



HAL
open science

Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus

Marie-Laure Fogeron, Vlastimil Jirasko, Susanne Penzel, David Paul, Roland Montserret, Clément Danis, Denis Lacabanne, Aurélie Badillo, Jérôme Gouttenoire, Darius Moradpour, et al.

► To cite this version:

Marie-Laure Fogeron, Vlastimil Jirasko, Susanne Penzel, David Paul, Roland Montserret, et al.. Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus. *Journal of Biomolecular NMR*, 2016, 65 (2), pp.87-98. 10.1007/s10858-016-0040-2 . hal-03347523

HAL Id: hal-03347523

<https://hal.science/hal-03347523v1>

Submitted on 7 Jun 2024

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

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

Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus

Journal Article

Author(s):

Böckmann, Anja; Fogeron, Marie-Laure; Jirasko, Vlastimil; Penzel, Susanne; Paul, David; Montserret, Roland; Danis, Clément; Lacabanne, Denis; Badillo, Aurélie; Gouttenoire, Jérôme; Moradpour, Darius; Bartenschlager, Ralf; Penin, François; Meier, Beat H.

Publication date:

2016-06

Permanent link:

<https://doi.org/10.3929/ethz-b-000118305>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Originally published in:

Journal of Biomolecular NMR 65(2), <https://doi.org/10.1007/s10858-016-0040-2>

Funding acknowledgement:

146757 - NMR studies in the Solid State (SNF)

159707 - NMR studies in the Solid State (SNF)

Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus

Marie-Laure Fogeron^{1‡}, Vlastimil Jirasko^{2,3‡}, Susanne Penzel^{2‡}, David Paul³, Roland Montserret¹, Clément Danis¹, Denis Lacabanne¹, Aurélie Badillo^{1,5}, Jérôme Gouttenoire⁴, Darius Moradpour⁴, Ralf Bartenschlager³, François Penin¹, Beat H. Meier^{2*}, and Anja Böckmann^{1*}

¹*Institut de Biologie et Chimie des Protéines, Bases Moléculaires et Structurales des Systèmes Infectieux, Labex Ecofect, UMR 5086 CNRS, Université de Lyon, 7 passage du Vercors, 69367 Lyon, France*

²*Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland*

³*Department of Infectious Diseases, Molecular Virology, Heidelberg University, Im Neuenheimer Feld 345, 69120 Heidelberg, Germany and German Centre for Infection Research (DZIF), partner site Heidelberg, Heidelberg, Germany*

⁴*Division of Gastroenterology and Hepatology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland*

⁵*RD-Biotech, Recombinant Protein Unit, 3 rue Henri Baigue, 25000 Besançon, France*

* Corresponding authors: AB: a.boeckmann@ibcp.fr, BHM: beme@ethz.ch

‡ These authors contributed equally

Keywords

Cell-free protein expression - Isotope labeling - Lipid reconstitution - Integral membrane protein - NS4B - Solid-state NMR

Abstract

We describe the expression of the hepatitis C virus (HCV) nonstructural protein 4B (NS4B), which is an integral membrane protein, in a wheat germ cell-free system, the subsequent purification and characterization of NS4B and its insertion into proteoliposomes in amounts sufficient for multidimensional solid-state NMR spectroscopy. First spectra of the isotopically [$^2\text{H}, ^{13}\text{C}, ^{15}\text{N}$]-labeled protein are shown to yield narrow ^{13}C resonance lines and a proper, predominantly α -helical fold. Clean residue-selective leucine, isoleucine and threonine-labeling is demonstrated. These results evidence the suitability of the wheat germ-produced integral membrane protein NS4B for solid-state NMR. Still, the proton linewidth under fast magic angle spinning (MAS) is broader than expected for a perfect sample and possible causes are discussed.

Introduction

Solid-state NMR is an emerging method to investigate the structure of membrane proteins in a native-like lipid environment (for recent contributions see for example (McDermott 2009; Renault et al. 2010; Kaur et al. 2015; Hansen et al. 2015; Shahid et al. 2015; Mehler et al. 2015; Mandal et al. 2015; Fu et al. 2015)). Still, as for other techniques in structure biology, sample preparation of membrane proteins remains challenging, as it is often difficult to express the proteins in a well-folded form, with the isotopic labeling and in the milligram amounts needed for NMR spectroscopy. This is especially true for eukaryotic membrane proteins, which are of particular interest as many of them are linked to human diseases. For protein expression in general, the simplest and most cost-effective method is based on *in vitro* systems using *E. coli*. For eukaryotic proteins, it can be advantageous to use eukaryotic expression systems, e.g. *P. pastoris* (Emami et al. 2013), insect cells (van Oers et al. 2015; Opitz et al. 2015) and mammalian cells (Andréll and Tate 2013). However, the membrane protein may still be found in inclusion bodies and folded in a non-native and heterogeneous manner, and expression levels may be low, in particular when the protein is toxic to the organisms. In that case, cell-free production is an interesting alternative (Betton 2004; Vinarov et al. 2004; Klammt et al. 2012), as in particular it offers the opportunity for complex labeling schemes for NMR (Kainosho et al. 2006; Tonelli et al. 2011). While cell-free expression using *E. coli* extracts remains the most widely used and successful approach, an interesting alternative for eukaryotic proteins is the wheat germ cell-free expression system (Sawasaki et al. 2002; Endo and Sawasaki 2006; Takai et al. 2010; Noirot et al. 2011; Fogeron et al. 2015a). Indeed, it combines the advantages of cell-free expression with the benefits of the slow translation rates inherent to eukaryotic ribosomes, thus being more apt in supporting adequate folding, although it has been shown that also in *E. Coli* lysates the translation rates are slower than *in vivo*.

We here demonstrate wheat germ cell-free expression of the non-structural protein 4B (NS4B) from the hepatitis C virus (HCV), a 27-kDa integral membrane protein of 261 amino acid residues, which is only poorly expressed in bacteria. While there is no 3D structure available, a topology model, as shown in Figure 1, has been proposed (Gouttenoire et al. 2014). In this model, membrane association of NS4B is mediated not only by the transmembrane domain in its central part but the protein equally shows determinants for membrane association in the N- and C-terminal regions (Moradpour and Penin 2013). Indeed, as shown in Figure 1A, the N-terminal region is composed of two amphipathic α -helices, AH1 (amino acids 4 to 32; Gouttenoire et al. 2014) and AH2 (amino acids 42 to 66; Gouttenoire et al. 2009a). In the C-terminal domain, a highly conserved α -helix is predicted between amino acids 201 and 212 and the structure of an amphipathic α -helix spanning amino acids 229 to 253 was resolved in isolated form (Gouttenoire et al. 2009a).

NS4B contributes to the formation of the membranous web, a specific membrane alteration consisting of locally confined membranous vesicles that serves as a scaffold for the HCV replicase (Egger et al. 2002; Gosert et al. 2003; Lundin et al. 2006; Gouttenoire et al. 2009a; Gouttenoire et al. 2014). As a master organizer of HCV replication complex formation (Paul et al. 2011), NS4B has been identified as a target for possible treatment modalities of HCV infection (Rai and Deval 2011). Structural information on full-length NS4B would allow to gain further insight into the multiple functions of this protein, which might be governed by distinct membrane topologies and/or interactions with other viral and cellular proteins (Moradpour and Penin 2013; Briggs et al. 2015).

We here show that chemically pure and homogenous samples of NS4B reconstituted in liposomes can be obtained using wheat germ cell-free expression, and that the preparations do yield high-quality ^{13}C NMR spectra showing features characteristic of a protein with the expected secondary-structure signatures predicted for this protein. The resolution in the

proton dimension in fast MAS experiments is however still not as good as found in model systems (Huber et al. 2011; Barbet-Massin et al. 2014; Agarwal et al. 2014; Penzel et al. 2015) and possible reasons are discussed.

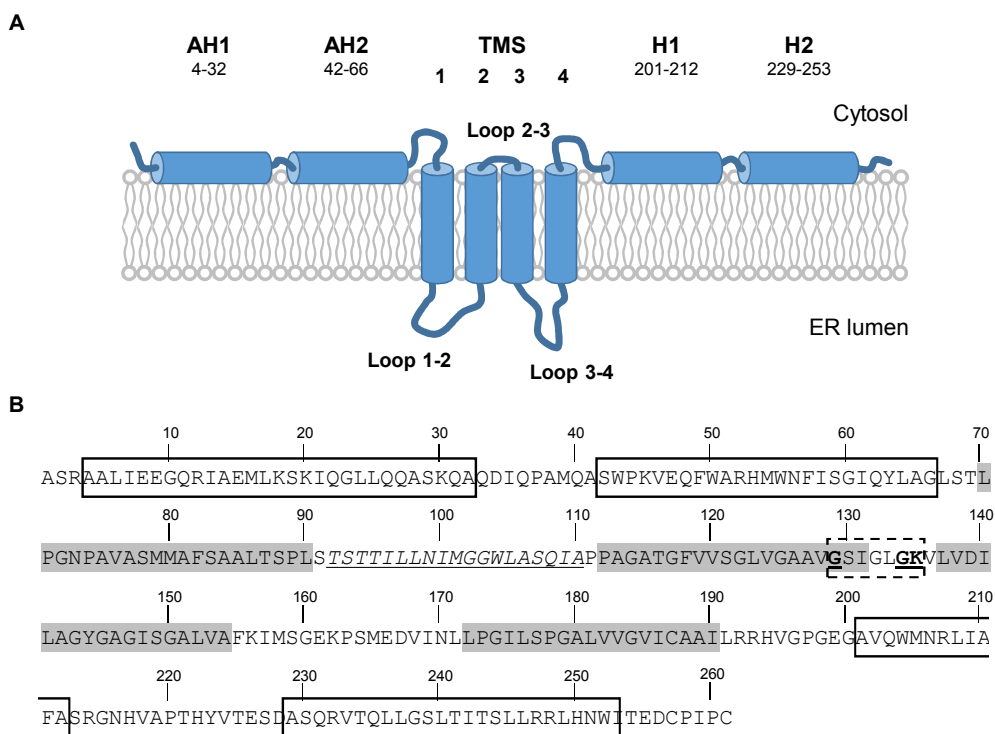


Figure 1: Putative topology model and primary structure of the NS4B protein of the HCV isolate JFH-1 (Gouttenoire et al. 2014). (A) NS4B is composed of three subdomains: N-terminal (aa 1-69), central (aa 70-190) and C-terminal subdomain (aa 191-261). In this model, the N terminus that resides on the cytoplasmic side of the ER membrane contains two amphipathic α -helices AH1 (aa 4-32, PDB accession - 2LVG) and AH2 (aa 42-66, 2JXF). The central part contains four predicted transmembrane segments (Gouttenoire et al. 2010). The C terminus harbors two other likely amphipathic helices: H1, predicted between aa 201-212, and H2, which structure has been determined (aa 229-253, PDB entry 2KDR). (B) NS4B

primary structure. Alpha helices analyzed by solution-state NMR in a context of short peptides (Gouttenoire et al. 2009b) and predicted H1 helix are labeled by solid boxes. Predicted transmembrane segments are shown with grey-highlighted background (Gouttenoire et al. 2010). Loop 1-2 which was predicted to be a putative transmembrane domain is underlined and in italic (Palomares-Jerez et al. 2012). The phosphate binding loop motif in the region of loop 2-3 is indicated with the dotted box (Einav et al. 2004). The amino-acid composition of the protein is shown in Fig. S1.

Materials and methods

Plasmids

cDNA of full-length NS4B from HCV strain JFH-1(2a), GenBank accession number AB047639, was amplified by PCR and inserted into the pEU-E01-MCS vector (CellFree Sciences, Japan). A thrombin cleavage site followed by a Gly-Ser-Ala linker and a *Strep*-tag II, yielding to the final amino-acid sequence LVPRGSAWSHPQFEK, are fused at the C terminus of the full-length NS4B proteins. For the NMR sample preparation, a NS4B_Twin-*Strep*-Tag protein containing a tandem of *Strep*-tag II sequences separated by flexible linker (LVPRGSAWSHPQFEKGGGSGGGSGGSAWSHPQFEK) was used. As detailed in (Fogeron et al. 2015b; Fogeron et al. 2015a), the resulting plasmids were transformed into *Escherichia coli* TOP10 chemically competent cells (Life Technologies) and DNA was prepared using a NucleoBond Xtra Maxi kit (Macherey-Nagel, France). Plasmids were further purified by a phenol/chloroform extraction according to CellFree Sciences (Yokohama, Japan) recommendations.

Wheat germ cell-free protein expression

Wheat germ cell-free protein expression and sample preparation were performed as described in (Takai et al. 2010; Fogeron et al. 2015b; Fogeron et al. 2015a). In short, home-made wheat germ extracts (Noirot et al. 2011) were prepared according to Takai et al. (Takai et al. 2010). Untreated durum wheat seeds from different varieties, namely Dakter, Miradoux or Clovis wheat strains, were purchased from Sud Céréales (ARTERRIS), France, and stored at room temperature for up to one year. Main steps of WGE preparation included grinding and sieving of the wheat seeds, solvent flotation, eye selection of the embryo particles followed by extensive washing, extraction, fractionation and buffer exchange. Twenty kilograms wheat seeds led to the isolation of about 20 g wheat germs, allowing the production of 30 mL WGE within 3 working days, including all steps detailed above. Eye selection of the wheat germs takes about 1 h/g, and is done about every second month during a lab meeting (approx. 8 people), resulting in 5 h work/person and a final yield of 40 g, enough for two subsequent preparations of 30 ml WGE.

Transcription and translation were carried out separately, as described in (Takai et al. 2010) and (Noirot et al. 2011), and translation was performed using the so-called bilayer method. When present, MNG-3 detergent was added to both the reaction mix and the feeding buffer. The reaction mixture was incubated at 22 °C for 16 h without shaking. Translation was performed either in 96-well plates for small-scale expression test or 6-well plates for larger production. The volume of the bottom layer corresponding to the translation mix was 20 µL and 500 µL per well in 96-well and 6-well plates, respectively, whereas the volume of the upper layer corresponding to the feeding buffer was 200 µL and 5.5 mL per well in 96-well and 6-well plates, respectively.

SDS-PAGE and western blotting

SDS-PAGE and western blotting analysis was performed as described in (Fogeron et al. 2015a).

Purification of full-length NS4B by affinity chromatography

NS4B was expressed in presence of 0.1% MNG-3 using the bilayer method as described previously (Fogeron et al. 2015a) adapted to wells of 6-well plates (i.e., 0.5 mL of reaction mix containing 0.25 mL mRNA and 40 μ g/ml creatine kinase and 0.25 mL of wheat germ extract for the bottom layer, and 5.5 mL of feeding buffer for the upper layer in each well). After protein synthesis, the CFS (total cell-free sample) of 18 wells was pooled (108 mL) and incubated with benzonase at RT on a rolling wheel for 30 min. In addition, DDM was added to CFS at a final concentration of 0.25% (all concentrations are in w/w), allowing buffer exchange from MNG-3 to DDM, a detergent standardly used in lipid reconstitution protocols. CFS was then centrifuged at 20.000 g, 4 °C for 30 min and the supernatant was loaded on a 5-mL *Strep*-Tactin high capacity gravity column (IBA Lifesciences, Germany). Purification was performed as specified by the manufacturer; all buffers contained 0.1% DDM. Full-length NS4B was eluted in 100 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 0.1% DDM and 2.5 mM desthiobiotin. Purification of NS4B_*Twin-Strep*-Tag protein was performed in the presence of 0.2% DDM, and washing steps were carried out at 150 mM and 300 mM NaCl. The final elution buffer contained 100 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 0.2% DDM and 25 mM desthiobiotin.

Size-exclusion chromatography

Size-exclusion chromatography of full-length NS4B previously purified by affinity chromatography was performed using a Superdex™ 200 Increase 3.2/100 column (GE Lifesciences). A 50 mM phosphate buffer pH 7.4 containing 0.1% DDM was used as eluent and 150 μ g protein were loaded on the column. A calibration curve of the size exclusion chromatography column was established according to le Maire et al. (le Maire et al. 1989) using the following protein standards: cytochrome C, 12.3 kDa ; myoglobin, 17.8 kDa, trypsin inhibitor, 22.1 kDa ; β -lactoglobulin, 35 kDa ; transferrin, 81 kDa ; aldolase, 158 kDa ; ferritin, 440 kDa ; thyroglobulin, 669 kDa. A 50 mM phosphate buffer pH 7.4 containing 0.1% DDM was used as eluent and 100 μ g of each protein standard were run separately on the column.

Circular dichroism

Spectra were recorded on a Chirascan spectrometer (Applied Photophysics) calibrated with 1S-(+)-10-camphorsulfonic acid. Measurements were carried out at 25 °C in a 0.1-cm path-length quartz cuvette. Spectra were measured in a 180–260 nm wavelength range with an increment of 0.2 nm, bandpass of 0.5 nm and integration time of 1 s. Spectra were processed, baseline corrected, smoothed and converted with the Chirascan software. Spectral units were expressed as the mean molar ellipticity per residue using the protein concentration determined at 280 nm using a NanoDrop spectrometer. Estimation of the secondary structure content was carried out using the CDSSTR, CONTIN and SELCON3 approaches available on the DICHROWEB server (dichroweb.cryst.bbk.ac.uk/).

Reconstitution of NS4B into proteoliposomes

Either egg PC (L- α -phosphatidylcholine, 99% pure) or a mixture of egg PC and cholesterol (99% pure, 70/30, w/w) were used for lipid reconstitution (Sigma). The lipids were first

solubilized using Triton X-100 with a detergent-to-lipid ratio of 10 (mol/mol) (Kunert et al. 2014). Purified full-length NS4B, solubilized in 0.2% DDM, was then mixed with the lipids at a defined lipid-to-protein ratio (w/w). Protein-lipid-detergent samples were dialyzed at 4 °C for 18 days against a buffer containing 100 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 1 mM DTT and hydrophobic polystyrene beads (Bio-Beads SM-2 adsorbents, Bio-Rad). A dialysis membrane with a molecular weight cut-off of 6-8000 Da (Spectra/Por) was used. A ratio Bio-Beads-to-detergent of 100 (w/w) was applied. Detergent removal by the Bio-Beads leads to the formation of proteoliposomes (Rigaud et al. 1995).

Analysis of reconstituted NS4B by sucrose density-gradient ultracentrifugation

Twenty five μ g of DDM-solubilized NS4B_Twin-Strep-Tag protein was reconstituted into PC or PC/cholesterol lipids solubilized in Triton X-100 with various lipid-to-protein ratios (LPRs). Proteoliposome samples were applied on the discontinuous sucrose gradient (7-70%) and centrifuged 18 h at 200.000 g. Eleven fractions of each gradient (collected from top to bottom) together with a pellet were analyzed by SDS PAGE followed by silver staining. The densities of harvested fractions were calculated from refractive index values, measured using a refractometer.

Sample preparation for solid-state NMR spectroscopy

NS4B_Twin-Strep-Tag was expressed in presence of 0.1% MNG-3 for 16 h at 22 °C in a bilayer reaction of a total of 144 ml (24 bilayer reactions in carried out in 4 6-well plates, containing a total of 6 mL of wheat germ extract). To obtain uniformly [^2H , ^{13}C , ^{15}N]-labeled protein, a mixture of 20 [^2H , ^{13}C , ^{15}N]-labeled amino acids (“Cell-free” amino acid mix 20 AA, Cambridge Isotope Laboratories, Inc.) in final concentration 6 mM (concentrations of individual amino acids are given in Table S3) were added to reaction and feeding buffer. To

prepare selectively labeled Leu-Ile-Thr (LIT) protein, [²H,¹³C,¹⁵N]-labeled leucine, isoleucine and threonine (Cambridge Isotope Laboratories, Inc.) were mixed with the remaining 17 ²H-labeled amino acids with carbon and nitrogen at natural isotopic abundance (Cambridge Isotope Laboratories, Inc.) and added to the reaction and feeding buffer in final concentration of 6 mM (concentrations of individual amino acids see Table S3). Concentrations of each individual amino acid for selectively labeled sample corresponded to the molar ratio of amino acids in NS4B sequence. Amide and exchangeable side-chain protons of amino acids were 100% protonated, the other protons were partially deuterated (vide infra).

After translation, the protein was incubated with benzonase and 0.25% DDM, followed by *Strep*-Tactin column purification in presence of 0.2% DDM. In addition to a standard IBA protocol, a washing step with 5x CV with 100 mM Tris-HCl pH 8, 300 mM NaCl, 1 mM EDTA and 0.2% DDM was included. The final elution buffer contained 100 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 0.2% DDM and 25 mM desthiobiotin. The estimated cost of the 6 mg triply labeled protein preparation used for the NMR sample amounts to 1091 € (see Table S4).

Purified NS4B protein in DDM was reconstituted at LPR 0.25 with PC and cholesterol lipids (70/30, w/w) solubilized in Triton X-100 (lipid detergent molar ratio 1:10) using dialysis in the presence of Bio-Beads SM-2. Detergent removal was performed for 20 days in a buffer containing 100 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT.

Reconstituted proteoliposomes were purified and concentrated by sedimentation onto a 90% OptiPrep™ ($\rho = 1.3 \text{ g/cm}^3$, corresponding to 62% sucrose) cushion overlaid by 30% OptiPrep™ ($\rho = 1.1 \text{ g/cm}^3$, corresponding to 25% sucrose), which separate (i) the lipid fraction, floating on the top of 30% layer, (ii) the proteoliposomes (LPR 0.25, $\rho = 1.19 \text{ g/cm}^3$), which migrate through the 30% layer and remain on the top of the 90% layer, and (iii) aggregated proteins sedimented at the bottom of the centrifugation tube. After sedimentation

on the cushion, the proteoliposomal fraction was harvested, washed with dialysis buffer to remove the OptiPrep™ medium and sedimented into a 1.3 mm Bruker NMR rotor using custom-made tools (Böckmann et al. 2009).

Electron microscopy

The material was adsorbed to carbon- and pioloform-coated 300 mesh copper grids (Science Services GmbH) for 10 min at RT, and remaining liquid was drained using a whatman paper. Immunolabeling against the C-terminal *Strep*-tag II was performed as described in (Fogeron et al. 2015b). Samples were examined by using a Zeiss EM-10 transmission electron microscope (Zeiss, Goettingen, Germany) with a built in Mega View camera (Olympus; Tokyo, Japan).

Solid-state MAS NMR experiments

Experiments were performed on Bruker AVANCE III wide-bore spectrometer operating at 850 MHz ^1H Larmor frequency while employing magic-angle spinning at 60 kHz using a Bruker 1.3 mm MAS probe. DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was used for referencing as internal standard. The sample temperature was based on the supernatant water resonance frequency (Böckmann et al. 2009) and adjusted to ~ 16 °C. The spectra were then processed with TOPSPIN (Bruker, Biospin) and analyzed in CCPN (Vranken et al. 2005).

The solid-state NMR ^{13}C - ^{13}C DREAM correlation spectra (Verel et al. 2001) for the two different NMR samples were acquired for 20 and 10 ms in the direct and indirect dimension, respectively, with 1024 scans for the fully labeled sample and 876 scans for the selectively labeled sample. For a complete set of the experimental parameters, see Table S1. The effective acquisition times (TDEFF) after processing with a quadratic sine bell function

(QSINE) shifted by 2.5, were chosen to be 10 and 8 ms in the direct and indirect dimension, respectively.

Proton-detected CO-N-H correlation experiments on the LIT sample were performed using CP transfer for all steps (Penzel et al. 2015). The total experimental time was about 5 days (118 hours) with 400 scans, a recycle delay of 1 s and acquisition times of 30, 5.2, and 3.0 ms in the ^1H , ^{15}N , and ^{13}C dimension, respectively. Details are given in Table S1.

The proton-detected (H)NH 2D spectrum at 90 kHz MAS was on a 0.8 ppm MAS probe constructed by Ago Samoson and described in (Agarwal et al. 2014) with parameters specified in Table S2.

Results and discussion

Production and purification of full-length NS4B

Full-length NS4B construct from HCV strain JFH-1 (GenBank accession number AB047639) with a *Strep*-tag II and a thrombin cleavage site fused at the C terminus was successfully expressed in a solubilized form in the presence of lauryl maltose neopentyl glycol (MNG-3) (**Fig. 2A**). Due to its hydrophobic nature, NS4B is mainly insoluble in the absence of detergent and therefore detected in the pellet, although a minor portion is found in the supernatant, possibly due to the presence of some residual lipids in the WGE (Fogeron et al. 2015a). However, in the presence of 0.1% MNG-3, fully solubilized JFH-1 full-length NS4B was obtained (**Fig. 2A**). Indeed, the tagged protein was mainly detected in the fraction concentrated by binding of the NS4B C-terminal *Strep*-tag II to magnetic *Strep*-Tactin beads (SN-Beads).

After expression in the presence of MNG-3, the *Strep*-tag II allowed us to obtain > 95% pure NS4B protein (HCV strain JFH-1) using a single affinity purification step starting from the supernatant fraction containing the solubilized protein (**Fig. 2B**). Although full-length NS4B

could be efficiently purified in the presence of MNG-3 (Fogeron et al. 2015a), the addition of 0.25% n-dodecyl β -D-maltoside (DDM) to the CFS treated with benzonase before the affinity purification step allowed us to exchange MNG-3 to DDM, a detergent we routinely use for lipid reconstitution. On average, the purified protein yield was about 1 mg/mL WGE, with a maximum yield of 1.5 mg NS4B per mL WGE being reached.

To further characterize the purified NS4B protein, size exclusion chromatography was performed. A first peak containing NS4B was observed (**Fig. 2C**), while the following peak resulted from reagents present in the *Strep*-Tactin elution buffer. NS4B protein-detergent complexes eluted with an apparent molecular mass of about 50 kDa (**Fig. 2D**), consistent with the NS4B monomer (29.2 kDa) emdedded into detergent. The amount of DDM associated with the protein corresponds to a seizable fraction of a DDM micelle (60-75 kDa).

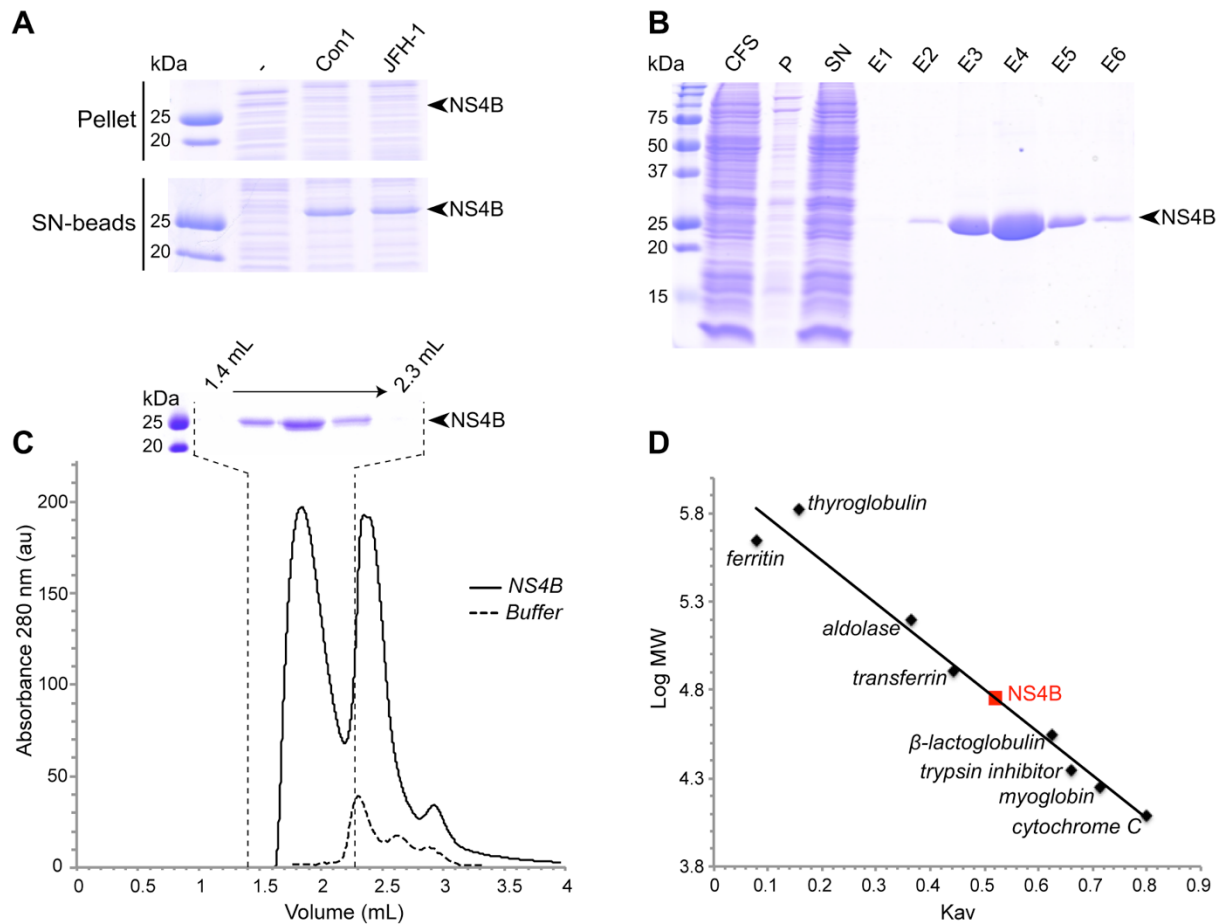


Figure 2: Production of detergent-solubilized full-length NS4B in a wheat germ cell-free expression system and one-step affinity purification. (A) Small scale expression test of full-length NS4B constructs in the presence of MNG-3 detergent. The full-length NS4B construct from HCV strain JFH-1 was expressed in the presence of 0.1% MNG-3. Protein samples were analyzed by SDS-PAGE. -, negative control (no NS4B mRNA); Pellet, pellet obtained after centrifugation of CFS; SN-beads, sample enriched in tagged NS4B by incubation of CFS supernatant with *Strep*-Tactin magnetic beads. Comparable amounts were loaded on the gel for Pellet and SN-beads (*i.e.* 27% of the total 220 μ L reaction mix from a single 96-well-plate reaction). Black arrowheads indicate the bands corresponding to NS4B. (B) Full-length C-terminally *Strep*-tag II tagged NS4B (HCV strain JFH-1) was expressed using 4.5 mL WGE in the presence of 0.1% MNG-3 and purified on a 5-mL *Strep*-Tactin column in the presence of 0.1% DDM. Protein samples were analyzed by SDS-PAGE. CFS, total cell-free sample (8

μL from 108 mL loaded on the gel); P, pellet obtained after centrifugation of CFS (equivalent to 80 μL from 108 mL loaded on the gel); SN, supernatant obtained after centrifugation of CFS and loaded on the affinity column (8 μL from 108 mL loaded on the gel); E1 to E6, affinity elution fractions (8 μL from 2.5 mL loaded on the gel). The black arrowhead indicates the bands corresponding to NS4B. (C) Size exclusion chromatography of the elution fraction E4. A 50 mM phosphate buffer [pH 7.4] containing 0.1% DDM was used as eluent and a flow rate of 75 $\mu\text{L}/\text{min}$ was applied. The elution profile of NS4B is drawn as a solid line, the affinity elution buffer as a dotted line. SDS-PAGE and Coomassie staining analysis of the collected fractions is shown above the chromatographic profile. Elution volume of NS4B is 1.85 mL. (D) Calibration curve of the size exclusion chromatography column established by using protein standards as indicated in the Figure (V_0 is 1.25 mL). NS4B is indicated in red on the standard curve.

Secondary structure analysis

The far UV circular dichroism (CD) spectrum of full-length NS4B eluted from size exclusion chromatography in 0.1% DDM (**Fig. 3**) displays the molar ellipticity per residue in the range that can be expected for such a protein. The two minima at 208 and 222 nm, together with a maximum at 192 nm, are typical of α -helical folding. Spectral deconvolution indicated an α -helix content of 72-79%, while turns represent 7-10%, and β -sheet content is almost nil (0-1%). The high α -helical content is consistent with the predictions shown in Fig.1.

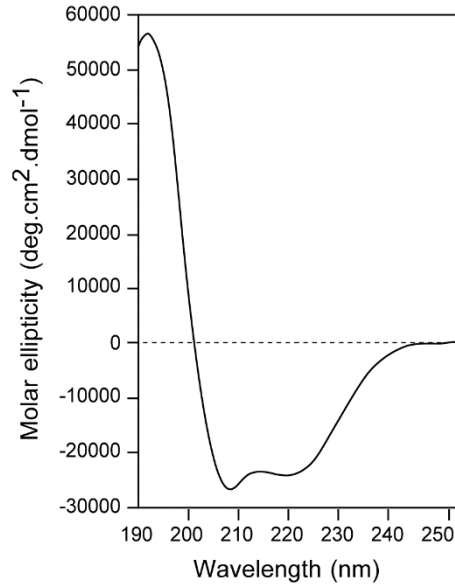


Figure 3: Far UV circular dichroism spectrum of full-length NS4B after purification by affinity chromatography and size exclusion chromatography.

NS4B lipid reconstitution at different LPRs

Proteoliposome reconstitution was performed with a classical dialysis method (see Materials and Methods section) using a PC/cholesterol mixture (70/30, w/w). LPRs between 0 and 1 (weight/weight) were tested. We analyzed the proteoliposomes obtained using a sucrose density gradient. **Fig. 4A** shows the resulting SDS-PAGE/ silver staining analysis of samples harvested after density gradient centrifugation of NS4B reconstituted with LPRs between 0 and 1, where 1 corresponds approximately to a molar LPR of 45. While the majority of the protein is retained in the gradient, a minor fraction (estimated to approximately 10-20%) forms aggregates and is found in the pellet. With decreasing LPRs the proteoliposomes become denser and migrate further into the gradient at higher LPRs, a difference between the density as a function of the lipid composition, PC or PC/cholesterol, is observed (**Fig. 4B**). Reconstitution in the absence of lipids (LPR = 0) causes aggregation of the majority of NS4B that ends up in the pellet and a minor fraction with a density of 1.29 g/cm³ (fraction 9,

corresponding to 60% sucrose), which is probably due to the recruitment of residual lipids from the wheat germ extract, which are not removed during the purification. Fig. 4C shows the electron micrographs of immunogold-labeled NS4B obtained for LPRs of 0.25, 0.1 and 0. Most of the protein is found to be concentrated in distinct electron dense structures, which likely represent proteoliposomes. Their membranous nature is best visible at LPR 0.25, but less well discernable at lower LPRs, i.e. for NS4B proteoliposomes containing fewer lipids. Based on these experiments, all LPRs shall yield samples containing proteoliposomes, which can be separated from aggregated NS4B located in the pellet. For NMR, the highest possible concentration of protein in the liposomes potentially yields the highest signal-to-noise ratio, if lines remain narrow. For the NMR sample we choose to work with an LPR of 0.25 in order to obtain a good separation from the aggregated pellet fraction, as only reconstituted protein shall be filled into the NMR rotor. Still, other LPRs might be promising to test as well, in particular to optimize the proton line width (*vide infra*).

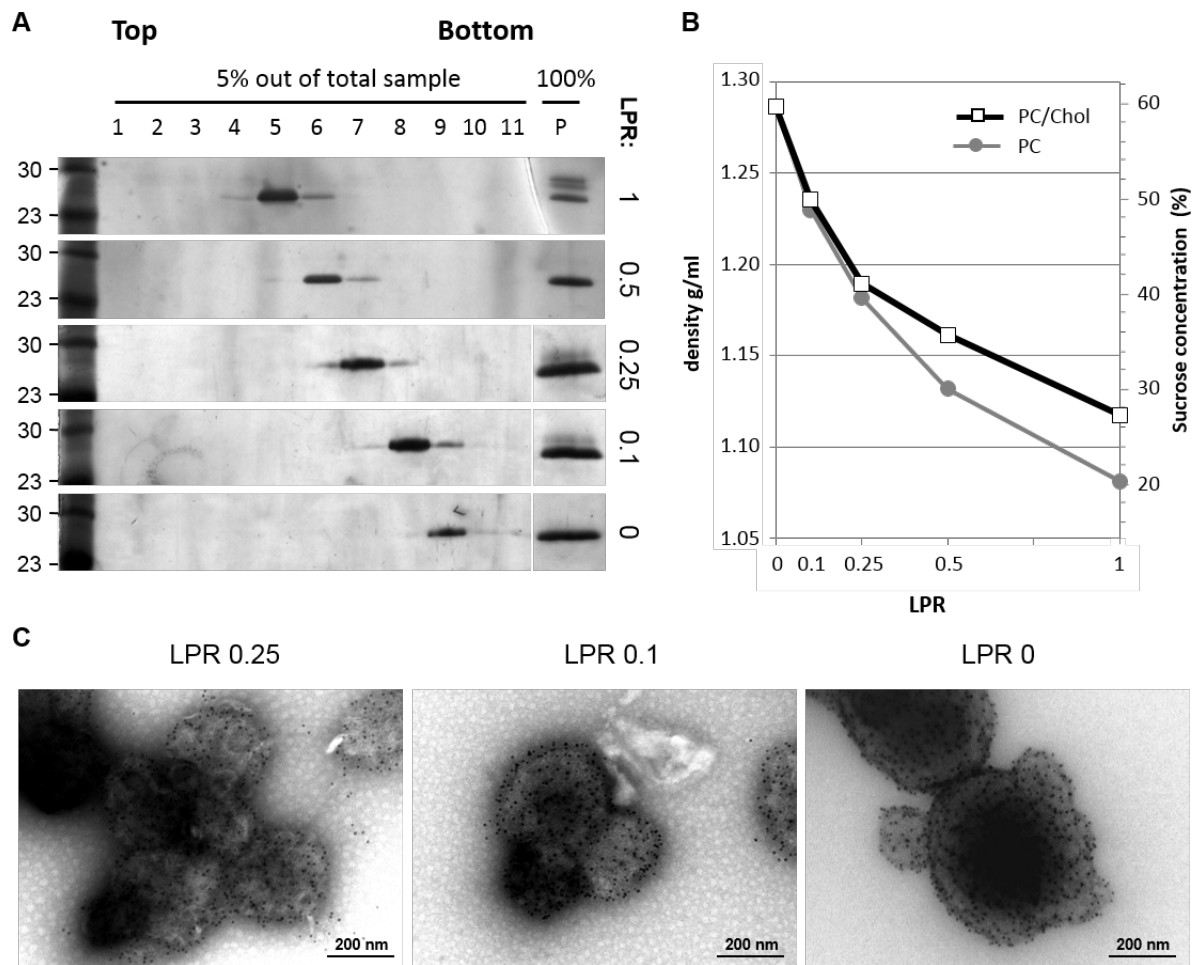


Figure 4: Discontinuous sucrose density gradient centrifugation of NS4B proteoliposomes. (A) NS4B reconstituted into PC/cholesterol liposomes at different LPRs was loaded onto a discontinuous sucrose gradient and separated by ultracentrifugation. Fractions were analyzed by SDS-PAGE and silver staining. “1” to “11” refers to the fraction number from the top to the bottom of the gradient; 5% of each fraction was loaded on the gel. “P” – whole pellet fraction loaded on the gel. (B) Densities (left y axis) and sucrose concentrations (right y axis) of protein positive fractions were calculated from refractive index values. (C) Electron-microscopy analysis of reconstituted samples in PC/cholesterol at LPR 0.25, 0.1 and 0. Immunogold-labeling against the C-terminal *Strep*-tag II indicates the localization of NS4B and samples were negatively stained using uranyl acetate.

Optimized NS4B samples yield high-quality ¹³C NMR spectra

For NMR, a uniformly [^2H , ^{13}C , ^{15}N]-labeled full-length NS4B sample was prepared using H_2O in the buffer medium. H^{N} protons and exchangeable side chains were there protonated to 100% while the other protons should nominally be fully deuterated. Reconstitution was performed with PC/cholesterol at an LPR of 0.25, and proteoliposomes were harvested by sedimentation onto a 90% OptiPrepTM cushion in order to avoid sucrose in the NMR sample. The proteoliposomes were pelleted in an 1.3 mm NMR rotor (Böckmann et al. 2009). A 2D DREAM spectrum was recorded in order to evaluate sample quality, which is shown in Fig. 5 (for full aliphatic regions see Figure S2). DREAM was chosen because PDS or DARR spectra show few crosspeaks in deuterated proteins. The DREAM spectrum is expected to show predominantly intraresidue ^{13}C - ^{13}C crosspeaks and allows estimating the ^{13}C line width, which is a proxy for the homogeneity of the sample. The spectrum shows signals typical for a mainly α -helical protein, with a high content of hydrophobic residues. Indeed, Ala, Ile, Leu and Val represent approximately 35 % of the residues in the primary sequence (Table S2). Several more isolated signals from non-helical residues are observed, and notably the Ser residues, for which 14 out of 29 residues are predicted to be located outside the helices, show a good spectral dispersion.

Figure 5B shows the corresponding spectrum for a sample where only Thr, Ile and Leu were ^{13}C and ^{15}N labeled. Only peaks corresponding to these amino acids are observed in this spectrum, indicating clean labeling without isotopic scrambling. (As a side remark, it might be worth mentioning that scrambling can be an issue when using other amino acids as D, E, and Q but can be reduced by the addition of inhibitors (Morita et al. 2004; Tonelli et al. 2011)). It can be seen in the Leu region that most signals are concentrated in the α -helical region, but four signals are clearly non-helical. These signals might correspond to the Leu residues which are predicted to be located outside α -helical regions (Fig. 1). In Fig. 5C, an extract of the Thr $\text{C}\beta$ - $\text{C}\gamma$ region is shown, and an isolated peak with a line width of 0.5 ppm

(100 Hz), as determined without apodization. While there are only a limited number of resolved peaks, a line width of 0.5-1 ppm appears representative for the entire spectrum. This line width is thought to be typical for solid-state NMR spectra of well-ordered membrane proteins in lipidic environment (Ward et al. 2015; Brown and Ladizhansky 2015).

Having used the established methods of ^{13}C spectroscopy to ascertain the proper reconstitution and fold of the protein, we also explored the resolution in proton-detected spectra, using an (H)NH CP-2D spectrum and a (H)CONH 3D spectrum (see Figure 6). Using the LIT sample, where we expect a total of 63 amide HN correlation peaks and 19 HNCO correlation peaks, the resolution was however not as good as observed by others in some highly deuterated membrane proteins. From the 3D spectrum, which was fitted by 21 peaks, we find a ^1H linewidth of 290 ± 95 Hz while values of 100 Hz and 130-180 Hz were obtained for the M2 channel and OmpG, respectively (Barbet-Massin et al. 2014). The linewidth in NS4B is partially refocussable: the transverse relaxation time of the amide peaks under a Hahn echo is measured as $T_2' = 2.8$ ms corresponding to a refocussed linewidth of 114 Hz. This value lies between the reference values for deuterated and protonated ubiquitin of 8.5 ms and 1.3 ms at the same spinning frequency. Reasons for the limited resolution may include (i) insufficient deuteration of the protein at the non-exchangeable sites, (ii) dynamical effects, and/or (iii) static structural heterogeneities to which the protons may be more sensitive than the ^{13}C . An obvious choice to improve homogeneity might be to increase the LPR to reduce protein-protein interactions which might induce linebroadening. While the resulting loss in signal would be problematic for the carbon-detected experiments, it would be considerably less so for the proton-detected spectra (total acquisition times of 2h for the spectrum shown in Figure 6A) and would at least partially be compensated by the gain in line-width.

To check the degree of deuteration, a (H)(C)H 1D spectrum was recorded with two CP-transfer steps of 1.5 ms which allows for efficient transfer not only for the directly ^{13}C bonded

protons but also for the amide protons (via the C α) and sidechain carbons via their protonated neighbors, if existing. The spectrum is shown in Figure 7 and indicates sizable protonation not only at the amide protons but also at the C α and methyl positions. This is further corroborated in Figure S3, where a 2D ^1H - ^{13}C HSQC spectrum measured along the bond via J-couplings is shown. Reprotonation, e.g. by transaminases, has been observed before in wheat-germ cell-free expression (Tonelli et al. 2011) and we are presently further investigating the reduction of the proton content which might indeed be the dominant reason for the observed line broadening (Böckmann et al. 2015) leading to refocussable as well as not Hahn-echo-refocussable contributions to T_2 relaxation. Another line to follow for obtaining narrower proton line widths might lie in the use of higher LPRs. Of course this approach will reduce the amount of protein in the sample and lead to lower signal integrals. In case that it is indeed possible to reduce linewidth, this will partially, or even fully, compensate this effect in terms of peak signal-to-noise ratios. The present acquisition time of the 2D HN spectrum is modest with 2h.

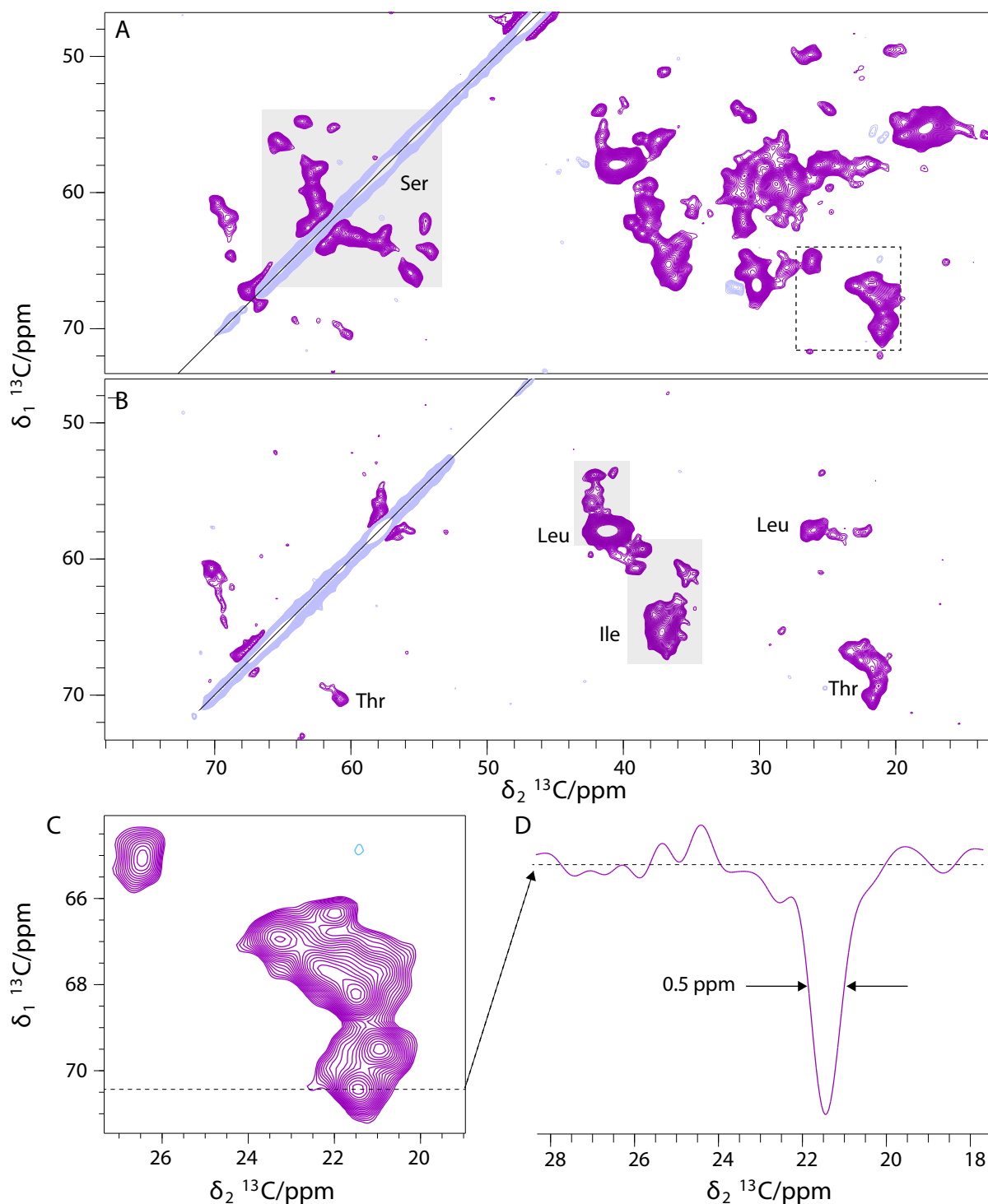


Figure 5: A 2D ^{13}C - ^{13}C DREAM spectrum recorded at 60 kHz MAS and $\sim 16^\circ\text{C}$ sample temperature for (A) 100%- H^{N} , uniformly [^2H - ^{13}C - ^{15}N]-labeled and (B) 100%- H^{N} , uniformly ^2H and LIT- $[^2\text{H}, ^{15}\text{N}, ^{13}\text{C}]$ selectively labeled NS4B in proteoliposomes (PC/cholesterol, LPR 0.25). (C) shows a magnified extract of the Thr $\text{C}\beta$ - $\text{C}\gamma$ region of (A). (D) shows the 1D trace of an isolated peak with a line width of 0.5 ppm which was determined without using an

apodization function while spectra A-C are processed with QSINE 2.5. The spectra of (A) and (B) displayed with their full aliphatic spectral width can be found in the Supporting Information, Figure S2.

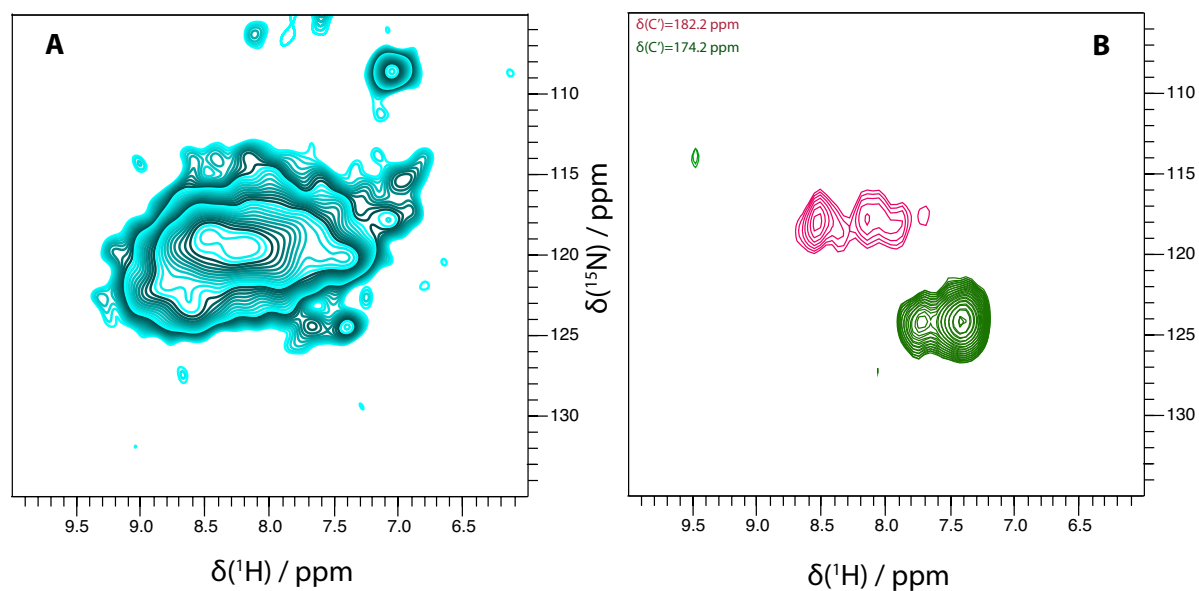


Figure 6: Proton-detected (H)NH 2D spectrum at 90 kHz MAS (A) and two slices (in red and green) out of the proton-detected (H)CONH 3D spectrum at 60 kHz MAS (B) both measured with the LIT-labeled NS4B sample. Details are given in Table S2. In (B), the ^{15}N and ^{13}C linewidths are sampling limited.

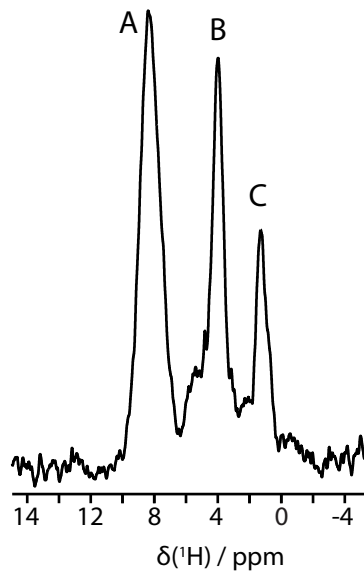


Figure 7: Proton spectrum of uniformly [^2H - ^{13}C - ^{15}N]-labeled NS4B after a ^1H - ^{13}C - ^1H using CP polarization transfer. The spectrum shows signal from ^1H in the protein. The signal marked A comes almost exclusively from amide-protons (expected to be ^1H), at B, many ^1H -C α resonate, while resonances below 3 ppm (C) are predominantly sidechains. The B and C signals would be missing with perfect deuteration. Even if the intensities are not quantitative, significant back-protonation is detected in particular for the H α .

Conclusions

Our results demonstrate that the “difficult” protein NS4B from HCV can be expressed in isotopically labeled form using a wheat germ cell-free expression system and can be properly reconstituted into proteoliposomes with high efficiency at low lipid-to-protein ratios such as LPR 0.25 (w/w). The expression yield was good (1 mg/1mL WGE) and amino acid selective isotopic labeling was demonstrated. The ILT labelling with ^{13}C was found to be very clean, while the degree of deuteration was lower than expected. The samples were shown to yield narrow resonance lines in the ^{13}C spectrum indicating a folded protein with mostly α -helical conformation, as predicted. The resolution of the ^{13}C NMR spectrum is estimated to be 0.5-1.0 ppm (100 Hz), slightly higher as for well-order microcrystalline proteins, but typical for

membrane proteins. These findings are a good proxy for the structural homogeneity and the proper folding state of the protein. The experimental approach described here opens the way for detailed NMR investigations using uniformly or selectively labeled protein reconstituted into proteoliposomes. Further work to understand the proton linewidth is needed.

Acknowledgements

This work was supported by the CNRS and by grants from the French ANRS (France Recherche, Nord & Sud, Sida-HIV et Hépatites), an autonomous agency at INSERM, France, and the ANR (ANR-14-CE09-0024B), the Swiss National Science Foundation (200020_146757, 200020_159707 and 31003A-156030), as well as the DFG TRR83-TP13.

References

- Agarwal V, Penzel S, Székely K, et al (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
- Andréll J, Tate CG (2013) Overexpression of membrane proteins in mammalian cells for structural studies. *Molecular Membrane Biology*. doi: 10.3109/09687688.2012.703703
- Barbet-Massin E, Pell AJ, Retel JS, 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
- Betton J-M (2004) High throughput cloning and expression strategies for protein production. *Biochimie* 86:601–605. doi: 10.1016/j.biochi.2004.07.004
- Böckmann A, Ernst M, Meier BH (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, 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
- Briggs ELA, Gomes R, Elhussein M, et al (2015) Interaction between the NS4B amphipathic helix, AH2, and charged lipid headgroups alters membrane morphology and AH2 oligomeric state — Implications for the Hepatitis C virus life cycle. *BBA - Biomembranes* 1848:1–8. doi: 10.1016/j.bbamem.2015.04.015
- Brown LS, Ladizhansky V (2015) Membrane proteins in their native habitat as seen by solid-state NMR spectroscopy. *Protein Science* 24:1333–1346. doi: 10.1002/pro.2700
- Egger D, Wölk B, Gosert R, et al (2002) Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *Journal of Virology* 76:5974–5984.

- Einav S, Elazar M, Danieli T, Glenn JS (2004) A nucleotide binding motif in hepatitis C virus (HCV) NS4B mediates HCV RNA replication. *Journal of Virology* 78:11288–11295. doi: 10.1128/JVI.78.20.11288-11295.2004
- Emami S, Fan Y, Munro R, et al (2013) Yeast-expressed human membrane protein aquaporin-1 yields excellent resolution of solid-state MAS NMR spectra. *J Biomol NMR* 55:147–155. doi: 10.1007/s10858-013-9710-5
- Endo Y, Sawasaki T (2006) Cell-free expression systems for eukaryotic protein production. *Curr Opin Biotech* 17:373–380. doi: 10.1016/j.copbio.2006.06.009
- Fogeron M-L, Badillo A, Jirasko V, et al (2015a) Wheat germ cell-free expression: Two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins. *Protein Express Purif* 105:39–46. doi: 10.1016/j.pep.2014.10.003
- Fogeron M-L, Paul D, Jirasko V, et al (2015b) Functional expression, purification, characterization, and membrane reconstitution of non-structural protein 2 from hepatitis C virus. *Protein Express Purif* 1–6. doi: 10.1016/j.pep.2015.08.027
- Fu R, Gill RL, Gill RL Jr., et al (2015) Spherical Nanoparticle Supported Lipid Bilayers for the Structural Study of Membrane Geometry-Sensitive Molecules. *J Am Chem Soc* 137:14031–14034. doi: 10.1021/jacs.5b08303
- Gosert R, Egger D, Lohmann V, et al (2003) Identification of the Hepatitis C Virus RNA Replication Complex in Huh-7 Cells Harboring Subgenomic Replicons. *Journal of Virology* 77:5487–5492. doi: 10.1128/JVI.77.9.5487-5492.2003
- Gouttenoire J, Castet V, Montserret R, et al (2009a) Identification of a Novel Determinant for Membrane Association in Hepatitis C Virus Nonstructural Protein 4B. *Journal of Virology* 83:6257–6268. doi: 10.1128/JVI.02663-08
- Gouttenoire J, Montserret R, Kennel A, et al (2009b) An Amphipathic α -Helix at the C Terminus of Hepatitis C Virus Nonstructural Protein 4B Mediates Membrane Association. *Journal of Virology* 83:11378–11384. doi: 10.1128/JVI.01122-09
- Gouttenoire J, Montserret R, Paul D, et al (2014) Aminoterminal Amphipathic α -Helix AH1 of Hepatitis C Virus Nonstructural Protein 4B Possesses a Dual Role in RNA Replication and Virus Production. *PLoS Pathog* 10:e1004501–17. doi: 10.1371/journal.ppat.1004501
- Gouttenoire J, Penin F, Moradpour D (2010) Hepatitis C virus nonstructural protein 4B: a journey into unexplored territory. *Rev Med Virol* 20:117–129. doi: 10.1002/rmv.640
- Hansen SK, Bertelsen K, Paaske B, et al (2015) Solid-state NMR methods for oriented membrane proteins. *Prog Nucl Mag Res Sp* 88-89:48–85. doi: 10.1016/j.pnmrs.2015.05.001
- Huber M, Hiller S, Schanda P, et al (2011) A Proton-Detected 4D Solid-State NMR Experiment for Protein Structure Determination. *Chemphyschem* 12:915–918. doi: 10.1002/cphc.201100062
- Kainosho M, Torizawa T, Iwashita Y, et al (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440:52–57. doi: 10.1038/nature04525
- Kaur H, Lakatos A, Spadaccini R, et al (2015) The ABC exporter MsbA probed by solid state NMR – challenges and opportunities. *Biological Chemistry*. doi: 10.1515/hsz-2015-0119
- Klammt C, Maslennikov I, Bayrhuber M, et al (2012) Facile backbone structure determination of

- human membrane proteins by NMR spectroscopy. *Nature Methods* 9:834–839. doi: 10.1038/nmeth.2033
- Kunert B, Gardienet C, Lacabanne D, et al (2014) Efficient and stable reconstitution of the ABC transporter BmrA for solid-state NMR studies. *Front mol biosci* 1:1–11. doi: 10.3389/fmolb.2014.00005
- le Maire M, Viel A, Moller JV (1989) Size exclusion chromatography and universal calibration of gel columns. *Analytical Biochemistry* 177:50–56. doi: 10.1016/0003-2697(89)90012-2
- Lundin M, Lindström H, Grönwall C, Persson MAA (2006) Dual topology of the processed hepatitis C virus protein NS4B is influenced by the NS5A protein. *J Gen Virol* 87:3263–3272. doi: 10.1099/vir.0.82211-0
- Mandal A, Hoop CL, DeLucia M, et al (2015) Structural Changes and Proapoptotic Peroxidase Activity of Cardiolipin-Bound Mitochondrial Cytochrome c. *Biophys J* 109:1873–1884. doi: 10.1016/j.bpj.2015.09.016
- McDermott A (2009) Structure and Dynamics of Membrane Proteins by Magic Angle Spinning Solid-State NMR. *Annu Rev Biophys* 38:385–403. doi: 10.1146/annurev.biophys.050708.133719
- Mehler M, Eckert CE, Busche A, et al (2015) Assembling a correctly folded and functional heptahelical membrane protein by protein trans-splicing. *J Biol Chem jbc.M115.681205–12*. doi: 10.1074/jbc.M115.681205
- Moradpour D, Penin F (2013) Hepatitis C virus proteins: from structure to function. *Curr Top Microbiol Immunol* 369:113–142. doi: 10.1007/978-3-642-27340-7_5
- Morita EH, Shimizu M, Ogasawara T, et al (2004) A novel way of amino acid-specific assignment in (1)H-(15)N HSQC spectra with a wheat germ cell-free protein synthesis system. *J Biomol NMR* 30:37–45. doi: 10.1023/B:JNMR.0000042956.65678.b8
- Noirot C, Habenstein B, Bousset L, et al (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
- Opitz C, Isogai S, Grzesiek S (2015) An economic approach to efficient isotope labeling in insect cells using homemade N-15-, C-13- and H-2-labeled yeast extracts. *J Biomol NMR* 62:373–385. doi: 10.1007/s10858-015-9954-3
- Palomares-Jerez F, Nemesio H, Villalaín J (2012) The membrane spanning domains of protein NS4B from hepatitis C virus. *Biochimica et Biophysica Acta* 1818:2958–2966. doi: 10.1016/j.bbamem.2012.07.022
- Paul D, Romero-Brey I, Gouttenoire J, et al (2011) NS4B Self-Interaction through Conserved C-Terminal Elements Is Required for the Establishment of Functional Hepatitis C Virus Replication Complexes. *Journal of Virology* 85:6963–6976. doi: 10.1128/JVI.00502-11
- 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* 63:1–22. doi: 10.1007/s10858-015-9975-y
- Rai R, Deval J (2011) New opportunities in anti-hepatitis C virus drug discovery: targeting NS4B. *Antiviral Res* 90:93–101. doi: 10.1016/j.antiviral.2011.01.009
- Renault M, Cukkemane A, Baldus M (2010) Solid-State NMR Spectroscopy on Complex

- Biomolecules. *Angew Chem Int Ed Engl* 49:8346–8357. doi: 10.1002/anie.201002823
- Rigaud J-L, Pitard B, Lévy D (1995) Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. *BBA - Bioenergetics* 1231:223–246.
- Sawasaki T, Ogasawara T, Morishita R, Endo Y (2002) A cell-free protein synthesis system for high-throughput proteomics. *P Natl Acad Sci USA* 99:14652–14657. doi: 10.1073/pnas.232580399
- Shahid SA, Nagaraj M, Chauhan N, et al (2015) Solid-state NMR Study of the YadA Membrane-Anchor Domain in the Bacterial Outer Membrane. *Angew Chem Int Ed Engl* 54:12602–12606. doi: 10.1002/anie.201505506
- Takai K, Sawasaki T, Endo Y (2010) Practical cell-free protein synthesis system using purified wheat embryos. *Nat Protoc* 5:227–238. doi: 10.1038/nprot.2009.207
- Tonelli M, Singarapu KK, Makino S-I, et al (2011) Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J Biomol NMR* 51:467–476. doi: 10.1007/s10858-011-9575-4
- van Oers MM, Pijlman GP, Vlak JM (2015) Thirty years of baculovirus–insect cell protein expression: from dark horse to mainstream technology. *J Gen Virol* 96:6–23. doi: 10.1099/vir.0.067108-0
- Verel R, Ernst M, Meier BH (2001) Adiabatic dipolar recoupling in solid-state NMR: The DREAM scheme. *J Magn Reson* 150:81–99.
- Vinarov DA, Lytle BL, Peterson FC, et al (2004) Cell-free protein production and labeling protocol for NMR-based structural proteomics. *Nature Methods* 1:149–153. doi: 10.1038/nmeth716
- 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
- Ward ME, Wang S, Munro R, et al (2015) In Situ Structural Studies of Anabaena Sensory Rhodopsin in the E. coli Membrane. *Biophys J* 108:1683–1696. doi: 10.1016/j.bpj.2015.02.018

Supporting material for:

**Cell-free expression, purification, and membrane reconstitution for NMR
studies of the nonstructural protein 4B from hepatitis C virus**

Marie-Laure Fogeron^{1‡}, Vlastimil Jirasko^{2,3‡}, Susanne Penzel^{2‡}, David Paul³, Roland Montserret¹, Denis Lacabanne¹, Aurélie Badillo^{1,5}, Jérôme Gouttenoire⁴, Darius Moradpour⁴, Ralf Bartenschlager³, François Penin¹, Beat H. Meier^{2*}, and Anja Böckmann^{1*}

¹*Institut de Biologie et Chimie des Protéines, Bases Moléculaires et Structurales des Systèmes Infectieux, Labex Ecofect, UMR 5086 CNRS, Université de Lyon
7 passage du Vercors, 69367 Lyon, France*

²*Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland*

³*Department of Infectious Diseases, Molecular Virology, Heidelberg University, Im Neuenheimer Feld
345, 69120 Heidelberg, Germany and German Centre for Infection Research (DZIF), partner site
Heidelberg, Heidelberg, Germany.*

⁴*Division of Gastroenterology and Hepatology, Centre Hospitalier Universitaire Vaudois, University of
Lausanne, 1011 Lausanne, Switzerland*

⁵*RD-Biotech, Recombinant Protein Unit, 3 rue Henri Baigue, 25000 Besançon, France*

* Corresponding authors: AB: a.boeckmann@ibcp.fr, BHM: beme@ethz.ch


```

      10      20      30      40      50      60
MASRAALIEE GQRIAEMLKS KIQGLLQQAS KQAQDIQPAM QASWPKVEQF WARHMWNFIS

      70      80      90     100     110     120
GIQYLAGLST LPGNPAVASM MAFSAALTSP LSTSTTILLN IMGGWLASQI APPAGATGFV

      130     140     150     160     170     180
VSGLVGAAVG SIGLGKVLVD ILAGYGAGIS GALVAFKIMS GEKPSMEDVI NLLPGILSPG

      190     200     210     220     230     240
ALVVGVICAA ILRRHVGPGE GAVQWMNRLI AFASRGNHVA PTHYVTESDA SQRVTQLLGS

      250     260     270     280     290
LTITSLLRRL HNWITEDCPI PCLVPRGSAW SHPQFEKGGG SGGGSGGSAW SHPQFEK

```

Number of amino acids: 297

Molecular weight: 31057.9

Theoretical pI: 9.10

Amino acid composition:

| | | |
|----------------|-----------|--------------|
| Ala (A) | 35 | 11.8% |
| Arg (R) | 11 | 3.7% |
| Asn (N) | 7 | 2.4% |
| Asp (D) | 5 | 1.7% |
| Cys (C) | 3 | 1.0% |
| Gln (Q) | 16 | 5.4% |
| Glu (E) | 11 | 3.7% |
| Gly (G) | 36 | 12.1% |
| His (H) | 7 | 2.4% |
| Ile (I) | 21 | 7.1% |
| Leu (L) | 30 | 10.1% |
| Lys (K) | 9 | 3.0% |
| Met (M) | 10 | 3.4% |
| Phe (F) | 8 | 2.7% |
| Pro (P) | 17 | 5.7% |
| Ser (S) | 29 | 9.8% |
| Thr (T) | 12 | 4.0% |
| Trp (W) | 8 | 2.7% |
| Tyr (Y) | 3 | 1.0% |
| Val (V) | 19 | 6.4% |

Figure S1: Amino-acid sequence of recombinant NS4B (strain JFH-1) protein. N-terminal methionine for translation initiation as well as C-terminal thrombin cleavage site and *Strep*-Tag are colored in grey.

Bold letters indicate the addition for the Twin-*Strep*-Tag which was used for the NMR sample preparation.

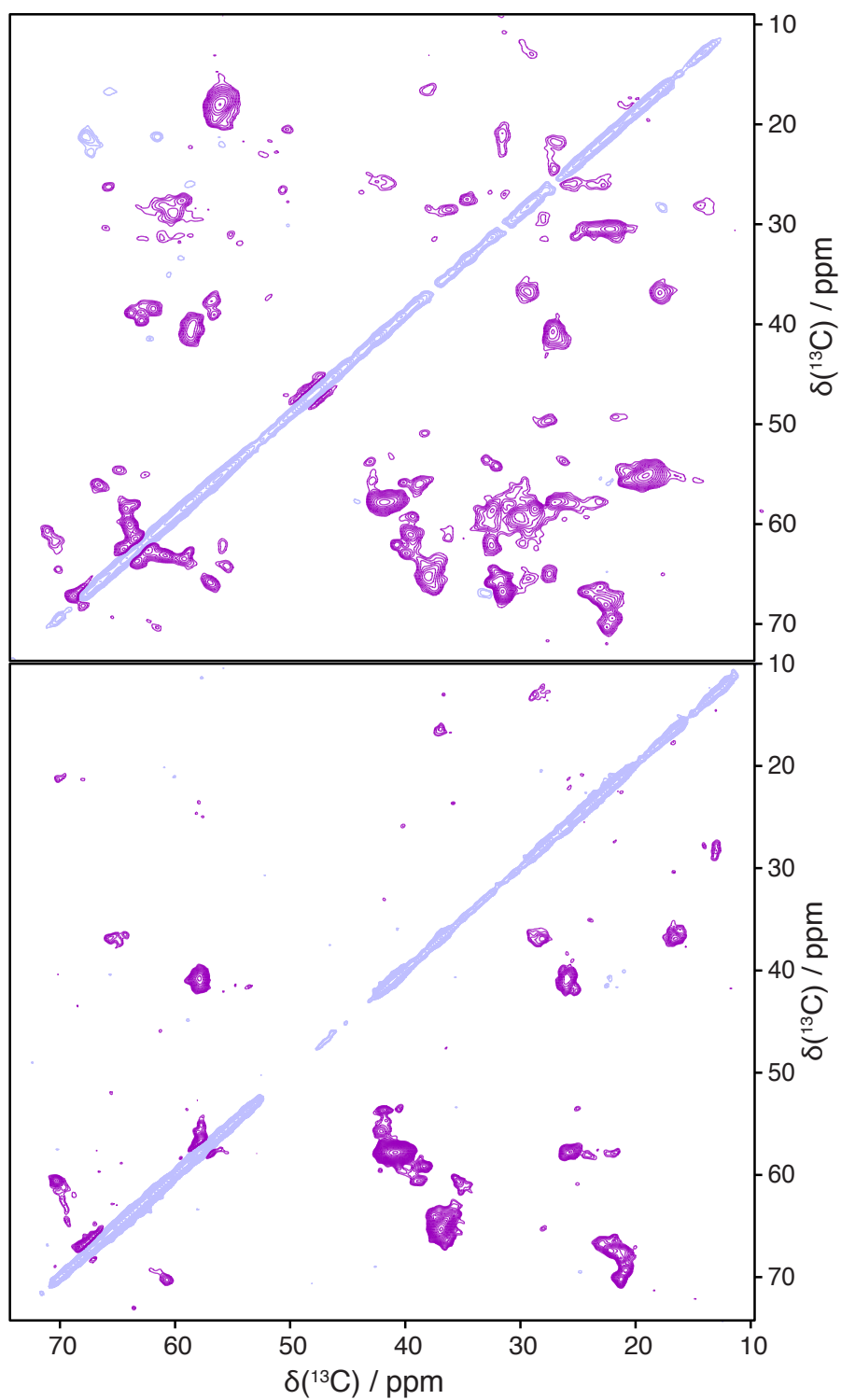


Figure S2: Full aliphatic region of the 2D ^{13}C - ^{13}C CA-CB DREAM spectrum at 60 kHz MAS and $\sim 16^\circ\text{C}$ sample temperature of the 100%-HN, uniformly [^2H , ^{13}C , ^{15}N]-labeled (top) and of the 100%-HN, uniformly [^2H] and LIT- ^{13}C , ^{15}N] selectively labeled (bottom) NS4B in proteoliposomes (PC/cholesterol, LPR 0.25) samples of which extracts are shown in Figure 5.

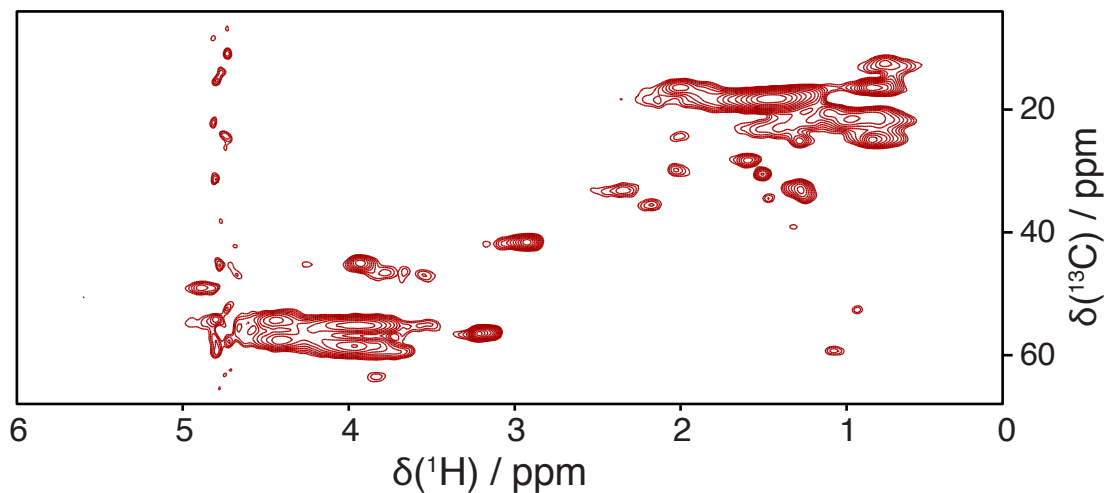


Figure S3: J-coupling based ^1H - ^{13}C HSQC spectrum of the 100%-HN, uniformly [^2H , ^{13}C , ^{15}N]-labeled NS4B acquired at 60 kHz MAS and a sample temperature of ~ 16 $^\circ\text{C}$. The J-coupling evolution delay of the refocused INEPT periods employed in this HSQC was optimized to highest transfer with 1.3 and 1.6 ms during ^1H and ^{13}C transverse magnetization, respectively. Pulses were set to 100 kHz RF-field strength on ^{13}C and ^1H , and the same parameters for MISSISSIPPI and decoupling were used as described in Table S2 as well as 5 kHz WALTZ on ^{13}C in addition during the acquisition. The spectrum was acquired for 3.5 and 2.6 ms in the direct and indirect dimension respectively with a spectral width of 160 ppm in the indirect dimension. The total number of scans was 1184.

| | CA-CB DREAM | CA-CB DREAM | (H)NCO 3D |
|---------------------|-------------|-------------|-----------|
| labeling scheme | fully | LIT | LIT |
| MAS frequency / kHz | 60 | 60 | 60 |
| Field / T | 20 | 20 | 20 |
| Interscan delay / s | 1.0 | 1.0 | 1.0 |
| Number of scans | 1024 | 876 | 400 |

| 1st transfer | H-CA CP | H-CA CP | H-CO CP |
|--|-------------------------|-------------------------|-------------------------|
| ¹ H field/ kHz | 42 | 75 | 80 |
| ¹³ C field / kHz | 18 | 15 | 20 |
| Shape | tangent ¹ H | tangent ¹ H | tangent ¹ H |
| ¹³ C carrier / ppm | 54 | 54 | 180 |
| Time / ms | 2.5 | 1.8 | 2.0 |
| 2nd transfer | CA-CB DREAM | CA-CB DREAM | CO-N CP |
| ¹ H field/ kHz | - | - | - |
| ¹³ C field / kHz | 30 | 30 | 20 |
| ¹⁵ N field / kHz | - | - | 40 |
| Shape | tangent ¹³ C | tangent ¹³ C | tangent ¹³ C |
| ¹³ C carrier / ppm | 28.5 | 28.5 | 180 |
| Time / ms | 4.0 | 4.0 | 6.25 |
| 3rd transfer | - | - | N-H CP |
| ¹ H field/ kHz | - | - | 49 |
| ¹³ C field / kHz | - | - | - |
| ¹⁵ N field / kHz | - | - | 11 |
| Shape | - | - | tangent ¹ H |
| Time / ms | - | - | 1.0 |
| direct dimension | | | |
| ¹³ C carrier / ppm | 40 | 40 | 180 |
| ¹ H decoupling | swf-TPPM | swf-TPPM | - |
| ¹ H decoupling field / kHz | 18 | 18 | - |
| ¹⁵ N decoupling | - | - | waltz64 |
| ¹⁵ N decoupling field / kHz | - | - | 5 |
| Number of increments | 2992 | 2992 | 2376 |
| Sweep width / ppm | 350.8 | 350.8 | 46.67 |
| Acquisition time / ms | 20 | 20 | 30 |
| indirect dimension | | | (¹³ C) |
| ¹³ C carrier / ppm | 40 | 40 | 180 |
| ¹ H decoupling | swf-TPPM | swf-TPPM | swf-TPPM |
| ¹ H decoupling field / kHz | 18 | 18 | 12 |
| Number of increments | 300 | 300 | 26 |
| Sweep width / ppm | 70.0 | 70.0 | 20 |
| Acquisition time / ms | 10 | 10 | 3.0 |
| indirect dimension | - | - | (¹⁵ N) |
| ¹⁵ N carrier / ppm | - | - | 117.5 |
| ¹ H decoupling | - | - | swf-TPPM |
| ¹ H decoupling field / kHz | - | - | 12 |
| Number of increments | - | - | 36 |
| Sweep width / ppm | - | - | 40 |
| Acquisition time / ms | - | - | 5.2 |
| Total acquisition time /h | 85 | 73 | 104 |

Table S1: Experimental parameters of the CA-CB DREAM with the fully labeled NS4B (first column), the CA-CB DREAM with the LIT-labeled NS4B (second column) and 3D (H)NCOH, also with the LIT-labeled NS4B (last column).

| | LIT-labeled |
|--------------------------|-------------|
| MAS frequency / kHz | 90 |
| Field / T | 20 |
| Sample temperature / °C | 18 |
| Interscan delay / s | 1.0 |
| Number of scans | 64 |
| Total acquisition time/h | 2 |

| Dimensions | ¹ H | ¹⁵ N |
|------------------------------|----------------|-----------------|
| Spectral widths (ppm) | 46.68 | 120 |
| Acq. times (ms) | 70 | 8 |
| Data points | 5550 | 116 |
| Phase-sensitive detection | DQD | States-TPPI |
| Carrier (ppm) | 4.8 | 115 |
| RF-fields, hard pulses (kHz) | 100 kHz | 50 kHz |

| Decoupling/Water suppression | ¹ H decoupling | ¹⁵ N decoupling | Solvent suppression |
|------------------------------|---------------------------|----------------------------|---------------------|
| Scheme | Frequency-swept TPPM | WALTZ64 | MISSISSIPPI |
| RF-field (kHz) | 10 | 5 | 15 |
| Pulse length (μs) | 48 | 102 | 60 000 |
| Placement (timing) | indirect dimensions | acquisition | - |

| Transfers | Specification | RF-field ¹ H (kHz) | RF-field ¹⁵ N (kHz) | Contact time (ms) | Shape |
|-----------|---------------|-------------------------------|--------------------------------|-------------------|------------------|
| H-N CP | DQ (n=1) | 78 | 15 | 1 | 40-60 tangential |
| N-H CP | DQ (n=1) | 78 | 15 | 1 | 60-40 tangential |

Table S2: Experimental parameters for the proton-detected (H)NH CP-based 2D spectrum at 90 kHz MAS. For the transfer steps and decoupling, as well as water suppression the same parameters as in (ref: Penzel et al. 2015 of the main text) were used. The experiment was performed on a 100 kHz triple-channel fast MAS probe built in Tallinn, Estonia using 0.8 mm rotors.

| | AA concentration of uniformly [2H.13C.15N]-labeled protein (mM) | AA concentration of selectively -labeled protein (mM) |
|-------------------------------------|---|--|
| Alanine | 0.36 | 0.71 |
| Arginine | 0.36 | 0.22 |
| Asparagine | 0.30 | 0.14 |
| Aspartic acid | 0.48 | 0.10 |
| Cysteine | 0.18 | 0.06 |
| Glutamic acid | 0.54 | 0.22 |
| Glutamine | 0.30 | 0.32 |
| Glycine | 0.30 | 0.73 |
| Histidine | 0.06 | 0.14 |
| Isoleucine | 0.18 | 0.42 |
| Leucine | 0.54 | 0.61 |
| Lysine | 0.72 | 0.18 |
| Methionine | 0.06 | 0.20 |
| Penylalanine | 0.24 | 0.16 |
| Proline | 0.30 | 0.34 |
| Serine | 0.24 | 0.59 |
| Threonine | 0.24 | 0.24 |
| Tryptophane | 0.18 | 0.16 |
| Tyrosine | 0.18 | 0.06 |
| Valine | 0.24 | 0.38 |
| | | |
| Average AA concentration | 0.30 | 0.30 |
| | | |
| Total concentration of all AA | 6.00 | 6.00 |

Table S3: Concentration of amino acids used for expression of uniformly and selectively labeled samples (uniformly labeled sample: [http://www.isotope.com/userfiles/files/assetLibrary/cell-free%20protein%20expression%20\(1\).pdf](http://www.isotope.com/userfiles/files/assetLibrary/cell-free%20protein%20expression%20(1).pdf)). For the AA concentration of the selectively labeled protein the amino acid composition as given in Figure S1 was chosen.

| | Used amount | Price per unit | Price of used material in € |
|---|-------------|-----------------------------------|-----------------------------|
| Transcription reaction: | | | |
| Transcription buffer 5X (CellFree Sciences) | 1.2 mL | 6 €/mL | 7 |
| NTP mix 25 mM (Promega) | 0.6 mL | 200 €/1.6 mL | 75 |
| RNAsin 80U/μL (CellFree Sciences) | 75 uL | 750 €/mL | 56 |
| SP6 RNA Polymerase 80U/μL (CellFree Sciences) | 75 uL | 750 €/mL | 56 |
| | | total: | 194 |
| Translation reaction: | | | |
| SUB-AMIX NA 10X (CellFree Sciences) | 14.4 mL | 280 €/50 mL | 80 |
| [2H,13C,15N]-labeled amino acids (Cambridge Isotope Laboratories) | 125 mg | 2000 €/0,5 g | 500 |
| MNG-3 5% (Anatrace) | 2.8 mL | 275 €/5 g | 8 |
| Creatine kinase 4 mg/mL (Roche) | 0.12 mL | 88 €/100 mg | 1 |
| WGE (home made) | 6 mL | - | - |
| | | total: | 589 |
| Purification: | | | |
| DDM 10% (Carl-Roth) | 4.4 mL | 440 €/10g | 19 |
| StrepTactin high affinity 5 mL column (IBA) (can be recycled several times) | 1 piece | 289 € | 289 |
| | | total: | 308 |
| | | Total price of NMR sample: | 1091 |

Table S4: Price of material used for preparation of uniformly [²H, ¹³C, ¹⁵N]-labeled NMR sample for 1.3 mm rotor (6 ml wheat germ extract). Commercial wheat-germ extract would be 3900 € (650€/mL). The cost of home-made wheat-germ extract is difficult to estimate as salaries should likely be included to make a fair comparison. Wheat seeds are about 20 €/25 kg. The preparation yields about 6-9 mg of protein (1-1.5 mg/mL WGE)