Multi-scale studies of Measeles virus nucleocapsid assembly

Serafima Guseva

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

Serafima GUSEVA

Thèse dirigée par **Rob RUIGROK**,Professeur
et codirigée par **Martin BLACKLEDGE**, CEA

préparée au sein du **Laboratoire Institut de Biologie Structurale**
dans **l'École Doctorale Chimie et Sciences du Vivant**

Etudes multi-échelles de l'assemblage de la nucléocapside du virus de la rougeole

Multi-scale studies of Measeles virus nucleocapsid assembly

Thèse soutenue publiquement le **18 mai 2021**,
devant le jury composé de :

Monsieur ROB RUIGROK
PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES, Directeur de thèse
Monsieur MARTIN BLACKLEDGE
CADRE SCIENTIFIQUE DES EPIC, CEA CENTRE DE GRENOBLE, Co-directeur de thèse
Madame TANJA MITTAG
DOCTEUR EN SCIENCES, ST. JUDE CHILDREN'S RESEARCH HOSPITAL, Examinatrice
Monsieur MARKUS ZWECKSTETTER
PROFESSEUR, UNIVERSITE DE GÖTTINGEN, Rapporteur
Monsieur WINFRIED WEISSENHORN
PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES, Président
Madame MARIE-ANNE RAMEIX-WELTI
PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER, UNIVERSITE DE VERSAILLES - SAINT QUENTIN, Rapporteure

Remerciements

Foremost, I would like to thank my thesis directors, Martin Blackledge and Rob Ruigrok, for the opportunity to work in their laboratories. I cannot imagine another place where I would have been able to grow more within 3 years. Thank you for all the support and the guidelines on the path to the thesis. You have always been keen on listening to my interests and providing me with freedom of choice. And most importantly, you transferred me the enthusiasm to make science, motivating me to go further to the core.

I would like to thank Sigrid Milles, who was carrying the measles project before. Thank you for all the knowledge, encouragement and patience, for teaching me multiple techniques and for always being available for any problem. I really appreciated working with you.

I am very glad to work in such a fantastic group. Therefore, I would like to thank Luiza Mamigonian Bessa and Aldo Camacho Zarcho for all the support on a daily basis in the lab and in the office. It was a pleasure to work as a team on one of the projects. I cannot imagine better colleagues.

I would like to thank Nicola Salvi and Malene Jensen for all the help with NMR spectroscopy and all the advice. Thank you for being ready to help in any emergency situation. I also thank Damien Maurin for all the help in the lab and useful tricks.

I would like to thank the following people for helping with this research project :
Emmanuelle Neumann, for the help with electron microscopy
Caroline Mass, for providing guidance on the biophysical platform
Jean-Philippe Kleman, for introducing me to the fluorescent microscopy techniques
Alicia Vallet, for all the help with NMR spectrometers.

And my biggest thanks to my family and friends, for all the support you have shown me through this research.

Finally, I would like to dedicate my thesis to my nanny Tanja, who was telling me stories about viruses during my childhood.

Introduction générale

Introduction générale.

La séparation de phases est aujourd'hui considérée comme l'un des sujets les plus brûlants de la biologie, susceptible de transformer la biologie et la biomédecine. Malgré de multiples exemples et des preuves indiscutables de leur existence dans les cellules, la communauté scientifique discute encore pour savoir s'il s'agit d'une "science à la va-vite ou d'une idée révolutionnaire". La séparation de phases est responsable de la formation d'organites sans membranes qui sont impliqués dans l'arrangement du contenu cellulaire. Le cytoplasme et le noyau sont remplis de telles structures. Et même si plusieurs processus se déroulent dans les cytoplasmes, ils peuvent être séparés spatialement par des organites sans membranes. La découverte des biocondensats dans les cellules peut enfin répondre à la question fondamentale de savoir comment les cellules organisent les molécules nécessaires au bon endroit et au bon moment afin de réaliser une fonction particulière.

Les bases du phénomène de séparation de phases ne sont pas encore totalement comprises, en particulier pour les systèmes biologiques. Cependant, l'importance des protéines intrinsèquement désordonnées (IDP) est soulignée dans de nombreuses études. Le rôle ces protéines intrinsèquement désordonnées dans la biologie est sous-estimé en général, et les programmes universitaires ignorent le plus souvent leur existence.

Pour comprendre comment les organites sans membranes contrôlent les fonctions cellulaires, il est important de définir les propriétés du système. Une série de techniques biophysiques peuvent être utilisées pour caractériser les propriétés macroscopiques des gouttelettes. Cependant, la RMN est probablement la seule technique qui permette d'étudier la séparation de phases au niveau moléculaire et d'obtenir des informations de résolution atomique sur les interactions impliquées dans leur formation, les changements dynamiques et structurels associés à la transition de phase.

Un des exemples de séparation de phases en biologie est la formation de biocondensats de protéines virales. Il apparaît que les composants viraux ne sont pas dissous dans le cytoplasme mais concentrés dans ce que l'on appelle des usines virales. Les avantages d'une telle organisation sont multiples : augmentation de la vitesse des réactions chimiques grâce à l'augmentation de la concentration, contrôle des conditions environnementales, protection contre la reconnaissance du système immunitaire de la cellule hôte. Dans cette thèse, les phénomènes de séparation de phase sont étudiés à travers la machinerie de réplication du virus de la rougeole. Cette machinerie nécessite la polymérase et deux protéines accessoires : La Nucléoprotéine et la Phosphoprotéine. La nucléoprotéine se lie à l'ARN viral et forme la capside, la phosphoprotéine accompagne cette réaction ainsi que l'activité de la polymérase, la formation des usines virales et probablement beaucoup d'autres qui ne sont pas encore identifiées. La phosphoprotéine possède de multiples sites de liaison avec la nucléoprotéine qui sont actives à différentes étapes du cycle viral. La manière dont la phosphoprotéine régule ses interactions avec la nucléoprotéine et contrôle ainsi la réplication virale reste une question ouverte. 'étude présente est importante pour comprendre le mécanisme de réplication de la rougeole dans les cellules ainsi que les virus apparentés. Elle permet également d'avancer dans la description des principes de base de la séparation de phase liquide-liquide des protéines, reliant ainsi la physique à la biologie cellulaire.

Cette thèse est consacrée à plusieurs sujets :

Tout d'abord, nous avons cherché à tester la capacité des nucléoprotéines et des phosphoprotéines du virus de la rougeole à se séparer en phase et à déterminer le mécanisme de formation des gouttelettes pour les protéines virales. De même, déterminer la fonction des biocondensats viraux - ceci est présenté dans le chapitre 3.

Ailleurs, pour étudier la phosphorylation de la phosphoprotéine de la rougeole *in vitro* et comprendre le rôle fonctionnel de la phosphorylation - c'est l'objet du chapitre 4.

Enfin, afin de mieux comprendre le processus de séparation des phases des protéines, nous avons comparé la structure et la dynamique des protéines dans les phases diluées et condensées - ce qui est présenté au chapitre 5.

General introduction

General introduction.

Phase separation is now considered to be one of the hottest topics in biology which might transform our understanding of biology and biomedicine. Despite multiple examples and undoubtable proofs of their existence in cells, the scientific community still discusses whether it is a "Sloppy science or ground-breaking idea". Phase separation is responsible for the formation of membraneless organelles which are involved in the arrangement of cell content. The cytoplasm and the nucleus are shown to be full of such structures. And even if several process take place in cytoplasms, they can be spatially separated with membraneless organelles. The discovery of biocondensates in the cell can finally answer the fundamental question of how cells organise the necessary molecules at the right place and time to carry out a particular function.

The basics of the phase separation phenomena are still not fully understood especially for biological systems. However, the importance of intrinsically disordered proteins (IDPs) is highlighted in many studies. IDPs role in biology is underestimated in general, while school and university programs mostly ignore their existence.

To understand how membraneless organelles control the cellular functions it is important to define the properties of the system. A range of biophysical techniques can be used to characterise macroscopic features of the droplets. However, NMR is probably the only technique which allows to study phase separation at the molecular level and obtain atomic resolution information about interactions involved in their formation, in terms of dynamic and structural changes associated with the phase transition.

One of the examples of phase separation in biology is the formation of viral proteins biocondensates. It appears that viral components are not dissolved in the cytoplasm but concentrated in so-called viral factories. Multiple advantages for such organisation can be found : Increase in the rate of chemical reactions due to the increase of concentration, control of the environmental conditions, protection from the host cell immune system recognition. In this thesis, phase separation phenomena is studied on the basis of measles virus replication machinery. The machinery requires the polymerase and two accessory proteins : Nucleoprotein and Phosphoprotein. The Nucleoprotein binds to the viral RNA and forms nucleocapsids, the Phosphoprotein controls this reaction as well as polymerase activity, formation of viral factories and probably many others which are not identified yet. The Phosphoprotein has multiple binding sites with the Nucleoprotein which are present at different stages of viral cycle. How the Phosphoprotein regulates its interactions with the Nucleoprotein and thus controls viral replication is an open question.

The present study is important for understanding the mechanism of measles replication in cell as well as related viruses. And also provides the advances in the description of the basics of liquid-liquid phase separation of proteins thereby linking physics with cellular biology.

This thesis is dedicated to several subjects :

First, we aimed to test the ability of measles virus Nucleoprotein and Phosphoprotein to phase separate and to determine the mechanism of droplet formation. Also, to determine the function of viral biocondensates - this is presented in Chapter 3.

In addition, I aimed to study phosphorylation of measles Phosphoprotein *in vitro* and unders-

tand the functional role of phosphorylation - this is the subject of Chapter 4.

Finally, in order to better understand the process of protein phase separation, we compared the protein structure and dynamics in dilute and condensed phases what as shown in Chapter 5.

Première partie

Measles Virus Replication Machinery : Viral factories and Assembly Regulation via Phosphorylation.

1 Introduction

An introduction on liquid-liquid phase separation (LLPS) is discussed here: The phenomenon has been well described by polymer physics for many decades, however, its role and importance in biology was revealed only recently. Different classes of phase separating mixtures are reviewed with the biological examples for each of them. In the second part, I discuss measles virus replication machinery: The proteins involved in this process, three dimensional structure for the folded parts and dynamics description for disordered regions, their interactions with each other, and viral RNA. Next, I present a review on studies of viral replication machinery forming membraneless organelles via LLPS in cellulo. Also, cellular proteins which were found to be present in viral factories are summarised and possible roles of viral droplets are introduced.
The final part of this chapter is dedicated to the description of intrinsically disordered proteins. LLPS is often maintained by proteins containing intrinsically disordered regions, which represent a separate class of proteins. The characteristics of such proteins predetermine their ability to phase separate.

1.1 Phase separation as a zeitgeist

Intercellular compartmentalization is essential for maintaining various biochemical processes as each of them requires particular components and specific conditions: pH, ion strength, etc. Membrane organelles are known for centuries, for instance, the nucleus was first mentioned by Antonie van Leeuwenhoek in 1719. Due to the destructive effects of the commonly used techniques to study cells, fragile membranless organelles remained concealed from scientists. Recently more invasive methods to study living cells in details started to become available, resulting in a wave of discoveries of membranless organelles. A range of such subcompartments have been known for more than 20 years: Nuclear membranless organelles [121, 161, 190], and cytoplasmic P-bodies [10] and stress granules [110]. Nowadays, many other membraneless organelles are discovered and characterized, the information about them is summarized in the phase separation database (PhaSepDB http://db.phasep.pro).

Now it is well established that such biomolecular condensates originate via liquid-liquid phase separation (LLPS), a spontaneous demixing usually observed with polymers. In cells, it is a reversible process that brings together proteins and nucleic acids in non-stoichiometrical dynamic assembles.

Membraneless organelles in cells are multicomponent structures and different proteins, nucleic acids and small molecules were found inside them. In a simplistic description of system, these components can be divided into those forming the scaffold (those that form the network and, thus, essential for the phase separation) and molecules that can be considered as clients (those attracted inside droplets via interaction with the scaffold). The second type may participate uniquely in the functional role of the condensate, essential for the cell and may change the properties of the droplet, however, it may not be necessary for the initial formation so, in the following parts we will focus on the scaffold part.

During the next sections the physics underlying the phenomenon of phase separation is discussed, together with simple models for its description. After this, some functional roles of phase separation in cells are presented.

1.1.1 The description of LLPS phenomena

Liquid-liquid phase separation (LLPS) is a well-known phase transition that can occur in multicomponent mixtures and depends on the interactions between elements in the system. It is a reversible energetically favorable (what means spontaneous) process of demixing and, results in division into two phases (Fig. 1.1), which differ by component concentrations: One is polymer rich, the other is solvent rich [174].

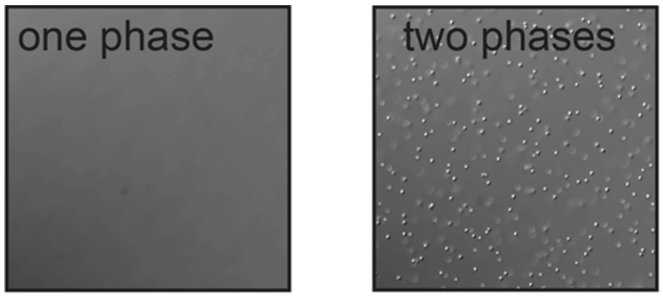


Figure 1.1 – Differential interference contrast (DIC) microscopy image of phase separation of biopolymers [133].

The global free energy of the system contains entropic and enthalpic contributions.

$$\Delta G_{free} = \Delta H - T\Delta S$$

Spontaneous demixing is associated with a decrease of the free energy that can be achieved by either decreasing enthalpy (enthalpic driven phase separation) or entropy (entropic driven phase separation). So, phase separation can be divided into two types by the driving force.

1.1.2 Enthalpy driven phase separation

In the ideal mixture, the interaction energy is not changed, so only an entropic contribution exists and the mixture is always homogeneous. For a binary system, for example, polymer-solvent, the energy of mixing per site is:

$$\Delta F_{mix} = kT(\frac{\phi}{N} ln\phi + (1-\phi) ln(1-\phi))$$

where k is the Boltzmann constant, T is temperature, ϕ is volume fraction and N is the polymer length relative to the solvent. As ϕ is always <1, $ln(\phi)$ is always negative, ΔF_{mix} has only negative contributions, favouring demixing. In non-ideal mixtures interactions between components exist and they can be either attractive or repulsive. In order to make the phase separation energetically favorable enthalpy contribution has to eliminate the contribution of the entropy to minimise the total energy [180].

The classical model to describe phase separation thermodynamically was proposed independently in 1941-1942 by Flory and Huggins [62, 96]. It is a mean-field lattice-base description that assumes that molecules in the solution are linear, constant in size and are placed on the nodes of the lattice, each node is occupied by one of the components, in the simplest case A (solution) and B (polymer) as presented on Fig. 1.2. Polymer B has a length N_B and occupies volume V_B, then for solution $N_A = 1$, and volume V_A. Volume fractions for each of the components are $\phi_A = \frac{V_A}{V_A+V_B}$ and $\phi_B = \frac{V_B}{V_A+V_B}$.

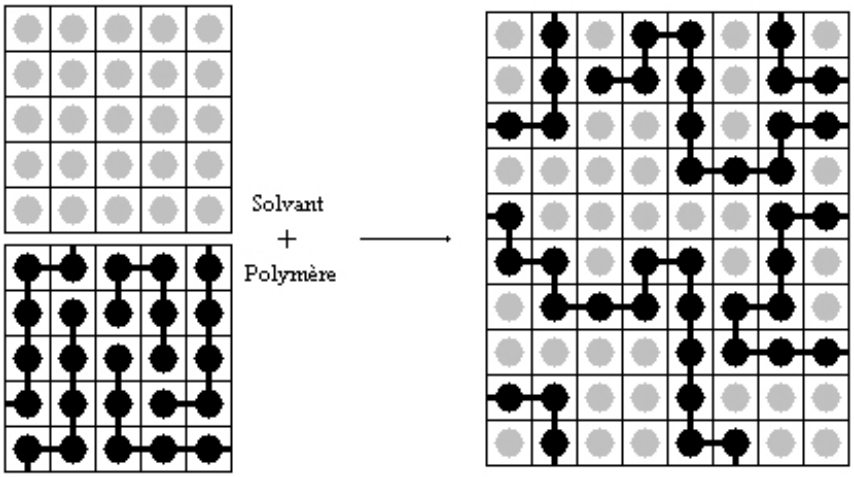


Figure 1.2 – Schematic illustration of the on-lattice model of mixing process. The mixture of two components: polymers (black) and solvent (grey) is shown.

The main invention of the Flory-Huggins model is Flory interacting parameter:

$$\chi = \frac{z}{2} \frac{2u_{AB} + u_{AA} + u_{BB}}{kT}$$

where u_{XX} are interacting energies between A and B components, z is the number of neighbors. This parameter can be used to calculate the additional contribution to the energy of mixing which is $\chi\phi(1-\phi)kT$. It represents the average energy associated with polymer-solvent interactions. A detailed analysis of the Flory-Haggins model, the application to experimental data and determination of the Flory parameter is presented in Chapter 2.

One can notice that this model is dependant on the volume fraction of the components in a non-trivial way. Actually, the energy of mixing along ϕ value varies for phase separating system and demixing is spontaneous only in a particular range of ϕ. Plotting it together with the temperature dependency summarises the mixture behaviors into a phase diagram, an example is presented on Fig. 1.3

Based on a shape of a phase diagram two types of phase separating systems can be distinguished: With an upper critical solution temperature (UCST) and a lower critical solution temperature (LCST). In Flory-Huggins theory they are distinguished by the sign of the Flory parameter: UCST has a positive one, LCST has a negative one (Fig. 1.3).

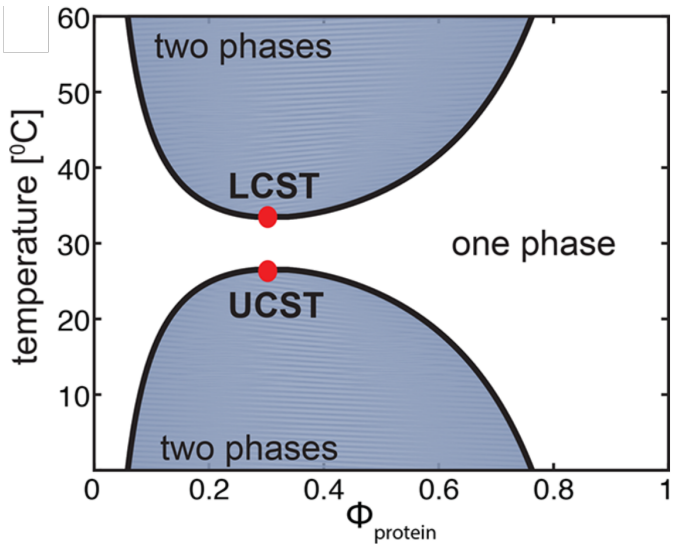


Figure 1.3 – Phase diagran with an upper or lower critical solution temperature (UCST or LCST, respectively) [133].

Enthalpic phase separation in biological systems

In biological systems phase separation is triggered by heteropolymers: Proteins and nucleic acids (RNA and DNA). Proteins which are involved in the formation of droplets share some common features. First, such proteins usually contain disordered regions, if not fully disordered, second, they represent multivalency.

Going through the known examples of protein systems which phase separate one can notice that either the determined multiple binding sites between the components exist (heterotypic domain motif systems) or it is organised out of IDPs which phase separate via homotypic interactions with each other.

Heterotypic systems

Such systems are usually composed of two or more scaffold molecules. The interaction between protein/RNA/DNA chains is essential for the phase transition and in order to form a network at least three are required. Participants in biomolecular condensates can be classified into separate classes by the number of binding sites (Fig. 1.4). Those which have a valency higher or equal to three are classified as "nodes" as they actually form a network. Those which have valency equal to two are classified as "bridges" as they can connect the higher valency particles or groups but do not increase the valency of the system, one binding site particles are classified as "caps" as they terminate the binding site and reduce interacting capacity. And finally, those which do not have interacting sites are classified as "bystanders" [185].

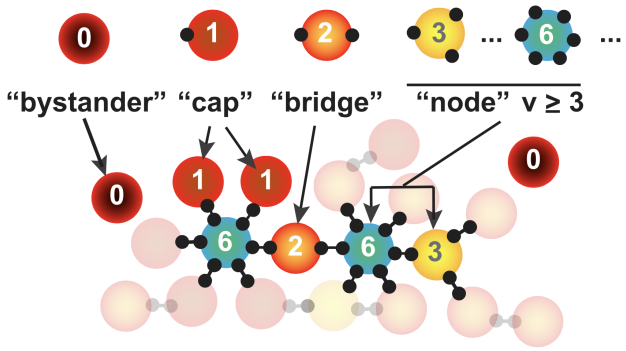


Figure 1.4 – Network-based classification of the droplet components. Valence (v) of the biocondensate member: v = 0 (bystander), v = 1 (cap); v = 2 (bridge), v>2 (node) [185].

Heterotypic interactions are represented with a specific interaction between molecules and often involve folded domains, one of the example is RNA-binding domain (RRM) or protein binding domain SH3 [87], but disordered motifs which perform specific binding also exist in such systems (RGG motif) [36]. In order to maintain multivalency several binding sites must be present on the molecule. This can be either achieved by sequential or a parallel connections of binding sites. An example of two distinct nucleolar subcompartments with parallel connections [58] and cytoplasmic biocondensates formed by sequentially connected interacting domains (Fig. 1.5) [87].

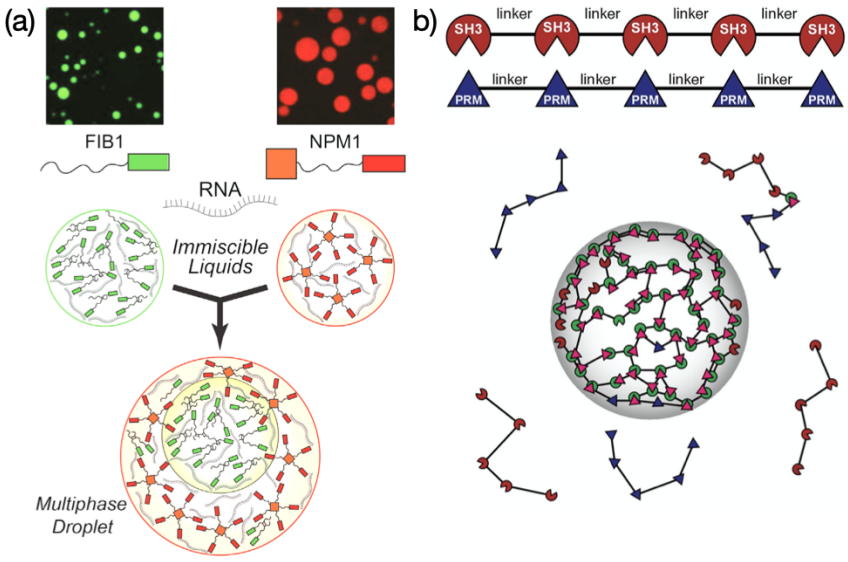


Figure 1.5 – (a) An example of parallel binding site connection [58]; (b) sequential binding site connection [87].

Binding sites comprising molecule multivalency can be either identical or different in terms of their affinity or even interact with different partners. All this affects phase separation and defines the phase diagram [218].

In order to to better understand phase separation, experimental data can be liked to the physical model. The concept of a "sticker-spacer" representation, identifying interacting and non-interacting components of the system, of phase separation proteins is becoming popular. A software LASSI (LAttice simulation engine for Sticker and Spacer Interactions) was specially designed for this concept and allows to calculate phase diagrams for multicomponent systems [35]. LASSI allows to vary architecture of the molecule, the affinity of binding sites, linkers length

and also its solvation volume what can modulate phase separation as well and allows to describe experimental data [87].

Homotypic systems

Single IDPs usually act as a scaffold for homotypic systems, the most studied examples are FUS [208] and Ddx4 [21]. The amino acid content is usually rich with charged, aromatic amino acids and glycine [133]. Four groups of interactions can be distinguished: electrostatic interactions which include charged and dipolar interactions, π-π and cation-π interactions (Fig. 1.6) [159]. Of course, hydrophobic contacts can be also involved in the formation of the network, in particular phase separating proteins are rich with hydrophobic amino acids, and their properties will be discussed further.

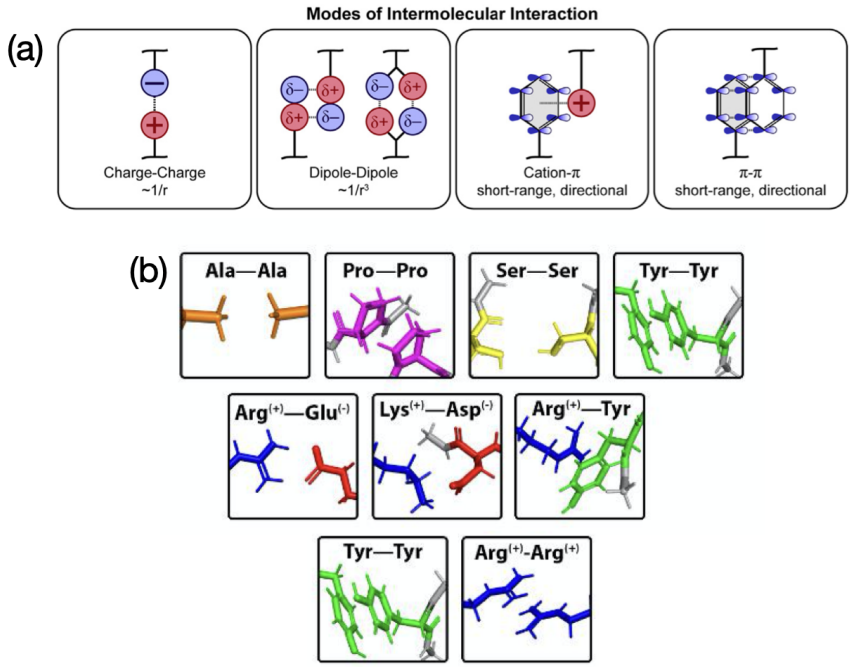


Figure 1.6 – (a) Types of molecular interactions underlying intracellular phase transitions [159]; (b) possible amino acid contacts which stabilise phase separation [116].

Different contacts between amino acids can be formed: Via oppositely charged residues (K/R and E/D) ; P,Y,H and T are involved in π-π and cation -π interactions, and polar contacts with (A-A, P-P, S–S, and Y–Y) (Fig. 1.6). Different pairs of amino acids make different strength contacts, for instance, replacing tyrosine with phenylalanine and arginine to lysine in FUS protein increases the saturation concentration, the critical concentration after which the phase separation occurs. A similar tendency was observed by decreasing the number of tyrosine and arginine residues [208].

Such systems can also be represented with a "sticker-sticker" model where interacting amino acids are stickers with a small radius and non-interacting amino acids such as glycin represent spacers [116]. The "Sticker-spacer" model does not take into account interaction with solvent, this could be overcome by modeling solvent as another interacting molecule with a low valency. However, it is impossible to introduce the contribution of entropy in such a model [19].

1.1.3 Entropy driven phase separation

Even though macromolecule content in droplets is very high (protein concentration is estimated to be around ~400 mg/ml) water concentration is still more than 60% which corresponds to at

least 33 M [21]. The entropic contribution to the system is mostly coming from the solvent and underlining the importance of solvent in phase separation.

Water around macromolecules can be subdivided into groups by layers (Fig. 1.7): The layer closer to the solute (hydration H_2O), the second hydration shells (extended hydration H_2O) and the rest further away (bulk H_2O) which, following the most widely accepted model, binds four other molecules in a tetrahedral structure via H bonds in the liquid state making a 3D network, however, the average lifetime of the network is in the range 1–8 ps, Fig. 1.7 [174].

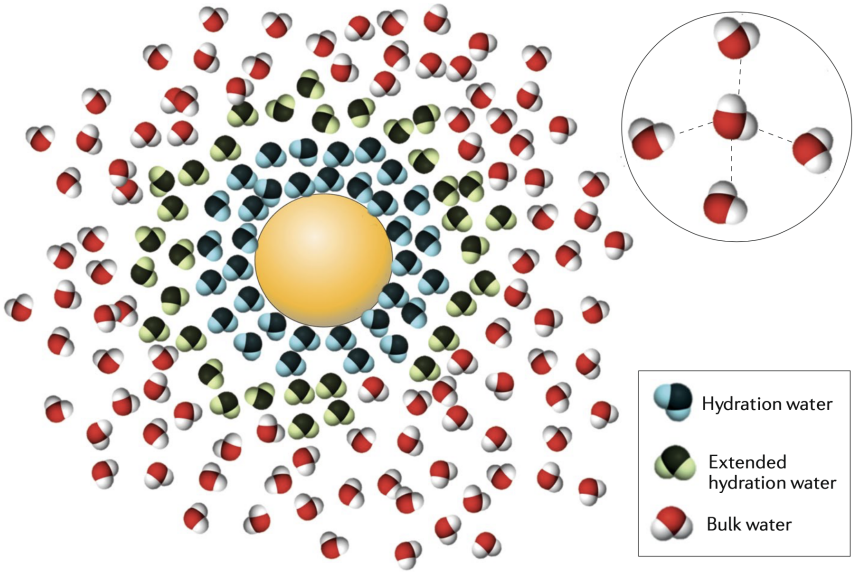


Figure 1.7 – Water classification in the solution. Hydration and bulk water around the solute [174].

Bulk water has a major contribution in the entropy in the dilute solute regime, however, in droplets the protein concentration is so high that the contribution that is classically identified as "hydration water" is becoming considerable as well.

An alternative to Flory-Huggins theory of phase separation is the theory of H_2O structure alteration originally formulated by Zaslavsky [224]. It proposes that polymers alter the structure and dynamics of H_2O to the point that it forms two different and mutually incompatible H_2O phases. In general, H_2O is more ordered around a protein's surface to the bulk, increasing the entropy and promotes demixing [205]. Also, hydrated water dynamics was shown to be slower (9 ps) in comparison with the bulk (2 ps) [202]. All this shows a decrease in entropy of water molecules which hydrate the polymer.

Some proteins with a high content of hydrophobic amino acids but low charged and aromatic amino acid content can also phase separate and interestingly, phase diagrams associated with such systems tend to exhibit an LCST. One of such polymers is elastin-like polypeptide (ELP). It is proposed that water molecules form a hydration shell around hydrophobic side chains and the release of the solvent into the bulk results in the increase of entropy and the collapse of polymer chains. The increase of the temperature weakens H-bonds making this process more favorable at higher temperature resulting in an LCST phase diagram [133]. ELP is an interesting example on which the transition between LCST to UCST was shown by the replacement of hydrophobic residues with charged residues [166]. This proves the hypothesis described above.

To conclude, both enthalpy and entropy can be responsible for phase separation, probably the characteristics of systems and the shape of their phase diagram differ because of the driving forces which initiate phase separation. It should be mentioned, that the entropy loss due to the restrictions on the polymer chain structure and dynamics were not discussed here. In order to be able to predict phase separation conditions this contribution has to be taken into account as well. That requires an

understanding of the difference in protein state between two phases what is the subject of Chapter 5.

1.1.4 The role of phase separation in biology

Membraneless organelles provide an alternative mechanism of intracellular compartmentalisation. In comparison with the membrane organelles they can be smaller (less than 1 μm), can be assembled and disassembled, gelates and recover dynamic liquid behavior, for example, on the response of chemical stress [80]. Inside droplets chemical reactions can be accelerated due to the increase of local concentration of the components, or sequester by selectively excluding some components. Using the selective permeability, which can be changed, the activity of the reaction is regulated and altered [149]. As it was already mentioned such organelles can become more or less dynamic, for instance in the nucleus ; chromatin undergoes such transitions, which allows regulating gene activity [3]. Chemical reactions in the cytoplasm are affected by biocondensates as well. Signaling molecules or some of the components can be harvested inside droplets and be released in response of stress [89]. As droplets are formed at the critical concentration of the component, they can also be used by the cell to buffer concentrations of proteins in the cytoplasm. One of the recent studies showed that after protein concentration in the cell reaches a critical level, the noise in cytoplasmic concentration variation is becoming much weaker [113].

A number of reviews which discuss biocondensates roles exist [5, 6, 149]. Their vital role in viral cycles is discussed below in section 1.4 and is the subject of Chapter 3.

1.2 Measles virus

1.2.1 The disease

Measles is one of the most contagious infectious airborne diseases caused by the Measles virus (MeV), which infects memory B and T lymphocytes and macrophage cells. Before vaccination was introduced in 1960s/1970s, virtually everyone experienced the disease during childhood [39]. The incubation period of measles is about 10-12 days after which initial symptoms of the disease appear: high fever, runny nose, cough, and skin rash, which is a signature of measles [215]. Measles is at its highest contagious level from 4 days before to 4 days after the rash onset. Despite a relatively low mortality rate, there is a list of serious side effects caused by measles: Immunosuppression, pneumonia, gastroenteritis, blindness, measles inclusion body encephalitis, and subacute sclerosing panencephalitis (SSPE). The latter is a rare complication of measles, which causes chronic progressive brain inflammation after 7-10 years of an initial infection, death occurring within 1 to 3 years [177].

In the pre-vaccination era, 30 million cases and >2 million deaths occurred each year globally. Besides, MeV was associated with up to 50% of all childhood deaths, mostly because of other infectious diseases, possibly due to the immune suppression caused by measles [139]. Despite the existence of a vaccine, measles is still common in many regions, including Europe, and affects about 7 million people every year with approximately 100 000 deaths globally [215]. Measles contagious rate is extremely high. The estimated reproductive number R_0, the average number of secondary cases of infectious disease, is usually estimated between 12 to 18, however, the value can vary between 4 and 50 depending on the way it is estimated. This means that each person with measles, on average, infects 12–18 other people. To stop viral transmission, the population has to be covered more than the herd-immunity threshold, which can be found with an equation: $p = 1 - 1/R_0$ which is higher than 90% for MeV [79]. Currently, there is no specific antiviral

treatment against measles.

Measles is believed to have evolved from an ancestral zoonotic virus that was transmitted from cattle around the 6th century BCE. Animal domestication occurred several centuries earlier, and the transmission of this ancestral virus between cattle and humans was possible, however, apparently a critical community size is required to support continuous MeV transmission. Measles virus separation from a zoonotic ancestor coincided with the rise of urban centers throughout Eurasia, South and East Asia which reached a critical size around the 6th century BCE. For measles the critical community size is estimated to be between 250,000 to 500,000 [52].

One of the consequences of MeV infection is a loss of immune memory. Following a contraction of the disease, a decrease in interferon signaling, cytokine response, lymphopenia, and suppression of lymphocyte proliferation was shown [120, 177]. What is more, immune memory cells seem to be affected as well. In a recent study, antibody levels before and after patients experienced measles were measured, and between 11 and 73% of the antibody repertoires were shown to be eliminated. During the disease, up to 70% of memory cells, including both B cells and T cells can be infected where measles replication impairs immune memory, resulting in "immunological amnesia". Notably, such effects were not observed on vaccinated individuals [139].

Measles virus has been studied for decades, the symptoms, virion structure and the life cycle are known. Its proteins are used as models to characterise Paramyxoviruses, which include Hendra and Nipah - deadly human pathogens, and to better understand general mechanisms of viruses. However, many processes remain unknown, especially viral replication. In order to understand the process of viral genome reproduction and its regulation, not only the structure of the protein has to be solved, but it should be put in the context of other proteins and nucleic acids, the kinetics of the processes should be revealed. Finally, viral regulation in cells has to be investigated. In Chapters 3 and 4 the progress in understanding these questions is presented.

1.2.2 Viral cycle

MeV is a member of the Paramyxoviridae family, Mononegavirales order. Its genome consists of non-segmented negative-strand RNA, which encodes eight proteins from six genes: N, P, L, H, F and M (Fig.1.8). The P gene encodes two additional proteins whose expression involves different frameshifts : V protein and C protein. The virion is an enveloped particle which contains the genome, covered with copies of N protein, in complex with P and L. M is associated with the membrane, H and the F are membrane proteins which are exposed on the surface of the virion (Fig. 1.9) [106] [177].

Figure 1.8 – Structure of viral genome. Linear negative sense single-strand RNA starts with a leader sequence on the 3'-end, followed by protein genes and terminated with a trailer sequence on 5'-end.

H recognises CD150/SLAM receptor which is expressed on the surface of T and B cells, macrophages and dendritic cells. Nectin 4, which is expressed at adherens junctions of epithelia, is a second major cellular receptor which H protein recognises [200] ; [143]. Then F protein fuses viral and cellular membranes and initiates the infection. F protein can also infect neighboring cells by fusing them together and results in multinucleated cells, termed "syncytia", which is a hallmark of many other paramyxoviruses [160]. After nucleocapsids (NC) are released into the cytoplasm of the target cells they serve as a template for both transcription and replication. The viral cycle is summarized in Fig. 1.9 [106].

The MeV replication machinery consists of viral RNA covered by nucleoproteins (N), termed the nucleocapsid (NC) that protects the genome against the host cell immune system, the Large

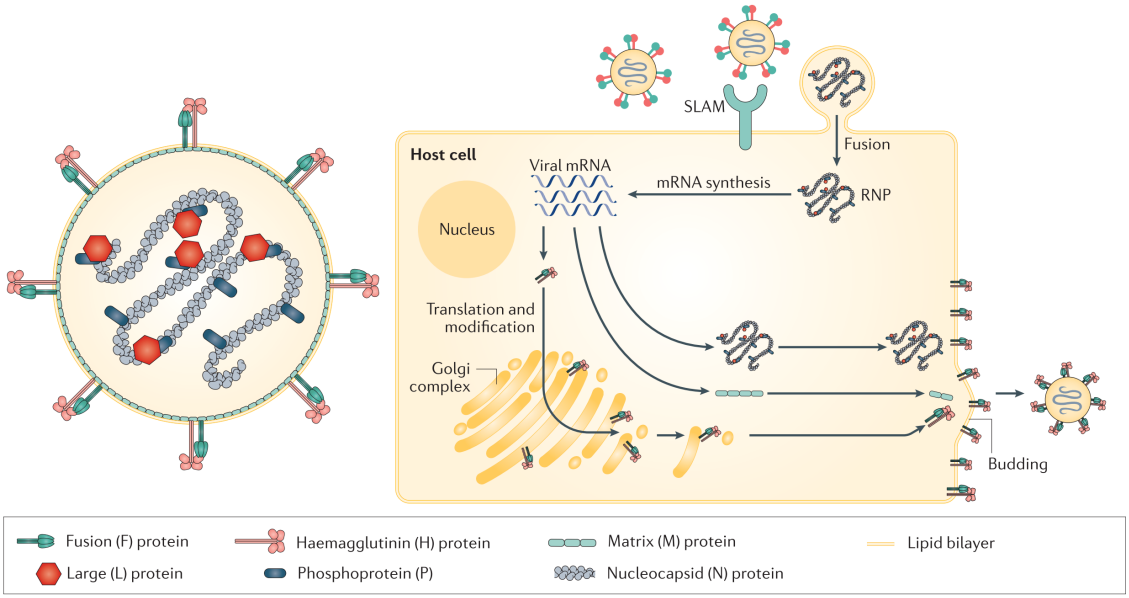


Figure 1.9 – The structure of MeV virion (left) and the schematic representation of viral life cycle (rigth). MeV genome together with N form a helical nucleocapsid (NC) that is associated with L and P proteins. H and F proteins are incorporated into the lipid envelope that is derived from the host cell membrane. M protein interacts with both the NC and the cytoplasmic tails of F, and promotes virion assembly. MeV viral life cycle: Following the binding of the H protein to the host receptor and membrane fusion with the help of F, viral NC releases into the host cytoplasm. Replication and transcription of the viral genome take place in the cytoplasm. Newly synthesized NC and other viral proteins are transported to the plasma membrane. M protein interacts with NC, the cytoplasmic tails of the H protein and the F protein, the cell membrane and actin filaments what promote virus assembly [177].

protein (L), the viral RNA-dependent RNA-polymerase (RdRP), and its essential co-factor, Phosphoprotein (P) [176]. As the replication process plays a vital role in viral reproduction, a detailed description of its molecular mechanisms is indispensable for the design of anti-viral treatments.

Initially, viral polymerase synthetises mRNAs for translation of new viral proteins. At some point it switches to the replication of viral anti-genomic full length RNAs, followed by a second cycle resulting in genomic copies. Once new copies of a viral genome are synthesised and encapsidated they are transported to the plasma membrane for M protein mediated virion assembly and budding. In contrast to many viruses, MeV does not use cellular ESCRT system for budding [106], [177].

Interaction with host factors

Viral metabolism requires the involvement of multiple host proteins. For instance, Hsp72 is implicated in the replication process [30]. Casein kinase II and PIM 3 kinase are responsible for MeV phosphorylation of N and P proteins [43, 199]. Also the cellular cytoskeleton is essential for viral replication and trafficking, assembly and budding [142]. In addition, cellular mRNA translation is inhibited by MeV N via its interaction with eIF3. This consequently promotes stress granules (SG) formation and silencing cellular mRNAs [186].

Cellular immune response

Uncapped ssRNA (with a 5-triphosphate) is a virus-specific feature which is used by the cell to recognise the infection. The retinoic acid-inducible gene I protein (RIG-I)-like receptors and

melanoma differentiation-associated protein 5 (MDA5) are two main sensors which recognise MeV RNAs. RIG-I is an activating kinase which phosphorylates interferon regulatory factors. Interferon activates the Janus kinase (JAK)–signal transducer and activator of transcription (STAT) signaling pathway, stimulating the transcription of various antiviral genes. P, V and C proteins were shown to be involved in the inhibition of the host interferon responses, probably by interacting with initial sensory proteins and blocking JAK-STAT signaling pathways [177].

1.3 MeV replication machinery

As described above MeV replication requires three proteins : L - large protein, which is responsible for the polymerisation reaction, N - nucleoprotein, which forms a template together with viral genome and P - phosphoprotein, an essential co-factor. In the following sections, the structure of all the components is discussed, together with their possible conformational rearrangements associated with function. In addition, interactions between components of the replication machinery are introduced together with their functional roles.

Phosphoprotein

P is a multifunctional protein whose sequence varies significantly over the Mononegavirales order, but whose architecture and functional roles are conserved. P of MeV is a 507 amino acid protein which has a multi-domain structure with alternating ordered and disordered regions (Fig. 1.10 (a)). It starts with a long disordered tail (304 amino acids, P_{tail}) followed by an oligomerisation domain (OD) of four long parallel helices between residues 304 and 377. At the C-terminus of P there is another folded domain (XD) (residues 457-507) which is linked to the tetramerization domain via the unfolded P_{loop} (residues 377-457). Structures of folded domains were previously solved using X-ray crystallography and P_{tail} was recently characterised in our laboratory using NMR and SAXS [137].

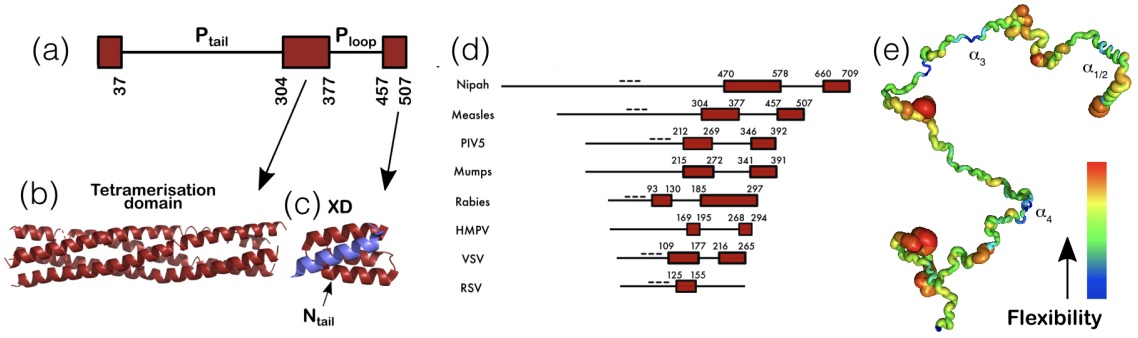


Figure 1.10 – (a) A schematic representation of the structure of MeV P protein. Bars corresponds to structural regions, lines to disordered regions; (b) crystal structure of MeV tetramerisation domain (PDB 3ZDO); (c) crystal structure of a complex between the linear helical motif of N_{tail} and the XD of P (PDB 1T6O); (d) a comparison of P structures from different viruses. Bars corresponds to structural regions, lines to disordered regions, "-" indicates acidic region in P_{tail}. (e) Localization of secondary structure propensity and representation of flexibility/rigidity of P_{tail} with colors from red (flexible) to blue (rigid) and the transient α-helices from the N-terminal helix (α1 to α4).

The oligomerisation domain is composed of four parallel α-helices interacting via hydropho-

bic contacts and was solved using X-ray crystallography (Fig. 1.10 (b)). The sequence consists of repetitions of the classical coiled-coil pattern comprising segments of seven amino acids, "a-b-c-d-e-f-g", in which the residues in positions "a" and "d" are usually hydrophobic and establish interaction between helices. This long hydrophobic contact determines high stability of the complex (melting temperature was estimated to be of 85°C), probably to avoid the presence of monomer-tetramer equilibrium. The helices are interrupted by a kink which appears near three leucine residues (L339 L340 L341) [37]. This kink was shown to be required to support viral gene expression and is conserved in oligomerisation domains across the Paramyxoviridae family [16].

The C-terminal domain, called XD, is a folded domain formed by a three helical bundle. Its structure was solved in complex with a C-terminal helix of N (see section 1.1.7, Fig. 1.10 (c)). Using NMR, XD was also shown to form three -helices in solution in the absence of N_{tail}.

P_{tail} is the largest domain of P (304 amino acids). The length of the architecturally conserved disordered N-terminal domain varies from 93 (for Rabies) to 470 (for Nipah) across Mononegalovirus order (Fig. 1.10 (d)) MeV P_{tail} was described using solution NMR and was shown to be mostly disordered, however, the first 37 residues form two transient α-helices which interact with N_{core} (see section 1.1.7). Using spin relaxation two other regions along P_{tail} were found to have propensities to form short helices: residues 87-93 (α3) and 189-198 (α4), (Fig. 1.10 (e)) [137]. The region between α3 to α4 is also highly charged making a long acidic loop. Such sequence features are conserved along P_{tail} in other viruses, for example, Nipah, PIV5, VSV, RSV, Rabies (Fig. 1.10 (d)).

Finally, another disordered region is situated between the oligomerisation domain and XD - P_{loop}. This was the only part of P which structure was not characterised prior to the work presented here, and its characterisation using NMR can be found in Chapter 3.

1.3.1 L protein structure

L protein is the longest protein of measles, but the least abundant, as its mRNA is transcribed last. The level of protein abundance is controlled by gradient transcription: Polymerase always starts with a 3'-end of the genome and in between genes it releases from the template in the junctions with a certain probability, for example for VSV transcript level is reduced by 20% at each junction [99]. So, the later the gene is placed on the genome, the less it is transcribed. MeV L is a 2183 amino acid long protein which structure was not solved yet, however, the structures of homologous viruses are known.

L is responsible for the synthesis of viral genomic RNA and mRNA what requires not only RNA transcription but also such modifications as capping, methylation and polyadenylation. By homology with recently determined structures (see below) we can hypothesize that these functions are all carried out by different domains of the enzyme: the RNA-dependent RNA polymerase (RdRp), the mRNA capping domain, and a methyltransferase domain (MT). In addition, L is thought to comprise a connector domain (CD) and C-terminal domain (CTD) and two disordered regions acting as linkers (Fig. 1.11) [127].

L is responsible for genome replication which is initiated at a certain viral life cycle stage and requires some changes in the polymerase activity: Avoidance of capping, ignorance of junction regions to synthesise a full-length anti-genome and later genome RNA [1]. To sum it up, L is a multifunctional protein which must exist in several conformations depending on whether mRNA or a genomic RNA is synthesised. Each of these processes also requires L to be highly dynamic to switch between initiation, elongation and termination stages.

Until recently, the structure of L had only been solved for vesicular stomatitis virus (VSV), an NNRV of the Rhabdovirus family, but lately several other structures have been determined on an atomic revolution in complex with P by Cryo-Electron Microscopy (Cryo-EM) [1, 27, 73, 94, 101, 127, 154]. L has a right-hand, "fingers-palm-thumb" structure similar to other RNA and DNA polymerases. Interestingly, different domains of P have been located in the complex with L

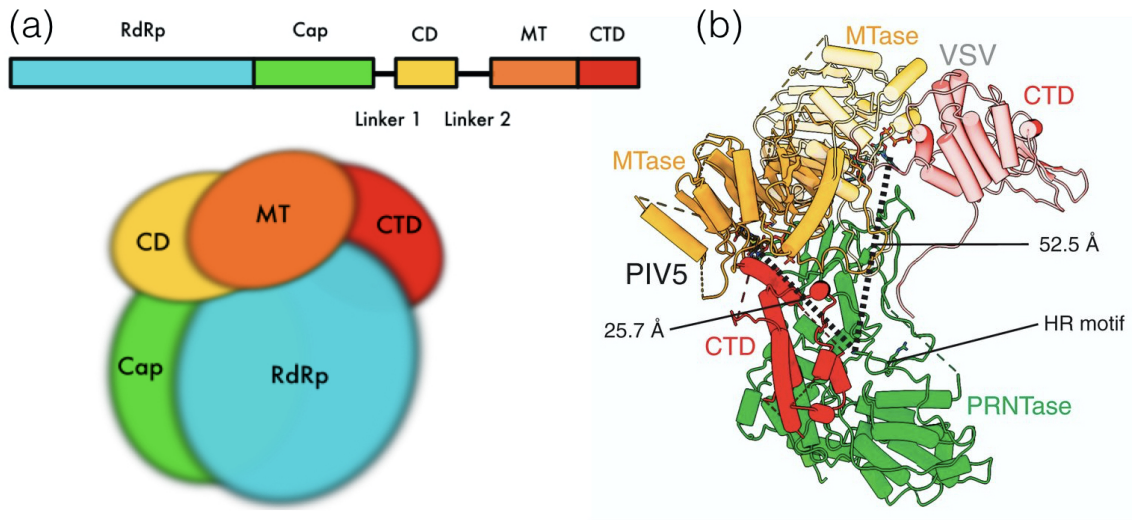


Figure 1.11 – (a) A schematic representation of the structure of MeV L protein. (b) The comparison of L protein structure of PIV5 and VSV. PIV5 structure corresponds to the transcription mode, VSV - to the replication mode [1].

for different viruses: For VSV and rabies virus (RABV), the full length L protein was visualized in complex with the N-terminal part of P with the resolutions of the structures being 3.0 Å and 3.3 Å, respectively [94, 101, 127]. The binding of P occurred on the C-terminal of L, which has also been described to be the binding site of SeV P by mutational analysis ([194]). Such alternating interactions of P may stabilise L in different sub-conformations and regulate activity.

For VSV and RABV the phosphoprotein has been observed to lock the CD, MT, and CTD domains into a stabilized conformation with respect to the RdRp-CAP module, leading to a reorganization of the structure (Fig. 1.11). When L alone is visualized by negative stain EM, the polymerase architecture resembles a doughnut, the C-terminal domains (CD, MT and CTD) were seen hovering on top of it. Binding of P to the linker segments at the ends of the CD leads to a rearrangement of the overall conformation and the capping domain folds back on the ring-like domain [127, 167, 168]. Overall, the structures of VSV L in complex with N-terminal of P was assumed to correspond to its initiation state and for RABV to elongation state by the conformation of the RdRp domain [95, 127].

Despite, it has been observed that the N-terminal half of the L protein is responsible for the L-P interaction while the C-terminal half has been found not essential for the polymerase structure stabilization for PIV5, MeV, SeV, HPIV3, hMPV and RSV [27, 33, 73, 93, 130, 154, 156], for other viruses, L has been found in complex with the C-terminal of P. For parainfluenza virus 5 (PIV5), the full length L has been observed with 4.3 Å resolution in complex with the C-terminal fragments of the phosphoprotein including the oligomerization and XD domains. But due to the poor observed density, a crystal structure of P OD was modeled to fill in the lack of information ([1]). The MTase active site is placed on top of the PRNTase domain, resulting in a much closer distance between the conserved HR motif and the active site of 25.7 Å, rather than 52.5 Å in the VSV structure (grey dashed lines on Fig. 1.11 (b)). Because of the proximity of the MTase active site to the HR motif, a conserved motif required for polyribonucleotidyltransferase activity, the conformation observed in PIV5 structure was speculated to represent a transcriptionally competent form of the polymerase [1], while VSV structure could correspond to the replication form.

The longest modeled fragment of P bound on the surface of L was around 100 residues for RSV and 80 for hMPV [27, 73, 154]. Due to the intrinsically disordered nature of P, the fragments that are not bound to L are not structured and cannot be visualized successfully. Therefore, there is no information for the whole organization of the phosphoprotein in relation to L.

Despite all existing data about viral replication machinery structure and interactions, the regulation of its activity, switch between transcription and replication, the mechanism by which L uncoats N in order to interact with the RNA remain unclear [1, 27, 94, 101, 127]. Probably, P is strongly involved in the conformational regulation of L what controls its activity. Multiple interactions between different parts of P and L are shown, for instance, a bipartite interaction has been found for MeV L and P where OD and XD bind at two independent sites of L ([162]). What controls and regulates L-P interactions still has to be explored.

1.3.2 Nucleoprotein

N can be present in an RNA-free, or monomeric, form and in an RNA-bound form – NC. The monomeric structure can be subdivided into a folded N_{core} formed by N-terminal 400 amino acid and disordered N_{tail}, C-terminal 125 amino acid which additionally contains a linear motif with α-helical propensity (residues 488-499) (Fig. 1.12). In solution protein sequence folds into several helices of different length that exchange with each other and with a purely unfolded form on timescales faster than microseconds [37, 102, 103].

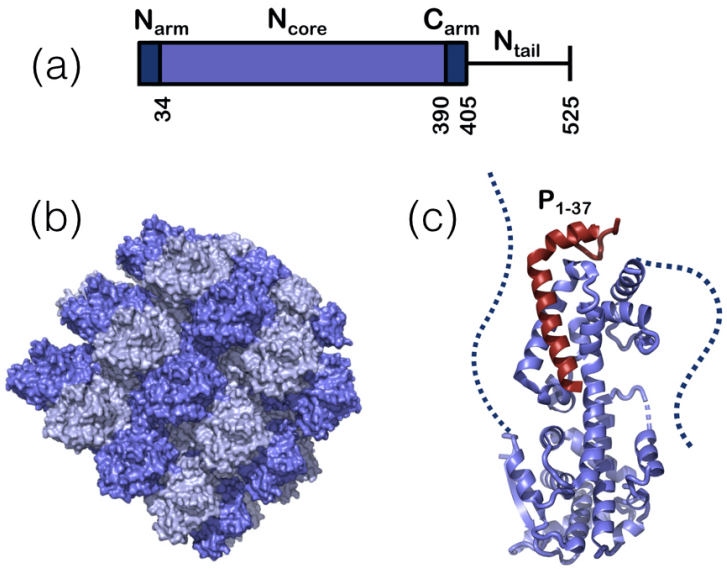


Figure 1.12 – (a) A schematic representation of the structure of MeV N protein. (b) MeV NC structure solved by Cryo-EM (PDB 4UFT), alternating colors correspond to different copies of N; (c) the structure of N^0P (PDB number 4CO6) Narm (left) and the C-arm (right) are shown with the dash, the region of P is shown in red.

N_{core} is flanked by disordered N and C-arms in the RNA-free form of the protein and becomes more structured upon binding to the neighboring N_{core} subunits in NC.

Once N protein is expressed in cells it directly binds to the cellular RNA and forms NC [134]. A chimeric protein was designed in order to obtain a monomeric N: It contains the first 50 residues of P followed by N (N_{core} or N_{full} can be used) with a TEV-cleavage site in between. Such construct is called N^0P [136]. As mentioned above, the N-terminus of P forms two α-helices which bind to N_{core} and chaperone it in a monomeric form. Expressing such a chimeric construct, the structure of RNA-free N was solved using crystallography (N and C-arms were also deleted to enhance crystallisation). The folded part of N is composed of two lobes with the RNA binding site located in a groove in between these lobes [222].

1.3.3 N and RNA interaction: nucleocapsid formation

N, together with RNA assembles into long helical structures - NC. The structure of NCs assembled around random cellular RNA was solved using Cryo-EM, however, because they were assembled around heterogeneous cellular RNA the electron density of the different RNA bases was averaged. In addition, the disordered N_{tail} was digested by trypsin, which resulted in straight NCs which was essential for solving the structure [83]. The comparison of this structure with the structure of monomeric N_{core} in complex with N-terminal P shows a global conformational change and reveals the mechanism of $P_{1\text{-}50}$ chaperone activity: The binding site of $P_{1\text{-}50}$ on N overlaps with the binding sites of N-arm and C-arm, which are involved in NC stabilisation. Despite the relatively high resolution (3.5 Å) of the folded part of capsid, it lacked information about RNA bases and N_{tail}.

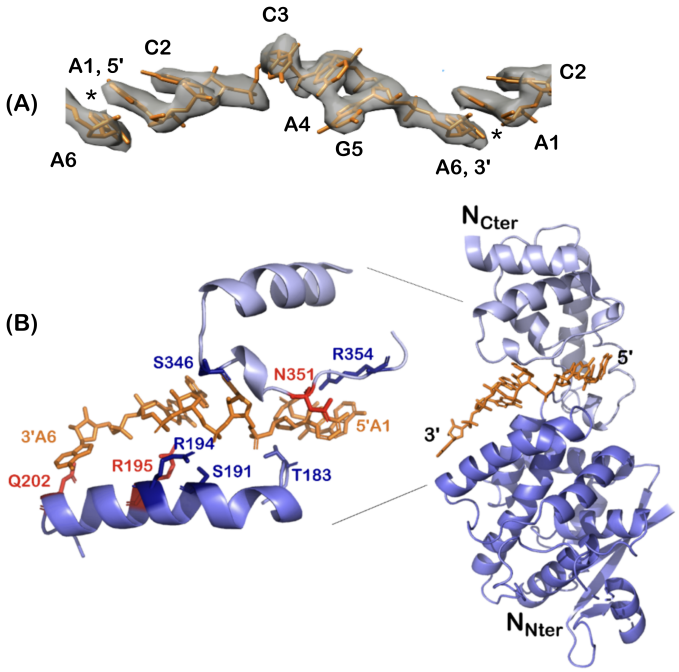


Figure 1.13 – Structure of the 5 viral genomic RNA (HO-ACCAGA-OH) inside the NC. (a) The Cryo-EM electrostatic potential of the six bases at the 5 end of the viral RNA inside the NC (EM data base EMD-0142) which forms A-helix with a turn every three nb. (b) RNA is squeezed between the two halves of N. The figure shows some of the residues binding the RNA, in red those that bind the bases [82].

An engineered N^0P construct can therefore be exploited to prevent the assembly process during protein expression. N^0P can be purified in a nucleotide-free form, and addition of RNA leads to assembly of N onto the RNA and therefore NC formation [136]. *In vitro* NC assembly from N^0P allows control of the sequence of RNA inside NCs and also controls assembly initiation,contributing to a better understanding of the assembly process on a molecular level. Using *in vitro* assembly one can obtain NC with homogeneous RNA sequence, using Cryo-EM and helical reconstruction its structure was obtained at higher resolution with electron density that was also better resolved, allowing identification of contacts between nucleotides and amino acids of N_{core} [47].

First, a global structural change during RNA binding was found: The angle between the upper and the lower lobes of N_{core} differ by 20°, assembled N adopts a "closed" conformation squeezing RNA in between. In addition it is suggested that the α6 helix which forms a lower lip of RNA binding pocket, exhibits high intrinsic flexibility, facilitating accommodation of different RNA sequences.

Using a known RNA sequence for structure determination revealed a number of contacts between RNA and N. K180, T183, R194, and R195 residues were identified previously [83], although the nature of the interactions were revealed, for example, bifurcation between backbone and base interactions with the sidechain R195, additionally, Q202, E263 and N351 were proposed to be involved in interaction [47]. The contacts are shown on Fig. 1.13.

In the NC structure each N subunit binds to exactly six nucleobases, explaining why MeV genome length is a multiple of 6. While binding RNA backbone adapts a certain conformation: Nucleotides on 1st and 2nd positions pointing out of the NC, making their bases solvent accessible, then first flip accrues, bases 3–5 are pointing toward the protein surface and after the second flip base six is pointing out. Bases in between flips (6-1-2 and 3-4-5) stack in the A-form conformation.

In the structure presented by Desfosses et al. NC was formed in complex with six base RNA lacking the 6th phosphate group, so the structure also verified that the binding register was maintained throughout the entire capsid. One can notice that in the shown NC structure the first three nucleotides at the 3'-end are exposed to the solvent, making them accessible for the interaction with the polymerase. This can be a possible initial recognition and binding site followed by RNA uncoating and transcription [47].

1.3.4 Kinetics of NCs assembly

The monomeric form of N allows the study of the kinetics of the NC assembly process which is illustrated in Fig. 1.14. Several methods can be used to study it from the side of RNA or from the side of the protein.

Using fluorescently labeled RNA the kinetics can be studied as a function of changing fluorescent anisotropy. Short RNA molecules have a fast rotational diffusion exhibiting low anisotropy. Once they are bound to N they experience an increase in anisotropy with further growth during assembly. So, by measuring fluorescence at a certain wavelength with alternating filter polarization before detection, the anisotropy can be calculated and the assembly rate extracted. An example of how such an approach can be applied to study the assembly is presented in [136].

To study the assembly from the side of the protein, NMR spectroscopy can be used. This allows to follow the process at residue specific resolution. As a reminder, the N^0P complex which is used for *in vitro* assembly is composed of a short region of P (1-50) bound to N_{core}, which are both not observed in ^{1}H-^{15}N HSQC spectra and disordered arms which are present on the spectra (a construct with N_{tail} can be used as well, in this case N_{tail} is also present on the spectra) [47, 136]. The example of such spectra is shown on Fig. 1.14. During assembly, $P_{1\text{-}50}$ releases, becoming observable on HSQC spectra, while the arms bind to the neighboring N_{core}s such that their intensity drops (as a population weighted average between visible and invisible signals). Measuring a series of HSQC spectra during the assembly process allows to follow the intensities of peaks corresponding to $P_{1\text{-}50}$ and to the arms and to determine the assembly rates.

Finally, NC imaging can be done with negative stain EM. Samples incubated for a different time with RNA show an increase in the length of NCs.

1.3.5 Multiple interactions between N and P

Several interactions between N and P are known in the literature, they include different domains and are present during different viral life cycle stages. The interactome of N and P is presented on Fig. 1.15.

First, as described above, the N-terminus of P (1-37) binds to N_{core} and maintains it in a RNA-free form by preventing self-interaction between N and C-arms and N_{core}. This interaction is present through Mononegavirales order via a conserved N-terminal motif, and has been called "soyuz1" [108]. This motif is placed in a hypervariable region (P_{tail}), that remains hydrophobic with a high helical propensity.

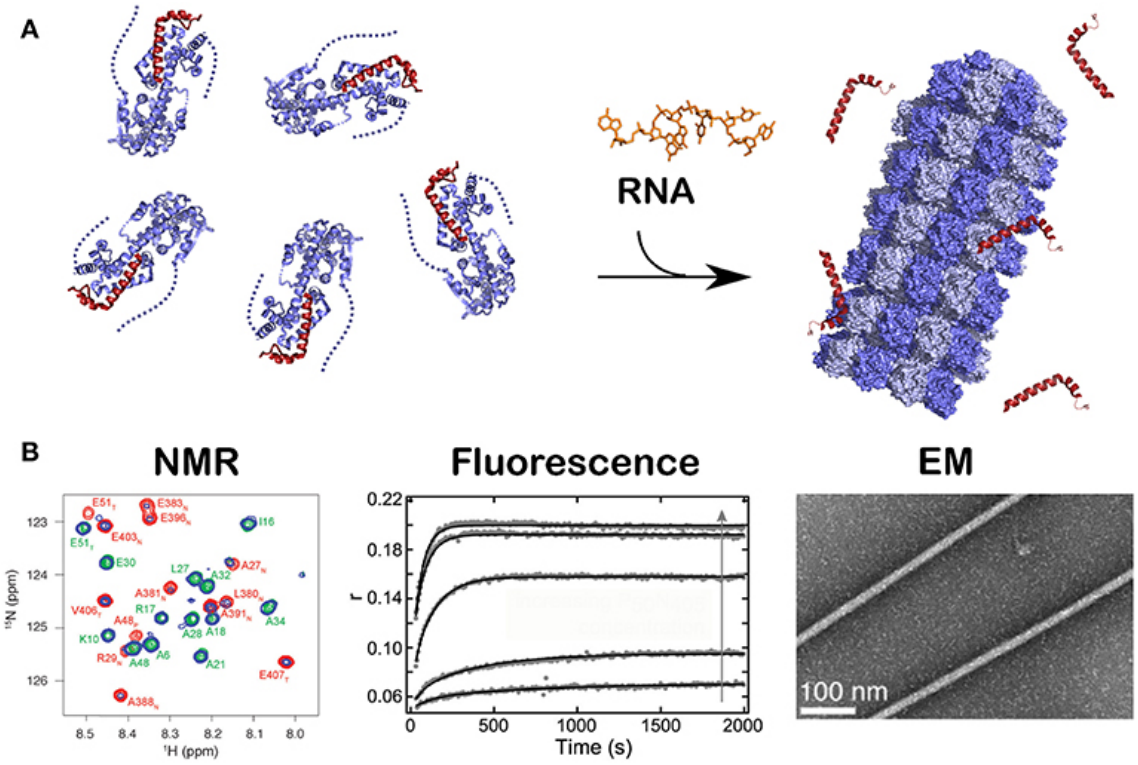


Figure 1.14 – Kinetics of the assembly of NC and methods to follow it: NMR, fluorescence and EM. When RNA was added to the N^0P complex, NC assembly could be followed by NMR in real time because the NMR signals for the P peptide appear over time, by fluorescence anisotropy using fluorescein amidite-labeled RNAs, and by negative staining EM showing the nucleocapsids [82].

As mentioned above, the C terminal domain of P (XD) also interacts with N via a disordered N_{tail} through a linear recognition element with a helical propensity. XD binding region of N_{tail} is separated from N_{core} by approximately 90 amino acids and appears as a transient helical structure in the free form, and folds upon binding to XD [1]. This interaction between P and N was speculated to be required for bringing L polymerase, which binds to the tetramerisation domain of P, in close proximity to NCs and therefore to the RNA [15]. As P does not interact with RNA, it probably plays the role of a connector between replication machinery components. This interaction is conserved along viruses highlighting its importance, although the actual role remains unclear [108].

Recently, a novel interaction between P_{tail} and N_{core} was found. It includes transient helix $\alpha 4$ (^{191}HELL^{194}) from the side of P_{tail} and the lower lobe of N. This second interaction between P_{tail} and N_{core} was shown to be required for maintaining proper function of N in MeV. The deletion of the interaction abolished viral transcription and replication in the cell. As this interaction is extremely weak (around 600 µM) it is difficult to detect using classical biophysical techniques. However, sequence similarities where found across paramyxoviridae, including Nipah and Sendai virus [137].

1.3.6 Phosphorylation

Phosphorylation modifications of both N and P are known in measles. P_{tail} is phosphorylated at the positions S86 and S151 and regulates transcription efficiency. Initially, the kinase involved in phosphorylation was identified - Casein Kinase 2 (CK2). These three serines were predicted to be phosphorylated by CK2 : S86, S151 and S180. These three serines were tested for phophorylation by mutagenesis using autoradiography and ^{32}P-labeled proteins, only the first two were confirmed

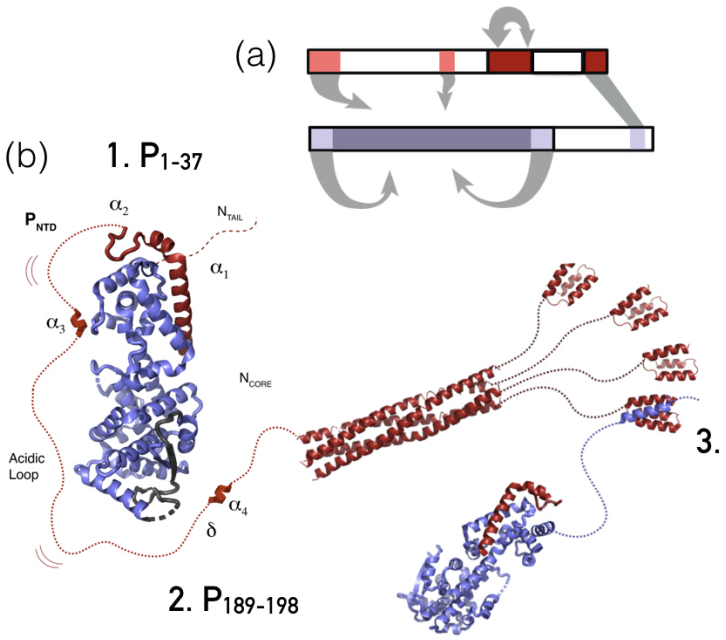


Figure 1.15 – N : P interactome. (a) A schematic illustration of interactions between P (red) and N (blue) regions. Folded domains are shown with red and blue color, regions with a transient structure that folds upon binding to the partner are shown with light red and blue, IDRs are not colored. (b) Structure of different P:N complexes (PDB 3ZDO, IT6O, 4CO6), three interactions are highlighted.

[43]. Phosphorylation of these sites downregulates the viral transcriptional activity [198]. Phosphorylation of N_{tail} at the position S479 and S510 was shown also using ^{32}P-labeled proteins. Early studies propose that these sites are phosphorylated by CK2 [84], however, recent studies show that they are PIM3 kinase dependant. Phosphorylation of these sites is involved in the regulation of viral gene expression and stability of viral genome RNA. PIM3 kinase co-localises with N in the cellular cytoplasm in speckles [199]. Furthermore, N_{core} domain was found to be phosphorylated at T279. Mutation T279A *in cellulo* caused malfunctions in nucleocapsid assembly [197]. Nothing is known about the molecular mechanisms underlying the regulatory role of phosphorylation in N or P.

MeV P phosphorylation is one of the subjects of this thesis and the recent progress is presented in Chapter 4.

1.4 Viral factories

Membraneless organelles play an important role in cellular biochemistry and are used for cell compartmentalisation without a need of a membrane. They are formed via liquid-liquid phase separation (LLPS) of scaffold proteins which usually contain IDRs and have multivalent interactions what allow formation of a network where other proteins (client) can be attracted [192]. In case of (-)ssRNA viral membranless organelles two proteins, apparently, play a role of a scaffold: nucleoprotein (N) and phosphoprotein (P) what was recently shown for rabies, HMPV, HRSV and SV5 [46, 69, 147, 164, 176]. At the beginning of my thesis the formation of such droplets had not been proven in the case of MeV, although co-localization of N and P in the cytosol was a known phenomenon since the 1970s [44]. Our work *in vitro* will be described in this thesis (see Chapter

3), but simultaneously to this work, N:P droplets were observed *in vivo* by [227].

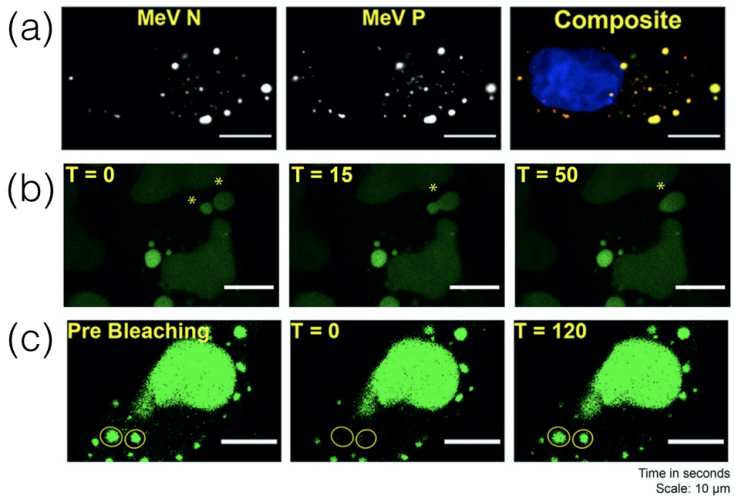


Figure 1.16 – MeV biocondensates exhibit liquid properties in cellulo [227]. (a) N and P co-localisation; (b) fusion between two droplets; (c) an example of FRAP experiment of MeV droplets.

Mononegavirales order includes 11 families and its members are characterised by linear non-segmented negative-sense single-stranded RNA [8]. Their viral particles are enveloped structures and are composed of 5 to 10 proteins. Viral membranless organelles were shown in cellulo for most Mononegavirales families (Bornaviridae, Filoviridae, Paramyxoviridae, Pneumoviridae, Rhabdoviridae). The three proteins mentioned above (N, L and P), that together form the viral replication machinery were found to co-localised in infected cells inside membraneless organelles [46, 90, 118, 151, 164, 176, 227]. This led to speculation that they are involved in the viral transcription and replication processes.

1.4.1 Viral droplets require N and P for the formation and exhibits liquid properties

The necessity of N and P to form viral droplets was shown for a range of viruses: Rabies, measles, Nipah, RSV [147, 151, 176, 227]. Overexpression of both N and P triggered droplet formation which grew in size and number over time [118, 147, 227], an example of measles droplets in the cellular environment is shown in Fig. 1.16 [227]. Such droplets exhibited properties characteristic for the liquid phase : spherical shape, ability to fuse associated with relaxation into a spherical particle (Fig.1 1.16), exchange of components with the surrounding environment what was shown using fluorescence recovery after photobleaching (Fig. 1.16 C). Similar characteristics were shown *in cellulo* for rabies and VSV [90, 147].

The requirement of different N and P domains for droplet formation was tested on rabies and it appeared that the minimum phase separating system was composed of a full length N together with P starting with oligomerization domain, followed by P_{loop} and XD. While P_{tail} was apparently not required for droplet formation, it plays an important role in N chaperoning (see above) and interacts with cellular proteins which could selectively attract them inside droplets what is discussed below. For some viruses (measles [227], Nipah [151]) over-expression of N on its own displays phase separation as well, however, this can be also related to the instability of N protein without chaperoning P_{tail}. The size of droplets was significantly smaller, so P appears to have a strong positive effect. In the case of VSV, L protein was also required for droplets formation, however, they were independent of the polymerization activity of L polymerase [90].

1.4.2 Multiple interactions between N and P

It was speculated that the XD-N_{tail} interaction is required for bringing L, which also binds to P, in a closer proximity to nucleocapsids and thus closer to RNA prior to transcription or replication. The necessity of this interaction for the formation of viral droplets formation was shown for rabies and measles [147]. Deletion of the XD domain suppressed phase separation *in cellulo* for rabies. Deletions of XD [147] or N_{tail} transient helix [69, 226, 227] were previously shown to inhibit droplet formation. This implies that XD-N_{tail} interaction is involved in the formation of a droplet scaffold and replication is suppressed as a result of a dissolvent of viral factories.

Apart from N: P interactions, the oligomeric state of P is probably also essential for droplet formation making interactions with P multivalent. There are a number of examples of phase separating proteins with sequential repeats of binding motives [164]. In the case of P, XD domains are linked in parallel, what can be an advantage in terms of saving genome space, as the same binding motif does not have to be coded in the genome several times. N also seems to have an ability to oligomerise, it was mentioned that N can form inclusion bodies *in cellulo* on its own. Additional evidence of N oligomerisation is present in S. Zhang et al. article, where the presence of a small amount of non-defective N rescued the ability to form droplets and synthetize RNA, and the defective N (interaction with P was suppressed by single point mutation) was present in droplets as well [226].

1.4.3 Viral factories - replication in droplets

Even thought only N and P are required for droplets formation, other viral proteins were shown to be present inside droplets as well: L, C, V and also viral RNAs (genomic, antigenomic and mRNA) [109, 118, 227] . In contrast to M and H which were shown to be excluded from the droplets [56, 90, 109, 118].

There is evidence that viral replication and transcription do take place in droplets. First, as mentioned above all the components of replication machinery are present and concentrated in droplets: N, P and L. Second, all types of viral RNA were found inside [69, 118, 226] and third, the concentration of viral RNA increased in droplets after infection and was only detected in droplets [176]. In addition, viral RNA was shown to be protected from RNAse action while it is inside droplets, N concentration was proposed to be higher on the periphery of the droplet, forming a potential mode of protection of the RNA [118]. In RSV protein concentration was also shown to be inhomogeneously distributed in droplets. And once viral mRNA was synthesised, it accumulates in a central part of the droplet together with M protein and translation initiation proteins (PABP and eIF4G), N, P, L and genomic RNA are excluded from this region. Probably, this sub-compartment is required for mRNA storage, by analogy with stress granules [176].

1.4.4 Client proteins

Viral droplets exhibit selective permeability. Some cellular proteins were found to accumulate inside viral factories, for instance WD5 [227], Hsp70 [118], Hsp72 [29, 109, 145], FAK [64], PABP, eIF4G [176], PI4KB [126], ubiquitinated proteins [118], others are excluded - TIA-1, G3BP, RNAses [176]. As proteins which were found in droplets are associated with cellular immune response, a second functional role of viral factories was proposed: Trap chambers for cellular immune factors. These traps would act by hooking signal proteins thereby dampening the cellular response as well as excluding cellular proteins such as RNAses, in order to protect viral components. Viral factories were also shown to protect RNA from interaction with stress granule proteins, which can adsorb viral RNA and suppress viral reproduction. Interestingly, Negri bodies were shown to be able to fuse with SGs but do not mix [147]. Scaffold proteins of SGs were also shown to be excluded from viral droplets for RSV [176]. Also, PI4KB binding is present on N for HRSV and on P for HPIV3 [126]. Some cellular factors accumulated in viral droplets are known

to accelerate replication and transcription: WDR5 [227] and Hsp72 [29] for measles, Hsp70 for rabies [117] ; Hsp72 for RSV [145]. Moreover, co-expression of N and P increases Hsp72 expression level as was shown for mumps virus [109]. Cellular phosphatase PP1 was also shown to be present in viral droplets. It participates in regulation of replication via phosphorylation of viral transcription factor M2-1 [175].

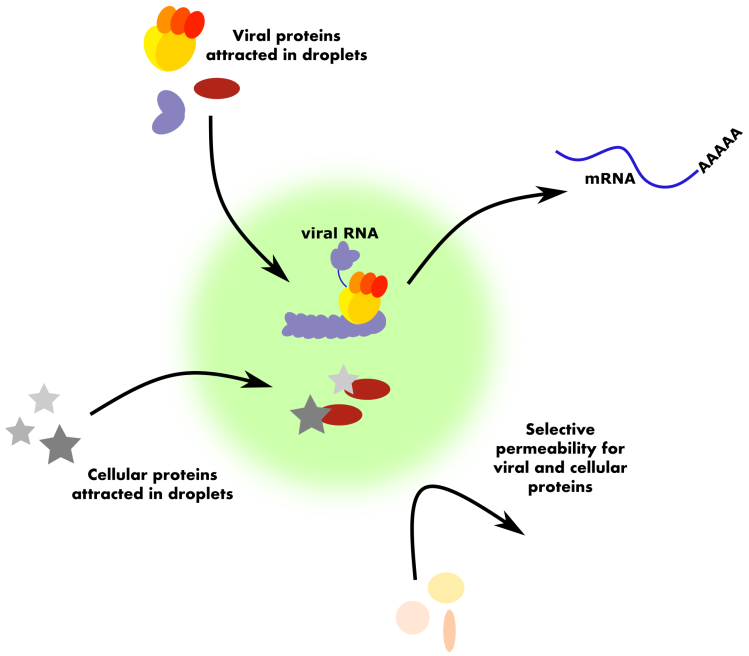


Figure 1.17 – An illustration of possible roles of viral factories.

1.4.5 Microfilaments

The cytoskeleton was shown to stabilise membraneless organelles and their distortion triggered the formation of a single droplet which sedimented to the bottom of the cell [57]. A similar effect was shown for rabies droplets: after cytoskeleton depolymerisation with nocodazole, single droplets were observed in case of rabies, however in the case of non-treated cells droplets size was in general smaller and they were distributed in the cell [117, 147]. The formation of droplets is cytoskeleton independent, actin, microtubules and dynein are not required for viral factory formation however, it is required for budding and transport along a microtubule is thought to facilitate distribution in the cell. Dynamics of microtubules apparently do not play role in budding as microtubules stabilisation with paclitaxel did not influence their formation and distribution [118]. In addition the ejection of ribonucleoprotein particles from the droplet following transport along the microtubules was observed during live cell imaging. Again the integrity of the microtubule network but not dynamics was required for the transfer [147].

1.4.6 Double membrane vesicles (DMVs)

Viral replication factories of some viruses are shown to be surrounded by a double membrane which originate from ER, Golgi apparatus or mitochondria. These are known as double membrane vesicles. The relationship between the membrane and non-membrane viral factories is not clear, however for certain negative sense RNA viruses it seems that membraneless organelles evolve into viral factories surrounded with a membrane during the infection cycle. Such evolution was demonstrated for rabies virus: Over time viral factories started to be associated with double membrane and finally completely surrounded. In addition, viral particles were observed to bud from these structures [118].

The existence of both membrane and non-membrane viral factories was also shown for RSV and HPIV3, however the co-existence of both structures at the same time was not found.

The benefit of the membrane in these cases (rabies, RSV, HPIV3) are not known, however, the attraction of some membrane proteins was shown, for instance PI4P which is essential for viral replication via regulation of L protein [126].

1.4.7 Multiple roles of viral factories

To conclude, the formation of viral factories display a range of advantages for the virus. First, it concentrates all of the molecular components involved in replication and transcription in a confined volume, thereby increasing the efficiency of the interaction processes. In addition, selective permeability of the droplets could protect viral RNA and proteins from degradation and immune system recognition. At the same time, essential cellular proteins, for instance, kinases can be selectively attracted into droplets. Also, some signaling proteins seems to be present in droplets, possibly they are silenced in viral factories in order to prevent activation of anti-viral cellular response. The summary of putative viral factory functions is illustrated in Fig. 1.17.

1.5 Intrinsically disordered proteins (IDPs)

Intrinsically disordered proteins (IDPs) are characterised by a lack of a constant 3D structure and instead represent a conformation ensemble in the solution, but nevertheless remain functional. The limits of the traditional paradigm "structure-means-function" were made even more evident with the discovery of IDPs [220]. The primary sequence of IDPs is rich in polar and charged amino acids, making their identification on the basis of sequence relatively unambiguous. Surprisingly, more than 40% of human proteins are predicted to contain disordered regions longer than 30 amino acids, pointing to the importance of such regions [150]. Indeed, many biochemical paradigms are now known where IDPs play an essential role: Viral disordered proteins, signaling pathways, stress response, transcription and translation [221].

The difference between a folded protein and an IDP can be illustrated on the basis of their energy landscapes. Folded protein energy can be minimised to the global minimum (or several) which corresponds to the stable structured conformation. Disordered proteins exist as dynamic polymers that rapidly interconvert between different conformations, and are better described by a relatively flat energy landscape without deep minima. Due to their highly dynamic nature, IDPs can be presented as an ensemble of conformers rather than a unique structure.

1.5.1 Functional classification of IDPs

There are multiple ways to describe the diverse biochemical roles that IDPs have been ascribed. As an example one can classify them into six classes by their function:
1. Entropic chains - disordered linkers which modulate distances between domains.
2. Display sites - disordered regions which contain sites for post-translational modification. As they are exposed to the solution, they are easily accessible.
3. Chaperoning - controlling the structural folding of other proteins and nucleic acids, can be involved in the conformations adaptation required for the complex formation.
4. Effectors - assist functions of other proteins. Via interaction with other IDPs can modify their activity and can work in both directions: Promote and sequester the activity.
5. Assemblers - recruit binding partners, work as a scaffold for biochemical processes.
6. Scavengers - store and neutralise small ligands, or other proteins. [207].

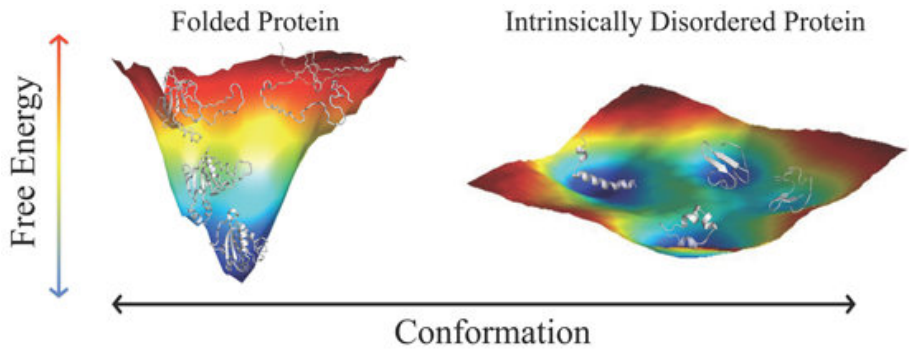


Figure 1.18 – Free energy landscapes of structured proteins characterized by a stable three-dimensional structure are dominated by a single deep funnel. Disordered proteins are characterized by shallow energy landscapes with several local minima of comparable free energy, separated by low barriers allowing fast interconversion kinetics between conformations [170].

IDPs often combine several functions which are maintained by spatially distinct parts of the protein. Another interesting feature of IDPs is that the same protein region can adopt different conformations (depending on the binding partner or post-translational modification or the environment) resulting in modulation of function [53]. In addition, interacting regions on an IDP can overlap allowing it to rapidly switch between binding partners and create a dynamic network [71].

1.5.2 IDP structure characterisation

IDPs can remain fully disordered or combine short structural elements, for example, transient helices or even folded domains, the examples of such proteins are presented above (MeV N and P proteins [82]). A complete folded protein and a fully disordered protein represent the extremities while most of the proteins represent the continuum of conformational states in between them.

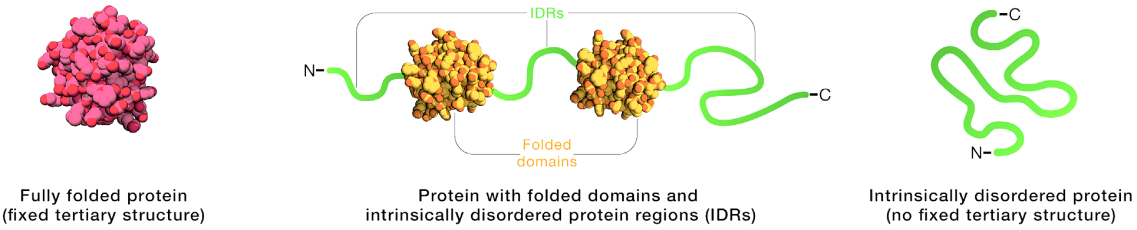


Figure 1.19 – A continuum of protein structures from fully folded protein to disordered [203].

Different degrees of disorder can be present along the IDP. Transient structural elements are frequently present and act as molecular recognition elements which may undergo a disorder-to-order transition upon binding to a partner one such example is measles N arms which fold into helices during assembly [136]. Some regions display structural propensity even in a free form which can increase upon interaction with the partner [102, 207].

1.5.3 Mechanisms of interaction

Two distinct mechanisms of IDP binding are proposed: Conformational selection and induced binding. The first, IDP is presented as an ensemble of conformations where one of the populations has an ability to bind to a partner. Once it is bound it shifts the population towards the binding-accessible. Following the second mechanism, a disordered region adopts a conformation while interacting with a partner [53]. IDP interactions often show both aspects at the same time. A good example of such combination is Sendai N_{tail}-XD interaction [189].

1.5.4 IDPs in viruses

The size of the genome of human viruses is limited due to the limitations on the virion size. ss(-)RNA virus genomes is in the range between 10 and 15 kilobase pair. As it was shown above, eight proteins are encoded by measles, however, various function has to be maintained [82]. Using IDPs the same protein can adopt different conformations and perform different functions during the viral cycle. For example, P_{tail} of MeV is involved in N chaperoning, but also binds to cellular proteins [136, 137]. Next, through molecular mimicry, viral IDPs could invade the host IDPs (or folded proteins) position involved in various cell regulatory processes and hijack host cell machinery.

In addition, IDPs exhibit a high mutation rate and thus fast evolution [22]. A single amino acid mutation can cause major changes in a viral function as it can affect several proteins at the same time if proteins are encoded with a frameshift. RNA virus mutation rates are 10^6 and 10^4 substitutions per nucleotide site per cell infection which is much higher than for eukariotic genomes [50]. Such rapid evolution is utile for fast adaptation and gives an advantage to continue to spread.

1.5.5 Structural characterisation of IDPs

The protein chain is composed of α-L amino acids covalently linked through amide bonds. The backbone chain arrangement can be depicted by two dihedral angles ϕ and ψ which adopts only certain combinations. The allowed angles areas are shown on a Ramachandran plot [169]. This constricts available conformational space [12].

In case of IDPs, where the protein is expected to continuously interchange between very different conformations on a rapid timescale, the most appropriate description of backbone dihedral angle sampling is to use a probability density function that maps the free energy landscape on a residue specific level (see below).

The overall compactness of the protein can be characterised by the radius of gyration (Rg) which is the root mean square distance of the protein atoms from the centre of mass:

$$R_g^2 = \frac{r_1^2 + r_2^2 + ... + r_n^2}{n}$$

For IDPs it is a probability distribution of R_g values rather than a single number. The probability distribution of end-to-end distances is also a characteristic parameter for IDPs.

Several models have been proposed to describe the behavior of IDPs, the simplest is probably a "random-coil" model, which assumes that the chain is composed of the beads freely joined and equivalent to a mathematical Gaussian chain model. In an ideal chain the end-to-end distance is [180]:

$$R^2 = Nb^2$$

and R_g is:

$$Rg = \frac{Nb^2}{6}$$

However, as was mentioned above, not all the values for the dihedral angles are allowed, so the actual protein is not a random coil chain. Flory suggested a power law dependency of the polymer size on a number of monomers:

$$R_g = R_0 N^{\nu}$$

, ν of an ideal chain is 0.5, for real polymers it depends on the properties of the polymer and solvent, for swollen linear polymer the scaling factor was estimated to be 0.59 [59]. Using experimental data ν was estimated to be (0.598±0.028) for denatured proteins [114] and 0.522±0.01 for IDPs [12]. So, IDPs show more compactness in comparison with denatured proteins.

In addition to the backbone dihedral angles restrictions, the nature of the side chains also affects the conformational space of the protein. The structural characteristics and populations of

the individual states in the conformational ensemble and the degree of compaction of the chain are determined by amino acids content and their distribution in the IDP sequence.

In general, IDPs are usually highly charged and rich with polar residues such as glutamine, asparagine, serine, glycine, and proline. A diagram-of-states can represent predicted conformational properties of IDPs depending on their sequence compositions (Fig. 1.20). The three axes denote the fraction of positively, negatively charged residues and hydropathy. The more hydrophobic residues are present the more globular and compact structure is predicted. Highly charged (both negative and positive) chains are represented as swollen coil. IDPs which are enriched in charged residues but have roughly equal numbers of positive and negative are called polyampholytes. Although they are placed in the middle of the diagram and thus predicted to be folded, in reality their dimensions depend on the distribution of the charges along the chain what can result in long-range attractive or repulsive interaction. For instance, it was shown that the alternation of charges along the sequence results in an extended chain while placing all positive charges on one end and negative on the other end of the chain decreases the R_g [42, 207]. Although these rather simplistic notions provide useful overall description of the physical nature of the disordered ensemble, additional aspects, such a polar interactions, interaction with solvent, ionic composition and temperature will all play important roles that will lead to a more complex behavior of a given sequence in solution.

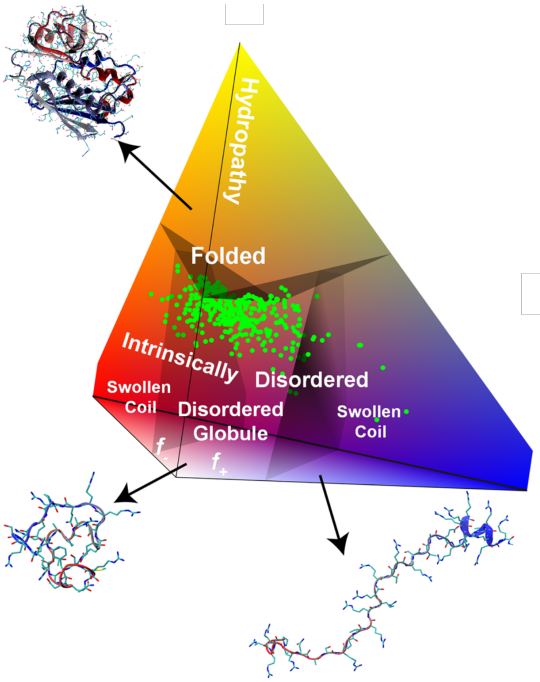


Figure 1.20 – Diagram-of-states to classify predicted conformational properties of IDPs [207].

1.5.6 Ensemble representation of IDPs

IDPs remain unstructured and highly dynamic in solution making it impossible to characterise with a single conformation. Instead the structural description of IDPs is a map of probabilities populating different regions of conformational space. A representative ensemble is the one which defines the behaviour of the protein and correlates the conformational ensemble with the function of the protein.

Due to the very large number of degrees of freedom available to such a disordered system, the problem of defining conformational space is highly underdetermined, requiring extensive experimental data to delimit the structural propensities of a given protein. Flexible meccano approach

uses the experimental information about the dihedral angle potential wells for each amino acid type derives from high-resolution X-ray crystallographic structures. While generating a protein ensemble, dihedral angles are randomly sampled using potential energy wells, each amino acid is added one by one. In order to avoid steric clashes amino acid-specific hard-spheres are used. As a result an ensemble of conformers is generated which represents a random coil [152].

Using additional experimental data short- and long range conformational restrictions or structural elements can be identified, by modulation of local conformational propensities as a function of experimental data, several such methods exist, for instance ASTEROIDS [13, 105].

Last, but not least, the role of the dynamics of IDPs should not be underestimated. IDPs backbone motions determine their behaviour and probably control protein ability to phase separate. The methods to study protein dynamics are presented in Chapter 2. Understanding how protein dynamics changes during phase separation is one of the goals of the thesis and is discussed in Chapter 5.

2
Methods

In this chapter methods to study IDPs and their interactions are discussed with an accent on NMR spectroscopy. In addition, applications to study LLPS using biophysical methods and especially NMR spectroscopy are introduced. Also, NMR methods to study protein dynamics are presented. I discuss recent advances in identifications of minimal protein scaffold required for liquid droplet formation, characterisation of their properties and interactions between proteins. Next, methods which were used during the experimental work presented in Chapter 3, 4 and 5 are presented.

2.1 Structural and dynamic characterisation of IDPs

Even though IDPs lack a unique structure they can be characterised by the overall compactness, conformational stability, shape, transient long-range contacts, regions of restricted or enhanced mobility, residual secondary structure. The most popular methods in structural biology, X-ray crystallography and Cryo-EM cannot be used due to the impossibility of crystallisation and the lack of electron density. However, solution state techniques can be applied for a long-range and short-range conformational description.

Circular dichroism (CD)

CD is the difference in the absorption of a right- and left-handed polorized light. This is a popular method to study secondary structure of proteins as the α-helices, β-sheets and random coil have a characteristic absorbtion spectra. Also the percentage of the protein structure or the population can be determined from the intensity of the characteristic peaks. Following the intensity of these peaks while changing the buffer or temperature, protein stability can be studied as well. However, CD does not provide residue specific information about the protein. Global secondary structure can be also studied by Fourier transform infrared spectroscopy of proteins (FTIR) and Raman spectroscopy [34].

Electron paramagnetic resonance (EPR)

The physical basis of EPR is similar to NMR (see later in this chapter), however, instead of nuclear spin, the electron spin is observed. This requires an unpaired electron for the detection, most proteins can be studied only with spin labelling. Site-directed spin-labelling allows local environment, ac-cessibility and spin mobility to be measured. Using double labelling the distances between labels can be determined. To conclude, using site directed spin labelling the information about motions and interactions of particular residues is obtained, however, this requires mutagenesis, chemical modifications with labels, each residue requires a separate sample and measurement [213].

Fluorescence anisotropy (FA)

Fluorescence anisotropy allows the study of dynamics and interactions between proteins and nucleic acids using fluorescence labels. The kinetics of slow processes can be followed as well, for example, nucleocapsid assembly was studied by measuring FA of RNA-FAM which increased during the process of encapsidation and NCs particles grow [136]. The principal of the method is the following: The fluorophore is excited with a polarised light and only those molecules whose excitation dipole moment is parallel to the polarised light will go to a higher energy state. After a delay (fluorescence lifetime) the photon will be emitted also parallel to an emission dipole moment. Two fluorescence intensities are measured with a vertical (I_V) and horizontal (I_H) polorizer placed before the detector to determine the anisotropy using the following equation:

$$r = \frac{I_V - I_H}{I_V + 2I_H}$$

The factor of 2 is indicating that the rotation in both clockwise and anticlockwise directions [119].

Fluorescence anisotropy depends on the viscosity of the solution, size and shape of the protein, which influences its rotational diffusion. The choise of a dye also plays a role first, because of the length of the linker between the protein and fluorescent group: The longer it is, the smaller is the anisotropy. In addition fluorescence lifetime affects anisotropy as it is the delay during which the molecule is rotating. So, to compare the anisotropy of the protein between different conditions, one must make sure that the fluorescence lifetime of the dye is not changing – a factor that can be influenced by changes in the environment [100].

Fluorescence lifetime and time-resolved fluorescence anisotropy

Fluorescence lifetime is the average time of a fluorescence molecule life in an excited state. It depends on the nature of a fluorophore, temperature and quenching.

It is measured as a time-correlated single-photon counting which is based on the detection of the arrival times of individual photons after excitation of a sample. The histogram of these times is fitted with an exponential decay (single or a linear combination of several) to determine the lifetime.

Measuring fluorescence time-correlated single-photon counting with vertical ($I(t)_V$) and horizontal ($I(t)_H$) polarisation allows to determine time-resolved fluorescence anisotropy:

$$r(t) = \frac{I(t)_V - I(t)_H}{I(t)_V + 2I(t)_H}$$

From the anisotropy decay curve rotational correlation time (r_c) can be found:

$$r(t) = Ae^{-t/r_c}$$

Again, if several rotations are present, the linear combination of exponents can be used to determine separate correlation times. The rotational correlation time characterises rotational dynamics of the molecule, in the case of IDP it can be sensitive to a chain like dynamics or segmental motions, as well as local dihedral angle fluctuations [41, 63, 100, 119]. Changes in the size of molecule, its rigidity or the viscosity of the environment will affect correlation time. In comparison with steady-state fluorescence anisotropy, changes in fluorescence lifetime do not influence r_c, allowing them to be compared between different samples even if the environment of the dye changes.

Förster energy transfer (FRET)

The efficency of energy transfer between chromophores depends on the distance between them. Using double labelled proteins the distance between labelled sites can be determined. Measuring a series of experiments with the labels placed at different positions allows the conformational dynamics of the molecule to be investigated, the extendness or the compactness of the molecule, and the presence of long-range contacts [119].

Single molecule FRET (smFRET) provides a direct observation of transient populations of sub-states within molecular ensembles as single molecule methods avoid ensemble averaging, thereby giving information about the complex distributions and dynamics of this important class of proteins [122].

Small angle X-ray/neutron scattering (SAXS/SANS)

In biological sciences, both techniques were initially applied to study folded proteins and complexes, however can be optimised for IDPs as well. These are low resolution methods which provides information about the size distribution and shape of the molecule. The scattered intensity I(q) is the Fourier Transform of g(r), the correlation function of the electronic density r(r), which corresponds to the probability to find a scatterer at position r in the sample if another scatterer is located at position 0 : elastic x-ray scattering experiments reveal the spatial correlations in the sample.

In the case of IDPs, the SAXS curve represents the average over the conformational ensemble. Unstructured proteins, due to the presence of extended conformations, are characterized by large average sizes compared to globular proteins. The radius of gyration, Rg, which can be directly obtained from a SAXS curve using a classical Guinier approximation, is the most common descriptor to quantify the overall size of molecules in solution [81].

As SAXS provides only averaged data, in order to extract molecular scale insights it is usually used in combination with molecular simulations. A pool of random conformations of the protein is used to generate individual scattering curves which are then used to fit experimental data and select a representative ensemble, similar to the ASTEROIDS approach. Different algorithms for generating initial pools can be used and different experimental data are used to apply the constrains on the conformations space (SAXS, SANS, smFRET, NMR). Experimental data which provides information about short- and long-range structural elements can be fitted [75].

2.2 IDPs by NMR

All the methods mentioned above experience limitations : either they do not give a residue-specific resolution or require labels that can only be used at a single site at the time. Nuclear magnetic resonance (NMR) provides structural and dynamics information at atomic resolution which measures the nucleus of the protein directly and not via an artificially added dye or spin label. A range of different NMR methods have been developed during recent decades which allow dynamic systems to be studied at different spatial and time scales.

2.2.1 Theoretical introduction to NMR

Each nucleus is characterised by its nuclear spin, or nuclear magnetic moment (μ) which produces magnetic interactions with an environment. The Zeeman interaction which is the basis of EPR and NMR is the interaction between the static magnetic field B_0 and the magnetic moment of the nucleus or electron. This interaction can be represented as Zeeman Hamiltonian (H_z) acting on (μ) and producing eigenstates. The Schrödinger equation can be used to determine the nuclear magnetic moment state :

$$i\hbar\frac{\delta\psi}{\delta t} = H\chi$$

The number of eigenstates depends on the nuclear spin number, for example for the proton it is 1/2 with two eigenstates :

$$\chi = (c_\alpha(t)c_\beta(t))$$

and produced with energies $E_{\alpha/\beta} = \pm 1/2\hbar\gamma B_0$, γ is the gyromagnetic ratio of the nucleus. Resonance absorption of electromagnetic radiation which corresponds to the difference between eigenstates is called NMR. The frequency of the electromagnetic radiation which is absorbed by the nucleus is called the Larmor frequency (ω_0) which depends on the external magnetic field and the gyromagnetic ratio of the nucleus :

$$\omega_0 = -\gamma B_0$$

This phenomenon of nuclear magnetic resonance was discovered by Rabi in 1938 and Purcell and Bloch in 1946 [165] ; [74]. Each nucleus type has its own gyromagnetic ratio. In addition, the local magnetic field of each nucleus is slightly different due to the electrons orbiting around the nucleus what shield them from B_0. So the effective magnetic field is

$$B_i = B_0(1-\sigma_i)$$

and so the Larmor frequency for each nucleus is

$$\omega_i = -\gamma B_0(1-\sigma_i)$$

This allows to resolve not only different types of nucleus, for instance ^{1}H from ^{13}C, but also the same nuclei experiencing different chemical environment, having different positions in the molecule.

Resonating frequency is measured in Hertz and usually displayed on a spectrum in units of chemical shift, parts per million (ppm), as it is field dependant value calculated relative to a reference the Larmor frequency as:

$$\delta = \frac{\nu_{sample} - \nu_{ref}}{\nu_{ref}}$$

Even though the chemical environment of each nucleus is slightly different, working with proteins it is impossible to distinguish all the protons due to resonance overlap. Multidimensional NMR is usually used to study proteins using nitrogen or carbon Larmor frequency in the second dimension. ^{14}N and ^{12}C spin quantum numbers are 1 and 0, however, their isotopes ^{15}N and ^{13}C spin numbers are 1/2 what makes them suitable to study by NMR. Protein expression in isotope labelled media is commonly used nowadays. Isotope labelling is not only useful for multidimensional NMR but also allows selective labelling to additionally decrease crowding, label particular amino acid types or groups, for example, methyl labelling, if necessary [31].

2.2.2 Comparison of IDP and folded proteins for NMR application

Working with IDPs brings both complications and simplifications for using NMR. Because of the disordered nature the average over all of the different chemical environments sampled by the different amino acids in IDPs does not differ a lot, this makes the differences in amide ^{1}H chemical shifts very subtle. Apart from the chemical shift difference between peaks, spectral resolution also depends on the linewidth of the peaks. Fortunately, IDPs have much more narrow peaks in comparison with folded proteins due to the fast rotational correlation time. The reason for this is introduced below [115].

2.2.3 Spin relaxation

When magnetic field is applied to the spin system the Hamiltonian contains only static components such as Zeeman interaction, chemical shifts and scolar coupling interactions and the system reaches an equilibrium state where longitudinal coherent component (M_z) is non-zero and transverse comonents (M_xy) are vanished. Each NMR experiment contains electromagnetic pulses which can be presented as time-dependent radio frequency field Hamiltonians ($H_{rf}(t)$) which perturb the equilibrium so M_z change and M_{xy} becomes non-zero. After the end of the pulse the system returns to equilibrium within a certain delay, the ensemble of mechanisms which brings the system to this static state is called spin relaxation. There are several time-dependant Hamiltonians acting during the relaxation delay: for reasons of clarity we will only discuss chemical shift anisotropy (CSA) and dipole-dipole interactions (DIPS) :

$$H(t) = H_{RDM} + H_{CSA} + H_{DIPS} + H_{JA}$$

Using NMR spectroscopy it is possible to observe how M_z and M_{xy} evolve. The characteristic time constant during which M_z reaches the equilibrium value is called longitudinal, or spin-lattice relaxation (R_1). Measuring signal intensities in a series of experiments with different delay during which the system is relaxing reveals the relaxation delay using:

$$M_z(t) = M_{z,\mathrm{eq}}\left(1 - e^{-tR_1}\right)M_z$$

The transverse, or spin-spin relaxation time constant (R_2) is the decay constant for the M_{xy} component. Relaxation rate can be determined similarly to R_1, however during the evolution delay M_z component evolution has to be eliminated what is usually done by a long rf pulse.

$$M_{xy}(t) = M_{xy}(0)e^{-tR_2}$$

Relaxation constants are dependent on the gyromagnetic ratio, magnetic field and rotational correlation time, as well as the strength of the relaxation-active interaction (for example the distance between nuclear spins for the dipole-dipole interaction). Gyromagnetic ratio of the proton is high, for this reason magnetisation can be transferred to the ^{15}N or ^{13}C which have lower gyromagnetic ratio. The second influence relaxation as a square, for this reason increasing the magnetic field does not always improve signal-to-noise ratio. The third affector is coming from the nature of the relaxation mechanisms: microscopic fluctuations causing the relaxation-active interactions which are proportional to e^{-t/τ_c}, where τ_c is correlation time [31]. More explicit analysis of relaxation is going to be presented below in this chapter.

Fast transverse relaxation of IDPs, due to the rapid correlation times, brings a huge advantage to their study by NMR. Peaks line width is dependant on transverse relaxation as :

$$\Delta\nu_{1/2} = R_2/\pi$$

, where $\Delta\nu_{1/2}$ is the peak width at half height. So, as R_2 of IDPs is shorter, the peaks are more sharp what increases the resolution of the spectra [211]. For instance, a spectrum of 304 amino acid long P_{tail} can be assigned despite the fact that the ^{1}H chemical shifts of the amide protons only span < 1 ppm [137].

2.3 NMR methods for IDPs characterisation

Proteins including IDPs are usually studies by 2D amide ^{1}H-^{15}N correlation NMR, several types of experiments can be used to obtain ^{1}H-^{15}N spectra: Single quantum (SQ) correlation experiments such as (HSQC and TROSY, or zero and double quantum correlation experiments (HZQC and HDQC), multiple quantum correlation experiments HMQC. The only amino acid type which is not present in the ^{1}H-^{15}N HSQC spectrum is proline because it does not contain an amino group. The correspondence of particular amino acids of the protein to the peaks on the spectra can be determined using 3D assignment experiments [32].

Assignment experiments

For protein backbone sequential assignment is usually done via several 3D ^{1}H-^{15}N-^{13}C experiments. They starts with ^{1}H-^{15}N and then magnetisation can be transferred on CO, $C\alpha$ and $C\beta$ of the same residue or the previous one. Next, comparing chemical shifts between CO_i and CO_{i-1}, C_i^α and C_{i-1}^α, C_i^β and C_{i-1}^β the sequence of connected peaks can be determined. Each amino acid type has a range of possible chemical shifts for H, N, CO, C^α and C^β. Some amino acid types have very characteristic chemical shifts, for instance glycine, serine, thyrosine. So, the correspondence of the peaks to an amino acid type can be done which helps to fit the determined sequence of peaks to the protein sequence.

IDPs can be rich in prolines which complicates the assignment as the sequence of connected peaks is disturbed. In addition, regions with repeats of amino acids are very similar in chemical shifts making it difficult to distinguish from each other [24, 115].

Structural propensities from chemical shifts

The chemical shifts are sensitive to the local backbone geometry and as it was mentioned above IDPs are much less disperse in the chemical shifts in comparison with folded proteins. Secondary structure and further folding increases chemical environment difference, thus, induce chemical

shift variation. Secondary structure has a particular effect on the chemical shift, in particular of ^{13}C backbone shifts and so-called secondary chemical shifts :

$$\delta = \delta_{observed} - \delta_{RandomColi}$$

provide information about the secondary structural elements propensities [195].

The random coil value for each nucleus type and amino acid is different, reference values arise from the experimental data. The reliability to distinguish secondary structural propensity from the random coil is in the order $^{13}C^{\alpha} > {}^{13}CO > {}^{1}H^{\alpha} > {}^{13}C^{\beta} > {}^{15}$N > ^{1}HN to distinguish an α-helix from a random coil; and $^{1}H^{\alpha} > {}^{13}C^{\beta} > {}^{1}$HN ~$^{13}C^{\alpha}$ ~^{13}CO ~^{15}N for β-strand. Algorithms which combine chemical shifts from different nuclei can also be used for a better structure propensity predictions [131].

For IDPs the values usually fluctuate around zero and differ significantly only in the regions of transient structures. Measuring secondary chemical shifts in free protein and in a complex with a binding partner can reveal the regions which experience folding upon binding transition.

Residual dipolar couplings (RDCs)

RDCs can provide an information about structure propensities and conformational behavior of IDPs. It is based on a through space dipolar coupling between two spins as:

$$D_{IS}(\theta) = \frac{\hbar\gamma_I\gamma_S}{4\pi^2 r_{IS}^3}[1 - 3\cos^2\theta]$$

, where γ_I and γ_S are gyromagnetic ratios, r_{IS} is the inter-spin distance, Ω is the angle between the inter-spin vector and the magnetic field B_0. In the isotropic solution due to fast tumbling and the average over all orientations in the ensemble, the net dipolar coupling is averaged to zero. Dissolution in an aligned medium results in a small percentage of the molecules adopting a preferential alignment breaking the isotropic averaging resulting in an observable dipolar coupling that depends on the population distribution of internuclear vectors with respect to the magnetic field.

For IDPs usually several RDCs are measured: D_{H-N}, $D_{H\alpha-C\alpha}$, $D_{H\alpha-C'}$, $D_{N-C'}$, $D_{H-C'}$, and D_{H-H}. For this HSQC type experiments are measured on a sample with aligned media and non-aligned. Then, peak splittings are measured, in a non-aligned medium only scalar couplings (J) between the nuclei of interest are observed. In the aligned media both J-coupling and dipolar coupling contribute. Comparing two sets of experiments dipolar coupling can be extracted. RDCs provide powerful structural constraints for NMR structure determination of folded proteins, while in IDPs, it is necessary to take into account the averaging of the alignment properties of an entire ensemble of conformers, which provides important insight into the local and long-range conformational behaviour of the IDP in solution [20, 104].

Paramagnetic Relaxation Enhancement (PRE)

PRE can be used to study long-range contacts (up to 25 Å) and solvent accessibility in IDPs. R_2 relaxation increase has a sixth power inverse dependency on the distance between the unpaired electron in the paramagnetic probe and the observed nucleus due to dipolar relaxation. For the first application paramagnetic spin labels at defined positions on the protein is used. Chemical modification is usually is usually introduced by mutating a cysteine into the sequence, and HSQC type experiments are recorded to extract peak intensities and additionally transverse relaxation is measures to directly detect the enhancement. As a reference sample with reduced PRE label (so non-active) is used – as the chemical modification can affect chemicals shifts and possibly cause additional shifts or broadening [144]. To study solvent accessibility paramagnetic probes are introduced directly into solution. Series of HSQC spectra with various percentages of paramagnetic probes in the solution are recorded to follow the intensity of the protein. The less broadening would be observed in the regions of less solvent accessibility [88].

Diffusion NMR

Translational diffusion of the molecule can be also studied by NMR. For this pulsed field gradient (PFG) spin-echo NMR techniques can be used. To measure the diffusion 1D or 2D experiments can be used, with two z-gradients applied to the sample. During the delay between the gradients molecules diffuse and magnetisation of those which position changed is not recovered, so signal intensity drops. In order to calculate diffusion coefficient several experiments with varying gradient strength are recorded. Then using following equation:

$$I(g) = I(0)e^{-D(\gamma g\delta)^2(\Delta-\gamma/3)}$$

, where $I(g)$ is the peak intensity at particular gradient strength, $I(0)$ - is the reference intensity at minimal gradient strength which is used to take into account losses in the intensity due to the relaxation. Δ is the diffusion delay, δ is the gradient time, g is the gradient strength and γ is a gyromagnetic ratio and D is a diffusion coefficient [25].

Diffusion coefficient is dependent on the temperature, viscosity and most importantly on the hydrodynamic (Stokes) radius. This can give an information about protein oligomeric state, binding, assembly can be followed, and for IDPs the extension or the compactness of the chain can be examined [51].

ASTEROIDS

Experiments described above give different type of information about IDP structure and dynamics. At the same time all the experimentally measured values are the average over the comformational ensemble of the protein under study. Molecular simulation provides a molecular characterisation of protein populations and combine all experimentally obtained protein characteristics.

One such approach is called ASTEROIDS (A Selection Tool for Ensemble Representations Of Intrinsically Disordered States). It is an algorithm which selects from a pool of a random coil conformations of the protein a representative ensemble which fit and complement experimental data.

An example of combining chemical shift information with RDCs and PRE data is shown in [183]. SAXS data were additionally included to NMR parameters [191].

Relaxation

^{15}N NMR relaxation (^{15}N-T_1,T_2 and ^{15}N–^{1}H NOEs) is a common method to characterise local conformational dynamics. CSA and DD are generally considered to be the same along the protein chain, so the variations in ^{15}N relaxation are associated with the difference in the local motions of the backbone. It is sensitive to a fast (ps–ns) time scale and different modes of motions can be extracted. For IDPs the dynamics has recently been interpreted in terms of three modes : Fast (50 ps) local librations, intermediate (ns) conformational sampling of backbone dihedral angles, and slow (10 ns) chain-like motions. Different ^{15}N relaxation rates are sensitive to the different modes of motions [115, 184].

NMR captures biomolecular motions at atomic resolution and supports understanding of the biomolecular dynamics-function paradigm. ^{15}N spin relaxation experiments map amide bond motions thought their dependency on a spectral density function (introduced below). So, the problem of NH bond dynamic characterisation reduces to the determination of a spectral density function. Relaxation rates are combinations of spectral density function values at several frequencies and cannot determine what these values are, it required a use of a theoretical model. The most commonly used one is called "model-free approach" proposed by Lipari and Szabo [128] is described further.

In the next sections correlation function and spectral density function are described, then relaxation experiments commonly used for protein dynamics study are discussed together with their dependency on spectral density function. And finally the theory of model-free analysis is expained.

2.3.1 The correlation between relaxation rates and protein dynamics

Correlation function describes random fluctuations in the system by measuring the difference between the states after different delays:

$$C(t,\tau) = \langle F(t)F^*(t+\tau)\rangle$$

where t is an initial time point, τ is the delay after which system state is compared and F is an operator defining an parameter which is compared. In other words, correlation function defines how fast the system loses its memory about the initial state and tends to 0. Fourier transform of the correlation function represents the speed of the evolution of the system at different frequencies and is called spectral density function:

$$J(\omega) = \int_0^\infty C(\tau)cos(\omega\tau)d\tau$$

In case of spin dynamics F operator represents fluctuations of the magnetic field which spin is experiencing, fluctuations are caused by rotational diffusion of the spin. Thus, using spectral density function rotational diffusion of the molecule can be determined.

As it was mentioned above relaxation rates are functions of spectral density as several frequencies and thus, store molecular dynamics information. In case of NH^N spin system five frequencies can be mapped on a density function: $J(0)$, $J(\omega_N)$, $J(\omega_H)$, $J(\omega_N - \omega_H)$ and $J(\omega_N + \omega_H)$, on Fig. 2.1 samplings of $J(\omega)$ by these frequencies and their contributions to different relaxation rates are shown.

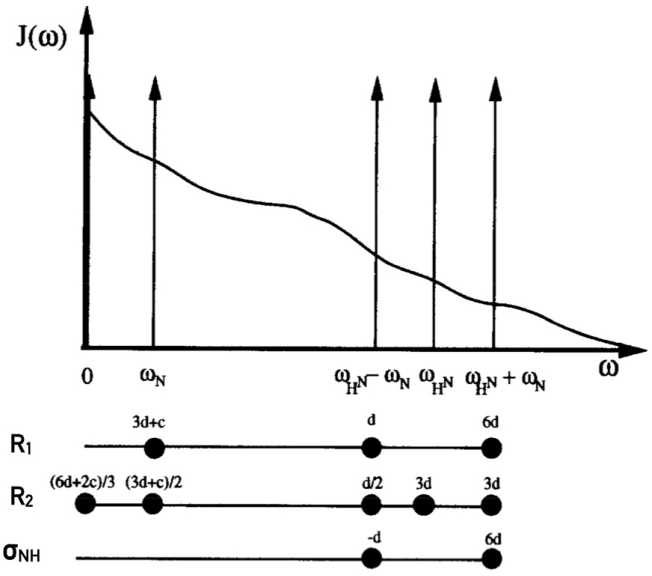


*Figure 2.1 – Schematic representation of the samplings of $J(\omega)$ by three relaxation parameters. Adapted from [***Peng:1991***]*

Contributions of spectral density function values at these five frequencies to longitudinal relaxation (R_1), transverse relaxation (R_2) and cross-relaxation (σ_{NH}) rates are the following:

$$R_1 = \frac{d^2}{4}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 j(\omega_N)$$

$$R_2 = \frac{d^2}{8}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_{H+N})] + \frac{c^2}{6}[4J(0) + 3J(\omega_N)]$$

$$\sigma_{NH} = 1 + \frac{\gamma_H}{\gamma_N} d_{NH}^2 (12J(\omega_H + \omega_N) - 2J(\omega_H - \omega_N)$$

where $d^2 = (\mu_0 h \gamma_H \gamma_N / 8\pi^2 r_{NH}^3)$ and $c = \Delta\sigma\omega_N/\sqrt{3}$. γ_H and γ_N are gyromagnetic ratios of H and N, h is a Planck's constant, r_{NH} is the length of the bond, $\Delta\sigma_{NH}$ is CSA of the spin, μ_0 is the permeability of free space. Cross-relaxation process is measured as a steady- state heteronuclear NOE:

$$NOE = 1 + \frac{\gamma_H}{\gamma_N} \frac{\sigma_{NH}}{R_1}$$

In total five parameters has to be determined from three equations what requires a theoretical model of motion. Lapari and Szabo model-free formalism relies two considerations: Modes of motions are independent and the motions are characterized by only two physical parameters: S^2 a generalized order parameter and τ_i an effective correlation time. Initially, two modes were proposed to be essential for the characterisation of protein dynamics: local and global. However, for IDP dynamic description three modes are required [72]:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_s}{1+\omega^2 b\tau_s^2} + \frac{(S_f^2 - S^2)\tau_i}{1+\omega^2\tau_i^2} + \frac{(1-S_f^2)\tau_f}{1+\omega^2\tau_f^2}\right)$$

where S^2 is the square of the generalized order parameter for the slow motion, S_f^2 for the fast motion and $S_i^2 = S^2/S_f^2$ for the intermediate motion. τ_s, τ_i and τ_f are rotational correlation times for the slow, intermediate and fast motions correspondingly. These motions corresponds to the local librations, backbone dihedral angle fluctuations, and chain-like segmental motions. To conclude, such analysis, allows to determine the influence of viscosity, temperature or other parameters on the dynamics of different motional modes.

2.4 Coupling of IDP dynamics to solvent

Recently a model which allows to predict IDP relaxation rates under various crowded conditions was proposed. It requires measurement of ^{15}N relaxation rates for the description of backbone motion. Then using microviscosity measurements chain dynamics can be extrapolated to the new environment using following equation:

$$\tau_k(C) = \tau_{k,\infty}(\varepsilon_k \rho(C) + 1)$$

where τ is the correlation time at given concentration (C), or ∞ - infinite dilution ans ε is a friction coefficient which should be determined for each motional mode. Microviscosity is proposed to be measured using water longitudinal relaxation:

$$\rho(C) = \frac{R_1^C - R_1^0)}{R_1^0}$$

Such model was successfully applied to predict relaxation rates in cell and potentially can be used to predict protein behaviour in droplets [4]. Water content and viscosity change dramatically in a biocondensate, and protein backbone dynamics are probably affec-ted as well.

2.4.1 Water relaxation as a proxy for effective viscosity of the protein

Solvent environment predetermines protein rotational and translational diffusion and several studies links protein hydrodynamics to water dynamics and shows correlation between water and protein rotational correlation [85]. Moreover, microviscosity influence the timescales of different motional modes of protein. So, water dynamics can be used as a proxy for local viscosity. Using NMR water rotational diffusion can me measured by $_1$H longitudinal relaxation.

Water is involved in hydrogen bonding with protein molecules and can interact with a NH group of the backbone and with aromatic and charged residues via their side chains. The residence time of the water molecule at hydration site is in a range off 10-50 ps at room temperature [85]. In supercrowded media hydrogen exchange with a NH peptide group was shown to remain the same , however side chains should be examined as well. So, water relaxation in droplet can be used as a tool to follow local viscosity.

An alternative to water relaxation measurements to determine effective viscosity in biological aqueous samples via ^{1}H can be measurement of ^{17}O which does not have a contribution of the R_{ex}. However, ^{17}O natural abundance is 0.037% and also dissolved molecular oxygen has to be eliminated from the sample by degassing before measurements. [173].

Recently, ^{23}Na was proposed as a reference molecule to measure internal fluidity by measuring its rotational transnational diffusion. Both parameters longitudinal relaxation rates and transnational diffusion coefficient are more more sensitive to subtle viscosity change for ^{23}Na then for ^{17}O. What is interesting, in this study 1D ^{23}Na spectrum of a condense phase showed double peak what suggests the existence of two microenvironment and the chemical exchange between them was confirmed [65]

2.5 Protein interaction by NMR

Using NMR to study protein interactions brings a lot of advantages in comparison with other methods. First, NMR is extremely sensitive allowing observation of ultra weak binding with a Kd of 10 s of millimole . Second, residual information about the involvement in the binding can be obtained with NMR as in a spectrum each peak can be treated independantly. NMR therefore allows distinct interactions to be followed using the same set of experiments [137]. All this gives enough information about the interface of binding and helps to build a model of a complex. Apart from the structural information about binding the affinity and on and off rates of binding can be determined using NMR [189]. It provides a range of different techniques sensitive to different affinity ranges and different time scales.

2.5.1 Chemical exchange

Protein interaction can be interpreted as a chemical exchange when the spin nucleus in the binding region experiences differences in the chemical environment. This exchange is observed even in the system which reached an equilibrium state. Two main characteristics of the exchange are constant of dissociation (K_d) and the exchange rate k_ex depend on the reaction rates k_1 and k_{-1}:

$$\mathrm{A} \underset{\mathrm{k}_{-1}}{\overset{\mathrm{k}_1}{\rightleftharpoons}} \mathrm{B}$$

Then

$$K_d = \frac{k_1}{k_{-1}}$$

$$k_{ex} = k_1 + k_{-1}$$

The first characterise the affinity of binding, the second is indicating the kinetics of the interaction [61].

It should be mentioned that as the spin experiences different chemical environments in different states, there is also the difference in chemical shift, thus in the Larmor frequency ($\Delta\omega$) and ppm. The effect on NMR spectra is dependant on the difference between k_{ex} and $\Delta\omega$ and different methods exist to study this exchange by NMR (Fig. 2.3). Chemical exchange affects resonance positions, intensities, and linewidths in NMR spectra [11].

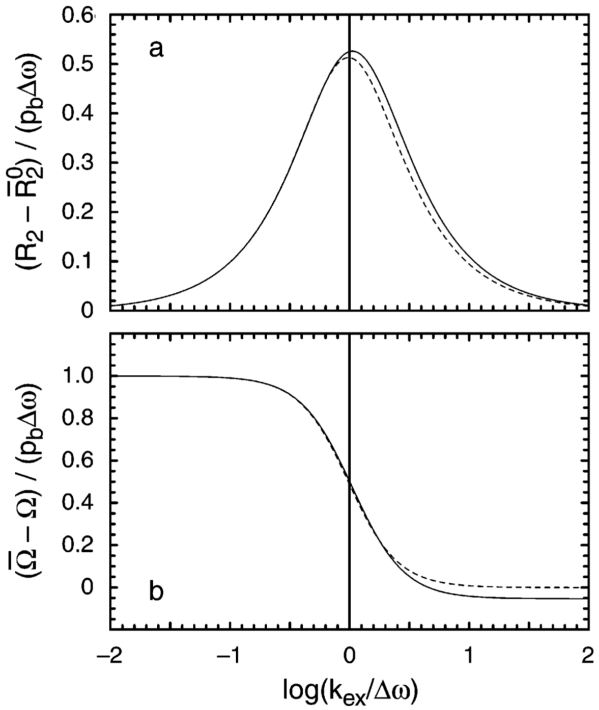


Figure 2.2 – (a) R_{ex} transverse relaxation rate. (b) Resonance offset for two-site chemical exchange constants [153].

Apart from the chemical shift dependency on the rate of the chemical exchange, the contribution of R_{ex} to the measurement of transverse relaxation is also affected. On the Fig. 2.2 the dependancy of R_{ex} and $\Delta\omega$ on the ratio $k_{ex}/\Delta\omega$ are shown. One can notice that the R_{ex} curve is at its maximum when $k_{ex} = \Delta\omega$ and so the maximum broadening is observed. From the dependency of R_{ex} and $\Delta\omega$ several regimes can be extracted: Fast, intermediate and slow exchange [153]. How NMR spectra are affected by them and which methods can be applied to study the interaction is described below.

Slow exchange

$k_{ex} \ll \Delta\omega$

The exchange is slow enough to reveal two peaks as initial frequencies ω_A and ω_B (Fig. 2.3). The intensity ratio between the peaks is dependant on the affinity and the concentration of binding partner, thus on the populations [61]. The intensity of the peaks can be interpreted as an estimate of the bound-fraction only as long as the transverse relaxation of the bound state is known to be the same as the free state, or can be accurately estimated, which is rarely the case. Alternatively, in the case of millisecond to sub-second timescale exchange, the so-called zz-exchange approaches can be used, which allow exchange of magnetisation to be measured as a cross-peak in a 2D-NOESY type experiment.

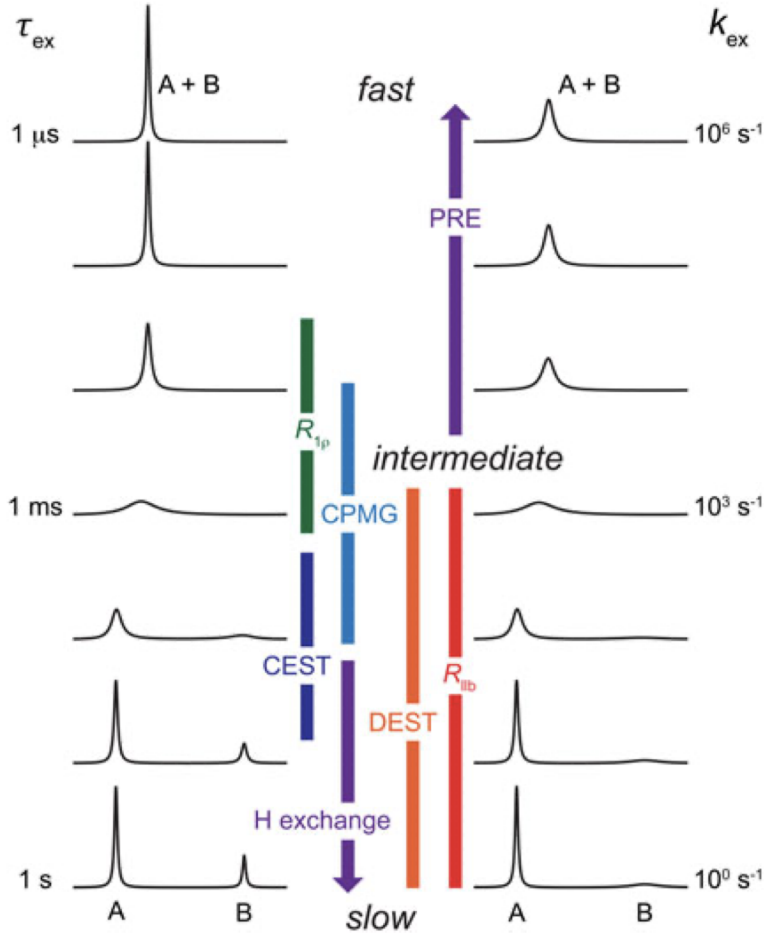


Figure 2.3 – Chemical exchange regimes ranked by the comparison between k_{ex} and $\Delta\omega$ cause different types of resonance in exchange between two states (A and B). NMR methods for characterizing the minor states in exchange with a visible species at different timescales with a rough range of timescales over which they can be applied. In the slow exchange regime, two separate peaks correspond to states A and B, although in practice state B may be invisible due to a low signal/noise ratio. In the fast exchange regime, a single peak is observed with a population-averaged chemical shift and apparent R_2. In the intermediate exchange regime, the peaks undergo extreme chemical shift broadening, and the apparent R_2 is greatly enhanced due to R_{ex}. The three methods shown toward the right (PRE, lifetime line broadening (Rllb), and DEST) are used to visualize a minor state with a greatly enhanced R_2 compared with the major state. The other methods shown, toward the left, depend on either a difference in chemical shift between states A and B (rotating frame relaxation dispersion ($R_{1\rho}$), CPMG relaxation dispersion, and CEST) or a difference in rates of hydrogen exchange [11].

If one of the populations is low or has a much slower tumbling in comparison with the other state, the peak can be absent. Saturation transfer experiments can be used to study the so called "dark" state. In the case of slow exchange and an exchange with a large size complex, dark state exchange saturation transfer (DEST) experiment can be used, the timescale range is 500 μs - 1 s. When the linewidth of the states differs enough, it is possible to saturate the resonances of an invisible state and the exchange rate is faster than the longitudinal relaxation rate (R_1), the magnetisation is transferred to the visible state and observed as a loss of the intensity [55].

Slow-intermediate exchange

$k_{ex} < \Delta\omega$

Broadening affects the peaks and also the peaks are present at the intermediate frequencies between ω_A and ω_B. The broadening arises from the increase in the transverse relaxation:

$$R_2 = R_2^0 + R_{ex}$$

The "invisible" state in this case can be studied using chemical exchange saturation transfer (CEST) experiment. Continuous-wave (CW) RF field at the invisible state resonance will cause a loss in intensity of the corresponding resonance in the NMR-visible major species, it is similar to DEST although unlike DEST, it requires significant chemical shift difference between the two states [206]. Next, the data can be fit to the McConnell equations, providing kinetic parameters (k_1 , k_{-1}) and populations p_A and p_B [135].

Intermediate exchange

$k_{ex} \simeq \Delta\omega$

The most broadening is observed at this regime that can lead to an apparent disappearance of the observed signal. As R_{ex} is significant at this regime and can be used to extract information about binding, one of such methods is Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion. During this experiment the apparent transverse relaxation is measured which has a contribution on R_2^0 and R_{ex}. The contribution of R_{ex} depends on a frequency of an applied spin-echo pulses and B_0 field. The data recorded at several fields (usually two are used) and by sampling the apparent relaxation at different spin-echo pulses frequencies (0 to 1000 Hz is a common range) the data can be fitted to the the McConnell equations, and similar to CEST, kinetic parameters and populations are estimated [86]. The faster the exchange the higher the effective frequency (CPMG or spin-lock) is required to sample the exchange events.

Fast-intermediate exchange

Fast exchange $k_{ex} > \Delta\omega$

The transverse relaxation is decreasing, which narrows the peak, however the intensity is still broadened and only one peak is observed.

Fast exchange

$k_{ex} \gg \Delta\omega$

Only one averaged signal is observed. The frequency at which the peak appears is closer to the Larmor frequency of the dominating population: $\omega_{AB} = p_A\omega_A + p_B\omega_B$ K_d can be determined from measurement of the peak positions on a HSQC spectra. The observed chemical shift (δ_{obs}) is the weighted average of the shifts in the free and bound states in this regime:

$$\delta_{obs} = \delta_f f_F + \delta_b f_B$$

where F goes for "free" and B for "bound". After certain mathematical transformations the dependency of δ_{obs} on component concentration and the K_d can be obtained:

$$\Delta\delta_{obs} = \frac{\Delta\delta_{max}([A]+[B]+K_d) - (([A]+[B]+K_d)^2 - 4[A][B])^{1/2}}{2[A]}$$

where $Delta\delta_{obs}$ is the change in the observed shift from the free state, and $\Delta\delta_{max}$ is the maximum shift change on saturation. The saturation curve where the concentration of one of the components ([B]) is varied is fitted with the equation and K_d is determined [217].

In case if the interaction is weak and the saturation cannot be reached, K_d can be extracted from the transverse relaxation measurements. In the fast exchange regime the R_{ex} contribution is negligiable so the measured value of R_2 is a combination of transverse relaxation at both states:

$$R_2 = \frac{AB}{AB+A} R_2^{AB} + (1 - \frac{AB}{AB+A}) R_2^A$$

where R_2^{AB} is a transverse relaxation of a complex and R_2^A of a free A component. This equation can be rewritten in the form containing K_d:

$$R_2 = \frac{B+AB}{K_d+A+AB}(R_2^A B - R_2^A) + R_2^A$$

Measuring R_2 values at several B concentrations allows to fit the curve and estimate the K_d. It should be mentioned that this approximation works only for the fast regime, where R_{ex} is negligible and the interaction is weak so the fitted curve is linear [138].

2.6 Biophysical methods to study phase separation

During recent years multiple protocols to study phase separating systems and the methodology for their characterisation were proposed. A range of methods for qualitative and quantitative analysis of phase separation can be applied. Below an overview of several of the most popular methods is presented.

2.6.1 Turbidity

Phase separation in the solution can be detected even by eye at certain concentration, as the solution is becoming milky because of light scattering by colloids. In order to estimate it quantitatively, the absorption at 400 nm, 500 nm or 600 nm can be registered. The higher the value, the higher the turbidity and so the phase separation is more pronounced. Usually the value of 0.1 is taken as a threshold of phase separation [17, 155].

This method is quick and does not require specific instrumentation, small volumes are required especially if a nanodrop is used, so a detailed screen of conditions can be easily obtained to plot a phase diagram. However, light scattering can be caused by many factors apart from phase separation: aggregation, precipitation, amyloid formation, so, additional methods are necessary to confirm phase separation.

2.6.2 Microscopy Light and Fluorescent

Using microscopy, the presence of colloids in solution can be imaged to confirm phase separation and not aggregation (in case of phase separation, spherical particles are observed). Size distribution, number of droplets and the morphology of droplets can be calculated [171], also microscopy is more sensitive than turbidity test and allows to verify the phase diagram (Fig. 2.4).

If the droplets are highly hydrophobic, they can wet the surface of the glass slip and can cover all the surface which makes the observation of the phase separation more complicated. Pegylated glass or addition of BSA in the solution can be used to avoid droplet wetting of the surface [7, 70, 97, 182]. A small volume and low concentration requirement are the advantages of the method.

Several parameters of the droplets can be followed using light micropscopy: Size distribution of droplets and number per cell [14], following fusion events to determine viscosity/surface tension

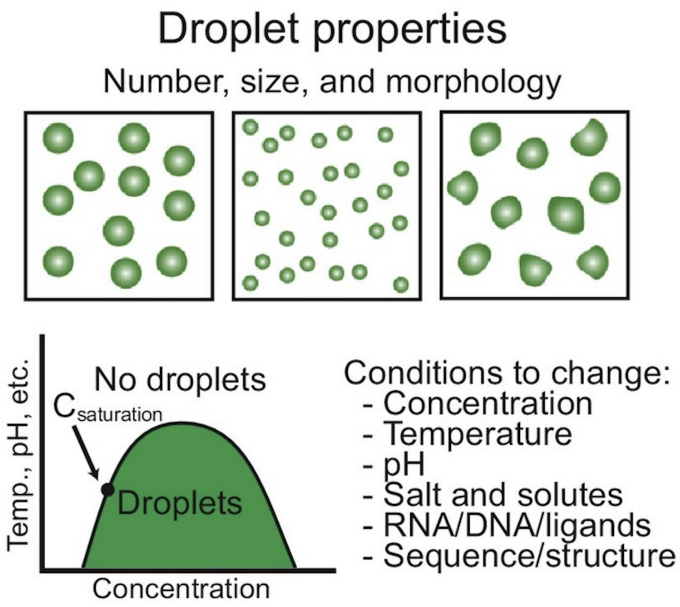


Figure 2.4 – Light and fluorescent microscopy provides information about droplets number, size and morphology (top panel). The phase diagram indicates regions of environmental conditions with single and double phase. Phase diagram can be done as a function of various parameters (lower panel) [68].

and viscoelasticity [58]. Using confocal microscopy droplet surface tension can be determined from the contact angle or using right angle imaging (Fig. 2.5) [58] [97]. All these proves the liquid

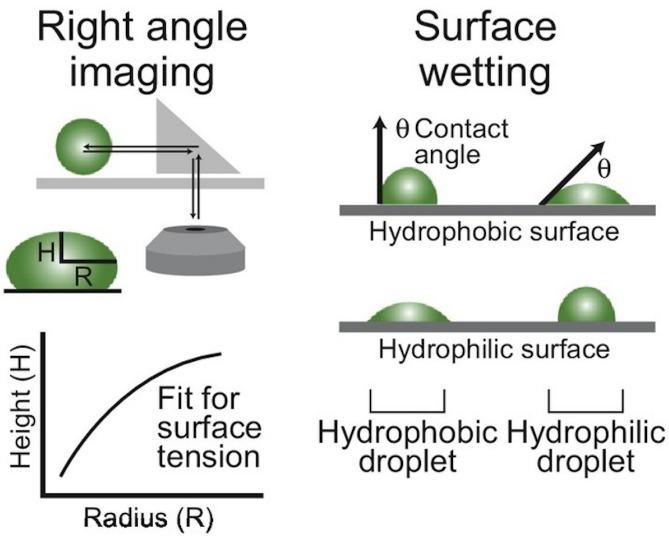


Figure 2.5 – Droplets surface tension determination. Right angle imaging technique measures the height and the radius of droplet from which surface tension can be obtained. Surface wetting technique provides surface tension value from the contact angle of the droplet. This requires information about the hydrophobicity of the surface [68].

nature of the droplets and additionally provides individual characteristics.

Fluorescent microscopy allows selective labelling of components of the solution and following its distribution in between phases. Several components can be followed simultaneously using multilabels that can reveal co-localisation or inhomogeneous distribution (Fig. 2.6) [58, 185].

In addition, protein concentration can be determined using fluorescence intensity *in vitro* or *in*

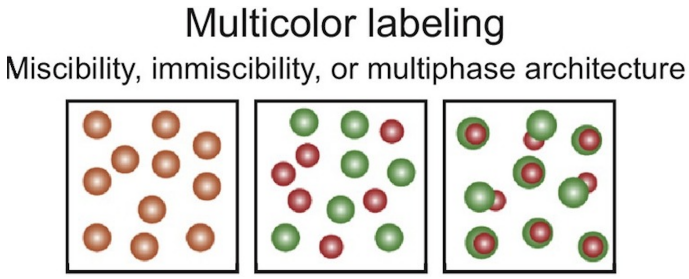


Figure 2.6 – Fluorescence microscopy provides a source to determine molecular distribution in droplets. Protein co-localisation or demixing can be determined, also several phase co-existence can be imaged [68].

vivo and even be tracked during the cell life cycle [113, 121, 209].

2.6.3 Phase diagram analysis

Using protein concentrations in co-existing values a full phase diagram can be estimated with a Flory-Huggins model. It is a mean-field model which is traditionally used to describe the behaviour of phase separation systems with two components. It is based on Helmholtz free energy change upon polymer and solvent mixing and introduce a mean parameter of polymer-solvent interaction χ which defines the difference in interacting energies between polymer and solvent u_{AB}, polymer-polymer u_{AA} and solvent-solvent u_{BB} as :

$$\chi = \frac{z(2u_{AB} - u_{BB} - u_{AA})}{2kT} \tag{2.1}$$

Then the energy of mixing is:

$$\Delta U_{mix} = \chi\phi(1-\phi)kT \tag{2.2}$$

And the Helmholtz free energy of mixing:

$$\Delta F_{mix} = \Delta U_{mix} - T\Delta S_{mix} = kT(\frac{\phi}{N_A} ln\phi + \frac{1-\phi}{N_B} ln(1-\phi) + \chi\phi(1-\phi) \tag{2.3}$$

The entropic contribution to the energy, which is small for polymers, is promoting mixing, the last one can be positive or negative depending of the sign of χ. Thus, the energy of mixing can have an attractive or repulsive effect depending on the χ parameter, and the degree of polymerisation of molecules. Obviously, the energy of mixing also depends on temperature and volume fractions of the components. Determination of the χ parameter is important to understand the properties of the solution. It depends on the temperature, chain length and composition of the polymer in a non-trivial way, empirically, it is depends linearly on temperature:

$$\chi(T) = A + \frac{B}{T} \tag{2.4}$$

, where A is an entropic and B is an enthalpic part.

In order to determine this interacting parameter a temperature–composition phase diagram has to be constructed. The boundaries of the phase diagram can be determined by measuring polymer composition in co-existing phases at different temperatures ($\phi^{|}$ and $\phi^{||}$). F_{mix} dependency on the composition is asymmetric in case if $N_B \neq N_A$, so the common tangent line is non-horizontal and $(\frac{\delta^2 \Delta F_{mix}}{\delta \phi^2})_{\phi=\phi^{|}} = (\frac{\delta^2 \Delta F_{mix}}{\delta \phi^2})_{\phi=\phi^{||}}$ correspond to two equilibrium phases. By equating the second derivative of free energy to zero:

$$\frac{\delta^2 \Delta F_{mix}}{\delta \phi^2} = kT(\frac{1}{N_A \phi} + \frac{1}{N_B(1-\phi)} - 2\chi) = 0 \tag{2.5}$$

This curve corresponds to the boundary between unstable and metastable regions (spinodal). And if χ can be extracted:

$$\chi = \frac{1}{2}(\frac{1}{N_A \phi} + \frac{1}{N_B(1-\phi)} \tag{2.6}$$

, so

$$T_s = \frac{B}{0.5(1/N_A\phi) + 1/(N_B(1-\phi))) - A} \tag{2.7}$$

, which can be fitted to phase diagram to determine A, B and χ.

The critical point corresponds to $(\frac{\delta^3 \Delta F_{mix}}{\delta \phi^3}) = 0$. For a polymer-solvent solution, $N_B = 0$ the critical composition is $\phi_C = \frac{1}{\sqrt{N}}$, critical interacting parameter $\chi_C = \frac{1}{2} + \frac{1}{\sqrt{N}}$ and the critical temperature $T_C = \frac{B}{\chi_C - A}$.

2.6.4 FRAP

Fluorescence recovery after photobleaching is a common method to prove the liquid nature of droplets and protein exchange between phases. FRAP is summarised in Fig. 2.7. Briefly, the area is illuminated with a high power laser, bleaching fluorescent labels in this region. The recovery of fluorescent intensity is associated with a diffusion of the labeled molecules. Application of FRAP to a phase separated sample requires fluorescent label and the use of the confocal microscope. Protein exchange between phases can be followed by bleaching an entire droplet, the diffusion inside droplet can be followed by bleaching a small area inside the droplet. Using multiple labeles FRAP can be applied simultaneously to several components of the droplet.

In order to analyse quantitatively FRAP the profile has to be fitted with a model. In the simplest case if protein exchange between phases is only dependant on the molecular diffusion two models are usually used, first with a simple exponent:

$$I(t) = A(1 - e^{t/\tau})$$

, where I(t) is the intensity, t is time, A is the mobile fraction which is rarely equal to 1 and τ is the recovery rate. Also the combination of exponents can be applied for the fit. The other one is the infinite boundary model which has the following dependency in the 2D case [201]:

$$I(t) = e^{-\tau/2t} I_0(-\tau/2t) + I_1(-\tau/2t)$$

The recovery curve fits to the infinite boundary model in case of diffusion limitation. In case of reaction limited recovery for interactions between components, are required k_{on} and k_{off} have to be taken into account [196]. In the simplest case it has the following dependency:

$$I(t) = A(1 - e^{-k_{off}t})$$

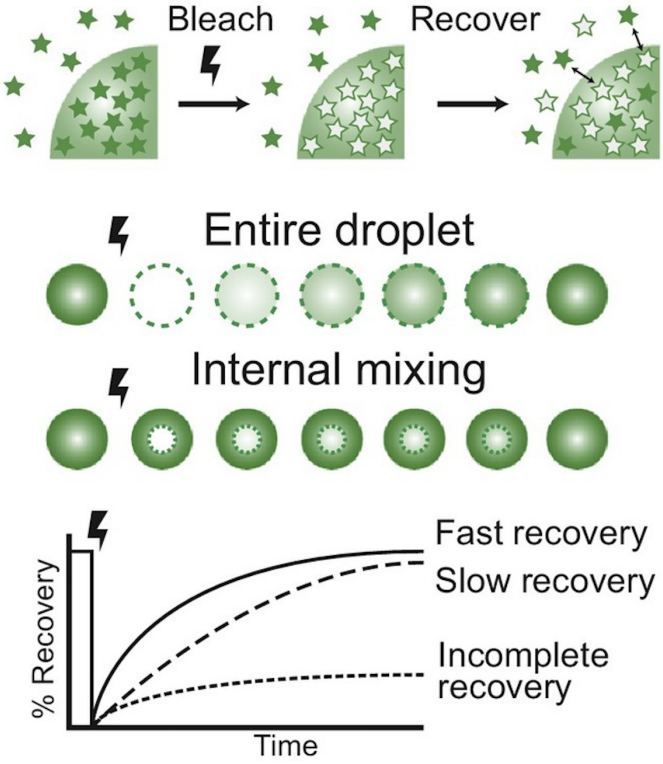


Figure 2.7 – Illustration of fluorescence recovery after photobleaching principle (top panel). Protein exchange between phases can be studied by bleaching an entire droplet, protein diffusion inside droplet can be measured by bleaching a small region inside droplet (middle panel). Fluorescence recovery profile representing fast (solid line) and slow diffusion cases (long dashed line), the case with a small mobile fraction is illustrated with short dashed line [68].

2.6.5 Atomic force microscopy (AFM)

AFM imaging can be also applied to phase separating systems. The advantage of this method is that it allows to work at nanomolar concentrations, however, if higher concentrations are required, the background signal will be too high. An example of AFM application to study FUS protein phase separation around DNA damage sites was published recently, allowed determination of DNA and protein molecules and even the DNA damage sites [193].

2.6.6 Electron microscopy

Another destructive method which can be used is Cryo electron microscopy what allows imaging of droplets less then 100 nm in diameter [125]. Next, Cryo-electron tomography allows determination of ultrastructural organization inside droplets, which was done on a tripartite droplets with two proteins and lipid vesicles [218].

2.6.7 Fluorescence techniques

Steady-state and time-resolved fluorescence anisotropy theory were introduced above in application to study IDPs. In the condensed phase, anisotropy is expected to increase as the mobility is dampened due to the increase in viscosity, however, it is also dependent on protein conformation and the mobility of the region where the dye is attached to. Time-resolved anisotropy decay usually has a biexponentional dependency: The fast one corresponds to the rotation of the dye, the slow component is associated with protein. In case of IDP, it depends on backbone segmental

torsional mobility what depends on the compactness of the chain [129].

The affinity between proteins and nucleic acids in droplets can be determined using smFRET. Measuring time-dependent distance changes between the dyes reveals the dynamics of the interaction, which can be compared with the dilute state [54].

Fluorescence correlation spectroscopy (FCS) measurements can be used to infer molecular diffusion coefficients, droplet nanorheology via probe-size-dependent permeability. For this, several radius particles and their hydrodynamic radius can be used to calculate an apparent viscosity (Fig. 2.8) [212]. However, microrheology can be also studied using confocal microscope by tracking

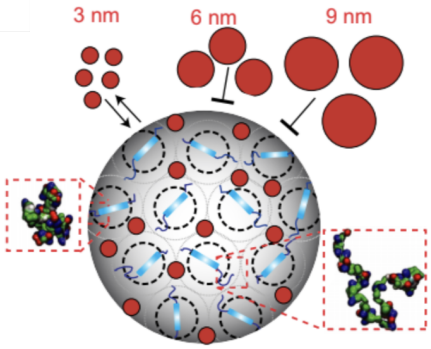


Figure 2.8 – Schematic illustration of the droplets probe-size-dependent permeability [212].

bead diffusion in the droplet [54].

Next, using FCS protein concentration in a dilute and condense phases allows to determine a full coexistence curve (binodals) in a phase diagram [132]. Finally, with fluorescence cross-correlation spectroscopy (FCCS) complex formation and interactions between the components in droplets can be studied [92].

Using fluorescent labelled proteins, the solvent accessibility can be studied by quenching of fluorescein with potassium iodide (KI) measuring fluorescence lifetimes as a function of KI and concentration [129].

2.6.8 SAXS, SANS and DLS

Comparing SAXS [67] and SANS [140] measurements carried out in a dilute and condense phase allows to reveal long-range conformational changes in the protein which occur during the transition between phases. From R_g the degree of chain extension can be determined. Similarly, the hydrodynamic diameter measured by DLS can be used [67].

2.7 NMR spectroscopy to study phase separation

Both, solution- and solid-state NMR can be applied to study LLPS, however, here only solution-state is going to be discussed. The theoretical introduction to NMR and methods for IDP characterisation were presented in section 2. In this section practical aspects to study phase separating systems are reviewed.

2.7.1 Sample preparation for NMR measurements

Dispersed phase

Using just the dilute phase for the measurements already provides useful information for phase separation characterisation. Individual proteins and interaction between them can be studied. There are many examples of phase separating proteins for which chemical shifts do not change much between phases, so the initial assignment can be done in a disperse phase and adapted for the condense phase spectra [21, 181]. Next, NMR measurements can be performed on an over-saturated sample which passed the region of phase separation, however, conformational transitions and interactions are hypothesised to happen under such conditions [214].

Colloid sample

In such samples two phases coexist. As the condense phase has a higher density, it sediments overtime to the bottom, which restricts the measurements to a short time period. In general, the signals coming from the dilute phase molecules are much more intense due to the faster tumbling, which masks the signal from the condensed phase [204]. However, successful examples where the signals from both phases are present on HSQC spectra exist. It should be mentioned, that chemical shifts of the protein in two phases were different and the signal from the condense phase was significantly weaker. The signal coming from the dilute phase can be removed by applying a pulse field gradient diffusion filter. The basics of this method were described above in section 2.1.4 to study protein diffusion. This removes the signal from molecules which diffuses faster than 10^{-7} cm^2 s^{-1} [172].

Single droplet sample

Filling the NMR tube sensitive volume with only condensed phase requires a lot of sample, as the protein concentration is estimated to be around 400 mg/ml. 3 mm diameter tubes require 130 μl, using Shigemi 3 mm allows to decrease volume, 1 mm diameter tubes enables the use of 20 μl samples. As the phase is highly viscous additional complications arise with sample transfer into the NMR capillary, but different protocols exist [21, 26, 146, 181].

2.7.2 Conformational changes induced by LLPS

In the condensed phase broadening in ^{1}H dimension causes peaks overlap and complicates determination of separate peaks. First, D_2O can be placed in a separate capillary tube to avoid contributions of solvent exchange [9]. Also additional improvment can be achieved using apodisation function for line sharpening [26]. In some cases, pH can be decreased or ^{13}C-direct detected experiments can be applied to avoid line broadening [111].

Protein conformation in the condensed phase is interesting in the context of transitions which occur during phase separation, the driving forces and the physics of the process. So, a comparison between protein conformation and dynamics in dilute and condensed phase is particularly important.

To begin, secondary structure changes can be examined by comparing secondary chemical shifts which can be compared if protein spectra in both phases are assigned [38].

Long-range conformational changes and contact within a single molecule can be examined by intramolecular PRE, which is sensitive to the transient weak contacts up to 25 Å[141]; [181]. In combination with ASTEROIDS approach, the information about protein conformations space changes during the transition between phases can be investigated.

Finally, protein dynamics also indicates structural changes in the protein by highlighting the regions which are becoming more rigid. Backbone dynamics in a pico-to-nanosecond timescales can be examined by R_1, R_2, and heteronuclear ^{1}H-^{15}N NOE. In general, due to increased viscosity

and protein concentrations spin relaxation increase all along the protein chain [21, 172, 181]. Additional contributions can be associated with an increase in a structural propensity in some regions [146]. And of course conformational exchange (R_{ex}) can contribute to R_2 (presented below in section 2.2.3). Chain dynamics in condensed phase can be further analysed to reveal the influence of a crowded environment which is discussed later in this chapter.

2.7.3 Protein interactions and transient contacts in condence phase

One of the driving forces which mediates LLPS is interaction between chains. NMR provides multiple techniques which can be applied to study interactions at different strengths and exchange regimes which were successfully applied to study phase separation.

Intermolecular PRE provides information about dynamic interactions, however, the interpretation can be challenging due to the variability in partitioning of the PRE labeled protein and the extent of labeling [141, 181].

NOESY can reveal residues involved in contacts mediating phase separation. Contacts up to 6 Å can be determined on NOESY spectra. To differentiate between intra- and intermolecular contacts, differential isotopic labeling schemes can be combined with heteronuclear filtering (^{13}C, ^{15}N). For example, using ^{13}C edited ^{1}H-^{1}H NOESY alanine, valine, glycine and proline type residues were shown to be involved in contacts during phase separation on elastin-like polypeptides [172]. For Ddx4 phenylalanine and arginine were shown to be most prone to form contacts [21].

Increased R_2 spin relaxation can indicate an chemical exchange associated with particular residue or linear region. For instance, again multiple regions of enhanced relaxation centered at aromatic residues were found for hnRNPA1 which are important for phase separation [132].

2.8 Sample preparation and experimental details

2.8.1 Protein expression and purification

MeV P and N constructs were cloned between the Nde I and Xho I into pET41c(+) finishing with a C-terminal 8His-tag [137]. All proteins (if different is not stated) were expressed in Escherichia coli Rosetta TM(lDE3)/pRARE (Novagen) with incubation overnight at 20° after the induction with with 1 mM isopropyl-β-D-thiogalactopyranoside at an optical density of $I_{OD} = 0.6$. $N_{1\text{-}525}$ was co-transformed with $P_{1–507}$ and purified in complex. Measles $N_{401\text{-}525}$ (N_{tail}) gene (strain Edmonston B) was cloned into the expression vector pET22b with a N-terminal 8His-tag and transformed into Escherichia coli BL21 (Novagen) for expression. After cells were centrifuged at 5000 rpm for 20 min, the pellets were resuspended in the lysis buffer (20 mM Tris (pH 8.0),150 mM NaCl or 500 mM NaCl for XD containing constructs), sonicated and centrifuged again at 18000 rpm, 50 min. The supernatant was subjected to a Ni-purification. The protein was eluted with 20 mM Tris (pH 8), 150/500 mM NaCl, and 500 mM imidazole. TEV cleavage was performed overnight at 4° C in the presence of 2 mM DTT. The sample was concentrated and subjected to size exclusion chromatography (SEC, Superdex 75/200) in NMR buffer (50 mM Na-phosphate (pH 6.0), 150/500 mM NaCl, and 2 mM dithiothreitol (DTT)). The protocols of $P_{50}N_{525}$, $P_{50}N_{405}$, $P_{300}N_{525}$, and $P_{300}N_{405}$ purification are described in more details in (13, 14). Mutation of S491L in N and of ^{409}VLK^{411}->AAA in P were inserted by site-directed mutagenesis, and $P_{50}N_{525}$ (S491L), $N_{460–525}$ (S491L), $P_{304–507}$ (^{409}VLK$^{411}\rightarrow$AAA), and $P_{377–507}$ (^{409}VLK$^{411}\rightarrow$AAA) were produced and purified in the same way as wild-type proteins.

Expression of unlabeled protein was performed in LB medium. Proteins labeled for NMR (^{15}N and ^{13}C) were expressed under the same conditions in M9 minimal medium.

2.8.2 Fluorescent and PRE labelling of protein samples

Cysteine mutations for fluorescence labeling were performed by site-directed mutagenesis into the constructs where previously all native cysteines were exchanged to alanines. Mutant proteins were expressed and purified in the same way as wild type and concentrated to around 100 μM. The proteins were then labeled using maleimide chemistry [124]. Briefly, the protein was incubated overnight with 10 mM DTT, then dialyzed against 50 mM Na-phosphate buffer (pH 7.0) and 150/500 mM NaCl in two steps, and mixed with 5 times excess of maleimide dye (fluorescein, Alexa Fluor 594 or TEMPO for PRE) solubilized in dimethyl sulfoxide. After 2 hours of incubation at room temperature, the labeling reaction was stopped by adding 5 mM DTT, and proteins were separated from the excess dye by SEC (Superdex 75/200) in NMR buffer. Only second step was used in case of TEMPO labelling.

2.8.3 Protein phosphorylation

P_{304}, $P_{304}N_{405}$ and $P_{304}N_{525}$ were phosphorylated with CK2 kinase (Merck, Germany) in the presence of 100 ATP and 10 mM $MgCl_2$. Sample was incubated at least 24 hours before purification on SEC column (S75 or S200), if different is not stated.

2.8.4 RNA samples

RNA samples were purchased from Integrated DNA Technologies (Leuven, Belgium), they were terminated either with OH at both 3′ and 5′ or with fluorescein amidite at the 3′ end.

2.8.5 Protein phase separation and turbidity measurements

First P and N constructs were mixed and then diluted with a buffer, so that the final NaCl concentration was 150 mM (unless different is stated). Before the turbidity measurements samples were incubated at room temperature for 5 min, and the absorbance at 395 nm was collected three times on a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA). An absorbance of 0.1 was used as a turbidity threshold indicating the presence of phase separation in the sample.

Concentration measurement

N_{tail} concentration determination in condense solution at different NaCl concentration and temperature was performed as following: Protein was mixed with PEG solution with various NaCl concentrations, so that final $N_{tail}$was 300 μM, PEG 200 mg/ml what triggered phase separation, sedimented the droplet down over night in PCR machine at given temperature. Next, volumes of both phases were measured and then protein concentration in the presence of guanidinium chloride was determined by the absorbance at 280 nm. Condense phase was diluted 40 times before the measurement, dilute phase twice. The protocol described in [21] was used, density of 1 was used to calculate void volume portion of protein.

2.8.6 Phase diagram fitting with the Flory-Haggins model

Based on protein concentration measurements of a dilute and condense phases, phase diagrams of coexistence curves (temperature vs. concentration) were plotted for several salt concentrations (118 mM, 181 mM, 244 mM, 307 mM and 377 mM NaCl). Measured concentrations were transferred into volume fraction using density of 1. Next, phase diagram was fitted with:

$$T_s = \frac{B}{0.5(1/N_A\phi) + 1/(N_B(1-\phi))) - A} \tag{2.8}$$

, which A and B are fitted parameters, $N_A = 1$ representing solvent, $N_B = 130$ representing N_{tail}, ϕ is the volume fraction. Curve for each salt concentration was fitted independently, χ is determined for each temperature using:

$$\chi = A + \frac{B}{T}$$

2.8.7 Light and fluorescence microscopy imaging

Differential interference contrast (DIC) or epifluorescence was used for light imaging on an Olympus IX81 inverted microscope. Fluorescence signal was acquired using a 485(20)-nm (for fluorescein) or a 560(25)-nm (for Alexa Fluor 594) excitation filter combined with a 525(30)-nm or a 607(36)-nm emission filter, respectively. Excitation light (Metal-Halide X-Cite 120) and emission wavelengths are split using a 4X4M-B quadriband dichroic mirror (Semrock). Oil immersion objectives 60x (Olympus PlanAPON60, 1.42 numerical aperture (NA)) or 100x (Olympus UPLFLN, 1.3 NA) were used for imaging, and a Hamamatsu Orca Flash4.0-V2 sCMOS (scientific Complementary Metal-Oxyde Semiconductor) 2048 x 2048 detection camera was used. The acquisition and further data analysis and processing were performed with Volocity (Quorum Technologies) software.

2.8.8 Fluorescence recovery after photobleaching (FRAP)

FRAP experiments were measured on an Olympus IX81 spinning-disk confocal microscope equipped with a Yokogawa CSUX1 spinning disk head, diode-pumped solid-state lasers 488 nm (50 mW Coherent Sapphire) and 561 nm (100 mW Cobolt JIVE4). Emission light is filtered through a 520(28) nm filter for fluorescein. Acquisition is performed with an iXon Ultra EMCCD (Electron Multiplying Charge-Coupled Device) (Andor) 512 x 512 pixel detection camera. Bleaching was restricted to a region of interest, defined graphically, with a galvanometric FRAP/PA device (Andor) using the same lasers used for imaging with AOTF (Acousto-Optic Tunable Filter) 15% and with a dwell time of 50 ms. 10 images are taken before bleaching and 40 repeats after with a 2 s interval. The laser power and the dwell time during bleaching were optimized to bleach 50% of the initial fluorescence intensity. FRAP profiles were extracted and analyzed using FIJI and python plugin ImageFRAP (https://imagej.net/Analyze_FRAP_movies_with_a_Jython_script).

2.8.9 Steady-state fluorescence anisotropy

NC assembly was followed by fluorescence anisotropy as described in [136] using various concentration of protein (tested constructs of N are described in Chapter 4) and 0.5 μM 10-base RNA-FAM (polyA or polyU) which was added prior to the kinetic acquisition. RNA fluorescence kinetics were recorded using excitation light at 470 nm and emission was recorded at the wavelength of 520 nm. Parallel and perpendicular polarization directions were recorded alternatively and used pairwise to calculate the fluorescence anisotropy.

2.8.10 Time-resolved fluorescence anisotropy

Time resolved fluorescence anisotropy decays were measured using a Mini-Tau fluorescence lifetime spectrometer (Edinburgh Instruments, UK) equipped with a 450 nm picosecond pulsed diode laser and a TCC2 electronics module for time-correlated single photon counting. Photons detection was used with high-speed photon multiplier tube (Hamamatsu H10720-01). The sample was excited with vertically polarized light, parallel and perpendicular polarized fluorescence light was detected during 120 s sequentially. Photon arrival times histograms were calculated from each measurement, and anisotropy decays were calculated as described above in this section. The G factor was measured for each buffer condition using the free dye.

2.8.11 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry for protein was measured with iMicroCal iTC200 at 298 K. The titration experiment was performed by sequential addition of $2\mu l$ aliquots of 500 μM P_{XD} into the microcalorimeter cell filled with 50 μM N_{tail}. In total 20 injections of XD with 180 sec intervals in between were done through titration; the reaction mixture in the microcalorimeter cell was continuously stirred at 750 rpm. The experimental titration curve was fitted to the 1:1 binding model using the Origin version 7.0 software from MicroCal.

2.8.12 Negative-stain electron microscopy

For NC structures, imaging with negative-stain electron microscopy 50 μM RNA (HO-A6-OH) was added to 20 μM of protein ($P_{50}N_{525}$, $P_{304}N_{525}$, $P_{50}N_{405}$, $P_{304}N_{405}$ phosphorylated or non-phosphorylated) or to droplets formed by mixing 20 μM $P_{50}N_{525}$ and 80 μM $P_{304–507}$ in a final volume of 20 μl. After the incubation for 1 hour at 37° C, 2 μl of the sample were applied to the clean side of carbon on mica (carbon/mica interface), and stained with 2% sodium silicotungstate (pH 7.0). Images were taken with a T12 FEI microscope at 120 kV and 30,000 magnification.

NMR spectroscopy used in Chapters 3 and 4

2.8.13 Assignment

All experiments were acquired in NMR buffer (150 or 500 mM NaCl) at 25° C. The spectral assignment of $P_{50}N_{525}$ at pH 6.0 was adapted from the previously published assignment at pH 7.0 (13) using ^{13}C, ^{15}N-labeled samples using a set of BEST-TROSY triple resonance experiments.

The spectral assignments of ^{13}C- and ^{15}N-labeled $P_{377–457}$ and $P_{457–507}$ constructs were obtained using sets of BEST-TROSY triple resonance experiments correlating C^{α}, C^{β}, and C_O resonances at 600 MHz. Phosphorylated residues of P_{304} were assigned using triple resonance experiments. Spectra were processed with NMRPipe [45], and automatic assignment with further manual verification was performed with the program MARS [107]. Secondary chemical shifts (scs) for C^{α}, C^{β}, CO, N and HN were calculated using the random coil values from refDB [225].

2.8.14 Ensemble analysis of P_{loop}

The genetic algorithm ASTEROIDS selected representative ensembles using experimental ^{1}H, ^{15}N, and ^{13}C(C_{α}, C_{β} and CO) backbone chemical shifts. A initial 10000-member ensemble of P_{loop} conformations was generated using the statistical coil model flexible-meccano [152], next, 200-conformer ensembles matching experimental backbone chemical shifts were selected using the ASTEROIDS genetic algorithm as previously described [148]; [105]. In total, four iterations were used to generate the final representative ensemble.

2.8.15 Determination of apparent Kd values from NMR experiments for the interaction between N and P_{loop}/P_{XD}

^{15}N R_1 relaxation rates were obtained by sampling the decay of magnetization with seven delays between 0 and 1.71 s. $R_{1\rho}$ relaxation rates were measured using seven delay times between 0.001 and 0.24 s. The spin-lock field was 1500 Hz, and R was calculated from R_1 and $R_{1\rho}$, as described above. The residue-specific apparent K_d values for $P_{377–457}$ were obtained for each residue from ^{15}N relaxation as described in section 2.2. $P_{377–507}$ concentration remained constant (100 μM), and N concentration was sampled from 0 to 140 μM. The rotational correlation time of the complex was assumed to be equal to τ_c of $P_{50}N_{525}$ alone and was estimated from its molecular weight. ^{15}N relaxation dispersion experiments was measured for ^{15}N labelled $P_{457–507}$ of 100 μM in the presence of 70 and 30 μM unlabeled N_{tail} at 700 and 850 MHz using 14 points

at CPMG frequencies between 31 and 1000 Hz with a constant-time relaxation delay of 32 ms [86]. CPMG profiles for nine residues and two fields were fitted simultaneously using ChemEx (https://github.com/gbouvignies/chemex) and a two-state exchange model. The bound population of $P_{457–507}$ to $N_{460–525}$ (p_B) and the exchange rate (k_{ex}) were calculated from the fit, and the K_d was calculated assuming a 1:1 stoichiometry. The error for the K_d calculation was estimated by the error propagation from the fitting error of p_B and assuming 5% error in the concentration determination of both proteins.

Relaxation experiments ($R_{1\rho}$, R_2, ^{1}H-^{15}N, hetNOE) of P_{304} and $P_{304}N_{525}$ phosphorylated and non-phosphorylated were measured using 100 μM concentration.

2.8.16 NMR measurements of phase-separating samples

^{15}N-labeled $P_{50}N_{525}$ and $P_{304–507}$ were mixed at several ratios and then loaded in a 3-mm NMR tube. ^{1}H-^{15}N HSQCs and $R_{1\rho}$ experiments were measured as described above. Pulsed field gradient NMR diffusion experiments were measured using ^{15}N-edited stimulated-echo diffusion experiments using a 300-ms diffusion delay with 16 or 20 z-gradient strengths ranging linearly from 0 to 0.55 T m^1 [210]. ^{1}H-^{15}N HSQCs were measured using both $P_{50}N_{525}$ and $P_{304–507}$ ^{15}N-labeled samples at a constant $P_{304–507}$ concentration of 75 μM with $P_{50}N_{525}$ 5, 10, 20, 50, and 100%. The concentration of $P_{50}N_{525}$ and $P_{304–507}$ in the dilute and condensed phase was estimated on the basis of intensities of 23 for $P_{50}N_{525}$ and 9 for $P_{304–507}$ non-overlapping and non-interacting residues. Their intensities were summed and normalized by the intensities in the spectra of the individual proteins. The amount of each protein in condensed phase was then estimated by subtraction. That allows to estimate the concentration of droplet-associated N and P in the sensitive NMR volume. The ratio between N and P in the condense phase was estimated as a function of total the concentration of N in the sample.

2.8.17 Kinetic measurements by NMR

Kinetics of $P_{50}N_{525}$ (with and without $P_{304–507}$ or $P_{377–507}$) nucleocapsid assembly were followed using NMR spectroscopy using the procedure introduced previously [136]. Time resolution was approximately 8 min (time of HSQC spectra measurement). The assembly process was triggered by adding six-nucleotide RNA (OH-A_6-OH), and the kinetics were followed by measuring intensities of resonances from $P_{1–50}$ on SOFAST HMQC spectra. Assembly rates were calculated by averaging intensities of 18 peaks and fitting the profile with a biexponential function. Mean assembly rates were calculated as follows: k1*a1 + k2*a2, where k1 and k2 are time rates, a1 and a2 are the amplitudes.

Phosphorylation kinetics were measured on P_{304} of 100 μM concentration by measuring SOFAST HMQC spectra after addition of 8 ng/μl CK2, 10 mM $MgCl_2$ and 100 μM ATP. The intensity decay of non-phosphorylated peaks and the increase on newly appeared peaks were followed and fitted together for each residue separately in order to determine phosphorylation rates.

Kinetics of $P_{304}N_{405}$ phosphorylation followed by assembly was measured in the presence of 8 ng/μl CK2, 10 mM $MgCl_2$, 100 μM ATP and 2mM six-base polyA RNA. Phosphorylation and assembly rates were followed as described above.

NMR spectroscopy used in Chapter 5

2.8.18 Supercondense NMR sample preparation

Mixture of ^{15}N-^{13}C N_{tail} with non-labelled were used for sample preparation (from 100% to 1% labelled). Protein sample of around 2 mM concentration was mixed 1:1 with PEG solution of 400 mg/ml in NMR buffer and centrifuged at 1000 rpm, 25° for at least 3 hours. Condense phase

formed on the bottom of the eppendorf was transferred to a 3 mm Sigemi tube using a glass pipet and hand centrifuge.

NMR measurements

All experiments were acquired in NMR buffer at 25°, if different is not stated, on Bruker cryoprobe spectrometers at ^{1}H frequencies of 600, 700, 850, or 950 MHz using TopSpin software. All spectra were processed and analysed with NMRPipe [45] and NMRFAM-sparky [123], 1D spectra of water longitudinal relaxation were analysed using Python library nmrglue [91].

2.8.19 Assignment

Backbone assignment of N_{tail} in super-condense phase was verified using a ^{15}N-^{13}C labelled sample. For this, a set of BEST triple resonance experiments were acquired (HNCO, HN-CACO, HNCA and HNCOCA). The comparison of secondary chemical shifts between dilute and condense phase protein were done using random coil values from refDB [225].

2.8.20 ^{15}N relaxation measurement

Four ^{15}N relaxation experiments were aquired: $R_{1\rho}$, R_1, ^{1}H-^{15}N heteronuclear NOE and transverse cross-correlated relaxation [157]. $R_{1\rho}$ and R_1 were measured with ten delays, the exact range was optimised depending on the viscosity of the sample. R_2 was calculated using $R_{1\rho}$ and R_1 measurements, a spin lock 1500 Hz, resonance offset was calculated for each residue using its ^{15}N chemical shift.

Pulsed field gradient (PFG)

PFG NMR diffusion experiments were measured using ^{15}N-edited stimulated-echo diffusion experiments using a 300-ms diffusion delay with 16 or 20 z-gradient strengths ranging linearly from 0 to 0.55 T m^{1} [210].

2.8.21 Water relaxation measurement

Microviscosity was measured using R_1 of water with 20 delays. Water signal was integrated and the extracted intensity dependency on the delay was fitted with a single exponent using Python 3.

2.8.22 Model-free analysis

^{15}N relaxation rates extracted for each residue of N_{tail}, data for two fields (850 and 600 MHz) and three peg concentrations (0, 37.5 and 75 mg/ml) were fitted together using model-free analysis:

$$R_1 = \frac{d^2}{4}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 j(\omega_N)$$

$$R_2 = \frac{d^2}{8}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + {}_N)] + \frac{c^2}{6}[4J(0) + 3J(\omega_N)]$$

$$NOE = 1 + \frac{\gamma_H}{\gamma_N}\frac{\sigma_{NH}}{R_1}$$

$$\sigma_{NH} = 1 + \frac{\gamma_H}{\gamma_N} d_{NH}^2 (12J(\omega_H + \omega_N) - 2J(\omega_H - \omega_N)$$

$$\eta_{xy} = \frac{1}{15} P_2(cos\Theta(\frac{\mu_0 h \gamma_H \gamma_N}{4\pi r_{NH}^3})(\sigma_{||} - \sigma\perp)\omega_N(4J(0) + 3J(\omega_N))$$

where $d^2 = (\mu_0 h \gamma_H \gamma_N / 8\pi^2 r_{NH}^3)$ and $c = \Delta\sigma\omega_N/\sqrt{3}$. γ_H and γ_N are the gyromagnetic ratios of H and N, h is a Planck's constant, r_{NH} is the length of the bond, $\Delta\sigma_{NH}$ is CSA of the spin, μ_0 is the permeability of free space. An expression introduced in [4] for rotational correlation time dependency on the viscosity was implicated:

$$\tau_k(C) = \tau_{k,\infty}(\varepsilon_k \rho(C) + 1)$$

where $\tau_{k,\infty}$ is the correlation time at infinite dilution, ρ is the viscosity at given peg concentration (C), ε is a friction coefficient. In total seven parameters were fitted with an in-house script for each residue: S^2, S_f^2, τ_i, τ_s, ε_f, ε_i and ε_f. τ_f was fixed to 0.045 ns.

Droplet sample was fitted independently using data at to fields (600 and 850 MHz) with four parameters: S^2, S_f^2, τ_i, τ_s. τ_f was fixed to 0.1 ns.

2.8.23 PRE

PRE effects were measured using ^{1}H-^{15}N HSQC of 300 μM ^{15}N-N_{tail} labelled with TEMPO either alone or in supercondense phase made with non-labelled N_{tail}, so that 1% of the sample was labelled in order to eliminate inter-residual interactions. Reference diamagnetic sample that was incubated with 3 mM of ascorbic acid before acquisition. For the super-condense sample a buffer containing ascorbic acid was added on top, the total concentration for the entire volume was 3 mM and incubated over night. Peak intensities were extracted from the HSQCs using NMRFAM-sparky [123] and their ratios between paramagnetic and diamagnetic samples were calculated.

2.8.24 Generation of N_{tail} boxes at different concentrations

Ensembles of Ntail conformations were generated using a flexible meccano [152] modified script which generates backbone chains by randomly sampling amino acid-specific backbone dihedral angle potential wells and using a self-avoiding walk. Four boxes were generated: infinite dilution (separate Ntail chains), 15 mM, 27.5 mM and 40 mM. Boxes of 50x50x50 nm dimentions were generated, the starting points of Ntail chains generation with flexible meccano were placed on a lattice in a box (shifts in xyz by 4.42, 3.81 and 3.19 nm for 15 mM, 27.5 mM and 40 mM cases). Next, side chains are added to the complete chain with pulchra [178] and missing protons with reduce [219]. Coordinates of atoms are added to the "avoid list" which is used during the generation of the following chains in order to avoid clashes. R_g and end-to-end distances are calculated for each chain. Dihedral angles of the generated chains are calculated using torsion library from Bioplib [163] and visualised with Python.

3

Measles Virus Nucleo- and Phosphoproteins form Liquid-like Phase-Separated Compartments that Promote Nucleocapsid Assembly

In this chapter measles viral factories formed by nucleo- and phosphoproteins are discussed. It starts with the article where biophysical characteristics of viral factories reproduced in vitro were studied (section 3.2). A minimum scaffold system was found by screening a panel of different mutants. A disordered region of P called P_{loop} was characterised using NMR and a new interaction with N was found and its role in phase separation is proposed in section 3.4. Next, an interaction between N and P essential for phase separation was discovered. Using NMR spectroscopy protein distribution between two phases was measured as a function of N:P total ratio. Based on these experimental data a thermodynamic model of N and P phase separation is proposed in section 3.3. Next, the influence of phase separation on nucleotide assembly was shown using a combination of fluorescence microscopy, NMR spectroscopy and negative stain electron microscopy. Supplementary materials for the article can be found in Annexe 1. In addition, preliminary results on the involvement of cellular client proteins in viral factories are shown (section 3.5). And MeV droplets formed by full length of N and P are discussed in section 3.6.

3.1 Introduction

Viral replication machinery has been shown to form bio-condensates in host cell cytoplasm (for example in Rabies virus), all essential components required for replication co- localise: N, P, L, nucleic acids. During the course of the work I carried out *in vitro* during this thesis, similar observations were made *in vivo* for measles virus. Phase separation of viral proteins have never been studied *in vitro* previously. MeV N and P both contains long disordered regions and several interactions between them are known. In addition, P has a tetrameric structure that further increases the multivalency between proteins. All these suggests N and P as possible scaffolds for MeV phase separation. In section 3.2 an article which characterises viral factories is presented.

Further analysis of droplets is still essential. Models capable of describing the behaviour of droplets can help to understand the mechanism of viral phase separation and reveal interactions crucial for its formation and regulation.

Small molecule drugs are shown to partition into biocondensates that has broad implications for drug discovery. Many anticancer drugs were shown to concentrate in membraneless organelles [112]. This discovery makes it impossible to ignore phase separation in drug design from now on. On the opposite, we should learn how to profit from this. The field of antiviral drug design could be revolutionized if viral factory disassembly could be used as a target.

3.2 Article

Supporting information to the article is presented in Annex 1.

VIROLOGY

Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly

Serafima Guseva*, Sigrid Milles*, Malene Ringkjøbing Jensen, Nicola Salvi, Jean-Philippe Kleman, Damien Maurin, Rob W. H. Ruigrok†, Martin Blackledge†

Many viruses are known to form cellular compartments, also called viral factories. Paramyxoviruses, including measles virus, colocalize their proteomic and genomic material in puncta in infected cells. We demonstrate that purified nucleoproteins (N) and phosphoproteins (P) of measles virus form liquid-like membraneless organelles upon mixing in vitro. We identify weak interactions involving intrinsically disordered domains of N and P that are implicated in this process, one of which is essential for phase separation. Fluorescence allows us to follow the modulation of the dynamics of N and P upon droplet formation, while NMR is used to investigate the thermodynamics of this process. RNA colocalizes to droplets, where it triggers assembly of N protomers into nucleocapsid-like particles that encapsidate the RNA. The rate of encapsidation within droplets is enhanced compared to the dilute phase, revealing one of the roles of liquid-liquid phase separation in measles virus replication.

INTRODUCTION

Measles virus (MeV) is a highly contagious virus that is responsible for more than 100,000 deaths every year. No treatment currently exists for MeV infection, and the impact of MeV throughout the world has recently been amplified by weakening levels of vaccine coverage (*1*). MeV is a nonsegmented, single-stranded negative-sense RNA virus (Mononegavirales) of the family Paramyxoviridae, whose members carry their own replication machinery comprising nucleoprotein (N), polymerase (L), and phosphoprotein (P) components. N binds to the viral genome, assembling into helical nucleocapsids, while P chaperones N before encapsidation and is thought to bring N and L in close spatial proximity during replication and transcription (*2*, *3*). The replication process represents a promising but hitherto relatively unexplored therapeutic target.

N and P both comprise domains exhibiting high levels of intrinsic disorder, with respectively 30 and 75% of their primary sequences lacking tertiary structure. N has a 125–amino acid disordered C-terminal domain (N_{TAIL}) (*4*) and a flexible 30–amino acid N-terminal arm (N_{ARM}), flanking the RNA binding domain N_{CORE}, while P comprises disordered domains of 304 amino acids (P_{TAIL}) (*5*) and 80 amino acids (P_{LOOP}), on either side of a coiled-coil tetrameric domain (*6*), and is capped by a C-terminal three-helix bundle (XD) (Fig. 1A) (*7*). N and P exhibit numerous interactions that are essential during different stages of the viral cycle: Two distant linear motifs in P_{TAIL}, $P_{1–50}$ and $P_{186–198}$, participate in chaperoning the monomeric form of N (*5*, *8*), preventing binding to cellular RNA. In an N^0P heterodimer, $P_{1–50}$ binds to the same surface of N as N_{ARM} of the succeeding protomer in assembled nucleocapsids (*9*), so that the N^0P complex efficiently inhibits assembly. XD is known to bind to N_{TAIL} (*10*, *11*), an interaction thought to be important for colocalizing N and L during the replication process and was recently shown to interact with L (*12*). The process of nucleocapsid assembly on viral RNA sequences was recently studied in real time, using nuclear magnetic resonance (NMR) and fluorescence (*13*), highlighting the role of N_{ARM} and $P_{1–50}$, and leading to high-resolution structures revealing N-RNA interactions in assembled nucleocapsids (*14*).

In recent years, liquid-liquid phase separation (LLPS), driven by weak interactions often involving intrinsically disordered proteins or RNA, has been shown to offer a highly efficient means to spatially and temporally control a remarkable array of cellular processes (*15*). The physicochemical properties of these microenvironments are thought to provide optimal conditions to enhance specific molecular interactions within the condensate, while maintaining immiscibility with respect to other excluded molecules (*16*–*18*). A number of viruses have been shown to form cytoplasmic compartments in infected cells, known as "viral factories," that have been associated with the viral replication process (*19*, *20*). Colocalization of N and P in vivo is a characteristic feature of mononegaviruses such as rabies (*21*), vesicular stomatitis virus (*22*), respiratory syncytial virus (*23*), human parainfluenza types 3 (*24*) and 5 (*25*), human metapneumovirus (*26*), Nipah (*27*), and MeV (*28*, *29*). Inclusion bodies from MeV have been shown to colocalize host proteins such as WDR5 (WD repeat-containing protein 5) (*28*) and other viral proteins [the polymerase L as well as C that is involved in controlling host immune response (*30*)]. Recent confirmation of the liquid-like nature of Negri bodies formed in rabies (*31*) and colocalization of N and P in infection-induced cytoplasmic inclusion bodies in MeV (*32*) suggests that LLPS represents a widespread mechanism exploited by mononegaviruses during infection.

Here, we demonstrate that mixing MeV N and P proteins in vitro triggers LLPS, allowing us to characterize the physical nature of the droplets, including the modulation of the dynamic behavior of N and P upon droplet formation, the molecular basis of the stabilizing interactions, and the thermodynamics of droplet formation. We are able to show that RNA molecules preferentially localize to the droplets and, once within the droplets, trigger assembly of nucleocapsid-like particles. The process of nucleocapsid assembly in LLPS is observed in real time using NMR spectroscopy. The rate of assembly is significantly enhanced compared to non–phase-separating conditions,

Institut de Biologie Structurale, Université Grenoble Alpes-CEA-CNRS, 71, Avenue des Martyrs, Grenoble, France.
*These authors contributed equally to this work.
†Corresponding author. Email: martin.blackledge@ibs.fr (M.B.); rob.ruigrok@ibs.fr (R.W.H.R.)

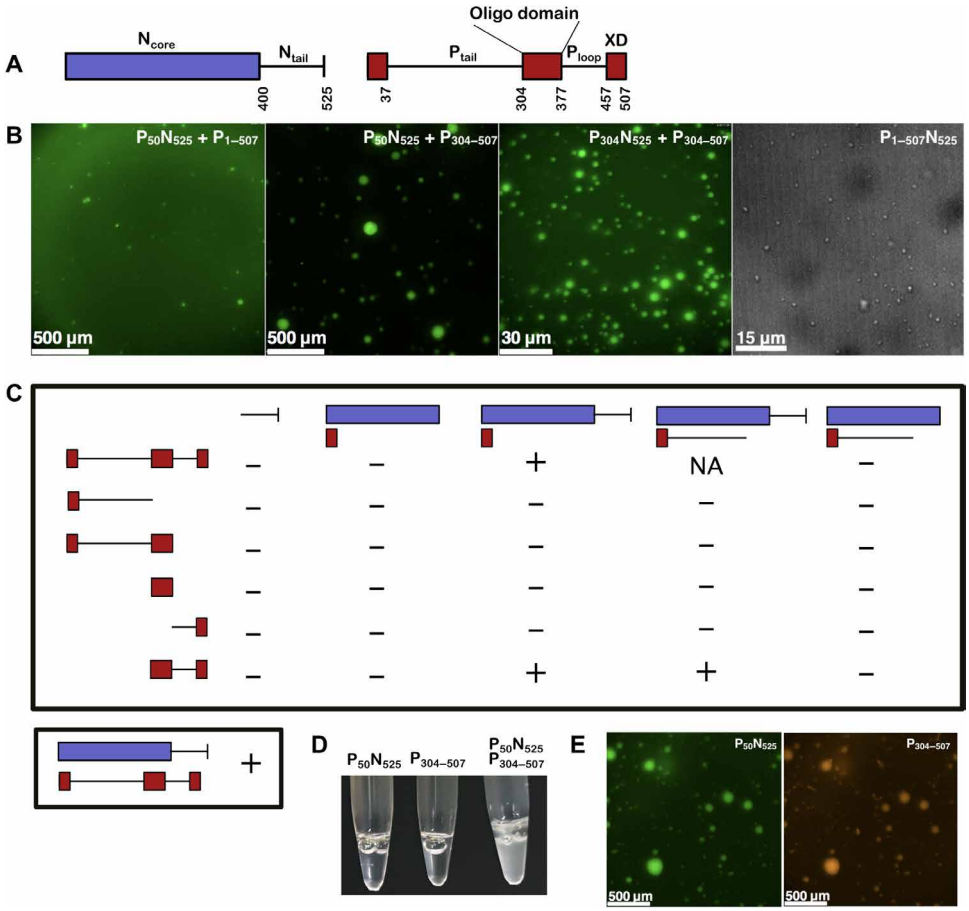


Fig. 1. Macroscale physical characterization of MeV droplets in vitro. (**A**) Schematic diagram of the structure of N and P; intrinsically disordered regions are presented as lines and folded domains are presented as boxes. (**B**) Fluorescence or DIC images of mixtures where phase separation events were observed: $P_{50}N_{525}$ (80 µM, 1% labeled) + $P_{1–507}$ (66 µM), $P_{50}N_{525}$ (80 µM, 1% labeled) + $P_{304–507}$ (200 µM), $P_{300}N_{525}$ + $P_{304–507}$ (290 µM), and $N_{1–525}$ + $P_{1–507}$ (coexpressed in *E. coli*). (**C**) The phase separation scaffold requires $P_{50}N_{525}$ and $P_{304–507}$ domains. Mixtures of different N and P truncation mutants were tested for the ability to trigger phase separation using DIC microscopy. "–" no phase separation events were observed; "+" phase separation events were observed; "NA" not tested. All mixtures were tested at several concentrations higher than 50 µM for each protein and several ratios including 1:1. Bottom panel represents coexpression of the full-length N and P proteins. (**D**) Picture of tubes with $P_{50}N_{525}$ and $P_{304–507}$. Sample with N and P mixed together becomes turbid. (**E**) Colocalization of $P_{50}N_{525}$-fluorescein (80 µM, 1% labeled) and $P_{304–507}$-Alexa Fluor 594 (200 µM, 0.5% labeled) droplets.

providing clear evidence that LLPS has the ability to enhance viral function and supporting previous suggestions that LLPS is exploited by negative-strand RNA viruses to create viral replication factories.

RESULTS

MeV N and P proteins phase separate upon mixing in vitro

Recombinant full-length P ($P_{1–507}$) and an RNA-free N^0P complex ($P_{50}N_{525}$) were separately expressed and mixed together and observed to form micron-sized spherical structures (Fig. 1B). Phase separation of $P_{1–507}$ and full-length N was also demonstrated by coexpressing both proteins using separate plasmids in *Escherichia coli*. The full N^0P complex was purified in a high-salt buffer to prevent phase separation (vide infra) and diluted into low salt, resulting in phase separation as observed from differential interference contrast (DIC) microscopy (Fig. 1B).

Identification of N and P domains that are required for phase separation

To identify which domains of MeV N and P are essential for phase separation, we tested a panel of different N and P constructs for their ability to trigger phase separation (Fig. 1, A to C, and fig. S1). Several mixtures formed condensed phases: $P_{50}N_{525}$ + $P_{1–507}$, $P_{304}N_{525}$ + $P_{304–507}$, and $P_{50}N_{525}$ + $P_{304–507}$ (Fig. 1B). $P_{50}N_{525}$ + $P_{304–507}$ was thus identified as the minimal phase-separating system, forming spherical bimolecular condensates as measured by fluorescence microscopy (Fig. 1B) and showing increased turbidity (Fig. 1D). Further investigations of MeV droplets were performed using this minimal system ($P_{50}N_{525}$ and $P_{304–507}$) unless otherwise stated. $P_{50}N_{525}$ and $P_{304–507}$ colocalize in droplets (Fig. 1E), and their phase transition was shown to be salt dependent: Beyond 400 mM, fluorescent speckles are no longer observed (fig. S2).

The phase diagram of LLPS was established on the basis of turbidity (Fig. 2A), identifying threshold concentrations of 10 µM for both proteins, although fluorescence microscopy and negative-stain electron microscopy reveal phase separation at lower protein concentrations. Fluorescence microscopy indicates that proteins are homogeneously distributed within droplets (fig. S3).

The liquid-like behavior of droplets was confirmed by DIC microscopy via observation of fusion events with relaxation into spherical droplets (Fig. 2B and movie S1) and further demonstrated

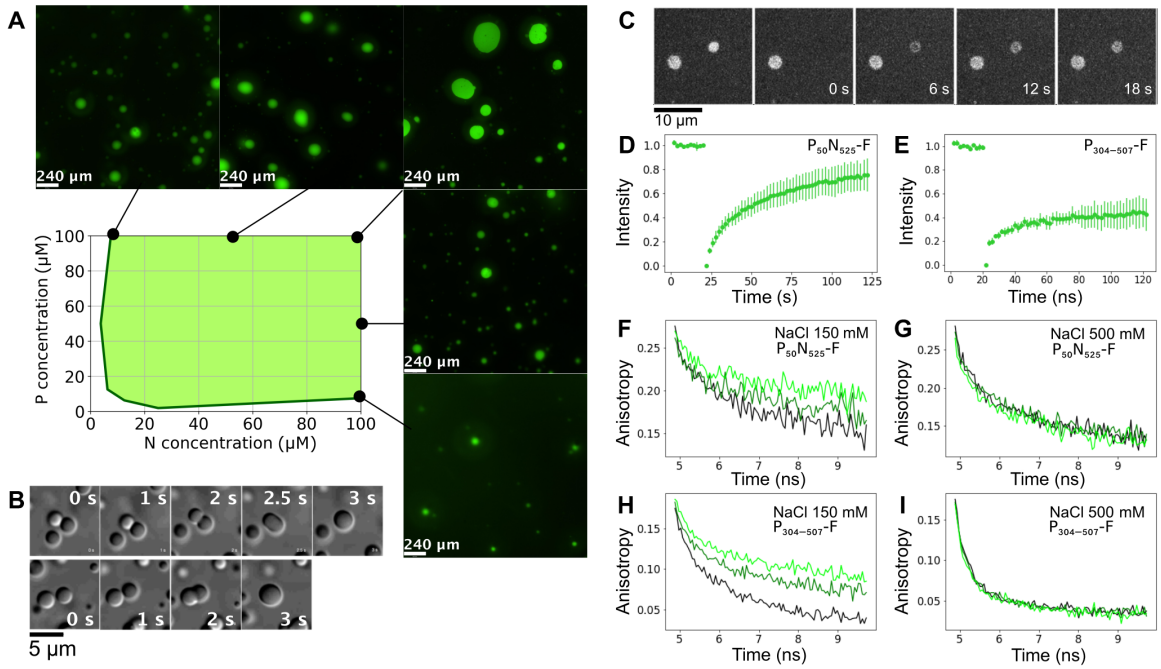


Fig. 2. MeV droplets exhibit liquid-like behavior. (**A**) Turbidity-based phase diagram as a function of $P_{50}N_{525}$ and $P_{304–507}$ concentrations. Green shading represents ranges of combinations of protein concentration that form droplets. Examples of fluorescent images at several N and P concentrations that show phase separation are presented around the phase diagram. (**B**) Fusion events followed by relaxation into spherical droplets in a mixture comprising $P_{304–507}$ (100 μM) and $P_{50}N_{525}$ (10 μM). (**C**) Images of FRAP shows exchange of $P_{50}N_{525}$–fluorescein amidite between phases. (**D**) FRAP kinetics of $P_{50}N_{525}$ (25 μM with 10 nM $P_{50}N_{525}$-fluorescein) in droplets formed with $P_{304–507}$ at 1:2 ratio. (**E**) FRAP kinetics of $P_{304–507}$ (50 μM with 50 nM $P_{304–507}$-fluorescein) in droplets formed with $P_{50}N_{525}$ at 2:1 ratio. (**F** and **G**) Fluorescence anisotropy of $P_{50}N_{525}$-fluorescein (10 μM $P_{50}N_{525}$, 500 nM $P_{50}N_{525}$-fluorescein) in the dilute phase (black) and in the presence of $P_{304–507}$ under phase-separating conditions [N:P = 1:2 (dark green) and N:P = 1:4 (light green)]. Fluorescence lifetimes were measured at room temperature, (F) in buffer containing 150 mM NaCl and (G) in buffer containing 500 mM NaCl that suppresses phase separation but not the interaction between proteins. (**H** and **I**) Fluorescence anisotropy of $P_{304–507}$-fluorescein (10 μM, 500 nM $P_{304–507}$-fluorescein) in the dilute phase (black) and in the presence of $P_{50}N_{525}$ at phase-separating conditions [P:N = 4:1 (dark green) and P:N 2:1 (light green)]. Fluorescence lifetimes were measured at room temperature, (H) in buffer containing 150 mM NaCl and (I) in buffer containing 500 mM NaCl that suppresses phase separation but not the interaction between proteins.

by fluorescence recovery after photobleaching (FRAP) of fluorescein-labeled $P_{50}N_{525}$, showing recovery within seconds (Fig. 2, C and D). FRAP of P protein is measurably slower (Fig. 2E), suggesting that the branched multimeric P acts as a dynamic scaffold for the liquid droplets (*33*). Rotational diffusion of proteins within droplets and in the surrounding phase was measured using time-resolved fluorescence anisotropy. The rotational correlation time of $P_{50}N_{525}$ in the presence of $P_{304–507}$ was significantly slowed down compared to $P_{50}N_{525}$ in free solution (Fig. 2F), while there was no difference under conditions of high-salt buffer, which suppresses phase transition but maintains interactions between the proteins (Fig. 2G). Similar observations were measured for $P_{304–507}$ in the free protein and in the presence of $P_{50}N_{525}$ (Fig. 2, H and I). Both proteins therefore remain liquid in the condensed phase, although their rotational dynamics are slowed, possibly due to increased viscosity in the droplets. In summary, these results confirm that LLPS occurs under near-physiological protein and salt concentrations upon mixing of full-length N- and the C-terminal domains of P in vitro, complementing recent in vivo observations (*32*).

P_{LOOP} is intrinsically disordered

The conformational behavior of the region between the tetramerization domain and XD, $P_{377–457}$ (P_{LOOP}), was analyzed using NMR spectroscopy. ^{13}C, ^{15}N, and ^{1}H chemical shifts were combined with statistical coil sampling and ensemble selection approach ASTEROIDS (fig. S4), revealing that this domain is essentially devoid of significantly populated secondary structure with the exception of two regions with 20 and 10% α-helical propensity around residues 406 and 433, respectively. ^{15}N relaxation data also show that this domain is highly dynamic (fig. S5), although relaxation rates around regions 410 and 433 increased in constructs containing the tetrameric or XD domains ($P_{377–507}$ and $P_{304–507}$), probably due to slower rotational tumbling. Carbon chemical shifts of $P_{304–507}$ did not change significantly in comparison with P_{LOOP}, demonstrating that P_{LOOP} remains disordered in the context of the construct that links the tetramerization domain and XD and that is required for phase separation (see fig. S6).

The N_{TAIL}:P_{XD} interaction is essential for droplet formation

The C-terminal domains of N and P are known to interact via residues 487 to 501 of N_{TAIL}, which populates a conformational equilibrium of partially folded helices in solution (*34*) and the XD domain of P. The affinity of the MeV N_{TAIL} and P_{XD} (Phosphoprotein XD domain) interaction (*7*, *11*) under our experimental conditions is approximately 5 μM (Fig. 3, A to C). This affinity was not measurably dependent on salt concentration. The sole intermolecular hydrogen bonding interaction observed in a chimeric crystal structure of $N_{486–505}$:P_{XD} involves S491 of N_{TAIL} interacting with K489 and D493 of P_{XD} (*7*). Mutation of N (S491) has been shown to abrogate the N_{TAIL}:P_{XD} interaction—this is also the case under our experimental conditions (Fig. 3D)—and to significantly decrease viral transcription in vivo (*11*). The critical nature of this interaction for LLPS was

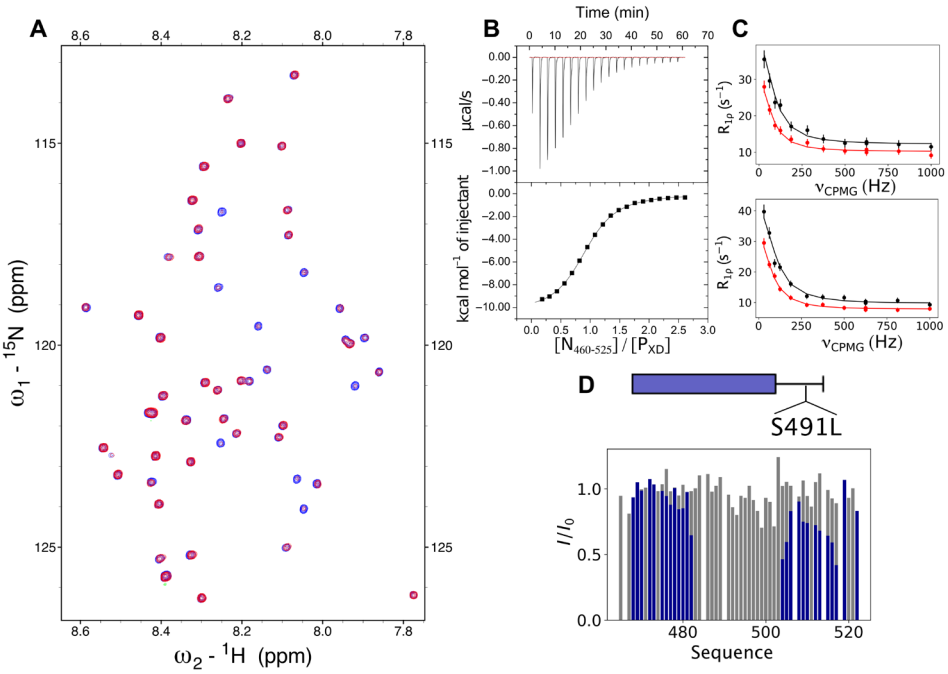


Fig. 3. Interaction between P_{XD} and $N_{460–525}$ is essential for MeV LLPS. (**A**) ^{1}H-^{15}N HSQC spectrum of free P_{XD} (blue) and upon interaction with unlabeled $N_{460–525}$ (red). Concentration of P_{XD} was 100 µM in both cases, and the concentration of $N_{460–525}$ was 20 µM. Spectra were recorded at 850 MHz ^{1}H frequency and 298 K. See Materials and Methods for additional experimental conditions. (**B**) Isothermal titration calorimetry of P_{XD} and $N_{460–525}$ interaction. Raw injection heats (top) and the corresponding specific binding isotherms (bottom) measured at 25°C are shown. P_{XD} (845 µM) was sequentially injected to $N_{460–525}$ (70 µM) in a microcalorimeter cell from 0 to 2.5 ratio. Fitted parameters: $\Delta S = -10.7$ cal mol^{-1} deg^{-1}, $\Delta H = -10{,}460 \pm 46.47$ cal mol^{-1}, $N = 0.927$ sites. (**C**) ^{15}N CPMG relaxation dispersion of P_{XD} (100 µM) for residues 504 (top) and 509 (bottom) in the presence of 46% $N_{460–525}$. CPMG frequencies were in the range from 31 to 1000 Hz. The experiments were recorded at a ^{1}H frequency of 700 MHz (black) and 950 MHz (red) at 25°C. Data were fitted simultaneously for residues R465, S466, I468, R472, R477, K478, Y480, K496, I504, and E509 with the two-state model, $p_b = 23\%$, exchange rate (k_{ex}) = 285 ± 50 s^{-1}. (**D**) Intensity ratio profiles of free versus P_{XD}-bound $N_{460–525}$ [wild type (WT)] (blue) and $N_{460–525}$(S491L) (gray). The ratio between N and P was 1:1 at 200 µM concentration.

demonstrated via the S491L mutation, which resulted in suppression of phase separation.

The P_{LOOP}-N interaction modulates the dynamics of MeV droplets

An additional interaction between P_{LOOP} and N was identified using NMR spectroscopy (Fig. 4, A and B), suggesting that the role of this domain, which is present in all constructs involved in phase separation, involves a direct interaction with N rather than acting as a simple spacer (*35*). Interactions centered on residues ^{409}VLK^{411} and ^{433}LLK^{435} exhibit distinct affinities [estimated dissociation constant (K_d) 264 and 606 µM, respectively]. Mutation of ^{409}VLK^{411} to AAA inhibited this interaction (Fig. 4C) with N and modulated macroscale dynamic properties of droplets, as measured from FRAP, but did not completely inhibit LLPS (Fig. 4D). Both ^{409}VLK^{411} and ^{433}LLK ^{435} exhibit high sequence conservation over MeV phosphoproteins (fig. S7).

Investigating the thermodynamics of droplet formation using NMR

To study interactions between proteins under phase-separating conditions, we compared ^{15}N heteronuclear single-quantum coherence (HSQC) spectra of $P_{304–507}$ and $P_{50}N_{525}$ separately and in different stoichiometric mixtures. Residues from 377 to 507 were observable in the $P_{304–507}$ spectrum, while N_{TAIL} and flexible residues corresponding to N_{ARM} and $P_{1–50}$ were observable in $P_{50}N_{525}$ (Fig. 5). Overall peak intensity was initially observed to decrease for both N and P with increasing stoichiometric ratio (N/P), with no noticeable chemical shift differences. Similar observations were recently made in mixtures of two interacting translational regulators (*36*). Apparent rotational and transverse diffusion rates were similar to those measured for free proteins (fig. S8). Rotational correlation times would be expected to increase significantly within droplets due to increased viscosity (*37*). This result therefore suggests that observable signal derives from N and P in the dilute phase.

Decrease in resonance intensity of N and P is accompanied by broadening of resonances from P_{XD} and $N_{483–505}$, due to the expected interactions between P and N. Increasing the N:P ratio further results in a minimum of total peak intensity at a ratio of approximately 0.15, followed by continual increase (Fig. 5). To follow the apparent thermodynamics of droplet formation, intensities of 23 (N) and 9 (P) nonoverlapping peaks corresponding to residues that are not affected by the interaction under non–phase-separating conditions were followed as a function of the concentration of N (concentration of P was held constant—see Materials and Methods). The ratio of N and P in droplets was estimated according to

$$r_D = \frac{N_D}{P_D} = (N_{TOT} - N_{OBS})/(P_{TOT} - P_{OBS}) \quad (1)$$

N_D and P_D refer to the apparent concentration of droplet-associated protein, N_{TOT} and P_{TOT} refer to the total concentrations of P and N,

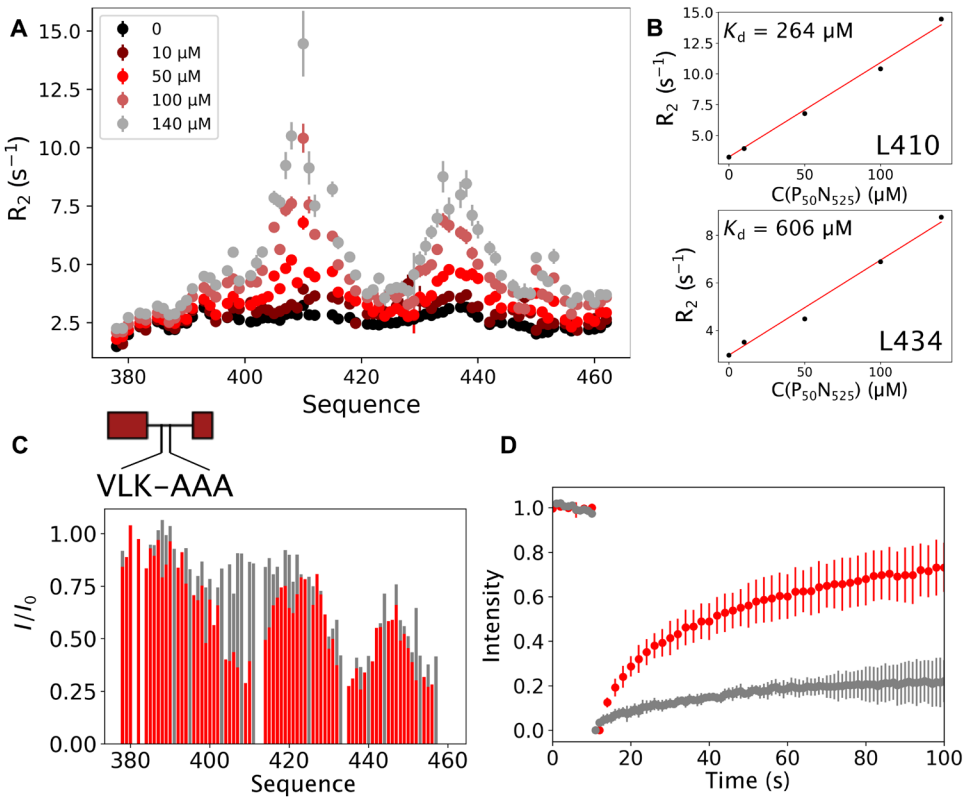


Fig. 4. Interaction of P_{LOOP} with N modulates the dynamics of MeV liquid droplets. (**A**) ^{15}N R_2 of P_{LOOP} (100 μM) in the presence of $P_{50}N_{525}$ at different concentrations at 600 MHz, 25°C. (**B**) Transverse ^{15}N relaxation rates of the central residues of the two interaction regions in P_{LOOP} (centered on residues 410 and 434, respectively). The rates increase linearly as a function of concentration of added $P_{50}N_{525}$. (**C**) Intensity ratios of free versus $P_{50}N_{525}$-bound P_{LOOP} (red) and P_{LOOP} ^{409}VLK^{411}-AAA mutant (gray). The ratio between N^0P and P was 1:1; concentration was 100 μM. (**D**) FRAP of $P_{50}N_{525}$-fluorescein with $P_{304–507}$ (WT) (red) and $P_{304–507}$ ^{409}VLK^{411}-AAA (gray) in droplets. In both cases, the concentration of $P_{50}N_{525}$ was 25 μM; concentrations of $P_{304–507}$ WT and ^{409}VLK^{411}-AAA mutant were 50 μM, and entire droplets were photobleached.

and N_{OBS} and P_{OBS} refer to the apparent concentrations in the liquid phase that are observed by NMR. This analysis reveals a clear dependence of r_D on the total concentration of N, varying by nearly a factor of 3 for total concentrations of N that vary between 7.5 and 75 μM (Fig. 5 and table S1). This dependence indicates that a simple two-state exchange mechanism cannot explain the experimental observation and suggests that at least two steps are necessary to describe the process of droplet formation.

RNA colocalizes to MeV droplets and triggers nucleocapsid formation in guttala

The functional role of MeV LLPS was examined by adding short, fluorescently labeled RNA decamers to samples containing droplets. Within seconds of adding the RNA, the fluorescence intensity of the droplets increased, starting from the perimeter (Fig. 6A and movie S2). This demonstrates that RNA localizes preferentially to the membraneless organelles where it presumably interacts with the N:P complex. After incubation for 1 hour at 37°C, the samples were examined using negative-stain electron microscopy, revealing the presence of assembled nucleocapsid-like structures (Fig. 6B).

To determine the rate of nucleocapsid assembly, we used real-time NMR. This approach has recently been used to observe the kinetics of assembly of nucleocapsid-like particles in the dilute phase (*13*). Nonfluorescently labeled RNA hexamers comprising six adenines were added to the droplet suspension comprising ^{15}N-labeled N, resulting in the assembly of nucleocapsid-like particles. Assembly can be followed by detecting the appearance and increasing intensity of NMR signals from the $P_{1–50}$ peptide that is released upon assembly of the nucleocapsids (Fig. 6, C and D). Comparison of the rate of increase of the NMR signals corresponding to $P_{1–50}$ in the condensed and dilute phase (in the absence of $P_{304–507}$) unambiguously shows that nucleocapsid formation is accelerated in droplets at the same global concentration of $N_{1–525}$.

DISCUSSION

Nonsegmented negative-sense RNA viruses are known to form inclusion bodies comprising proteins involved in the replication machinery. The recent observation of colocalization of N and P of both rabies (*31*) and measles (*32*) into liquid-like membraneless puncta following infection suggests that this phenomenon may play a role in viral replication or transcription and that the phenomenon may be widespread throughout the Mononegavirales order and possibly related viruses. The formation of such membraneless organelles has been proposed to enhance reaction rates by concentrating interaction partners in a confined volume, for storage purposes or, in the case of viral organelles, to provide protection from the host immune system. Investigating the role of this poorly understood phenomenon is essential to understanding the viral cycle. Describing the physical mechanisms that underlie their formation and stability

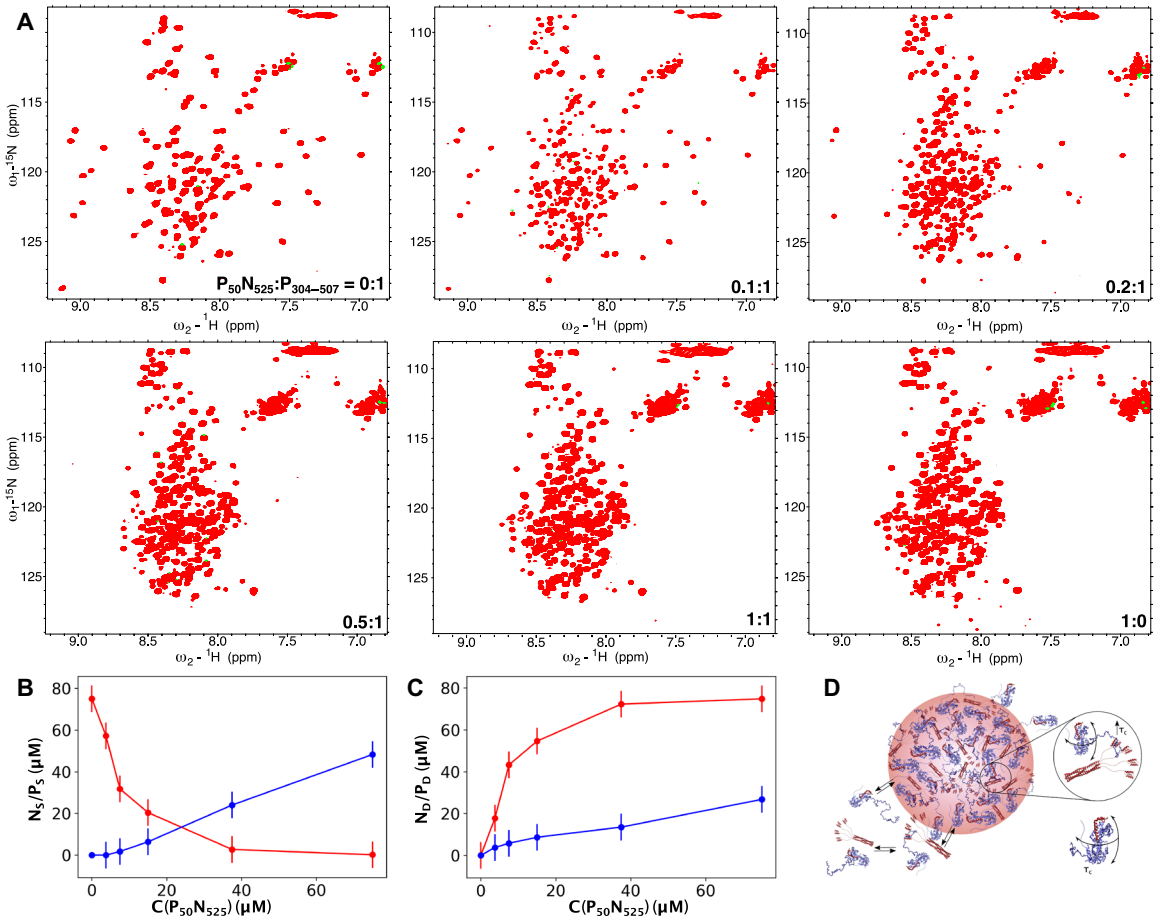


Fig. 5. Thermodynamics of MeV NP phase separation. (**A**) ^{1}H-^{15}N HSQC of ^{15}N-labeled $P_{50}N_{525}$ and $P_{304–507}$ at different admixtures of the two proteins; all spectra were recorded at 850 MHz, 25°C. $P_{304–507}$ concentration was 75 μM, and $P_{50}N_{525}$ concentration varies from 5 to 100% of 75 μM. (**B**) Concentration of $P_{50}N_{525}$ and $P_{304–507}$ proteins in the dilute phase derived from the spectra shown in (A). (**C**) Comparison of concentrations of droplet-associated $P_{50}N_{525}$ and $P_{304–507}$ in the sensitive volume, estimated on the basis of observation of protein in the dilute phase [derived from analysis of spectra shown in (A)]. Total concentration of P in the mixture was 75 μM. (**D**) Schematic representation of droplet formation upon mixing $P_{50}N_{525}$ and $P_{304–507}$.

represents a significant challenge, particularly if one considers their potential as targets of viral inhibition. To provide molecular insight into these aspects, we aimed to recreate the phenomenon of LLPS involving MeV N and P proteins in vitro.

We first demonstrated LLPS of N and P under conditions where neither protein undergoes phase separation alone. Expression of independent domains and combinations of domains then allowed us to identify the regions of N and P that are necessary for droplet formation in vitro. This revealed that apart from the first 50 amino acids of P that chaperone N, the rest of the 300–amino acid disordered P_{TAIL} is dispensable for LLPS, while the tetramerization domain of P, the hitherto uncharacterized disordered domain P_{LOOP}, and the three-helix bundle XD are all essential. N cannot be purified in an isolated form, owing to the propensity of the protein to bind to cellular RNA and form nucleocapsid-like structures. We therefore purified a chaperoned form of N bound to the N-terminal peptide of P, which maintains N in a monomeric, unassembled form (N^0P). LLPS only occurs when N^0P comprising the full-length N protein, including both N_{CORE} and the 125–amino acid disordered N_{TAIL}, is present. Full-length N and P were also coexpressed, and the integral N^0P complex was shown to undergo LLPS. Overall, these results verify that this minimal in vitro system reproduces the characteristic colocalization of N and P in membraneless organelles that has been observed in infected cells.

We also identified the interaction between the helical motif of N_{TAIL} and XD to be essential for phase separation. These results align closely with data measured in vivo on cells infected with rabies (*31*) and more recently MeV (*32*). The N_{TAIL}:P_{XD} interaction has previously been associated with transport of the polymerase complex to the nucleocapsid before replication and transcription (*10*, *12*, *38*, *39*). Here, we demonstrate that at least one role of this interaction involves the free-energy stabilization of LLPS. Mutation of a single residue (S491L) that is known to mediate the interaction between N_{TAIL} and P_{XD}, and to strongly affect replication in vivo (*11*), results in abrogation of LLPS.

Additional weak interactions between P_{LOOP} and N are also shown to modulate the fluidity of the droplets, suggesting that polyvalency of N:P interactions plays a role in droplet formation. The importance of both disordered and folded, or partially folded domains in the polyvalent complex that stabilizes the MeV droplets invokes their possible roles as "stickers" and "spacers" (*40*), which assure the noncovalent binding and physical separation of interacting components, respectively. Nevertheless, the role of P_{LOOP} appears to be more

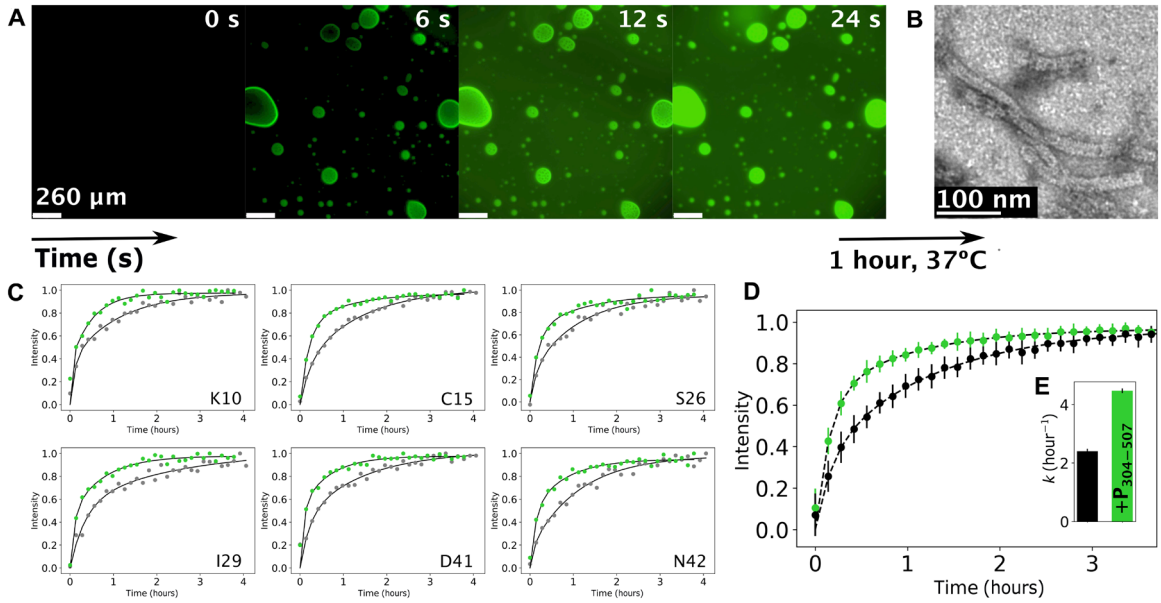


Fig. 6. Nucleocapsid assembly occurs in liquid droplets and is enhanced by phase separation. (**A**) Fluorescence microscopy images showing fluorescently labeled RNA diffusing into droplets preformed by mixing $P_{50}N_{525}$ (20 µM) and $P_{304–507}$ (80 µM); RNA of 200 µM (with 10% fluorescently labeled) was added on one side of the coverslip and spontaneously diffused in the sample. (**B**) Negative-stain electron microscopy of the sample shown in (A), after 1 hour of incubation at 37°C. (**C**) Kinetic traces of peak intensities corresponding to individual residues of $P_{1–50}$ in SOFAST ^{1}H-^{15}N HMQC spectra acquired at a resolution of 8.5 min (8 scans, 200 ^{15}N increments, and 1024 ^{1}H), following addition of OH-A_6-OH RNA to $P_{50}N_{525}$ (gray) and under phase-separating conditions with $P_{304–507}$ (green) at 293 K. Data were fitted with a biexponential function as previously described (*13*). (**D**) Averaged intensities for residues E4, K10, N11, G12, E14, C15, I16, G25, S26, L27, I29, A32, M33, A35, D41, N42, G44, and Q45 of $P_{1–50}$ in $P_{50}N_{525}$ (gray) and under phase-separating conditions with $P_{304–507}$ (green). Dots with SD represent experimental data, and lines represent biexponential fit. (**E**) Rate of assembly of nucleocapsids as measured by the appearance of resonances from P_{50} when assembled using $P_{50}N_{525}$ (gray) and when using $P_{50}N_{525} + P_{304–507}$ (green) having formed droplets. The concentration of $P_{50}N_{525}$ and RNA was constant in both cases.

complex than this as the impact of this interaction on the nature of MeV droplets suggests that it is capable of fulfilling both roles.

Using NMR spectroscopy, we are able to investigate the thermodynamic equilibrium between dilute and condensed phases in the presence of MeV LLPS by comparing the rate of increase in signal intensity from N_{TAIL} and the decrease in intensity of signal from P_{LOOP} and XD, as a function of the concentration of N. Assuming that all signal derives from the dilute phase (as evidenced by rotational and translational diffusion measurements) and that signal drop results uniquely from slow exchange with an invisible form of each protein in droplets, our observations suggest that the stoichiometry of N and P within droplets depends on the ratio of the two proteins in the mixture. As the experimental data cannot be explained using a simple one-step reaction, it is interesting to speculate on the nature of the second interaction. One possible explanation involves N interacting with P to form droplets, with additional recruitment of N to existing N:P droplets. Such a model would be in agreement with the critical presence of the tetramerization domain and the polyvalency of the four branched dynamic XD interaction sites, which may form a dynamic scaffold in MeV LLPS to which additional N could be recruited. Note that both P and N remain dynamic in the context of the droplets, as evidenced by the rotational diffusion measured by fluorescence anisotropy. The proposed role of P as a dynamic scaffold is supported by the observation of differential translational mobility of P within droplets with respect to N, as reported by FRAP. More experimental measurements on this system are required to better characterize these molecular processes.

Replication and transcription of negative-strand RNA viruses are thought to occur within inclusion bodies comprising N, P, and L. The colocalization of these proteins within liquid droplets may act to increase local concentration of the components of the viral replication machinery, thereby increasing replication efficiency. We have therefore investigated whether the liquid droplets formed upon mixing of N and P in vitro are capable of supporting nucleocapsid assembly, which is an essential part of the viral replication process. We previously demonstrated that assembly can be observed in the dilute phase, by adding RNA to monomeric N^0P and using NMR spectroscopy and fluorescence anisotropy to monitor the rate of assembly (*13*). By adding RNA to a suspension of N:P organelles, we show that RNA molecules preferentially localize to the droplets and, once within the droplets, trigger assembly of nucleocapsid-like particles. Notably, the rate of assembly is significantly increased compared to non–phase-separating conditions, providing evidence that LLPS has the ability to enhance viral function and offering a first glimpse of these viral factories in action. This also indicates that the condensed nature of the droplets does not restrict the ability to assemble these very large superstructures and demonstrates that these minimalist N:P droplets are sufficient to promote one of the essential processes of viral replication. In this context, we note that nucleocapsid structures were observed to be ejected from liquid droplets formed in cells infected with rabies virus (*31*).

Phosphorylation of P and N have both been reported to affect MeV function (*41*, *42*) and, more recently, to influence the nature of liquid droplet formation in vivo (*32*). Although both N and P were purified from *E. coli* in their unmodified forms, the technology

described here will allow us to investigate the impact of phosphorylation in future studies. In particular, the role of the 300–amino acid disordered P_{TAIL} domain may become clearer as this comprises numerous phosphorylation sites but is not essential for LLPS. This experimental system will also allow us to investigate the role of host proteins that are thought to enter inclusion bodies and the viral proteins C and L.

In conclusion, we have demonstrated that purified N and P MeV proteins form liquid-like membraneless organelles upon mixing in vitro. We identify weak interactions involving intrinsically disordered domains of N and P that are implicated in this process and identify one interaction between the C-terminal domains as essential to phase separation. P protein appears to play a role as a multivalent scaffold stabilizing the droplets. The droplets exhibit both physical and functional features characteristic of paramyxoviral inclusion bodies: RNA colocalizes to the droplets and thereby triggers assembly of N protomers into nucleocapsid-like particles that encapsidate the RNA. The ability to reproduce mechanistic behavior of MeV factories in vitro will allow us to further examine the role of additional host and viral partners at the molecular level. The observation that similar puncta or inclusion bodies have been found for numerous negative-strand RNA viruses, including filoviruses and pneumoviruses, suggests that the observations made here may be of general interest in understanding the molecular basis of mononegaviral infection.

MATERIALS AND METHODS

Protein expression, purification, and mutagenesis

All P and N constructs were cloned into pET41c(+) between the Nde I and Xho I cleavage sites, where the Xho I site was ligated with a cleaved Sal I site of the insert, yielding a construct with a C-terminal 8His-tag (*5*). All constructs were expressed in *E. coli* Rosetta TM(lDE3)/pRARE (Novagen) overnight at 20°C after induction at an optical density of 0.6 with 1 mM isopropyl-β-D-thiogalactopyranoside. N_{525} was cotransformed with $P_{1–507}$ and purified in complex. Cells were lysed by sonication and subjected to standard Ni purification in 20 mM tris (pH 8.0) and 150 mM NaCl or 500 mM NaCl for XD containing constructs. The protein was eluted with 20 mM tris (pH 8), 150/500 mM NaCl, and 500 mM imidazole and then concentrated and subjected to size exclusion chromatography (SEC, Superdex 75/200) in 50 mM Na-phosphate (pH 6.0), 150/500 mM NaCl, and 2 mM dithiothreitol (DTT) buffer (NMR buffer). The $P_{50}N_{525}$, $P_{50}N_{405}$, $P_{300}N_{525}$, and $P_{300}N_{405}$ were produced as described previously (*13*, *14*). Mutation of S491L in N and of ^{409}VLK^{411}→AAA in P were inserted by site-directed mutagenesis, and $P_{50}N_{525}$ (S491L), $N_{460–525}$ (S491L), $P_{304–507}$ (^{409}VLK^{411}→AAA), and $P_{377–507}$ (^{409}VLK^{411}→AAA) were produced and purified in the same way as wild-type proteins.

Expression of unlabeled protein was performed in LB medium. Proteins labeled for NMR (^{15}N and ^{13}C) were expressed under the same conditions in M9 minimal medium. Cysteine mutations for fluorescence labeling were performed by site-directed mutagenesis into the constructs where previously all native cysteines were exchanged to alanines. Mutant proteins were expressed and purified in the same way as wild type and concentrated to around 100 μM. The proteins were then labeled using maleimide chemistry (*43*). Briefly, the protein was incubated overnight with 10 mM DTT, then dialyzed against 50 mM Na-phosphate buffer (pH 7.0) and 150/500 mM NaCl in two steps, and mixed with 5× excess of maleimide dye (fluorescein or Alexa Fluor 594) solubilized in dimethyl sulfoxide. After 2 hours of incubation at room temperature, the labeling reaction was stopped by adding 5 mM DTT, and proteins were separated from the excess dye by SEC (Superdex 75/200) in NMR buffer. SDS–polyacrylamide gel electrophoresis (PAGE) gels are shown in fig. S1.

RNA

All RNAs were purchased from Integrated DNA Technologies (Leuven, Belgium), and all were terminated with OH at both 3′ and 5′ or with fluorescein amidite at the 3′ end.

Droplet assembly and turbidity assay

P and N constructs were first mixed and then diluted with buffer; thus, the final NaCl concentration was 150 mM (unless different NaCl concentration is stated). Samples were incubated at room temperature for 5 min, and absorbance at 395 nm was read on a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA). An absorbance value higher than 0.1 was interpreted as a presence of phase separation in the system.

Microscopy imaging

Samples were imaged using either DIC or epifluorescence on an Olympus IX81 inverted microscope. Fluorescence signal was acquired using a 485(20)-nm (fluorescein) or a 560(25)-nm (Alexa Fluor 594) excitation filter combined with a 525(30)-nm or a 607(36)-nm emission filter, respectively. Excitation light (Metal-Halide X-Cite 120) and emission wavelengths are split using a 4X4M-B quadriband dichroic mirror (Semrock). Oil immersion objectives 60× [Olympus PlanAPON60, 1.42 numerical aperture (NA)] or 100× (Olympus UPLFLN, 1.3 NA) were used for imaging, and a Hamamatsu Orca Flash4.0-V2 sCMOS (scientific Complementary Metal-Oxyde Semiconductor) 2048 × 2048 camera was used for detection. The acquisition and data analysis were performed with Volocity (Quorum Technologies) software.

Fluorescence recovery after photobleaching

FRAP was measured using an Olympus IX81 spinning-disk confocal microscope equipped with a Yokogawa CSUX1 spinning disk head and using diode-pumped solid-state lasers 488 nm (50 mW Coherent Sapphire) and 561 nm (100 mW Cobolt JIVE4). Emission is filtered through a 520(28)-nm filter for fluorescein. Acquisition is performed with an iXon Ultra EMCCD (Electron Multiplying Charge-Coupled Device) (Andor) 512 × 512 pixel camera. FRAP was restricted to a region of interest, defined graphically, with a galvanometric FRAP/PA device (Andor) using the same lasers used for imaging with AOTF (Acousto-Optic Tunable Filter) 15% and with a dwell time of 50 ms for 10 repeats before bleaching and with a 2-s interval for 40 repeats after. The laser power and the dwell time during bleaching were optimized to bleach 50% of the initial fluorescence intensity. Images were analyzed using FIJI and python plugin ImageFRAP (https://imagej.net/Analyze_FRAP_movies_with_a_Jython_script); the recovery was fitted with a single exponent.

Fluorescence anisotropy

Fluorescence anisotropy decays were measured using a Mini-Tau fluorescence lifetime spectrometer (Edinburgh Instruments, UK) equipped with a 450-nm picosecond pulsed diode laser and a TCC2 electronics module for time-correlated single photon counting. Photons were detected using a high-speed photon multiplier tube (Hamamatsu H10720-01). The sample was excited with vertically polarized light, and parallel and perpendicular polarized fluorescence light was detected during 120 s sequentially. Histograms of photon

arrival times were generated from every measurement, and anisotropy decays were calculated according to $r(t) = (I_{par}(t) - GI_{per}(t))/(I_{par}(t) + 2GI_{per}(t))$, with I_{par}/I_{per} being the intensities of parallel/perpendicular polarized light per time unit and G being the correcting factor for detection differences between the two detection pathways. The G factor was measured for each buffer condition using the free dye.

Isothermal titration calorimetry

Isothermal titration calorimetry was measured using iMicroCal iTC200 at 25°C. The titration experiment was performed by adding 2 µl of aliquots of 500 µM P_{XD} into the microcalorimeter cell filled with 50 µM N_{TAIL}. P_{XD} was titrated through 20 injections with 180-s intervals in between; the reaction mixture was continuously stirred at 750 rpm. The titration curve was fitted to the experimental data using the Origin version 7.0 software from MicroCal.

NMR spectroscopy

All experiments were acquired in NMR buffer at 25°C. The spectral assignment of $P_{50}N_{525}$ at pH 6.0 was verified on the basis of published assignment at pH 7.0 (*13*) using ^{13}C, ^{15}N-labeled samples using a set of BEST-TROSY triple resonance experiments.

The spectral assignments of ^{13}C- and ^{15}N-labeled $P_{377–457}$ and $P_{457–507}$ constructs were obtained using sets of triple resonance experiments correlating Ca, Cb, and CO resonances at a ^{1}H frequency of 600 MHz. NMRPipe (*44*) was used to process the spectra, and automatic assignment was performed with the program MARS and manually verified (*45*). Secondary chemical shifts were calculated using the random coil values from refDB (*46*).

Ensemble analysis of P_{LOOP}

The genetic algorithm ASTEROIDS selected representative ensembles [fitting ^{1}H, ^{15}N, and ^{13}C(Ca, Cb, and CO) backbone chemical shifts]. A 10,000-member ensemble of P_{LOOP} was generated using the statistical coil model Flexible Meccano (*47*), and 200-conformer ensembles matching experimental backbone chemical shifts were selected from this ensemble using the ASTEROIDS genetic algorithm as previously described (*48*, *49*).

Calculation of apparent K_d values from NMR experiments for the interaction between N and P_{LOOP}/P_{XD}

^{15}N R_1 relaxation rates were obtained by sampling the decay of magnetization (seven delays between 0 and 1.71 s). $R_{1\rho}$ relaxation rates were measured using seven delay times between 0.001 and 0.24 s. The spin-lock field was 1500 Hz, and R_2 was calculated from R_1 and $R_{1\rho}$, considering the resonance offset. The residue-specific apparent K_d values for $P_{377–457}$ and N were calculated from ^{15}N relaxation as described elsewhere (*50*). $P_{377–507}$ concentration remained constant (100 µM), and N concentration varied from 0 to 140 µM. The rotational correlation time of the complex was assumed to be equal to τ_c of $P_{50}N_{525}$ alone and was estimated from its molecular weight.

^{15}N relaxation dispersion was measured at a $P_{457–507}$ concentration of 100 µM in the presence of 70 and 30 µM unlabeled N_{TAIL} at 700 and 850 MHz and using 14 points at CPMG frequencies between 31 and 1000 Hz with a constant-time relaxation of 32 ms (*51*). Data for nine residues and two fields were fitted simultaneously using ChemEx (https://github.com/gbouvignies/chemex) and a two-state exchange model. The population of $P_{457–507}$ bound to $N_{460–525}$ (p_B) and the exchange rate (k_{ex}) were obtained from the fit, and the K_d was calculated assuming a 1:1 binding stoichiometry. The error of the calculated K_d value was estimated by error propagation from the fitting error of p_B and assuming a 5% error in concentration determination of both proteins.

NMR measurements of phase-separating samples

^{15}N-labeled $P_{50}N_{525}$ and $P_{304–507}$ were premixed at different ratios and diluted in buffer so that the final NaCl concentration was 150 mM and then loaded in a 3-mm NMR tube. ^{1}H-^{15}N HSQCs and $R_{1\rho}$ experiments were measured as described for non–phase-separating samples. Pulsed field gradient NMR diffusion experiments were measured using ^{15}N-edited stimulated-echo diffusion experiments using a 300-ms diffusion delay at 16 to 20 gradient strengths ranging linearly from 0 to 0.55 T m^{-1} (*52*). ^{1}H-^{15}N HSQCs used for the development of the kinetic model were measured using both $P_{50}N_{525}$ and $P_{304–507}$ ^{15}N-labeled samples at a constant $P_{304–507}$ concentration of 75 µM and $P_{50}N_{525}$ 5, 10, 20, 50, and 100%. The concentration of $P_{50}N_{525}$ and $P_{304–507}$ in the dilute and condensed phase was estimated on the basis of intensities of 23 for $P_{50}N_{525}$ and 9 for $P_{304–507}$ nonoverlapping and noninteracting peaks, which were summed and normalized by the intensities of the peaks in the spectra of the individual proteins. The amount of each protein in droplets was then estimated by subtraction. This allows us to estimate the concentration of droplet-associated N and P in the sensitive NMR volume (but not in droplets as we do not know the volume of droplets). We can then estimate the ratio of N and P associated to droplets as a function of concentration of N.

Nucleocapsid assembly by NMR

Kinetics of $P_{50}N_{525}$ (alone or in the presence of $P_{304–507}$ or $P_{377–507}$) nucleocapsid assembly were followed using NMR spectroscopy using the same procedure as previously described (*13*). Time resolution was approximately 8 min. The assembly process was initiated by adding six-nucleotide RNA (OH-A_6-OH), and its time trace was followed by measuring intensities of resonances from $P_{1–50}$ on SOFAST HMQC spectra. Assembly rates were obtained by averaging intensities of 18 peaks and fitting them with a biexponential function; mean assembly rates were calculated as follows: $k_1{*}a_1 + k_2{*}a_2$, where k_1 and k_2 are rates, a_1 and a_2 are their amplitudes.

Negative-stain electron microscopy

The nucleocapsid-like particles used for negative-stain electron microscopy were obtained by adding 50 µM RNA (HO-A_6-OH) to droplets formed by mixing 20 µM $P_{50}N_{525}$ and 80 µM $P_{304–507}$ in a final volume of 20 µl. Samples were incubated for 1 hour at 37°C, then applied to the clean side of carbon on mica (carbon/mica interface), and stained with 2% sodium silicotungstate (pH 7.0). Micrographs were taken with a T12 FEI microscope at 120 kV and a magnification of 30,000.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/14/eaaz7095/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. C. I. Paules, H. D. Marston, A. S. Fauci, Measles in 2019—Going backward. *N. Engl. J. Med.* **380**, 2185–2187 (2019).
2. J. Curran, D. Kolakofsky, Replication of paramyxoviruses. *Adv. Virus Res.* **54**, 403–422 (1999).
3. R. Cox, R. K. Plemper, The paramyxovirus polymerase complex as a target for next-generation anti-paramyxovirus therapeutics. *Front. Microbiol.* **6**, 459 (2015).

4. S. Longhi, V. Receveur-Bréchot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, R. Yeo, S. Finet, B. Canard, The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J. Biol. Chem.* **278**, 18638–18648 (2003).
5. S. Milles, M. R. Jensen, C. Lazert, S. Guseva, S. Ivashchenko, G. Communie, D. Maurin, D. Gerlier, R. W. H. Ruigrok, M. Blackledge, An ultraweak interaction in the intrinsically disordered replication machinery is essential for measles virus function. *Sci. Adv.* **4**, eaat7778 (2018).
6. G. Communie, T. Crépin, D. Maurin, M. R. Jensen, M. Blackledge, R. W. H. Ruigrok, Structure of the tetramerization domain of measles virus phosphoprotein. *J. Virol.* **87**, 7166–7169 (2013).
7. R. L. Kingston, D. J. Hamel, L. S. Gay, F. W. Dahlquist, B. W. Matthews, Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8301–8306 (2004).
8. S. G. Guryanov, L. Liljeroos, P. Kasaragod, T. Kajander, S. J. Butcher, Crystal structure of the measles virus nucleoprotein core in complex with an N-terminal region of phosphoprotein. *J. Virol.* **90**, 2849–2857 (2015).
9. I. Gutsche, A. Desfosses, G. Effantin, W. L. Ling, M. Haupt, R. W. H. Ruigrok, C. Sachse, G. Schoehn, Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* **348**, 704–707 (2015).
10. S. Gely, D. F. Lowry, C. Bernard, M. R. Jensen, M. Blackledge, S. Costanzo, J.-M. Bourhis, H. Darbon, G. Daughdrill, S. Longhi, Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. *J. Mol. Recognit.* **23**, 435–447 (2010).
11. L.-M. Bloyet, J. Brunel, M. Dosnon, V. Hamon, J. Erales, A. Gruet, C. Lazert, C. Bignon, P. Roche, S. Longhi, D. Gerlier, Modulation of re-initiation of measles virus transcription at intergenic regions by P_{XD} to N_{TAIL} binding strength. *PLOS Pathog.* **12**, e1006058 (2016).
12. V. Du Pont, Y. Jiang, R. K. Plemper, Bipartite interface of the measles virus phosphoprotein X domain with the large polymerase protein regulates viral polymerase dynamics. *PLOS Pathog.* **15**, e1007995 (2019).
13. S. Milles, M. R. Jensen, G. Communie, D. Maurin, G. Schoehn, R. W. H. Ruigrok, M. Blackledge, Self-assembly of measles virus nucleocapsid-like particles: Kinetics and RNA sequence dependence. *Angew. Chem. Int. Ed. Engl.* **55**, 9356–9360 (2016).
14. A. Desfosses, S. Milles, M. R. Jensen, S. Guseva, J.-P. Colletier, D. Maurin, G. Schoehn, I. Gutsche, R. W. H. Ruigrok, M. Blackledge, Assembly and cryo-EM structures of RNA-specific measles virus nucleocapsids provide mechanistic insight into paramyxoviral replication. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4256–4264 (2019).
15. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
16. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
17. T. J. Nott, T. D. Craggs, A. J. Baldwin, Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. *Nat. Chem.* **8**, 569–575 (2016).
18. M.-T. Wei, S. Elbaum-Garfinkle, A. S. Holehouse, C. C.-H. Chen, M. Feric, C. B. Arnold, R. D. Priestley, R. V. Pappu, C. P. Brangwynne, Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat. Chem.* **9**, 1118–1125 (2017).
19. R. R. Novoa, G. Calderita, R. Arranz, J. Fontana, H. Granzow, C. Risco, Virus factories: Associations of cell organelles for viral replication and morphogenesis. *Biol. Cell* **97**, 147–172 (2005).
20. C. Netherton, K. Moffat, E. Brooks, T. Wileman, A guide to viral inclusions, membrane rearrangements, factories, and viroplasm produced during virus replication. *Adv. Virus Res.* **70**, 101–182 (2007).
21. X. Lahaye, A. Vidy, C. Pomier, L. Obiang, F. Harper, Y. Gaudin, D. Blondel, Functional characterization of Negri Bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication. *J. Virol.* **83**, 7948–7958 (2009).
22. B. S. Heinrich, Z. Maliga, D. A. Stein, A. A. Hyman, S. P. J. Whelan, Phase transitions drive the formation of vesicular stomatitis virus replication compartments. *MBio* **9**, e02290-17 (2018).
23. V. Rincheval, M. Lelek, E. Gault, C. Bouillier, D. Sitterlin, S. Blouquit-Laye, M. Galloux, C. Zimmer, J.-F. Eleouet, M.-A. Rameix-Welti, Functional organization of cytoplasmic inclusion bodies in cells infected by respiratory syncytial virus. *Nat. Commun.* **8**, 563 (2017).
24. S. Zhang, L. Chen, G. Zhang, Q. Yan, X. Yang, B. Ding, Q. Tang, S. Sun, Z. Hu, M. Chen, An amino acid of human parainfluenza virus type 3 nucleoprotein is critical for template function and cytoplasmic inclusion body formation. *J. Virol.* **87**, 12457–12470 (2013).
25. B. Precious, D. F. Young, A. Bermingham, R. Fearns, M. Ryan, R. E. Randall, Inducible expression of the P, V, and NP genes of the paramyxovirus simian virus 5 in cell lines and an examination of NP-P and NP-V interactions. *J. Virol.* **69**, 8001–8010 (1995).
26. A. Derdowski, T. R. Peters, N. Glover, R. Qian, T. J. Utley, A. Burnett, J. V. Williams, P. Spearman, J. E. Crowe Jr., Human metapneumovirus nucleoprotein and phosphoprotein interact and provide the minimal requirements for inclusion body formation. *J. Gen. Virol.* **89**, 2698–2708 (2008).
27. M. Omi-Furutani, M. Yoneda, K. Fujita, F. Ikeda, C. Kai, Novel phosphoprotein-interacting region in Nipah virus nucleocapsid protein and its involvement in viral replication. *J. Virol.* **84**, 9793–9799 (2010).
28. D. Ma, C. X. George, J. L. Nomburg, C. K. Pfaller, R. Cattaneo, C. E. Samuel, Upon infection, cellular WD repeat-containing protein 5 (WDR5) localizes to cytoplasmic inclusion bodies and enhances measles virus replication. *J. Virol.* **92**, e01726-17 (2018).
29. D. Spehner, R. Drillien, P. M. Howley, The assembly of the measles virus nucleoprotein into nucleocapsid-like particles is modulated by the phosphoprotein. *Virology* **232**, 260–268 (1997).
30. Y. Nakatsu, M. Takeda, S. Ohno, Y. Shirogane, M. Iwasaki, Y. Yanagi, Measles virus circumvents the host interferon response by different actions of the C and V proteins. *J. Virol.* **82**, 8296–8306 (2008).
31. J. Nikolic, R. Le Bars, Z. Lama, N. Scrima, C. Lagaudrière-Gesbert, Y. Gaudin, D. Blondel, Negri bodies are viral factories with properties of liquid organelles. *Nat. Commun.* **8**, 58 (2017).
32. Y. Zhou, J. M. Su, C. E. Samuel, D. Ma, Measles virus forms inclusion bodies with properties of liquid organelles. *J. Virol.* **93**, e00948-19 (2019).
33. E. Boke, M. Ruer, M. Wühr, M. Coughlin, R. Lemaitre, S. P. Gygi, S. Alberti, D. Drechsel, A. A. Hyman, T. J. Mitchison, Amyloid-like self-assembly of a cellular compartment. *Cell* **166**, 637–650 (2016).
34. M. R. Jensen, G. Communie, E. A. Ribeiro Jr., N. Martinez, A. Desfosses, L. Salmon, L. Mollica, F. Gabel, M. Jamin, S. Longhi, R. W. H. Ruigrok, M. Blackledge, Intrinsic disorder in measles virus nucleocapsids. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9839–9844 (2011).
35. J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, A. A. Hyman, A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699.e16 (2018).
36. T. H. Kim, B. Tsang, R. M. Vernon, N. Sonenberg, L. E. Kay, J. D. Forman-Kay, Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science* **365**, 825–829 (2019).
37. W. Adamski, N. Salvi, D. Maurin, J. Magnat, S. Milles, M. R. Jensen, A. Abyzov, C. J. Moreau, M. Blackledge, A unified description of intrinsically disordered protein dynamics under physiological conditions using NMR spectroscopy. *J. Am. Chem. Soc.* **141**, 17817–17829 (2019).
38. S. Longhi, L.-M. Bloyet, S. Gianni, D. Gerlier, How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication. *Cell. Mol. Life Sci.* **74**, 3091–3118 (2017).
39. J. Brunel, D. Chopy, M. Dosnon, L.-M. Bloyet, P. Devaux, E. Urzua, R. Cattaneo, S. Longhi, D. Gerlier, Sequence of events in measles virus replication: Role of phosphoprotein-nucleocapsid interactions. *J. Virol.* **88**, 10851–10863 (2014).
40. A. E. Posey, A. S. Holehouse, R. V. Pappu, Phase separation of intrinsically disordered proteins. *Methods Enzymol.* **611**, 1–30 (2018).
41. A. Sugai, H. Sato, M. Yoneda, C. Kai, Phosphorylation of measles virus phosphoprotein at S86 and/or S151 downregulates viral transcriptional activity. *FEBS Lett.* **586**, 3900–3907 (2012).
42. A. Sugai, H. Sato, M. Yoneda, C. Kai, Phosphorylation of measles virus nucleoprotein affects viral growth by changing gene expression and genomic RNA stability. *J. Virol.* **87**, 11684–11692 (2013).
43. E. A. Lemke, Site-specific labeling of proteins for single-molecule FRET measurements using genetically encoded ketone functionalities. *Methods Mol. Biol.* **751**, 3–15 (2011).
44. F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: A multidimensional spectral processing system based on UNIX Pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
45. Y.-S. Jung, M. Zweckstetter, Mars—Robust automatic backbone assignment of proteins. *J. Biomol. NMR* **30**, 11–23 (2004).
46. H. Zhang, S. Neal, D. S. Wishart, RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **25**, 173–195 (2003).
47. V. Ozenne, F. Bauer, L. Salmon, J.-r. Huang, M. R. Jensen, S. Segard, P. Bernadó, C. Charavay, M. Blackledge, *Flexible-meccano*: A tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **28**, 1463–1470 (2012).
48. G. Nodet, L. Salmon, V. Ozenne, S. Meier, M. R. Jensen, M. Blackledge, Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.* **131**, 17908–17918 (2009).
49. M. R. Jensen, L. Salmon, G. Nodet, M. Blackledge, Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J. Am. Chem. Soc.* **132**, 1270–1272 (2010).
50. S. Milles, D. Mercadante, I. V. Aramburu, M. R. Jensen, N. Banterle, C. Koehler, S. Tyagi, J. Clarke, S. L. Shammas, M. Blackledge, F. Gräter, E. A. Lemke, Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell* **163**, 734–745 (2015).

51. D. F. Hansen, P. Vallurupalli, L. E. Kay, An improved ^{15}N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J. Phys. Chem. B* **112**, 5898–5904 (2008).
52. C. A. Waudby, M. D. Mantle, L. D. Cabrita, L. F. Gladden, C. M. Dobson, J. Christodoulou, Rapid distinction of intracellular and extracellular proteins using NMR diffusion measurements. *J. Am. Chem. Soc.* **134**, 11312–11315 (2012).

Acknowledgments: IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (CEA). **Funding:** This work was supported by the European Research Council Advanced Grant DynamicAssemblies to M.B., GRAL (ANR-10-LABX-49-01), Finovi, and the Fondation Recherche Medicale (FRM) (Equipe DEQ20170336754). The work used the platforms of the Grenoble Instruct-ERIC Centre (ISBG; UMS 3518 CNRS-CEA-UJA-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology. S.M. acknowledges funding from an EMBO long-term fellowship (ALTF 468-2014) and European Commission (EMBOCOFUND2012, GA-2012-600394) via Marie Curie Action. The electron microscope facility was supported by the Rhône-Alpes Region, FRM, the fonds FEDER, the Centre National de la Recherche Scientifique (CNRS), the CEA, the University of Grenoble, EMBL, and the GIS-Infrastructures en Biologie Sante et Agronomie. **Author contributions:** S.G. carried out experiments and analyzed the data and wrote the manuscript. S.M. supervised experimental and analytical aspects. J.-P.K. assisted in the imaging experiments. D.M. prepared proteins and participated in imaging experiments. M.R.J. supervised experimental and analytical aspects. N.S. provided expertise in analytical and experimental aspects. R.W.H.R. and M.B. conceptualized and supervised the project and wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 2 October 2019
Accepted 9 January 2020
Published 1 April 2020
10.1126/sciadv.aaz7095

Citation: S. Guseva, S. Milles, M. R. Jensen, N. Salvi, J.-P. Kleman, D. Maurin, R. W. H. Ruigrok, M. Blackledge, Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. *Sci. Adv.* **6**, eaaz7095 (2020).

3.3 Model N : P droplets

The experimental data reporting on protein ratios in condensed and dilute states (see article Fig. 5) were further analysed. MeV N:P phase separation is an example of a heterotypic system where protein-protein interactions play a key role. So, in this case, phase separation can be described as a system of chemical equations where each equation characterises one of the interactions. We assume that the NMR experiments were performed on samples which reached equilibrium, so that thermodynamic equilibrium modelling can be applied. The analysis is based on an equilibrium constant for predicting the composition of the system with the initial reagents and the products of the reaction.

Using the law of mass action, equilibrium concentrations (C_{eq}) of the components can be found from equilibrium constants (K_d) and initial concentrations of the components (C_0, before the reaction starts). A general idea of a dynamic approach to solve the equilibrium state can be explained by a simple one-step model:

$$\mathrm{a[A]} + \mathrm{b[B]} \overset{\mathrm{K_d}}{\rightleftharpoons} \mathrm{[C]}$$

$$\frac{dx}{dt} = k(A - ax)^a (B - bx)^b$$

$$\frac{dx}{dt} = k'(C + x)$$

$$(A - ax)^a (B - bx)^b = \frac{k'}{k}(C + x)$$

where k and k' are chemical rate constants for the forward and backward reactions, $\frac{k'}{k} = K_d$ is an equilibrium constant. x is an active mass, which is an infinitesimally small number at the equilibrium. Starting from initial guess values for A_0, B_0 and $C_0 = 0$ concentrations, x can be found, next new values for N_i, P_i and NP_i can be calculated. For chemical reactions, concentrations of mixed components are used as initial guess values. Recalculation of components concentrations continues while x is higher than an accepted threshold value, then the system is assumed to reach the equilibrium. The procedure is summarised in Algorithm 1.

Experimental data measured by NMR provide information about N and P in solution (Fig.3.1a), initial values of the components are also known (P is constant, 75 μM, N varied from 0 to 75 μM). Using a dynamic approach explained just above, K_d can be found, however, the thermodynamic model for fitting is necessary. We tested the following models and the corresponding systems of equations:

One-step model

$$\mathrm{a[N]} + \mathrm{[P]} \overset{\mathrm{K_d}}{\rightleftharpoons} \mathrm{[D]}$$

$$\frac{(N - ax)^a (P - x)}{D + x} = K_d$$

N and P are protein concentrations in a dilute state, so observable by NMR, a is the stoichiometry between proteins in the reaction, D represents concentration of the complex $\mathrm{N_aP}$.

Two-step model

$$\mathrm{a[N]} + \mathrm{[P]} \overset{\mathrm{K_{d1}}}{\rightleftharpoons} \mathrm{[D]} + \mathrm{b[N]} \overset{\mathrm{K_{d2}}}{\rightleftharpoons} \mathrm{[D_2]}$$

$$\begin{cases} \frac{(N-ax-by)^a (P-x)}{D_1+x-y} = K_{d1} \\ \frac{(N-ax-by)^b (D_1+x-y)}{D_2+y} = K_{d2} \end{cases}$$

Algorithm 1 How to find equilibrium concentrations

Generate a model (aA + bB $\leftrightarrow$ C)
System of equations explaining the model ($\frac{(A-ax)^a(B-bx)^b}{(C+x)} = K_d$)
Solve equation and find x.
for `i in` $n_{attempts}$ **do**
    Random values for initial x (low=0, high=max(A,B))
    Find x using modified Powell method
    Check boundaries conditions:
    if x **then** < 0
        continue
    if A **then** < 0
        continue
    if B **then** < 0
        continue
    if C **then** < 0
        continue
    else
        break
    solution found
Update values
($A_{eq} = A_0 - x$
$B_{eq} = B_0 - x$
$C_{eq} = C_0 + x$)

This model comprises an additional step representing additional recruitment of N into the droplets with stoichiometry of b, so that D_2 corresponds to the complex $N_{a+b}P$.

Three-step model

$$a[N] + [P] \overset{K_{d1}}{\rightleftharpoons} [D] + b[N] \overset{K_{d2}}{\rightleftharpoons} [D_2] + c[N] \overset{K_{d3}}{\rightleftharpoons} [D_3]$$

$$\begin{cases} \frac{(N-ax-by-cz)^a(P-x)}{D_1+x-y} = K_{d1} \\ \frac{(N-ax-by-cz)^b(D_1+x-y)}{D_3+y-z} = K_{d2} \\ \frac{(N-ax-by-cz)^c(D_3+y-z)}{D_3+z} = K_{d3} \end{cases}$$

This model comprises an more step explaining N transfer into the droplets with a third affinity (K_{d3}) and stoichiometry (c), D_3 corresponds to the complex $N_{a+b+c}P$.

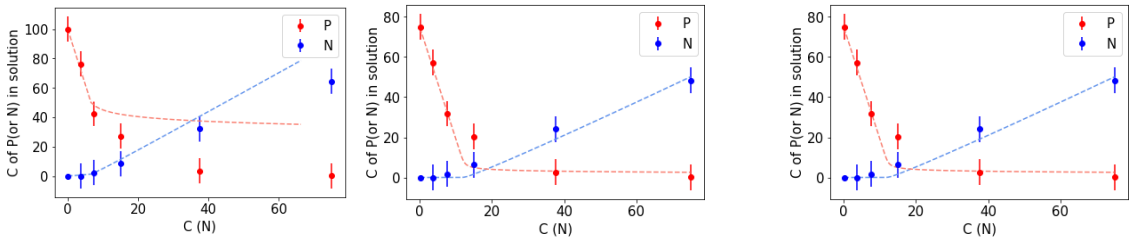
It was obvious that the "one-step model" would not provide a good solution of the fitting, as inspection of the data suggests that the stoichiometry between N_0 and P_0 is changing and this cannot be explained by one first-order chemical reaction. The best fit to the "one-step model" is presented in Fig. 3.1a. At low $N_0 : P_0$ ratio the fit follows experimental values, however at higher N_0 concentrations the ratio changes and both N_{eq} and P_{eq} concentrations in the solution are overestimated.

The two-step model includes an additional attraction of N into droplets, and as shown on Fig.3.1b better explains the behaviour of the system. Fitted parameters are:

$a = 1.16$
$b = 4.42$
$K_{d1} = 10\ \mu M$

$K_{d2} = 100\ \mu M$

During the first reaction N and P at 1:1 ratio are transferred into droplets, while during the second, four additional N molecules are attracted. It should be mentioned that the K_d of the first reaction is in the same range as the XD:N_{tail} affinity, and the second one is similar to P_{loop}:N binding constant.



(a) *(b)* *(c)*

Figure 3.1 – Fitting of the thermodynamic models. (a) 1 step model, (b) 2 step model, (c) 3 step model.

Including the third step did not improve the fit and fitting parameters for the step 2 and 3 are very similar:
$a = 3.27$
$b = 1.37$
$c = 1.14$
$K_{d1} = 50\ \mu M$
$K_{d2} = 200\ \mu M$
$K_{d3} = 200\ \mu M$

We propose that two elementary reactions mostly control droplet formation. However, using a thermodynamic description only a limited amount of reactions can be included. Also, spatial information is ignored. Fluctuations in local concentration are proposed to be one of the mechanisms of phase separation. Using molecular dynamics and statistical mechanics models, more complex chemical reactions can be simulated. There are several statistical mechanics softwares available for modeling phase separation which could be applied to describe N:P measles phase separating system. This we plan to do and compare the results of different approaches.

3.4 Role of P_{loop} interaction with N in droplets dynamics

As it was shown in section 3.2, inhibiting P_{loop}:N interaction did not suppress droplet formation. However, macrodynamic characteristics of droplets changed: FRAP was significantly slower for the mutant in comparison with the wild type. We repeated the experiment using $P_{50}N_{525}$ with $P_{304\text{-}507}$ VLK->AAA (Fig. 3.2) then added P_{loop} into the solution and surprisingly, the recovery rate increased again. We propose that the P_{loop}:N weak interaction perturbs a stronger P:N interaction and this makes the complex more dynamic and allows faster protein exchange between phases.

To better understand how exactly P_{loop} interacts with N and influences droplet dynamics, further investigation is required. In particular, measurements of protein state in droplets could reveal

roles of both interactions on the formation of phase separation.

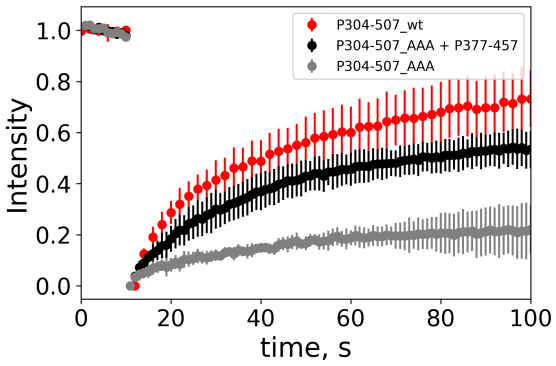


Figure 3.2 – FRAP profile of fluorescently labelled $P_{50}N_{525}$ in droplets formed with (red) $P_{304\text{-}507}$, (grey) $P_{304\text{-}507}$VLK->AAA, (black) $P_{304\text{-}507}$VLK->AAA and P_{loop}.

3.5 Interaction with importin α5

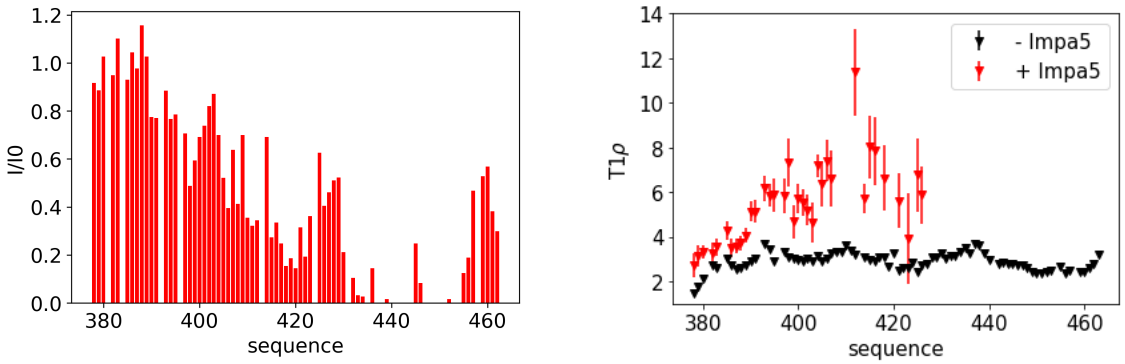
Apart from N and P, droplets are known to accumulate other viral and cellular proteins (client proteins). A review on proteins which were found in viral factories and their possible functions is presented in section 1.4. Client proteins are attracted into droplets through interactions with the scaffold or with other client proteins.

Phosphoprotein is known to interact with multiple cellular factors via its disordered tail. P_{loop} discussed above was not studied in detail previously. Recently, we found a new interaction which involves P_{loop} and importin α, a cellular factor responsible for nuclear import. Importin α recognises a specific sequence on target proteins called nuclear localisation signal (NLS). The human importin α family comprises at least seven isoforms, here we used importin α5 with a truncated disordered domain which is responsible for importin β binding. Classical NLS are proposed to be K-K/R-X-K/R, which is not present in P_{loop}, however, many exceptions are known in the literature and the common feature for all the known examples is a cluster of basic amino acids.

The intensity ratio profile presented on Fig. 3.3a shows a strong interaction between the C-terminal half of the P_{loop} with the folded part of importin α5 as the intensity of the peaks decrease due to the broadening. Four lysines spaced by four amino acids are present in the region where the intensity drops which we assume to be responsible for the interaction with importin. Relaxation profiles of P_{loop} alone and in the presence of importin on Fig. 3.3b show elevation in the N-terminal part of P_{loop} suggesting additional interaction, which, appear to be weaker. Residues 411-412 are lysines which could also bind to importin. As the intensity of the C-terminal half of P_{loop} dropped, the information about relaxation rates for these residues is absent.

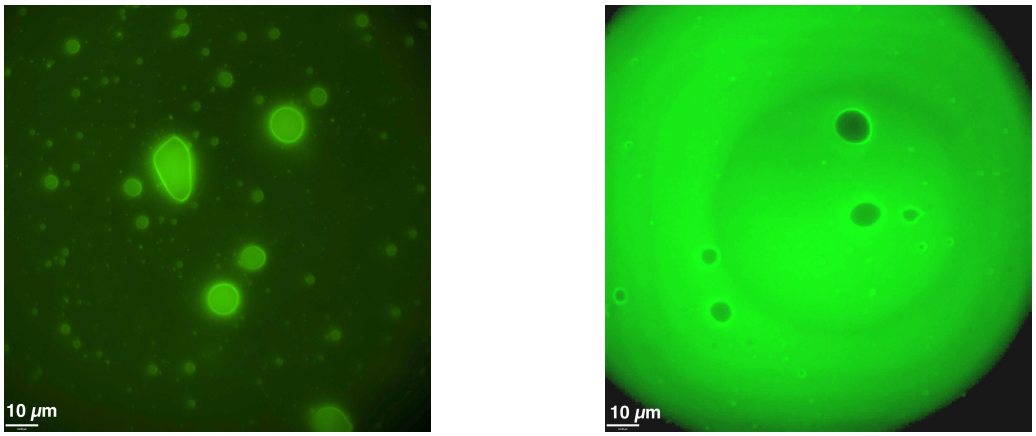
Using fluorescently labeled importin we could show its accumulation into pre-existing droplets formed with $P_{304\text{-}507}$ and $P_{50}N_{525}$ (Fig. 3.4a). In order to demonstrate some element of selectivity for this protein, we used YFP as a control (Fig. 3.4b), it is excluded from the droplet.

Importin α1 is know to be present in stress granules together with others nuclear transport proteins. Their accumulation in SGs during stress conditions is proposed to inhibit nuclear transport. I speculate that viral factories could absorb nuclear transport factors in order to sequester transcription factors involved in immunity import into the nucleus. In addition, P_{tail} is known to interact with STAT protein, which is itself a transcription factor, activating the cellular innate immune system [66]. Whether it is attracted into droplets has yet to be tested.



(a) *(b)*

Figure 3.3 – P_{loop} interaction with importin α5. (a) Intensity ratio of peaks on an HSQC spectra between P_{loop} and P_{loop} with importin α5 at 1:1 ratio (100 µM). (b) $R_{1\rho}$ profile of P_{loop} and P_{loop} with importin α5 at 1:1 ratio (100 µM).



(a) *(b)*

Figure 3.4 – Fluorescence microscopy imaging of $P_{50}N_{525}$:$P_{304\text{-}507}$ droplets with (a) importin α5-FAM, (b) YFP. Fluorescent proteins were added into pre-formed droplets.

The role of accumulation of importin into droplets was not studied yet and further investigations are required to characterize the interaction and how importin influences droplet behaviour.

3.6 N^0P chaperoned with the complete P_{tail}

As it was shown in section 3.2 Fig.1, $P_{304}N_{525}$ with $P_{304\text{-}507}$ also triggers phase separation. Such mixtures contain the full N and full P, truncated between P_{tail} and the tetramerization domain. P_{tail} is known to interact with cellular proteins and, probably, can act as a hub, attracting client proteins into droplets.

While characterizing droplets formed with $P_{304}N_{525}$ and $P_{304\text{-}507}$ we could show that they exhibit liquid behaviour: fuse, have spherical shape, and protein exchanges between phases. We compared the assembly process between $P_{50}N_{525}$ and $P_{304}N_{525}$ and it appeared, that $P_{304}N_{525}$ hardly assembles, either alone or in the presence of $P_{304\text{-}507}$ (Fig. 3.5), at least using similar conditions as for $P_{50}N_{525}$. $P_{304}N_{405}$ was also tested for the ability to assemble and behaved similarly to

$P_{304}N_{525}$.

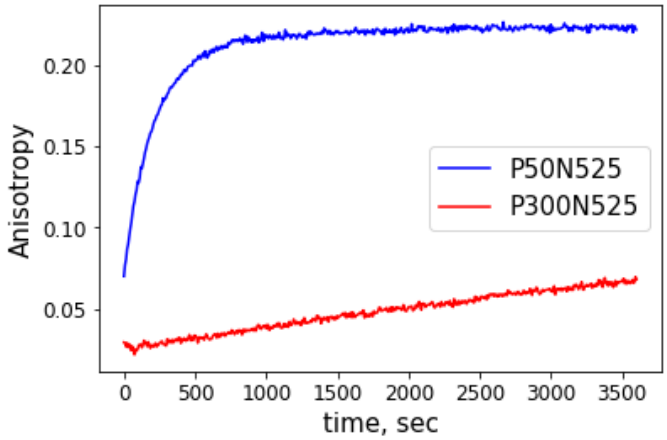


Figure 3.5 – Kinetics of fluorescence anisotropy of RNA-FAM after addition of $P_{50}N_{525}$ and $P_{304}N_{525}$.

In the introduction (section 1) interactions between P_{tail} and N_{core} were discussed. There are two interactions: $P_{1\text{-}37}$ and $P_{180\text{-}190}$ involved in chaperoning. The second interaction is placed on the opposite side of N_{core}, suggesting that P_{tail} wraps around N_{core}, while remaining disordered. Possibly, the region of P_{tail} between interacting sites shields the RNA binding surface, thus, inhibiting assembly process.

How the full length complex of NP assembles in the cell is not clear, probably an additional mechanism of regulation exists which controls the efficiency of N chaperoning by P_{tail}.

3.7 Conclusion

P is known to be an essential co-factor for viral replication and was speculated to be involved in the spatial organisation of the replication complex. Various models were proposed about how P can bring L into a closer proximity to nucleocapsids. Also, N interaction with L and P was speculated to be essential to encapsidate newly synthesised RNA, so it can be directly coated. In particular, XD:N_{tail} interaction was thought to be involved in a cartweel model. Indeed, this interaction is essential for droplet formation and thus vital for replication. The discovery of viral factories reveals a physical mechanism whereby all the components can be concentrated and provides an explanation for the the essential nature of P in viral transcription and replication processes.

The ratio between N and P in droplets depends on a total concentration of protein in both phases, which can possibly play role in regulation of protein concentration in cytoplasm. Indeed, the transition from transcription to replication in many single-strand negative (and more recently positive) sense RNA viruses has long been known to depend on the total concentration of the N protein. This, allied with the known transcription gradient of N and P in MeV, suggests that the formation of liquid droplets may represent, or simply control the previously unknown switch in polymerase function. Similar effects were shown for ddx4 protein. The increase on total concentration triggered the formation of more droplets, however, the concentration of a protein in the rest of the cytoplasm was constant. In addition, the noise of protein concentration fluctuation was significantly reduced in the presence of droplets [113]. To conclude, an additional role of viral droplets could be the control of viral proteins concentration in the cytoplasm what can be used to hide from cell-mediated immune responses, for instance. It was proposed that viral factories

can silence replication and transcription and reactivate after a prolonged period and cause another round of replication and new virion production [56].

Though P_{tail} was shown not to be sufficient for phase separation, it contains multiple binding sites with other proteins and might act as a hub which attracted client proteins inside droplets for instance STAT1/2. In case of STAT1/2 proteins, it was shown to inhibit alpha/beta interferon (IFN-α/β) and IFN-γ signaling [48, 77]. N is also known to interact with cellular proteins. What is interesting is that binding site to the same cellular protein can be switched between N and P for different viruses, for instance importin-binding nuclear localization sequence in rabies is present of C-terminus of P [179], and for canine distemper virus and rinderpest virus are on N [187].

Next, N protein assembly in the presence of P_{full} has to be investigated, alone and in droplet. Also, protein states and interactions inside droplets were not studied yet - requiring either optimisation of conditions for NMR or change of method for example, smFRET. Next, the effect of L protein on phase separation is intriguing and the effect of condensation on a polymerisation reaction and modulation of droplets in the presence of RNA (genomic or mRNA) are fascinating.

4
Phosphorylation

Measles N and P are both known to be phosphorylated. Here, *in vitro*. description of P_{tail} phosphorylation is presented. New phosphorylation sites were identified and their kinetics were measured using NMR (section 4.2). Next, the impact of phosphorylation on the assembly was discovered and studied using NMR, fluorescence anisotropy and negative stain microscopy (section 4.3). To understand the mechanism how phosphorylation controls the assembly we studied the effect of phosphorylation on P_{tail} backbone dynamics using NMR relaxation (section 4.4). Finally, in section 4.5 N phosphorylation and possible impact on nucleocapsid assembly is briefly discussed.

Some of this work was carried out by Sigrid Milles before and during my thesis.

4.1 Introduction

Measles virus phosphoprotein received its name for the ability to be phosphorylated at multiple sites. Two of them are known in the literature and are present on P_{tail} (S86 and S151) [43], [23] several more were predicted.

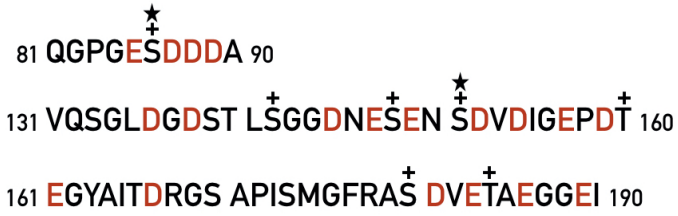


Figure 4.1 – Regions of P_{tail} which are known or predicated to be phosphorylated. Acidic amino acids are shown in red. ⋆ indicates phosphorylation sites (S86 and S151) [43]. + indicates amino acids predicted to be phosphorylated by CK2.

These serines and threonines are phosphorylated by Casein Kinase 2 (CK2), a protein kinase which recognises S/T-x-x-E/D sequence and phophorylates the first S/T residue. Acidic residues (E/D) could be substituted with a phosphorylated residue, so the actual recognition site is S/T-x-x-E/D/pX. Based on this, eight residues are predicted to be phosphorylated (Fig. 4.1), including those known in the literature: T49, S86, S142, S151, T160, S180, T184 and S148 which is dependent on S151 phosphorylation. All phosphorylated residues are surrounded by acidic residues, three of them are predicted to appear in a highly negatively charged region that we have previously termed the acidic loop of P_{tail} (Fig. 4.1).

Although, several phosphorylation sites were proven to exist experimentally, the role of this post-translational modification is unknown.

4.2 P_{tail} phosphorylation *in vitro*

First, we tested whether P_{tail} can be phosphorylated *in vitro*. by CK2. Phosphorylation can be studied by NMR. On an HSQC spectrum peaks which correspond to phosphorylated amino acids have a specific chemical shifts and appear, in particular in the 1H dimension. On Fig.4.2 an overlap of P_{tail} ^{1}H-^{15}N HSQC spectrum before and after phosphorylation is presented. One can notice five new peaks appearing in the region between 1H 8.5 ppm and 9 ppm and several peaks disappearing in the area which corresponds to the serine-threonine region (labelled on Fig.4.2). There are more than five peaks which are affected by phosphorylation as not only the phosphorylated amino acids change their chemical environment but also their neighbours. Changes in the position for some of them are highlighted with arrows on an HSQC.

New peaks were reassigned using 3D experiments as explained in Chapter 2. Phosphorylated serines will be indicated by the name with "P" after the number, for example, S151P.

Next, we studied kinetics of phosphorylation using NMR. For this, a series of SOFAST ^{1}H-^{15}N HSQC spectra of P_{tail} were recorded during 36 hours after addition of CK2 and essential components for phosphorylation (see Chapter 2). Each spectrum was recorded during 36 minutes. Following the intensity of individual peaks allowed us to study the kinetics of phosphorylation form the perspective of each amino acid individually. On Fig.4.10 a zoom into the region of ^{1}H-^{15}N HSQC where phosphorylated amino acids appear is shown for several time points during the phosphorylation. S151P peak appears first, then S86P and S148P start to appear. Phosphorylation

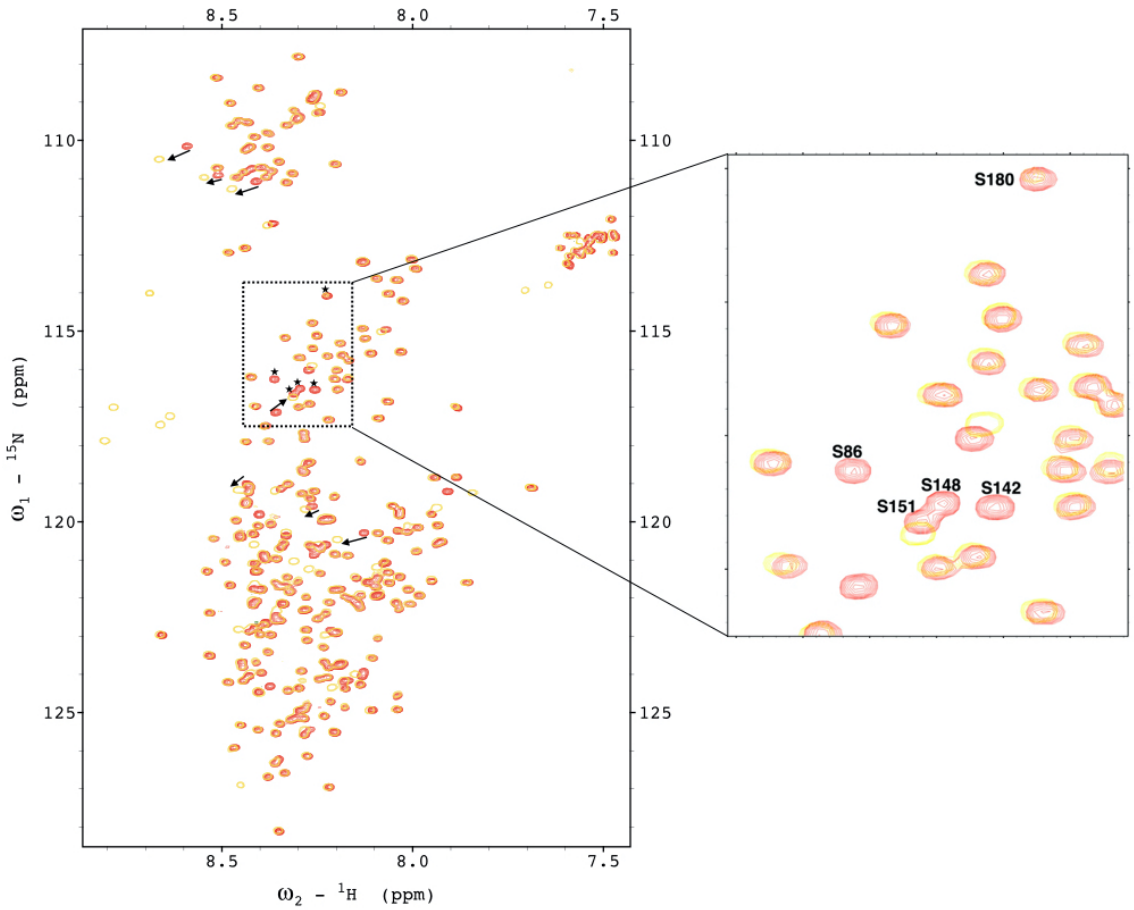


Figure 4.2 – 1H-^{15}N HSQC spectra of P_{tail} non-phosphorylated (red) and phosphorylated (yellow).

of S148 influences S151P, two peaks corresponding to it (S151P and S151PP): One with non-phosphorylated (S151P) and the other with phosphorylated S148 which also affects the chemical shift of S151. S151P intensity increase initially, then, as it activated S148 phosphorylation, its intensity is decreasing, while S151PP grows.

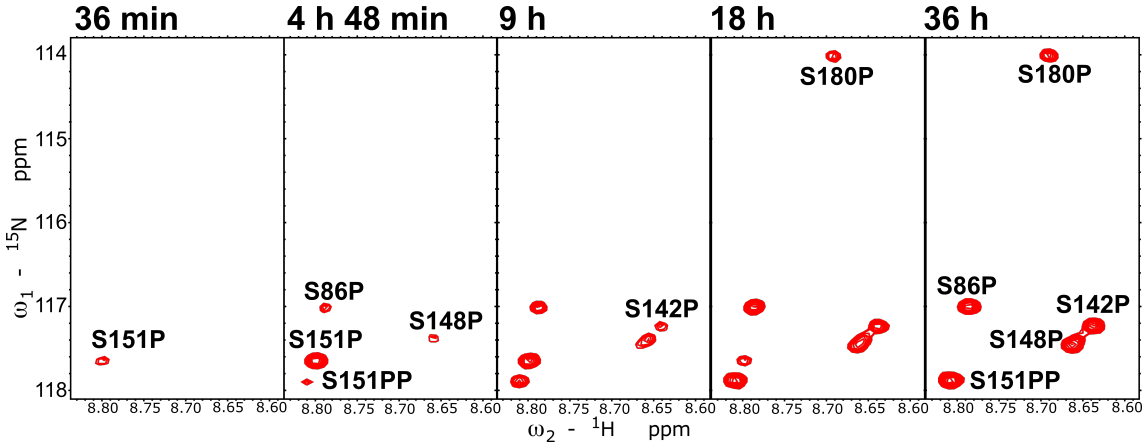


Figure 4.3 – ^{1}H-^{15}N HSQC spectra of P_{tail} phosphorylated amino acids at different time points.

On Fig.4.4 the profiles of intensity decrease for non-phophorylated serines and intensity growth for phosphorylated ones are shown. Phosphorylation rates were extracted by global fit of appearing and disappearing peak with single exponent except for S151.

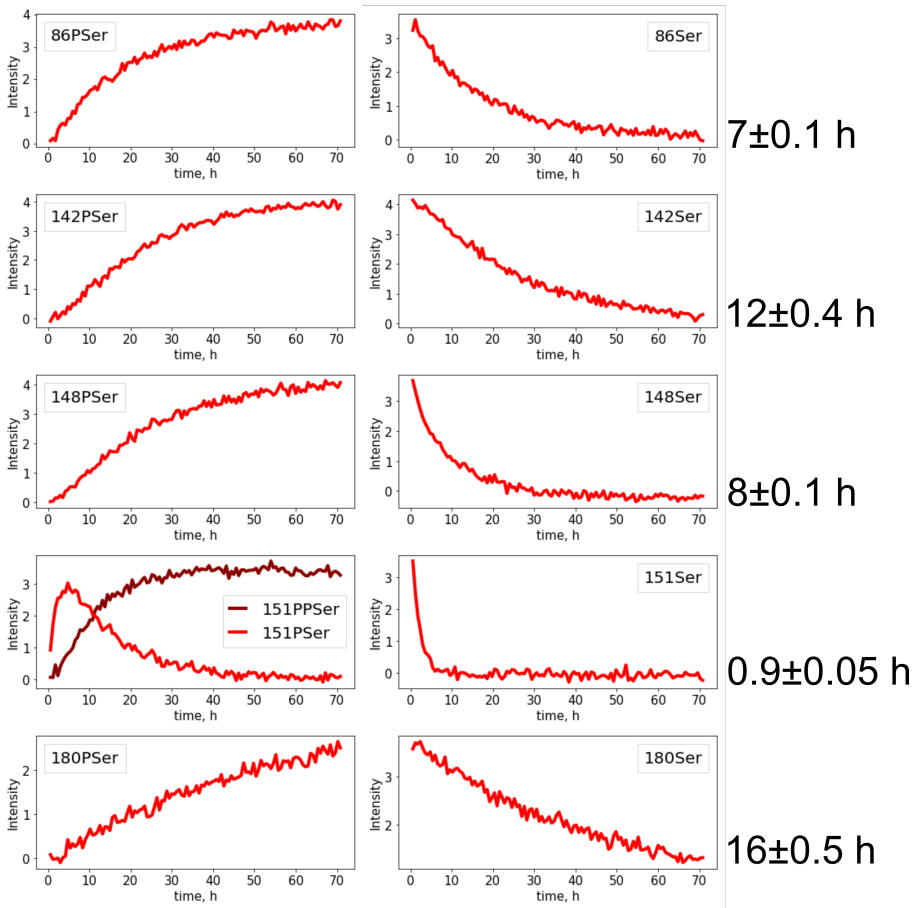


Figure 4.4 – ^{1}H-^{15}N HSQC spectra of P_{tail} phosphorylated amino acids at different time points.

To conclude, two phosphorylation sites which are known from *in vivo* studies are phosphorylated *in vitro*. In addition, we identify two new phosphorylation sites. Previously, three phosphorylation sites were identified using mutagenesis: S86, S151 and S180, the last had a minor phosphorylation [43]. The mutation of S151 affects S148 phosphorylation which may explain why this site was not previously found. S142 is slower than S151, S148 and S86 and could be dependant on one of phosphorylation sites. This hypothesis can be tested by phosphorylation of the S151A mutant.

Even though mutation of S86 and S151 were shown to affect transcription [198] the role of phosphorylation and the mechanism of regulation have not been previously reported and its identification was one of the goals of the current research.

4.3 Assembly of $P_{304}N_{525}$

We investigated whether phosphorylation of $P_{304}N_{525}$ can play a role in nucleocapsid assembly. As shown in section 3.6, in the presence of P_{tail} bound to N_{core}, assembly on RNA is extremely inefficient, suggesting that P_{tail} interferes with the assembly process. We repeated the fluorescence anisotropy experiment with RNA-FAM for $P_{300}N_{525}$ which was phosphorylated with CK2 and then purified from the kinase, ATP and $MgCl_2$, using gel filtration, so that all the compared samples were in the same buffer condition.

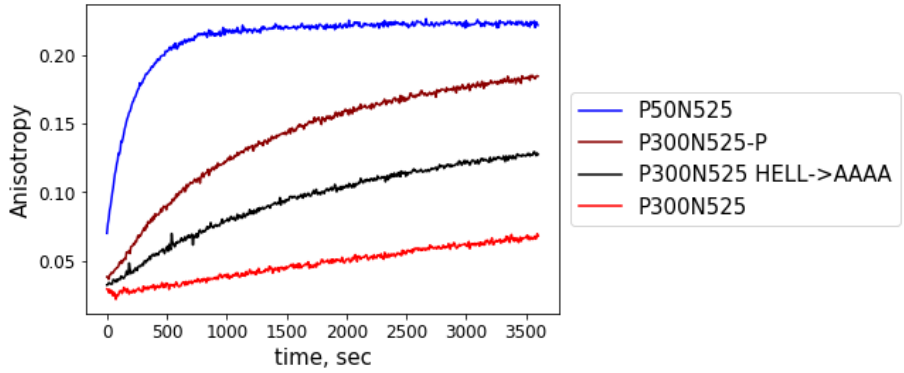


Figure 4.5 – Kinetics of RNA-FAM fluorescence anisotropy for $P_{50}N_{525}$, $P_{304}N_{525}$, phosphorylated $P_{304}N_{525}$ and $P_{304}N_{525}$ with a mutation of HELL motive.

The assembly rate of phosphorylated construct is significantly higher, however still lower than the one of $P_{50}N_{525}$ (Fig. 4.5). In order to prove that anisotropy increases due to the assembly process and not because of aggregation or any other process, we tested the sample with NS EM (Fig. 4.6).

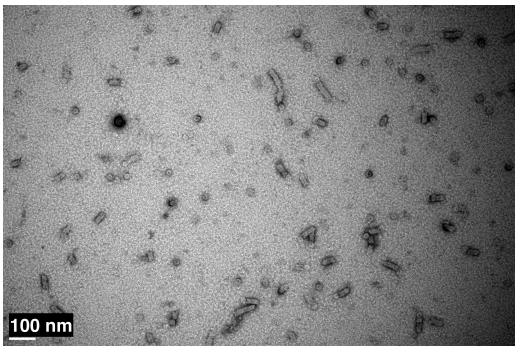


Figure 4.6 – NS imaging of phosphorylated $P_{304}N525$ assembled with polyA RNA.

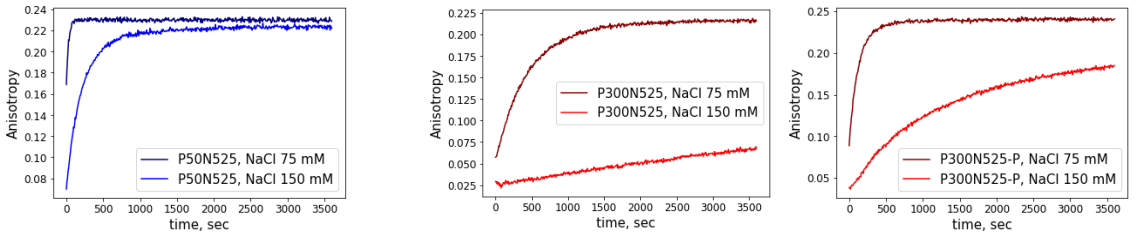
It should be mentioned that $P_{304}N_{525}$ with a mutated HELL motif, so that the second interaction is inhibited, has an ability to assemble, but the efficiency is not high (Fig. 4.5).

The mechanism how phosphorylation can control assembly and RNA binding is not clear. Adding phosphorous groups to serine adds negative charges to the chain and can influence electrostatic interactions. As all phosphorylation sites are surrounded with acidic residues, phosphorylation of these regions makes it even more acidic. In order to test whether strength of electrostatics along P_{tail} can control nucleocapsid assembly we repeated fluoresence anisotropy of RNA at lower NaCl concentration (75 mM instead of 150 mM).

For all tested constructs the assembly was faster at lower salt, even for $P_{50}N_{525}$, probably, due to the electrostatic nature of RNA:N interaction. $P_{304}N_{525}$ was assembling at low salt even without phosphorylation (Fig. 4.7). The rate of the assembly of the HELL mutant is higher than the one of the wild type which indicates that not only the acidic loop controls the assembly but HELL interaction is involved as well (Fig. 4.8).

Using NS EM we confirmed nucleocapsid assembly for $P_{304}N_{405}$: At low salt and with phosphorylated P_{tail} (Fig. 4.9). No major difference was detected between these three samples, however a more detailed analysis is required. NS microscopy of $P_{304}N_{525}$ construct at different conditions should be imaged as well.

The kinetics of the assembly can be followed with NMR as was previously explained in Chapter 1 which also provides information on how individual residues are involved in the nucleocapsid



(a) *(b)* *(c)*

Figure 4.7 – RNA-FAM fluorescence anisotropy kinetics comparison at law NaCl concentration (75 mM)

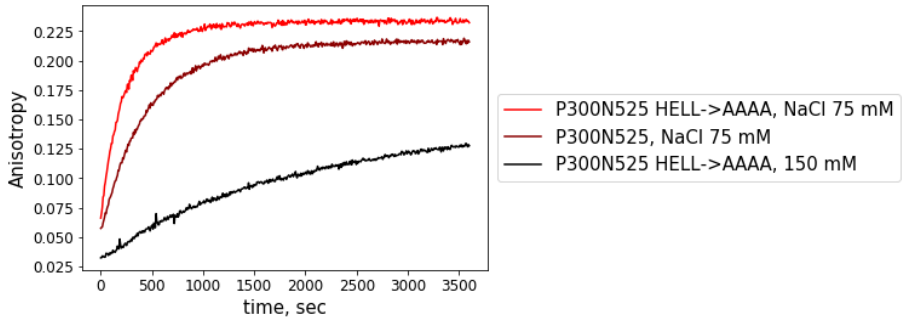
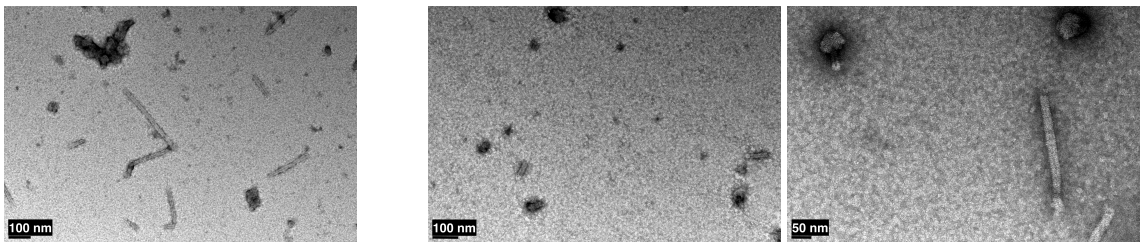


Figure 4.8 – Kinetics of RNA-FAM fluorescence anisotropy for $P_{304}N_{525}$ and $P_{304}N_{525}$ HELL->AAA and low and high salt.



(a) *(b)* *(c)*

Figure 4.9 – NS microscopy of $P_{304}N_{405}$ (a) at 75 mM NaCl, (b) phosphorylated at NaCl 150 mM, (c) phosphorylated at NaCl 75 mM.

formation. First, we studied $P_{304}N_{525}$ assembly at 75 mM NaCl without phosphorylation. HSQC spectra before and after assembly are presented on Fig. 4.10. During the assembly, the intensity of P_{tail} peaks increased, most effect was observed for residues 1-50 which were not detectable before assembly, as shown on Fig. 4.11 (preassambled intensities (Pre) are shown with red line, assembled intensities in yellow bars). This indicates that $P_{1\text{-}50}$ releases from N_{core}. The intensity of HELL motif was not recovered after the assembly, probably it stays bound to the N_{core} or in an exchange with those particles which did not assemble. The kinetics of intensity change during the assembly for several peaks are presented in Fig. 4.10 (b), and the distribution of extracted rates are presented on Fig. 4.10 (c), showing the mean value to be around 0.7 h.

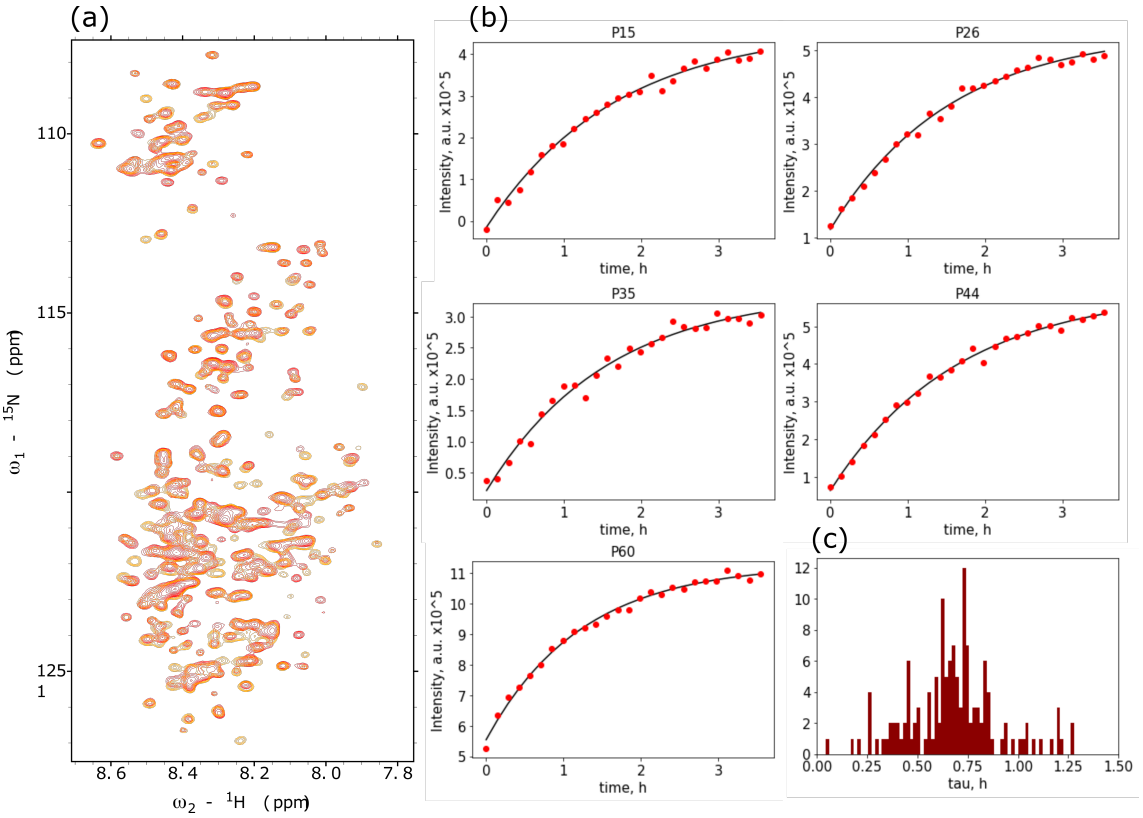


Figure 4.10 – (a) HSQC spectra of $P_{304}N_{525}$ before assembly (red) and after (yellow) at 75 mM NaCl triggered with polyA RNA. (b) Kinetics of extracted peak intensities during the assembly. They were fitted with exponentional curve. (c) Histogram distribution of assembly rates obtained by fitting individual peak intensities kinetics.

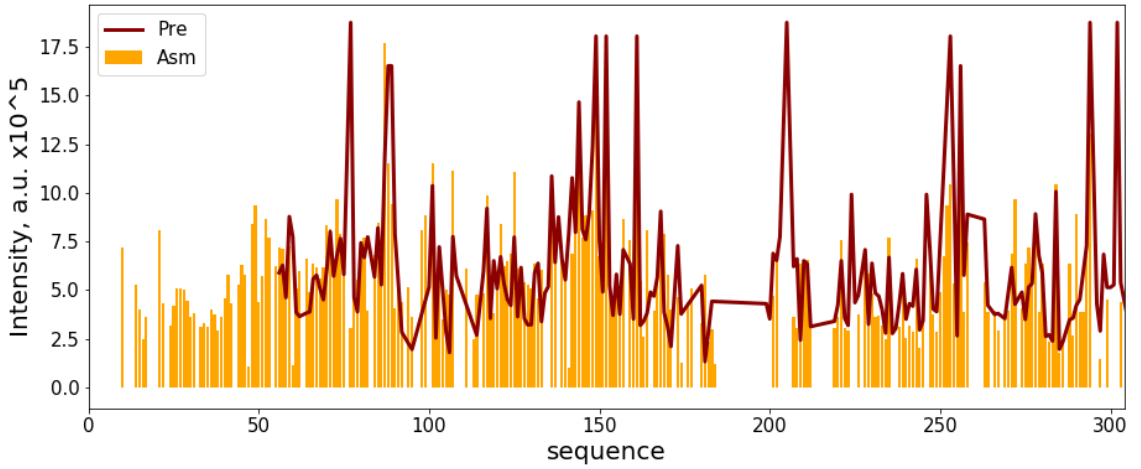


Figure 4.11 – Intensity of peaks of P_{304} before and after $P_{304}N_{525}$ assembly.

4.4 Coupled phosphorylation and assembly

Next, we studied the assembly of phosphorylated constructs. It should be mentioned that in the presence of N (N_{405} or N_{525}) phosphorylation is less efficient and S180 was not observed to be phosphorylated (Fig. 4.12). S180 is in a close proximity to the HELL binding motif and is likely less accessible to the kinase because of the interaction with N_{core}.

As the spectra of $P_{304}N_{405}$ have fewer overlapped peaks (contains 125 fewer peaks), we used

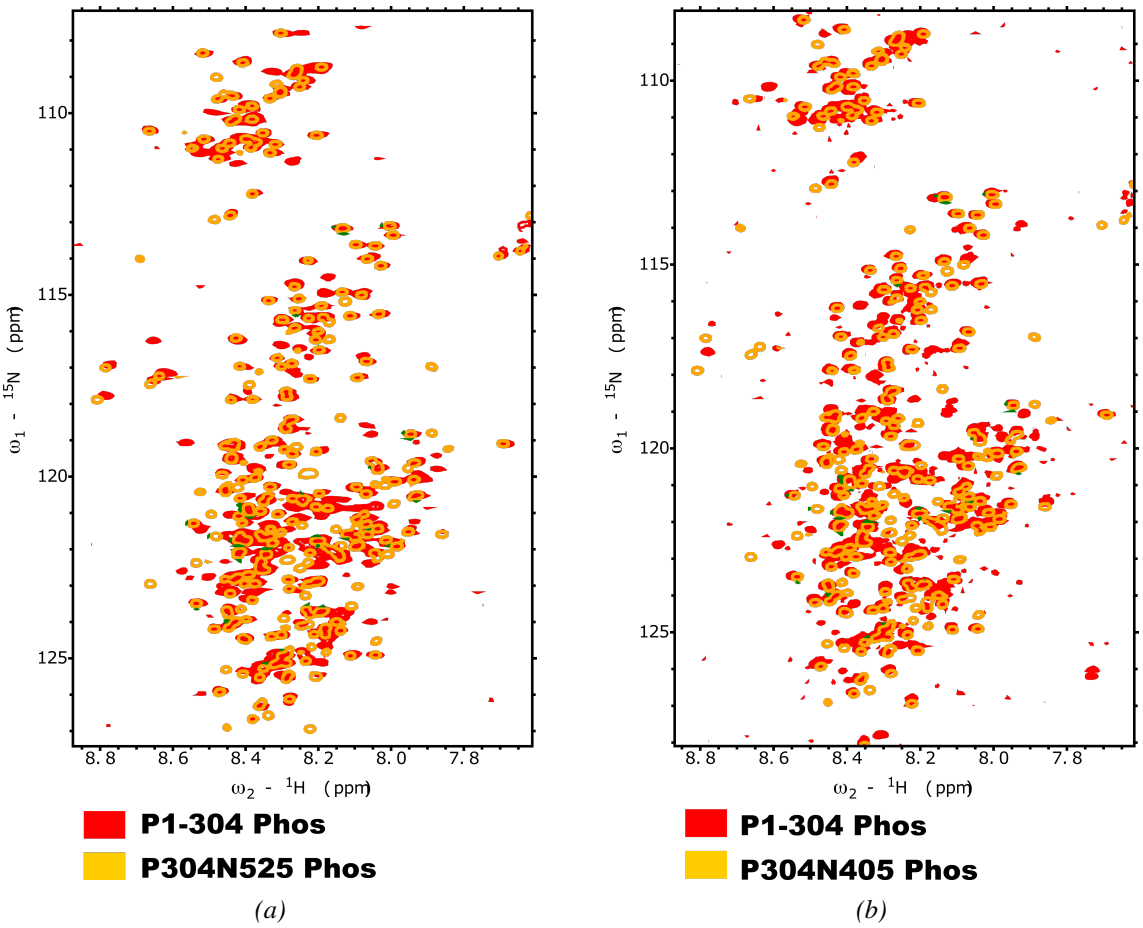
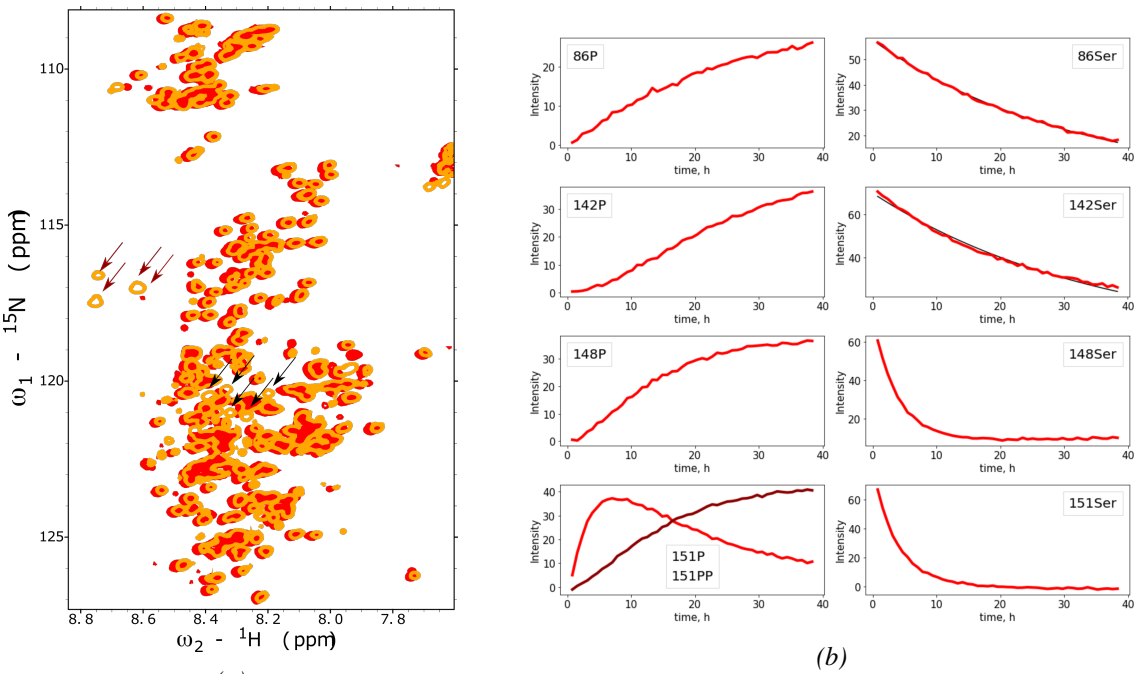


Figure 4.12 – HSQC spectra of (a) $P_{304}N_{525}$ phosphorylation and (b) $P_{304}N_{405}$ in comparison with P_{304}.

this construct to track simultaneous phosphorylation and nucleocapsid assembly. For this, RNA and all the components essential for phosphorylation were added to the protein and transferred into the NMR tube. CK2 kinase was added just before recording a series of HSQCs.

The comparison of 0 point and the last recorded HSQCs (after 40 hours) are presented on Fig. 4.13. Red arrows highlight phosphorylated peaks, black indicates P peaks which appear as a result of the assembly. The kinetics of phosphorylated peaks are on Fig. 4.13. Ser180 was not phosphorylated. The decay of non-phosphorylated peaks was fitted with a single exponentional curve, Ser148 was not fitted with a single exponent as its phosphorylation is dependent on Ser151 phosphorylation.

The assembly process was followed by the increase in the intensity of peaks which correspond to P residues from 1 to 50 (Fig. 4.14). The extracted rates are in a range of 0.1 - 0.5 h^{-1}. The only residue whose rate of phosphorylation is faster is S151, which suggests that assembly activation is dependent only on the phosphorylation of S151. If phosphorylation is regulating assembly, as appears to be the case, this suggests that the rate-limiting phosphorylation site is S151.



(a) *(b)*

Figure 4.13 – HSQC spectra of (a) $P_{304}N_{525}$ phosphorylation and (b) $P_{304}N_{405}$ in comparison with P_{304}.

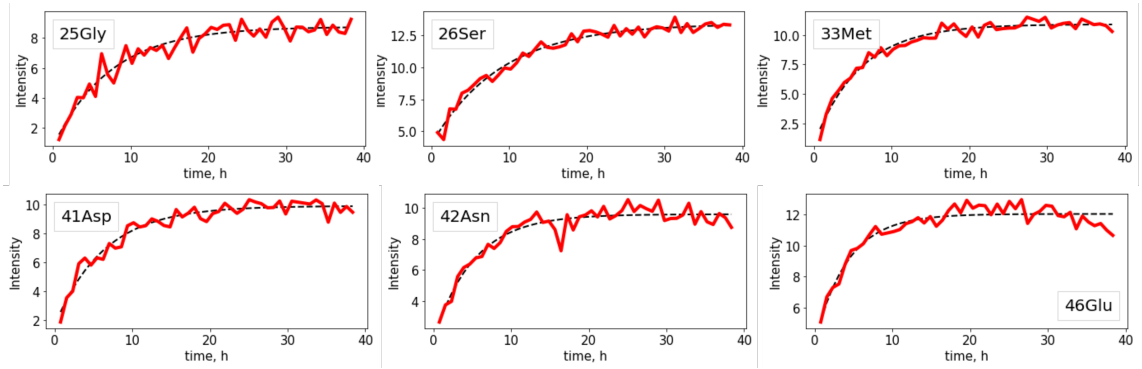


Figure 4.14 – Intensity of peaks of P_{304} during the assembly.

4.5 Effect of phosphorylation on backbone dynamics

The effect of P_{tail} phosphorylation on the dynamics of the chain studied by measuring relaxation experiments: R_1, $R_{1\rho}$ and het-NOE. The comparison of relaxation rates between P_{304} before and after phosphorylation are on the Fig. 4.15.

Although, the overall dynamics of the chian are essentially identical, there are local differences around the phosphorylation sites. Around Ser86 there is a slight elevation in $R_{1\rho}$ profile. The most significant difference is around residues 140-150, not surprisingly as there are three phosphorylation sites. R_1, $R_{1\rho}$ and NOE values are increasing what indicates an influence on backbone dynamics. In addition, the difference in NOE profile can be noticed for the region around the residue 190.

Relaxation profiles of P_{tail} in the complex with N_{525} in a phosphorylated and non-phosphorylated state were also compared (Fig. 4.16). Residues around phosphorylation sites Ser142, Ser148 and Ser151 are elevated in the phosphorylated form which corresponds to the observation in the free P_{tail}. We also measured $P_{304}N_{525}$ relaxation at low salt conditions. The main difference between

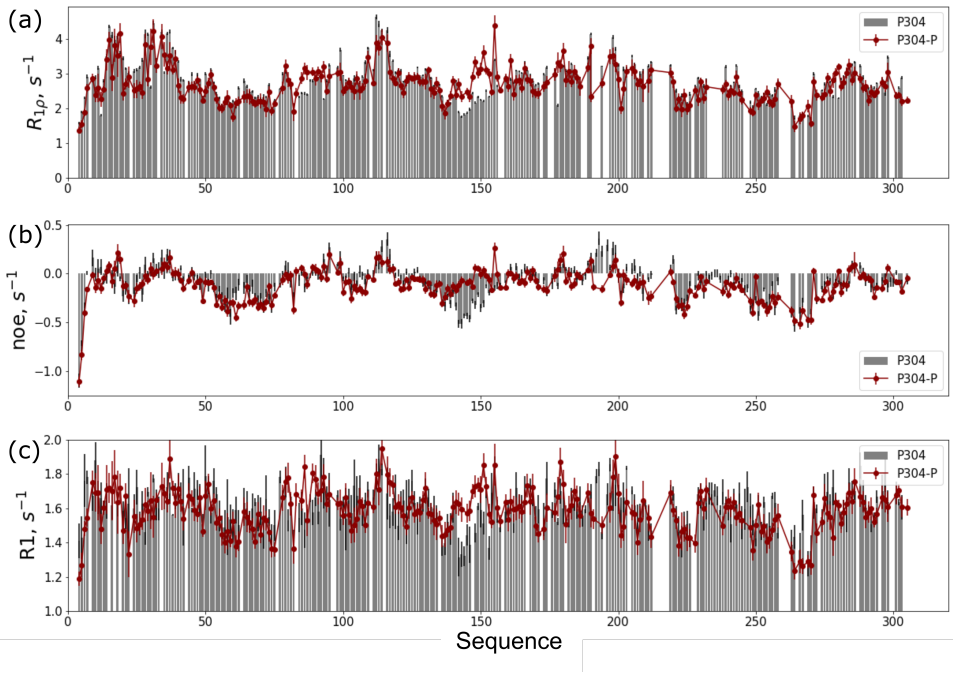


Figure 4.15 – Relaxation profiles (a) R_1, (b) $R_{1\rho}$ and (c) het-NOE for P_{tail} with and without phosphorylation.

150 mM NaCl and 75 mM NaCl is in the region of acidic loop which also contains a triple phosphorylation site (Fig. 4.16). Analogous to phosphorylation at high salt, lowering salt concentration elevates $R_{1\rho}$ values. However, in order to fully understand how charges influence P_{tail} dynamics and control chaperoning, more experiments have to be acquired.

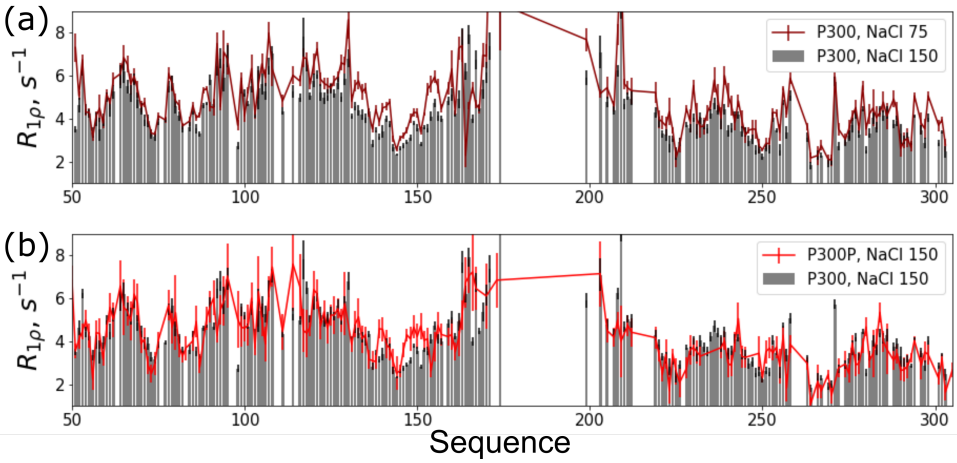


Figure 4.16 – Relaxation profiles (a) R_1, (b) $R_{1\rho}$ and (c) het-NOE for P_{tail} with and without phosphorylation.

4.6 Nucleoprotein phosphorylation

There are two phosphoryaltion sites along N_{tail} which are known in the literature (S497 and S510) and one in N_{core} T279. The last one is known to be involved in nucleocapsid assembly and is positioned close to the first binding site ($P_{1\text{-}37}$: under the kink between two helices (Fig. 4.18).

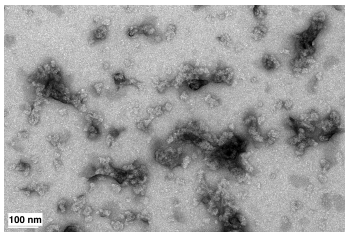


Figure 4.17 – NS imaging of N_{core} T279E after TEV cleavage.

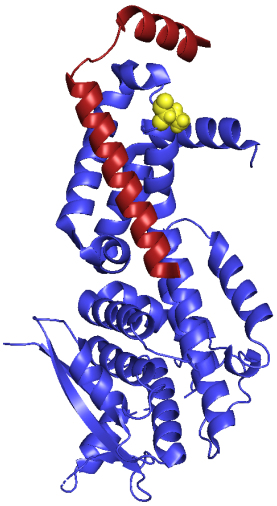


Figure 4.18 – Structure of N^0P complex. T279 is shown in yellow, P in red, N_{core} in blue.

A phosphorylation mimic mutant T279E was produced to study the role of this phosphorylation site and it appeared that after TEV cleavage between P1-50 and N_{core} this mutant is not stable in a monomeric form. NS imaging show some oligomeric structures in which flexible and badly ordered nucleocapsids may be present (Fig. 4.17).

The kinase which phosphorylates this site is unknown, the epitope T-X-Y-X is present in measles and Nipah could be a phosphorylation site on TTBK1. However, phosphorylation predictions propose high probability of non-specific phosphorylation including with CK2. In order to exclude its phosphorylation with CK2 a control experiment has to be done. Otherwise, a T279A mutant can be used to confirm the role of N_{core} phosphorylation in assembly control.

4.7 Conclusion

We confirmed P_{tail} phosphorylation with CK2 *in vitro*. and, in addition, identified new phosphorylation sites. Using NMR we followed phosphorylation kinetics of P_{tail} residues separately, allowing us to distinguish the phosphorylation rates for each of them and determine the residue which is phosphorylated first: S151.

Previously, in cell studies show the effect of phosphorylation on viral transcription efficiency (see Chapter 1, section 1.3.6). Here, we show that P_{tail} phosphorylation is required for the initiation of the NC assembly *in vitro*. The phosphorylation sites are surrounded by acidic amino acids, to which the phosphate group brings additional negative charge which, probably, modulates the chaperoning strength. I speculate that phosphorylation weakens P_{tail} association with N_{core} and allows RNA binding that, in turn, triggers NC assembly. In order to understand the mechanism of NC assembly regulation the complex should be studied and compared in phosphorylated and non-phosphorylated forms. Using NMR we could not observe any signal coming from N_{core}, probably, due to the big size and fast relaxation of this domain. How P_{tail} phosphorylation affects N_{core} should be studied by direct observation of N_{core}. Methyl-transverse relaxation optimized spectroscopy (methyl-TROSY) is a NMR technique which allows to observe particles up to 1 MDa. We plan to apply this technique to study phosphorylation effect on N_{core} structure, dynamics and the exposure of the RNA-binding site. In addition, we suggest to perform SAXS measurements to examine long-range changes of the complex caused by phosphorylation.

Finally, it should be mentioned that MeV phosphorylation also affects the nature of phase separated viral factories [18]. Using *in vitro* formed droplets we plan to test the effect of P_{tail} phosphorylation on the biophysical properties of N:P droplets and possibly link it with the NC formation.

5

Ntail dynamics in droplets

In the current chapter I aim to compare the dynamics of an IDP in the dilute and condensed phases. In order to achieve this I needed to first identify a system that phase separates and gives high quality NMR spectra in both phases. For this MeV N_{tail} was used. I start with a classical characterisation of a liquid behaviour and determination of a phase diagram which was fit with a mean-field Flory-Huggins model. Next, sample preparation for studying condensed N_{tail} by NMR is explained. Protein secondary structure was compared in both states and the protein was shown to remain disordered in the droplet. Next, backbone dynamics using spin relaxation was studied in both states. Using a model-free approach three time scale contributions to the molecular motions were separated and compared between the dilute and condensed phases. Finally, long-range conformational changes are predicted using a modified flexible meccano approach. Calculated N_{tail} structures are more extended in a crowded environment, although dihedral angle distribution do not change significantly between two states.

5.1 Introduction

Cellular biochemical reaction have been studied *in vitro* for decades and thousands of enzymatic reactions characterised in buffers designed to reproduce physiological conditions. However, recent discoveries show that the environment in cell may differ from these, especially due to the crowded conditions [78]. In addition, it appears that many biochemical reactions occur in membraneless organelles where the concentration of the reagents can be several orders of magnitude higher than the average concentration in cell [21]. Thus, understanding of biochemical reactions in biologically relevant environment requires understanding of the impact of the droplet environment. This in turn requires protein characterisation in biocondensates which can differ from the dilute state. IDPs are common members of condensates, and their functions are determined by their dynamics, which is highly sensitive to the changes in the environment, especially viscosity. Following the dynamics-to-function paradigm, the problem reduces to the understanding of how protein dynamics changes in biocondensates and modulates protein function.

Protein dynamics in supercrowded environment was studied previously for several proteins [4]. However, such studies still remain exceptional. Protein backbone motions were shown to be restricted (see below), however, a quantitative description and comparison with a dilute protein state is required in order to understand the mechanisms of phase separation. Besides, while this phenomenon is known for a long time in polymer physics field, the physics of it is not known. Two main models are presented in the introduction. Working with proteins makes the study even more complicated as bio-polymers are hetero-polymers and each amino acid type has a specific reactivity. For now, most studies on phase separating proteins and nucleic acids suggest that interactions between polymers are important (the review can be found in Chapter 1 as well). However, there is no doubt that solvent plays an important role in phase separation and probably both polymer and solvent contributions need to be taken into account. In addition, a gap between macro- and micro-scale studies on phase separation exists. The roles of different factors in phase separation are debated: protein dynamics, short-range structure and interactions, long-range structure and interactions, protein:solvent interactions, protein diffusion.

Solution NMR is a leading biophysical technique to study IDPs which provides an atomic resolution information. All processes mentioned above can be studied using NMR: protein structure and dynamics, interactions with solvent and other proteins and even protein translational diffusion. In particular, NMR provides a range of powerful techniques for the description of backbone dynamics behavior in a wide range of time-scales from picoseconds to seconds.

5.1.1 Protein and solvent concentration in condensed phase

Until now several studies were carried out on protein in the supercondensed phase. Below a short review on protein concentration, structure and dynamics in droplets is presented.

Protein concentration inside droplets was measured in several studies by different techniques (UV-absorbance, FCS) and in all cases concentration was three orders of magnitude higher than in a co-existing dilute phase. For LCD from hnRNPA1 protein concentration was estimated to be 30 mM, which corresponds to around 400 mg/ml [132]. For hnRNPA2 LCD protein concentration was slightly higher (440 mg/ml) [181]. Similar values were found for ddx4 (380 mg/ml) and protein concentration dependency on NaCl in the buffer was shown: the increase of NaCl concentration decreased protein concentration. Also, a phase diagram for ddx4 phase separation as a function of NaCl and temperature was measured and fitted globally, which allowed to determine protein density and calculate void volume, thus, solvent fraction as well, giving in this case is 0.75 (41.5 M H_2O) [21].

5.1.2 Protein secondary structure comparison between dilute and condensed phases

Most phase separating proteins exhibit a high content of disorder, some proteins are know to be fully disordered in the dilute state (fus, ddx4, LCD of hnRNPA1). The role of charged and aromatic residues in triggering phase separation was shown using NMR and mutagenesis. Nevertheless, transient interactions between chains apparently do not affect the secondary structure and proteins remain disordered in the condensed phase as was shown for ddx4 [21], hnRNPA2 [181] or fus [26]. However, in the case of tau protein, β-hairpin-like conformations [9] or α-helix is formed upon dimerization with TDP-43 in the droplet [38].

5.1.3 Protein conformational changes during the transition between phases

It is a debated question whether an IDP becomes more extended in a crowded medium or rather collapses. Even for the same protein (fus) contradictory data are available. Using PRE labelling Monahan et al. showed that fus becomes more extended and has fewer intramolecular contacts in a phosphorylated form or with phosphorylation mimicking mutation which are less prone to phase separate [141]. On the other hand, using pyrene emission tau K18 was shown to exhibit more extended conformations in the droplet which, as suggested, allows it to participate in a multitude of intermolecular interactions to form a network [129].

5.1.4 Spin relaxation of proteins in condensed phase

NMR relaxation is exquisitely sensitive to the dynamics of macromolecules occurring on timescales from picoseconds to tens of nanoseconds, and has been largely employed to study the motion of IDPs, from the dilute state, to the study of interactions, through the effects of crowding (see Chapter 2) to the impact of the cellular environment [72, 184]. While there are clear correlations between measured relaxation rates and, for example, the formation of transient secondary structure or interactions, it has not always been clear which motional modes are probed by the different relaxation rates that are normally measured. Recent studies from our laboratory have employed multi-field relaxation measured as a function of temperature and crowding to shed more light on this problem. This revealed that high field ^{15}N relaxation is sensitive to three components of IDP dynamics: fast local libration (50 ps), intermediate backbone dihedral angle motion (1 ns) and slow segmental motions (10 ns) [2, 4]. The last are shown to be coupled to water dynamics which probe local microviscosity [4]. Water content and viscosity change dramatically in a biocondensate, and protein backbone dynamics is therefore probably affected as well.

Comparison of R_2, R_1 and hetNOE were used in several studies to explore how protein dynamics change between phases.

Slowed motion in the condensed phase was shown for hnRNPA2 [181]: A dramatic increase in NOE (from -1 to 0.5) and R_1 (from 0.75 to 1.5) and minor changes in R_2 indicated that slow-intermediate motions were affected more than fast ones.

For fus a different tendency was shown: R_2 values increase from 3 to 5 s^{-1} in the dispersed state to 15 - 35 s^{-1} in the condensed phase, R_1 values decrease from 1.3 to 1.7 s^{-1} to 0.75 to 1.2 s^{-1}, and heteronuclear NOE values do not change much (in condensed phase fluctuate around 0.5 s^{-1}, in dilute state between 0.2–0.6 s^{-1}). In general, the changes in three relaxation experiments were interpreted to report on protein restricted mobility in comparison with the dilute state. However, only a qualitative analysis of protein dynamics is available [26].

Spin relaxation of ddx4 was also measured in the dilute state (7 mg/ml), condensed state which was formed by LLPS (380 mg/ml) and ddx4 mutant which does not phase separate but was concentrated till 370 mg/ml and compared. The average R_2 value for the dilute state was 4.3 s^{-1} which was significantly lower than both condensed phase samples. Surprisingly, even though protein concentration between phase separated and non-phase separated condensed samples does not differ a lot, R_2 values change from 9.8 s^{-1} (non-phase separating ddx4 mutant), to 16.4 s^{-1} (phase

separating ddx4). However, NOE values are higher for the non-phase separated sample. Using these data a generalised order parameter for backbone amide (S2), and residue-specific overall tumbling time, (τ_C) were calculated using model-free approach (see methods). The average values for $S2/\tau_C$ are presented on Fig. 5.1. The two points at highest concentration correspond to phase separating samples at different NaCl concentrations, other three to non-phase separating samples. S^2/τ_C has a step function between points corresponding to the non-phase separating and phase

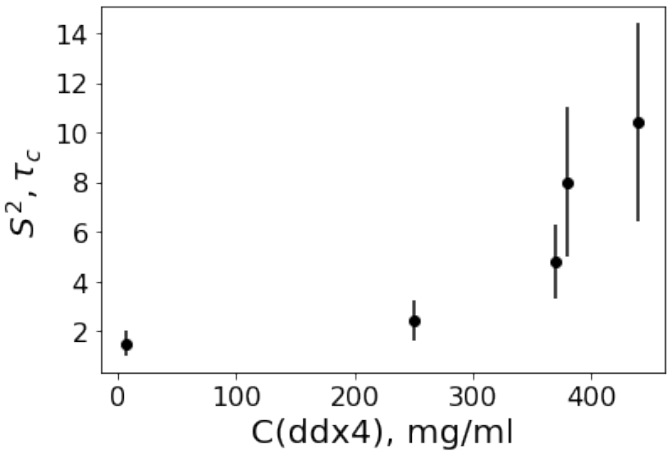


Figure 5.1 – The dependency of $S2\tau_C$ on a protein concentration in the sample. Values are taken from Table S1 [21]

separating measurements. All this suggests that apart from the effect of the viscosity and crowding on protein dynamics an additional slowing of backbone motions in ddx4 is, potentially, the result of transient intermolecular contacts that mediate phase separation [21]. However, this study does not provide residue-specific information about how relaxation rates and dynamics change. In addition, as mentioned above it was shown that model-free analysis of IDPs require three modes at different timescales contributions, and here two were used.

5.2 Open questions

It is clear from the preceding chapters that it is essential to understand the structural and dynamic behaviour of component IDPs in liquid droplets if we are to understand the mechanisms of stabilisation. In order to do this we need to understand the transformation of the free-energy landscape when transitioning from the dilute to the condensed phase. Solution NMR is a leading biophysical technique to study IDPs which provides atomic resolution information. In particular, NMR provides a range of powerful techniques for the description of backbone dynamics behavior in a wide range of time-scales from picoseconds to seconds. Important contributions have already been made in this field and it is clear that recently developed approaches from our laboratory to describe the properties of IDPs in solution as a function of physically meaningful dynamic modes are well-adapted to examine this important paradigm (see section 2.3 and 2.4) [2, 4, 184].

Until now, most studies on phase separating proteins and nucleic acids suggest that interactions between polymers are important for stabilizing liquid droplets, however, there is no doubt that solvent plays an important role in phase separation and probably both polymer and solvent contributions have to be taken into account. In addition, a gap between macro- and micro-scale studies on phase separation exists. The roles of different factors in phase separation are debated:

protein dynamics, short-range structure and interactions, long-range structure and interactions, protein:solvent interactions, protein diffusion.

In Chapter 3 we presented a study on measles N :P droplets. However, using NMR we could not observe any signal coming from the protein in the condensed phase. There are several reasons why this could happen. In order to investigate the dynamic properties at high resolution using NMR relaxation, and to understand the limitations limit our ability to observe N:P droplets by NMR we require a model phase separating system. This would allow us to test multiple conditions and better understand what are the triggering factors for phase transition and how droplets are stabilised.

In this chapter I will show how we used the model recently developed in the lab to quantify the three contributions to spin relaxation that are parameterised by the model-free approach, with particular focus on the comparison between dilute, crowded and condensed phases. Using an approach proposed recently (see Chapter 2, section 2.4) to predict relaxation rates in various environments, it is intriguing to test it for the ability to predict protein dynamics in condensed phase.

Here we combine several biophysical techniques with the main accent on NMR to give a detailed description of a model phase separating system and determine the changes which MeV N_{tail} undergoes during the transition between the two states.

5.3 MeV N_{tail} as a model phase separating protein

5.3.1 Confirmation of N_{tail} droplet liquid behaviour

As a model to study protein conformational and dynamic behaviour in the condensed phase we used measles virus N from 405 to 525 aa (N_{tail}) which phase separates after being mixed with PEG 10'000 solution and forms a biphasic colloid solution. With time it evolves in a two macro-phase solution: One rich in N_{tail} - yellow dense part and the second with a low N_{tail} concentration (Fig.5.2(a)). Using fluorescent labelled N_{tail} we could image the colloid solution, as one can see on Fig. 5.2(b) N_{tail} is concentrated inside droplets while the surrounding solution is not fluorescent, indicating low N_{tail} content. N_{tail} is in exchange between both phases even after droplets are stabilised as shown using fluorescence recovery after photobleaching (Fig. 5.2(c)). Also fusion events of two droplets with a subsequent relaxation into spherical particles were observed indicating liquid nature of the condensed phase.

5.3.2 Indicating PEG content in both phases

As phase separation was only observed in the presence of PEG we first studied its distribution between phases using IR-spectroscopy as PEG gives a characteristic peak at 2978 cm^{-1}. The presence of PEG in the condensed phase was minor in comparison with protein and also lower then the average concentration in the total volume, indicating that it is excluded from the condensed phase. Fig. 5.3 shows the absorption in IR region for several concentrations of PEG, which were used as references, condensed phase and dilute phases.

In addition, on a 1D ^{1}H NMR spectra PEG peak gives a characteristic peak around 4 ppm. The intensity of this peak measured on a condensed phase was much lower than in the buffer containing 75 mg/ml PEG, however, this could be due to faster relaxation of the molecule because of slower dynamics or interactions (Fig. 5.4).

To conclude, as PEG is mostly excluded from droplets, we consider that only N_{tail} is required for phase separation and forms a self-scaffold, and PEG acts as a trigger, probably by replacing

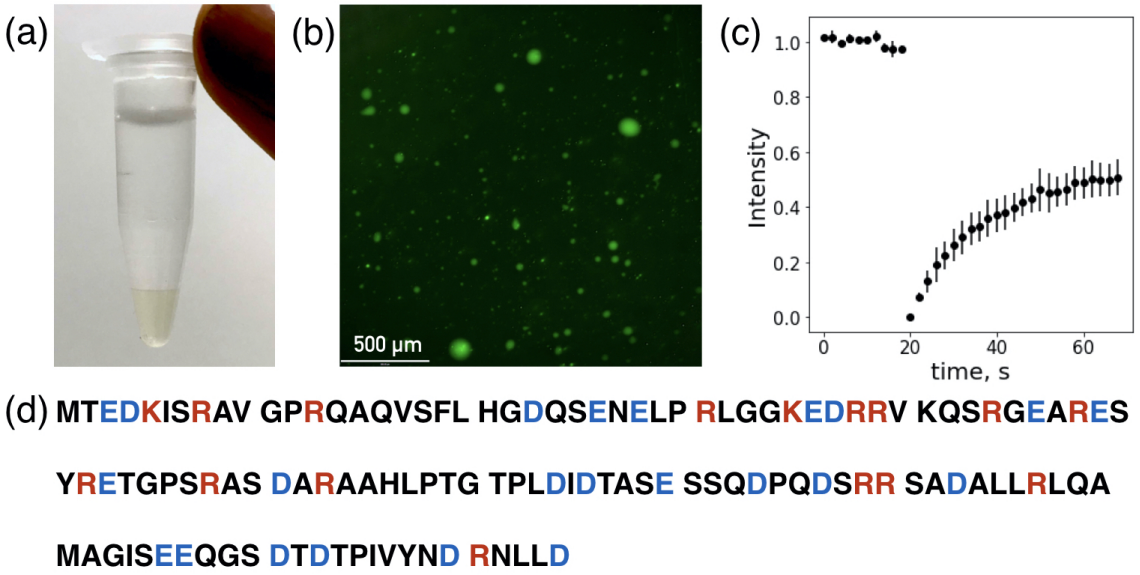


Figure 5.2 – N_{tail} phase separation. (a) After incubation solution separates into two phases: N_{tail} condensed and dilute ; (b) Fluorescent microscopy imaging of a colloid system where N_{tail} is labelled with FAM ; (c) FRAP of N_{tail} ; (d) N_{tail} amino acid sequence.

water and helping the system to reach a critical water volume fraction.

5.4 Sequence analysis

A high proportion of glycine and prolines is typical for IDPs which tend to phase separate. N_{tail} is also highly charged and around 10% of the sequence is represented by arginine residues. Several examples of phase separating proteins with a high content of arginine are known, and in some cases the replacement of arginines with lysines suppressed their ability to phase separate. However, in contrast to a general observation for proteins which are prone to phase separate, low aromatic amino acid content is present in N_{tail} (Fig. 5.2(d)). Only three aromatic residues are present: 418Phe, 451Tyr, 517Tyr. The sequence could also not be described as low-complexity.

5.5 Phase diagram and protein concentration in droplets

5.5.1 N_{tail} phase diagram

In order to measure the phase diagram we used the protocol described in [21]. Protein was mixed with PEG 10'000 solution at various NaCl concentration, so that final N_{tail} was 300 μM and PEG 200 mg/ml. PEG and protein solutions were gently mixed by pipetting until homogeneous turbid solution was obtained. The condensed phase was sedimented down by incubating tubes over night in PCR machine at given temperature and briefly sedimented at 11'000 rpm. Next, we measured volumes of both phases and then measured protein concentration in the presence of guanidinium chloride. The condensed phase was diluted 40 times before the measurement, dilute phase twice. This procedure was performed at four temperatures (282 K, 288 K, 298 K, 308 K)

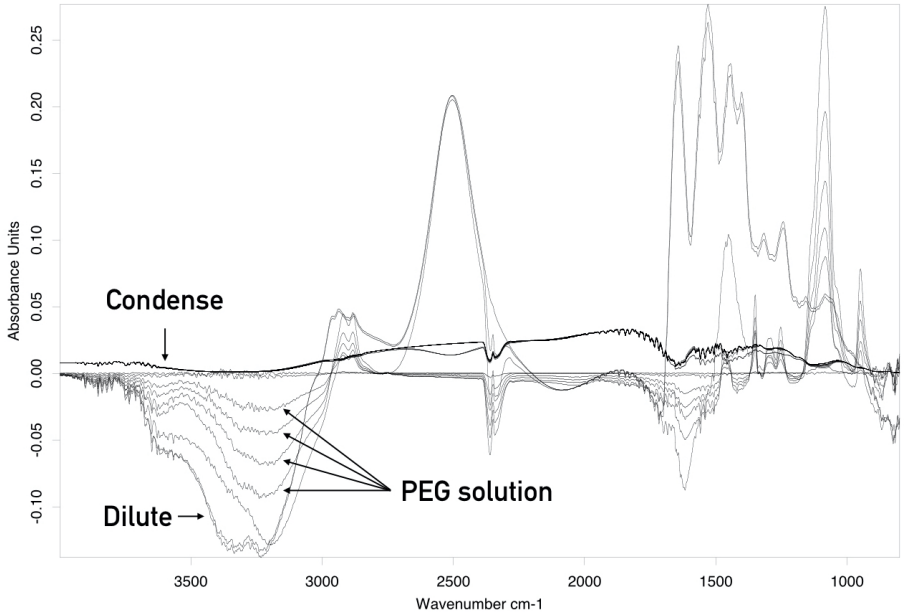


Figure 5.3 – IR absorption spectra, condensed phase spectra are in black

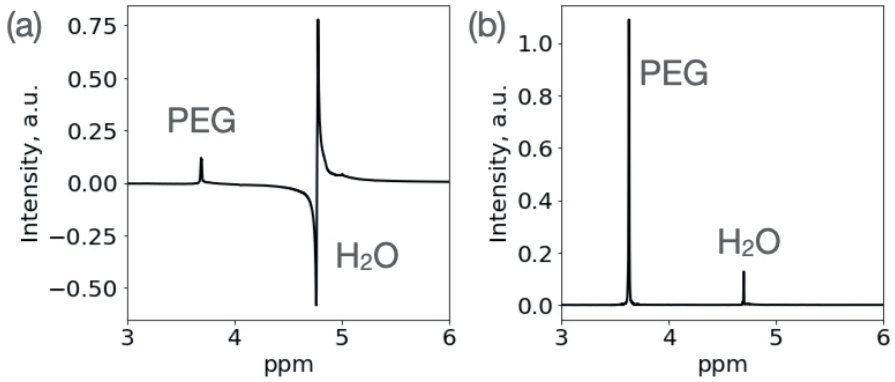


Figure 5.4 – 1H NMR spectra of (a) condensed solution, (b) in the presences of 75 mg/ml PEG.

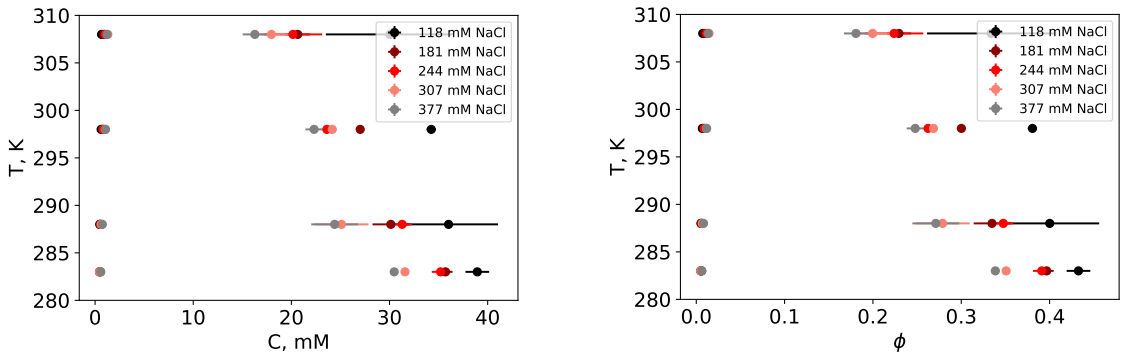
and five NaCl concentrations (118 mM NaCl, 181 mM NaCl, 244 mM NaCl, 307 mM NaCl, 377 mM NaCl).

Protein concentration was measured using nanodrop, three measurements for each sample were recorded and averaged. Extinction coefficients $\varepsilon = 2980 M^{-1} cm^{-1}$ was calculated with ExPasy [216] and used to determine concentration from the absorbance at 280 nm. Calculated N_{tail} concentrations as a function of salt and temperature are presented on Fig.5.5a in molar units and 5.5b in volume units. Error bars represent standard deviation of three measurements.

Protein density of 1.35 g/cm^3 [60] was used to calculate volume fraction ϕ of the protein in both phases.

As shown on Fig.5.5a and Fig.5.5b the phase diagram has an upper critical solution temperature, the highest protein concentration was estimated to be 38.9 mM what corresponds to 584 mg/ml at 288 K and 118 mM NaCl. Thus, 0.42 of volume fraction corresponds to solvent. Considering negligible presence of PEG, water concentration is estimated to be around 23.3 M at taken experimental conditions.

While salt concentration is increasing, the protein fraction in the condensed phase is decreasing due to the lower protein concentration and, in addition, the volume of the condensed phase is changing. Using fluorescence microscopy we could show that increasing of NaCl concentration



(a) Concentration variation *(b) Volume fraction variation*

Figure 5.5 – N_{tail} phase diagram as a dependency on temperature. Different colors correspond to different salt concentrations.

progressively dissolve droplets (Fig. 5.6).

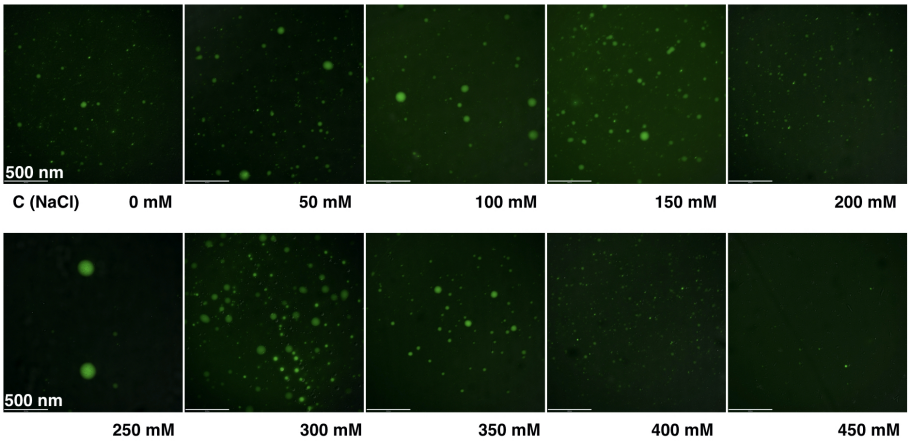


Figure 5.6 – Fluorescent microscopy images of N_{tail} droplets taken at different NaCl concentrations.

5.5.2 Fitting N_{tail} phase diagram with Flory-Huggins model

Using equation 2.7 we fit phase diagram at different NaCl concentrations (see Chapter 2, section 2.5.3). We assumed that $N_{Ntail} = 130$ by the number of amino acids and $N_{solvent} = 1$, the results of the fit are plotted together with experimental data on Figure 5.7a. Critical composition in this case is $\phi_c = 0.08$, critical temperature points for each curve are summarized in Table 5.1 together with the fitted parameters A and B.

χ was calculated using equation 2.4 and plotted in 5.7b, all have a positive value which corresponds to an attractive interaction promoting phase separation. Critical χ is lower than all experimentally obtained values in agreement with observations of phase separation in the solution.

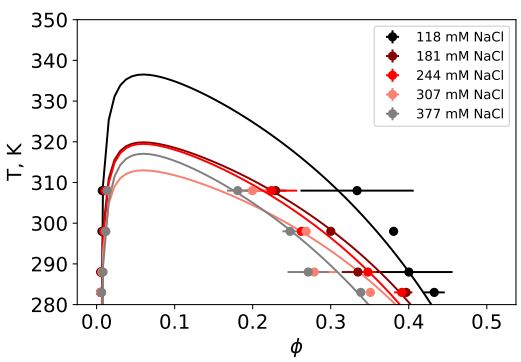
Equation 2.3 allows to estimate free energy of mixing and $\Delta F_{mix}/\phi = kT\chi(1-\phi)$ is an average interacting energy per one amino acid, the configurational entropic component of free energy of mixing depends only on volume fraction

$$F_{mix}/\phi = kT\frac{ln\phi}{N_A} + kT\frac{1-\phi}{\phi}ln(1-\phi)$$

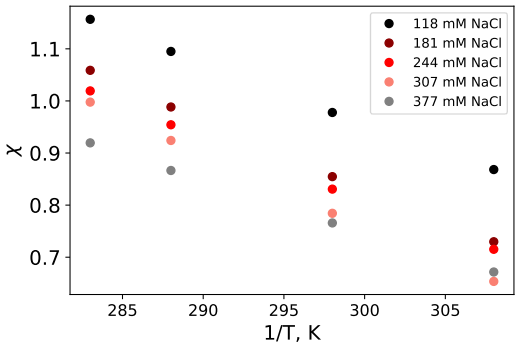
Values for both contributions together with the free energy of mixing are summarised in Table 5.2. One can notice that they have opposite signs: The entropic part is negative, thus promoting mixing,

Tableau 5.1 – Fitted parameters for each salt concentration and corresponding critical temperature.

NaCl, mM	A	B	T_c, K
118	-2.396	1005.484	336.960
181	-2.994	1147.174	320.208
244	-2.726	1060.044	319.850
307	-3.242	1199.898	313.295
377	-2.134	864.163	317.498



(a) Fitted phase diagram with equation 2.7.



(b) Calculated χ parameter for all temperature and salt conditions.

Figure 5.7 – The results of Flory-Huggins model fit.

interacting part is positive, thus promote demixing. The module of interacting contribution is higher in all presented cases and the total energy of mixing is positive, indicating phase separation. In case of a pure N_{tail} solution ($\phi = 1$) maximum restriction on the chain is achieved and the entropy can be calculated as follows: $\Delta S = kT$ (0.562 kcal mol $^{-1}$, 0.572 kcal mol $^{-1}$, 0.592 kcal mol $^{-1}$, 0.612 kcal mol $^{-1}$ for used temperatures). Taken it together with mixing entropy, conformational entropy of the mixture can be found (Table 5.3).

Tableau 5.2 – Energy of mixing: entropic, interacting and total. Energies are in kcal mol-1.

	283, K			288, K			298, K			308, K		
NaCl, mM	ΔH_{mix}	ΔS_{mix}	ΔF_{mix}	ΔH_{mix}	ΔS_{mix}	ΔF_{mix}	ΔH_{mix}	ΔS_{mix}	ΔF_{mix}	ΔH_{mix}	ΔS_{mix}	ΔF_{mix}
118	0.36	-0.23	0.13	0.37	-0.24	0.13	0.35	-0.25	0.10	0.35	-0.27	0.07
181	0.35	-0.24	0.11	0.37	-0.25	0.11	0.35	-0.27	0.08	0.34	-0.29	0.04
244	0.34	-0.24	0.10	0.35	-0.25	0.10	0.35	-0.28	0.08	0.33	-0.29	0.04
307	0.35	-0.24	0.11	0.37	-0.26	0.11	0.33	-0.28	0.06	0.31	-0.30	0.01
377	0.33	-0.25	0.09	0.35	-0.27	0.09	0.33	-0.28	0.05	0.33	-0.30	0.03

Using a simple Flory-Huggins model we could already describe the behaviour of the system via an average interaction parameter. Changes in free energy of mixing are dominated by interacting energy over repulsive configurational entropic contributions. However, this approach does not take into account possible differences in N_{tail} chain conformation and assumes that it is the same in both phases.

During the following sections we discuss how N_{tail} short- and long-range conformational elements change between dilute and condensed states as studied using NMR and molecular modelling.

5.6 Free energy of mixing

The difference in free energy of the system during phase separation consists of two main contributions: Energy of mixing and chain deformation. (As current research is focused on a protein characterisation, solvent contribution is ignored here).

$$\Delta G = \Delta G_{mix} + \Delta G_{def} \tag{5.1}$$

In a condensed phase the protein chains interact with each other, triggering phase transition and, in order to maintain such interactions, adopt new conformations. In addition it may be that packaging N_{tail} chains so densely can require some conformational adaptation. N_{tail} is represented not as a single conformation but an ensemble, however, the characteristics of the ensemble in dilute and condensed phase may differ significantly.

5.6.1 Sample preparation

Protein state in the condensed phase was studied by NMR. The measurements were made on a macroscopic droplet sample. N_{tail} and PEG were mixed, as described in Chapter 2, section 2.7, then sedimented at 1000 rpm 20°C, around 80 μl of dense N_{tail} phase was obtained, and transferred into a 3 mm Shigemi tube using glass capillary. Use of the Shigemi tube helped to achieve a better shim during the NMR measurements and resulting in narrower peaks in ^{1}H dimension (Fig. 5.8(c)). The comparison of ^{1}H-^{15}N HSQC spectra recorded in 3 mm Shigemi and normal tubes are presented on (Fig.5.8). We used 10% ^{15}N-^{13}C labelled N_{tail}, and 90% of non-labelled N_{tail}.

Almost all the peaks which are observed in dilute N_{tail} (300 μM) were present in droplet-sample spectra (Fig. 5.8(b) apart from the region between 488-499 aa which correspond to a transient helix. The absence of this region could be due to a slow tumbling, as it was already more rigid in the dilute state. Alternatively, this region could be involved in the interaction between N_{tail} molecules what could be essential for phase separation.

5.6.2 Translational diffusion

As in a droplet sample macromolecular concentration is much higher, and, thus, the solution is more viscous, molecules diffusion is slowed down. Translational diffusion of N_{tail} was measured using pulse field NMR experiment. Between the endoding and decoding gradients the molecules were diffusing during 1 sec. A series of 1D ^{1}H with a varying z-gradient strength from 0 to 0.5 G/m were recorded. Signal intensity is dependant on the gradient strength and diffusion: The faster the molecule diffuses, the less magnetisation is recovered after the decoding gradient, the lower the signal is. The maximum difference in the intensity is around 10% in the tested region of gradient strength which confirms slow translational diffusion and a highly viscous state of droplet phase (Fig.5.9).

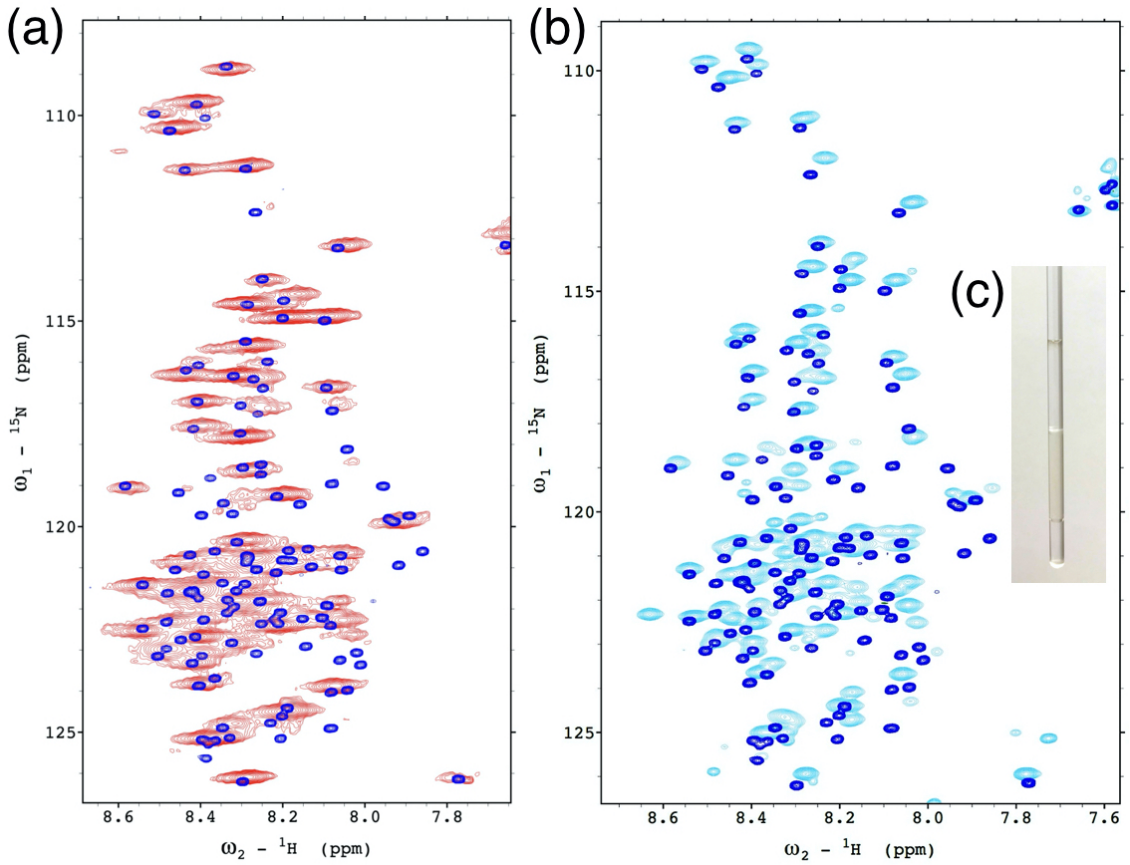


Figure 5.8 – ^{1}H-^{15}N HSQC spectra of N_{tail} recorded using 3 mm normal tube (a, red) and Shigemi (b, cyan), N_{tail} spectra in a dilute state is shown in blue on both (a) and (b) ; (c) image of a Shigemi tube with a droplet sample

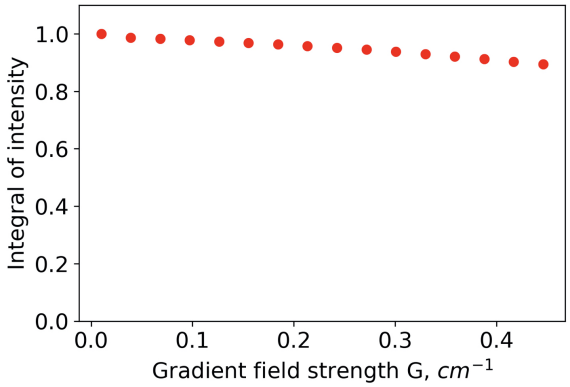


Figure 5.9 – Integral of 1D ^{1}H spectra signal intensity of N_{tail} droplet sample as a function of z-gradient strength.

5.6.3 Protein secondary structure propensity does not differ between dilute and condensed states

Despite the advantages in spectral resolution which were obtained using Shigemi tube, still N_{tail} peaks in the condensed phase are broadened especially in the ^{1}H dimension resulting in a massive overlap in the ^{1}H-^{15}N HSQC spectrum as the central region is highly crowded. In addition, some peaks are missing. Nevertheless, the HSQC spectra show a high level of similarity between dilute and condensed phase N_{tail}, indicating that protein conformation should not differ a lot. We

measured CO and C^{α} chemical shifts which are more sensitive to the difference in secondary structure propensity. In addition, recorded 3D experiments also allowed to verify the assignment of N_{tail} in droplet . The comparison of secondary C^{α} and CO together with N and NH chemical shifts of the droplet sample with those recorded on a dilute N_{tail} are presented on (Fig. 5.10). Apart from the missing helical region, minor differences are observed. To conclude, despite slow translational diffusion (see FRAP and pulse field NMR measurements) most of the protein chain remains intrinsically disordered.

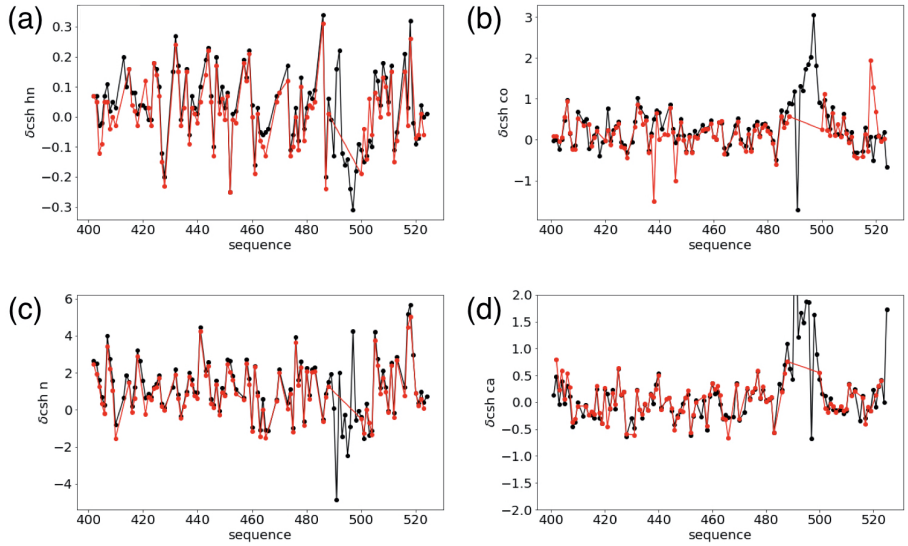


Figure 5.10 – Secondary chemical shift (a) ^{1}H, (b) CO, (c) ^{15}N, (d) Cα. Black corresponds to N_{tail} in dilute state, red to N_{tail} in condensed phase.

Similar observations were made previously on other phase separating proteins: ddx4 and fus also remain disordered even in droplet environment.

5.7 N_{tail} backbone dynamics is various crowded conditions

Although we could confirm that protein structure is not changed as a result of the phase transition, the dynamic characteristics could be affected. IDPs features are determined by their motional dynamics, which can be studied using NMR spin relaxation experiments.

5.7.1 Relaxation measurements on N_{tail} in a dilute state.

In order to characterize N_{tail} dynamics a set of NMR experiments was measured: ^{15}N longitudinal relaxation (R_1), ^{15}N transverse relaxation (R_2), which was measured using $R_{1\rho}$ experiment, cross relaxation measured with ^{15}N-^{1}H NOE and CSA/dipole-dipole cross-correlated transverse relaxation.

N_{tail} samples (300 μM) were measured at two fields (600 and 850 MHz) and three PEG concentrations (0, 37.5 mg/ml and 75 mg/ml). It should be mentioned that at these chosen experimental conditions N_{tail} still remained in a single phase even at 75 mg/ml PEG concentration. The extracted relaxation rates are presented on Fig. 5.11.

As expected, the percentage of crowding agent in solution is increasing, transverse relaxation and NOE are increasing as well. The effect of a crowder present in the system on longitudinal

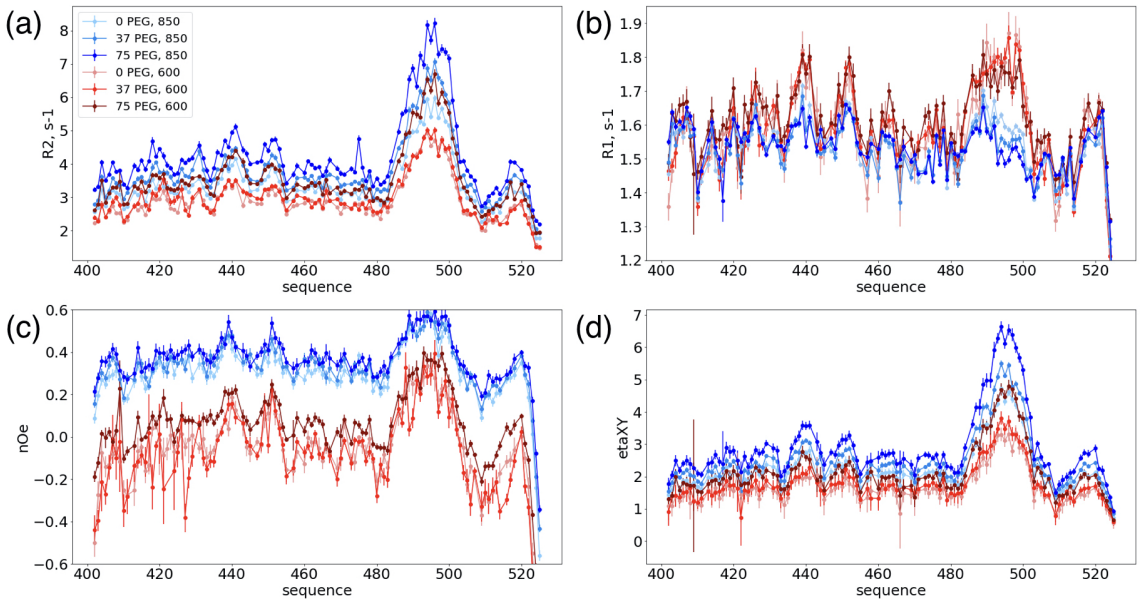


Figure 5.11 – Experimental relaxation rates of N_{tail} in a dilute state. Red plots correspond to measurements at 600 MHz, blue - 850 MHz.

relaxation values is more subtle probably because at chosen experimental conditions relaxation rates are close to the maximum R_1 values.

5.7.2 Model-free analysis of N_{tail} in the dilute state

Measured relaxation rates indicate significant backbone mobility along N_{tail} chain apart from the short helical region between 488-499 amino acids where relaxation values are elevated and show a higher rigidity. These data were used for model-free analysis to determine amplitudes and timescales of motions (for a more detailed description see Chapter 2). For the simultanious fit of all the data nanoviscosity of the samples was measured with longitudional water relaxation. The extracted values are shown in Fig. 5.12 and were used for N_{tail} analysis.

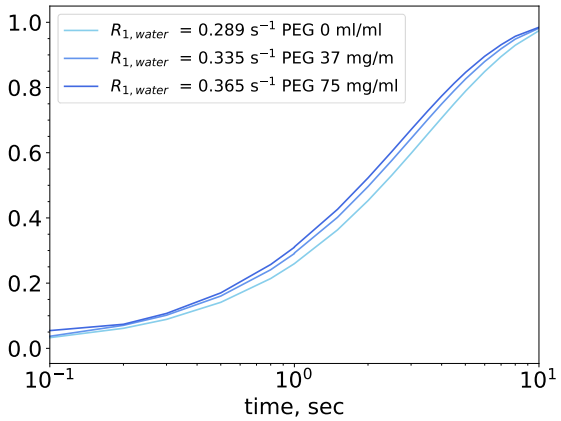


Figure 5.12 – Water relaxation rates in the presance of 0, 37.5 mg/ml and 75 mg/ml PEG and 300 μM N_{tail}.

Extracted order parameters and correlation times are shown on Fig. 5.13, 5.14.

As was described in Chapter 2, the correlation time of the fast motion generally varies very weakly along the sequence with similar conformational properties, and was fixed to previously

optimized values of 45 ps along the disordered region and 0 for the helical region. Fitted intermediate motions are in the range of 0.8-1.6 ns, slow are in the range of 4 to 15 ns. The distribution of amplitudes of the different components of the motion, their correlation times and the friction coefficients of each amino acid fall in very similar ranges compared to those of N^{tail} from Sendai virus and the disordered domain of MKK4 that were published recently [4].

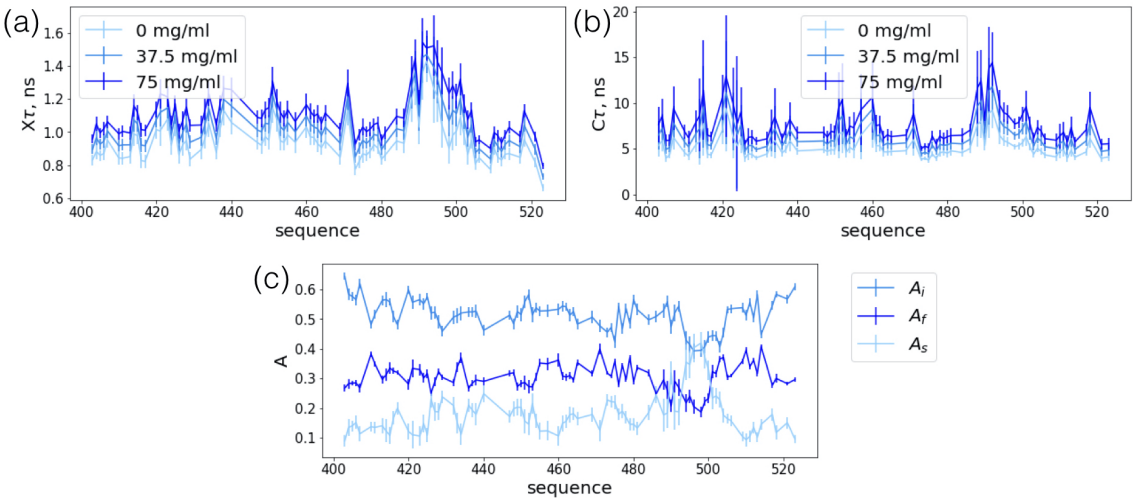


Figure 5.13 – Extracted parameters of backbone motions. (a) Intermediate correlation times, (b) slow correlation times, (c) amplitudes of motions for all three correlation times.

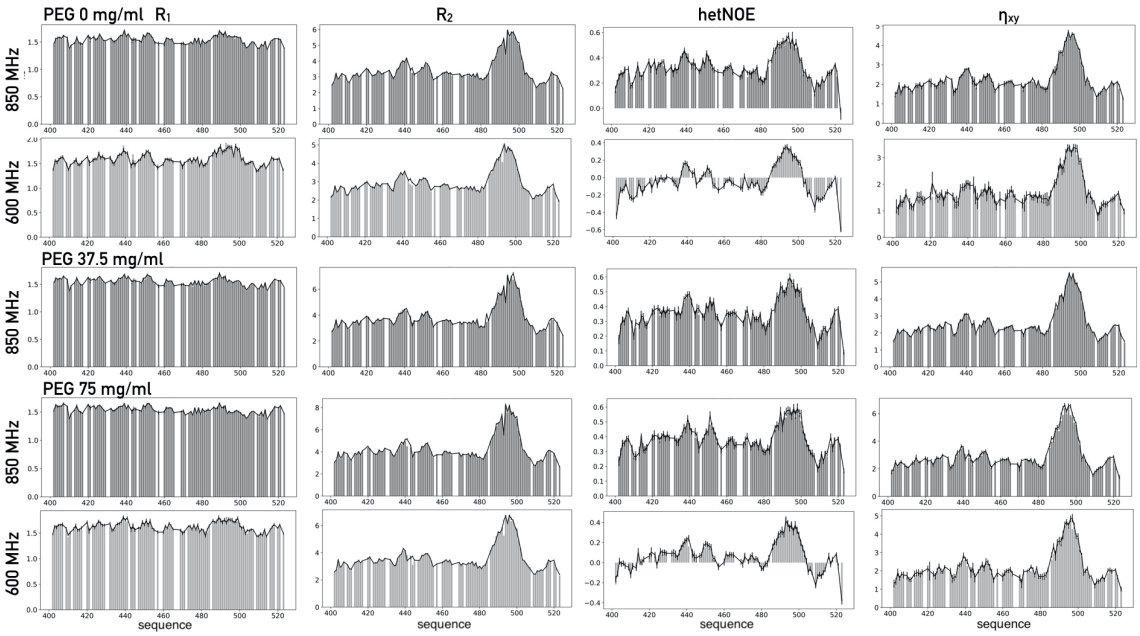


Figure 5.14 – Fits of experimental spin relaxation values.

5.7.3 Relaxation measurements on N_{tail} in a condensed state

A similar set of experiments of a droplet sample where protein concentration was estimated around 28 mM was measured. N_{tail} in a dilute sample was 300 μM. Relaxation rates in the droplet sample are plotted together with a dilute samples measurements (Fig. 5.15) to represent the effect of a supercrowded condition on a spin relaxation. First, as it was mentioned above, the helical

region (residues between 488 and 499) was not observed in the spectrum, so relaxation values for these residues are absent. R_2 in the disordered regions increased from 2-4 s^{-1} in the dilute state to 10-20 s^{-1} in the droplet. Similarly, transverse cross-correlated relaxation rates increased from 1-3 s^{-1} to 4-11 s^{-1}. R_1 values are reduced in the crowded condition and at 850 MHz drop by 0.4 s^{-1}. NOE values increased by 0.3-0.4. One could also notice that the profile (local peaks and depressions) are surprisingly well conserved between the dilute and the condensed states. In general, with an increase of sample viscosity as a result of an increase of PEG or protein concentration, backbone motions appear to be effected at all timescales. In the droplet sample backbone motions seems to be changed, but in order to understand which timescales are mostly affected, a model-free analysis has to be performed.

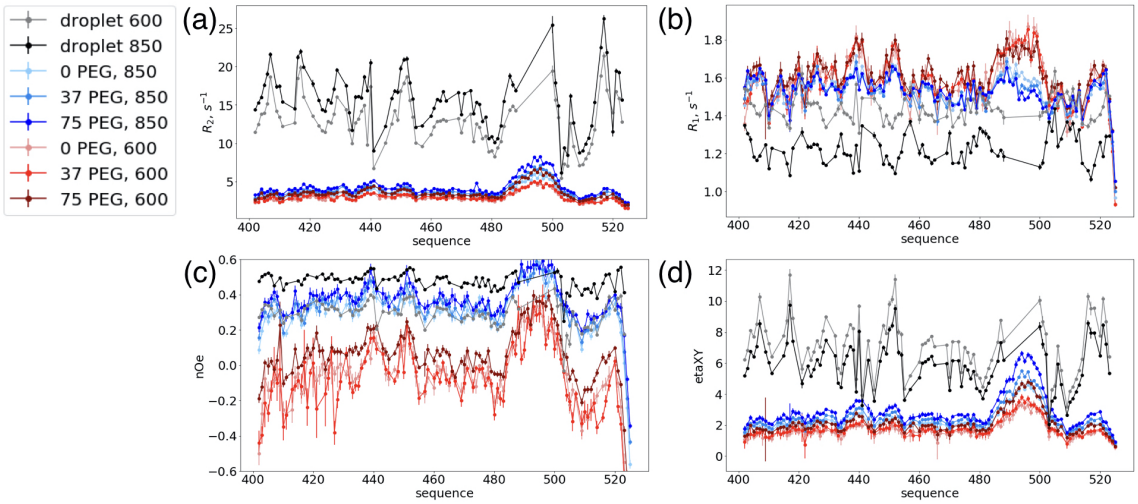


Figure 5.15 – Experimental relaxation rates of N_{tail} in a condensed state. Red plots correspond to measurements of dilute N_{tail} at 600 MHz, blue - dilute N_{tail} at 850 MHz, black-grey to condenced dilute N_{tail}.

Water relaxation was again measured in the droplet, in an attempt to measure the nanoviscosity of the system and to relate this to dynamic correlation times as in the case of the crowded and *in cellulo* investigations. In order to faithfully interpret the water R_1 in terms of a rotational correlation time, it is essential that exchange terms do not contribute to the measurement. This can be tested by measuring the water relaxation as a function of magnetic field strength. In MeV N_{tail} droplets studied here the water relaxation shows a clear and strong field dependence, precluding continuation of this line of investigation. It is possible that relaxometry measurements may clarify the origin of the field dependent contribution and this may be an avenue worth exploring. Despite not being able to link solvent viscosity in dilute, crowded and condensed phases any further, we can still compare the dynamic behaviour of the protein in the different conditions.

5.7.4 Model free analysis of N_{tail} in a condensed state and comparison with a dilute state.

Relaxation data measured on condensed N_{tail} were fitted alone using the model-free approach. In this case four parameters were fitted, the results of the fit are plotted on Fig. 5.16.

As previously for the dilute phase, the fast correlation time was optimised for the entire molecule. In this case the optimal value was in the range of 100 ps, significantly slower than in the absence of phase separation. In the case of the condensed phase, it was noted that η_{xy} and R_2 gave slightly different distributions of J(0). We therefore removed R_2 from the fit and back-calculated it. By subtracting this back-calculated value from the experimental value, we could estimate the contribution from exchange to R_2. Interestingly, this value is quantitatively very similar at both

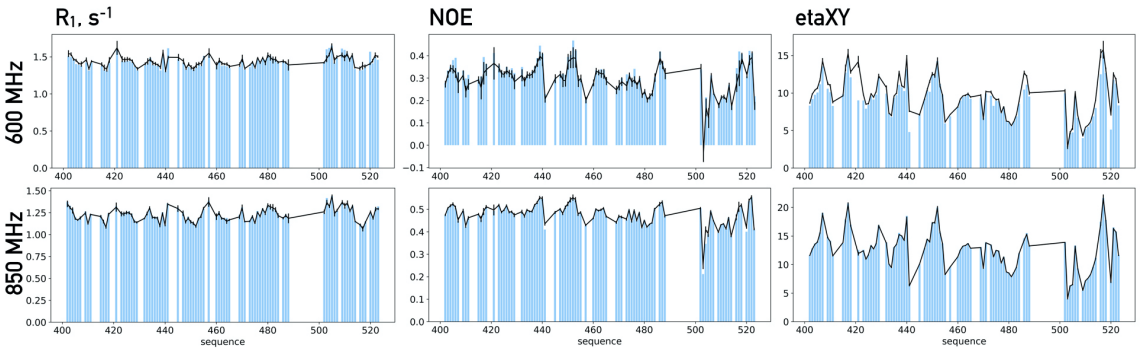


Figure 5.16 – Fitted relaxation rates of N_{tail} in a condensed state.

magnetic fields: 600 and 850 MHz, strongly suggesting that this is a contribution from slow exchange, reporting on the association rate of an interaction, possibly the stabilizing interaction for the formation of liquid-liquid phase separation. The contribution is localized in the N-terminal region of N_{tail}, although of course we cannot rule out such a contribution in the helical region, as we do not have relaxation measurements in this region in the condensed phase. Extracted order parameters and site specific intermediate and slow correlation times of motion of the droplet sample are presented together with the values obtained for dilute samples Fig. 5.17.

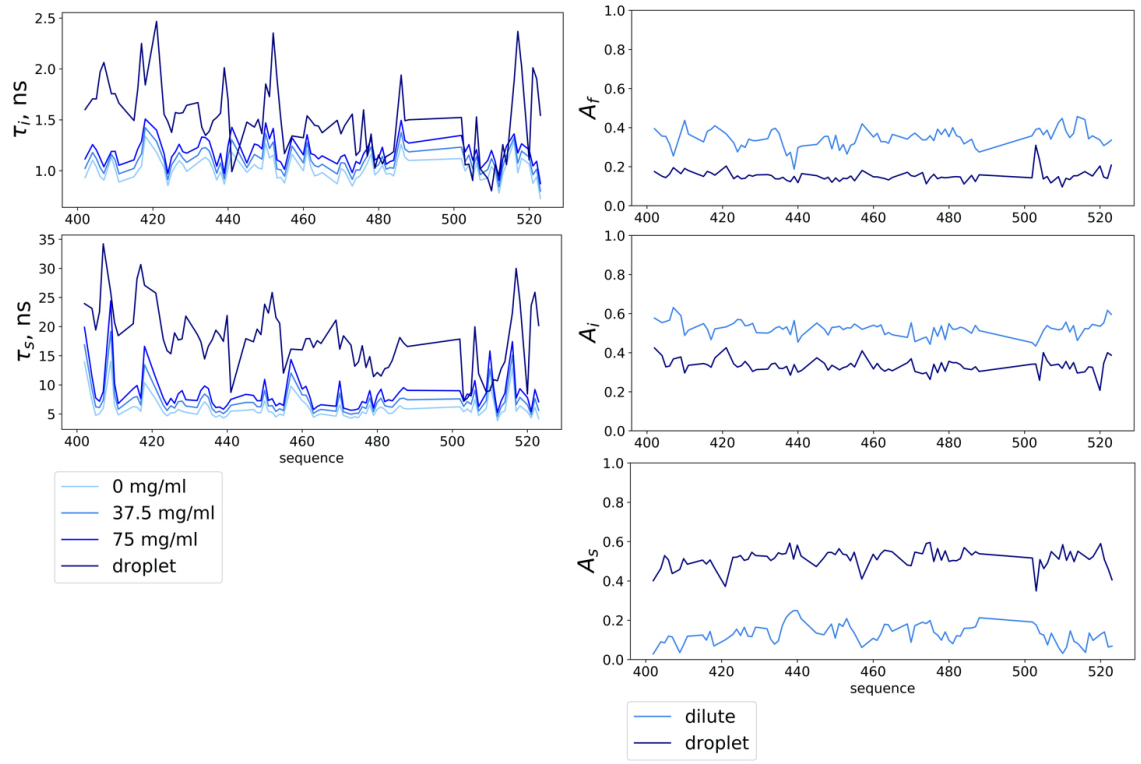


Figure 5.17 – Fitted relaxation rates of N_{tail} in a condensed state.

Correlation times of slow motions (τ_s) increase by around 15 ns along the N_{tail} sequence, while correlation times of intermediate backbone motions (τ_i) are about 40% slower in general with local increases that are much more significant, around residues 420, 451, 517 and close to the helical region. Surprisingly, the only three aromatic residues N_{tail} contains are Phe418, Tyr451, Tyr517, in addition His421. Why intermediate motions are more affected around aromatic residues still has to be studied, but one could speculate that these residues could be involved into a inter-chain contacts which organise droplet network.

The amplitudes of the three contributions to the motions of N_{tail} backbone also change significantly between the dilute and the condensed states. While the fast motions are slowed down, their amplitude is quenched by a factor of about two. The contribution of the intermediate motion also decreases from approximately 0.6 to 0.4. As a consequence, the remainder of the correlation function is accounted for by the slow motional component, which is much bigger: it increases from the average value of 0.1 to 0.5. To conclude, in condense phase the contribution of the slow motion mode in protein dynamics is much higher and the amplitudes of fast and intermediate motions decrease. Some similarity can be noticed on Fig. 5.13 for the helical region: the only region along N_{tail} sequence where the amplitudes for slow motions are higher then for other two. Taking into account that the secondary structure of N_{tail} does not change during the transition between phases, we speculate that such an effect on the protein dynamics can be caused by the intermolecular interactions between protein chains. Multiple N_{tail} molecules form a network which keeps them together and stabilises the condensed phase.

This is one of the first quantitative, site-specific comparison of motional amplitudes and timescales between IDPs in condensed and non-condensed phases and it will be interesting to compare these results with molecular simulation to provide more insight into the physical origin of these changes.

5.8 Long-range conformational changes

Short-range structure of N_{tail} was shown to be the same in both states: dilute and condense. However, N_{tail} could still experience long-range chain deformation during the transition between dilute and condensed states. This could also contribute to the entropy and the total free energy of the system.

In order to study long-range conformational changes we generated an *in silico* N_{tail} droplet using flexible-meccano, an approach which was developed for an explicit ensemble description of IDPs. This algorithm samples backbone dihedral angles to generate protein chains according to a random coil model. A Boltzmann distribution of dihedral angles is derived from non-structured elements in a database of structures determined by X-ray crystallography. While the chain is generated it follows a self-avoiding walk. In order to generate a supercondensed phase we added additional restriction to the algorithm to avoid all previously generated chains as well. Each next chain was started to be generated with a shift in the xyz coordinates, so that starting points form a lattice in the virtual box.

N_{tail} box generation:
1. Generate self avoiding chain using flexible-meccano.
2. Add side chains using pulchra.
3. Protonate a chain using reduce (3.23).
4. Add new atoms to avoid list.
5. Shift starting point for generating the next chain.

We generated an ensemble of single N_{tail} chains, which represents an infinite dilution case and three boxes which represent condensed phase at different positions on phase diagram and has 16, 27 and 40 mM N_{tail} concentration. An example of 40 mM N_{tail} box is present on Fig. 5.18.

First, dihedral angles of generated ensembles were compared and we could not find any concentration dependant differences in dihedral angles distribution (Fig. 5.19).

This indicates that despite the tight packing of the protein chain backbone dihedral angles can sample the same Boltzmann distribution as in the dilute case. However, possible attractive interactions between chains are not taken into account here.

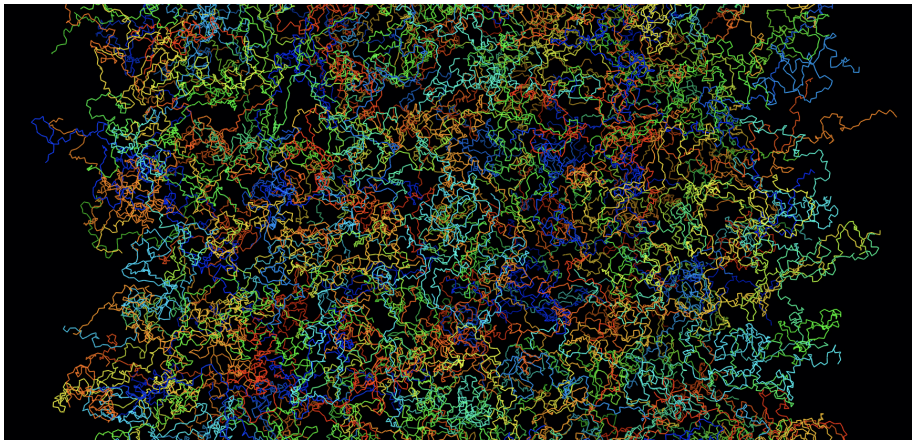


Figure 5.18 – Fitted relaxation rates of N_{tail} in a condensed state.

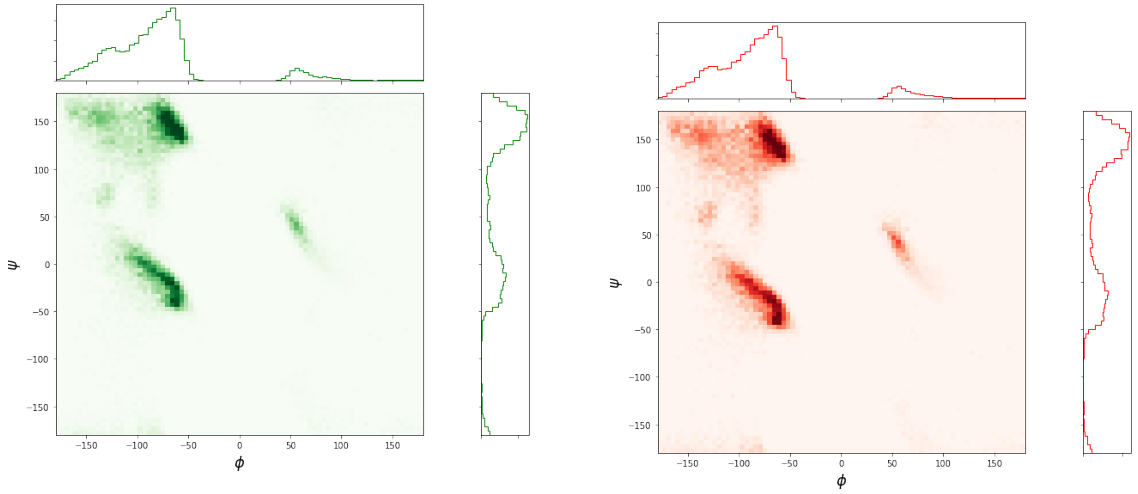


Figure 5.19 – Ramachandran plots for (a) dilute, (b) 40 mM N_{tail} boxes.

Next, we compared the long-range characteristics of the chain: R_g and end-to-end distances. R_g is increasing with concentration from 31 nm to 39 nm. The effect on end-to-end distance is even more noticeable: It is increasing from 108 nm to 149.6 nm. Distributions of R_g and end-to-end distances are presented on Fig 5.20. I should state that we have no experimental data that confirms this extension of the chain.

5.9 Conclusions

In this Chapter a new phase separating model system is characterised. We first show the proves of its liquid behaviour and perform a routine analysis of the phase diagram using Flory-Huggins model.

Protein structure in the condensed phase is shown to be the same as in the dilute state in terms of local backbone sampling over the ensemble of states, however, information about the helical region is missing. This strongly suggests that the populations of different conformations are essentially maintained.

We continued with the characterisation of protein dynamics in condensed phase. All the mo-

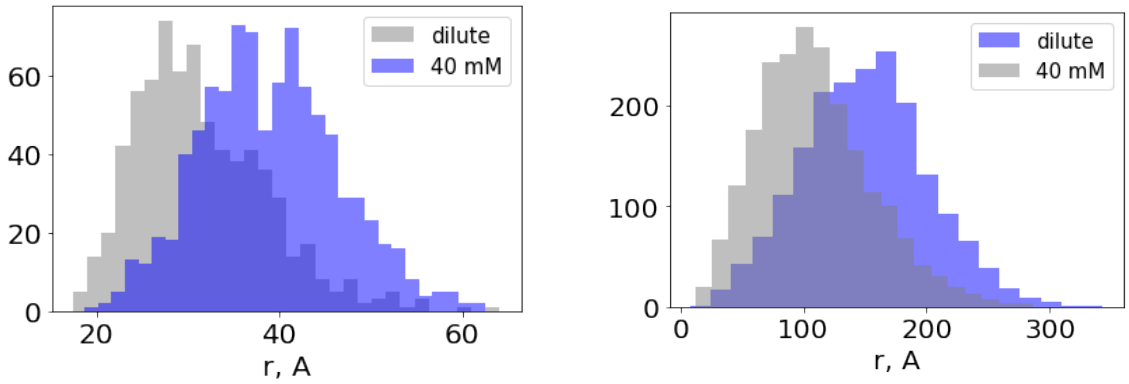


Figure 5.20 – The comparison of long-range characteristics. (right) R_g, (left) end-to-end distance for N_{tail} dilute (grey) and 40 mM box (blue).

tional modes are shown to be slower in the droplet sample and what is intriguing is that the amplitudes of motions change and show increased contributions of the slow mode. This means that N_{tail} spectral density function does not just shifts but changes the entire shape. Combination of these experimental data with molecular simulation studies may help understand how protein dynamics are affected by phase separation.

The are several indications of intermolecular interactions between N_{tail} chains in the condensed phase: slowed dynamics, particular regions are more affected then others, an additional contribution to R_2 relaxation rates. Direct observations of such contacts are required in order to first identify them and evaluate the importance. For this NMR and mutagenesis can be combined.

Long-range structural changes are predicted with molecular modelling, however, experimental data are necessary to confirm the computational predictions. We plan to test them using SAXS and PRE measurments.

Finally, solvent in condensed phase is poorly studied and requires new approaches to be developed to characterise its concentration, translational and rotational diffusion, exchange with protein.

Deuxième partie

SARS-CoV2

6
SARS-CoV-2

In this chapter preliminary results on SARS-COV-2 Nucleoprotein are presented. They were obtained after the SARS-COV-2 pandemic started both in the context of the covid19-nmr consortium and as an independent project in our laboratory.

6.1 Introduction to the SARS-CoV-2

The ongoing coronoviral pandemic is caused by a Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which is a positive-sense single-stranded RNA virus corresponding to the family Coronaviridae, order Nidovirales. Its genome, similarly to negative-sense RNA viruses, which were discussed in the previous chapters, is associated with the nucleoprotein (N). However, structure and probably some functions differ.

6.1.1 SARS-CoV-2 nucleoprotein

SARS-CoV-2 N is known to be an important cofactor of the viral replication complex, interacting and protecting the viral genome from the host cell environment, and regulating viral mRNA transcription and also involved in virus budding and envelope formation.

The structure of N is illustrated on Fig. 6.1.

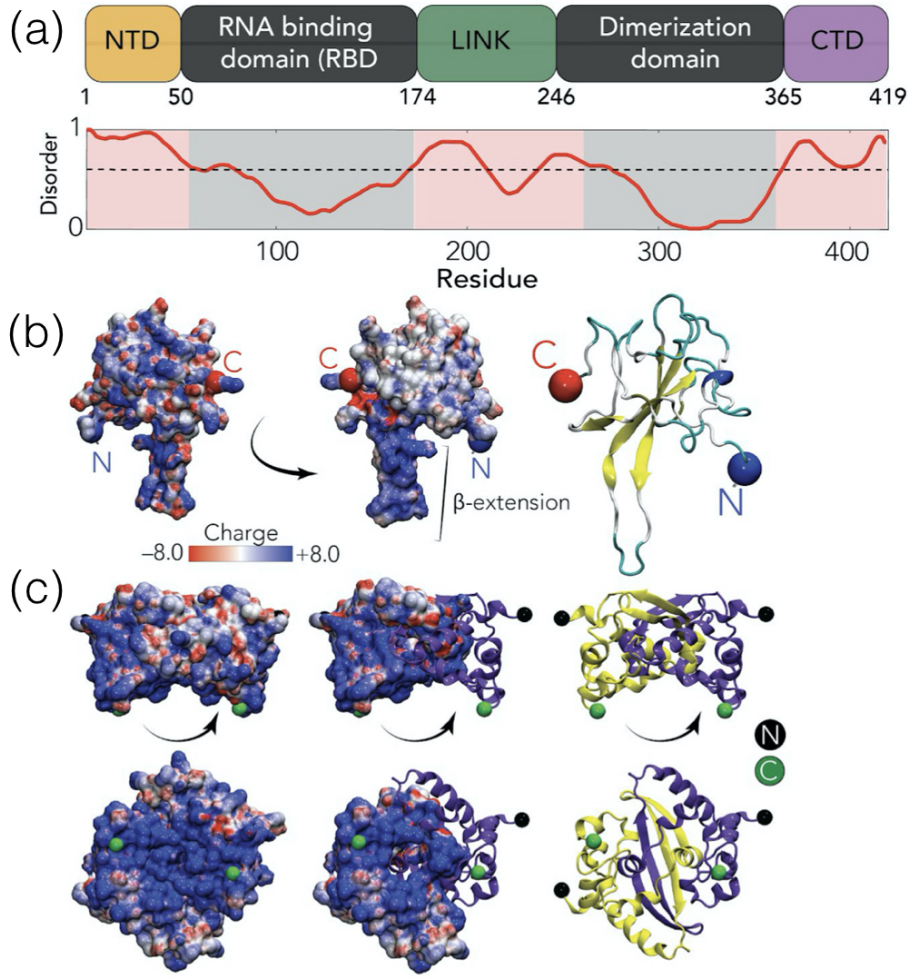


Figure 6.1 – Structure of SARS-CoV-2 N protein. (a) Domain architecture of the SARS-CoV-2 N protein. Two folded domains: RNA binding domain and dimerization domain. And three disordered: N-terminal domain (NTD), link and C-terminal domain (CTD). (b) Structure of the SARS-CoV-2 RNA binding domain (PBD: 6YI3). Center and left: coloured based on surface potential. Right: ribbon structure with N- and C-termini highlighted with spheres. (c) Dimer structure of the SARS-CoV-2 dimerization domain (PDB: 6YUN). Center and left: coloured based on surface potential. Right: ribbon structure, yellow and purple for separate monomers, with N- and C-termini highlighted with spheres.

N is a highly dynamic protein, comprising five domains, three of which are predicted to be intrinsically disordered. The structural features of coronaviral N are strongly conserved, with 90% sequence identity between SARS-CoV-1 and SARS-CoV-2. The structural domains: RNA-binding domain and dimerisation domain, are solved by X-ray crystallography (Fig. 6.1).

The central linker domain comprises a serine-arginine (SR) rich domain that is predicted to be phosphorylated at multiple sites, the impact of this modification in the structure of N is unknown.

N of SARS-CoV-1 was studied by NMR, N-terminal folded domain (RNA-binding domain) has been investigated in isolation using NMR spectroscopy [49]. The disordered domains (1, 3 and 5 on Fig. 6.1) have so far not been investigated at a residue specific level, although recent studies used single molecule FRET and molecular modelling propose the existence of transient helices in domains 1 and 3 [40]. A recent study also investigated the conformational behaviour of a 15 amino acid peptide representing the SR region of domain 3 in its phosphorylated and non-phosphorylated forms [188].

6.1.2 SARS-CoV-2 liquid-liquid phase separation

LLPS of SARS-CoV-2 N upon mixing with RNA was shown recently in several parallel studies. Phosphorylation of the central linker domain modulates the liquid nature of the droplets. The molecular basis of the stabilization of this phenomenon remains unknown in the case of SARS-CoV-2.

Recently, it has been shown that full length N protein undergoes liquid-liquid phase separation upon binding long RNA (more than 300 nucleotides), that phosphorylation impacts this process by affecting the liquid-like nature of the droplet, and that the viral polymerase colocalizes to these membraneless organelles [98] ; [158] ; [40] ; [188] [28]. LLPS upon interaction with RNA was found to be sequence specific, an observation that may be related to the sequence-dependent encapsidation observed for MeV N:RNA [136]. LLPS was also found to span physiologically relevant temperature ranges [98]. Interestingly, a recent investigation of the interactome of Sars-CoV-2 found that N interacts with G3BP1/2 [76], a known component of granules that store mRNA in response to stress [223]. A recent investigation of N:RNA condensates revealed that N colocalizes to G3BP1/2 stress granules [188]. Finally it has recently been shown that Sars-CoV-2 RNA-dependent RNA-polymerase localizes to N:RNA droplets, supporting the suggestion that the virus exploits condensation to enhance genome replication. It is not yet known which interactions are responsible for LLPS in Sars-CoV-2 and the process is not yet understood at atomic resolution.

6.2 Results

We optimised expression and purification of SARS-COV-2 nucleoprotein. Based on the predictions of structural propensity, we subdivided N into five domains: N_1, N_2, N_3, N_4 and N_5. The following consctructs were designed: N_{12}, N_3, N_{45}, N_{123}, N_{full}. Expression and purification were optimised.

The results are presented in the article in Annexe 2.

Next, NMR signal assignment of N_3 backbone was performed. For this 3D experiments HNCO, HNCACO, HNCA and HNCOCA were recorded and the assignment was done manually.

The results are presented in the article in Annexe 3.

Finally, we optimized expression and purification of full length N and tested conditions for its *in vitro* phase separation. Surprisingly, we could observe phase separation without adding RNA (Fig. 6.2). Either N indeed can phase separate on its own and in the presence of RNA the process is enhanced, or during the purification RNA from *E. coli* remains in the sample.

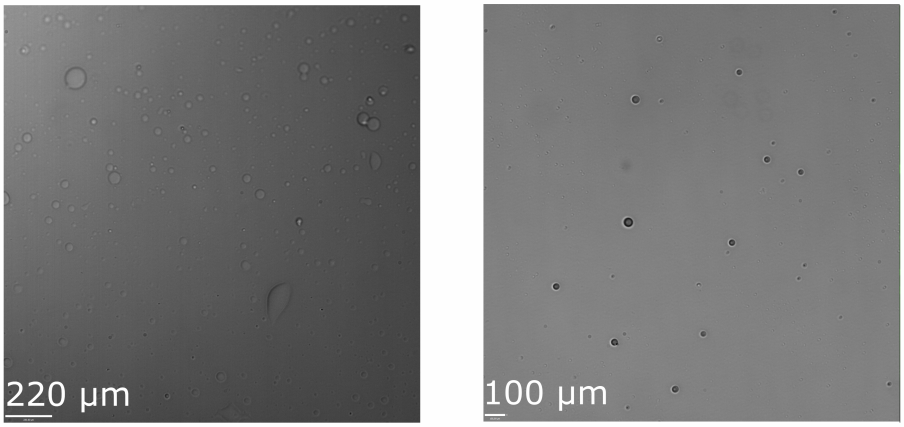


Figure 6.2 – DIC microscopy of SARS-COV-2 N.

6.3 Conclusions

Taking into account the importance of the SARS-CoV-2 study, we plan to continue SARS-CoV-2 N characterisation with a main focus on its behaviour in liquid-liquid phase separation.

Bibliographie

1. ABDELLA, R., AGGARWAL, M., OKURA, T., LAMB, R. A. & HE, Y. Structure of a paramyxovirus polymerase complex reveals a unique methyltransferase-CTD conformation. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 4931-4941. ISSN : 10916490 (2020).
2. ABYZOV, A., SALVI, N., SCHNEIDER, R., MAURIN, D., RUIGROK, R. W. H., JENSEN, M. R. & BLACKLEDGE, M. Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation. *Journal of the American Chemical Society* **138,** 6240-6251. ISSN : 0002-7863. https://doi.org/10.1021/jacs.6b02424 (2016).
3. ACKERMANN, B. E. & DEBELOUCHINA, G. T. Heterochromatin Protein HP1 Gelation Dynamics Revealed by Solid-State NMR Spectroscopy. *Angewandte Chemie International Edition* **58,** 6300-6305. eprint : https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201901141. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201901141 (2019).
4. ADAMSKI, W., SALVI, N., MAURIN, D., MAGNAT, J., MILLES, S., JENSEN, M. R., ABYZOV, A., MOREAU, C. J. & BLACKLEDGE, M. A Unified Description of Intrinsically Disordered Protein Dynamics under Physiological Conditions Using NMR Spectroscopy. *Journal of the American Chemical Society* **141,** 17817-17829. ISSN : 0002-7863. https://doi.org/10.1021/jacs.9b09002 (2019).
5. ALBERTI, S. The wisdom of crowds : Regulating cell function through condensed states of living matter. *Journal of Cell Science* **130,** 2789-2796. ISSN : 14779137 (2017).
6. ALBERTI, S., GLADFELTER, A. & MITTAG, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. eng. *Cell* **176,** 419-434. ISSN : 1097-4172. https://pubmed.ncbi.nlm.nih.gov/30682370%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6445271/ (2019).
7. ALBERTI, S., SAHA, S., WOODRUFF, J. B., FRANZMANN, T. M., WANG, J. & HYMAN, A. A. A User ' s Guide for Phase Separation Assays with Purified Proteins. *Journal of Molecular Biology* **430,** 4806-4820. ISSN : 0022-2836. https://doi.org/10.1016/j.jmb.2018.06.038 (2018).
8. AMARASINGHE, G. K. *et al.* Taxonomy of the order Mononegavirales : update 2019. eng. *Archives of virology* **164,** 1967-1980. ISSN : 1432-8798 (Electronic) (2019).
9. AMBADIPUDI, S., REDDY, J. G., BIERNAT, J., MANDELKOW, E. & ZWECKSTETTER, M. Residue-specific identification of phase separation hot spots of Alzheimer's-related protein tau. *Chem. Sci.* **10,** 6503-6507. http://dx.doi.org/10.1039/C9SC00531E (26 2019).
10. ANDREI, M. A., INGELFINGER, D., HEINTZMANN, R., ACHSEL, T., RIVERA-POMAR, R. & LÜHRMANN, R. A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. eng. *RNA (New York, N.Y.)* **11,** 717-727. ISSN : 1355-8382 (Print) (2005).
11. ANTHIS, N. J. & CLORE, G. M. Visualizing transient dark states by NMR spectroscopy. *Quarterly Reviews of Biophysics* **48,** 35-116. ISSN : 14698994 (2015).

12. BERNADÓ, P. & BLACKLEDGE, M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophysical Journal* **97,** 2839-2845. ISSN : 15420086. http://dx.doi.org/10.1016/j.bpj.2009.08.044 (2009).

13. BERNADÓ, P. & BLACKLEDGE, M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophysical Journal* **97,** 2839-2845. ISSN : 15420086. http://dx.doi.org/10.1016/j.bpj.2009.08.044 (2009).

14. BERRY, J., WEBER, S. C., VAIDYA, N., HAATAJA, M. & BRANGWYNNE, C. P. RNA transcription modulates phase transition-driven nuclear body assembly. *Proceedings of the National Academy of Sciences* **112,** E5237-E5245. ISSN : 0027-8424. eprint : https://www.pnas.org/content/112/38/E5237.full.pdf. https://www.pnas.org/content/112/38/E5237 (2015).

15. BLOYET, L. M., BRUNEL, J., DOSNON, M., HAMON, V., ERALES, J., GRUET, A., LAZERT, C., BIGNON, C., ROCHE, P., LONGHI, S. & GERLIER, D. Modulation of Re-initiation of Measles Virus Transcription at Intergenic Regions by PXDto NTAILBinding Strength. *PLoS Pathogens* **12,** 1-39. ISSN : 15537374 (2016).

16. BLOYET, L.-M., MORIN, B., BRUSIC, V., GARDNER, E., ROSS, R. A., VADAKKAN, T., KIRCHHAUSEN, T. & WHELAN, S. P. J. Oligomerization of the Vesicular Stomatitis Virus Phosphoprotein Is Dispensable for mRNA Synthesis but Facilitates RNA Replication. *Journal of Virology* **94,** 1-12. ISSN : 0022-538X (2020).

17. BOEYNAEMS, S., BOGAERT, E., KOVACS, D., KONIJNENBERG, A., TIMMERMAN, E., VOLKOV, A., GUHAROY, M., DE DECKER, M., JASPERS, T., RYAN, V. H., JANKE, A. M., BAATSEN, P., VERCRUYSSE, T., KOLAITIS, R. M., DAELEMANS, D., TAYLOR, J. P., KEDERSHA, N., ANDERSON, P., IMPENS, F., SOBOTT, F., SCHYMKOWITZ, J., ROUSSEAU, F., FAWZI, N. L., ROBBERECHT, W., VAN DAMME, P., TOMPA, P. & VAN DEN BOSCH, L. Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Molecular Cell* **65,** 1044-1055.e5. ISSN : 10974164. http://dx.doi.org/10.1016/j.molcel.2017.02.013 (2017).

18. BOKE, E., RUER, M., WÜHR, M., COUGHLIN, M., LEMAITRE, R., GYGI, S. P., ALBERTI, S., DRECHSEL, D., HYMAN, A. A. & MITCHISON, T. J. Amyloid-like Self-Assembly of a Cellular Compartment. *Cell* **166,** 637-650. ISSN : 0092-8674. https://doi.org/10.1016/j.cell.2016.06.051 (2016).

19. BORCHERDS, W., BREMER, A., BORGIA, M. B. & MITTAG, T. How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation ? arXiv : 2008.02917 [q-bio.BM] (2020).

20. BOUVIGNIES, G., BERNADÓ, P., MEIER, S., CHO, K., GRZESIEK, S., BRÜSCHWEILER, R. & BLACKLEDGE, M. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 13885-13890. ISSN : 00278424 (2005).

21. BRADY, J. P., FARBER, P. J., SEKHAR, A., LIN, Y.-H., HUANG, R., BAH, A., NOTT, T. J., CHAN, H. S., BALDWIN, A. J., FORMAN-KAY, J. D. & KAY, L. E. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences* **114,** E8194-E8203. ISSN : 0027-8424. eprint : https://www.pnas.org/content/114/39/E8194.full.pdf. https://www.pnas.org/content/114/39/E8194 (2017).

22. BROWN, C. J., JOHNSON, A. K. & DAUGHDRILL, G. W. Comparing Models of Evolution for Ordered and Disordered Proteins. *Molecular Biology and Evolution* **27,** 609-621. ISSN : 07374038 (2010).

23. BRUHN, J. F., HOTARD, A. L., SPIROPOULOU, C. F., LO, M. K. & SAPHIRE, E. O. A Conserved Basic Patch and Central Kink in the Nipah Virus Phosphoprotein Multimerization Domain Are Essential for Polymerase Function. *Structure* **27,** 1-9 (2019).

24. BRUTSCHER, B., FELLI, I. C., GIL-CABALLERO, S., HOŠEK, T., KÜMMERLE, R., PIAI, A., PIERATTELLI, R. & SÓLYOM, Z. NMR Methods for the Study of Instrinsically Disordered Proteins Structure, Dynamics, and Interactions : General Overview and Practical Guidelines. eng. *Advances in experimental medicine and biology* **870,** 49-122. ISSN : 0065-2598 (Print) (2015).

25. BUEVICH, A. V. & BAUM, J. Residue-specific real-time NMR diffusion experiments define the association states of proteins during folding. *Journal of the American Chemical Society* **124,** 7156-7162. ISSN : 00027863 (2002).

26. BURKE, K. A., JANKE, A. M., RHINE, C. L. & FAWZI, N. L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Article Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Molecular Cell* **60,** 231-241. ISSN : 1097-2765. http://dx.doi.org/10.1016/j.molcel.2015.09.006 (2015).

27. CAO, D., GAO, Y., ROESLER, C., RICE, S., D'CUNHA, P., ZHUANG, L., SLACK, J., DOMKE, M., ANTONOVA, A., ROMANELLI, S., KEATING, S., FORERO, G., JUNEJA, P. & LIANG, B. Cryo-EM structure of the respiratory syncytial virus RNA polymerase. *Nature Communications* **11,** 1-9. ISSN : 20411723. http://dx.doi.org/10.1038/s41467-019-14246-3 (2020).

28. CARLSON, C. R., ASFAHA, J. B., GHENT, C. M., HOWARD, C. J., HARTOONI, N., SAFARI, M., FRANKEL, A. D. & MORGAN, D. O. Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. *Molecular Cell* **80,** 1092-1103.e4. ISSN : 1097-2765. https://www.sciencedirect.com/science/article/pii/S1097276520308030 (2020).

29. CARSILLO, T., TRAYLOR, Z., CHOI, C., NIEWIESK, S. & OGLESBEE, M. hsp72, a Host Determinant of Measles Virus Neurovirulence. *Journal of Virology* **80,** 11031-11039. ISSN : 0022-538X (2006).

30. CARSILLO, T., TRAYLOR, Z., CHOI, C., NIEWIESK, S. & OGLESBEE, M. hsp72, a host determinant of measles virus neurovirulence. eng. *Journal of virology* **80,** 11031-11039. ISSN : 0022-538X (Print) (2006).

31. CAVANAGH, J., FAIRBROTHER, W. J., PALMER, A. G., RANCE, M. & SKELTON, N. J. *CHAPTER 2 - THEORETICAL DESCRIPTION OF NMR SPECTROSCOPY* Second Edition (éd. CAVANAGH, J., FAIRBROTHER, W. J., PALMER, A. G., RANCE, M. & SKELTON, N. J.) 29-113. ISBN : 978-0-12-164491-8. https://www.sciencedirect.com/science/article/pii/B978012164491850004X (Academic Press, Burlington, 2007).

32. CAVANAGH, J., FAIRBROTHER, W. J., PALMER, A. G., RANCE, M. & SKELTON, N. J. in *Protein NMR Spectroscopy (Second Edition)* (éd. CAVANAGH, J., FAIRBROTHER, W. J., PALMER, A. G., RANCE, M. & SKELTON, N. J.) Second Edition, 271-332 (Academic Press, Burlington, 2007). ISBN : 978-0-12-164491-8. https://www.sciencedirect.com/science/article/pii/B9780121644918500063.

33. CEVIK, B., HOLMES, D. E., VROTSOS, E., FELLER, J. A., SMALLWOOD, S. & MOYER, S. A. The phosphoprotein (P) and L binding sites reside in the N-terminus of the L subunit of the measles virus RNA polymerase. *Virology* **327,** 297-306. ISSN : 00426822 (2004).

34. CHEMES, L. B., ALONSO, L. G., NOVAL, M. G. & de PRAT-GAY, G. *Circular Dichroism Techniques for the Analysis of Intrinsically Disordered Proteins and Domains* (éd. UVERSKY, V. N. & DUNKER, A. K.) 387-404. ISBN : 978-1-61779-927-3. https://doi.org/10.1007/978-1-61779-927-3_22 (Humana Press, Totowa, NJ, 2012).

35. CHOI, J. M., DAR, F. & PAPPU, R. V. *LASSI : A lattice model for simulating phase transitions of multivalent proteins* 1-39. ISBN : 1111111111 (2019).

36. CHONG, P. A., VERNON, R. M. & FORMAN-KAY, J. D. RGG/RG Motif Regions in RNA Binding and Phase Separation. *Journal of Molecular Biology* **430.** Phase Separation in Biology and Disease, 4650-4665. ISSN : 0022-2836. https://www.sciencedirect.com/science/article/pii/S0022283618306107 (2018).

37. COMMUNIE, G., HABCHI, J., YABUKARSKI, F., BLOCQUEL, D., SCHNEIDER, R., TARBOURIECH, N., PAPAGEORGIOU, N., RUIGROK, R. W., JAMIN, M., JENSEN, M. R., LONGHI, S. & BLACKLEDGE, M. Atomic Resolution Description of the Interaction between the Nucleoprotein and Phosphoprotein of Hendra Virus. *PLoS Pathogens* **9.** ISSN : 15537366 (2013).

38. CONICELLA, A. E., ZERZE, G. H., MITTAL, J. & FAWZI, N. L. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. eng. *Structure (London, England : 1993)* **24,** 1537-1549. ISSN : 1878-4186 (Electronic) (2016).

39. CONIS, E. Measles and the Modern History of Vaccination. *Public Health Reports* **134,** 118-125. ISSN : 14682877 (2019).

40. CUBUK, J., ALSTON, J. J., INCICCO, J. J., SINGH, S., STUCHELL-BRERETON, M. D., WARD, M. D., ZIMMERMAN, M. I., VITHANI, N., GRIFFITH, D., WAGONER, J. A., BOWMAN, G. R., HALL, K. B., SORANNO, A. & HOLEHOUSE, A. S. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature Communications* **12,** 1936. ISSN : 2041-1723. https://doi.org/10.1038/s41467-021-21953-3 (2021).

41. DAS, D. & MUKHOPADHYAY, S. Studying backbone torsional dynamics of intrinsically disordered proteins using fluorescence depolarization kinetics. *Journal of Biosciences* **43,** 455-462. ISSN : 09737138 (2018).

42. DAS, R. K. & PAPPU, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 13392-13397. ISSN : 00278424 (2013).

43. DAS, T., SCHUSTER, A., SCHNEIDER-SCHAULIES, S. & BANERJEE, A. K. Involvement of Cellular Casein Kinase II in the Phosphorylation of Measles Virus P Protein : Identification of Phosphorylation Sites. *Virology* **211,** 218-226. ISSN : 0042-6822. http://www.sciencedirect.com/science/article/pii/S0042682285713943 (1995).

44. DE BRITO, T., DE FÁTIMA ARAUJO, M. & TIRIBA, A. Ultrastructure of the Negri body in human rabies. *Journal of the Neurological Sciences* **20,** 363-372. ISSN : 0022-510X. https://www.sciencedirect.com/science/article/pii/0022510X73901706 (1973).

45. DELAGLIO, F., GRZESIEK, S., VUISTER, G. W., ZHU, G., PFEIFER, J. & BAX, A. NMRPipe : a multidimensional spectral processing system based on UNIX pipes. eng. *Journal of biomolecular NMR* **6,** 277-293. ISSN : 0925-2738 (Print) (1995).

46. DERDOWSKI, A., PETERS, T. R., GLOVER, N., QIAN, R., UTLEY, T. J., BURNETT, A., WILLIAMS, J. V., SPEARMAN, P. & CROWE, J. E. Human metapneumovirus nucleoprotein and phosphoprotein interact and provide the minimal requirements for inclusion body formation. *Journal of General Virology* **89,** 2698-2708. ISSN : 00221317 (2008).

47. DESFOSSES, A., MILLES, S., JENSEN, M. R., GUSEVA, S., COLLETIER, J. P., MAURIN, D., SCHOEHN, G., GUTSCHE, I., RUIGROK, R. W. & BLACKLEDGE, M. Assembly and cryo-EM structures of RNA-specific measles virus nucleocapsids provide mechanistic insight into paramyxoviral replication. *Proceedings of the National Academy of Sciences of the United States of America* **116,** 4256-4264. ISSN : 10916490 (2019).

48. DEVAUX, P., PRINISKI, L. & CATTANEO, R. The measles virus phosphoprotein interacts with the linker domain of STAT1. eng. *Virology* **444,** 250-256. ISSN : 1096-0341 (Electronic) (2013).

49. DINESH, D. C., CHALUPSKA, D., SILHAN, J., KOUTNA, E., NENCKA, R., VEVERKA, V. & BOURA, E. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLOS Pathogens* **16,** 1-16. https://doi.org/10.1371/journal.ppat.1009100 (2020).

50. DRAKE, J. W. & HOLLAND, J. J. Mutation rates among RNA viruses. *Proceedings of the National Academy of Sciences* **96,** 13910-13913. ISSN : 0027-8424. eprint : https://www.pnas.org/content/96/24/13910.full.pdf. https://www.pnas.org/content/96/24/13910 (1999).

51. DUDÁS, E. F. & BODOR, A. Quantitative, Diffusion NMR Based Analytical Tool to Distinguish Folded, Disordered, and Denatured Biomolecules. *Analytical Chemistry* **91,** 4929-4933. ISSN : 15206882 (2019).

52. DÜX, A., LEQUIME, S., PATRONO, L. V., VRANCKEN, B., BORAL, S., GOGARTEN, J. F., HILBIG, A., HORST, D., MERKEL, K., PREPOINT, B., SANTIBANEZ, S., SCHLOTTERBECK, J., SUCHARD, M. A., ULRICH, M., WIDULIN, N., MANKERTZ, A., LEENDERTZ, F. H., HARPER, K., SCHNALKE, T., LEMEY, P. & CALVIGNAC-SPENCER, S. Measles virus and rinderpest virus divergence dated to the sixth century BCE. *Science* **368,** 1367 LP -1370. http://science.sciencemag.org/content/368/6497/1367.abstract (2020).

53. DYSON, H. J. Making Sense of Intrinsically Disordered Proteins. *Biophysical Journal* **110,** 1013-1016. ISSN : 15420086. http://dx.doi.org/10.1016/j.bpj.2016.01.030 (2016).

54. ELBAUM-GARFINKLE, S., KIM, Y., SZCZEPANIAK, K. & CHEN, C. C.-h. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. **112** (2015).

55. FAWZI, N. L., YING, J., GHIRLANDO, R., TORCHIA, D. A. & CLORE, G. M. protofibrils probed by solution NMR. *Nature* **480,** 268-272 (2012).

56. FEARNS, R., YOUNG, D. F. & RANDALL, R. E. Evidence that the paramyxovirus simian virus 5 can establish quiescent infections by remaining inactive in cytoplasmic inclusion bodies. *Journal of General Virology* **75,** 3525-3539. ISSN : 00221317 (1994).

57. FERIC, M. & BRANGWYNNE, C. P. A nuclear F-actin scaffold stabilizes ribonucleoprotein droplets against gravity in large cells. eng. *Nature cell biology* **15,** 1253-1259. ISSN : 1476-4679. https://pubmed.ncbi.nlm.nih.gov/23995731%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3789854/ (2013).

58. FERIC, M., VAIDYA, N., HARMON, T. S., MITREA, D. M., ZHU, L., RICHARDSON, T. M., KRIWACKI, R. W., PAPPU, R. V. & BRANGWYNNE, C. P. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165,** 1686-1697. ISSN : 10974172. http://dx.doi.org/10.1016/j.cell.2016.04.047 (2016).

59. FETTERS, L. J., HADJICHRISTIDIS, N., LINDNER, J. S. & MAYS, J. W. Molecular Weight Dependence of Hydrodynamic and Thermodynamic Properties for Well-Defined Linear Polymers in Solution. *Journal of Physical and Chemical Reference Data* **23,** 619-640. eprint : https://doi.org/10.1063/1.555949. https://doi.org/10.1063/1.555949 (1994).

60. FISCHER, H., POLIKARPOV, I. & CRAIEVICH, A. F. Average protein density is a molecular-weight-dependent function. *Protein Science* **13,** 2825-2828. eprint : https://onlinelibrary.wiley.com/doi/pdf/10.1110/ps.04688204. https://onlinelibrary.wiley.com/doi/abs/10.1110/ps.04688204 (2004).

61. FISHER, J. *Modern NMR techniques for synthetic chemistry* 1-330. ISBN : 9781466592254 (2014).

62. FLORY, P. J. Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics* **10,** 51-61. eprint : https://doi.org/10.1063/1.1723621. https://doi.org/10.1063/1.1723621 (1942).

63. FOOTE, A. K., MANGER, L. H., HOLDEN, M. R., MARGITTAI, M. & GOLDSMITH, R. H. Time-resolved multirotational dynamics of single solution-phase tau proteins reveals details of conformational variation. *Phys. Chem. Chem. Phys.* **21,** 1863-1871. http://dx.doi.org/10.1039/C8CP06971A (4 2019).

64. FOUQUET, B., NIKOLIC, J., LARROUS, F., BOURHY, H., WIRBLICH, C., LAGAUDRIÈRE-GESBERT, C. & BLONDEL, D. Focal Adhesion Kinase Is Involved in Rabies Virus Infection through Its Interaction with Viral Phosphoprotein P. *Journal of Virology* **89,** 1640-1651. ISSN : 0022-538X (2015).

65. FUENTES-MONTEVERDE, J. C., BECKER, S. & REZAEI-GHALEH, N. Biomolecular phase separation through the lens of sodium-23 NMR. *Protein Science* **n/a.** eprint : https://onlinelibrary.wiley.com/doi/pdf/10.1002/pro.4010. https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4010.

66. FUJIMURA, K., SUZUKI, T., YASUDA, Y., MURATA, M., KATAHIRA, J. & YONEDA, Y. Identification of importin 1 as a novel constituent of RNA stress granules. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1803,** 865-871. ISSN : 0167-4889. https://www.sciencedirect.com/science/article/pii/S0167488910000972 (2010).

67. GABRYELCZYK, B., CAI, H., SHI, X., SUN, Y., SWINKELS, P. J. M., SALENTINIG, S., PERVUSHIN, K. & MISEREZ, A. Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides. *Nature Communications.* ISSN : 2041-1723. http://dx.doi.org/10.1038/s41467-019-13469-8.

68. GANSER, L. R. & MYONG, S. Methods to Study Phase-Separated Condensates and the Underlying Molecular Interactions. *Trends in Biochemical Sciences* **45,** 1004-1005. ISSN : 13624326. https://doi.org/10.1016/j.tibs.2020.05.011 (2020).

69. GARCÍA-BARRENO, B., DELGADO, T. & MELERO, J. A. Identification of protein regions involved in the interaction of human respiratory syncytial virus phosphoprotein and nucleoprotein : significance for nucleocapsid assembly and formation of cytoplasmic inclusions. *Journal of virology* **70,** 801-808. ISSN : 0022-538X (1996).

70. GHOSH, A., MAZARAKOS, K. & ZHOU, H.-x. Three archetypical classes of macromolecular regulators of protein liquid – liquid phase separation. **116** (2019).

71. GIBSON, T. J. Cell regulation : determined to signal discrete cooperation. *Trends in Biochemical Sciences* **34,** 471-482. ISSN : 09680004 (2009).

72. GILL, M. L., BYRD, R. A. & PALMER, A. G. Dynamics of GCN4 facilitate DNA interaction : A model-free analysis of an intrinsically disordered region. *Physical Chemistry Chemical Physics* **18,** 5839-5849. ISSN : 14639076 (2016).

73. GILMAN, M. S. A., LIU, C., FUNG, A., BEHERA, I., JORDAN, P., RIGAUX, P., YSEBAERT, N., TCHERNIUK, S., SOURIMANT, J., ELÉOUËT, J.-F., SUTTO-ORTIZ, P., DECROLY, E., ROYMANS, D., JIN, Z. & MCLELLAN, J. S. Structure of the Respiratory Syncytial Virus Polymerase Complex. *Cell* **179,** 193-204.e14. ISSN : 0092-8674. https://doi.org/10.1016/j.cell.2019.08.014 (2019).

74. GIUNTA, C. J. & MAINZ, V. V. Discovery of Nuclear Magnetic Resonance : Rabi, Purcell, and Bloch. *ACS Symposium Series* **1349,** 3-20. ISSN : 19475918 (2020).

75. GOMES, G. N., KRZEMINSKI, M., MARTIN, E. W., MITTAG, T., HEAD-GORDON, T., FORMAN-KAY, J. D. & GRADINARU, C. C. Integrating smFRET, SAXS and NMR data to infer structural ensembles of an intrinsically-disordered protein. *bioRxiv,* 1-25 (2020).

76. GORDON, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583,** 459-468. ISSN : 1476-4687. https://doi.org/10.1038/s41586-020-2286-9 (2020).

77. GOTOH, B., KOMATSU, T., TAKEUCHI, K. & YOKOO, J. The C-terminal half-fragment of the Sendai virus C protein prevents the gamma-activated factor from binding to a gamma-activated sequence site. eng. *Virology* **316,** 29-40. ISSN : 0042-6822 (Print) (2003).

78. GRIMA, R. Intrinsic biochemical noise in crowded intracellular conditions. *The Journal of Chemical Physics* **132,** 185102. eprint : https://doi.org/10.1063/1.3427244. https://doi.org/10.1063/1.3427244 (2010).

79. GUERRA, F. M., BOLOTIN, S., LIM, G., HEFFERNAN, J., DEEKS, S. L., LI, Y. & CROWCROFT, N. S. The basic reproduction number (R0) of measles : a systematic review. *The Lancet Infectious Diseases* **17,** e420-e428. ISSN : 14744457. http://dx.doi.org/10.1016/S1473-3099(17)30307-9 (2017).

80. GUILLÉN-BOIXET, J., KOPACH, A., HOLEHOUSE, A. S., WITTMANN, S., JAHNEL, M., SCHLÜSSLER, R., KIM, K., TRUSSINA, I. R., WANG, J., MATEJU, D., POSER, I., MAHARANA, S., RUER-GRUSS, M., RICHTER, D., ZHANG, X., CHANG, Y.-T., GUCK, J., HONIGMANN, A., MAHAMID, J., HYMAN, A. A., PAPPU, R. V., ALBERTI, S. & FRANZMANN, T. M. RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181,** 346-361.e17. ISSN : 0092-8674. https://www.sciencedirect.com/science/article/pii/S0092867420303421 (2020).

81. GUINIER, ANDRÉ. La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys.* **11,** 161-237. https://doi.org/10.1051/anphys/193911120161 (1939).

82. GUSEVA, S., MILLES, S., BLACKLEDGE, M. & RUIGROK, R. W. H. The Nucleoprotein and Phosphoprotein of Measles Virus. *Frontiers in Microbiology* **10,** 1832. ISSN : 1664-302X. https://www.frontiersin.org/article/10.3389/fmicb.2019.01832 (2019).

83. GUTSCHE, I., DESFOSSES, A., EFFANTIN, G., LING, W. L., HAUPT, M., RUIGROK, R. W., SACHSE, C. & SCHOEHN, G. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* **348,** 704-707. ISSN : 10959203 (2015).

84. HAGIWARA, K., SATO, H., INOUE, Y., WATANABE, A., YONEDA, M., IKEDA, F., FUJITA, K., FUKUDA, H., TAKAMURA, C., KOZUKA-HATA, H., OYAMA, M., SUGANO, S., OHMI, S. & KAI, C. Phosphorylation of measles virus nucleoprotein upregulates the transcriptional activity of minigenomic RNA. *PROTEOMICS* **8,** 1871-1879. eprint : https://onlinelibrary.wiley.com/doi/pdf/10.1002/pmic.200701051. https://onlinelibrary.wiley.com/doi/abs/10.1002/pmic.200701051 (2008).

85. HALLE, B. & DAVIDOVIC, M. Biomolecular hydration : From water dynamics to hydrodynamics. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 12135-12140. ISSN : 00278424 (2003).

86. HANSEN, D. F., VALLURUPALLI, P. & KAY, L. E. An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *Journal of Physical Chemistry B* **112,** 5898-5904. ISSN : 15206106 (2008).

87. HARMON, T. S., HOLEHOUSE, A. S., ROSEN, M. K. & PAPPU, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *bioRxiv,* 1-31. ISSN : 2050-084X (2017).

88. HARTLMÜLLER, C., SPREITZER, E., GÖBL, C., FALSONE, F. & MADL, T. NMR characterization of solvent accessibility and transient structure in intrinsically disordered proteins. *Journal of Biomolecular NMR* **73,** 305-317. ISSN : 15735001. https://doi.org/10.1007/s10858-019-00248-2 (2019).

89. HARTMAN, N. C. & GROVES, J. T. Signaling clusters in the cell membrane. *Current Opinion in Cell Biology* **23.** Membranes and organelles, 370-376. ISSN : 0955-0674. https://www.sciencedirect.com/science/article/pii/S095506741100072X (2011).

90. HEINRICH, B. S., MALIGA, Z., STEIN, D. A., HYMAN, A. A. & WHELAN, S. P. Phase transitions drive the formation of vesicular stomatitis virus replication compartments. *mBio* **9,** 1-10. ISSN : 21507511 (2018).

91. HELMUS, J. J. & JARONIEC, C. P. Nmrglue : an open source Python package for the analysis of multidimensional NMR data. *Journal of Biomolecular NMR* **55,** 355-367. ISSN : 1573-5001. https://doi.org/10.1007/s10858-013-9718-x (2013).

92. HOISCHEN, C., MONAJEMBASHI, S., WEISSHART, K. & HEMMERICH, P. Multimodal Light Microscopy Approaches to Reveal Structural and Functional Properties of Promyelocytic Leukemia Nuclear Bodies. *Frontiers in Oncology* **8,** 125. ISSN : 2234-943X. https://www.frontiersin.org/article/10.3389/fonc.2018.00125 (2018).

93. HORIKAMI, S. M., SMALLWOOD, S., BANKAMP, B. & MOYER, S. A. An amino-proximal domain of the L protein binds to the P protein in the measles virus RNA polymerase complex. eng. *Virology* **205,** 540-545. ISSN : 0042-6822 (Print) (1994).

94. HORWITZ, J. A., JENNI, S., HARRISON, S. C. & WHELAN, S. P. Structure of a rabies virus polymerase complex from electron cryo-microscopy. *bioRxiv* (2019).

95. HORWITZ, J. A., JENNI, S., HARRISON, S. C. & WHELAN, S. P. Structure of a rabies virus polymerase complex from electron cryo-microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 2099-2107. ISSN : 10916490 (2020).

96. HUGGINS, M. L. Some Properties of Solutions of Long-chain Compounds. *The Journal of Physical Chemistry* **46,** 151-158. ISSN : 0092-7325. https://doi.org/10.1021/j150415a018 (1942).

97. IJAVI, M., STYLE, R. W., EMMANOUILIDIS, L., KUMAR, A., MEIER, S. M., TORZYNSKI, A. L., ALLAIN, F. H. T., BARRAL, Y., STEINMETZ, M. O. & DUFRESNE, E. R. *Surface tensiometry of phase separated protein and polymer droplets by the sessile drop method* 2020. arXiv : 2007.13502 [cond-mat.soft].

98. ISERMAN, C., RODEN, C., BOERNEKE, M., SEALFON, R., MCLAUGHLIN, G., JUNGREIS, I., PARK, C., BOPPANA, A., FRITCH, E., HOU, Y. J., THEESFELD, C., TROYANSKAYA, O. G., BARIC, R. S., SHEAHAN, T. P., WEEKS, K. & GLADFELTER, A. S. Specific viral RNA drives the SARS CoV-2 nucleocapsid to phase separate. eng. *bioRxiv : the preprint server for biology,* 2020.06.11.147199. https://pubmed.ncbi.nlm.nih.gov/32587965%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7310621/ (2020).

99. IVERSON, L. E. & ROSE, J. K. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. *Cell* **23,** 477-484. ISSN : 0092-8674. https://www.sciencedirect.com/science/article/pii/0092867481901434 (1981).

100. JAIN, N., BHATTACHARYA, M. & MUKHOPADHYAY, S. Chain collapse of an amyloidogenic intrinsically disordered protein. *Biophysical Journal* **101,** 1720-1729. ISSN : 00063495. http://dx.doi.org/10.1016/j.bpj.2011.08.024 (2011).

101. JENNI, S., BLOYET, L. M., DIAZ-AVALOS, R., LIANG, B., WHELAN, S. P., GRIGORIEFF, N. & HARRISON, S. C. Structure of the Vesicular Stomatitis Virus L Protein in Complex with Its Phosphoprotein Cofactor. *Cell Reports* **30,** 53-60.e5. ISSN : 22111247. https://doi.org/10.1016/j.celrep.2019.12.024 (2020).

102. JENSEN, M. R., COMMUNIE, G., RIBEIRO, E. A., MARTINEZ, N., DESFOSSES, A., SALMON, L., MOLLICA, L., GABEL, F., JAMIN, M., LONGHI, S., RUIGROK, R. W. & BLACKLEDGE, M. Intrinsic disorder in measles virus nucleocapsids. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 9839-9844. ISSN : 00278424 (2011).

103. JENSEN, M. R., HOUBEN, K., LESCOP, E., BLANCHARD, L., RUIGROK, R. W. H. & BLACKLEDGE, M. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings : application to the molecular recognition element of Sendai virus nucleoprotein. *J Am Chem Soc* **130,** 8055-61 (2008).

104. JENSEN, M. R., MARKWICK, P. R. L., MEIER, S., GRIESINGER, C., ZWECKSTETTER, M., GRZESIEK, S., BERNADÓ, P. & BLACKLEDGE, M. Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure* **17,** 1169-1185. ISSN : 09692126 (2009).

105. JENSEN, M. R., SALMON, L., NODET, G. & BLACKLEDGE, M. Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *Journal of the American Chemical Society* **132,** 1270-1272. ISSN : 15205126 (2010).

106. JIANG, Y., QIN, Y. & CHEN, M. Host-pathogen interactions in measles virus replication and anti-viral immunity. *Viruses* **8.** ISSN : 19994915 (2016).

107. JUNG, Y.-S. & ZWECKSTETTER, M. Mars - robust automatic backbone assignment of proteins. *Journal of Biomolecular NMR* **30,** 11-23. ISSN : 1573-5001. https://doi.org/10.1023/B:JNMR.0000042954.99056.ad (2004).

108. KARLIN, D. & BELSHAW, R. Detecting remote sequence homology in disordered proteins : Discovery of conserved motifs in the N-termini of mononegavirales phosphoproteins. *PLoS ONE* **7.** ISSN : 19326203 (2012).

109. KATOH, H., NAKATSU, Y., KUBOTA, T., SAKATA, M., TAKEDA, M. & KIDOKORO, M. Mumps Virus Is Released from the Apical Surface of Polarized Epithelial Cells, and the Release Is Facilitated by a Rab11-Mediated Transport System. *Journal of Virology* **89,** 12026-12034. ISSN : 0022-538X (2015).

110. KEDERSHA, N. & ANDERSON, P. Stress granules : Sites of mRNA triage that regulate mRNA stability and translatability. *Biochemical Society Transactions* **30,** 963-969. ISSN : 03005127 (2002).

111. KIM, T. H., TSANG, B., VERNON, R. M., SONENBERG, N., KAY, L. E. & FORMAN-KAY, J. D. Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. eng. *Science (New York, N.Y.)* **365,** 825-829. ISSN : 1095-9203 (Electronic) (2019).

112. KLEIN, I. A. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science* **368,** 1386 LP -1392. http://science.sciencemag.org/content/368/6497/1386.abstract (2020).

113. KLOSIN, A., OLTSCH, F., HARMON, T., HONIGMANN, A., JÜLICHER, F., HYMAN, A. A. & ZECHNER, C. Phase separation provides a mechanism to reduce noise in cells. *Science* **367,** 464-468. ISSN : 0036-8075. eprint : https://science.sciencemag.org/content/367/6476/464.full.pdf. https://science.sciencemag.org/content/367/6476/464 (2020).

114. KOHN, J. E., MILLETT, I. S., JACOB, J., ZAGROVIC, B., DILLON, T. M., CINGEL, N., DOTHAGER, R. S., SEIFERT, S., THIYAGARAJAN, P., SOSNICK, T. R., HASAN, M. Z., PANDE, V. S., RUCZINSKI, I., DONIACH, S. & PLAXCO, K. W. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences* **101,** 12491-12496. ISSN : 0027-8424. eprint : https://www.pnas.org/content/101/34/12491.full.pdf. https://www.pnas.org/content/101/34/12491 (2004).

115. KOSOL, S., CONTRERAS-MARTOS, S., CEDEÑO, C. & TOMPA, P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **18,** 10802-10828. ISSN : 14203049 (2013).

116. KRAINER, G., WELSH, T., JOSEPH, J., ESPINOSA, J., WITTMANN, S., de CSILLÉRY, E., SRIDHAR, A., TOPRAKCIOGLU, Z., GUDIŠKYTĖ, G., CZEKALSKA, M., ARTER, W., GEORGE-HYSLOP, P. S., HYMAN, A., COLLEPARDO-GUEVARA, R., ALBERTI, S. & KNOWLES, T. Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions, 1-22 (2020).

117. LAHAYE, X., VIDY, A., FOUQUET, B. & BLONDEL, D. Hsp70 Protein Positively Regulates Rabies Virus Infection. *Journal of Virology* **86,** 4743-4751. ISSN : 0022-538X (2012).

118. LAHAYE, X., VIDY, A., POMIER, C., OBIANG, L., HARPER, F., GAUDIN, Y. & BLONDEL, D. Functional Characterization of Negri Bodies (NBs) in Rabies Virus-Infected Cells : Evidence that NBs Are Sites of Viral Transcription and Replication. *Journal of Virology* **83,** 7948-7958. ISSN : 0022-538X (2009).

119. LAKOWICZ, J. R. *Principles of fluorescence spectroscopy* 3rd ed. eng. ISBN : 9780387312781. http://lib.ugent.be/catalog/rug01:001953566 (New York : Springer, 2006).

120. LAKSONO, B. M., de VRIES, R. D., VERBURGH, R. J., VISSER, E. G., de JONG, A., FRAAIJ, P. L., RUIJS, W. L., NIEUWENHUIJSE, D. F., van den HAM, H. J., KOOPMANS, M. P., van ZELM, M. C., OSTERHAUS, A. D. & de SWART, R. L. Studies into the mechanism of measles-associated immune suppression during a measles outbreak in the Netherlands. *Nature Communications* **9,** 1-10. ISSN : 20411723. http://dx.doi.org/10.1038/s41467-018-07515-0 (2018).

121. LAMOND, A. I. & SPECTOR, D. L. Nuclear speckles : A model for nuclear organelles. *Nature Reviews Molecular Cell Biology* **4,** 605-612. ISSN : 14710072 (2003).

122. LEBLANC, S. J., KULKARNI, P. & WENINGER, K. R. Single molecule FRET : A powerful tool to study intrinsically disordered proteins. *Biomolecules* **8.** ISSN : 2218273X (2018).

123. LEE, W., TONELLI, M. & MARKLEY, J. L. NMRFAM-SPARKY : enhanced software for biomolecular NMR spectroscopy. eng. *Bioinformatics (Oxford, England)* **31,** 1325-1327. ISSN : 1367-4811. https://pubmed.ncbi.nlm.nih.gov/25505092%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393527/ (2015).

124. LEMKE, E. A. Site-specific labeling of proteins for single-molecule FRET measurements using genetically encoded ketone functionalities. eng. *Methods in molecular biology (Clifton, N.J.)* **751,** 3-15. ISSN : 1940-6029 (Electronic) (2011).

125. LI, P., BANJADE, S., CHENG, H.-C., KIM, S., CHEN, B., GUO, L., LLAGUNO, M., HOLLINGSWORTH, J. V., KING, D. S., BANANI, S. F., RUSSO, P. S., JIANG, Q.-X., NIXON, B. T. & ROSEN, M. K. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483,** 336-340. ISSN : 1476-4687. https://doi.org/10.1038/nature10879 (2012).

126. LI, Z., GUO, D., QIN, Y. & CHEN, M. PI4KB on Inclusion Bodies Formed by ER Membrane Remodeling Facilitates Replication of Human Parainfluenza Virus Type 3. eng. *Cell reports* **29,** 2229-2242.e4. ISSN : 2211-1247 (Electronic) (2019).

127. LIANG, B., LI, Z., JENNI, S., RAHMEH, A. A., MORIN, B. M., GRANT, T., GRIGORIEFF, N., HARRISON, S. C. & WHELAN, S. P. J. Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell* **162,** 314-327. ISSN : 0092-8674. https://doi.org/10.1016/j.cell.2015.06.018 (2015).

128. LIPAD, G. & SZABO, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. Theory and Range of Validity. *JACS* **104,** 45464559. ISSN : 0092-8674. https://www.sciencedirect.com/science/article/pii/0092867481901434 (1982).

129. MAJUMDAR, A., DOGRA, P., MAITY, S. & MUKHOPADHYAY, S. Liquid Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered Protein Molecules (2019).

130. MALUR, A. G., GUPTA, N. K., DE, B. P. & BANERJEE, A. K. Analysis of the mutations in the active site of the RNA-dependent RNA polymerase of human parainfluenza virus type 3 (HPIV3). *Gene Expression* **10,** 93-100. ISSN : 10522166 (2002).

131. MARSH, J. A., SINGH, V. K., JIA, Z. & FORMAN-KAY, J. D. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein : Implications for fibrillation. *Protein Science* **15,** 2795-2804. ISSN : 09618368 (2006).

132. MARTIN, E. W., HOLEHOUSE, A. S., PERAN, I., FARAG, M., INCICCO, J. J., BREMER, A., GRACE, C. R., SORANNO, A., PAPPU, R. V. & MITTAG, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367,** 694-699. ISSN : 10959203 (2020).

133. MARTIN, E. W. & MITTAG, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* **57,** 2478-2487. ISSN : 15204995 (2018).

134. MASTERS, P. S. & BANERJEE, A. K. Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J Virol* **62,** 2658-64 (1988).

135. MCCONNELL, H. M. Reaction Rates by Nuclear Magnetic Resonance. *The Journal of Chemical Physics* **28,** 430-431. eprint : https://doi.org/10.1063/1.1744152. https://doi.org/10.1063/1.1744152 (1958).

136. MILLES, S., JENSEN, M. R., COMMUNIE, G., MAURIN, D., SCHOEHN, G., RUIGROK, R. W. & BLACKLEDGE, M. Self-Assembly of Measles Virus Nucleocapsid-like Particles : Kinetics and RNA Sequence Dependence. *Angewandte Chemie - International Edition* **55,** 9356-9360. ISSN : 15213773 (2016).

137. MILLES, S., JENSEN, M. R., LAZERT, C., GUSEVA, S., IVASHCHENKO, S., COMMUNIE, G., MAURIN, D., GERLIER, D., RUIGROK, R. W. & BLACKLEDGE, M. An ultraweak interaction in the intrinsically disordered replication machinery is essential for measles virus function. *Science Advances* **4,** 1-11. ISSN : 23752548 (2018).

138. MILLES, S., MERCADANTE, D., ARAMBURU, I. V., JENSEN, M. R., BANTERLE, N., KOEHLER, C., TYAGI, S., CLARKE, J., SHAMMAS, S. L., BLACKLEDGE, M., GRÄTER, F. & LEMKE, E. A. Plasticity of an Ultrafast Interaction between Nucleoporins and Nuclear Transport Receptors. *Cell* **163,** 734-745. ISSN : 10974172 (2015).

139. MINA, M. J., KULA, T., LENG, Y., LI, M., VRIES, R. D. D., KNIP, M., SILJANDER, H., REWERS, M., CHOY, D. F., WILSON, M. S., LARMAN, H. B., NELSON, A. N., GRIFFIN, D. E., SWART, R. L. D. & ELLEDGE, S. J. other pathogens. **606,** 599-606 (2019).

140. MITREA, D. M., CIKA, J. A., GUY, C. S., BAN, D., BANERJEE, P. R., STANLEY, C. B., NOURSE, A., DENIZ, A. A. & KRIWACKI, R. W. Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA, 1-33 (2016).

141. MONAHAN, Z., RYAN, V. H., JANKE, A. M., BURKE, K. A., RHOADS, S. N., ZERZE, G. H., O'MEALLY, R., DIGNON, G. L., CONICELLA, A. E., ZHENG, W., BEST, R. B., COLE, R. N., MITTAL, J., SHEWMAKER, F. & FAWZI, N. L. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. eng. *The EMBO journal* **36,** 2951-2967. ISSN : 1460-2075. https://pubmed.ncbi.nlm.nih.gov/28790177%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5641905/ (2017).

142. MOYER, S. A., BAKER, S. C. & HORIKAMI, S. M. Host cell proteins required for measles virus reproduction. *Journal of General Virology* **71,** 775-783. ISSN : 00221317 (1990).

143. MÜHLEBACH, M. D., MATEO, M., SINN, P. L., PRÜFER, S., KATHARINA, M., LEONARD, V. H. J., NAVARATNARAJAH, C. K., FRENZKE, M., XIAO, X., SAWATSKY, B., RAMACHANDRAN, S., JR, P. B. M. & CICHUTEK, K. receptor for measles virus. **480,** 530-533 (2012).

144. MUKRASCH, M. D., BIBOW, S., KORUKOTTU, J., JEGANATHAN, S., BIERNAT, J., GRIESINGER, C., MANDELKOW, E. & ZWECKSTETTER, M. Structural polymorphism of 441-residue Tau at single residue resolution. *PLoS Biology* **7,** 0399-0414. ISSN : 15449173 (2009).

145. MUNDAY, D. C., WU, W., SMITH, N., FIX, J., NOTON, S. L., GALLOUX, M., TOUZELET, O., ARMSTRONG, S. D., DAWSON, J. M., ALJABR, W., EASTON, A. J., RAMEIX-WELTI, M.-A., de OLIVEIRA, A. P., SIMABUCO, F. M., VENTURA, A. M., HUGHES, D. J., BARR, J. N., FEARNS, R., DIGARD, P., ELÉOUËT, J.-F. & HISCOX, J. A. Interactome Analysis of the Human Respiratory Syncytial Virus RNA Polymerase Complex Identifies Protein Chaperones as Important Cofactors That Promote L-Protein Stability and RNA Synthesis. *Journal of Virology* **89,** 917-930. ISSN : 0022-538X (2015).

146. MURTHY, A. C., DIGNON, G. L., KAN, Y., ZERZE, G. H., PAREKH, S. H., MITTAL, J. & FAWZI, N. L. Molecular interactions underlying liquid liquid phase separation of the FUS low-complexity domain. *Nature Structural & Molecular Biology* **26.** ISSN : 1545-9985. http://dx.doi.org/10.1038/s41594-019-0250-x (2019).

147. NIKOLIC, J., LE BARS, R., LAMA, Z., SCRIMA, N., LAGAUDRIÈRE-GESBERT, C., GAUDIN, Y. & BLONDEL, D. Negri bodies are viral factories with properties of liquid organelles. *Nature Communications* **8,** 1-12. ISSN : 20411723. http://dx.doi.org/10.1038/s41467-017-00102-9 (2017).

148. NODET, G., SALMON, L., OZENNE, V., MEIER, S., JENSEN, M. R. & BLACKLEDGE, M. Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *Journal of the American Chemical Society* **131,** 17908-17918. ISSN : 0002-7863. https://doi.org/10.1021/ja9069024 (2009).

149. O'FLYNN, B. G. & MITTAG, T. The role of liquid–liquid phase separation in regulating enzyme activity. *Current Opinion in Cell Biology* **69,** 70-79. ISSN : 18790410. arXiv : 2010.11343 (2021).

150. OATES, M. E., ROMERO, P., ISHIDA, T., GHALWASH, M., MIZIANTY, M. J., XUE, B., DOSZTÁNYI, Z., UVERSKY, V. N., OBRADOVIC, Z., KURGAN, L., DUNKER, A. K. & GOUGH, J. D2P2 : database of disordered protein predictions. *Nucleic Acids Research* **41,** D508-D516. ISSN : 0305-1048. eprint : https://academic.oup.com/nar/article-pdf/41/D1/D508/3693476/gks1226.pdf. https://doi.org/10.1093/nar/gks1226 (2012).

151. OMI-FURUTANI, M., YONEDA, M., FUJITA, K., IKEDA, F. & KAI, C. Novel Phosphoprotein-Interacting Region in Nipah Virus Nucleocapsid Protein and Its Involvement in Viral Replication. *Journal of Virology* **84,** 9793-9799. ISSN : 0022-538X (2010).

152. OZENNE, V., SCHNEIDER, R., YAO, M., HUANG, J. R., SALMON, L., ZWECKSTETTER, M., JENSEN, M. R. & BLACKLEDGE, M. Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *Journal of the American Chemical Society* **134,** 15138-15148. ISSN : 00027863 (2012).

153. PALMER, A. G. NMR characterization of the dynamics of biomacromolecules. *Chemical Reviews* **104,** 3623-3640. ISSN : 00092665 (2004).

154. PAN, J., QIAN, X., LATTMANN, S., EL SAHILI, A., YEO, T. H., JIA, H., CRESSEY, T., LUDEKE, B., NOTON, S., KALOCSAY, M., FEARNS, R. & LESCAR, J. Structure of the human metapneumovirus polymerase phosphoprotein complex. *Nature* **577,** 275-279. ISSN : 1476-4687. https://doi.org/10.1038/s41586-019-1759-1 (2020).

155. PARK, S., BARNES, R., LIN, Y., jin JEON, B., NAJAFI, S., DELANEY, K. T., FREDRICKSON, G. H., SHEA, J. E., HWANG, D. S. & HAN, S. Dehydration entropy drives liquid-liquid phase separation by molecular crowding. *Communications Chemistry* **3,** 1-12. ISSN : 23993669 (2020).

156. PARKS, G. D. Mapping of a region of the paramyxovirus L protein required for the formation of a stable complex with the viral phosphoprotein P. *Journal of Virology* **68,** 4862-4872. ISSN : 0022-538X (1994).

157. PELUPESSY, P., ESPALLARGAS, G. M. & BODENHAUSEN, G. Symmetrical reconversion : measuring cross-correlation rates with enhanced accuracy. *Journal of Magnetic Resonance* **161,** 258-264. ISSN : 1090-7807. https://www.sciencedirect.com/science/article/pii/S1090780702001908 (2003).

158. PERDIKARI, T. M., MURTHY, A. C., RYAN, V. H., WATTERS, S., NAIK, M. T. & FAWZI, N. L. SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *The EMBO Journal* **39,** e106478. eprint : https://www.embopress.org/doi/pdf/10.15252/embj.2020106478. https://www.embopress.org/doi/abs/10.15252/embj.2020106478 (2020).

159. PERRY, S. L. Phase separation : Bridging polymer physics and biology. *Current Opinion in Colloid and Interface Science* **39,** 86-97. ISSN : 18790399. https://doi.org/10.1016/j.cocis.2019.01.007 (2019).

160. PLATTET, P., ALVES, L., HERREN, M. & AGUILAR, H. C. Measles virus fusion protein : Structure, function and inhibition. *Viruses* **8.** ISSN : 19994915 (2016).

161. POMBO, A., CUELLO, P., SCHUL, W., YOON, J. B., ROEDER, R. G., COOK, P. R. & MURPHY, S. Regional and temporal specialization in the nucleus : A transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO Journal* **17,** 1768-1778. ISSN : 02614189 (1998).

162. PONT, V. D., JIANG, Y. & PLEMPER, R. K. Bipartite interface of the measles virus phosphoprotein X domain with the large polymerase protein regulates viral polymerase dynamics. *PLoS Pathogens* **15,** 1-27. ISSN : 15537374 (2019).

163. PORTER, C. T. & MARTIN, A. C. R. BiopLib and BiopTools–a C programming library and toolset for manipulating protein structure. eng. *Bioinformatics (Oxford, England)* **31,** 4017-4019. ISSN : 1367-4811. https://pubmed.ncbi.nlm.nih.gov/26323716%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4673973/ (2015).

164. PRECIOUS, B., YOUNG, D. F., BERMINGHAM, A., FEARNS, R., RYAN, M. & RANDALL, R. E. Inducible expression of the P, V, and NP genes of the paramyxovirus simian virus 5 in cell lines and an examination of NP-P and NP-V interactions. *Journal of virology* **69,** 8001-8010. ISSN : 0022-538X (1995).

165. PURCELL, E. M., TORREY, H. C. & POUND, R. V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **69,** 37-38. https://link.aps.org/doi/10.1103/PhysRev.69.37 (1-2 1946).

166. QUIROZ, F. G. & CHILKOTI, A. Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. eng. *Nature materials* **14,** 1164-1171. ISSN : 1476-1122. https://pubmed.ncbi.nlm.nih.gov/26390327%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4618764/ (2015).

167. RAHMEH, A. A., MORIN, B., SCHENK, A. D., LIANG, B., HEINRICH, B. S., BRUSIC, V., WALZ, T. & WHELAN, S. P. Critical phosphoprotein elements that regulate polymerase architecture and function in vesicular stomatitis virus. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 14628-14633. ISSN : 00278424 (2012).

168. RAHMEH, A. A., SCHENK, A. D., DANEK, E. I., KRANZUSCH, P. J., LIANG, B., WALZ, T. & WHELAN, S. P. Molecular architecture of the vesicular stomatitis virus RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 20075-20080. ISSN : 10916490 (2010).

169. RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. eng. *Journal of molecular biology* **7,** 95-99. ISSN : 0022-2836 (Print) (1963).

170. RASKATOV, J. A. & TEPLOW, D. B. Using chirality to probe the conformational dynamics and assembly of intrinsically disordered amyloid proteins. *Scientific Reports* **7,** 1-7. ISSN : 20452322. http://dx.doi.org/10.1038/s41598-017-10525-5 (2017).

171. RAY, S., SINGH, N., PANDEY, S., KUMAR, R., GADHE, L., DATTA, D., PATEL, K., MAHATO, J., NAVALKAR, A., PANIGRAHI, R., CHATTERJEE, D., MAITI, S., BHATIA, S., MEHRA, S., SINGH, A., GEREZ, J., CHOWDHURY, A., KUMAR, A., PADINHATEERI, R., RIEK, R., KRISHNAMOORTHY, G. & MAJI, S. K. Liquid-liquid phase separation and liquid-to-solid transition mediate -synuclein amyloid fibril containing hydrogel formation. *bioRxiv*. eprint : https://www.biorxiv.org/content/early/2019/04/26/619858.full.pdf. https://www.biorxiv.org/content/early/2019/04/26/619858 (2019).

172. REICHHELD, S. E., MUIZNIEKS, L. D., KEELEY, F. W. & SHARPE, S. Direct observation of structure and dynamics during phase separation of an elastomeric protein, 1-8 (2017).

173. REZAEI-GHALEH, N., MUNARI, F., BECKER, S., ASSFALG, M. & GRIESINGER, C. A facile oxygen-17 NMR method to determine effective viscosity in dilute, molecularly crowded and confined aqueous media. *Chemical Communications* **55,** 12404-12407. ISSN : 1364548X (2019).

174. RIBEIRO, S. S., SAMANTA, N., EBBINGHAUS, S. & MARCOS, J. C. The synergic effect of water and biomolecules in intracellular phase separation. *Nature Reviews Chemistry* **3,** 552-561. ISSN : 23973358 (2019).

175. RICHARD, C.-A., RINCHEVAL, V., LASSOUED, S., FIX, J., CARDONE, C., ESNEAU, C., NEKHAI, S., GALLOUX, M., RAMEIX-WELTI, M.-A., SIZUN, C. & ELÉOUËT, J.-F. RSV hijacks cellular protein phosphatase 1 to regulate M2-1 phosphorylation and viral transcription. eng. *PLoS pathogens* **14,** e1006920. ISSN : 1553-7374 (Electronic) (2018).

176. RINCHEVAL, V., LELEK, M., GAULT, E., BOUILLIER, C., SITTERLIN, D., BLOUQUIT-LAYE, S., GALLOUX, M., ZIMMER, C., ELEOUET, J. F. & RAMEIX-WELTI, M. A. Functional organization of cytoplasmic inclusion bodies in cells infected by respiratory syncytial virus. *Nature Communications* **8,** 1-11. ISSN : 20411723. http://dx.doi.org/10.1038/s41467-017-00655-9 (2017).

177. ROTA, P. A., MOSS, W. J., TAKEDA, M., DE SWART, R. L., THOMPSON, K. M. & GOODSON, J. L. Measles. *Nature Reviews Disease Primers* **2.** ISSN : 2056676X (2016).

178. ROTKIEWICZ, P. & SKOLNICK, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. eng. *Journal of computational chemistry* **29,** 1460-1465. ISSN : 1096-987X (Electronic) (2008).

179. ROWE, C. L., WAGSTAFF, K. M., OKSAYAN, S., GLOVER, D. J., JANS, D. A. & MOSELEY, G. W. Nuclear Trafficking of the Rabies Virus Interferon Antagonist P-Protein Is Regulated by an Importin-Binding Nuclear Localization Sequence in the C-Terminal Domain. eng. *PloS one* **11,** e0150477. ISSN : 1932-6203 (Electronic) (2016).

180. RUBINSTEIN, M. & COLBY, R. H. *M_Rubinstein,_Ralph_H_Colby,_Polymer.pdf* 2003.

181. RYAN, V. H., DIGNON, G. L., ZERZE, H. & BURKE, K. A. Mechanistic View of hnRNPA2 Low-Complexity Domain Structure , Interactions , and Phase Separation Altered by Mutation and Arginine Methylation Article Mechanistic View of hnRNPA2 Low-Complexity Domain Structure , Interactions , and Phase Separation Altered by Mutation and Arginine Methylation, 465-479 (2018).

182. SALEH, O. A., JEON, B.-j. & LIEDL, T. Enzymatic degradation of liquid droplets of DNA is modulated near the phase boundary. **117** (2020).

183. SALMON, L., NODET, G., OZENNE, V., YIN, G., JENSEN, M. R., ZWECKSTETTER, M. & BLACKLEDGE, M. NMR characterization of long-range order in intrinsically disordered proteins. eng. *Journal of the American Chemical Society* **132,** 8407-8418. ISSN : 1520-5126 (Electronic) (2010).

184. SALVI, N., ABYZOV, A. & BLACKLEDGE, M. Analytical Description of NMR Relaxation Highlights Correlated Dynamics in Intrinsically Disordered Proteins. *Angewandte Chemie - International Edition* **56,** 14020-14024. ISSN : 15213773 (2017).

185. SANDERS, D. W., KEDERSHA, N., LEE, D. S., STROM, A. R., DRAKE, V., RIBACK, J. A., BRACHA, D., EEFTENS, J. M., IWANICKI, A., WANG, A., WEI, M. T., WHITNEY, G., LYONS, S. M., ANDERSON, P., JACOBS, W. M., IVANOV, P. & BRANGWYNNE, C. P. Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **181,** 306-324.e28. ISSN : 10974172. http://dx.doi.org/10.1016/j.cell.2020.03.050 (2020).

186. SATO, H., MASUDA, M., KANAI, M., TSUKIYAMA-KOHARA, K., YONEDA, M. & KAI, C. Measles Virus N Protein Inhibits Host Translation by Binding to eIF3-p40. *Journal of Virology* **81,** 11569-11576. ISSN : 0022-538X (2007).

187. SATO, H., MASUDA, M., MIURA, R., YONEDA, M. & KAI, C. Morbillivirus nucleoprotein possesses a novel nuclear localization signal and a CRM1-independent nuclear export signal. eng. *Virology* **352,** 121-130. ISSN : 0042-6822 (Print) (2006).

188. SAVASTANO, M. C., GAMBINI, G., SAVASTANO, A., FALSINI, B., DE VICO, U., SANGUINETTI, M., CATTANI, P., MARCHETTI, S., LARICI, A. R., FRANCESCHI, F., SANTOLIQUIDO, A., MORONI, R., CAMBIERI, A., BELLANTONE, R., LANDI, F., SCAMBIA, G. & RIZZO, S. Evidence-based of conjunctival COVID-19 positivity : An Italian experience : Gemelli Against COVID Group. eng. *European journal of ophthalmology,* 1120672120976548. ISSN : 1724-6016 (Electronic) (2020).

189. SCHNEIDER, R., MAURIN, D., COMMUNIE, G., KRAGELJ, J., HANSEN, D. F., RUIGROK, R. W., JENSEN, M. R. & BLACKLEDGE, M. Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR. *Journal of the American Chemical Society* **137,** 1220-1229. ISSN : 15205126 (2015).

190. SCHUL, W., GROENHOUT, B., KOBERNA, K., TAKAGAKI, Y., JENNY, A., MANDERS, E. M., RAŠKA, I., VAN DRIEL, R. & DE JONG, L. The RNA 3 cleavage factors CstF 64 kDa and CPSF 100 kDa are concentrated in nuclear domains closely associated with coiled bodies and newly synthesized RNA. *EMBO Journal* **15,** 2883-2892. ISSN : 02614189 (1996).

191. SCHWALBE, M., OZENNE, V., BIBOW, S., JAREMKO, M., JAREMKO, L., GAJDA, M., JENSEN, M. R., BIERNAT, J., BECKER, S., MANDELKOW, E., ZWECKSTETTER, M. & BLACKLEDGE, M. Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α-synuclein in solution from NMR and small angle scattering. *Structure* **22,** 238-249. ISSN : 09692126 (2014).

192. SHIN, Y. & BRANGWYNNE, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357.** ISSN : 10959203 (2017).

193. SINGATULINA, A. S., HAMON, L., SUKHANOVA, M. V., BOUHSS, A., LAVRIK, O. I. & PASTRE, D. PARP-1 Activation Directs FUS to DNA Damage Sites to Form PARG-Reversible Compartments Enriched in Damaged DNA Article PARP-1 Activation Directs FUS to DNA Damage Sites to Form PARG-Reversible Compartments Enriched in Damaged DNA, 1809-1821 (2019).

194. SMALLWOOD, S., RYAN, K. W. & MOYER, S. A. Deletion analysis defines a carboxyl-proximal region of Sendai virus P protein that binds to the polymerase L protein. **202,** 154-163. ISSN : 00426822 (1994).

195. SPERA, S. & BAX, A. Empirical Correlation between Protein Backbone Conformation and Cα and Cβ 13C Nuclear Magnetic Resonance Chemical Shifts. *Journal of the American Chemical Society* **113,** 5490-5492. ISSN : 15205126 (1991).

196. SPRAGUE, B. L., PEGO, R. L., STAVREVA, D. A. & MCNALLY, J. G. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophysical Journal* **86,** 3473-3495. ISSN : 00063495. http://dx.doi.org/10.1529/biophysj.103.026765 (2004).

197. SUGAI, A., SATO, H., HAGIWARA, K., KOZUKA-HATA, H., OYAMA, M., YONEDA, M. & KAI, C. Newly Identified Minor Phosphorylation Site Threonine-279 of Measles Virus Nucleoprotein Is a Prerequisite for Nucleocapsid Formation. *Journal of Virology* **88,** 1140-1149. ISSN : 0022-538X (2014).

198. SUGAI, A., SATO, H., YONEDA, M. & KAI, C. Phosphorylation of measles virus phosphoprotein at S86 and/or S151 downregulates viral transcriptional activity. *FEBS Letters* **586,** 3900-3907. ISSN : 00145793. http://dx.doi.org/10.1016/j.febslet.2012.09.021 (2012).

199. SUGAI, A., SATO, H., YONEDA, M. & KAI, C. PIM 3 kinase, a proto-oncogene product, regulates phosphorylation of the measles virus nucleoprotein tail domain at Ser 479 and Ser 510. *Biochemical and Biophysical Research Communications* **531,** 267-274. ISSN : 10902104. https://doi.org/10.1016/j.bbrc.2020.06.002 (2020).

200. TATSUO, H., ONO, N., TANAKA, K. & YANAGI, Y. Slam (CDw150) is a cellular receptor for measles virus. *Nature* **406,** 893-897. ISSN : 00280836 (2000).

201. TAYLOR, N. O., WEI, M. T., STONE, H. A. & BRANGWYNNE, C. P. Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophysical Journal* **117,** 1285-1300. ISSN : 15420086. https://doi.org/10.1016/j.bpj.2019.08.030 (2019).

202. TROS, M., ZHENG, L., HUNGER, J., BONN, M., BONN, D., SMITS, G. J. & WOUTERSEN, S. Picosecond orientational dynamics of water in living cells. *Nature Communications* **8,** 1-7. ISSN : 20411723. http://dx.doi.org/10.1038/s41467-017-00858-0 (2017).

203. TSANG, B., PRITIŠANAC, I., SCHERER, S. W., MOSES, A. M. & FORMAN-KAY, J. D. Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell* **183,** 1742-1756. ISSN : 0092-8674. https://www.sciencedirect.com/science/article/pii/S0092867420316226 (2020).

204. TURNER, A. L., WATSON, M., WILKINS, O. G., CATO, L., TRAVERS, A., THOMAS, J. O. & STOTT, K. Highly disordered histone H1 DNA model complexes and their condensates. **115** (2018).

205. URRY, D. W. Free energy transduction in polypeptides and proteins based on inverse temperature transitions. *Progress in Biophysics and Molecular Biology* **57,** 23-57. ISSN : 00796107 (1992).

206. VALLURUPALLI, P., BOUVIGNIES, G. & KAY, L. E. Studying "invisible" excited protein states in slow exchange with a major state conformation. *Journal of the American Chemical Society* **134,** 8148-8161. ISSN : 00027863 (2012).

207. VAN DER LEE, R., BULJAN, M., LANG, B., WEATHERITT, R. J., DAUGHDRILL, G. W., DUNKER, A. K., FUXREITER, M., GOUGH, J., GSPONER, J., JONES, D. T., KIM, P. M., KRIWACKI, R. W., OLDFIELD, C. J., PAPPU, R. V., TOMPA, P., UVERSKY, V. N., WRIGHT, P. E. & BABU, M. M. Classification of intrinsically disordered regions and proteins. *Chemical Reviews* **114,** 6589-6631. ISSN : 15206890 (2014).

208. WANG, J., CHOI, J. M., HOLEHOUSE, A. S., LEE, H. O., ZHANG, X., JAHNEL, M., MAHARANA, S., LEMAITRE, R., POZNIAKOVSKY, A., DRECHSEL, D., POSER, I., PAPPU, R. V., ALBERTI, S. & HYMAN, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174,** 688-699.e16. ISSN : 10974172 (2018).

209. WANG, Z., ZHANG, G. & ZHANG, H. Protocol for analyzing protein liquid–liquid phase separation. *Biophysics Reports* **5,** 1-9. ISSN : 2364-3439. https://doi.org/10.1007/s41048-018-0078-7 (2019).

210. WAUDBY, C. A., MANTLE, M. D., CABRITA, L. D., GLADDEN, L. F., DOBSON, C. M. & CHRISTODOULOU, J. Rapid Distinction of Intracellular and Extracellular Proteins Using NMR Diffusion Measurements. *Journal of the American Chemical Society* **134,** 11312-11315. ISSN : 0002-7863. https://doi.org/10.1021/ja304912c (2012).

211. WAUDBY, C. A., OUVRY, M., DAVIS, B. & CHRISTODOULOU, J. Two-dimensional NMR lineshape analysis of single, multiple, zero and double quantum correlation experiments. *Journal of Biomolecular NMR* **74,** 95-109. ISSN : 15735001. https://doi.org/10.1007/s10858-019-00297-7 (2020).

212. WEI, M.-T., ELBAUM-GARFINKLE, S., HOLEHOUSE, A. S., CHEN, C. C.-H., FERIC, M., ARNOLD, C. B., PRIESTLEY, R. D., PAPPU, R. V. & BRANGWYNNE, C. P. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. eng. *Nature chemistry* **9,** 1118-1125. ISSN : 1755-4349 (Electronic) (2017).

213. WEICKERT, S., CATTANI, J. & DRESCHER, M. *Intrinsically disordered proteins (IDPs) studied by EPR and in-cell EPR* 1-37. ISBN : 978-1-78801-372-7. http://dx.doi.org/10.1039/9781788013888-00001 (The Royal Society of Chemistry, 2019).

214. WHITE, M. R., MITREA, D. M., ZHANG, P., TOLBERT, M., TAYLOR, J. P., KRIWACKI, R. W., WHITE, M. R., MITREA, D. M., ZHANG, P., STANLEY, C. B., CASSIDY, D. E., NOURSE, A., PHILLIPS, A. H., TOLBERT, M., TAYLOR, J. P. & KRIWACKI, R. W. C9orf72 Poly (PR) Dipeptide Repeats Disturb Biomolecular Phase Separation and Disrupt Nucleolar Function C9orf72 Poly (PR) Dipeptide Repeats Disturb Biomolecular Phase Separation and Disrupt Nucleolar Function. *Molecular Cell* **74,** 713-728.e6. ISSN : 1097-2765. https://doi.org/10.1016/j.molcel.2019.03.019 (2019).

215. WHO. Global Measles and Rubella Update. *World Health Organisation, www.who.int/immunization/diseases/m* (2019).

216. WILKINS, M. R., GASTEIGER, E., BAIROCH, A., SANCHEZ, J. C., WILLIAMS, K. L., APPEL, R. D. & HOCHSTRASSER, D. F. Protein identification and analysis tools in the ExPASy server. eng. *Methods in molecular biology (Clifton, N.J.)* **112,** 531-552. ISSN : 1064-3745 (Print) (1999).

217. WILLIAMSON, M. P. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* **73,** 1-16. ISSN : 00796565. http://dx.doi.org/10.1016/j.pnmrs.2013.02.001 (2013).

218. WONG, L. E., BHATT, A., ERDMANN, P. S., HOU, Z., MAIER, J., PIRKULIYEVA, S., ENGELKE, M., BECKER, S., PLITZKO, J., WIENANDS, J. & GRIESINGER, C. Tripartite phase separation of two signal effectors with vesicles priming B cell responsiveness. *Nature Communications* **11,** 848. ISSN : 2041-1723. https://doi.org/10.1038/s41467-020-14544-1 (2020).

219. WORD, J. M., LOVELL, S. C., RICHARDSON, J. S. & RICHARDSON, D. C. Asparagine and glutamine : using hydrogen atom contacts in the choice of side-chain amide orientation. eng. *Journal of molecular biology* **285,** 1735-1747. ISSN : 0022-2836 (Print) (1999).

220. WRIGHT, P. & DYSON, H. Intrinsically unstructured proteins : re-assessing the protein structure-function paradigm. *Journal of molecular biology* **293 2,** 321-31 (1999).

221. WRIGHT, P. E. & DYSON, H. J. Intrinsically disordered proteins in cellular signalling and regulation. eng. *Nature reviews. Molecular cell biology* **16,** 18-29. ISSN : 1471-0080. https://pubmed.ncbi.nlm.nih.gov/25531225%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4405151/ (2015).

222. YABUKARSKI, F., LAWRENCE, P., TARBOURIECH, N., BOURHIS, J. M., DELAFORGE, E., JENSEN, M. R., RUIGROK, R. W., BLACKLEDGE, M., VOLCHKOV, V. & JAMIN, M. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nature Structural and Molecular Biology* **21,** 754-759. ISSN : 15459985. http://dx.doi.org/10.1038/nsmb.2868 (2014).

223. YANG, P., MATHIEU, C., KOLAITIS, R.-M., ZHANG, P., MESSING, J., YURTSEVER, U., YANG, Z., WU, J., LI, Y., PAN, Q., YU, J., MARTIN, E. W., MITTAG, T., KIM, H. J. & TAYLOR, J. P. G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181,** 325-345.e28. ISSN : 0092-8674. https://www.sciencedirect.com/science/article/pii/S0092867420303391 (2020).

224. ZASLAVSKY, B., BAGIROV, T., BOROVSKAYA, A., GULAEVA, N., MIHEEVA, L., MAHMUDOV, A. & RODNIKOVA, M. Structure of water as a key factor of phase separation in aqueous mixtures of two nonionic polymers. *Polymer* **30,** 2104-2111. ISSN : 0032-3861. https://www.sciencedirect.com/science/article/pii/0032386189903017 (1989).

225. ZHANG, H., NEAL, S. & WISHART, D. S. RefDB : A database of uniformly referenced protein chemical shifts. *Journal of Biomolecular NMR* **25,** 173-195. ISSN : 1573-5001. https://doi.org/10.1023/A:1022836027055 (2003).

226. ZHANG, S., CHEN, L., ZHANG, G., YAN, Q., YANG, X., DING, B., TANG, Q., SUN, S., HU, Z. & CHEN, M. An Amino Acid of Human Parainfluenza Virus Type 3 Nucleoprotein Is Critical for Template Function and Cytoplasmic Inclusion Body Formation. *Journal of Virology* **87,** 12457-12470. ISSN : 0022-538X (2013).

227. ZHOU, Y., SU, J. M., SAMUEL, C. E. & MA, D. Measles Virus Forms Inclusion Bodies with Properties of Liquid Organelles. *Journal of Virology* **93.** ISSN : 0022-538X (2019).

Table des figures

Liste des tableaux

Publications

Guseva S, Perez LM, Camacho-Zarco A, Bessa LM, Salvi N, Malki A, Maurin D, Blackledge M. 1H, 13C and 15N Backbone chemical shift assignments of the n-terminal and central intrinsically disordered domains of SARS-CoV-2 nucleoprotein. Biomol NMR Assign. (2021) 17:1–6.

Salvi N, Bessa LM, Guseva S, Camacho-Zarco A, Maurin D, Perez LM, Malki A, Hengesbach M, Korn SM, Schlundt A, Schwalbe H, Blackledge M. 1H, 13C and 15N backbone chemical shift assignments of SARS-CoV-2 nsp3a. Biomol NMR Assign. (2021) 15(1):173-176.

Altincekic, Korn, Qureshi, Dujardin, Ninot-Pedrosa, Abele, Abi Saad, Alfano, Almeida, Alshamleh, Cardoso de Amorim, Anderson, Ano Bom, Anorma, Bains, Bax, Blackledge, Blechar, Böckmann, Brigandat, Bula et al. Large-scale recombinant production of the SARS-CoV-2 proteome for high-throughput and structural biology applications. Front. Mol. Biosci. (2021) Accepted.

Guseva S, Milles S, Jensen MR, Schoehn G, Ruigrok RW, Blackledge M, Structure, dynamics and phase separation of measles virus RNA replication machinery. Curr Opin Virol. (2020) 41:59-67.

Guseva S, Milles S, Jensen MR, Salvi N, Kleman J-P, Maurin D, Ruigrok RW, Blackledge M. Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. Sci. Adv. (2020) 6:eaaz7095.

Guseva S, Milles S, Blackledge M, Ruigrok RW. The Nucleoprotein and Phosphoprotein of Measles Virus. Frontiers in Microbiol. (2019) 10:1832.

Desfosses A, Milles S, Jensen MR, Guseva S, Colletier J-P, Maurin D, Schoehn G, Gutsche I, Ruigrok R. WH, Blackledge M. Assembly and Cryo-EM structures of RNA-specific measles virus nucleocapsids provide mechanistic insight into paramyxoviral replication, Proc Natl Acad Sci USA (2019) 116:4256–4264.

Milles S, Jensen MR, Lazert C, Guseva S, Ivashchenko S, Communie G, Maurin D, Gerlier D, Ruigrok RW, Blackledge M. An ultraweak interaction in the intrinsically disordered replication machinery is essential for measles virus function. Sci Adv. (2018) 4:eaat7778.

A

Annexe 1: Supporting information to the article "Measles Virus Nucleo- and Phosphoproteins form Liquid-like Phase-Separated Compartments"

Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/14/eaaz7095/DC1

Supplementary Materials for

Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly

Serafima Guseva, Sigrid Milles, Malene Ringkjøbing Jensen, Nicola Salvi, Jean-Philippe Kleman, Damien Maurin, Rob W. H. Ruigrok*, Martin Blackledge*

*Corresponding author. Email: martin.blackledge@ibs.fr (M.B.); rob.ruigrok@ibs.fr (R.W.H.R.)

Published 1 April 2020, *Sci. Adv.* **6**, eaaz7095 (2020)
DOI: 10.1126/sciadv.aaz7095

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

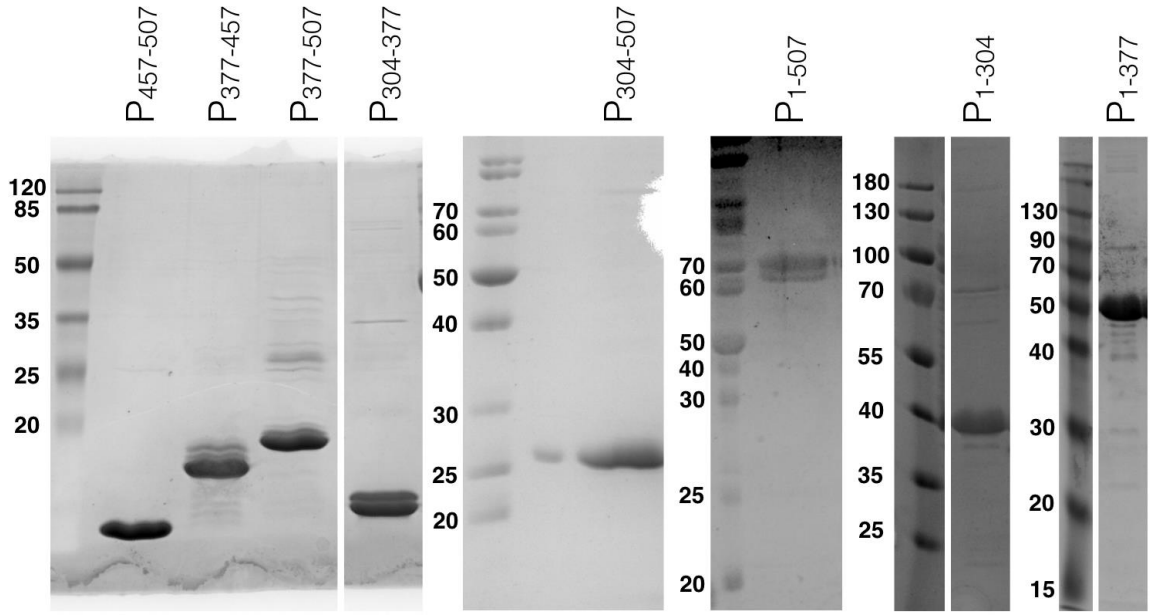
(available at advances.sciencemag.org/cgi/content/full/6/14/eaaz7095/DC1)

Table S1

Ptot (μM)	Ntot (μM)	P_D (μM)	N_D (μM)	ratio
75	7,5	43,23	5,77	0,13
75	15	54,60	8,64	0,16
75	37,5	72,3	13,49	0,19
75	75	75	26,74	0,36

Calculated ratio of [N]/[P] in liquid droplets as determined from NMR (see main text)

Fig. S1 SDS-PAGE (15%) with Coomassie staining of purified truncated constructs of P and P_{1-507}.



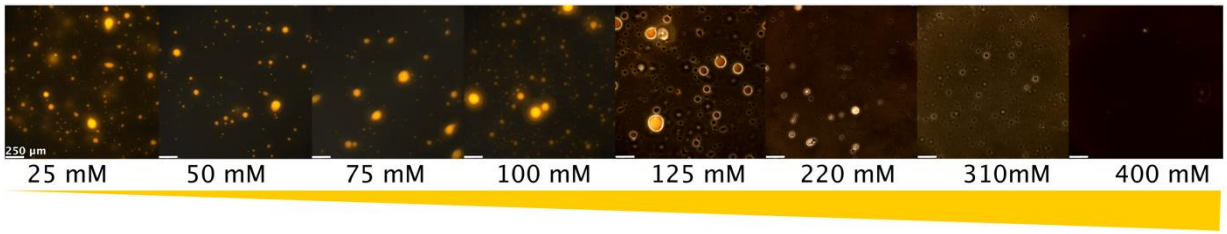


Fig. S2

Fluorescent images overlaid with DIC show the salt dependence of LLPS formation. $P_{50}N_{525}$ (100 µM, 1% fluorescently labelled) and $P_{304\text{-}507}$ (100 µM) were used, NaCl concentration in buffer was increased linearly from 25 mM to 400 mM.

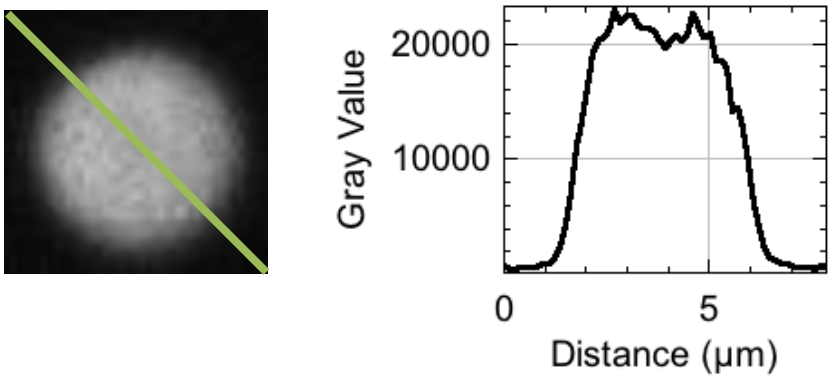


Fig. S3

Fluorescence image of a droplet formed by $P_{50}N_{525}$ (25 μM, 0.04% labelled) and $P_{304\text{-}507}$ (50 μM) and its intensity profile represent homogeneous protein distribution within the droplet.

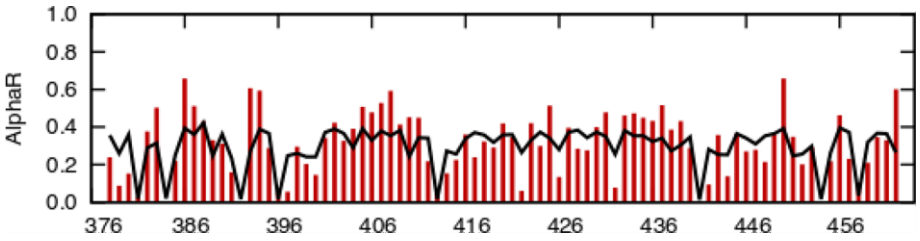


Fig. S4. Secondary structural propensity of P_{LOOP}

Population of α-helical secondary structure of P_{LOOP} showing 15% and 10% difference between the statistical coil ensemble (black line) and the selected ensemble using ASTEROIDS (red bars) for regions 405-411 and 433-438 respectively.

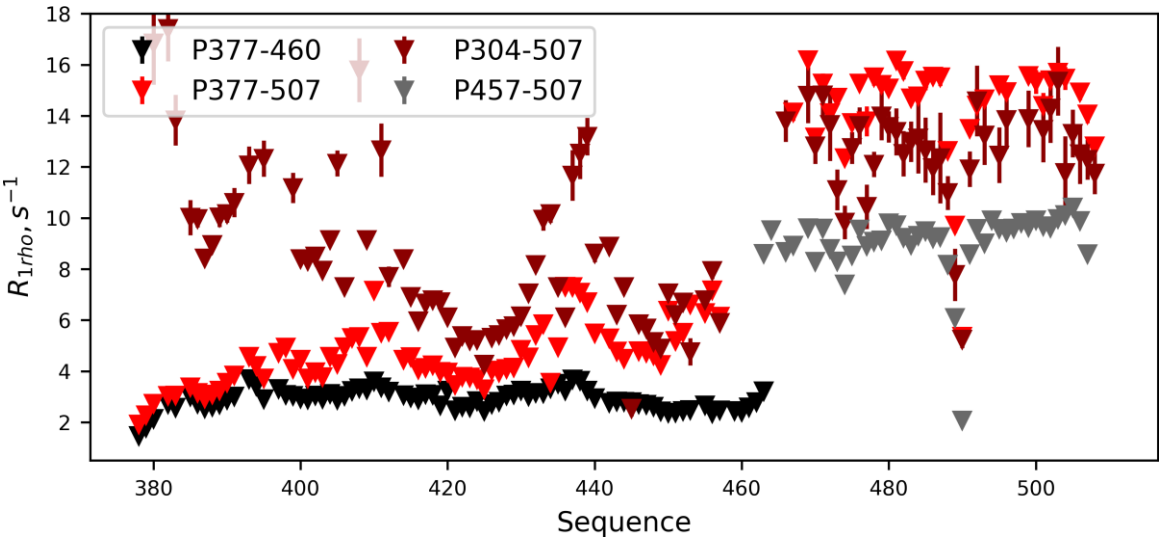


Fig. S5. Dynamic behaviour of P_{LOOP}

^{15}N $R_{1\rho}$ of P_{LOOP} (concentration 500µM), P_{XD} (concentration 500µM), $P_{377\text{-}507}$ (concentration 100 µM) and $P_{304\text{-}507}$ (concentration 100 µM) measured at 25ºC, ^{1}H frequency of 850MHz. Information about the tetramerization domain is not available as resonances for this region are not observable under these experimental conditions.

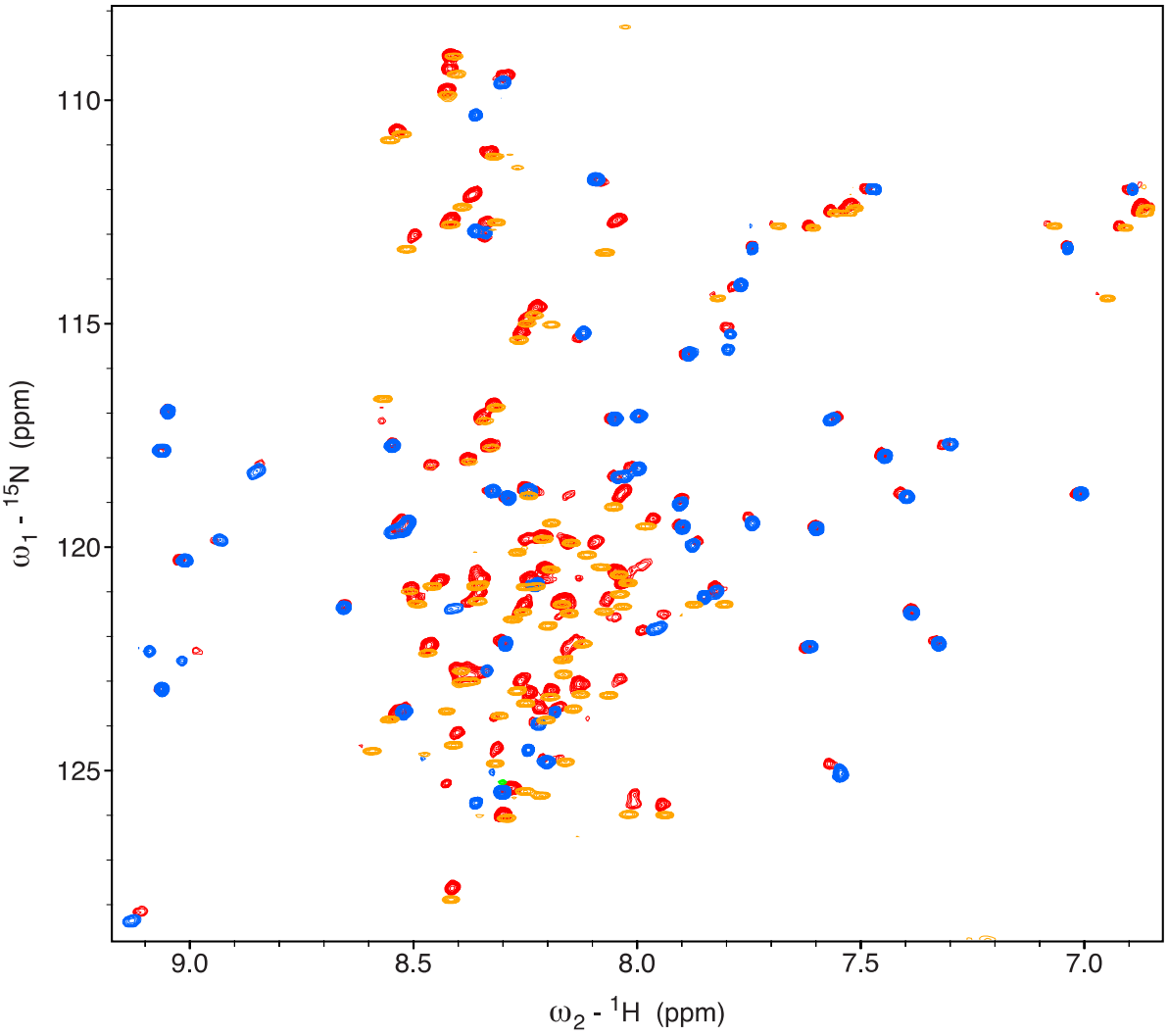


Fig. S6 P_{LOOP} remains unfolded in the context of phase separating protein constructs

^{1}H-^{15}N HSQC spectrum of $P_{304\text{-}507}$ (red) (298K, phosphate buffer, 150mM salt and pH 6.0, 240 µM, 200 increments, 20 scans) overlapped with P_{LOOP} (orange, 100 µM, 200 increments, 20 scans) and P_{XD} (blue, 100 µM, 400 increments, 16 scans), all spectra were recorded at ^{1}H frequency of 700 MHz.

Fig. S7 Sequence analysis of P_{LOOP}

Measles virus phosphoprotein homologs were found using The Basic Local Alignment Search Tool (BLAST) from the Swiss-Prot database, non-truncated sequences were chosen for further analysis (110 sequences) and visualised using https://weblogo.berkeley.edu for residues 377-460.

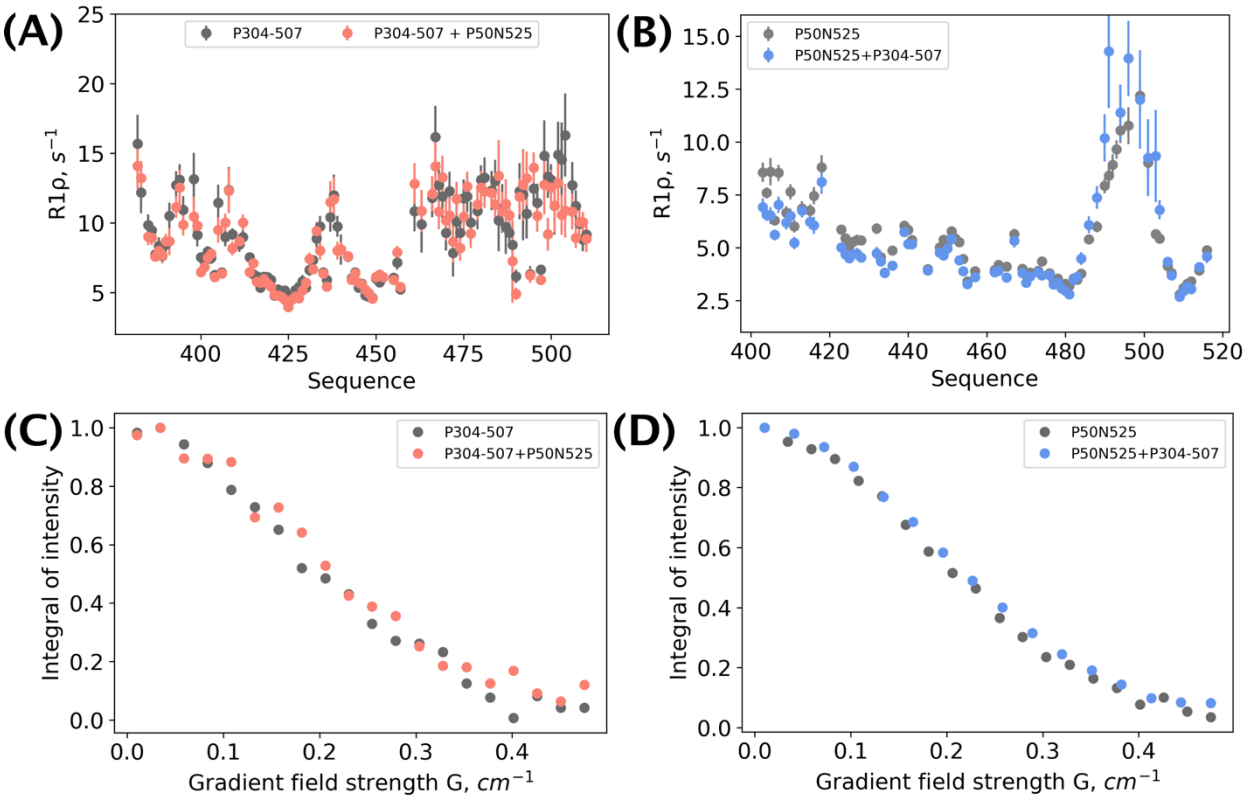


Fig. S8 NMR signals of both $P_{50}N_{525}$ and $P_{304\text{-}507}$ at phase separating conditions exhibit similar characteristics as for the same proteins in the dilute phase. $R_{1\rho}$ profiles at 25ºC 700 MHz for (A) $P_{304\text{-}507}$ alone (100 μM, grey) and with $P_{50}N_{525}$ (10 μM, salmon pink); (B) $P_{50}N_{525}$ alone (100 μM, grey) and with $P_{304\text{-}507}$ (10 μM, blue) measured at 25ºC 850 MHz. Translational diffusion for (C) $P_{304\text{-}507}$ alone (75 μM, grey) and with $P_{50}N_{525}$ (7.5 μM salmon pink); (D) $P_{50}N_{525}$ alone (75 μM, grey) and with $P_{304\text{-}507}$ (37.5 μM, blue).

Movie S1.

Droplets formed by $P_{50}N_{525}$ (10 μM) and $P_{304\text{-}507}$ (100 μM) highlighting examples of fusion upon touching within several seconds. The video speed is accelerated by a factor of 2. Scale bar is 24 μM.

Movie S2.

Fluorescence microscopy images showing RNA diffusing into droplets pre-formed by mixing $P_{50}N_{525}$ (20 μM) and $P_{304\text{-}507}$ (80 μM). Concentration of RNA was 200 μM (with 10% labelled RNA-FAM). RNA was added on one side of the coverslip and spontaneously diffused in the sample, the time gap between images is 2 seconds. Scale bar is 240 μm.

Data S1. Chemical shift assignment of P_{LOOP}

Phosphate buffer 50 mM, pH 6.0
NaCl 150 mM
DTT 2 mM
298K

	N	H	CO	CA
378	110.825	8.547	173.598	45.1
379	120.357	8.075	176.122	55.6725
380	124.46	8.585	-	51.967
381	-	-	176.947	63.866
382	116.607	8.56	174.73	53.2185
383	121.187	7.868	174.841	51.963
384	-	-	177.501	63.791
385	112.628	8.307	174.57	62.221
386	125.898	7.931	177.289	52.525
387	120.012	8.261	176.157	54.3155
388	119.453	7.975	175.915	62.016
389	124.335	8.404	176.204	56.325
390	122.421	8.158	175.570	60.769
391	123.769	8.544	175.475	50.978
392	-	-	176.839	63.799
393	118.767	8.237	176.206	54.675
394	121.208	7.801	176.957	54.8565
395	123.226	8.057	174.311	54.424
396	-	-	176.76	62.97
397	121.527	8.27	176.414	61.1525
398	125.462	8.209	176.664	61.0285
399	113.247	8.508	-	45.27
400	120.772	8.211	175.369	56.342
401	121.18	8.484	176.527	54.325
402	116.791	8.308	174.497	59.009
403	110.679	8.52	-	45.643
404	120.718	8.015	176.277	56.2955
405	124.762	8.311	178.064	52.741
406	121.392	8.145	177.412	55.471
407	123.543	8.137	178.092	52.8755
408	119.372	8.186	176.854	56.8525
409	120.959	8.032	176.299	62.8285
410	124.725	8.154	177.172	55.149
411	122.084	8.115	176.176	56.105
412	123.781	8.196	174.404	54.235
413	-	-	176.956	63.005
414	120.81	8.238	176.162	62.334
415	127.802	8.409	177.804	52.554
416	115.276	8.258	174.830	58.399
417	122.887	8.355	176.364	56.477
418	121.129	8.35	176.079	56.0555
419	123.42	8.248	177.496	55.307
420	120.78	8.343	176.515	56.2045

421	109.802	8.418	174.26	45.393
422	119.745	8.205	176.704	55.594
423	114.736	8.223	174.476	61.8445
424	120.912	8.501	175.683	53.414
425	108.94	8.403	174.228	53.7865
426	120.423	8.187	176.710	56.241
427	114.908	8.237	174.672	61.821
428	118.005	8.367	174.731	58.2235
429	118.119	8.451	174.817	58.525
430	122.69	8.385	177.03	56.721
431	109.341	8.398	174.29	45.52
432	119.83	8.141	176.09	55.967
433	123.153	8.263	177.379	55.401
434	122.777	8.158	177.498	55.2965
435	121.688	8.193	176.799	56.7215
436	120.792	8.358	176.455	57.0605
437	120.096	8.107	175.580	57.8815
438	121.372	8.069	175.441	55.6295
439	123.226	8.121	176.98	55.143
440	123.287	8.188	174.287	54.016
441	-	-	176.959	63.176
442	121.382	8.249	176.902	61.241
443	112.707	8.41	173.911	53.479
444	121.208	8.16	176.546	56.162
445	122.953	8.381	176.632	56.408
446	122.294	8.461	176.3	47.0785
447	117.099	8.336	174.536	58.301
448	117.699	8.32	174.152	58.23
449	125.997	8.287	177.706	52.516
450	119.024	8.046	176.698	62.557
451	112.3	8.384	173.454	53.363
452	120.551	8.035	175.076	57.809
453	125.899	8.014	173.687	59.259
454	-	-	176.44	62.95
455	120.785	8.452	176.577	54.386
456	113.322	8.066	174.994	61.418
457	111.182	8.317	-	44.798
458	-	-	177.126	63.412
459	123.569	8.419	177.996	52.58
460	114.92	8.183	174.749	58.295

B

Annexe 2 : Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: XX 2021
doi: 10.3389/fmolb.2021.653148

OPEN ACCESS

Edited by:
Qian Han, Hainan University, China
Reviewed by:
David Douglas Boehr, Pennsylvania State University, United States
Luis G. Brieba, National Polytechnic Institute of Mexico (CINVESTAV), Mexico

***Correspondence:**
Anja Böckmann
a.bockmann@ibcp.fr
Harald Schwalbe
schwalbe@nmr.uni-frankfurt.de
Martin Hengesbach
hengesbach@nmr.uni-frankfurt.de
Andreas Schlundt
schlundt@bio.uni-frankfurt.de

†These authors have contributed equally to this work and share first authorship
‡These authors share last authorship

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: 13 January 2021
Accepted: 04 February 2021
Published: XX 2021

Citation:
Altincekic N, Korn SM, Qureshi NS, Dujardin M, Ninot-Pedrosa M, Abele R, Abi Saad MJ, Alfano C, Almeida FCL, Alshamleh I, Cardoso de Amorim G, Anderson TK, Ano Bom C, Anorma C, Bains JK, Bax A, Blackledge M, Blechar J, Böckmann A, Brigandat L, Bula A, Bütikofer M, Camacho Zarco A, Carlomagno T, Caruso IP, Ceylan B, Chaikuad A, Chu F, Cole L, Crosby MG, Dhamotharan K, Felli IC, Ferner J, Fleischmann Y, Fogeron M-L, Fourkiotis NK, Fuks C, Fürtig B, Gallo A, Gande SL, Gerez JA, Ghosh D, Gomes-Neto F, Gorbatyuk O, Guseva S, Hacker C, Häfner S, Hargittay B, Henzler-Wildman K, Hoch JC, Hohmann K, Hutchison MT, Jaudzems K, de Jesus V, Jovic K, Kaderli J, Kalniņš G, Kaņepe I, Kirchdoerfer RN, Kirkpatrick J, Knapp S, Krishnathas R, Kutz F, zur Lage S, Lambertz R, Lang A, Laurents D, Lecoq L, Linhard V, Löhr F, Malki A, Mamigonian Bessa L, Martin RW, Matzel T, Maurin D, McNutt SW, Mebus-Antunes NC, Meier BH, Meiser N, Mompeán M, Monaca E, Montserret R, Moreno Perez L, Moser C, Muhle-Goll C, Neves-Martins TC, Ni X, Norton-Baker B, Pierattelli R, Pontoriero L, Pustovalova Y, Ohlenschläger O, Orts J, T. Da Poian A, Pyper DJ, Richter C, Riek R, Robertson A, S. Pinheiro A, Sabbatella R, Salvi N, Saxena K, Schulte L, Schiavina M, Schwalbe H, Silber M, da Silva Almeida M, Sprague-Piercy MA, Spyroulias GA, Sreeramulu S, Tants J-N, Tārs K, Torres F, Töws S, Treviño MÁ, Trucks S, Tsika AC, Varga K, Wang Y, Weber ME, Weigand JE, Wiedemann C, Wirmer-Bartoschek J, Wirtz Martin MA, Zehnder J, Hengesbach M and Schlundt A (2021) Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. Front. Mol. Biosci. 8:653148. doi: 10.3389/fmolb.2021.653148

Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications

Nadide Altincekic[1,2†], *Sophie Marianne Korn*[2,3†], *Nusrat Shahin Qureshi*[1,2†], *Marie Dujardin*[4†], *Martí Ninot-Pedrosa*[4†], *Rupert Abele*[5], *Marie Jose Abi Saad*[6], *Caterina Alfano*[7], *Fabio C. L. Almeida*[8,9], *Islam Alshamleh*[1,2], *Gisele Cardoso de Amorim*[8,10], *Thomas K. Anderson*[11], *Cristiane Ano Bom*[8,12], *Chelsea Anorma*[13], *Jasleen Kaur Bains*[1,2], *Adriaan Bax*[14], *Martin Blackledge*[15], *Julius Blechar*[1,2], *Anja Böckmann*[4*‡], *Louis Brigandat*[4], *Anna Bula*[16], *Matthias Bütikofer*[6], *Aldo Camacho Zarco*[15], *Teresa Carlomagno*[17,18], *Icaro Putinhon Caruso*[8,9,19], *Betül Ceylan*[1,2], *Apirat Chaikuad*[20,21], *Feixia Chu*[22], *Laura Cole*[4], *Marquise G. Crosby*[23], *Karthikeyan Dhamotharan*[2,3], *Isabella C. Felli*[24,25], *Jan Ferner*[1,2], *Yanick Fleischmann*[6], *Marie-Laure Fogeron*[4], *Nikolaos K. Fourkiotis*[26], *Christin Fuks*[1], *Boris Fürtig*[1,2], *Angelo Gallo*[26], *Santosh L. Gande*[1,2], *Juan Atilio Gerez*[6], *Dhiman Ghosh*[6], *Francisco Gomes-Neto*[8,27], *Oksana Gorbatyuk*[28], *Serafima Guseva*[15], *Carolin Hacker*[29], *Sabine Häfner*[30], *Bruno Hargittay*[1,2], *K. Henzler-Wildman*[11], *Jeffrey C. Hoch*[28], *Katharina Hohmann*[1,2], *Marie T. Hutchison*[1,2], *Kristaps Jaudzems*[16], *Vanessa de Jesus*[1,2], *Katarina Jovic*[22], *Janina Kaderli*[6], *Gints Kalniņš*[31], *Iveta Kaņepe*[16], *Robert N. Kirchdoerfer*[11], *John Kirkpatrick*[17,18], *Stefan Knapp*[20,21], *Robin Krishnathas*[1,2], *Felicitas Kutz*[1,2], *Susanne zur Lage*[18], *Roderick Lambertz*[3], *Andras Lang*[30], *Douglas Laurents*[32], *Lauriane Lecoq*[4], *Verena Linhard*[1,2], *Frank Löhr*[2,33], *Anas Malki*[15], *Luiza Mamigonian Bessa*[15], *Rachel W. Martin*[13,23], *Tobias Matzel*[1,2], *Damien Maurin*[15], *Seth W. McNutt*[22], *Nathane Cunha Mebus-Antunes*[8,9], *Beat H. Meier*[6], *Nathalie Meiser*[1], *Miguel Mompeán*[32], *Elisa Monaca*[7], *Roland Montserret*[4], *Laura Moreno Perez*[15], *Celine Moser*[34], *Claudia Muhle-Goll*[34], *Thaís Cristtina Neves-Martins*[8,9], *Xiamonin Ni*[20,21], *Brenna Norton-Baker*[13], *Roberta Pierattelli*[24,25], *Letizia Pontoriero*[24,25], *Yulia Pustovalova*[28], *Oliver Ohlenschläger*[30], *Julien Orts*[6], *Andrea T. Da Poian*[9], *Dennis J. Pyper*[1,2], *Christian Richter*[1,2], *Roland Riek*[6], *Angus Robertson*[14], *Anderson S. Pinheiro*[8,12], *Raffaele Sabbatella*[7], *Nicola Salvi*[15], *Krishna Saxena*[1,2], *Linda Schulte*[1,2], *Marco Schiavina*[24,25], *Harald Schwalbe*[1,2*‡], *Mara Silber*[34], *Marcius da Silva Almeida*[8,9], *Marc A. Sprague-Piercy*[23], *Georgios A. Spyroulias*[26], *Sridhar Sreeramulu*[1,2], *Jan-Niklas Tants*[2,3], *Kaspars Tārs*[31], *Felix Torres*[6], *Sabrina Töws*[3], *Miguel Á. Treviño*[32], *Sven Trucks*[1], *Aikaterini C. Tsika*[26], *Krisztina Varga*[22], *Ying Wang*[17], *Marco E. Weber*[6], *Julia E. Weigand*[35], *Christoph Wiedemann*[36], *Julia Wirmer-Bartoschek*[1,2], *Maria Alexandra Wirtz Martin*[1,2], *Johannes Zehnder*[6], *Martin Hengesbach*[1*‡] and *Andreas Schlundt*[2,3*‡]

[1]Institute for Organic Chemistry and Chemical Biology, Goethe University Frankfurt, Frankfurt am Main, Germany, [2]Center of Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, Frankfurt am Main, Germany, [3]Institute for Molecular Biosciences, Goethe University Frankfurt, Frankfurt am Main, Germany, [4]Molecular Microbiology and Structural Biochemistry, UMR 5086, CNRS/Lyon University, Grenoble, France, [5]Institute for Biochemistry, Goethe University Frankfurt, Frankfurt am Main, Germany, [6]Swiss Federal Institute of Technology, Laboratory of Physical Chemistry, ETH Zurich, Zurich, Switzerland, [7]Structural Biology and Biophysics Unit, Fondazione Ri.MED, Palermo, Italy, [8]National Center of Nuclear Magnetic Resonance (CNRMN, CENABIO), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, [9]Institute of Medical Biochemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, [10]Multidisciplinary Center for Research in Biology (NUMPEX), Campus Duque de Caxias Federal University of Rio de Janeiro, Duque de Caxias, Brazil, [11]Institute for Molecular Virology, University of Wisconsin-Madison, Madison, WI, United States, [12]Institute of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, [13]Department of Chemistry, University of California, Irvine, CA, United States, [14]LCP, NIDDK, NIH, Bethesda, MD, United States, [15]Univ. Grenoble Alpes, CNRS, CEA, IBS, Grenoble, France, [16]Latvian Institute of Organic Synthesis, Riga, Latvia, [17]BMWZ and Institute of Organic

Chemistry, Leibniz University Hannover, Hannover, Germany, [18]Group of NMR-Based Structural Chemistry, Helmholtz Centre for Infection Research, Braunschweig, Germany, [19]Multiuser Center for Biomolecular Innovation (CMIB), Department of Physics, São Paulo State University (UNESP), São José do Rio Preto, Brazil, [20]Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt am Main, Germany, [21]Structural Genomics Consortium, Buchmann Institute for Molecular Life Sciences, Frankfurt am Main, Germany, [22]Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH, United States, [23]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, United States, [24]Magnetic Resonance Centre (CERM), University of Florence, Sesto Fiorentino, Italy, [25]Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy, [26]Department of Pharmacy, University of Patras, Patras, Greece, [27]Laboratory of Toxicology, Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, Brazil, [28]Department of Molecular Biology and Biophysics, UConn Health, Farmington, CT, United States, [29]Signals GmbH & Co. KG, Frankfurt am Main, Germany, [30]Leibniz Institute on Aging—Fritz Lipmann Institute (FLI), Jena, Germany, [31]Latvian Biomedical Research and Study Centre, Riga, Latvia, [32]"Rocasolano" Institute for Physical Chemistry (IQFR), Spanish National Research Council (CSIC), Serrano, Spain, [33]Institute of Biophysical Chemistry, Goethe University Frankfurt, Frankfurt am Main, Germany, [34]IBG-4, Karlsruhe Institute of Technology, Karlsruhe, Germany, [35]Department of Biology, Technical University of Darmstadt, Darmstadt, Germany, [36]Institute of Biochemistry and Biotechnology, Charles Tanford Protein Centre, Martin Luther University Halle-Wittenberg, Halle/Saale, Germany

The highly infectious disease COVID-19 caused by the *Betacoronavirus* SARS-CoV-2 poses a severe threat to humanity and demands the redirection of scientific efforts and criteria to organized research projects. The international *COVID19-NMR* consortium seeks to provide such new approaches by gathering scientific expertise worldwide. In particular, making available viral proteins and RNAs will pave the way to understanding the SARS-CoV-2 molecular components in detail. The research in *COVID19-NMR* and the resources provided through the consortium are fully disclosed to accelerate access and exploitation. NMR investigations of the viral molecular components are designated to provide the essential basis for further work, including macromolecular interaction studies and high-throughput drug screening. Here, we present the extensive catalog of a holistic SARS-CoV-2 protein preparation approach based on the consortium's collective efforts. We provide protocols for the large-scale production of more than 80% of all SARS-CoV-2 proteins or essential parts of them. Several of the proteins were produced in more than one laboratory, demonstrating the high interoperability between NMR groups worldwide. For the majority of proteins, we can produce isotope-labeled samples of HSQC-grade. Together with several NMR chemical shift assignments made publicly available on *covid19-nmr.com*, we here provide highly valuable resources for the production of SARS-CoV-2 proteins in isotope-labeled form.

Keywords: COVID-19, SARS-CoV-2, nonstructural proteins, structural proteins, accessory proteins, intrinsically disordered region, cell-free protein synthesis, NMR spectroscopy

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, SCoV2) is the cause of the early 2020 pandemic coronavirus lung disease 2019 (COVID-19) and belongs to *Betacoronaviruses*, a genus of the Coronaviridae family covering the α–δ genera (Leao et al., 2020). The large RNA genome of SCoV2 has an intricate, highly condensed arrangement of coding sequences (Wu et al., 2020). Sequences starting with the main start codon contain an open reading frame 1 (ORF1), which codes for two distinct, large polypeptides (pp), whose relative abundance is governed by the action of an RNA pseudoknot structure element. Upon RNA folding, this element causes a −1 frameshift to allow the continuation of translation, resulting in the generation of a 7,096-amino acid 794 kDa polypeptide. If the pseudoknot is not formed, expression of the first ORF generates a 4,405-amino acid 490 kDa polypeptide. Both the short and long polypeptides translated from this ORF (pp1a and pp1ab, respectively) are posttranslationally cleaved by virus-encoded proteases into functional, nonstructural proteins (nsps). ORF1a encodes eleven nsps, and ORF1ab additionally encodes the nsps 12–16. The downstream ORFs encode structural proteins (S, E, M, and N) that are essential components for the synthesis of new virus particles. In between those, additional proteins (accessory/auxiliary factors) are encoded, for which sequences partially overlap (Finkel et al., 2020) and whose identification and classification are a matter of ongoing research (Nelson et al., 2020; Pavesi, 2020). In total, the number of identified peptides or proteins generated from the viral genome is at least 28 on the evidence level, with an additional set of smaller proteins or peptides being predicted with high likelihood.

High-resolution studies of SCoV and SCoV2 proteins have been conducted using all canonical structural biology approaches, such as X-ray crystallography on proteases (Zhang et al., 2020) and methyltransferases (MT) (Krafcikova et al., 2020), cryo-EM of the RNA polymerase (Gao et al., 2020; Yin et al., 2020), and liquid-state (Almeida et al., 2007; Serrano et al., 2009; Cantini et al., 2020; Gallo et al., 2020; Korn et al., 2020a; Korn et al., 2020b;

TABLE 1 | SCoV2 protein constructs expressed and purified, given with the genomic position and corresponding PDBs for construct design.

Protein genome position (nt)[a]	Trivial name construct expressed	Size (aa)	Boundaries	MW (kD[a])	Homol. SCoV (%)[b]	Template PDB[c]	SCoV2 PDB[d]
nsp1	Leader	180		19.8	84		
266–805							
	Full-length	180	1–180	19.8	83		
	Globular domain (GD)	116	13–127	12.7	85	2GDT	7K7P
nsp2		638		70.5	68		
806–2,719							
	C-terminal IDR (CtDR)	45	557–601	4.9	55		
nsp3		1,945		217.3	76		
2,720–8,554							
[a]	Ub-like (Ubl) domain	111	1–111	12.4	79	2IDY	7KAG
[a]	Ub-like (Ubl) domain + IDR	206	1–206	23.2	58		
[b]	Macrodomain	170	207–376	18.3	74	6VXS	6VXS
[c]	SUD-N	140	409–548	15.4	69	2W2G	
[c]	SUD-NM	267	409–675	29.5	74	2W2G	
[c]	SUD-M	125	551–675	13.9	82	2W2G	
[c]	SUD-MC	195	551–745	21.5	79	2KQV	
[c]	SUD-C	64	680–743	7.3	73	2KAF	
[d]	Papain-like protease PL^{pro}	318	743–1,060	36	83	6W9C	6W9C
[e]	NAB	116	1,088–1,203	13.4	87	2K87	
Y	CoV-Y	308	1,638–1,945	34	89		
nsp5	Main protease (M^{pro})	306		33.7	96		
10,055–10,972							
	Full-length[e]	306	1–306	33.7	96	6Y84	6Y84
nsp7		83		9.2	99		
11,843–12,091							
	Full-length	83	1–83	9.2	99	6WIQ	6WIQ
nsp8		198		21.9	98		
12,092–12,685							
	Full-length	198	1–198	21.9	97	6WIQ	6WIQ
nsp9		113		12.4	97		
12,686–13,024							
	Full-length	113	1–113	12.4	97	6W4B	6W4B
nsp10		139		14.8	97		
13,025–13,441							
	Full-length	139	1–139	14.8	97	6W4H	6W4H
nsp13	Helicase	601		66.9	100		
16,237–18,039							
	Full-length	601	1–601	66.9	100	6ZSL	6ZSL
nsp14	Exonuclease/	527		59.8	95		
18,040–19,620	methyltransferase						
	Full-length	527	1–527	59.8	95	5NFY	
	Mtase domain	240	288–527	27.5	95		
nsp15	Endonuclease	346		38.8	89		
19,621–20,658							
	Full-length	346	1–346	38.8	89	6W01	6W01
nsp16	Methyltransferase	298		33.3	93		
20,659–21,552							
	Full-length	298	1–298	33.3	93	6W4H	6W4H
ORF3[a]		275		31.3	72		
25,393–26,220							
	Full-length	275	1–275	31.3	72	6XDC	6XDC
ORF4	Envelope (E) protein	75		8.4	95		
26,245–26,472							
	Full-length	75	1–75	8.4	95	5 × 29	7K3G
ORF5	Membrane	222		25.1	91		
26,523–27,387	glycoprotein (M)						
	Full-length	222	1–222	25.1	91		
ORF6		61		7.3	69		
27,202–27,387							
	Full-length	61	1–61	7.3	69		
ORF7[a]		121		13.7	85		
27,394–27,759							

(Continued on following page)

TABLE 1 | (*Continued*) SCoV2 protein constructs expressed and purified, given with the genomic position and corresponding PDBs for construct design.

Protein genome position (nt)[a]	Trivial name construct expressed	Size (aa)	Boundaries	MW (kD[a])	Homol. SCoV (%)[b]	Template PDB[c]	SCoV2 PDB[d]
	Ectodomain (ED)	66	16–81	7.4	85	1XAK	6W37
ORF7[b]		43		5.2	85		
27,756–27,887							
	Full-length	43	1–43	5.2	85		
ORF8		121		13.8	32		
27,894–28,259							
ORF8	Full-length	121	1–121	13.8	32	7JTL	7JTL
ΔORF8	W/o signal	106	16–121	12	41	7JTL	7JTL
ORF9[a]	Nucleocapsid (N)	419		45.6	91		
28,274–29,533							
	IDR1-NTD- IDR2	248	1–248	26.5	90		
	NTD-SR	169	44–212	18.1	92		
	NTD	136	44–180	14.9	93	6YI3	6YI3
	CTD	118	247–364	13.3	96	2JW8	7C22
ORF9[b]		97		10.8	72		
28,284–28,574							
	Full-length	97	1–97	10.8	72	6Z4U	6Z4U
ORF14		73		8	n.a		
28,734–28,952							
	Full-length	73	1–73	8	n.a		
ORF10		38		4.4	29		
29,558–29,674							
	Full-length	38	1–38	4.4	29		

[a]*Genome position in nt corresponding to SCoV2 NCBI reference genome entry NC_045512.2, identical to GenBank entry MN908947.3 (2).*
[b]*Sequence identities to SCoV are calculated from an alignment with corresponding protein sequences based on the genome sequence of NCBI Reference NC_004718.3.*
[c]*Representative PDB that was available at the beginning of construct design, either SCoV or SCoV2.*
[d]*Representative PDB available for SCoV2 (as of December 2020).*
[e]*Additional point mutations in fl-construct have been expressed.*
n.a.: not applicable.

Kubatova et al., 2020; Tonelli et al., 2020) and solid-state NMR spectroscopy of transmembrane (TM) proteins (Mandala et al., 2020). These studies have significantly improved our understanding on the functions of molecular components, and they all rely on the recombinant production of viral proteins in high amount and purity.

Apart from structures, purified SCoV2 proteins are required for experimental and preclinical approaches designed to understand the basic principles of the viral life cycle and processes underlying viral infection and transmission. Approaches range from studies on immune responses (Esposito et al., 2020), antibody identification (Jiang et al., 2020), and interactions with other proteins or components of the host cell (Bojkova et al., 2020; Gordon et al., 2020). These examples highlight the importance of broad approaches for the recombinant production of viral proteins.

The research consortium *COVID19-NMR* found in 2020 seeks to support the search for antiviral drugs using an NMR-based screening approach. This requires the large-scale production of all druggable proteins and RNAs and their NMR resonance assignments. The latter will enable solution structure determination of viral proteins and RNAs for rational drug design and the fast mapping of compound binding sites. We have recently produced and determined secondary structures of SCoV2 RNA *cis*-regulatory elements in near completeness by NMR spectroscopy, validated by DMS-MaPseq (Wacker et al., 2020), to provide a basis for RNA-oriented fragment screens with NMR.

We here compile a compendium of more than 50 protocols (see **Supplementary Tables SI1–SI23**) for the production and purification of 23 of the 30 SCoV2 proteins or fragments thereof (summarized in **Tables 1**, **2**). We defined those 30 proteins as existing or putative ones to our current knowledge (see later discussion). This compendium has been generated in a coordinated and concerted effort between >30 labs worldwide (**Supplementary Table S1**), with the aim of providing pure mg amounts of SCoV2 proteins. Our protocols include the rational strategy for construct design (if applicable, guided by available homolog structures), optimization of expression, solubility, yield, purity, and suitability for follow-up work, with a focus on uniform stable isotope-labeling.

We also present protocols for a number of accessory and structural E and M proteins that could only be produced using wheat-germ cell-free protein synthesis (WG-CFPS). In SCoV2, accessory proteins represent a class of mostly small and relatively poorly characterized proteins, mainly due to their difficult behavior in classical expression systems. They are often found in inclusion bodies and difficult to purify in quantities adequate for structural studies. We thus here exploit cell-free synthesis, mainly based on previous reports on production and purification of viral membrane proteins in general (Fogeron et al., 2015b; Fogeron et al., 2017; Jirasko et al., 2020b). Besides yields compatible with structural studies, ribosomes in WG extracts further possess an increased folding capacity (Netzer and Hartl, 1997), favorable for those more complicated proteins.

TABLE 2 | Summary of SCoV2 protein production results in *Covid19-NMR*.

Construct expressed	Yields [mg/L][a] or (mg/mL)[b]	Results	Comments	BMRB	Supplementary Material
nsp1					SI1
fl	5	NMR assigned	Expression only at >20°C; after 7 days at 25°C partial proteolysis	50,620[c]	
GD	>0.5	HSQC	High expression; mainly insoluble; higher salt increases stability (>250 mM)		
nsp2					SI2
CtDR	0.7–1.5	NMR assigned	Assignment with His-tag shown in (Mompean et al., 2020)	50,687[c]	
nsp3					SI3
Ubl	0.7	HSQC	Highly stable over weeks; spectrum overlays with Ubl + IDR		
Ubl + IDR	2–3	NMR assigned	Highly stable for >2 weeks at 25°C	50,446[d]	
Macrodomain	9	NMR assigned	Highly stable for >1 week at 25°C and > 2 weeks at 4°C	50,387[d] 50,388[d]	
SUD-N	14	NMR assigned	Highly stable for >10 days at 25°C	50,448[d]	
SUD-NM	17	HSQC	Stable for >1 week at 25°C		
SUD-M	8.5	NMR assigned	Significant precipitation during measurement; tendency to dimerize	50,516[d]	
SUD-MC	12	HSQC	Stable for >1 week at 25°C		
SUD-C	4.7	NMR assigned	Stable for >10 days at 25°C	50,517[d]	
PL^{pro}	12	HSQC	Solubility-tag essential for expression; tendency to aggregate		
NAB	3.5	NMR assigned	Highly stable for >1 week at 25°C; stable for >5 weeks at 4°C	50,334[d]	
CoV-Y	12	HSQC	Low temperature (<25°C) and low concentrations (>0.2 M) favor stability; gradual degradation at 25°C; lithium bromide in final buffer supports solubility		
nsp5					SI4
fl	55	HSQC	Impaired dimerization induced by artificial N-terminal residues		
nsp7					SI5
fl	17	NMR assigned	Stable for several days at 35°C; stable for >1 month at 4°C	50,337[d]	
nsp8					SI6
fl	17	HSQC	Concentration dependent aggregation; low concentrations favor stability		
nsp9					SI7
fl	4.5	NMR assigned	Stable dimer for >4 months at 4°C and >2 weeks at 25°C	50,621[d] 50,622[d] 50,513	
nsp10					SI8
fl	15	NMR assigned	Zn2+ addition during expression and purification increases protein stability; stable for >1 week at 25°C	50,392	
nsp13					SI9
fl	0.5	HSQC	Low expression; protein unstable; concentration above 20 μM not possible		
nsp14					SI10
fl	6	Pure protein	Not above 50μM; best storage: with 50% (v/v) glycerol; addition of reducing agents		
Mtase	10	Pure protein	As fl nsp14; high salt (>0.4 M) for increased stability; addition of reducing agents		
nsp15					SI11
fl	5	HSQC	Tendency to aggregate at 25°C		
nsp16					SI12
fl	10	Pure protein	Addition of reducing agents; 5% (v/v) glycerol favorable; highly unstable		
ORF3[a]					SI13
fl	0.6	Pure protein	Addition of detergent during expression (0.05% Brij-58); stable protein		
E protein					SI14
fl	0.45	Pure protein	Addition of detergent during expression (0.05% Brij-58); stable protein		
M Protein					SI15
fl	0.33	Pure protein	Addition of detergent during expression (0.05% Brij-58); stable protein		
ORF6					SI16

(Continued on following page)

TABLE 2 | (*Continued*) Summary of SCoV2 protein production results in *Covid19-NMR*.

Construct expressed	Yields [mg/L][a] or (mg/mL)[b]	Results	Comments	BMRB	Supplementary Material
fl	0.27	HSQC	Soluble expression without detergent; stable protein; no expression with STREP-tag at N-terminus		
ORF7[a]					SI17
ED	0.4	HSQC	Unpurified protein tends to precipitate during refolding, purified protein stable for 4 days at 25°C		
ORF7[b]					SI18
fl	0.6	HSQC	Tendency to oligomerize; solubilizing agents needed		
fl	0.27	HSQC	Addition of detergent during expression (0.1% MNG-3); stable protein		
ORF8					SI19
fl	0.62	HSQC	Tendency to oligomerize		
ΔORF8	0.5	Pure protein			
N protein					SI20
IDR1-NTD-IDR2	12	NMR assigned	High salt (>0.4 M) for increased stability	50,618, 50,619[c]	
NTD-SR	3	HSQC			
NTD	3	HSQC		34,511	
CTD	2	NMR assigned	Stable dimer for >4 months at 4°C and >3 weeks at 30°C	50,518[d]	
ORF9[b]					SI21
fl	0.64	HSQC	Expression without detergent, protein is stable		
ORF14					SI22
fl	0.43	HSQC	Addition of detergent during expression (0.05% Brij-58); stable in detergent but unstable on lipid reconstitution		
ORF10					SI23
Fl	2	HSQC	Tendency to oligomerize; unstable upon tag cleavage		

[a]*Yields from bacterial expression represent the minimal protein amount in mg/L independent of the cultivation medium.*
[b]*Yields from CFPS represent the minimal protein amount in mg/mL.*
[c]COVID19-nmr *BMRB depositions yet to be released.*
[d]COVID19-nmr *BMRB depositions.*

We exemplify in more detail the optimization of protein production, isotope-labeling, and purification for proteins with different individual challenges: the nucleic acid–binding (NAB) domain of nsp3e, the main protease nsp5, and several auxiliary proteins. For the majority of produced and purified proteins, we achieve >95% purity and provide 15 N-HSQC spectra as the ultimate quality measure. We also provide additional suggestions for challenging proteins, where our protocols represent a unique resource and starting point exploitable by other labs.

MATERIALS AND METHODS

Strains, Plasmids, and Cloning

The rationale of construct design for all proteins can be found within the respective protocols in SI1-23. For bacterial production, *E. coli* strains and expression plasmids are given; for WG-CFPS, template vectors are listed. Protein coding sequences of interest have been obtained as either commercial, codon-optimized genes or, for shorter ORFs and additional sequences, annealed from oligonucleotides prior to insertion into the relevant vector. Subcloning of inserts, adjustment of boundaries, and mutations of genes have been carried out by standard molecular biology techniques. All expression plasmids can be obtained upon request from the *COVID19-NMR* consortium (https://covid19-nmr.de/), including information about coding sequences, restriction sites, fusion tags, and vector backbones.

Protein Production and Purification

For SCoV2 proteins, we primarily used heterologous production in *E. coli*. Detailed protocols of individual full-length (fl) proteins, separate domains, combinations, or particular expression constructs as listed in **Table 1** can be found in the (**Supplementary Tables SI1–SI23**).

The ORF3a, ORF6, ORF7b, ORF8, ORF9b, and ORF14 accessory proteins and the structural proteins M and E were produced by wheat-germ cell-free protein synthesis (WG-CFPS) as described in the **Supplementary Material**. In brief, transcription and translation steps have been performed separately, and detergent has been added for the synthesis of membrane proteins as described previously (Takai et al., 2010; Fogeron et al., 2017).

NMR Spectroscopy

All amide correlation spectra, either HSQC- or TROSY-based, are representative examples. Details on their acquisition parameters and the raw data are freely accessible through https://covid19-nmr.de/or upon request.

Data Availability Statement

Assignments of backbone chemical shifts have been deposited at BMRB for proteins, as shown in **Table 2**, indicated by their

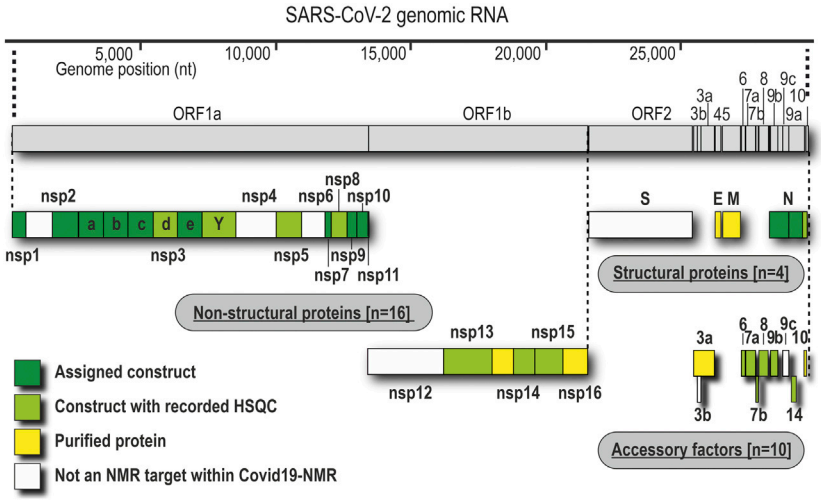


FIGURE 1 | Genomic organization of proteins and current state of analysis or purification. Boxes represent the domain boundaries as outlined in the text and in **Table 1**. Their position corresponds with the genomic loci. Colors indicate whether the pure proteins were purified (yellow), analyzed by NMR using only HSQC (lime), or characterized in detail, including NMR resonance assignments (green).

respective BMRB IDs. All expression constructs are available as plasmids from https://covid19-nmr.de/.

RESULTS

In the following, we provide protocols for the purification of SCoV2 proteins sorted into 1) nonstructural proteins and 2) structural proteins together with accessory ORFs. **Table 1** shows an overview of expression constructs. We use a consequent terminology of those constructs, which is guided by domains, intrinsically disordered regions (IDRs) or other particularly relevant sequence features within them. This study uses the SCoV2 NCBI reference genome entry NC_045512.2, identical to GenBank entry MN908947.3 (Wu et al., 2020), unless denoted differently in the respective protocols. Any relevant definition of boundaries can also be found in the SI protocols.

As applicable for a major part of our proteins, we further define a standard procedure for the purification of soluble His-tagged proteins that are obtained through the sequence of IMAC, TEV/Ulp1 Protease cleavage, Reverse IMAC, and Size-exclusion chromatography, eventually with individual alterations, modifications, or additional steps. For convenient reading, we will thus use the abbreviation IPRS to avoid redundant protocol description. Details for every protein, including detailed expression conditions, buffers, incubation times, supplements, storage conditions, yields, and stability, can be found in the respective **Supplementary Tables SI1–SI23** (see also **Supplementary Tables S1, S2**) and **Tables 1**, **2**. Entries in the tables are linked to the respective protocols for convenient download.

Nonstructural Proteins

We have approached and challenged the recombinant production of a large part of the SCoV2 nsps (**Figure 1**), with great success (**Table 2**). We excluded nsp4 and nsp6 (TM proteins), which are little characterized and do not reveal soluble, folded domains by prediction (Oostra et al., 2007; Oostra et al., 2008). The function of the very short (13 aa) nsp11 is unknown, and it seems to be a mere copy of the nsp12 amino-terminal residues, remaining as a protease cleavage product of ORF1a. Further, we left out the RNA-dependent RNA polymerase nsp12 in our initial approach because of its size (>100 kDa) and known unsuitability for heterologous recombinant production in bacteria. Work on NMR-suitable nsp12 bacterial production is ongoing, while other expert labs have succeeded in purifying nsp12 for cryo-EM applications in different systems (Gao et al., 2020; Hillen et al., 2020). For the remainder of nsps, we here provide protocols for fl-proteins or relevant fragments of them.

nsp1

nsp1 is the very N-terminus of the polyproteins pp1a and pp1ab and one of the most enigmatic viral proteins, expressed only in α- and β-CoVs (Narayanan et al., 2015). Interestingly, nsp1 displays the highest divergence in sequence and size among different CoVs, justifying it as a genus-specific marker (Snijder et al., 2003). It functions as a host shutoff factor by suppressing innate immune functions and host gene expression (Kamitani et al., 2006; Narayanan et al., 2008; Schubert et al., 2020). This suppression is achieved by an interaction of the nsp1 C-terminus with the mRNA entry tunnel within the 40 S subunit of the ribosome (Schubert et al., 2020; Thoms et al., 2020).

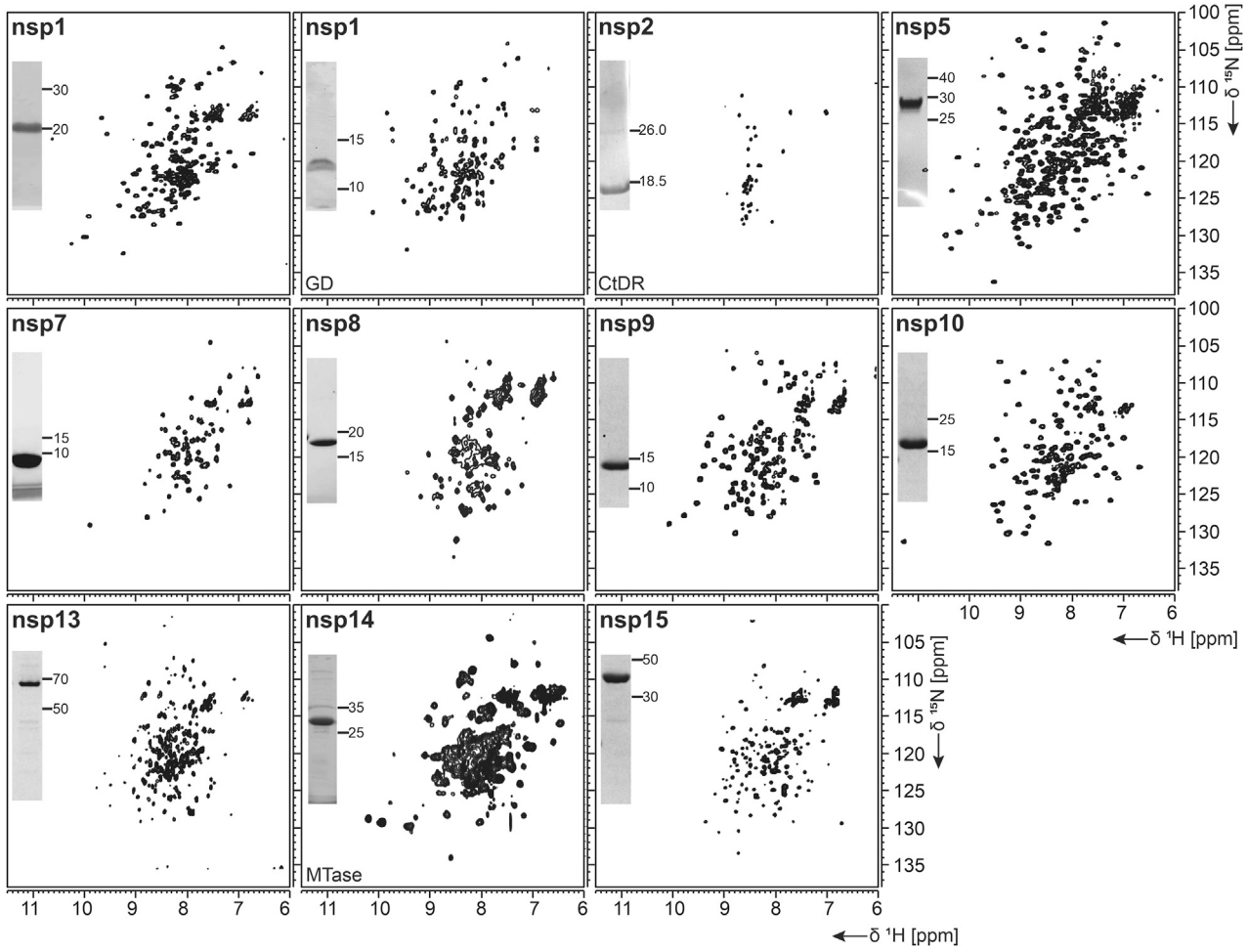


FIGURE 2 | ^{1}H, ^{15}N-correlation spectra of investigated nonstructural proteins. Construct names according to **Table 1** are indicated unless fl-proteins are shown. A representative SDS-PAGE lane with final samples is included as inset. Spectra for nsp3 constructs are collectively shown in **Figure 3**.

As summarized in **Table 1**, fl-domain boundaries of nsp1 were chosen to contain the first 180 amino acids, in analogy to its closest homolog from SCoV (Snijder et al., 2003). In addition, a shorter construct was designed, encoding only the globular core domain (GD, aa 13–127) suggested by the published SCoV nsp1 NMR structure (Almeida et al., 2007). His-tagged fl nsp1 was purified using the IPRS approach. Protein quality was confirmed by the available HSQC spectrum (**Figure 2**). Despite the flexible C-terminus, we were able to accomplish a near-complete backbone assignment (manuscript submitted).

Interestingly, the nsp1 GD was found to be problematic in our hands despite good expression. We observed insolubility, although buffers were used according to the homolog SCoV nsp1 GD (Almeida et al., 2007). Nevertheless, using a protocol comparable to the one for fl nsp1, we were able to record an HSQC spectrum proving a folded protein (**Figure 2**).

nsp2

nsp2 has been suggested to interact with host factors involved in intracellular signaling (Cornillez-Ty et al., 2009; Davies et al., 2020). The precise function, however, is insufficiently understood. Despite its potential dispensability for viral replication in general, it might be a valuable model to gain insights into virulence due to its possible involvement in the regulation of global RNA synthesis (Graham et al., 2005). We provide here a protocol for the purification of the C-terminal IDR (CtDR) of nsp2 from residues 557 to 601, based on disorder predictions [PrDOS (Ishida and Kinoshita, 2007)]. The His-Trx-tagged peptide was purified by IPRS. Upon dialysis, two IEC steps were performed: first anionic and then cationic, with good final yields (**Table 1**). Stability and purity were confirmed by an HSQC spectrum (**Figure 2**) and a complete backbone assignment (Mompean et al., 2020; **Table 2**).

nsp3

nsp3, the largest nsp (Snijder et al., 2003), is composed of a plethora of functionally related, yet independent, subunits. After cleavage of nsp3 from the fl ORF1-encoded polypeptide chain, it displays a 1945-residue multidomain protein, with individual functional entities that are subclassified from nsp3a to nsp3e

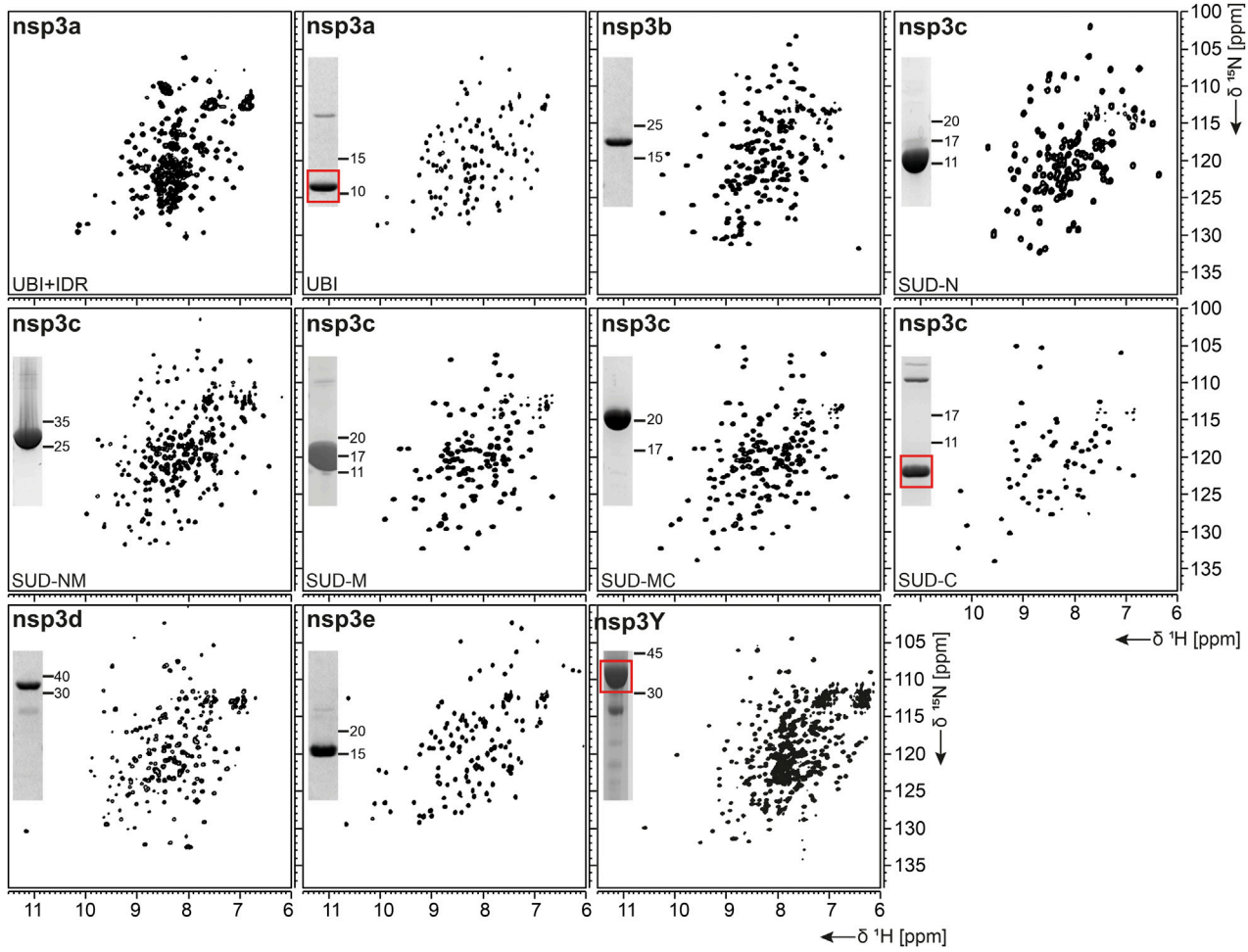


FIGURE 3 | ^{1}H, ^{15}N-correlation spectra of investigated constructs from nonstructural protein 3. Construct names of subdomains according to **Table 1** are indicated unless fl-domains are shown. A representative SDS-PAGE lane with final samples is included as inset. Red boxes indicate protein bands of interest.

followed by the ectodomain embedded in two TM regions and the very C-terminal CoV-Y domain. The soluble nsp3a-3e domains are linked by various types of linkers with crucial roles in the viral life cycle and are located in the so-called viral cytoplasm, which is separated from the host cell after budding off the endoplasmic reticulum and contains the viral RNA (Wolff et al., 2020). Remarkably, the nsp3c substructure comprises three subdomains, making nsp3 the most complex SCoV2 protein. The precise function and eventual RNA-binding specificities of nsp3 domains are not yet understood. We here focus on the nsp3 domains a–e and provide elaborated protocols for additional constructs carrying relevant linkers or combinations of domains (**Table 1**). Moreover, we additionally present a convenient protocol for the purification of the C-terminal CoV-Y domain.

nsp3a

The N-terminal portion of nsp3 is comprised of a ubiquitin-like (Ubl) structured domain and a subsequent acidic IDR. Besides its ability to bind ssRNA (Serrano et al., 2007), nsp3a has been reported to interact with the nucleocapsid (Hurst et al., 2013; Khan et al., 2020), playing a potential role in virus replication. We here provide protocols for the purification of both the Ubl (aa 1–111) and fl nsp3a (aa 1–206), including the acidic IDR (Ubl + IDR **Table 1**). Domain boundaries were defined similar to the published NMR structure of SCoV nsp3a (Serrano et al., 2007). His-tagged nsp3a Ubl + IDR and GST-tagged nsp3a Ubl were each purified via the IPRS approach. nsp3a Ubl yielded mM sample concentrations and displayed a well-dispersed HSQC spectrum (**Figure 3**). Notably, the herein described protocol also enables purification of fl nsp3a (Ubl + IDR) (**Tables 1**, **2**). Despite the unstructured IDR overhang, the excellent protein quality and stability allowed for near-complete backbone assignment [**Figure 3**, (Salvi et al., 2021)].

nsp3b

nsp3b is an ADP-ribose phosphatase macrodomain and potentially plays a key role in viral replication. Moreover, the de-ADP ribosylation function of nsp3b protects SCoV2 from antiviral host immune response, making nsp3b a promising drug target (Frick et al., 2020). As summarized in **Table 1**, the domain

boundaries of the herein investigated nsp3b are residues 207–376 of the nsp3 primary sequence and were identical to available crystal structures with PDB entries 6YWM and 6YWL (unpublished). For purification, we used the IPRS approach, which yielded pure fl nsp3b (**Table 2**). Fl nsp3b displays well-dispersed HSQC spectra, making this protein an amenable target for NMR structural studies. In fact, we recently reported near-to-complete backbone assignments for nsp3b in its apo and ADP-ribose–bound form (Cantini et al., 2020).

nsp3c

The SARS unique domain (SUD) of nsp3c has been described as a distinguishing feature of SCoVs (Snijder et al., 2003). However, similar domains in more distant CoVs, such as MHV or MERS, have been reported recently (Chen et al., 2015; Kusov et al., 2015). nsp3c comprises three distinct globular domains, termed SUD-N, SUD-M, and SUD-C, according to their sequential arrangement: N-terminal (N), middle (M), and C-terminal (C). SUD-N and SUD-M develop a macrodomain fold similar to nsp3b and are described to bind G-quadruplexes (Tan et al., 2009), while SUD-C preferentially binds to purine-containing RNA (Johnson et al., 2010). Domain boundaries for SUD-N and SUD-M and for the tandem-domain SUD-NM were defined in analogy to the SCoV homolog crystal structure (Tan et al., 2009). Those for SUD-C and the tandem SUD-MC were based on NMR solution structures of corresponding SARS-CoV homologs (**Table 1**) (Johnson et al., 2010). SUD-N, SUD-C, and SUD-NM were purified using GST affinity chromatography, whereas SUD-M and SUD-MC were purified using His affinity chromatography. Removal of the tag was achieved by thrombin cleavage and final samples of all domains were prepared subsequent to size-exclusion chromatography (SEC). Except for SUD-M, all constructs were highly stable (**Table 2**). Overall protein quality allowed for the assignment of backbone chemical shifts for the three single domains (Gallo et al., 2020) amd good resolved HSQC spectra also for the tandem domains (**Figure 3**).

nsp3d

nsp3d comprises the papain-like protease (PL^{pro}) domain of nsp3 and, hence, is one of the two SCoV2 proteases that are responsible for processing the viral polypeptide chain and generating functional proteins (Shin et al., 2020). The domain boundaries of PL^{pro} within nsp3 are set by residues 743 and 1,060 (**Table 1**). The protein is particularly challenging, as it is prone to misfolding and rapid precipitation. We prepared His-tagged and His-SUMO-tagged PL^{pro}. The His-tagged version mainly remained in the insoluble fraction. Still, mg quantities could be purified from the soluble fraction, however, greatly misfolded. Fusion to SUMO significantly enhanced protein yield of soluble PL^{pro}. The His-SUMO-tag allowed simple IMAC purification, followed by cleavage with Ulp1 and isolation of cleaved PL^{pro} via a second IMAC. A final purification step using gel filtration led to pure PL^{pro} of both unlabeled and 15N-labeled species (**Table 2**). The latter has allowed for the acquisition of a promising amide correlation spectrum (**Figure 3**).

nsp3e

nsp3e is unique to *Betacoronaviruses* and consists of a nucleic acid–binding domain (NAB) and the so-called group 2-specific marker (G2M) (Neuman et al., 2008). Structural information is rare; while the G2M is predicted to be intrinsically disordered (Lei et al., 2018); the only available experimental structure of the nsp3e NAB was solved from SCoV by the Wüthrich lab using solution NMR (Serrano et al., 2009). We here used this structure for a sequence-based alignment to derive reasonable domain boundaries for the SCoV2 nsp3e NAB (**Figures 4A,B**). The high sequence similarity suggested using nsp3 residues 1,088–1,203 (**Table 1**). This polypeptide chain was encoded in expression vectors comprising His- and His-GST tags, both cleavable by TEV protease. Both constructs showed excellent expression, suitable for the IPRS protocol (**Figure 4C**). Finally, a homogenous NAB species, as supported by the final gel of pooled samples (**Figure 4D**), was obtained. The excellent protein quality and stability are supported by the available HSQC (**Figure 3**) and a published backbone assignment (Korn et al., 2020a).

nsp3Y

nsp3Y is the most C-terminal domain of nsp3 and exists in all coronaviruses (Neuman et al., 2008; Neuman, 2016). Together, though, with its preceding regions G2M, TM 1, the ectodomain, TM2, and the Y1-domain, it has evaded structural investigations so far. The precise function of the CoV-Y domain remains unclear, but, together with the Y1-domain, it might affect binding to nsp4 (Hagemeijer et al., 2014). We were able to produce and purify nsp3Y (CoV-Y) comprising amino acids 1,638–1,945 (**Table 1**), yielding 12 mg/L with an optimized protocol that keeps the protein in a final NMR buffer containing HEPES and lithium bromide. Although the protein still shows some tendency to aggregate and degrade (**Table 2**), and despite its relatively large size, the spectral quality is excellent (**Figure 3**). nsp3 CoV-Y appears suitable for an NMR backbone assignment carried out at lower concentrations in a deuterated background (ongoing).

nsp5

The functional main protease nsp5 (Mpro) is a dimeric cysteine protease (Ullrich and Nitsche, 2020). Amino acid sequence and 3D structure of SCoV [PDB 1P9U (Anand et al., 2003)] and SCoV2 (PDB 6Y2E [Zhang et al., 2020)] homologs are highly conserved (**Figures 5A,B**). The dimer interface involves the N-termini of both monomers, which puts considerable constraints on the choice of protein sequence for construct design regarding the N-terminus.

We thus designed different constructs differing in the N-terminus: the native N-terminus (wt), a GS mutant with the additional N-terminal residues glycine and serine as His-SUMO fusion, and a GHM mutant with the amino acids glycine, histidine, and methionine located at the N-terminus with His-tag and TEV cleavage site (**Figure 5C**). Purification of all proteins via the IPRS approach (**Figures 5D,E**) yielded homogenous and highly pure protein, analyzed by PAGE (**Figure 5G**), mass

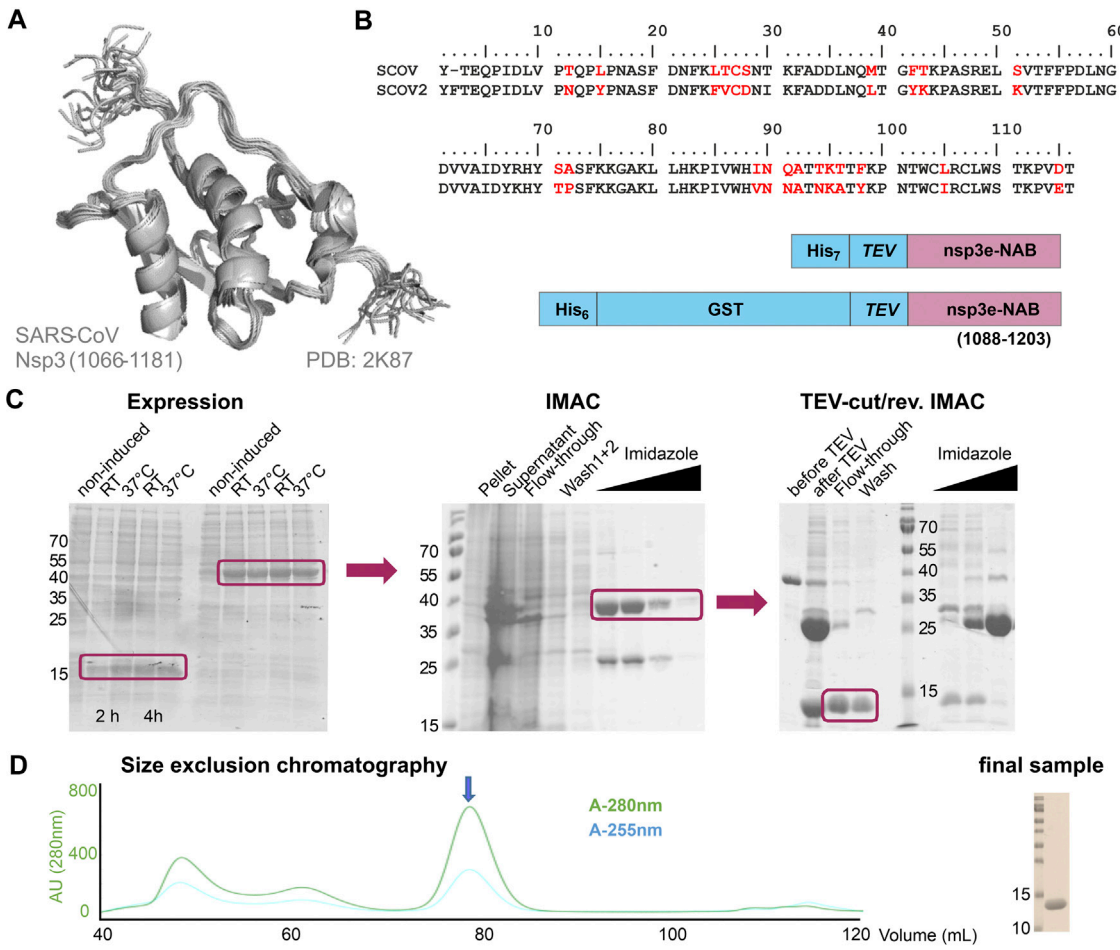


FIGURE 4 | Rationale of construct design, expression, and IPRS purification of the nsp3e nucleic acid–binding domain (NAB). **(A)** NMR structural ensemble of the homologous SCoV nsp3e (Serrano et al., 2009). The domain boundaries as displayed are given. **(B)** Sequence alignment of SCoV and SCoV2 regions representing the nsp3e locus. Arrows indicate the sequence stretch as used for the structure in panel **(A)**. The analogous region was used for the design of the two protein expression constructs shown **(C)**. Left, SDS-PAGE showing the expression of nsp3e constructs from panel **(B)** over 4 h at two different temperatures. Middle, SDS-PAGE showing the subsequent steps of IMAC. Right, SDS-PAGE showing steps and fractions obtained before and after TEV/dialysis and reverse IMAC. Boxes highlight the respective sample species of interest for further usage **(D)** SEC profile of nsp3e following steps in panel **(C)** performed with a Superdex 75 16/600 (GE Healthcare) column in the buffer as denoted in **Supplementary Table SI3**. The arrow indicates the protein peak of interest containing monomeric and homogenous nsp3e NAB devoid of significant contaminations of nucleic acids as revealed by the excellent 280/260 ratio. Right, SDS gel shows 0.5 µL of the final NMR sample used for the spectrum in **Figure 3** after concentrating relevant SEC fractions.

spectrometry, and 2D [^{1}H, ^{15}N]-BEST TROSY spectra (**Figure 5H**). Final yields are summarized in **Table 2**.

nsp7 and nsp8

Both nsp7 and nsp8 are auxiliary factors of the polymerase complex together with the RNA-dependent RNA polymerase nsp12 and have high sequence homology with SCoV (100% and 99%, respectively) (Gordon et al., 2020). For nsp7 in complex with nsp8 or for nsp8 alone, additional functions in RNA synthesis priming have been proposed (Tvarogova et al., 2019; Konkolova et al., 2020). In a recent study including an RNA-substrate-bound structure (Hillen et al., 2020), both proteins (with two molecules of nsp8 and one molecule of nsp7 for each nsp12 RNA polymerase) were found to be essential for polymerase activity in SCoV2. For both fl-proteins, a previously established expression and IPRS purification strategy for the SCoV proteins (Kirchdoerfer and Ward, 2019) was successfully transferred, which resulted in decent yields of reasonably stable proteins (**Table 2**). Driven by its intrinsically oligomeric state, nsp8 showed some tendency toward aggregation, limiting the available sample concentration. The higher apparent molecular weight and limited solubility are also reflected in the success of NMR experiments. While we succeeded in a complete NMR backbone assignment of nsp7 (Tonelli et al., 2020), the quality of the spectra obtained for nsp8 is currently limited to the HSQC presented in **Figure 2**.

nsp9

The 12.4 kDa ssRNA-binding nsp9 is highly conserved among *Betacoronaviruses*. It is a crucial part of the viral replication

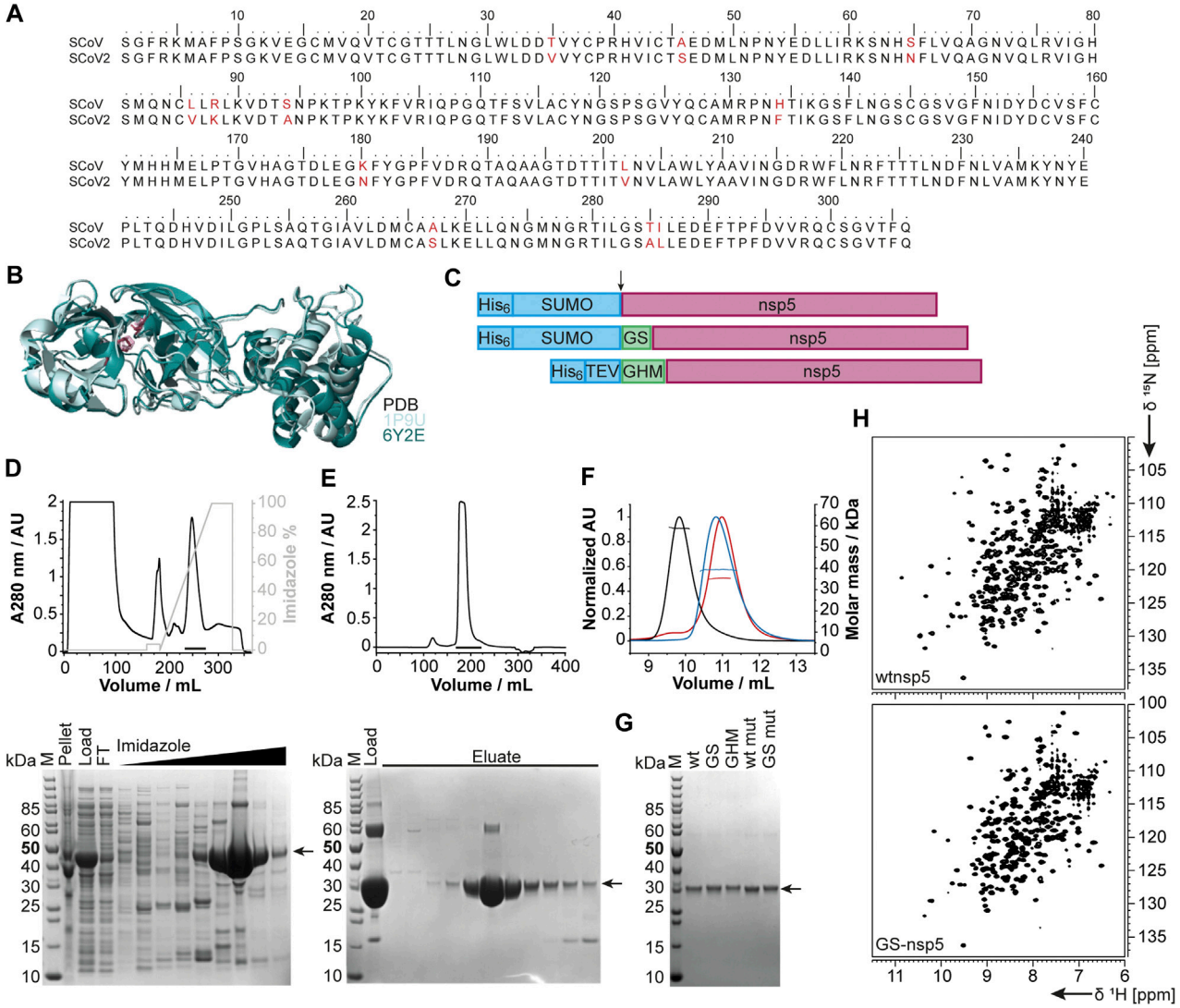


FIGURE 5 | Rationale of construct design, expression, and purification of different nsp5 constructs. **(A)** Sequence alignment of SCoV and SCoV2 fl nsp5. **(B)** X-ray structural overlay of the homologous SCoV (PDB 1 P9U, light blue) and SCoV2 nsp5 (PDB 6Y2E, green) in cartoon representation. The catalytic dyad (H41 and C145) is shown in stick representation (magenta). **(C)** Schematics of nsp5 expression constructs involving purification and solubilization tags (blue), different N-termini and additional aa after cleavage (green), and nsp5 (magenta). Cleavage sites are indicated by an arrow. **(D, E)** An exemplary purification is shown for wtnsp5. IMAC **(D)** and SEC **(E)** chromatograms (upper panels) and the corresponding gels (lower panels). Black bars in the chromatograms indicate pooled fractions. Gel samples are as follows: M: MW standard; pellet/load: pellet/supernatant after cell lysis; FT: IMAC flow-through; imidazole: eluted fractions with linear imidazole gradient; eluate: eluted SEC fractions from input (load). **(F)** SEC-MALS analysis with ~0.5 µg of wtnsp5 without additional aa (wtnsp5, black) with GS (GS-nsp5, blue) and with GHM (GHM-nsp5, red)) in NMR buffer on a Superdex 75, 10/300 GL (GE Healthcare) column. Horizontal lines indicate fractions of monodisperse nsp5 used for MW determination. **(G)** A SDS-PAGE showing all purified nsp5 constructs. The arrow indicates nsp5. **(H)** Exemplary [^{15}N, ^{1}H]-BEST-TROSY spectra measured at 298 K for the dimeric wtnsp5 (upper spectrum) and monomeric GS-nsp5 (lower spectrum). See **Supplementary Table SI4** for technical details regarding this figure.

machinery (Miknis et al., 2009), possibly targeting the 3'-end stem-loop II (s2m) of the genome (Robertson et al., 2005). nsp9 adopts a fold similar to oligonucleotide/oligosaccharide-binding proteins (Egloff et al., 2004), and structural data consistently uncovered nsp9 to be dimeric in solution (Egloff et al., 2004; Sutton et al., 2004; Miknis et al., 2009; Littler et al., 2020). Dimer formation seems to be a prerequisite for viral replication (Miknis et al., 2009) and influences RNA-binding (Sutton et al., 2004), despite a moderate affinity for RNA *in vitro* (Littler et al., 2020).

Based on the early available crystal structure of SCoV2 nsp9 (PDB 6W4B, unpublished), we used the 113 aa fl sequence of nsp9 for our expression construct (**Table 1**). Production of either His- or His-GST-tagged fl nsp9 yielded high amounts of soluble protein in both natural abundance and ^{13}C- and ^{15}N-labeled form. Purification *via* the IPRS approach enabled us to separate fl nsp9 in different oligomer states. The earliest eluted fraction represented higher oligomers and was contaminated with nucleic acids and it was not possible to concentrate above 2 mg/ml. This

was different for the subsequently eluting dimeric fl nsp9 fraction, which had a A260/280 ratio of below 0.7 and could be concentrated to >5 mg/ml (**Table 2**). The excellent protein quality and stability are supported by the available HSQC (**Figure 2**), and a near-complete backbone assignment (Dudas et al., 2021).

nsp10

The last functional protein encoded by ORF1a, nsp10, is an auxiliary factor for both the methyltransferase/exonuclease nsp14 and the 2′-O-methyltransferase (MTase) nsp16. However, it is required for the MTase activity of nsp16 (Krafcikova et al., 2020), it confers exonuclease activity to nsp14 in the RNA polymerase complex in SCoV (Ma et al., 2015). It contains two unusual zinc finger motifs (Joseph et al., 2006) and was initially proposed to comprise RNA-binding properties. We generated a construct (**Table 1**) containing an expression and affinity purification tag on the N-terminus as reported for the SCoV variant (Joseph et al., 2006). Importantly, additional Zn^{2+} ions present during expression and purification stabilize the protein significantly (Kubatova et al., 2020). The yield during isotope-labeling was high (**Table 2**), and tests in unlabeled rich medium showed the potential for yields exceeding 100 mg/L. These characteristics facilitated in-depth NMR analysis and a backbone assignment (Kubatova et al., 2020).

nsp13

nsp13 is a conserved ATP-dependent helicase that has been characterized as part of the RNA synthesis machinery by binding to nsp12 (Chen et al., 2020b). It represents an interesting drug target, for which the available structure (PDB 6ZSL) serves as an excellent basis (**Table 1**). The precise molecular function, however, has remained enigmatic since it is not clear whether the RNA unwinding function is required for making ssRNA accessible for RNA synthesis (Jia et al., 2019) or whether it is required for proofreading and backtracking (Chen et al., 2020b). We obtained pure protein using a standard expression vector, generating a His-SUMO-tagged protein. Following Ulp1 cleavage, the protein showed limited protein stability in the solution (**Table 2**).

nsp14

nsp14 contains two domains: an N-terminal exonuclease domain and a C-terminal MTase domain (Ma et al., 2015). The exonuclease domain interacts with nsp10 and provides part of the proofreading function that supports the high fidelity of the RNA polymerase complex (Robson et al., 2020). Several unusual features, such as the unusual zinc finger motifs, set it apart from other DEDD-type exonucleases (Chen et al., 2007), which are related to both nsp10 binding and catalytic activity. The MTase domain modifies the N7 of the guanosine cap of genomic and subgenomic viral RNAs, which is essential for the translation of viral proteins (Thoms et al., 2020). The location of this enzymatic activity within the RNA synthesis machinery ensures that newly synthesized RNA is rapidly capped and thus stabilized. As a strategy, we used constructs, which allow coexpression of both nsp14 and nsp10 (pRSFDuet and pETDuet, respectively). Production of isolated fl nsp14 was successful, however, with limited yield and stability (**Table 2**). Expression of the isolated MTase domain resulted in soluble protein with 27.5 kDa mass that was amenable to NMR characterization (**Figure 2**), although only under reducing conditions and in the presence of high (0.4 M) salt concentration.

nsp15

The poly-U-specific endoribonuclease nsp15 was one of the very first SCoV2 structures deposited in the PDB [6VWW, (Kim et al., 2020)]. Its function has been suggested to be related to the removal of U-rich RNA elements, preventing recognition by the innate immune system (Deng et al., 2017), even though the precise mechanism remains to be established. The exact role of the three domains (N-terminal, middle, and C-terminal catalytic domain) also remains to be characterized in more detail (Kim et al., 2020). Here, the sufficient yield of fl nsp15 during expression supported purification of pure protein, which, however, showed limited stability in solution (**Table 2**).

nsp16

The MTase reaction catalyzed by nsp16 is dependent on nsp10 as a cofactor (Krafcikova et al., 2020). In this reaction, the 2'-OH group of nucleotide +1 in genomic and subgenomic viral RNA is methylated, preventing recognition by the innate immune system. Since both nsp14 and nsp16 are in principle susceptible to inhibition by methyltransferase inhibitors, a drug targeting both enzymes would be highly desirable (Bouvet et al., 2010). nsp16 is the last protein being encoded by ORF1ab, and only its N-terminus is formed by cleavage by the Mpro nsp5. Employing a similar strategy to that for nsp14, nsp16 constructs were designed with the possibility of nsp10 coexpression. Expression of fl nsp16 resulted in good yields, when expressed both isolated and together with nsp10. The protein, however, is in either case unstable in solution and highly dependent on reducing buffer conditions (**Table 2**). The purification procedures of nsp16 were adapted with minor modifications from a previous X-ray crystallography study (Rosas-Lemus et al., 2020).

Structural Proteins and Accessory ORFs

Besides establishing expression and purification protocols for the nsps, we also developed protocols and obtained pure mg quantities of the SCoV2 structural proteins E, M, and N, as well as literally all accessory proteins. With the exception of the relatively well-behaved nucleocapsid (N) protein, SCoV2 E, M, and the remaining accessory proteins represent a class of mostly small and relatively poorly characterized proteins, mainly due to their difficult behavior in classical expression systems.

We used wheat-germ cell-free protein synthesis (WG-CFPS) for the successful production, solubilization, purification, and, in part, initial NMR spectroscopic

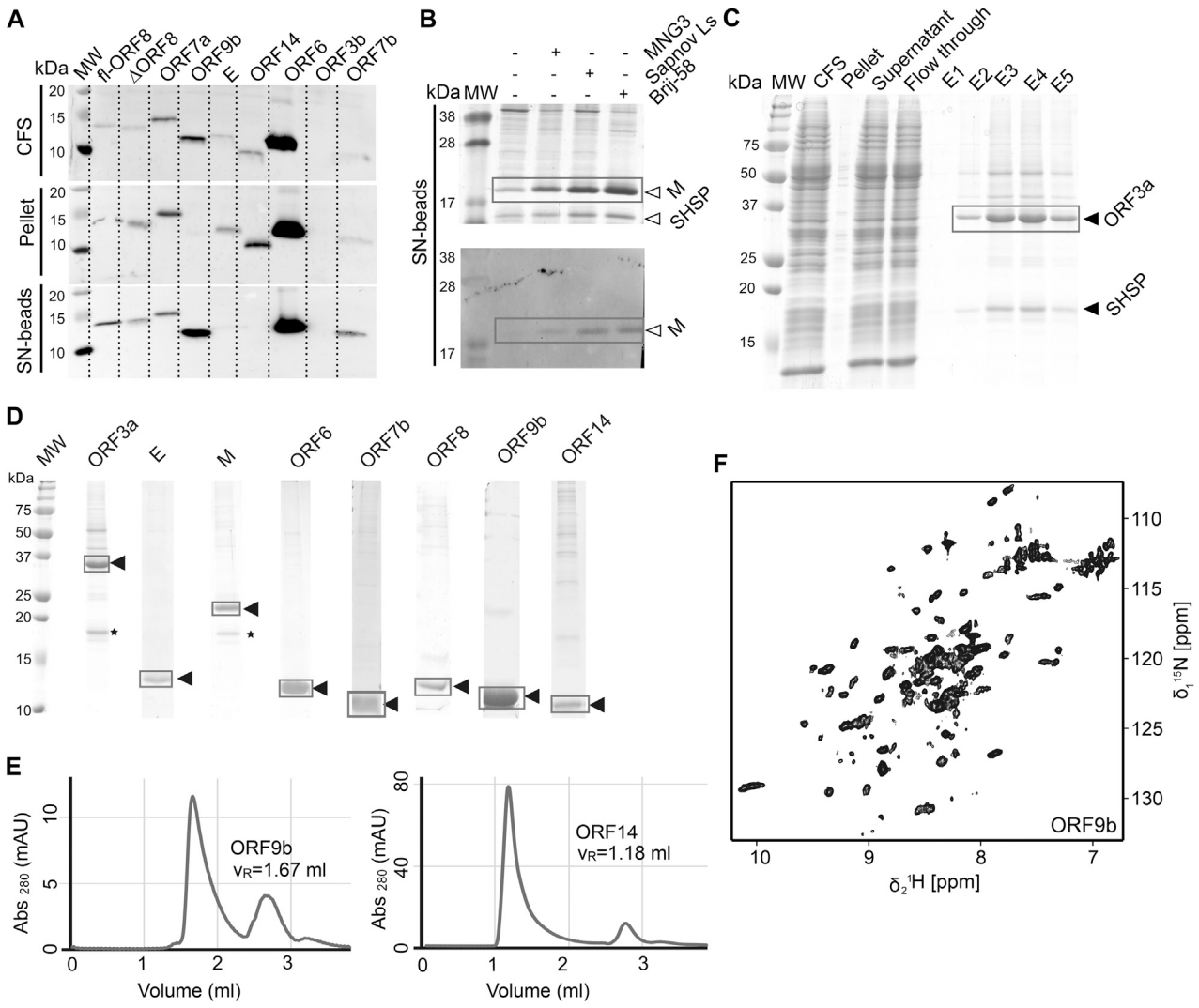


FIGURE 6 | Cell-free protein synthesis of accessory ORFs and structural proteins E and M. **(A)** Screening for expression and solubility of different ORFs using small-scale reactions. The total cell-free reaction (CFS), the pellet after centrifugation, and the supernatant (SN) captured on magnetic beads coated with Strep-Tactin were analyzed. All tested proteins were synthesized, with the exception of ORF3b. MW, MW standard. **(B)** Detergent solubilization tests using three different detergents, here at the example of the M protein, shown by SDS-PAGE and Western Blot. **(C)** Proteins are purified in a single step using a Strep-Tactin column. For ORF3a (and also for M), a small heat-shock protein of the HSP20 family is copurified, as identified by mass spectrometry (see also * in Panel **D**). **(D)** SDS-PAGE of the ^{2}H, ^{13}C, ^{15}N-labeled proteins used as NMR samples. Yields were between 0.2 and 1 mg protein per mL wheat-germ extract used. **(E)** SEC profiles for two ORFs. Left, ORF9b migrates as expected for a dimer. Right, OFR14 shows large assemblies corresponding to approximately 9 protein units and the DDM detergent micelle. **(F)** 2D [^{15}N, ^{1}H]-BEST-TROSY spectrum of ORF9b, recorded at 900 MHz in 1 h at 298 K, on less than 1 mg of protein. See **Supplementary Tables SI13–SI19** and **Supplementary Tables SI19, SI20** for technical and experimental details regarding this figure.

investigation of ORF3a, ORF6, ORF7b, ORF8, ORF9b, and ORF14 accessory proteins, as well as E and M in mg quantities using the highly efficient translation machinery extracted from wheat-germs (**Figures 6A–D**).

ORF3a

Protein from ORF3a in SCoV2 corresponds to accessory protein 3a in SCoV, with homology of more than 70% (**Table 1**). It has 275 amino acids, and its structure has recently been determined (Kern et al., 2020). The structure of SCoV2 3a displays a dimer, but it can also form higher oligomers. Each monomer has three TM helices and a cytosolic β-strand rich domain. SCoV2 ORF3a is a cation channel, and its structure has been solved by electron microscopy in nanodiscs. In SCoV, 3a is a structural component and was found in recombinant virus-like particles (Liu et al., 2014), but is not explicitly needed for their formation. The major challenge for NMR studies of this largest accessory protein is its size, independent of its employment in solid state or solution NMR spectroscopy.

As most other accessory proteins described in the following, ORF3a has been produced using WG-CFPS and was expressed in soluble form in the presence of Brij-58

(**Figure 6C**). It is copurified with a small heat-shock protein of the HSP20 family from the wheat-germ extract. The protocol described here is highly similar to that of the other cell-free synthesized accessory proteins. Where NMR spectra have been reported, the protein has been produced in a ^{2}H, ^{13}C, ^{15}N uniformly labeled form; otherwise, natural abundance amino acids were added to the reaction. The proteins were further affinity-purified in one step using Strep-Tactin resin, through the Strep-tag II fused to their N- or C-terminus. For membrane proteins, protein synthesis and also purification were done in the presence of detergent.

About half a milligram of pure protein was generally obtained per mL of extract, and up to 3 ml wheat-germ extract has been used to prepare NMR samples.

ORF3b

The ORF3b protein is a putative protein stemming from a short ORF (57 aa) with no homology to existing SCoV proteins (Chan et al., 2020). Indeed, ORF3b gene products of SCoV2 and SCoV are considerably different, with one of the distinguishing features being the presence of premature stop codons, resulting in the expression of a drastically shortened ORF3b protein (Konno et al., 2020). However, the SCoV2 nucleotide sequence after the stop codon shows a high similarity to the SCoV ORF3b. Different C-terminal truncations seem to play a role in the interferon-antagonistic activity of ORF3b (Konno et al., 2020). ORF3b is the only protein that, using WG-CFPS, was not synthesized at all; i.e., it was neither observed in the total cell-free reaction nor in supernatant or pellet. This might be due to the premature stop codon, which was not considered. Constructs of ORF3b thus need to be redesigned.

ORF4 (Envelope Protein, E)

The SCoV2 envelope (E) protein is a small (75 amino acids), integral membrane protein involved in several aspects of the virus' life cycle, such as assembly, budding, envelope formation, and pathogenicity, as recently reviewed in (Schoeman and Fielding, 2020). Structural models for SCoV (Surya et al., 2018) and the TM helix of SCoV2 (Mandala et al., 2020) E have been established. The structural models show a pentamer with a TM helix. The C-terminal part is polar, with charged residues interleaved, and is positioned on the membrane surface in SCoV. E was produced in a similar manner to ORF3a, using the addition of detergent to the cell-free reaction.

ORF5 (Membrane Glycoprotein, M)

The M protein is the most abundant protein in the viral envelope and is believed to be responsible for maintaining the virion in its characteristic shape (Huang et al., 2004). M is a glycoprotein and sequence analyses predict three domains: A C-terminal endodomain, a TM domain with three predicted helices, and a short N-terminal ectodomain. M is essential for viral particle assembly. Intermolecular interactions with the other structural proteins, N and S to a lesser extent, but most importantly E (Vennema et al., 1996), seem to be central for virion envelope formation in coronaviruses, as M alone is not sufficient. Evidence has been presented that M could adopt two conformations, elongated and compact, and that the two forms fulfill different functions (Neuman et al., 2011). The lack of more detailed structural information is in part due to its small size, close association with the viral envelope, and a tendency to form insoluble aggregates when perturbed (Neuman et al., 2011). The M protein is readily produced using cell-free synthesis in the presence of detergent; as ORF3a, it is copurified with a small heat-shock protein of the HSP20 family (**Figure 6B**). Membrane-reconstitution will likely be necessary to study this protein.

ORF6

The ORF6 protein is incorporated into viral particles and is also released from cells (Huang et al., 2004). It is a small protein (61 aa), which has been found to concentrate at the endoplasmic reticulum and Golgi apparatus. In a murine coronavirus model, it was shown that expressing ORF6 increased virulence in mice (Zhao et al., 2009), and results indicate that ORF6 may serve an important role in the pathogenesis during SCoV infection (Liu et al., 2014). Also, it showed to inhibit the expression of certain STAT1-genes critical for the host immune response and could contribute to the immune evasion. ORF6 is expressed very well in WG-CFPS; the protein was fully soluble with detergents and partially soluble without them and was easily purified in the presence of detergent, but less efficiently in the absence thereof. Solution NMR spectra in the presence of detergent display narrow but few resonances, which correspond, in addition to the C-terminal STREP-tag, to the very C-terminal ORF6 protein residues.

ORF7a

SCoV2 protein 7a (121 aa) shows over 85% homology with the SCoV protein 7a. While the SCoV2 7a protein is produced and retained intracellularly, SCoV protein 7a has also been shown to be a structural protein incorporated into mature virions (Liu et al., 2014). 7a is one of the accessory proteins, of which a (partial) structure has been determined at high resolution for SCoV2 (PDB 6W37). However, the very N-terminal signal peptide and the C-terminal membrane anchor, both highly hydrophobic, have not been determined experimentally yet.

Expression of the ORF7a ectodomain (ED) with a GB1 tag (Bogomolovas et al., 2009) was expected to produce reasonable yields. The IPRS purification resulted in a highly stable protein, as evidenced by the NMR data obtained (**Figure 7**).

ORF7b

Protein ORF7b is associated with viral particles in a SARS context (Liu et al., 2014). Protein 7b is one of the shortest ORFs with 43 residues. It shows a long hydrophobic stretch, which might correspond to a TM segment. It shows over 93% sequence homology with a bat coronavirus 7b protein (Liu et al., 2014). There, the cysteine residue in the C-terminal part is not conserved, which might facilitate structural studies. ORF7b has

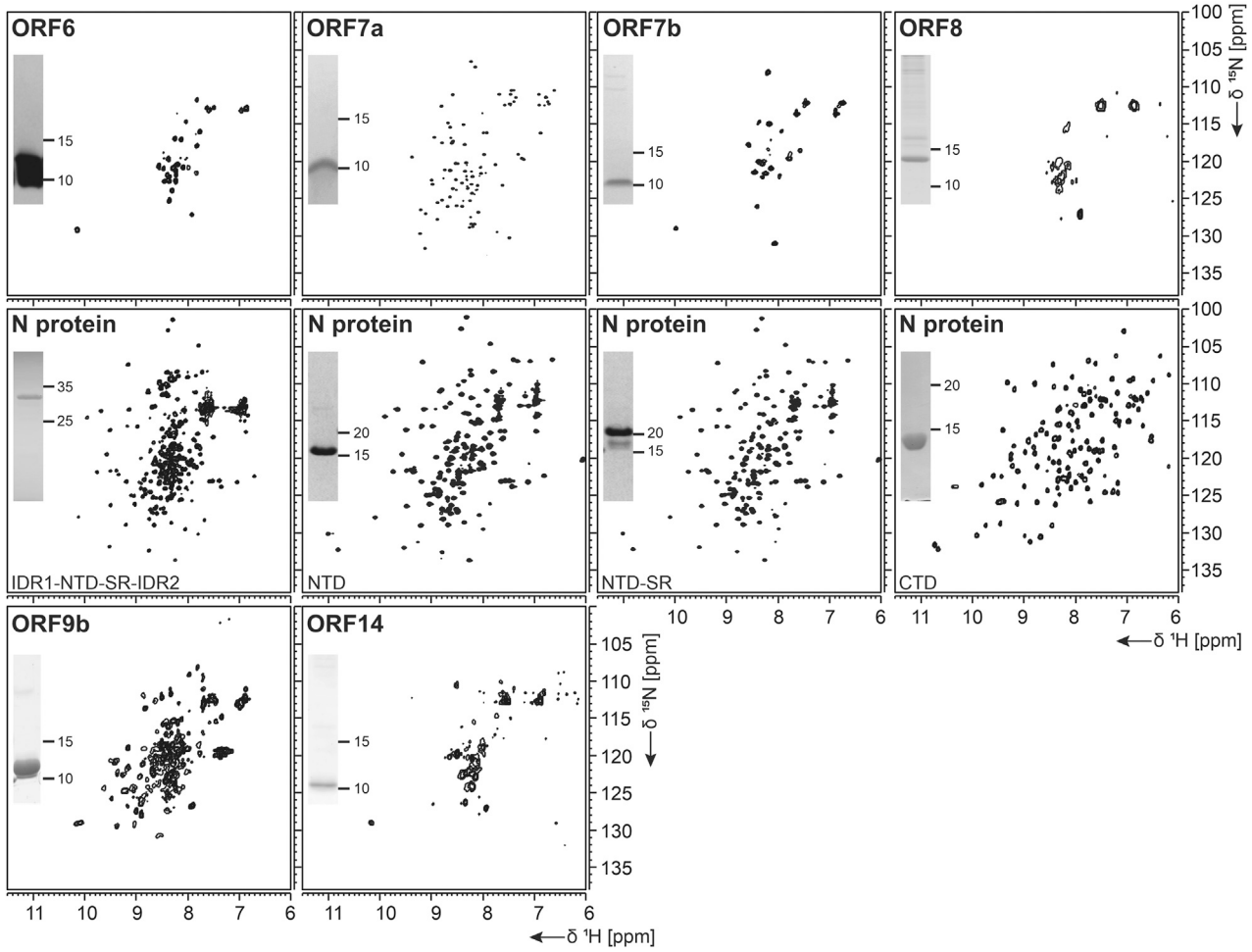


FIGURE 7 | ^{1}H, ^{15}N-correlation spectra of investigated structural and accessory proteins. Construct names according to **Table 1** are indicated unless fl-proteins are shown. A representative SDS-PAGE lane with final samples is included as inset.

been synthesized successfully both from bacteria and by WG-CFPS in the presence of detergent and could be purified using a STREP-tag (**Table 2**). Due to the necessity of solubilizing agent and its obvious tendency to oligomerize, structure determination, fragment screening, and interaction studies are challenging. However, we were able to record the first promising HSQC, as shown in **Figure 7**.

ORF8

ORF 8 is believed to be responsible for the evolution of *Betacoronaviruses* and their species jumps (Wu et al., 2016) and to have a role in depressing the host response (Tan et al., 2020). ORF 8 (121 aa) from SCoV2 does not apparently exist in SCoV on the protein level, despite the existence of a putative ORF. The sequences of the two homologs only show limited identity, with the exception of a small 7 aa segment, where, in SCoV, the glutamate is replaced with an aspartate. It, however, aligns very well with several coronaviruses endemic to animals, including Paguma and Bat (Chan et al., 2020). The protein comprises a hydrophobic peptide at its very N-terminus, likely corresponding to a signal peptide; the remaining part does not show any specific sequence features. Its structure has been determined (PDB 7JTL) and shows a similar fold to ORF7a (Flower et al., 2020). In this study, ORF8 has been used both with (fl) and without signal peptide (ΔORF8). We first tested the production of ORF8 in *E. coli*, but yields were low because of insolubility. Both ORF8 versions have then been synthesized in the cell-free system and were soluble in the presence of detergent. Solution NMR spectra, however, indicate that the protein is forming either oligomers or aggregates.

ORF9a (Nucleocapsid Protein, N)

The nucleocapsid protein (N) is important for viral genome packaging (Luo et al., 2006). The multifunctional RNA-binding protein plays a crucial role in the viral life cycle (Chang et al., 2014) and its domain architecture is highly conserved among coronaviruses. It comprises the N-terminal intrinsically disordered region (IDR1), the N-terminal RNA-

binding globular domain (NTD), a central serine/arginine- (SR-) rich intrinsically disordered linker region (IDR2), the C-terminal dimerization domain (CTD), and a C-terminal intrinsically disordered region (IDR3) (Kang et al., 2020).

N represents a highly promising drug target. We thus focused our efforts not exclusively on the NTD and CTD alone, but, in addition, also provided protocols for IDR-containing constructs within the N-terminal part.

N-Terminal RNA-Binding Globular Domain

The NTD is the RNA-binding domain of the nucleocapsid (Kang et al., 2020). It is embedded within IDRs functions of which have not yet been deciphered. Recent experimental and bioinformatic data indicate involvement in liquid-liquid phase separation (Chen et al., 2020a).

For the NTD, several constructs were designed, also considering the flanking IDRs (**Table 1**). In analogy to the available NMR [PDB 6YI3, (Dinesh et al., 2020)] and crystal [PDB 6M3M, (Kang et al., 2020)] structures of the SCoV2 NTD, boundaries for the NTD and the NTD-SR domains were designed to span residues 44–180 and 44–212, respectively. In addition, an extended IDR1-NTD-IDR2 (residues 1–248) was designed, including the N-terminal disordered region (IDR1), the NTD domain, and the central disordered linker (IDR2) that comprises the SR region. His-tagged NTD and NTD-SR were purified using IPRS and yielded approx. 3 mg/L in 15N-labeled minimal medium. High protein quality and stability are supported by the available HSQC spectra (**Figure 7**).

The untagged IDR1-NTD-IDR2 was purified by IEC and yielded high amounts of 13C, 15 N-labeled samples of 12 mg/L for further NMR investigations. The quality of our purification is confirmed by the available HSQC (**Figure 7**), and a near-complete backbone assignment of the two IDRs was achieved (manuscript submitted). Notably, despite the structurally and dynamically heterogeneous nature of the N protein, the mentioned N constructs revealed a very good long-term stability, as shown in **Table 2**.

C-Terminal Domain

Multiple studies on the SCoV2 CTD, including recent crystal structures (Ye et al., 2020; Zhou et al., 2020), confirm the domain as dimeric. Its ability to self-associate seems to be necessary for viral replication and transcription (Luo et al., 2006). In addition, the CTD was shown to, presumably nonspecifically, bind ssRNA (Zhou et al., 2020).

Domain boundaries for the CTD were defined to comprise amino acids 247–364 (**Table 1**), in analogy to the NMR structure of the CTD from SCoV (PDB 2JW8, [Takeda et al., 2008)]. Gene expression of His- or His-GST-tagged CTD yielded high amounts of soluble protein. Purification was achieved via IPRS. The CTD eluted as a dimer judged by its retention volume on the size-exclusion column and yielded good amounts (**Table 2**). The excellent protein quality and stability are supported by the available HSQC spectrum (**Figure 7**) and a near-complete backbone assignment (Korn et al., 2020b).

ORF9b

Protein 9b (97 aa) shows 73% sequence homology to the SCoV and also to bat virus (bat-SL-CoVZXC21) 9b protein (Chan et al., 2020). The structure of SCoV2 ORF9b has been determined at high resolution (PDB 6Z4U). Still, a significant portion of the structure was not found to be well ordered. The protein shows a β-sheet-rich structure and a hydrophobic tunnel, in which bound lipid was identified. How this might relate to membrane binding is not fully understood at this point. The differences in sequence between SCoV and SCoV2 are mainly located in the very N-terminus, which was not resolved in the structure (PDB 6Z4U). Another spot of deviating sequence not resolved in the structure is a solvent-exposed loop, which presents a potential interacting segment. ORF9b has been synthesized as a dimer (**Figure 6E**) using WG-CFPS in its soluble form. Spectra show a well-folded protein, and assignments are underway (**Figure 6F**).

ORF14 (ORF9c)

ORF14 (73 aa) remains, at this point in time, hypothetical. It shows 89% homology with a bat virus protein (bat-SL-CoVZXC21). It shows a highly hydrophobic part in its C-terminal region, comprising two negatively charged residues and a charged/polar N-terminus. The C-terminus is likely mediating membrane interaction. While ORF14 has been synthesized in the wheat-germ cell-free system in the presence of detergent and solution NMR spectra have been recorded, they hint at an aggregated protein (**Figure 6E**). Membrane-reconstitution of ORF14 revealed an unstable protein, which had been degraded during detergent removal.

ORF10

The ORF10 protein is comprised of 38 aa and is a hypothetical protein with unknown function (Yoshimoto, 2020). SCoV2 ORF10 displays 52.4% homology to SCoV ORF9b. The protein sequence is rich in hydrophobic residues, rendering expression and purification challenging. Expression of ORF10 as His-Trx-tagged or His-SUMO tagged fusion protein was possible; however, the ORF10 protein is poorly soluble and shows partial unfolding, even as an uncleaved fusion protein. Analytical SEC hints at oligomerization under the current conditions.

DISCUSSION

The ongoing SCoV2 pandemic and its manifestation as the COVID-19 disease call for an urgent provision of therapeutics that will specifically target viral proteins and their interactions with each other and RNAs, which are crucial for viral propagation. Two “classical” viral targets have been addressed in comprehensive approaches soon after the outbreak in December 2019: the viral protease nsp5 and the RNA-dependent RNA polymerase (RdRp) nsp12. While the latter turned out to be a suitable target using the repurposed compound Remdesivir (Hillen et al., 2020), nsp5 is undergoing a broad structure-based screen against a battery of inhibitors in

multiple places (Jin et al., 2020; Zhang et al., 2020), but with, as of yet, the limited outcome for effective medication. Hence, a comprehensive, reliable treatment of COVID-19 at any stage after the infection has remained unsuccessful.

Further viral protein targets will have to be taken into account in order to provide inhibitors with increased specificity and efficacy and preparative starting points for following potential generations of (SARS-)CoVs. Availability of those proteins in a recombinant, pure, homogenous, and stable form in milligrams is, therefore, a prerequisite for follow-up applications like vaccination, high-throughput screening campaigns, structure determination, and mapping of viral protein interaction networks. We here present, for the first time, a near-complete compendium of SCoV2 protein purification protocols that enable the production of large amounts of pure proteins.

The *COVID19-NMR* consortium was launched with the motivation of providing NMR assignments of all SCoV2 proteins and RNA elements, and enormous progress has been made since the outbreak of COVID-19 for both components [see **Table 2** and (Wacker et al., 2020)]. Consequently, we have put our focus on producing proteins in stable isotope-labeled forms for NMR-based applications, e.g., the site-resolved mapping of interactions with compounds (Li and Kang, 2020). Relevant to a broad scientific community, we here report our protocols to suite perfectly any downstream biochemical or biomedical application.

Overall Success and Protein Coverage

As summarized in **Table 2**, we have successfully purified 80% of the SCoV2 proteins either in fl or providing relevant fragments of the parent protein. Those include most of the nsps, where all of the known/predicted soluble domains have been addressed (**Figure 1**). For a very large part, we were able to obtain protein samples of high purity, homogeneity, and fold for NMR-based applications. We would like to point out a number of CoV proteins that, evidenced by their HSQCs, for the first time, provide access to structural information, e.g., the PL^{pro} nsp3d and nsp3Y. Particularly for the nsp3 multidomain protein, we here present soluble samples of almost the complete cytosolic region with more than 120 kDa in the form of excellent 2D NMR spectra (**Figure 3**), a major part of which fully backbone-assigned. We thus enable the exploitation of the largest and most enigmatic multifunctional SCoV2 protein through individual domains in solution, allowing us to study their concerted behavior with single residue resolution. Similarly, for nsp2, we provide a promising starting point for studying the so far neglected, often uncharacterized, and apparently unstructured proteins.

Driven by the fast-spreading COVID-19, we initially left out proteins that require advanced purification procedures (e.g., nsp12 and S) or where *a priori* information was limited (nsp4 and nsp6). This procedure seems justified with the time-saving approach of our effort in favor of the less attended proteins. However, we are in the process of collecting protocols for the missing proteins.

Different Complexities and Challenges

The compilation of protein production protocols, initially guided by information from CoV homologs (**Table 1**), has confronted us with very different levels of complexity. With some prior expectation toward this, we have shared forces to quickly "work off" the highly conserved soluble and small proteins and soon put focus into the processing of the challenging ones. The difficulties in studying this second class of proteins are due to their limited sequence conservation, no prior information, large molecular weights, insolubility, and so forth.

The nsp3e NAB represents one example where the available NMR structure of the SCoV homolog provided a *bona fide* template for selecting initial domain boundaries (**Figure 4**). The transfer of information derived from SCoV was straightforward; the transferability included the available protocol for the production of comparable protein amounts and quality, given the high sequence identity. In such cases, we found ourselves merely to adapt protocols and optimize yields based on slightly different expression vectors and *E. coli* strains.

However, in some cases, such transfer was unexpectedly not successful, e.g., for the short nsp1 GD. Despite intuitive domain boundaries with complete local sequence identity seen from the SCoV nsp1 NMR structure, it took considerable efforts to purify an analogous nsp1 construct, which is likely related to the impaired stability and solubility caused by a number of impacting amino acid exchanges within the domain's flexible loops. In line with that, currently available structures of SCoV2 nsp1 have been obtained by crystallography or cryo-EM and include different buffers. As such, our initial design was insufficient in terms of taking into account the parameters mentioned above. However, one needs to consider those particular differences between the nsp1 homologs as one of the most promising target sites for potential drugs as they appear to be hotspots in the CoV evolution and will have essential effects for the molecular networks, both in the virus and with the host (Zust et al., 2007; Narayanan et al., 2015; Shen et al., 2019; Thoms et al., 2020).

A special focus was put on the production of the SCoV2 main protease nsp5, for which NMR-based screenings are ongoing. The main protease is critical in terms of inhibitor design as it appears under constant selection, and novel mutants remarkably influence the structure and biochemistry of the protein (Cross et al., 2020). In the present study, the expression of the different constructs allowed us to characterize the protein in both its monomeric and dimeric forms. Comparison of NMR spectra reveals that the constructs with additional amino acids (GS and GHM mutant) display marked structural differences to the wild-type protein while being structurally similar among themselves (**Figure 5H**). The addition of two residues (GS) interferes with the dimerization interface, despite being similar to its native N-terminal amino acids (SGFR). We also introduced an active site mutation that replaces cysteine 145 with alanine (Hsu et al., 2005). Intriguingly, this active site mutation C145A, known to stabilize the dimerization of the main protease (Chang et al., 2007), supports dimer formation of the GS added construct (GS-nsp5 C145A) shown by its 2D NMR spectrum overlaying with the one of wild-type nsp5 (**Supplementary Table SI4**). The NMR

results are in line with SEC-MALS analyses (**Figure 5F**). Indeed, the additional amino acids at the N-terminus shift the dimerization equilibrium toward the monomer, whereas the mutation shifts it toward the dimer despite the N-terminal aa additions. This example underlines the need for a thorough and precise construct design and the detailed biochemical and NMR-based characterization of the final sample state. The presence of monomers vs. dimers will play an essential role in the inhibitor search against SCoV2 proteins, as exemplified by the particularly attractive nsp5 main protease target.

Exploiting Nonbacterial Expression

As a particular effort within this consortium, we included the so far neglected accessory proteins using a structural genomics procedure supported by wheat-germ cell-free protein synthesis. This approach allowed us previously to express a variety of difficult viral proteins in our hands (Fogeron et al., 2015a; Fogeron et al., 2015b; Fogeron et al., 2016; Fogeron et al., 2017; Wang et al., 2019; Jirasko et al., 2020a). Within the workflow, we especially highlight the straightforward solubilization of the membrane proteins through the addition of detergent to the cell-free reaction, which allowed the production of soluble protein in milligram amounts compatible with NMR studies. While home-made extracts were used here, very similar extracts are available commercially (Cell-Free Sciences, Japan) and can thus be implemented by any lab without prior experience. Also, a major benefit of the WG-CFPS system for NMR studies lies in the high efficiency and selectivity of isotopic labeling. In contrast to cell-based expression systems, only the protein of interest is produced (Morita et al., 2003), which allows bypassing extensive purification steps. In fact, one-step affinity purification is in most cases sufficient, as shown for the different ORFs in this study. Samples could be produced for virtually all proteins, with the exception of the ORF3b construct used. With new recent insight into the stop codons present in this ORF, constructs will be adapted, which shall overcome the problems of ORF3b production (Konno et al., 2020).

For two ORFs, 7b and 8, we exploited a paralleled production strategy, i.e., both in bacteria and via cell-free synthesis. For those challenging proteins, we were, in principle, able to obtain pure samples from either expression system. However, for ORF7b, we found a strict dependency on detergents for follow-up work from both approaches. ORF8 showed significantly better solubility when produced in WG extracts compared to bacteria. This shows the necessity of parallel routes to take, in particular, for the understudied, biochemically nontrivial ORFs that might represent yet unexplored but highly specific targets to consider in the treatment of COVID-19.

Downstream structural analysis of ORFs produced with CFPS remains challenging but promising progress is being made in the light of SCoV2. Some solution NMR spectra show the expected number of signals with good resolution (e.g., ORF9b). As expected, however, most proteins cannot be straightforwardly analyzed by solution NMR in their current form, as they exhibit too large objects after insertion into micelles and/or by inherent oligomerization. Cell-free synthesized proteins can be inserted into membranes through reconstitution (Fogeron et al., 2015a; Fogeron et al., 2015b; Fogeron et al., 2016; Jirasko et al., 2020a; Jirasko et al., 2020b). Reconstitution will thus be the next step for many accessory proteins, but also for M and E, which were well produced by WG-CFPS. We will also exploit the straightforward deuteration in WG-CFPS (David et al., 2018; Wang et al., 2019; Jirasko et al., 2020a) that circumvents proton back-exchange, rendering denaturation and refolding steps obsolete (Tonelli et al., 2011). Nevertheless, the herein presented protocols for the production of non-nsps by WG-CFPS instantly enable their employment in binding studies and screening campaigns and thus provide a significant contribution to soon-to-come studies on SCoV2 proteins beyond the classical and convenient drug targets.

Altogether and judged by the ultimate need of exploiting recombinant SCoV2 proteins in vaccination and highly paralleled screening campaigns, we optimized sample amount, homogeneity, and long-term stability of samples. Our freely accessible protocols and accompanying NMR spectra now offer a great resource to be exploited for the unambiguous and reproducible production of SCoV2 proteins for the intended applications.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

NA, SK, NQ, MD, MN, AB, HS, MH, and AS designed the study, compiled the protocols and NMR data, and wrote the manuscript. All authors contributed coordinative or practical work to the study. All authors contributed to the creation and collection of protein protocols.

FUNDING

This work was supported by Goethe University (Corona funds), the DFG-funded CRC: "Molecular Principles of RNA-Based Regulation," DFG infrastructure funds (project numbers: 277478796, 277479031, 392682309, 452632086, 70653611), the state of Hesse (BMRZ), the Fondazione CR Firenze (CERM), and the IWB-EFRE-program 20007375. This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 871037. AS is supported by DFG Grant SCHL 2062/2-1 and by the JQYA at Goethe through project number 2019/AS01. Work in the lab of KV was supported by a CoRE grant from the University of New Hampshire. The FLI is a member of the Leibniz Association (WGL) and financially supported by the Federal Government of Germany and the State of Thuringia. Work in the lab of RM was supported by NIH (2R01EY021514) and NSF (DMR-2002837). BN-B was supported by the NSF GRFP. MC was supported by

NIH (R25 GM055246 MBRS IMSD), and MS-P was supported by the HHMI Gilliam Fellowship. Work in the labs of KJ and KT was supported by Latvian Council of Science Grant No. VPPCOVID 2020/1-0014. Work in the UPAT's lab was supported by the INSPIRED (MIS 5002550) project, which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure," funded by the Operational Program "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014–2020) and cofinanced by Greece and the EU (European Regional Development Fund) and the FP7 REGPOT CT-2011-285950–"SEE-DRUG" project (purchase of UPAT's 700 MHz NMR equipment). Work in the CM-G lab was supported by the Helmholtz society. Work in the lab of AB was supported by the CNRS, the French National Research Agency (ANR, NMR-SCoV2- ORF8), the Fondation de la Recherche Médicale (FRM, NMR-SCoV2-ORF8), and the IR-RMN-THC Fr3050 CNRS. Work in the lab of BM was supported by the Swiss National Science Foundation (Grant number 200020_188711), the Günthard Stiftung für Physikalische Chemie, and the ETH Zurich. Work in the labs of AB and BM was supported by a common grant from SNF (grant 31CA30_196256). This work was supported by the ETH Zurich, the grant ETH 40 18 1, and the grant Krebsliga KFS 4903 08 2019. Work in the lab of the IBS Grenoble was supported by the Agence Nationale de Recherche (France) RA-COVID SARS2NUCLEOPROTEIN and European Research Council Advanced Grant DynamicAssemblies. Work in the CA lab was supported by Patto per il Sud della Regione Siciliana–CheMISt grant (CUP G77B17000110001). Part of this work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-05-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003).

ACKNOWLEDGMENTS

The authors thank Leonardo Gonnelli and Katharina Targaczewski for the valuable technical assistance. IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG CEA). They acknowledge the Advanced Technologies Network Center of the University of Palermo to support infrastructures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.653148/full#supplementary-material.

REFERENCES

Almeida, M. S., Johnson, M. A., Herrmann, T., Geralt, M., and Wüthrich, K. (2007). Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. *J. Virol.* 81 (7), 3151–3161. doi:10.1128/JVI.01939-06

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., and Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300 (5626), 1763–1767. doi:10.1126/science.1085658

Bogomolovas, J., Simon, B., Sattler, M., and Stier, G. (2009). Screening of fusion partners for high yield expression and purification of bioactive viscotoxins. *Protein Expr. Purif.* 64 (1), 16–23. doi:10.1016/j.pep.2008.10.003

Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., et al. (2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* 583 (7816), 469–472. doi:10.1038/s41586-020-2332-7

Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E. J., Canard, B., et al. (2010). *In vitro* reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog.* 6 (4), e1000863. doi:10.1371/journal.ppat.1000863

Cantini, F., Banci, L., Altincekic, N., Bains, J. K., Dhamotharan, K., Fuks, C., et al. (2020). (1)H, (13)C, and (15)N backbone chemical shift assignments of the apo and the ADP-ribose bound forms of the macrodomain of SARS-CoV-2 non-structural protein 3b. *Biomol. NMR Assign.* 14 (2), 339–346. doi:10.1007/s12104-020-09973-4

Chan, J. F., Kok, K. H., Zhu, Z., Chu, H., To, K. K., Yuan, S., et al. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* 9 (1), 221–236. doi:10.1080/22221751.2020.1719902

Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., and Huang, T. H. (2014). The SARS coronavirus nucleocapsid protein--forms and functions. *Antivir. Res.* 103, 39–50. doi:10.1016/j.antiviral.2013.12.009

Chang, H. P., Chou, C. Y., and Chang, G. G. (2007). Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride. *Biophys. J.* 92 (4), 1374–1383. doi:10.1529/biophysj.106.091736

Chen, H., Cui, Y., Han, X., Hu, W., Sun, M., Zhang, Y., et al. (2020a). Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res.* 30, 1143. doi:10.1038/s41422-020-00408-2

Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P. M. M., Olinares, P. D. B., et al. (2020b). Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* 182 (6), 1560–1573. doi:10.1016/j.cell.2020.07.033

Chen, P., Jiang, M., Hu, T., Liu, Q., Chen, X. S., and Guo, D. (2007). Biochemical characterization of exoribonuclease encoded by SARS coronavirus. *J. Biochem. Mol. Biol.* 40 (5), 649–655. doi:10.5483/bmbrep.2007.40.5.649

Chen, Y., Savinov, S. N., Mielech, A. M., Cao, T., Baker, S. C., and Mesecar, A. D. (2015). X-ray structural and functional studies of the three tandemly linked domains of non-structural protein 3 (nsp3) from murine hepatitis virus reveal conserved functions. *J. Biol. Chem.* 290 (42), 25293–25306. doi:10.1074/jbc.M115.662130

Cornillez-Ty, C. T., Liao, L., Yates, J. R., 3rd, Kuhn, P., and Buchmeier, M. J. (2009). Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *J. Virol.* 83 (19), 10314–10318. doi:10.1128/JVI.00842-09

Cross, T. J., Takahashi, G. R., Diessner, E. M., Crosby, M. G., Farahmand, V., Zhuang, S., et al. (2020). Sequence characterization and molecular modeling of clinically relevant variants of the SARS-CoV-2 main protease. *Biochemistry* 59 (39), 3741–3756. doi:10.1021/acs.biochem.0c00462

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. Engl.* 57 (17), 4787–4791. doi:10.1002/anie.201712091

Davies, J. P., Almasy, K. M., McDonald, E. F., and Plate, L. (2020). Comparative multiplexed interactomics of SARS-CoV-2 and homologous coronavirus non-structural proteins identifies unique and shared host-cell dependencies. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.07.13.201517

Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., et al. (2017). Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc. Natl. Acad. Sci. U.S.A.* 114 (21), E4251–E4260. doi:10.1073/pnas.1618310114

Dinesh, D. C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., et al. (2020). Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.* 16 (12), e1009100. doi:10.1371/journal.ppat.1009100

Dudas, F. D., Puglisi, R., Korn, S. M., Alfano, C., Kelly, G., Monaca, E., et al. (2021). Backbone chemical shift spectral assignments of coronavirus-2 non-structural protein nsp9. *Biomol. NMR Assign.* 2021, 1–10. doi:10.1007/s12104-020-09992-1

Egloff, M. P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., et al. (2004). The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc. Natl. Acad. Sci. U.S.A.* 101 (11), 3792–3796. doi:10.1073/pnas.0307877101

Esposito, D., Mehalko, J., Drew, M., Snead, K., Wall, V., Taylor, T., et al. (2020). Optimizing high-yield production of SARS-CoV-2 soluble spike trimers for serology assays. *Protein Expr. Purif.* 174, 105686. doi:10.1016/j.pep.2020.105686

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., et al. (2020). The coding capacity of SARS-CoV-2. *Nature* 589, 125. doi:10.1038/s41586-020-2739-1

Flower, T. G., Buffalo, C. Z., Hooy, R. M., Allaire, M., Ren, X., and Hurley, J. H. (2020). Structure of SARS-CoV-2 ORF8, a rapidly evolving coronavirus protein implicated in immune evasion. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.08.27.270637

Fogeron, M. L., Badillo, A., Jirasko, V., Gouttenoire, J., Paul, D., Lancien, L., 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 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

Fogeron, M. L., Jirasko, V., Penzel, S., Paul, D., Montserret, R., Danis, C., et al. (2016). Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus. *J. Biomol. NMR* 65 (2), 87–98. doi:10.1007/s10858-016-0040-2

Fogeron, M. L., Paul, D., Jirasko, V., Montserret, R., Lacabanne, D., Molle, J., et al. (2015b). Functional expression, purification, characterization, and membrane reconstitution of non-structural protein 2 from hepatitis C virus. *Protein Expr. Purif.* 116, 1–6. doi:10.1016/j.pep.2015.08.027

Frick, D. N., Virdi, R. S., Vuksanovic, N., Dahal, N., and Silvaggi, N. R. (2020). Molecular basis for ADP-ribose binding to the Mac1 domain of SARS-CoV-2 nsp3. *Biochemistry* 59 (28), 2608–2615. doi:10.1021/acs.biochem.0c00309

Gallo, A., Tsika, A. C., Fourkiotis, N. K., Cantini, F., Banci, L., Sreeramulu, S., et al. (2020). 1H, 13C and 15N chemical shift assignments of the SUD domains of SARS-CoV-2 non-structural protein 3c: "the N-terminal domain-SUD-N". *Biomol. NMR Assign.* 2020, 1–5. doi:10.1007/s12104-020-09987-y

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., et al. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 368 (6492), 779–782. doi:10.1126/science.abb7498

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583 (7816), 459–468. doi:10.1038/s41586-020-2286-9

Graham, R. L., Sims, A. C., Brockway, S. M., Baric, R. S., and Denison, M. R. (2005). The nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral replication. *J. Virol.* 79 (21), 13399–13411. doi:10.1128/JVI.79.21.13399-13411.2005

Hagemeijer, M. C., Monastyrska, I., Griffith, J., van der Sluijs, P., Voortman, J., van Bergen en Henegouwen, P. M., et al. (2014). Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology* 458–459, 125–135. doi:10.1016/j.virol.2014.04.027

Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. *Nature* 584 (7819), 154–156. doi:10.1038/s41586-020-2368-8

Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P., et al. (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* 280 (35), 31257–31266. doi:10.1074/jbc.M502577200

Huang, Y., Yang, Z. Y., Kong, W. P., and Nabel, G. J. (2004). Generation of synthetic severe acute respiratory syndrome coronavirus pseudoparticles: implications for assembly and vaccine production. *J. Virol.* 78 (22), 12557–12565. doi:10.1128/JVI.78.22.12557-12565.2004

Hurst, K. R., Koetzner, C. A., and Masters, P. S. (2013). Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *J. Virol.* 87 (16), 9159–9172. doi:10.1128/JVI.01275-13

Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464. doi:10.1093/nar/gkm363

Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., et al. (2019). Delicate structural coordination of the severe acute respiratory syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res.* 47 (12), 6538–6550. doi:10.1093/nar/gkz409

Jiang, H. W., Li, Y., Zhang, H. N., Wang, W., Yang, X., Qi, H., et al. (2020). SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. *Nat. Commun.* 11 (1), 3581. doi:10.1038/s41467-020-17488-8

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582 (7811), 289–293. doi:10.1038/s41586-020-2223-y

Jirasko, V., Lakomek, N. A., Penzel, S., Fogeron, M. L., Bartenschlager, R., Meier, B. H., et al. (2020a). Proton-detected solid-state NMR of the cell-free synthesized α-helical transmembrane protein NS4B from hepatitis C virus. *Chembiochem.* 21 (10), 1453–1460. doi:10.1002/cbic.201900765

Jirasko, V., Lends, A., Lakomek, N. A., Fogeron, M. L., Weber, M. E., Malär, A. A., et al. (2020b). Dimer organization of membrane-associated NS5A of hepatitis C virus as determined by highly sensitive 1 H-detected solid-state NMR. *Angew. Chem. Int. Ed.* 60 (10), 5339–5347. doi:10.1002/anie.202013296

Johnson, M. A., Chatterjee, A., Neuman, B. W., and Wüthrich, K. (2010). SARS coronavirus unique domain: three-domain molecular architecture in solution and RNA binding. *J. Mol. Biol.* 400 (4), 724–742. doi:10.1016/j.jmb.2010.05.027

Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Brooun, A., Griffith, M., et al. (2006). Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *J. Virol.* 80 (16), 7894–7901. doi:10.1128/JVI.00467-06

Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., et al. (2006). Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc. Natl. Acad. Sci. U.S.A.* 103 (34), 12885–12890. doi:10.1073/pnas.0603144103

Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., et al. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* 10 (7), 1228–1238. doi:10.1016/j.apsb.2020.04.009

Kern, D. M., Sorum, B., Mali, S. S., Hoel, C. M., Sridharan, S., Remis, J. P., et al. (2020). Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *bioRxiv* 17, 156554. doi:10.1101/2020.06.17.156554

Khan, M. T., Zeb, M. T., Ahsan, H., Ahmed, A., Ali, A., Akhtar, K., et al. (2020). SARS-CoV-2 nucleocapsid and Nsp3 binding: an in silico study. *Arch. Microbiol.* 203, 59. doi:10.1007/s00203-020-01998-6

Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., et al. (2020). Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci.* 29 (7), 1596–1605. doi:10.1002/pro.3873

Kirchdoerfer, R. N., and Ward, A. B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* 10 (1), 2342. doi:10.1038/s41467-019-10280-3

Konkolova, E., Klima, M., Nencka, R., and Boura, E. (2020). Structural analysis of the putative SARS-CoV-2 primase complex. *J. Struct. Biol.* 211 (2), 107548. doi:10.1016/j.jsb.2020.107548

Konno, Y., Kimura, I., Uriu, K., Fukushi, M., Irie, T., Koyanagi, Y., et al. (2020). SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is increased by a naturally occurring elongation variant. *Cell Rep.* 32 (12), 108185. doi:10.1016/j.celrep.2020.108185

Korn, S. M., Dhamotharan, K., Fürtig, B., Hengesbach, M., Löhr, F., Qureshi, N. S., et al. (2020a). 1H, 13C, and 15N backbone chemical shift assignments of the nucleic acid-binding domain of SARS-CoV-2 non-structural protein 3e. *Biomol. NMR Assign.* 14 (2), 329–333. doi:10.1007/s12104-020-09971-6

Korn, S. M., Lambertz, R., Fürtig, B., Hengesbach, M., Löhr, F., Richter, C., et al. (2020b). 1H, 13C, and 15N backbone chemical shift assignments of the C-terminal dimerization domain of SARS-CoV-2 nucleocapsid protein. *Biomol. NMR Assign.* 2020, 1–7. doi:10.1007/s12104-020-09995-y

Krafcikova, P., Silhan, J., Nencka, R., and Boura, E. (2020). Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat. Commun.* 11 (1), 3717. doi:10.1038/s41467-020-17495-9

Kubatova, N., Qureshi, N. S., Altincekic, N., Abele, R., Bains, J. K., Ceylan, B., et al. (2020). 1H, 13C, and 15N backbone chemical shift assignments of coronavirus-2 non-structural protein Nsp10. *Biomol. NMR Assign.* 2020, 1–7. doi:10.1007/s12104-020-09984-1

Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L., and Hilgenfeld, R. (2015). A G-quadruplex-binding macrodomain within the "SARS-unique domain" is essential for the activity of the SARS-coronavirus replication-transcription complex. *Virology* 484, 313–322. doi:10.1016/j.virol.2015.06.016

Leao, J. C., Gusmao, T. P. L., Zarzar, A. M., Leao Filho, J. C., Barkokebas Santos de Faria, A., Morais Silva, I. H., et al. (2020). Coronaviridae-old friends, new enemy! *Oral Dis.* 2020, 13447. doi:10.1111/odi.13447

Lei, J., Kusov, Y., and Hilgenfeld, R. (2018). Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antivir. Res.* 149, 58–74. doi:10.1016/j.antiviral.2017.11.001

Li, Q., and Kang, C. (2020). A practical perspective on the roles of solution NMR spectroscopy in drug discovery. *Molecules* 25 (13), 2974. doi:10.3390/molecules25132974

Littler, D. R., Gully, B. S., Colson, R. N., and Rossjohn, J. (2020). Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. *iScience* 23 (7), 101258. doi:10.1016/j.isci.2020.101258

Liu, D. X., Fung, T. S., Chong, K. K., Shukla, A., and Hilgenfeld, R. (2014). Accessory proteins of SARS-CoV and other coronaviruses. *Antivir. Res.* 109, 97–109. doi:10.1016/j.antiviral.2014.06.013

Luo, H., Chen, J., Chen, K., Shen, X., and Jiang, H. (2006). Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry* 45 (39), 11827–11835. doi:10.1021/bi0609319

Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., et al. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc. Natl. Acad. Sci. U.S.A.* 112 (30), 9436–9441. doi:10.1073/pnas.1508686112

Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., and Hong, M. (2020). Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat. Struct. Mol. Biol.* 27, 1202. doi:10.1038/s41594-020-00536-8

Miknis, Z. J., Donaldson, E. F., Umland, T. C., Rimmer, R. A., Baric, R. S., and Schultz, L. W. (2009). Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *J. Virol.* 83 (7), 3007–3018. doi:10.1128/JVI.01505-08

Mompean, M., Trevino, M. A., and Laurents, D. V. (2020). Towards targeting the disordered SARS-CoV-2 nsp2 C-terminal region: partial structure and dampened mobility revealed by NMR spectroscopy. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.11.09.374173

Morita, E. H., Sawasaki, T., Tanaka, R., Endo, Y., and Kohno, T. (2003). A wheat germ cell-free system is a novel way to screen protein folding and function. *Protein Sci.* 12 (6), 1216–1221. doi:10.1110/ps.0241203

Narayanan, K., Huang, C., Lokugamage, K., Kamitani, W., Ikegami, T., Tseng, C. T., et al. (2008). Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. *J. Virol.* 82 (9), 4471–4479. doi:10.1128/JVI.02472-07

Narayanan, K., Ramirez, S. I., Lokugamage, K. G., and Makino, S. (2015). Coronavirus nonstructural protein 1: common and distinct functions in the regulation of host and viral gene expression. *Virus. Res.* 202, 89–100. doi:10.1016/j.virusres.2014.11.019

Nelson, C. W., Ardern, Z., Goldberg, T. L., Meng, C., Kuo, C. H., Ludwig, C., et al. (2020). Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic. *Elife* 9, 59633. doi:10.7554/eLife.59633

Netzer, W. J., and Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388 (6640), 343–349. doi:10.1038/41024

Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A., et al. (2008). Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J. Virol.* 82 (11), 5279–5294. doi:10.1128/JVI.02631-07

Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., et al. (2011). A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* 174 (1), 11–22. doi:10.1016/j.jsb.2010.11.021

Neuman, B. W. (2016). Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles. *Antivir. Res.* 135, 97–107. doi:10.1016/j.antiviral.2016.10.005

Oostra, M., Hagemeijer, M. C., van Gent, M., Bekker, C. P., te Lintelo, E. G., Rottier, P. J., et al. (2008). Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J. Virol.* 82 (24), 12392–12405. doi:10.1128/JVI.01219-08

Oostra, M., te Lintelo, E. G., Deijs, M., Verheije, M. H., Rottier, P. J., and de Haan, C. A. (2007). Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. *J. Virol.* 81 (22), 12323–12336. doi:10.1128/JVI.01506-07

Pavesi, A. (2020). New insights into the evolutionary features of viral overlapping genes by discriminant analysis. *Virology* 546, 51–66. doi:10.1016/j.virol.2020.03.007

Robertson, M. P., Igel, H., Baertsch, R., Haussler, D., Ares, M., Jr., and Scott, W. G. (2005). The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol.* 3 (1), e5. doi:10.1371/journal.pbio.0030005

Robson, F., Khan, K. S., Le, T. K., Paris, C., Demirbag, S., Barfuss, P., et al. (2020). Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol. Cel* 79 (5), 710–727. doi:10.1016/j.molcel.2020.07.027

Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N. L., Kiryukhina, O., Wiersum, G., et al. (2020). The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.04.17.047498

Salvi, N., Bessa, L. M., Guseva, S., Camacho-Zarco, A., Maurin, D., Perez, L. M., et al. (2021). 1H, 13C and 15N backbone chemical shift assignments of SARS-CoV-2 nsp3a. *Biomol. NMR Assign.* 2021, 1–4. doi:10.1007/s12104-020-10001-8

Schoeman, D., and Fielding, B. C. (2020). Is there a link between the pathogenic human coronavirus envelope protein and immunopathology? A review of the literature. *Front. Microbiol.* 11, 2086. doi:10.3389/fmicb.2020.02086

Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L. A., et al. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat. Struct. Mol. Biol.* 27 (10), 959–966. doi:10.1038/s41594-020-0511-8

Serrano, P., Johnson, M. A., Almeida, M. S., Horst, R., Herrmann, T., Joseph, J. S., et al. (2007). Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. *J. Virol.* 81 (21), 12049–12060. doi:10.1128/JVI.00969-07

Serrano, P., Johnson, M. A., Chatterjee, A., Neuman, B. W., Joseph, J. S., Buchmeier, M. J., et al. (2009). Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. *J. Virol.* 83 (24), 12998–13008. doi:10.1128/JVI.01253-09

Shen, Z., Wang, G., Yang, Y., Shi, J., Fang, L., Li, F., et al. (2019). A conserved region of nonstructural protein 1 from alphacoronaviruses inhibits host gene expression and is critical for viral virulence. *J. Biol. Chem.* 294 (37), 13606–13618. doi:10.1074/jbc.RA119.009713

Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., et al. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* 587 (7835), 657–662. doi:10.1038/s41586-020-2601-5

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., et al. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331 (5), 991–1004. doi:10.1016/s0022-2836(03)00865-9

Surya, W., Li, Y., and Torres, J. (2018). Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim. Biophys. Acta Biomembr.* 1860 (6), 1309–1317. doi:10.1016/j.bbamem.2018.02.017

Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., et al. (2004). The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure* 12 (2), 341–353. doi:10.1016/j.str.2004.01.016

Takai, K., Sawasaki, T., and Endo, Y. (2010). Practical cell-free protein synthesis system using purified wheat embryos. *Nat. Protoc.* 5 (2), 227–238. doi:10.1038/nprot.2009.207

Takeda, M., Chang, C. K., Ikeya, T., Güntert, P., Chang, Y. H., Hsu, Y. L., et al. (2008). Solution structure of the c-terminal dimerization domain of SARS

coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J. Mol. Biol.* 380 (4), 608–622. doi:10.1016/j.jmb.2007.11.093

Tan, J., Vonrhein, C., Smart, O. S., Bricogne, G., Bollati, M., Kusov, Y., et al. (2009). The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *Plos Pathog.* 5 (5), e1000428. doi:10.1371/journal.ppat.1000428

Tan, Y., Schneider, T., Leong, M., Aravind, L., and Zhang, D. (2020). Novel immunoglobulin domain proteins provide insights into evolution and pathogenesis of SARS-CoV-2-related viruses. *mBio* 11 (3). doi:10.1128/mBio.00760-20

Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., et al. (2020). Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* 369 (6508), 1249–1255. doi:10.1126/science.abc8665

Tonelli, M., Singarapu, K. K., Makino, S., Sahu, S. C., Matsubara, Y., Endo, Y., et al. (2011). Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J. Biomol. NMR* 51 (4), 467–476. doi:10.1007/s10858-011-9575-4

Tonelli, M., Rienstra, C., Anderson, T. K., Kirchdoerfer, R., and Henzler-Wildman, K. (2020). 1H, 13C, and 15N backbone and side chain chemical shift assignments of the SARS-CoV-2 non-structural protein 7. *Biomol. NMR Assign.* 2020, 1–5. doi:10.1007/s12104-020-09985-0

Tvarogová, J., Madhugiri, R., Bylapudi, G., Ferguson, L. J., Karl, N., and Ziebuhr, J. (2019). Identification and characterization of a human coronavirus 229E nonstructural protein 8-associated RNA 3′-terminal adenylyltransferase activity. *J. Virol.* 93 (12), e00291–e00319. doi:10.1128/JVI.00291-19

Ullrich, S., and Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* 30 (17), 127377. doi:10.1016/j.bmcl.2020.127377

Vennema, H., Godeke, G. J., Rossen, J. W., Voorhout, W. F., Horzinek, M. C., Opstelten, D. J., et al. (1996). Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* 15 (8), 2020–2028. doi:10.1002/j.1460-2075.1996.tb00553.x

Wacker, A., Weigand, J. E., Akabayov, S. R., Altincekic, N., Bains, J. K., Banijamali, E., et al. (2020). Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* 48, 12415. doi:10.1093/nar/gkaa1013

Wang, S., Fogeron, M. L., Schledorn, M., Dujardin, M., Penzel, S., Burdette, D., et al. (2019). Combining cell-free protein synthesis and NMR into a tool to study capsid assembly modulation. *Front. Mol. Biosci.* 6, 67. doi:10.3389/fmolb.2019.00067

Wolff, G., Limpens, R. W. A. L., Zevenhoven-Dobbe, J. C., Laugks, U., Zheng, S., de Jong, A. W. M., et al. (2020). A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* 369 (6509), 1395–1398. doi:10.1126/science.abd3629

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579 (7798), 265–269. doi:10.1038/s41586-020-2008-3

Wu, Z., Yang, L., Ren, X., Zhang, J., Yang, F., Zhang, S., et al. (2016). ORF8-related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J. Infect. Dis.* 213 (4), 579–583. doi:10.1093/infdis/jiv476

Ye, Q., West, A. M. V., Silletti, S., and Corbett, K. D. (2020). Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* 29, 1890. doi:10.1002/pro.3909

Yin, W., Mao, C., Luan, X., Shen, D. D., Shen, Q., Su, H., et al. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 368 (6498), 1499–1504. doi:10.1126/science.abc1560

Yoshimoto, F. K. (2020). The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J.* 39 (3), 198–216. doi:10.1007/s10930-020-09901-4

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* 368 (6489), 409–412. doi:10.1126/science.abb3405

Zhao, J., Falcón, A., Zhou, H., Netland, J., Enjuanes, L., Pérez Breña, P., et al. (2009). Severe acute respiratory syndrome coronavirus protein 6 is required for optimal replication. *J. Virol.* 83 (5), 2368–2373. doi:10.1128/JVI.02371-08

Zhou, R., Zeng, R., Von Brunn, A., and Lei, J. (2020). Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. *Mol. Biomed.* 1 (2), 1–11. doi:10.1186/s43556-020-00001-4

Züst, R., Cervantes-Barragán, L., Kuri, T., Blakqori, G., Weber, F., Ludewig, B., et al. (2007). Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines. *PLoS Pathog.* 3 (8), e109. doi:10.1371/journal.ppat.0030109

Conflict of Interest: CH was employed by Signals GmbH & Co. KG. Q19

The remaining 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.

GLOSSARY

aa Amino acid

BEST Band-selective excitation short-transient

BMRB Biomagnetic resonance databank

CFPS Cell-free protein synthesis

CoV Coronavirus

CTD C-terminal domain

DEDD Asp-Glu-Glu-Asp

DMS Dimethylsulfate

E Envelope protein

ED Ectodomain

fl Full-length

GB1 Protein G B1 domain

GD Globular domain

GF Gel filtration

GST Glutathione-S-transferase

His Hisx-tag

HSP Heat-shock protein

HSQC Heteronuclear single quantum coherence

IDP Intrinsically disordered protein

IDR Intrinsically disordered region

IEC Ion exchange chromatography

IMAC Immobilized metal ion affinity chromatography

IPRS IMAC-protease cleavage-reverse IMAC-SEC;

M Membrane protein

MERS Middle East Respiratory Syndrome

MHV Murine hepatitis virus

Mpro Main protease

MTase Methyltransferase

N Nucleocapsid protein

NAB Nucleic acid–binding domain

nsp Nonstructural protein

NTD N-terminal domain

PL^{pro} Papain-like protease

RdRP RNA-dependent RNA polymerase

S Spike protein

SARS Severe Acute Respiratory Syndrome

SEC Size-exclusion chromatography

SUD SARS unique domain

SUMO Small ubiquitin-related modifier

TEV Tobacco etch virus

TM Transmembrane

TROSY Transverse relaxation-optimized spectroscopy

Trx Thioredoxin

Ubl Ubiquitin-like domain

Ulp1 Ubiquitin-like specific protease 1

WG Wheat-germ.

C

Annexe 3 : ^{1}H, ^{13}C and ^{15}N Backbone chemical shift assignments of the N-terminal and central intrinsically disordered domains of SARS-CoV-2 nucleoprotein

Biomolecular NMR Assignments
https://doi.org/10.1007/s12104-021-10014-x

ARTICLE

^{1}H, ^{13}C and ^{15}N Backbone chemical shift assignments of the n-terminal and central intrinsically disordered domains of SARS-CoV-2 nucleoprotein

Serafima Guseva[1] · Laura Mariño Perez[1] · Aldo Camacho-Zarco[1] · Luiza Mamigonian Bessa[1] · Nicola Salvi[1] · Anas Malki[1] · Damien Maurin[1] · Martin Blackledge[1]

Received: 11 November 2020 / Accepted: 7 March 2021

Abstract
The nucleoprotein (N) from SARS-CoV-2 is an essential cofactor of the viral replication transcription complex and as such represents an important target for viral inhibition. It has also been shown to colocalize to the transcriptase-replicase complex, where many copies of N decorate the viral genome, thereby protecting it from the host immune system. N has also been shown to phase separate upon interaction with viral RNA. N is a 419 amino acid multidomain protein, comprising two folded, RNA-binding and dimerization domains spanning residues 45–175 and 264–365 respectively. The remaining 164 amino acids are predicted to be intrinsically disordered, but there is currently no atomic resolution information describing their behaviour. Here we assign the backbone resonances of the first two intrinsically disordered domains (N1, spanning residues 1–44 and N3, spanning residues 176–263). Our assignment provides the basis for the identification of inhibitors and functional and interaction studies of this essential protein.

Keywords SARS-CoV-2 · Nucleoprotein · Intrinsically disordered protein · NMR · COVID-19 · Dynamics · Relaxation

Biological context

The spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) represents a serious threat to the stability of human societies throughout the world. In order to eradicate the disease, it is necessary to foster research aimed at the development of vaccines or effective antiviral inhibitors. The establishment of effective therapeutic strategies requires the identification of molecular targets against which inhibitory strategies can be developed.

The nucleoprotein (N) of SARS-CoV-2 is a crucial component of the replication machinery, localizing to the viral replicase–transcriptase complex (4, 5, 7–9). Beta-coronaviral N encapsidates the viral genome within a nucleocapsid, formed from multiple copies of the protein (Zúñiga et al. 2007; McBride et al. 2014), thereby providing protection from the host immune system, and is essential for regulation of viral gene transcription (Wu et al. 2014). N is also produced in abundance in infected cells and is an important marker for infection. In order to identify possible avenues targeting the essential roles of N in the viral cycle, it is important to understand the structure and function of this complex protein at atomic resolution.

N is a 419 amino acid multi-domain protein, comprising folded, RNA-binding (N2) and dimerization (N4) domains spanning residues 45–175 and 266–365 respectively (Chang et al. 2014). N2 has been described using both X-ray crystallography (Kang et al. 2020; Peng et al. 2020) and NMR spectroscopy (Dinesh et al. 2020). The dimerization domain N4 has been described using X-ray crystallography (Ye et al. 2020; Peng et al. 2020) and N has been shown to form higher order oligomers via N4 (Ye et al. 2020; Zeng et al. 2020; Lutomski et al. 2020). The remaining 164 amino acids comprising N1, N3 and N5 are predicted to be intrinsically disordered, but despite evidence that these domains are essential for function in the related Mouse Hepatitis virus (Keane and Giedroc 2013), there is currently no atomic resolution information describing their behaviour in solution. Nuclear

Serafima Guseva and Laura Mariño Perez have contributed equally to this work.

✉ Martin Blackledge
martin.blackledge@ibs.fr

[1] University Grenoble Alpes, CNRS, CEA, IBS, 38000 Grenoble, France

Published online: 17 March 2021

magnetic resonance is the ideal tool for studying the conformational behaviour of highly dynamic or intrinsically disordered proteins (Jensen et al. 2014).

Here we present the assignment of the backbone resonances of the first two intrinsically disordered domains (N1, spanning residues 1–44 and N3, the central disordered domain, spanning residues 176–265). The N3 IDR (175–263) comprises a serine-arginine (SR) rich domain that is phosphorylated in virions, a modification that plays a role in both function and localization of SARS-CoV-1 N (Peng et al. 2008). N5 is thought to contribute to the oligomerization of N in SARS-CoV-1 (Lo et al. 2013).

Recent studies used single molecule FRET and molecular modelling (Cubuk et al. 2020) to describe the global conformational behaviour of N1 and N3. Peptides representing the SR region of N3 in its phosphorylated and non-phosphorylated forms were also described using NMR spectroscopy (Savastano et al. 2020). Mass spectrometry has also revealed a number of auto-catalytic sites in N, two of which are present in N3 (Lu et al. 2020; Lutomski et al. 2020).

N from SARS-CoV-2 has been shown to undergo liquid–liquid phase separation (LLPS) upon mixing with RNA (Cubuk et al. 2020; Savastano et al. 2020; Chen et al. 2020; Iserman et al. 2020; Perdikari et al. 2020; Carlson et al. 2020; Lu et al. 2020; Jack et al. 2020). The formation of membraneless condensates by colocalization of components involved in viral replication in liquid droplets is prevalent in negative sense single strand RNA viruses, such as rabies (Nikolic et al. 2017), measles (Zhou et al. 2019; Guseva et al. 2020), and Nipah (Ringel et al. 2019). It is not yet known which interactions are responsible for LLPS in SARS-CoV-2, but it is highly likely that the intrinsically disordered domains are involved in this process. Moreover, N2 and N3 were shown to be essential for phase separation with RNA (Lu et al. 2020).

Here, we present backbone resonance assignment and NMR relaxation of two domains (N1 and N3) of N, providing data prerequisite for characterisation of the functional modes of this enigmatic protein, screening of interaction partners and detailed mapping of interactions.

Methods and experiments

Sample preparation

The primary sequence of N from SARS-CoV-2 was extracted from NCBI genome entry NC_045512.2 [GenBank entry MN908947.3 (Wu et al. 2014)]. Commercially synthesized genes (Genscript Biotech) were codon-optimized for expression in *Escherichia coli* and subcloned in a pET21b(+) vector. Hexa-histidine and TEV-cleavage tags were included at the N-terminus to facilitate protein purification. After protease cleavage, the proteins contain N-terminal GRR- extensions. For ^{15}N and ^{13}C isotope labelling, cells were grown in M9-minimal medium supplemented with ^{15}N–NH_4–Cl and $^{13}C_6$–D–glucose (1 and 2 g/L respectively).

Both nucleoprotein constructs (N123, comprising residues 1–263 and N3, comprising residues 175–263) were cloned into a pESPRIT vector between the AatII and NotI cleavage sites with His_8-tag and TEV cleavage sites at the N-terminus (GenScript Biotech Netherlands). Transformation was performed by heat-shock and proteins were expressed in *E. coli* BL21(DE3) (Novagen) for 5 h at 37°C after induction at an optical density of 0.6 with 1 mM isopropyl–b–D–thiogalactopyranoside. Cells were harvested by centrifugating at 5000 rpm, resuspended in

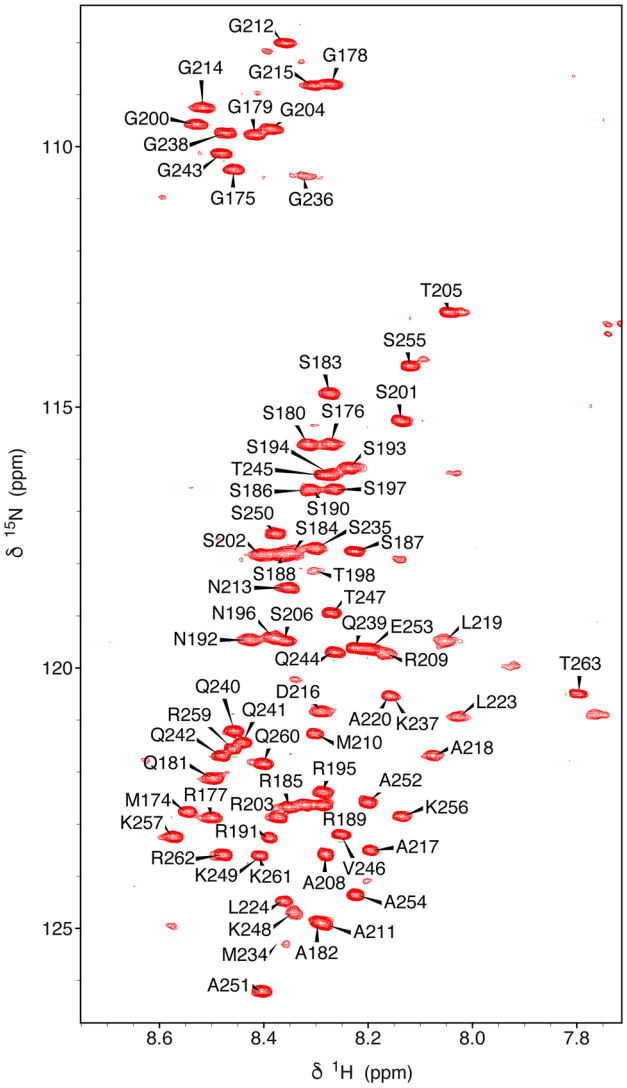


Fig. 1 ^{1}H–^{15}N HSQC of N3 (175–263) from SARS-CoV-2. The spectrum was acquired on an 850 MHz spectrometer at 298 K at a concentration of 150 µM in 50 mM Na-phosphate, pH 6.0, 250 mM NaCl. Assigned backbone ^{1}H–^{15}N peaks are labelled

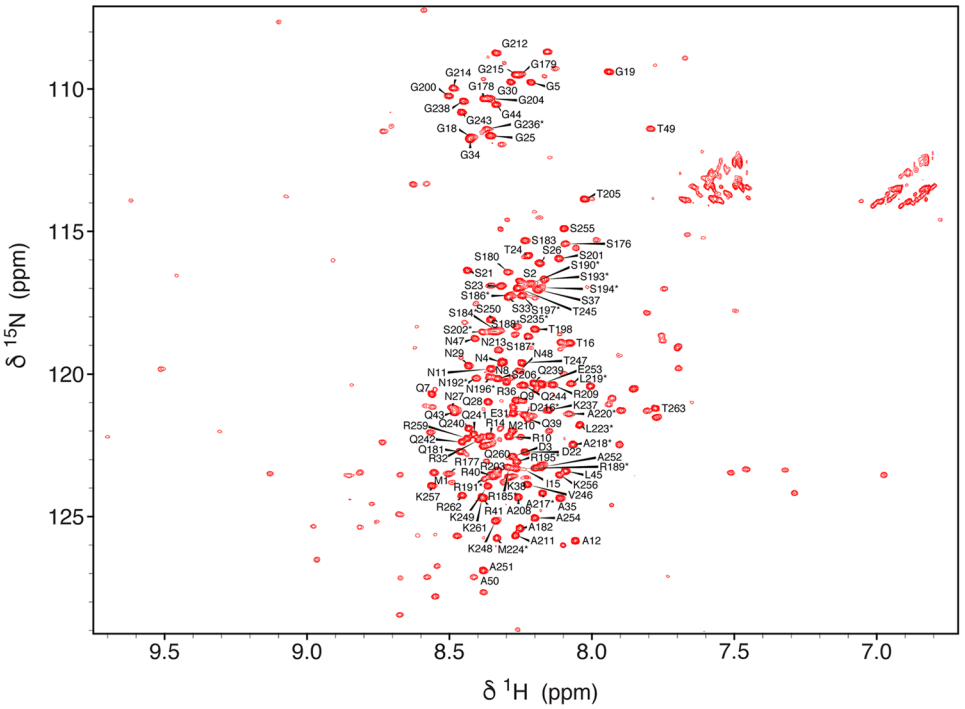


Fig. 2 ^{1}H–^{15}N BEST-TROSY of N123 (1–263) from SARS-CoV-2. The spectrum was acquired on a 950 MHz spectrometer at 298 K at a concentration of 150 µM in 50 mM Na-phosphate, pH 6.5, 250 mM NaCl. Backbone ^{1}H–^{15}N peaks assigned in this study are labelled, resonances transferred from assignment of N3 (Fig. 1) are starred. The disperse signals from N2 for which an assignment is already available (Dinesh et al. 2020) are present but have not been labelled

20 mM Tris (pH 8.0) and 500 mM NaCl buffer, lysed by sonication, and centrifugated again at 18,000 rpm at 4°C. The supernatant was subjected to standard Ni purification. Proteins were eluted with 20 mM Tris (pH 8), 500 mM NaCl and 500 mM imidazole. Samples were then dialysed against 20 mM Tris (pH 8), 500 mM NaCl, 2 mM DTT at room temperature overnight. Following TEV cleavage, samples were concentrated and subjected to size exclusion chromatography (SEC, Superdex 75/200) in 50 mM Na-Phosphate (pH 6.5), 250 mM NaCl 2 mM DTT buffer (NMR buffer). Proteins were studied at 600 and 150 µM for N3 and 0.91 mM for N123.

NMR experiments

BEST and BEST-TROSY (BT) double and triple resonance assignment experiments, including BEST-HNCA, BEST HN(CO)CA, BT-HNCO (Lescop et al. 2007; Solyom et al. 2013), were recorded on ^{15}N,^{13}C-labeled samples at 298 K using Bruker Avance III spectrometers equipped with a cryoprobe at ^{1}H frequencies of 600 and 850 MHz. R1rho relaxation experiments (Lakomek et al. 2012) were recorded at 150 µM protein concentration and 298 K in a 50 mM Na-Phosphate (pH 6.5), 250 mM NaCl 2 mM DTT buffer at 950 MHz.

Residues 217–224 and 231–235 of N3 were assigned using a deuterated ^{13}C ^{15}N labelled sample at 300 µM. BEST-HNCA, BEST-HN(CO)CA with one increment in ^{15}N dimension were recorded at 850 MHz, allowing detection of weak additional peaks of the helical region and connection with those already assigned. All spectra were processed using NMRPipe (Delaglio et al. 1995) and analyzed using CCPNMR Analysis Assign (Skinner et al. 2016) and NMR-FAM-SPARKY (Lee et al. 2015). Assignment was further assisted by comparison to BMRB entry 34511 (Dinesh et al. 2020) of the N-terminal RNA-binding domain of the SARS-CoV-2 nucleoprotein, specifically nucleoprotein residues 44 to 180 (N2).

Assignments and data deposition

The HSQC of N3 is typical of an intrinsically disordered protein (Fig. 1). Most peaks in N3 are reproduced in the spectrum of N123, that also reveals the presence of the folded RNA-binding domain N2, in addition to disordered peaks from N1 (intensity of peaks from the first 15 residues of N3 were reduced in N123, probably due to the proximity with N2). The assignment of N2 has been published (Dinesh et al. 2020), and assignment of N1 in N123 (Fig. 2) and N3 in its isolated form and in the context of N123 were accomplished using standard triple resonance approaches. A high percentage of resonances could be assigned in N3 (84% $^{1}H^{N}$, 84% $^{15}N^{H}$, 82% $^{13}C\alpha$ and 76% ^{13}C) and N1 in the context of N123 (94% $^{1}H^{N}$, 94% $^{15}N^{H}$, 100% $^{13}C\alpha$, 57% $^{13}C\beta$ and 100% ^{13}C). These assignments have been deposited in the Biological Magnetic Resonance Databank (BMRB ID: 50557, comprising backbone resonance assignment of N3 and 50558, comprising backbone resonance assignment of

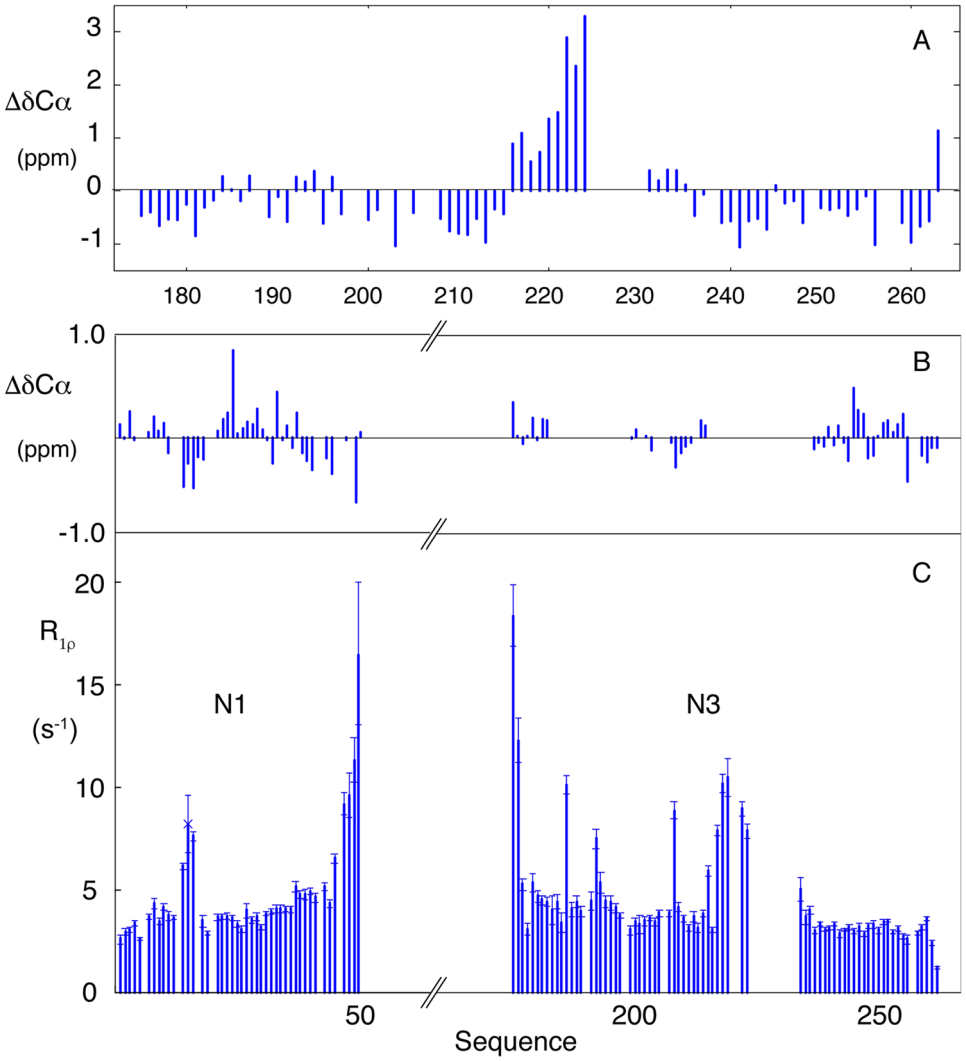


Fig. 3 Analysis of chemical shifts and relaxation rates for N1 and N3 from SARS-CoV-2. **a** Secondary C^{α} chemical shifts for the disordered residues in N3. **b** Secondary C^{α} chemical shifts for the disordered residues in N123 that are observed in triple resonance assignment experiments. **c** Rotating frame relaxation rates measured in N123. Measurements were made at 950 MHz at 298 K using sample concentration of 150 µM. Residues from N2 are not shown

N1 and of resolved resonances in N3, both in the context of N123).

Secondary structural propensity and dynamics

Resonance assignment of the two domains confirms the disordered nature of N1 and N3 (Figs. 1 and 2). The linker region (N3) connecting the two folded domains (N2 and N4), comprises intrinsically disordered SR-rich and polar termini, flanking a central hydrophobic strand that exhibits a pronounced helical propensity (> 30% from 216 to 224 with near 100% helical population from position 220) (Fig. 3). Assignment of 5 residues at the centre of this region was not possible, possibly due to dimerization mediated by the helix. This region is also predicted to form a hydrophobic helix in SARS-CoV (Chang et al. 2009). Recent simulations (Cubuk et al. 2020) also proposed the presence of a weakly (< 20%) helical motif in the SR region at the N-terminal region of N (176–185), which is not seen experimentally, although there is overlap of predicted helical propensity (< 30% helix predicted to stretch from 213 to 225) in the vicinity of experimentally observed helical propensity. Studies of isolated peptides (Savastano et al. 2020) also suggested a helical propensity in the SR region which is not seen here.

Assignment of the backbone resonances of the N-terminal domain of N (N1) within the N123 construct (Fig. 2) reveals the presence of an intrinsically disordered chain with no detectable secondary structural propensity and only very slight differences around residue 248 compared to free N3. Note that a number of resonances that were visible in N3 were not assigned in N123, most probably due to short relaxation times experienced in the vicinity of the helical motif that limits transfer of magnetization in triple resonance experiments. Nevertheless putative transfer of assignment was proposed on the basis of the ^{15}N–^{1}H correlation spectra of N3 and N123 (starred resonances in Fig. 2). Spin

relaxation measured in N123 confirms the highly dynamic nature of N1 and N3 in the context of the folded RNA binding domain N2 (Fig. 3), with relaxation rates in the expected range for a disordered domain (Adamski et al. 2019), with the exception of the helical element in N3 and three consecutive residues in N1 (R14-I15-T16) that both show elevated rates.

In conclusion, NMR backbone assignment and preliminary relaxation studies provide the basis for further NMR studies of this important drug target, providing the tools necessary for the identification of inhibitors and for detailed functional studies of this essential protein.

Acknowledgements This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-05-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG CEA). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG CEA). The authors acknowledge discussions and encouragement from the COVID-19 NMR consortium under the coordination of Prof. Harald Schwalbe and colleagues. Financial support from the TGIR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged.

References

Adamski W, Salvi N, Maurin D et al (2019) A unified description of intrinsically disordered protein dynamics under physiological conditions using NMR spectroscopy. J Am Chem Soc 141:17817–17829. https://doi.org/10.1021/jacs.9b09002

Carlson CR, Asfaha JB, Ghent CM et al (2020) Phosphorylation modulates liquid-liquid phase separation of the SARS-CoV-2 N protein. Bio Rxiv. https://doi.org/10.1101/2020.06.28.176248

Chang C-K, Hsu Y-L, Chang Y-H et al (2009) Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. J Virol 83:2255–2264. https://doi.org/10.1128/JVI.02001-08

Chang C, Hou M-H, Chang C-F et al (2014) The SARS coronavirus nucleocapsid protein—forms and functions. Antivir Res 103:39–50. https://doi.org/10.1016/j.antiviral.2013.12.009

Chen H, Cui Y, Han X et al (2020) Liquid–liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. Cell Res 30:1143–1145. https://doi.org/10.1038/s41422-020-00408-2

Cubuk J, Alston JJ, Incicco JJ et al (2020) The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. Bio Rxiv. https://doi.org/10.1101/2020.06.17.158121

Delaglio FD, Grzesiek S, Vuister G et al (1995) NMRPIPE–A multidimensional spectral processing system based on unix pipes. J biomol nmr 6:277–293

Dinesh DC, Chalupska D, Silhan J et al (2020) Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. Bio Rxiv. https://doi.org/10.1101/2020.04.02.022194

Guseva S, Milles S, Jensen MR et al (2020) Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. Sci Adv. https://doi.org/10.1126/sciadv.aaz7095

Iserman C, Roden C, Boerneke M et al (2020) Specific viral RNA drives the SARS CoV-2 nucleocapsid to phase separate. Bio Rxiv. https://doi.org/10.1101/2020.06.11.147199

Jack A, Ferro LS, Trnka MJ et al (2020) SARS CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. Bio Rxiv. https://doi.org/10.1101/2020.09.14.295824

Jensen MR, Zweckstetter M, Huang J, Blackledge M (2014) Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. Chem Rev 114:6632–6660. https://doi.org/10.1021/cr400688u

Kang S, Yang M, Hong Z et al (2020) Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. Acta Pharm Sin B. https://doi.org/10.1016/j.apsb.2020.04.009

Keane SC, Giedroc DP (2013) Solution structure of mouse Hepatitis Virus (MHV) nsp3a and determinants of the interaction with MHV nucleocapsid (N) protein. J Virol 87:3502–3515

Lakomek N-A, Ying J, Bax A (2012) Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. J Biomol NMR 53:209–221. https://doi.org/10.1007/s10858-012-9626-5

Lee W, Tonelli M, Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 31:1325–1327. https://doi.org/10.1093/bioinformatics/btu830

Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. J Magn Reson 187:163–169. https://doi.org/10.1016/j.jmr.2007.04.002

Lo Y-S, Lin S-Y, Wang S-M et al (2013) Oligomerization of the carboxyl terminal domain of the human coronavirus 229E nucleocapsid protein. FEBS Lett 587:120–127. https://doi.org/10.1016/j.febslet.2012.11.016

Lu S, Ye Q, Singh D et al (2020) The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. Bio Rxiv. https://doi.org/10.1101/2020.07.30.228023

Lutomski CA, El-Baba TJ, Bolla JR, Robinson CV (2020) Autoproteolytic products of the SARS-CoV-2 nucleocapsid protein are primed for antibody evasion and virus proliferation. Bio Rxiv. https://doi.org/10.1101/2020.10.06.328112

McBride R, van Zyl M, Fielding BC (2014) The coronavirus nucleocapsid is a multifunctional protein. Viruses 6:2991–3018. https://doi.org/10.3390/v6082991

Nikolic J, Le Bars R, Lama Z et al (2017) Negri bodies are viral factories with properties of liquid organelles. Nat Commun 8:58. https://doi.org/10.1038/s41467-017-00102-9

Peng T-Y, Lee K-R, Tarn W-Y (2008) Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. FEBS J 275:4152–4163. https://doi.org/10.1111/j.1742-4658.2008.06564.x

Peng Y, Du N, Lei Y et al (2020) Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. EMBO J 39(20):e105938. https://doi.org/10.15252/embj.2020105938

Perdikari TM, Murthy AC, Ryan VH et al (2020) SARS-CoV-2 nucleocapsid protein undergoes liquid-liquid phase separation stimulated by RNA and partitions into phases of human ribonucleoproteins. Bio Rxiv. https://doi.org/10.1101/2020.06.09.141101

Ringel M, Heiner A, Behner L et al (2019) Nipah virus induces two inclusion body populations: Identification of novel inclusions at the plasma membrane. PLoS Pathog 15:e1007733. https://doi.org/10.1371/journal.ppat.1007733

Savastano A, de Opakua AI, Rankovic M, Zweckstetter M (2020) Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. Bio Rxiv. https://doi.org/10.1101/2020.06.18.160648

Skinner SP, Fogh RH, Boucher W et al (2016) CcpNmr analysis assign: a flexible platform for integrated NMR analysis. J Biomol NMR 66:111–124. https://doi.org/10.1007/s10858-016-0060-y

Solyom Z, Schwarten M, Geist L et al (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR 55:311–321. https://doi.org/10.1007/s10858-013-9715-0

Wu C-H, Chen P-J, Yeh S-H (2014) Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. Cell Host Microbe 16:462–472. https://doi.org/10.1016/j.chom.2014.09.009

Ye Q, West AMV, Silletti S, Corbett KD (2020) Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. Bio Rxiv. https://doi.org/10.1101/2020.05.17.100685

Zeng W, Liu G, Ma H et al (2020) Biochemical characterization of SARS-CoV-2 nucleocapsid protein. Biochem Biophys Res Commun 527:618–623. https://doi.org/10.1016/j.bbrc.2020.04.136

Zhou Y, Su JM, Samuel CE, Ma D (2019) Measles virus forms inclusion bodies with properties of liquid organelles. J Virol. https://doi.org/10.1128/JVI.00948-19

Zúñiga S, Sola I, Moreno JL et al (2007) Coronavirus nucleocapsid protein is an RNA chaperone. Virology 357:215–227. https://doi.org/10.1016/j.virol.2006.07.046

"Multi-scale studies of Measles virus nucleocapsid assembly"

Résumé

La séparation de phases liquide-liquide est un phénomène crucial dans toute la biologie, impliqué dans de multiples processus cellulaires et responsable de la formation d'organelles sans membrane qui sont essentielles pour l'organisation spatiale et temporelle intracellulaire. Il a été proposé que de tels organites soient impliqués dans la formation des usines de réplication virale, qui résultent de l'infection par un certain nombre de virus à ARN simple brin négatif.
Ici, nous étudions la séparation de phase de la machinerie de réplication de la rougeole *in vitro*, nous identifions la nature et la localisation des interactions requises et montrons que certains processus essentiels sont accélérés, en particulier l'assemblage de la nucléocapside où la nucléoprotéine de la rougeole se lie à l'ARN génomique pour former des capsides hélicoïdales.
Les protéines impliquées dans la réplication de la rougeole sont connues pour être phosphorylées dans la cellule, cependant, le rôle fonctionnel de cette modification post-traductionnelle n'était pas compris auparavant. Au cours de ma thèse, nous avons découvert que la phosphorylation de la phosphoprotéine de la rougeole déclenche l'assemblage de la nucléocapside.
Pour mieux comprendre l'origine physique de la séparation de phase des protéines, nous avons étudié la structure et la dynamique des protéines en utilisant un système modèle. En utilisant une partie désordonnée de la nucléoprotéine de la rougeole, nous comparons le comportement de la protéine entre les états dilué et condensé. En utilisant la spectroscopie RMN, nous effectuons une comparaison spécifique au site des amplitudes de mouvement et des échelles de temps de la protéine entre les phases.
Enfin, il a également été démontré que le SARS-COV-2 forme des condensats viraux ne nécessitant qu'une seule protéine virale : la nucléoprotéine. Nous caractérisons ici les régions intrinsèquement désordonnées du SARS-COV-2 N et démontrons sa séparation de phase *in vitro*.

Mots-clés : Séparation de phases liquide-liquide, RMN, rougeole, usines virales, dynamique des protéines, SARS-COV-2

Abstract

Liquid-liquid phase separation is a crucial phenomenon throughout biology involved in multiple cellular processes and responsible for forming membraneless organelles that are essential for the intracellular spatial and temporal organisation. Such organelles have been proposed to be involved in the formation of so-called viral replication factories, which result from infection by a number of negative single-strand RNA viruses.
Here, we study measles replication machinery phase separation *in vitro*, we identify the nature and location of the required interactions and show that certain essential processes are accelerated, in particular nucleocapsid assembly where the measles nucleoprotein binds to the genomic RNA to form helical capsids. Proteins involved in measles replication are known to be phosphorylated in the cell, however, the functional role of this post-translational modification was not previously understood. During my thesis we discovered that phosphorylation of measles phosphoprotein triggers nucleocapsid assembly.
To better understand the physical origin of protein phase separation, we studied protein structure and dynamics using a model system. Using a disordered part of measles nucleoprotein, we compare protein behaviour between dilute and condensed states. Using NMR spectroscopy, we perform a site-specific comparison of motional amplitudes and timescales of the protein between phases.
Finally, SARS-COV-2 has also been shown to form viral condensates requiring only one viral protein: the nucleoprotein. Here, we characterise the intrinsically disordered regions of SARS-COV-2 N and demonstrate its phase separation *in vitro*.

Keywords : Liquid-liquid phase separation, NMR, measles, viral factories, protein dynamics, SARS-COV-2