage of deuteration obtained in linewidths must therefore be weighed against the disadvantage of incomplete back exchange, which obscures information on residues at the center of the capsid spike.

➢ Conclusions

We have presented here sequential assignments of the amide proton resonances of the fully- protonated HBV capsid using 100 kHz fast MAS solid-state NMR methods. Sequential assignments were obtained from a set of four spectra recorded within 15 days, and yielded 85 % of backbone assignments. We compared this assignment to the $^{13}C/^{15}N$ assignments obtained previously and found that the ^{13}C-detection method led to ~10 % more ^{13}C-^{15}N assigned residues compared to ^{1}H-detection, with the additional advantage that it enabled sidechains assignment using a similar experimental time investment. We have compared the sensitivity of the amide proton resonances to conformational variations as imposed by the icosahedral symmetry, and found that ^{1}H chemical shifts are a valuable and sensitive addition to the analysis of $^{13}C/^{15}N$ chemical shifts. We showed that, while for the 149-residue HBV core protein a protonated sample allows for assignment of the amide protons, the spectral resolution is still clearly better in a deuterated sample, at both 60 and 100 kHz MAS. This gain in resolution is however concomitant with the loss of a set of amide resonances due to incomplete back-exchange. The success of amide-proton assignments shall open the way for the analysis of the HBV capsid in the presence of partner molecules, notably those only available in small quantities.

Author contributions

MS, SSP, LL, AM, and MC conducted the NMR experiments. SW and LL generated protein samples. LL, MS and AM analyzed the data. MN designed the plasmid and expression/purification protocols. LL, BHM and AB designed and supervised the study, and wrote the manuscript. All authors contributed to the manuscript and approved the submitted version.

Funding

This work was supported by the French ANR (ANR- 14-CE09-0024B), the LABEX ECOFECT (ANR-11-LABX-0048) within the Université de Lyon program Investissements d'Avenir (ANR-11-IDEX-0007), by the Marie Skłodowska-Curie program (H2020-MSCA-IF-2016 748516) by the the Swiss National Science Foundation (Grant 200020_159707), and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement n° 741863, FASTER).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

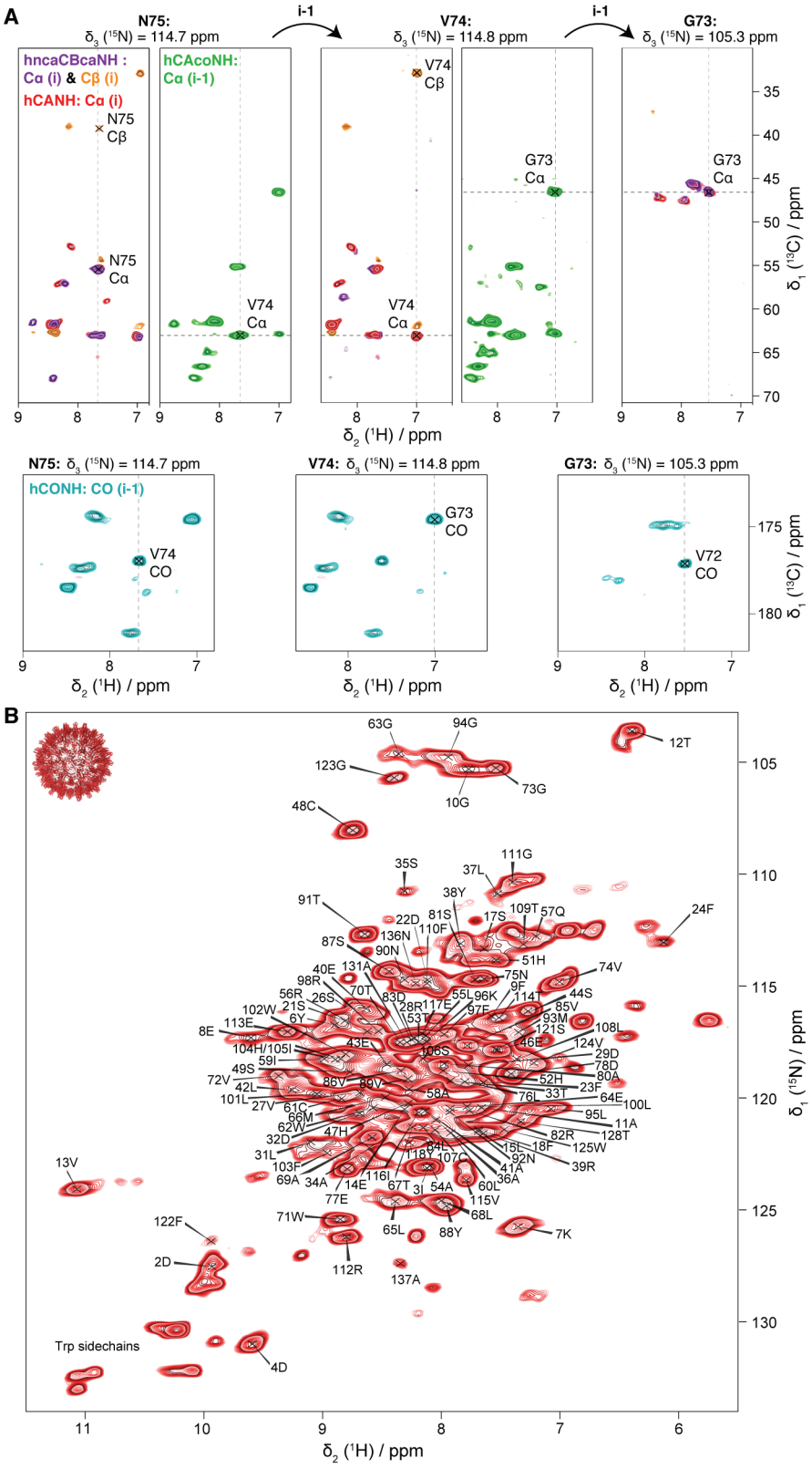
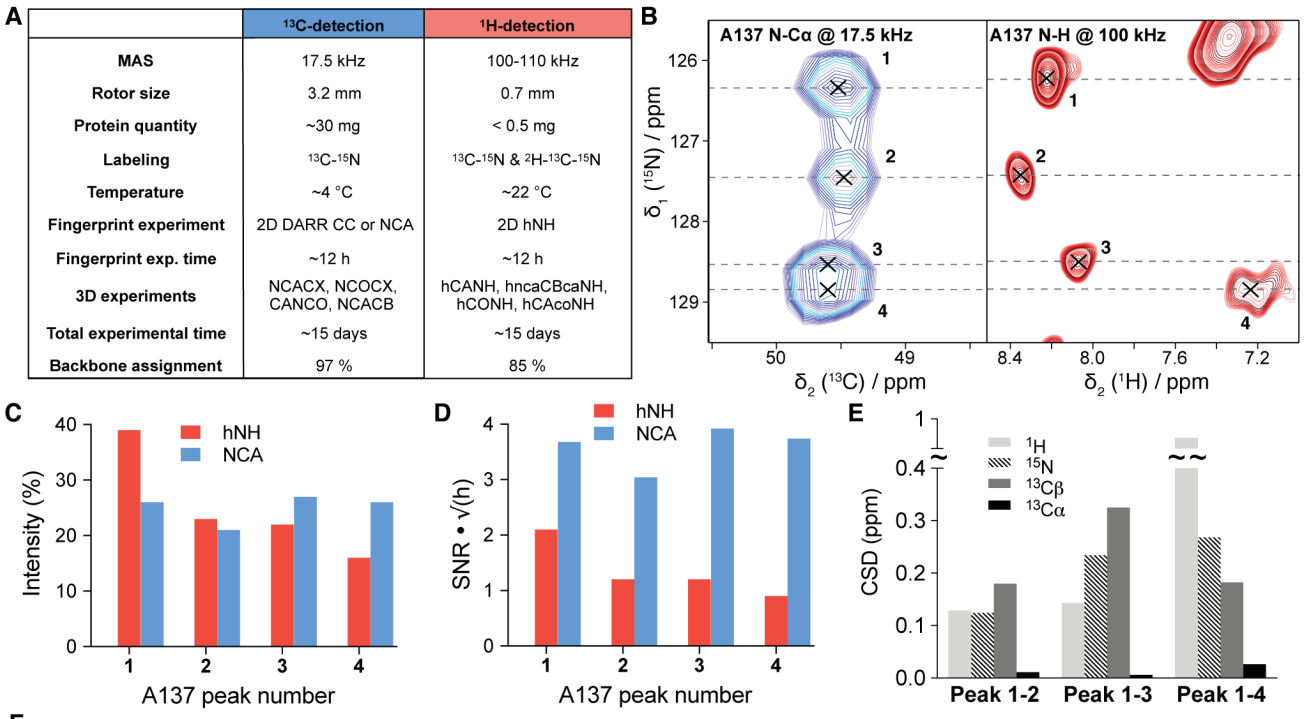


Fig. 1 (A) 3D planes of pCp149 capsids spectra showing a sequential walk for residues N75, V74 and G73. Four spectra were recorded: hncaCBcaNH in purple (positive peaks: Cα) and orange

(negative peaks: Cβ); hCANH in red (Cα); hCAcoNH in green (Cα of the preceding residue); and hCONH in cyan (CO of the preceding residue). **(B)** Assigned hNH spectrum of pCp149 capsids, with resonance frequencies deposited in the BMRB under accession number 27845. The spectrum was processed with QSINE 3 in both dimensions.

A

	^{13}C-detection	^{1}H-detection
MAS	17.5 kHz	100-110 kHz
Rotor size	3.2 mm	0.7 mm
Protein quantity	~30 mg	< 0.5 mg
Labeling	^{13}C-^{15}N	^{13}C-^{15}N & ^{2}H-^{13}C-^{15}N
Temperature	~4 °C	~22 °C
Fingerprint experiment	2D DARR CC or NCA	2D hNH
Fingerprint exp. time	~12 h	~12 h
3D experiments	NCACX, NCOCX, CANCO, NCACB	hCANH, hncaCBcaNH, hCONH, hCAcoNH
Total experimental time	~15 days	~15 days
Backbone assignment	97 %	85 %



F

^{13}C-detection				
Protonated Cp149 @ 17.5kHz	Spectrum	Exp time (h)	SNR	SNR/√(h)
	2D NCA	8	28.2	**10.0**
	2D DARR	15	28.7	**7.4**
	3D NCACB	60	27.9	**3.6**
	3D CANCO	92	13.0	**1.4**
	3D NCACX	72	11.2	**1.3**
	3D NCOCX	141	10.4	**0.9**

^{1}H-detection							
Protonated Cp149 @ 100 kHz				*Deuterated Cp149 @ 110 kHz*			
Spectrum	Exp time (h)	SNR	SNR/√(h)	Spectrum	Exp time (h)	SNR	SNR/√(h)
2D hNH	12	14.5	**4.2**	2D hNH	12	17.3	**5.0**
3D hCANH	22	3.8	**0.8**	3D hCANH	23	5.4	**1.1**
3D hCONH	66	4.8	**0.6**	*Deuterated Cp149 @ 60 kHz*			
3D hCAcoNH	130	3.4	**0.3**	2D hNH	11.2	26.3	**7.8**
3D hncaCBcaN	142	2.5	**0.2**	3D hCANH	23	3.8	**0.8**

Fig. 2 (A) Comparison of acquisition parameters and assignments of Cp149 capsids (residues 1-139) using ^{13}C-detected experiments recorded in a 3.2 mm rotor (thin-wall) at 17.5 kHz MAS, and ^{1}H-detected experiments recorded in a 0.7 mm rotor at 100 kHz MAS. The backbone assignment percentages include ^{15}N, ^{13}Cα and ^{13}C' resonances. **(B)** Extract of hNH (in red) and NCA (in blue) spectra showing peak splitting for residue A137, which displays four resonances corresponding to the four subunits of the pCp149 capsid. Both 2D spectra were recorded in ~8 h, with 32 scans for the hNH and 8 scans for the NCA. The NCA extract is adapted from previous data (Lecoq et al., 2018b). **(C)** Relative intensities of the four signals of A137 in the hNH and NCA spectra. **(D)** Signal to noise ratio of the four signals of A137 in the hNH and NCA spectra extracted by dividing the

peak's height by the noise rmsd and by the square root of the experimental time in hours. A factor of 1.0625 was applied to the NCA spectrum to take into account the field difference (850 MHz for proton-detection *versus* 800 MHz for carbon-detection). **(E)** ^{1}H, ^{15}N, ^{13}Cα and ^{13}Cβ CSPs between the four subunit peaks of A137, with $CSD = \left(\frac{\gamma_X}{\gamma_H}\right)|\Delta\delta_X|$ (x, nucleus of interest; Δδ, chemical shift difference between the peak of interest and the reference peak, here number 1). **(F)** Signal-to-noise ratio (SNR) divided by the square root of the experimental time in hours for ^{13}C-detected and ^{1}H-detected experiments. SNR was calculated based on the intensities of 14 representative and isolated peaks (D2, D4, G10, T12, V13, D22, E46, C48, W71, D78, T91, C107, G111 and I116) and compared to the global noise of each individual spectrum as estimated in CcpNmr (Vranken et al., 2005). For the DARR, the probe was set to double-resonance mode (^{15}N-insert was removed). All carbon-detected spectra were recorded on the same thin-wall 3.2 mm rotor as described in (Lecoq et al., 2018b). All proton-detected spectra of pCp149 and dCp149 were recorded on the same rotors. The four rotors were full. A factor of 1.0625 was applied to carbon-detected spectra and a factor of 1.214 to proton-detected spectra at 60 kHz to take into account the field difference (850 MHz for proton-detection at 100-110 kHz *versus* 800 MHz for carbon-detection and 700 MHz for proton-detection at 60 kHz). For 3D spectra with multiple correlation peaks, the intensity of the following resonances was taken into account for the SNR calculation: H-N-Cα for the hncaCBcaNH, N-Cα-Cβ for the NCACX (except for Glycines: N-Cα-C'), and N-[C'-Cα]i-1 for the NCOCX. For the deuterated samples, C107 and I116 were excluded from the calculation as they were not back-exchanged.

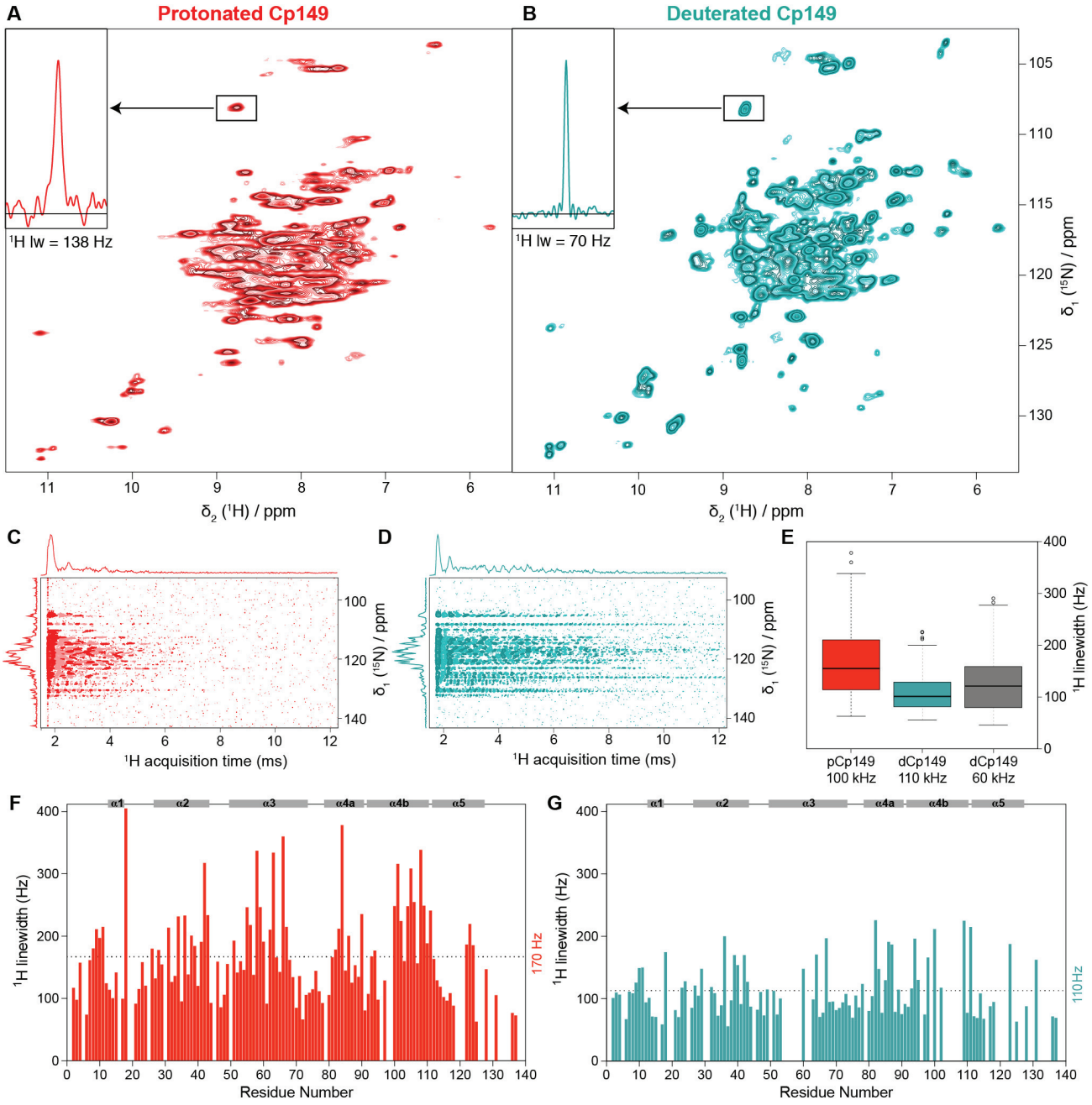


Fig. 3 hNH spectra of **(A)** pCp149, and **(B)** dCp149 capsids. The proton line of C48 is shown in the inserts. Both spectra were recorded with identical acquisition parameters (100 kHz MAS, 40 scans, 12 hours total experimental time, VTU 273 K, and 25 ms acquisition in the ^{15}N dimension) and processing parameters (with no apodization function, cut at 12.9 ms acquisition in ^{1}H dimension and zero-filled to 4096 points and 1024 points in ^{1}H and ^{15}N dimensions, respectively). (**C-D**) FIDs from hNH spectra processed in the time domain using the xf1 command in TopSpin for Cp149 protonated (red) and deuterated capsids (cyan). (**E**) Dispersion of total proton linewidths measured on all assigned residues in 3D hCANH spectra of pCp149 at 100 kHz (red) and dCp149 capsids at 110 kHz (cyan) and 60 kHz (grey). Median values are indicated as a black line within each box and outliers as circles, defined as exceeding 1.5 times the interquartile range above the third or below

the first quartile. (**F-G)** Total proton linewidths for assigned residues observed in the 3D hCANH spectra of pCp149 (red) and dCp149 capsids (cyan) using parabolic fit in CcpNmr (Vranken et al., 2005). Both 3D spectra were run with a similar experimental time of about 22 h and were processed identically (with no apodization function, cut at 12.9 ms acquisition in ^{1}H dimension and zero-filled to 2048 points in ^{1}H dimension and 128 points in ^{13}C and ^{15}N dimensions). The average linewidth value for both samples is indicated as dotted lines and secondary structure elements are identified on the top of each graph.

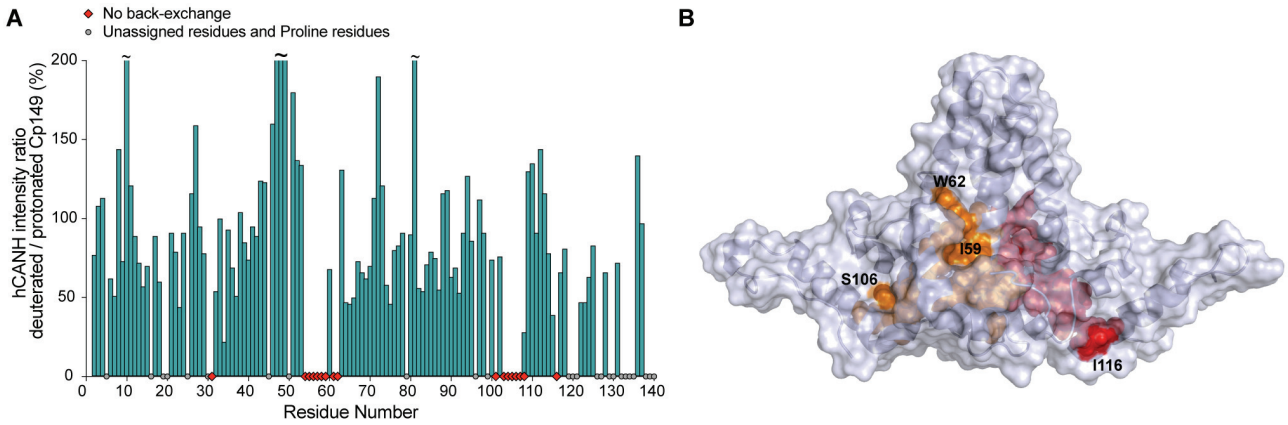


Fig. 4 (A) Intensity ratio of NMR signals in the 3D hCANH spectra of back-exchanged deuterated Cp149 capsids *versus* protonated Cp149 capsids. Unassigned and proline residues are shown as grey circles and residues with no signal in the deuterated sample are shown as red diamonds. The noise was adjusted to the same level for both spectra in CcpNmr before fitting the peaks intensity and no normalization was applied. **(B)** Residues whose NH could not be back-exchanged in the deuterated capsids are shown in orange for chain C and in red for chain D on the X-ray structure 1QGT (Wynne et al., 1999).

➢ References

Abramov, G., Morag, O., and Goldbourt, A. (2015). Magic-angle spinning NMR of intact bacteriophages: Insights into the capsid, DNA and their interface. *J. Magn. Reson.* 253, 80–90. doi:10.1016/j.jmr.2015.01.011.

Andreas, L. B., Jaudzems, K., Stanek, J., Lalli, D., Bertarello, A., Le Marchand, T., et al. (2016). Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9187–9192. doi:10.1073/pnas.1602248113.

Andreas, L. B., Le Marchand, T., Jaudzems, K., and Pintacuda, G. (2015). High-resolution proton-detected NMR of proteins at very fast MAS. *J. Magn. Reson.* 253, 36–49. doi:10.1016/j.jmr.2015.01.003.

Barbet-Massin, E., Pell, A. J., Jaudzems, K., Franks, W. T., Retel, J. S., Kotelovica, S., et al. (2013). Out-and-back ^{13}C-^{13}C scalar transfers in protein resonance assignment by proton-detected solid-state NMR under ultra-fast MAS. *J. Biomol. NMR* 56, 379–386. doi:10.1007/s10858-013-9757-3.

Barbet-Massin, E., Pell, A. J., Retel, J. S., Andreas, L. B., Jaudzems, K., Franks, W. T., et al. (2014). Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J. Am. Chem. Soc.* 136, 12489–12497. doi:10.1021/ja507382j.

Birnbaum, F., and Nassal, M. (1990). Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. *J. Virol.* 64, 3319–3330.

Böckmann, A., Ernst, M., and Meier, B. H. (2015). Spinning proteins, the faster, the better? *J. Magn. Reson.* 253, 71–79. doi:10.1016/j.jmr.2015.01.012.

Böckmann, A., Gardiennet, C., Verel, R., Hunkeler, A., Loquet, A., Pintacuda, G., et al. (2009). Characterization of different water pools in solid-state NMR protein samples. *J. Biomol. NMR* 45, 319–327. doi:10.1007/s10858-009-9374-3.

Cala-De Paepe, D., Stanek, J., Jaudzems, K., Tars, K., Andreas, L. B., and Pintacuda, G. (2017). Is protein deuteration beneficial for proton detected solid-state NMR at and above 100 kHz magic-angle spinning? *Solid State Nuclear Magnetic Resonance* 87, 126–136. doi:10.1016/j.ssnmr.2017.07.004.

Gallina, A., Bonelli, F., Zentilin, L., Rindi, G., Muttini, M., and Milanesi, G. (1989). A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. *J. Virol.* 63, 4645–4652.

Gottlieb, H. E., Kotlyar, V., and Nudelman, A. (1997). NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J. Org. Chem.* 62, 7512–7515.

Han, Y., Ahn, J., Concel, J., Byeon, I.-J. L., Gronenborn, A. M., Yang, J., et al. (2010). Solid-State NMR Studies of HIV-1 Capsid Protein Assemblies. *J. Am. Chem. Soc.* 132, 1976–1987. doi:10.1021/ja908687k.

Higman, V. A. (2018). Solid-state MAS NMR resonance assignment methods for proteins. *Prog. Nucl. Mag. Res. Sp.* 106-107, 37–65. doi:10.1016/j.pnmrs.2018.04.002.

Lakomek, N.-A., Penzel, S., Lends, A., Cadalbert, R., Ernst, M., and Meier, B. H. (2017). Microsecond Dynamics in Ubiquitin Probed by Solid-State 15 N NMR Spectroscopy R1ρ Relaxation Experiments under Fast MAS (60-110 kHz). *Chemistry* 23, 9425–9433. doi:10.1002/chem.201701738.

Lalli, D., Idso, M. N., Andreas, L. B., Hussain, S., Baxter, N., Han, S., et al. (2017). Proton-based structural analysis of a heptahelical transmembrane protein in lipid bilayers. *J. Am. Chem. Soc.*, jacs.7b05269–9. doi:10.1021/jacs.7b05269.

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

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

Linser, R. (2017). Solid-state NMR spectroscopic trends for supramolecular assemblies and protein aggregates. *Solid State Nuclear Magnetic Resonance* 87, 45–53. doi:10.1016/j.ssnmr.2017.08.003.

Mandala, V. S., and Hong, M. (2019). High-sensitivity protein solid-state NMR spectroscopy. *Current Opinion in Structural Biology*. doi:10.1016/j.sbi.2019.03.027.

Marchanka, A., Stanek, J., Pintacuda, G., and Carlomagno, T. (2018). Rapid access to RNA resonances by proton-detected solid-state NMR at >100 kHz MAS. *Chemical Communications* 54, 8972–8975. doi:10.1039/C8CC04437F.

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. *J. Virol.* 66, 4107–4116.

Penzel, S., Oss, A., Org, M.-L., Samoson, A., Böckmann, A., Ernst, M., et al. (2019). Spinning faster: protein NMR at MAS frequencies up to 126 kHz. *J. Biomol. NMR* 128, 12620–11. doi:10.1007/s10858-018-0219-9.

Penzel, S., Smith, A. A., Agarwal, V., Hunkeler, A., Org, M.-L., Samoson, A., et al. (2015). Protein resonance assignment at MAS frequencies approaching 100 kHz: a quantitative comparison of J-coupling and dipolar- coupling-based transfer methods. *J. Biomol. NMR*, 1–22. doi:10.1007/s10858-015-9975-y.

Penzel, S., Smith, A. A., Ernst, M., and Meier, B. H. (2018). Setting the magic angle for fast magic-angle spinning probes. *J. Magn. Reson.* 293, 115–122. doi:10.1016/j.jmr.2018.06.002.

Samoson, A., Tuherm, T., Past, J., Reinhold, A., Heinmaa, I., Anupõld, T., et al. (2010). *Fast Magic-Angle Spinning: Implications*. ed.John Wiley Sons, Ltd Encyclopedia of Magnetic Resonance doi:10.1002/9780470034590.emrstm1017.

Schubeis, T., Le Marchand, T., Andreas, L. B., and Pintacuda, G. (2018). ^{1}H magic-angle spinning NMR evolves as a powerful new tool for membrane proteins. *J. Magn. Reson.* 287, 140–152. doi:10.1016/j.jmr.2017.11.014.

Smith, A. A., Ravotti, F., Testori, E., Cadalbert, R., Ernst, M., Böckmann, A., et al. (2017). Partially-deuterated samples of HET-s(218-289) fibrils: assignment and deuterium isotope effect. *J. Biomol. NMR* 67, 109–119. doi:10.1007/s10858-016-0087-0.

Stanek, J., Andreas, L. B., Jaudzems, K., Cala, D., Lalli, D., Bertarello, A., et al. (2016). NMR Spectroscopic Assignment of Backbone and Side-Chain Protons in Fully Protonated Proteins: Microcrystals, Sedimented Assemblies, and Amyloid Fibrils. *Angew. Chem. Int. Ed.* 128, 15730–15735. doi:10.1002/ange.201607084.

Stevens, T. J., Fogh, R. H., Boucher, W., Higman, V. A., Eisenmenger, F., Bardiaux, B., et al. (2011). A software framework for analysing solid-state MAS NMR data. *J. Biomol. NMR* 51, 437–447. doi:10.1007/s10858-011-9569-2.

Stöppler, D., Macpherson, A., Smith-Penzel, S., Basse, N., Lecomte, F., Deboves, H., et al. (2018). Insight into small molecule binding to the neonatal Fc receptor by X-ray crystallography and 100 kHz magic-angle-spinning NMR. *PLoS Biol* 16, e2006192–27. doi:10.1371/journal.pbio.2006192.

Struppe, J., Quinn, C. M., Lu, M., Wang, M., Hou, G., Lu, X., et al. (2017). Expanding the horizons for structural analysis of fully protonated protein assemblies by NMR spectroscopy at MAS frequencies above 100 kHz. *Solid State Nuclear Magnetic Resonance* 87, 117–125. doi:10.1016/j.ssnmr.2017.07.001.

Vasa, S. K., Singh, H., Grohe, K., and Linser, R. (2019). Assessment of a Large Enzyme-Drug Complex by Proton-Detected Solid-State NMR Spectroscopy without Deuteration. *Angew. Chem. Int. Ed.* doi:10.1002/anie.201811714.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., et al. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687–696. doi:10.1002/prot.20449.

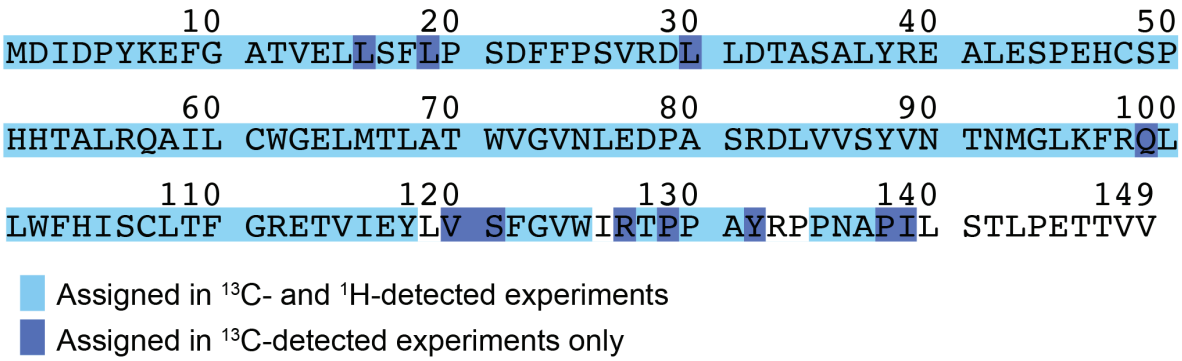
Watts, N. R., Conway, J. F., Cheng, N., Stahl, S. J., Belnap, D. M., Steven, A. C., et al. (2002). The morphogenic linker peptide of HBV capsid protein forms a mobile array on the interior surface. *EMBO J.* 21, 876–884. doi:10.1093/emboj/21.5.876.

Webb, A. G. (1997). Radiofrequency microcoils in magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *31*, 1–42.

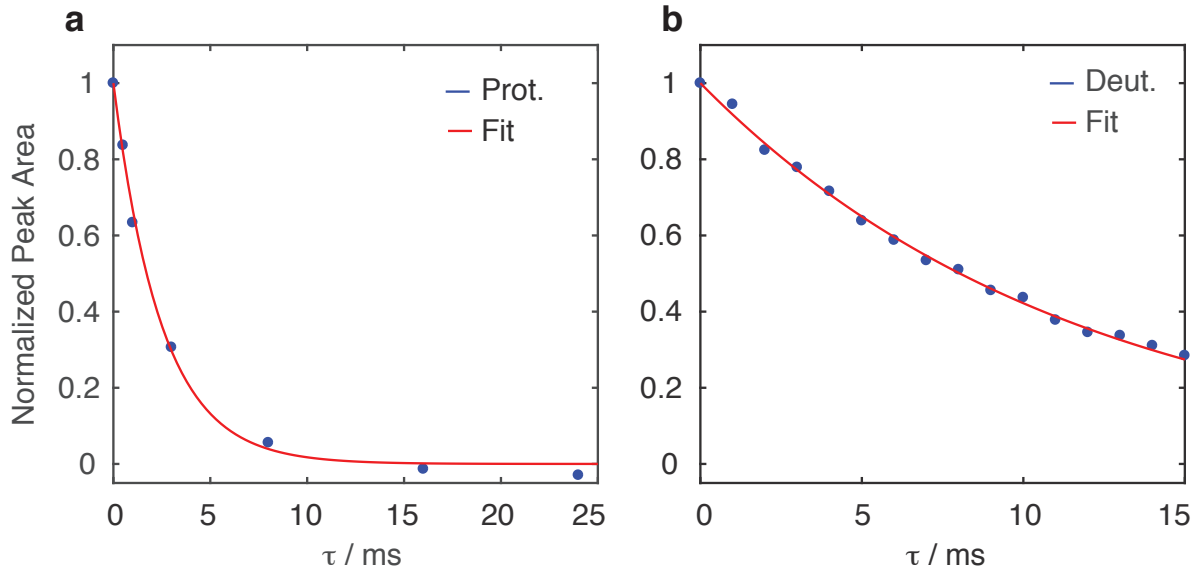
Wynne, S. A., Crowther, R. A., and Leslie, A. G. (1999). The crystal structure of the human hepatitis B virus capsid. *Mol. Cell* 3, 771–780.

Xue, K., Sarkar, R., Motz, C., Asami, S., Camargo, D. C. R., Decker, V., et al. (2017). Limits of Resolution and Sensitivity of Proton Detected MAS Solid-State NMR Experiments at 111 kHz in Deuterated and Protonated Proteins. *Sci. Rep.*, 1–7. doi:10.1038/s41598-017-07253-1.

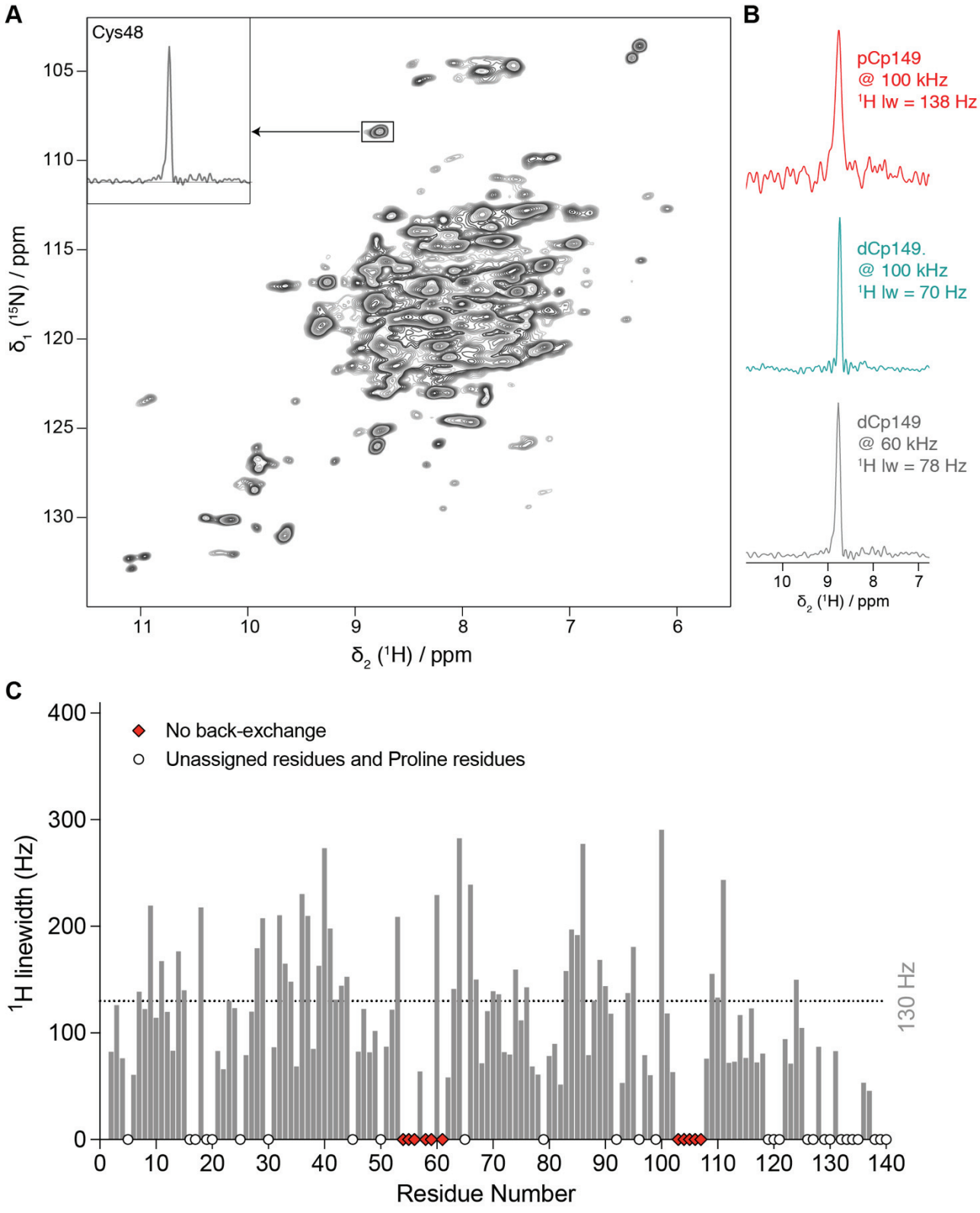
➢ Supplementary Material



Supplementary Figure 1. Cp149 sequence showing the residues assigned in both ^{13}C- and ^{1}H-detected experiments (light blue) and in ^{13}C-detected experiments only (dark blue). Residues with at least 2 assigned atoms are colored, including prolines in ^{1}H-detected experiments for which carbon resonances could be assigned from the hCONH and hCAcoNH 3D spectra. Unassigned residues are not colored.



Supplementary Figure 2. T2' decay for the **a)** pCp149 capsids, fitted to a value of $T_2' = 2.5 \pm 0.1$ ms, and for **b)** dCp149 capsids, fitted to a value of 11.6 ± 0.2 ms. Note that these values give the homogeneous, but not the full contribution to the signal decay for the two samples, as depicted in the main text in Figure 3c/d.



Supplementary Figure 3. (A) hNH spectrum recorded in a 1.3 mm rotor of dCp149 capsids, with the proton linewidth of the isolated C48 residue shown in the insert. The spectrum was recorded at 60 kHz spinning frequency at a field of 700 MHz with 80 scans, with similar acquisition and processing parameters as data recorded at 100 kHz MAS (11.2 hours total experimental time, 25 ms acquisition in

the ^{15}N dimension, processed with no apodization function, cut at 12.9 ms acquisition in ^{1}H dimension and zero-filled to 4096 points and 1024 points in ^{1}H and ^{15}N dimensions, respectively). **(B)** Comparison of C48 proton linewidths for pCp149 at 100 kHz MAS (in red), dCp149 at 110 kHz (in cyan) and at 60 kHz MAS (in grey). **(C)** Total proton linewidths for assigned residues observed in the 3D hCANH spectra of dCp149 capsids at 60 kHz MAS using parabolic fit in CcpNmr (Vranken et al., 2005). The 3D hCANH spectrum was run with a similar experimental time (22 h) as the hCANH spectra recorded at 100-110 kHz, and was processed identically (with no apodization function, cut at 12.9 ms acquisition in ^{1}H dimension and zero-filled to 2048 points in ^{1}H dimension and 128 points in ^{13}C and ^{15}N dimensions). The average linewidth value is indicated as dotted lines. Unassigned and proline residues are indicated as white circles, and residues with no back-exchange (invisible in the 3D spectrum) as red diamonds.

Supplementary Table 1: NMR parameters of 2D and 3D spectra recorded on protonated Cp149 capsids and deuterated Cp149 capsids at 100 kHz and 110 kHz.

			pCp149			dCp149	
Experiment	**CP hNH 2D**	**hCANH 3D**	**hCONH 3D**	**hCAcoNH 3D**	**hncaCBcaNH 3D**	**CP hNH 2D**	**hCANH 3D**
MAS frequency / kHz	100	100	100	100	100	100	110
Field / T	20.0	20.0	20.0	20.0	20.0	20.0	20.0
t1 increments	776	64	42	82	112	776	64
Sweep width (t1) / ppm	180	30	16	30	55	180	30
Acquisition time (t1) / ms	25	5	6.1	6.4	4.8	25	5
t2 increments	1024	38	38	48	64	1024	43
Sweep width (t2) / ppm	46.7	35	35	35	35	46.7	40
Acquisition time (t2) / ms	13	6.3	6.3	8.0	10.6	13	6.2
t3 increments	-	1024	1024	1024	1024	-	1024
Sweep width (t3) / ppm	-	46.7	46.7	46.7	46.7	-	46.7
Acquisition time (t3) / ms	-	12.9	12.9	12.9	12.9	-	12.9
^{1}H dec. (swfTPPM) / kHz	10	10	10	10	10	10	10
^{15}N dec. (WALTZ64) / kHz	5	5	5	5	5	5	5
Water sup. (120 ms) / kHz	20	20	20	20	20	20	20
Interscan delay / s	1.2	2.0	1.0	1.3	1.3	1.2	1.1
Number of scans	40	16	128	80	48	40	24
Experiment time	**12 h**	**23 h**	**2 days 18 h**	**5 days 10 h**	**5 days 22 h**	**12 h**	**23 h**
Transfer I	**HN (dipolar)**	**HC (dipolar)**	**HC (dipolar)**	**HC (dipolar)**	**HN (dipolar)**	**HN (dipolar)**	**HC (dipolar)**
^{1}H field / kHz	78	72	80	77	79	69	76
X field / kHz	17	16	15	15	15	15	16.5
Shape	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H
Carrier ^{13}C / ppm $^{(*)}$	-	56	174	56	-	-	51
Time / ms	1.0	1.2	0.8	1.0	1.0	2.5	4.4
Transfer II	**NH (dipolar)**	**CN (dipolar)**	**CN (dipolar)**	**CC (DREAM)**	**NCA (scalar)**	**NH (dipolar)**	**CN (dipolar)**
^{1}H field / kHz	78	-	-	-	-	69	-
^{13}C field / kHz	-	71	66	49	-	-	63
^{15}N field / kHz	17	25	35	-	-	15	36
Shape	Tangent ^{1}H	Tangent ^{13}C	Tangent ^{13}C	Tangent ^{13}C	-	Tangent ^{1}H	Tangent ^{13}C
Carrier ^{13}C / ppm	-	56	174	130	42	4.8	51
Time / ms	1.0	10.0	20.0	9.0	22	2.5	18.0
Transfer III		**NH (dipolar)**	**NH (dipolar)**	**CN (dipolar)**	**CACB (scalar)**		**NH (dipolar)**
^{1}H field / kHz		74	84	-	-		74
^{13}C field / kHz		-	-	60	-		-
^{15}N field / kHz		15	15	36	-		16
Shape		Tangent ^{1}H	Tangent ^{1}H	Tangent ^{13}C	-		Tangent ^{1}H
Carrier ^{13}C / ppm		-	-	174	42		-
Time / ms		1.0	1.6	12.0	7		2.5
Transfer IV				**NH (dipolar)**	**CBCA (scalar)**		
^{1}H field / kHz				77	-		
^{15}N field / kHz				15	-		
Shape				Tangent ^{1}H	-		
Carrier ^{13}C / ppm				-	42		
Time / ms				1.2	7		
Transfer V					**CN (scalar)**		
Carrier ^{13}C / ppm					42		
Time / ms					22		
Transfer VI					**NH (dipolar)**		
^{1}H field / kHz					79		
^{15}N field / kHz					15		
Shape					Tangent ^{1}H		
Time / ms					1.0		

$^{(*)}$ Carriers for ^{1}H and ^{15}N were 4.8 and 117.5 ppm respectively, for all experiments.

Supplementary Table 2: NMR parameters of 2D and 3D spectra recorded on deuterated and back-exchanged Cp149 capsids at 60 kHz.

	dCp149	
Experiment	**CP hNH 2D**	**hCANH 3D**
MAS frequency / kHz	60	60
Field / T	16.4	16.4
t1 increments	248	58
Sweep width (t1) / ppm	70	30
Acquisition time (t1) / ms	25	5.4
t2 increments	1800	40
Sweep width (t2) / ppm	99.2	40
Acquisition time (t2) / ms	13	7
t3 increments	-	1800
Sweep width (t3) / ppm	-	99.2
Acquisition time (t3) / ms	-	13
^{1}H dec. (WALTZ16) / kHz	10	10
^{15}N dec. (WALTZ16) / kHz	5	9
^{13}C dec. (WALTZ16) / kHz	10	10
Water sup. (100 ms) / kHz	15	15
Interscan delay / s	2	1
Number of scans	80	32
Experiment time	**11.2 h**	**23 h**
Transfer I	**HN (dipolar)**	**HC (dipolar)**
^{1}H field / kHz	100	85
X field / kHz	40	50
Shape	Ramp ^{1}H	Ramp ^{1}H
Carrier ^{13}C / ppm $^{(*)}$	-	54
Time / ms	0.6	5
Transfer II	**NH (dipolar)**	**CN (dipolar)**
^{1}H field / kHz	100	-
^{13}C field / kHz	-	34
^{15}N field / kHz	43.8	26
Shape	Ramp ^{1}H	Tangent ^{15}N
Carrier ^{13}C / ppm	4.8	54
Time / ms	0.6	7
Transfer III		**NH (dipolar)**
^{1}H field / kHz		95
^{13}C field / kHz		-
^{15}N field / kHz		44
Shape		Ramp ^{1}H
Carrier ^{13}C / ppm		-
Time / ms		0.6

$^{(*)}$ Carriers for ^{1}H and ^{15}N were 4.8 and 120 ppm respectively.

Supplementary Table 3: Rotors volumes and weights. The active volume of each rotor was calculated using the inner rotor area based on the inner diameter as indicated by Bruker, multiplied by the coil length. For the 3.2 mm rotor weight, an average was calculated on 3 empty rotors and 4 full rotors filled with similar core protein samples which give a standard deviation of 3 mg. For the weight of 0.7 mm and 1.3 mm rotors, the original rotor samples were used. The mass ratios with respect to the protonated 0.7 mm rotors are given in the right column. All samples were filled using overnight ultracentrifugation at 200 000 g, and residual protein was observed in the funnel for all samples, indicating that the rotors were full. Note that the actual protein quantity required to fill each rotor is about half of the total sample weight indicated, as the sedimented capsids contains solvent. Typical starting amounts of capsids used to fill a rotor are in our hands ~30 mg for a thin-wall 3.2 mm rotor, ~2 mg for a 1.3 mm rotor and ~0.4 mg for a 0.7 mm rotor.

Cp149	Active volume (μl)	Empty weight / mg	Full weight / mg	**Total sample weight / mg**	**Mass ratio to 0.7 mm protonated**
0.7 mm, protonated	0.37	6.10 ± 0.04	6.66 ± 0.01	**0.56 ± 0.04**	**1**
0.7 mm, deuterated	0.37	6.12 ± 0.01	6.67 ± 0.01	**0.55 ± 0.01**	**1**
1.3 mm, deuterated	1.45	35.6 ± 0.1	39.9 ± 0.1	**4.3 ± 0.1**	**7.8**
3.2 mm thin-wall, protonated	33	394 ± 2	449 ± 3	**55 ± 3**	**98.2**

Supplementary References

Lecoq, L., Wang, S., Wiegand, T., Bressanelli, S., Nassal, M., Meier, B. H., et al. (2018). Solid-state [13C-15N] NMR resonance assignment of hepatitis B virus core protein. *Biomol. NMR Assign.* 12, 205–214. doi:10.1007/s12104-018-9810-y.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., et al. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687–696. doi:10.1002/prot.20449.

- **Combining cell-free protein synthesis and NMR into a tool to study capsid assembly modulation**

This work was done with:

Shishan Wang, Marie-Laure Fogeron, Maarten Schledorn, Marie Dujardin, Susanne Penzel, Dara Burdette, Jan Martin Berke, Michael Nassal, Lauriane Lecoq, Beat H. Meier & Anja Böckmann.

Contribution: carried out protein bacterial and cell-free syntheses and analyses, generated NMR samples, recorded and analyzed electron micrographs, analyzed results, wrote the paper.

➢ Abstract

Modulation of capsid assembly by small molecules has become a central concept in the fight against viral infection, as proper capsid assembly is crucial to form the high molecular weight structures that protect the viral genome and, for various viruses in concert with the envelope, allow for cell entry and fusion. Atomic details underlying assembly modulation are generally studied using preassembled protein complexes, while the activity of assembly modulators during assembly remains largely open and poorly understood, as necessary tools are lacking. We here use the full-length hepatitis B virus (HBV) capsid protein (Cp183) as a model to present a combination of cell-free protein synthesis and solid-state NMR as an approach which shall open the possibility to produce and analyze the formation of higher-order complexes directly on exit from the ribosome. We demonstrate that assembled capsids can be synthesized in amounts sufficient for structural studies, and show that addition of assembly modulators to the cell-free reaction produces objects similar to those obtained by addition of the compounds to preformed Cp183 capsids. These results establish the cell-free system as a tool for the study of capsid assembly modulation directly after synthesis by the ribosome, and they open the perspective of assessing the impact of natural or synthetic compounds, or even enzymes that perform post-translational modifications, on capsids structures.

➢ Introduction

The hepatitis B virus (reviewed in (Nassal, 2008; Seeger and Mason, 2015)) is a small enveloped DNA virus whose genomic information encodes few genes: the envelope proteins S, M and L (collectively known as hepatitis B surface antigen/HBsAg), the core protein (Cp), the polymerase (P), and the X protein (HBx). The icosahedral HBV capsid is formed by Cp, the different functions of which are driven by phosphorylation / dephosphorylation of its C-terminal domain (Blondot et al., 2016; Gazina et al., 2000; Heger-Stevic et al., 2018b; Kann and Gerlich, 1994; Ludgate et

al., 2016). Cp is a 183-residue protein with two domains: the assembly domain that forms the contiguous capsid shell, and the C-terminal domain (CTD, residues 150-183) that amongst other functions is responsible for RNA packaging (Birnbaum and Nassal, 1990). The two domains are connected by a linker (residues 141-149). In infected cells, the core proteins pack the pregenomic (pg) RNA on assembly (Nassal, 1992), as well as a copy of the viral polymerase (Bartenschlager et al., 1990). Inside the capsid, the pgRNA is then transcribed to double-stranded relaxed circular (rc) DNA, generating mature capsids ready for envelopment.

The core protein thus plays essential roles at different stages of the virus life cycle and currently emerges as a promising drug target (Zlotnick et al., 2015) (recently reviewed in (Nijampatnam and Liotta, 2019; Yang and Lu, 2018)), with development of corresponding, effective antiviral agents well under way. Molecules targeting Cp are often called capsid assembly modulators or core protein allosteric modulators (CAMs (Zlotnick et al., 2015), CpAMs (Zlotnick et al., 2015)). Their major mechanism has been described either to be the acceleration of capsid assembly kinetics, which promotes the formation of morphologically normal capsid structures, but results in a failure to package pgRNA, as observed for example for AT-130. Or induction of aberrant oversized Cp structures (Diab et al., 2018), sequestering capsids from their functions, as observed for heteroaryldihydropyrimidines (HAP). To avoid the confusion in the literature as to which mechanism of action is to be called class I vs. class II (Lahlali et al., 2018; Yang et al., 2019) we herein use a tentative new nomenclature whereby CAM-N refers to modulators causing normal and CAM-A to modulators causing abnormal capsid structures. CAMs interfere with several central steps in the viral life cycle. They have been shown to prevent nuclear transport of capsids blocking *de novo* formation of covalently closed circular (ccc) DNA (Nassal, 2015); they are active pan-genotypic, and active against nucleoside analogue resistant virus mutants. Several CAMs of both classes are being evaluated in clinical trials (Durantel and Zoulim, 2016; Feng et al., 2018; Schinazi et al., 2018).

The capsid structure has been investigated by a range of structural-biology techniques. With the exception of a 3.3 Å X-ray structure (Wynne et al., 1999) of the N-terminal assembly domain, structures of the full-length capsid have been determined

by cryo-electron microscopy (cryo-EM) (Bottcher et al., 1997; Crowther et al., 1994), the latest to date at 2.7 Å resolution (Böttcher and Nassal, 2018). The different cryo-EM structures have mostly been described as similar to the X-ray structure, although small differences have been attributed to the absence/presence of the CTD (Yu et al., 2013), the presence of RNA as opposed to DNA (Roseman et al., 2005), or to drug binding (Schlicksup et al., 2018). Importantly, these structures have excluded the highly positively charged CTD and therefore a complete structure has yet to be determined (Patel et al., 2017; Zlotnick et al., 1997).

Eight structures of capsids with antiviral compounds bound have been determined (Katen et al., 2013; Klumpp et al., 2015; Qiu et al., 2016; Schlicksup et al., 2018; Venkatakrishnan et al., 2016; Zhou et al., 2017) (Bourne et al., 2006). All characterized Cps carried mutations, and none contained the CTD. The most commonly used constructs were Cp150 carrying an unnatural C-terminal cysteine plus triple Cys to Ala mutations depleting the protein from all endogenous cysteine residues (3CA-Cp150C). These mutations maintain the symmetry used in cryo-EM reconstruction of capsid structures, as the covalent intra-capsid cross linking via the C-terminal cysteine counterbalances the drastic destabilizing effects of the investigated HAP1 CAM. The increased stability of the modified capsids enabled X-ray structures with a decent resolution of around 4 Å (Venkatakrishnan et al., 2016). A higher resolution (1.7 Å) structure was obtained by X-ray crystallography employing the Cp Y132A mutation (Qiu et al., 2016) that abrogates capsid formation. Instead, Y132A induces flat hexameric structures (trimers of dimers) which form excellent crystals but clearly do not reflect the structure of the assembled capsid which is important when considering CAM action (Schlicksup et al., 2018).

Overall it remains unclear, at a molecular level, whether assembly modulators act similarly on preassembled versus nascent capsids. One approach to address this issue is to disassemble capsids into dimeric Cp subunits which then are incubated with the CAM under assembly-favoring conditions, e.g. high concentrations of salt (Schlicksup et al., 2018). Such non-physiological conditions could possibly interfere with assembly modulation. Studying assembly modulation instead directly at the exit from the ribosome, under conditions close to the cellular environment, is thus of high interest.

This can principally be achieved using cell-free protein synthesis (CFPS). CFPS of the HBV capsid has been described early-on by Lingappa and coworkers (Lingappa et al., 1994) who produced viral capsids in wheat-germ extract cell-free system (WGE-CF). Their study was motivated by the question how capsid assembly is influenced, under near-physiological concentrations, by cellular proteins, the cytoplasmic environment, and organelles (Lingappa et al., 1994). Indeed, in cells, the concentration of capsid protein is relatively low (an estimate of the steady-state HBc concentration in stably transfected hepatoma cells established ca. 300 nM (Ludgate et al., 2016)). Another important point is that assembly and its modulation with purified protein differs from that in cells where capsid formation is linked to Cp translation (Lingappa et al., 2005) and occurs in the presence of chaperones. A rabbit reticulocyte lysate (RRL) cell-free system has been recently applied to study HBV capsid assembly under more physiological conditions; Cp is expressed with low concentrations and assembles under near-physiological conditions (Liu and Hu, 2018; Ludgate et al., 2016). However, this system generally does not yield quantities (about 250 ng per mL reaction (Ludgate et al., 2016)) sufficient for structural studies, notably by NMR.

While cell-free expression can provide a means to sample capsid modulation directly at the exit from the ribosome, the approach remains limited without a means to structurally analyze the products at atomic resolution. Solid-state NMR can study full-length, wild-type capsids simply as sediments resulting from ultracentrifugation (Andreas et al., 2016; Goldbourt et al., 2007; Han et al., 2010). Notably, NMR has low requirements on sample properties: they neither need to be crystalline, nor show symmetry, only local order. This allows comparisons of the NMR signals of a variety of preparations and forms, and to conclude about structural and dynamic differences. NMR can in principle provide spectral fingerprints relating to structural features for both normal and abnormal capsid induced by modulators, including for capsids carrying modifications like phosphorylation. As the necessary basis for further studies, we have recently assigned the NMR signals of the HBV capsid (Lecoq et al., 2018b), revealing residues which conformationally adapt to allow for the dimer-to-capsid transition. Also, we identified the residues of the core protein which form the hinges that accommodate formation of the quasi-equivalent five-fold and quasi-six-fold vertices in the capsid (Lecoq et al., 2018a).

The classical approach to solid-state NMR involving carbon-13 detection is difficult to apply to the milligram quantities CFPS can easily produce. The recent development of proton-detection techniques opens the way for such studies, as it reduces the necessary protein amount by almost two orders of magnitude, to submilligram quantities (Böckmann et al., 2015) (Lecoq et al., under review). We have recently shown that the duck hepatitis B virus (DHBV) subviral particles can auto-assemble in the cell-free system and be analyzed by NMR (David et al., 2018). We show here that this same system can be used to produce wild-type full-length Cp HBV capsids, and do so in amounts compatible with solid-state NMR structural investigations, including the recording of 3D spectra with sufficient resolution and sensitivity. We show that the phenotypes produced by CAM-N and CAM-A are similar to those produced using purified capsids from *E. coli*. Hence, WGE-CF synthesis of capsids combined with solid-state NMR provides a valuable tool to study the effects of capsid assembly modulation on proteins directly at the exit of the ribosome.

➢ Material and Methods

Plasmids

The genes corresponding either to the full-length core protein (Cp183) or to its truncated form Cp149 were cloned into the pEU-E01-MCS vector (CellFree Sciences, Japan) for WGE-CF expression. The plasmids were amplified in DH5α bacteria, and purified using a NucleoBond Xtra Maxi kit (Macherey-Nagel, France). An additional purification step was performed with a phenol/chloroform extraction to ensure the purity of the plasmid according to the recommendations of CellFree Sciences (Yokohama, Japan).

mRNA transcription

Transcription was performed according to (Takai et al., 2010) in 1.5 mL Eppendorf tubes using 100 μg/mL plasmid, 2.5 mM NTP mix (Promega), 1U/μL SP6 RNA Polymerase (CellFree Sciences) and 1U/μL RNase inhibitor (CellFree Sciences) in transcription buffer (CellFree Sciences) containing 80 mM Hepes-KOH pH 7.6, 16 mM

magnesium acetate, 10 mM DTT and 2 mM spermidine. After incubation for 6 h at 37 °C, mRNA was used directly for translation.

Wheat germ cell-free protein synthesis

Non-treated durum wheat seeds (Sud Céréales, France) were used to prepare home-made WGE as described in (Fogeron et al., 2017), according to the protocol of Takai and colleagues (Takai et al., 2010) with minor modifications. Translation was performed using the bilayer method as described in (Fogeron et al., 2017; Takai et al., 2010) for small scale expression tests in the presence of compounds, or using the dialysis mode as described in (David et al., 2018) for larger scale production followed by isolation on a sucrose density gradient. For the bilayer method, the bottom layer (20 µL) corresponding to the translation mixture contains per well 10 µL of mRNA, 10 µL of WGE, 40 ng/µL of creatine kinase and 6 mM of amino-acid mix (0.3 mM per amino acid, average concentration). The upper layer (200 µL) corresponding to the feeding buffer contains SUB-AMIX NA (CellFree Sciences; 30 mM Hepes-KOH pH 7.6, 100 mM potassium acetate, 2.7 mM magnesium acetate, 16 mM creatine phosphate, 0.4 mM spermidine, 1.2 mM ATP, 0.25 mM GTP and 4 mM DTT), and 6 mM of amino acid mix (0.3 mM per amino acid, average concentration). For Cp183 expression in the presence of different compounds, 10 nmol of antiviral (dissolved in DMSO at a concentration of 10 mM) was added into 200 µL feeding buffer and translation was performed at 22 °C for 16 h.

For large-scale production, dialysis cassettes with a volume of either 500 µL or 3 mL, depending on the production scale, and a MWCO of 10 kDa were used. The translation mixture contained ½ by volume of feeding buffer, 1/3 of mRNA, 1/6 of WGE, 40 ng/µL of creatine kinase, 0.3 mM of amino-acid mix. The feeding buffer (either 20 mL or 124 mL for a 500-µL or a 3-mL dialysis cassette, respectively) contains SUB-AMIX NA (CellFree Sciences) as described above, supplemented with 0.3 mM of amino-acid mix. The dialysis cassette containing the translation mix was soaked in the feeding buffer, and incubated for 16 h under shaking at 60 rpm, 22°C. A mix containing all twenty isotopically labeled amino acids (Cambridge Isotope Laboratory) was used for the production of ^{13}C-^{15}N-Cp183 for NMR studies in a 3 mL-translation reaction experiment.

Isolation of the capsids on a sucrose density gradient

The total cell-free reaction mixture (CFS) was treated with 25,000 units/mL of benzonase for 30 min at room temperature before centrifugation at 20,000 g, 4 °C for 30 min. The supernatant (SN) was loaded onto a discontinuous sucrose gradient with layers of 10%, 20%, 30%, 40%, 50% and 60% sucrose (w/v), each with a volume of 350 µL for a production in a 500-µL cassette. For the production of a ^{13}C-^{15}N-Cp183 sample in a 3-mL dialysis cassette, the supernatant (SN) was split into two fractions and loaded onto two sucrose gradients with layers of 10%, 20%, 30%, 40%, 50% and 60% sucrose (w/v), each with a volume of 1.5 mL. The gradients were centrifuged at 200,000 g, 4 °C for 12 h. After centrifugation, the different sucrose fractions were harvested and analyzed by SDS-PAGE and Western blotting, as well as by electron microscopy after negative staining as described below.

Capsids from E. coli

Cp183 capsids used as reference for negative stain EM with CAMs were obtained from BL21*-CodonPlus (DE3) cells using plasmid pRSF-T7-HBc183opt. Expression and purification were done as previously reported (Heger-Stevic et al., 2018a; Lecoq et al., 2018b). In brief, protein was expressed overnight after induction with1 mM IPTG at 20 °C, and cell lysate was separated with 10-60% sucrose gradient. Cp183 capsids were precipitated after the sucrose gradient by 40% saturation ammonium sulfate, and resuspended in final buffer (50 mM Tris pH 7.5, 5 mM DTT, 1 mM EDTA, 5% sucrose). The interaction between preformed capsids and compounds was performed with a molar ratio of Cp183 monomer : compound of 1:4, at 37 °C for 2 h.

Rotor filling and NMR data acquisition

Four different Cp183 NMR samples were prepared: two from cell-free protein synthesis, one synthesized using $^{13}C/^{15}N$, and the other one $^{2}H/^{13}C/^{15}N$ amino acids, resulting in a protnated sample, and a deuterated, but 100% protonated on exchanging protons, as synthesis is carried out in H_2O; and for reference two samples from *E. coli* expression, one deuterated and back exchanged on exchangeable sites, and one protonated (Heger-Stevic et al., 2018a; Lecoq et al., 2018b). NMR samples were filled

into 0.7 mm rotors as sediment obtained by ultracentrifugation directly into the rotor (Böckmann et al., 2009) at 200,000 g for approximately 16 h at 4 °C, yielding approximately 0.5 mg of sediment. As an internal chemical-shift reference, about 30 μL of saturated (0.3 M) 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was added to the protein solution before sedimentation.

On each of the samples a two-dimensional (2D) fingerprint hNH spectrum was recorded. On the protonated, uniformly ^{13}C-^{15}N labeled cell-free produced sample, an hCANH 3D (Penzel et al., 2015) was recorded in addition. All spectra were acquired on a wide-bore 850 MHz Bruker Avance III spectrometer with a 0.7 mm triple-resonance MAS probe (Bruker Biospin) operated at 100 kHz MAS. Magic angle and shim for this probe were set using a 0.7 mm rotor with glycine ethylester by optimizing the intensity and *J*-coupling based splitting of the CO resonance (Penzel et al., 2018). The sample was cooled with a BCU (Bruker Cooling Unit) gas flow of 400 l/h with a VT (Variable Temperature) set to 272 K, corresponding to a sample temperature of approximately 22 °C, extrapolated from the water chemical shift in a ^{1}H 1D (Böckmann et al., 2009; Gottlieb et al., 1997). Detailed acquisition parameters can be found in Supplementary Table 1.

NMR data processing

TopSpin 4.0.3 (Bruker Biospin) was used for the data acquisition and processing. 2D hNH spectra were processed with 1024 points in ^{1}H dimension (corresponding to 12.9 ms of acquisition time) and zero filling was applied to respectively 4096 points in ^{1}H and 1024 points in ^{15}N dimension. The 3D hCANH was processed with zero filling to respectively 2048 points in ^{1}H, 128 points in ^{15}N and 256 points in ^{13}C dimensions. All spectra were apodized with a shifted sine-bell window function using SSB = 3.5 in TopSpin. Linear prediction to 770 points was applied in the ^{15}N dimension for 2D hNH spectra of the protonated capsids produced by CFPS, in order to reach a similar resolution as acquired for the other samples. Spectral analyses were performed using the CcpNmr Analysis package 2.4.2 (Stevens et al., 2011). The proton linewidths were obtained using the parabolic fit function integrated on CcpNmr on six isolated peaks in the hNH spectra. The errors given represent the standard deviations between the six values. Signal-to-noise ratio were calculated on the bulk signals from 1D hNH spectra

recorded and processed with similar parameters and divided by the square root of the number of scans.

SDS-PAGE & Western blotting analysis

The expression of Cp183 was assessed by 15% Coomassie blue stained SDS–PAGE and Western blotting as described in (Fogeron et al., 2015). A polyclonal rabbit antiserum against the N-terminal domain of the HBV core protein (a-c149) was used to detect both Cp149 and Cp183 on blots.

Negative staining electron micrographs

Samples for electron microscopy were negatively stained as described in (Lecoq et al., 2018b). Briefly, 5 µL of each fraction were loaded on a carbon-coated grid (EMS Microscopy) and incubated for 2 min at room temperature. Remaining liquid was drained using Whatman paper. Grids were negatively stained on a 50-µL drop of 2% phosphotungstic acid (pH=7) for 2 min at room temperature and observed with a JEM-1400 transmission electron microscope operating at 100 kV.

➢ Results

1. Full-length Cp183 but not CTD-less Cp149 self-assembles upon cell-free protein synthesis

CFPS of the core protein was performed for both Cp149 and Cp183. The protein was found mainly in the soluble fraction after centrifugation, as indicated in Western blots in Figure 1a and b. The protein band is partly visible in the total CFS fraction of the Coomassie blue gel. Enrichment via a sucrose gradient reveals that Cp149 stays mainly in the load and in the 10% sucrose fraction, indicating the protein remained in an unassembled, probably dimeric state. Accordingly, the electron micrograph of the 10% fraction (Figure 1a, blue asterisk), showed only very few capsids. In contrast, Cp183 sedimented largely into the 50% and 60% sucrose fractions (Figure 1b, red asterisk), as expected when capsids have been formed. EM inspection revealed

numerous auto-assembled Cp183 capsids with a diameter of about 30 nm, as also observed for capsids assembled in *E. coli* (Gallina et al., 1989; Lecoq et al., 2018b).

Upon expression in *E. coli*, both Cp183 and the CTD-less Cp149 variant auto-assemble into capsids. Only full-length protein packages RNA, while Cp149 capsids remain empty (Birnbaum and Nassal, 1990). Both types of capsids can be isolated from bacteria by a set of purification steps (Heger-Stevic et al., 2018b; Lecoq et al., 2018b), with Cp149 giving particularly high yields (100 mg per liter of culture, compared to 20 mg/L for Cp183). The capsids can be disassembled using either urea (Cp149) or guanidinium chloride (Cp183) (Porterfield et al., 2010; Zlotnick et al., 1997). Reassembly is concentration dependent, and *in vitro* assembly of the full-length protein needs addition of nucleic acids which are non-sequence specifically packaged (Porterfield et al., 2010). Failure of Cp149 to assemble upon WGE-CF synthesis is likely due to the higher concentrations this protein needs for assembly, while the interaction between the positively-charged Cp183 CTD with the negatively-charged nucleic acids enables Cp183 assembly at concentrations as low as 5 nM (Klein et al., 2004). Failure of Cp149 to assemble has also been observed in rabbit reticulocyte extract (Ludgate et al., 2016).

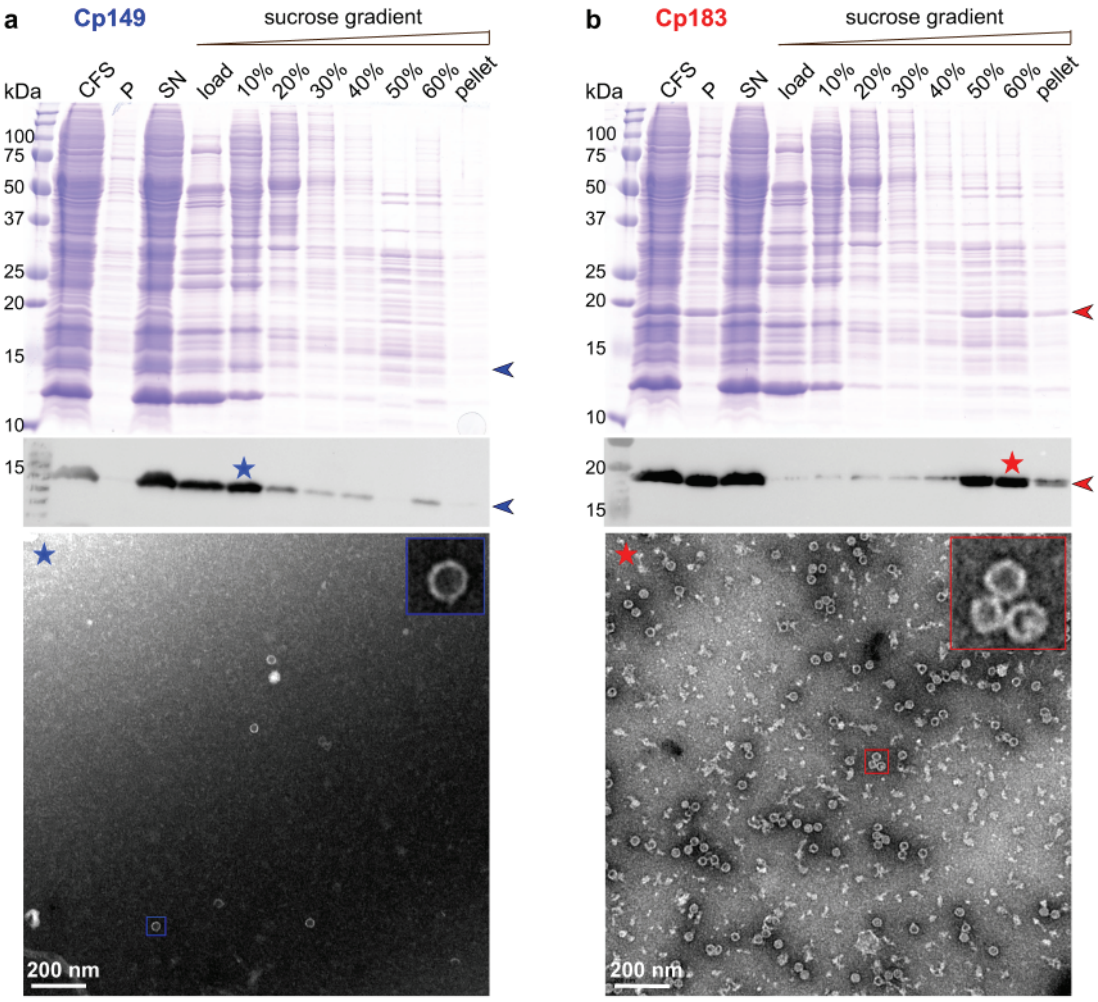


Figure 1: CFPS and sucrose gradient isolation of Cp149 (a) and Cp183 (b). Shown are from top to bottom, Coomassie blue stained gels, western blots, and negative staining electron micrographs of protein-containing fractions. CFS: total cell-free reaction mixture; P and SN: pellet and supernatant obtained after centrifugation of the CFS at 20,000 g, 4 °C for 30 min; 0%-60%: fractions from the sucrose gradient. Scale bar = 200 nm.

2. Milligram amounts of $^{13}C/^{15}N$ labeled Cp183 capsids can be produced in protonated and deuterated form

For large-scale production (approx. 1 milligram) needed for NMR sample preparation, CFPS was carried out in dialysis reactions, as described for the duck HBV envelope subviral particle synthesis (David et al., 2018). Either protonated or deuterated, HN protonated Cp183 was prepared, with the latter referred to in the following as dCp183. As for the small-scale synthesis, the protein was mainly found in the soluble fraction, and on sucrose gradient isolation, migrated to the 60 % fraction (Figure 2a, b). The

preparation using the deuterated amino acids shows higher purity, which might be due to a slightly different migration behavior of the deuterated protein in the sucrose gradient. EM inspection revealed abundant capsids in both preparations (Figure 2c, d).

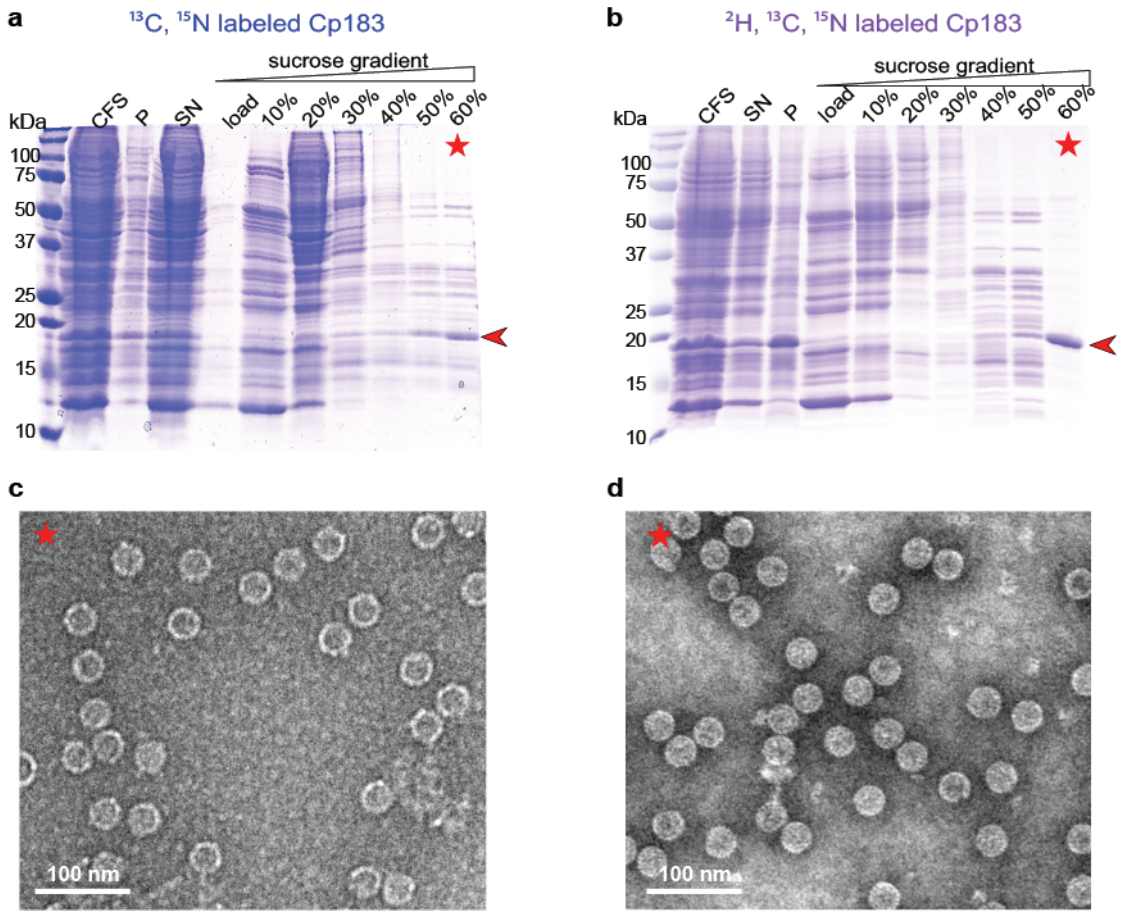


Figure 2: Analysis of CFPS and sucrose gradient isolation on 15% SDS-PAGE gels of a) ^{13}C, ^{15}N and b) ^{2}H, ^{13}C, ^{15}N isotopically labeled dCp183. CFS: total cell-free reaction mixture; P and SN: pellet and supernatant obtained after centrifugation of the CFS at 20,000 g, 4°C for 30 min; 0%-60%: fractions from the sucrose gradient. Negative staining electron micrographs display the ^{13}C, ^{15}N labeled (c) and ^{2}H, ^{13}C, ^{15}N labeled (d) capsids from the 60% sucrose fractions. Scale bar = 100 nm.

3. Cell-free synthesized capsids can be analyzed by NMR

Conformational details can be revealed by NMR in so-called fingerprint spectra, which show either in two (2D) or three dimensions (3D) the typical signature of the protein preparation. Structural variations can be sensitively identified by comparing spectra recorded under different conditions, and analyzing the differences in the observed

chemical shifts, i.e. the NMR frequencies (Williamson, 2013). An opportunity of the combination of CFPS and NMR is the fact that only the synthesized protein, which is the sole isotopically labelled protein, will be observed in the spectra. The use of a simple sucrose gradient concentration step thus might not produce perfectly pure protein; still, only the protein of interest will produce signal in the spectra. A possible drawback might lie in a loss of signal-to-noise ratio (SNR) in the spectra, since the NMR sample container (rotor) also might contain residual contaminating proteins (Figure 2a). It is thus important to establish whether protein samples prepared by CFPS are indeed compatible with the recording of 2D and in particular 3D spectra in a reasonable amount of time.

The hNH 2D correlation spectrum recorded in 16 h on the protonated cell-free Cp183 displays a highly similar spectrum to the one recorded on the capsids purified from *E. coli* (Figure 3a and S1) in 10 h. The NMR signal amplitude of the sample from CFPS is about 35% of the spectra obtained on the preparation from purified *E. coli* protein recorded under the same experimental conditions. As both rotors were full with protein sediment, this means that the contaminating unlabeled proteins from the WGE fill almost 2/3 of the rotor. A 3D hCANH spectrum was recorded on the sample in 4 days and 15h, and an overlay of all 3D NH planes onto the 2D NH plane shows that most signals in the 2D hNH spectrum are also observed in the 3D (Figure 3b).

The 2D spectrum recorded on the deuterated sample is shown in Figure 3c-d. SNR is very favorable in this sample, since the deuterated protein surprisingly showed better purity (Figure 2b). The spectrum reveals narrower lines than the spectrum from the protonated sample, as also observed in model systems (Penzel et al., 2019) and, in particular, also in capsid preparations purified from *E. coli* (Lecoq et al. under revision): 140 Hz on average for the protonated versus 100 Hz for the deuterated sample, as measured on six isolated resonances. The SNR and proton linewidths for the four samples are summarized in Figures 3e and f, respectively. It reveals that CFPS samples show a greater variability in sample amounts than the well-established *E. coli* samples; further experience is needed to evaluate parameters allowing reproducible sample preparation using CFPS. The proton linewidths are virtually similar between the two protonated and two deuterated samples, indicating that production by CFPS

or *E. coli* expression does not make a difference with respect to linewidth and therefore conformational homogeneity.

Importantly, several peaks are present in the cell-free synthesized dCp183 which could not be observed in the deuterated sample purified from *E. coli*, as emphasized in Figure 3d. The origin of this observation lies in the incomplete back-exchange in *E. coli* produced samples. Indeed, when deuterated protein is expressed in *E. coli*, synthesis takes place in D_2O, and exchange of deuterons to protons is achieved during the subsequent purification steps, carried out in H_2O. Still, solvent-inaccessible deuterons can remain in the protein over long periods of time, and often denaturation/renaturation of the protein is applied to complete proton exchange important for NMR observation. However, this step can be very difficult for more complex proteins, and the present experiment highlights this interesting feature of CFPS, where the protein is synthesized from the beginning in H_2O, and deuteration is achieved not via metabolism, but by addition of deuterated amino acids to the cell-free reaction. This results in fully protonated amide (and exchangeable sidechain) protons in the synthesized protein, which is essential for the recording of NMR spectra showing resonances for all amino acids.

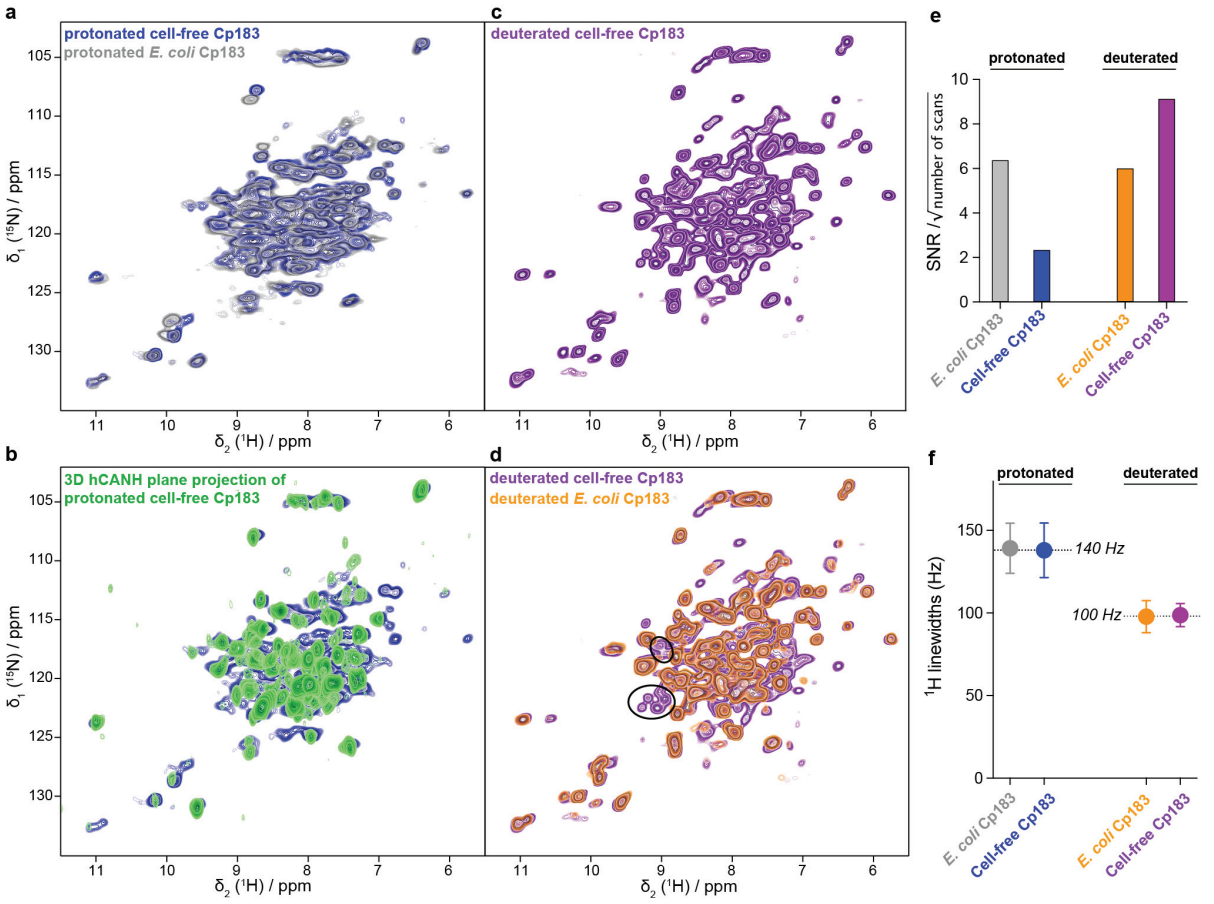


Figure 3. Comparisons of NMR spectra between the capsids from CFPS and capsids purified from *E. coli*. a) Overlay of the 2D hHN spectra of the protonated Cp183 capsids from CFPS (in blue) and purified from *E. coli* (in grey); b) overlay of the 2D planes from the 3D hCANH spectrum recorded on the protonated Cp183 CFPS capsids; c) 2D hNH spectrum of the deuterated Cp183 CFPS capsids; d) overlay of the 2D hNH spectra of the deuterated capsids from CFPS (in purple) and purified from *E. coli* (in orange), with resonances not observed in the *E. coli* sample highlighted as black circles. Spectra are shown individually in Figure S1. e) Comparison of signal-to-noise ratios of the different samples; f) comparison of proton linewidths in the different samples. The averages over the two protonated and the two deuterated samples are indicated.

4. Capsids can be synthesized in the presence of antiviral compounds

CFPS proceeds in an open system, and a variety of substances can be added to the reaction mixture. We added different capsid assembly modulators to the reaction, in order to analyze whether this produces comparable phenotypes to those observed on capsids purified from *E. coli*. Figure 4a shows the Coomassie blue stained gels of the cell-free solutions without compounds, in the presence of DMSO used for solubilization

of the antiviral, and in presence of AT-130, JNJ-623 (CAM-N), and JNJ-890 (CAM-A). The corresponding Western blots are shown in Figure 4b. None of the compounds inhibited protein synthesis. We analyzed the total cell-free solutions, without any concentration or purification, under the electron microscope, and compared the observed capsids with the ones obtained from addition of compounds to capsids purified from *E. coli*. One can see in the micrographs that the resulting objects closely resemble those obtained by addition to preformed capsids: DMSO vehicle and CAM-Ns produced no visible effect, whereas CAM-As showed the typical disruption of capsids also reported in the literature (Berke et al., 2017; Lahlali et al., 2018). Notably, the presence of AT-130 lead to poorer contrast in the EM micrographs of both preparations.

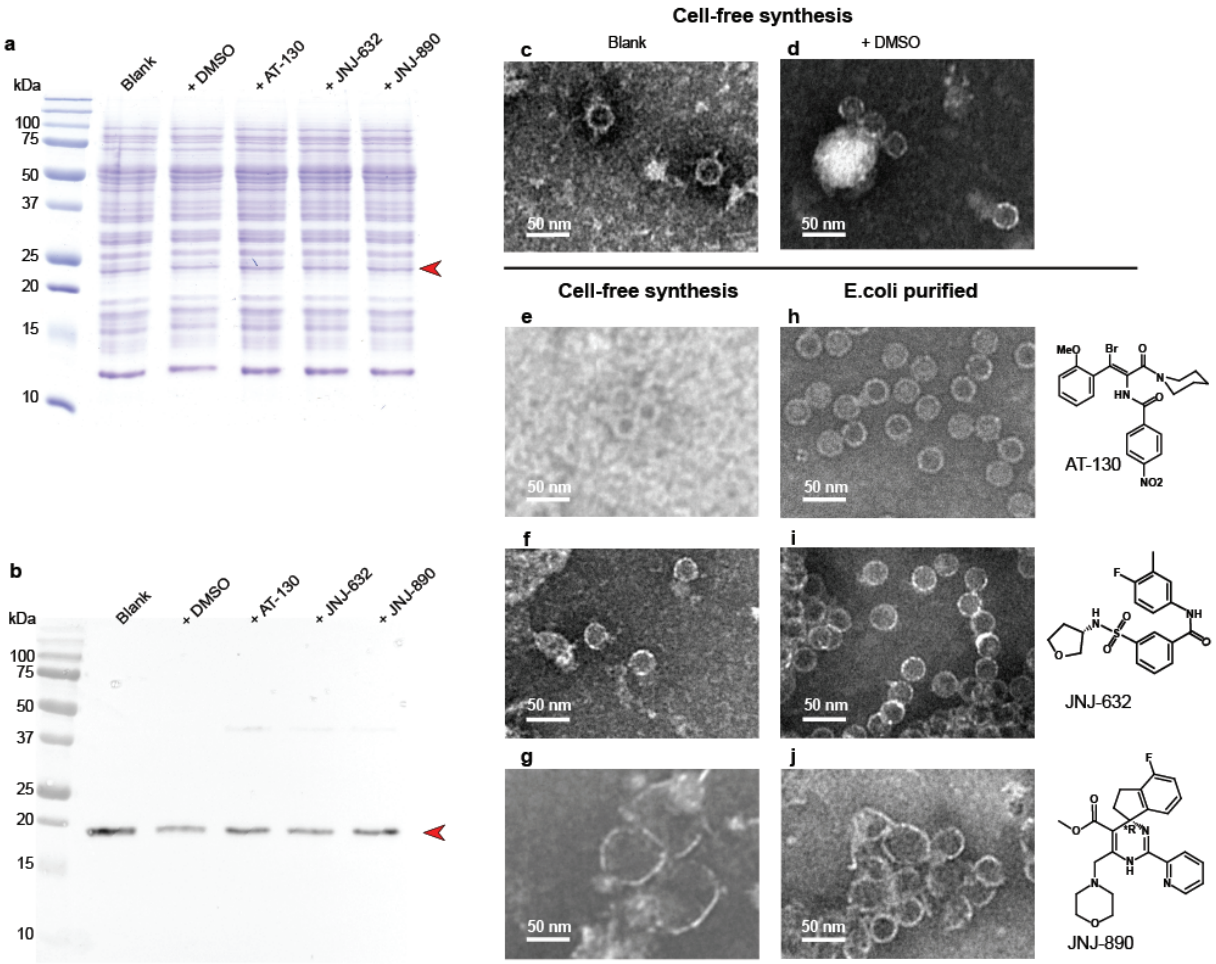


Figure 4: CFPS of Cp183 in presence of different compounds. a-b) 15% SDS-PAGE (a) and Western blot (b) analysis on Cp183 protein produced in 220-µL CFPS system with 10 nmol drugs dissolved in 1 µL DMSO. The lanes of two controls, referred as *blank* and *+DMSO*, and productions

with compounds are highlighted on the top. Electron microscopy (negative staining) analysis of cell-free products of the *blank* (c), *+DMSO* (d), *+AT-130* (e), *+JNJ-632* (f), *+JNJ-890* (g), as well as capsids from *E. coli* incubated with AT-130 (h), JNJ-632 (i), JNJ-890 (j), are shown on the right. Scale bar = 50 nm.

➢ Conclusions

We have synthesized HBV viral capsids in a eukaryotic wheat germ cell-free system in sufficient amounts for structural analyses, including by solid-state NMR. We have shown that the full-length Cp183 protein auto-assembles in the cell-free system to form icosahedral capsids virtually identical to those obtained upon bacterial expression. This finding opens the possibility to produce isotope labeled samples, both in protonated and deuterated forms, for advanced proton-detected NMR experiments. The spectra recorded on the samples showed sufficient signal-to-noise to analyze 2D and 3D spectral fingerprints and thus conformational changes. Importantly, this enables investigations of capsid interactions directly on synthesis with assembly modulators, other natural compounds such as lipids, or chaperones and enzymes that might be relevant *in vivo*. We demonstrated this at the example of three capsid assembly modulators from different chemical classes, which induced similar structural changes in capsids synthesized and assembled in presence of the compounds and in preformed capsids isolated from *E. coli*. Hence the influence of small molecules on the capsid can now also be assessed on assembly after exit from the ribosome, on the relevant full-length protein, without extensive purification steps, and in the presence of nucleic acids.

Author Contributions

SW, MLF and MD carried out protein syntheses and analyses, and generated NMR samples. MS, SP and LL conducted the NMR experiments. DB and JMB provided antiviral compounds, and contributed expert insight to CAMs. MN designed the plasmid and established bacterial expression/purification protocols, and contributed expert insight to HBV. MLF, LL, BHM and AB designed and supervised the study, and wrote the manuscript. All authors contributed to the manuscript and approved the submitted version.

Funding

This work was supported by the French ANRS (ECTZ71388), the LABEX ECOFECT (ANR-11-LABX-0048) within the Université de Lyon program Investissements d'Avenir (ANR-11-IDEX-0007), by the Marie Skłodowska-Curie program (H2020-MSCA-IF-2016 748516) by the Swiss National Science Foundation (Grant 200020_159707), and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement n° 741863, FASTER). We thank the Centre d'Imagerie Quantitative Lyon-Est (CIQLE) for technical support with the electron microscope.

➢ References

Andreas, L. B., Jaudzems, K., Stanek, J., Lalli, D., Bertarello, A., Le Marchand, T., et al. (2016). Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9187–9192. doi:10.1073/pnas.1602248113.

Bartenschlager, R., Junker-Niepmann, M., and Schaller, H. (1990). The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. *J. Virol.* 64, 5324–5332.

Berke, J. M., Dehertogh, P., Vergauwen, K., Van Damme, E., Mostmans, W., Vandyck, K., et al. (2017). Capsid Assembly Modulators Have a Dual Mechanism of Action in Primary Human Hepatocytes Infected with Hepatitis B Virus. *Antimicrob. Agents Chemother.* 61, e00560–17–14. doi:10.1128/AAC.00560-17.

Birnbaum, F., and Nassal, M. (1990). Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. *J. Virol.* 64, 3319–3330.

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

Bottcher, B., Wynne, S. A., and Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 386, 88–91. doi:10.1038/386088a0.

Bourne, C. R., Finn, M. G., and Zlotnick, A. (2006). Global structural changes in hepatitis B virus capsids induced by the assembly effector HAP1. *J. Virol.* 80, 11055–11061. doi:10.1128/JVI.00933-06.

Böckmann, A., Ernst, M., and Meier, B. H. (2015). Spinning proteins, the faster, the better? *Journal of Biomolecular NMR* 253, 71–79.

Böckmann, A., Gardiennet, C., Verel, R., Hunkeler, A., Loquet, A., Pintacuda, G., et al. (2009). Characterization of different water pools in solid-state NMR protein samples. *J. Biomol. NMR* 45, 319–327. doi:10.1007/s10858-009-9374-3.

Böttcher, B., and Nassal, M. (2018). Structure of Mutant Hepatitis B Core Protein Capsids with Premature Secretion Phenotype. *J. Mol. Biol.* 430, 4941–4954. doi:10.1016/j.jmb.2018.10.018.

Crowther, R. A., Kiselev, N. A., Bottcher, B., Berriman, J. A., Borisova, G. P., Ose, V., et al. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* 77, 943–950.

David, G., Fogeron, M.-L., Schledorn, M., Montserret, R., Haselmann, U., Penzel, S., et al. (2018). Structural Studies of Self-Assembled Subviral Particles: Combining Cell-Free Expression with 110 kHz MAS NMR Spectroscopy. *Angew. Chem. Int. Ed.* 57, 4787–4791. doi:10.1002/anie.201712091.

Diab, A., Foca, A., Zoulim, F., Durantel, D., and 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. doi:10.1016/j.antiviral.2017.11.015.

Durantel, D., and Zoulim, F. (2016). New antiviral targets for innovative treatment concepts for hepatitis B virus and hepatitis delta virus. *J Hepatol* 64, S117–S131. doi:10.1016/j.jhep.2016.02.016.

Feng, S., Gao, L., Han, X., Hu, T., Hu, Y., Liu, H., et al. (2018). Discovery of Small Molecule Therapeutics for Treatment of Chronic HBV Infection. *ACS Infect Dis* 4, 257–277. doi:10.1021/acsinfecdis.7b00144.

Fogeron, M.-L., Badillo, A., Jirasko, V., Gouttenoire, J., Paul, D., Lancien, L., et al. (2015). Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins. *Protein Expr Purif* 105, 39–46. doi:10.1016/j.pep.2014.10.003.

Fogeron, M.-L., Badillo, A., Penin, F., and Böckmann, A. (2017). Wheat Germ Cell-Free Overexpression for the Production of Membrane Proteins. *Methods Mol. Biol.* 1635, 91–108. doi:10.1007/978-1-4939-7151-0_5.

Gallina, A., Bonelli, F., Zentilin, L., Rindi, G., Muttini, M., and Milanesi, G. (1989). A recombinant hepatitis

B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. *J. Virol.* 63, 4645–4652.

Gazina, E. V., Fielding, J. E., Lin, B., and Anderson, D. A. (2000). Core protein phosphorylation modulates pregenomic RNA encapsidation to different extents in human and duck hepatitis B viruses. *J. Virol.* 74, 4721–4728.

Goldbourt, A., Gross, B. J., Day, L. A., and McDermott, A. E. (2007). Filamentous Phage Studied by Magic-Angle Spinning NMR: Resonance Assignment and Secondary Structure of the Coat Protein in Pf1. *J. Am. Chem. Soc.* 129, 2338–2344. doi:10.1021/ja066928u.

Gottlieb, H. E., Kotlyar, V., and Nudelman, A. (1997). NMR chemical shifts of common laboratory solvents as trace impurities. *The Journal of organic ….*

Han, Y., Ahn, J., Concel, J., Byeon, I.-J. L., Gronenborn, A. M., Yang, J., et al. (2010). Solid-State NMR Studies of HIV-1 Capsid Protein Assemblies. *JACS* 132, 1976–1987. doi:10.1021/ja908687k.

Heger-Stevic, J., Kolb, P., Walker, A., and Nassal, M. (2018a). Displaying Whole-Chain Proteins on Hepatitis B Virus Capsid-Like Particles. *Methods Mol. Biol.* 1776, 503–531. doi:10.1007/978-1-4939-7808-3_33.

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

Kann, M., and Gerlich, W. H. (1994). Effect of core protein phosphorylation by protein kinase C on encapsidation of RNA within core particles of hepatitis B virus. *J. Virol.* 68, 7993–8000.

Katen, S. P., Tan, Z., Chirapu, S. R., Finn, M. G., and Zlotnick, A. (2013). Assembly-directed antivirals differentially bind quasiequivalent pockets to modify hepatitis B virus capsid tertiary and quaternary structure. *Structure* 21, 1406–1416. doi:10.1016/j.str.2013.06.013.

Klein, K. C., Polyak, S. J., and Lingappa, J. R. (2004). Unique features of hepatitis C virus capsid formation revealed by de novo cell-free assembly. *J. Virol.* 78, 9257–9269. doi:10.1128/JVI.78.17.9257-9269.2004.

Klumpp, K., Lam, A. M., Lukacs, C., Vogel, R., Ren, S., Espiritu, C., et al. (2015). High-resolution crystal structure of a hepatitis B virus replication inhibitor bound to the viral core protein. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15196–15201. doi:10.1073/pnas.1513803112.

Lahlali, T., Berke, J. M., Vergauwen, K., Foca, A., Vandyck, K., Pauwels, F., et al. (2018). Novel Potent Capsid Assembly Modulators Regulate Multiple Steps of the Hepatitis B Virus Life Cycle. *Antimicrob. Agents Chemother.* 62, 672–15. doi:10.1128/AAC.00835-18.

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

Lecoq, L., Wang, S., Wiegand, T., Bressanelli, S., Nassal, M., Meier, B. H., et al. (2018b). Solid-state [13C-15N] NMR resonance assignment of hepatitis B virus core protein. *Biomol NMR Assign* 253, 80–10. doi:10.1007/s12104-018-9810-y.

Lingappa, J. R., Martin, R. L., Wong, M. L., Ganem, D., Welch, W. J., and Lingappa, V. R. (1994). A eukaryotic cytosolic chaperonin is associated with a high molecular weight intermediate in the assembly of hepatitis B virus capsid, a multimeric particle. *The Journal of Cell Biology* 125, 99–111.

Lingappa, J. R., Newman, M. A., Klein, K. C., and Dooher, J. E. (2005). Comparing capsid assembly of primate lentiviruses and hepatitis B virus using cell-free systems. *Virology* 333, 114–123. doi:10.1016/j.virol.2004.12.024.

Liu, K., and Hu, J. (2018). Host-regulated Hepatitis B Virus Capsid Assembly in a Mammalian Cell-free System. *Bio Protoc* 8. doi:10.21769/BioProtoc.2813.

Ludgate, L., Liu, K., Luckenbaugh, L., Streck, N., Eng, S., Voitenleitner, C., et al. (2016). Cell-Free Hepatitis B Virus Capsid Assembly Dependent on the Core Protein C-Terminal Domain and Regulated by Phosphorylation. *J. Virol.* 90, 5830–5844. doi:10.1128/JVI.00394-16.

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. *J. Virol.* 66, 4107–4116.

Nassal, M. (2008). Hepatitis B viruses: reverse transcription a different way. *Virus Res.* 134, 235–249. doi: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, 1972–1984. doi:10.1136/gutjnl-2015-309809.

Nijampatnam, B., and Liotta, D. C. (2019). Recent advances in the development of HBV capsid assembly modulators. *Current Opinion in Chemical Biology* 50, 73–79. doi:10.1016/j.cbpa.2019.02.009.

Patel, N., White, S. J., Thompson, R. F., Bingham, R., Weiß, E. U., Maskell, D. P., et al. (2017). HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. *Nature Microbiology*, 1–10. doi:10.1038/nmicrobiol.2017.98.

Penzel, S., Oss, A., Org, M.-L., Samoson, A., Böckmann, A., Ernst, M., et al. (2019). Spinning faster: protein NMR at MAS frequencies up to 126 kHz. *J. Biomol. NMR* 128, 12620–11. doi:10.1007/s10858-018-0219-9.

Penzel, S., Smith, A. A., Agarwal, V., Hunkeler, A., Org, M.-L., Samoson, A., et al. (2015). Protein resonance assignment at MAS frequencies approaching 100 kHz: a quantitative comparison of J-coupling and dipolar-coupling-based transfer methods. *J. Biomol. NMR* 63, 165–186. doi:10.1007/s10858-015-9975-y.

Penzel, S., Smith, A. A., Ernst, M., and Meier, B. H. (2018). Setting the magic angle for fast magic-angle spinning probes. *J. Magn. Reson.* 293, 115–122. doi:10.1016/j.jmr.2018.06.002.

Porterfield, J. Z., Dhason, M. S., Loeb, D. D., Nassal, M., Stray, S. J., and Zlotnick, A. (2010). Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *J. Virol.* 84, 7174–7184. doi:10.1128/JVI.00586-10.

Qiu, Z., Lin, X., Zhou, M., Liu, Y., Zhu, W., Chen, W., et al. (2016). Design and Synthesis of Orally Bioavailable 4-Methyl Heteroaryldihydropyrimidine Based Hepatitis B Virus (HBV) Capsid Inhibitors. *J. Med. Chem.* 59, 7651–7666. doi:10.1021/acs.jmedchem.6b00879.

Roseman, A. M., Berriman, J. A., Wynne, S. A., Butler, P. J. G., and Crowther, R. A. (2005). A structural model for maturation of the hepatitis B virus core. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15821–15826. doi:10.1073/pnas.0504874102.

Schinazi, R. F., Ehteshami, M., Bassit, L., and Asselah, T. (2018). Towards HBV curative therapies. *Liver Int* 38 Suppl 1, 102–114. doi:10.1111/liv.13656.

Schlicksup, C. J., Wang, J. C.-Y., Francis, S., Venkatakrishnan, B., Turner, W. W., VanNieuwenhze, M., et al. (2018). Hepatitis B virus core protein allosteric modulators can distort and disrupt intact capsids. *Elife* 7, 13046. doi:10.7554/eLife.31473.

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

Stevens, T. J., Fogh, R. H., Boucher, W., Higman, V. A., Eisenmenger, F., Bardiaux, B., et al. (2011). A software framework for analysing solid-state MAS NMR data. *J. Biomol. NMR* 51, 437–447. doi:10.1007/s10858-011-9569-2.

Takai, K., Sawasaki, T., and Endo, Y. (2010). Practical cell-free protein synthesis system using purified wheat embryos. *Nat Protoc* 5, 227–238. doi:10.1038/nprot.2009.207.

Venkatakrishnan, B., Katen, S. P., Francis, S., Chirapu, S., Finn, M. G., and Zlotnick, A. (2016). Hepatitis B Virus Capsids Have Diverse Structural Responses to Small-Molecule Ligands Bound to the Heteroaryldihydropyrimidine Pocket. *J. Virol.* 90, 3994–4004. doi:10.1128/JVI.03058-15.

Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Progr NMR Spectr* 73, 1–16. doi:10.1016/j.pnmrs.2013.02.001.

Wynne, S. A., Crowther, R. A., and Leslie, A. G. W. (1999). The Crystal Structure of the Human Hepatitis B Virus Capsid. *Mol Cell* 3, 771–780.

Yang, L., and Lu, M. (2018). Small Molecule Inhibitors of Hepatitis B Virus Nucleocapsid Assembly: A New Approach to Treat Chronic HBV Infection. *Curr. Med. Chem.* 25, 802–813. doi:10.2174/0929867324666170704121800.

Yang, L., Liu, F., Tong, X., ORCID: 0000-0003-2973-7869, D. H., Zuo, J., and ORCID: 0000-0003-4287-9941, M. L. (2019). Treatment of Chronic Hepatitis B Virus Infection Using Small Molecule Modulators of Nucleocapsid Assembly: Recent Advances and Perspectives. *ACS Infect Dis*, 1–12. doi:10.1021/acsinfecdis.8b00337.

Yu, X., Jin, L., Jih, J., Shih, C., and Zhou, Z. H. (2013). 3.5Å cryoEM structure of hepatitis B virus core assembled from full-length core protein. *PLoS ONE* 8, e69729. doi:10.1371/journal.pone.0069729.

Zhou, Z., Hu, T., Zhou, X., Wildum, S., Garcia-Alcalde, F., Xu, Z., et al. (2017). Heteroaryldihydropyrimidine (HAP) and Sulfamoylbenzamide (SBA) Inhibit Hepatitis B Virus Replication by Different Molecular Mechanisms. *Sci. Rep.*, 1–12. doi:10.1038/srep42374.

Zlotnick, A., CHENG, N., Stahl, S. J., Conway, J. F., Steven, A. C., and 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. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9556–9561.

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

➢ Supplementary Material

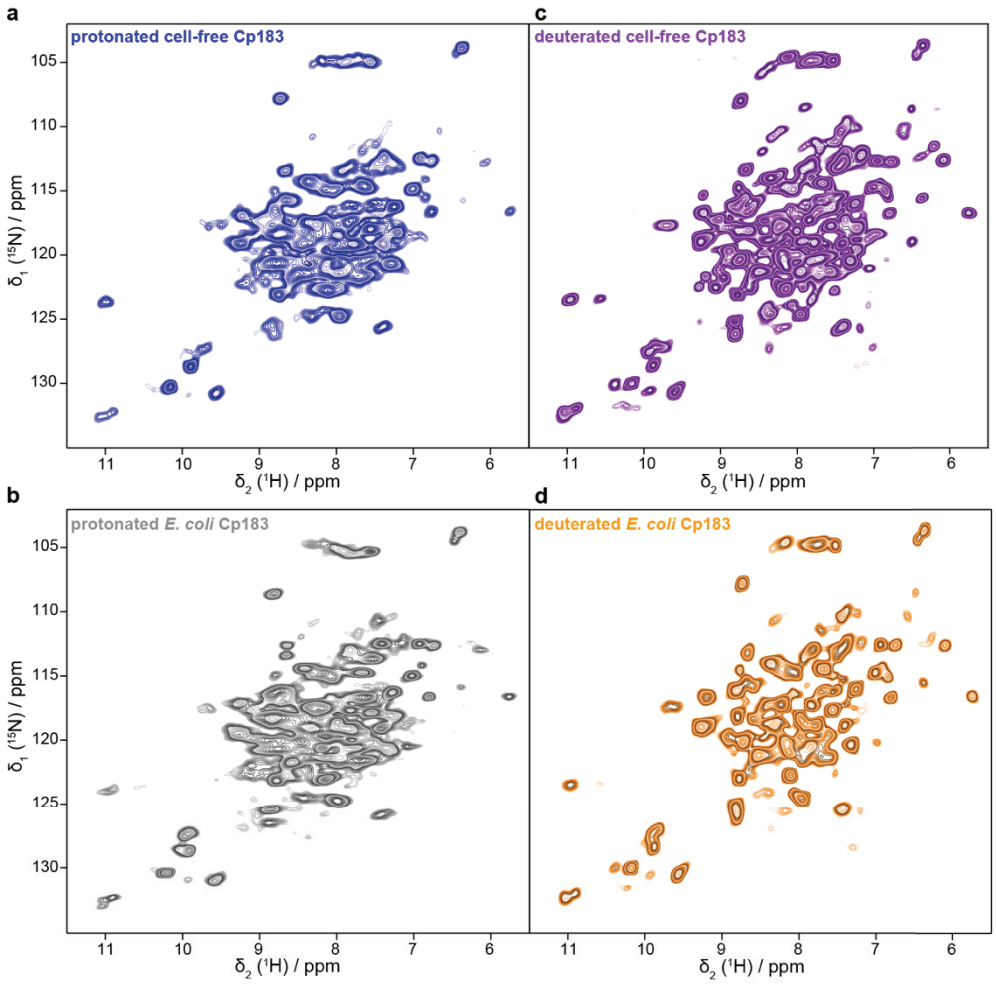


Figure S1. Individual spectra from Figure 3.

Supplementary Table 1: NMR experimental parameters for the presented data acquired on Cp183 capsids.

Sample production	***E. coli***	***E. coli***	**Cell-free**	**Cell-free**	
Sample labeling	**^{2}H ^{13}C ^{15}N**	**^{1}H ^{13}C ^{15}N**	**^{2}H ^{13}C ^{15}N**	**^{1}H ^{13}C ^{15}N**	
Experiment	**hNH 2D**	**hNH 2D**	**hNH 2D**	**hNH 2D**	**hCANH 3D**
MAS frequency / kHz	100	100	100	100	100
Field / T	20	20	20	20	20
t1 increments	776	774	620	330	104
Sweep width (t1) / ppm	180	180	180	180	36
Acquisition time (t1) / ms	25	25	20	11	6.8
t2 increments	5550	5550	2048	2048	44
Sweep width (t2) / ppm	47	47	47	40	34
Acquisition time (t2) / ms	70	70	26	30	7.5
t3 increments	-	-	-	-	2048
Sweep width (t3) / ppm	-	-	-	-	47
Acquisition time (t3) / ms	-	-	-	-	26
Proton decoupling (swfTPPM) / kHz	11	12	10	10	10

Nitrogen decoupling (WALTZ64) / kHz	6	6	5	5	5
Carbon decoupling (WALTZ64) / kHz	-	-	-	-	5
Water suppression (120 ms MISS.) / kHz	22	24	20	20	20
Interscan delay / s	2.0	1.2	2.0	1.2	1.2
Number of scans	32	32	64	128	64
Measurement time / dd:hh:mm	00:15:34	00:09:50	00:16:41	00:15:59	04:14:32
Carrier ^{1}H / ppm	4.8	4.8	4.8	4.8	4.8
Carrier ^{15}N / ppm	117.5	117.5	117.5	117.5	117.5
Transfer I	**HN CP**	**HN CP**	**HN CP**	**HN CP**	**HC CP**
^{1}H field / kHz	72	77	78	77	82
X field / kHz	17	16	15	15	15
Shape	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H
Carrier ^{13}C / ppm	-	-	-	-	56
Time / ms	1.0	1.0	1.8	1.6	0.6
Transfer II	**NH CP**	**NH CP**	**NH CP**	**NH CP**	**CN CP**
^{1}H field / kHz	72	77	78	77	-
^{13}C field / kHz	-	-	-	-	63

^{15}N field / kHz	17	16	15	15	35
Shape	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{1}H	Tangent ^{13}C
Carrier ^{13}C / ppm	-	-	-	-	56
Time / ms	1.0	1.0	2.4	1.6	20.0
Transfer III	**-**	**-**	**-**	**-**	**NH CP**
^{1}H field / kHz	-	-	-	-	78
^{15}N field / kHz	-	-	-	-	15
Shape	-	-	-	-	Tangent ^{1}H
Time / ms	-	-	-	-	1.4

Conclusion and perspective

Conclusion and perspective

We produced HBV core protein (Cp) samples using both cell-based and cell-free expression systems. The *E. coli* expression system is the most common tool for structural studies due to its low cost and high yield. Preformed capsids isolated from *E. coli* cells can be directly used or be dissociated and reassembled for solid-state NMR detection. With these samples, we characterized different subunits of HBV capsids, structural differences between Cp dimers and capsids, as well as a new capsid conformation, which might correlate with viral maturation and envelopment. We have introduced the wheat germ cell-free (WGE-CF) system to the HBV capsid as a promising tool to study capsid assembly modulation on exit from the ribosome. Indeed, compared with bacterial expression systems, WGE-CF system not only can produce preformed capsids with efficient isotopic labeling and high purity, but also make it possible to apply solid-state NMR to investigate the interactions of capsid with capsid assembly modulators in simulating eukaryotic cell environment.

Only matured capsids can be enveloped and secreted *in vivo*, but the details of HBV maturation mechanism are still not fully understood. In this process, the preS region of surface protein L is considered to interact with matured capsid. In this context, it is interesting to analyze the interaction between both proteins by NMR in a future work. The extensive knowledge in sample preparation acquired during this work shall form the basis for investigations of the different capsid forms analyzed here. The current work opens the possibility to investigate samples, prepared from purified preS peptides and preformed capsids, or addition of preS peptides to the WGE-CF system, or even co-expression of Cp and preS in the WGE-CF system. In another direction, characterization of capsid-nucleic acid interactions involving the C-terminal domain will be a central theme in the lab in coming years, and shall use the established bases in sample preparation for the investigation of this interaction by NMR.

Conclusion et perspectives

Nous avons produit des échantillons de la protéine core (Cp) du VHB en utilisant des systèmes d'expression cellulaires et acellulaires.

Le système d'expression en bactéries *E. coli* est l'outil le plus commun pour les études structurales grâce à son bas coût et son rendement élevé. Des capsides auto-assemblées dans *E. coli* peuvent être utilisées directement, ou bien être dissociées puis réassemblées pour être ensuite être étudiées par RMN du solide. Avec ces échantillons, nous avons identifié les différentes sous-unités des capsides du VHB, caractérisé les différences entre les protéines core sous forme de dimère ou de capside, et découvert une nouvelle conformation de la capside, qui pourrait être corrélée à l'enveloppement et la maturation du virus. Nous avons introduit le système d'expression acellulaire à base de germes de blé pour produire les capsides du VHB comme un outil prometteur pour étudier la modulation de l'assemblage des capsides directement à la sortie du ribosome. En effet, comparé aux systèmes d'expression en bactéries, le système acellulaire peut produire des capsides avec un marquage isotopique efficace et une grande pureté, mais rend également possible l'étude en RMN du solide des interactions entre les capsides et les modulateurs d'assemblage dans un environnement simulant une cellule eucaryote.

Seules les capsides matures peuvent être enveloppées et secrétées *in vivo*, mais les détails concernant les mécanismes de maturation du VHB ne sont pas encore compris. Durant ce processus, la région preS de la protéine de surface L interagirait avec la capside mature. Dans ce contexte, il serait intéressant dans le futur d'analyser l'interaction entre ces deux protéines par RMN. Les vastes connaissances acquises en préparation d'échantillons tout au long de cette thèse constituent la base des études montrées ici sur toutes les différentes formes de capsides. Ce travail ouvre la voie vers l'étude d'autres échantillons, préparés à partir de peptides de preS et de capsides auto-assemblées, ou bien de l'ajout de preS directement dans le système acellulaire,

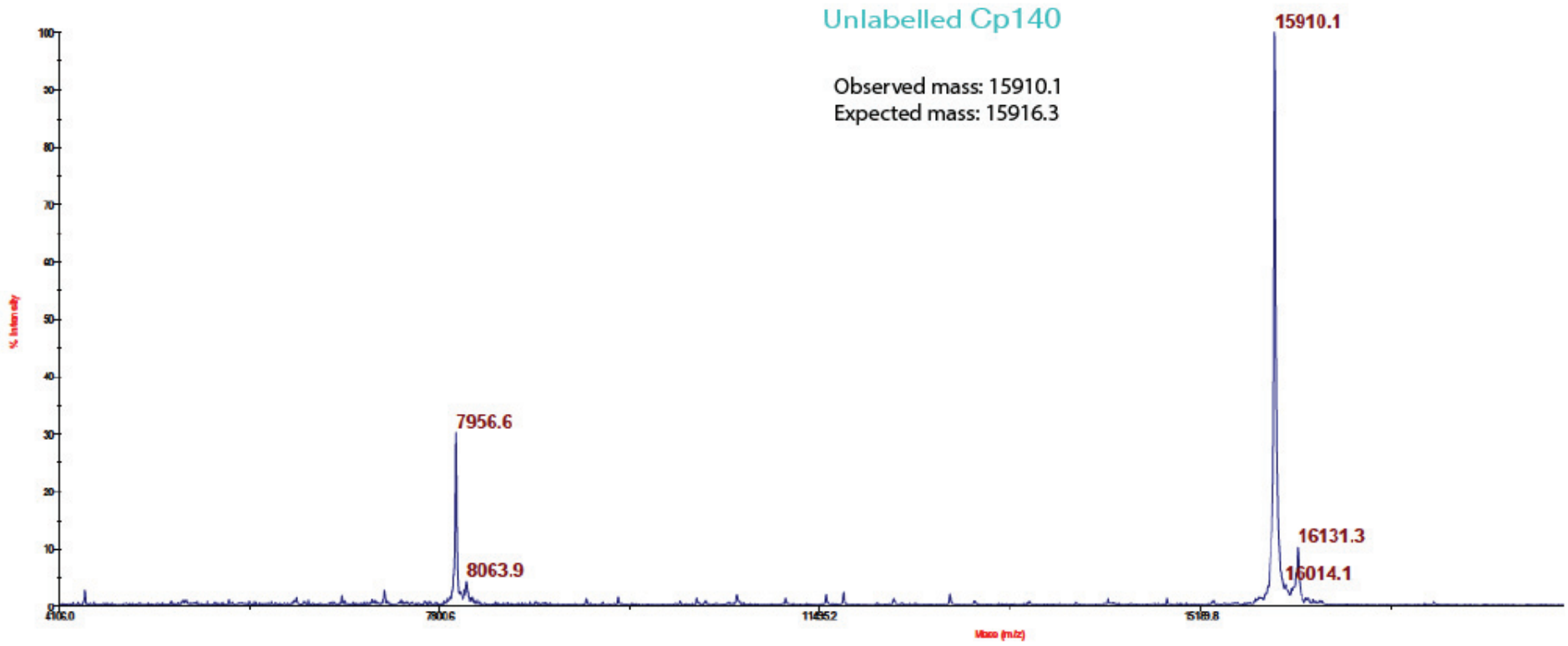
ou encore de la co-expression de capsides et preS dans ce système. Dans une autre direction, la caractérisation des interactions entre les acides nucléiques et le domaine C-terminal de la capside sera un thème central au laboratoire dans les prochaines années, et les bases établies durant la thèse présentée ici seront cruciales pour atteindre ces nouveaux objectifs.

Annexes

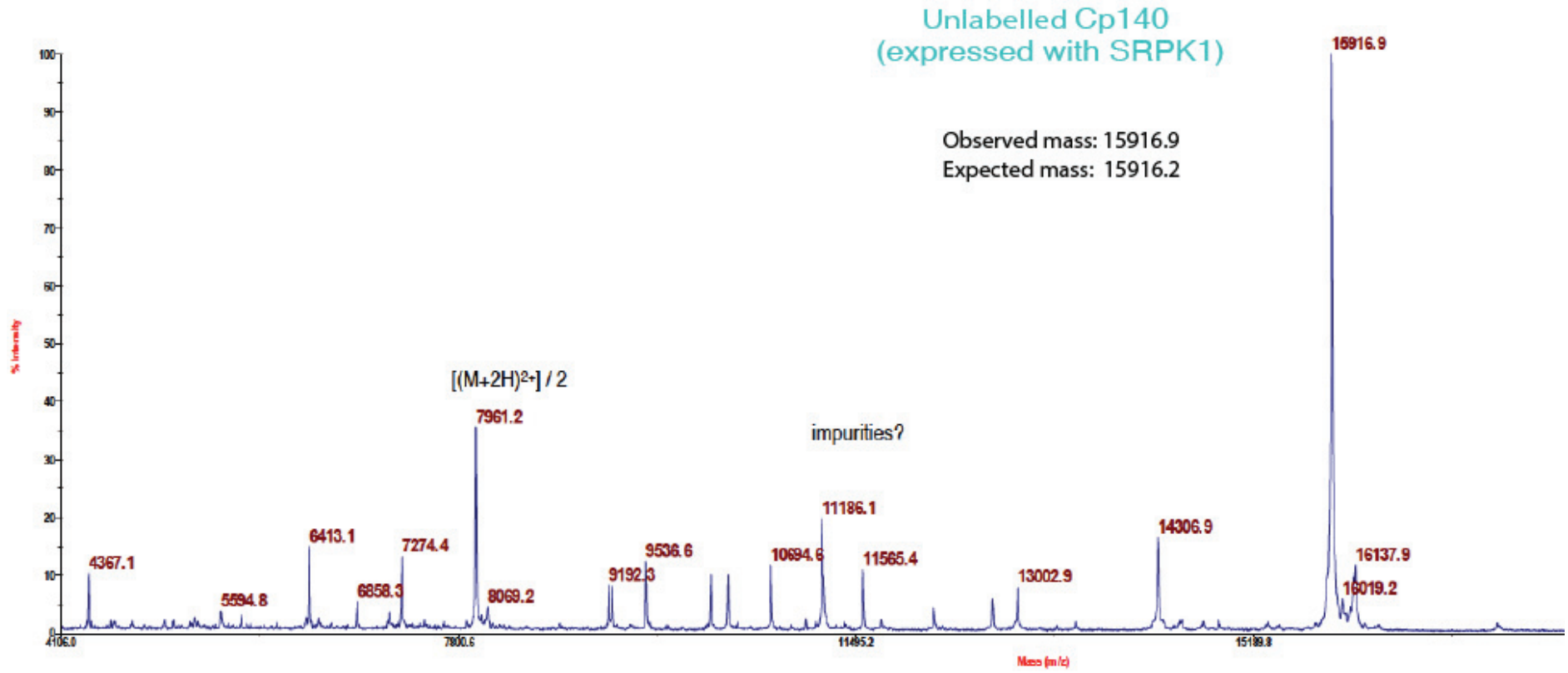
Cps	Sequences
Cp140	MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSP HHTALRQAILCWGELMTLATWVGVNLEDPASRDLVVSYVNTNMGLKFR QLLWFHISCLTFGRETVIEYLVSFGVWIRTPPAYRPPNAPIL
Cp140 C61A	MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSP HHTALRQAILAWGELMTLATWVGVNLEDPASRDLVVSYVNTNMGLKFR QLLWFHISCLTFGRETVIEYLVSFGVWIRTPPAYRPPNAPIL
Cp149	MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSP HHTALRQAILCWGELMTLATWVGVNLEDPASRDLVVSYVNTNMGLKFR QLLWFHISCLTFGRETVIEYLVSFGVWIRTPPAYRPPNAPILSTLPETTVV
Cp183	MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSP HHTALRQAILCWGELMTLATWVGVNLEDPASRDLVVSYVNTNMGLKFR QLLWFHISCLTFGRETVIEYLVSFGVWIRTPPAYRPPNAPILSTLPETTVV RRRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQC
P-Cp183	MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSP HHTALRQAILCWGELMTLATWVGVNLEDPASRDLVVSYVNTNMGLKFR QLLWFHISCLTFGRETVIEYLVSFGVWIRTPPAYRPPNAPILSTLPETTVV RRRGR**S**PRRR**T**P**S**PRRRRSQ**S**PRRRRSQSRESQC
Cp183 F97L	MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSP HHTALRQAILCWGELMTLATWVGVNLEDPASRDLVVSYVNTNMGLK**L**R QLLWFHISCLTFGRETVIEYLVSFGVWIRTPPAYRPPNAPILSTLPETTVV RRRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQC

Cp183 EEE	MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSP HHTALRQAILCWGELMTLATWVGVNLEDPASRDLVVSYVNTNMGLKFR QLLWFHISCLTFGRETVIEYLVSFGVWIRTPPAYRPPNAPILSTLPETTVV RRRGREPRRRTPEPRRRRSQEPRRRRSQSRESQC
HBV/G Cp	MGMDRTTLPYGLFGLDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASA LYRESLESSDHCSPHHTALRQAILCWGELMTLATWVGNNLEDPASRDL VVNYVNTNMGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPAYRP PNAPILSTLPETTVVRRRGRSPRRRTPSPRRRRSQSPRRRRSASPASQ CLEHHHHHH

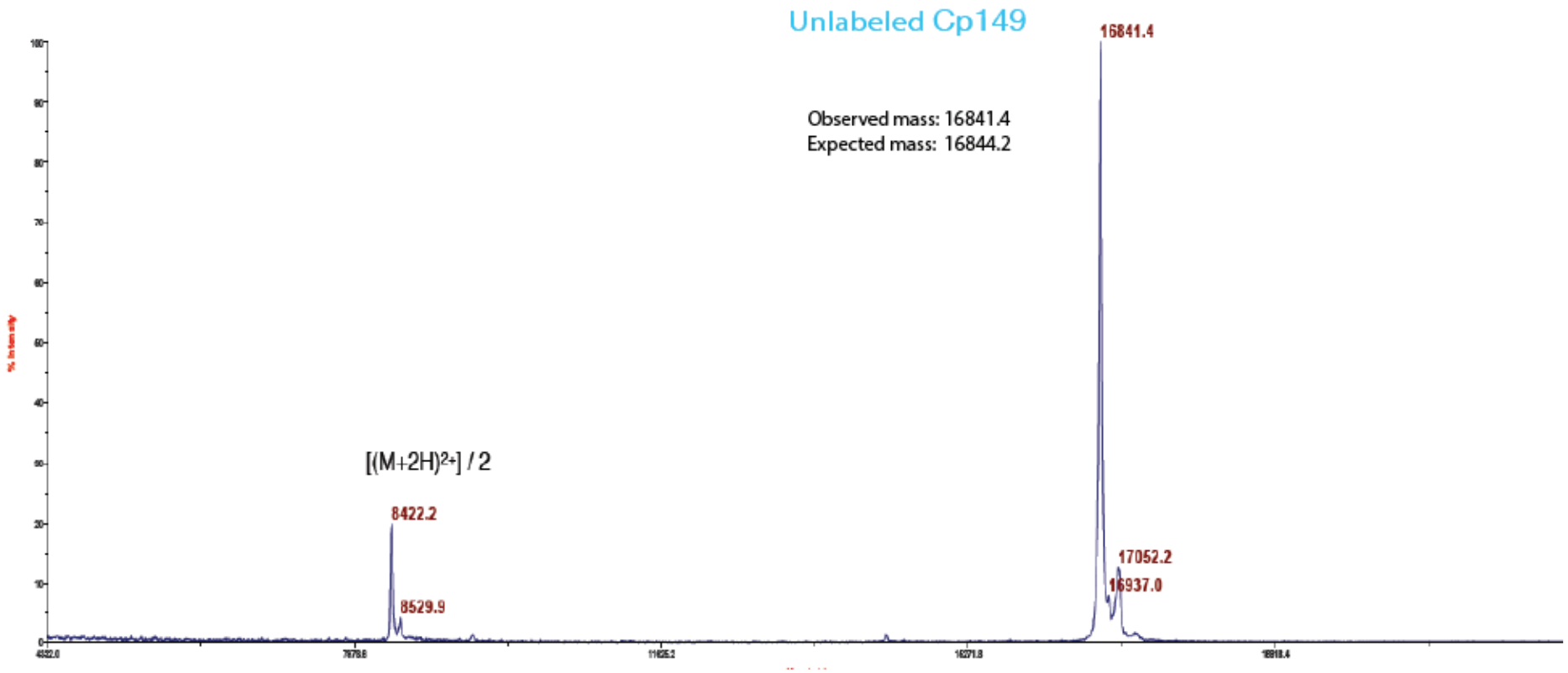
Annexe 1: Sequence of HBV core proteins used in the thesis. The amino acid mutation sites are highlighted in red. The 3 major (bold) and 4 minor phosphorylation sites of P-Cp183 are also highlighted by red. For HBV/G Cp, the redundant sequences including a His-tag derived from plasmid are indicated in yellow.



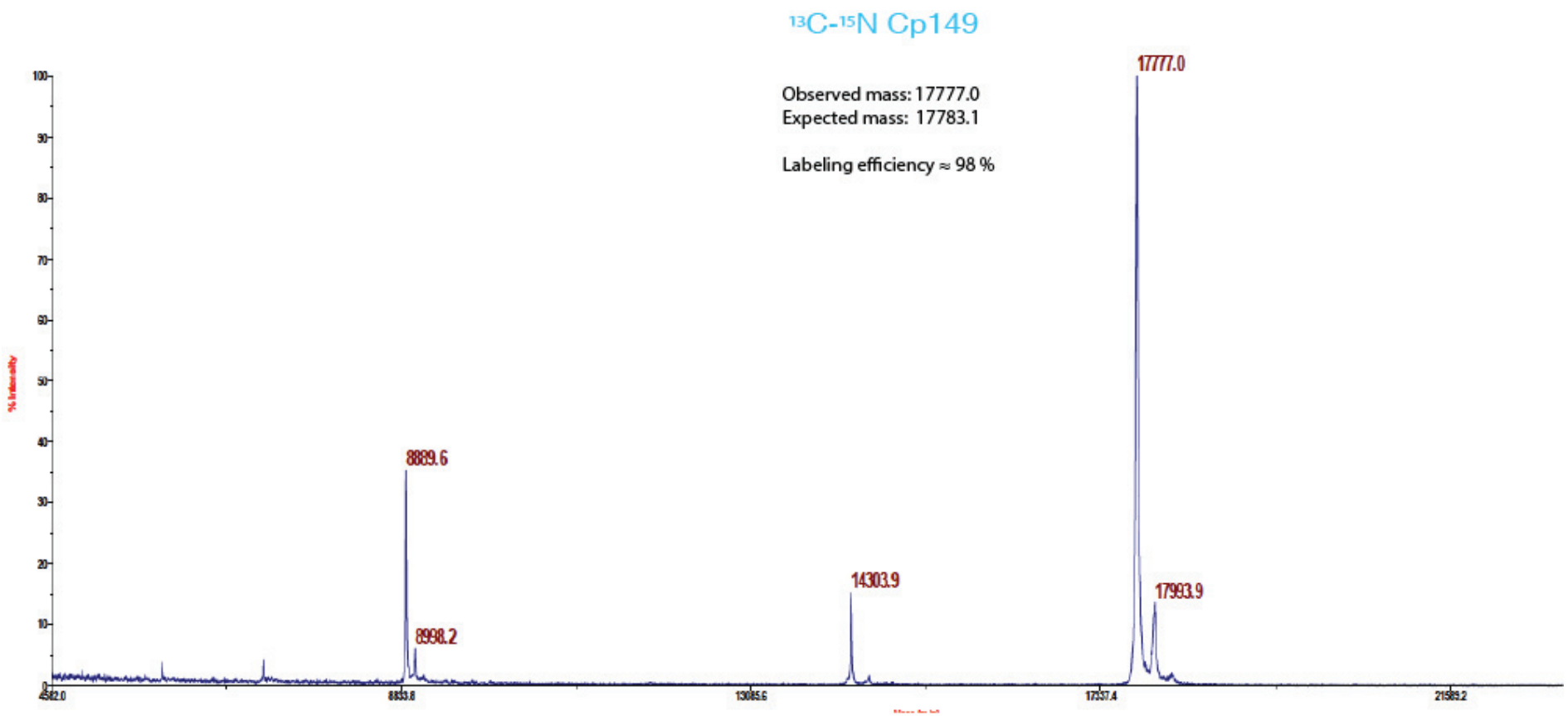
Annexe 2: Matrix Assisted Laser Desorption Ionisation - Time of Flight (MALDI-TOF) spectrum of natural abundance Cp140.



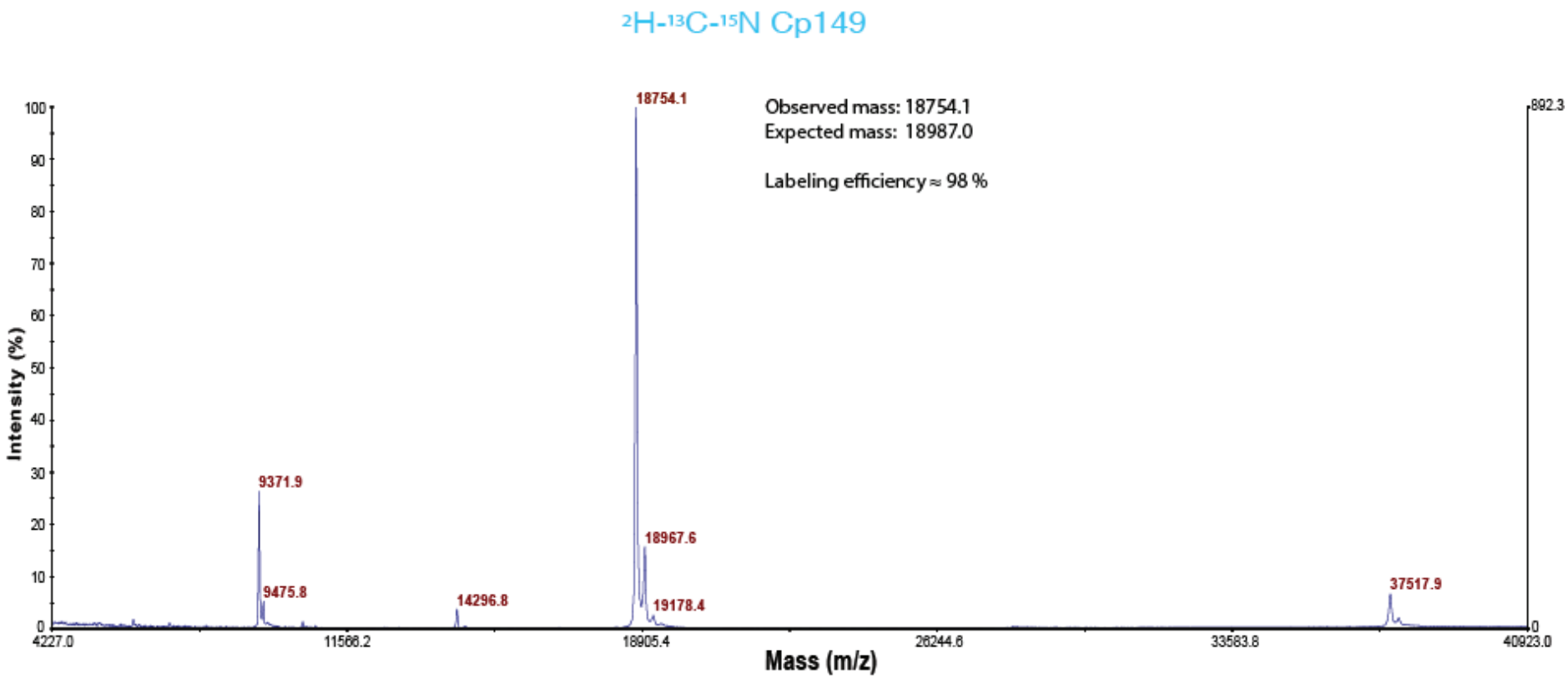
Annexe 3: MALDI-TOF spectrum of natural abundance Cp140 co-expressed with SRPK1 enzyme. The protein co-expressed with SRPK1 matched well with the mass of Cp140, which means that Cp140 is not phosphorylated by SRPK1.



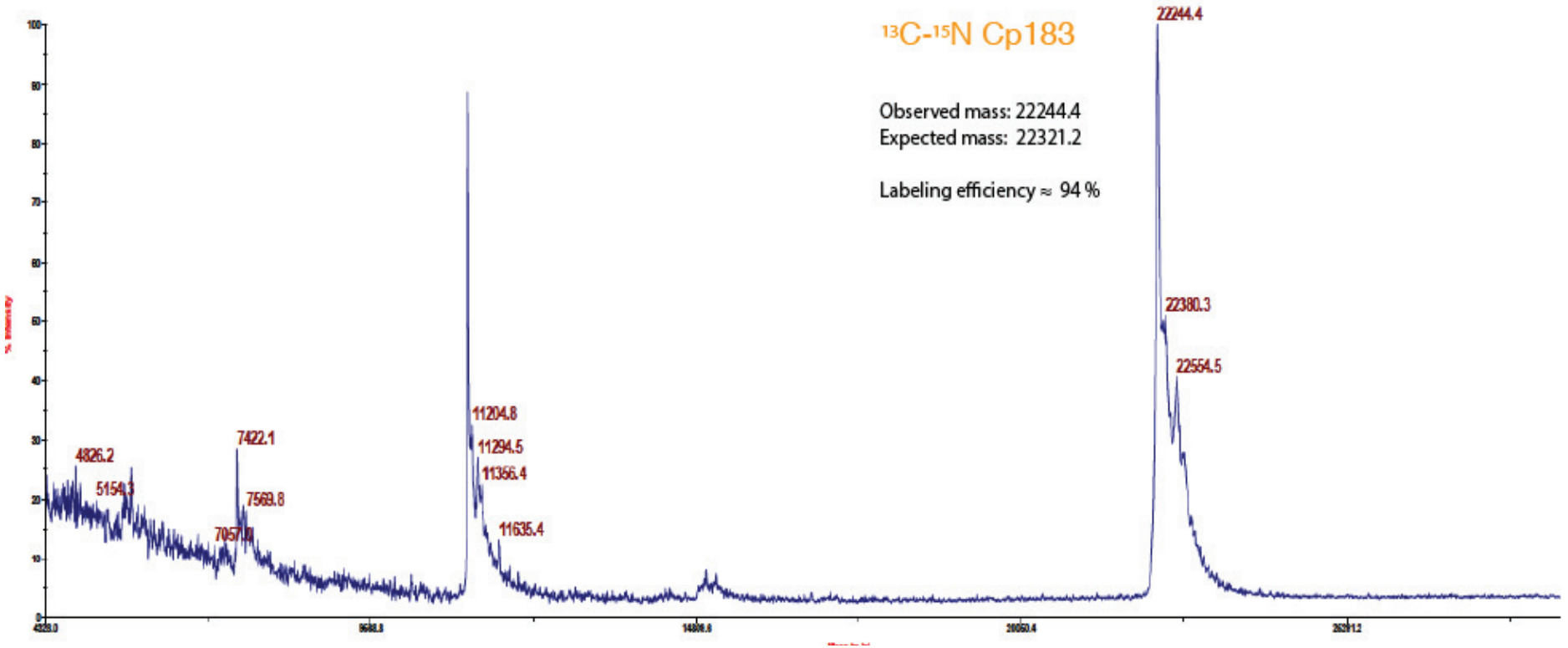
Annexe 4: MALDI-TOF spectrum of natural abundance Cp149.



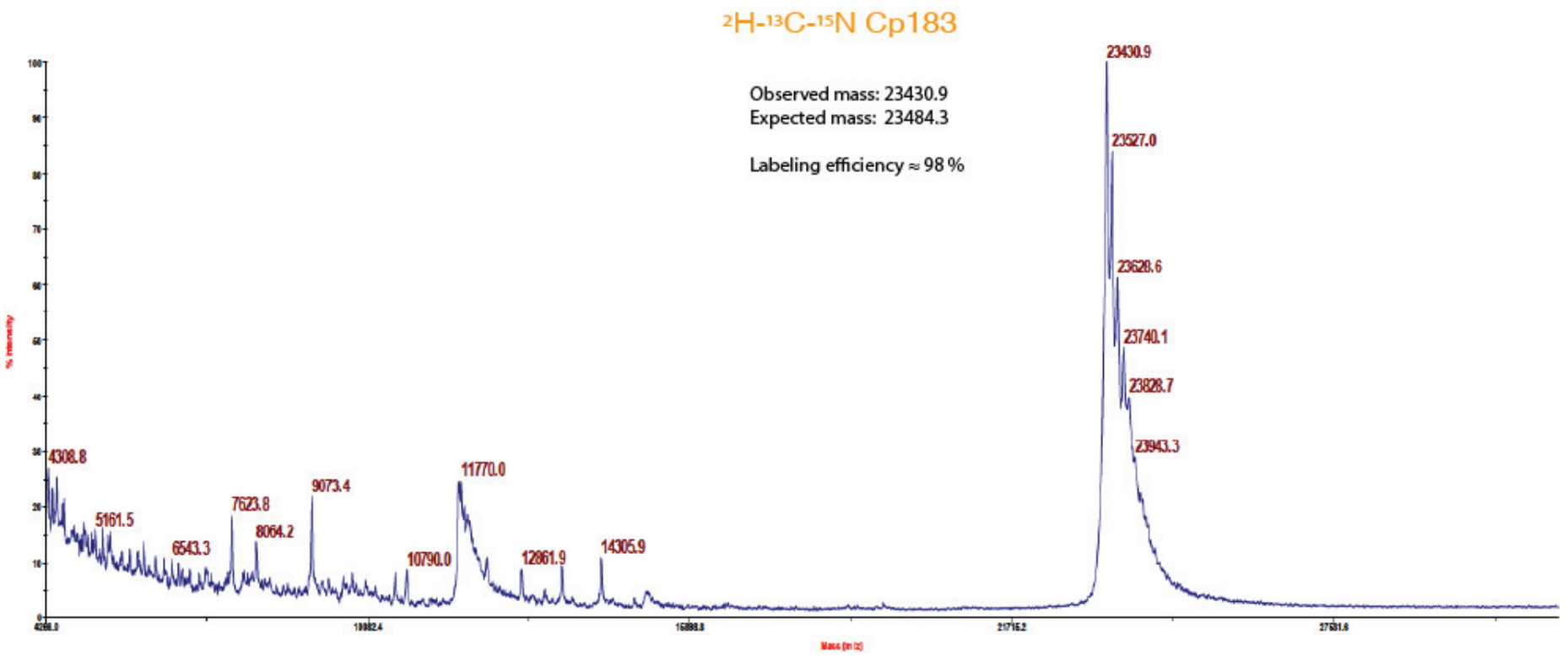
Annexe 5: MALDI-TOF spectrum of ^{13}C-^{15}N Cp149.



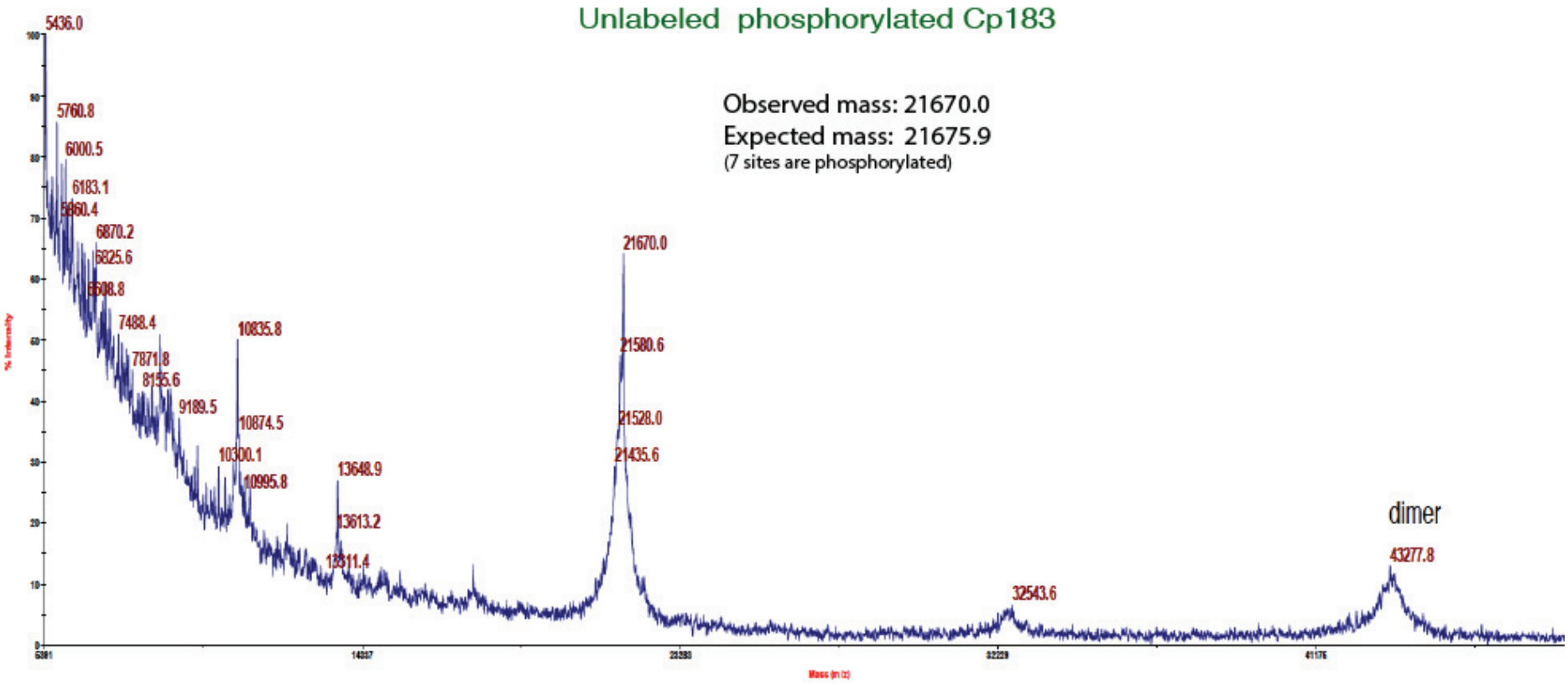
Annexe 6: MALDI-TOF spectrum of ^{2}H-^{13}C-^{15}N Cp149.



Annexe 7: MALDI-TOF spectrum of ^{13}C-^{15}N Cp183.

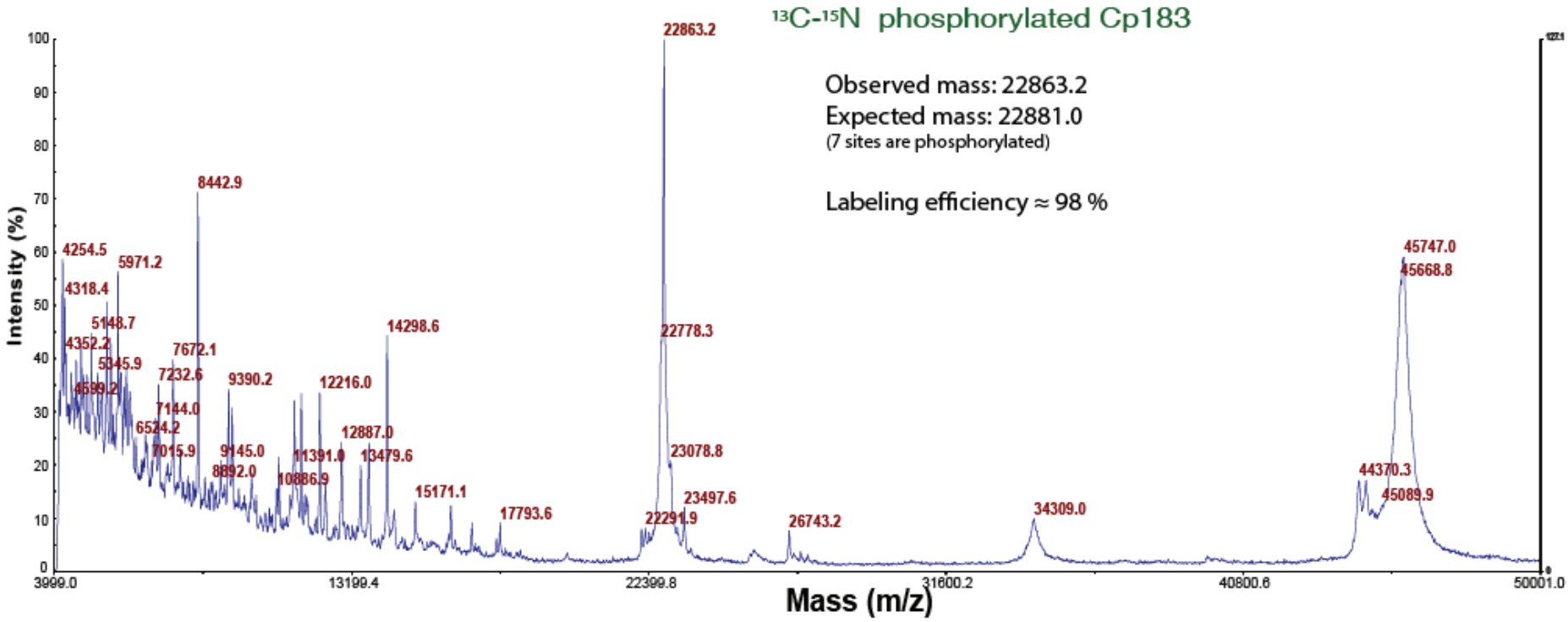


Annexe 8: MALDI-TOF spectrum of ^{2}H $^{-13}$C-^{15}N Cp183.

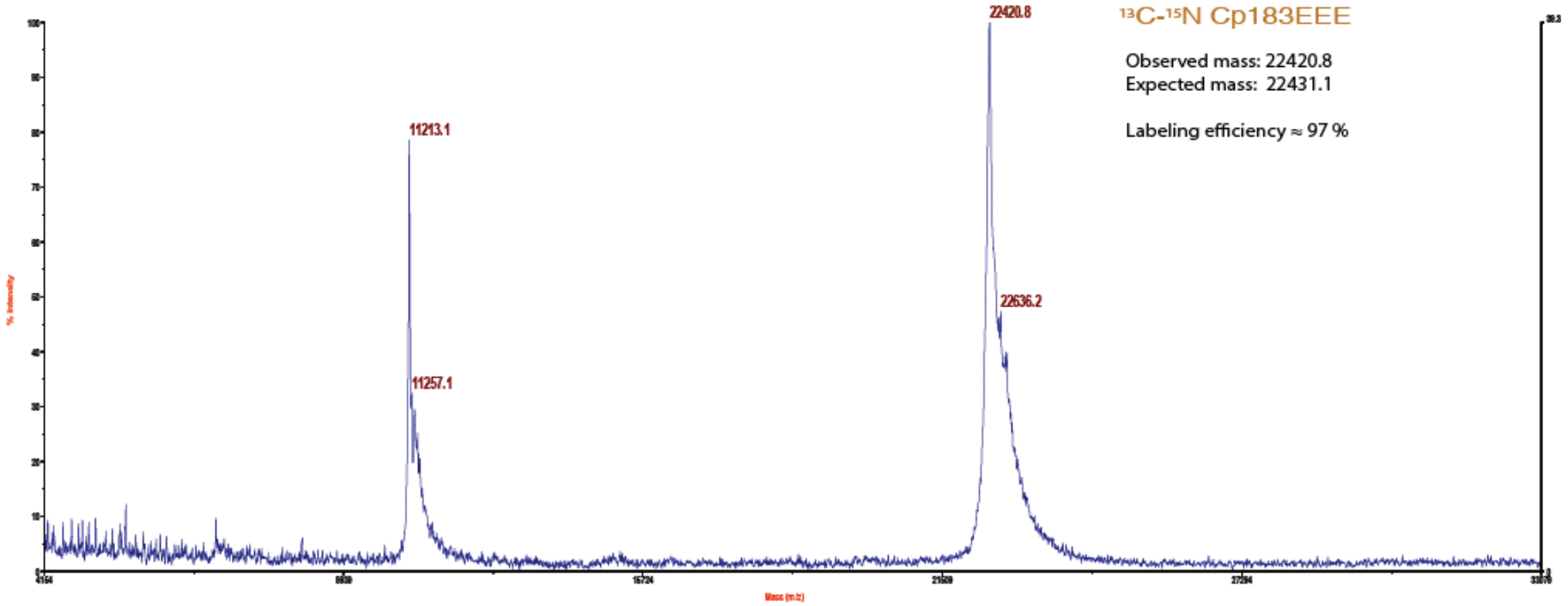


Annexe 9: MALDI-TOF spectrum of natural abundance P-Cp183. The phosphorylated Cp183 was produced by coexpression of wide-type Cp183 and SRPK1 enzyme.

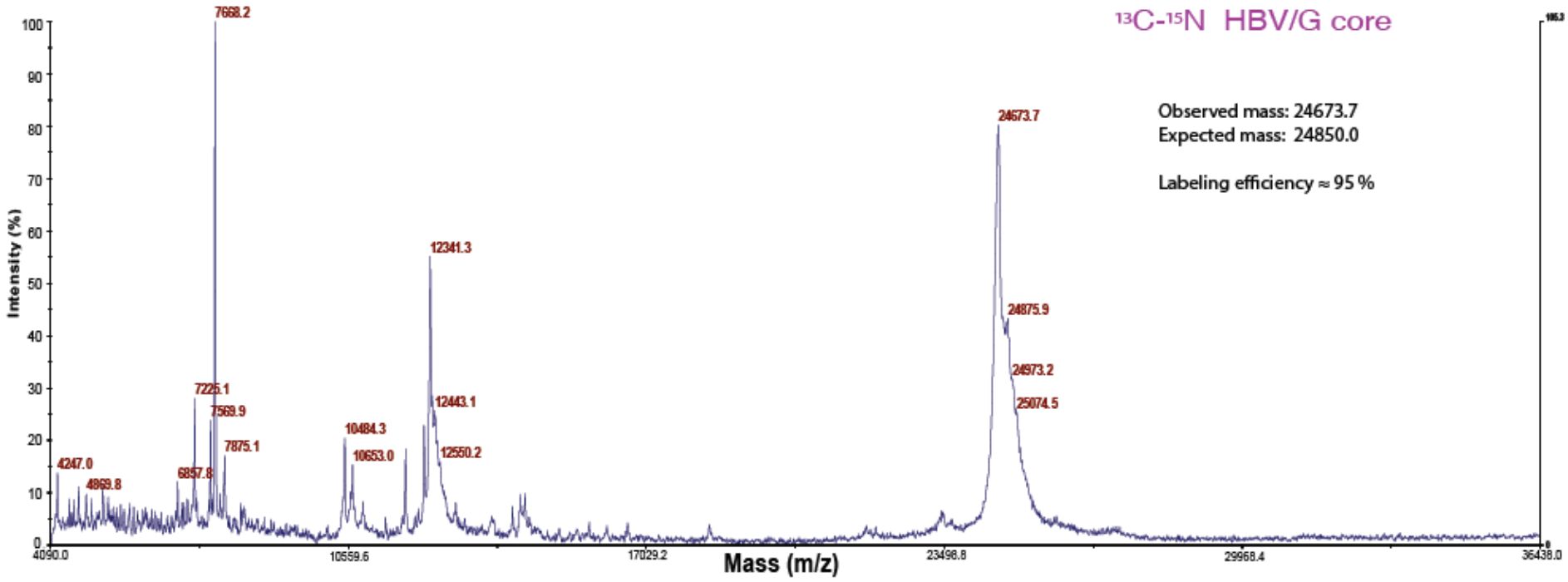
According to the mass spectrum, 7 phosphorylation sites on the CTD are phosphorylated.



Annexe 10: MALDI-TOF spectrum of ^{13}C-^{15}N phosphorylated Cp183. The 7 phosphorylation sites are phosphorylated.



Annexe 11: MALDI-TOF spectrum of ^{13}C-^{15}N Cp183 EEE.



Annexe 12: MALDI-TOF spectrum of ^{13}C-^{15}N HBV/G core protein.