Investigating the role of intrinsic conformational disorder in mumps virus proteins

Stefaniia Ivashchenko

To cite this version:

Stefaniia Ivashchenko. Investigating the role of intrinsic conformational disorder in mumps virus proteins. Virology. Université Grenoble Alpes, 2019. English. NNT : 2019GREAY026 . tel-02448300v1

HAL Id: tel-02448300
https://theses.hal.science/tel-02448300v1

Submitted on 22 Jan 2020 (v1), last revised 28 Jan 2020 (v2)

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **PHYSIQUE POUR LES SCIENCES DU VIVANT**

Arrêté ministériel : 25 mai 2016

Rédigée par

Stefaniia IVASHCHENKO

Thèse dirigée par **Martin BLACKLEDGE**

préparée au sein de l'équipe de **Flexibilité et Dynamiques des Protéines de l'Institut de Biologie Structurale**
dans **l'École Doctorale de Physique de Grenoble**

Rôle du désordre conformationnel dans les protéines du virus des oreillons

Investigating the role of conformational disorder in mumps virus proteins

Thèse soutenue publiquement le **1 juillet 2019** devant le jury composé de :

Dr. Sonia LONGHI
Laboratoire de l'Architecture et Fonction des Molécules Biologiques, Marseille, Rapporteur
Dr. Christina SIZUN
Institut de Chimie des Substances Naturelles, Gif-sur-Yvette, Rapporteur
Dr. Pau BERNADO
Centre de Biochimie Structurale, Montpellier, Examinateur
Prof. Marc JAMIN
Institut de Biologie Structurale, Grenoble, Président
Dr. Martin BLACKLEDGE
Institut de Biologie Structurale, Grenoble, Directeur de thèse

TABLE OF CONTENTS

ABBREVIATIONS

A adenosine
ASTEROIDS a selection tool for ensemble representations of intrinsically disordered states
BCA bicinchoninic acid assay
BCE before the common era
BEST band-selective excitation short-transient
CDC Centre for disease control
CKII casein kinase II
CSA chemical shift anisotropy
DNA deoxyribonucleic acid
DTT dithiothreitol
EBOV Ebola virus
EDTA ethylenediaminetetraacetic acid
EM electron microscopy
FID free induction decay
G guanosine
GST glutathione *S*-transferase
HeV Hendra virus
HMPV human metapneumovirus
HPIV3 human parainfluenza 3
HSQC heteronuclear single quantum coherence
ICTV International committee on taxonomy of viruses
IDP intrinsically disordered protein
IPTG isopropyl β-D-1-thiogalactopyranoside
JL Jeryll Lynn (vaccine)
MALLS multiangle laser light scattering
MARV Marburg virus
MD molecular dynamics
MeV measles virus
MMR measles-mumps-rubella (vaccine)
MoRE molecular recognition element
mRNA messenger ribonucleic acid
MuN mumps nucleoprotein
MuP mumps phosphoprotein
MuV mumps virus
L large protein
L-3 Leningrad 3 (vaccine)

LB Luria Broth

N nucleoprotein

NC nucleocapsid

NDV Newcastle disease virus

NiV Nipah virus

NMR nuclear magnetic resonance

NOE nuclear Overhauser effect

ORF open reading frame

P phosphoprotein

PBD Protein Data Bank

PEG polyethylene glycol

PIV5 parainfluenza virus 5

PLK1 Polo-Kinase 1

PRE paramagnetic relaxation enhancement

PTM post-translational modification

RDC residual dipolar coupling

RF radio-frequency

RLP ring-like particle

RNA ribonucleic acid

RSV respiratory syncytial virus

SAXS small-angle X-ray scattering

SDS-PAGE sodium dodecyl sulphate–polyacrylamide gel electrophoresis

SeV Sendai virus

SOFAST-HMBC heteronuclear multiple bond correlation

STAT1/STAT2 signal transducer and activator of transcription 1/2

TCEP tris(2-carboxyethyl)phosphine

TEMPO (2,2,6,6-tetramethylpiperidin-1-yl)oxyl

TEV *Tobacco Etch Virus* (protease)

TMAO trimethylamine N-oxide

TROSY transverse relaxation-optimized spectroscopy

SCS secondary chemical shifts

SSP secondary structure propensity

VSV vesicular stomatitis virus

WHO World Health Organisation

UNICEF United Nations International Children's Emergency Fund

ZFD zinc finger domain

CHAPTER 1.
INTRODUCTION

1. INTRODUCTION

Mumps is a highly contagious infectious disease that was first described by one of the fathers of medicine Hippocrates in his work *Epidemics* in 5th century BC. The epidemic of mumps affected people from Thasos island in Ancient Greece with specified symptoms as fever, headache, swelling near the ears, sore throat, and, more rarely, swollen testicles [Tsoucalas et al., 2013].

The word "mumps" likely comes from an old English verb (17th century) that means to grimace or to mumble [Online Etymology Dictionary]. The French version of "oreillons" derives from the clinical signs of the pain in the area of ears (*oreilles*) [Dictionnaire électronique de l'AF]. In the 18th century, the physician Hamilton found an effect of mumps disease on the central nervous system [T.E.C., 1970]. But only in 1945 the infecting agent, mumps virus, was isolated and the year later the first inactivated vaccine was developed.

Nowadays, despite the fact that mumps disease mostly affects children, and the use of vaccines has become routinary, mumps outbreaks among highly vaccinated populations still occur. In my opinion, this leads to review the concept of prevention and control of viral infections and, emphasizes the need for searching for highly effective medicines that take into the account complicated viral cycle mechanisms.

In this chapter, I will discuss basic concepts of the mumps disease and existing treatments against it, describe the molecular structure of the pathogen and the most significant objectives of this scientific study, i.e. the structure and interactions of the viral proteins and the functional role of their intrinsically disordered regions.

1.1. Mumps disease

Mumps disease is caused by mumps virus, a member of the *Paramyxoviridae* family, the morphology of which will be discussed later. Primarily, the virus affects the salivary glands (parotitis) that usually starts with headache, fever, muscle pain, malaise. After 2-10 days the symptoms disappear, in some cases (50-60% of patients) following with asymptomatic cerebrospinal fluid alterations, rarely with aseptic meningitis (15% of cases) and encephalitis (0.02-0.3%). 10-20% of postpubertal males suffer from orchitis (inflammation of the testicles) that can lead to permanent subfertility. Pregnant women (gestational age below 12 weeks) infected with mumps virus have spontaneous abortion in 25% of cases. In addition, there are no reported cases of the effect of mumps disease on the foetal development. Although the fatality rate from mumps disease is quite low, complications concerning the central nervous system and reproductive organs disorders are non-negligible. Humans are the only natural host for mumps virus. It is believed that the virus is transmitted from human to human through the respiration of salivary droplets or direct contact. Mumps is a childhood disease affecting children from about 5 to 9 years. Nevertheless, adolescents and adults can also develop mumps. The incubation times varies from 16-18 days to 2-4 weeks [Rubin et al., 2015].

1.1.1. Prevention and treatment

No specific therapy against mumps exists. Usually, the basic treatment is symptomatic. Once infection with mumps virus gives a natural lifelong protection. Consequently, the use of vaccines in order to prevent and control the virus spreading, remains the only way of mumps prophylactic treatment.

The generation of the first mumps vaccine that consists of inactivated virus in 1946 was followed by its licensure in 1948 in the USA. It was used in the period of 1950 - 1978 but did not provoke efficient immune protection. Since then other countries developed a variety of attenuated (cultivated under conditions disabling its virulence) mumps vaccines that differed historically in cell substrates used for vaccine preparation and manufacturing, and are effective in successful reduction of mumps cases. Table 1.1 represents the list of vaccines in current use worldwide.

In the USA, the recommended vaccination procedure consists of the introduction of two vaccine doses to children: first dose is given at the age of 12-18 months and second at least one month later, usually, before going to school, at 5-6 years. In France, first dose is introduced for 12 months old infants and second between 13 and 24 months. Contraindications are mostly based on allergic reactions, thus, it is not recommended to introduce the vaccine to patients with immune suppression or deficiency and pregnant women [Mumps virus vaccines. Wkly Epidemiol Rec, 2007].

Jeryl Lynn vaccine

The strain of mumps virus used in this vaccine is named after Jeryl Lynn Hillerman, daughter of Merck scientist Maurice Hillerman, that first collected the virus from her throat in 1963 and transferred first to embryonated hen's eggs and after to chicken embryo cultures. The vaccine was brought to market in 1967. Currently, Jeryl Lynn is a part of combinational MMR (measles-mumps-rubella) vaccine. Serological studies on antibody titers showed evidence of 80-100% efficacy (proportionate reduction in cases among vaccinated persons compared to an unvaccinated group). However, during outbreaks it ranges between 63 and 96% [Hilleman et al., 1967; Weibel et al., 1967].

Table 1.1. Available mumps vaccines.

Vaccine strain	***Genotype***	***Manufacturer***	***Name***	***Main distribution area***
Jeryl Lynn	A	Merck / Aventis Pasteur	MMRII® MMR-Vaxpro® ProQuad®	Worldwide Europe Worldwide
		GlaxoSmithKline (RIT 4385 strain)	Priorix® PriorixTetra®	Worldwide Europe
		Netherlands Vaccine Inst.	BMR vaccine®	Netherlands
		Sevapharma Inc.	Pavivac® Trivivac®	Chech Republic Slovak Republic
		Dalian Jinjang-Andi Bioproducts	S79	China
Leningrad-3	-	Moscow State Facility for Bacterial preparations	Leningrad-3	Russia
Leningrad-Zagreb	-	Inst. Of Immunology, Zagreb Serum Institute of India	Leningrad-Zagreb	Croatia, Slovenia Worldwide
Urabe Am9	B	Sanofi-Pasteur	Trimovax®	Worldwide
		Biken	Pluserix®	Japan
Torii		Takeda Pharmaceutical	Torii	Japan
Hoshino		Kitasato Institute	Hokken Hoshino	Japan, Korea
S-12	H	Razi State Serum and Vaccine Institute	S-12	Iran
		Crucell Vaccines	BBM-18®	Europe

It was found that Jeryl Lynn vaccine consists of two distinct viral substrains JL-5 and JL-2 [Afzal et al., 1993] that are different from each other for 414 nucleotides and exist in 5:1 ratio [Amexis et al., 2002]. Mechanisms responsible for the generation of these substrains are unclear. The newer strain RIT 4385, derived from JL-5 vaccine, showed almost the same seroconversion rates of about 96% against 97% of Jeryl Lynn. Being a biological medical product, Jeryl Lynn vaccine has some adverse effects that are mostly based on the allergic reactions to the vaccine components [MMR – Vaccine use and strategies, 1998]. No relationship between the use of vaccine and development of encephalitis, optic neuritis and orchitis was found. However, mild parotitis under the administration of Jeryl Lynn was observed in 1% of vaccinated patients [Fescharek et al., 1990].

Leningrad-3 and Leningrad-Zagreb vaccines

Developed in former Soviet Union, Leningrad-3 (L-3) vaccine was used in a large scale nation immunisation program since 1981 [Medynicin, 2003]. It was elaborated in guinea-pig kidney cell culture, further passaged in Japanese quail embryos [Smorodintsev et al., 1970]. L-3 has protective efficacy of 92-99% [Unanov, 1977] and only a limited number of clinical trials identify side effects after vaccination. However, it was reported about asymptomatic transmission of L-3 vaccine from healthy vaccinated to previously vaccinated patients [Atrasheuskaya et al., 2005].

The further attenuation of L-3 strain was obtained in Croatia with a different passage procedure in chick embryo fibroplast cell cultures. Leningrad-Zagreb (L-Z) vaccine has a rate of protection from 97 to 100% [Beck, 1989]. Nevertheless, recent studies showed evidence of L-Z association with aseptic meningitis [da Cunha, 2002].

Urabe Am9 vaccine

The vaccine was first licensed in Japan in 1979, and later in Belgium, France and Italy. Urabe is usually prepared in the amnion of embryonated hen's eggs or in chick embryo cell cultures [Vesikari, 1983]. Immunogenic properties were reported to be quiet different depending on the country where the clinical trial investigations were made and the age of patients: from 75% in India to 100% in South Africa for 15 months old children [Singh, 1994]. The safety of Urabe vaccine is under discussion. Being a mixture of viruses, it was found to provoke the same adverse effects as L-3 [Sawada et al., 1993].

Other vaccines

Among the vaccines that were officially declared to be ineffective and not recommended to routine use by the World Health Organisation (WHO) is Rubini (Switzerland). Torii, Hoshino, Miyahara and NKM-46 have limited scale use only in Japan and Korea. Their immunological properties are approximately equal to those of Urabe [Mumps virus vaccines. Wkly Epidemiol Rec, 2007].

1.1.2. Mumps outbreaks

After the introduction of vaccine, mumps disease incidences in industrialised countries decreased by more than 95%. However, the early 2000s were characterised by several notable outbreaks that occurred in the USA, Belgium, United Kingdom, Sweden, Netherlands, Canada, Czech Republic [Cohen et al., 2007; Sartorius et al., 2005; Vandermeulen et al., 2004; Brockhoff et al., 2010; Boxall et al., 2008]. In the USA, where the mumps epidemiological surveillance was followed from the beginning of the 20th century on, during post-vaccinated era, mumps epidemics with relatively large number of cases started to concern national and international health authorities (Figure 1.1) [Principi and Esposito, 2018].

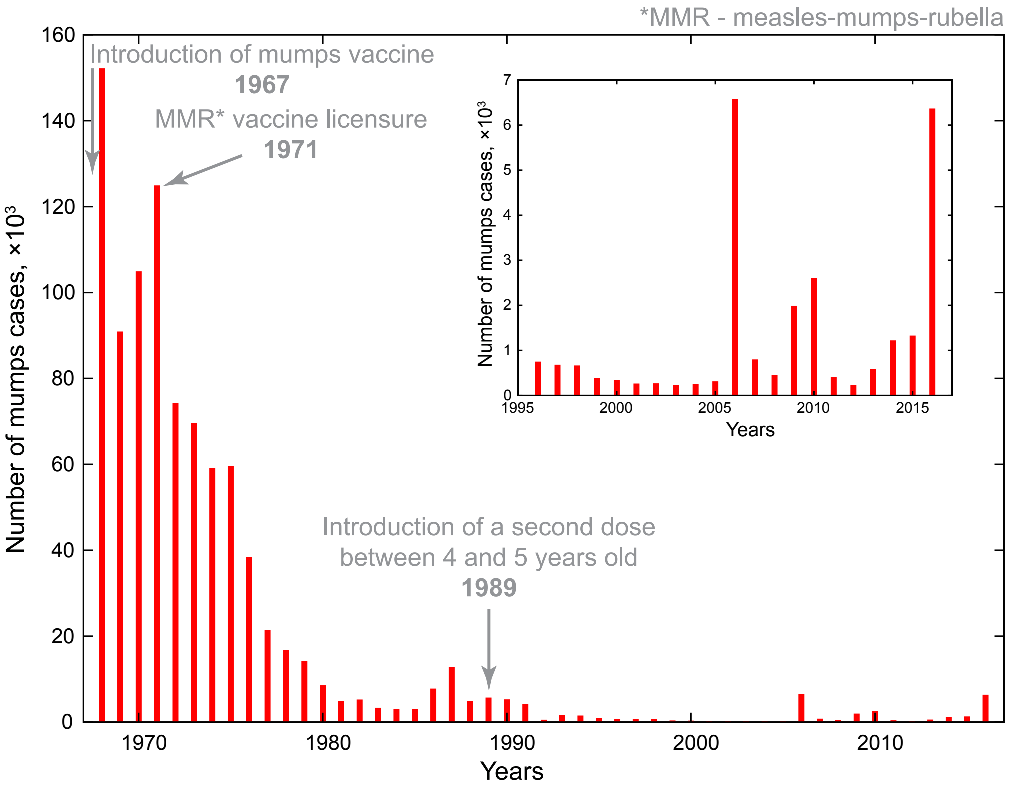


Figure 1.1. Mumps cases reported in the USA from the year of the introduction of mumps vaccine in 1967 to nowadays (2016) with the time scale of 1996-2016 [NNDS system].

After the addition of Jeryl Lynn to the MMR vaccine and launch of second dose, reported cases of mumps disease decreased to an annual average of 300 in the end of the 90s and beginning of the 2000s. The outbreak of 2006 with 6584 registered cases in a majority of Midwest states was unexpected, as the overall vaccination coverage of concerned population was almost 90%

(62,5% two doses and 24,8% one dose) and affected mostly young adults (18-22 years old) in crowded environments such as schools, university campuses and dormitories [Dayan et al., 2008]. In 2009-2010 two large outbreaks occurred in New York city and the US Territory of Guam. The next epidemic peak happened in 2016, located in Iowa and Illinois states, with 6369 cases [CDCP report 2016]. In 2017 the number of preliminary confirmed cases was more than 5500 [NNDS system].

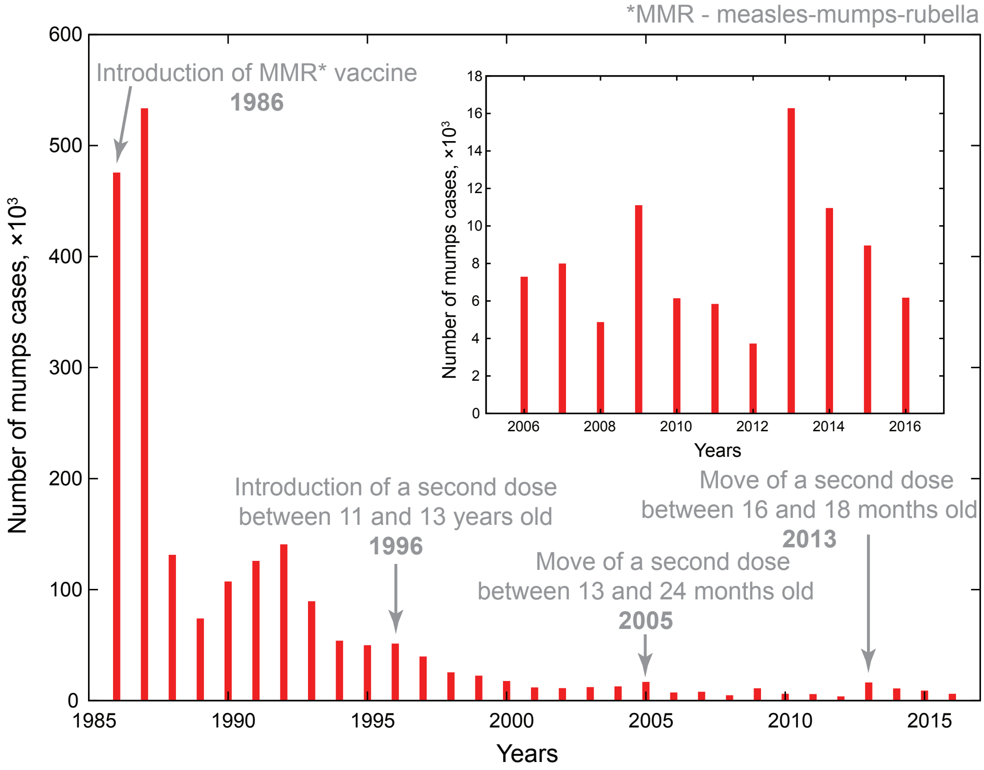


Figure 1.2. The mumps epidemiological situation in France from 1985 to 2016, with scale to the last decade (2006-2016) [Sentinelles, 2016].

In France, the *Sentinelle* network has performed epidemiological data collection from 1986 on (Figure 1.2.), when the MMR vaccine was introduced to the national immunisation program. Mumps is claimed to be not a notifiable disease (disease that needs to be reported to state authorities) in France [Maladies à declaration obligatoire]. Despite a recent outbreak occurred in 2013, 16,281 mumps cases were reported. As an example, at that time sixty-two cases were identified among students of an engineering school in Grenoble (average age about 21 years). Vaccination status with one or two doses was confirmed for 27 patients (43.5%) [Maillet et al., 2015]. However, the current critical situation with measles outbreak in Europe has led to the adoption of a new legislation, which obliges all children born after 1st of January 2018 to be vaccinated against 11 diseases (including mumps) in order to be accepted in child care facilities;

this regulation is also applied to medical and army staff [République française – Loi n° 2017-1836].

In Ukraine, my homeland country, the number of mumps cases tends to progressively diminish after the introduction of the second vaccine dose in 2000 (Leningrad-3 was replaced by Priorix®, Belgium) [Hrynash et al., 2008] (Figure 1.3).

Compared to the measles and rubella epidemical situation with higher disease occurrence and mortality, there is a global tendency that mumps is not considered to be important and prioritised in research investments. However, mumps outbreaks among healthy and vaccinated populations become an alarm bell to turn the international attention to the increasing risk of the appearance of infectious diseases that would be difficult to prevent and control in very near future [Ramanathan et al., 2018].

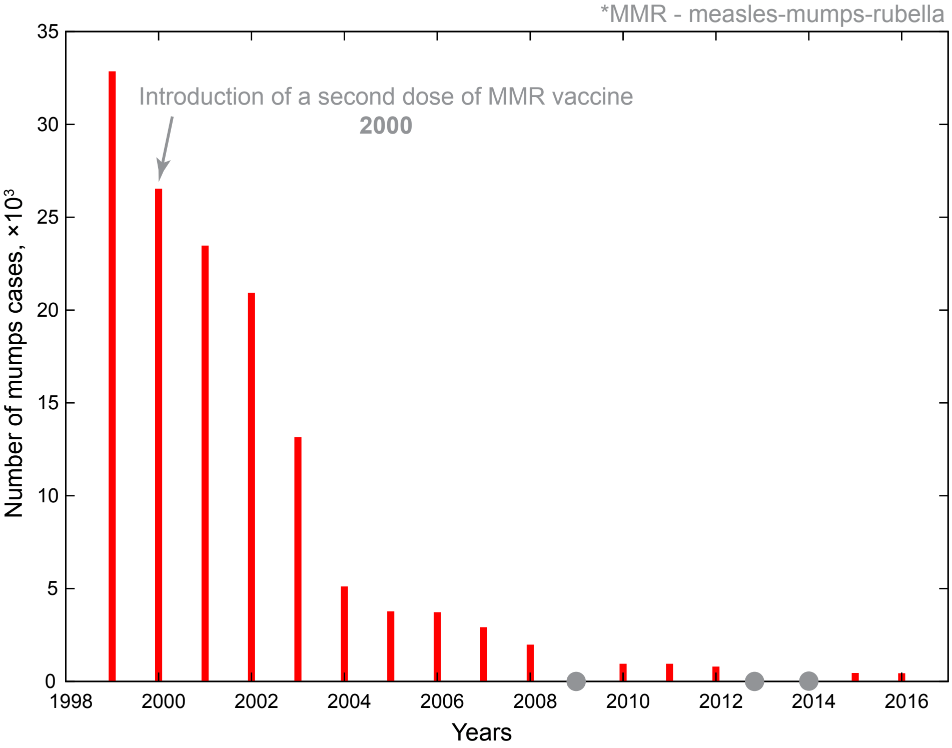


Figure 1.3. Ukrainian mumps situation for 1999-2016 [WHO/IVB database]. The grey circles indicate the absence of data for corresponding years.

1.1.3. Factors influencing mumps outbreaks

Despite the fact that in several countries mumps disease is declared to be vanquished, several aspects affecting appearance and development of mumps epidemics still exist. It is critical to highlight low vaccination coverage, vaccine immune escape and waning immunity (progressive loss of protective antibodies against mumps virus) as leading factors causing mumps outbreaks in the post-vaccine period [Ramanathan et al., 2018; Principi and Esposito, 2018].

Low and incomplete vaccination coverage

Vaccine controversy and misinformation historically accompany public health immunisation campaigns. One of the examples of a major vaccine scare is the scandal created in 1999 by the fraudulent paper of the British doctor Andrew Wakefield claiming the existence of a link between the use of MMR vaccine and the appearance of symptoms of "autistic enterocolitis", an inexistent disease that combines gastroenteric and development disorders in children, who received the above mentioned vaccine [Wakefield, 1998]. This publication heightened the critics of the safety and vaccine hesitancy among the society. Despite a vast number of scientific documentations and reports contradicting the harmfulness of MMR vaccine [Stratton et al., 2001; Immunisation safety review, 2004 etc.] and further retraction of Wakefield's paper from *The Lancet* journal in 2010, anti-vaccination movement still has a major impact on the declining confidence and acceptance of the vaccination procedure [Chang, 2018].

In addition to the pseudoscientific arguments against vaccine use, there are individual's freedom violation reasons (religious and civil), conflicts of interests and financial motives. As follows, Japan banned the MMR vaccine and permanently introduced instead the national monovalent mumps Urabe vaccine [Why Japan banned MMR vaccine, Dly Mail Online, 2013] that was previously withdrawn by Canadian authorities due to its side effects (aseptic meningitis) [Prevention and control of mumps outbreaks in Canada, 2009].

Statistical data showed evidence that during outbreaks most patients suffering from mumps did not even get a single dose of mumps vaccine. The observation was made in the UK, Sweden and Canada [Mumps epidemic – UK, 2006; Sartorius et al., 2005; Public Health Agency of Canada, 2010]. Ukraine, which is currently suffering from a measles epidemic crisis, according to WHO and UNICEF, is the country in Europe with the lowest immunization range that in the last decades (2006-2016) has decreased to less than 50% [Ukraine – estimation of the immunisation coverage, 2017]. The access to Ukrainian educational institutions is restricted by the Law on protection of the population against infectious diseases [Verkhovna Rada of Ukraine, Law 1645-III, 2000]. This means that non-vaccinated children can not be accepted to kinder gardens, schools and universities which is brutally ignored by the corrupt medical system. Parents in doubt of vaccine safety for their children due to its low quality and under anti-vaccine movement, passively refuse the vaccination by falsifying health certificates. Of course, these events give false data to the statistics when "officially" vaccinated patients still develop preventable diseases [Suprun, 2018].

However, comparing the percentage of vaccinated people and the extent of mumps outbreaks in developed European countries and the USA in last decades, the vaccine coverage is not the only evident problem that needs a solution.

Antigenic variation and immune escape

Another factor that can explain the mumps resurgence is the phylogenetic diversity of mumps virus circulating over the outbreaks which can differ from the viral genotype used for vaccination, and, therefore, the levels of the viral neutralization are not same. Currently, twelve mumps virus genotypes exist: from A to L. The classification and epidemiological survey is based on sequencing of mumps SH protein gene that varies with the strain.

Several studies failed to reveal evidence of immune escape by showing high levels of neutralising activity of antibodies stimulated by Jeryl-Lynn mumps vaccine against different virus genotypes [Santak et al., 2013]. However, there are findings suggesting good *in vitro* cross-neutralisation rates between the strains for F viral protein and significantly low rates for HN protein [Rubin et al., 2012]. In general, identifying the viral clusters based on sequencing of genes encoding HN and F proteins, that are immunodominant antigens for the host antibodies production, is one of the promising methods to analyse and track the evolution and transmission of mumps virus [Gouma et al., 2016].

Vaccine failure and waning immunity

All available vaccines against mumps are highly immunogenic, which has been proven clinically in the first years after their introduction. This can be different after a long period of use due to the vaccine waning immunity. On the contrary, during outbreaks, the effectiveness of the Jeryl Lynn vaccine was found to be very low. Due to measles outbreaks in 1980s, the second dose of MMR vaccine was introduced in 1989 in most European countries and the USA. At the same time, the effect of the second boost of mumps vaccine on induced immunity was not statistically followed. The secondary effects of several vaccines that are also considered as one of the causes of the vaccine malfunction should not be underestimated.

In addition to the primary vaccine failure, waning immunity and time-dependent protection after the injection of vaccine were demonstrated. This means that the risk of infection by mumps increases over time after vaccination, possibly more than 10 years later [Dayan et al., 2008]. This may explain why recent outbreaks mostly concerned adult population.

1.1.4. Solutions to the mumps prevention problem

The influence and contribution of the multiple factors in recurrent mumps outbreaks discussed above could not be estimated because of the lack of statistical, epidemiological and biochemical data. Large scale vaccination remains the best defence for prevention and eradication of mumps. It clearly appears that the encouragement of vaccination, its strict control by legislation and overall education can solve the problem of insufficient vaccine coverage in order to avoid outbreaks. Meanwhile, vaccine non-confidence among the global public can be explained by its partial inefficiency and rare but severe side effects. This problem needs to result in the development of novel harmless and efficient vaccine which assures the society in a safety and tolerability of vaccination.

The problem of vaccine immunisation properties declining over time can be addressed by the introduction of a third MMR vaccine dose - a promising preventive action was shown during acute mumps outbreaks at the University of Iowa, USA, in 2015 [Cardemil et al., 2017]. However, the immune stimulating effect of the introduction of the additional booster is temporary and rapidly decreased over time. Certainly, the immune escape is a second argument for the correction, improvement and optimisation of existent vaccines. Strategies of introduction of specifically structured and polyvalent vaccines for each epidemical situation were proposed [May et al., 2017] as emergent solutions.

The first mumps vaccine that saved thousands of lives was invented more than 50 years ago with limited knowledge about the mechanisms of viral invasion and host immune response - it simply worked. Nowadays, it is important to fill the gaps between the epidemiological analysis carried out manually and discoveries about viral behaviour under human infection at the molecular level. In order to develop a targeting vaccine or any other antiviral medicine, more research needs to be dedicated, firstly, to the study of virus components (proteins and RNA) and their interactions with host factors, also in the context of mumps genotype diversity.

1.2. Mumps virus

1.2.1. Paramyxoviridae family

Mumps virus (MuV) belongs to the family *Paramyxoviridae*, genus *Rubulavirus*. *Paramyxoviridae* family members, from Greek *para* (alternate) and *myxa* (mucus) due to the virus binding to host mucoproteins, are classified in *Mononegavirales* order of membrane enveloped nonsegmented single negative strand RNA viruses (Figure 1.4). The classification is based on common morphology criteria, genome organisation and biological functions of viral proteins and nucleic acids. Until 2016, the *Pneumoviridae* family was formally a part of *Paramyxoviridae* and was named as subfamily *Pneumovirinae* [Afonso et al., 2016]. Having some structural and functional similarities with *Paramyxoviridae* family, *Rhabdoviridae, Filoviridae* and *Pneumoviridae* members will also be discussed in this manuscript.

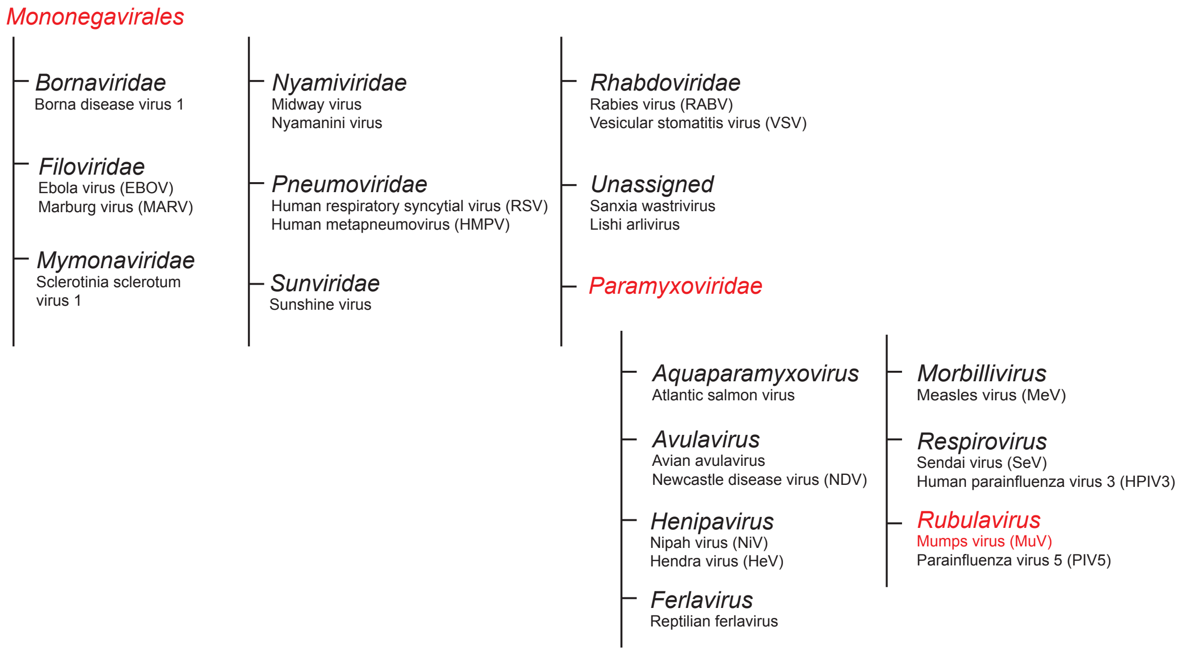


Figure 1.4. Virus classification of *Mononegavirales* order according to the recent taxonomy [ICTV, 2017] with examples of viral species and the position of mumps virus in red.

1.2.2. *Virion structure*

Paramyxoviridae viral particles (virions) are mostly spherical. However, they can be pleomorphic (different in size and shape) as well as filamentous. For example, the diameter of these particles varies from 110 to 540 nm for Sendai virus (SeV), 50-510 nm for measles virus (MeV), and the size for MuV virions ranges from 100 to 600 nm [Cox & Plemper, 2017].

Virions enveloped by a lipid bilayer derived from the host cell membrane from which the virus buds (Figure 1.5). Glycoproteins are inserted in the membrane and are easily visualised by electron microscopy as “spikes” positioned 8-10 nm from the virus surface. Two glycoprotein complexes hemagglutinin-neuraminidase (HN) and fusion (F) proteins are essential for virus attachment to cell receptors and fusion with the host plasma membrane. This process occurs under neutral pH conditions. Inside the virus, the viral RNA is enveloped by the nucleoprotein (N) in filamentous structures called nucleocapsids (NCs). NCs together with the phosphoprotein P (a polymerase cofactor) and large protein L (the viral RNA-dependent RNA-polymerase) form the transcription-replication complex. The matrix protein (M) connects external and internal parts of the virion by associating with N and glycoproteins. *Rubulavirus* members, such as MuV and parainfluenza virus 5 (PIV5), also possess a small hydrophobic protein (SH) with unclear function [Knipe & Howley, 2013].

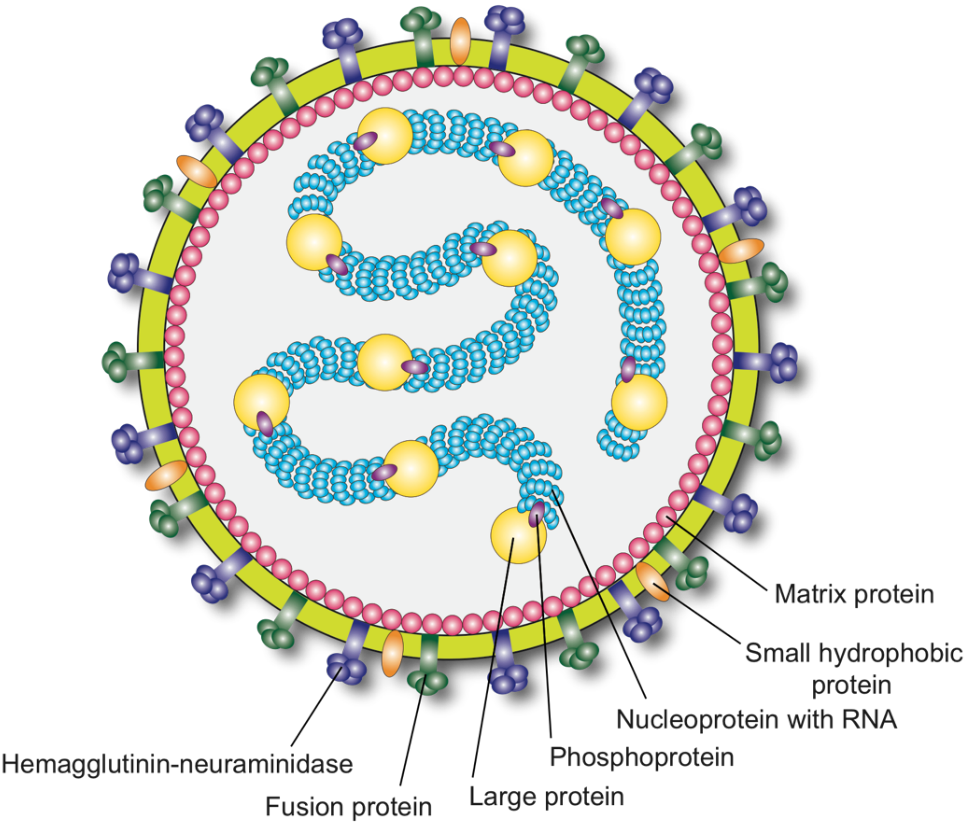


Figure 1.5. Schematic representation of MuV particle. Illustrated relative molecule size, protein abundance and relative localisation are drawn approximately [ViralZone, 2009].

1.2.3. Viral genome

The size of Paramyxoviral RNA genomes ranges between 15,000 to 19,000 nucleotide base pairs. Its short non-coding 3′ extremity (“leader”) of about 50 nucleotides is a signal for initiation of RNA synthesis by the viral polymerase and transcription to subgenomic, capped and polyadenylated messenger RNAs (mRNAs). Another control region is located in the 5′ end (from 50 to 161 nucleotides) and is called “trailer” (Figure 1.6). After protein biosynthesis, the trailer functions as a signal for genome packaging. The antigenome trailer becomes the 3′ promoter for negative polarity genome replication [Noton & Fearns, 2015].

The viral genome encodes for 5-9 proteins. The order of genes that starts with N, follows with P/V/I, M, F, HN and ends with L (with additional gene for SH between F and HN) is highly conserved among *Paramyxoviridae* family members. However, length, nucleotide composition and some protein names are different (Figure 1.6). Between the boundaries of protein genes there are intergenic regions with length of 3 nucleotides for MeV and SeV, and quite variable (1-47 base pairs) for MuV and PIV5. The viral RNA has an extended coding capacity by the mRNA editing through overlapping of open reading frames (ORFs): one coding for P protein and two others leading to the further generation of different proteins: V/I, V/W etc [Field's Virology, 2013].

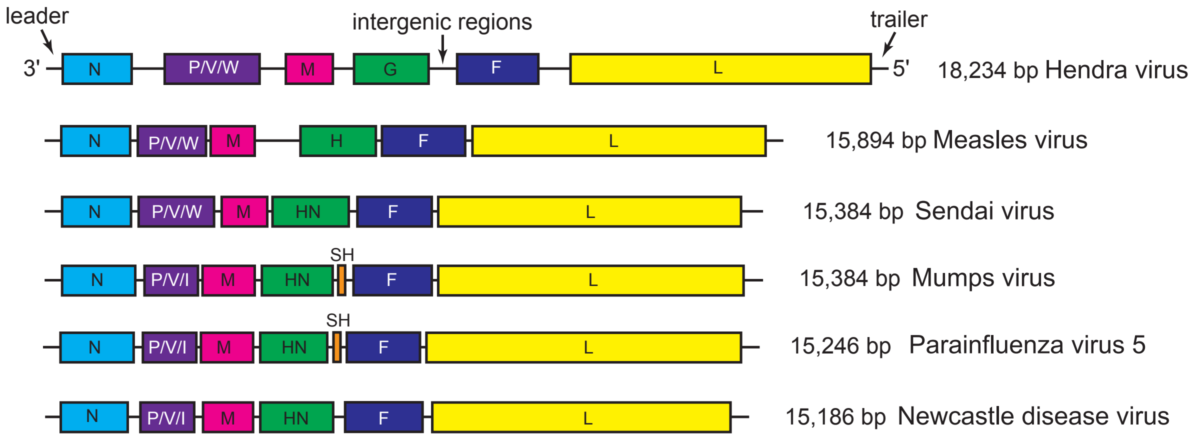


Figure 1.6. Schematic representation of the viral RNA genes of typical members of *Paramyxoviridae* family with indication of genome length in bp.

Inside the host cell the viral genome is always found associated with N that protects it from ribonuclease digestion (Figure 1.7). In addition, capsids rather than naked RNA serve as template for mRNA and for the positive strand (antigenome) synthesis [Acheson, 2011]. Each nucleoprotein unit binds to six RNA residues [Gutsche et al., 2015; Alayyoubi et al., 2015] by following the "rule of six" [Calain & Roux, 1993]. Under the replication of RNA chains at 5′ end, first 6 nucleotides become immediately packaged into one N molecule. According to this rule, for further polymerase recognition, the 3′ end of nucleotide sequence needs to be a multiple of 6. The rule could be evolved from the mechanism of prevention of RNA edited genes replication and encapsidation, and, therefore, generation of erroneous virions [Acheson, 2011].

1.2.4. *Nucleoprotein*

The gene encoding nucleoprotein is the first gene to be transcribed in *Paramyxoviridae* genome. Its size ranges from 489 to 553 amino acids throughout the paramyxoviral family. Structurally N is composed of two distinct regions: conserved among *Mononegavirales* - a structured Ncore and a non-conserved Ntail (Figure 1.8), which is disordered, i.e. without stable three-dimensional structure. Ncore is separated in two subdomains - the N- and C-lobes terminated by flexible loops to N- and C-arms respectively.

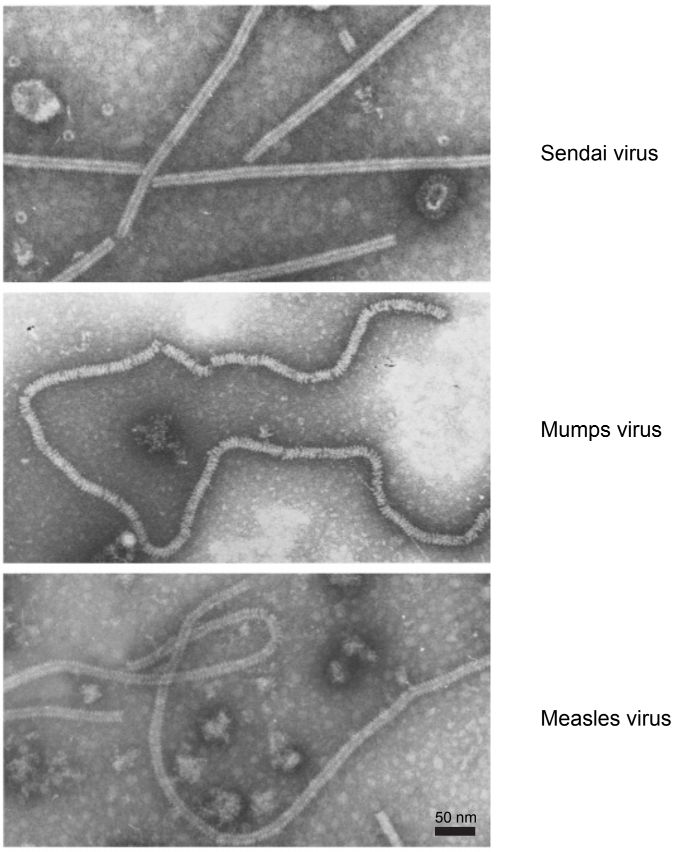


Figure 1.7. Electron micrographs of nucleocapsids isolated from cells infected by different *Paramyxoviridae* [Finch & Gibbs, 1970].

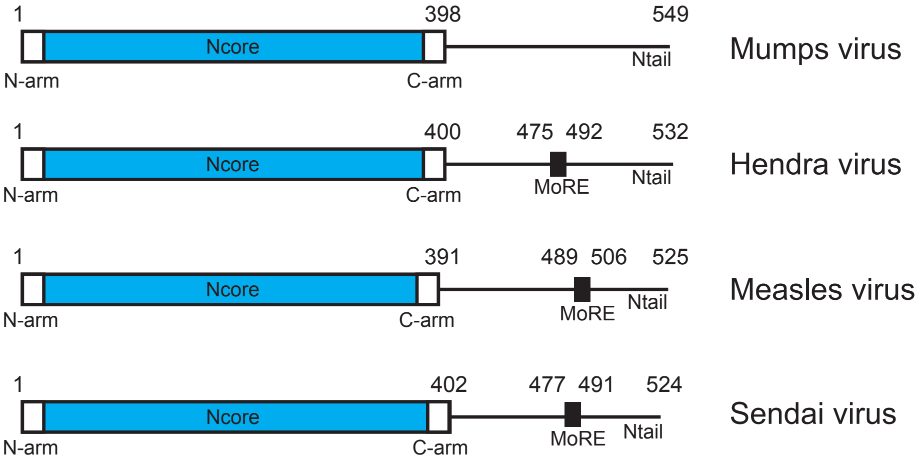


Figure 1.8. Schematic representation of some Paramyxoviral N sequences with important functional regions. Folded domain Ncore is represented in cyan rectangles, N- and C-arms in

white rectangles, disordered Ntail in bold line and transient secondary structure MoRe in black rectangle.

Generally, Ncore is responsible for binding to the viral RNA. For MeV, the function of Ntail, especially, of its molecular recognition element (MoRE), a transient secondary structure region, is to associate with the phosphoprotein, as part of the integral polymerase complex [Longhi et al., 2003; Chan et al., 2004; Houben et al., 2007; Jensen et al., 2008; Jensen et al., 2011; Communie et al., 2013; Habchi et al., 2011; Johansson et al., 2003]. Contrary to other representative *Paramyxoviridae*, MuV Ntail was not found to interact with P, and its function therefore remains unknown [Kingston et al., 2004]. In infected cells, N exists in two forms: associated with RNA in NCs and in monomeric soluble N^o form.

Nucleocapsids and ring-like particles

N has a high affinity for the RNA and encapsidates both negative and positive viral genomes. Apart from protection against RNA degradation, the N envelope limits annealing of mRNA to complementary genomic RNA, and also participates in assembly of NC into virions by interacting with matrix proteins [Field's Virology, 2013]. Practically, bacterial expression of N alone leads to the formation of NCs or/and ring-like particles (RLPs), also called N-rings, always with incorporation of random RNA sequence.

Known *Paramyxoviridae* NCs have a left-handed helical symmetry of diameter about 20 nm and length of about 1 μm, around 13 nucleoprotein units incorporated per turn. Pitch and central hole are about 5 nm [Egelman et al., 1989; Schoehn et al., 2004; Cox et al., 2014]. For the moment (September 2018), only one high resolution cryo-EM reconstructed structure of MeV NC is available [PDB 4UFT] (Figure 1.9B, C).

According to Gutsche et al., 2015, besides the RNA interaction that occurs between the Ncore N- and C-lobes, oligomerisation of nucleoprotein units is mediated by N- and C-arms interaction with neighbouring N_{i+1} and N_{i-1} units respectively. In more detail, the N-arm is inserted into the groove of C-lobe of N_{i+1} while the C-arm remains bound on top of the C-lobe by leaving the intrinsically disordered Ntail to go between NC helical turns (Figure 1.9E, F) [Jensen et al., 2011]. The presence of Ntail allows the NCs to be more flexible. Consequently, NC trypsination leads to the cleavage of Ntail and the appearance of more coiled rigid NCs with low pitch between turns [Desfosses et al., 2010]. The same effect was discovered under both low and high salt conditions [Haggeness et al., 1980].

Concerning MuV, several structural models were proposed for MuV nucleocapsid in low resolution [Cox et al., 2014; Severin et al., 2016]. In the first case, NCs were purified from virions formed in transfected cells. These NCs showed a significant level of flexibility. Consequently, the structure of the nucleocapsids was solved at relatively low (18 Å) resolution. Furthermore, the authors investigated a 25 Å structure of the complex NC-P_{CTD} which appeared having larger diameter than NC alone (Figure 1.10). In the study by Severin et al., 2016, the resolution of the structure was improved to 10,4 Å by coexpressing full length N and P with further proteolysis and RNAse I treatment. As a result, the NCs were empty (RNA free).

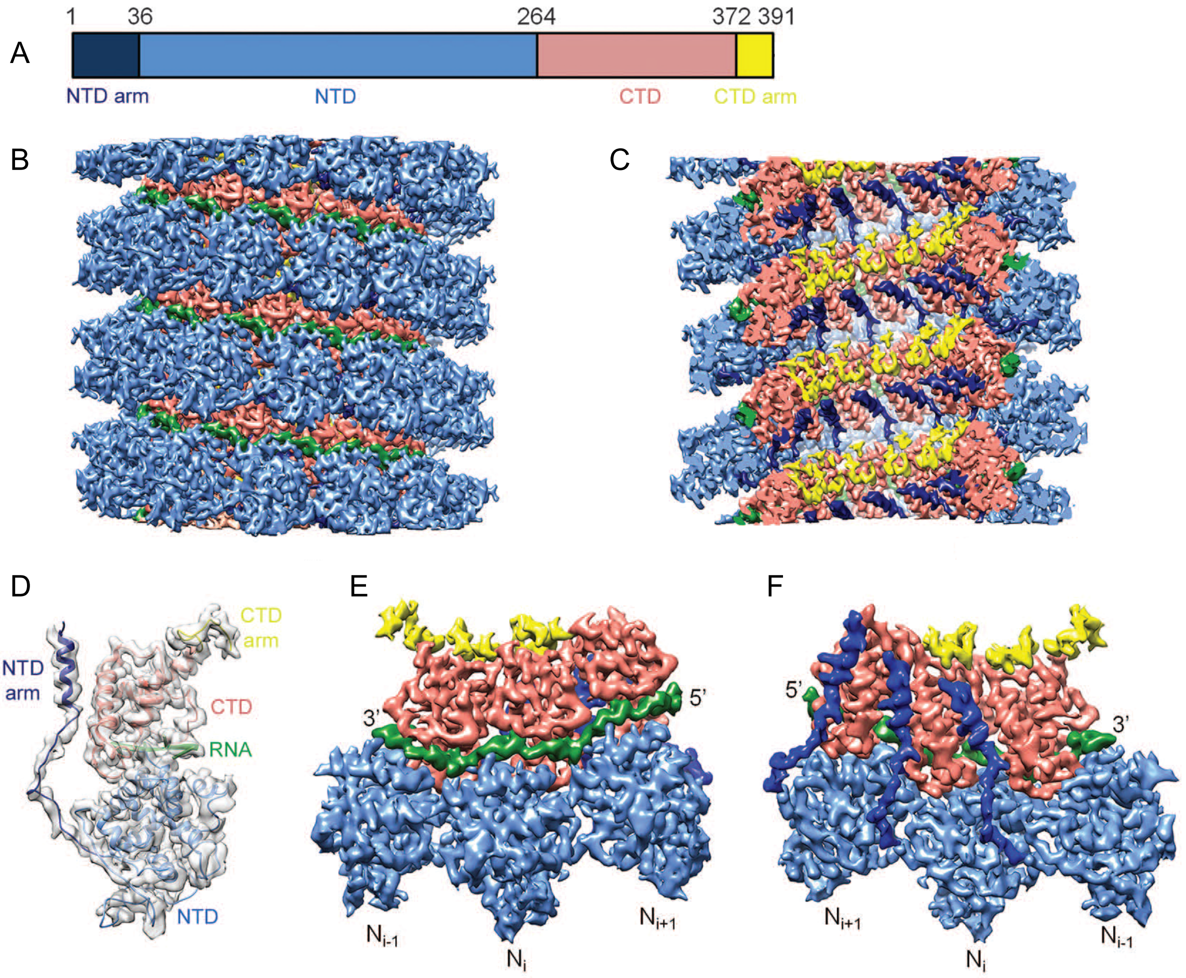


Figure 1.9. Cryo-EM reconstructed structure of MeV NC. Schematic protein sequence of MeV Ncore with important regions (A). Isosurface representation of 3D NC in front (B) and in cutaway (C) with RNA in green. Ncore monomer represented in ribbon (D). The way N protomers interact with each other from exterior (E) and interior (F) of NC helix [Gutsche et al., 2015].

RLPs formation was observed for numerous *Mononegavirales*: RSV, VSV, rabies, PIV5 and MuV [Tawar et al., 2009; Green et al., 2006; Albertini et al., 2006; Alayyoubi et al., 2015; Cox et al. 2009] (Figure 1.11). Consistently, N-rings do not exist under viral infection. However, they can be considered as building blocks for NCs and are usually more amenable to structural biology studies.

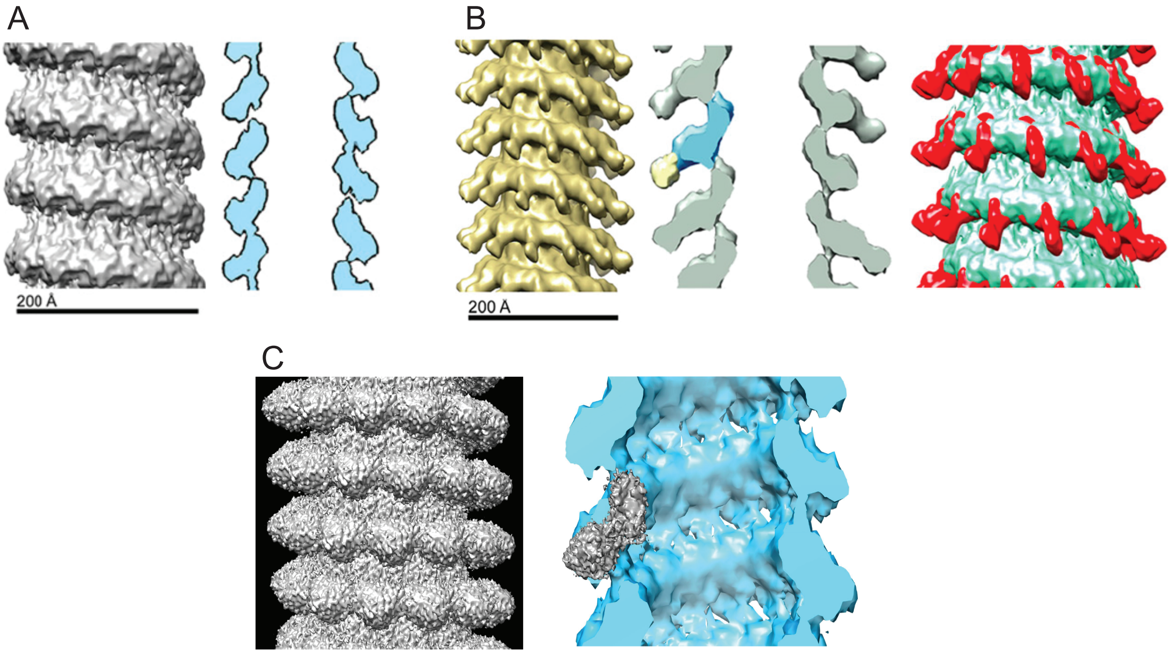


Figure 1.10. Cryo-electron microscopy reconstruction of MuV NCs (A) and complex P_{CTD}-NCs (B) from Cox et al. In the complex of N (in blue/green) and P (yellow/red), it is still difficult to confirm as the last could be interpreted as part of N C-terminal domain. Recombinant RNA-free NCs (in grey) were compared to authentic nucleocapsids (in blue) [Severin et al., 2016].

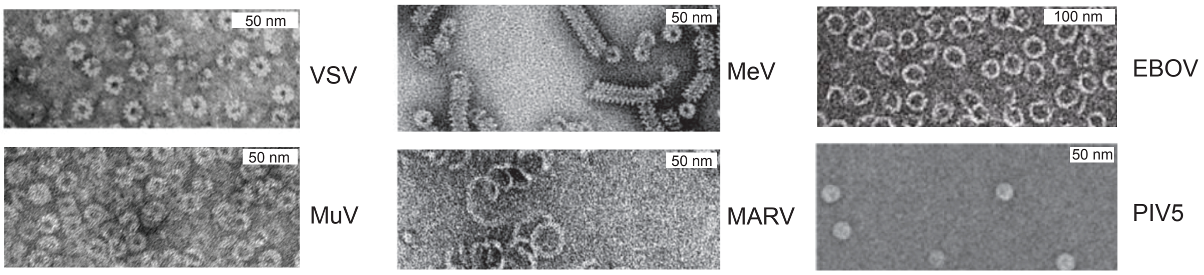


Figure 1.11. *Mononegavirales* intact recombinant nucleocapsids and ring-like particles produced from bacterial cell culture. Adapted from [Alayyoubi et al., 2015; Cox et al. 2009; et al., Ruigrok et al., 2011; Green et al., 2010; Peng et al., 2016].

The crystal structure of PIV5 N-rings [PDB 4XJN] reveals 13 nucleoprotein molecules enveloping 78 nucleotides. As for measles virus NC, the diameter of RLP is measured to be 20 nm (Figure 1.12). PIV5 rings have narrow outside top which is positively charged and wide inside bottom which is negatively charged [Alayyoubi et al., 2015], probably, by attracting them to each other, this can be modulated by Ntail proteolysis and ionic conditions [Cox et al., 2009]. The stabilisation of Ncore lobes and arms through N protomer oligomerisation is essentially identical to that described above for MeV NC.

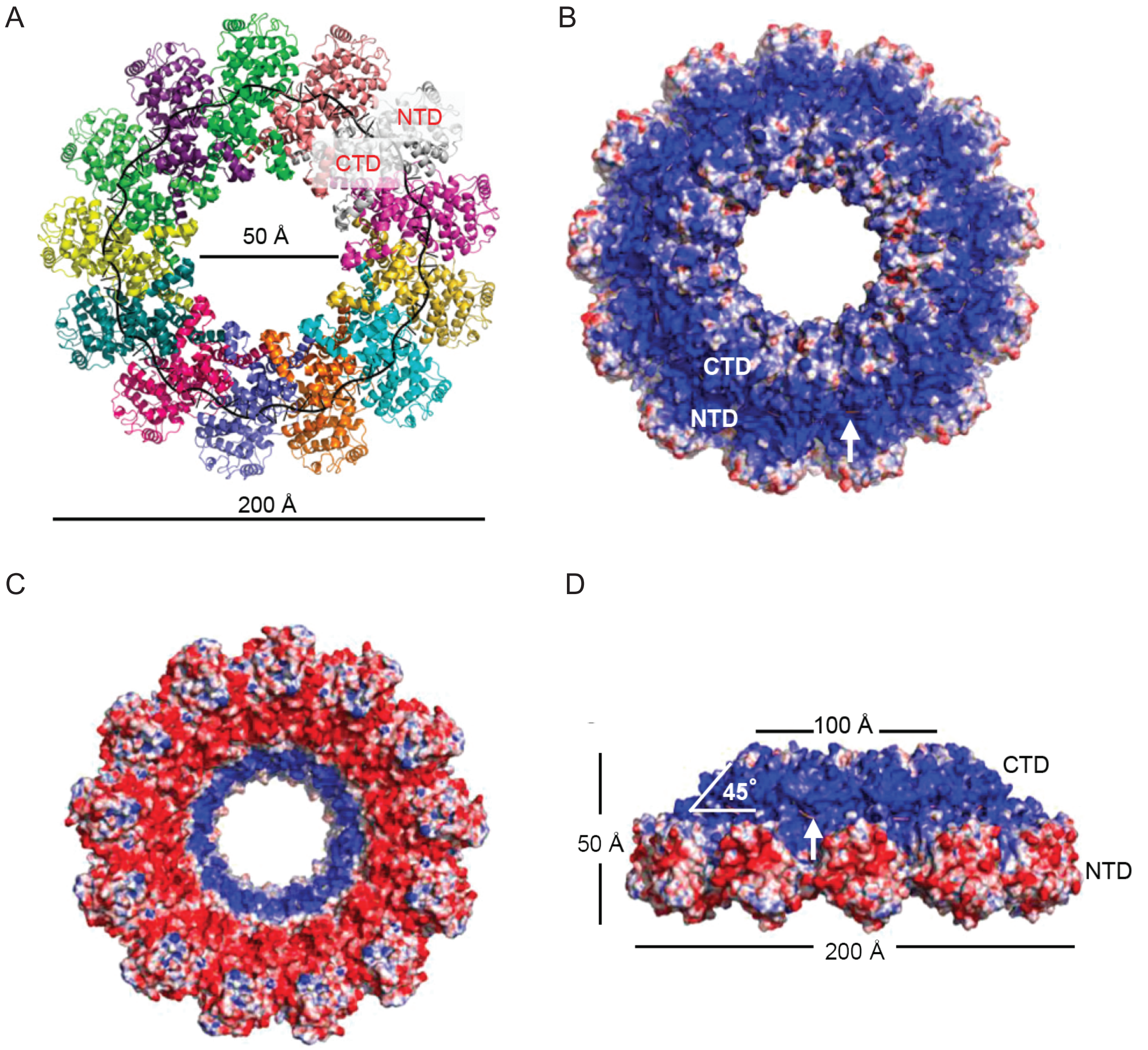
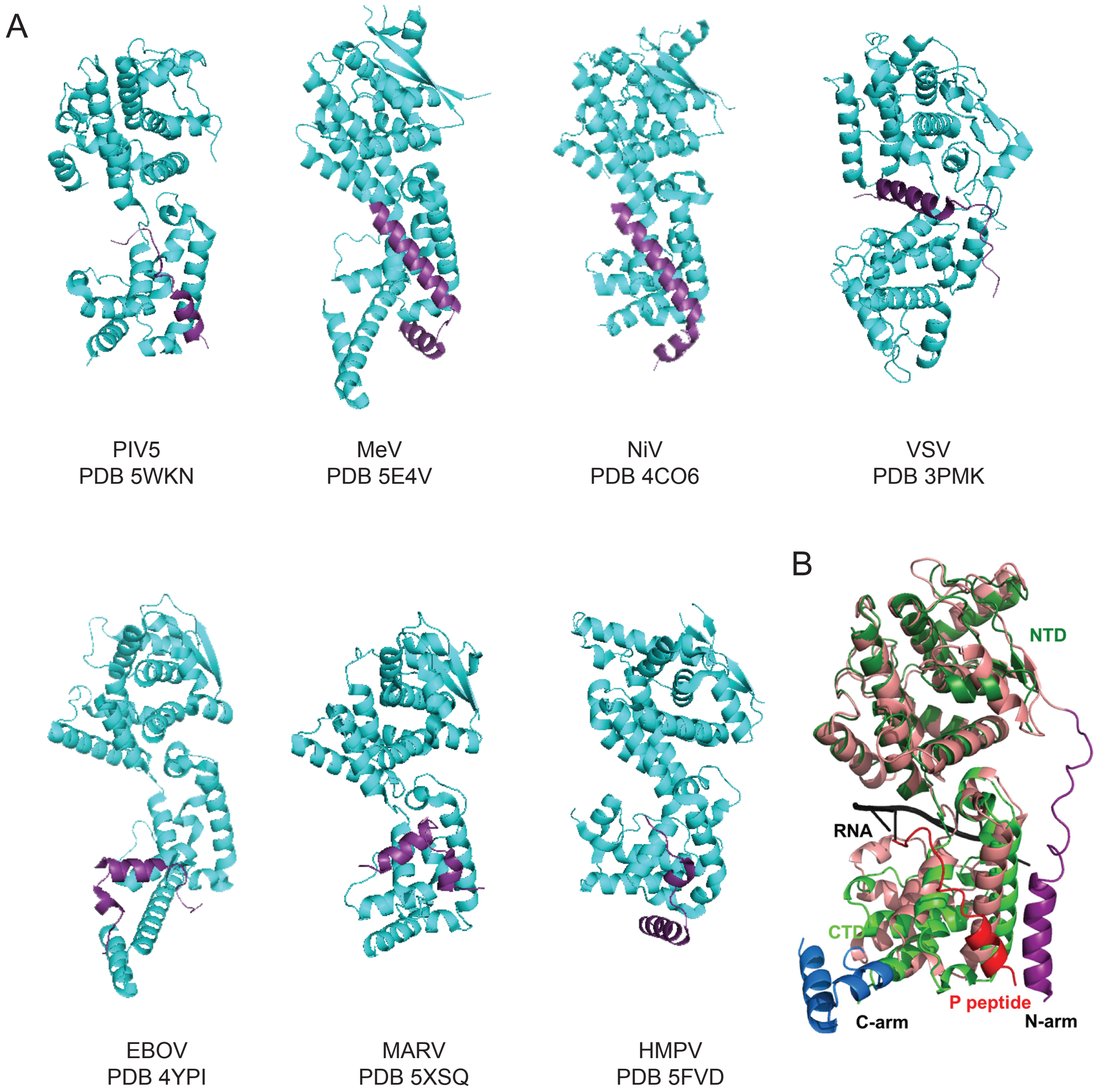


Figure 1.12. Crystal structure of PIV5 RLPs in cartoon representation which contains 13 N units marked in different colours and RNA strand in black, white rectangles indicate NTD and CTD of Ncore (A). Molecular surface representation of N-rings coloured in the respect of electrostatic potential ranging from −10 kT/e(red) to +10 kT/e (blue) from the top (B), bottom (C) and side views (D). RNA is marked in salmon and with white arrow [Aggarwal et al., 2017].

For *Mononegavirales*, post-translational modifications (PTMs) of N, in particular, phosphorylation, can influence viral genome transcription, replication and stability [Hagiwara et al., 2008; Huang et al., 2011; Yang et al., 1999]. For MuV N, a major site of phosphorylation was identified at the position of S439 located in Ntail [Zengel et al. 2015]. The authors show, that mutation of S439 to A increased viral RNA synthesis, which is unique among other *Mononegavirales* members.

N^{0}P complex

Another functional form of nucleoprotein is the N^{0}P complex in which an N-terminal region of P chaperones N in its monomeric RNA-free state. During viral replication, formation of N^{0}P complex is necessary in order to prevent assembly of newly synthetized nucleoprotein molecules which serve for nascent RNA encapsidation. The increasing number of available atomic resolution N^{0}P structures of *Paramyxoviridae* and related viruses (NiV, PIV5, MeV, VSV, EBOV, HMPV, MARV) allows the determination of the structural role of P peptide in N^{0}P (Figure 1.13A).



PIV5
PDB 5WKN

MeV
PDB 5E4V

NiV
PDB 4CO6

VSV
PDB 3PMK

EBOV
PDB 4YPI

MARV
PDB 5XSQ

HMPV
PDB 5FVD

Figure 1.13. N^{0}P complexes of different *Mononegavirales*. P peptide is in violet and Ncore is in cyan (A). Overlay of PIV5 N in monomeric N^{0}P (in green and its CTD in light green) and

polymeric (in salmon) NC form (B). RNA is marked in black, C-arm in blue and N-arm in violet, P peptide in red [Alayyoubi et al., 2015] (B).

Superimposition of nucleoprotein structures in monomeric N^0 and oligomeric NC states shows the existence of two switchable Ncore conformations. Open conformation is adopted by N^{0}P, while in the presence of RNA, the N- and C-lobes are in a closed conformation (Figure 1.13B), meaning that, probably, P hinders the closure of the N molecule and prevents RNA encapsidation.

The common feature for P peptide is its binding mostly to C-lobe of Ncore at the place of the neighbouring N-arm from N$_{i-1}$ protomer in NC. However, there are some differences in P and N binding. For *Rhabdoviridae*, in particular VSV [Leyrat et al., 2011], P was found to occupy the RNA-binding pocket and keeps N's closed conformation. For NiV N^{0}P, P peptide interacts with the surface on N that, in the capsid, interacts with both N- and C-arms [Yabukarski et al., 2014]. However, for example, in HMPV N^{0}P, C-arm of N blocks the RNA binding site [Renner et al., 2016].

Practically, in order to keep MeV N in RNA-free state, N was coexpressed with N-terminal region of P [Guryanov et al., 2015], which makes possible NC self-assembly experiments under addition of short viral RNA to N^{0}P [Milles et al., 2016]. However, this conversion was not observed for EBOV N^{0}P [Kirchdoerfer et al., 2015] which can be explained by the fact that assembly depends on RNA sequence. Equally, the deletion of nucleoprotein N- and/or C-arm parts from PIV5, NiV and EBOV N^{0}Ps which was done for complex stabilization prevents it from oligomerising in presence of RNA [Aggarwal et al., 2017; Yabukarski et al;, 2014; Kirchdoerfer et al., 2015].

1.2.5. P gene proteins

As mentioned before, for several subfamilies of *Paramyxoviridae* three different proteins can be generated from the P gene upon mRNA editing – P, V and I (W). This phenomenon is based on the insertion of one or two non-templated guanosines (G) when viral polymerase slides back on the mRNA sequence (Figure 1.14). For measles and Sendai viruses expression of P protein is a result of non-edited gene sequence, whereas in MuV and PIV5, phosphoprotein ORF is altered by addition of two G-residues (Table 1.2) [Vidal et al., 1990; Paterson & Lamb, 1990; Curran et al., 1991; Cattaneo et al., 1989; Lo et al., 2009].. P, V and I (W) share the same N-terminal and have different C-terminal regions.

Phosphoprotein

Despite various phosphoprotein lengths (from 391 amino acids for MuV to 709 for NiV) and low sequence homology in the intrinsically disordered region, the architecture of *Paramyxoviridae* P proteins is very similar (Figure 1.15). The N-terminal domain (NTD) predicted to be in a random coil conformation. The last 50 residues of the C-terminal domain (CTD) are structured (also named XD) and found to be conserved among viral family members. And finally, for several viruses (MuV, MeV, SeV, NiV, MARV etc), the P oligomerisation domain (OD), situated in the middle of polypeptide sequence, adopts a coiled coil structure, whose conformation has been resolved by X-ray crystallography. As concerns functional properties, in addition to the formation of the complex with L, P interacts with N as well as some host cellular proteins.

Table 1.2. Different *Paramyxoviridae* P gene ORFs.

Examples	*mRNA insertion*		
	0	*+ 1G*	*+ 2G*
PIV5, MuV	V	I	P
MeV, SeV, NiV, HeV	P	V	W
NDV	P	V	I

Generally, approximately the first 60 residues of intrinsically disordered P_{NTD} are responsible for interaction with Ncore in order to form the N^oP complex [Guryanov et al., 2015; Renner et al., 2016; Leyrat et al., 2011; Yabukarski et la., 2014; Aggarwal et al., 2018]. Recently, by replacing the short peptide by full-length MeV P N-terminal domain in N^oP complex, another region was found in the middle of NTD that weakly interacts with Ncore and is important for viral RNA transcription-replication processes [Milles et al., 2018]. Concerning MuV P_{NTD}, it was reported to bind RLPs [Cox et al., 2013] and uncoil authentic nucleocapsids [Cox et al., 2014].

Connected to OD by a flexible linker, the C-terminal domain of P forms a 3-helix bundle (Figure 1.16). Crystal and NMR structures are available for MeV, HeV and SeV, as well as for MuV [Johansson et al., 2008; Blanchard et al., 2004; Kingston et al., 2008]. However, under physiological conditions recombinant MuV XD does not have a stable tertiary structure in solution, for this reason, crystallisation occurred under the addition of TMAO (trimethylamine N-oxide) [Kingston et al., 2008]. In contrast to morbillivirus, respirovirus and henipavirus, where XD binds to Ntail [Longhi et al., 2003; Johansson et al., 2003; Chan et al., 2004;

Houben et al., 2007; Habchi et al., 2011], there is a clear evidence, that MuV P C-terminal domain interacts with Ncore when Ntail was deleted [Kingston et al., 2004]. Another interaction site in the position between two neighbouring nucleoprotein units in NCs was discovered for VSV [Green et al., 2009].

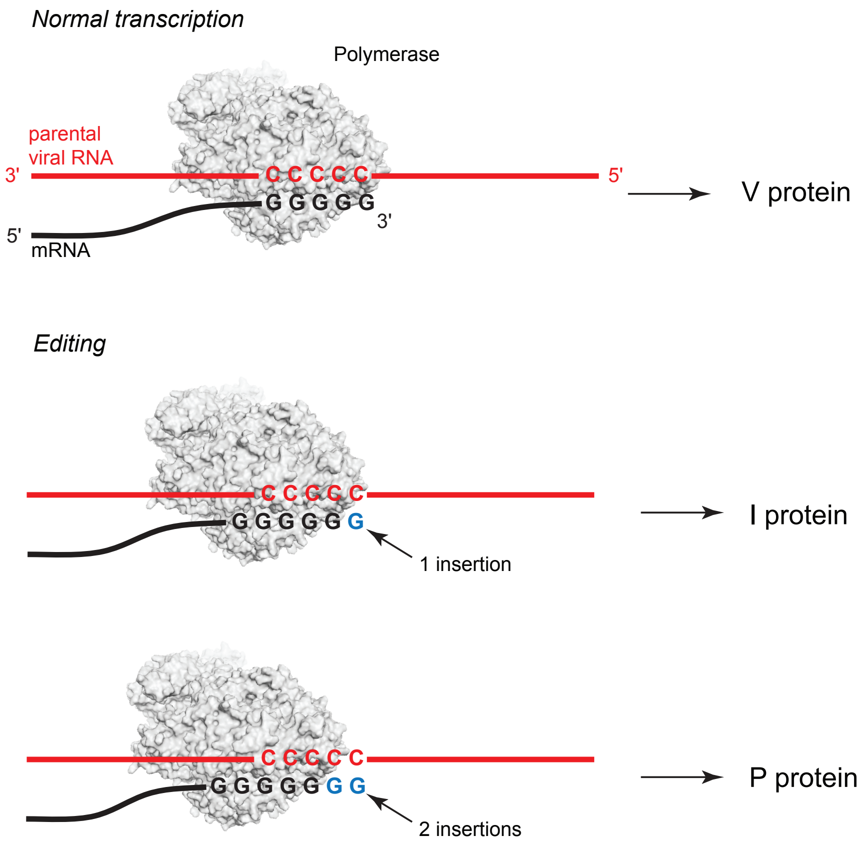


Figure 1.14. Schematic representation of MuV mRNA editing process in order to get V, I and P proteins.

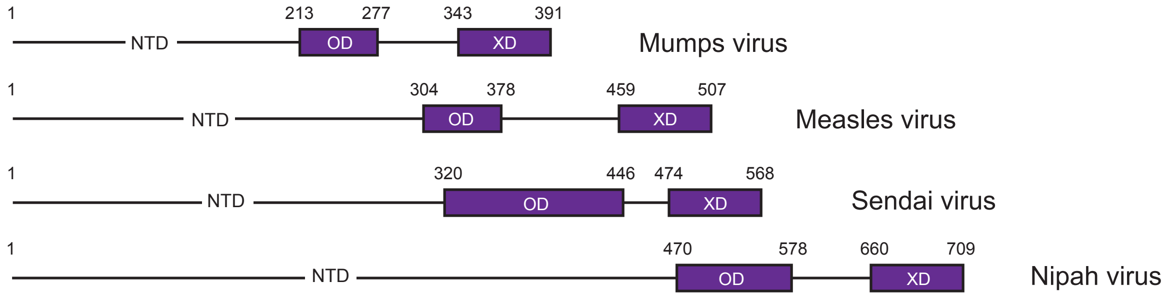
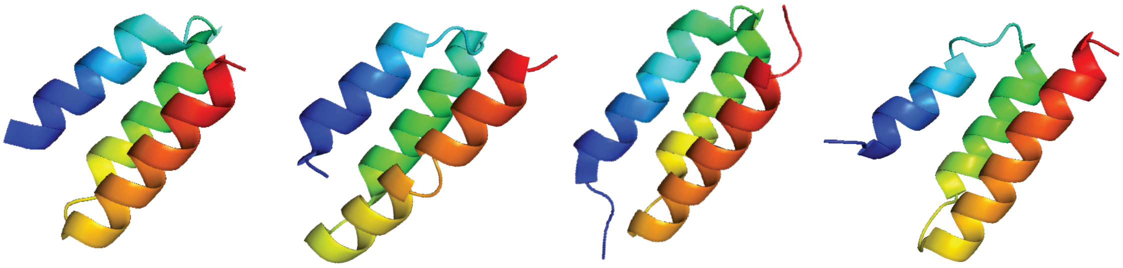


Figure 1.15. Schematic representation of some *Paramyxoviridae* P sequences with important functional regions. Folded domains are represented with violet rectangles, disordered domains with bold lines.



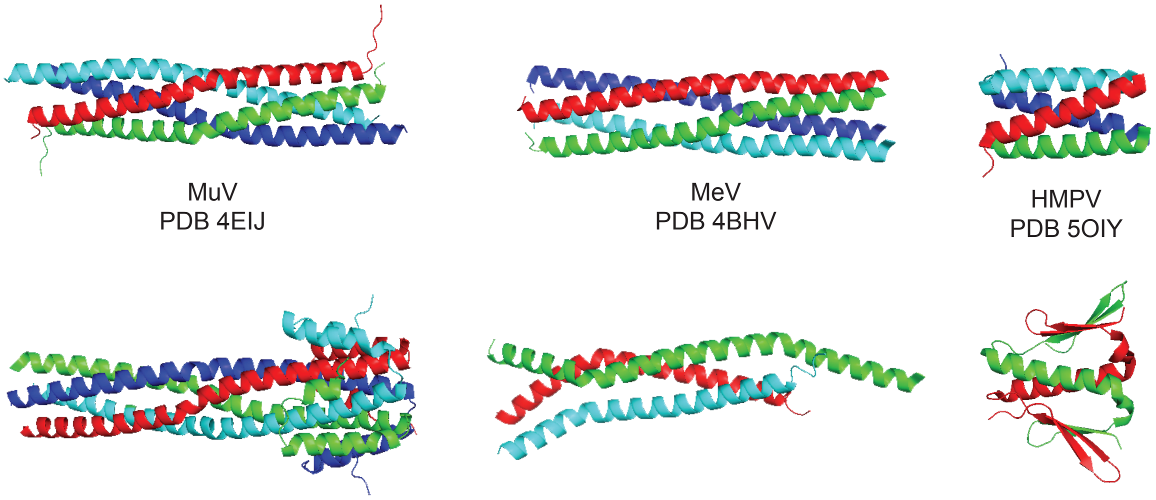
MuV
PDB 3BBZ

MeV
PDB 1OKS

HeV
PDB 4HEO

SeV
PDB 1R4G

Figure 1.16. Conserved three-helix bundle structures of *Paramyxoviridae* phosphoprotein C-terminal domain (XD), that starts from blue N- and ends by red C-termini. SeV XD is an NMR-solved structure, only one conformation from ensemble is visualised.



MuV
PDB 4EIJ

MeV
PDB 4BHV

HMPV
PDB 5OIY

NiV
PDB 4GJW

MARV
PDB 5TOH

VSV
PDB 2FQM

Figure 1.17. Crystal structures of *Mononegavirales* phosphoprotein (or polymerase cofactor analogues) oligomerisation domains. All ODs have parallel coiled coil orientation except MuV which is an antiparallel dimer of parallel dimers.

Phosphoprotein is a self-associated oligomer. For all known *Mononegavirales*, the OD is represented by parallel orientation of two, three or four monomeric units bound to each other by hydrophobic surfaces (Figure 1.17). The oligomerisation mode is different and unique for MuV: a pair of parallel α-helices interacts with a second one in the opposite direction by making 2 NTD and 2 CTD extremities on both sides of the P protein [Cox et al., 2013; Pickar et al., 2015]. This finding is in agreement with the tendency of MuV P N- and C-terminal domains to interact with NC. The proposed mechanism is based on the hypothesis that under viral genome replication, P NTD and CTD act together by inducing changes in NC. MuV P_{CTD} (XD) specifically recognises NC, while P_{NTD} uncoils NC by disrupting N-N interactions and by opening the “gate” to the RNA for viral polymerase. In addition, uncoiling of NC may facilitate and enhance viral RNA synthesis, as was suggested from minigenome activity in the presence of an excess of isolated P_{NTD} [Cox et al., 2014].

Rich in threonines and serines, especially in the disordered regions, P protein was found to be heavily phosphorylated in infected cells. In general, *Paramyxoviridae* protein phosphorylation plays a major role in regulation of viral RNA synthesis [Fuentes et al., 2010]. Post-translational modification occurs with the help of host kinases: casein kinase II (CKII) for RSV and MeV P [Villanueva et al;, 1994; Das et al., 1995], protein kinase C isoform zeta (PKC-ζ) for HPIV3 and SeV P [De et al;, 1995; Huntley et al., 1997], Polo-like kinase 1 (PLK1) for PIV5 P [Sun et al., 2009]. Having structural similarities with the sequentially closest PIV5, MuV P phosphorylation with PLK1 was tested [Pickar et al., 2015]. The authors found that interaction with N enhances P phosphorylation which involves residues S292 and S294 located in P_{CTD} linker; the binding site for associated kinase was identified to be in P_{NTD} region (T147). Overall phosphorylation decreases MuV viral gene expression.

V protein

In addition to the unfolded part of P N-terminal region, *Paramyxoviridae* V protein has a conserved C-terminal region composed of 7 cysteins and 1 tryptophane that bind two Zn^{2+} ions and forms a zinc-finger domain (ZFD) (Figure 1.18A, B).

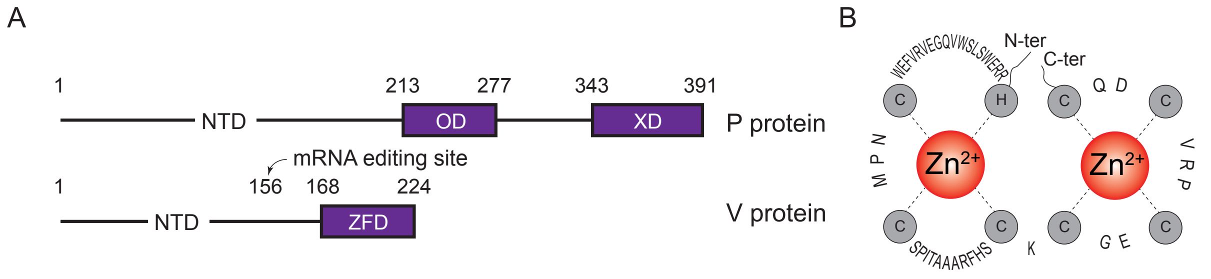


Figure 1.18. Schematic representation of MuV V and P structural domains (A) and ZFD (B) which consists of two Zn-cations (in red) coordinated with 1 histidine H and 7 cysteins C (in grey), other residues are shown in one letter code.

V protein plays an important role in viral pathogenesis by supressing host immune response. It blocks several cellular signalling pathways early after infection. V antagonises the antiviral activity of MDA5 (melanoma differentiation-associated gene 5) which detects viral cytosolic RNA and induces the signalling cascade of interferon production [Andrejeva et al., 2004]. The ZFD of PIV5 V in complex with MDA5 was recently crystallised [Motz et al., 2013].

V interacts with STAT1/STAT2 (signal transducer and activator of transcription), host transcription factors for α/β interferon gene expression, and mediates their degradation [Rosas-Murrieta et al., 2010; Caignard et al., 2009]. Targeting DDB1 (DNA damage-binding protein 1) is the way V protein induces STAT2 ubiquitination. It is interesting that despite having a disordered region on its N-terminal, full-length PIV5 V protein was crystalized with DDB1 in order to hijack the ubiquitin ligase complex [Li et al., 2006]. This result leads to the hypothesis that having already a structured cysteine-rich domain, V_{NTD} could be entirely folded upon the interaction with viral and host cell partners.

As about interaction with viral proteins, association of V and N was described for PIV5 and SeV [Randall & Bermingham, 1996; Horikami et al., 1996]. The ability of PIV5 V_{NTD} to chaperone N in a monomeric N^o complex that is more stable than with P, might be required for RNA transcription inhibition by regulating a switch between the viral transcription and replication which will be described further. V_{ZFD} also binds to N and serves to inhibit RNA synthesis by an unclear mechanism [Yang et al., 2015].

I protein

The function of the I protein, also known as W and D, is not clearly defined. I comprises on average less than 5% of total P/V/I transcripts. Practically, I protein is a truncated P_{NTD} protein (for MuV I is 171 amino acids long with its first 156 residues shared with P) and in theory, it can be considered to potentially interact with N and together with P and V control viral RNA synthesis.

1.2.6. Large protein

The viral polymerase L is the largest protein among other viral structural units (molecular mass of about 250 kDa), it is found in low amounts inside the virions. For the moment, no *Paramyxoviridae* L structure is available. However, the structure of polymerase of vesicular stomatitis virus (VSV) with fragment of P, was resolved recently [Liang et al., 2015] (Figure 1.19A). The viral polymerase protein sequence is mostly conserved among *Mononegavirales* and is structurally composed of six functional domains (Figure 1.19B). The first three domains (RdRp region) form a catalytic core. In particular, domain II is involved in RNA binding by positively charged regions, domain III has a conserved active site GDNQ motif for RNA polymerisation. Domains IV and V (capping region) participate in polyadenylation and capping of mRNA. Methyl-transferase function is performed by domain VI and partially by C-terminal domain (CTD). Connector between V and VI domains might be responsible for structural reconfiguration [Liang et al., 2015; Fearns & Plemper, 2017]. Binding of polymerase to phosphoprotein, which actually constitutes the polymerase complex, serves as a bridge to link the nucleocapsid and L, A more detailed role of polymerase in viral genome transcription and replication will be discussed later.

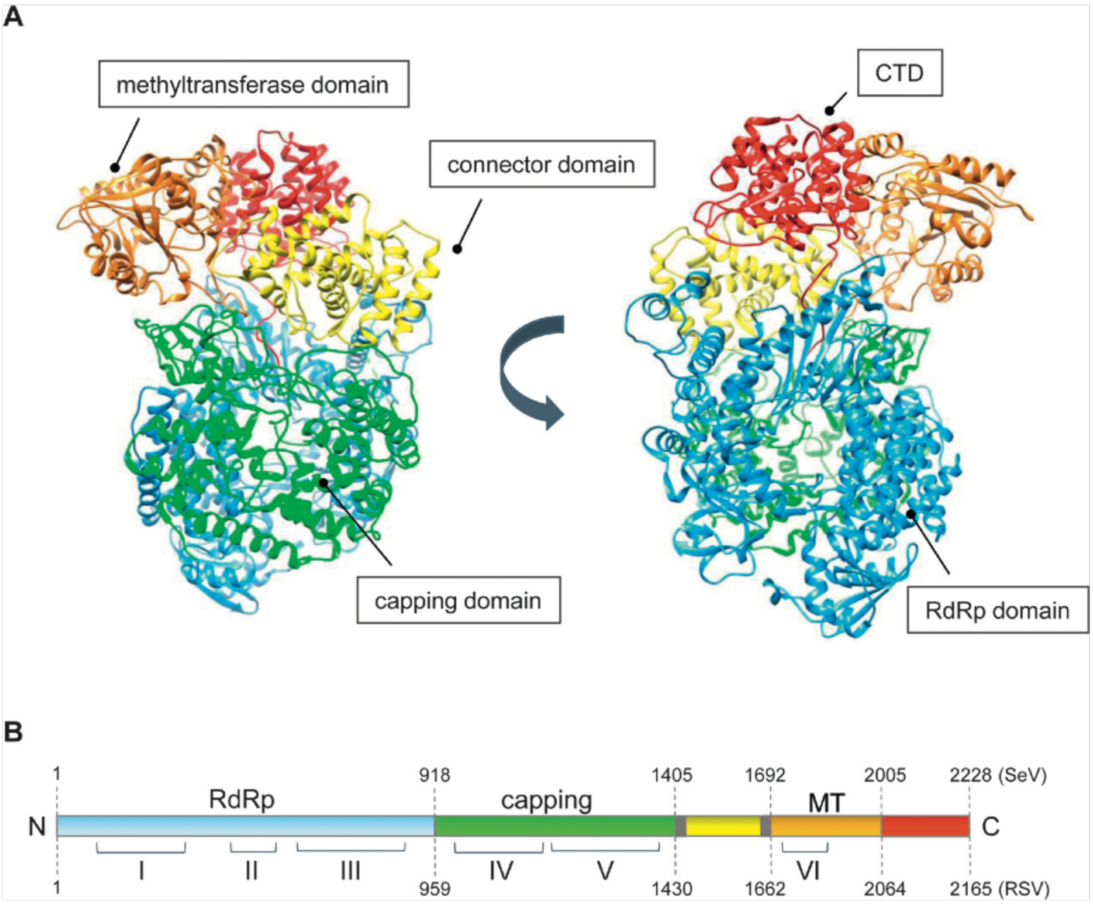


Figure 1.19. Structure of VSV polymerase [PDB 5A22] with identified domains in blue, green, yellow, orange and red (A). Linear representation of polymerase sequence for two viral species RSV and SeV from *Pneumoviridae* and *Paramyxoviridae* families respectively showing the amino acids boundaries of the domains (B) [Fearns & Plemper, 2017].

1.2.7. Other viral proteins

Glycoproteins

Being an important antigenic species that is detected by the host immune system, for the moment, viral glycoproteins are the most intensively studied among *Paramyxoviridae* viral proteins. The HN glycoprotein (also known as H or G) is responsible for the virus attachment to the host cell surface receptors. It appears as a homotypic dimer and is classified as a type II integral membrane protein. HN has a propeller-like head fold with receptor-binding ability and the enzymatic activity is carried out by the C-terminal part of the protein, while the N-terminus is a transmembrane part with cytoplasmic tails. After attachment, another glycoprotein named F protein mediates the viral envelope fusion with the host cell membrane. This process is pH independent and is one of the crucial criteria for *Paramyxoviridae* family classification. Unlike HN, the homotrimeric fusion protein is a type I integral membrane protein, its C-terminal part is placed inside the plasma membrane. The long external N-terminal part consists of two subunits connected to each other by a disulphide bridge which is cleaved under fusion of membranes and virion entry to the host cell. Characteristic for rubulavirus, SH protein is supposed to block virus induced apoptosis by forming ion channels in the membrane [Cox & Plemper, 2017].

Matrix protein

In virions, M multimerises and creates grids which interact with lipid membranes and cytoplasmic tails of HN and F proteins. At the same time, M dimer associates with nucleocapsids through positively charged regions [Battisti et al., 2012]. For several *Paramyxoviridae*, transient expression of M resulted in formation of empty virus-like particles (VLPs) [Harrison et al., 2010]. However, in the case of rubulavirus, in particular, MuV, VLP assembly can occur only co-expressing M, N and glycoproteins [Schmitt et al., 2009]. It is believed that budding and viral particle release is triggered by MuV M-N interaction. Several crystal structures of dimeric M are available in the PDB database: HeV (PDB 6BK6), NDV (PDB 4G1G), RSV (PDB 4D4T) and VSV (PDB 1LG7). All of them have similar structural properties.

1.2.8. Viral life cycle

To summarise the way paramyxoviral proteins act together under infection of a host cell, a detailed description of viral life cycle will be provided in this subchapter which was written with the help of Acheson, 2011; Knipe & Howley, 2013; Ortín & Martín-Benito, 2015, Fearns & Plemper, 2017.

After attachment of the virus HN protein to cellular receptors, the viral membrane fuses with the host cell membrane with the help of the F protein that goes through a conformational change on its coiled coil domain and brings the two membranes together. Due to different pH levels inside the virion and the cell, M-N contacts disrupt which leads to the release of the nucleocapsid and P and L proteins into the cytoplasm where all processes of viral gene expression occur (Figure 1.20). As all the processes of gene synthesis occurs in the host cell cytoplasm, the virion polymerase uses cellular ribonucleoside triphosphates. It is assumed that *Mononegavirales* genome is not infectious by itself, only when it is encapsidated by N and associated with P and L.

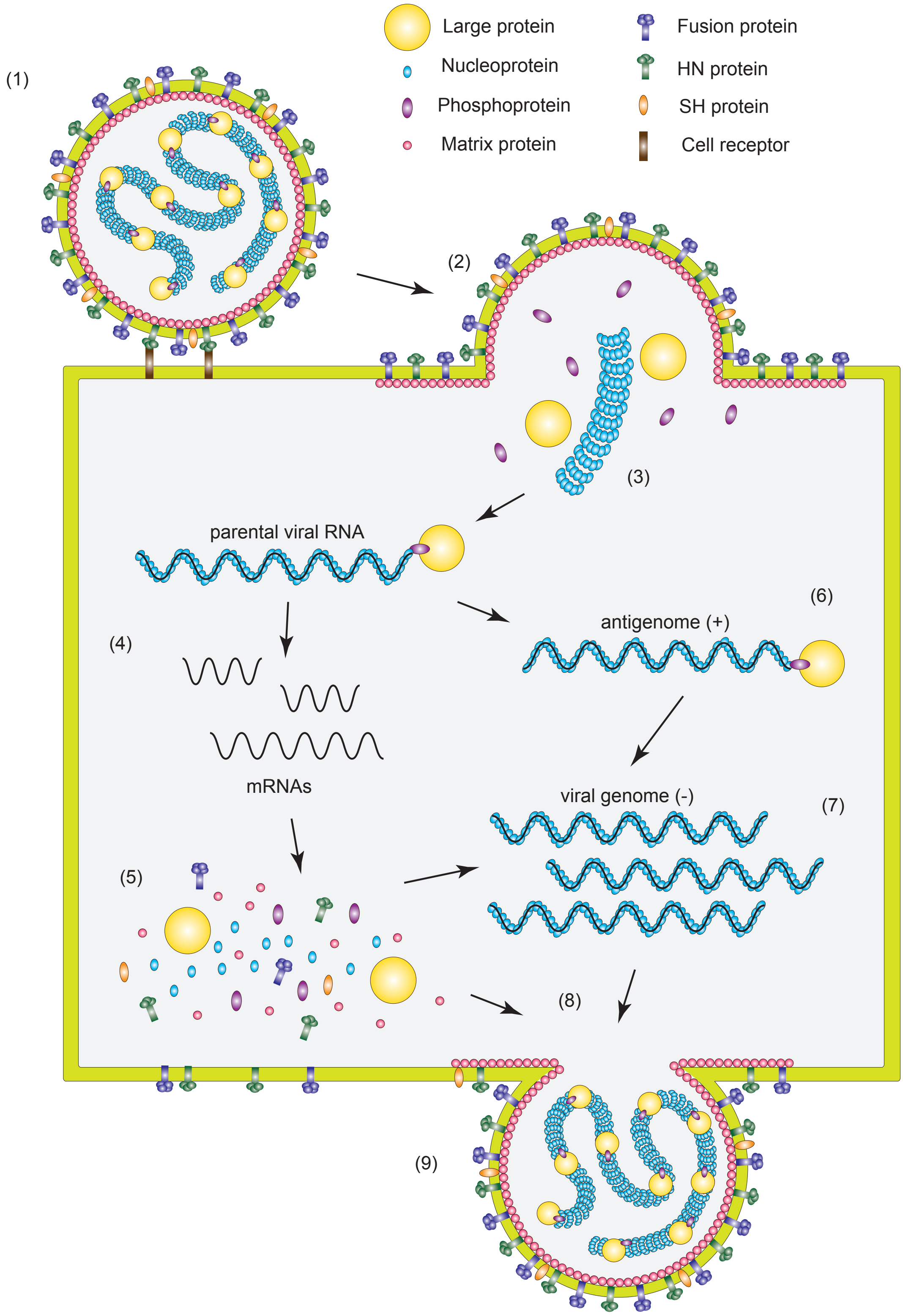


Figure 1.20. Schematic representation of *Paramyxoviridae* virus life cycle (example of MuV) that occurs in host cell cytoplasm and starts from virion attachment to cell receptors (1), its fusion with the cellular membrane (2) and release of the nucleocapsid (3). With the help of viral polymerase, the encapsidated parental RNA is transcribed into mRNAs (4) in order to allow for translation (5) of viral proteins necessary for further replication (6). HN, F and SH

proteins after translation are transported from endoplasmic reticulum to the Golgi apparatus and after to the plasma membrane (not shown). Primary RNA is first replicated to the antigenome (6) and then copied to the negative strands RNA copies (7). Resulting encapsidated genome and proteins are assembled to the virion (8) which buds from cellular membrane (9).

The viral genome has negative polarity, in order to be multiplied, it needs to be first copied into the positive complementary antigenome nucleotide sequence. But first, parental RNA is transcribed into the mRNAs for the corresponding protein genes. After accumulation of the necessary viral proteins, genome amplification can be started.

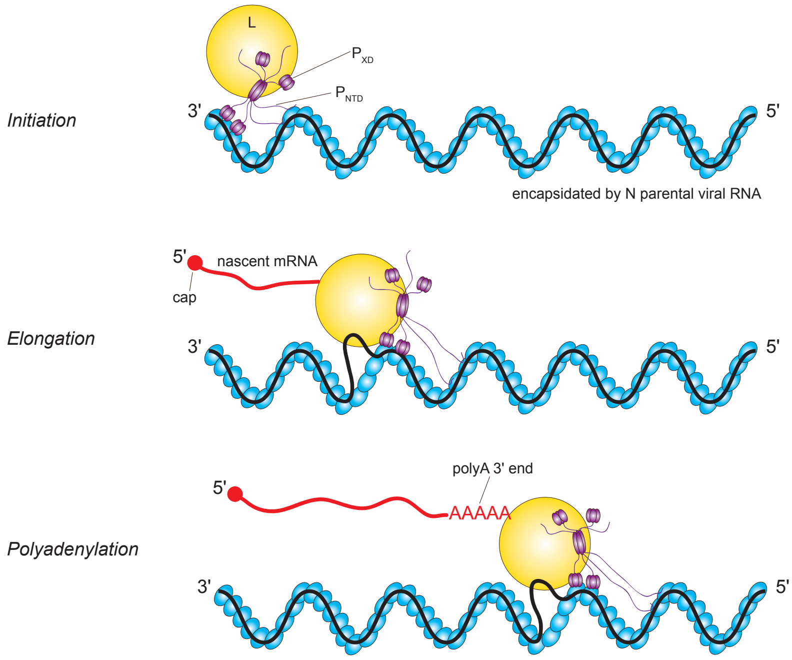


Figure 1.21. Proposed model of viral transcription for MuV. The initiation occurs under the polymerase recognition of NC by phosphoprotein XD and its further uncoiling by disordered NTD. Elongation starts by adding methylated cap to the nascent mRNA and ends by polyadenine sequence. After transcribing the first gene, the polymerase can continue sequester the next gene or switch to the replication mode.

Transcription begins from polymerase template recognition following by generation of initiation complex at 3′ leader region of viral RNA. L in cooperation with P transcribes a short region of 40-60 bp, after termination releases an uncapped leader RNA and starts mRNA synthesis of first protein gene. L adds methylation G cap to the 5′ end of the growing mRNA, elongates it and reaches the polyadenylation site in 3′ end which is post-transcriptionally generated from complementary polyuracil part. The same process is repeated for the second

RNA region in the sequence of protein genes. The mRNAs produced in this way are not encapsidated by N. After release of mRNA the polymerase can dissociate from the substrate and associate back to the leader. This is why, more proteins are produced close to the leader region of parental RNA: N, P and M, rather than L, gene of which lies last on viral genome sequence (Figure 1.21).

The switch between transcription and replication is regulated by nucleoprotein concentration, as N is needed to envelope all the products of RNA multiplication and avoid their annealing with viral mRNAs. During genomic replication, L changes the mode of synthesis, does not recognise termination signals and continues to make the full-length antigenome which is immediately covered with N with the help of the earlier described $N^{o}P$ complex. The next step is L and P association with the trailer end of the positive strand RNA and its further replication into negative strand viral RNAs, which are then encapsidated by N (Figure 1.22). It is important to note, that the first 6 residues on the 5′ antigenome and 5′ genome RNA are detected by N which upon interaction with them induces assembly of other N-units in NC without specific RNA sequence recognition.

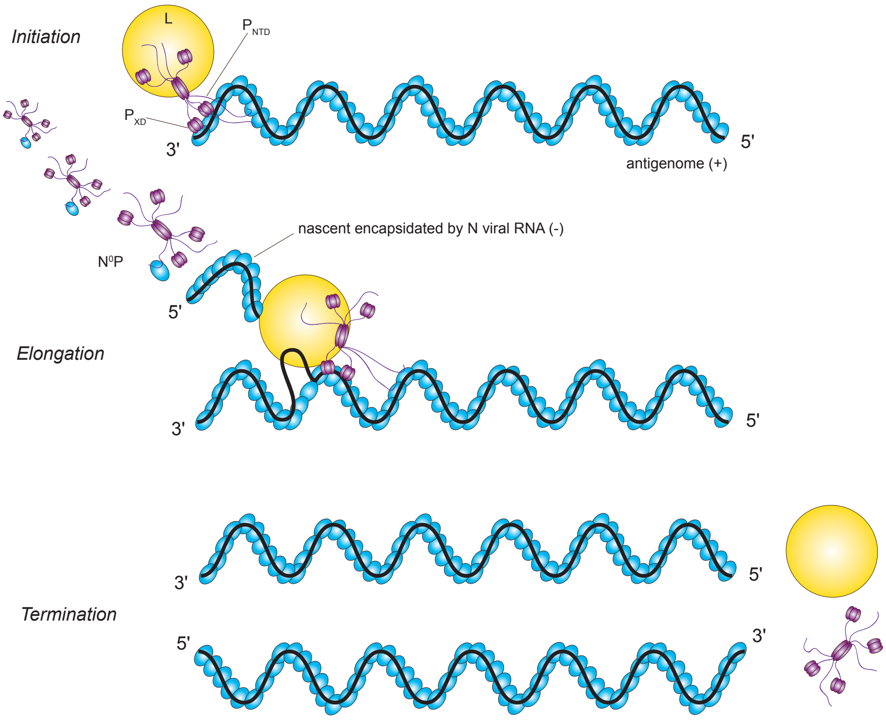


Figure 1.22. Model of viral genome amplification after generation of antigenome for MuV. Replication can be possible when there are enough produced nucleoproteins chaperoned by phosphoprotein ($N^{o}Ps$). Elongation occurs with the help of $N^{o}P$ and results in formation of negative strand RNA.

During synthesis, envelope proteins are inserted into the membrane of endoplasmic reticulum and are transported via the Golgi apparatus to the cell membrane. After replication, newly synthetized nucleocapsid also migrates to the plasma membrane and interacts with M which associates with HN and F tails. When the whole complex of proteins is formed, the virion starts budding from the membrane (Figure 1.20). The effect of interferon inhibition by V can occur during the viral RNA replication process.

Understanding the behaviour of the virus inside the human cell is important for development of antiviral drugs. Therefore, each step of viral life cycle envolving host and viral proteins can be considered as a potential therapeutic target. The nature and mechanisms of antivirals ranges from small molecules to short peptides and RNAs, that either block the protein conformation or perturb the interaction substrate. Targeting proteins could not be a great idea due to continuous mutagenesis [Aguilar & Lee, 2012]. However, immunostimulants as a part of host-directed therapy are now widely used during infectious diseases outbreaks.

The most promising inhibition approach was made for viral polymerase catalytic activities [Fearns & Plemper, 2017]. Nucleoside analogue inhibitors, especially for RSV and VSV treatment, act as chain terminators during viral replication [Aljabr et al., 2015]. Non-nucleoside antivirals are mostly based on binding to N and perturbing viral replication complex formation [Challa et al., 2015]. Therefore, intrinsically disordered protein regions (IDRs) of P and N can be also considered for drug development strategies.

1.3. Intrinsically disordered proteins

Proposed about 130 years ago by Emil Fischer, the concept describing the “lock-and-key” mechanism of enzyme catalysis is not applicable to intrinsically disordered proteins (IDPs), which became the subject of scientific interest in the last two decades. IDPs and IDRs do not adopt a stable secondary and tertiary structure under physiological conditions and in the absence of an interaction partner, they are usually described as an ensemble of several structural conformations. The amino acid composition determines their inability to fold. However, they modulate interactions with other molecules [Tompa, 2002; Dunker et al., 2002].

IDPs are predicted to compose one third of eukaryotic proteomes, and were found to play a crucial role in cellular signalling and regulation processes which were linked to the development of neurogenerative disorders, several cancers and type II diabetes [Dyson & Wright, 2005; Uversky, 2013].

The absence of stable structure can be viewed as an important advantage of IDPs. The ability to bind several partners, opens the opportunity for certain multifunctionality enabling several interaction sites along the polypeptide chain. It is also common for IDPs to contain molecular recognition sites and undergo disorder-order transition upon binding. However, it is also suggested, that upon interaction IDPs form so called "fuzzy complexes" by remaining mostly unfolded in bound state. In addition, IDPs are able to self-associate, sometimes in the presence of nucleic acids, and form separate protein-dense droplets. These properties are widely used in viral transcription-replication machineries: the organelles are also called varial factories, or Negri bodies, well studied for RABV [Lahaye et al., 2009; Nikolic et al., 2017]. IDPs are accessible for PTM enzymes which can induce protein structural changes, their stabilisation and destabilisation [Theillet et al., 2012].

Of course, due to IDP heterogeneous and fluctuating structure, classical drug discovery methods cannot be used for targeting them, as they almost do not reach nanomolar affinities with therapeutic agents which is one of the main requirements for drug licensure [Tsafou et al., 2018]. Therefore, the design of the drug needs to be changed for these challenging systems, but this would be possible only after clear understanding of the nature of dynamic conformations describing IDPs behaviour.

1.3.1. Characterisation of IDPs

The convenient way to represent protein conformations is to use free energy landscapes (Figure 1.23). In statistical mechanics, for folded proteins, the most populated conformational (native) state corresponds to the minimum which can often be determined experimentally (by crystallization or NMR, for example), resulting in its three dimensional coordinates being made available in the PDB. In IDPs with a large number of local minima, conformations are described by their statistical weight and also by transition rates between these conformations, they can be represented as structural ensembles. Some states may be undetectable by NMR because of their low population below the sensitivity of a given experiment. [Felli & Pieratelli, 2015].

In order to provide structural and functional interrogation of IDPs, theoretical (bioinformatic) and experimental methods are usually combined. Information about protein covalent bond lengths, dihedral angles and rotameric states of side chains can be collected through the use of force fields of molecular dynamics (MD) simulations or protein structure databases. Protein sequence based tools can predict IDRs from disorder signature algorithms, for example, IUPred, DisEMBL, PONDR, DisPROT, AUCpreD, RONN [Dosztányi et al., 2005; Linding et al., 2003; Todorova, 2012; O-Piovesan et al., 2017; Wang et al., 2016; Yang et al., 2005]. The predictors can use different methods but the principle of disorder prediction is based on the observation that IDRs have specific amino acid sequences designated for disorder and the capacity to form stabilizing contacts. For example, protein regions with a large number of highly polar residues, low hydrophobicity, low aromaticity, rich in prolines and overall do not show sequence complexity. These characteristics suggest proteins to adopt random coil conformations.

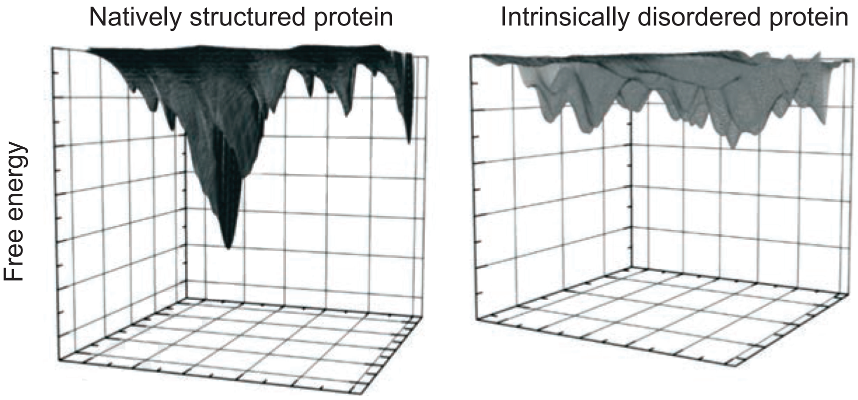


Figure 1.23. Examples of free energy landscapes for folded and unfolded proteins [Howton et al., 2016].

Because IDPs have fluid conformational properties, their behaviour is best described using solution state techniques. Experimental data can be derived from a number of biophysical methods such as nuclear magnetic resonance (NMR), small angle X-ray scattering (SAXS), single-molecule fluorescence Förster resonance energy transfer (smFRET), surface plasmon resonance (SPR), circular dichroism (CD), ion-mobility mass spectrometry (IM-MS).

1.3.2. IDRs in viral proteins and objects of this work

The tendency for conformational disorder in *Paramyxoviridae* proteins can be correlated with extreme pathogenicity of these viruses. Under evolution, they elaborate the production of multifunctional and adaptable proteins, which are essential for the effective and optimised amplification of viral genetic material. In most cases, IDRs are alternated with folded regions which probably are able to modulate conformational preferences of flexible parts. Despite the structural complexity and challenging characterisation, viral IDPs constitute a good system that could be used in studying functional mechanisms of disordered proteins of other species and systems.

As mentioned before, *Paramyxoviridae* P and N are composed of structured and unstructured regions. For the moment, only several small regions have established functional regions, the role of other parts in the polypeptide sequence are unknown. However, recent scientific findings revealed the existence of conserved small transient helices in RSV P_{NTD} being in long-range contacts between each other, they can be recognised as potential interaction sites [Pereira et al., 2017]. Referring to [Milles et al., 2018], the short motif HELL conserved among *Paramyxoviridae* and situated in the middle of P_{NTD} sequence was found to bind to the N-terminal lobe of monomeric Ncore. The interesting feature of this interaction is that although it was characterised as ultra-weak, the mutation to AAAA abolished the viral transcription in minigenome activity test by limiting the interaction between P and N which is essential for the viral RNA synthesis machinery.

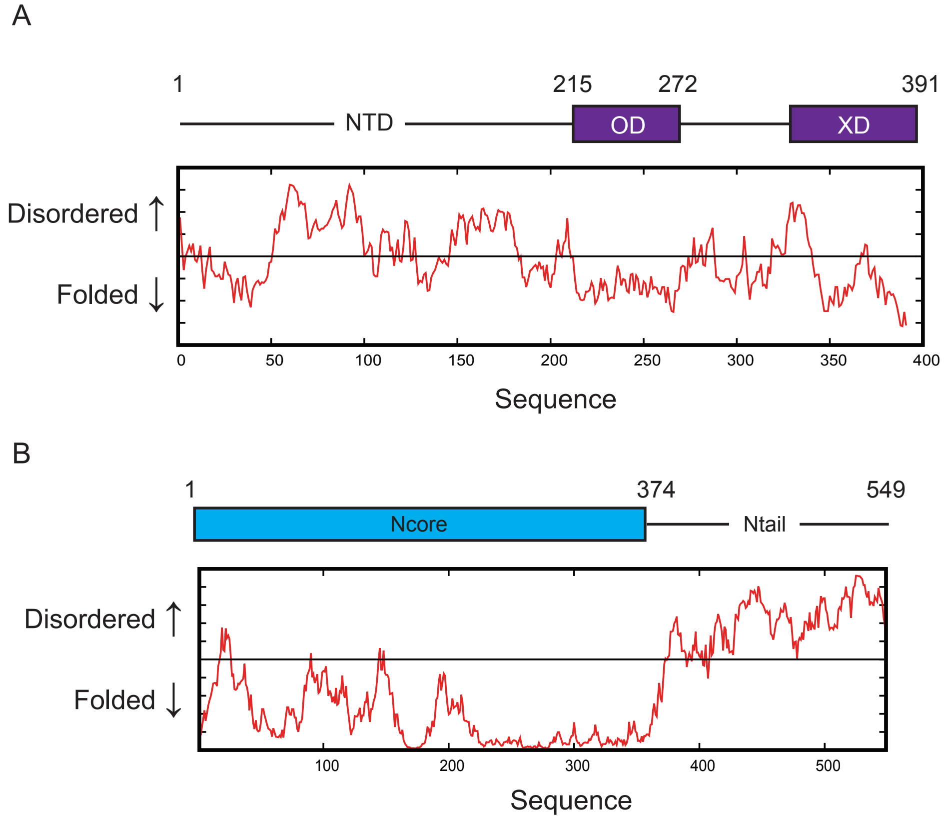


Figure 1.24. IUPred disorder prediction for MuV phosphoprotein (A) and nucleoprotein (B). Large unfolded regions are present in P_{NTD}, between OD and XD and in Ntail regions.

As for structural organisation of MuV P and N, apart from crystal structures of P folded domains (OD and XD), IDRs remain unstudied. Several protease digestion assays and disorder prediction software suggest that the disordered P_{NTD} region comprises first 215 amino acid residues, the OD is located roughly between amino acids 215 and 272, followed P_{CTD} region with unfolded linker and XD (343-391) (Figure 1.24A). The short N-terminal region of P is believed to be sufficient for N chaperoning in the monomeric state, although this has never been shown for MuV. No findings are available for MuV N^oP. Also it is remains unclear why

P_{XD} crystallisation failed and required a stabilising agent for three dimensional folding. Information is missing also from the long disordered linker (272-342) connecting OD and XD of phosphoprotein.

For the moment, although MuV N has amino acid sequences that are similar to those of PIV5 (52% identity mostly in Ncore region), there are some differences, in particular, the sequence of Ntail is not conserved and has a different amino acid length. Consequently, MuV nucleocapsids can still have different structural characteritics, so it appears essential. To characterize them in detail. As mentioned before, MuV Ntail does not interact with P_{XD} as, for example, MeV, SeV, HeV that have MoRE interaction site, the existence and interactions of which have been characterized in detail, for example by NMR. There is no available structural information for MuV N in NC and N^oP and the function of disordered Ntail (Figure 1.24B). No clear and obvious research was carried out concerning post-translational modification of N and P, particularly, of their disordered regions.

In this work, I will perform the first studies of MuV P and N disordered and structured regions and discuss the idea of their functional relevance during viral life cycle. I will try to answer the following questions that can be defined as goals of this project:

- What is the role of conformational disorder in MuV viral proteins N and P? In particular, what is the role of MuV Ntail and what are the potential partners it interacts with? Does this domain interact with P, as is the case for other *Paramyxoviridae*, and if so how?
- What are the dynamic properties of the disordered N-terminal domain of MuV P? What is the functional role of this domain and how does it interact with other viral proteins during the replication cycle?
- How does phosphorylation of N and P influence the protein structure, dynamics and interactions of the domains, and the interactions between MuV N and P?
- What is the role of the disordered linker between OD and XD of phosphoprotein? Can the apparent interaction between P_{XD} and Ncore be described at atomic resolution?
- Is it possible to study MuV NCs with RNA issued from bacterial cell expression? Is it possible to develop optimal protocols for MuV RLPs purification, and to investigate the structure of N-subunit in oligomeric NC?
- Can a peptide issued from P_{NTD} stabilise N in the monomeric state in N^oP complex? If yes, what are the conditions to achieve NC assembly from N^oP *in vitro*? Can the structure of MuV N^oP be characterized?

These problems and other challenges will appear along the project realisation. In the next chapters I will describe methods and materials I have used to get a clear illustration of MuV proteins structure and dynamics, and further important results and conclusions.

CHAPTER 2.

MATERIALS AND METHODS. NUCLEAR MAGNETIC RESONANCE

2. NUCLEAR MAGNETIC RESONANCE

Nuclear magnetic resonance (NMR) spectroscopy is one of the most essential techniques used in this project. It allows not only to get atomic resolution structural information, it is also crucial for the study of the molecular dynamics and flexibility on time scales ranging from picoseconds (ps) to seconds (s). As this manuscript is mostly focused on protein intrinsically disordered regions, the use of NMR, offering residue-specific resolution coupled with other techniques, is the most suitable method to describe their conformational behaviour, and to attempt to understand their function.

2.1. General outlines

2.1.1. Phenomenon

An NMR experiment is based on placing an NMR active nuclear spin in a static magnetic field, thus generating spin polarization, and on its subsequent perturbation by application of electromagnetic radio-frequency (RF) pulses.

Not all atomic nuclei can be studied by NMR. To identify which nucleus is active for NMR, firstly, we need to refer to atomic quantum numbers, in particular, to the spin number I which is dependent on the number of protons and neutrons of the nucleus $^{protons+neutrons}X_{protons}$. Only atoms having an odd number of protons and neutrons and any number of protons with $I = 1/2$ are easily observed by NMR such as $^{1}H_{1}$ (proton), $^{13}C_{6}$, $^{15}N_{7}$, $^{31}P_{15}$. These nuclei are commonly used in biomolecular NMR. Nuclei with $^{even}X_{odd}$ having $I = 1$ (or more generally $I > 1/2$) are considered to be quadrupolar nuclei, for example, $^{2}H_{1}$ (deuterium), whose detection is complicated by spectral line broadening. Another characteristic nuclear property is its natural abundance. Table 2.1 represents the isotope distribution of the elements found in biological molecules most commonly studied by NMR. Due to the low natural abundance of ^{13}C and ^{15}N in biological samples, investigation of a particular protein requires production in media enriched with ^{13}C and ^{15}N isotopes.

In the presence of a static magnetic field B_o, the $I = 1/2$ nucleus gives rise to two energy levels represented by the Zeeman splitting (Figure 2.1A), and characterised by the spin quantum number m_S of +½ (α state) and -½ (β state). In general, the number of states is given by $2I + 1$. In such manner, for quadrupolar nuclei $^{2}H_{1}$ with, for example, $I = 1$, the energy splitting will lead to the appearance of three energy levels.

Table 2.1. The most used nuclear isotopes in biomolecular NMR, their natural abundance and gyromagnetic ratios.

Nucleus	*Natural abundance, %*	*Gyromagnetic ratio γ, 10^7 rad s^{-1} T^{-1}*	*Nucleus*	*Natural abundance, %*	*Gyromagnetic ratio γ, 10^7 rad s^{-1} T^{-1}*
^1H	99,985	26,751	^{13}C	1,07	6,728
^2H	0,015	4,106	^{14}N	99,632	1,933
^{12}C	98,93	-	^{15}N	0,368	-2,711

Coming back to the two states, at thermal equilibrium, they are degenerate, and therefore, they have the same energy and eqial population. Interaction of nuclear spins with B_0 leads to the difference between two energy levels and in their population N_α and N_β (Figure 2.1B). As the lower energy α state is more populated (this population difference is very small, about 1 in 10.000 for ^1H in 11.74 T magnetic field), it results in a net magnetization M_0 oriented in the direction of the applied B_0 field (Figure 2.1B). It is important to note that here we do not talk about single spin, about net magnetisation instead, which is over all the spins in the ensemble.

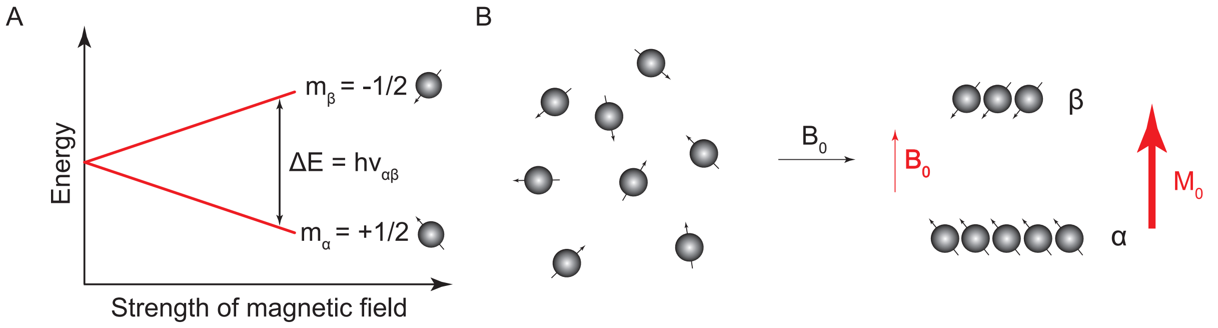


Figure 2.1. Zeeman splitting of energy levels for a nucleus with spin number ½ (A). In the presence of B_0, α and β states become different in magnetic moments population (number of spins) according to the Boltzmann distribution (B). The small population difference gives the net magnetization M_0 in the direction of B_0. The higher the field used, the greater will be the population difference between states and M_0.

According to the selection rule (possibility of system transition from one quantum state to another), only energy transitions having a Δm_S equal to ±1 are allowed (equation 2.1). This makes it possible to extract the frequency of transition $\nu_{\alpha\beta}$ from the Planck relation which depends on m:

$$\Delta m_S = m_\alpha - m_\beta = \pm 1 \tag{2.1}$$

$$\Delta E = h(m_\alpha \nu_\alpha - m_\beta \nu_\beta) \qquad \nu_{\alpha\beta} = \nu_\alpha - \nu_\beta \tag{2.2}$$

Alternatively, the energy difference can be expressed in terms of the nuclear magnetic moment μ of the individual spin, which, in the presence of B_o, precesses with angular frequency ν_o called Larmor frequency:

$$\vec{\mu} = \hbar m \gamma \vec{I} \qquad E = -\vec{\mu}\vec{B}_0 = -\hbar m \gamma B_0 \qquad \Delta E = \hbar \gamma B_0 = \frac{h}{2\pi}\gamma B_0 \tag{2.3}$$

$$\Delta E = h\nu_{\alpha\beta} = \hbar \gamma B_0 \qquad \nu_0 = -\frac{1}{2\pi}\gamma B_0 \tag{2.4}$$

where γ is the gyromagnetic ratio which varies for each nucleus (Table 2.1) and defines the sense of spin precession, sensitivity, relaxation, strength of NMR-measurable interactions, h – Planck constant $6{,}62\times10^{-34}$ J×s, $\hbar$ – reduced Planck constant, equal to $h/2\pi$, $1/2\pi$ was used to convert rad/s in Hz. ν_o is negative as the sign indicates that the spin precesses around B_o in a clockwise direction.

The principle of the NMR experiment is based on the perturbation of M_o by applying an electromagnetic field with the frequency of the resonance ν_o and further detection of M_o while returning to equilibrium along the z-axis (Figure 2.2).

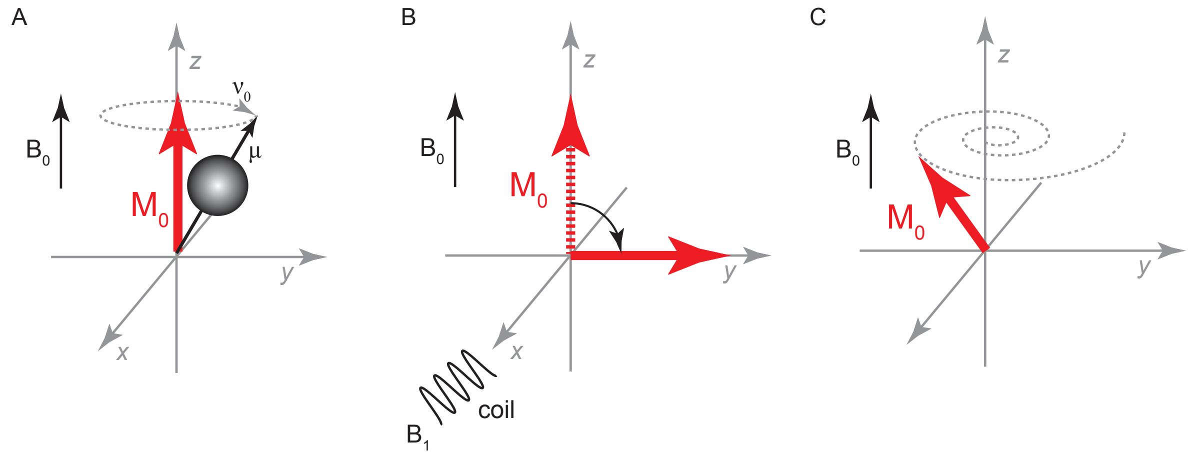


Figure 2.2. Vector model of magnetization M_o, and the individual magnetic moment precession at ν_o Larmor frequency of an isolated spin ½ (A). The application of an electromagnetic field B_1 (pulse) via the radio-frequency coil (designated along x-axis) generates the energy transition which corresponds to M_o tilt. The duration of the pulse determines the angle of M_o tilt (B), after M_o returns to the equilibrium (C) which is actually detected in an NMR experiment.

2.1.2. *Rotating frame*

The strong B_0 field supplied by a powerful superconducting magnet maintains M_0 along the *z*-axis, so the only way to move the vector from *z*-axis is to apply the small magnetic field B_1 along *x*-axis which is oscillating with the same (resonant) Larmor frequency ω_0 ($\omega_0 = 2\pi\nu_0$). In order to understand how the B_1 field works, we need to introduce the rotating frame, which moves with the same frequency $\omega_{rot.\,fr.}$ as the one of magnetization ω_0, by resulting the apparent frequency (offset) Ω which will be equal to zero:

$$\Omega = \omega_0 - \omega_{rot.\ fr.} \tag{2.5}$$

In the rotating frame, the apparent reduced field ΔB will become small and B_1 field will be dominant:

$$\Delta B = -\frac{\Omega}{\gamma} \tag{2.6}$$

From the perspective of the rotating frame, there are two magnetic fields being applied B_1 and ΔB_0. The resulting effective field B_{eff} can be simply calculated from the Pythagorian theorem (Figure 2.3). For practical reasons, here we work with the frequency units where ω_1 is the frequency of B_1 field (transmitter frequency):

$$B_{eff} = \sqrt{{B_1}^2 + \Delta B^2} \qquad \omega_{eff} = \sqrt{{\omega_1}^2 + \Omega^2} \tag{2.7}$$

So, to achieve B_{eff} equal to B_1 the transmitter frequency is applied and tilts magnetization at the angle θ = 90° (π/2) during a time *t*, which means to apply a pulse:

$$\omega_1 = \frac{\theta}{t} \tag{2.8}$$

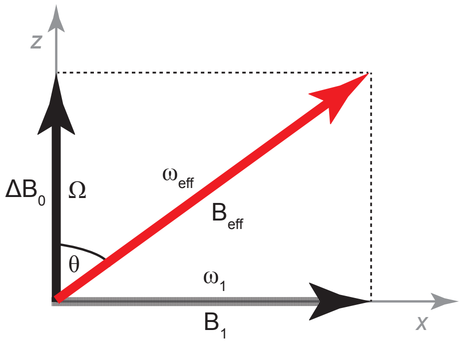


Figure 2.3. In the rotating frame the effective field B_{eff} is the vector sum of the reduced static field ΔB and the RF B_1 field. Practically, the magnetic field can be expressed in the frequency units.

2.1.3. Signal detection and Fourier transformation

Magnetization can be measured in the *xy*-plane. Generally, the coil detects the *x*- and *y*-component of M_o (Figure 2.4A), so M_X can be expressed by the simple geometry, where β - angle of tilted M_o towards the *x*-axis by M_x = M_o sin β, while M_y is equal to zero. After introducing the time component t with the frequency ω_o (in rad/s), the M_x and M_y dependency over time appears as (Figure 2.4B, C):

$$M_x = M_0 \sin\beta \cos(\omega_0 t) \qquad M_y = -M_0 \sin\beta \sin(\omega_0 t) \tag{2.9}$$

In the rotating frame system, ω_o is replaced by the offset frequency Ω:

$$M_x = M_0 \sin\beta \cos(\Omega t) \qquad M_y = M_0 \sin\beta \sin(\Omega t) \tag{2.10}$$

By assuming that β = 90° (hard pulse), the M_x and M_y components can be described by:

$$M_x = M_0 \cos(\Omega t) \qquad M_y = M_0 \sin(\Omega t) \tag{2.11}$$

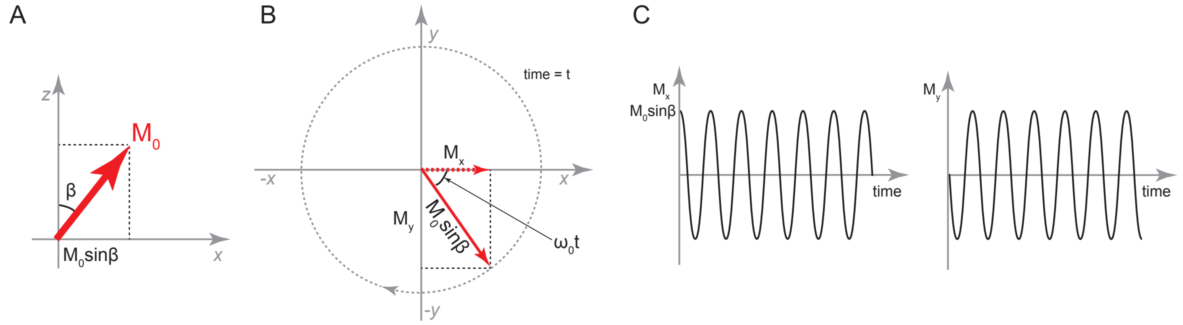


Figure 2.4. View of M_o tilt in *xz*-plane (A) and M_o oscillation over t in the *xy*-plane with trigonometric extraction of *x*- M_x and *y*-components M_y of M_o (B). The plot of the M_x and M_y showed as the sine and cosine functions of time (C).

Due to the action of spin relaxation, magnetization (signal) will decay away over time T_2 *(*$t = T_2$, $T_2 = 1/R_2$*)*. For this reason, it is called free induction decay (FID). The easier way to interpret the NMR data is to transform the time dependent FID into the frequency domain function. This is possible via the Fourier transformation (FT) (Figure 2.5).

If the signal $S(t)$ is proportional to the M_x and M_y precessing in the *xy*-plane of the rotating frame, it can be expressed as complex number:

$$S(t) = S_x(t) + iS_y(t) = S_0 \cos(\Omega t) + iS_0 \sin(\Omega t) = S_0 \exp(i\Omega t) \tag{2.12}$$

Resulting exponential decay with constant T_2, or rate $R_2 = 1/T_2$ will be equal to:

$$S(t) = S_0 \exp(i\Omega t) \exp\left(-\frac{t}{T_2}\right) = S_0 \exp(i\Omega t) \exp(-R_2 t) \tag{2.13}$$

Further Fourier transformation of *S(t)* to *S(ν)* leads to the absorption Lorentzian line shape (NMR spectrum peak):

$$S(t) \overset{FT}{\rightarrow} S(\nu) = \frac{R_2}{4(\pi\Delta\nu)^2 + R_2^2} \tag{2.14}$$

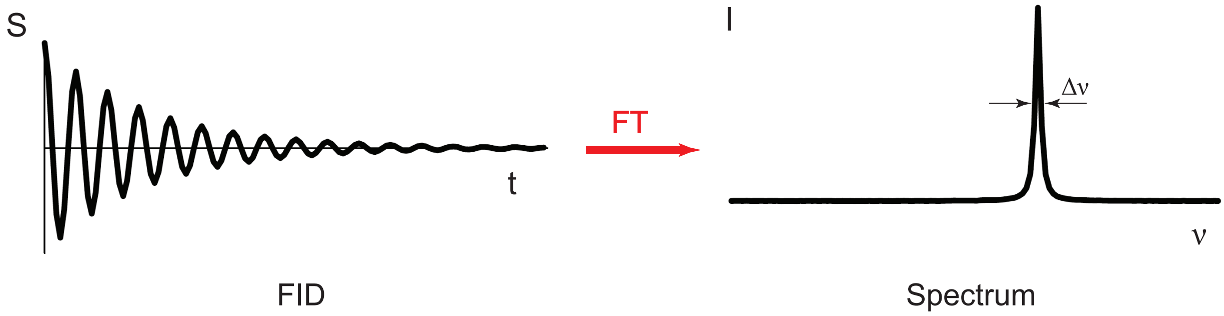


Figure 2.5. Fourier transformation of signal evolution *S* as a function of time *t* gives a spectrum peak which is a frequency *ν* domain function.

The shape of the peak is dependent of the homogeneity of the field (so it needs to be well calibrated and "shimmed"), as well as of T_2. More precisely half-height width $\Delta\nu$ of NMR line (Figure 2.5) is related to T_2 by the equation $\Delta\nu = 1/(\pi T_2)$ (Figure 2.5). However, other factors can contribute to the line broadening.

2.2. *NMR parameters*

2.2.1. *Chemical shift*

Chemical shifts (CSs) are the most frequently used parameters obtained from NMR spectroscopy for structural studies of IDPs. According to the previous description, the NMR spectrum of protein composed from active ^1H, ^{15}N and ^{13}C spins would only have 3 peaks with different Larmor frequency. However, in addition to the action of B_0 field, nuclear spins can be influenced by local magnetic fields induced by electronic magnetic moments currents circulating on the molecular orbitals. Thus, an atom in close proximity to another nucleus can perturb its NMR frequency which, consequently, is sensitive to the nuclear chemical environment.

Practically, from equation 2.4, it is clear, that the stronger the magnet used to generate B_0, the bigger sthe value of ν_0 will be. To exclude the instrument dependency, the chemical shift δ is introduced. δ is equal to the resonance frequency ν of the nucleus relative to the resonance frequency of a reference compound ν_{ref}, both measured in the same field:

$$\delta = 10^6 \times \frac{\nu - \nu_{ref}}{\nu_{ref}} \tag{2.15}$$

Usually, the values of δ are relatively small. Due to this they are expressed in parts per million (ppm). As ν_{ref} is often in MHz, while substraction ν - ν_{ref} in Hz, therefore, ratio Hz/MHz gives ppm.

Among the factors influencing nuclear CSs, shielding (increasing) and deshielding (decreasing), it is possible to distinguish:

- electronegativity (ability of an atom to attract a shared pair of electrons) of the neighbouring chemical groups will increase the δ value of the surrounding groups, for example, an atom of nitrogen decreases the electron density of the connected hydrogen through the hydrogen bond and shields its chemical shifts;
- anisotropy refers to the induced magnetic field generated by the circulation of electrons. Often the π-electrons from alkene, alkyne bonds, aromatics or any other conjugated structures at the molecule can have paramagnetic or diamagnetic effects depending of their orientation (parallel or opposite) relative to the applied magnetic field.

Therefore, chemical shifts are very sensitive to the electronic and anisotropic environment of the nucleus. Perturbation in CSs it is not only the change in covalent bonds but also non-covalent interactions with solvent or upon binding to a partner. A variety of the frequencies from all the protein nuclei can be used in the comparison to those protein CSs associated with structures of folded proteins that have been determined. This can help to predict protein secondary structure, as constraints for molecular structure calculation and sometimes to investigate protein dynamic properties.

Secondary chemical shifts

The presence of α-helix and β-sheet conformations can be extracted by comparing measured chemical shifts for a given residue with those theoretically predicted from complete random coil structures. This difference $\Delta\delta$ is called secondary chemical shifts (SCS): it is mostly used for carbon values, as ^{15}N and ^{1}H$_N$ chemical shifts are less informative due to their rigid positioning in peptide bond planes.

Negative secondary shifts for C$_\alpha$ correspond to β-sheet structures, while positive SCS are observed for α-helices, while for C$_\beta$ positive SCS are characteristic of protein α-structures and negative are characteristic of β-sheets. By reference to March *et al*, 2006, different SCSs are combined and re-compared with the chemical shifts for α- and β-structures (derived from reference data bases of CSs from resolved protein structures by NMR) resulting in the secondary structure propensity (SSP) score. In the case of IDPs, the SSP can be used to identify partially structured regions that can, in some cases, correlate with regions of the protein that are found to bind to partners.

Chemical shift anisotropy

The chemical shift also depends on the local magnetic fields induced by atomic electrons. The effect of the electron density on the total magnetic field experienced by the nucleus is correlated with the spin orientation relative to the external magnetic field. In general, the B_o field is reduced by shielding factor σ which can be described by a second rank tensor representing the different local field variations in three orthogonal directions:

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \tag{2.16}$$

Nevertheless, for liquid samples due to orientational averaging over the time, the chemical shift value δ is an average of different nucleus orientations. In the case when shielding is anisotropic, depending on the molecule reorientation over the time (therefore, impact to the relaxation), the spin resonance frequency ω_s will be affected by the local magnetic fields. Here we are talking about chemical shift anisotropy (CSA).

$$\delta = 1 - \sigma \qquad \nu = -\frac{1}{2\pi}\gamma(1-\sigma)B_0 \quad \text{or} \quad \omega_s = \gamma(1-\sigma)B_0 \tag{2.17}$$

Usually, CSA is defined from the chemical shift tensor, and in the principal axis system with axial symmetry appears as:

$$\Delta\delta = \delta_{\parallel} - \delta_{\perp} \tag{2.18}$$

where $\delta_{//}$ - tensor element parallel to the symmetry axis $\delta_{\perp} = \delta_{zz}$, $\delta_{//}$ - two equivalent perpendicular tensors $\delta_{\perp} = (\delta_{xx} + \delta_{yy})/2$.

2.2.2. Scalar and dipolar couplings

Besides chemical shift, structural information can be extracted from scalar and dipolar couplings between nuclear spins.

In the case of two or more covalently-bound spins, nuclear spins can experience scalar couplings. On the spectrum, the presence of the coupling constant gives rise to the appearance of multiplets (Figure 2.6A). *J*-coupling is independent on the external static field and is measured in Hz. The mechanism of the coupling is based on spin polarization. One nuclear magnetic moment of spin *I* polarizes the spins of another neighbouring atom electrons and influences the energy of the neighbouring nucleus *S*. The value of the *J*-coupling can be positive or negative, on Figure 2.6B the most important scalar coupling constants for biomolecular NMR analysis are shown.

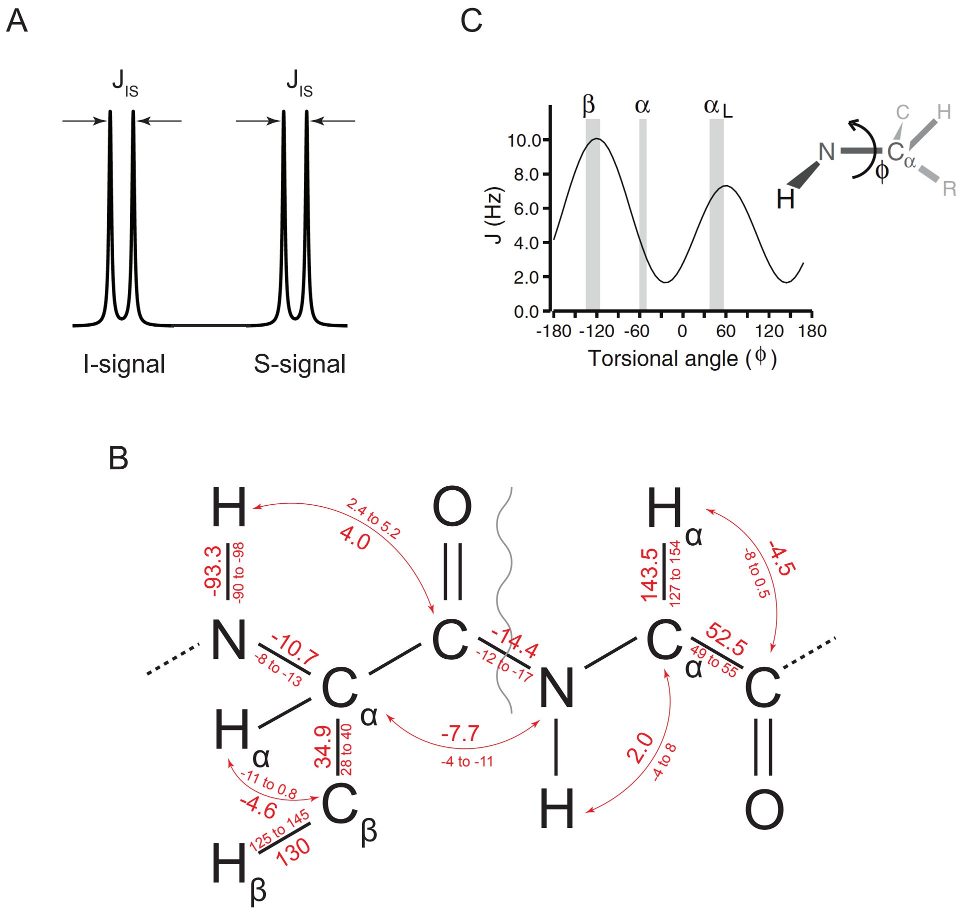


Figure 2.6. The *J* couplings between two spins *I* and *S* can be calculated from the peak splitting (multiplet) (A). Schematic representation of average one- and two-bond *J*-couplings in a polypeptide chain (B). The plot of the relationship of $^3J_{HN\text{-}H\alpha}$-coupling and φ torsion angle, grey bars indicate *J* and φ values for different protein secondary structures such as β-sheet (β) and α-helices (α and αL as left-handed helix) (C). Adapted from [Rule and Hitchens, 2006].

Multiple bond J-couplings can be affected by the conformation of interacting spins. For example, for the bond $H-N-C_\alpha-H_\alpha$ the range of ^3J-coupling constant is from 2 to 10 Hz, which has been shown to follow the Karplus relationship:

$$J = A\cos^2\phi + B\cos\phi + C \tag{2.19}$$

where *A, B, C* are empirical constants, φ is the angle between planes defined by $C-N-C_\alpha$ and $N-C_\alpha-C$ (torsional angle) (Figure 2.6C).

Scalar couplings are used for determination of small molecule structure, dihedral angles for secondary structure of folded and disordered (creation of structural ensembles) proteins and for transfer of magnetization from one to another coupled nucleus, which will be discussed further.

In contrast to scalar couplings, the dipolar coupling constant (*D*) is much larger and is observed in solids and very viscous liquids (anisotropic systems). In isotropically reorienting liquids *D* is efficiently averaged to zero. Dipolar coupling is based on the direct interaction between the spins *I* and *S* close to each other in space. Nuclear spin precession generates a magnetic field that disturbs the magnetic field of other spins. Consequently, B_{eff} of the spin *I* will be affected depending on the relative position of both magnetic dipoles. The dipolar coupling interaction can provide information about the distance between the spins r and orientation of the inter-atomic vector relative to the B_0 field (through θ angle):

$$D_{IS} = -\frac{\gamma_I \gamma_S}{8\pi^2 r_{IS}^3}\left[3\cos^2\theta\,(t) - 1\right] \tag{2.20}$$

2.2.3. Relaxation rates

Relaxation is the process of the return of the magnetization to equilibrium after its perturbation. It involves the transitions between the energy levels that are caused by fluctuating magnetic fields. The rate of the energy transition will depend on interactions of spins with each other and the static magnetic field, molecular rotational motions and external fluctuations. In greater detail, after excitation, two types of relaxation can be considered:

- T_1 relaxation time is responsible for the magnetization recovery along the direction of B_0 field (equilibrium value) by regenerating the population of energy levels α and β according to the Boltzmann distribution. It is often called longitudinal or spin-lattice relaxation as the energy transfer occurs from the excited state to the surroundings. The rate of relaxation can be characterized by $R_1 = 1/T_1$ (Figure 2.7A);
- T_2 or transverse or spin-spin relaxation appears when there is loss of spin coherence, more precisely, dephasing with the contribution from energy transition, which doesn't change the population of the excited states. It gives the FID decay and the peak line-broadening. The rate of this relaxation is described as $R_2 = 1/T_2$ (Figure 2.7B).

Several mechanisms have a direct contribution to the relaxation rates R_1 and R_2, such as the dipolar interaction D (applied to both rates), chemical shift anisotropy CSA (often applicable to heteronuclei ^{13}C and ^{15}N), chemical exchange ex (influencing R_2). In general:

$$R_1 = \frac{1}{T_1} = R_1^D + R_1^{CSA} \tag{2.21}$$

$$R_2 = \frac{1}{T_2} = R_2^D + R_2^{CSA} + R_2^{ex} \tag{2.22}$$

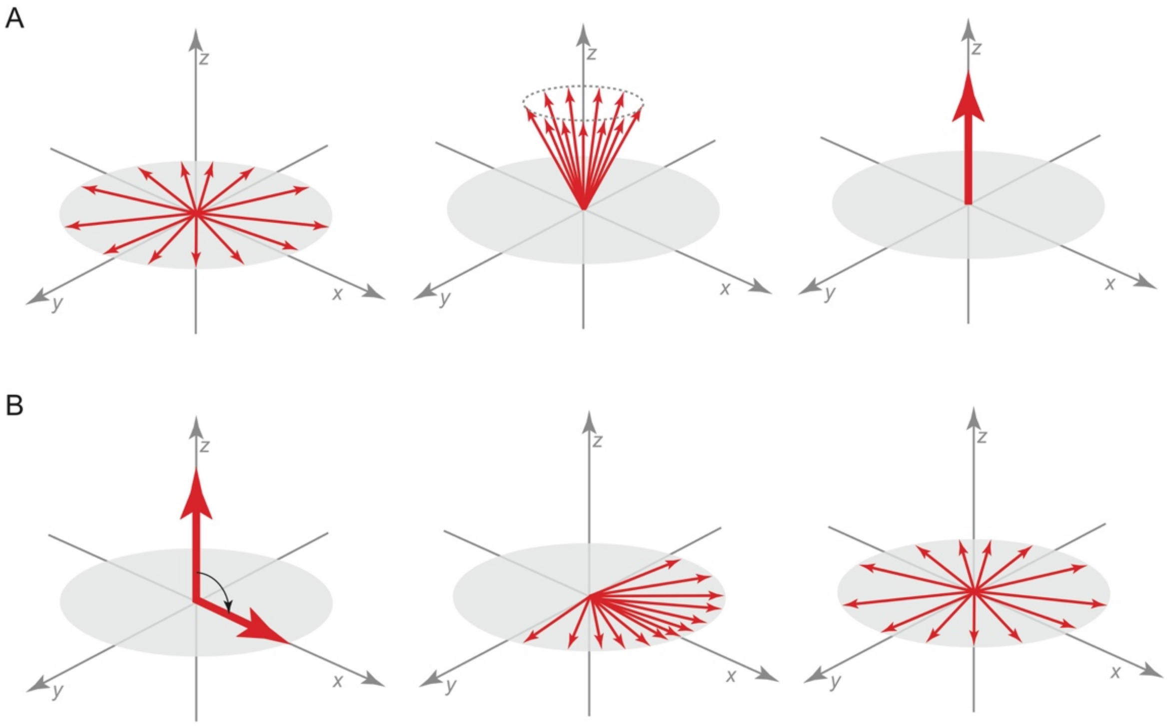


Figure 2.7. After applying RF pulse the system relaxes back to equilibrium through T_1 and T_2 relaxation processes that can be represented with the vector model. T_1 is a recovery of longitudinal relaxation of magnetization M_z along z-axis meaning there is the energy transfer from the nucleus to the lattice or solvent (A). T_2 is a loss (dephasing) of transverse magnetization M_{xy} on xy-plane with no energy change, only the energy exchange occurs between the excited nucleus and low energy state nucleus (B).

Correlation function and spectral density

It appears clear that both nuclear spin relaxations T_1 and T_2 processes – restore of thermal equilibrium and loss in the coherence transfer magnetisation – are caused by energy transitions between excited and ground states. The transition can occur by two different mechanisms: spontaneous and stimulated emission. The frequency of spontaneous emission is much higher than the NMR detectable absorption range: consequently, its rate is very small, making frequencies of oscillating electromagnetic fields from stimulated emission a major source of NMR relaxation. These fields are created by random rotational motion of the molecule or by internal motions within the molecule by generating time dependent fluctuations in the magnetic field upon anisotropic electron shielding (CSA) and dipolar interaction (*D*). If field fluctuations contain frequencies that match the Larmor frequency ω_o that cause the transitions, efficient relaxation can occur. The intensity of the changes in magnetic field as a function of the frequency ω is represented by the spectral density function $J(\omega)$. In order to understand, how to quantify the effect of $J(\omega)$ on relaxation, the correlation function needs to be used.

The correlation function $G(\tau)$ refers to the memory of the system on the specific spin arrangement, in other words, it describes the random molecular motion as a function of time τ. For times τ that are much shorter than the time needed for the system to rearrange itself, $G(\tau)$ is at its most elevated (Figure 2.8A), and conversely, upon evolution of τ, the memory function diminishes, $G(\tau)$ exponentially falls and decays to zero over the long times τ:

$$G(\tau) = G(0)\exp\left(-\frac{\tau}{\tau_c}\right) \tag{2.23}$$

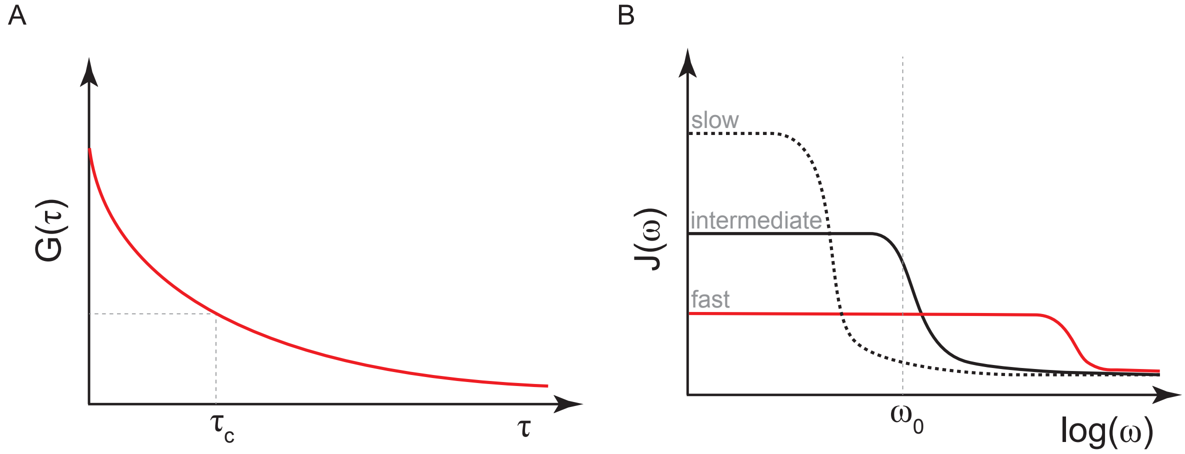


Figure 2.8. Correlation time τ_c can be determined from the sum of exponential plots of correlation functions (A). $G(\tau)$ can be Fourier transformed to the spectral density function having Lorenzian curve (B). Different types of molecular motion can be extracted from $J(\omega)$ function: from slow (dotted line) which corresponds to the higher τ_c to fast (red line) which is characteristic for lower τ_c (small molecules and IDPs).

From equation 2.23, it is clear that the exponential slope will be dependent on the value of τ_c (correlation time). τ_c is an average time that takes for the molecule to rotate one radian (a rigid spin attached to a rigid rotor). The value of τ_c is affected by the size of the molecule, more precisely, by its radius r, and viscosity η. According to Stokes-Einstein equation correlation time is equal to:

$$\tau_c = \frac{4\pi\eta r^3}{3k_B T} \tag{2.24}$$

where k_B – Bolzmann constant 1,38×10^{-23} J×K^{-1} and T – temperature. Small molecules and IDPs have a correlation time in the range from ps to ns. As an estimate, folded proteins at 25°C have τ_c as half of their molecular weight expressed in kD, for example, for the protein of 50 kDa τ_c is approximately equal to 25 ns.

As the correlation function is time dependent, we can apply the Fourier transformation on it, resulting in the spectral density function:

$$G(\tau) \xrightarrow{FT} J(\omega) = \frac{2}{5}\frac{\tau_c}{1+\omega^2\tau^2} \tag{2.25}$$

Plotting $J(\omega)$ against ω gives rise to the Lorenzian curve (Figure 2.8B) and presents the density of the field fluctuations at different frequencies. The maximum contribution of spectral density will be found when $\tau_c = 1/\omega_0$ (T_1 minimum). It can be seen that for molecules with short correlation times ($\tau_c << 1/\omega_0$) being in fast motion, spectral density is almost identical at $\omega = 0$ and at higher frequencies. In contrast, for slow motion with $\tau_c >> 1/\omega_0$ we observe a high spectral density at $\omega = 0$ that decreases with higher frequencies. Therefore, there is an effect of molecular weight on rotational motion and spectral density.

In case of complex molecules, such as proteins, the local field fluctuations are affected by the internal molecular motions. According to "model-free" approach [Lipari et Szabo, 1982; Clore et al., 1990], in addition to the overall correlation time τ_c, there is a time constant for internal motion τ_i and order parameter S^2 to be introduced to the spectral density function (equation 2.26). S^2 can be 1 meaning no internal motion present and 0 for random motions of very large amplitude.

$$J(\omega) = (1 - S^2)\frac{\tau_{mix}}{1+\omega^2\tau_{mix}^2} + S^2\frac{\tau_c}{1+\omega^2\tau^2} \qquad \tau_{mix} = \frac{\tau_i\tau_c}{\tau_i+\tau_c} \tag{2.26}$$

^{15}N relaxation

In proteins, protons are involved in numerous dipolar interactions, therefore, it is complicated to interpret their relaxation in terms of protein dynamics. For this reason, ^{15}N relaxation is mostly used to characterise the intramolecular dynamics and overall molecular diffusion of folded proteins and IDPs. Especially, T_2 is important to describe protein backbone dynamics (ps to ns time scale) as well as conformational exchange processes that occur on the µs to ms time scale and significantly influence R_2. However, the influence of molecular size and magnetic field on T_1 and T_2 is not negligible.

To summarize, ^{15}N R_1 and R_2 relaxation rates depend on a linear combination of the spectral density functions *J(ω)* describing local fluctuation fields at the proton and nitrogen frequencies for each protein residue with dipolar *d*, CSA *c* and exchange R^{ex} contribution to the final relaxation equation:

$$R_1 = \frac{d^2}{20}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + \frac{c^2}{15}J(\omega_N) \tag{2.27}$$

$$R_2 = \frac{d^2}{20}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 3J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{15}[J(0) + 3J(\omega_N)] + R^{ex}$$

$$d = \frac{\mu_0 \gamma_H \gamma_N \hbar}{4\pi {r_{NH}}^3} \quad c = \omega_N \Delta\sigma_N \tag{2.28}$$

The effect of τ_c on nitrogen T_1 and T_2 relaxation can be described as shown in Figure 2.9. From the spectral density function, it can be derived that T_1 is minimal when the molecular tumbling rate is equal to the Larmor frequency. Fast motion gives the low intensity of spectral density function at ω, and long T_1. After passing T_1 minimum, the longer τ_c gives longer spin-lattice relaxation for large molecules. Since *J(ω)* decreases with increasing ω, the rise in magnetic field strength will make T_1 value higher.

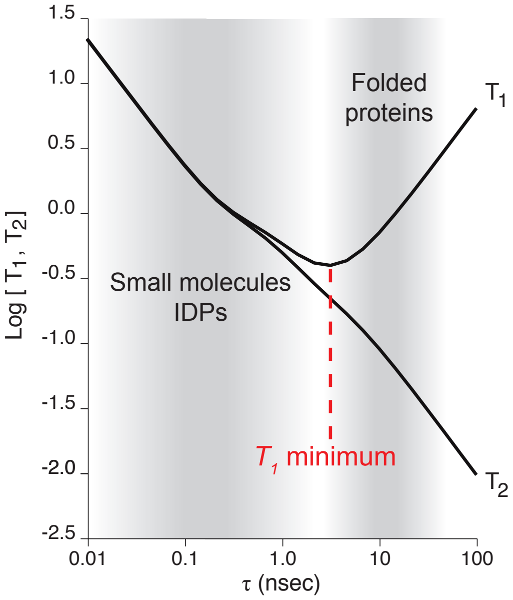


Figure 2.9. Effect of the molecular weight on nitrogen T_1 and T_2 relaxation. The grey bars indicate correlation time range for small molecules/IDPs and the folded proteins that can be studied by NMR. Adapted from [Rule and Hitchens, 2006].

Unlike T_1 parabolic dependence of τ_c, spin-spin relaxation becomes higher over the decrease in the rotational molecular motion (Figure 2.9). This is due to the contribution of $J(0)$ (spectral density at $\omega = 0$) to the R_2 relaxation rate. Hence, T_2 is inversely proportional to the molecular size. For larger molecules it is difficult to transfer magnetization between the spins which leads to greater R_2 values and therefore, to peaks broadening. Ideally, the value of the magnetic field does not affect much spin-spin relaxation. However, in order to have an idea about the exchange contribution to the final R_2, it is recommended to record transverse relaxation rate at two distinct magnetic fields.

2.2.4. *Chemical exchange*

Intra- (isomerisation, protein side chain motions and unfolding) and intermolecular (small molecules binding to macromolecules, isotope exchange processes, protonation/deprotonation) conformational changes play an important role in protein dynamics. Conformational changes are often related to chemical exchange which means that the chemical environment for one nucleus is exchanging with another. Chemical exchange is also the important contribution factor to protein transverse relaxation.

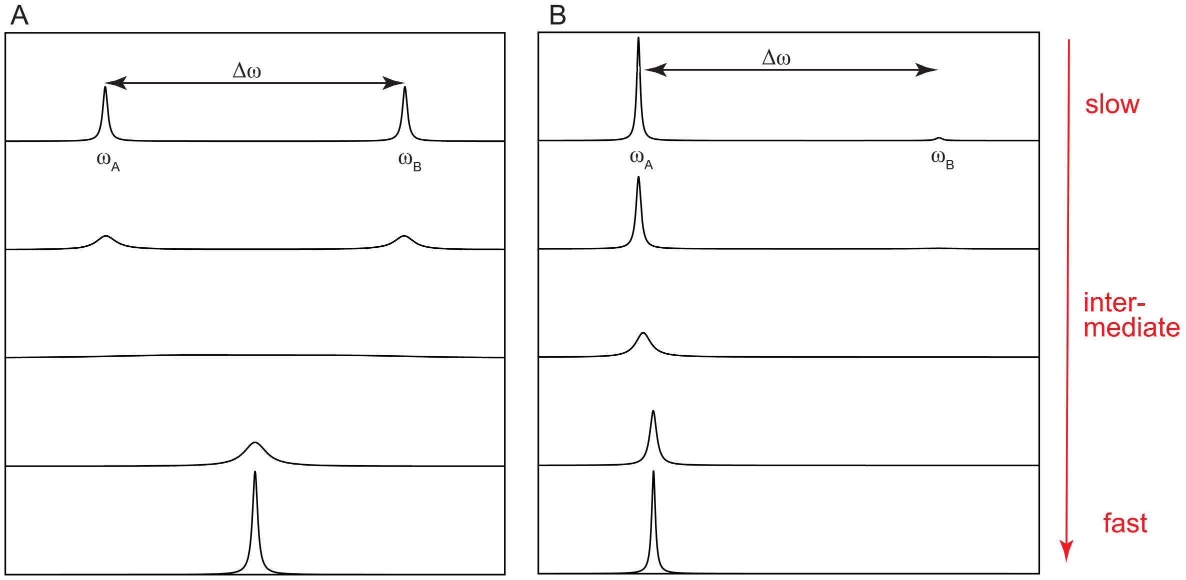


Figure 2.10. Spectra of spins undergoing chemical exchange at different rates between the states A and B as a function of k_{ex} when population of states A and B is equal ($p_A = 0{,}5$ and $p_B = 0{,}5$) (A) and widely different with A as major state ($p_B = 0{,}95$ and $p_B = 0{,}05$) (B). Chemical shift is defined by ω_A and ω_B and difference between them is $\Delta\omega$.

The conformational exchange can be defined as an equilibrium driven with first order forward k_{AB} and reverse k_{BA} rate constants:

$$A \leftrightharpoons B \quad \Delta\nu = \nu_A - \nu_B \text{ or } \Delta\omega = \omega_A - \omega_B \quad k_{ex} = k_{AB} + k_{BA} \tag{2.29}$$

$$p_B = \frac{k_{AB}}{k_{AB}+k_{BA}} \quad p_A + p_B = 1 \tag{2.30}$$

with $\Delta\nu$ or $\Delta\omega$ used for the chemical shift difference between states A and B, k_{ex} as the exchange constant at which system returns to equilibrium and p_A and p_B fractional population of states A and B.

According to the Bloch-McConnell equations (not shown), the chemical exchange regime can be defined from the value of k_{ex} as compared to $\Delta\omega$:

- when $\Delta\omega << k_{ex}$, the system is in fast exchange, we observe one peak, as the molecule is interconverting between two states and the magnetization is evolving with the weighted average chemical shift depending on the populations of A and B;
- for slow exchange $\Delta\omega >> k_{ex}$, two peaks appear, corresponding to states A and B. Sometimes due to the low population on the second state, the peak can be too low in intensity, and therefore, absent in the NMR spectrum;
- $\Delta\omega \sim k_{ex}$ gives rise to intermediate exchange and the peak is broadened and can be almost non-observable (Figure 2.10).

The chemical exchange can be characterised by the study of the spectral line width which is linked to the protein dynamics and has a non-negligible contribution to the transverse relaxation. In the case of fast exchange, the influence of exchange R^{ex} to the T_2 relaxation depends (among already mentioned $\Delta\omega$, k_{ex}, p_A and p_B) on the field strength used for NMR acquisition, as $\Delta\omega \propto B_0$. For slow exchange rate the major peak line width will depend only on k_{ex} and p_B. In addition, a temperature increase can shift the system to the fast exchange. By summarising in the limits of slow and fast exchange respectively:

$$R^{ex} = p_B k_{ex} \quad \text{when} \quad \frac{k_{ex}}{\Delta\omega} \to 0 \tag{2.31}$$

$$R^{ex} = \frac{\Delta\omega^2 p_A p_B}{k_{ex}} \quad \text{when} \quad \frac{k_{ex}}{\Delta\omega} \to \infty \tag{2.32}$$

Depending on the time scale and nature of the chemical exchange, it can be followed firstly, by the R_2 measurements at different magnetic fields and secondly, by specific experiments such as Carr-Purcell-Meiboom-Gill sequence (CPMG) [Carr & Purcell, 1954] and $T_{1\rho}$ (relaxation in the rotating frame) [Blackledge et al., 1993] for the fast exchange (from ms to μs), Chemical Exchange Saturation Transfer (CEST) [Vallurupalli et al., 2012] and ZZ exchange [Montelione

& Wagner, 1989] for slow exchange [Hansen et al., 2008; Loria et al., 1999; Tjandra et al., 1996; Palmer et al., 2001].

2.3. *NMR measurements and applications*

2.3.1. *NMR time range and limitations*

The study of NMR spectra can provide a lot of useful information about physical and chemical processes which take place in the sample. All previously described NMR parameters (chemical shift, scalar and dipolar couplings, peak intensity, T_1 and T_2 relaxation etc) are sensitive to a wide range of time scales (Figure 2.11) from picoseconds to days.

However, the use of NMR spectroscopy is limited by low sensitivity compared to other analytical methods and by the size of studied molecule. The sensitivity of NMR experiment can be described by the signal-to-noise ratio S/N:

$$S/N \propto \frac{n}{V}\gamma_{exc}\sqrt{{\gamma_{det}}^3{B_0}^3(NS)} \tag{2.33}$$

where n – net Bolzmann distribution of observed spins (number of spins), V - sample volume, γ_{exc} and γ_{det} gyromagnetic ratio of spins being excited and observed respectively, NS – number of times the pulse sequence is repeated (number of scans). From equation 2.33 it is obvious that utilisation of higher magnetic fields, higher number of scans which leads to longer acquisition time, nucleus with higher gyromagnetic ratio and bigger amount of working material will increase the S/N.

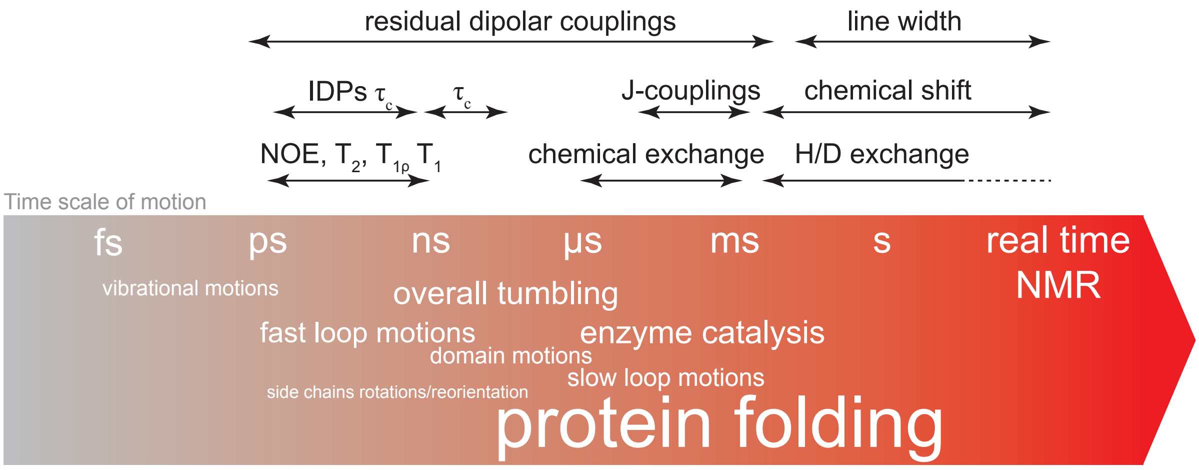


Figure 2.11. Accessible time range for NMR spectroscopy. Protein motional behaviour (dynamics) can be described by different NMR parameters.

Another important feature for sensitivity is the NMR probe which arranges the sample in the homogeneous field and is responsible for RF excitation and signal detection. Over the last years, the use of cryogenic probes (NMR magnets are operating at cryogenic temperatures) allows decreasing the sample volume and reducing the thermal noise.

When it comes to the molecular size, a folded protein has, higher τ_c and shorter T_2. Larger molecules show slower response to the small solvent molecules encountering under Brownian motion, and, therefore, reorient themselves slowly. Depending on the studied system, folding properties and conditions, it is difficult to specify the limit of molecular weight for globular proteins to be studied by NMR. In general, very big proteins and oligomers result in a loss of resolution due to the spectral line width enlargement. However, recent progress in biochemical methods such as segmental, residue-specific labelling and deuteration can help to work even with very large oligomeric structures [Ohki & Kainosho, 2008]. Some NMR methods were developed in order to minimize the effect of transverse relaxation and the increase polarization transfer. TROSY (transverse relaxation-optimized spectroscopy) technique is one of the widely used [Pervushin et al., 1998].

In the case of the objects of this project, IDPs, there is almost no limits in the amino acid number to be analysed by NMR. Due to the absence of the non-covalent inter-residual interactions (hydrophobic, hydrogen bonds) and the high flexibility (fast interconversion between conformational states), it is difficult to define the rotational correlation time of IDPs. Instead of an overall molecular τ_c, residue-specific correlation times can be introduced, and they correspond to those of small molecules. Consequently, for large molecules, fast molecular motion and longer relaxation rate R_2 gives rise to broad NMR signals. At the same time, the contribution of the similar chemical environment to the chemical shift values of different residues of IDPs leads to low signal dispersion and overlapping which can be overcome by high order dimensional NMR experiments.

2.3.2. *Two-dimensional NMR*

Due to the complexity of biomolecular one-dimensional ^{1}H spectra (Figure 2.12A) caused by overlapping signals, the addition of a second dimension simplifies the spectrum by resolving the overlapped peaks. In other words, if the 1D spectrum is a plot of the intensity of the signal as a function of the frequency, for the 2D experiment the intensity (contour levels) is plotted over two frequency axes f_1 and f_2 of two same (homonuclear) or different (heteronuclear spectrum) nuclei (Figure 2.12B).

Concerning Fourier transformation of two-dimensional spectrum, since the FID signal depends on two times, t_1 and t_2, the resulting spectrum is obtained as a function of two frequency variables resulting from two Fourier transforms. The basic 2D experiment pulse sequence (Figure 2.12C) consists of 4 periods:

- preparation is started by spin excitation usually by one or several 90° pulses;
- during evolution the magnetization precesses over certain time t_1;
- following mixing period uses further pulses for the magnetization transfer (scalar or dipolar coupling mechanisms, for example);
- the NMR signal is acquired during detection over time t_2.

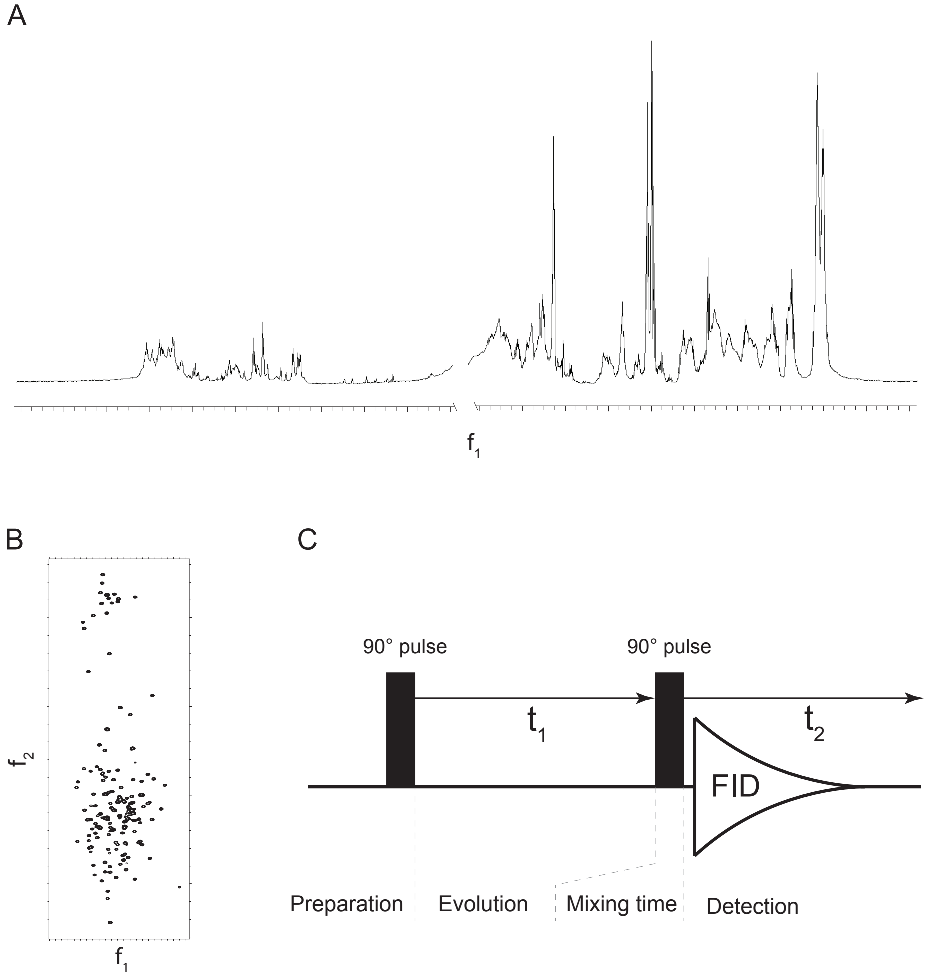


Figure 2.12. Protein one-dimensional spectrum (A) does not provide enough resolution and structural information due to the signals merging. Typical two-dimensional heteronuclear spectrum gives one correlation peak from two different nuclei corresponding to two frequency axes (B). The pulse sequence for simplest homonuclear 2D experiment (COSY) as an example of 4 periods (C).

In more details, after setting t_1 to 0 and performing the pulse sequence, the signal is recorded. Following that, spins are allowed to return to the equilibrium and t_1 is set to Δ_1 (t_1 sampling interval). The pulse sequence is executed for the second time, but the recorded FID during t_2 time is stored separately. The same procedure is repeated for increasing values $2\Delta_1$, $3\Delta_1$ and so on, until there are enough of t_1 increments to achieve the required resolution in the indirect dimension(s).

Depending on the requested information, a variety of homo- (COSY, NOESY, TOCSY) and heteronuclear (HSQC, HMBC, TROSY) 2D experiments have been developed. In this project we will mostly focus on HSQC (heteronuclear single quantum coherence transfer) and other multidimensional NMR spectra based on it [Cavanagh et al., 2006; Chary & Govil, 2008].

HSQC spectrum

The ^{1}H-^{15}N HSQC spectrum is one of the most commonly used NMR experiment for proteins. It can be set up under nitrogen natural abundance but due to the low natural abundance of ^{15}N, uniform ^{15}N protein labelling is normally required. The spectrum is based on magnetization transfer from ^{1}H to ^{15}N via *J*-couplings. After chemical shift evolution on nitrogen, magnetization is transferred back to proton for detection.

As ^{1}H-^{15}N HSQC displays correlations between proton and nitrogen separated by one bond: a single cross-peak appears from ^{1}H-^{15}N of the amide bond, corresponding to one protein residue (except proline). Moreover, in the spectrum there are peaks from the side chains of tryptophane, arginine and doublets of asparagine and glutamine (Figure 2.13A).

Together with the 1D spectrum, the HSQC can be used to identify whether a protein is folded or unfolded. Peaks that cluster in the middle of the spectrum usually from about 7,8 to 8,8 ppm in the ^{1}H dimension indicate that the protein is in random coil conformation, which is an average of ensemble of structures, while structured proteins have well-dispersed peaks in the hydrogen dimension (Figure 2.13B).

The HSQC spectrum constitutes the protein “fingerprint”, which means that protein folding, screening of optimal buffer conditions (pH and ionic strength), ligand binding (small molecule, protein, nucleic acid), post-translational modifications (phosphorylation, *O*-glycosylation) and actual protein state (absence of degradation) and temperature can be monitored. The spectrum is the starting point for backbone assignment and can answer the question about the relevance of more expensive ^{13}C labelling. In addition, before performing relaxation experiments useful information can be extracted from the peak intensities.

In IDPs, the amide protons are freely exposed to the solvent, making hydrogen exchange with water possible, which can lead to extensive line broadening sometimes with complete disappearance of corresponding NMR signals. Hydrogen exchange is highly dependent on the sample pH and temperature. Under physiological conditions (pH 7 and 37°C) several amino acids cross-peaks (for example, serines and threonines due to their exchangeable with water hydroxyl group) are often not visible at 2D ^{1}H-^{15}N HSQC. All measurements in this project were carried out at pH 6 and 25°C.

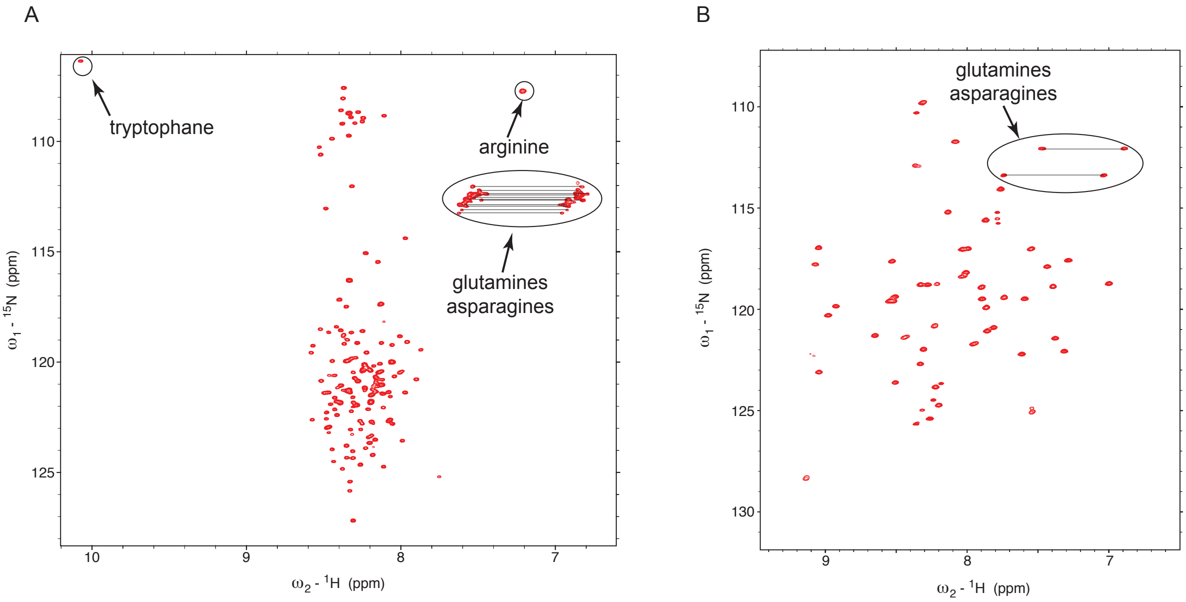


Figure 2.13. 2D HSQC spectra of unfolded (A) (225 residues of MuV P_{NTD}) and folded (B) proteins (58 residues of MeV XD). Each peak corresponds to one amino acid of the polypeptide backbone chain. Amide side chain peaks are marked with arrows and circles. Glutamine and asparagine doublet peaks are connected with lines.

All the liquid-state NMR measurements from this project were performed on Bruker 600, 700, 850, 950 MHz spectrometers with cryo- and room temperature probes and an Agilent 600 MHz spectrometer. The samples contained of 10% D_2O to lock the spectrometer frequency. The resulting spectra were transformed with the help of NMRPipe program [Delaglio et al., 1995], visualized and analyzed by the Sparky software [Goddard and Kneller, University of California].

2.3.3. *Protein backbone assignment*

From abovementioned HSQC spectrum, it is clear that each cross-peak corresponds to one protein residue. In order to determine, which peak corresponds to which amino acid pair N-H in the protein sequence, sequence specific resonance assignment needs to be performed. All the amino acids are chemically distinct, and, therefore, they have different chemical shift values for ^{1}H, ^{13}C and ^{15}N nuclei. The large chemical shift dispersion for carbon and nitrogen is the key for successful IDP assignment

Several three-dimensional BEST (band-selective excitation short-transient) experiments were recorded [Lescop et al., 2007]. 3D spectra with $^{1}H_N$ detection and correlation through one- and two-bonds intra- and inter-residual scalar couplings allow to achieve sequential assignment walk from one residue *i* to another *(i - 1)*:

- HNCO is the most sensitive experiment, it correlates $^{1}H_N$ and ^{15}N chemical shifts of residue *i* to $^{13}C'$ chemical shifts of the residue *(i - 1)*. iHN(CA)CO spectrum completes these data with intra-residual chemical shifts for H_N, N and C′ for the residue *i*;
- in the same manner, HN(CO)CA shows inter-residual correlations for $^{13}C_\alpha$ for amino acid *(i - 1)*, iHNCA is an intra-residual experiment;
- finally, the inter- and intra-residual $^{13}C_\beta$ chemical shifts are extracted from HN(CO)CACB and iHNCACB spectra (Figure 2.14).

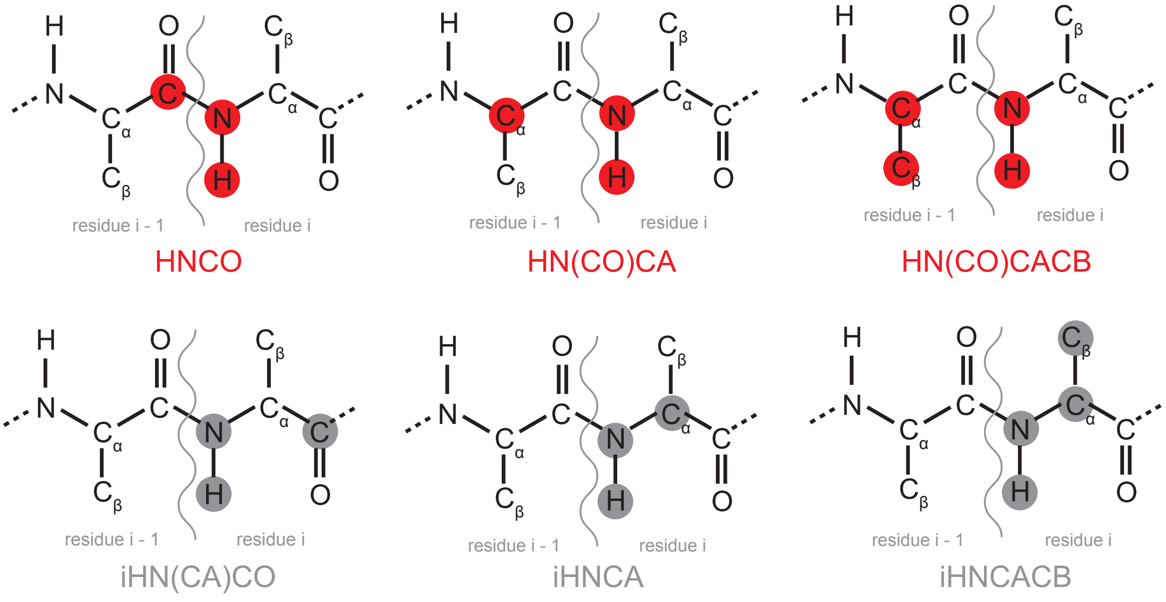


Figure 2.14. Summary of 3D experiments used for C′, C_α, C_β by sequential inter- (in red) and intraresidual (in grey) assignments.

Experiments iHN(CA)CO, iHNCA and iHNCACB can be set up for both transfers to occur from H^N and N to carbons of *i* and *i-1* residues, resulting in two peaks with different intensities which depend on the the coupling constant between connected N and C atoms. This allows sequential assignment, residue by residue, by following coherence transfers from one slice of 3D intra-residue spectrum to another slice of inter-residue experiment and so on (Figure 2.15).

Automated backbone protein assignments were performed with the help of MARS software [Jung and Zweckstetter, 2004]. The required input for this program consists of the chemical shifts table extracted from peaks (H^N, N, C′, C_α, C_β for *(i – 1)* and *i* residues), protein sequence and secondary structure prediction file from PSIPRED server [Buchan et al, 2013]. The assignment by MARS needs to be followed by manual verification and validated.

2.3.4. ^{15}N $R_{1\rho}$ and R_1 measurements

R_1 and $R_{1\rho}$ experiments are based on HSQC spectra. They are recorded at different relaxation times *t*. The peak intensities *I* are then plotted as a function of *t*, and R_i and extracted via a fit to a single exponential decay:

$$I = I_0(e^{-tR_i}) \tag{2.34}$$

Rotating frame relaxation $R_{1\rho}$ and R_1 were used to derive R_2. Technically, the measurement of $R_{1\rho}$ is based on spin-locking the magnetization byapplication of an RF field at certain carrier frequency ω_{RF} (spin-lock field). ω_{RF} becomes equal to ω_1, notably, fixed on the x-axis [Guenneugues et al., 1999]. R_2 can be calculated back after R_1 has been measured:

$$R_{1\rho} = R_1 \cos^2 \theta + R_2 \sin^2 \theta \qquad \theta = \tan^{-1} \frac{\omega_{RF}}{\Omega} \tag{2.35}$$

where Ω - offset frequency.

2.1.1. Heteronuclear NOE

Two nuclei undergoing dipole-dipole interaction, do not relax independently, since their dipolar coupling is responsible for the cross-correlated relaxation called Nuclear Overhauser Effect (NOE). NOE can be observed between two heteronuclei, N and H^N, and then, called heteronuclear NOE (hetNOE). The size of NOE η depends on the gyromagnetic ratios of the interacting nuclei, correlation time and static field strength (R_1 contribution):

$$\eta_{N(H)} = 1 + \frac{d^2}{20}\frac{\gamma_H}{\gamma_N}[6J(\omega_H + \omega_N)]\frac{1}{R_1} \tag{2.36}$$

For $^{15}N\{^1H\}$ pair of spins, it is rather ^1H that is being saturated due to the higher γ. In practice, two experiments are recorded, one as a reference and another as a saturation, after their intensities are then compared.

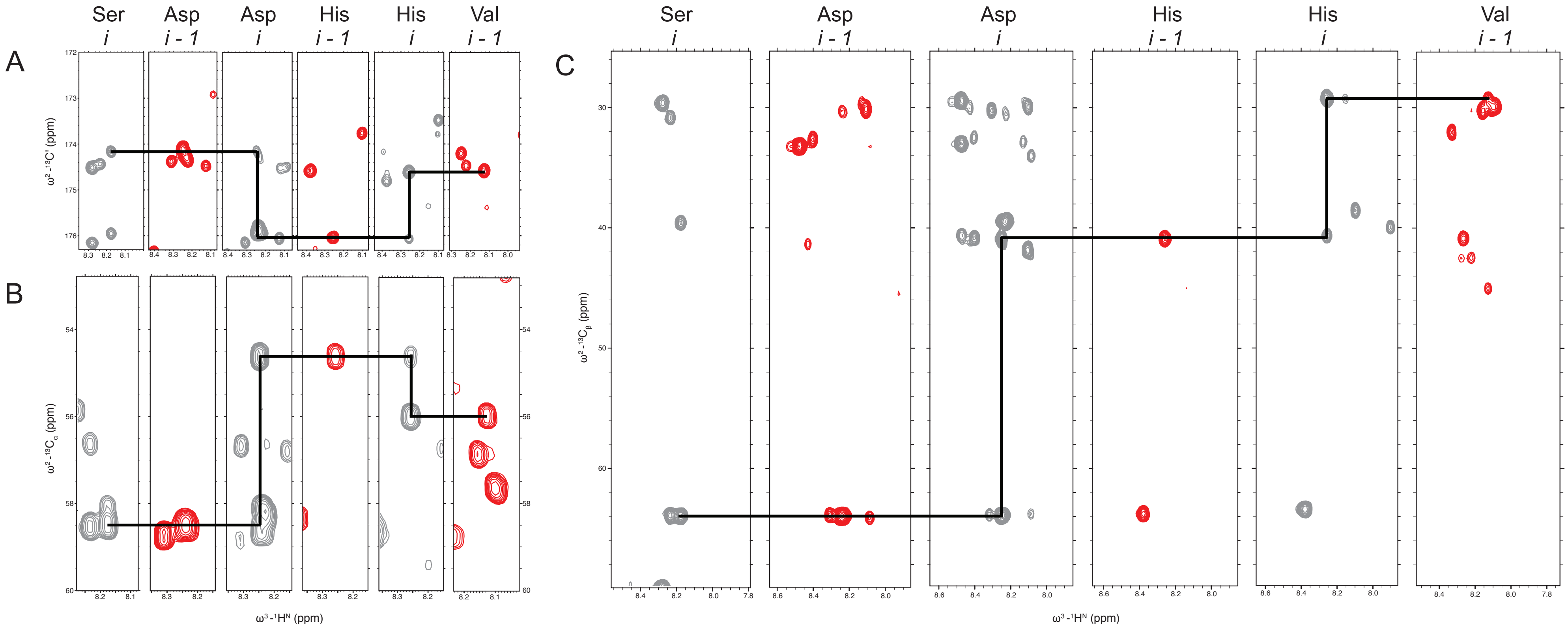


Figure 2.15. The example of sequential assignment for the peptide SerAspHisVal coming from MuV P$_{NTD}$. Peaks from inter- and intra-residual spectra are coloured in red and grey respectively, showing the chemical shifts for carbons C′ (A), C$_\alpha$ (B), C$_\beta$ (C). Correlations inside the residue *i* show additional peaks from *i – 1* amino acid by making possible to follow the protein sequence (black line).

Having negative value γ for ^{15}N, measured at 600 MHz field negative NOEs come from flexible protein regions and positive from structured and rigid regions. Polarization transfer from amide H to N can be an additional tool to describe the motion of individual bond vectors.

2.3.5. *Real-time NMR*

Real-time NMR can be used for monitoring of protein folding, post-translational modifications (phosphorylation), assembly of proteins and can be followed with the help of fast-pulsing SOFAST-HMQC spectra [Schanda *et al.*, 2005]. The experiment is based on recording of the set of individual high-speed and optimal-sensitivity HSQCs (about 5 min of SOFAST-HMQC against tens of minutes for a standard HSQC) over the time. Peak intensities are monitored as a function of time by allowing to extract, for example, kinetic constants for each protein residue that was affected in the interaction.

2.3.6. *Paramagnetic relaxation enhancement*

Probing long-range contacts between protein residues spaced by about 15-20 Å requires the chemical modification of the protein by covalently attaching a paramagnetic (spin) label to the thiol group of a cysteine. Unpaired electrons from the spin label generate fields ω_e that can interact with the nuclei of interest. As the gyromagnetic ratio of electron γ_e (1,76×10^{11} rad×s^{-1}×T^{-1}) is much higher than, that of proton, according to equation 2.28, the dipolar interaction between the proton and the electron becomes a source of T_2 relaxation:

$$R_2^D = \frac{1}{8}\left[\frac{\hbar\mu_0\gamma_H\gamma_e}{4\pi r^3}\right]^2 [4J(0) + 6J(\omega_H) + J(\omega_H - \omega_e) + 3J(\omega_e) + 6J(\omega_H + \omega_e)] \quad (2.37)$$

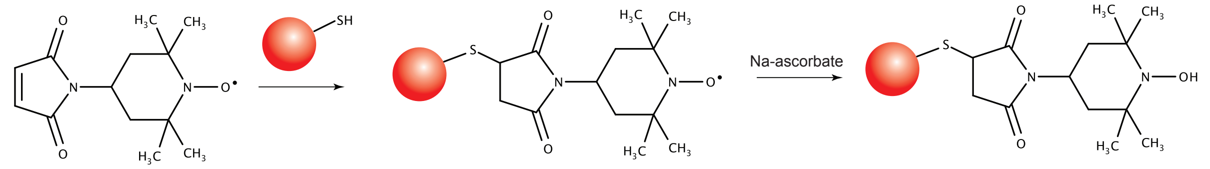


Figure 2.16. Procedure of protein TEMPO labelling via cysteine and mild reduction with ascorbate. The red sphere indicates the protein molecule.

According to equation 2.37, the dipole-dipole interaction contribution to the relaxation rate is dependent on the distance between the spin and the electron. Here we can talk about paramagnetic relaxation enhancement (PRE). The choice of the paramagnetic label is crucial, it is important to control whether the protein secondary structure is impacted by the labeling, and to ensure stability of both protein conformational behaviour and labelling. In the case of this project, we have used TEMPO-maleimide (Figure 2.16) labelling via cysteine, followed by chemical reduction with the help of sodium ascorbate.

2.3.7. *Residual dipolar couplings*

Another experiment can be performed for additional protein structural and dynamic information. When the protein is dissolved in an anisotropic « alignment » medium, it becomes partially ordered with respect to the applied magnetic field. Therefore, the restricted overall averaging causes non-zero residual dipolar couplings (RDCs). In practice, they are measured as the difference between couplings in isotropic (J_{NH}) and anisotropic ($J_{NH} + D_{NH}$) media (Figure 2.17).

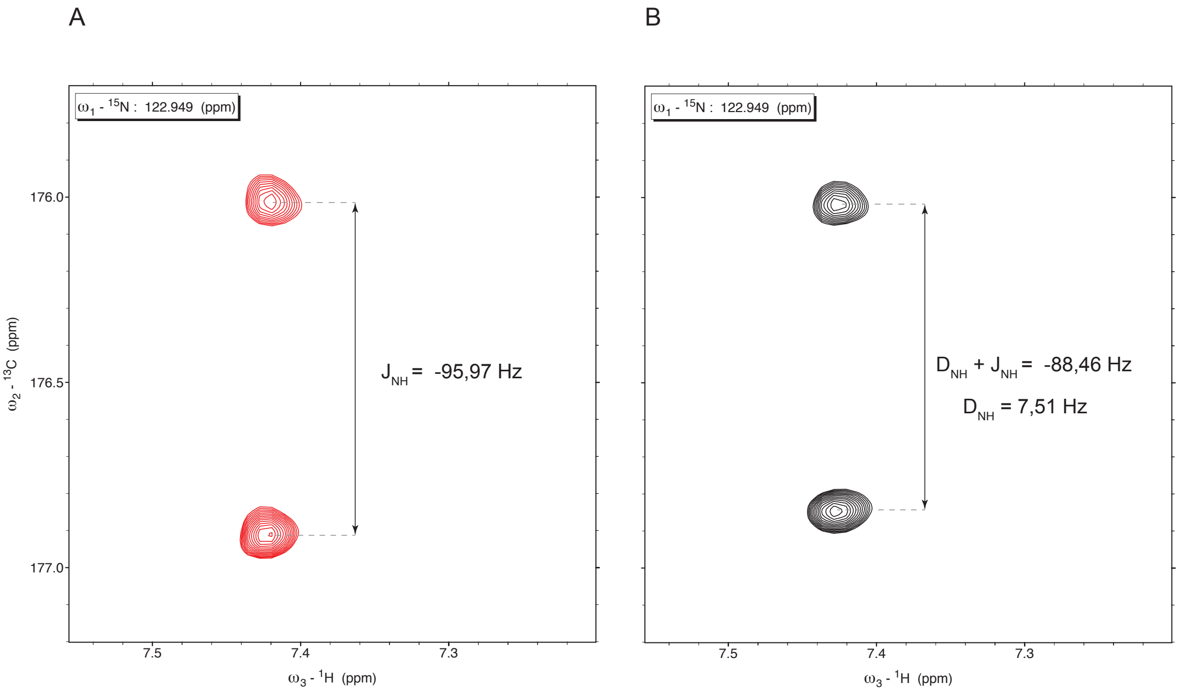


Figure 2.17. The extraction of RDC from $^{1}J_{NH}$ of MuV P_{CTD}. ^{15}N slices of two 3D spectra are compared, in absence (A) and presence (B) of anisotropic medium with further calculation of dipolar coupling constant for particular protein residue (B).

The easiest way to align the protein molecules is to place it directly in an anisotropic medium. Only a very small fraction is actually aligned due to the electrostatic or steric interactions with the medium molecules which have a large anisotropic dipole. Among different anisotropic conditions, one can mention filamentous bacteriophages and liquid crystals [Clore et al., 1998; Rückert & Otting, 2000], which will be both used in this manuscript, as well as lipid bicelles, polyacrylamide gels etc [Prosser et al., 1998; Ishii et al., 2001]. However, phage medium implicates the presence of protein electrostatic interactions with the negative surface of phages. In all cases, the degree of the alignment is usually checked by measuring the deuterium quadrupolar residual coupling of HDO molecule.

Practically, the experiment is based on the recording of two blocks of spectra without and with orienting medium and further extraction of RDCs from the sum of overall coupling that is a contribution of scalar J and dipolar D interactions. If the HSQC spectrum on its own doesn't give enough spectral resolution, one can also run 3D HNCO and HN(CO)CA based experiments with coupling evolution (intra-residual $N{-}H^N$, $C'{-}C_\alpha$ and double bond $C'{-}N{-}H^N$, inter-residual $C_\alpha{-}H_\alpha$) in the carbon dimension (Figure 2.17).

2.4. Structural ensemble calculation

Because of the flexibility, it is not appropriate to describe the structure of IDPs using a single structure. In the case of IDPs that have random coil conformation and only transiently populated secondary structures, only an ensemble (set) of structures can be used to describe their dynamic behaviour. In general, NMR parameters are needed to build up the descriptors for ensemble molecular models.

The protein structure can be represented by a distribution of dihedral angles φ and ψ connecting each peptide bond plane in the long polymer chain. *in silico* statistical coil generators, as for example, Flexible Meccano [Bernado et al., 2005, Ozenne et al., 2012], are based on this approach. The algorithm uses amino acid specific dihedral angles extracted from loop regions of known PDB protein structures, and adds residue per residue, avoiding steric clashes. From the generated pool of structures NMR parameters are calculated, they can be compared to the experimental ones.

The second step is to determine a subset of conformers that together describe the experimental data (chemical shifts, RDCs, PREs, R_g radius of gyration from SAXS experiments); this can be done with the genetic algorithm ASTEROIDS [Nodet et al., 2009]. It takes the experimental input as restraints to select a subset of conformers that on average agrees with the experimental data. Usually, a few repeated iterations (regeneration of pool and following selection) are required until the back-calculated data fits the experimental one and no further improvement in the fitting is observed.

CHAPTER 3.

MATERIALS AND METHODS. OTHER TECHNIQUES

3. OTHER TECHNIQUES

Due to the low sensitivity and the molecular size limitations, NMR spectroscopy needs additional techniques that instead of the atomic description, can provide information about the exact molecular size, shape, oligomerisation state. Techniques such as SAXS, EM and MALLS do not require large amounts of sample and usually, are rapid and robust. In addition, biochemical methods as definition of protein constructs, clone generation, recombinant protein expression and purification are described in this chapter.

3.1. SAXS

Small-angle X-ray scattering (SAXS) is currently a widely used method for the determination of biomolecular shape, oligomerisation state and mass. The technique relies on elastic scattering (without changing of the wavelength) of X-rays and due to the inhomogeneity of the molecule in the nanometer scale, the distribution of the scattering signal is measured at very small angles (0.1 – 10°). Despite its low resolution range (1-25 nm), SAXS can be complementary for structure determination using for example, NMR.

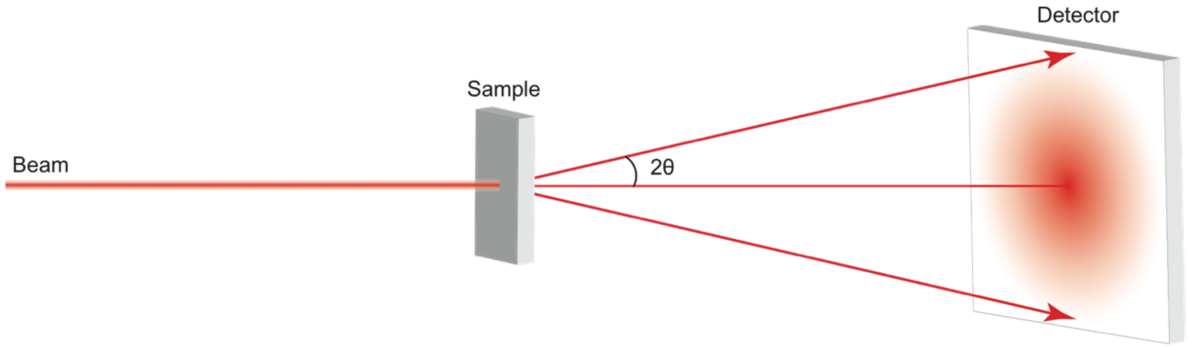


Figure 3.1. Physical basis of SAXS experiment. The sample in quartz capillary is illuminated by monochromatic X-ray beam, after intensity of scattered radiation is recorded by the detector as a function of 2θ angle.

A standard SAXS experiment is based on the irradiation of the sample in solution by X-rays and the measurement of the beam scattering azimuthal pattern on the detector (Figure 3.1). In more details, the signal $\Delta\rho$ is derived from the difference between the average electron density of the biomolecule ρ_r and the solvent ρ_s in which it was dissolved:

$$\Delta\rho = \rho_r - \rho_s \tag{3.1}$$

In the scattering curve (Figure 3.2A, B), the intensity of the scattered light *I(q)* is recorded as a function of the scattering vector q which can be defined from the scattering angle 2θ and the irradiating wavelength λ of X-ray beam:

$$q = \frac{4\pi sin\theta}{\lambda} \tag{3.2}$$

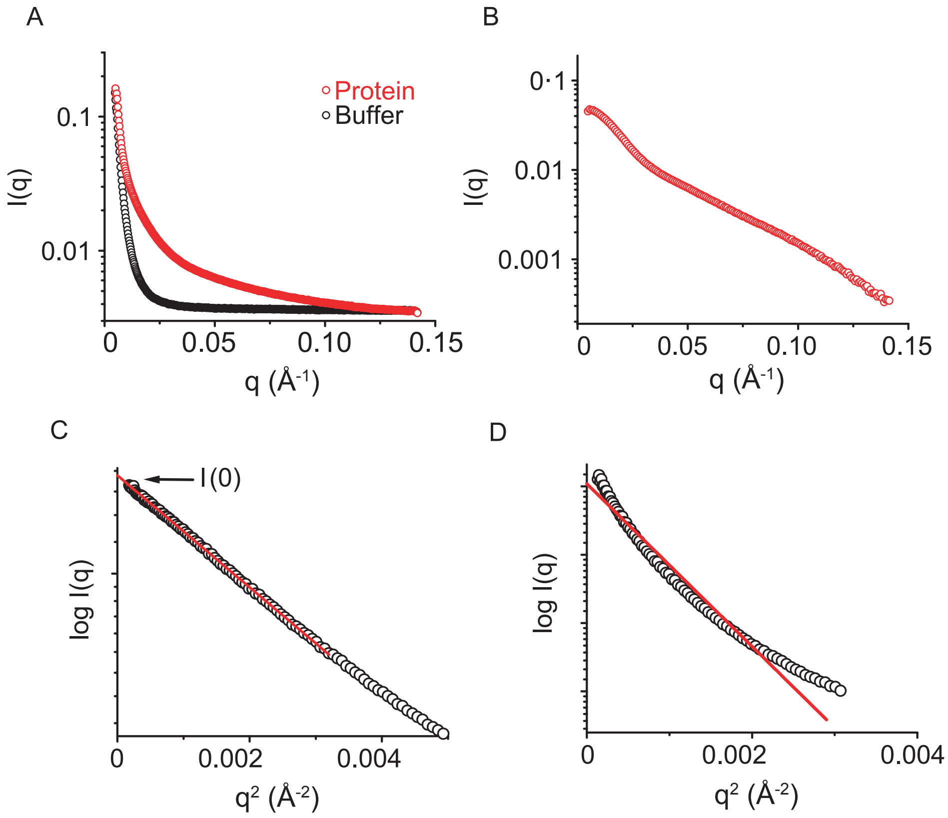


Figure 3.2. Typical SAXS curves from sample and buffer (A) and the one with the subtracted buffer (B), ideally, it is recommended to record several protein concentrations. Guinier plots can be used to distinguish non-aggregated protein (C) from aggregated sample (D) and to determine the R_g. Red line indicates the Guinier fit. Adapted from [Puthnam et al., 2007].

Ideally, several protein concentrations are measured, to account for any protein aggregation, inhomogeneity, polydispersity and long-range contacts. The size parameter R_g, radius of gyration, can be easily identified from the Guinier approximation. By plotting *ln[I(q)]* against q^2 and by extrapolating the straight line to the zero angle $q = 0$, *I(0)* can be extracted (Figure 3.2B).

$$I(q) = I(0)e^{\left(\frac{-q^2 {R_g}^2}{3}\right)} \tag{3.3}$$

The Guinier plot can be used for $qR_g < 1{,}3$ Å^{-1} (globular proteins). In the case of non-linearity, the sample can be considered as aggregated or elongated, therefore, either the sample quality needs to be improved, or other methods for R_g estimation need to be incorporated (Figure 3.2C, D).

I(0) depends on the square of the number of electrons, and consequently SAXS is useful for the identification of the protein molecular weight and oligomerisation state.

$$M_W = \frac{I(0)\mu^2}{N_A\left(1-\frac{\rho_S}{\rho_r}\right)^2} \tag{3.4}$$

where M_W – molecular weight, N_A – Avogadro's number, μ - average mass per number of electrons.

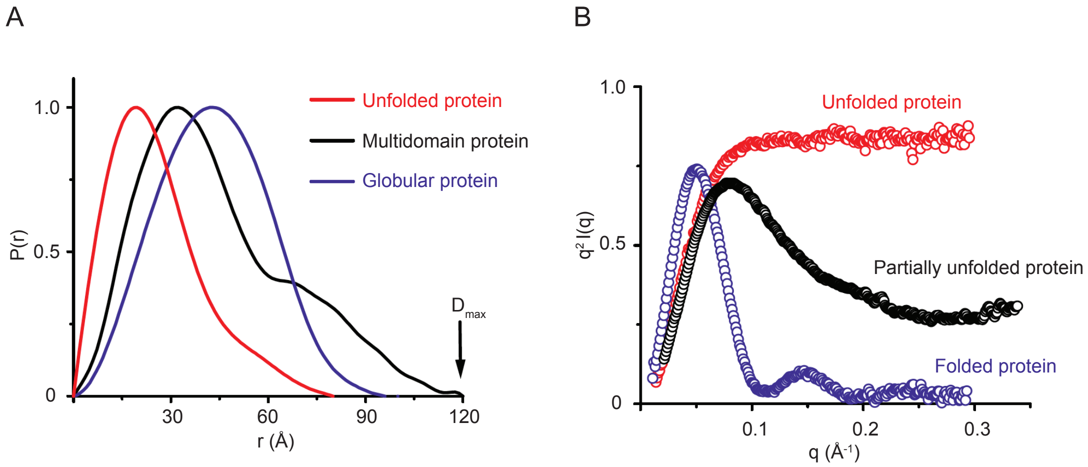


Figure 3.3. The Kratky plot is a method to identify unfolded and folded proteins (A). Depending on the shape of the pair distribution function *P(r)*, elongated and globular proteins can be distinguished (B). Adapted from [Puthnam et al., 2007].

The whole range of *q* can be monitored through the pair distribution function *P(r)*. Several indirect Fourier methods were developed to transform the scattering curve:

$$P(r) = \frac{r}{2\pi^2}\int_0^\infty I(q)q\sin(qr)\partial q \tag{3.5}$$

Hereby, the *I(0)* and R_g can be identified from *P(r)* by introducing the D_{max} parameter, the maximum linear dimension of the scattering particle, when *P(r)* = 0. Moreover, the pair distribution function can be used to recognise extended or unfolded proteins.

$$I(0) = 4\pi\left(\int_0^{D_{max}} P(r)\partial r\right) \tag{3.6}$$

$$R_g{}^2 = \frac{\int_0^{D_{max}} r^2 P(r)\partial r}{\int_0^{D_{max}} P(r)\partial r} \tag{3.7}$$

In addition, IDPs can easily be detected from the Kratky plot (Figure 3.3) which is also one of the most suitable ways to identify the shape of the molecule. According to the Porod law, the volume of the molecule can be calculated from the Porod invariant Q:

$$V = \frac{2\pi^2 I(0)}{Q} \tag{3.8}$$

$$Q = \int_0^\infty q^2 I(q) \partial q \tag{3.9}$$

Globular proteins follow the Porod law and show parabola-like curves, the intensity of the scattering decreases with increasing q. In contrast, an extended protein curve appears as the plateau over the higher q values (Figure 3.3B). This subchapter was based on the review of Puthnam et al., 2007.

For this project, the coupled SAXS and size-exclusion chromatography (SEC) experiment was used, the acquired data were measured at ESRF BM29 beam line and further analysed by ATSAS software [Franke et al., 2017].

3.2. Electron microscopy

Visualisation of big oligomeric structures (> 0,8 MDa) was performed by electron microscopy (EM). Transmission electron microscopy (TEM) is based on the physical phenomenon that the electron beam crosses the thin specimen and after is registered by appropriate detectors. The source of free electrons can be either thermionic or field emission. The imaging occurs when electrons interact with the sample, by generating contrast which depends on the number of atoms: the more the sample is condensed in atoms (and therefore, in electrons), the more the electron scattering will result. As proteins are composed mostly of low electron number atoms (carbon, hydrogen, nitrogen), they need to be stained by heavy metals [Egerton, 2005].

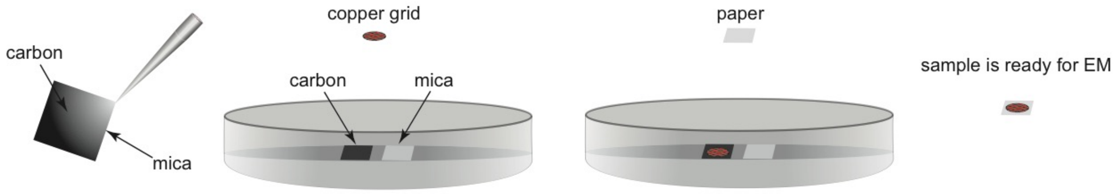


Figure 3.4. The technique of negative stain used in this work. A few micro-litters of the sample are injected between carbon and mica, then the plate is put to a solution of heavy metals. After the copper grid grabs the carbon and gets collected by the piece of paper.

As the microscope is operating under vacuum conditions, the samples also need to be specifically prepared and dehydrated. The technique of negative stain-mica-carbon flotation was used for sample preparation (Figure 3.4). Samples were absorbed to the clean side of a carbon film on mica, stained and transferred to a 400-mesh copper grid. For the staining, I have used Sodium Silico Tungstate (SST) $Na_4O_4SiW_{12}$ at 2% in distilled water (pH 7-7,5). The images were taken under low dose conditions ($<10\ e^-/Å^2$) with defocus values between 1.2 and 2.5 µm on a Tecnai 12 LaB6 electron microscope at 120 kV accelerating voltage using CCD Camera Gatan Orius 1000.

Cryogenic electron microscopy (cryo-EM) allows more detailed structural analysis in the native biological environment and further 3D structural reconstruction. In negative staining samples can be dehydrated which can generate artefacts from removed water, while aqueous samples prepared in cryogenic conditions do not form ice crystals and have stable vapour pressure of water, and, therefore, proteins keep their hydrated native state. In addition, unlike negative staining which is based on the stain penetration to the molecule, cryo-EM permits to obtain full electronic density distribution, which in combination with 3D reconstruction computational methods, is an ideal tool to solve protein structure at near-atomic resolution [Dokland, 2009].

The protocol of sample preparation is as follows. 4 µL of sample (about 1mg/ml) were frozen with a Vitrobot Mark IV (FEI) at 100% humidity and 2 s of blotting time. Quantifoil holey carbon grids were used and glow discharged for 45 s and 25 mA. A Tecnai Polara (FEI) electron microscope working at 300 kV was used to collect movies with a K2-summit direct electron detector (Gatan Inc) with a magnification of 31500x in counting mode. The pixel size at the detector level was 1.21 Å. Automatic data collection have been carried out using Latitude S software (Gatan Inc). 1500 movies of 40 frames (0.15 s each) have been collected. Selection of particles have been done using helix boxer manually and classification carried out using the standard Relion protocol [Scheres, 2012]. Different helical parameters (http://rico.ibs.fr/helixplorer/) have been tested and helicoidal reconstruction performed using Relion. Resolution was limited to about 15-20 Å because of the flexibility of the NC helices and because of the low number of segments.

3.3. MALLS

The multiangle laser light scattering (MALLS) technique coupled with SEC allows the determination of the protein concentration and its average size, and as a result, to determine the molar mass and oligomerisation state of the molecule. The physical background is based on the detection of light scattering of biomolecules in solution at different angles. The detection is achieved using several detectors (Figure 3.5).

The scattered rate $R(\theta)$ can be identified for each detector depending on the angle θ relative to the emitted light beam (Rayleigh ratio):

$$R(\theta) = \frac{I(\theta)r^2}{I_0 \Delta V} \tag{3.10}$$

where initial I_o and scattered $I(\theta)$ light intensity, ΔV – scattering volume, r – distance from the point of scattering to the detector.

When the light interacts with matter, it induces oscillating polarization. The number of oscillating charges is directly proportional to the proton number, and therefore, to the molar mass, that can be calculated from the Zimm equation.

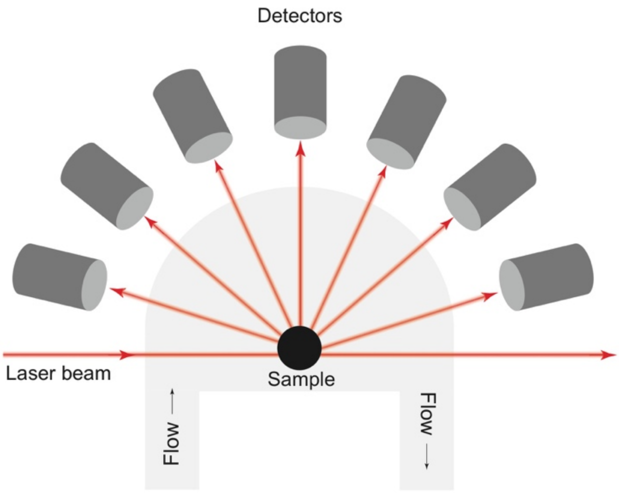


Figure 3.5. The principle of the MALLS experiment. Usually, MALLS is coupled with SEC to allow continuous flow for more accurate reference baseline. Interaction of the laser beam with the sample molecule leads to the scattering off the axis of light beam which is recorded simultaneously by detectors placed at different angles. In addition, the system can contain a refractive index detector for precise concentration determination.

$$\frac{K^* c}{R(\theta)P(\theta)} = \frac{1}{M_W} + 2A_2 c \tag{3.11}$$

where c – protein concentration, A – second virial coefficient describing the interaction straight between the molecule and the solvent, $P(\theta)$ – shape factor, K^* - optical constant which can be expressed by:

$$K^* = \frac{4\pi^2\left(\frac{\partial n}{\partial c}\right)^2 {n_0}^2}{N_A {\lambda_0}^4} \tag{3.12}$$

where λ_o – vacuum wavelength of the light, dn/dc – variation of the scattered light with the respect to $sin^2\theta/2$, n_o – refractive index of the solvent [Wyatt, 1993].

In this thesis, I recorded MALLS measurements at sample concentrations of the proteins of interest in a range from 2 to 10 mg/mL. The choice of the SEC column depended on the previous purification conditions. The experiments were recorded at the biophysical platform (www.isbg.fr) on Wyatt/Hitachi equipment, visualised and analysed by ASTRA software.

3.4. Biochemical materials and methods

3.4.1. Molecular biology

Constructs

Clones encoding for different N and P constructs of Jeryl Lynn mumps virus strain were subcloned to the expression vector from the full-length synthetic gene optimised for *Escherichia coli* (*E. coli*) provided by GenScript company. The fragments of N, P and all N°P fusion complexes were generated:

- N_{FL} (residues 1-549) and $N_{1\text{-}445}$
- Ntail (373-549 and 398-549)
- Ncore ($N_{261\text{-}374}$, $N_{45\text{-}374}$ with GST protein and $N_{1\text{-}261}$, $N_{1\text{-}374}$, $N_{1\text{-}398}$)
- P_{FL} (1-391)
- P_{NTD} (1-215)
- P_{CTD} (272-391)
- P_{XD} (343-391)
- N°Ps ($P_{1\text{-}141}$, $P_{1\text{-}215}$, $P_{1\text{-}90}$, $P_{1\text{-}150}$, $P_{90\text{-}150}$, $P_{90\text{-}215}$, $P_{1\text{-}65}$ all with N_{FL} and $P_{1\text{-}65}$ $N_{32\text{-}549}$)
- V

Vector

The expression vector used for all clones was the DNA plasmid pET-41 c(+). It has the size of 5933 nucleotide base pairs (bp), T7 polymerase promoter/terminator and multiple cloning sites for incorporation of the gene of interest, Kanamycin bacterial resistance gene and solubility GST-tag. All the constructs were cloned into *NdeI/SpeI* and *NcoI/XhoI* sites. The resulting protein ends with a non-cleavable 8-His tag (Figure 3.6). Only V protein was ordered from GenScript to subclone to the same vector, it had TEV-cleaved His-tag on its N-terminus and stop codon before *XhoI* site.

Subcloning procedure

First, the inserts were generated and amplified by polymerase chain reaction (PCR) from the gene-template and DNA primers. Then, the protocol for the PCR reaction was applied as follows: mix the template of about 50 ng, forward and reverse primers of about 0,1 μM (from IDT company), 1x of Phusion Master Mix (Thermo Scientific) and incubate in the thermocycler with the following program (Table 3.1).

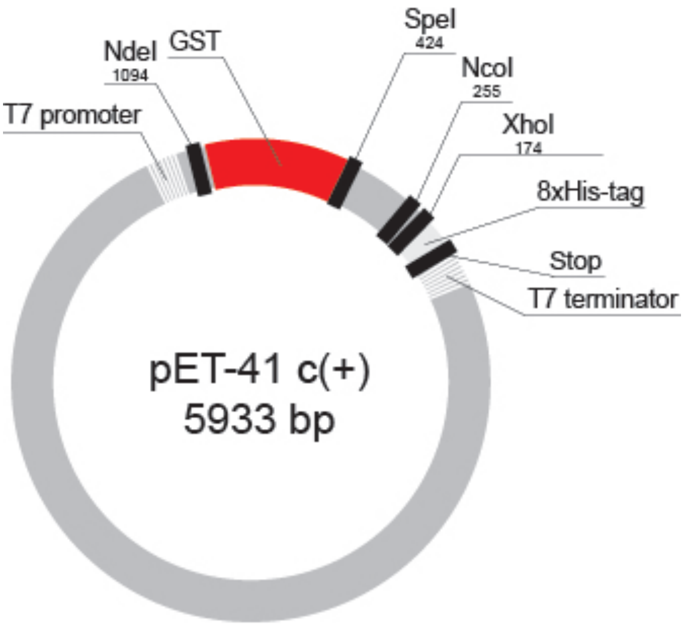


Figure 3.6. Schematic representation of the vector pET-41 c(+) with restriction sites and stop codon (in black rectangles) and their relative position in bp.

Afterwards the generated PCR products were treated with *DpnI* enzyme (all enzymes from Thermo Scientific company) during 1 hour at 37°C in order to remove the template. *DpnI* has the cleavage recognition sequence GATC when A is methylated, which is present in most plasmids transformed to *E. coli* strains. The resulting insert was purified by Nucleospin Gel and PCR Clean Up kit distributed by Macherey Nagel company.

Insert and vector were digested with *NdeI* and *XhoI* restriction enzymes at 37°C during 1 hour. For the constructs that needed the GST tag, the cleavage occurred between *SpeI* and *XhoI* restriction sites. The pET-41 c(+) digestion leads to the generation of two fragments of 5019 and 920 bp for *NdeI/XhoI*, 5094 and 839 bp for *NdeI/NcoI* and 5263 and 670 bp for *SpeI/XhoI* that were separated by the agarose gel migration and purified from gel with the kit.

Table 3.1. Thermocyclic program for the inserts generation.

Step	*Temperature, °C*	*Time*	*Cycle*
Initial Denaturation	98	30 s	1
Denaturation	98	7 s	30
Annealing	$T_m^* + 3$	30 s	
Elongation	72	30 s/1kb	
Final Elongation	72	10 min	1

---

* T_m is the melting temperature for DNA primer

The ligation reaction by *T4 ligase* occurred at room temperature (RT) over night with the specific ratio between the insert and vector which can be determined by:

$$[insert] = \frac{S_{vector}}{S_{insert}} \times [vector] \tag{3.13}$$

where [] - concentration of the insert/vector (in ng/μl), S - size of the insert/vector (in bp). The resulting ligation product was transformed to *E. coli* XL-1 Blue strain, incubated at 37°C overnight on Lysogeny Broth (LB) and agar Petri plate with the addition of 30 mg/l of Kanamycin. Bacterial colonies were picked and grown in LB during one night, and then the plasmids were purified with the help of NucleoSpin Plasmid kit from Macherey Nagel. To verify the successful subcloning, DNA sequencing was performed by the company Genewiz (www.genewiz.com).

The N^{o}Ps were generated first by introduction of the P peptide into the vector between *NdeI/NcoI* sites, and then the N fragment with a TEV protease cleavage site on its N-terminus by *NcoI/XhoI.*

3.4.2. Protein expression

Three types of protein labelling were used for NMR experiments: natural abundance $^{14}N/^{15}N$ and $^{12}C/^{13}C$ (unlabelled), single isotope labelling by ^{15}N and double labelling by ^{15}N and ^{13}C.

Expression vectors were transformed into the *E. coli* strain BL21 (DE3). Depending of the labelling, the protein expression occurred in unlabelled LB medium or in $^{15}N/^{15}N$-^{13}C labelled M9 media.

For unlabelled protein expression, the over-night bacterial preculture of 25 mL was diluted in 1L of LB with the addition of the antibiotic (Kanamycin 30mg/l). After 4-5 hours of bacterial growth at 37°C and reaching the optical density about 0.7-0.8 at 600nm (Thermo Scientific absorbance spectrometer), the culture was induced by 1mM IPTG (isopropyl-D-1-thiogalactopyranoside) and incubated for over-night expression at 20°C for about 15-18 hours. The resulting bacterial pellet after cell harvesting was frozen at -80°C.

M9 minimal medium, used for labelled protein expression, is composed of 22 mM KH_2PO_4, 48 mM Na_2HPO_4, 8,5 mM NaCl, 18 mM ^{15}N-labeled NH_4Cl, 1x MEM Vitamins solution (Gibco), 2 mM $MgSO_4$, 100 μM $CaCl_2$, Kanamycin 30 mg/l, 2g/l of unlabelled or ^{13}C-labeled glucose.

A specific protocol was used for isotope protein labelling. Initially, the cells were separated from over-night maxi-preculture of 100 mL and resuspended in 1L of M9. To reach bacterial optical density at 600 nm of about 0,6-0,8, usually the bacterial growth in M9 takes over 8 hours. The conditions of the IPTG induction and expression were the same as for LB medium.

3.4.3. Protein purification

The bacterial pellet was resuspended in 50 ml of 20 mM Tris-HCl 150 mM/500mM NaCl 0,2 mM TCEP (tris(2-carboxyethyl)phosphine) pH 8 buffer supplemented with EDTA-free (ethylenediaminetetraacetic acid) complete anti-protease (50x) from Roche. Cells were lysed using a sonicator Sonics, then centrifuged at 18000 rpm at Thermo Scientific Sorvall Lynx 4000/6000 centrifuge. The supernatant was filtered with 0,2 μm pore filter and introduced to the Ni-NTA column equilibrated with the lysis buffer. After the washing step, the protein bound to the resin was eluted with 20 mM Tris-HCl 150 mM/500 mM NaCl 400 mM imidazole 0,2 mM TCEP pH 8.

The imidazole was eliminated by dialysis against lysis buffer over night at 4°C or few hours at RT. After the sample was concentrated with the help of Amicon filters, and injected to the size-exclusion chromatography (SEC) column equilibrated with NMR buffer: 50mM Na_2HPO_4/NaH_2PO_4, 150 mM/500 mM NaCl, 2mM DTT pH 6.

The fractions containing the protein were pooled together and concentrated until 200 μM or more and stored at -80°C. The purity of each purification step was verified by SDS-PAGE.

Purification of N_{FL} ring-like structures and nucleocapsids

All the purification steps occurred with the addition of 500 mM NaCl. Having high molecular mass (62.8 kDa of the single unit x 13 + RNA 78-mer = about 855 kDa), the N-rings after Ni-affinity step were purified with Superose Increase 6 10/300 GL column that is able to separate the oligomers from 5kDa to 5MDa. Protein concentration was measured with the help of the Pierce BCA kit (Thermo Scientific).

In order to form nucleocapsids, the N-rings were treated by addition of trypsin (for every 200 mg of the protein 1 mg of trypsin) at 37°C during 1 hour. According to the literature [Severin et al., 2016], the digestion occurs between Ncore and Ntail.

Purification of Ntail

The lysis, Ni-affinity and gel-filtration was performed under low salt (150 mM NaCl). This condition was chosen to get better resolution on NMR measurements. For SEC purification, the Superdex S75 colum was used.

Purification of other N constructs

Constructs $N_{1\text{-}445}$ and $N_{1\text{-}374}$ were purified with the same purification protocol as N_{FL}. Ni-affinity purification of $N_{1\text{-}261}$, $N_{45\text{-}374}$, $N_{261\text{-}374}$ was completed with high salt buffers (500 mM NaCl). $N_{1\text{-}261}$ was further purified with the help of Superdex S200 column.

Purification of P constructs

All P fragments (P_{FL}, P_{NTD}, P_{CTD} and P_{XD}) were purified under 500 mM NaCl, to increase the stability of the proteins, as under low NaCl concentration they undergo the fast precipitation. Superose Increase 6 was again used for tetrameric P_{FL} purification. P_{NTD} was purified using Superdex S200, P_{CTD} and P_{XD} using Superdex S75.

Purification of N^oP

Several N^oP constructs were produced in order to purify stable monomeric complexes. The P_{NTD} truncations were coexpressed with full-length N. All purification buffers were supplemented with 500 mM NaCl to improve the purification profile and protein stability. The dialysis against lysis buffer was preceded with a TEV cleavage reaction at RT in order to covalently separate P sequence from N. The gel-filtration purification was carried out on a Superdex S200 column.

Purification of V protein

The specific protocol of denaturating conditions was used for V protein purification based on the work of Salladini et al., 2017. First, the bacterial pellet was resuspended in 20 mM Tris-HCl, 500 mM NaCl, 6M guanidine-HCl pH 8. Elution of Ni-affinity tagged protein was carried out with 20 mM Tris-HCl, 500 mM NaCl, 3M guanidine-HCl pH8 supplemented with 250 mM imidazole. Several steps of over-night dialysis to the native conditions were performed at 4°C to buffers containing 20 mM Tris-HCl, 500 mM NaCl, 1μM $ZnSO_4$, 1 mM BME (β-mercaptoethanol) 1,5 mM urea, 2 mM cysteine, 0,2 mM cysteine, 10% glycerol pH 8 and then next day to 20 mM Tris-HCl, 500 mM NaCl, 1μM $ZnSO_4$, 1 mM BME (β-mercaptoethanol) pH 8. As V protein had a removable His-tag, it was further cleaved with TEV at RT and purified with the help of Superdex S200 column to the high salt NMR buffer with 1 μM $ZnSO_4$.

3.4.4. Protein assays

NC assembly from N^oP

For NMR experiments, to the sample containing about 100-150 μM N^oP, necessary amount of 1 mM RNA was added to reach 1:1 molar ratio. For EM imaging, 20 μM of N^oP was incubated over-night at 25°C with 50 μM of RNA. Two RNA sequences were used in NC assembly (both purchased from IDT): poly-adenosine-6 or PolyA6 (rArArArArArA), as most of *Paramyxoviridae* 5′-ends are A-rich, and 5′-end of MuV genome (rArCrCrArArG).

N and P phosphorylation

After recording reference NMR experiment of 150 μM protein supplemented with 50mM MgCl2 and 5 mM ATP (adenosine triphosphate), the necessary amount of kinase (about 6 μl for 130 μl total sample volume) was added to the sample and mixed. Ntail phosphorylation experiments were carried out with CKII kinase (1 μl, from Thermo Scientific) and P_{CTD} with PLK1 (Merck). Preliminarily, kinase susceptibility was predicted using on-line phosphorylation prediction software GPS and NetPhos Server [Xue et al., 2011; Blom et al. 2004].

TEMPO-labelling of P

First P_{CTD} was diluted to about 50 μM in 1 mL, DTT was added to reach about 10 mM final concentration. Sample was then dialysed to the non-reductant containing buffer 50 mM Na-phosphate, 500mM NaCl, pH 7. Buffer was changed until DTT reached less than 1 μM concentration. Sample was concentrated up to 400 μl. TEMPO-maleimide ((2,2,6,6-tetramethylpiperidin-1-yl)oxyl) was prepared from aliquots (1 mmol) under nitrogen atmosphere with addition of about 20 μl of DMSO (dimethyl sulfoxide). In continuous Vortex mixing mode TEMPO-label was added to the protein, the mixture was left for 2 hours and further purified by SEC and transferred back to pH6 buffer with no DTT to get 130 μl final volume. Reduction of the label occurred under the addition of 1,75 μl of Na-ascorbate (from stock of 1M).

Alignment for RDCs

Two types of alignment were used in this work: phages and liquid crystals. Phages (Pf1 from ASLA Biotech company) were added gradually to reach about 10 mg/ml. Liquid crystals were prepared from 112,5 µl of protein, 6,5 µl of polyethylene glycol (PEG) and gradual addition of 0,3 µl of hexanol in order to pass through phase separation and get a homogeneous transparent solution with significant viscosity which was further tested by NMR in order to determine the alignment level.

CHAPTER 4.
RESULTS. NUCLEOPROTEIN

4. RESULTS. NUCLEOPROTEIN

Mumps virus nucleoprotein plays an important role in successful viral infection by encapsidating and protecting the RNA genome, a process specifically performed by its structured part (Ncore). However, the function of its long predicted to be disordered region (Ntail) remains unknown. No atomic resolution structures or ensembles of Ncore and Ntail are available for the moment (December 2018). In this chapter, I will try to investigate the structure of isolated Ntail, Ncore and full-length nucleoprotein.

4.1. Structural regions of nucleoprotein

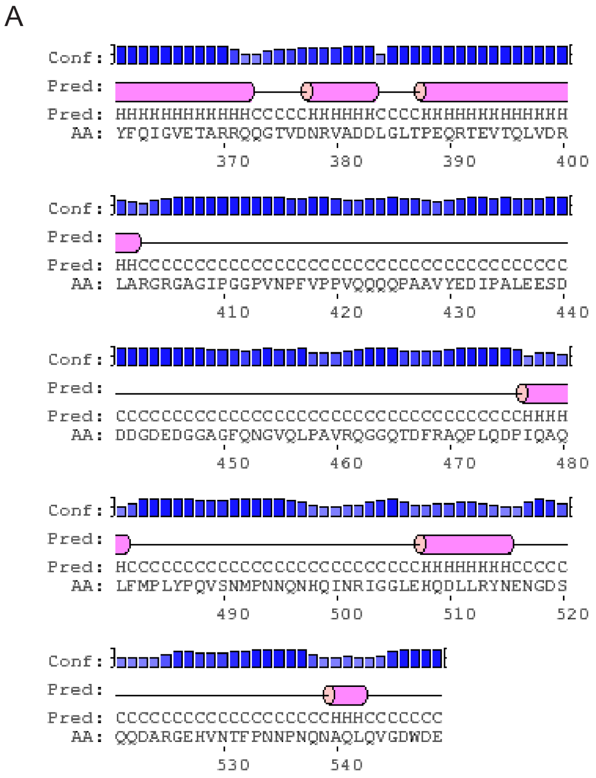
Mumps virus nucleoprotein (MuN) is composed of 549 amino acids. According to the IUPred disorder prediction (Figure 1.24) and to the literature, structured Ncore is defined between residues 1 to 373, and residues 373 to 549 are believed to be intrinsically disordered (Ntail) with unknown function. Other protein structure predictors such as PSIPRED [Buchan et al., 2013], reveal in detail the helical propensity in Ntail (Figure 4.1). Despite the fact that Ntail does not bind to phosphoprotein XD [Kingston et al., 2014], it might still have transient regions with helical propensity which could be responsible for the potential interaction with other viral proteins, and therefore, reveal Ntail function.

However, alignment of MuN sequence with PIV5 N, which has 50,6% sequence identity, showed evidence of the Ncore ending in the position about 398. This means that the region from 373 to 398 likely corresponds to the C-arm by analogy to PDB PIV5 nucleoprotein oligomeric structure. Both Ntail truncations 373-549 and 398-549 were expressed, purified and examined by NMR. In the following results Ntail was studied in the isolated form and in the context of oligomeric structures: ring-like structures and nucleocapsids.

4.2. Ncore

MuV Ncore itself has about 65% sequence identity with PIV5 Ncore which can therefore be a good model for further structural analysis, while Ntail has different length and less similarity. Expression of Ncore alone (1-374 and 1-398) lead to the formation of big unstable oligomeric structures (observed in EM) that were difficult to purify. Nevertheless, in the literature there are several cases describing the expression of soluble Ncore truncations [Milles et al., 2018; Pereira et al., 2017]. Several truncations of Ncore were cloned into an expression vector: N-lobe (residues from 1 to 261), C-lobe with GST (261-374), N- and C-lobes without N- and C-arms with GST (45-374).

Only N-lobe was expressed in a soluble fraction and purified. Unlabelled N_{1-261} was tested for interaction with ^{15}N-labelled phosphoprotein N-terminal part but no signs of binding were observed. In addition, peak distribution on ^{1}H-^{15}N HSQC spectrum of N-lobe was more characteristic of a disordered protein (data not shown). This indicates a possible protein misfolding that probably lost its functional activity.



B

MGTVDNRVAD	382
DLGLTPEQRT	392
EVTQLVDRLA	402
RGRGAGIPGG	412
PVNPFVPPVQ	422
QQQPAAVYED	432
IPALEESDDD	442
GDEDGGAGFQ	452
NGVQLPAVRQ	462
GGQTDFRAQP	472
LQDPIQAQLF	482
MPLYPQVSNM	492
PNNQNHQINR	502
IGGLEHQDLL	512
RYNENGDSQQ	522
DARGEHVNTF	532
PNNPNQNAQL	542
QVGDWDE	549

Figure 4.1. PSIPRED structure prediction for MuN (shown only for Ntail 373-549) (A). Blue bars indicate the level of confidence of prediction, magenta cylinders and letter H – predicted α-helices, lines and letters C – random coil, AA is for the amino acid protein composition. Helices were predicted in the regions 377-383 and 388-402 (C-arm), 477-481 and 508-515 can be considered as the potential MoRE. (B) The amino acid composition of Ntail, rich in glutamines (Q in red) and prolines (P in green).

4.3. *Intrinsically disordered Ntail*

4.3.1. *Ntail NMR spectral characterisation*

As mentioned before, two Ntail fragments were studied in parallel. Both had HSQC spectra characteristic of IDPs. It was decided to focus on the longer form (373-549) to which Iwill refer to as Ntail. After the addition of 2 residues from the cloning restriction site (residues M on N-terminal, L and E on C-terminal) and 8 residues Histidine-tag in C-terminal end, the Ntail protein counts 185 amino acids.

The purified construct was assigned using standard triple-resonance assignment experiments as described earlier (Figure 4.2). The first NMR spectra were acquired in physiological conditions (buffer with 150 mM NaCl). However, for practical reasons (P_{CTD} is not stable at low salt) interactions with potential partners, Ntail was further dialysed to high salt buffer containing increased NaCl concentration (500 mM) and assigned. The assignment of N-rings was also performed in 500 mM buffer.

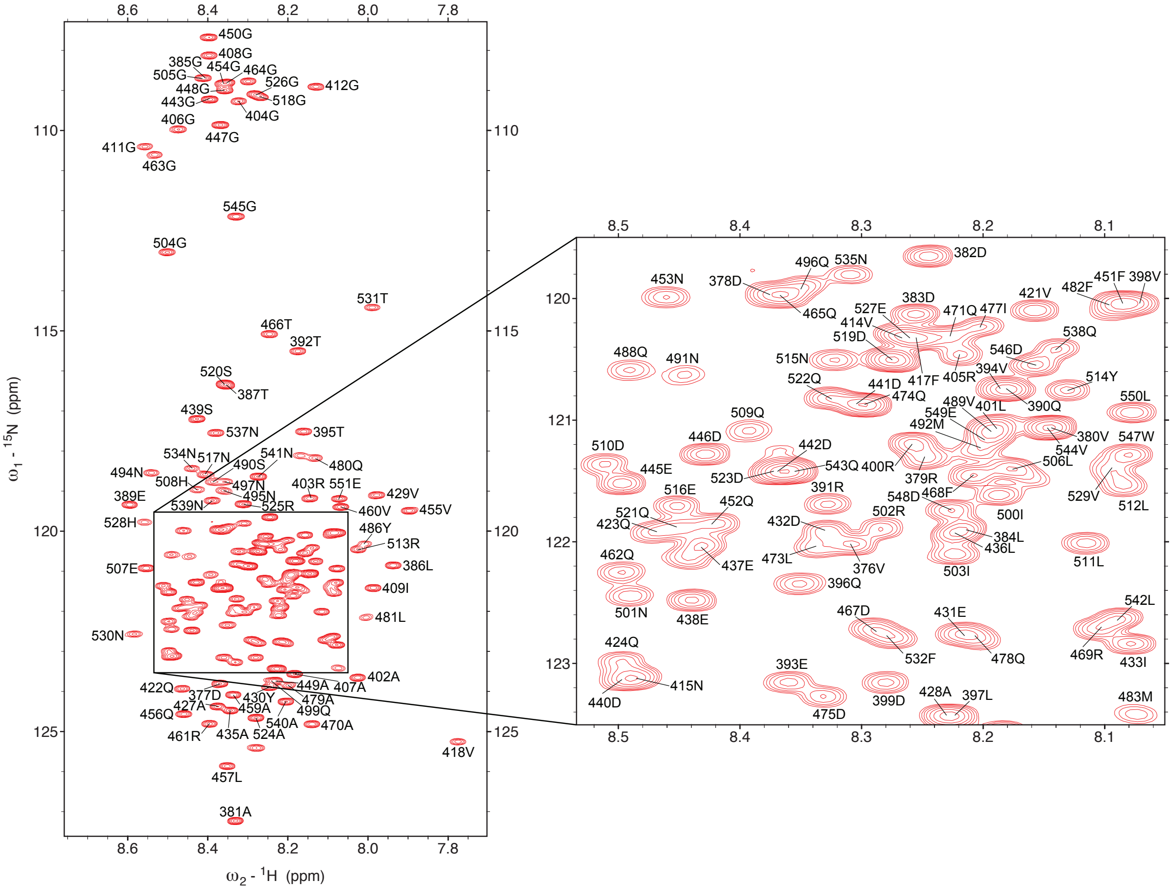


Figure 4.2. Spectral assignment of ^{1}H-^{15}N HSQC of Ntail (373-549) at 25°C and 600 MHz in 50 mM Na-phosphates, 500 mM NaCl, 2 mM DTT at pH 6. DDT was added to all samples with or without cysteins in order to use them for interaction essays. Peaks from glutamine, asparagine, arginine and tryptophan side chains are not shown.

Secondary structure propensities (SSP) were calculated from C_α and C_β chemical shifts. The region with elevated SSPs between residues 373 and 405 revealed a helical conformation (Figure 4.3A) which can be explained theoretically by the close proximity of these amino acids to the Ncore and probable formation of C-arm. The last 7 residues were found having higher than 0,2 SSPs that can be explained by the presence of histidine-rich tag or the formation of transient binding site for a partner such as the matrix protein. Other regions had negligible SSPs. The presence of the helix in the N-terminal part of Ntail was visualised on the ensemble calculation with ASTEROIDS (Figure 4.3B). Only the CS from C_α are shown, as C_β, C′, N and H_N

were less informative. Ramachandran space that indicates the distribution of protein dihedral torsion angles ϕ/ψ can be derived from conformational sampling of the selected ensemble. The population of each possible secondary structure can be ploted as a function of residue (Figure 4.3C): β-strand (βS), left-handed polyproline II helix (βP), α-helix (αR) and left-handed helix (αL). It is possible to highlight the highly αR populated regions discussed before: on N- and C-terminal part of Ntail, less populated α-helix were found for residues in the potential concerved MoRE domain (440-530), supposedly located in the middle of Ntail sequence.

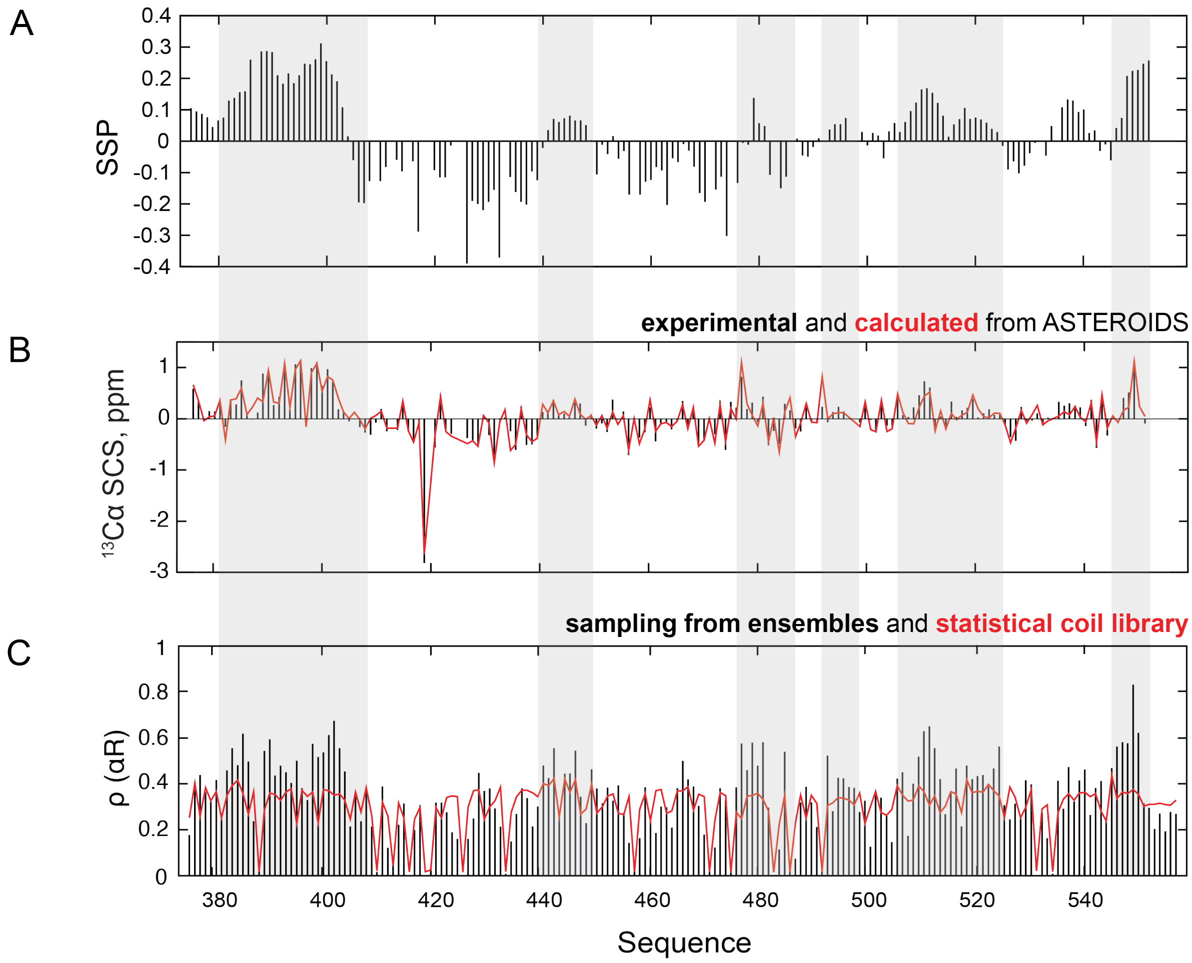


Figure 4.3. Ntail SSPs (A) and C_α secondary chemical shifts (SCS) calculated from ASTEROIDS ensemble (B) based on chemical shifts as an input (the initial pool contained 10000 structures, the selected ensembles 200 conformers). Population plot ρ (C) corresponding to α-helix (αR) derived from the ASTEROIDS ensemble and selected from statistical coil library (database of dihedral angles used during protein statistical coil generation). Grey rectangles indicate protein regions with potential secondary structure conformation.

Measuring $R_{1\rho}$ did not show the presence of "strong" secondary structure conformations. However, there are slightly higher relaxation rates on residues from 385 to 408, from 450 to 464, from 474 to 482, from 505 to 518 and few residues on Ntail C-terminal end. This complements the previous experimental and computational data about C-arm and potential MoRE domains. In addition, similar values measured at higher magnetic field (at 850 MHz) didn't show any evidence of exchange contributions. In contrast, R_1 relaxation rates are similar throughout the sequence (Figure 4.4).

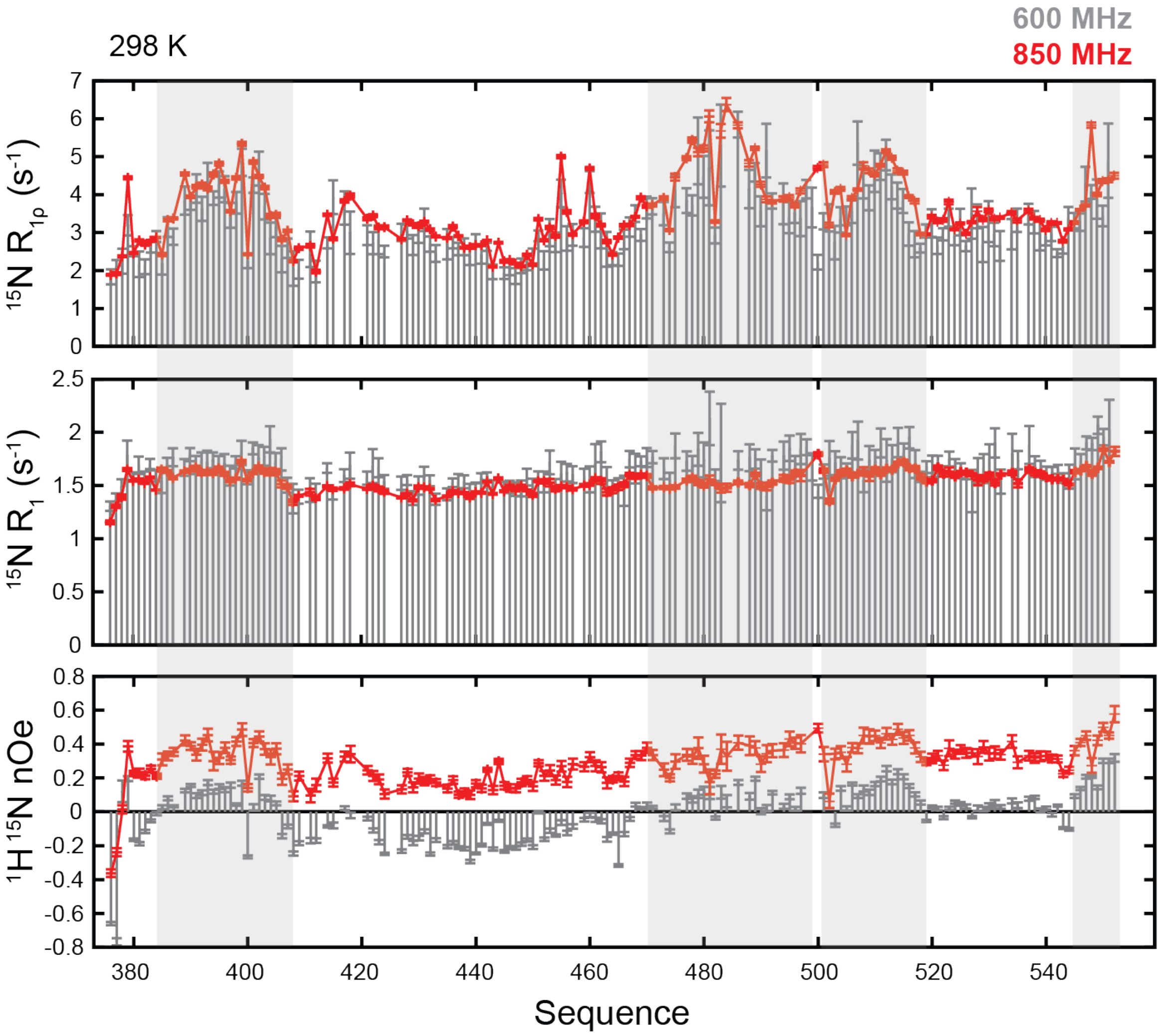


Figure 4.4. Dynamics of Ntail. $R_{1\rho}$, R_1 and hetNOE profiles were measured at two different magnetic fields, they describe the conformational behaviour in the ps to ns time scale. Grey rectangles indicate the position of potential transiently/partly structured regions.

Positive hetNOE values (at 600 MHz) indicates a decrease in flexibility in pico to nano seconds timescale in regions 385-405, 475-500, 500-518, 545-557 (Figure 4.4) which is caused by the presence of helical propensity. Other protein sequence regions display low hetNOE which is usually observed for high flexibility of random coil conformation.

Additional information was extracted from RDCs after Ntail alignment (Figure 4.5). Three types of RDCs were measured. Surprisingly, high $^1D_{NH}$ values were observed for residues 468-498 with sufficiently high dipolar waves in the N-terminal part of Ntail (C-arm) (373-406) and relatively small on the other region I suggested to be a potential MoRE (residues 500-512). Dipolar waves (variations in the dipolar constant indicating the different from parallel orientation of the helix vectors in the direction of the magnetic field [Mesleh et al., 2002]), were also observed on $^1D_{C_\alpha H_\alpha}$ for C-arm and residues 468-498, only positive values were present at the region with 500-512 amino acids (Figure 4.5). This finding is very promising in Ntail structural investigation to detect strong helical propensity of potential MoRE domains. The "invisibility" of those regions can be explained by the presence of prolines (P) and glutamines (Q) in Ntail protein sequence (Figure 4.1). The last RDC experiment $^1D_{C_\alpha C}$ was less informative because of signal to noise (not shown). The Ntail C-terminal end was not affected much in the alignment.

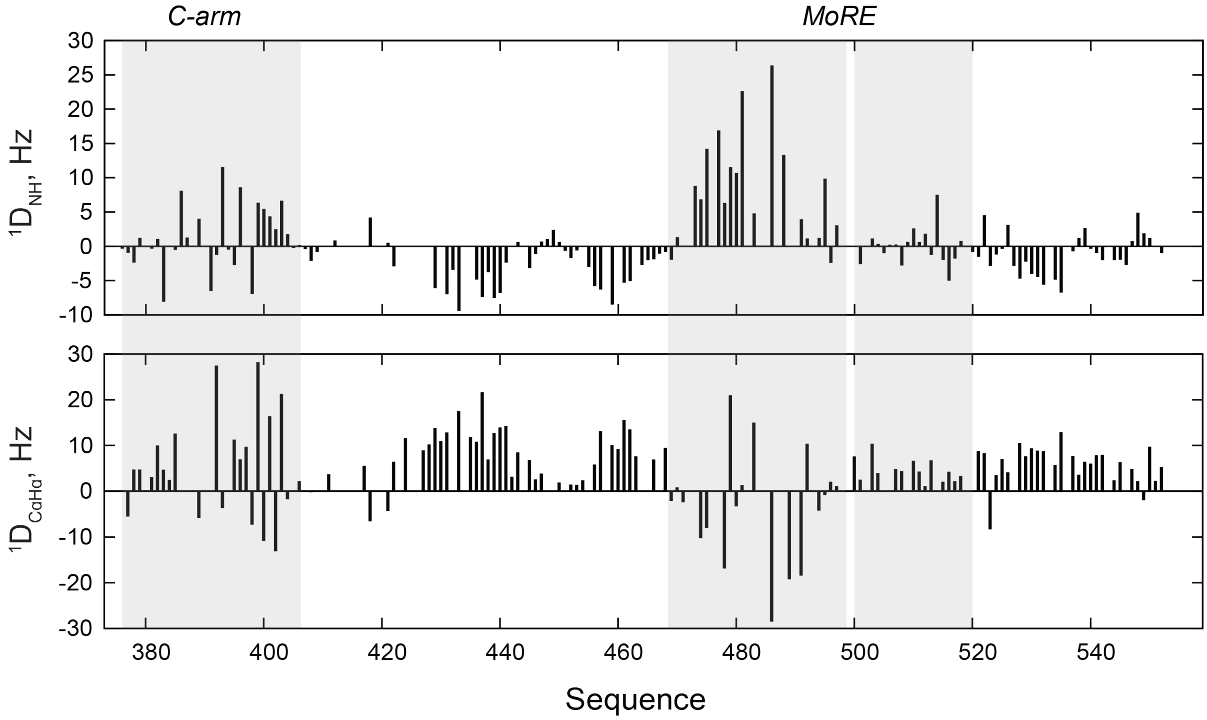


Figure 4.5. $^1D_{NH}$ and $^1D_{C_\alpha H_\alpha}$ of Ntail after alignment in PEG-hexanol liquid crystals at 25°C. ^2H quadrupolar splitting of water was 25 Hz. Grey rectangles indicate protein regions with potential secondary structure conformation: C-arm and MoRE domains.

Finally, several studies probing Ntail interaction with different truncations of P were performed. No affinity was found for Ntail to bind phosphoprotein.

4.3.2. Ntail phosphorylation in vitro

According to Zengel et al., MuN is phosphorylated in MuV-infected cells. This was shown by mass spectrometry and radioactive labelling, and further by *in silico* studies. The major phosphorylation site was identified to be located at serine 439. However, the kinase(s) responsible for the modification was not known.

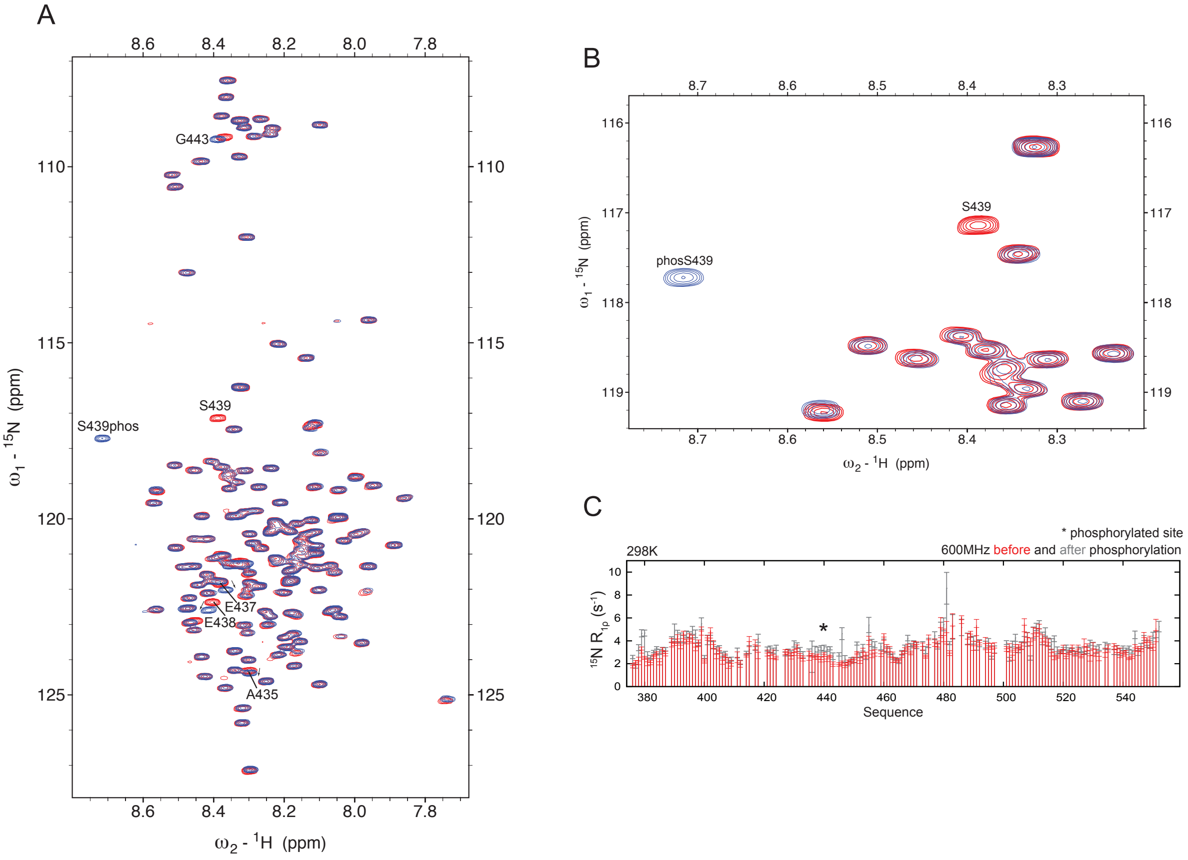


Figure 4.6. Phosphorylation profile of Ntail. Overlay of ^{1}H-^{15}N HSQC spectra before (in red) and after (in blue) phosphorylation (A). Zoom on spectra with phosphorylated site S439 (B). $R_{1\rho}$ profile of Ntail before (in red) and after (in grey) phosphorylation, asterisk indicates the position of the phosphorylated residue (C).

With the knowledge of potential phosphorylation sites obtained from predictors, CKII was used for *in vitro* phosphorylation of the recombinant protein. The residue S439 was phosphorylated as expected (Figure 4.6A, B). The phosphorylated amino acid changed the chemical environment of the close residues which was observed on the HSQC spectra before and after modification. From the $R_{1\rho}$ relaxation profile, the peptide chain in the region close to S439 has slightly higher relaxation rate values (Figure 4.6C) without dramatic change in overall protein conformation and dynamics. As in the publication of to Zengel et al. it was suggested about the creation of the new binding site after Ntail phosphorylation, I performed tests of phosphorylated Ntail with P_{CTD} interaction but it did not show any evidence to support this hypothesis.

4.4. Ntail in the context of RLP and NC

4.4.1. Purification of RLPs

The full-length MuN (N_{FL}) was expressed in the form of oligomeric 13N-units ring-like structures (N-rings) with the incorporated random RNA (Figure 4.7A, B). N-rings appear to form nucleocapsids by stacking onto each other and forming helical structures (Figure 4.7A).

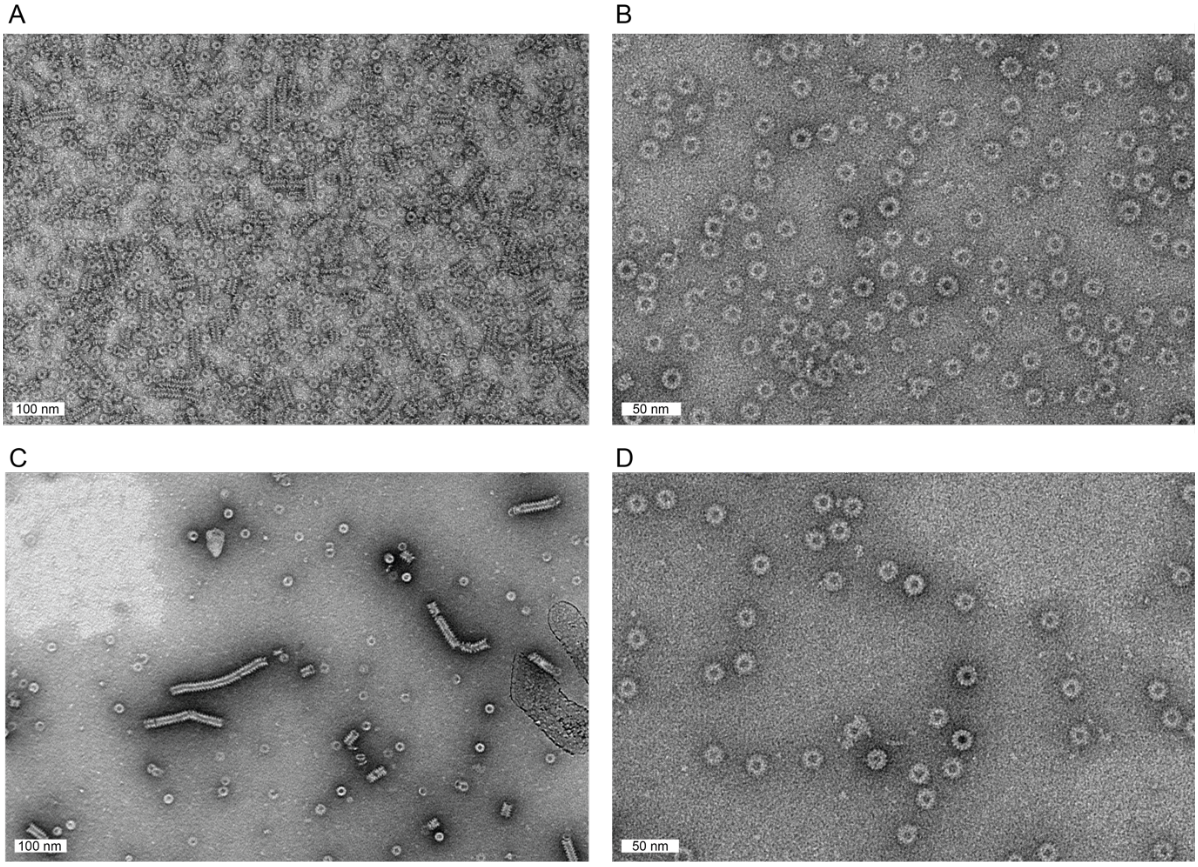


Figure 4.7. Electron micrographs of N_{FL} after bacterial lysis (A), last step of purification (B), treatment with trypsin (C) and RNAse I (D).

Usually, this happens during long storage at 4°C, or much faster at RT. The protein degrades and aggregates by releasing Ntail, as was described in the literature [Severin et al., 2016] and followed under SDS-PAGE analysis (data not shown). The same effect was observed after N-rings proteolysis with trypsin (Figure 4.7C). For this reason, I always used the N_{FL} from frozen aliquots at -80°C. High salt conditions prevent the rings from aggregating with each other. Consequently, all the purification steps were performed with the addition of 500 mM NaCl which allowed the collection of N-rings in a separate fraction (Figure 4.8). N_{FL} after RNAse I treatment did not change the morphology of ring-like oligomeric structure (Figure 4.7D).

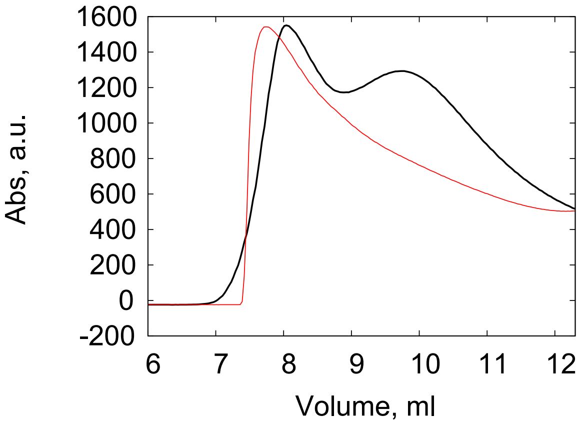


Figure 4.8. Superose Increase 6 10/300 GL column size-exclusion chromatography profiles of RLPs purification in buffer containing 150 mM (in red) and 500 mM NaCl (in black). Fractions from 9 to 11 mL were collected for the N-rings analysis.

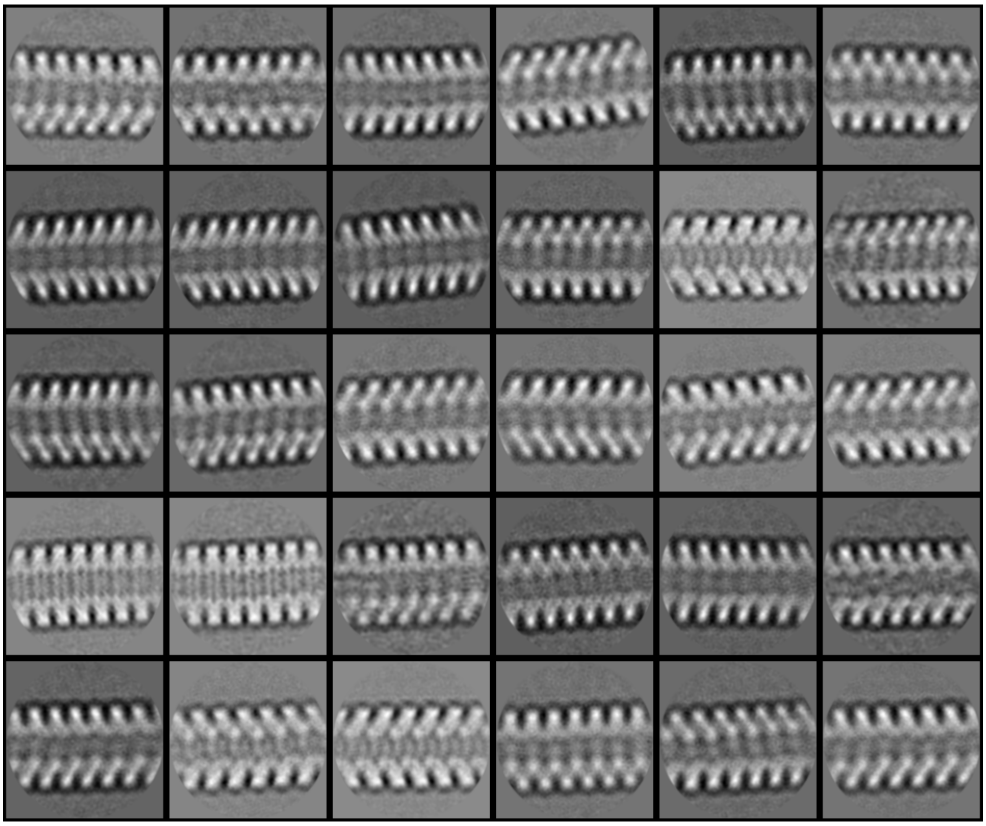


Figure 4.9. 30 classes showing the variability of sizes and shapes of the truncated N_{FL} NC-like structures collected from cryo-EM. Each class is an image of 153*153 pixels and each pixel represents 2.42 Å, 370.26 Å in total. To make each class, about 250 particles were averaged.

N-rings alone and capsids formed from N_{FL} proteolysis were also visualised by cryo-EM. Unfortunately, RLPs images did not have a high enough resolution to be able to draw any significant conclusions from this study. I tried to improve the resolution by decreasing the salt conditions (up to 50 mM NaCl), but the resolution was still very poor. Better resolution was obtained with trypsinated rings that then formed NC-like structures. However, after analysis, about 30 classes of different helical orientations were found (Figure 4.9). This variation decreased the chance of obtaining a high-resolution structure with this protocol of capsid preparation.

Scattering of large N-rings particles and the presence of RNA prevented determination of the exact concentration by UV absorbance. Therefore, I have measured the concentration with the help of BCA assay. The cratio between the concentration inferred from the absorbance and tht inferred from BCA measurements was constant and equal to 3,8.

4.4.2. N-rings NMR characterisation

Despite the size of N-rings, they still give rise to NMR signal due to the flexibility of the attached Ntail domain, as previously shown for MeV nucleocapsids [Jensen et al., 2011; Milles et al., 2016].

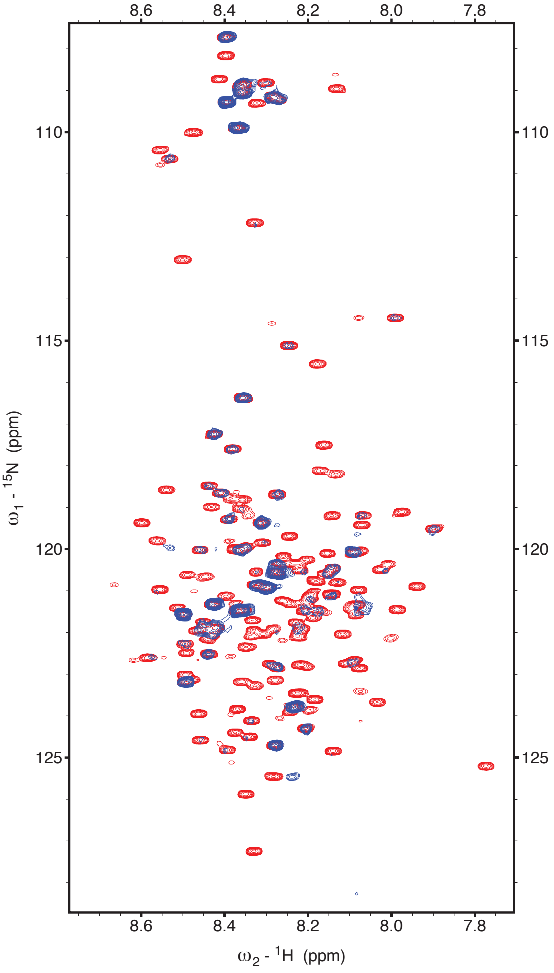


Figure 4.10. Superposition of the ^{1}H-^{15}N HSQC spectrum of isolated Ntail (in red) and N-rings (in blue), both spectra were recorded at 600 MHz at 25°C.

Figure 4.10 represents the overlay of two HSQC spectra from Ntail and N_{FL} allowing me to assign the superposing N-rings signals. The intensity was extracted from visible peaks (residues 430-470 and 495-549) and plotted it against the Ntail sequence (Figure 4.11). Signals from residues close to the Ncore (373-430) and potential MoRE (470-495 and 500-510) have either very small intensities or are absent on the HSQC spectrum.

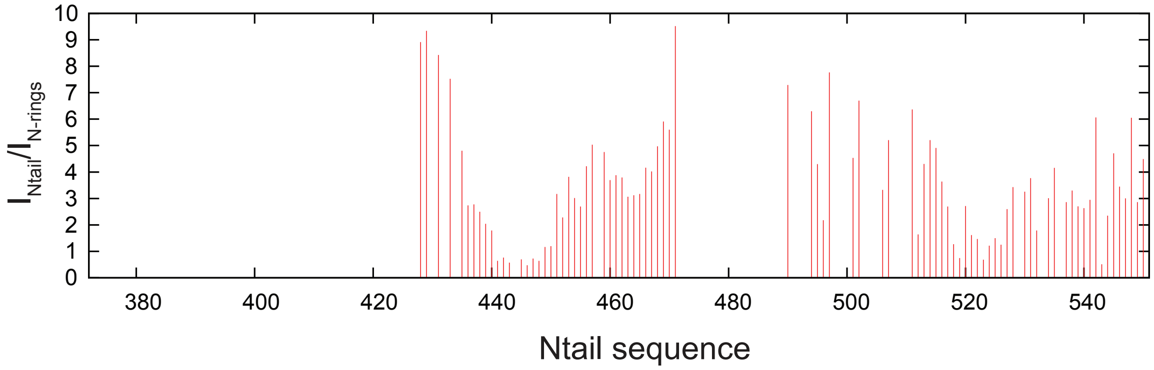


Figure 4.11. Peak intensity profile extracted from HSQC spectrum of N_{FL}, calculated as a ratio between the peak intensities from N_{FL} and Ntail.

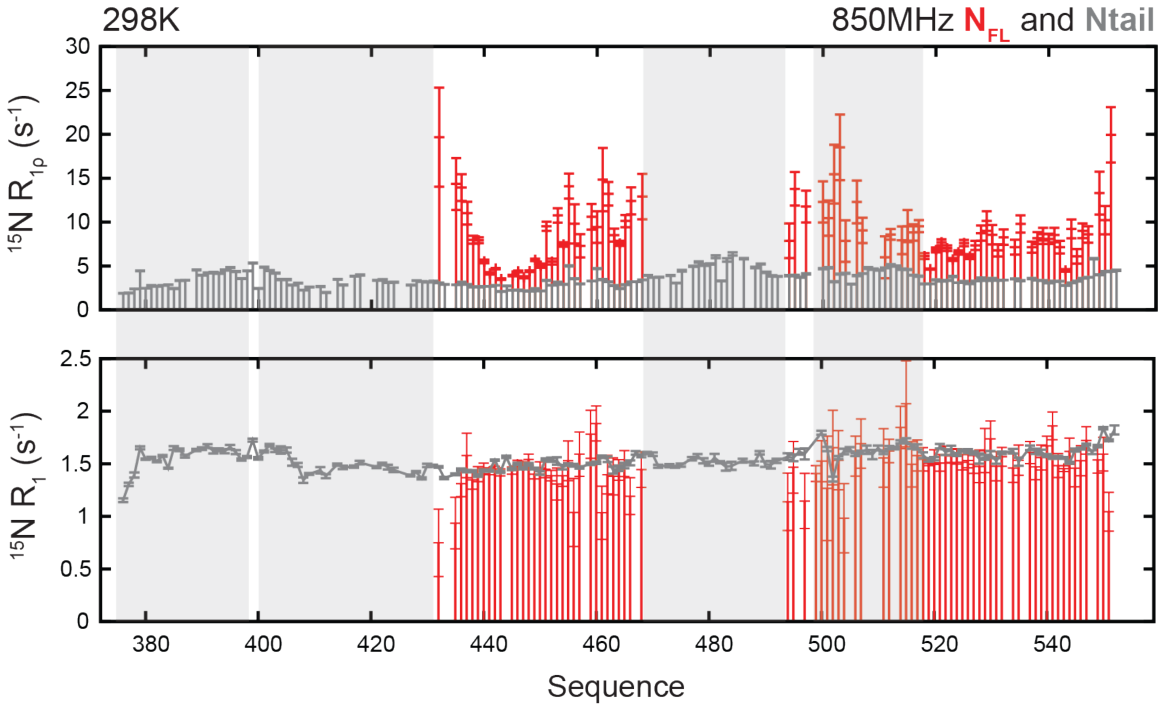


Figure 4.12. Relaxation profile ($R_{1\rho}$ and R_1) of N_{FL} compared to the isolated Ntail recorded at the same magnetic field and temperature. Grey rectangles indicate the position of potential structural regions. Second rectangle is the domain of Ntail in close proximity to the big oligomeric Ncore that, due to the slow molecular tumblinsg, is not visible for NMR.

I have measured N_{FL} $R_{1\rho}$ and R_1 relaxation in order to probe RLPs conformational dynamics and compare it to the isolated Ntail (Figure 4.12). In the similar way to peak intensities, regions supposedly corresponding to C-arm and MoRE are identifiable by the peaks being too broad and hence non-detectable.

4.4.3. Interaction of RLPs with Ntail

To extend the hypothesis of the Ntail MoREs interaction with the surface of N-rings, a mixture containing 10 µM of ^{15}N labelled Ntail and 100 µM of unlabelled N-rings was measured by NMR. Intensities extracted from peaks in the HSQC spectrum (Figure 4.13A) showed the region of 470-495 disappearing apparently because of a contact with N_{FL} (Figure 4.14).

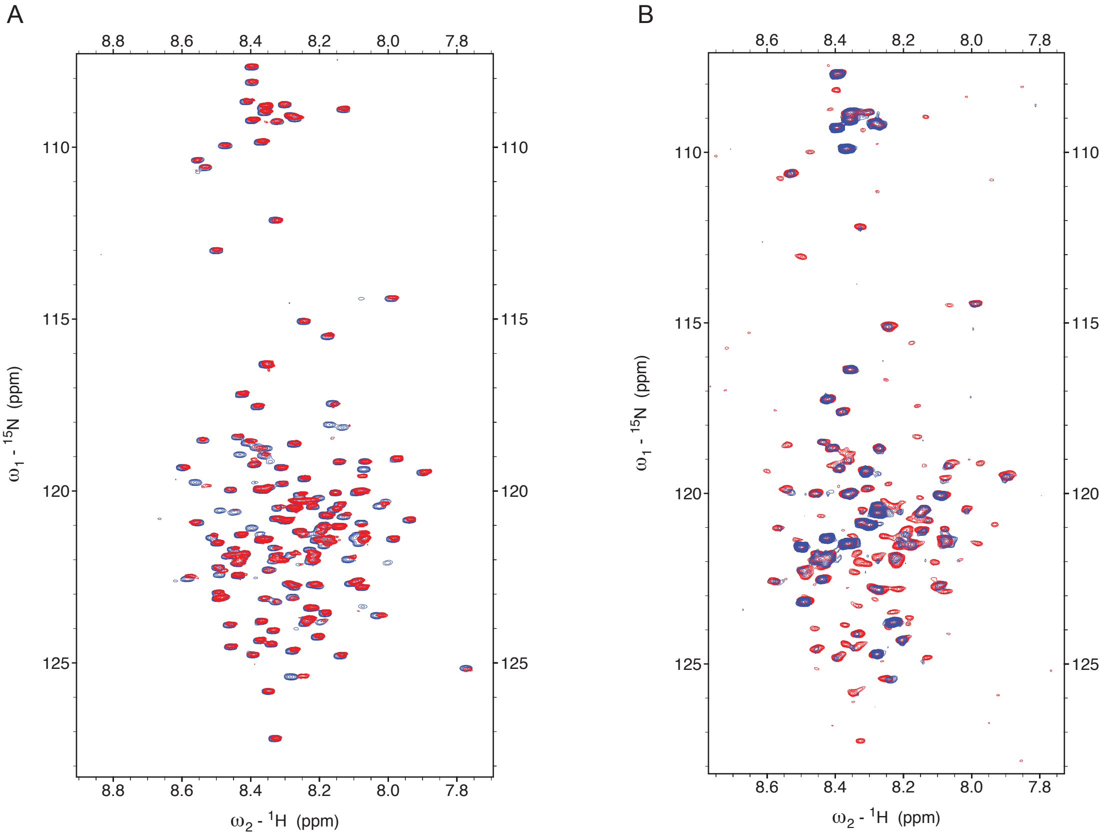


Figure 4.13. Superposition of the ^{1}H-^{15}N HSQC spectrum of isolated Ntail alone (in blue) and the mixture of labelled Ntail and unlabelled N-rings (in red) (A), of labelled N-rings alone (in blue) and with Ntail (in red) (B).

The mirror experiment was acquired by probing the interaction of ^{15}N labelled N_{FL} and unlabelled Ntail with a molar ratio of 1:10. According to the appearance of the new peaks on N_{FL} HSQC spectrum (Figure 4.13B) and their intensities (Figure 4.14), in the presence of an excess of isolated Ntail, Ntail from RLPs gets liberated, confirming a direct interaction of the first MoRE region (470-495) with Ncore. In addition, the region close to Ncore of N-rings Ntail from 373 to 430 also was affected in the interaction. This could be explained by the involvement of C-arm of free Ntail (its N-terminus) in the assembly of N-rings.

In this focus, the intact and trypsin-truncated N_{FL} was incubated with a large fraction (at least 10 times higher) of Ntail. A tendency of uncoiling capsids was observed on EM images (Figure 4.15A, B). In parallel, a new construct N_{1-445} was generated without MoRE region. The protein was unstable and precipitated very quickly, so the purification procedure was stopped with Ni-affinity, indicating the rapid formation of aggregating NCs (Figure 4.15C).

4.4.4. Interaction of RLPs with phosphoprotein and kinase

In the literature, the Ncore region of MuV NCs and N-rings were suggested to bind both N- and C-terminal domains of the phosphoprotein. Some NMR titrations were performed here, but as the N_{FL} HSQC spectrum was not informative for this kind of study due to the absence of Ncore cross-peaks, detailed description about these interactions in the context of P will be provided in the next chapter.

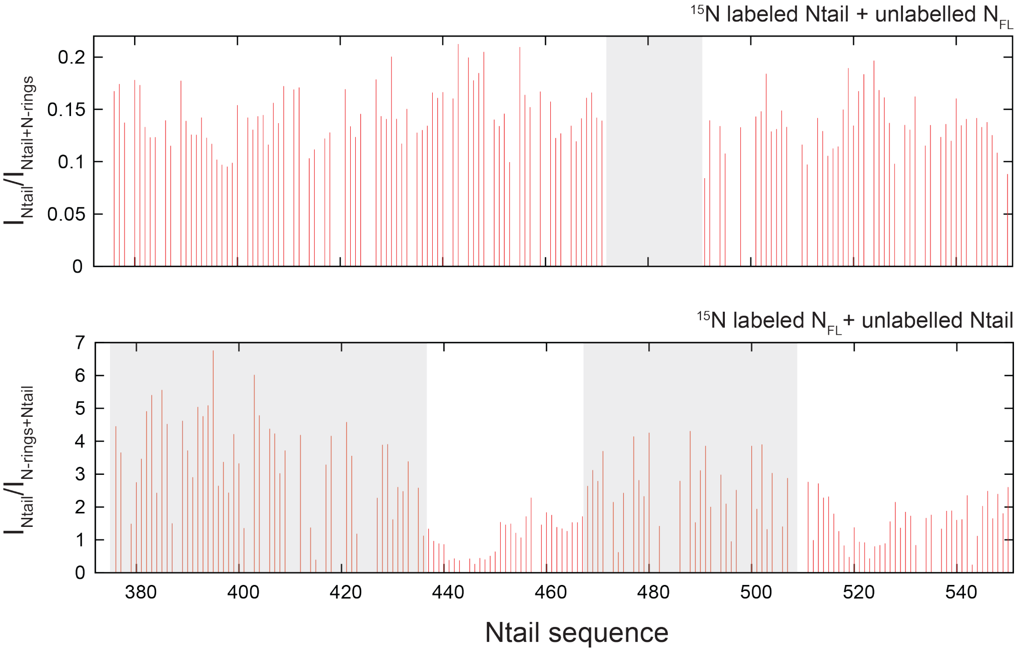


Figure 4.14. Intensity profiles of Ntail and N_{FL} from HSQC spectra from Figure 4.13. They were calculated as the ratio of peak intensities between Ntail and Ntail with unlabelled N_{FL} and between N_{FL} and N_{FL} with unlabelled Ntail.

At the same time, N_{FL} interaction with P was visualised in EM. The binding to P N-terminal or/and C-terminal domains did not show significant changes on N-rings structure. However, adding to the mixture of N_{FL}, P_{NTD} and P_{CTD} RNAse I completely broke RLP structure (Figure 4.15D), while incubation of N_{FL} with RNAse I alone does not influence the N-ring morphology.

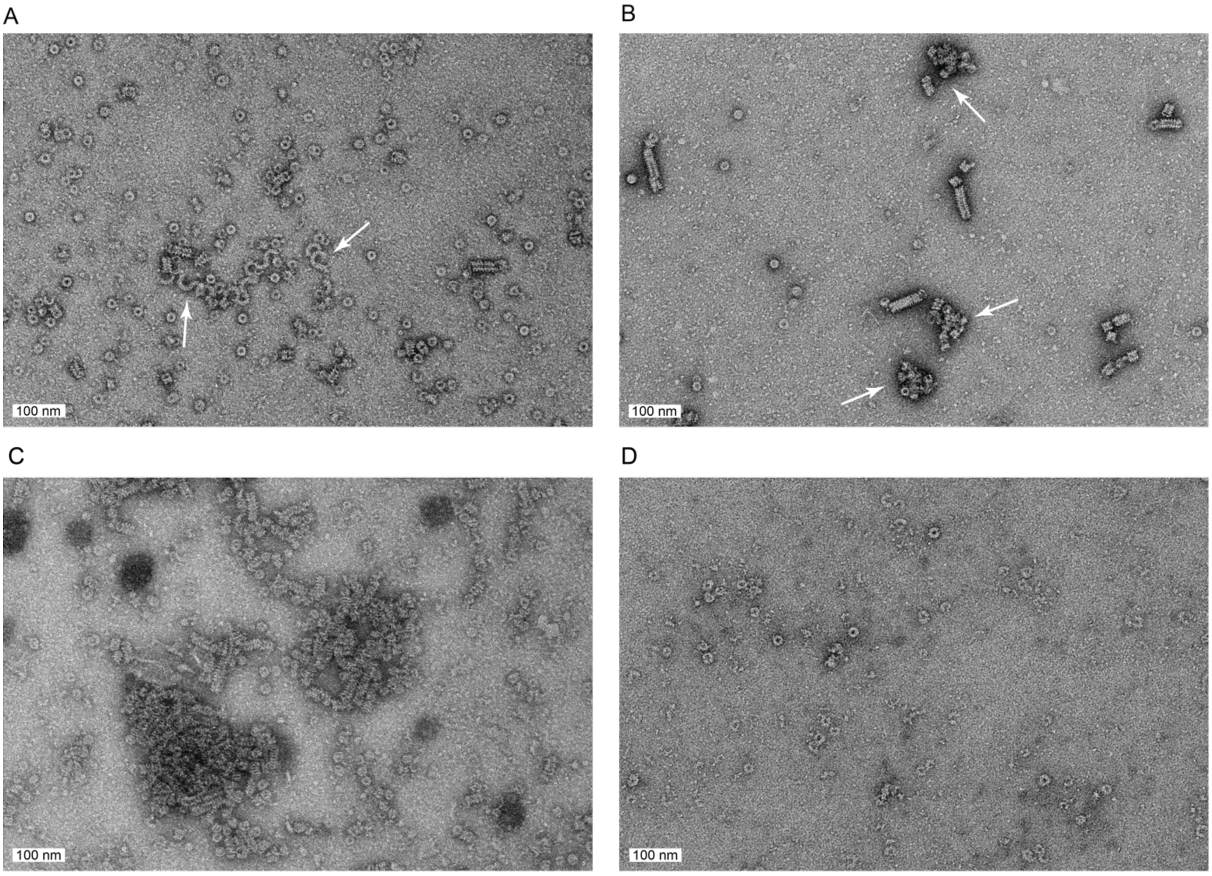


Figure 4.15. Electron micrographs of N_{FL} treated with excess of Ntail (A), with trypsin and Ntail (B) and $N_{1\text{-}445}$ alone (C). N_{FL} after incubation with P_{NTD}, P_{CTD} and RNAse I (D). Uncoiled capsids are indicated by white arrows.

In a similar way as Ntail, N_{FL} was also tested for phosphorylation with CKII. The same residue S439 was modified, however, no other changes in the HSQC spectrum after phosphorylation were observed, showing that the dynamic behaviour and environment of Ntail was not impacted by phosphorylation.

4.4.5. SAXS of RLPs

The principle of coupled SAXS-SEC experiment is the direct visualisation of the protein elution profile in terms of X-ray scattering. The coupled device had a long tube connecting the injection valve and the column, for this reason, the resolution decreased and did not allow separation of the elution peak of N-rings from the void volume in the SAXS profile. Instead both peaks on SEC-chromatography profile were merged, as was observed for N_{FL} purification under low salt conditions (Figure 4.8). In view of the overall quality of the data no further analysis was performed.

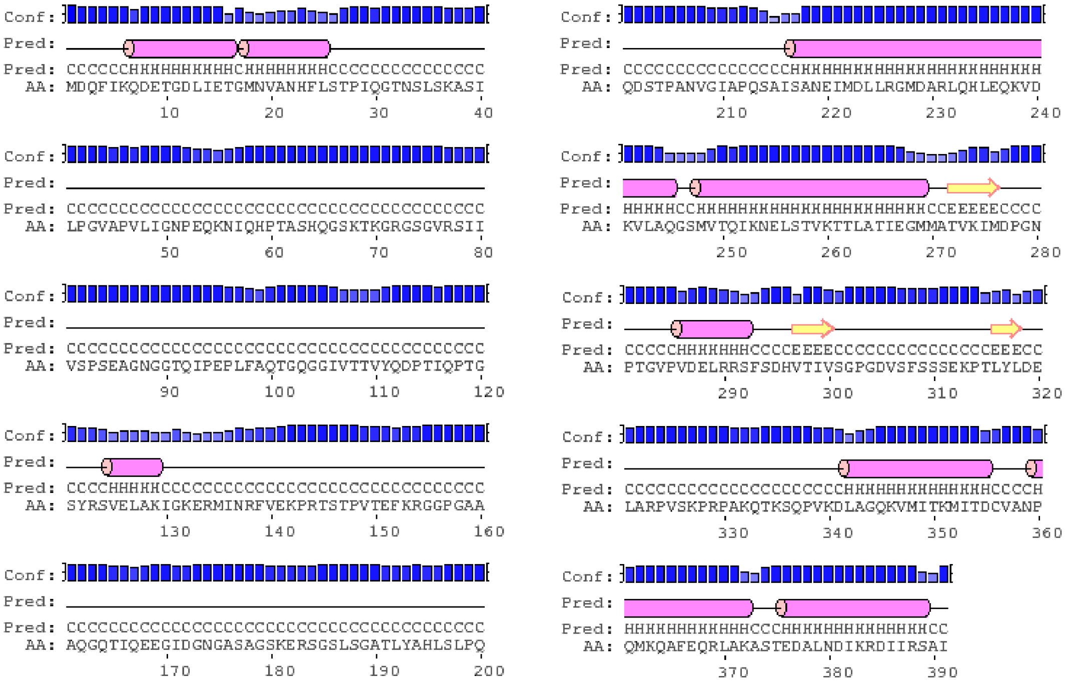
CHAPTER 5.
RESULTS. PHOSPHOPROTEIN

5. RESULTS. PHOSPHOPROTEIN

Although known crystal structures are available for OD and XD domains of mumps virus P, nothing is known about disordered regions and their function. This chapter will be dedicated to the study of these unknown regions of phosphoprotein structure, in particular isolated N- and C-terminal domains of P, and their interaction with N. In addition, for the first time I will introduce findings from mumps virus V protein purification and its NMR characterisation.

5.1. *Phosphoprotein functional domains*

Mumps virus phosphoprotein (MuP) consists of 391 amino acids, and functionally is divided into three parts: the N-terminal domain (P_{NTD}, residues 1-215), the oligomerization domain (OD, residues 215-272) and the C-terminal domain (P_{CTD}, residues 272-391) that includes XD (residues 343-391). According to the structure prediction from IUPred (Figure 1.24A) and PSIPRED (Figure 5.1), P_{NTD} is predicted to adopt a disordered conformation with some transiently structured regions. A disordered linker was identified between folded OD and XD (residues 272-343). The structures of OD and XD were previously solved by X-ray crystallography [Cox et al., 2013; Kingston et al., 2008].



```
Pred: CCCCCCHHHHHHHHHHCHHHHHHHHCCCCCCCCCCCCCCC
  AA: MDQFIKQDETGDLIETGMNVANHFLSTPIQGTNSLSKASI
               10        20        30        40

Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: LPGVAPVLIGNPEQKNIQHPTASHQGSKTKGRGSGVRSII
               50        60        70        80

Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: VSPSEAGNGGTQIPEPLFAQTGQGGIVTTVYQDPTIQPTG
               90        100       110       120

Pred: CCCCHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: SYRSVELAKIGKERMINRFVEKPRTSTPVTEFKRGGPGAA
               130       140       150       160

Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: AQGQTIQEEGIDGNGASAGSKERSGSLSGATLYAHLSLPQ
               170       180       190       200

Pred: CCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHH
  AA: QDSTPANVGIAPQSAISANEIMDLLRGMDARLQHLEQKVD
               210       220       230       240

Pred: HHHHHCCHHHHHHHHHHHHHHHHHHHHHHCCEEEEECCCC
  AA: KVLAQGSMVTQIKNELSTVKTTLATIEGMMATVKIMDPGN
               250       260       270       280

Pred: CCCCCHHHHHHHCCCCEEEECCCCCCCCCCCCCCCEEECC
  AA: PTGVPVDELRRSFSDHVTIVSGPGDVSFSSSEKPTLYLDE
               290       300       310       320

Pred: CCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCCH
  AA: LARPVSKPRPAKQTKSQPVKDLAGQKVMITKMITDCVANP
               330       340       350       360

Pred: HHHHHHHHHHHHCCCHHHHHHHHHHHHHHCC
  AA: QMKQAFEQRLAKASTEDALNDIKRDIIRSAI
               370       380       390
```

Figure 5.1. PSIPRED profile for P_{FL} indicating regions with potential secondary structure. Blue bars indicate the level of prediction confidence, magenta cylinders and letter H –

predicted α-helices, yellow arrows and letter E - β-sheets, lines and letters C – random coil, AA is for the amino acid protein composition. Without taking into the account the coiled coil OD and three-helix bundle of XD, the rest of the protein chain seems to adopt random coil conformation with α-helical (residues 7-16, 18-25, 125-129, 286-292) and β-sheet (residues 396-300, 315-318) structural propensities.

With the aim of discovering the structural and functional role of the folded and unfolded parts of MuP, different phosphoprotein fragments were expressed: $P_{1\text{-}215}$ (NTD), $P_{1\text{-}272}$ (NTD+OD), $P_{215\text{-}391}$ (OD+CTD), $P_{272\text{-}391}$ (CTD), $P_{1\text{-}391}$ (full-length) and $P_{341\text{-}391}$ (XD). All the constructs with OD aggregated during purification and had two bands on SDS-PAGE. $P_{1\text{-}272}$ and P_{FL} were analysed by Western blot using anti-hisantibodies (data not shown). The presence of the same two bands for each P construct excludes the protein degradation that occurs from the C-terminal end where His-tag is attached. Further analysis by mass-spectrometry did not explain the reason of proteins aggregation and the major protein bands on the gel after purification. For this reason, only P_{NTD} and P_{CTD} were further studied in detail.

5.2. Phosphoprotein N-terminal domain (P_{NTD})

5.2.1. NMR assignments and dynamics of P_{NTD}

P_{NTD} contained 225 amino acids including 2 residues from the restriction site and the 8-histidine tag. Due to precipitation at low salt conditions (150 mM NaCl), P_{NTD} was kept and purified in high salt buffer containing 500 mM NaCl in addition to 50 mM Na-phosphates and 2 mM DTT. The ^{1}H-^{15}N HSQC spectrum (Figure 5.2) is characteristic of a disordered protein. The protein residue assignment was performed in a similar way as for Ntail.

As in the case of Ntail, structural propensities and relaxation data of P_{NTD} do not reveal regions of high helical probability. P_{NTD} shows lower SSPs (Figure 5.3A) extracted from C_α and C_β chemical shifts compared to Ntail. The SSPs revealed regions that are transiently structured in the N-terminal part (10-37) named $\alpha 1$, residues 114-137 in the middle of P_{NTD} peptide chain are defined as $\alpha 2$, and finally in the C-terminal end comprising residues 169-195 named $\alpha 3$. ASTEROIDS ensemble calculation did not provide a lot of necessary information about secondary structure description of P_{NTD}. However, relatively high positive CS of $\alpha 2$ were revealed at the C_α SCS (Figure 5.3B). Not much information was collected from αR population. However, it was enough to confirm the presence of three regions with residual helical structure.

Similar to Ntail, $R_{1\rho}$ and R_1 relaxation profile did not show evidence of the higher relaxation rates at these regions (Figure 5.4). However, from ^{1}H-^{15}N hetNOE profile, SSPs and

ASTEROIDS the potential structured regions can be identified and validated (Figure 5.3, 5.4). First of all, positive NOE values were observed for the residues from 110 to 139 which is in agreement with their slightly higher $R_{1\rho}$ that is higher than 2 s^{-1} in average.

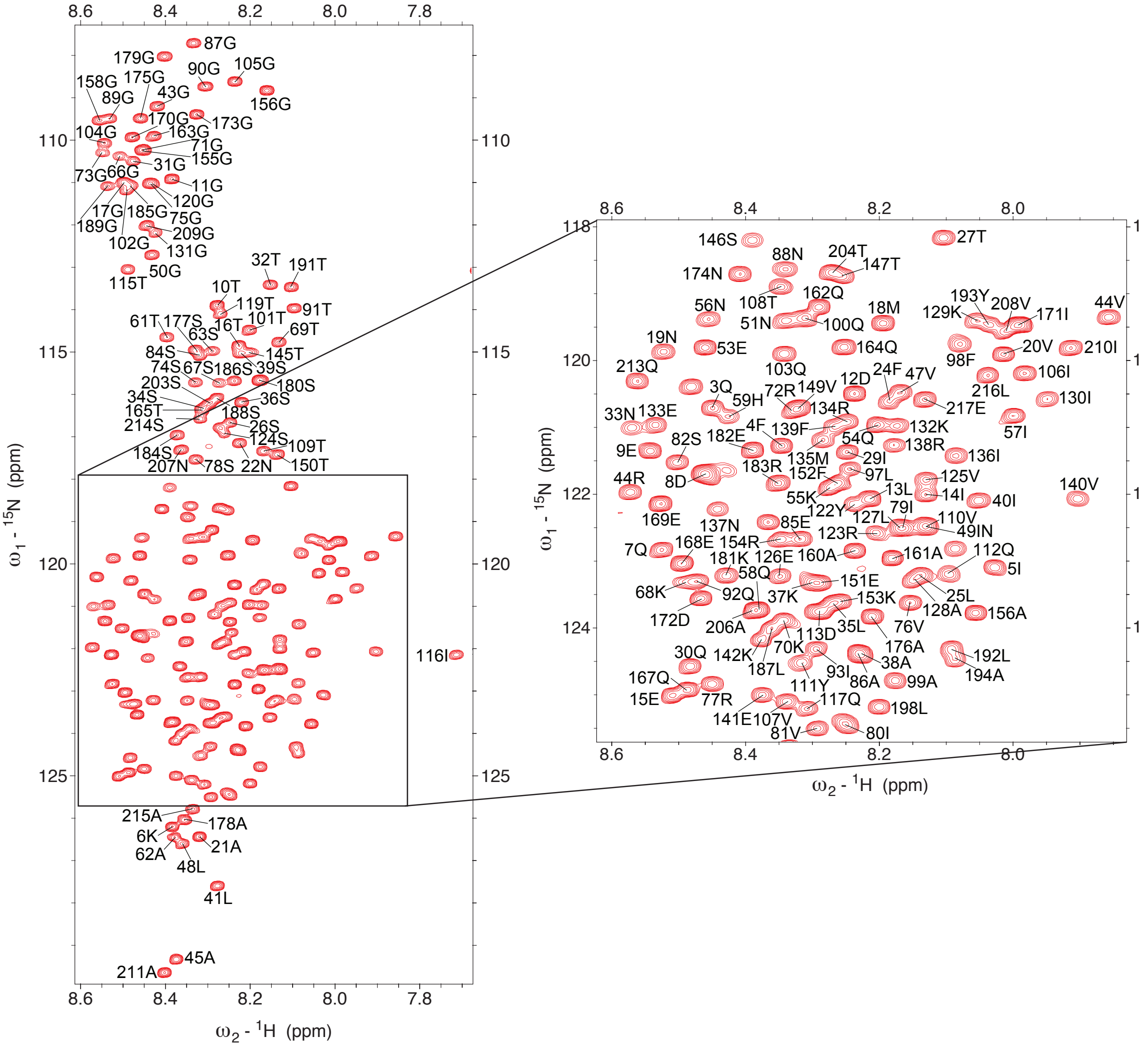


Figure 5.2. Residue specific assignment of P$_{NTD}$ displayed on ^{1}H-^{15}N HSQC spectrum at 25°C and 850 MHz in 50 mM Na-phosphates, 500 mM NaCl, 2 mM DTT pH6. Peaks from glutamines, arginines and asparagines side chains are not shown.

In order to extend the search for secondary structures, I tried to align P$_{NTD}$ in anisotropic medium (phages and liquid crystals), but it led to immediate protein precipitation. Concerning phosphorylation, P$_{NTD}$ was tested with PLK1 and CKII, but it did not lead to any effect on protein structure.

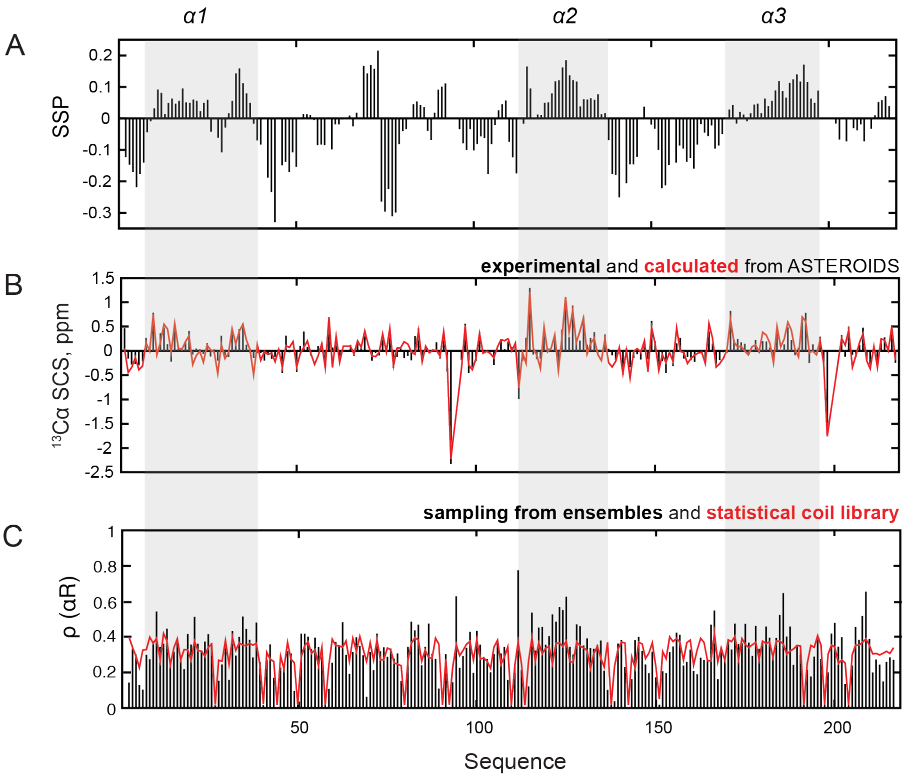


Figure 5.3. SSPs, ASTEROIDS ensemble and *αR* population data describing propensities in secondary structure of P_{NTD}, derived from chemical shifts. Grey rectangles indicate probable regions with residual structure.

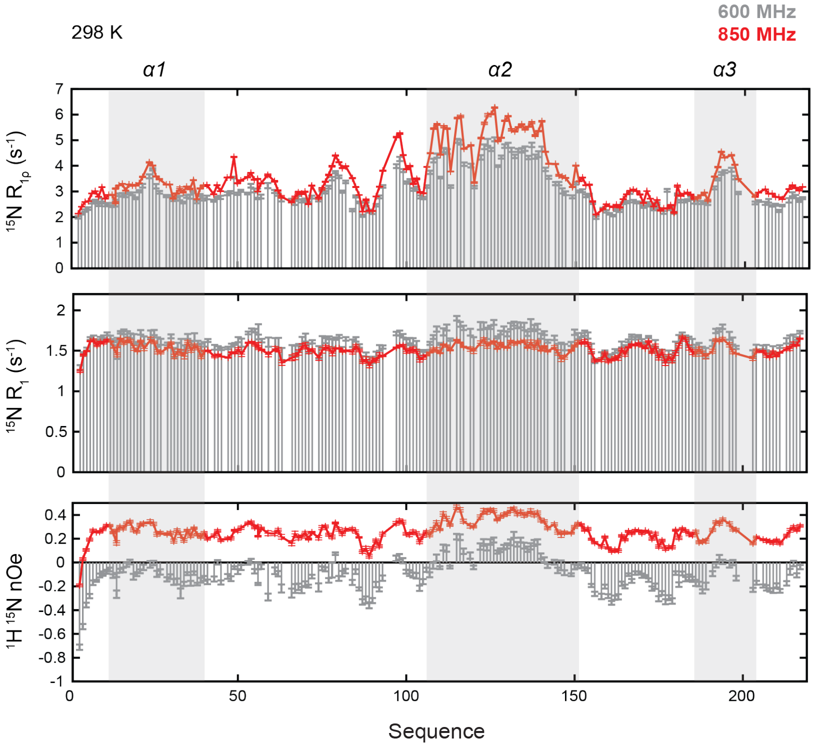


Figure 5.4. Dynamics of P_{NTD}. $R_{1\rho}$, R_1 and hetNOE profiles were measured at two different magnetic fields. Coloured rectangles indicate the regions with residual structure.

5.2.2. P_{NTD} interaction with RLPs

The interesting feature of P_{NTD} is its ability to bind RLPs and intact capsids with relatively low affinity compared to P_{CTD} [Cox et al., 2013]. In this respect, ^{15}N labelled P_{NTD} was tested for binding to unlabelled N-rings by NMR spectroscopy. A sample was prepared for a qualitative analysis of the interaction with a molar ratio of 1:0.5 (P_{NTD}:N_{FL}). Adding a higher concentration of N_{FL} led to sample precipitation. In addition, N_{FL} tends to precipitate upon the long storage at RT during NMR acquisitions. Furthermore, N_{FL} is difficult to purify in small volume analytical SEC column.

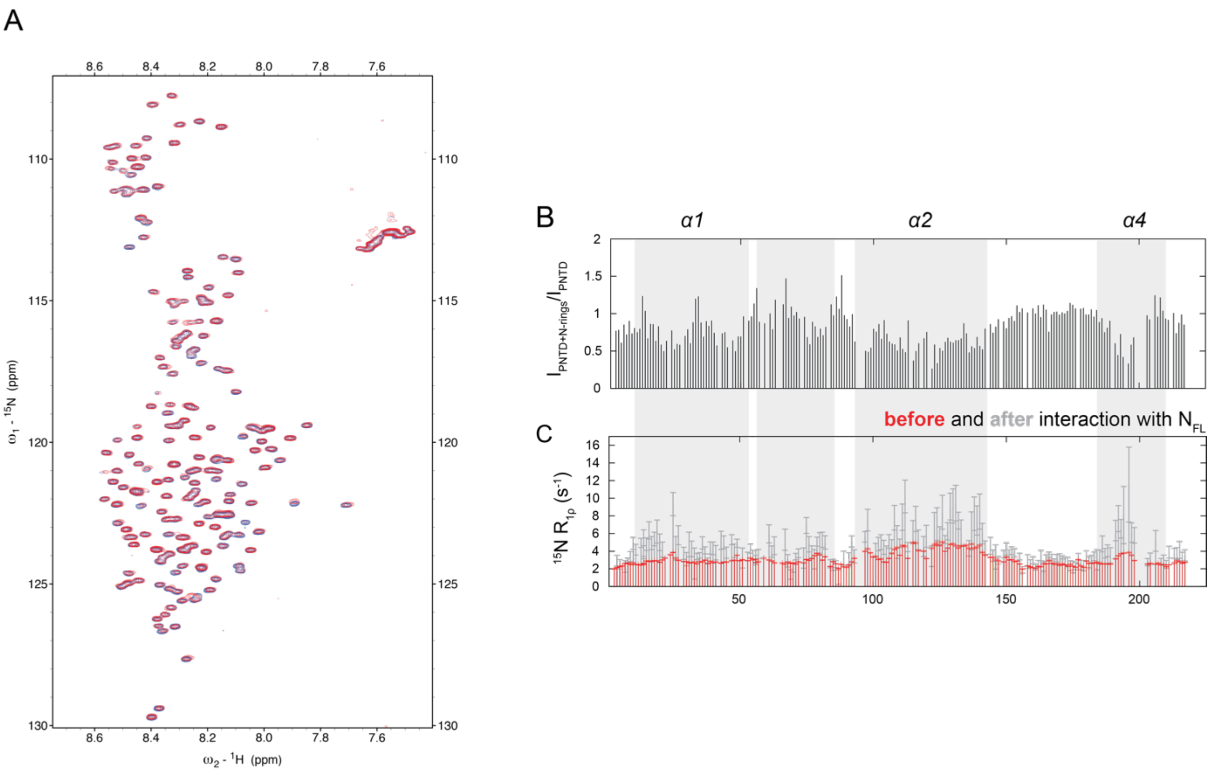


Figure 5.5. ^{1}H-^{15}N HSQC spectra superposition of P_{NTD} before (in blue) and after (in red) addition of N-rings (A). The intensity ratio between the peaks from of P_{NTD} in the presence of N_{FL} and P_{NTD} alone (B) indicates the three regions (in grey) of P_{NTD} being affected in weak interaction. The same profile was observed for increased relaxation rates for above mentioned parts of P_{NTD} protein (C). Another region between *α1* and *α2* (residues from 57 to 85) was also highlighted in grey as having less significative intensity decrease and relaxation rates increase.

Coming back to the interaction, there were no significant chemical shifts variations in the P_{NTD} spectrum (Figure 5.5A). Nevertheless, comparison of peak intensities before and after interaction with N-rings indicate regions with decreasing intensities (Figure 5.5B). They correspond to those regions identified to have transient secondary structure: helices *α1*, *α2* and *α3*. The same trend was observed in $R_{1\rho}$ relaxation measurements: the same regions show a decrease in flexibility due to contacts with N-rings (Figure 5.5C). Peak intensity decrease could be the sign of intermediate exchange. It is important to note that residues between *α1* and *α2* also had slightly decreased peak intensities and increased $R_{1\rho}$ relaxation rates.

5.3. Phosphoprotein C-terminal domain (P_{CTD})

5.3.1. XD

Before investigating P_{CTD}, the short XD was expressed and purified under high salt conditions. Its ^{1}H-^{15}N HSQC spectrum contained a lot of broad peaks that were impossible to analyse (Figure 5.6). This could be explained by the instability of the three-helix bundle. Adding TMAO with increased concentration which is supposed to stabilise the conformation according to Kingston et al., 2008, did not lead to any spectral improvement under the variation of temperature and concentration. Also XD was tested for interaction with Ntail, but no evidence of binding was observed.

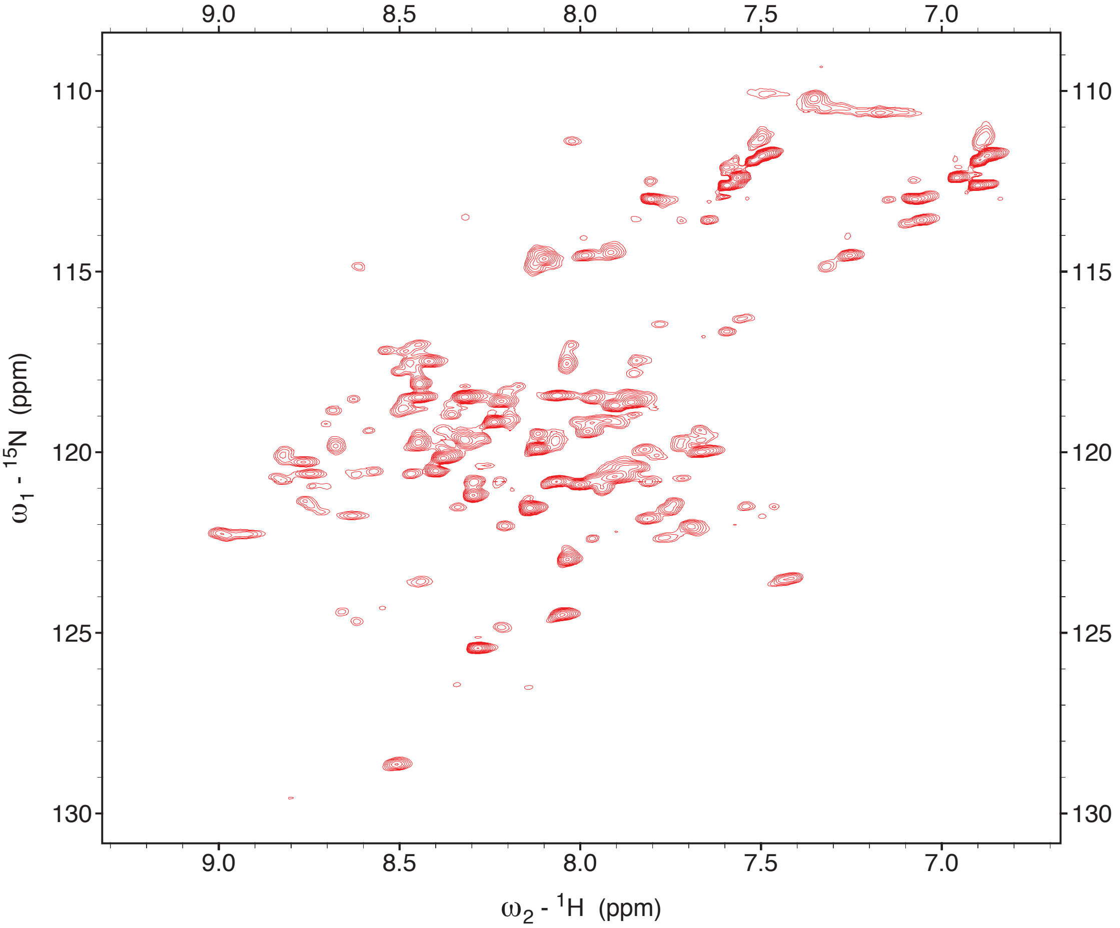


Figure 5.6. 2D ^{1}H-^{15}N HSQC spectrum of XD in 50 mM Na-phosphates, 500 mM NaCl, 2 mM DTT pH 6 at 25°C.

5.3.2. *NMR structural characterisation of* P_{CTD}

Following the hypothesis that the linker between the tetramerization domain and XD might be required for stabilizing the XD three-helix bundle, the P_{CTD} was expressed and purified under the same conditions as XD alone. However, several steps of the purification were optimised due to the formation of oligomers. In particular, upon SEC purification, one peak made of two unresolved peaks corresponded to the formation of trimeric structure which was identified by MALLS (data not shown). In order to separate the two peaks, I have injected P_{CTD} onto the SEC column at a rather low concentration after dialysing the sample against lysis buffer to remove the imidazole. I was then able to successfully collect the peak corresponding to the monomeric fraction of P_{CTD} which was further analysed by NMR.

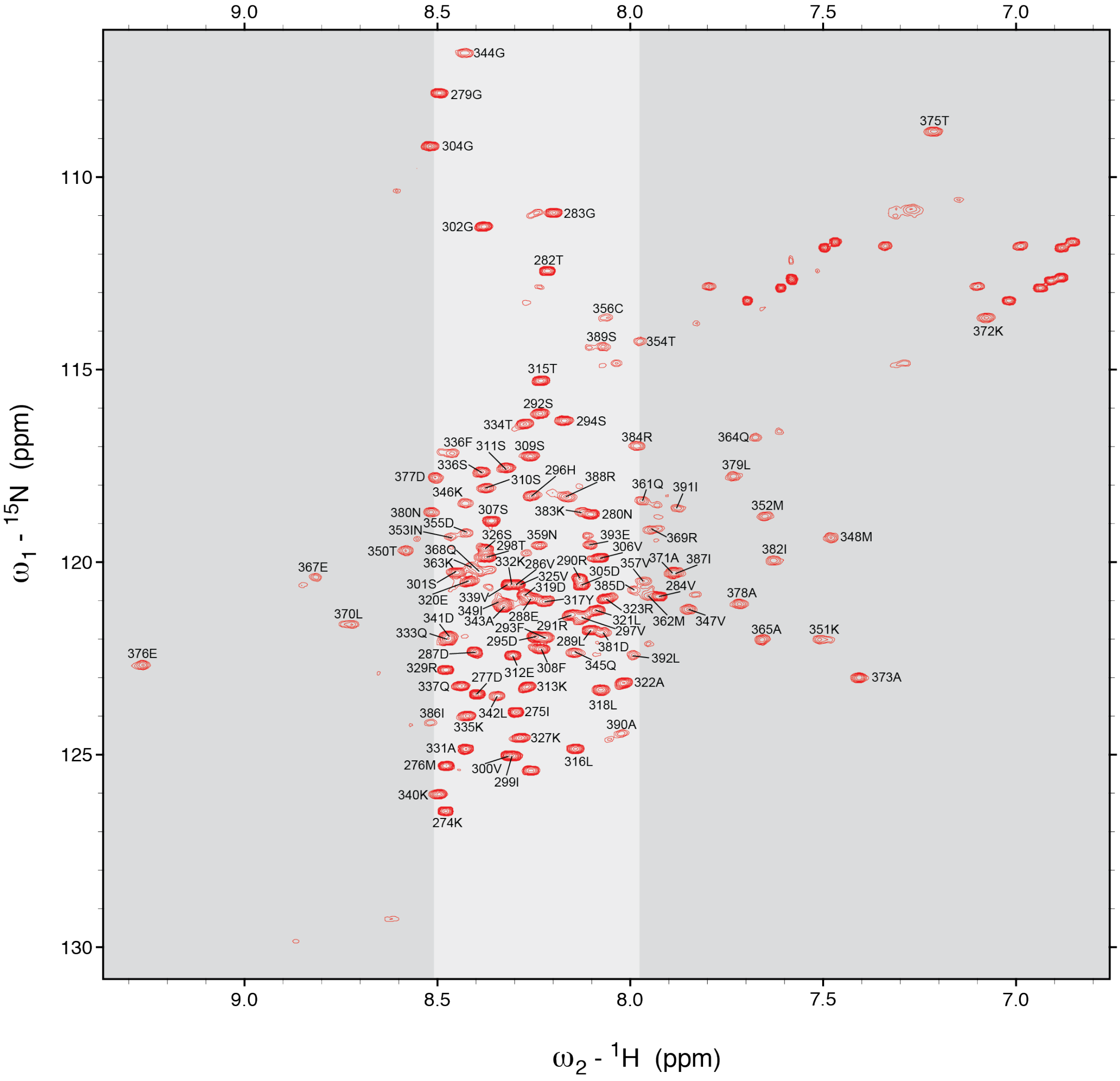


Figure 5.7. Protein residue spectral assignments of P_{CTD} recorded at 25°C and 700 MHz in 50 mM Na-phosphates, 500 mM NaCl, 2 mM DTT pH 6. The cross-peak distribution in terms of

proton CS distribution is characteristic of a protein having a mixture of folded (in grey) and unfolded (in light grey) conformation.

The ^{1}H-^{15}N HSQC spectrum of P_{CTD} showed a peak distribution in the ^{1}H dimension, which was characteristic for a protein that consists of unfolded and folded parts (Figure 5.7). A visible difference in peak intensities can be distinguished between the folded and the unfolded regions of the protein, as expected from the different correlation times of the two regions. The amino acid sequential assignment was performed. Furthermore, in the spectrum, there are several very low intensity peaks that were assigned as secondary peaks from residues close to *cis*- or *trans*-orientation of prolines.

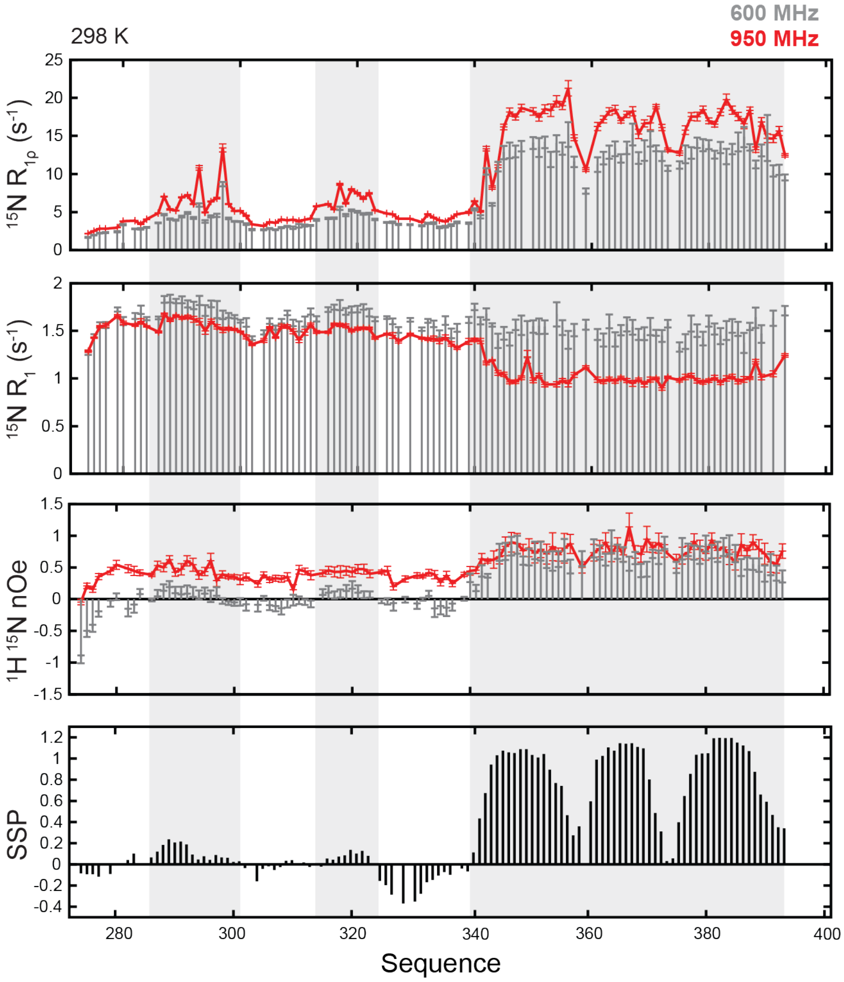


Figure 5.8. Dynamics profile of P_{CTD}. $R_{1\rho}$, R_1 and hetNOE were recorded at two different magnetic fields. Red and grey rectangles indicate two regions with transient secondary structure, green rectangle specifies the XD domain with three helix bundle.

SSPs and relaxation measurements, in particular, $R_{1\rho}$ and hetNOE were sensitive to the folded XD in its C-terminus with the three helical bundle being clearly distinguishable (Figure 5.8).

The R_1 relaxation profile was quite flat for the rates measured at lower magnetic field. However, for 950 MHz, it is possible to see the clear decrease by 0,5 s^{-1} for the folded part. As concerns the region of the unfolded linker (272-342), it has two potential structured sites with transient secondary structure (residues 286-301 and 315-324) with relatively high increase in relaxation rate $R_{1\rho}$ and in SSP values (Figure 5.8).

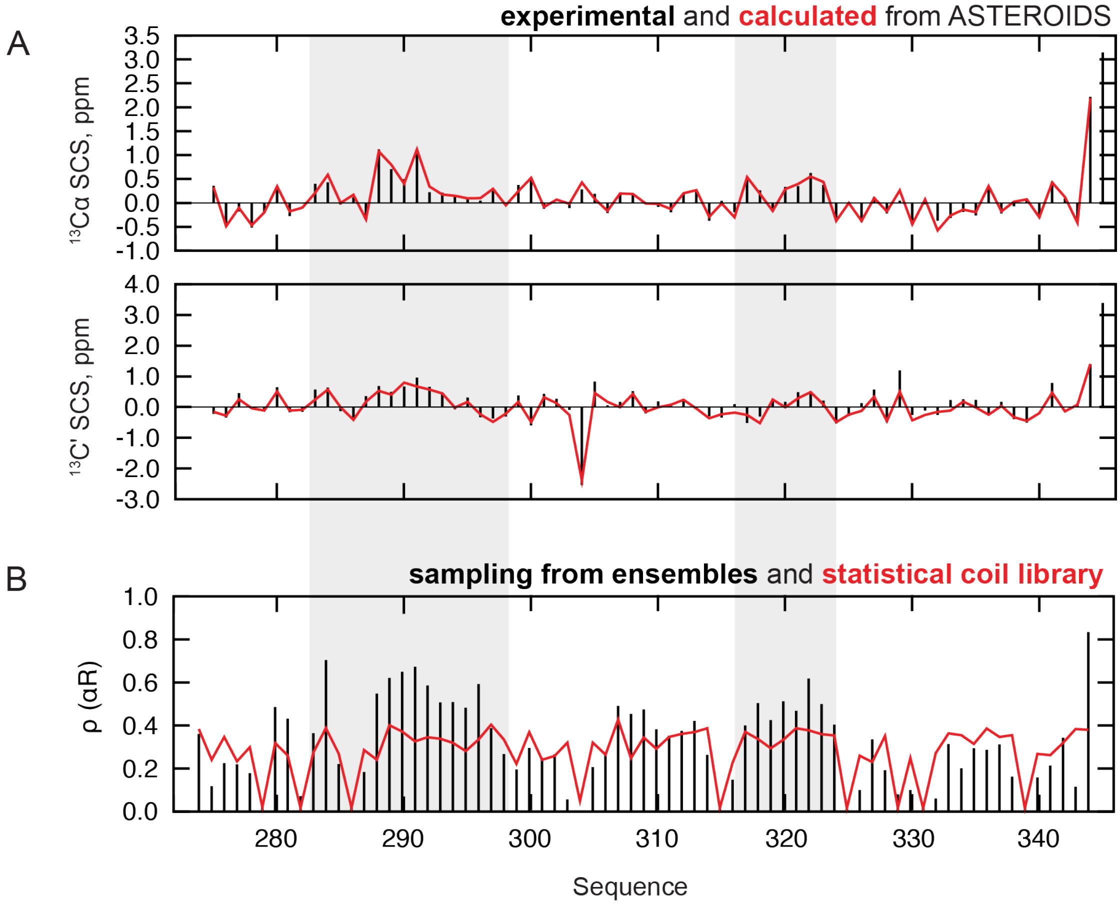


Figure 5.9. ASTEROIDS ensemble calculation (A) and *αR* population (B) for the flexible linker (272-342) confirms the existence of two secondary transient structures in the regions of 281-293 and 315-321 (in pink and green respectively).

ASTEROIDS ensemble calculation from C_α and C′ chemical shift was performed only for the flexible part of P_{CTD} (Figure 5.9A). Two regions with population of α-helix (Figure 5.9B), they contained the signature of two transient helical elements located in regions with residues 283-298 and 316-324.

To further probe P_{CTD} conformational dynamics, RDCs were measured on the protein aligned in phages. The XD domain appeared in positive $^1D_{HN}$ and dipolar waves on $^1D_{CC}$. In addition, small dipolar waves were also observed for the two mentioned before regions with transient helical propensity 286-301 and 315-324 (Figure 5.10).

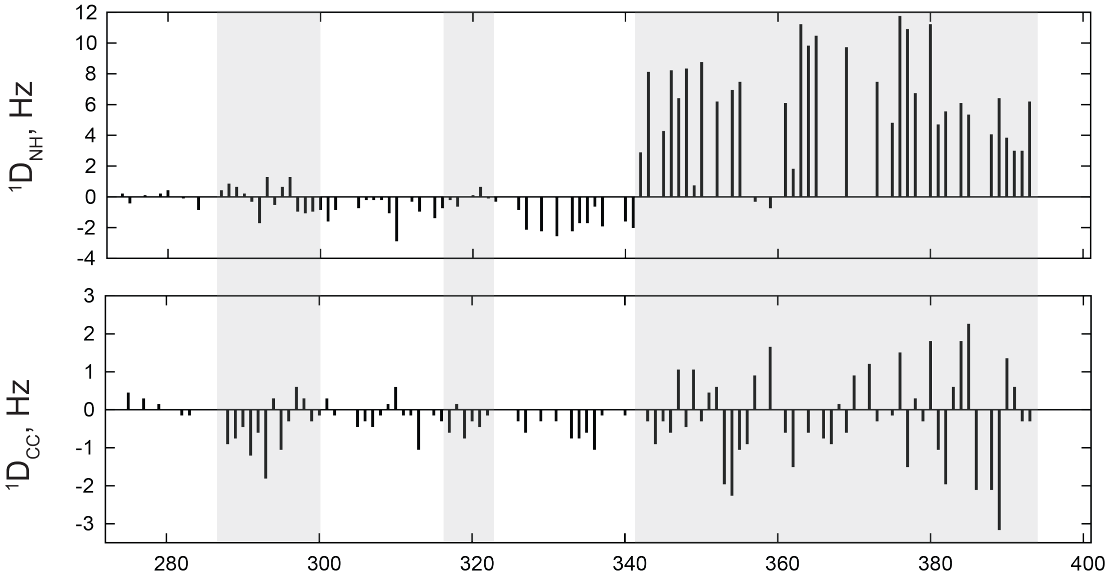


Figure 5.10. $^1D_{HN}$ and $^1D_{CC}$ RDCs extracted from P_{CTD} aligned in phages. ^2H quadrupolar splitting was 25 Hz.

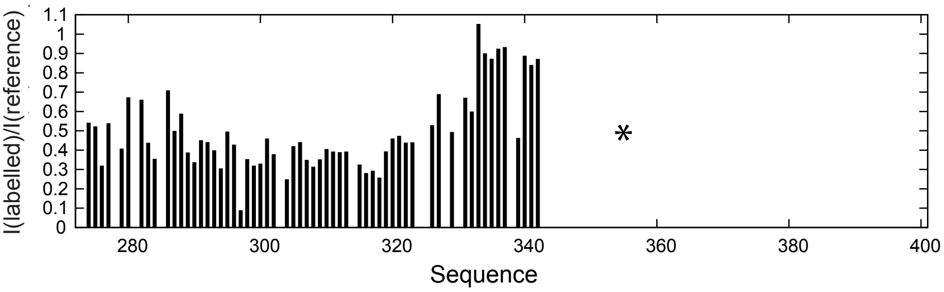


Figure 5.11. PRE from the P_{CTD} labelled with TEMPO at position C356 is expressed as the intensity ratio between the reference (reduced by Na-ascorbate) and labelled (oxidised) P_{CTD} protein.

Another question raised by this analysis was whether there were long-range transient contacts between the long linker and the folded domain, as XD alone was not observed to adopt a stable tertiary structure. The wild-type protein sequence P_{CTD} has a cysteine at position of 356, located in the first helix of XD. This allowed me to chemically modify the amino acid by adding a TEMPO-maleimide tag. Two spectra were recorded (labelled, or oxidized, and reduced protein) and the peak intensities were extracted and compared to each other. Being in close contact with the paramagnetic label, all three helices disappeared from the spectrum (Figure 5.11). In addition, peak intensities from the flexible linker decreased all-over the sequence, particularly, around residues 300 to 320. This finding does not allow to conclude the specific and direct interaction of the folded XD and transient helical domains of P_{CTD} but confirms their potential weak long-range contacts.

5.3.3. P_{CTD} interaction with RLPs

According to the literature, P_{CTD} was found to interact with RLPs with higher affinity than P_{NTD} [Cox et al., 2013], K_D of $4,26\cdot10^{-7}$ M against $4,95\cdot10^{-6}$ M. However, the exchange regime and the exact position of the interaction site had never been identified.

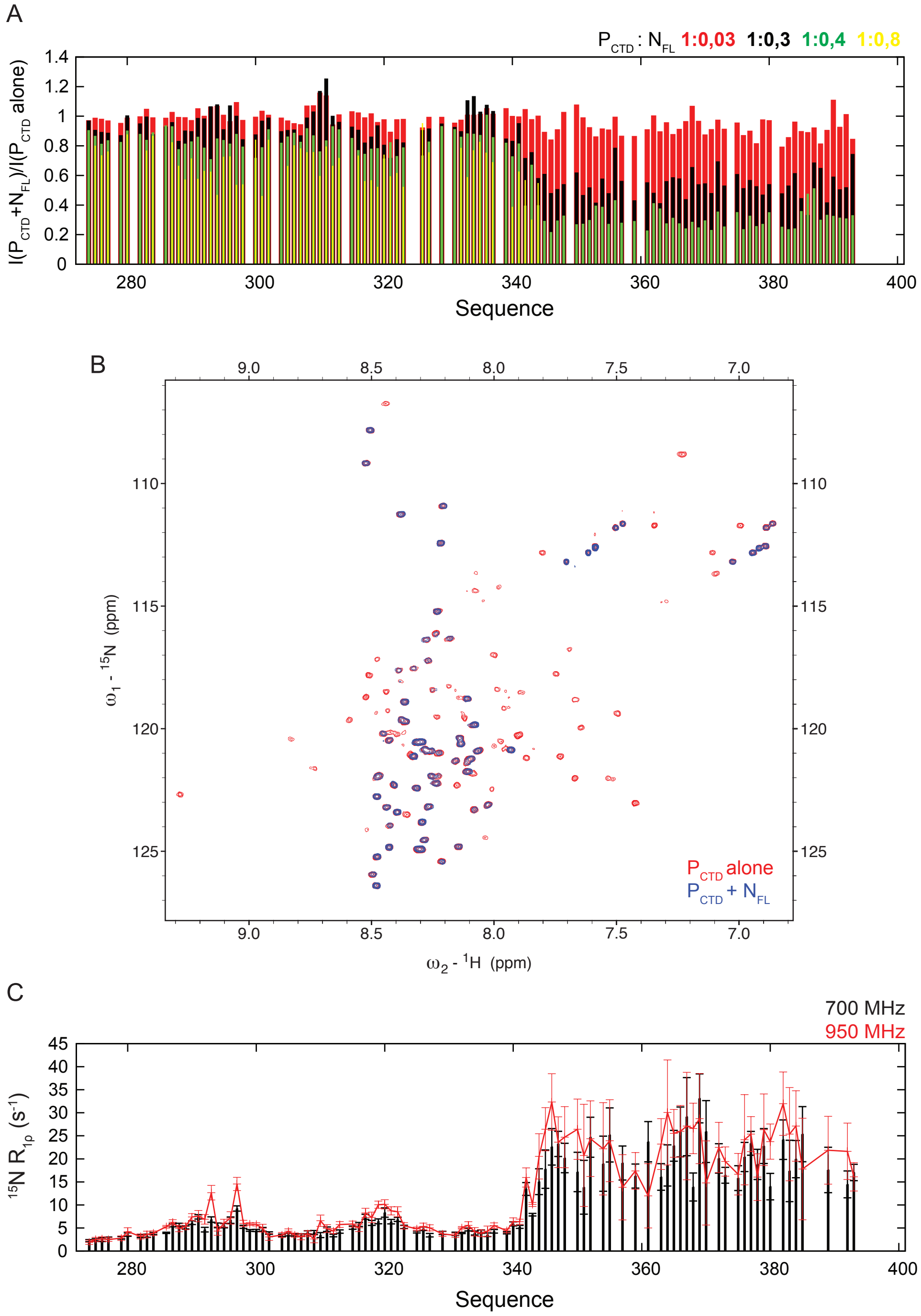


Figure 5.12. P_{CTD} interaction with N_{FL}. Intensity ratio between the mixture of P_{CTD} and N-rings and P_{CTD} alone at different molar ratios (A), each titration step was recorded as individual experiment. Overlay of 2D ^{1}H-^{15}N HSQC spectra of P_{CTD} alone (in red) and complex of P_{CTD} and RLPs (in blue) (1:0,8), peaks of XD disappeared completely (B). $R_{1\rho}$ relaxation measurements of the complex (1:0,3) at different magnetic fields (C).

In the case of this work, I have tried different molar ratios of P_{CTD} and N_{FL} (1:0,03; 1:0,3; 1:0,4; 1:0,8) in order to identify P_{CTD} interaction site with N_{FL} in case of fast exchange (Figure 5.12A). Due to the fast RLPs degradation, each titration step was prepared from the same stock and measured individually. Upon step-wise addition of N_{FL} to phosphoprotein, XD peaks disappear, indicating complex formation of P_{CTD} with big oligomeric N-rings (Figure 5.12B). Relaxation measurements were also performed (Figure 5.12C) of the complex 1:0,3 at different magnetic fields (700 and 950 MHz). Low stability of RLPs, however, was a major constraint to study in details P_{CTD} and N_{FL} interaction.

5.3.4. P_{CTD} phosphorylation

PLK1 kinase was found to modify MuP C-terminal domain by binding to P_{NTD} [Pickar et al., 2016], which perfectly agrees with the structure of OD and the anti-parallel oligomeric structure of P, when P_{NTD} and P_{CTD} can be in close proximity. The phosphorylated residues were identified to be S292/S294.

In vitro phosphorylation of P_{CTD} was performed by addition of PLK1. The residue S307 was modified first. After its full phosphorylation, S292 was phosphorylated. Probably, due to the absence of the P_{NTD} for the kinase fixation, to fully phosphorylate two residues of the protein, it took about 3 days at RT. Therefore, the kinetics was not followed by SOFAST experiments, only 2D HSQC were recorded in between assignment spectra acquisition. Both modifications led to significant changes on ^{1}H-^{15}N HSQC spectrum (Figure 5.13A). For this reason, the sequential assignment experiments were carried out. Further $R_{1\rho}$ measurements were also recorded. The relaxation experiment (Figure 5.13B) shows that phosphorylation does not change the overall chain flexibility from the linker (272-342). However, SSP analysis (Figure 5.13C) showed decreasing of the helical propensity on the region from 285 to 291 and from 301 to 311.

Phosphorylated P_{CTD} was also tested for interaction with N_{FL} in order to find another interaction site described earlier [Pickar et al., 2016], supposedly appearing after modification and binding to Ntail but only the expected changes on XD peaks were observed (data not shown).

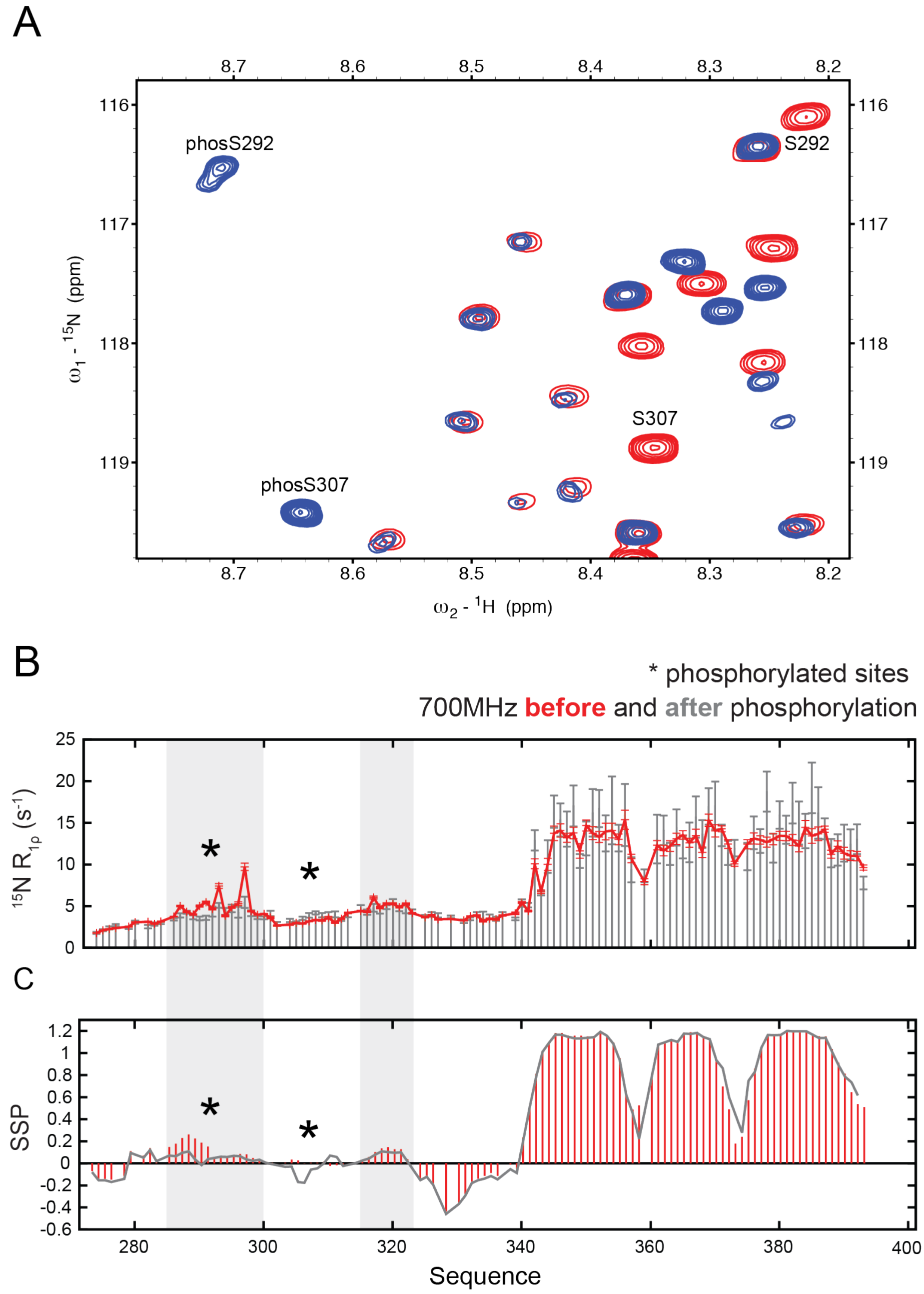


Figure 5.13. Phosphorylation profile of P$_{CTD}$ by PLK1. The zoom of 2D $_1$H-^{15}N HSQC spectrum before (in red) and after (in blue) phosphorylation (A). $T_{1\rho}$ relaxation (B) and SSP (C) of P$_{CTD}$ before and after phosphorylation.

5.4. *Full-length phosphoprotein*

Despite P$_{FL}$ being unstable and degraded protein, the ^{1}H-^{15}N HSQC spectrum of P$_{FL}$ was measured (Figure 5.14A). By overlaying it with P$_{NTD}$ and P$_{CTD}$, the disordered regions could be assigned (Figure 5.14B). The absolute intensities of the identified residues are shown on Figure 5.14C. Increased peak intensities in the position of residues from 149 to 190 could give some ideas about the localisation of the cleavage site and the reason of the presence of two bands on

SDS-PAGE gel (data not shown). Nevertheless, the sample could not be used for further analysis due to its instability and poor purification profile.

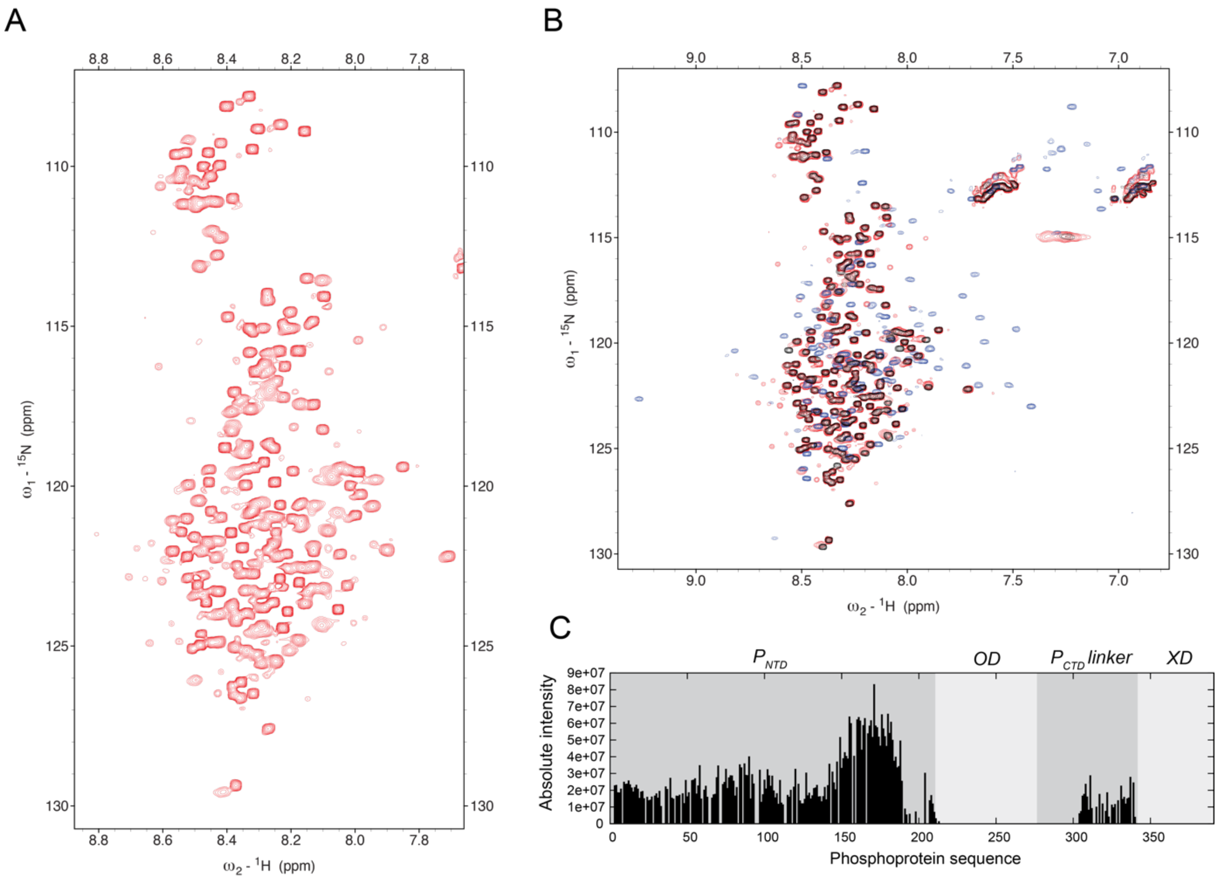


Figure 5.14. 2D ^{1}H-^{15}N HSQC spectrum of P$_{\mathrm{FL}}$ (A). Overlay of HSQCs of P$_{\mathrm{FL}}$ (in red), P$_{\mathrm{NTD}}$ (in black) and P$_{\mathrm{CTD}}$ (in blue) (B). Absolute intensities extracted from P$_{\mathrm{FL}}$ HSQC (C) with functional regions (P$_{\mathrm{NTD}}$, OD, P$_{\mathrm{CTD}}$ with XD) in grey and light grey.

5.5. V protein

For V protein, I was particularly interested in studying the structural and functional differences with respect to the phosphoprotein. In particular P$_{\mathrm{NTD}}$ is identical to V up to residue 156 (Figure 1.18), and it has the potential ability to form N^{0}V in a similar manner to N^{0}P as it was observed for PIV5 V [Randall & Bermingham, 1996]. According to the PSIPRED structure(Figure 5.15), MuV V protein has residues 170-224 on its C-terminus that contain a β-sheet conformation. Of course, there is some evidence of the presence of structured regions affecting the same residues as P$_{\mathrm{NTD}}$.

A new plasmid construct was generated for mumps virus V protein that had a TEV-cleavable His-tag at its N-terminal region in order to prevent Zn^{2+} ions to interact with histidine residues. Different expression, purification conditions and bacterial strains were used to get V in the lysis soluble fraction. As V consists of 2 zinc-fingers, made of 1 histidine and 7 cysteines, there is a high probability for V to aggregate or misfold during purification, which can be worsen by the presence of the IDR that tends to aggregate in the presence of the nucleation site as hydrophobic residues in IPDs are solvent exposed, and therefore, not protected from sticking to other hydrophobic surfaces.

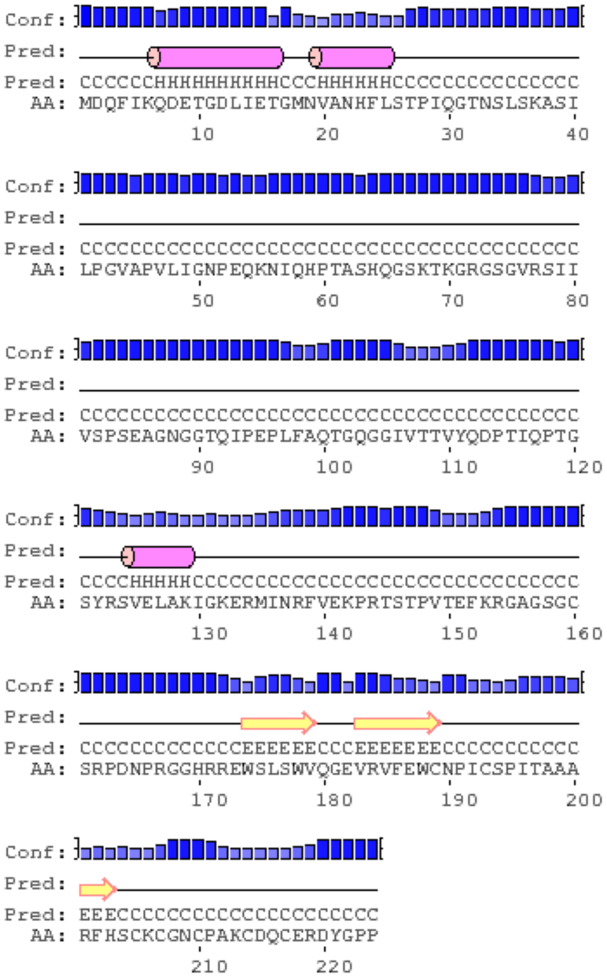


Figure 5.15. PSIPRED structure prediction for mumps virus V protein, indicates the presence of folded and unfolded regions. Blue bars indicate the level of prediction confidence, magenta cylinders and letter H – predicted α-helices, yellow arrows and letter E - β-sheets, lines and letters C – random coil, AA is for the amino acid protein composition.

Finally, I have tried denaturing conditions with mild dialysis into physiological buffer containing Zn-cations as described in the work of Saladini et al., for HeV and NiV Vs. A ^{1}H-^{15}N HSQC was measured on this sample showing the characteristic for intrinsic disorder, no sign of the zinc-finger was present (Figure 5.16A). This spectrum was compared to P_{NTD} allowing identification of the common residues. The extracted absolute intensities gradually decrease after residue 60 (Figure 5.16A) indicating them to be close to the probably misfolded and aggregated ZFD. Two out of three tryptophan residues are visible at around ^{1}H 10,1-10,2 ppm (Figure 5.16B) that come from the first loop of the zinc-finger (Figure 1.18B).

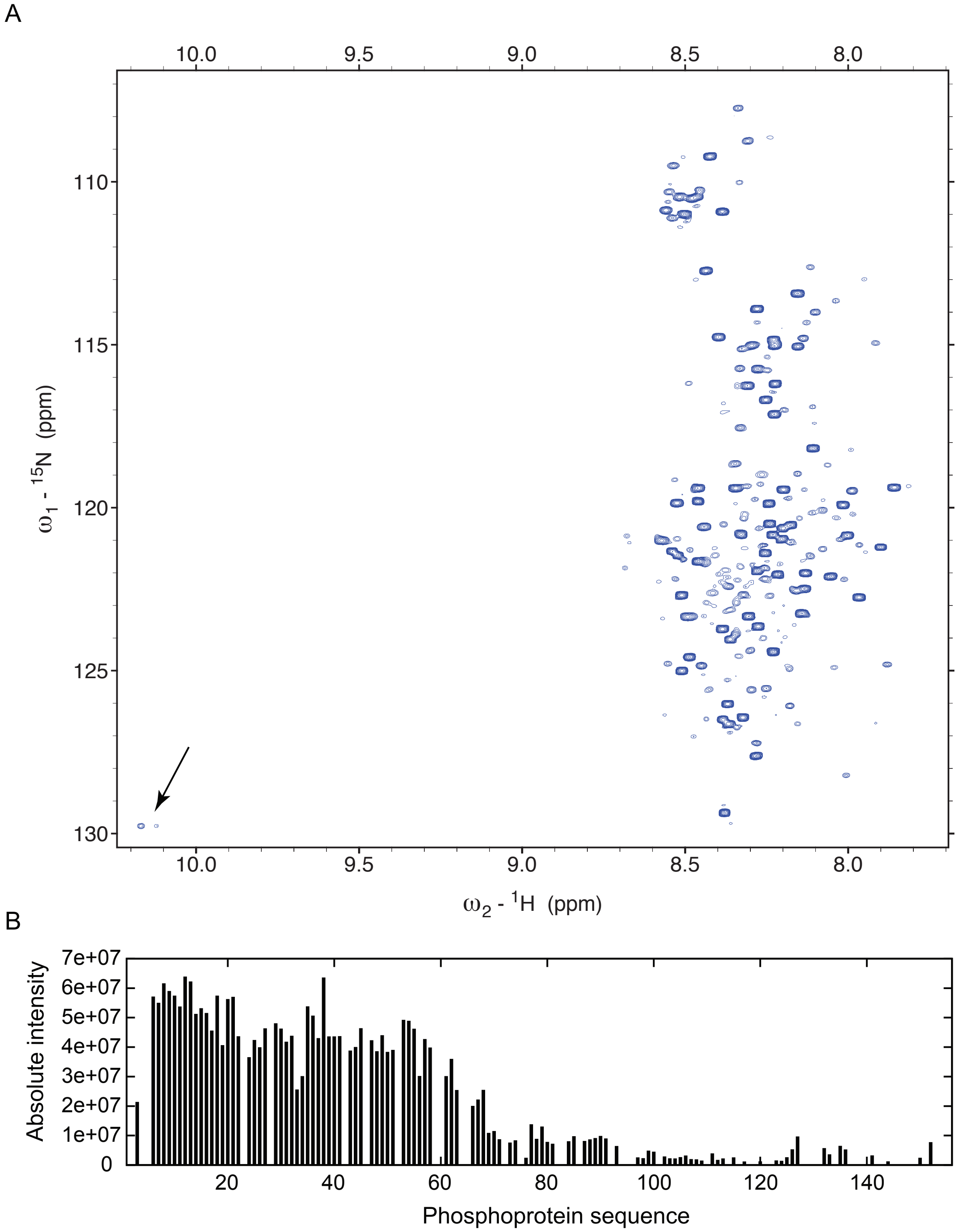


Figure 5.16. 2D ^{1}H-^{15}N HSQC spectrum of V protein (A) appears as an intrinsically disordered protein with minor and major intensity peaks, the arrow indicates two tryptophan residues located in the first loop of the Zn-finger. Comparison of V and P$_{NTD}$ spectra allowed to assign overlaid residues and extract peak intensities (B).

CHAPTER 6.
RESULTS. NºP COMPLEX

6. RESULTS. N^{o}P COMPLEX.

The N°P complex has never been studied for MuV, and this constitutes one of the aims of my work. The generation of a soluble N^{o}P complex is therefore an important part of this project, as it allows N to be studied in a monomeric form and with this, its potential interaction sites with P. In addition, the incubation of N^{o}P with short RNA has led to assembly into long NCs for the related measles virus [Milles et al., 2016]. I therefore expect that an N^{o}P construct from MuV proteins would allow me to study the assembly into oligomeric filamentous nucleocapsids upon addition of RNA also for this virus.

6.1. *Generation of different N^{o}P constructs*

As the exact region of P_{NTD} which is responsible for the interaction with N (in order to keep it monomeric) was not defined yet, two constructs differing in the length of P both fused with N_{FL} and separated by a TEV cleavage site were prepared: $P_{1\text{-}215}$ (P_{NTD}) and $P_{1\text{-}141}$ (further P_{141}). The first one did not express in the soluble fraction. However, the second one was successful, but a lot of the protein formed aggregates manifested by a large peak corresponding to the void volume of SEC column (data not shown), which decreased the overall purification yield. $P_{141}N_{549}$ was further analysed by MALLS, EM and NMR. For analysis, the void volume peak was excluded and I continued working with the heterodimeric fraction.

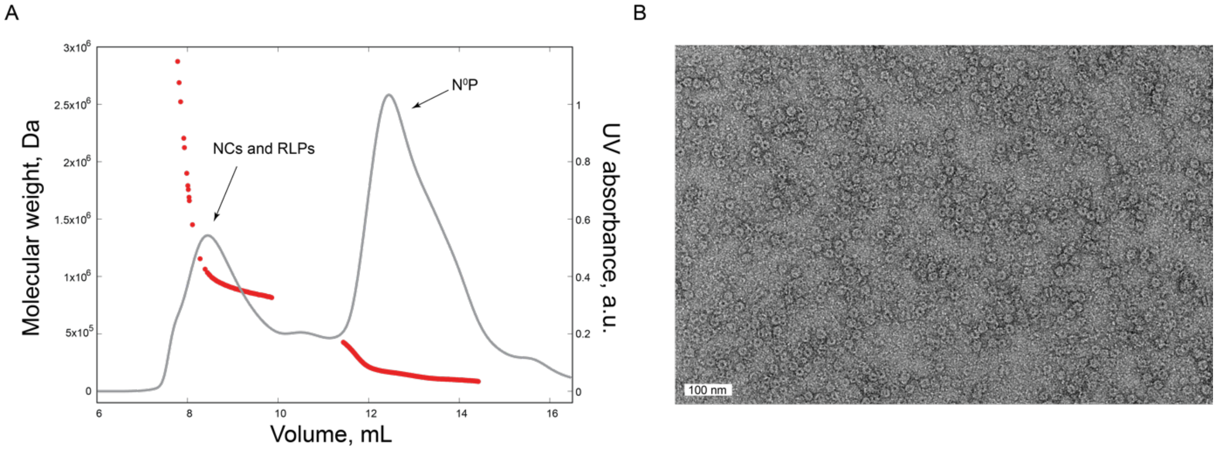


Figure 6.1. MALLS profile (in red) from $P_{141}N_{549}$ showing the presence of RLPs and N^{o}P under SEC purification (in grey) (A). Electron micrographs of the same sample after SEC purification confirmed the formation of N-rings with attached P peptide (B).

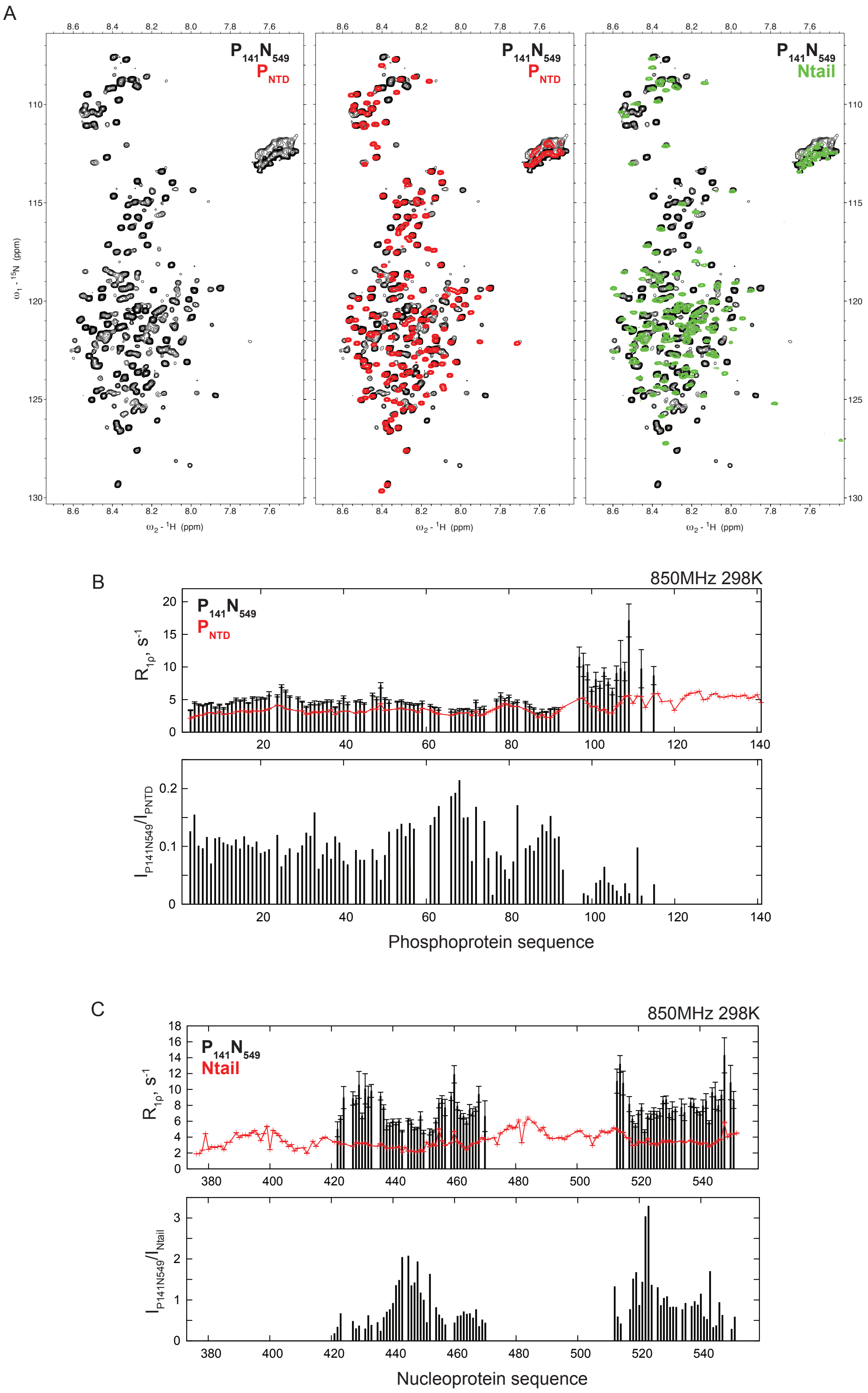


Figure 6.2. Spectrum of P$_{141}$N$_{549}$ alone (in black) and overlaid with P$_{NTD}$ (in red) and Ntail (in green) (A). Intensity ratio calculated from P$_{NTD}$ or Ntail over P$_{141}$N$_{549}$ and $R_{1\rho}$ profile of N^{0}P compared to Ntail and P$_{NTD}$ measured at the same temperature and magnetic field (B, C).

This purified, concentrated heterodimeric N^0P complex was then analysed by MALLS. Surprisingly, the MALLS profile showed again a peak corresponding to the void volume (fraction of about 26% of total injected volume) with a molecular weight corresponding to about 13 N^0P units (Figure 6.1A). The second peak, corresponding to the same elution volume as the one previously collected in the purification, corresponded to the N^0P dimer. This suggests that, over time, soluble heterodimeric N^0P complexes made of N_{FL} and $P_{1\text{-}141}$ associated into large molecular weight assemblies.

$P_{141}N_{549}$ was also visualised by electron microscopy showing the presence of N-rings in the same sample (Figure 6.1B). This is in agreement with the conclusion drawn from the MALLS experiment. After purification, N^0P seems to slowly transform into RLPs by keeping P bound to N. This hypothesis was expected to be confirmed by NMR measurements.

A 2D ^1H-^{15}N HSQC spectrum of $P_{141}N_{549}$ that appeared with the signature of disordered protein, was overlaid with spectra of Ntail and P_{NTD} to assign the common residues (Figure 6.2A). Peak intensity ratio indicated a decreased peak intensity and disappearance of P_{141} C-terminal peaks, however, residues of *NcoI* restriction site and the left-over from TEV cleavage site were not assigned. The relaxation measurements also showed the increase in $R_{1\rho}$ of the residues from 92 to 141. Unsuccessful TEV cleavage was excluded by appearance of lower N bands on SDS-PAGE gel (data not shown). Residues 1 to 92 of P have slightly higher relaxation rates compared to free P_{NTD} which can be explained by the attachment of P to the surface of N-rings (Figure 6.2B). Ntail within $P_{141}N_{549}$ has almost the same intensity and relaxation profile as RLPs described in Chapter 4.4.2: residues close to the Ncore and MoRE were not present in the spectrum (Figure 6.2C). In addition, the first residues of following the MoRE were also absent.

In conclusion, while searching for the interaction site responsible for the stable N^0P complex (about first 70 residues of P_{NTD}), another interaction site in P_{NTD} was identified. This site is located between the residues 92 and 141 and seems to be identical to the transient binding site of P_{NTD} with N-rings (Figure 6.2B). In addition, other constructs were cloned with full-length N and different P truncations: 1-90, 1-150, 150-215, 90-215. The first construct purified from small volume test-expression showed that the P peptide remained bound to N after successful TEV-cleavage (this was verified with SDS-PAGE). Therefore, in addition to the interaction site around residues 91-141, P must have another binding site further at its N-terminus that interacts with N. This is in agreement with N^0P binding modes of other *Paramyxoviridae* [Milles et al., 2018].

6.2. *Characterisation of $P_{90}N_{549}$*

Compared to $P_{141}N_{549}$, $P_{90}N_{549}$ showed higher protein yield and less aggregates. At the same time, EM revealed similar instability of the complex and gradual formation of N-rings over time (Figure 6.3A). The comparison of ^{1}H-^{15}N HSQC spectra of $P_{90}N_{549}$ and P_{NTD} showed that all 90 peaks of P peptide were present in the $N^{0}P$ spectrum. For Ntail, the same residues as for N-rings were absent (Figure 6.4A). An interesting feature was observed in the $R_{1\rho}$ relaxation profile of $P_{90}N_{549}$: the rates for the regions 13-30 and 39-51 within P were increased as compared to P_{NTD} alone (Figure 6.4B,C). I have also performed coupled SAXS-SEC analysis (in analogy to N-rings) with $P_{90}N_{549}$. Unfortunately, in the same manner as N_{FL}, the collected single peak was relatively heterogeneous that did not allow further data treatment.

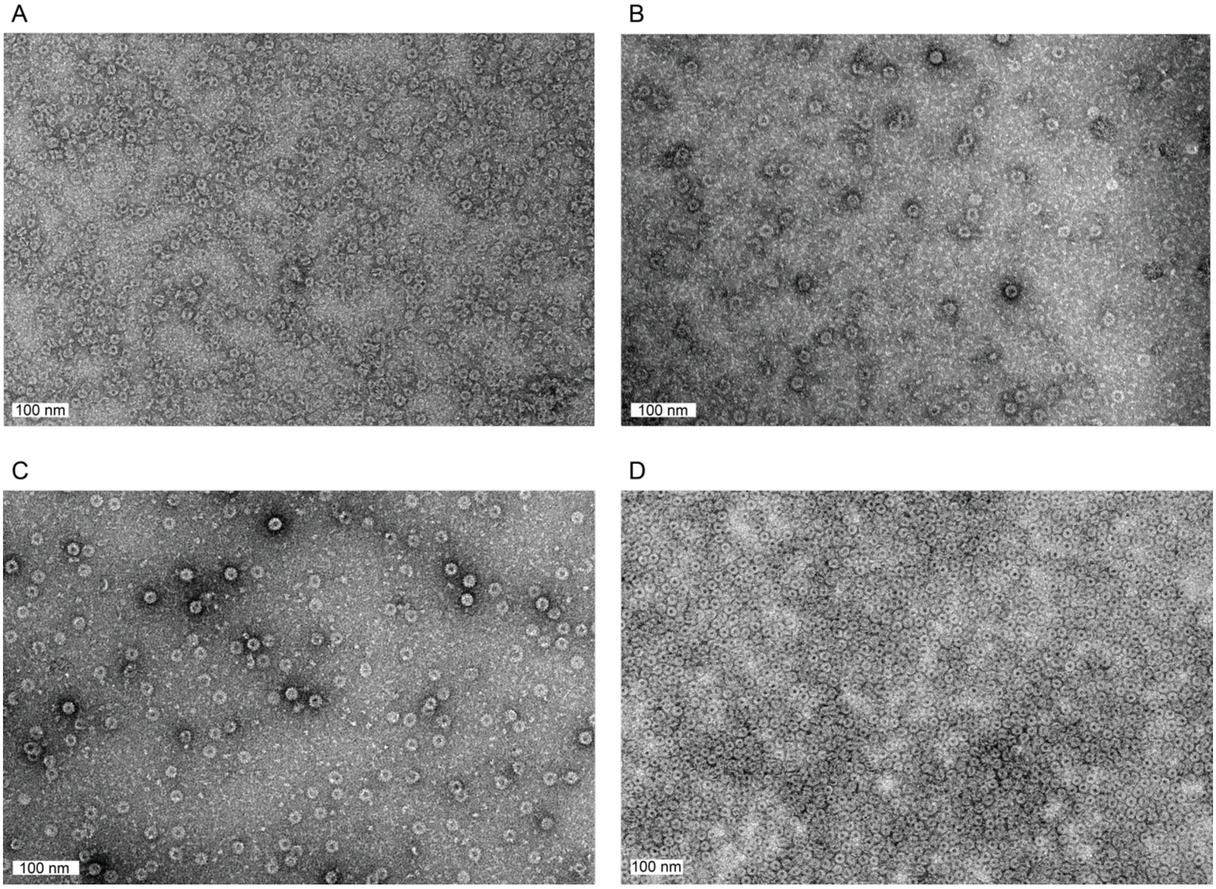


Figure 6.3. Electron micrographs of $P_{90}N_{549}$ alone (A), $P_{90}N_{549}$ without TEV-cleavage (B) and incubated O/N with TEV (C), $P_{90}N_{398}$ alone (D).

In order to obtain a monomeric form of $N^{0}P$ that does not assemble into rings over time, the TEV-cleavage step was omitted to force P into proximity with N and therefore stabilize the complex. The peak corresponding to the complex was observed in the SEC profile and was collected. A ^{1}H-^{15}N HSQC was recorded without further concentrating the sample. This spectrum appeared to have less peaks corresponding to the P peptide (Figure 6.5A), however,

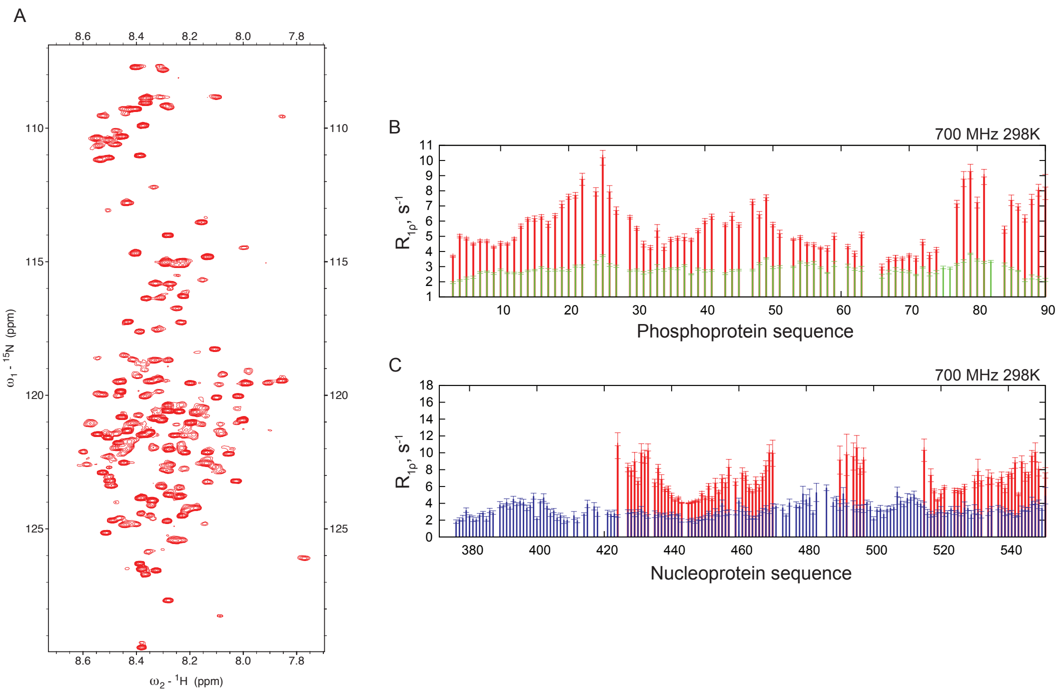


Figure 6.4. 2D ^{1}H-^{15}N HSQC spectrum of $P_{90}N_{549}$ (A), $R_{1\rho}$ profile of $N^{o}P$ (in red) compared to P_{NTD} (in green) and Ntail (in blue) measured at the same temperature and magnetic field (B and C respectively).

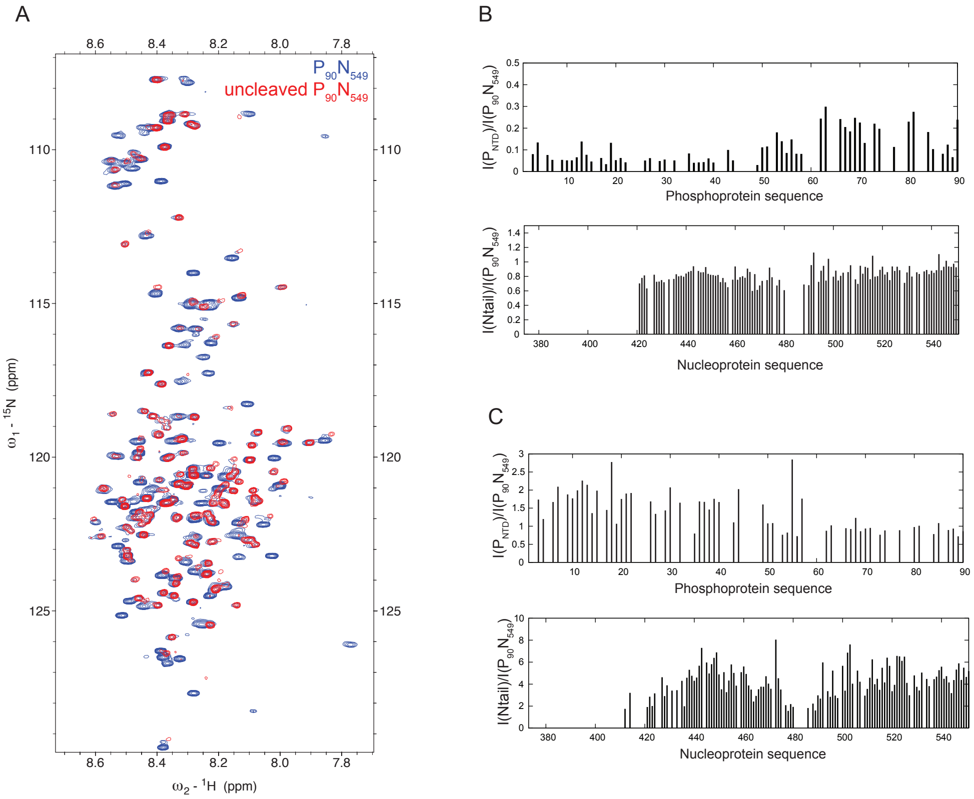


Figure 6.5. 2D ^{1}H-^{15}N HSQC spectrum of uncleaved $P_{90}N_{549}$ overlaid with cleaved one (A). Peak intensity ratio from P and Ntail extracted after purification (B) and after over-night acquisition at RT (C).

for Ntail the C-arm and a few peaks of the MoRE were not present. Peak intensity ratios illustrate this very clearly (Figure 6.5B). The sample was left for over-night NMR acquisition, but the next morning all P peptide peaks appeared again and Ntail peaks from structured regions decreased in intensity (Figure 6.5C). Electron micrographs from this sample showed the presence of N-rings (Figure 6.3B). The over-night cleavage was tested after, it enhanced the formation of RLPs (Figure 6.3C). Therefore, I needed to find another way to stabilise N^oP to investigate NC assembly and further structural studies.

6.3. Further assays on N^oP stabilisation

Several strategies were pursued in order to obtain a stable N^oP complex. My first strategy was to test whether the presence of Ntail can also influence the stability of N^oP. For this reason, I generated two constructs $P_{90}N_{374}$ and $P_{90}N_{398}$. Upon purification, TEV cleavage did not work. Probably, it was complicated by the limited access of protease to the TEV recognition site which could be blocked by the formation of RLPs. Further SEC purification did not lead to separation of the monomeric complex as a huge fraction was transformed into N-rings that were eluted at the column void volume. EM confirmed that RLPs were present in the sample (Figure 6.3C).

In addition, as observed for $P_{141}N_{549}$, the longer P peptide was found to decrease the protein yield and stability. Therefore, increased relaxation rates in P region from residues 75 to 90 could be linked to the exchange regime of the complex between P and N. I generated a shorter P peptide from 1 to 65 that was believed to be sufficient for the interaction. The resulting NMR measurements did not show any evidence of better complex stabilisation (Figure 6.6A). Relaxation rates were also very similar to those measured for $P_{90}N_{549}$ (Figure 6.6B, C).

My next strategy was to try to improve complex stability by changing the pH to 7, as the theoretically calculated pI of the fusion protein was very close to 6 (http://protcalc.sourceforge.net). However, P peaks responsible of the potential interaction with N on HSQC spectrum of $P_{65}N_{549}$ appeared with the same relative intensity profile as it was observed before for $P_{65}N_{549}$ at pH 6 (data not shown). In addition, due to the exchange processes, several residues disappeared from the spectrum (data not shown).

The final strategy was to generate an N^oP complex without N-arm to prevent assembly. This method has been used for several *Paramyxoviridae* N^oPs to facilitate crystallisation [Yabukarski et al., 2014; Aggarwal et al., 2017], but this complex theoretically could not be used for assembly assays for which the N-arm is essential. The construct $P_{1\text{-}65}N_{32\text{-}549}$ was cloned, expressed, purified and measured by NMR. According to the 2D ^{1}H-^{15}N HSQC spectrum and the corresponding plots of absolute peak intensities (Figure 6.7A, B), the very N-terminal region of P (residues from 1 to 31) disappeared completely. Residues from 32 to 65, which are also implicated in the interaction as concluded from increased $R_{1\rho}$ rates in $P_{90}N_{549}$, however, also seemed broadened, in particular in the region that showed the highest $R_{1\rho}$ in $P_{90}N_{549}$ (about residue 45). EM of this NMR sample did not show formation of N-rings, rather smaller oligomers that could be artefacts of the highly concentrated protein and its interaction with the stain (Figure 6.8A).

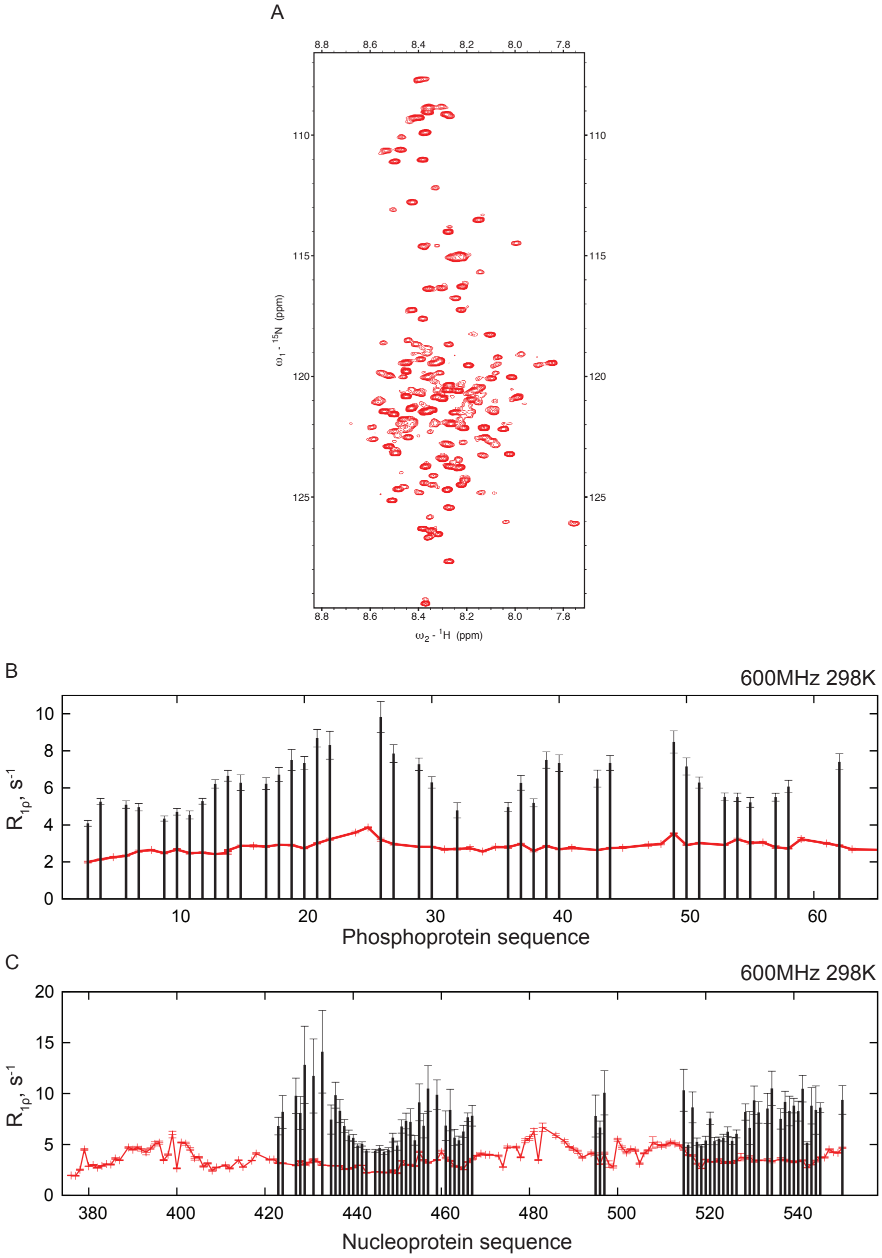


Figure 6.6. NMR measurements on $P_{65}N_{549}$. 2D ^{1}H-^{15}N HSQC spectrum of $N^{o}P$ (A). $R_{1\rho}$ profile of $N^{o}P$ (in black) compared to P_{NTD} (B) and Ntail alone both in red (C).

During analysis of $P_{90}N_{549}$ and $P_{65}N_{549}$ complexes, I noticed that, after initial polymerization into rings, a fraction of monomeric $N^{o}P$ remained stable during a long period of time (sample was kept several days at RT for NMR measurements, was frozen and unfrozen several times), as relaxation profile and peak intensities remained intact. This fact made me think about the possible stabilisation of the rest of N°P by free P that was liberated after the transformation of N°P to N-rings.

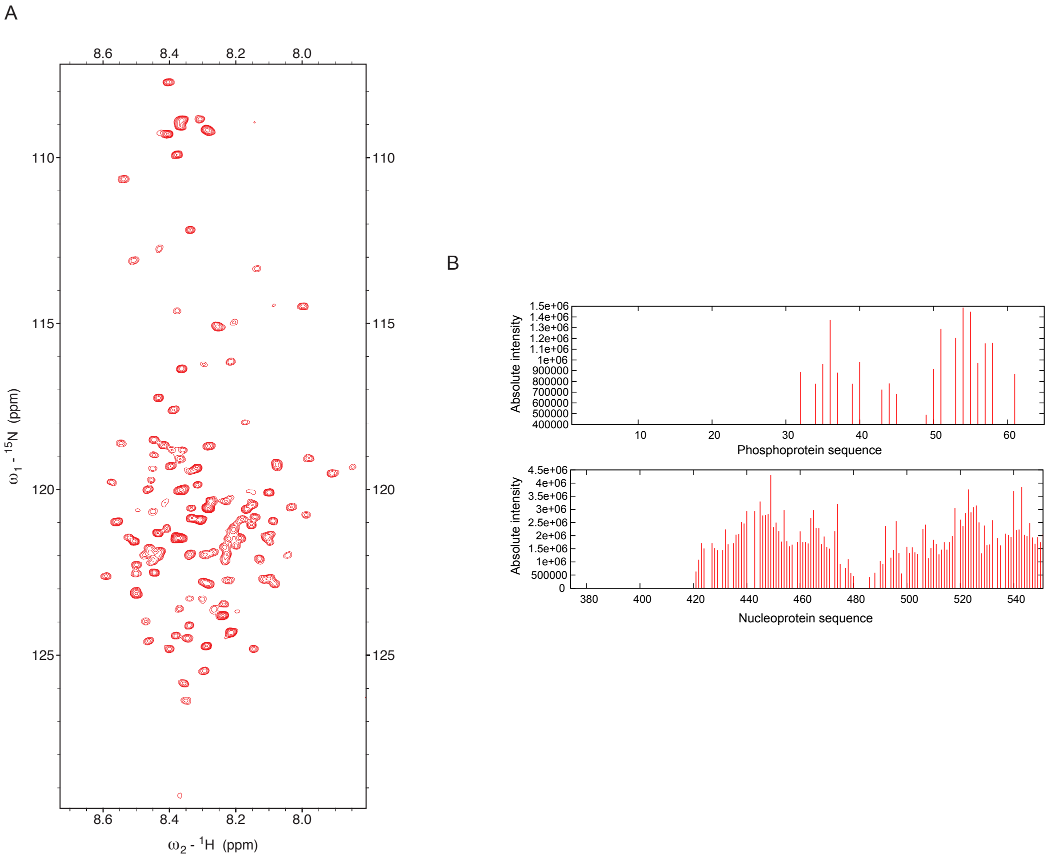


Figure 6.7. 2D ^{1}H-^{15}N HSQC spectrum of $P_{65}N_{32\text{-}549}$ (A). Extracted absolute intensities of P and N peaks from N°P (B).

To test the hypothesis that an excess of P could stabilize N°P, a fresh ^{15}N-labelled $P_{65}N_{549}$ freshly eluted from the SEC column purification in the single peak was immediately mixed with P_{NTD} in a molar ratio of about 1:3, further concentrated and visualised by EM. In this case, more P_{NTD} surrounds N°P, more N°P is present and more stable the latter is. Due to the complex stability dependence on the sample concentration the electron micrographs have a lot of protein background as the sample dilution can lead to the formation of RLPs (Figure 6.8B). As expected, P_{NTD} prevented N°P from oligomerisation and no N-rings were observed. 2D ^{1}H-^{15}N HSQC spectra of this sample did not show any spectral changes except few peaks that appeared or increased their intensity and remain unidentified (Figure 6.9A). However, relaxation profile showed the complete liberation of $P_{1\text{-}65}$ peptide which was in competition with P_{NTD} to bind N (Figure 6.9B).

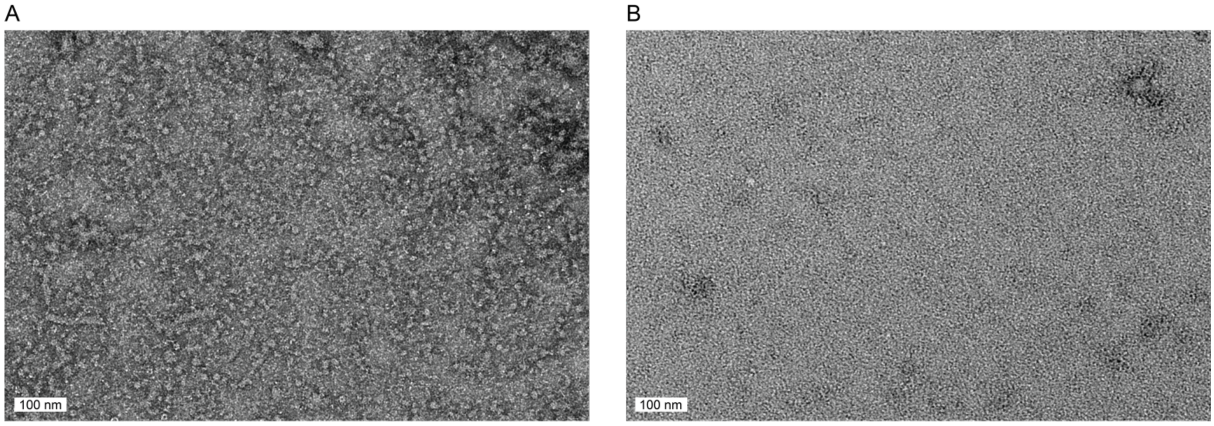


Figure 6.8. Electron micrograms of $P_{1--65}N_{32-549}$ from NMR sample (A) and $P_{65}N_{549}$ stabilised with P_{NTD} at high concentration and issued from NMR measurements (B).

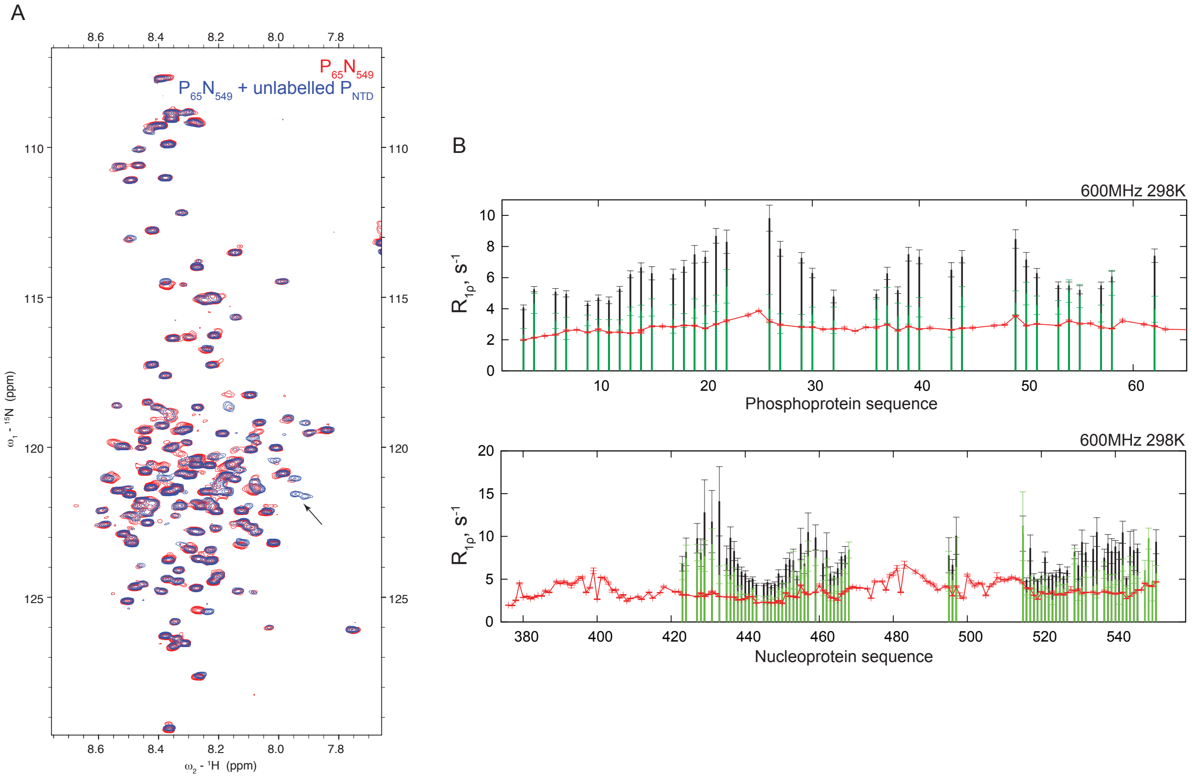


Figure. 6.9. Overlay of 2D ^{1}H-^{15}N HSQC spectra of $P_{65}N_{549}$ alone (in red) and stabilised by P_{NTD} (in blue) (A). Arrows indicate the same peaks as those that appeared after the addition of P_{NTD} to N^{o}P. The $R_{1\rho}$ profile of N^{o}P P and N before (in black) and after (in green) addition of P_{NTD} compared to P and N alone (in red) measured at the same magnetic field (B).

6.4. *Assembly of NC from N^{o}P*

All along the process of N^{o}P stabilisation, NC assembly was tested for almost every construct. Adding PolyA6 RNA (OH-AAAAAA-OH) to $P_{141}N_{549}$ did not show any spectral changes (including peak intensities) after over-night incubation at RT. This result was expected as the sample contained a very low concentration of monomeric N^0P, as it was transformed to RLPs after recording of relaxation experiment. This means that RLPs are pre-formed and do not incorporate RNA.

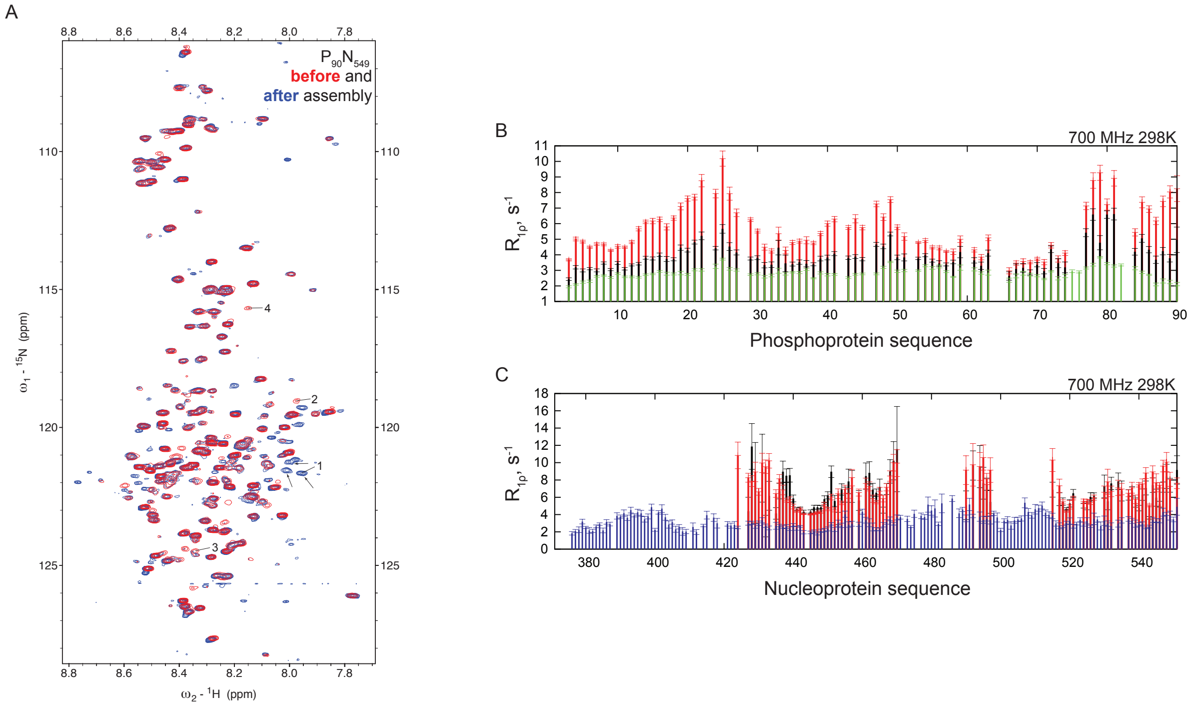


Figure 6.10. 2D ^{1}H-^{15}N HSQC spectrum of $P_{90}N_{549}$ before (in red) and after (in blue) assembly with PolyA6 RNA, arrows indicate peaks that also appeared during the addition of P_{NTD} to $P_{65}N_{549}$ (A). Numbers indicate the disappeared (2, 3, 4) and appeared (1) residues. The $R_{1\rho}$ profile of N^0P P (B) and N (C) before (in red) and after (in black) addition of RNA compared to P_{NTD} (in green) and N alone (in blue) measured at the same magnetic field.

However, interesting results were obtained from assembly of $P_{90}N_{549}$ with the same RNA sequence. 2D HSQC spectra before and after had several differences in the intensities: these are increasing for $P_{1\text{-}90}$ and decreasing for Ntail (Figure 6.10A). The assembly rates (*k*) were extracted from fitting. For P and N they were almost the same (1,37 and 1,03 s^{-1}), indicating a simultaneous P liberation and N assembly (Figure 6.11A, B). In the spectrum after assembly there were some peaks that appeared and disappeared (Figure 6.10A), three of them were the same found upon the addition of P_{NTD} to $P_{65}N_{549}$ (Figure 6.9A). They were not assigned and due to their relatively low intensity and resolution available from the 2D SOFAST experiment, the peak intensity variation was difficult to extract accurately and to fit to a kinetic profile. However, from several of them, it was possible to extract intensities and fit to the single exponential curve (Figure 6.11 C). The rate for unassigned residues was at least 2 times higher and equal to 2,62 for peaks with increasing (peak 1) and 2,04 s^{-1} for decreasing (peak 2, 3, 4) in intensity (Figure 6.11C).

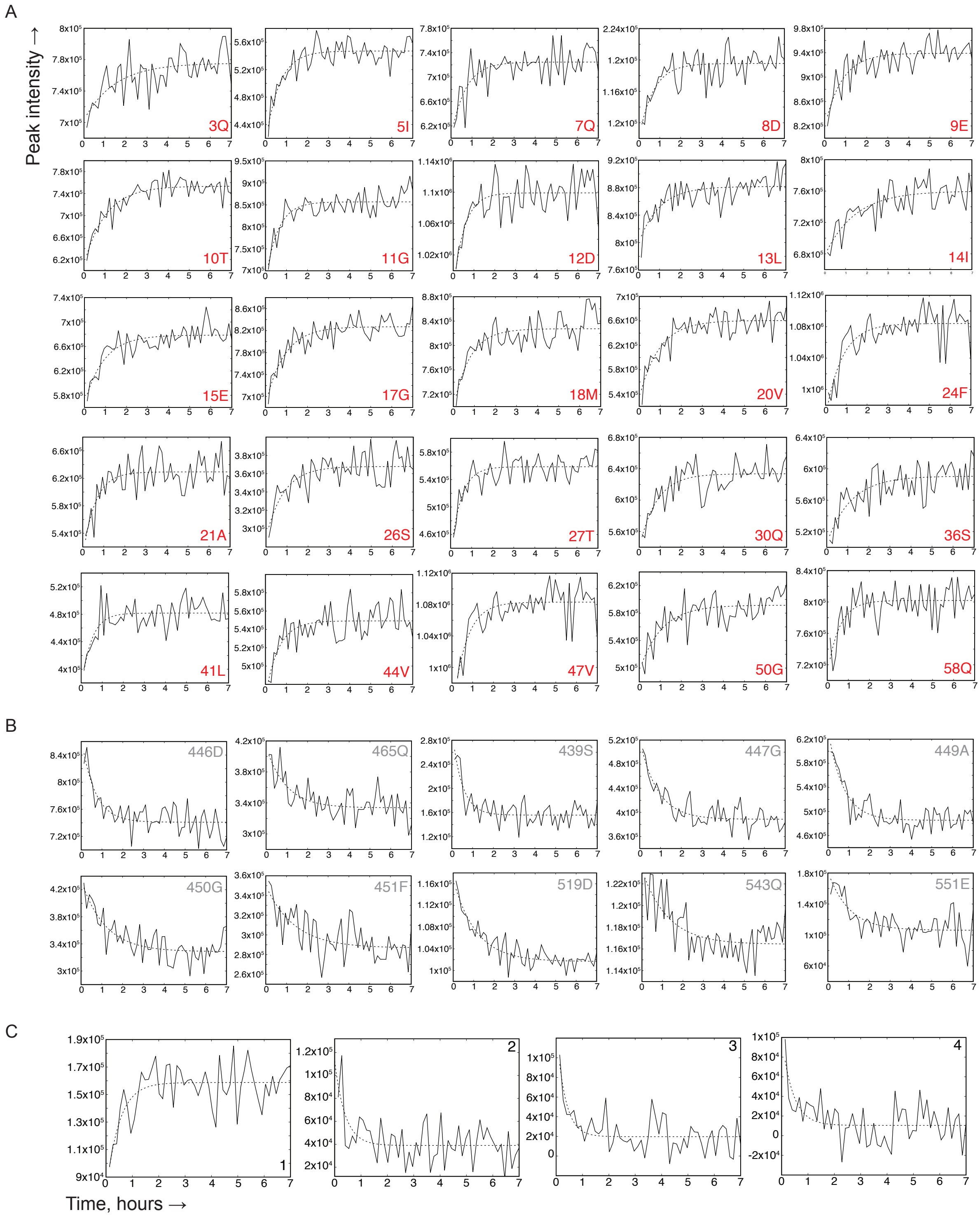


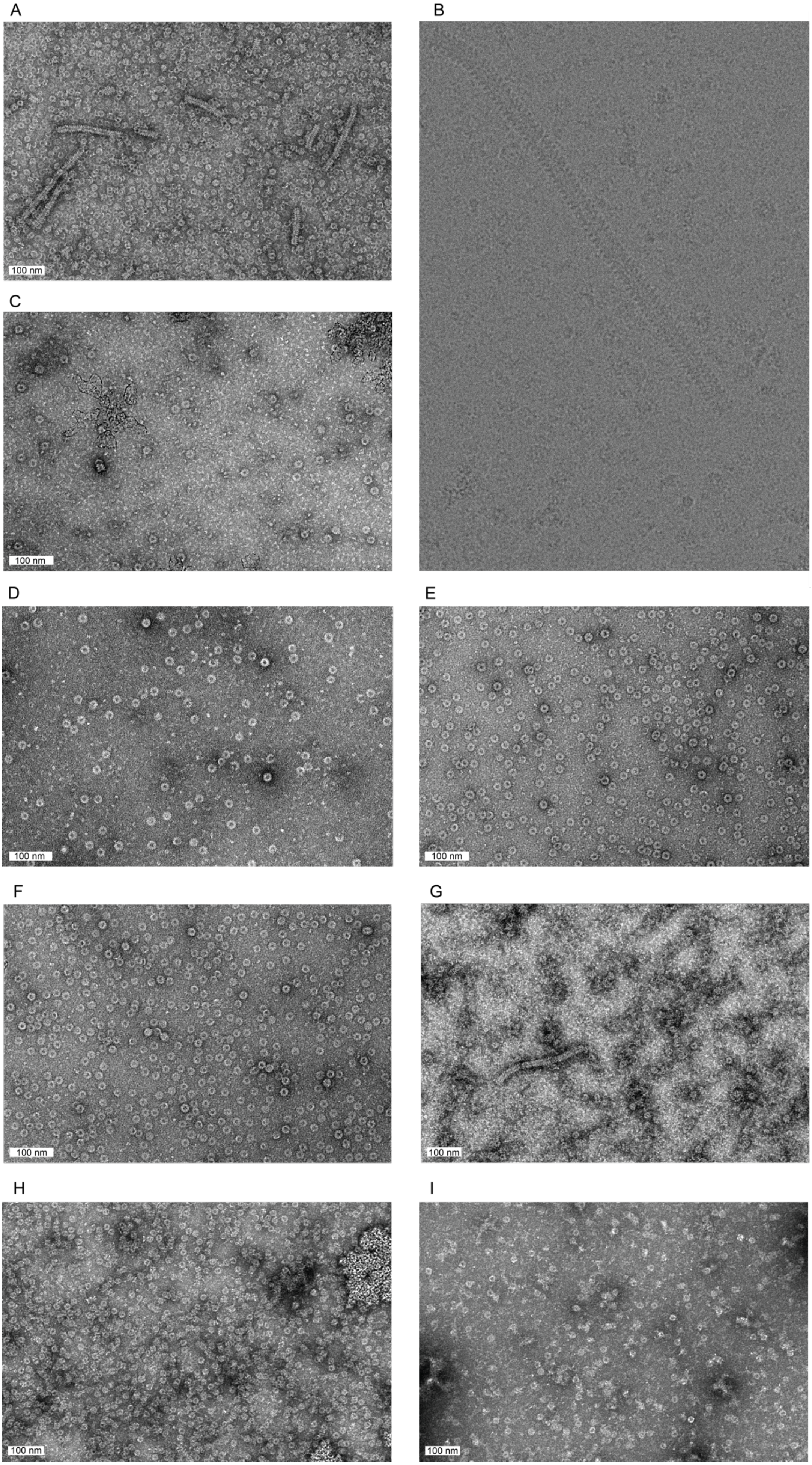
Figure 6.11. Kinetic profiles of each residue of phosphoprotein (A), nucleoprotein (B) and unassigned residues (C) of $P_{90}N_{549}$ changing in peak intensity after addition of RNA. The intensity of the last residues of P (from 66 to 89) were stable all over the experiment.

The profile of ^{15}N relaxation for P after assembly was quite different as compared to that measured before (Figure 6.10B), indicating the partial liberation of P_{1-90} from the rest of the $N^{0}P$, as the $R_{1\rho}$ rates still were increased in comparison with the rates from isolated PNTD measured at the same temperature and magnetic field. The Ntail relaxation profile, however did not change significantly (Figure 6.10C). However, only the region between C-arm and MoRE was affected and relaxation rates increased. This provides clear evidence of successful assembly and formation of oligomers. After assembly, the sample stayed for about one more extra night at RT for further relaxation recording and had some precipitation. This sample was homogenised and sent for EM analysis (Figure 6.12A). The mixture of RLPs and nucleocapsids with clean helical turns were observed also with cryo-EM (Figure 6.12B). However, NCs disappeared and probably were formed back to RLPs after storage for several days at 4°C, only N-rings were observed.

After this experiment, I tried to reproduce the experiments with freshly cleaved and uncleaved $P_{90}N_{549}$ (albeit at a different concentration). After eluting from the SEC column (concentration of about 10-20 μM), $N^{0}P$ was directly mixed with two types of 6-mer RNA: PolyA6 and mumps virus genome sequence MuV RNA and left for over-night incubation at RT. Unfortunately, this assay led only to the formation of N-rings instead of capsids (Figure 6.12C, D, E), even after keeping the sample for several days (assembly observed by NMR was carried out at a concentration of $N^{0}P$ at least 10 times higher and needed several days to form).

High sample concentrations were also tested with $P_{65}N_{549}$ and MuV RNA. Over-night incubation only led to the formation of RLPs (Figure 6.12F). The same sample with the addition of P_{NTD}, however, with poor appearance of N-rings, led to the formation of capsids after mixing with PolyA6 (Figure 6.12G). Unexpectedly, as the absence of N-arm would not favorise the formation of N oligomers, I observed N-rings that were formed with $P_{65}N_{32-549}$ and PolyA6 (Figure 6.12H). $P_{65}N_{32-549}$ supplemented with MuV RNA and 1 molar part of Ntail resulted the formation of disrupted N-rings (Figure 6.12I). No morphological differences were observed between the use of two different RNA sequences.

Figure 6.12 (next page). Electron micrographs of $P_{90}N_{549}$ (NMR sample) incubated with PolyA6 RNA (A), the same sample visualised with cryo-EM (B). EM images of uncleaved $P_{90}N_{549}$ incubated with PolyA6 RNA (C) and MuV RNA (D), $P_{90}N_{549}$ with MuV RNA and TEV (E), $P_{65}N_{549}$ with PolyA6 (F) and supplemented with P_{NTD} (G), $P_{1-65}N_{32-549}$ incubated with PolyA6 (H) and under addition of Ntail (I).


A
100 nm
B
C
100 nm
D
100 nm
E
100 nm
F
100 nm
G
100 nm
H
100 nm
I
100 nm

CHAPTER 7.
DISCUSSION

7. DISCUSSION

7.1. Structure and functional role of unfolded Ntail

The study of Ntail structure and function was one of the main priorities during the realisation of this project. In the beginning, it was quite frustrating and challenging that MuV Ntail does not interact with phosphoprotein, so several experiments were performed in order to identify Ntail's molecular behaviour or interaction with potential viral and host partners. Conformational dynamics were probed using different approaches: chemical shifts, $R_{1\rho}$, R_1, hetNOE, ASTEROIDS, RDCs, taking into the account the existence of conserved for *Paramyxoviridae* transient helical MoRE in Ntail.

Going deeper into the details, Ntail $T_{1\rho}$ relaxation rates measurements were informative enough to detect the regions with slightly higher rates: residues from 373 to 405, from 472 to 497, from 500 to 512 and 6 last residues before the His-tag on C-terminal end. However, MuV Ntail $R_{1\rho}$ in the mentioned regions were relatively low compared to partially structured regions in other *Paramyxoviridae* Ntails, that usually have at least 2 times higher transverse relaxation rates than those from random coil all over the protein sequence [Jensen et al., 2009; Abyzov et al., 2016]. In the mentioned regions, heteronuclear NOEs were also slightly increased and positive at 600 MHz magnetic field, designating the presence of the secondary structures. Chemical shift analysis and ASTEROIDS calculation showed these regions to be considered as having some rigidity and therefore, potential helical structure. Further RDC measurements indicated strong dipolar couplings at 373-405 and 472-497, confirming the suggestion that these regions are surely not a pure random coil. At the same time, amino acids from 500 to 512 were not affected in the alignment but showed higher relaxation rates.

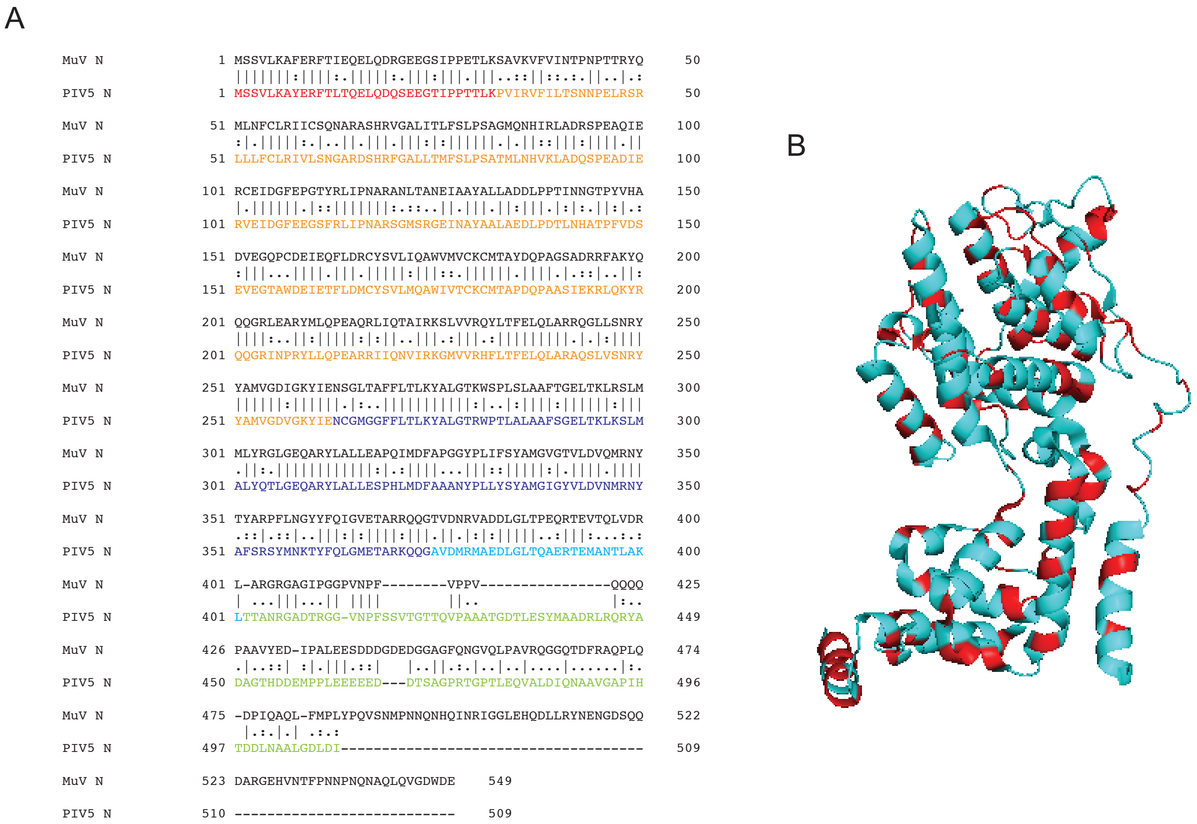
In addition, experiment with displacement of Ntail from N-rings by an excess of isolated Ntail revealed the disappearance of the region of 472-497 of free Ntail, providing evidence for the interaction between this helical region of Ntail and the surface of RLPs. This experiment has never been done before and actually, shows another important property of Ntail. However, during this interaction when N_{FL} was studied by NMR, Ntail was able to displace the whole Ntail from N_{FL} including residues 373-405. This experiment was repeated using EM and showed that free Ntail uncoiled the capsids formed from RLPs with and without the trypsination.

Starting from the protein sequence alignment studies between MuV and PIV5 nucleoproteins and having the crystal structure of PIV5 N oligomer, it was already possible to define the mumps virus residues from 373 to 405 as those corresponding to the Ncore C-arm (Figure

7.1A). Ntail interaction with N-rings with evidence of free Ntail binding to the Ncore surface of RLPs, increased $R_{1\rho}$ relaxation rates extracted from N_{FL} relaxation profile and missing residues from N^oP constructs (uncleaved $P_{90}N_{549}$ and $P_{65}N_{32\text{-}549}$ believed to be in the monomeric state) showed another important domain to consider 472-497, that could be called MoRE by analogy with other known *Paramyxoviridae* Ntail functional domain. The region after was called post-MoRE (Figure 7.2).

Both mentioned secondary structures MoRE and post-MoRE have high number of Qs and Ps in the sequence (Figure 4.1B) which are known to increase the flexibility of the protein. In addition, these two closely situated helical structures can be in exchange from one to another, or to be both in fast exchange, and, therefore, undetectable for adapted relaxation experiments, as for example $R_{1\rho}$ can quench the exchange contribution.

Concerning other interactions, the phosphorylation with CKII kinase was successful but why the protein needs to be modified and by which specific kinase still remained unclear. Except Ncore and kinase, it was not possible to find any other partner to interact with Ntail.



```
MuV N            1 MSSVLKAFERFTIEQELQDRGEEGSIPPETLKSAVKVFVINTPNPTTRYQ     50
                   |||||||:||||:.|||||:.|||:|||.|||..::||::.:.||..|.:
PIV5 N           1 MSSVLKAYERFTLTQELQDQSEEGTIPPTTLKPVIRVFILTSNNPELRSR     50

MuV N           51 MLNFCLRIICSQNARASHRVGALITLFSLPSAGMQNHIRLADRSPEAQIE    100
                   :|.|||||:.|..||.|||.|||:|:||||||.|.||::|||:||||.||
PIV5 N          51 LLLFCLRIVLSNGARDSHRFGALLTMFSLPSATMLNHVKLADQSPEADIE    100

MuV N          101 RCEIDGFEPGTYRLIPNARANLTANEIAAYALLADDLPPTINNGTPYVHA    150
                   |.||||||.|::|||||||:.::..||.|||.||:|||.|:|:.||:|.:
PIV5 N         101 RVEIDGFEEGSFRLIPNARSGMSRGEINAYAALAEDLPDTLNHATPFVDS    150

MuV N          151 DVEGQPCDEIEQFLDRCYSVLIQAWVMVCKCMTAYDQPAGSADRRFAKYQ    200
                   :|||...||||.|||.|||||:|||::.||||||.||||.|.::|..||:
PIV5 N         151 EVEGTAWDEIETFLDMCYSVLMQAWIVTCKCMTAPDQPAASIEKRLQKYR    200

MuV N          201 QQGRLEARYMLQPEAQRLIQTAIRKSLVVRQYLTFELQLARRQGLLSNRY    250
                   ||||:..||:|||||:|:||..|||.:|||.:|||||||||.|.|:||||
PIV5 N         201 QQGRINPRYLLQPEARRIIQNVIRKGMVVRHFLTFELQLARAQSLVSNRY    250

MuV N          251 YAMVGDIGKYIENSGLTAFFLTLKYALGTKWSPLSLAAFTGELTKLRSLM    300
                   ||||||:||||||.|:..|||||||||||:|..|:||||:||||||:|||
PIV5 N         251 YAMVGDVGKYIENCGMGGFFLTLKYALGTRWPTLALAAFSGELTKLKSLM    300

MuV N          301 MLYRGLGEQARYLALLEAPQIMDFAPGGYPLIFSYAMGVGTVLDVQMRNY    350
                   .||:.||||||||||||:|.:||||...|||::|||||:|.||||.||||
PIV5 N         301 ALYQTLGEQARYLALLESPHLMDFAAANYPLLYSYAMGIGYVLDVNMRNY    350

MuV N          351 TYARPFLNGYYFQIGVETARRQQGTVDNRVADDLGLTPEQRTEVTQLVDR    400
                   .::|.::|..|||:|:||||:|||.||.|:|:|||||..:|||:...:.:
PIV5 N         351 AFSRSYMNKTYFQLGMETARKQQGAVDMRMAEDLGLTQAERTEMANTLAK    400

MuV N          401 L-ARGRGAGIPGGPVNPF--------VPPV----------------QQQQ    425
                   | ...|||...|| ||||        ||..                |:..
PIV5 N         401 LTTANRGADTRGG-VNPFSSVTGTTQVPAAATGDTLESYMAADRLRQRYA    449

MuV N          426 PAAVYED-IPALEESDDDGDEDGGAGFQNGVQLPAVRQGGQTDFRAQPLQ    474
                   .|..::| :|.|||.::|   |..||.:.|..|..|....|......|:.
PIV5 N         450 DAGTHDDEMPPLEEEEED---DTSAGPRTGPTLEQVALDIQNAAVGAPIH    496

MuV N          475 -DPIQAQL-FMPLYPQVSNMPNNQNHQINRIGGLEHQDLLRYNENGDSQQ    522
                    |.:.|.| .:.:
PIV5 N         497 TDDLNAALGDLDI-----------------------------------    509

MuV N          523 DARGEHVNTFPNNPNQNAQLQVGDWDE    549

PIV5 N          510 ---------------------------    509
```

Figure 7.1. Nucleoprotein sequence alignment of MuV and PIV5 (A) was performed using the EMBOSS Needle server (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). The PIV5 N sequence coloured in red corresponds to N-arm, orange to NTD, blue to CTD, light blue to

C-arm, green to Ntail. The crystal structure of PIV5 N-unit (in red) with residues coloured in cyan being identical to MuN (B).

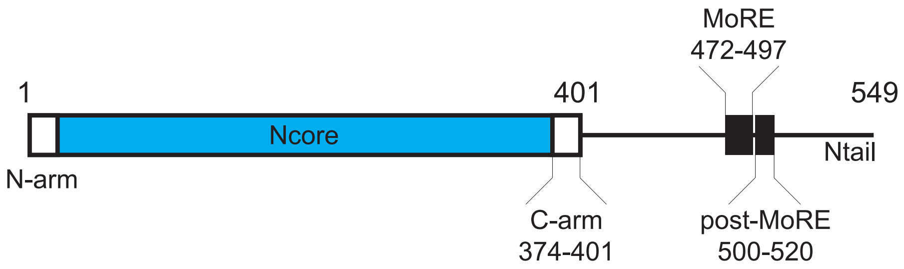


Figure 7.2. Schematic representation of Ntail structure according to the conformational behaviour of the protein.

7.2. Mumps virus N-rings as models to study N in the oligomeric state

N-rings are the major product of MuN expression in bacterial cells. It is clear that RLPs do not exist during mumps virus infection, they are artefacts formed due to the encapsidation of short random bacterial RNAs by nucleoprotein units. However, as I did not manage to produce separate Ncore truncations, N-rings can be a good model for further structural and functional analysis of N in the oligomeric state.

In this project, for the first time, N-rings were purified in a separate fraction on a SEC column. They tend to form capsids when kept at room temperature for long periods of time. The RLPs sticking between each other leasd to the formation of NCs and to the release of Ntail is supposed to that sterically protect the N-rings from interaction [Severin et al., 2016]. The expression of Ncore alone showing aggregates confirms this Ntail function. High salt conditions favour the stabilization of N-rings as well. The NC formation can be enhanced by Ntail proteolysis with trypsin, which was employed for getting capsids for further structure calculation using cryo-EM. Unfortunately, this method was not effective as formed NCs had different morphology and the helix pitch described by different interaction of N-rings between each other: several of them just stick to each other, other sticks and change the conformation of the part of the N-units to allow the formation of the helical turns by interacting of N molecule from one ring with N of another one and by forming capsids close structurally to those formed in the infected cells (Figure 4.9). This fact makes the NC structure calculation complicated and does not allow obtention of a high resolution structure.

Rings are composed of 13 N units which was shown before in the literature [Cox et al., 2009] and, in contrary to previous reports [Cox et al., 2009; Severin et al., 2016], N-rings could be obtained without any coexpression with P. Comparing the morphology of MuV N-rings to those of the evolutionarily closest PIV5, they appear to be quite similar which is expected from alignment of Ncore sequences (Figure 7.1). Conversely, PIV5 rings were obtained by proteolysing Ntail and were successfully crystallised with further structure determination by X-ray and did not form capsids [Alayyoubi et al., 2015]. This means that the release of PIV5 Ntail do not lead to the rings sticking and formation of NCs probably due to different functionality of their Ntails as protein sequence in the unfolded region do not align with each other (Figure 7.1A).

RLPs formation was also observed under the expression of N^oP complexes. Excluding the bacterial RNA contamination, the formation of N-rings is probably triggered by high affinity between N-units which is the main reason of N^oP instability and rapid P-peptide liberation over time. Generally, rings with and without RNA are energetically favourable. However, it was found that the excess of free P, more precisely, its first 65 residues were able to stabilise N in the monomeric state by out-completing interaction between N units. This particular feature will be discussed later.

Due to the size of N-rings, 2D HSQC NMR experiments of N-rings are limited by the partial visualisation only of Ntail peaks corresponding to the flexible regions alternating with those that are close to the Ncore (C-arm) and transiently structured MoRE and post-MoRE domains. These data are in agreement with the findings of MeV NCs [Jensen et al., PNAS 2011], NMR spectra of which also lacks the above-mentioned regions. For measles virus NCs, it was suggested that Ntail binds to Ncore in the fast regime which was also observed for MuV through interaction studies of ^{15}N-labelled N-rings with an excess of free unlabelled Ntail. Peaks corresponding to both regions that were not visible, increased in the intensity in this experiment. This observation can be explained by replacement of the Ntail of N_{FL} by free Ntail that interacts by its C-arm and MoRE with the surface of Ncore of RLPs.

2D HSQC of RLPs was not informative to get insight into interactions with P_{NTD} and P_{CTD} and EM images did not show any visible structural changes neither in the presence of P. However, when N-rings were treated with RNAse in order to get rid of encapsidated RNA, the mixture of P_{NTD} and P_{CTD} was able to break the RLPs. This observation suggests the dual role of the two P extremities to interact with Ncore in order to open the gate for RNA. This hypothesis needs more experiments to be carried out for further conclusions.

7.3. Disordered phosphoprotein N-terminal region interacts with monomeric and oligomeric N

Like Ntail, P_{NTD} also appeared highly disordered with no high propensities for secondary structure. However, summarizing all the dynamics measurements, simulations and also interactions with N, it was possible to identify three important regions that potentially can be considered as transient helices essential for interactions with oligomeric and monomeric N and host factors: $\alpha 1$ and $\alpha 1'$ (residues 13-60), $\alpha 2$ (117-142) and $\alpha 3$ (189-204) (Figure 7.3). Upon interaction of P_{NTD} with oligomeric N, these regions had increased relaxation rates. This binding could be considered as weak in agreement with findings from the literature [Cox et al., 2013].

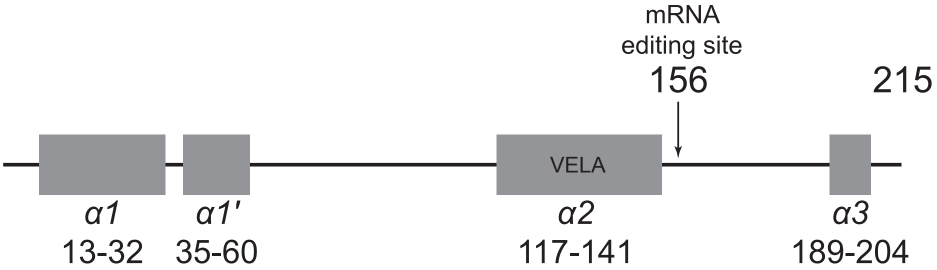


Figure 7.3. Schematic representation of the structural units of MuV P_{NTD}.

During the generation of N^0P constructs, three interaction sites for P_{NTD} with monomeric and oligomeric N were found. The first interaction site was located from residues about 117 to 141 (corresponds to $\alpha 2$) and could be involved in the formation of N^0P. This binding site could be comparable to the weak interaction site located in the middle of MeV P_{NTD} with anamino acid sequence HELL [Milles at al., 2018]. From MuP alignment with MeV P (not shown), this region is ^{125}VELA^{128} and perfectly fits with the helical propensity from the experimental data (Figure 5.3, 5.4).

In view of assessing the conservation of the VELA interaction site among *Rubulavirus* members, I have analysed the structure of the PIV5 V protein that has the same disordered N-terminal part as PIV5 P (Figure 7.4B). After alignment with MuP (Figure 7.4A), it seems that PIV5 V has a similar amino acid stretch VKLA which, in the crystal structure of the V protein in complex with the host protein DDB1 [Li et al., 2006], adopts an α-helix (Figure 7.4B). Two β-strands occurring before and after VKLA having almost identical residues as mumps virus P/V, also could be considered as precursors for the most significant secondary motif. In order to show the approximate localisation of the conserved P_{VELA} binding site within Ncore, I will take into the consideration the PIV5 crystal structure of N-rings and findings from MeV P about the position on Ncore where this weak binding occurs.

A

```
MuV P        1 MDQF---IKQDETGDLIETGMNVANHFLSTPIQGTNSLSKASILPGVAPV   47
               ||..   ...||...|||||:|...:|.|..:.||:||.|.:|.|||..:
PIV5 V       1 MDPTDLSFSPDEINKLIETGLNTVEYFTSQQVTGTSSLGKNTIPPGVTGL   50

MuV P       48 LIGNPEQKNIQHPTASHQGSK---TKGRGSGVRSIIVSPSEAGNGGTQIP   94
               |....|.| ||..|...:||.   .|.:....:..||...:....|..||
PIV5 V      51 LTNAAEAK-IQESTNHQKGSVGGGAKPKKPRPKIAIVPADDKTVPGKPIP   99

MuV P       95 EPLFAQTGQGGIVTTVYQDPTIQPTGSYRSVELAKIGKERMINRFVEKPR  144
               .||..........|.:.......|:|||:.|:|||.|||.::.||:|:||
PIV5 V     100 NPLLGLDSTPSTQTVLDLSGKTLPSGSYKGVKLAKFGKENLMTRFIEEPR  149

MuV P      145 -----TSTPVTEFKRG    155
                    ||:|: :||||
PIV5 V     150 ENPIATSSPI-DFKRG    164
```

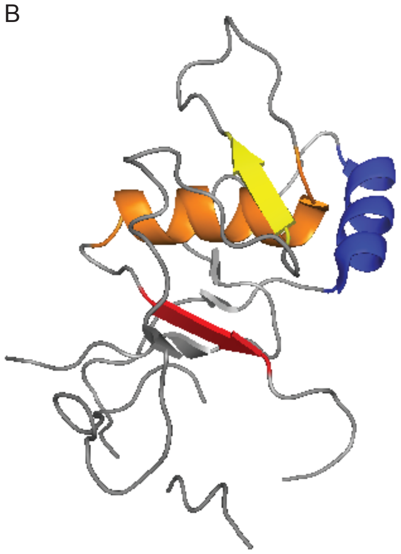


Figure 7.4. Sequence alignment of MuP NTD and PIV5 V N-terminal domains with coloured folded domains that appeared upon binding and are distinguishable from crystal structure. VELA region is in bold italic, PIV5 P peptide (from crystal structure of PIV5 N^{0}P) responsible for N$_{0}$P chaperoning is underlined (A). Crystal structure of PIV V protein [PDB 2HYE] without the visualisation of ZFD of V, DDB1, Cullin-4A and RING-box protein 1. Structural domains of V are coloured in the same way as in the sequence alignment (B).

From MALLS profile and NMR measurements, MuV RLPs were found to be bound to P$_{1\text{-}141}$ which agrees with the fact that the VELA interaction site is not affected by N oligomerisation from P$_{141}$N$_{549}$ and is probably located on the external side of N-rings. Alignment of the nucleoprotein sequences of MeV, MuV and PIV5 (Figure 7.5A) of residues from 90 to 125 showed the conserved amino acids that were actually identified as those participating in MeV P$_{\text{HELL}}$ interaction with N-lobe of Ncore. These residues were mapped on the structure of PIV5

N-rings which confirms the position of the interaction suggested from MeV (Figure 7.5B). However, the function of this binding for MuV life cycle could not be interpreted for the moment.

Other binding sites were revealed from increased $R_{1\rho}$ relaxation rates from the shorter P peptide (1-90) from interaction with monomeric N in N^{o}P complex: *α1* (13-32), *α1′* (35-60) and the region comprising residues 77-85 without adefined structure type. The first two are potentially responsible for chaperoning N in the monomeric RNA-free conformation, as the generation of N^{o}P with shorter P (1-65) did not change the binding with N as estimated from the overall P relaxation profile.

A

```
MuV N          90 ADRSPEAQIERCEIDGFEPGTYRLIPNARANLTAN     125
                  ||:||||.|||.||||||.|::|||||||:.::..
PIV5 N         90 ADQSPEADIERVEIDGFEEGSFRLIPNARSGMSRG     125
                  ....|:..|..:|:...::....|...:|......
MeV N          90 ITDDPDVSIRLLEVVQSDQSQSGLTFASRGTNMED     125
```

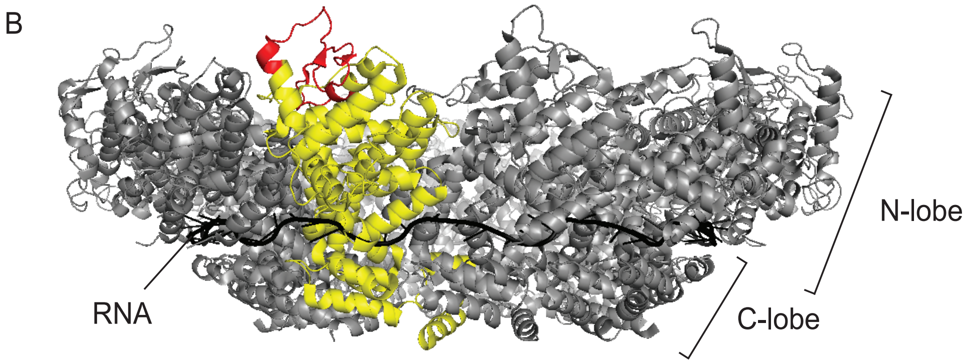


Figure 7.5. Sequence alignment of N-lobe region (residues from 90 to 125) of MeV, MuV and PIV5 (A). Approximate localisation of potential MUV P_{VELA} interaction site (in red) for one N-unit (in yellow), RNA is in black (B). Crystal structure of PIV5 N-RNA complex was used as a model [PDB 4XJN].

The N^{o}P crystal structure of other *Mononegavirales* members showed P peptide composed of two helices (Figure 1.13A). However, for PIV5 it is possible to see only the first P *α*-helix (Figure 1.13B, 7.6), which starts from residue Leu16 corresponding to MuP Leu13 in agreement with the relaxation measurements (Figure 7.4A). When N-arm was removed for MuV N^{o}P complex ($P_{65}N_{32\text{-}549}$) to prevent oligomerisation, by analogy to PIV5, it led to the disappearance of *α1* while peaks of *α1′* only decreased in intensity. This could be the direct evidence of a role of *α1* in the interaction with monomeric N, while P *α1′* and residues 77-85 are secondary interaction sites with Ncore (Figure 7.3). Concerning the position of the $P_{13\text{-}32}$ binding site, it could be suggested that it is located in the same place of Ncore as it was observed for PIV5 N^{0}P (Figure 7.6) by partially occupying the RNA pocket and by perturbing the access of the separate N-unit to the N-arm and C-lobe of Ncore. We can also imagine the number of transient helices that can have contacts with the surface of Ncore, as well as the VELA-interaction site (Figure 7.6). However, to get more evidence for this model, more additional binding experiments need to be performed.

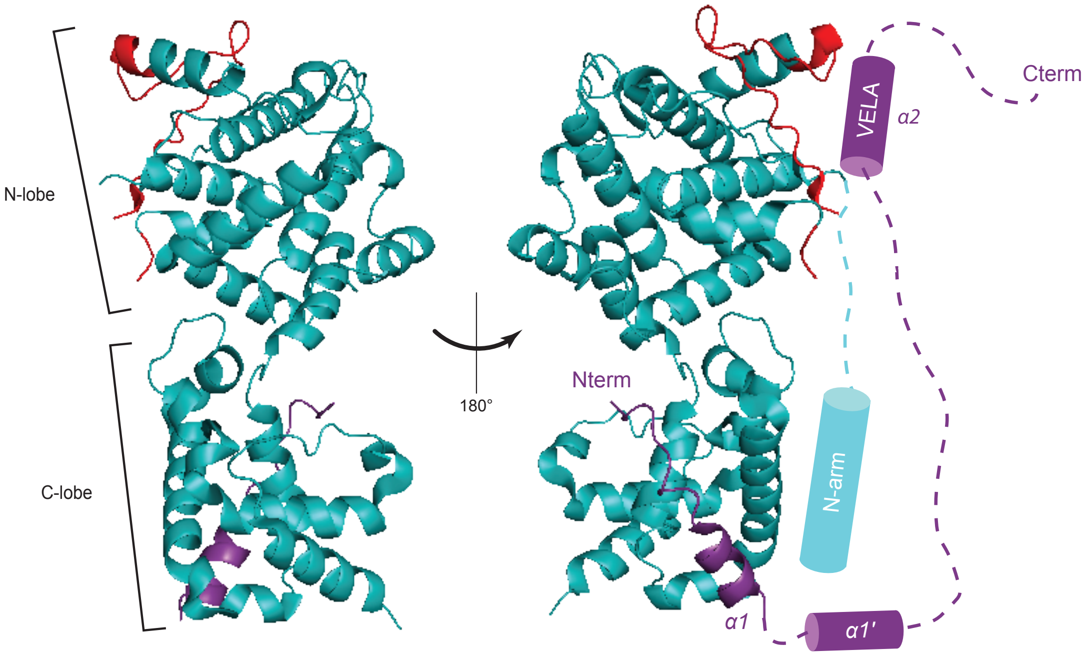


Figure 7.6. Model based on the crystal structure of PIV5 N^{0}P [PDB 5WKN], representing the approximate localisation of MuV P_{NTD} (in violet) and Ncore (in cyan) interaction sites. α-helices are shown as cylinders. Potential VELA interaction site within the N-lobe is coloured in red.

7.4. *The C-terminal domain of phosphoprotein has a flexible region with transient secondary structures*

It has been known that MuV P_{XD} expressed alone is not stable, as even its crystallization required the addition of stabilising agents [Kingston et al., 2008]. However, adding the flexible linker (residues 272-342) between OD and XD to the expression vector apparently helps XD to remain in a stable conformation. This could be probably explained by the presence of two secondary structures that have elevated $R_{1\rho}$ and long-range weak interactions with the three-helix bundle that assessedthrough PRE studies. In addition, the contacts with these regions were also observed under the titration of P_{CTD} with N_{FL} that led to a decrease in peak intensities and increase in relaxation rates in two regions of P named *α4* (286-301) and *α5* (316-323) (Figure 7.7). It was also interesting to discover that the phosphorylation by PLK1 occurred at positions of S307 and S292 which led to the change in the peptide chain structure. Consequently, these modifications could be necessary for the decrease of the affinity to oligomeric N which for the moment could not be confirmed.

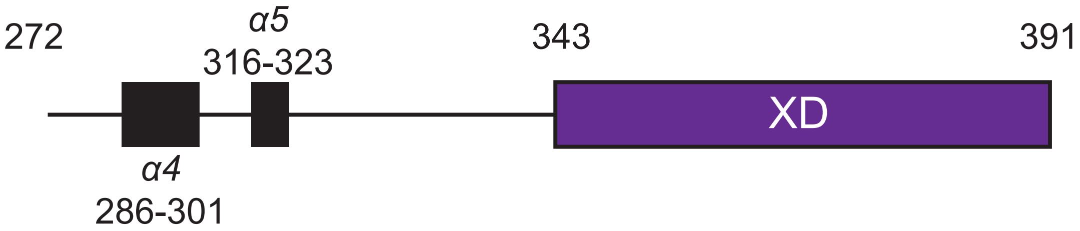


Figure 7.7. Structural organisation of MuV P_{CTD}, it comprises from two transient helices on the disordered linker and three-helix bundle on the C-terminal region.

7.5. *Interaction of XD of P_{CTD} with RLPs*

Apart from sequence differences, one of the MuV unique properties distinguishing it from other *Paramyxoviridae* family is its three helix bundle of P C-terminal that interacts with Ncore and not with Ntail. The binding affinity was much higher than that identified for P_{NTD} with N [Cox et al., 2013].

During the titration of P_{CTD} with N_{FL}, under the gradual increase of N-rings concentration, HSQC peaks corresponding to XD started to disappear. However, RLPs are challenging to study, as they tend to aggregate over time, especially, in high concentrations (up to 100 µM which was used for relaxation measurements), so further long experiments could not provide clear conclusions about the exchange regime describing the binding between XD and Ncore. Also the position of the binding site and its morphology could not be identified.

7.6. *Mumps virus N^oP complex is stabilised by an excess of P_{NTD}*

As mentioned before, the generation of a stable N^oP complex was very challenging due to its gradual transformation to N-rings. However, this fact added certain clarification to the functionality of MuV N^oP. Previously discussed interactions sites were identified, as well as the important factor leading to the N^oP stabilisation.

Th generation of a monomeric PIV5 N^oP construct was difficult too as it tended to form RLPs. Generally, N^oP expressed without N-arm could not be used for the assembly experiments.

However, they are essential to study the interactions between P and monomeric N. For this reason, I was trying to find the reasons of N^0P instability without losing its important functionality to assemble in NCs. Possible traces of RNA that could trigger the formation of N-rings were not identified, so it could be suggested that the main problem of MuV N^0P stability is a very high affinity between N units which is certainly induced by the presence of N- and C-arms.

In search of a stable complex, an interesting feature was observed. In the presence of an excess of free P peptide there is a stable fraction of N^0P present (Figure 7.8) which is certainly connected to the concentration dependence. This suggests that in the host cellular conditions, MuV can organise its RNA replication inside inclusion bodies formed with help of disordered extremities of P or V/I proteins which potentially have unstructured regions. The physico-chemical conditions inside these bodies can be modulated, including protein concentration. In addition, it was found that the excess of P_{NTD} can enhance viral gene expression [Cox et al., 2014], which correlates with the fact that more N^0P will be protected with P_{NTD}, more replicated viral RNA will be further successfully encapsidated with the help of new incoming N^0Ps.

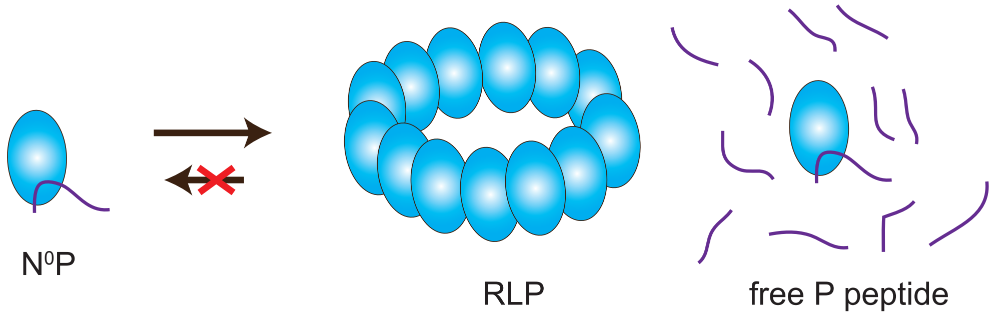


Figure 7.8. Schematic representation of the formation of the stable fraction of MuV N^0P by liberating P peptide from formed N-rings.

7.7. Assembly of N^0P in NCs is possible under high sample concentrations

$P_{90}N_{549}$ was described to interact with PolyA6 RNA which led to the formation of NCs. These capsids can be formed either from RLPs that were generated firstly from instable N^0P or from an assembly process. As following P liberation and N engagement in oligomeric structures could actually lead to the formation of either N-rings or NCs. At the same time, this NMR experiment was done with high concentration of N^0P which certainly favoured to have a stable monomeric N and the direct NC assembly (Figure 7.9). When I tried to repeat the same experiment with lower concentration, only N-rings were observed by EM. This can explain the fact that the high concentration of N^0P and therefore, the higher number of P peptides is able to protect N^0P from transformation to RLPs, is crucial for NC assembly.

On the contrary, the assembly needs to be improved in order to obtain pure capsids without a major fraction of N-rings in order to avoid the secondary formation of capsids from them. This may be possible if longer RNA sequences were used, that could disturb the preorganisation of N in rings and lead to the "natural" assembly. In addition, formed from assembly NCs were not stable upon storage at 4°C and degraded before I wanted to perform the cryo-EM measurements. The presence of free P peptide could potentially also influence the assembly by uncoiling formed NCs and having the competition with the N-arm.

On the same time, it was discovered that the N-arm cleaved N^0P ($P_{65}N_{32\text{-}549}$) eventually formed N-rings under the incubation with RNA which could suggest that the assembly process for MuV is regulated first, by affinity between N-units, and second, by the interaction with RNA, however, the order of these two processes is unknown.

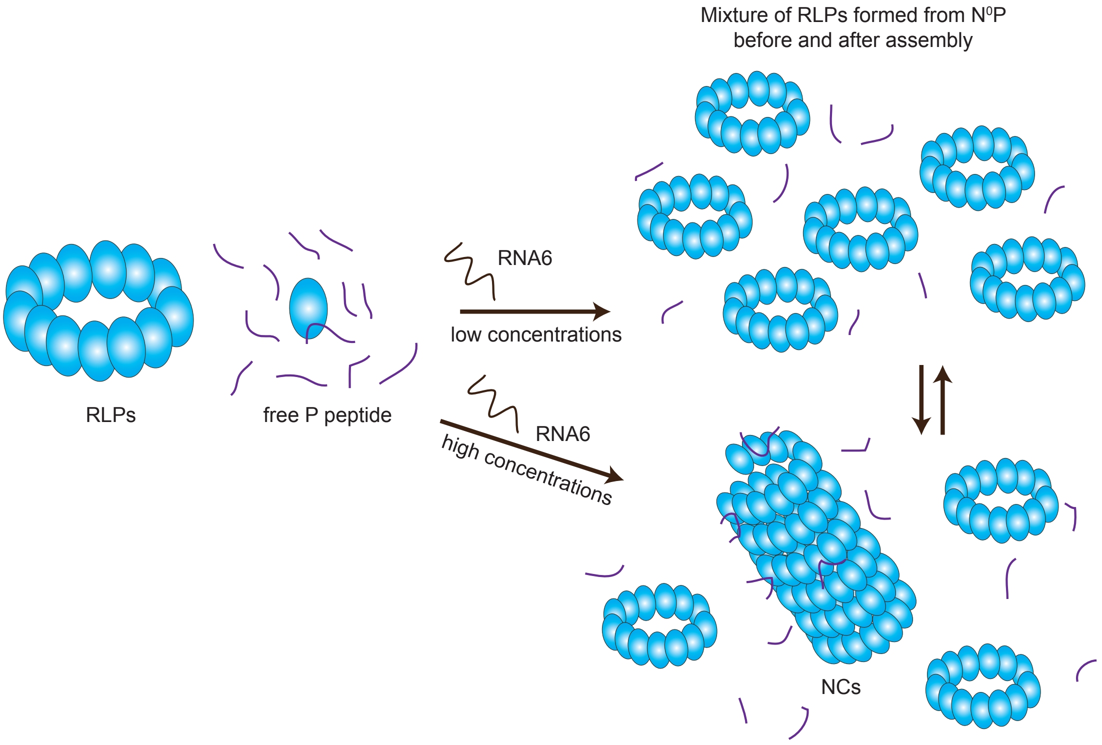


Figure 7.9. Schematic representation of the expected for NC assembly process for MuV.

CHAPTER 8.

CONCLUSIONS AND PERSPECTIVES

8. CONCLUSIONS AND PERSPECTIVES

Mumps virus is not the deadliest virus in the world, and there is an available vaccine for its prevention. However, it has been found that with time vaccines become inefficient, and, therefore, mumps outbreaks still occur. This process could be difficult to control in the near future. Furthermore, mumps disease could bring serious health consequences. Epidemiological studies would provide a clear vision of the current situation with mumps in the world but now the research, which is dedicated to drug discovery, needs to be argumented. Specific and highly effective medicines are the best solution for treatment of infectious diseases, as they can target important steps of viral life cycle, this includes viral genome transcription and replication. Mumps virus remains underexplored and poorly studied due to the fact that humans are only natural hosts for mumps, so other biological models could not be used to understand at least the spread and the circulation of the viral infection in the human body.

Taking together all the facts, the study of the mumps virus machineries at the molecular level is essential to prevent and treat the mumps disease. Thus multifunctional mumps proteins, nucleoprotein and phosphoprotein, mediating gene synthesis inside the host cell are the main object of this work. Belonging to the family of *Paramyxoviridae*, mumps virus has some shared features as well as differences making its genome transcription and replication, unique among known *Paramyxoviridae*.

In this project, NMR spectroscopy was an ultimate method that allowed to investigate disordered regions in mumps N and P proteins. In addition, this technique is able to detect ultra-weak interactions between disordered and folded proteins. Electron microscopy was used to visualise oligomeric structures, their morphology and some interactions between oligomeric and monomeric N with P and RNA. Additional methods such as SAXS and MALLS both coupled with SEC were less useful. However, they could be critical for routine MuV protein analysis during the further development of the mumps virus project.

Concerning N, despite the low sensitivity of standard NMR methods, it was possible to observe the localisation of MoRE along Ntail sequence. The Ntail partner responsible for the functional interaction was not identified. However, interesting feature was discovered for the MoRE that binds to the surface of N-rings. In addition, another transient helix was detected coming after MoRE and called post-MoRE, which is suggested to somehow destabilise the MoRE formation that could explain its relatively low population in secondary structure. Nevertheless, as a perspective, more detailed study of Ntail alone and connected to the Ncore needs to be performed. Adding RDCs data as an input for ASTEROIDS ensemble generation is essential to better describe the conformational behaviour of Ntail. The role of Ntail post-translational modification needs to be determined as well by some biological methods, possibly in collaboration with research groups from University of Alabama at Birmingham, USA.

For the first time MuV N-rings were purified and characterized by NMR. This helped to confirm the existence of MoRE region and its role in RLP stabilisation, as the N-rings without Ntail tend to stick to each other. Unfortunately, due to the sample degradation and transformation to stacked N-rings, or NCs, it was quite difficult to get the high resolution structure of capsids and rings. On the contrary, the problem of the unassigned Ncore could be solved using solid state NMR assuming that well structured NCs could be reproducibly produced. As a bonus, this method could help to find the localisation of P_{CTD} binding site within the Ncore.

MuV phosphoprotein structure can be described by different transient secondary structures that alternate with structured parts that have already been resolved by X-ray crystallography. Unfortunately, it was not possible to study the full-length phosphoprotein, so N- and C-terminal parts without oligomerisation domain were studied separately. However, the optimisation of stable P_{FL} purification is necessary in order to understand the simultaneous action of both extremities P_{NTD} and P_{CTD} being together in both sides of the P_{FL} molecule.

P_{NTD} shows signs of random coil conformation which can adopt helical propensities. They were found to interact with oligomeric and monomeric N. At least three regions with low helical population bind to the surface of N-rings and do not show any expected visible distortion of RLP structure. Furthermore, P_{NTD} interacts with N in order to keep it monomeric (N^oP complex) by its first 30 residues (helix *α1*) by analogy to the crystal structure of PIV5 N^oP. Other regions, such as helix *α1* and residues 77-85, are also binding sites but the role of these interactions is unclear. Another binding site *α2* with N was identified in the middle of P_{NTD} sequence and is believed to have the same binding feature as MeV P by having the similar to HELL amino acid scretch VELA. The confirmation of evidence of the exact localisation of interaction site on Ncore requires more experiments to be performed such as the generation of stable monomeric N either by the optimisation of protocols for Ncore and separate lobes expression or by the stabilisation of N^oP. The third region of potential helical region is located in the C-terminal part of P_{NTD}. With the exception increased relaxation rates under interaction with N_{FL}, it was poorly studied. Therefore, more NMR experiments needs to be done in order to reveal the dynamic and functional properties of all P_{NTD} structural elements.

Besides the already known unstable three helix bundle of XD, P_{CTD} contains a disordered linker connecting folded OD and XD. It has also two regions exhibiting transient secondary structure and long-range weak contacts with XD. It also comprises a recognition site for PLK1 kinase that was able to phosphorylate two serine residues. This modification could be suggested to decrease the affinity of XD and Ncore. However, more binding experiments need to be carried out.

One of the most challenging parts of this work was the generation of N^oP. After expression and study of a lot of different fusion complexes, it was concluded that mumps N^oP exists in stable conformation only in the presence of an excess of isolated free P_{NTD} which was known from the literature to enhance viral RNA synthesis. This finding could suggest a crucial role of membraneless organelles (viral factories) stabilised by disordered viral protein regions, in the processes of viral genome transcription and replication, and therefore, able to maintain high IDP concentrations. In the same time, more attention needs to be put to other methods of N^oP stabilisation which could allow to study N in the monomeric form without the addition of the excess of P. For example, the promising N^oP construct without N-arm can be further studied to identify the affinities between P and N.

N^oP was also tested in the assembly in NCs and showed significant dependence on sample concentration. Under the addition of short RNA, I was able to observe the liberation of P from N^oP and engagement of N in NCs. Nevertheless, it remains unclear, what was formed first RLPs or capsids. Consequently, this experiment needs to be optimised in order to mimic natural conditions where NCs are formed without the preformation of N-rings which are essentially, artefacts of N oligomerisation.

My initial experiments on V protein unfortunately did not lead to the formation of the monomeric protein due to the cysteine-rich C-terminal that is difficult to fold in Zn-finger conformation. Follow-up studies need to be performed using other *E. coli* strains that allow cysteines to keep in the reduced form and avoid intermolecular interactions. In the same time, other parameters of denaturant conditions could be optimised in order to get the proper ZFD folding. V protein was shown for PIV5 to have a stable complex with N. This ability could be used in assays to replace P peptide in N^oP complex by V. It would be also good to start the research on I protein that is also believed to be functional all over the viral gene synthesis process.

To conclude, mumps virus has several points that are common with other *Paramyxoviridae* members such as the existence of helical structure MoRE in Ntail, conserved transient structural domains in P_{NTD} and their function, morphology of helical bundle of XD, and different from *Paramyxoviridae* features like the function of XD and Ntail, stability of N^oP complex and the way it assembles in oligomeric structures. In the perspective, a promising approach will be to include to the study other subfamily *Rubulavirus* members such as PIV5, PIV2 and PIV4.

9. RÉSUMÉ EN FRANÇAIS

Ce manuscrit de thèse se compose de quatre parties et huit chapitres : l'introduction, deux chapitres des matériels et méthodes, quatre chapitres dédiés à la présentation des résultats et leur discussion et le dernier chapitre commentant les conclusions et perspectives du travail effectué.

9.1. Introduction

9.1.1. Les vaccins contre le virus des oreillons

Les oreillons sont une maladie infectieuse décrite pour la première fois par le père de la médecine, Hippocrate, dans son œuvre Epidémie en 5ème siècle avant JC quand la maladie a touché les habitants de l'île Thasos. Le mot « oreillons » provient de la douleur au niveau des oreilles pour la personne infectée par le virus ourlien.

Dans un premier temps, le virus affecte les glandes parotides (parotidite) provoquant des céphalées et de la fièvre. Après 2-10 jours les symptômes disparaissent mais dans la moitié des cas, une méningite aseptique ou/et une encéphalite peuvent se développer. Le taux de mortalité reste faible mais parmi les complications, des troubles du système nerveux et reproductif sont observés. L'homme étant l'unique hôte pour le virus, sa transmission n'existe qu'entre les humains. Les oreillons sont une maladie infantile, touchant les enfants de 5-9 ans. Toutefois, les adolescents et les adultes peuvent aussi être infectés.

Il n'existe pas de thérapie contre la maladie. Son traitement est symptomatique. L'unique solution pour contrôler la propagation du virus est l'utilisation des vaccins. Le premier vaccin mis au point en 1946 aux Etats-Unis, contenait un virus inactivé. Cependant, il n'induisait pas une protection immunologique efficace. En 1967, il a été remplacé par le vaccin Jeryl Lynn possédant une efficacité d'immunisation de 80-100%. Par la suite, il a été commercialisé sous la forme d'un vaccin trivalent, le ROR, permettant une protection contre les virus de la rougeole, des oreillons et de la rubéole (ROR). D'autres pays dans le monde ont eux aussi développé différentes variétés de vaccins qui diffèrent par les cellules substrats pour leur préparation et les procédés de fabrication : Leningrad (ex-USSR), Leningrad-Zagreb (Croatie), Urabe Am9, Torii, Hoshino (Japon), S-12 (Iran).

Après l'introduction du vaccin, le nombre de cas dela maladie a fortement diminué dans les pays développés. Néanmoins, le début des années 2000 a été caractérisé par plusieurs épidémies aux Etats-Unis, en Belgique, au Royaume-Uni, aux Pays-Bas, au Canada, en République Tchèque.

Par comparaison avec la situation épidémiologique de la rougeole qui présente une fréquence plus importante, les oreillons ne sont pas considérés comme une maladie grave, se reflétant sur les investissements scientifiques. Cependant, les épidémies apparues au sein des populations vaccinées et saines devraient attirer l'attention internationale sur l'apparition des maladies infectieuses qui seront difficile à prévenir et contrôler dans un avenir très proche.

Les facteurs influençant l'apparition des épidémies récentes s'expliquent par:

- une couverture vaccinale faible et incomplète liée au mouvement anti-vaccination et aux défauts de la législation d'obligation de vaccination ;
- la variation antigénique et l'évasion immunitaire associées aux vaccins issus d'un génotype viral pouvant être inefficace contre le virus d'un autre génotype circulant pendant l'épidémie ;
- l'échec du vaccin et la baisse de l'immunité expliquant la propagation de la maladie parmi des populations vaccinées à l'âge adulte.

Des solutions peuvent être apportées par des recherches statistiques, épidémiologiques et biochimiques plus profondes qui pourraient conduire à la conception d'un nouveau médicament efficace.

9.1.2. La structure du virus ourlien

Le virus des oreillons appartient à la famille *Paramyxoviridae*, genre *Rubulavirus*, ordre *Mononegavirales*, ordre regroupant tous les virus enveloppés possédant un génome d'acide ribonucléique (ARN) non-segmenté, de polarité négative.

Les particules virales (virions) sont principallement sphériques. Pourtant elles peuvent différer en taille et en forme. Les virions sont composés d'une bicouche lipidique dérivée de la membrane de la cellule hôte. Les glycoprotéines (HN et F) sont insérées dans la membrane et sont essentielles pour l'attachement du virion sur les récepteurs cellulaires. A l'intérieur de la particule virale, l'ARN est enveloppé par de multiples copies de nucléoprotéine (N) lui donnant une structure filamenteuse appelée nucléocapside (NC). La taille de l'ARN codant pour toutes les protéines est entre 15000 et 19000 bases, avec une molécule de N pour 6 bases d'ARN. Les NCs avec la phosphoprotéine (P, le cofacteur de l'ARN polymérase) et la protéine large (L, l'ARN polymérase ARN-dépendante) forment le complexe de transcription-réplication. Les NCs servent de matrice pour la synthèse de l'antigénome (l'ARN à simple brin à polarité positive). La protéine de matrice (M) connecte les parties externe et interne du virion en interagissant avec la N et les deux glycoprotéines.

La N est composée de deux parties : une partie repliée N-terminale (Ncore) et une partie C-terminale prédite désordonnée (Ntail). Ncore est également divisé en deux lobes (N- et C-) possédant chacun des extrémités, appelées N- et C-arms (bras). Ncore est responsable de l'interaction avec l'ARN. Ntail de certains *Paramyxoviridae* et du virus de la rougeole (MeV) en particulier, présente un motif de reconnaissance moléculaire (MoRE), impliqué dans l'interaction avec la phosphoprotéine. Le Ntail du virus des oreillons (MuV) ne possède pas cette caractéristique.

Dans les cellules infectées, N existe sous deux formes, oligomerique en NCs et monomerique en complexe N^oP. L'expression recombinante en bactérie de la nucléoprotéine seule conduit à la formation de NCs et/ou des particules annulaires (RLPs) par incorporation aléatoire d'ARNs. Les NCs de *Paramyxoviridae* ont une symétrie hélicoïdale gauche, avec un diamètre d'environ 20 nm et une longueur de 1 μm et avec 13 unités de N par tour. Pour le moment, une seule structure à résolution atomique est connue, celle de la NC du virus de la rougeole. La formation de RLPs a été observée pour MuV, mais pas lors de l'infection virale.

La seconde forme fonctionnelle de la N est le complexe N^oP, où le peptide N-terminal de phosphoprotéine chaperonne la N dans un état monomerique, sans ARN. Pendant la réplication virale, ce complexe est nécessaire afin de prévenir l'assemblage de nouvelles molécules de N qui serviront à l'encapsidation de l'ARN naissant. Plusieurs structures cristallines de N^oP ont été résolues, nottament pour les virus de la rougeole, parainfluenza 5 (PIV5), Nipah (NiV) etc. La superposition des structures de N monomerique et oligomerique suppose l'existence de deux conformations « switch », ouverte et fermée respectivement.

La Phosphoprotéine du virus des oreillons se compose de trois régions fonctionnelles: le domaine N-terminal (P_{NTD}) avec une conformation prédite dépliée, le domaine d'oligomerisation (P_{OD}) auto-associé (quatre molécules de P, 2 parallèles, 2 antiparallèles, conformation unique parmi les *Paramyxoviridae*) et le domaine C-terminal (P_{CTD}) comportant un faisceau de trois hélices-α (région XD). Les structures de P_{OD} et XD ont été résolues par diffraction des rayons X.

Trois protéines différentes peuvent être transcrites à partir du gène de P: P, V et I. Ce phénomène provenant de l'édition de l'ARN messager (ARNm), est basé sur l'insertion d'une ou deux guanosines (G) lorsque l'ARN polymérase virale glisse en arrière sur la séquence de l'ARNm. Pour MuV, V provient de la séquence non éditée, alorsque I et P se forment après l'insertion de Gs dans le codon codant pour l'acide aminé 156 de P. Les trois protéines partagent donc la même partie N-terminale.

Comme mentionné précédemment, les 60 premiers résidus de P_{NTD} participent à la formation du complexe N°P. Récemment, un autre motif au milieu de la séquence de P_{NTD} de MeV avec une affinité relativement faible pour le Ncore a été identifié. Concernant MuV, P_{NTD} interagit avec les surfaces des RLPs. En comparaison avec les autres *Paramyxoviridae*, où XD se lie avec Ntail, les trois hélices de P_{CTD} de virus des oreillons s'associent avec le Ncore, pouvant expliquer la morphologie unique de l'oligomère P possédant les extrémités de N- et C-terminal sur les deux côtés.

La protéine V possède deux doigts de zinc dans sa region C-terminale qui joueraient un rôle important dans la suppression de la réaction immunitaire cellulaire. Pour PIV5, V forme un complexe N°V plus stable que N°P, grâce aux mêmes NTDs. La structure et le rôle de la protéine I sont peu documentés pour le moment.

La nucléoprotéine et la phosphoprotéine, notamment dans les séquences désordonnées riches en sérines et thréonines, peuvent présenter des modifications post-traductionnelles (MPT), notamment *via* leur phosphorylation par les kinases cellulaires. Ces modifications sont censées contrôler la transition de la transcription vers la réplication, par perturbation des interactions entre N, P, L et l'ARN viral.

9.1.3. Les protéines intrinsèquement désordonnées

Les protéines intrinsèquement désordonnées (PIDs) n'ont pas de structure tridimensionnelle stable, dans des conditions physiologiques et en absence de partenaires potentiels. Elles peuvent adopter un ensemble de conformations. La composition en acides aminés détermine leur incapacité à se replier. Toutefois elles modulent les interactions avec les autres molécules. Les PIDs jouent un rôle crucial dans les processus cellulaires de régulation et de transmission des signaux et sont liées au développement de maladies neurodégénératives, de plusieurs cancers et du diabète de type II. L'absence de structure stable peut être considérée comme un avantage. La capacité de s'associer avec plusieurs partenaires ouvre l'opportunité de multifonctionnalité permettant d'avoir plusieurs sites d'interaction le long de la chaîne polypeptidique.

Il est également courant que les PIDs contiennent des sites de reconnaissance moléculaire et subissent une transition désordre-ordre lors de l'interaction. Les PIDs forment souvent des « complexes flous » en restant désordonnées dans l'état lié. De plus, les protéines dépliées peuvent s'auto-associer, parfois en présence d'acides nucléiques, en formant des gouttelettes de protéines. Ces propriétés sont largement utilisées dans les machineries-organelles de transcription-réplication virales (les corps de Negri). Les PIDs sont accessibles aux enzymes de MPT pouvant induire des changements structuraux des protéines, leurs stabilisation et déstabilisation.

9.1.4. Objectifs de cette thèse

Certainement, il existe une corrélation entre le désordre conformationnel des protéines des *Paramyxoviridae* et leur pathogénicité extrême. L'évolution a sélectionné des protéines multifonctionnelles et adaptables, indispensables à l'amplification efficace du matériel génétique viral. Malgré la complexité structurale et la caractérisation difficile, les PIDs virales constituent un bon système pour l'étude des mécanismes fonctionnels des protéines désordonnées.

Au cours de ce travail, j'ai réalisé les premières études sur les régions désordonnées et structurées des P et N du virus des oreillons et je discuterai leur pertinence fonctionnelle au cours du cycle de vie du virus. J'ai tenté de répondre aux questions suivantes qui peuvent être définies comme objectifs de ce projet :

- Quel est le rôle du désordre conformationnel des N et P ? En particulier, quel est le rôle du Ntail ? et avec quels partenaires potentiels interagit-il ? Ce domaine interagit-il avec P ? Comme c'est le cas pour les autres *Paramyxoviridae*, et si oui, comment ?
- Quelles sont les propriétés dynamiques du P_{NTD} ? Quel est le rôle fonctionnel de ce domaine et comment interagit-il avec d'autres protéines virales au cours de la réplication ?
- Comment la phosphorylation de N et P influence-t-elle la structure de la protéine, sa dynamique, les interactions entre domaines, ainsi que les interactions entre N et P ?
- Quel est le rôle du linker désordonné entre OD et XD de la phosphoprotéine ? L'interaction entre P_{XD} et Ncore peut-elle être décrite à la résolution atomique ?
- Est-il possible d'étudier les NCs du virus des oreillons avec de l'ARN issu de l'expression de cellules bactériennes ? Est-il possible de développer des protocoles de purification des RLPs ?
- Le peptide issu de P_{NTD} peut-il stabiliser la N monomerique dans le complexe N^oP ? Si oui, quelles sont les conditions pour obtenir l'assemblage des NCs à partir de N^oP *in vitro* ? Peut-on caractériser la structure N^oP du MuV ?

9.2. Matériels et méthodes

9.2.1. La résonance magnétique nucléaire

La résonance magnétique nucléaire (RMN) est une des techniques les plus utilisées dans ce projet. La RMN ne permet pas que d'avoir la résolution atomique de la structure mais elle est aussi cruciale pour les études de la dynamique moléculaire et la flexibilité en échelle du temps de la picoseconde (ps) à la seconde (s). Ce manuscrit est plutôt focalisé sur les PIDs, l'utilisation de la RMN qui offre une résolution à l'échelle par résidu, est bien adapté pour décrire le comportement conformationnel des ces protéines et tenter de comprendre leur fonction.

L'expérience de RMN la plus simple consiste à placer un spin nucléaire actif de RMN dans un champ magnétique statique, générant ainsi une polarisation de spin, ainsi que sa perturbation ultérieure par l'application d'impulsions électromagnétiques à radiofréquence. Parmi les paramètres de RMN les plus importants, il est possible de distinguer :

- le déplacement chimique (DC) apparaît sous l'action des champs magnétiques locaux (induits par le courant des moments électroniques) sur le noyau atomique. Donc chaque noyau actif en RMN (^{1}H, ^{15}N, ^{13}C) peut avoir son DC unique s'il a un environnement chimique, distinct des autres noyaux avec les champs locaux. En comparant les DCs mesurés pour les différentes structures protéiques (hélice α et feuillet β) avec ceux d'une pelote aléatoire, on peut extraire des déplacements chimiques secondaires décrivant la probabilité de la présence de la structure secondaire ;
- les couplages scalaires (pour liquides) et dipolaires (milieu anisotropique et solide) entre les noyaux par les liaisons chimiques et l'espace respectivement ;
- la vitesse de relaxation, ou le retour à l'équilibre de magnétisation du noyau après la perturbation : T_1 (temps de relaxation longitudinal) et T_2 (temps de relaxation transversale).

Les expériences de RMN utilisées dans ce travail sont bi- et tridimensionnelles : la HSQC (corrélation entre ^{1}H et ^{15}N amide de chaque acide aminé) permet d'avoir « le plan de la molécule » ; les mesures de la dynamique moléculaires R_1, $R_{1\rho}$ et ENO (NOE) hétéronucléaire ; spectres de l'augmentation de la relaxation paramagnétique (PRE) avec l'attachement d'une sonde paramagnétique ; mesures de couplages dipolaires résiduels (RDC) en utilisant des milieux d'alignement, spectres 3D nécessaires pour l'attribution de la chaine principale protéique, y compris l'analyse de propension de structure secondaire (SSP) calculée à partir des DCs.

9.2.2. Autres techniques

En raison de limitations de la RMN comme sa sensibilité faible et la taille moléculaire des objets biologiques étudiés, j'ai utilisé aussi d'autres techniques qui peuvent fournir des informations sur la taille exacte des molécules, leur forme, leur état oligomérique comme la diffusion des rayons X aux petits angles (SAXS), la diffusion de la lumière laser multi-angle (MALLS), la spectroscopie électronique (EM). De plus, les méthodes biochimiques (définition des constructions des protéines, génération des clones, expression et purification des protéines recombinantes) ont également été décrites dans ce chapitre.

9.3. Résultats et discussion

Les domaines structuraux Ncore et Ntail ont été étudiées séparément, mais également en présence de la séquence non-tronqué N_{FL} (549 acides aminés). Malheureusement, je n'ai pas réussi à obtenir les constructions du Ncore tronquées stables. Certaines n'ont pu être obtenues dans la fraction soluble, d'autres étaient mal repliées ou agrégées au cours de la purification. Donc dans ce travail, je me suis focalisée sur la structure et la fonction du Ntail et du N_{FL}.

Le Ntail (résidus 373-549) a été étudié par la RMN. Le spectre 2D ^{1}H-^{15}N HSQC est caractéristique d'une protéine intrinsèquement désordonnée. Après avoir attribué les signaux pour chaque acide aminé, j'ai effectué les analyses SSP et ASTEROIDS avec le calcul du taux de population en hélice-α (logiciel permettant de créer l'ensemble des structures décrivant le comportement dynamique de la protéine). Ces expériences ont montré la présence de structures secondaires au niveau de la partie N-terminale (373-405), qui à la suite de comparaison de la séquence N de MuV avec celle de PIV5 (membre du même genre *Rubulavirus* et dont la structure de N oligomerique a été résolue), s'est révélée faire partie du Ncore plus particulièrement du C-arm. Après mesure de relaxation, en particulier $R_{1\rho}$, j'ai découvert des taux de relaxation élevés en plus du C-arm mentionné précédemment dans les régions 474-500 et 505-518 qui ont été appelées comme le More et post-MoRE. Les RDC constituent la méthode la plus sensible pour détecter la propension α-hélicale du MoRE en étant aligné avec le milieu des cristaux liquides. Plusieurs tests d'affinité n'ont pas montré des interactions avec P_{NTD} et P_{CTD}. La phosphorylation *in vitro* a été effectuée avec l'aide de la kinase CKII qui modifie le résidu S439 sans impact sur la dynamique totale de la protéine. Le rôle de cette modification reste donc flou.

L'expression de la N non-tronquée conduit à la formation de RLPs avec l'ARN bactérien. Bien évidement, ces structures ne sont pas naturelles pour le virus mais peuvent constituer un bon modèle pour étudier la nucléoprotéine oligomerique. Grâce à sa taille (diamètre environ 20 nm) les anneaux sont facilement visualisés par l'EM. La trypsination (la protéolyse entrainant la coupure de Ntail) conduit à la formation de NCs. En raison de la morphologie variée de ces nucléocapsides, la structure cryo-EM à haute résolution n'a pu être obtenue. Plus d'informations extraites des spectres RMN, y compris la relaxation, qui montrent que les régions proches du C-arm et le MoRE ne sont pas detectables. Ce phénomène peut être expliqué par la proximité de ces domaines aux parties structurales plus grosses et donc plus « lentes » en fonction du temps de corrélation pour être visible par RMN. La validation de cette hypothèse a été effectuée par le traitement de RLPs avec un excès de Ntail conduisant à l'apparition de domaines non visibles précédemment. Le rajout en excès de N-anneaux au Ntail a eu l'effet inverse, à savoir la disparition du MoRE à la suite de l'interaction de Ntail avec la surface de N_{FL}. Par conséquent, cette expérience a montré que le MoRE s'associe avec le Ncore mais la fonction de cette interaction demeure inconnue.

Après l'échec de la purification la phosphoprotéine non-tronquée P_{FL}, et aussi des P_{NTD+OD} et P_{OD+CTD}, il a été décidé d'étudier les deux domaines P_{NTD} et P_{CTD} séparément. De la même manière que le Ntail, les spectres HSQC de P_{NTD} ont un profil caractéristique de PID. Les analyses des DCs et les mesures de la dynamique ont mis en évidence la présence de trois régions avec une propension hélicoïdale, localisées entre les résidus 13-60 ($\alpha 1$), 117-141 ($\alpha 2$) et 189-204 ($\alpha 3$). Ces structures sont particulièrement affectées lors de l'interaction avec les RLPs. J'ai notamment observé des taux de relaxation élevés, ce qui est en accord avec les recherches précédentes. Plus d'informations ont été extraites en générant le complexe de fusion N^oP. J'ai trouvé que les régions structurales $\alpha 1$ et $\alpha 2$ interagissent avec la nucléoprotéine monomérique. Dans le cas de $\alpha 1$, j'ai observé deux hélices 13-32 et 35-60, par analogie avec le N^oP du PIV5, la première chaperonnant la N monomérique. L'hélice $\alpha 2$ possède un motif conservé chez les *Paramyxoviridae*, de la séquence VELA chez MuV. D'ailleurs, la phosphoprotéine du MeV a aussi cette région nommée HELL qui s'associe faiblement avec le N-lobe du Ncore. Il est possible de localiser ce site sur la structure cristalline du PIV5, très proche en composition (50,6% d'identité de séquence).

Le domaine XD de P_{CTD} a été également étudié en RMN. Des signaux larges, résultant probablement de l'instabilité du faisceau d'hélices-α, ont rendu l'analyse du spectre impossible. Par contre, l'expression de P_{CTD} comportant les acides aminés entre OD et XD, a été plus efficace en terme de qualité du spectre HSQC. Les expériences de relaxation et RDCs avec l'alignement des bacteriophages et des analyses des SSPs, ASTEROIDS avec l'estimation de population des structures secondaires, ont confirmé la présence de trois hélices-α dans XD ainsi que deux structures transitoires du linker désordonné entre les OD et XD (résidus 286-301 $\alpha 4$ et 316-323 $\alpha 5$) qui présentent des contacts à grande distance avec le XD. De plus, j'ai effectué les titrations RMN pour étudier l'interaction entre P_{CTD} et N_{FL}. Sous l'effet de l'addition d'un excès de RLPs, j'ai observé la disparition graduelle des signaux correspondant aux acides aminés du XD. Ensuite j'ai effectué la phosphorylation avec la Polo-Kinase I, kinase spécifique pour le PIV5. L'enzyme a modifié deux résidus serines (292 et 307) situés dans la partie dépliée de P_{CTD}. La phosphorylation de ces deux régions entraîne une faible diminution de la propension structurale de deux régions structurées de manière transitoire mentionnées précédemment. Le rôle de ce changement n'est pas compris à ce jour.

Le défi le plus important de ce projet était de générer un complexe $N^{o}P$ stable. Plusieurs constructions P_{NTD}-TEV-N_{FL} ont donné des N en anneaux, suite à la libération du peptide P. En faisant plusieurs spectres RMN, j'ai trouvé que la présence de ce peptide libre peut stabiliser le complexe contre l'agrégation. Donc le rajout en excès de P_{NTD} dans des préparations de $N^{o}P$ fraichement purifiées peut augmenter la chance d'obtenir le complexe monomerique. L'autre axe de recherche sur le complexe $N^{o}P$ était l'étude de l'assemblage en NCs en présence d'un ARN court de 6 nucléotides. J'ai observé la formation de RLPs et de capsides mais leurs cinétiques d'apparition restent malcomprises. J'ai mis aussi en évidence une influence de la concentration initiale du $N^{o}P$ sur la formation des NCs.

9.4. Conclusions et perspectives

Dans ce projet, certains questions et problématiques définies au début ont été validées, d'autres sont restées sans réponse :

- Le Ntail possède des propriétés de protéine désordonnée avec formation de structures secondaires transitoires : les C-arm, MoRE et post-MoRE. Par contre il serait intéressant maintenant de comprendre la nature dynamique des hélices-α.
- La nouvelle fonction du MoRE est découvert : l'interaction avec le Ncore.
- Les structures de N-anneaux et NCs à haute résolution n'ont pas été menées car le protocole de leur génération nécessite une optimisation pour générer les structures homogènes.
- Comme le Ntail, le P_{NTD} est aussi une région dépliée mais elle nécessite une étude de la dynamique plus profonde, par exemple, par mesure des RDCs avec alignement en milieu anisotrope.
- Trois régions de P_{NTD} présentant des propensions hélice-α ont été trouvées. La fonction pour la première a été dévoilé : chaperonner N en $N^{o}P$. Par conséquent, il est nécessaire de trouver le(s) rôle(s) des deux autres.
- J'ai identifié des structures secondaires transitoires le long de la séquence non-structurée de P_{CTD}. Il serait donc important d'étudier leur(s) fonction(s).
- L'interaction entre P_{CTD} et N_{FL} n'a pas été étudiée suffisamment à cause de la forte dégradation des anneaux de N. Les protocoles biochimiques favorisant la stabilité des RLPs doivent être améliorés.
- Le protocole de stabilisation de $N^{o}P$ a été mis au point. Il reste cependant difficile d'étudier les affinités de $N^{o}P$ en présence d'un excès de P_{NTD}.
- L'assemblage de NCs a été réussi mais il faut étudier les mécanismes de formation des anneaux et capsides, mais aussi les facteurs influençant leur apparition.

Le virus des oreillons est peu étudié mais son appartenance à la famille des *Paramyxoviridae* permet de comprendre les mécanismes des interactions des protéines impliquées dans le processus de transcriptionet réplication viral. Dans le même temps, la faible conservation de séquence dans les régions désordonnées de ces protéines justifie d'étudier de manière plus approfondie d'autres membres du genre *Rubulavirus*, tels que PIV5, PIV2 et HPIV4.

BIBLIOGRAPHY

Abyzov A, Salvi N, Schneider R, Maurin D, Ruigrok RW, Jensen MR, Blackledge M. Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation. *J Am Chem Soc*, **2016**, 138(19): 6240-51.

Acheson NH. Fundamentals of molecular virology, 2nd edition, *Wiley*, **2011**.

Afonso CL, Amarasinghe GK, Banyai K, Bao Y, Basler CF, Bavari S, Bejerman N et al. Taxonomy of the order Mononegavirales: update 2016. *Arch. Virol*, **2016**, 161: 2351–2360.

Aggarwal M, Leser GP, Kors CA, Lamb RA. Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein in complex with an amino-terminal peptide of the phosphoprotein. *J Virol*, **2018**, 92.

Aguilar HC, Lee B. Emerging paramyxoviruses: molecular mechanisms and antiviral strategies. *Expert Rev Mol Med*, **2011**, 13: 6.

Alayyoubi M, Leser GP, Kors CA, Lamb RA. Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein – RNA complex. *Proc Natl Acad Sci USA*, **2015**, 112(14): 1792-9.

Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, Schoehn G, Weissenhorn W, Ruigrok RW. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science*, **2006**, 313(5785): 360–363.

Aljabr W, Touzelet O, Pollakis G, Wu W, Munday DC, Hughes M, Hertz-Fowler C, Kenny J, Fearns R, Barr JN, Matthews DA, Hiscox JA. Investigating the influence of ribavirin on human respiratory syncytial virus RNA synthesis using a high-resolution RNAseq approach. *J Virol*, **2016**, 90(10): 4876-4888.

Andrejeva J, Childs KS, Young DF, Carlos TS, Stock N, Goodbourn S, Randall RE. 2004. The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-β promoter. *Proc Natl Acad Sci USA*, **2004**, 101: 17264 –17269.

Atrasheuskaya AV, Blatun EM, Kulak MV, Atrasheuskaya A, Karpov IA, Rubin S, et al. Investigation of mumps vaccine failures in Minsk, Belarus, 2001–2003. *Vaccine*, **2007**, 25(24): 4651–8.

Atrasheuskaya AV, Neverov AA, Rubin S, Ignatyev GM. Horizontal transmission of the Leningrad-3 live attenuated mumps vaccine virus. *Vaccine*, **2006**, 24(10): 1530-6.

Battisti AJ, Meng G, Winkler DC, McGinnes LW, Plevka P, Steven AC, Morrison TG, Rossmann MG. Structure and assembly of a paramyxovirus matrix protein. *Proc Natl Acad Sci USA*, **2012**, 109: 13996-14000.

Beck M, Welsz-Malecek R, Mesko-Prejac M, Radman V, Juzbasic M, Rajninger-Miholic M, Prislin-Musklic M, Dobrovsak-Sourek V, Smerdel S, Stainer DW. Mumps vaccine L-Zagreb, prepared in chick fibroblasts: I. production and field trials. *Journal of biological standardization*, **1989**, 17: 85–90.

Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok R, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *PNAS*, **2005**, 102 (47): 17002-17007.

Blackledge MJ, Brüschweiler R, Griesinger C, Schmidt JM, Xu P, Ernst RR. Conformational backbone dynamics of the cyclic decapeptide antamanide. Application of a new multiconformational search algorithm based on NMR data. *Biochemistry*, **1993**, 32 (41): 10960-10974.

Blanchard L, Tarbouriech N, Blackledge M, Timmins P, Burmeister W, Ruigrok R, Marion D. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*, **2004**, 319: 201–11.

Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, **2004**, 4(6): 1633-49.

Boxall N, Kubinyiova M, Prikazsky V, Benes C, Castkova J. An increase in the number of mumps cases in the Czech Republic, 2005–2006. *Euro Surveill*, **2008**, 13(16).

Braeye T, Linina I, De Roy R, Hutse V, Wauters M, Cox P, et al. Mumps increase in Flanders, Belgium, 2012–2013: results from temporary mandatory notification and a cohort study among university students. *Vaccine*, **2014**, 32 (35): 4393–8.

Brockhoff HJ, Mollema L, Sonder GJ, Postema CA, van Binnendijk RS, Kohl RH, de Melker HE, Hahné SJ. Mumps outbreak in a highly vaccinated student population, The Netherlands, 2004. *Vaccine*, **2010**, 28(17): 2932–6.

Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*, **2013**, 41(1): 340-348.

Calain P, Roux L. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *J Virol*, **1993**, 67: 4822–4830.

Cardemil CV, Dahl RM, James L, Wannemuehler K, Gary HE, Shah M, Marin M, Riley J, Feikin DR, Patel M, Quinlisk P. Effectiveness of a third dose of MMR vaccine for mumps outbreak control. *New Engl J Med*, **2017**, 377(10): 947–56.

Carr HY, Purcell EM. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev*, **1954**, 94, 630.

Cattaneo R, Kaelin K, Baczko K, Billeter MA. Measles virus editing provides an additional cysteine-rich protein. *Cell*, **1989**, 56(5): 759-64.

Cavanagh J, Skelton NJ, Fairbrother WJ, Rance M, Palmer AG et al. Protein NMR spectroscopy. Principles and Practice. 2nd edition, *Elsevier*, **2006**.

CDCP Centers for Disease Control and Prevention. Mumps cases and outbreaks; 2017. <http://www.cdc.gov/mumps/outbreaks.html> [Date accessed: May 7, 2018].

Centers for Disease Control and Prevention. Measles, mumps, and rubella—Vaccine use and strategies for the elimination of measles, rubella, and congenital rubella syndrome and control of mumps: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep*, **1998**, 47: 1–48.

Challa S, Scott AD, Yuzhakov O, Zhou Y, Tiong-Yip CL, Gao N, Thresher J, Yu Q. Mechanism of action for respiratory syncytial virus inhibitor RSV604. *Antimicrob Agents Chemother*, **2015**, 59: 1080–1087.

Chan YP, Koh CL, Lam SK, Wang LF. Mapping of domains responsible for nucleocapsid protein–phosphoprotein interaction of Henipaviruses, *J Gen Virol*, **2004**, 85: 1675–1684.

Chang LV. Information, education and health behaviours: evidence from MMR vaccine autism controversy. *Health Econ*, **2018**, 27(7): 1043-1062.

Chary KVR, Govil G. NMR in Biological Systems, *Springer*, **2008**.

Clore GM, Driscoll PC, Wingfield PT, Gronenborn AM. Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear 15N-1H NMR spectroscopy. *Biochemistry*, **1990**, 29(32): 7387-401.

Clore GM, Starich MR, Gronenborn AM. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J Am Chem Soc*, **1998**, 120: 10571-10572.

Cohen C, White JM, Savage EJ, Glynn JR, Choi Y, Andrews N, Brown D, Ramsay ME. Vaccine effectiveness estimates, 2004–2005 mumps outbreak, England. *Emerg Infect Dis*, **2007**, 13(1): 12–7.

Communie G, Habchi J, Yabukarski F, Blocquel D, Schneider R, Tarbouriech N, Papageorgiou N, Ruigrok RWH, Jamin M, Jensen MR, Longhi S, Blackledge M. Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. *PLoS Pathog*, **2013**, 9(9): 1003631.

Cox R, Green TJ, Purushotham S, Deivanayagam C, Bedwell GJ, Prevelige PE, Luo M. Structural and functional characterization of the mumps virus phosphoprotein. *Journal of Virology*, **2013**, 87(13): 7558–68.

Cox R, Green TJ, Qiu S, Kang J, Tsao J, Prevelige PE, He B, Luo, M. Characterization of a mumps virus nucleocapsidlike particle. *Journal of Virology*, **2009**, 83(21): 11402–11406.

Cox R, Pickar A, Qiu S, Tsao J, Rodenburg C, Dokland T, Elson A, He B, Luo M. Structural studies on the authentic mumps virus nucleocapsid showing uncoiling by the phosphoprotein. *Proceedings of the National Academy of Sciences*, **2014**, 111(42): 15208–15213.

Cox RM, Plemper RK. Structure and organization of paramyxovirus particles. *Current Opinion in Virology*, **2017**, 24: 105–114.

Curran J, Boeck R, Kolakofsky D. The Sendai virus P gene expresses both an essential protein and an inhibitor of RNA synthesis by shuffling modules via mRNA editing. *EMBO J*, **1991**, 10(10): 3079-85.

da Cunha SS, Rodrigues LC, Barreto ML, Dourado I. Outbreak of aseptic meningitis and mumps after mass vaccination with MMR vaccine using the Leningrad-Zagreb mumps strain. *Vaccine*. **2002**, 20(7-8): 1106-12.

Das T, Schuster A, Schneider-Schaulies S, Banerjee AK. Involvement of cellular casein kinase II in the phosphorylation of measles virus P protein: identification of phosphorylation sites. *Virology*, **1995**, 211: 218–226.

Dayan GH, Quinlisk MP, Parker AA, Barskey AE, Harris ML, Schwartz JM, et al. Recent resurgence of mumps in the United States. *N Engl J Med*, **2008**, 358 (15): 1580–9.

De BP, Gupta S, Banerjee AK. Cellular protein kinase C isoform zeta regulates human parainfluenza virus type 3 replication. *Proc Natl Acad Sci USA*, **1995**, 92: 5204–5208.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, **1995**, 6(3): 277-93.

Desfosses A, Goret G, Farias Estrozi L, Ruigrok RW, Gutsche I. Nucleoprotein-RNA orientation in the measles virus nucleocapsid by three-dimensional electron microscopy. *J Virol*, **2011**, 85(3): 1391-5.

Dictionnaire électronique de l'Académie française.
<https://academie.atilf.fr/consulter/oreillon?page=1>

Dokland T. Back to the basics: The fundamentals of cryo-electron microscopy. *Microsc Microanal*, **2009**, 15(2).

Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **2005**, 21: 3433–3434.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. *J Mol Graph Model*, **2001,** 19(1): 26-59.

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, **2005**, 6(3): 197–208.

Egelman EH, Wu SS, Amrein M, Portner A, Murti G. The Sendai virus nucleocapsid exists in at least four different helical states. *J Virol*, **1989**, 63(5): 2233–2243.

Egerton R.F. Physical Principles of Electron Microscopy. *Springer*, **2005**.

Fearns R & Plemper RK. Polymerases of paramyxoviruses and pneumoviruses. *Virus Research*, **2017**, 234: 87–102.

Felli IC, Pieratelli R. Advances in Experimental Medicine and Biology. Intrinsically disordered proteins Studies by NMR. *Springer*, **2015**.

Fescharek R, Quast U, Maass G, Merkle W, Schwarz S. Measles–mumps vaccination in the FRG: an empirical analysis after 14 years of use. II. Tolerability and analysis of spontaneously reported side-effects. *Vaccine*, **1990**, 8: 446–456.

Finch JT, Gibbs AJ. Observations on the structure of the nucleocapsids of some Paramyxoviruses. *J Gen Virol*, **1970**, 6: 141-150.

Fuentes SM, Sun D, Schmitt AP, He B. Phosphorylation of paramyxovirus phosphoprotein and its role in viral gene expression. *Future Microbiology*, **2010**, 5(1): 9–13.

Gouma S, Cremer J, Parkkali S, Veldhuijzen I, van Binnendijk RS, Koopmans MP. Mumps virus F gene and HN gene sequencing as a molecular tool to study mumps virus transmission. *Infect Genet Evol,* **2016**, 45: 145–50.

Green TJ, Luo M. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc Natl Acad Sci USA*, **2009**, 106(28): 11713–11718.

Green TJ, Rowse M, Tsao J, Kang J, Ge P, Zhou ZH, Luo M. Access to RNA encapsidated in the nucleocapsid of vesicular stomatitis virus. *J Virol*, **2011**, 85(6): 2714-22.

Green TJ, Zhang X, Wertz GW, Luo M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science, **2006**, 313(5785): 357–360.

Guenneugues M, Berthault P, Desvaux H, Goldman M. Direct determination of the heteronuclear T_1/T_2 ratio by off-resonance steady-state magnetization measurement: Investigation of the possible application to fast exchange characterization of ^{15}N-labeled proteins. *Journal of Biomolecular NMR*, **1999**, 15(4): 295–307.

Guryanov SG, Liljeroos L, Kasaragod P, Kajander T, Butcher SJ. Crystal structure of the measles virus nucleoprotein core in complex with an N-terminal region of phosphoprotein. *J Virol*, **2015**, 90: 2849-2857.

Gutsche I, Desfosses A, Effantin G, Ling WL, Haupt M, Ruigrok RW, Sachse C, Schoehn G. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science*, **2015**, 348: 704-707.

Habchi J, Blangy S, Mamelli L, Jensen MR, Blackledge M, Darbon H, Oglesbee M, Shu Y, Longhi S. (2011) Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipaviruses. *Journal of Biological Chemistry*, **2011**, 286: 13583–13602.

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*, **2008**, 8: 1871–1879.

Hansen DF, Vallurupalli P, Kay LE. An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J Phys Chem B*, **2008**, 112 (19): 5898–5904.

Harrison MS, Sakaguchi T, Schmitt AP: Paramyxovirus assembly and budding: building particles that transmit infections. *Int J Biochem, Cell Biol*, **2010**, 42: 1416-1429.

Heggeness MH, Scheid A, Choppin PW. Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc Natl Acad Sci USA*, **1980**, 77(5): 2631–2635.

Horikami SM, Smallwood S, MOYER SA. The Sendai virus V protein interacts with the NP protein to regulate viral genome RNA replication. *Virology*, **1996**, 222: 383–390.

Houben K, Marion D, Tarbouriech N, Ruigrok, RW, Blanchard L. Interaction of the C-terminal domains of Sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J Virol*, **2007**, 81: 6807–6816.

Howton TC, Sun Y, Zhan AY, Mukhtar MS. Intrinsically disordered proteins: controlled chaos or random walk. *International Journal of Plant Biology*, **2016**, 6(1): 52-57.

Hrynash Y, Nadraga A, Dasho M. Effectiveness of a vaccination program against mumps in Ukraine. *European Journal of Clinical Microbiology & Infectious Diseases*, **2008**, 27(12): 1171–1176.

Huang M, Sato H, Hagiwara K, Watanabe A, Sugai A, Ikeda F, Kozuka-Hata H, Oyama M, Yoneda M, Kai C. Determination of a phosphorylation site in Nipah virus nucleoprotein and its involvement in virus transcription. *J Gen Virol*, **2011**, 92: 2133–2141.

Huntley CC, De BP, Banerjee AK. 1997. Phosphorylation of Sendai virus phosphoprotein by cellular protein kinase Cζ. *J Biol Chem*, **1997**, 272: 16578–16584.

ICTV (International Committee on Taxonomy of Viruses).
<https://talk.ictvonline.org/taxonomy/> [According to 2017].

Institute of Medicine (IOM). Immunization safety review: Vaccines and autism. *National Academies Press*, **2004**.

Ishii Y, Markus MA, Tycko R. Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. *J Biomol NMR*, **2001**, 21: 141–151.

Jensen MR, Communie G, Euripedes Ribeiro Jr. AR, Martinez N, Desfosses A, Salmon L, Mollica L, Gabel F, Jamin M, Longhi S, Ruigrok R, Blackledge M. Intrinsic disorder in measles virus nucleocapsids. *PNAS*, **2011**, 108(24): 9839-9844.

Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RW, 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*, **2008**, 130(25): 8055-61.

Johansson K, Bourhis J, Campanacci V, Cambillau C, Canard B, Longhi S. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem*, **2003**, 278: 44567–73.

Jung YS, Zweckstetter M. Mars - robust automatic backbone assignment of proteins. *Journal of Biomolecular NMR*, **2004**, 30(1): 11–23.

Kho CL, Tan WS, Tey BT, Yusoff K. Regions on nucleocapsid protein of Newcastle disease virus that interact with its phosphoprotein. *Arch Virol*, **2004**, 149: 997–1005.

Kingston RL, Baase WA, Gay LS. Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins. Journal of Virology, **2004**, 78(16): 8630–8640.

Kingston RL, Gay LS, Baase WS, Matthews BW. Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. *Journal of Molecular Biology*, **2008**, 379(4), 719–731.

Kirchdoerfer RN, Abelson DM, Li S, Wood MR, Saphire EO. Assembly of the Ebola Virus Nucleoprotein from a Chaperoned VP35 Complex. *Cell Reports*, **2015**, 12(1): 140–149.

Knipe DM, Howley P. Fields Virology, sixth edition, *LWW*, **2013**.

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. *J Virol*, **2009**, 83(16): 7948-58.

Lescop E, Schanda P, Brutscher B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson*, **2007**, 187(1): 163-9.

Leyrat C, Yabukarski F, Tarbouriech N, Ribeiro EA, Jensen MR, Blackledge M, Ruigrok RW, Jamin M. Structure of the vesicular stomatitis virus N^{o}-P Complex. *Plos Pathog*, **2011**, 7:

1002248-1002248.

Li M, Schmitt PT, Li Z, McCrory TS, He B, Schmitt AP. Mumps virus matrix, fusion, and nucleocapsid proteins cooperate for efficient production of virus-like particles. *J Virol*, **2009**, 83: 7261-7272.

Li T, Chen X, Garbutt KC, Zhou P, Zheng N. Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. *Cell*, **2006**, 124(1): 105-117.

Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, Grigorieff N, Harrison SC, Whelan SP. Structure of the L protein of vesicular stomatitis virus from electron cryomicroscopy. *Cell*, **2015**, 162, 314–327.

Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics, *Structure*, **2003**, 11: 1453–1459.

Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. *J Am Chem Soc*, **1982**, 104 (17): 4546–4559.

Lo MK, Harcourt BH, Mungall BA, Tamin A, Peeples ME, Bellini WJ, Rota PA. Determination of the henipavirus phosphoprotein gene mRNA editing frequencies and detection of the C, V and W proteins of Nipah virus in virus-infected cells. *J Gen Virol*, **2009**, 90(2): 398-404.

Longhi S, Receveur-Bréchot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S, Canard B. 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*, **2003**, 278: 18638–18648.

Loria JP, Rance M, and Palmer AG. A TROSY CPMG sequence for characterizing chemical exchange in large proteins. *J Biomol NMR*, **1999**, 15(2): 151–155.

Maillet M, Bouvat E, Robert N, Baccard-Longère M, Morel-Baccard C, Morand P, Vabret A, Stahl, J. Mumps outbreak and laboratory diagnosis. *Journal of Clinical Virology*, **2015**, 62(February 2013), 14–19.

Maladies à declaration obligatoire en France <http://invs.santepubliquefrance.fr/Espace-professionnels/Maladies-a-declaration-obligatoire/Liste-des-maladies-a-declaration-obligatoire>

May M, Rieder CA, Rowe RJ. Emergent lineages of mumps virus suggest the need for a polyvalent vaccine. *Int J Infect Dis*, **2018**, 66: 1–4.
Medynicin NV. Vaccines used for prophylactic of epidemiological mumps in Russia. *Bull "Vaccination"*, **2003**, 1–2.

Mesleh MF, Veglia G, DeSilva TM, Marassi FM, Opella SJ. Dipolar waves as NMR maps of protein structure. *J Am Chem Soc*, **2002**, 124(16): 4206–4207.

Milles S, Jensen MR, Communie G, Maurin D, Schoehn G, Ruigrok RW, Blackledge M. Communications self-assembly of measles virus nucleocapsid-like particles: kinetics and RNA sequence dependence, *Angew Chem Int Ed Engl*, **2016**, 9356–9360.

Motz C, Schuhmann KM, Kirchhofer A, Moldt M, Witte G, Conzelmann KK, Hopfner KP. Paramyxovirus V proteins disrupt the fold of the RNA sensor MDA5 to inhibit antiviral signaling. *Science*, **2013**, 339(6120): 690-3.

Mountcastle WE, Compans RW, Lackland H, Choppin PW. Proteolytic cleavage of subunits of the nucleocapsid of the paramyxovirus simian virus 5. *J Virol*, **1974**, 14(5): 1253–1261.

Mumps epidemic – United Kingdom, 2004–2005. *MMWR*, **2006**, 55(7): 173–5.

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*, **2017**, 8(1): 58.

NNDSS National Notifiable Diseases Surveillance System, annual tables for notifiable infectious deseases and condition data. <https://wwwn.cdc.gov/nndss/data-and-statistics.html> [Date accessed: June, 2018].

Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J Am Chem Soc*, **2009**, 131(49): 17908-18.

Noton SL, Fearns R. Initiation and regulation of paramyxovrus transcription and replication. *Virology*, **2015**, 479-480: 545-54.

Ohki S, Kainosho M. Stable isotope labeling methods for protein NMR spectroscopy. *Progress in Nuclear Magnetic Resonance*, 2008, 53: 208–226.

Online Etymology Dictionary.<https://www.etymonline.com/word/mumps>

Ortín J, Martín-Benito J. The RNA synthesis machinery of negative-stranded RNA viruses. *Virology*, **2015**, 479–480, 532–544.

Ozenne V, Bauer F, Salmon L, Huang, Jensen MR, Segard S, Bernadó P, Charavay C, Blackledge M. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, **2012**, 28(11): 1463–1470.

Palmer AG, Kroenke CD, and Loria JP. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol*, **2001**, 339: 204–238.

Paterson RG, Lamb RA. RNA editing by G-nucleotide insertion in mumps virus P-gene mRNA transcripts. *J Virol*, **1990**, 64(9): 4137-45.

Peng R, Zhu T, Olawale B, Abednego O, Musyoki M, Cui Y, Shi Y, Wang P, Fu Gao G. In vitro assembly of Ebola virus nucleocapsid-like complex expressed in E. coli. *Protein & Cell*, **2016**, 7(12): 888–898.

Pereira N, Cardone C, Lassoued S, Galloux M, Fix J, Assrir N, Lescop E, Bontems F, Eléouët JF, Sizun C. New Insights into Structural Disorder in Human Respiratory Syncytial Virus Phosphoprotein and Implications for Binding of Protein Partners, *J Biol Chem*, **2017**, 292(6): 2120–2131.

Pervushin KV, Wider G, Wüthrich K. Single Transition-to-single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. *J Biomol NMR*, **1998**, 12(2): 345-8.

Pickar A, Elson A, Yang Y, Xu P, Luo M, He B. Oligomerization of mumps virus phosphoprotein. *J Virol*, **2015**, 89(21), 11002–11010.

Pickar A, Zengel J, Xu P, Li Z, He B. Mumps virus nucleoprotein enhances phosphorylation of the phosphoprotein by polo-like kinase 1. *Journal of Virology*, **2015**, 90(3): 02160-15.

Piovesan D, Tabaro F, Mičetić I, Necci M, Quaglia F, Oldfield CJ, et al. DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res*, **2017**, 45: 219–227.

Prevention and Control of Mumps outbreaks in Canada.
<https://www.canada.ca/en/public-health/services/reports-publications/canada-communicable-disease-report-ccdr/monthly-issue/2010-36/guidelines-prevention-control-mumps-outbreaks-canada/immunization.html>

Principi N. & Esposito S. Mumps outbreaks: A problem in need of solutions. *Journal of Infection*, **2018**, 76(6): 503–506.

Prosser RS, Hwang JS, Vold RR. Magnetically aligned phospholipid bilayers with positive ordering: a new model membrane system. *Biophys J*, **1998**, 74: 2405–2418.
Public Health Agency of Canada. Supplement: guidelines for the prevention and control of mumps outbreaks in Canada. *Canada communicable disease report*, **2010**.
<http://www.phac-aspc.gc.ca/publicat/ccdr-rmtc/10pdf/36s1- eng.pdf> [Date accessed: May, 2018].

Ramanathan R, Voigt EA, Kennedy RB, & Poland GA. Knowledge gaps persist and hinder progress in eliminating mumps. *Vaccine*, **2018**, 36(26): 3721–3726.

Randall RE, Bermingham A. NP:P and NP:V interactions of the paramyxovirus simian virus 5 examined using a novel protein:protein capture assay. *Virology*, **1996**, 129: 121–129.

Renner M, Bertinelli M, Leyrat C, Paesen GC, Saraiva de Oliveira LF, Huiskonen JT, Grimes JM. Nucleocapsid assembly in pneumoviruses is regulated by conformational switching of the N protein. *Elife*, **2016**, 5: 12627-12627.

République française. Loi n° 2017-1836 du 30 décembre 2017 de financement de la sécurité sociale pour 2018, article 49. *Journal officiel*, 31 décembre **2017**.
<https://www.legifrance.gouv.fr/jo_pdf.do?id=JORFTEXT000036339090>

Rosas-Murrieta NH, Herrera-Camacho I, Palma-Ocampo H, Santos-Lopez G, Reyes-Leyva J. Interaction of mumps virus V protein variants with STAT1-STAT2 heterodimer: experimental and theoretical studies. *Virol J*, **2010**, 7: 263.

Rubin S, Eckhaus M, Rennick LJ, Bamford CG, Duprex WP. Molecular biology, pathogenesis and pathology of mumps virus. *J Pathol*, **2015**, 235(2): 242-52.

Rubin SA, Link MA, Sauder CJ, Zhang C, Ngo L, Rima BK, Paul Duprex W. Recent mumps outbreaks in vaccinated populations: no evidence of immune escape. *J Virol*, **2012**, 86(1): 615–20.

Rückert M, Otting G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc*, **2000**, 122 (32): 7793–7797.

Ruigrok RW, Crépin T, Kolakofsky D. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Current Opinion in Microbiology*, **2011**, 14(4): 504-510.

Rule GS, Hitchens TK. Fundamentals of Protein NMR Spectroscopy. *Springer*, **2006**.

Salladini E, Delauzun V, Longhi S. The Henipavirus V protein is a prevalently unfolded protein with a zinc-finger domain involved in binding to DDB1. *Mol. BioSyst*, **2017**, 13: 2254-2267.

Santak M, Lang-Balija M, Ivancic-Jelecki J, Kosutic-Gulija T, Ljubin-Sternak S, Forcic D. Antigenic differences between vaccine and circulating wild-type mumps viruses decreases neutralization capacity of vaccine-induced antibodies. *Epidemiol Infect*, **2013**, Jun; 141(6): 1298–309.

Sartorius B, Penttinen P, Nilsson J, Johansen K, Jonsson K, Arneborn M, Löfdahl M, Giesecke J. An outbreak of mumps in Sweden, February-April 2004. *Euro Surveill*, **2005**, 10 (9):191–3.

Sawada H, Yano S, Oka Y, et al. Transmission of Urabe mumps vaccine between siblings. *Lancet*, **1993**, 342-371.

Schanda P, Kupce E, Brutscher B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR*, **2005**, 33(4): 199-211.

Scheres SHW. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, **2012**, 180(3): 519-530.

Schoehn G, Mavrakis M, Albertini A, Wade R, Hoenger A, Ruigrok RW. The 12 Å structure of trypsin-treated measles virus N-RNA. *J Mol Biol*, **2004**, 339(2): 301–31.

Sentinelles network, INSERM/UPMC, <http://www.sentiweb.fr> [Date accessed: June, 2018].

Severin C, Terrell JR, Zengel JR, Cox R, Plemper RK, He B, Luo M. Releasing the genomic RNA sequestered in the mumps virus nucleocapsid. *Journal of Virology*, **2016**, 90: 10113–10119.

Singh R, John TJ, Cherian T, Raghupathy P. Immune response to measles, mumps & rubella vaccine at 9, 12 & 15 months of age. *Indian journal of medical research*, **1994**, 100: 155–159.

Smorodintsev AA, Nasibov MN, Jakovleva NV. Experience with live rubella virus vaccine combined with live vaccines against measles and mumps. *Bulletin of the World Health Organization*, **1970**, 42: 283–289.

Stratton K, Gable A, Shetty P, McCormick M. Immunization safety review: Measles-mumps-rubella vaccine and autism. *Institute of Medicine, National Academies Press*, **2001**.

Sun D, Luthra P, Li Z, He B. PLK1 down-regulates parainfluenza virus 5 gene expression. *PLoS Pathog*, **2009**, 5: 1000525.

Suprun U. To vanquish the Middle Age by vaccination. Article is written by Minister of Health of Ukraine, Uliana Suprun, *Pravda*, January 25, **2018**.
<https://life.pravda.com.ua/columns/2018/01/25/228642/>

T. D. Goddard, D. G. Kneller, SPARKY 3, University of California, San Francisco.

T. E. C. The first detailed description of mumps by Robert Hamilton, M.D. (1721-1793), published in 1790. *Pediatrics*, **1970**, 45 (2).

Tawar RG, Duquerroy S, Vonrhein C, Varela PF, Damier-Piolle L, Castagné N, MacLellan K, Bedouelle H, Bricogne G, Bhella D, Eléouët JF, Rey FA. Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science*, **2009**, 326(5957):

1279–1283.

Theillet FX, Smet-Nocca C, Liokatis S, Thongwichian R, Kosten J, Yoon MK, Kriwacki RW, Landrieu I., Lippens G, Selenko P. Cell signaling, post-translational protein modifications and NMR spectroscopy. *J Biomol NMR*, **2012**, 54(3): 217–236.

Tjandra N, Szabo A, Bax A. Protein backbone dynamics and N-15 chemical shift anisotropy from quantitative measurement of relaxation interference effects. *J Am Chem Soc*, **1996**, 118: 6986–6991.

Todorova R. Disorder structural predictions of the native EWS and its oncogenic fusion proteins in rapport with the function. *Advances in Bioscience and Biotechnology*, **2012**, 25–34.

Tompa P. Intrinsically disordered proteins. *Trends Biochem Sci*, **2002**, 27(10): 527-33.

Tsafou K, Tiwari PB, Metallo SJ, Toretsky JA. Targeting Intrinsically Disordered Transcription Factors: Changing the Paradigm. *Journal of Molecular Biology*, **2018**, 430(16): 2321–2341.

Tsoucalas G, Laios K, Karamanou M, Androutsos G. The Thasian epidemic of mumps during the 5th century BC. *Infez Med*, **2013**, 21(2): 149-50.

Ukraine: WHO and UNICEF estimates of immunization coverage: 2016 revision, **2016**. <https://data.unicef.org/wp-content/uploads/country_profiles/Ukraine/immunization_country_profiles/immunization_ukr.pdf> [Date accessed: June, 2018].

Unanov SS. Results of studying a live mumps vaccine from strain L-3 manufactured by the Moscow Research Institute of Viral Preparations: the epidemiological effectiveness of the vaccine]. *Voprosy Virusologii*, **1977**, 1: 59–61 (in Russian).

Uversky VN. A decade and a half of protein intrinsic disorder: Biology still waits for Physics. *Protein Sci*, **2013**, 22(6): 693–724.

Vallurupalli P, Bouvignies G, Kay LE. Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc*, **2012**, 134(19): 8148-61.

Vandermeulen C, Roelants M, Vermoere M, Roseeuw K, Goubau P, Hoppenbrouwers K. Outbreak of mumps in a vaccinated child population: a questionquestion of vaccine failure? Vaccin, **2004**, 22(21–22): 2713–6.

Verkhovna Rada of Ukraine. Law of Ukraine on protection of the population against infectious diseases from 06.04.2000, 1645-III. <http://zakon0.rada.gov.ua/laws/show/1645-14/conv> (in Ukrainian)

Vesikari T, André FE, Simoen E, Florent G, Ala-Laurila EL, Heikkinen A, Kuusinen H, Terho A. Evaluation in young children of the Urabe Am 9 strain of live attenuated mumps vaccine in comparison with the Jeryl Lynn strain. *Acta paediatrica scandinavica*, **1983**, 72: 37–40.

Vidal S, Curran J, Kolakofsky D. A stuttering model for paramyxovirus P mRNA editing. *EMBO J*, **1990**, 9(6): 2017-22.

Villanueva N, Navarro J, Méndez E, García-Albert I. Identification of a protein kinase involved in the phosphorylation of the C-terminal region of human respiratory syncytial virus P protein. *J Gen Virol*, **1994**, 75: 555–565.

ViralZone, **2009**. <https://viralzone.expasy.org/88?outline=complete_by_species>

Wakefield AJ, Murch SH, Anthony A, Linnell J, DM Casson DM, Malik M et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet*, **1998**, 351(9103): 637-641. [Retracted].

Wang S, Ma J, Xu J. AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. *Bioinformatics*, **2016**, 32: 672–679.

WHO/IVB database.
<http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tsincidenceMumps.html> [Date accessed: June, 2018].

Why Japan banned MMR vaccine. *Dly. Mail Online*. November 21, 2013.
<http://www.dailymail.co.uk/health/article-17509/Why-Japan-banned-MMR-vaccine.html>

Wyatt PJ. Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta*, **1993**, 272: 1–40.

Xue Y, Liu Z, Cao J, Ma Q, Gao X, Wang Q, Jin J, Zhou Y, Wen L, Ren J. GPS 2.1: enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. *Protein Engineering, Design and Selection*, **2011**, 24 (3): 255-260.

Yabukarksi F, Lawrence P, Tarbouriech N, Bourhis JM, Delaforge E, Jensen MR, Ruigrok RW, Blackledge M, Volchkov V, Jamin M. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nat Struct Mol Biol*, **2014**, 21: 754.

Yang J, Koprowski H, Dietzschold B, Fu ZF. Phosphorylation of rabies virus nucleoprotein regulates viral RNA transcription and replication by modulating leader RNA encapsidation. *J Virol*, **1999**, 73: 1661–1664.

Yang Y, Zengel J, Sun M, Sleeman K, Timani KA, Aligo J, Rota P, Wu J He, B. Regulation of viral RNA synthesis by the V Protein of parainfluenza virus 5. *J Virol*, **2015**, 89(23): 11845–11857.

Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, **2005**, 21: 3369–3376.

Zengel J, Pickar A, Pei X, Lin A, He B. The roles of phosphorylation of the nucleocapsid protein of mumps virus in regulating viral RNA transcription and replication. *Journal of Virology*, **2015**, 89(14): 7338–7347.

Communauté
UNIVERSITÉ Grenoble Alpes

Résumé

Les oreillons sont une maladie très contagieuse causée par le virus ourlien. La méthode préventive (le vaccin) contre ce virus a été déjà mise au point. Par contre, les épidémies récentes restent incontrôlables. Il est donc très important de comprendre le mécanisme moléculaire de son cycle de vie afin d'élaborer le traitement effectif et spécifique. Ce virus appartient à la famille des *Paramyxoviridae*. Son génome, l'ARN non segmenté monocaténaire de polarité négative, est protégé par la nucléoprotéine (N) en formant des structures filamenteuses nucléocapsides. N joue un rôle essentiel dans la synthèse du génome viral. En effet, cette protéine avec la polymérase et son cofacteur phosphoprotéine (P) constitue la machinerie de transcription-réplication du virus. La N et la P sont composées des régions pliées et dépliées. Malgré que la morphologie du virus ourlien est conservée parmi les autres membres de la famille, il existe quelques différences. Il a été démontré que la P est un oligomère antiparallel avec les deux extrémités d'un côté qui interagissent avec la partie structurale de N (Ncore). Tandis que la fonction de la région désordonnée (Ntail) est compliquée à identifier pour le moment. En comparant avec les autres paramyxovirus connus, Ntail n'interagit pas avec le domaine C-terminal de la P. Le rôle des régions déstructurées de P n'a pas été défini. Dans ce projet, nous dévoilons les mécanismes des interactions entre diverses régions de N et P et nous expliquons comment les domaines intrinsèquement désordonnés de N et P sont impliqués dans la régulation de la machine complexe de réplication virale. Nous avons utilisé la résonance magnétique nucléaire qui est la méthode la plus puissante afin de déterminer la structure, la dynamique et les partenaires d'interaction dont la fonction des protéines dépliées virales.

Abstract

Mumps is a highly contagious disease caused by the mumps virus. The prevention treatment (vaccine) against it is already in the routine use. However, recent outbreaks still remain uncontrollable. Therefore, it is important to understand the molecular mechanism of the mumps virus life cycle. This virus belongs to the family of *Paramyxoviridae*. Its genome, negative strand non-segmented RNA is protected by the nucleoprotein (N) by forming filamentous structures called nucleocapsids. N plays an important role in viral genome synthesis. Together with the polymerase and its cofactor phosphoprotein (P) they constitute the transcription-replication machinery. Both N and P contain folded and unfolded regions. Despite mumps virus common morphology with other paramyxovirus, there are some differences. It has been proposed that P is an antiparallel oligomer with two extremities on the one side being in interaction with the structural part of N (Ncore). The function of the disordered domain (Ntail) remains unclear, as it does not seem to bind to the C-terminal part of P, as is the case for other paramyxoviruses. The role of the disordered domains of P is also not known. In this project we revealed mechanisms of interaction between different regions of N and P and we explain how disordered regions of N and P are implicated in the regulation of the complex machinery of viral replication. We used the nuclear magnetic resonance which is the most powerful method to determine structure, dynamics and potential interaction partners, and therefore, function of disordered viral proteins.