Structural analysis of critical interactions in the replication machinery of rabies virus

By

Jingyu Zhan

0000-0001-8940-3663

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy January 2019

Department of Biochemistry and Molecular Biology

Bio21 Molecular Science and Biotechnology Institute

The University of Melbourne

Abstract

Rabies virus is a non-segmented negative sense RNA virus that causes encephalitis in humans with a 100% case fatality rate, resulting in > 61,000 deaths/year world-wide. There are currently no treatments for rabies disease, but a number of critical interactions of viral proteins provide potential targets to develop new antiviral compounds. Of particular interest are the interactions of viral nucleo (N), phospho (P), and large/polymerase (L) wherein the N protein encapsidates genomic RNA to form the helical nucleocapsid (N-RNA) that serves as the template for viral transcription and replication by the RNA-dependent polymerase complex, composed of the enzymatic L protein and non-catalytic polymerase cofactor P protein. P protein is critical in attaching L to the N-RNA template via an interaction between the P protein C-terminal domain (P_{CTD}) and the C-terminal trypsin-sensitive peptide of N protein (N-pep), which provides a potentially valuable target for anti-viral drug design. As a multifunctional protein, P also forms multiple interactions with host factors that underlie diverse roles in the virus-host interface, particular in immune evasion. Its C-terminal domain (P_{CTD}) alone bears binding sites for host factors including STAT1, microtubule and PML protein and two phosphorylation sites, apart from N-binding site. The strategies by which P mediates diverse functions in viral replication and immune evasion by coordinating its multiple interactions with viral protein and host cellular proteins remain unknown.

We have commenced a project to investigate the structure and dynamics of P_{CTD} and characterize its precise molecular interactions with N-pep. The P_{CTD} and N-pep have been expressed in *Escherichia coli*, generating protein at high yield and purity. NMR titrations of P_{CTD} with N-pep suggest a single binding mode with a micromolar affinity at the positive patch

of P_{CTD} and the flexible loop region of N-pep. A mutagenesis study indicates that the interaction between N-pep and P_{CTD} is mainly mediated through electrostatic and hydrophobic interactions. By introducing mutations and cyclizing N-pep, the binding affinity has been improved 200-fold. The structural and dynamic study of P_{CTD} and its phosphomimetics using X-ray crystallography and NMR suggests that there is no conformational change of P_{CTD} upon phosphorylation. NMR titration experiments further confirmed that the single binding site model of P_{CTD}/N-pep is not affected by the phosphorylation state of P_{CTD}.

Declaration

This is to certify that

i. The thesis comprises only my original work towards the PhD except where indicated in the Preface,

ii. Due acknowledgment has been made in the text to all other material used,

iii. The thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Jingyu Zhan

Preface

Publications arising from this thesis:

Zhan, J., Hossain, M. A., Sethi, A., Ose, T., Moseley, G. W., & Gooley, P. R. (2018). ^{1}H, ^{15}N and ^{13}C resonance assignments of the C-terminal domain of the P protein of the Nishigahara strain of rabies virus. *Biomolecular NMR assignments*, 1-4.

Where otherwise stated, work presented in this thesis was conducted by myself, the candidate, during my PhD candidature other than:

- Chapter 3, 3.3.6, Evidence that the highest affinity binder for P_{CTD} adopts a transient secondary structure.
  The ^{13}C,^{15}N-labelled sample, cyclized triple mutant (K376A/T386A/S389E) N-pep, was prepared by Siqiong Zheng[1]. Data analysis was performed by the PhD candidate.

[1] Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3010, Australia and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia.

Acknowledgements

First of all, I would like to thank my three supervisors Paul Gooley, Gregory Moseley and Michael Griffin for accepting me onto this project, and providing me guidance, encouragement and support throughout my PhD. I'm very thankful for all the time and effort they spent on helping me develop my skills. Also thank you to my committee chair Heung-Chin Cheng for the excellent advice and the help with guiding this project.

Thank you to all the previous and present members of the Gooley Lab, Moseley Lab and Griffin Lab, Patrick, Jesse, Alex, Ashish, Alamgir, Fei, Fengjie, Tasneem, Sam, Fabian, Shoni, Kim, Stephen, Cassandra, Aaron, Ericka, Angie, Kahenia and Riley. Thank you for all the help and suggestions in research, and for making it a great working place.

Of the other members of the Bio21 Institute, I would also like to thank Yee-Foong Mok for helping me with ITC, Ching-Seng Ang for teaching me to use RP-HPLC and Mass spectrometry, Shenggen Yao and David Keizer for helping me with NMR. Special thanks to my friends at Bio21, including Sunyuan, Tuo, Yilin and Xiaojing, who did PhD at the same time period as me and were great company.

Finally, I would like to thank my family and friends back at home in China. Despite the long distance for the past years, I still got endless understanding and love from them, which supported me through the ups and downs. Thank you.

Abbreviations

2D	Two Dimensional
3D	Three Dimensional
Å	Ångström
ABLV	Australian bat lyssavirus
ARAV	Aravan virus
BBLV	Bokeloh bat lyssavirus
BMRB	Biological Magnetic Resonance Bank
CK-II	Cellular casein kinase II
CPMG	Carr-Purcell-Meiboom-Gill
CRM-1	Chromosome region maintenance 1
CSP	Chemical shift perturbation
CTD	C-terminal domain
DEV	Duck embryo vaccine
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DUVV	Duvenhage virus
E. coli	*Escherichia coli*
EBLV-1	European bat lyssavirus 1
EBLV-2	European bat lyssavirus 2
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
FAK	Focal adhesion kinase
FDA	Food and Drug Administration
G	Glycoprotein, G protein
GBLV	Gannoruwa bat lyssavirus
GST	Glutathione S-transferase
HDCV	Human diploid cell vaccine
HEP	High egg passage
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSP70	Heat shock protein 70
HSQC	Heteronuclear Single Quantum Coherence
ICTV	International Committee on Taxonomy of Viruses

IDR_{CT}	C-terminal intrinsically disordered region of P
IDR_{NT}	N-terminal intrinsically disordered region of P
IDRs	Intrinsically disordered regions
IFN	Interferon
IKOV	Ikoma lyssavirus
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRKV	Irkut lyssavirus
ITC	Isothermal titration calorimetry
K	Association binding constant
kb	Kilobase
K_D	Dissociation constants
kDa	Kilodalton
KHUV	Khujand virus
k_{off}	Disassociation rate, off-rate constant
k_{on}	Association rate, on-rate constant
L	Large subunit of the RNA polymerase, L protein
LB	Luria Bertani
LBV	Lagos bat virus
LC8	Dynein light chain 8
le	Leader RNA
LEP	Low egg passage
LLEBV	Lleida bat lyssavirus
M	Matrix protein, M protein
MHz	Megahertz
μM	Micromolar
MALLS	Multi-angel laser light scattering
MOKV	Mokola virus
MoRE	Molecular recognition element
MWCO	Molecular weight cut-off
N	Nucleoprotein, N protein
N-5052	2YT auto-induction media
N-pep	The C-terminal flexible loop region of N
N-RNA	Nucleocapsid (genome encapsidated by N)
N^0	Nascent RNA-free N protein
nAChR	Nicotinic acetylcholine receptor

NBs	Negri bodies
NCAM	Neural cell adhesion molecule
NCL	Nucleolin
N_{CTD}	C-terminal domain of N protein
NES	Nuclear export sequence
NLS	Nuclear localization sequence
NMR	Nuclear Magnetic Resonance
N_{NTD}	N-terminal domain of N protein
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
NTV	Nerve tissue vaccine
NUS	Non-uniform sampling
OD_{600}	Optical density at 600 nm
P	Phosphoprotein, P protein
PCECV	Purified chicken embryo cell vaccine
P_{CED}	Central dimerization domain of P
PCR	Polymerase chain reaction
P_{CTD}	C-terminal domain of P
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PEP	Post-exposure prophylaxis
PHKCV	Primary hamster kidney cell vaccine
PKC	Protein kinase C
PML	Promyelocytic leukemia
ppm	Parts per million
PTM	Post-translational modification
PVCV	Purified Vero cell rabies vaccine
R_2	Transverse relaxation rate
$R_{2,eff}$	Effective transverse relaxation rate
RABV	Rabies virus
R_{ex}	Transverse relaxation term
RI	Refractometry
RMSD	Root-mean-square deviation of atomic positions
RNA	Ribonucleic acid
RNase	Ribonuclease

RNP	Ribonucleoprotein complex
RP-HPLC	Reverse phase high-performance liquid chromatography
RSV	Respiratory syncytial virus
RVA	Rabies Vaccine Adsorbed
RVPK	Rabies virus protein kinase
SAXS	Small angel X-ray scattering
SEC	Size-exclusion chromatography
SHIBV	Shimoni bat virus
SPR	Surface plasmon resonance
STAT	Signal transducer and activator of transcription
$\mathbf{T_2}$	Transverse relaxation rates
TBS	Tris Buffered Solution
$\mathbf{T_{cpmg}}$	Constant-time CPMG relexation period
TEV	Tobacco Etch Virus
TLS	Translation-Libration-Screw-rotation
TNFRSF16	Neurotrophin receptor $p75^{NTR}$
tr	Trailer RNA
Tris	Tris(hydroxymethyl)aminomethane
TBS	Tris-buffered saline
$\boldsymbol{\nu_{CPMG}}$	Frequency of the CPMG pulse
VSV	Vesicular stomatitis virus
WCBV	West Caucasian bat virus
WHO	World Health Organization
WT	Wild type
$\boldsymbol{\Delta G}$	Gibbs free energy
$\boldsymbol{\Delta H}$	Enthalpy
$\boldsymbol{\Delta S}$	Entropy

Amino acid name, three-letter code and one-letter code

Ala	**Alanine (A)**
Arg	**Arginine (R)**
Asn	**Asparagine (N)**
Asp	**Aspartate (D)**
Cys	**Cysteine (C)**
Gln	**Glutamine (Q)**
Glu	**Glutamate (E)**
Gly	**Glycine (G)**
His	**Histidine (H)**
Ile	**Isoleucine (I)**
Leu	**Leucine (L)**
Lys	**Lysine (K)**
Met	**Methionine (M)**
Phe	**Phenylalanine (F)**
Pro	**Proline (P)**
Ser	**Serine (S)**
Thr	**Threonine (T)**
Trp	**Tryptophan (W)**
Tyr	**Tyrosine (Y)**
Val	**Valine (V)**

Table of Contents

Tables

Figures

Supplementary Figures

Chapter 1:

Rabies virus and its replication machinery

1.1 Rabies: a public health threat and an economic burden

Rabies is one of the oldest zoonotic diseases with the earliest written description dated back to about 4,000 years ago [1], and it remains as one of the most deadly diseases with no treatment available and a huge threat to public health. Rabies is endemic throughout most parts of the world except Antarctica, resulting in > 61,000 human deaths per year with about 40% being children under fifteen years old, ranking as the seventh most important global infectious disease [2]. In Western Europe and North America, rabies is under epidemiological surveillance and controlled by mass vaccination programs, while in developing countries in Africa and Asia, rabies is still epidemic especially in rural areas, with more than 95% of rabies caused human deaths occurring in these two continents [3]. The actual death number is often underreported because of frequent misdiagnoses and lack of surveillance in these poor rural regions, making rabies also a neglected tropical disease.

Since there is no effective treatment for symptomatic rabies, widespread vaccination of domestic dogs and cats, mass oral vaccine baits for wildlife and administration of post-exposure prophylaxis (PEP) including human rabies immunoglobulin and vaccines are the most important ways to control rabies [4]. Although the number of human deaths in the developed countries are much less, the economic burden related to rabies is still huge. The worldwide annual cost of canine rabies is estimated to be about 8.6 billion US dollars [5], of which 40% is spent on post-exposure prophylaxis and directly related medical sector and patients, not to mention the impact on wildlife species (e.g. endangered Ethiopian wolves *Canis simensis* and African wild dogs *Lycaon pictus*) and losses in livestock industry [2]. The scale of the rabies disease is underestimated and there is an urgent need for the development of effective therapeutics for symptomatic rabies as well as high efficacy vaccines at low costs to prevent rabies.

1.2 Pathophysiology of Rabies

1.2.1 Transmission

Viruses in the *Lyssavirus* genus of the *Rhabdoviridae* family are the causative agent of rabies, among which the best characterized member is rabies virus. Lyssavirus is usually transmitted to humans through a bite or a scratch from the rabid animal (Figure 1-1), which allows the virus-laden saliva from the infected animal to enter into the human body through the non-intact skin or rarely through mucosal surfaces. Human-to-human transmission is extremely rare and is limited to exceptional cases reported in which the organ recipients were infected with rabies through cornea, kidney or liver transplantations from the unidentified infected donors [6-8]. Although all mammals can contract rabies, only species of the *Carnivora* order (including raccoons, foxes, dogs, skunks, mongooses) and *Chiroptera* order (bats) are important as the natural reservoir of lyssaviruses. Dogs are the principal vector of rabies responsible for up to 99% of the human infection cases especially in Asia and Africa where there is a large number of unvaccinated or stray dogs [2]. In North America, Western Europe and Australia where canine rabies is eliminated or controlled, bats become the principle transmitter of rabies [9].

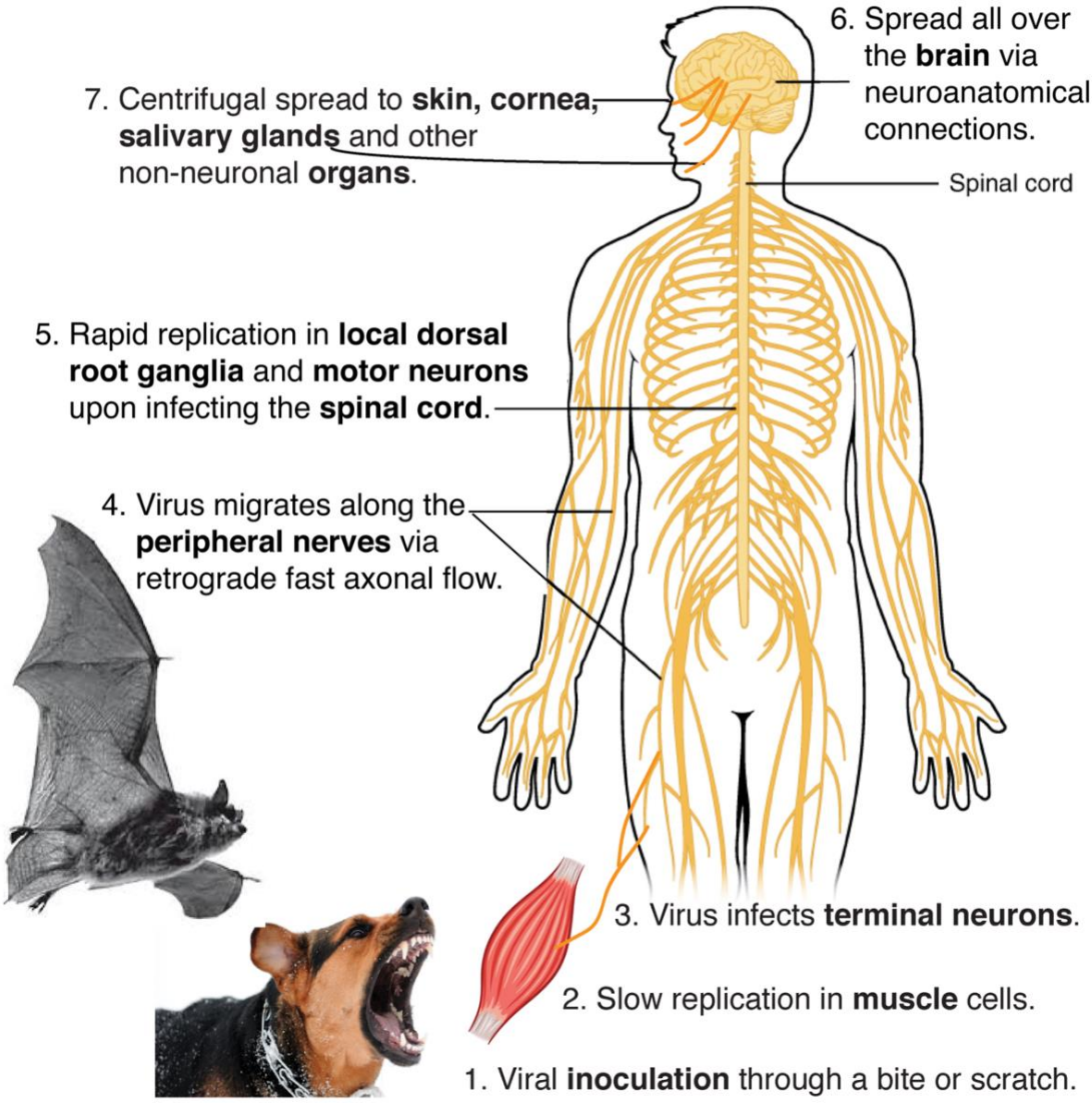


Figure 1-1.Schematic illustration of the major steps in lyssavirus infection.

This figure is modified from "Overview of Nervous System" by OpenStax, used under CC BY 4.0.

1.2.2 Incubation period

After the virus inoculation is established within the muscle tissues at the primary infection site, there is an incubation period with no symptoms of infection, during which virus replicates slowly in the muscle cells to reach terminal neurons like motor axons and motor endplates, then migrates through the peripheral nerves up to the central nervous system where extensive replication occurs, leading to the onset of the neurological symptoms and eventually almost inevitable death. This incubation period is a feature shared by lyssavirus infection and could expand from less than a week to more than a year, but the average time period is about three weeks to three months depending on the specific virus species, the severity of the wound, the viral dose inoculated within the wound, and the distance the virus has to travel from the primary site of infection to the central nervous system [10-12].

During the majority of the incubation period, virus keeps its replication at a very low level so that the infection is limited within the site of inoculation and is insufficient to be detected by the dendritic cells which could trigger the innate and adaptive immune system once activated [13]. The low replication level could be related to the endogenous RNA-silencing mechanisms, for example, skeletal muscle specific microRNA-133 is found to effectively suppress rabies viral protein expression, probably by interacting with the transcripts of viral nucleoprotein and glycoprotein [11]. As demonstrated in mouse models, the slow replication stage can be skipped in the case of high viral dose inoculation and/or deep penetrating bites, where the virus is able to infect neurons without previous accumulation within muscle cells or is even directly inoculated into the nervous system, leading to a substantially rapid disease progression [13-15].

On the contrary, bat lyssaviruses are found to be less efficient in invading the central nervous system, taking a prolonged time to induce immune gene expression and mortality compared with both the wild type and laboratory adapted rabies virus [16-18]. Another distinguishing feature of bat lyssaviruses is that their infectivity in epidermal cells at 34 ℃ (which is lower than normal human core body temperature) is much higher than rabies virus [19], which might be a critical factor that enables bat lyssaviruses to cause lethal rabies infection with only a small viral dose inoculated in the epidermis at body surface after trivial bat bites or scratches.

1.2.3 Prodrome

Once the lyssavirus reaches the terminal neurons at the primary infection site, it migrates along the peripheral nerves towards the central nervous system, with an estimated speed of about 8-20 mm per day via the retrograde fast axonal flow [20]. Upon infecting the spinal cord, virus replicates extensively in the motor neurons and dorsal root ganglia, and subsequently infects and spreads all over the brain via neuroanatomical connections.

The centripetal migration of virus from the peripheral nerves to the spinal ganglia and the central nervous system marks the end of the incubation period [21], and the initial prodromal symptoms of infection such as non-specific headache, fever, vomiting and fatigue show up and could persist up to ten days, which also indicates the progression of rabies has become irreversible and untreatable, and death is inevitable. The first rabies-specific clinical symptoms are neuropathic pain and burning sensation at the primary infection site, presumably due to the inflammation reaction in response to the viral replication in the local sensory ganglia [22]. This neuropathic pain is much more frequently found in bat lyssavirus infected patients (70%) compared with dog transmitted cases (30%) [21, 23], suggesting different preferences in sensory pathway.

1.2.4 Acute neurological phase, comatose phase and death

Acute neurological signs develop following the prodrome stage, with about 80% of the patients showing a classic furious/encephalitic form of rabies and the remaining 20% showing a dumb/paralytic form of rabies. The typical symptoms of the furious form of rabies includes hydrophobia, aerophobia, hypersalivation, seizures, episodes of agitation and hyperactivity alternating with lucid intervals, inspiratory spasms, contractions of the diaphragm on attempts to swallow, and dysfunction of autonomic nervous system [12]. The rapid neurologic deterioration soon leads to coma and eventually death by cardiorespiratory arrest.

Patients with dumb/paralytic form of rabies suffer a longer but less dramatic course, starting with flaccid muscle weakness at the primary infection site with later facial involvement. Quadriparesis, bladder incontinence and sphincter dysfunction are normally preceded by the slow development of coma and death resulting from respiratory failure. The symptoms of the paralytic rabies are often misdiagnosed as stroke, Horner syndrome or Guillain-Barré syndrome [2, 24], leading to an under-report and neglect of rabies.

After infection is developed within the brain, virus centrifugally spreads to and sheds in skin, salivary and serous glands, cornea, heart, liver and other non-neuronal organs except blood through the sensory and autonomic nerves using the same fast axonal transport mechanism, which serves as the eventual step to complete the viral infectious cycle [25]. The virus antigen in cutaneous nerves can be detected using immunofluorescent antibody in nuchal skin biopsy, which is used as a means of confirmatory diagnosis of the rabies infection [26]. The virus excreted in the salivary glands enables the infected mammals to spread rabies through bites,

and the viral infected organs from undiagnosed or misdiagnosed patients could cause human-to-human transmission if used in organ transplantation [8, 27].

1.3 Management, pre- and post-exposure treatment

Lyssaviruses infection in human stimulates the immune response and induces the production of a neutralizing antibody, but most patients do not have detectable neutralizing antibody in the serum on presentation with or in the early stage of infection because of the ability of the virus to evade and suppress the immune system [28, 29]. Later on, even if the neutralizing antibody become detectable as the infection is developed within the CNS, it still cannot penetrate the blood-brain barrier to enter into the CNS where the antibody is most needed, which also makes rabies untreatable from this point onwards.

1.3.1 Nerve tissue vaccines

Although rabies is inevitably fatal after symptom onset, it is preventable through pre- and post-exposure prophylaxis providing prompt vaccination [30] and immunoglobulin treatment are delivered in the case of a recognized exposure. The first rabies vaccine (Table 1-1) was developed by Louis Pasteur in 1885 by drying the spinal cords of rabid rabbits over caustic potash. This first ever rabies vaccine successfully saved hundreds of human lives, but the residual virus left after the incomplete inactivation procedure also brought the risk of occasional infection [31]. To achieve total inactivation, similar nerve tissue vaccines derived from infected sheep or goat brains were made by Claudio Fermi (1908) and David Semple (1911) but inactivated and preserved using chemicals such as phenol, thiomersal and beta propiolactone [32]. The major disadvantage of such nerve tissue vaccines is that its high myelin protein content may stimulate the immune system, to cause severe post-vaccinal encephalitis. Later, infant mouse brain with no or scant myelin present was used for vaccine production (Table 1-1) to reduce the neurological symptoms, although occasional

complications have been reported [33]. Due to the neurological side effects, rabies vaccine made from nerve tissues is marked as unsafe by the World Health Organization (WHO) and is recommended to be discontinued. Nevertheless, nerve tissue vaccines are still in use in some parts of Africa, Asia and Latin America because of the low cost and unsophisticated production process. According to WHO, the severe adverse reactions related to the use of rabies nerve tissue vaccines account for 0.8% of rabies caused human deaths [3].

1.3.2 Tissue culture and egg-based vaccines

The second generation of human rabies vaccine (Table 1-1), as the modern rabies vaccine, was developed using tissue culture or egg-based methods from the 1950s to 1990s. Fixed strain of rabies virus was cultured on human cells or chicken embryo fibroblasts, purified and concentrated through ultracentrifugation, inactivated by beta propiolactone, acetyl ethyleneimine or formaldehyde, and finally lyophilized except for Rabies Vaccine Adsorbed (RVA) which is adsorbed to aluminium phosphate ($AlPO_4$) thus being liquid (Table 1-1). These vaccines normally consist of three or five doses, given intramuscularly over a monthly period for pre-exposure or post-exposure treatment purpose respectively [3], and are used world-wide nowadays, The World Health Organization (WHO) has recommended to replace nerve tissue vaccines with modern embryonated eggs-based or cell culture-based vaccines, among which Rabavert®, Rabipur® and Verorab® (Table 1-1) have been prequalified by WHO [3]. Compared to the nerve tissue vaccines, the modern cell culture vaccines have the advantage of more immunogenic, higher potency and very few side effects. But the high cost raising from the virial purification step using ultracentrifugation, and the extensive vaccination regimens have limited its availability to the most needed people in less developed rural areas. Also due to the limited viral strains used for vaccine production, the protection is limited to certain lyssaviruses and cross protection from more distantly related lyssaviruses is not

available. In the case of post-exposure treatment for unvaccinated patients after a transdermal bite or scratch, rabies immunoglobulin which is highly expensive and hard to come by should also be instituted in combination with vaccination immediately after washing the wound [2].

1.3.3 Subunit vaccines

The new generation of subunit vaccines (Table 1-1) which uses virus vectors to overexpress rabies virus glycoprotein (G protein) is also under development [34]. Lyssavirus glycoprotein (G protein) is the only viral protein exposed on the virion surface, with multiple antigenic sites serving as the target for neutralizing antibodies. The third generation of rabies vaccine, which uses modified live viral vectors to express rabies G protein, has been proven to be effective in immunizing wildlife animals. Live vaccine bait such as Raboral V-RG® (Table 1-1) has been successfully used in the mass vaccination programmes in Europe and the U.S. to eradicate rabies from wild raccoons, coyotes and foxes [35]. Compared to the inactivated cell culture-based vaccines, attenuated vaccines have the obvious advantages of being able to preserve the original antigens, long duration of immunity, potential of oral administration, less or no requirement of adjuvant, and low cost for production. However, due to the safety concerns such as virus-vector infection, the attenuated live rabies vaccines have not been permitted for human use. Other plant-based vaccines, DNA vaccines and recombinant virus-vector vaccines are still in the experimental phase [36-39].

With the effective rabies control measures being developed and becoming available in developed countries, the effort put into rabies vaccine development has declined for the last thirty years. But in the meantime, the need for rabies vaccines in rabies-endemic areas has not decreased at all if not increased. As the current vaccines have the obvious disadvantages of extensive vaccination regimens, limited protection for only a number of lyssaviruses and high

production cost, development of safer, more affordable rabies vaccines with cross protection for human, domestic animals and wildlife as well as effective treatment for symptomatic rabies are in urgent need.

Table 1-1. The development of rabies virus vaccines.

First generation—Nerve tissue vaccines					
Vaccine	Year	Species	From	Viral strain and Inactivation	Notes
Pasteur vaccine [31] (by Louis Pasteur)	1885	Human	Rabbit spinal cord	Incomplete inactivation by drying over caustic potash.	Incomplete inactivation could cause infection. High myelin content could cause post-vaccinal encephalitis
Fermi vaccine (by Claudio Fermi)	1908	Human	Sheep or goat brain	Incomplete inactivation by phenol at 22 ℃.	
Semple vaccine (by David Semple)	1911	Human	Sheep or goat brain	Complete inactivation by phenol at 37 ℃.	High myelin content.
Beta Propiolactone vaccine		Human	Sheep or goat brain	Complete inactivation by beta propiolactone.	More antigenic than Semple vaccine. High myelin content.
Infant mouse brain vaccine (by Fuenzalida and Palacios)	1955-1964	Human	Suckling mouse brain	Complete inactivation by UV irradiation, beta propiolactone or phenol.	No or scanty myelin present to reduce neurological side effects.
Second generation—Tissue culture and egg-based vaccines					
Avian tissue vaccine (Flury LEP and HEP vaccine)	1950s	Dogs, cats and cattle	Chicken embryos	Virus from Flury strain was attenuated by passage through chicken embryos for 50^{th} generations as LEP or 180^{th} generations as HEP.	Low egg passage (LEP) live vaccine is used for dogs. High egg passage (HEP) vaccine is for cattle, cats and puppies.
Duck embryo vaccine [30] (DEV)	1956	Human	Duck embryos	Beta propiolactone	Poor immunogenicity
Primary hamster kidney cell vaccine (PHKCV)	1958	Dogs, cats and cattle	Primary hamster kidney cells	Fixed strain inactivated by beta propiolactone or acetyl ethyleneimine	Highly antigenic
Human diploid cell vaccine (HDCV-Imovax®)	1964	Human	Human diploid cells WI-38 or MRC-5	Flury HEP virus or Pitman-Moore strain virus was completely inactivated by beta propiolactone or formaldehyde.	Low yields of virus in WI-38 cells. Requires ultracentrifugation.

Purified chicken embryo cell vaccine (PCECV-Rabipur®)	1984	Human, domestic animals	Chicken fibroblasts	Flury LEP virus was completely inactivated by beta propiolactone	As effective as HDCV but cheaper. Requires ultracentrifugation.
Purified Vero cell rabies vaccine (PVCV-Verorab®)	1985	Human	Vero cell	Wistar strain virus was completely inactivated by beta propiolactone	The residual DNA can be oncogenic. Requires ultracentrifugation. Not approved by FDA yet.
Rabies vaccine absorbed (RVA)	1988	Human	Foetal rhesus lung diploid cell	CVS strain inactivated by beta propiolactone	Concentrated by adsorption to $AlSO_4$. Only used in Michigan, the USA.
Third generation—Subunit vaccines					
Vaccinia virus-vectored vaccine **[40]** (Raboral V-RG®)	1984	Wildlife. Raccoons and coyotes	Baby hamster kidney cells	Recombinant vaccinia virus expressing RABV G protein on the surface	The first and only oral live rabies vaccine bait to immunize wildlife approved by FDA.
Canarypox virus-vectored vaccine **[41]** (PureVax®)	1992	Cats and kittens	Chicken fibroblasts	RABV G protein expressed on the surface of canarypox virus	Attenuated live vaccine. Administered by subcutaneous route.
Canine adenovirus-vectored vaccine **[42]** (ONRAB®)	1996	Wildlife. Rodent, canine and skunk	Human embryonic kidney cell, Hela cells	Canine adenovirus expressing RABV G protein	Oral live vaccine bait for wildlife. First U.S. field trial of ONRAB® was conducted in 2011.
Recombinant virus expressing chimeric lyssavirus G protein **[43]**	1999	Tested in mice model	Neuroblastoma cells	Chimeric G protein with segments from European bat lyssavirus and Mokola virus	Recombinant vaccine virus expressing chimeric G protein provides broad cross protection from lyssaviruses.

1.4 The *Lyssavirus* genus

Members of *Lyssavirus* genus of the *Rhabdoviridae* family, in the order *Mononegavirales* are the causative agent of rabies. This order also contains the *Paramyxoviridae* (e.g. Respiratory syncytial virus and measles virus), *Filoviridae* (e.g. Ebola virus) and *Bornaviridae* (e.g. Borna disease virus) families. The negative-sensed single-strand RNA viruses within the *Mononegavirales* order are known to cause significant disease in human, and some are pathogenic to animals and plants.

Rabies virus (RABV) is the prototype virus of lyssaviruses, yet is just one member of this divergent virus genus. Based on genetic distances of the ectodomain of viral G protein and antigenic cross-reactivity, the currently identified sixteen species in *Lyssavirus* genus are categorized into three phylogroups (Table 1-2) [44, 45]. The virus species in phylogroup I are all maintained by bats, except that rabies virus (RABV) is also adapted to circulate in carnivores. Phylogenetic analysis indicated that the European bat lyssavirus 1(EBLV-1), Duvenhage virus (DUVV) and Irkut lyssavirus (IRKV) shared the same ancestor which is different from the common ancestor of RABV, EBLV-2, Aravan virus (ARAV), Khujand virus (KHUV), Bokeloh bat lyssavirus (BBLV), Gannoruwa bat lyssavirus (GBLV) and Australian bat lyssavirus (ABLV) [46]. The natural reservoir and epidemiology of some members such as RABV and EBLV are clearly documented, but for the newly identified ones species information need to be supplemented alongside future surveillance. Australia was considered "rabies free" until Australian bat lyssavirus (ABLV) was identified in 1994 in a flying fox and has caused three human deaths since then. ABLV was also identified on two horses which brought the concern of spill-over infection into domestic animals [47].

The three members in phylogroup II, Mokola virus (MOKV), Lagos bat virus (LBV) and Shimoni bat virus (SHIBV) share a separate monophyletic group. Both LBV and SHIBV are originated from bats, while the natural reservoir of MOKV remains to be identified. MOKV is widespread in Africa and is isolated mostly from domestic cats and sometimes from small mammals but never from bats. As MOKV is distantly-related to RABV, the current vaccines derived from RABV do not provide effective protection against MOKV infection, which is indicated by numerous cases in which MOKV was isolated from vaccinated domestic cats. So far, two human deaths due to MOKV infection have been reported in Nigeria [48].

In phylogroup III, West Caucasian bat virus (WCBV) and Lleida bat lyssavirus (LLEBV) are both bat-borne lyssaviruses, with a distinctly different ancestor from the lyssaviruses in phylogroup I or II. Ikoma lyssavirus (IKOV) was isolated from an African civet which is a rare lyssavirus reservoir. IKOV and MOKV are the only two lyssaviruses that have not been isolated from bats, which distinguishes IKOV and MOKV from all the other lyssaviruses. Nevertheless, the genetic relationship between IKOV and WCBV indicates that IKOV could actually result from a spill-over infection from bat-borne lyssavirus and its true reservoir is yet to be identified [49].

Table 1-2. Phylogroup classification of Lyssavirus

Species	Virus	Abbreviation by ICTV#	Reservoirs	Human Death
Phylogroup I				
Rabies lyssavirus	Rabies virus	RABV	All mammals	Y
Australian bat lyssavirus	Australian bat lyssavirus	ABLV	Bats	Y
European bat lyssavirus type 1	European bat lyssavirus 1	EBLV-1	Bats	Y
European bat lyssavirus type 2	European bat lyssavirus 2	EBLV-2	Bats	Y
Duvenhage lyssavirus	Duvenhage virus	DUVV	Bats	Y
Irkut lyssavirus	Irkut lyssavirus	IRKV	Bats	Y
Aravan lyssavirus	Aravan virus	ARAV	Bats	N
Bokeloh bat lyssavirus	Bokeloh bat lyssavirus	BBLV	Bats	N
Gannoruwa bat lyssavirus	Gannoruwa bat lyssavirus	GBLV	Bats	N
Khujand lyssavirus	Khujand lyssavirus	KHUV	Bats	N
Phylogroup II				
Mokola lyssavirus	Mokola virus	MOKV	Rodents, domestic animals	Y
Lagos bat lyssavirus	Lagos bat virus	LBV	Bats, domestics animals	N
Shimoni bat lyssavirus	Shimoni bat virus	SHIBV	Bats	N
Phylogroup III				
West Caucasian bat lyssavirus	West Caucasian bat virus	WCBV	Bats	Y
Ikoma lyssavirus	Ikoma lyssavirus	IKOV	African civet	N
Lleida bat lyssavirus	Lleida bat lyssavirus	LLEBV	Bats	N
#ICTV: International Committee on Taxonomy of Viruses				

1.5 Rabies virus

1.5.1 Genomic organization

Like the other Mononegaviruses, RABV has a non-segmented negative-sense RNA genome which is limited in size, and thus coding capacity [50, 51]. The 12 kb genome comprises five genes flanked by leader RNA (le) and trailer RNA (tr), encoding five viral proteins successively from the 3' end, which are the nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and the large subunit of the RNA polymerase (L) (Figure 1-2A) [51]. N protein encapsidates genomic RNA to form the helical nucleocapsid (N-RNA) in the virion. This N-RNA complex, instead of the naked RNA, serves as the functional template for viral transcription and replication by the RNA-dependent polymerase complex, composed of the enzymatic active L protein and the non-catalytic polymerase cofactor P protein [52]. The RNA-dependent polymerase complex associates with the N-RNA complex to form the internal capsid of the virion, termed as the ribonucleoprotein complex (RNP). Matrix protein (M) further condenses the RNP complex and forms the helix within the virion, which is then enveloped by the lipid bilayer membrane derived from the host cell. As a bridge, M protein connects the viral capsid, membrane, and the cytoplasmic side of the G protein. G protein exists as a trimer embedded in the viral membrane and plays vital roles in virus entry and release, as well as in membrane fusion [53]. Together, the viral RNA genome and five viral proteins assemble as the bullet-shaped virion, with a diameter of 90 nm and length of 180 nm [54] (Figure 1-2B).

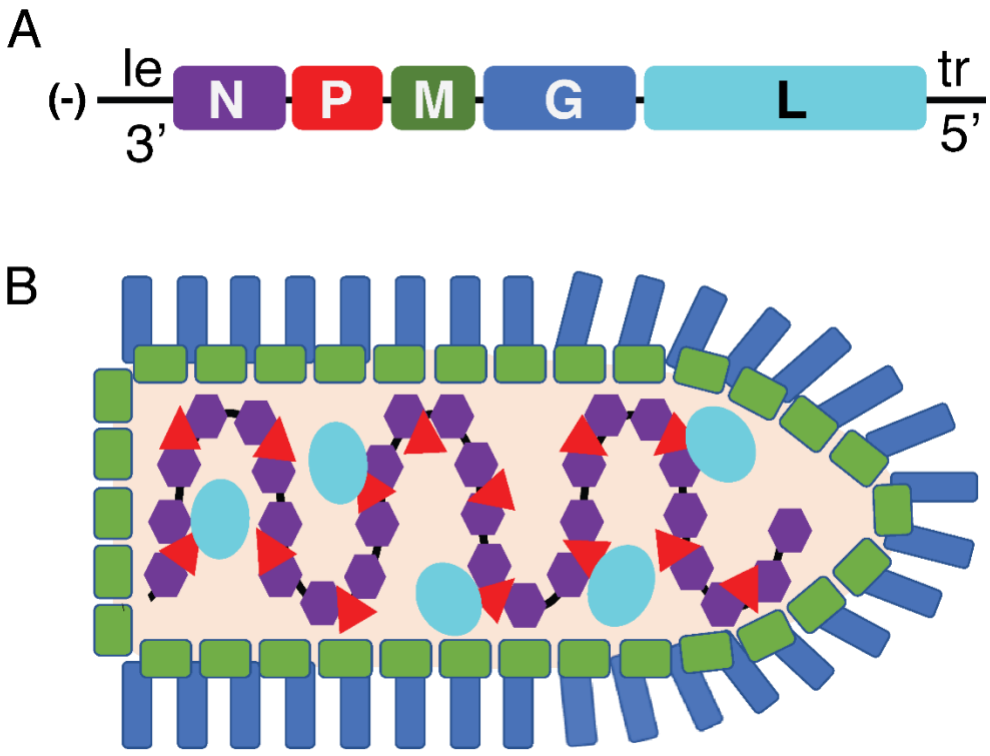


Figure 1-2. RABV genomic organization and virion structure [53].

A. Schematic organization of the RABV genome. RABV has a non-segmented negative sense RNA genome, which encodes N, P, M, G and L proteins successively from the 3' end.
B. N protein (purple) encapsidates genomic RNA (black) to form the helical nucleocapsid in the virion, which associates with the polymerase complex composed of P (red) and L (cyan) to form the internal capsid of the virion. The capsid is engulfed by the host cell-derived membrane, associating with M protein (green) and G protein (blue), forming the bullet-shaped virion.

1.5.2 Viral life cycle

The life cycle of RABV can be divided into three phases: virus internalization through endocytosis, transcription and replication within the cytoplasm, and final assembly and budding (Figure 1-3).

1.5.2.1 Virus internalization

After virus is inoculated into the muscle tissues via a bite or scratch, it slowly replicates in muscle cells to infect terminal neurons like motor axons and motor endplates (see 1.2.2). Entry of virus into the nerve cells is mediated by viral G protein binding to certain receptors on the host cell, and the virus is internalized through endocytosis. Since RABV can infect nearly all mammals, speculation about the host cellular receptors for RABV G protein is that the receptors should be highly conserved. Currently it is thought that the neurotrophin receptor $p75^{NTR}$ (TNFRSF16), nicotinic acetylcholine receptor (nAChR) and the neural cell adhesion molecule (NCAM) could all bind to G protein, and potentially other receptors are yet to be identified. Depending on the species of the host and stages of infection, different receptors might serve as needed for viral entry [55]. Once virus has penetrated into the host cell, its viral membrane merges with the membrane of the endocytic vesicle and the internal capsid, the RNP complex, is released into the host cell.

1.5.2.2 Viral transcription and replication

After the ribonucleoprotein complex (RNP complex) is delivered into the cytosol, transcription starts at the 3’ end of the negative-sense genome and produces one short leader RNA and five 5’-capped and 3’-polyadenylated positive-sense mRNAs encoding the five viral proteins [56]. Because the polymerase complex (L/P complex) initiates transcription at a single-entry site from 3’ and transcribes the five viral protein genes sequentially to 5’ in a stop-and-start manner,

the transcription of the downstream genes after a short intergenic sequence may not always re-initiate successfully, and thus the abundance of the leader RNA and the five mRNAs decrease sequentially along their gene positions on the genome (Figure 1-3) [57]. The mRNAs are then used to synthesize viral proteins by free ribosomes in the cytosol, except G protein which is translated on membrane-bound ribosomes and co-translationally inserted into the endoplasmic reticulum (ER) for post-translational modification of N-linked glycosylation before being processed by the Golgi [58]. P protein binds to the nascent RNA-free N^0 protein as the N^0/P complex (the 0 represents there is no RNA binding to nascent N), which confers N^0 protein the specificity of encapsidation of viral genomic RNA and delivers N^0 to the viral replication site for the assembly of the new nucleocapsids [59]. The accumulation of N^0/P is speculated to trigger the switch from transcription to replication [60, 61], but M protein is also found to be one regulatory signal which shifts the activity of viral polymerase complex from transcriptase to replicase [62, 63].

Replication starts at the 3' terminus of the negative-sense genome, and firstly generates a full-length complementary positive-sense anti-genome which is then encapsidated by N protein to serve as the substrate for the production of the negative-sense viral genome. The neo-synthesized genomic RNA is encapsidated by N protein as nucleocapsid, together with the attached polymerase complex (L/P complex) forming the ribonucleoprotein complex (RNP), and the assembly of sub-viral particles starts. The replication process heavily relies on ongoing viral protein synthesis to provide a continuous supply of nascent N^0 protein to encapsidate genomes and antigenomes so that the formation of double stranded RNA is avoided, also viral RNA does not get exposed to trigger the host immune system and is protected from cellular exonuclease [64].

Both transcription and replication are carried out in the cytoplasmic inclusion bodies termed as Negri bodies (NBs). First discovered in 1903 [65], Negri bodies are found in the brain tissues of rabid animals and infected patients, as well as in rabies virus-infected cells [66]. Negri bodies are viral factories with similar characteristics of liquid organelles in the cytosol, composed of viral N, P and L protein which are all essential for viral reproduction [66]. In a minimal system, similar cytoplasmic inclusions which recapitulate Negri-bodies properties can be formed by co-expressing RABV N and P proteins [67]. Some cellular proteins such as heat shock protein 70 (HSP70) [68] and focal adhesion kinase (FAK) [69] are also found in Negri bodies, probably hijacked by RABV to perform certain pro-viral functions. During transcription and replication, RABV uses a self-limiting strategy to keep its gene expression within the NBs to a submaximal level, such that the ongoing viral replication does not trigger apoptosis or any other counteracting response from the host cell. While in the case of vesicular stomatitis virus (VSV) infection from the same *Rhabdoviridae* family, the host gene expression is shut down completely for the efficient reproduction of VSV [70].

1.5.2.3 Assembly and Budding

In the last stage, the neo-synthesized viral ribonucleoprotein complex (RNP) is ejected from the replication centre (Negri bodies) and transported along the microtubules towards plasma membrane to get assembled [67]. M protein further condenses the RNP and recruits the relevant cellular machineries to form vesicles, which allows the new viral particle to bud out and start a new cycle of infection.

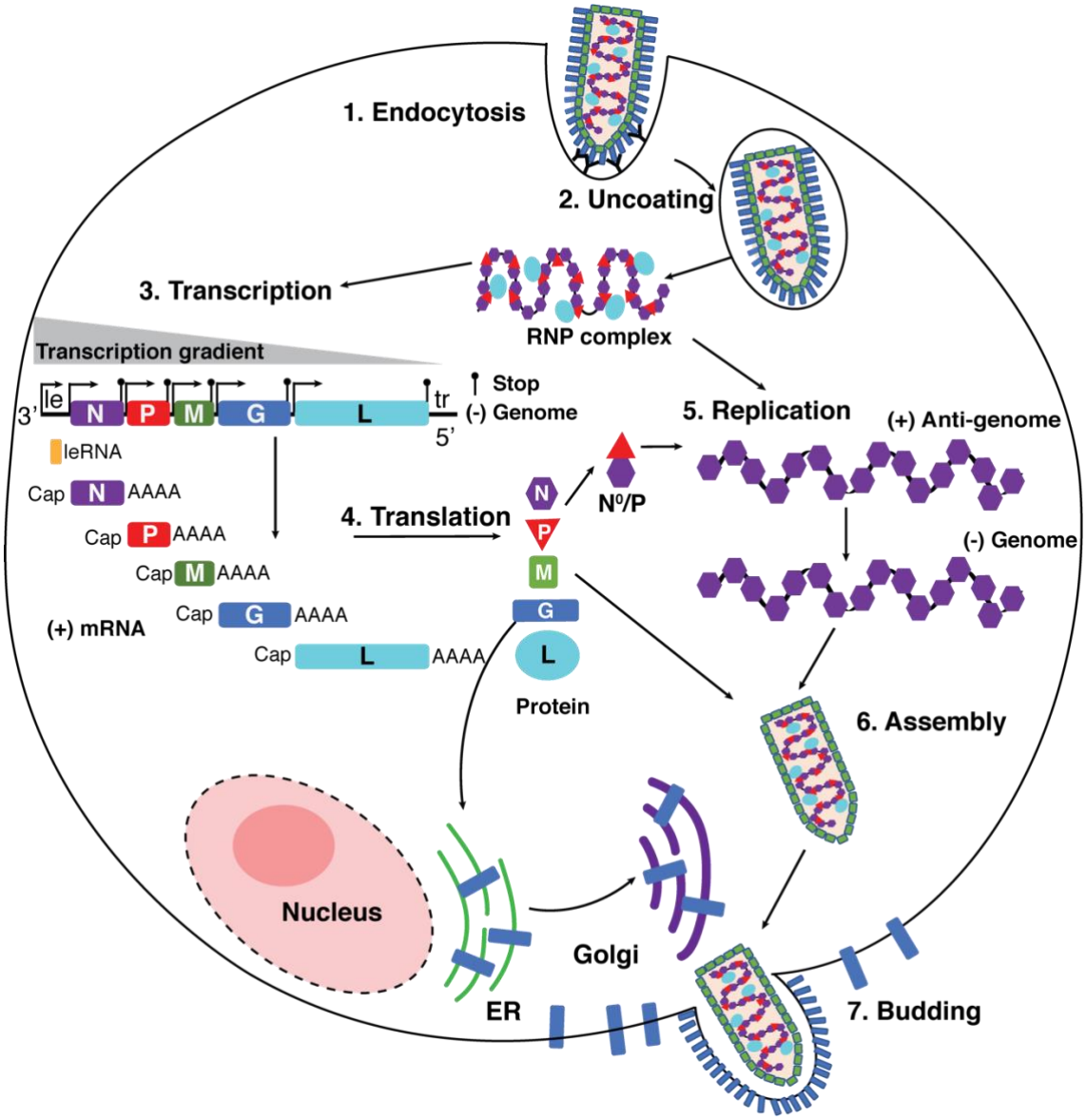


Figure 1-3. The replication cycle of rabies virus [68].

The rabies virus replication cycle can be divided into three main phases (i to iii). i) 1-2, Virus enters host cell through endocytosis and releases the inter capsid. ii) 3-5, Transcription and replication of viral genome and the production of viral proteins. iii) 6-7, Assembly of sub-viral particles and budding through membrane.

1.5.3 N protein

RABV N protein encapsidates genomic RNA and forms the helical nucleocapsid, serving as the template for viral replication and transcription. In infected cells, N protein specifically binds to viral genomic RNA, not host mRNA nor cellular RNA. Notably, when N protein is expressed in insect cells, recombinant N protein spontaneously binds to host cellular RNA non-specifically, forming helical nucleocapsid-like structures or circular N-RNA ring structures depending on the length of the RNA encapsidated [71]. The N-RNA rings usually contain 9 to 13 protomers, and these rings with different numbers of N protomers can be separated by preparative gel electrophoresis [72]. The recombinant helical structures and N-RNA ring structures have very similar properties as viral nucleocapsids, such as sensitivity towards trypsin treatment, resistance to RNase, density in CsCl gradient and the N-RNA stoichiometry [73].

1.5.3.1 Structure

The crystal structures of several *Mononegavirales* N-RNA rings have been solved. Although the sequence similarity is not very high (e.g. the sequence similarity between rabies virus N protein and vesicular stomatitis virus N protein is about 36%), their structures are homologous (Figure 1-4). N protein from rabies virus, vesicular stomatitis virus and respiratory syncytial virus have been crystallized as the recombinant N-RNA rings with 10 or 11 N monomers [64, 74, 75]. The N protomers in all the three ring structures have a similar bi-lobed appearance, consisting of an N-terminal domain (N_{NTD}) and C-terminal domain (N_{CTD}). Viral genomic RNA is bound in the positively charged groove between the N_{NTD} and N_{CTD}. In addition, two subdomains appear as extensions from the N_{CTD} and N_{NTD}. The N-terminal extension reaches out to the back of an adjacent N protomer to stabilize polymerization. For RABV and VSV, the

C-terminal extension also makes additional contacts with the neighbouring N protomer at the other side.

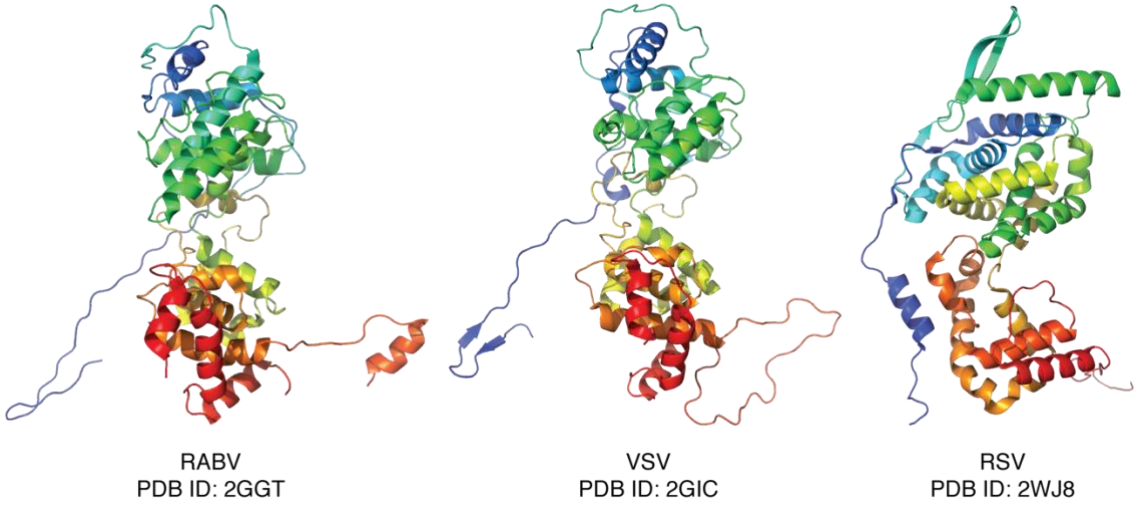


Figure 1-4. Structures of the N protein of RABV, VSV and RSV.

The ribbon diagrams are coloured from blue at the N-terminus to red at the C-terminus and are all in the same orientation.

For RABV N protein, both of its core domains are made up of α-helical bundles, with six α-helices in the N_{NTD} and 11 in the N_{CTD} [64]. Each core domain has a flexible loop region (N_{NTD} loop: 105-118; N_{CTD} loop: 376-397) which is missing in the crystal structure. The two core domains clamp down onto the genomic RNA, completely enwrap the RNA and hide it from the solvent. The tight RNA sequestering protects the RNA from being recognized by the intracellular receptors of the innate immune system, also hide the RNA from exonuclease activities (Figure 1-5) [64]. The C-terminal part (residues 377-450) can be cleaved off by trypsin treatment, leading to a reduction in molecular weight from around 50 kDa to 43 kDa. This trypsin sensitive peptide is located on the top end of the N-RNA ring, and the removal of this peptide does not have a major effect on the appearance of the ring structure, as observed by electron microscopy, or change to other structural parameters [73]. But cleaving off the C-terminal part (residues 377-450) by trypsin results in the loss of binding to P protein [76-78], which indicates that this C-terminal part might be involved in the interaction with P protein.

1.5.3.2 RABV N protein is phosphorylated

RABV N protein can be phosphorylated by cellular casein kinase II (CK-II) [79] and the phosphorylation site has been mapped to S389 in the N_{CTD} flexible loop region. The phosphorylation of S389 has been reported to modulate viral transcription and replication [80] by stabilizing the interaction between P protein and the viral nucleocapsid (N-RNA complex) [81], and it may also be crucial for binding to RNA or forming an ordered N-RNA complex as indicated by failed attempts to recombinantly express RABV N in *Escherichia coli* [72, 82]. As mentioned previously, N protein from RABV, VSV and RSV can be recombinantly expressed in insect cells to form N-RNA rings or helical nucleocapsid-like structures depending on the length of the cellular RNA encapsidated inside [71]. Currently the full-length RABV N proteins used in the published studies are mostly derived from insect cells [64, 72, 73, 83]. In

contrast, N protein from VSV and RSV are not phosphorylated, and can be recombinantly produced in both insect cells and bacteria [74, 75].

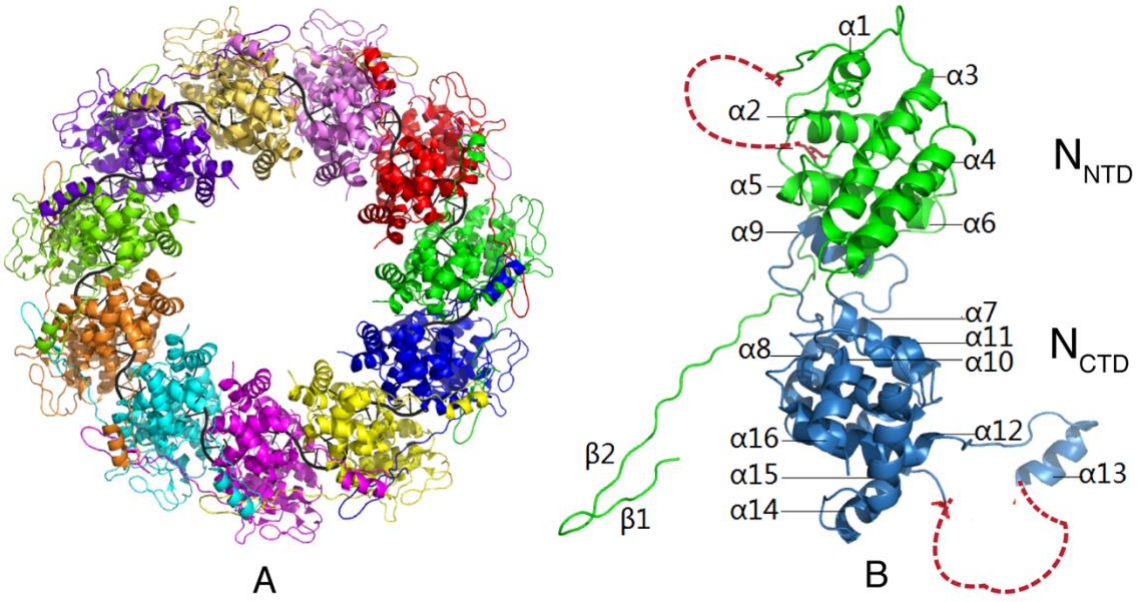


Figure 1-5. Structure of the rabies virus N-RNA complex and the N protomer [62].

A. Ribbon diagram of the N_{11}-RNA ring. The N_{11}-RNA ring is made of eleven N protomers with each of the N protomer colored differently. The RNA (black coil) is enwrapped inside the ring structure.
B. Ribbon diagram of the individual N protein. N has two core domains, N_{NTD} (green) and N_{CTD} (blue). Both domains are made up of α-helical bundles. The flexible loops in N_{NTD} and N_{CTD} are missing in the X-ray crystal structure and are marked as red dashed lines.

1.5.4 P protein

RABV P protein is a multi-domain protein that performs diverse roles in the virus-host interface as indicated by its interactions with multiple host factors, particularly in antagonizing host immune response [84, 85]. As the polymerase cofactor, P is also essential in viral replication and transcription [83, 86]. P protein self-assembles and oligomerizes both *in vitro* and *in vivo* through its central dimerization domain P_{CED} (residues 95-131, PDB: 3L32) to form a parallel dimer [87]. The structured P_{CED} is flanked by two intrinsically disordered regions (IDRs), a long IDR_{NT} in the N-terminus and a short IDR_{CT} which links to another well-structured domain P_{CTD} (residues 186-297, PDB: 1VYI) in the C-terminus. Similar structural organization is mapped to the corresponding regions of VSV P protein, suggesting a conserved modular organization (Figure 1-6).

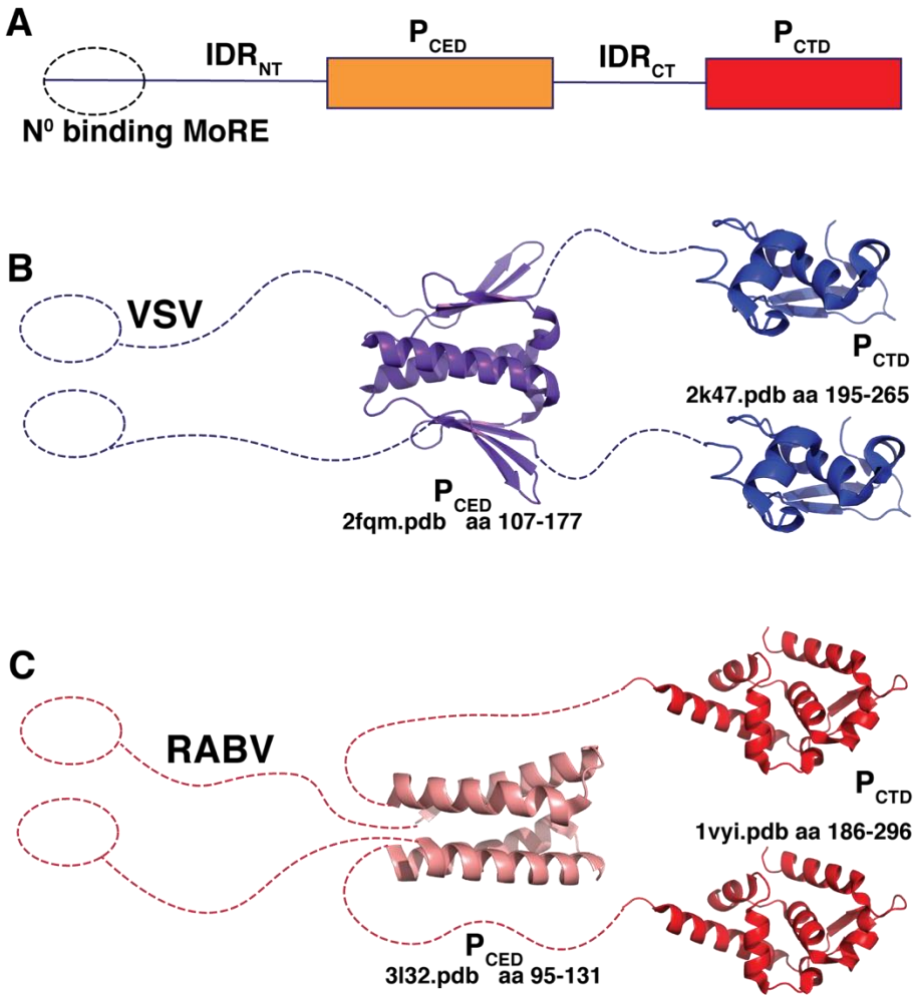


Figure 1-6. Structure of the phosphoproteins (P proteins) of the rhabdoviruses [85].

A. Schematic structure of the rhabdovirus P protein. It has a long disordered N-terminal region IDR_{NT}, a central dimerization domain P_{CED}, a short disordered region IDR_{CT} in the C-terminal and a structured domain P_{CTD} in the C-terminus.
B. The diagram structure of VSV P protein. The crystal structures of P_{CED} and P_{CTD} from VSV have been solved and are shown as cartoon. The dashed lines represent the intrinsically disorder regions (IDR_{NT} and IDR_{CT}), and the dashed circles correspond to the molecular recognition element (MoRE), the binding site for N^0.
C. The diagram structure of RABV P protein.

1.5.4.1 Structure, binding sites and functional modules

Overall, RABV P has well-structured domains separated by disordered regions, with multiple independently functional motifs and regions in both ordered and disordered regions. Its long acidic and presumably unstructured N-terminal region IDR_{NT}, spanning residues 1 to 95, contains the binding sites for the viral polymerase L protein [88] as the polymerase cofactor, and for nascent RNA-free N protein (N^0) as a chaperone to confer N^0 protein specificity to encapsidate viral genomic RNA [89]. Both P/L and P/N^0 interactions are critical for viral transcription and replication. There are also important functional regions identified within the IDR_{NT}, including a highly conserved CRM-1 exportin-dependent nuclear export sequence (N-NES) and two non-conserved phosphorylation sites S63 and S64, phosphorylation of which is catalyzed by a putative rabies virus protein kinase [90] (Figure 1-7).

Following the IDR_{NT} is a structured dimerization domain P_{CED} (PDB: 3L32), through which P forms a stable parallel dimer both *in vitro* and *in vivo* [87]. Focal adhesion kinase (FAK), a tyrosine kinase in the cytosol and a critical component in cellular signalling pathways regulating cell apoptosis, migration and proliferation, associates with P via the dimerization domain P_{CED} [69]. FAK is found in the cytoplasmic viral factories, Negri bodies, probably hijacked by P protein via the FAK/P_{CED} interaction to perform pro-viral functions [69].

The central dimerization domain P_{CED} links to another short disordered region IDR_{CT}, which bears a conserved PKC phosphorylation site S162 and a dynein light chain (LC8) binding motif. The interaction between P and the light chain of the microtubule minus-end directed motor dynein was initially assumed to be related to virus transportation in the peripheral nerves via retrograde fast axonal flow [91, 92], but has been more recently found to facilitate the microtubule-dependent NLS-mediated nuclear accumulation of P [93]. A nuclear localization

sequence is also found in the N-terminal region of P spanning from the second half of IDR_{NT} to IDR_{CT} (N-NLS, residues 53-174), which overlaps with and is inhibited by the N-NES (residues 49-58) in the full length P but can be activated when the first 52 residues including the N-NES are truncated [94].

The C-terminal domain P_{CTD} is an autonomous folding domain and a major interaction hub, which is conserved in the *Lyssaviruse* genus [95] and *Rhabdoviridae* family [96]. P_{CTD} has the shape of a sliced pear with a round face and a flat face, composed of seven helices (six α-helices and one 3_{10} helix) and an antiparallel β-sheet formed by two short β-strands. Several functionally important sites involved in immune evasion are identified in P_{CTD}, which include a dominate nuclear localization sequence (C-NLS, residues $_{211}$KKYK$_{214}$ and R_{260}) [97], a buried CRM-1 exportin-dependent nuclear export sequence (C-NES, residues 221-238) [98], two conserved PKC phosphorylation sites S210 and S271 [90], and binding sites for host factors including microtubules [99], nucleolin (NCL) [100], promyelocytic leukemia (PML) protein [101] and the signal transducers and activators of transcription (STAT) proteins [84, 102]. Among these host factors, microtubules together with other elements of the host cellular trafficking machinery exploited by P through NLS and NES enable P to go through nucleo-cytoplasmic trafficking, which is implicated in immune evasion; nucleolin expression is found to be important for RABV infection [100]; PML gene is one of the primary targets of interferons and could potentially be a part of the immune system although the biological function of the P/PML interaction is unclear [101]; STAT proteins are the key components in interferon signalling pathways of the cellular defence system and the P/STAT1 interaction is suggested to be part of the viral immune evasion mechanism and is critical for viral pathogenicity [84, 102]. Critical to viral replication and transcription, the P_{CTD} also incorporates the N-RNA binding site as the polymerase cofactor [73, 83, 96, 103], which is

localized to a conserved positively charged patch (residues $_{211}$KKYK$_{214}$, L$_{224}$, R$_{260}$) and is overlapped with the C-NLS.

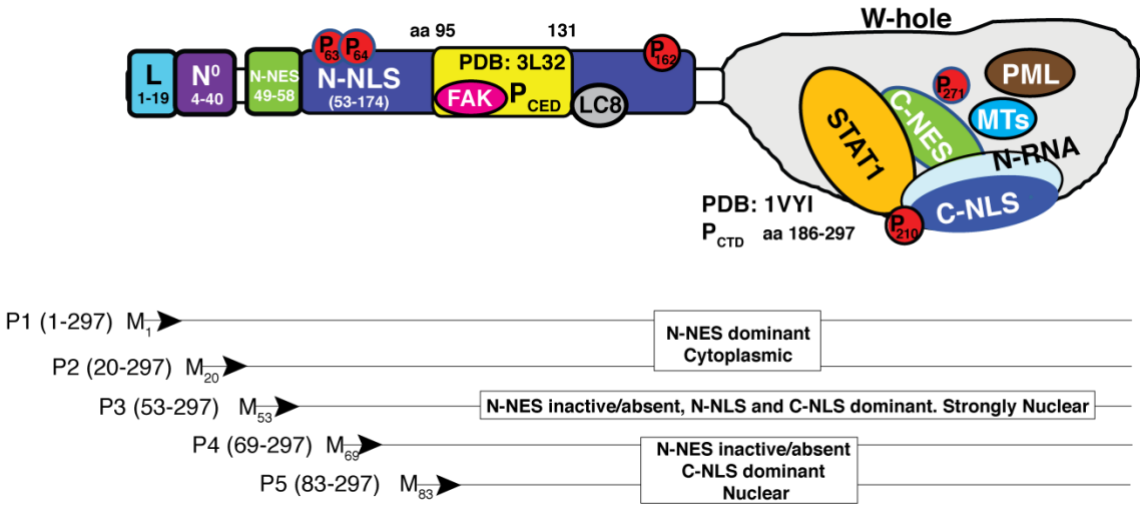


Figure 1-7. Schematic diagram of RABV P protein [98].

Upper panel: The NESs, NLSs, phosphorylation sites (P) and binding sites for protein partners are labelled on the schematic diagram of RABV P. L: large subunit of the polymerase complex; N^0: nascent RNA-free N protein; FAK: focal adhesion kinase; P_{CED}: central domain, and dimerization domain of RABV P (residues 95-131); LC8: dynein light chain 8; P_{CTD}: C-terminal domain (residues 186-297); STAT1: signal transducer and activator of transcription 1; N-RNA: nucleocapsid (genome encapsidated by N); PML: Promyelocytic leukemia protein; N-NES: N-terminal nuclear export sequence (NES); N-NLS: N-terminal nuclear localization sequence (NLS); C-NES: C-terminal nuclear export sequence (NES); C-NLS: C-terminal nuclear localization sequence (NLS); Serine phosphorylation sites: P63, P64, P162, P210 and P271.

Lower panel: There are five isoforms: one full length P (P1) and 4 truncated (P2-5) expressed from the P protein gene because of the leaky scanning mechanism. As indicated the isoforms vary in their cellular location. The sequence length differences of these isoforms are shown.

1.5.4.2 Functions

The first important role of RABV P protein is to chaperone the nascent N protein by forming the N^0/P complex, thus preventing N^0 binding to cellular RNA non-specifically and N polymerization [89, 104]. As mentioned before, N has a strong affinity to RNA and binds host cellular RNA non-specifically when expressed alone in insect cells. P binds to nascent N^0 molecules and delivers N^0 to the viral replication site for the assembly of the neo-synthesized genome and anti-genome, conferring N^0 specificity to encapsidating viral genomic RNA. Therefore, viral genome replication requires continuous production of the N^0/P complex and the accumulation of this complex is hypothesized to act as a switch between replication and transcription [105]. The binding site for N^0 is mapped to residues 4-40 of P by limited proteolysis [89].

Secondly, the non-catalytic polymerase cofactor P is essential for L to bind to the N-RNA substrate so that L can carry out all the enzymatic activity for viral genome replication and transcription, such as RNA synthesis, mRNA capping, methylation and polyadenylation [106, 107]. L is unable to bind to the N-RNA complex or access genomic RNA by itself [108], thus P binds to L and the N-RNA complex as the polymerase cofactor to keep the polymerase attached to its template during the translation and replication process. The binding site for L on P is mapped to the first 19 residues [109] whereas the binding site for N-RNA is localized to P_{CTD} [83, 96, 103].

Thirdly, P plays vital roles in interferon (IFN) antagonist activity. The principal antiviral response of host cells is the production of type I IFN, comprising IFN-α and IFN-β. IFNα/β binds to IFN receptors to activate STAT1 and STAT2 by phosphorylation. Activated STAT1 and STAT2 heterodimerize, translocate to the nucleus and activate IFN-stimulated genes

encoding antiviral proteins. The STATs are then dephosphorylated and exported to the cytoplasm. P binds to STAT, inhibits translocation of STATs to the nucleus, and consequently suppresses the innate immune response of host cells [84, 102]. The STATs binding site is mapped to the P_{CTD} [102], which is also the binding site for N-RNA complex and various cellular proteins like importin, CRM-1 exportin, microtubules, nucleolin, PML protein and potentially other unidentified proteins. The overlap of binding sites of the various proteins partners indicates that coordination of the N-RNA/P interaction and the diverse P-host cell interface is required.

1.5.4.3 Regulation and coordination of the diverse functions

As a multi-functional protein with multiple binding partners and overlapping functional sequences (Figure 1-7), how P coordinates these diverse functions and juxtapositioned binding interfaces remains largely unclear. One of the best understood mechanisms is the production of isoforms. The rabies virus genome, which is limited in size, encodes five viral proteins, among which the mRNA of P gene can be translated into 5 proteins due to a ribosomal leaky scanning mechanism, which includes a full-length P (P1) and four other N-terminally truncated P isoforms referred to as P2 to P5 (Figure 1-7). With the presence of a suboptimal Kozak sequence, translation of P can be initiated from the secondary downstream in-frame start codons encoding M20, M53, M69 and M83 (residues are numbered with respect to the full-length P) which results in the production of four N-terminally truncated P isoforms. These isoforms are found in purified virus, infected cells and P gene transfected cells with gradually lower abundance from P1 to P5, and their nucleo-cytoplasmic distribution and functions also varies because of the inactivation/truncation of the N-terminal functional sequences in the shorter isoforms [110].

Phosphorylation is a major reversible post-translational modification which may play important roles in regulating viral protein functions. *In vitro* experiments showed that P protein can be phosphorylated by a putative rabies virus protein kinase (RVPK) at non-conserved S63 and S64, and by protein kinase C (PKC) at conserved S162, S210 and S271 [90]. The phosphorylating state and the biological functions of these five putative phosphorylation sites remain unknown, except that PKC phosphorylation of S210 and its phosphomimetic mutation (S210D) are found to shift the nucleo-cytoplasmic distribution of P_{CTD} towards nuclear export [98]. Possibly phosphorylation of S210 induces conformational change to P_{CTD} and unmasks the originally buried C-NES, thus simultaneously regulates C-NLS and C-NES functions [98]. Since P_{CTD} interacts with multiple viral and host cellular proteins, change in its phosphorylation states and presumably the subsequent conformational change might impact on these interactions and ultimately, viral growth and pathogenicity.

1.6 Project background

The intermolecular interface between the C-terminal trypsin-sensitive peptide of N protein (N-pep) and P_{CTD} is an important component in the RABV replication machinery, also a potentially valuable target for anti-viral drug design and structure-based drug discovery. However, the interaction interface residues were only partially known from mutagenic studies which did not discount possible off-target effects. And the only data relating to the structure of the N-RNA/P_{CTD} complex came from a modelling study, based on a random mutagenic screen of P_{CTD} of the lyssavirus Mokola virus (MOKV) for N-binding residues [103], the crystal structures of RABV P_{CTD} [111] and circular N-RNA ring [64, 72], and small angel X-ray scattering (SAXS) data [83].

To identify P_{CTD} residues that may be critical for N-binding, MOKV P_{CTD} with random mutations were generated and their interaction with N were assessed using yeast two-hybrid system [103]. In this study, strong P_{CTD} mutants with defective N interaction were found to be more frequently mutated in a conserved basic patch ($_{211}$KKYK$_{214}$), although mutations in a hydrophobic pocket ('W-hole', which includes C261, W265, M287, see Figure 1-7) also appeared in some strong mutants almost simultaneously with basic patch mutations. Mavrakis et al. [111] re-analysed the random mutagenesis data [103] when they solved the crystal structure of RABV P_{CTD} [111], and suggested that both the basic patch and W-hole were important for N-RNA-binding [111].

Based on SAXS [83] and the available biochemical data [103, 111], a RABV N-RNA/P_{CTD} model was generated using a multistage modelling procedure [83]. Firstly, the flexible N_{CTD} loop was modelled based on SAXS data as it was missing in the N-RNA crystal structure [64, 72]. Then, initial N-RNA/P_{CTD} models were generated and filtered based on Mavrakis' analysis [111] of the random mutagenesis data [103], and only models with N_{CTD} loops binding to both the basic patch and W-hole of P_{CTD} were selected for further refinement. In the final model, P_{CTD} was proposed to lie on the C-terminal top of the N-RNA ring, with disordered N_{CTD} loops from two neighbouring N monomers acting as a pair of pincers that contact the P_{CTD} on two separate interfaces, the basic patch and the W-hole. And the affinity between P_{CTD} and N-RNA complex measured by surface plasmon resonance (SPR) was about 160 ± 20 nM for the circular N-RNA ring, or 150 ± 30 nM for the helical N-RNA. The final RABV N-RNA/P_{CTD} model showed a similar arrangement as the crystal structure of the Rhabdovirus VSV N-RNA/P_{CTD} complex [112] (Figure 1-8) despite the great variability in their protein sequences and binding interfaces [113]. One major difference is that in the VSV N-RNA/P_{CTD}

complex, each N protomer bound to one P_{CTD} [112], while the RABV N-RNA ring can only accommodate two P_{CTD} molecules without generating steric clashes [83].

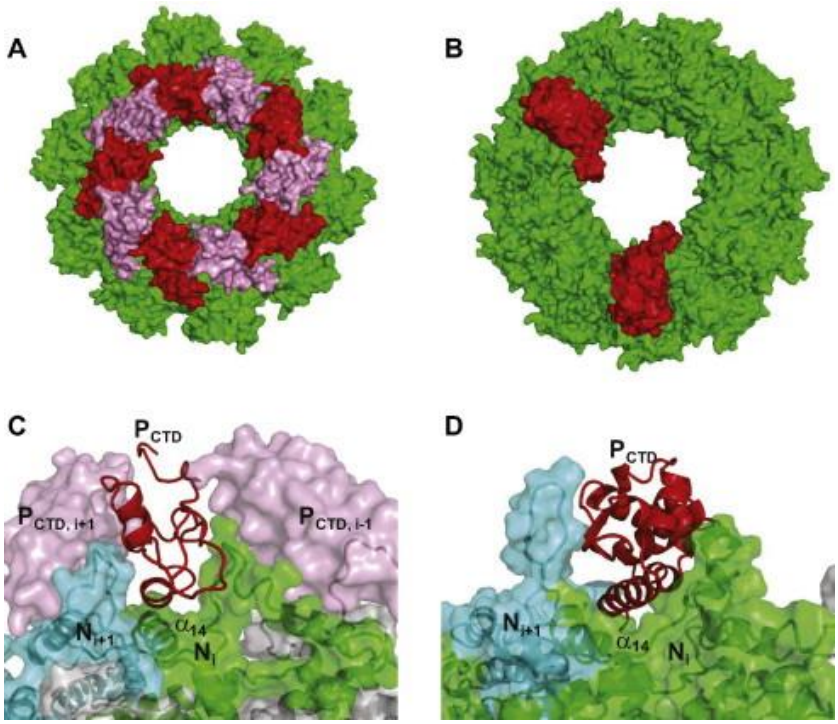


Figure 1-8. Structure of the N-RNA/P_{CTD} complex from VSV (X-ray structure) and RABV (SAXS model). Adapted from [84].

A. Overall structure of VSV N-RNA/P_{CTD} complex. Each P_{CTD} (pink or red) binds to the C-terminal top of one N protomer (green).

B. Structure of RABV N-RNA/P_{CTD} complex generated by molecular modelling from [83]. Two P_{CTD} (red) binds to the C-terminal top of one N-RNA ring (green).

C. Detailed structure of the VSV N-RNA/P_{CTD} complex.

D. Detailed structure of the RABV N-RNA/P_{CTD} complex.

However, the predictions in the RABV N-RNA/P_{CTD} models are controversial, as there are no direct data supporting a role for W-hole in N-binding. While the importance of the basic patch was validated in later targeted mutagenic study of MOKV P_{CTD} using yeast two-hybrid system [96] and in RABV P_{CTD} using mammalian cell-based assays [102], targeted W-hole mutations were found to have little impact on N-binding or viral replication in mammalian cell-based assays [102]. As the random mutagenic study [103] only generated multiple substitutions with rare W-hole mutations appearing almost simultaneously with the basic patch mutations, there is no evidence of independent function of the W-hole. Furthermore, although consistent effects of basic patch mutations were observed in previous assays [96, 102, 103], indirect off-target effects could not be discounted.

Moreover, as the putative reciprocal P_{CTD} binding interface, little was known about the N_{CTD} loop. This region was missing in the crystal structure probably due to flexibility, and was modelled based on SAXS data when generating the N-RNA/P_{CTD} complex model [83]. As unstructured loops can often go through disordered-to-ordered transition upon binding, e.g. the N_{CTD} loop of the Paramyxovirus Sendai virus and Hendra virus fold upon binding to P_{CTD}, the possible folding event was also predicted for RABV N [64], but has not been experimentally verified. Because N is highly sensitive to deletion [114] and is toxic to yeast, functional dissection studies cannot be performed on N [103]. Until now, no residues critical for P-binding were identified, except that CK-II phosphorylation of S389 of N protein [79] was found to be important for P-binding [81] and viral replication [80]. Targeted mutagenic analysis of MOKV N failed to identify any critical residues – mutating individually or in combination the two highly conversed acidic regions (residues $_{373}$EE$_{374}$ and $_{390}$DEED$_{393}$ of MOKV N) within the vicinity of S389 did not affect the N/P interaction [96].

The high nanomolar binding affinity measured by SPR reported together with the model [83] is also questionable. N and P were recently found to form cytoplasmic inclusion bodies (Negri bodies) with properties of liquid organelles, which is indicating a weak and dynamic interaction and is not compatible with the strong affinity reported [67]. Previous negative stain studies of the RABV N-RNA ring complexed with P dimers also indicated that P dimers bound to N-RNA ring in a flexible manner [73].

As efforts to generate crystallographic data for N-RNA/P_{CTD} complex have failed [83], the lack of direct high-resolution structural data supporting this model means it needs to be further examined and the exact nature of the interaction needs to be defined.

1.6.1 Hypothesis

P_{CTD} and its interaction with N-RNA represents a valuable target, and the structural details of the N-RNA/P_{CTD} interaction might help to unravel the replication machinery and the pathogenicity of rabies virus, providing new insights into the development of anti-viral drugs or vaccines.

1.6.2 Project aims

1. To fully assign the NMR spectra of P_{CTD} (Chapter 2).
2. To map the N-RNA/P_{CTD} binding interface and identify critical residues involved in the interaction (Chapter 3).
3. To investigate the potential effects of S210 phosphorylation on P_{CTD} structure, dynamics and its interaction with N-RNA (Chapter 4).

Chapter 2

^{1}H, ^{15}N and ^{13}C resonance assignments of the C-terminal domain of the P protein of the Nishigahara strain of rabies virus

^{1}H, ^{15}N and ^{13}C resonance assignments of the C-terminal domain of the P protein of the Nishigahara strain of rabies virus

Jingyu Zhan[1], Md. Alamgir Hossain[1], Ashish Sethi[1], Toyoyuki Ose[2], Greg W. Moseley[1,3] and Paul R. Gooley[1]

[1] Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3010, Australia and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia.

[2] Faculty of Advanced Life Science, Hokkaido University, Sapporo 060-0810, Japan.

[3] Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton Campus, Victoria 3800, Australia.

* Correspondence to:

Paul R. Gooley
The Department of Biochemistry and Molecular Biology and Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Victoria 3010, Australia
Ph: +61 3 9034 2273
Email: prg@unimelb.edu.au.

2.1 Introduction

Rabies virus is a species of the *Lyssavirus* genus of the *Rhabdoviridae* family in the order *Mononegavirales* [115, 116]. These viruses have a single-stranded negative-sense RNA genome that is limited in size (c. 12 kb) and, consequently, coding capacity. Viral pathogenicity depends on the ability of virus to enter cells, replicate, assemble progeny, and escape cells, for which lyssaviruses have a 'skeleton crew' of only five genes, all essential in the basic life cycle [86, 116]. However, as viruses are obligate intracellular parasites, their capacity to infect, spread and cause disease is also critically dependent on their ability to remodel host cells into efficient virus factories, by controlling elements such as innate antiviral responses, gene expression, cell survival and metabolism. To achieve such functional diversity, viruses express multifunctional proteins, which comprise overlapping functional sequences and domains [116, 117]. Mechanisms that enable and regulate this functional diversity include the expression of multiple isoforms of the protein from one gene, post-translational modification and intrinsically disordered regions [116-119]. The rabies phosphoprotein (P protein) is an archetypal example of this multifunctionality.

The P proteins of rabies virus and of other lyssavirus mediate multiple essential roles in replication and host cell manipulation [116, 117]. There are five isoforms, full-length P1 and P2-P5, which differ only through deletions at the N-terminal end [105]. The N-terminal region, which is likely to be disordered, is followed by a dimerization domain, another disordered region of about 30 residues and a globular 13-kDa C-terminal domain (P_{CTD}) [120]. The latter domain contains multiple interaction sites, including nuclear import and export sequences [98, 121]; two potential phosphorylation sites [90]; a basic-rich site that binds to the viral RNA-N protein complex [83]; and sites for interacting with host-cell proteins including the transcription factor STAT1 [102], nucleolin [100], PML protein [101] and microtubule

proteins [99, 122]. Importantly, several studies indicate that a number of key interactions are dynamic and regulated by processes such as dimerization and phosphorylation [98, 122], while other experiments using recombinant viruses derived from the pathogenic Nishigahara rabies virus strain have indicated critical roles in virulence for interactions of P protein with STATs [102], microtubules [99], and with nucleo-cytoplasmic trafficking machinery [123]. While an X-ray crystal structure of the P_{CTD} from the rabies CVS strain is available [111] we have initiated NMR studies of the P_{CTD} from the Nishigahara strain to define the sites of interaction and conformational changes that occur on post-translational modification and protein interaction. Here we report the ^{1}H, ^{13}C and ^{15}N chemical shift assignments of P_{CTD} which can be used for future protein-interactions and dynamics-based studies.

2.2 Methods and experiments

2.2.1 Protein expression and purification

To express P_{CTD}, the gene corresponding to residues 186-297 were PCR amplified and inserted into pET28a vector using *NdeI-EcoRI* restriction sites. The terminal cysteine (Cys297) was mutated to serine to avoid formation of intermolecular disulphide bonds. This pET28a vector was designed to have a His_{10}-tag on the N-terminal region of P_{CTD}. A TEV protease cleavage site was inserted between the His_{10}-tag and the P_{CTD} gene.

To express unlabelled P_{CTD}, vectors were transformed into BL21(DE3) using heat shock transformation protocols. 5 ml of overnight culture, prepared from a single colony in LB media supplemented with 0.5 ml of 30 mg/ml Kanamycin, was used to inoculate 500 ml of 2YT auto-induction media (N-5052), where the gene of interest is induced upon glucose-lactose metabolism [124]. At first the culture was grown at 37 °C for 4 hrs until the OD_{600} reached around 0.7-0.8. The culture was then transferred to a 16 °C incubator for overnight expression.

Cells were harvested by centrifugation at 3000 *g* for 10 min at 4 °C. Cell pellets were collected and stored at -20 °C until required for purification.

Expression of uniformly ^{15}N-labelled P_{CTD} was auto-induced by growing BL21 (DE3) in N-5052 containing 1 g/L of $^{15}NH_4Cl$ as a sole source of nitrogen [124]. Cells were grown and harvested similar to unlabelled P_{CTD} and stored at -20 °C until use. To label with ^{13}C and ^{15}N isotopes cells were grown in N-5052 [124] supplemented with 1 g/L of $^{15}NH_4Cl$ (Sigma-Aldrich) and 3 g/L of D-[^{13}C] glucose (Sigma-Aldrich) as sole sources of nitrogen and carbon. Cells were grown at 37 °C until an OD_{600} of 0.6-0.7, transferred to 16 °C and induced with 0.4 mM IPTG and protein expressed overnight. Biosynthetically directed 10% fractional ^{13}C labelling was achieved by growing cells in N-5052 [124] supplemented with 0.3 g/L of D-[^{13}C] glucose (Sigma-Aldrich) and 2.7 g/L of unlabelled glucose as the sole source of carbon.

To purify P_{CTD}, pellets from a 500 ml culture were resuspended in 50 ml of NaPi resuspension buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 7.4) with one complete-EDTA free protease inhibitor tablet (Sigma-Aldrich) dissolved. A hand-held homogenizer was used to completely resuspend the cells before applying to an Avestin EmulsiFlex C3 cell crusher. Cells were lysed with three cycles to ensure complete disruption. Insoluble debris was separated from the supernatant by centrifugation (13,000 *g* for 30 min at 4 °C). The clear supernatant was passed through 0.22 µm filters and applied to 5 ml Talon® metal affinity resin (Clontech, TAKARA) packed in a gravity flow column and pre-equilibrated in resuspension buffer. The resin was rocked for 90 mins at 4 °C, the column was washed twice (approximately 100 ml) with resuspension buffer and bound P_{CTD} was then eluted from the resin using ~80 ml of elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, pH 7.4). To remove

imidazole, the eluted sample was concentrated down to 5 ml using an Amicon Ultra-15 (MWCO 3 kDa) (Sigma-Aldrich) at 3,000 *g* at 4 °C and exchanged with 50 ml of buffer (50 mM Na_2HPO_4, 300 mM NaCl, pH 7.4). Following buffer exchange, and to cleave the His_{10}-tag, the protein sample was incubated overnight with TEV protease (0.5 ml of 1.8 mg/ml of purified TEV to 50 ml of protein sample) in the presence of 1 mM DTT. After cleavage, the protein was further concentrated to 1.5 ml using an Amicon Ultra-15 centrifugal unit (MWCO 3 kDa). P_{CTD} was further purified using a HiLoad™ 16/60 Superdex™ 75 prep grade column (GE Healthcare) pre-equilibrated with SEC buffer (50 mM Na_2HPO_4, 100 mM NaCl, 1 mM DTT, pH 6.8) and the flow rate was kept at 1 ml/min. Eluted 1 ml fractions with high purity were collected, pooled and concentrated in Amicon Ultra-4 centrifugal units (MWCO 3 kDa).

2.2.2 NMR spectroscopy

All NMR experiments were conducted at 298 K on Bruker Avance II 800 MHz or Avance IIIHD 700 MHz spectrometers equipped with 5-mm HCN (TCI) cryogenic probes. Assignments were accomplished with the following experiments: 2D ^{1}H-^{1}H TOCSY and NOESY recorded on 0.6 mM unlabelled P_{CTD} dissolved in 100% $^{2}H_2O$ and 50 mM phosphate buffer at pH 6.8; 2D ^{1}H-^{15}N HSQC, 3D HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, HBHA(CO)NH, CC(CO)NH, H(CCO)NH and ^{15}N-edited NOESY recorded on 0.6 mM ^{15}N P_{CTD} or 0.3 mM ^{13}C,^{15}N P_{CTD} dissolved in 10%/90% $^{2}H_2O/H_2O$ and 50 mM phosphate buffer at pH 6.8; two 2D constant-time ^{1}H-^{13}C HSQC (constant time periods of 13.3 ms (aliphatic) and 8.3 ms (aromatic)), 3D HCCH-TOCSY and two ^{13}C-edited NOESY (with ^{13}C offsets at 40 ppm (aliphatic) and 123 ppm (aromatic)) recorded on 0.3mM ^{13}C,^{15}N P_{CTD} dissolved in 100% $^{2}H_2O$ and 50 mM phosphate buffer at pH 6.8. A single high-resolution ^{1}H-^{13}C HSQC was recorded on a 0.32 mM 10% ^{13}C-labelled sample of P_{CTD} dissolved in 100% $^{2}H_2O$ and 50 mM phosphate buffer at pH 6.8. All 3D spectra were recorded using 10% non-uniform sampling (NUS) and

Poisson gap sampler [125], except the 3D ^{13}C and ^{15}N-edited NOESY spectra which were recorded with 25% NUS. Spectra were reconstructed with the compressed sensing algorithm using *qMDD* [126] and processed using *NMRPipe* [127]. All other non-NUS data were processed in *NMRPipe*. NMR data were usually processed with a Lorentz–Gaussian function in the direct dimension and cosine-squared bells in the indirect dimensions and zero-filling once. Preliminary auto-assignments of backbone atoms ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$) were obtained from the *PINE* server [128], and manually verified and extended to include Hα and sidechain atoms using *NMRFAM-Sparky* [129]. Preliminary ^{15}N transverse relaxation T_2 experiments were acquired at 800 MHz, processed in *NMRPipe* and analysed in *relax* [130]. Secondary structure analysis was performed using $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts with *SSP* [131].

2.3 Results

2.3.1 Backbone assignment

The 2D ^{1}H-^{15}N HSQC spectrum of P_{CTD} showed well-dispersed resonances from the backbone amides, Asn and Gln sidechain amides, and the $N^{\varepsilon 1}/H^{\varepsilon 1}$ of the indole ring of Trp, which is typical of a folded domain (Figure 2-1). Backbone assignments were done by manually verifying and supplementing the initial auto-assignment generated by the PINE server [128].

Preliminary auto-assignments of backbone atoms ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$) were obtained from the PINE server [128]. PINE is a novel algorithm that utilizes empirical estimates of the experimental data based on the available database, combined with consistency measures, to produce probabilistic assignments for the backbone chemical shifts. By uploading the unassigned peak lists of a set of NMR experiments covering the protein backbone (HSQC,

HNCO, HN(CA)CO, HNCACB, HN(CO)CACB) along with the P_{CTD} protein sequence to the PINE server, automated P_{CTD} backbone assignments were generated.

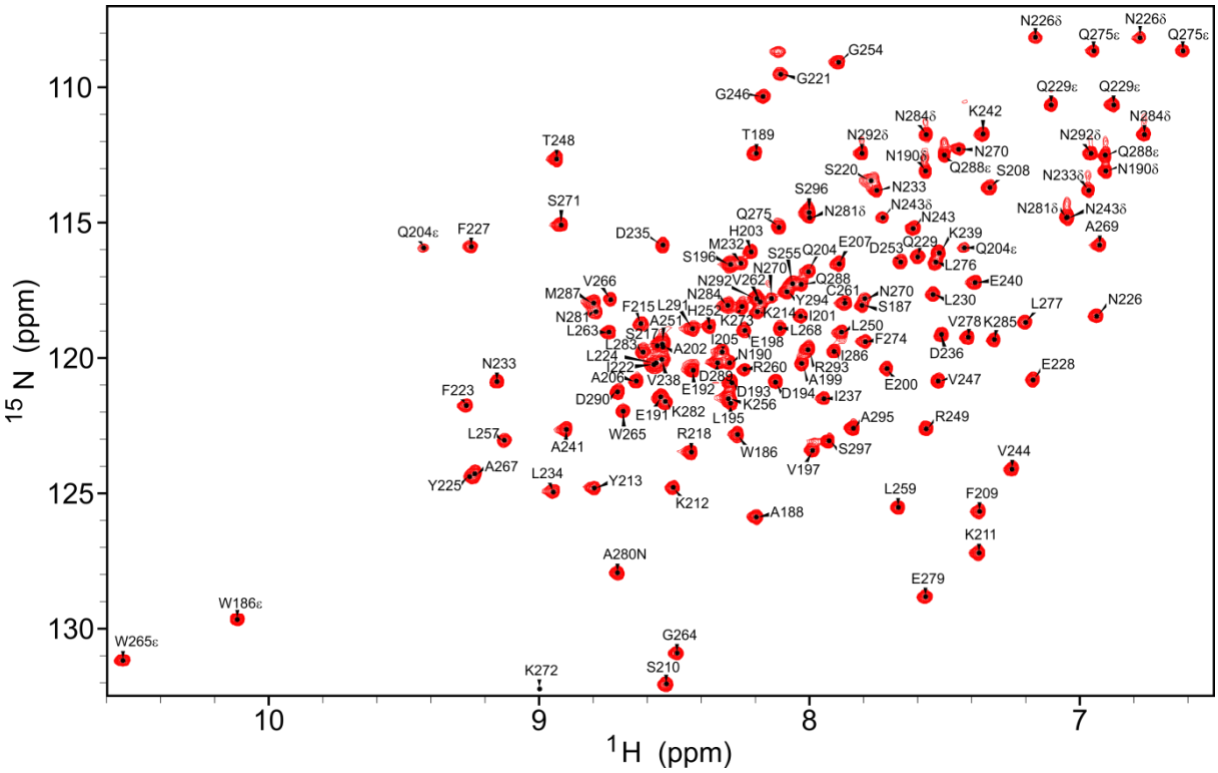


Figure 2-1. 700 MHz ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled P_{CTD} at pH 6.8 and 25 °C.

Assignments of backbone amide resonances are indicted by residue and number; side-chain amides for Gln and Asn residues and the indole of the two Trp residues are indicated by residue, number and atom.

To verify the automated backbone assignments from PINE, sequential connectivity between the spin systems was inspected manually. The NH strips of the complementary HNCACB and HN(CO)CACB spectra were arranged according to the protein sequence, and the sequential connectivity of the inter- and intra-residues was examined (Figure 2-2). In each NH strip of the HNCACB spectrum, the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ peaks from the same residue of the NH group ($^{13}C^{\alpha}_{(i)}$ and $^{13}C^{\beta}_{(i)}$) and the preceding residue ($^{13}C^{\alpha}_{(i-1)}$ and $^{13}C^{\beta}_{(i-1)}$) are usually all visible, but the

intensity of the intra-residue peaks is stronger than that of the preceding residue due to stronger intra-residue $^{1}J_{(N,C\alpha)}$ compared to inter-residue $^{2}J_{(N,C\alpha)}$. The NH strips of the HN(CO)CACB spectrum only have two strong peaks in opposite phases, which are from the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ of the preceding residue ($^{13}C^{\alpha}_{(i-1)}$ and $^{13}C^{\beta}_{(i-1)}$). Apart from the sequential connectivity, some amino acids with characteristic patterns of the backbone $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts could also be identified in this step, such as Ala, Gly, Ser and Thr (Figure 2-2).

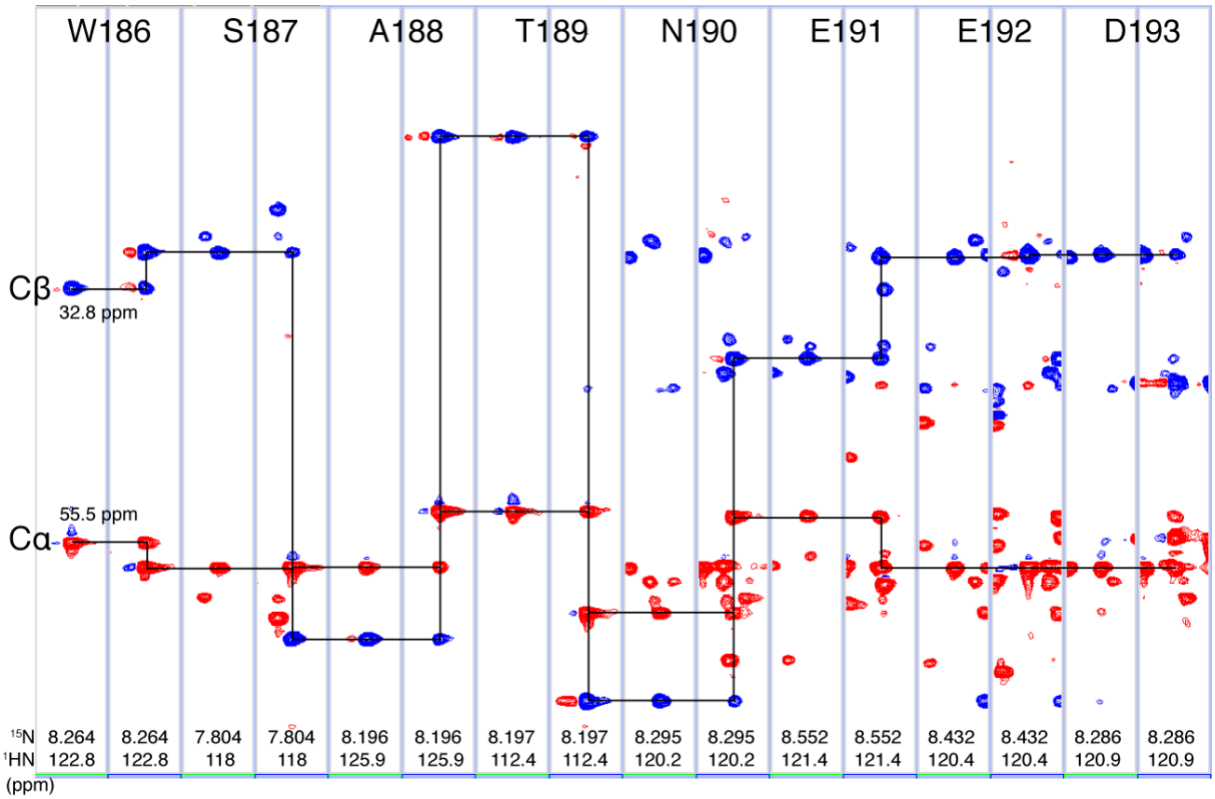


Figure 2-2. Sequential backbone assignment of P_{CTD}.

Sequential connectivity of the $^{13}C^{\alpha}$ (red) and $^{13}C^{\beta}$ (blue) are illustrated for residues W186 to D193. For each residue, strips from the complimentary HNCACB (right) and HN(CO)CACB (left) spectra at the amide ^{15}N frequencies are shown. S187 and T189 showed typical low C^{β} chemical shifts at about 63 ppm and 69 ppm respectively, while A188 showed typical high C^{β} chemical shifts at about 20 ppm; such characteristic patterns facilitate direct identification of the spin system types.

HNCACB, HN(CO)CACB together with ^{15}N,^{1}H-HSQC form the standard triple resonance experiments required for backbone assignment. However, for a medium-sized protein (P_{CTD} is 13 kDa) with overlapping peaks in certain regions, the HNCO and HN(CA)CO experiments were also conducted and analysed in a similar way to confirm the sequential connectivity. HNCO experiment has the best sensitivity among the three-dimensional experiments, and the C' chemical shifts obtained from this experiment can be used for secondary structure analysis and torsion angles prediction by TALOS+ [132]. The overall P_{CTD} backbone assignment completeness is 98.73%, among which 107 backbone NH resonances were assigned for the 109 non-proline P_{CTD} residues, except for Ser219 and Lys231 (Figure 2-1).

2.3.2 Chemical shift assignment of the sidechain atoms of P_{CTD}

Although PINE could perform automated assignment for sidechain atoms as well, the assignment of the sidechain resonances was performed manually considering the relatively low signal-to-noise and the complexity of the spectra. The assignment of the aliphatic sidechain carbon and proton atoms gradually supplemented and further confirmed the backbone assignment. With the defined backbone chemical shifts (see 2.3.1), the ^{13}C atoms in the sidechain can be identified from the CC(CO)NH spectra guided by the empirical estimates of the residue-specific chemical shifts distribution. Assignments of the H^{α} and H^{β} were made from HBHA(CO)NH and H(CCO)NH experiments, while the H^{γ}, H^{δ} and H^{ε} were assigned in H(CCO)NH spectrum. All the aliphatic ^{13}C and ^{1}H assignments were verified in the 3D HCCH-TOCSY dataset where all the aliphatic sidechain ^{1}H within the same ^{13}C spin system can be observed. The HCCH-TOCSY was also used to resolve overlapped peaks, and to assign the sidechain atoms of the residue preceding a Proline residue as its sidechain chemical shifts cannot be assigned through H(CCO)NH and CC(CO)NH experiments. Stereo-specific assignment of the methyl groups of Leu and Val are discussed later in 2.3.4.

The amide sidechain atoms of Gln and Asn were assigned using the 3D aliphatic ^{13}C and ^{15}N-edited NOESY experiments. The H^{δ} of Asn and H^{ε} of Gln were assigned based on their spatial correlation with H^{β} atoms in the NOESY spectra, and the chemical shift of the attached N was subsequently identified in the HSQC spectrum. All 24 of the amide sidechain resonances of Asn and Gln were present in the HSQC spectra and successfully assigned (Figure 2-1).

2.3.3 Aromatic sidechain assignment

The aromatic sidechain atoms were assigned by manually analysing the 2D ^{1}H-^{1}H TOCSY, ^{1}H-^{1}H NOESY, and the aromatic constant-time ^{1}H-^{13}C HSQC spectra. 2D TOCSY and NOESY spectra were acquired on unlabelled P_{CTD} dissolved in 100% $^{2}H_2O$ so that the amide protons are exchanged with ^{2}H in the solvent and will not show cross peaks. The suppression of the amide proton signals is important as amide protons have similar chemical shifts as aromatic sidechains. 2D aromatic ^{1}H-^{13}C HSQC was recorded with a constant time period of 8.3 ms, thus the ^{13}C nuclei coupled to odd number of carbons (e.g. $C^{\delta 2}/H^{\delta 2}$ of His) appear in opposite sign with those coupled to even number of carbons (e.g. $C^{\varepsilon 1}/H^{\varepsilon 1}$ of His), which provides additional information for the assignment.

The assignment of the aromatic sidechains started with the identification of the aromatic protons which are strongly correlated to H^{β} protons in the 2D NOESY spectrum, namely H^{δ} of Tyr, Phe, His and Trp. Tyr is relatively easy to assign as its H^{ε} chemical shift falls into a different region from H^{δ} and can be readily identified from its correlation to H^{δ}. Similarly, $H^{\varepsilon 1}$ of His can be identified from its correlation with $H^{\delta 2}$. The tautomeric states of the imidazole ring of His were not identified in this study. Trp has a complex sidechain, but the chemical shift

distributions of the aromatic protons are distinctly different from each other; upon identifying the $H^{\delta 1}$ of Trp, the rest could be identified by through-bond correlation using the 2D TOCSY dataset. The $N^{\varepsilon 1}/H^{\varepsilon 1}$ of Trp was assigned using the NOE connections between $H^{\delta 1}$, $H^{\zeta 2}$ and $H^{\varepsilon 1}$ in the 3D ^{15}N NOESY spectra. Phe is difficult to assign, because its $C^{\delta 2}/H^{\delta 2}$, $C^{\varepsilon 1}/H^{\varepsilon 1}$ and C^{ζ}/H^{ζ} pairs share very similar chemical shifts, causing strong coupling effects in the ^{13}C dimension and overlapping NOE signals. Except Phe209, Phe227 and Phe274, the aromatic sidechain resonances of P_{CTD} were all unambiguously assigned. The aromatic proton resonances of Phe209 and Phe227 were identified, but ^{13}C spins were not identified possibly due to their similar chemical shifts. Only one proton resonance of Phe274 was identified, which might be due to resonance broadening as a result of ring flipping at intermediate rate.

2.3.4 Stereo-specific assignment of the methyls

Reasonably comprehensive NMR resonance assignments are a prerequisite for NMR determinations of three-dimensional protein structures, and there are constant efforts made to improve the precision of the NMR structure calculation. Apart from the chemical shifts of the protein backbone and sidechains acquired using the basic sequential assignment strategy, stereo-specific assignment of the diastereotopic groups also greatly contributes to the precision of the final 3D structure without requiring more quantitative distance restraints. Stereo-specific assignment eliminates the use of pseudoatoms and pseudoatom corrections, which could otherwise weaken the accuracy of the quantitative experimental constraints [133]. Since the isopropyl methyl groups of Leu and Val have exceptionally large pseudoatom corrections [134] and usually make extensive NOE connections, their stereo-specific assignments can have an important improvement to the overall structure.

The stereo-specific assignment of the methyl groups of Leu and Val was achieved using 10% biosynthetic fractional ^{13}C labelling [135]. The bacteria were grown in minimal media supplemented with a mixture of 10% D-[^{13}C] glucose and 90% unlabelled glucose as sole carbon source. As the result of stereo-specific biosynthesis pathway, the *pro-R* methyl (γ^1 of Val and δ^1 of Leu) and the adjacent CH- group are always ^{13}C-labelled at the same time as they originate from the same pyruvate molecule; while the *pro-S* methyl (γ^2 of Val and δ^2 of Leu) and the adjacent CH- group only have a 1% chance (10% × 10%) to be ^{13}C-labelled simultaneously as they originate from different pyruvate molecules. The stereo-specific difference between the isopropyl methyl groups of the biosynthetic fractionally ^{13}C-labelled P_{CTD} is visible from its aliphatic ^{1}H-^{13}C HSQC spectra, where the *pro-S* methyls are shown as single peaks and the *pro-R* methyl signals are shown as doublets due to the one-bond ^{13}C-^{13}C coupling between the ^{13}C of *pro-R* methyl and the neighboring ^{13}CH- group (Figure 2-3). Using 10% biosynthetic fractional ^{13}C labelling, all the 39 identified diastereotopic methyls from Leu and Val were stereo-specific assigned.

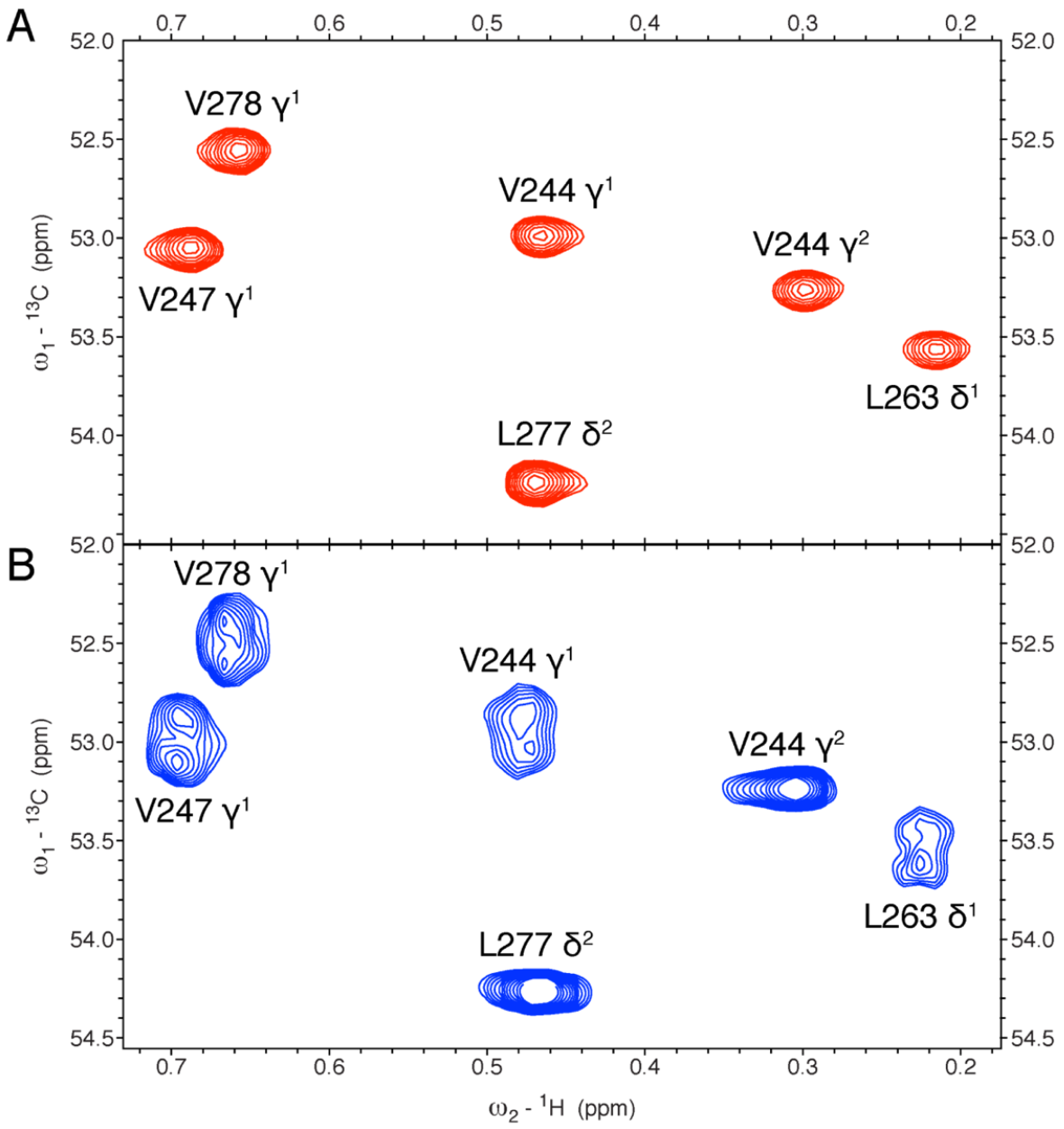


Figure 2-3. Segment of the ^{1}H-^{13}C HSQC spectra recorded on a uniformly ^{13}C-labelled (A) and a 10% ^{13}C-labelled (B) P_{CTD}.

Using biosynthetic fractionally ^{13}C labelling (B), the pro-S methyl (γ^2 of V244 and δ^2 of L277) peaks are singlets, while the pro-R methyl (γ^1 of V244, V247, V278, and δ^1 of L263) peaks are doublets with a splitting of about 33 Hz on the ^{13}C dimension (~0.19 ppm for spectra recorded on a 700 MHz spectrometer). The ^{13}C dimension of the methyl region showing here is folded and a spectra width of 32 ppm should be subtracted to get the accurate ^{13}C chemical shifts of the methyl groups.

2.4 Discussion – Extent of assignments and data deposition

The 2D ^{1}H-^{15}N HSQC spectrum of P_{CTD} showed well-dispersed resonances of a folded domain (Figure 2-1). As the first three residues (Gly-His-Met) represent non-native sequence of the RABV P protein, introduced from cloning, they are not included in the following statistics. The overall backbone assignment completeness is 98.73% among which 107 backbone NH resonances were assigned for the 109 non-proline residues, except for Ser219 and Lys231. All 24 of the amide side-chain resonances of Asn and Gln were present in the spectra and assigned (Figure 2-1). A total of 63 out of 64 methyl groups of Ala, Leu, Ile, Val, Met and Thr were assigned except for the $C^{\delta 2}/H^{\delta 2}$ of Leu257. All the 39 identified Leu and Val methyl groups were stereo-specifically assigned using biosynthetic fractional (10%) ^{13}C labelling. For the aromatic residues, the side-chain resonances of Trp186, Trp265, His203, His252, Tyr213, Tyr225 and Tyr294 were unambiguously assigned. The $C^{\delta 2}/H^{\delta 2}$, $C^{\varepsilon 1}/H^{\varepsilon 1}$ and C^{ζ}/H^{ζ} pairs of Phe215 and Phe223 were fully assigned while Phe209, Phe227 and Phe274 were partially assigned. Secondary structure analysis (Figure 2-4) reveals seven α-helices and two β-strands: residues 194-208, α1; 227-230, α2; 233-243, α3; 245-252, α4; 258-271 α5; 272-276 α6; 280-293 α7; 212-214 β1 and 222-225 β2. These secondary structures are in good agreement with the crystal structure (pdb ident: 1vyi) of the P_{CTD} from the CVS strain of rabies [111]. The only significant difference is that the first residue of helix α1 is 189 in the crystal structure. However, analysis of ^{15}N-R_2 suggests that the N-terminal of P_{CTD} is dynamic up to residue 195 (Figure 2-4). The chemical shift assignments for P_{CTD} have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number 27498.

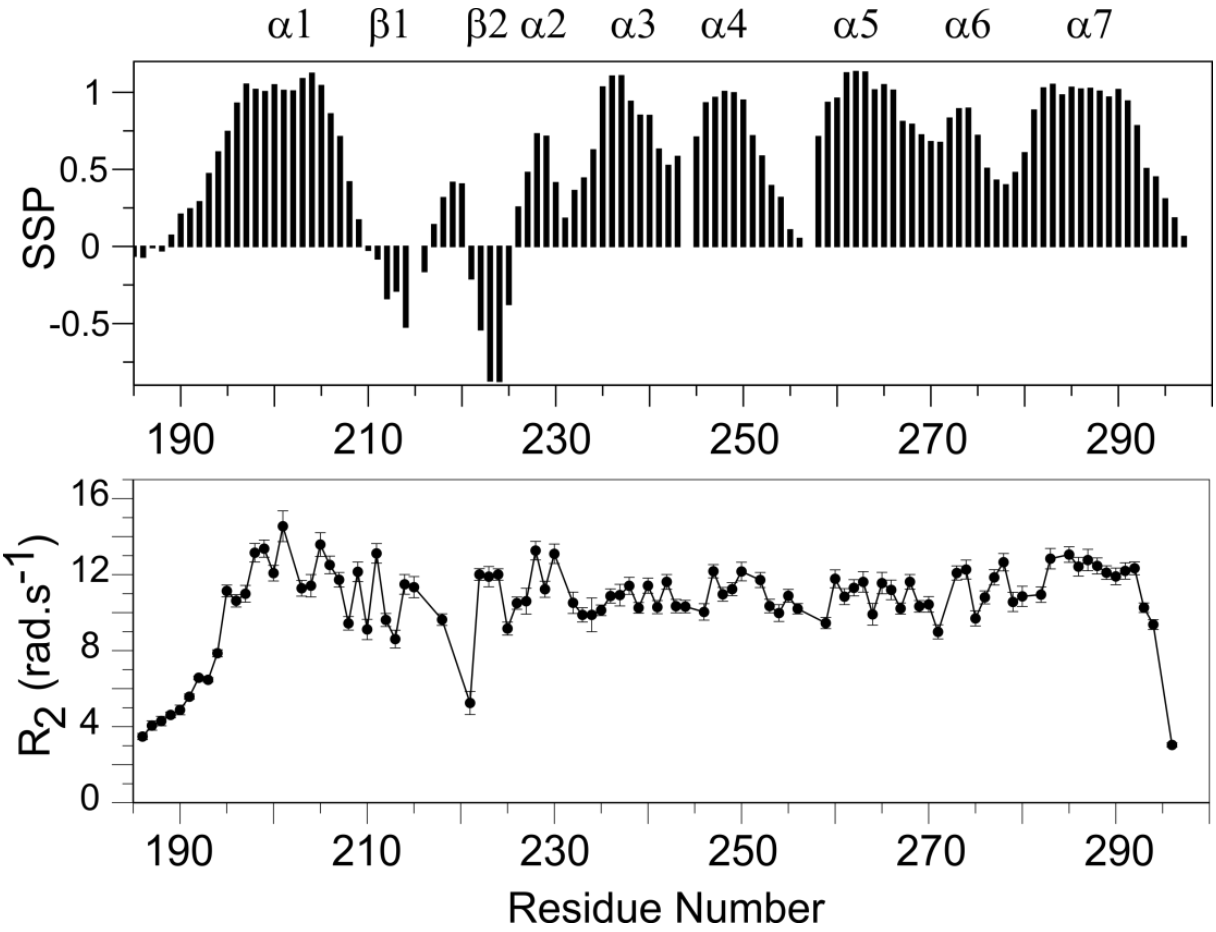


Figure 2-4. Secondary structure (A) and ^{15}N-R_2 analysis (B) of P_{CTD}.

Secondary structure was assessed by SSP [131] for the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift data. Seven α-helices and two β-strands are identified in agreement with the crystal structure for the same domain from the CVS rabies strain [111]. ^{15}N-R_2 analysis shows mobility in the N-terminus up to residue 195.

Chapter 3:

Structural analysis of critical interactions in the replication machinery of rabies virus: The N-pep/P_{CTD} interface

Structural analysis of critical interactions in the replication machinery of rabies virus: The N-pep/P_{CTD} interface

Jingyu Zhan[1], Md. Alamgir Hossain[1], Ashish Sethi[1], Siqiong Zheng[1], Gregory W. Moseley[1,2]*, Paul R. Gooley[1]*

[1] Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3010, Australia and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia.

[2] Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton Campus, Victoria 3800, Australia.

* Corresponding authors:

Paul R. Gooley
The Department of Biochemistry and Molecular Biology and Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Victoria 3010, Australia

Ph: +61 3 9034 2273
Email: prg@unimelb.edu.au.

Gregory W. Moseley
Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton Campus, Victoria 3800, Australia

Ph: +61 3 9905 1036
Email: greg.moseley@monash.edu

3.1 Introduction

Rabies virus (RABV) is a species of the *Lyssavirus* genus of the *Rhabdoviridae* family in the order *Mononegavirales* [136], which causes neurological disease in humans with a 100% case fatality rate, resulting in > 61,000 deaths/year world-wide [2, 5, 137]. Like the other viruses in the order *Mononegavirales,* RABV has a non-segmented negative sense RNA genome [50, 51], which is encapsidated by the nucleoprotein (N protein) to form the helical nucleocapsid (N-RNA). The N-RNA complex serves as the template for viral genome transcription and replication by the RNA-dependent polymerase complex [52], which is composed of the enzymatic L protein and non-catalytic polymerase cofactor phosphoprotein (P protein) [109, 138]. In addition to acting as the RNA-dependent RNA polymerase, L protein carries out other enzymatic activities including mRNA capping and polyadenylation. Importantly, L acts as a processive enzyme that has to remain attached and proceed along the N-RNA template [138, 139]. However, L protein alone is unable to bind to the N-RNA complex or to access genomic RNA, and so is absolutely dependent on the P protein, as the polymerase cofactor to attach to the N-RNA template [88]. The viral polymerase complex P/L together with N-RNA template accumulate in the infected cells, forming cytoplasmic inclusion bodies termed as Negri bodies (NBs). With all the replication machinery inside, these N/P rich inclusion bodies are the viral factories where all genome transcription and replication take place [66, 67]. The basic transcription and replication mechanism are shared by all members of *Mononegavirales* order, which includes all single-stranded negative sense RNA viruses.

P protein is a multi-domain protein that forms multiple interactions with host factors (Supplementary Figure 3-1), that underlie diverse roles in the virus-host interface, particularly in immune evasion [84, 85]. Interactions include several elements of the host cellular trafficking machinery [94, 97], enabling P protein to travel between the host cell cytoplasm and

nucleus, which is implicated in immune evasion [85, 116, 140]. As the viral polymerase cofactor, P also plays important roles in viral genome replication and transcription [83, 86]. The N-terminal region which is acidic and predicted to be globally disordered [105], contains binding sites for L protein [88] enabling P protein function as the polymerase cofactor, and for nascent RNA-free N protein (N^0), for chaperone function in conferring N^0 protein specificity to encapsidate viral genomic RNA [89]. Following a structured dimerization domain through which P forms parallel dimer (P_2), there is another short region of predicted disordered structure, and then a well-structured C-terminal globular domain, spanning residues 186-297, referred to as the P protein C-terminal domain (P_{CTD}) [111]. The P_{CTD} contains several functionally important sites including binding sites for host factors including microtubules [99], signal transducers and activators of transcription (STAT) proteins [84, 102], and promyelocytic leukemia (PML) protein [101], as well as a Nuclear Localization Sequence (NLS) [97] and a Nuclear Export Sequence (NES) [98]. Critical to viral replication/transcription, the P_{CTD} also incorporates the N-RNA binding site [73, 83].

When RABV N protein is expressed alone in insect cells, it binds to host cell RNA nonspecifically, forming helical nucleocapsid-like structures or circular N-RNA ring structures depending on the length of the RNA encapsidated [71, 73]. The crystal structure of the undecameric N-RNA ring complex has been solved [64, 72] and it reveals that N protein has a bi-lobal appearance, consisting of an N-terminal domain (N_{NTD}) and a C-terminal domain (N_{CTD}). The RNA is bound in the positively charged groove between the N_{NTD} and N_{CTD}. Two regions in N (N_{NTD} loop: 105-118 and N_{CTD} loop: 376-397) are absent in the crystal structure, probably due to flexibility. Removal of the C-terminal part of the flexible loop region (residues 377-450) by trypsin results in loss of binding to P protein [76-78], which indicates that this loop might be involved in the interaction with the P_{CTD}.

The interdomain interface between the C-terminal trypsin-sensitive peptide of N protein (N-pep) and P_{CTD} provides a potentially valuable target for anti-viral drug design and structure-based drug discovery. However, the only data relating to the structure of the N-RNA/P_{CTD} complex structure comes from a modelling study, based on a random mutagenic screen of P_{CTD} of the lyssavirus Mokola virus (MOKV) to identify residues involved in binding to N-RNA [96], the crystal structures of RABV P_{CTD} [111] and N-RNA [64, 72], and small angel X-ray scattering (SAXS) data [83]. In this study, P_{CTD} was proposed to lie on the C-terminal top of the N-RNA ring, where disordered N_{CTD} loops from adjacent N monomers act as pincers that contact the P_{CTD} on two separate interfaces, a positively-charged patch ($_{211}$KKYK$_{214}$) and a hydrophobic pocket ('W-hole', which includes C261, W265, M287, see Supplementary Figure 3-1). However, these predictions are controversial as there are no direct data supporting a role for W-hole in N binding [96]. Furthermore, while yeast two-hybrid screening performed with MOKV proteins [96] and mammalian cell-based assays performed with RABV proteins [102] validated a critical role for the positive patch, they also indicated that W-hole mutation did not impair N-binding or viral replication [102]. Importantly, in the random mutagenesis study of MOKV P_{CTD}, W-hole mutations were rare and found only in P_{CTD} proteins also mutated in the positive patch [103] or with other core packing destabilization mutations [111]; thus, there is no evidence of independent function of the W-hole. Furthermore, despite consistent effects of positive patch mutation, off-target effects could not be discounted. In addition, little is known about the N_{CTD} loop as the putative P_{CTD} binding site, as it was not resolved in the N-RNA crystal structure, and mutagenic analysis of MOKV N failed to identify any critical residues for P-binding [96]. Specifically, mutating individually or in combination the two highly conversed acidic regions (residues $_{373}$EE$_{374}$ and $_{390}$DEED$_{393}$ of MOKV N) in the C-terminal flexible loop region of MOKV N did not affect the

N/P interaction [96]. As efforts to generate crystallographic data for N-RNA/P_{CTD} complex have failed [83], the lack of direct high-resolution structural data supporting this model means it needs to be examined by other methods to define the exact nature of the interaction.

In the present study, we aimed to characterize the precise molecular interactions of P_{CTD} and N-pep. The P_{CTD} and N-pep were expressed in *Escherichia coli*, generating protein at high yield and purity; the structural properties of N-pep which were not characterized in the N-RNA crystal structure due to flexibility were investigated using chemical shift indexing, and NMR titrations of the proteins suggested a single binding mode with a micromolar affinity at the positive patch of P_{CTD} and the flexible loop region of N-pep. This provides the first direct confirmation that this represents the binding site. Using two-dimensional lineshape analysis, we measured the binding kinetics between various N-pep mutants and $P_{CTD,}$ and successfully identified the key N residues involved in P-binding for the first time, which revealed that the interaction between N-pep and P_{CTD} is mainly mediated through electrostatic and hydrophobic interactions. By introducing mutations and cyclizing N-pep, we found a N-pep variant which binds to P_{CTD} 200-fold more tightly than the wild type N-pep, and characterized the interaction by studying the energetics via isothermal titration calorimetry (ITC) and the kinetics via a combination of ITC and NMR experiments.

3.2 Materials and Methods

3.2.1 Constructs and mutagenesis

N-pep (residues 363-414, comprising the N_{CTD} loop and flanking helices $\alpha 13$ and $\alpha 14$) was cloned into a pGEX-6P-3 vector using *BamHI-XhoI* restriction sites, generating a construct to express N-pep with an N-terminal GST-tag followed by a PreScission protease cleavage site. P_{CTD} (residues 186-297) was cloned into pET28a vector using *NdeI-EcoRI* restriction sites,

producing a construct to express P_{CTD} as a His-tagged protein, with a TEV cleavage site between the His-tag and P_{CTD}. The last residue of P_{CTD} C297 was mutated to serine to avoid the formation of an intermolecular disulphide bond.

Mutations were introduced into the N-pep or P_{CTD} constructs using PrimeSTAR Max DNA Polymerase (Takara) according to the manufacturer's instructions, and using mutagenic primers designed according to Zheng et al. [141]. The amplified material was digested with DpnI for 1.5 hours at 37 °C to remove wild type bacterially expressed plasmid before transformation into Top10 *E. coli* cells. Mutation of the plasmid was confirmed by sequencing at the Micromon sequencing facility (Monash University).

3.2.2 Protein expression and purification

Unlabelled N-pep and P_{CTD} protein samples were expressed in *E. coli* BL21(DE3) in 2YT auto-induction media (N-5052) in which the expression of the target protein is induced upon the change of glucose-lactose metabolic state during the culture growth [124]. After OD_{600} reached 0.6~0.8 at 37 °C, the culture was transferred to 16 °C for overnight incubation. To produce uniformly ^{15}N-labelled N-pep and P_{CTD} protein samples for NMR, cells were grown in N-5052 supplemented with 1 g/L of $^{15}NH_4Cl$ as a sole source of nitrogen [124]. To label N-pep with ^{13}C and ^{15}N isotopes, cells were grown in 2YT media until OD_{600} reached 0.6~0.8 at 37 °C, then pelleted and resuspended in quarter the original culture volume of fresh N-5052 media supplemented with 2 g/L of $^{15}NH_4Cl$ (Sigma-Aldrich) and 4 g/L of D-[^{13}C] glucose (Sigma-Aldrich) as sole sources of nitrogen and carbon [142]. The resuspended cell culture was transferred to 16 °C and induced with 0.25 mM isopropyl β-D-1-thiogalactopyranoside after 1-hour incubation at 16 °C, harvested after a 15-hour induced period.

For purification of N-pep, bacterial pellets were resuspended in 10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl and 3 mM KCl (pH 7.4) and a protease inhibitor cocktail tablet (Roche). Cells were lysed using an Avestin EmulsiFlex-C3 homogenizer and the debris removed by centrifugation at 13,000 g, 4 °C for 45 minutes. The supernatant was applied to Glutathione Sepharose 4B resin (GE Healthcare), before washing with Phosphate-buffered saline (PBS) and incubation with Tris-buffered saline (TBS) solution containing PreScission protease (200 μL of 3 mg/ml purified PreScission protease in 25 mL of TBS) for 2 hours at 4 °C to remove the GST-affinity tag while bound to the Glutathione Sepharose matrix. The cleaved N-pep was washed out of the column with TBS buffer and concentrated using an Amicon Ultra 3 kDa centrifugation filter (Millipore). N-pep was then further purified by reversed-phase high performance liquid chromatography (RP-HPLC) using a 0 to 60% acetonitrile gradient (0.1% trifluoracetic acid) applied over 60 min at a flow rate of 5 ml/min using a C18 column (Agilent ZORBAX 300SB-C18, 5 μm, 9.4×250 mm). To generate cyclic N-pep variants, the cleaved linear precursor was concentrated and buffer-exchanged into 100 mM NH_4HCO_3 (pH 8.5) and oxidized overnight at room temperature before purification by RP-HPLC. The fractions collected from RP-HPLC were analysed by mass spectrometry to confirm the molecular weight, and the oxidation states for cyclic N-pep, before freeze drying. Typically, 3 to 4 mg of N-pep were purified from 1 L of bacterial culture.

The purification of P_{CTD} was performed as described previously [143]. Briefly, recombinant P_{CTD} was purified by TALON metal affinity chromatography (Clontech), and the His-tag was removed by overnight TEV treatment (0.5 ml of 1.8 mg/ml purified TEV per 50 ml of protein sample). Protein samples were further purified by size-exclusion chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare) in 50 mM Na_2HPO_4/NaH_2PO_4 (pH 6.8), 100 mM NaCl.

3.2.3 NMR experiments and Resonance Assignments

All NMR data were collected at 25 °C using a Bruker 700 MHz AVANCE III HD or a Bruker 800 MHz AVANCE II spectrometer, both are equipped with cryogenically cooled triple resonance probes to maximize sensitivity. Spectra were processed using *NMRPipe* [127] in which data were essentially Fourier-transformed after being multiplied by a Lorentz–Gaussian function in the direct dimension and cosine bells in the indirect dimensions and zero-filled once. Spectra were analysed using *NMRFAM-SPARKY* [129].

The spectra assignment of N-pep or its variant was performed using a 200 μM ^{13}C,^{15}N-labelled sample in 50 mM Na_2HPO_4/NaH_2PO_4 (pH 6.8) and 100 mM NaCl. A set of five triple resonance spectra were acquired: 3D HNCO, HN(CA)CO, HNCACB, HN(CO)CACB and CC(CO)NH. All the triple resonance spectra were recorded using 10% non-uniform sampling (NUS) [144] with multidimensional Poisson gap scheduling [125], and reconstructed using *qMDD* [145] with the compressed sensing algorithm before being processed in *NMRPipe* [127]. Assignments were done by manually verifying and supplementing the initial auto-assignment generated by the *PINE* server [128]. $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts were referenced using the random coil chemical shifts and secondary structure chemical shifts from *RefDB* [146], and were used to calculate secondary structure propensities using *SSP* [131]. Figures were generated using *PyMOL* [147] and *ESPript* [148]. Sequence alignment was performed with *Clustal X* [149] and the secondary structure elements were added using *DSSP* [150].

3.2.4 NMR titration experiments

The binding of N-pep variants to P_{CTD} was monitored using ^{1}H-^{15}N HSQC titration experiments. Both N-pep and P_{CTD} samples were dialyzed overnight against the same buffer

with 50 mM Na_2HPO_4/NaH_2PO_4 (pH 6.8) and 100 mM NaCl using the mini dialysis kit with 1 kDa cut-off (Amersham Biosciences). Concentrated unlabelled P_{CTD} was gradually added into the ^{15}N-labelled N-pep variant sample until reaching at least 5-fold molar excess. The ^{1}H-^{15}N HSQC spectra (2048×256 data points) were recorded at apo and at least six different titration points. The reverse titration of ^{15}N-labelled P_{CTD} with unlabelled N-pep (wild type) was conducted in a similar manner. Seven spectra were recorded at P_{CTD} : N-pep molar ratios of 1:0, 1:0.4, 1:0.8, 1:1.5, 1:3, 1:6 and 1:9.

Resonances showing obvious chemical shift changes yet remaining well-resolved were selected to calculate binding affinity by fitting the titration curve into a model for a protein with only one ligand binding site using *Xcrvfit* software [151]. Dilution through the titration was taken into account. The chemical shift perturbation (CSP) of each residue was calculated by the formula [152]:

$$\mathrm{CSP} = \sqrt{\Delta\delta_{\mathrm{H}}^2 + 0.154^2 \cdot \Delta\delta_{\mathrm{N}}^2} \qquad \text{Equation 1}$$

For all ^{15}N-labelled N-pep variants with P_{CTD} NMR titration data, two-dimensional lineshape analysis was done using the software *TITAN* [153]. The off-rate (k_{off}) and dissociation constant (K_D) were determined by simultaneously fitting the lineshape of selected shifted yet resolved peaks to a two-state single binding site model. The on-rate (k_{on}) was determined by the equation:

$$k_{on} = \frac{k_{off}}{K_D} \qquad \text{Equation 2}$$

3.2.5 Isothermal titration calorimetry

Thermodynamic analysis of the interaction between P_{CTD} and N-pep was conducted with a MicroCal iTC200 apparatus (GE Healthcare). Samples were dialyzed overnight against 50 mM

Na_2HPO_4/NaH_2PO_4 (pH 6.8) and 100 mM NaCl. The cell was loaded with 20 to 60 μM N-pep solution and the syringe was loaded with P_{CTD} (from 200 μM to 1.3 mM, dependent on the estimated affinity between P_{CTD} and different N-pep variants). The experiments were performed at 25 °C, with 16 total injections of 2.48 μL per injection. Each injection duration is 5 s with 180 s intervals. To correct the dilution heat of the P_{CTD} sample, a blank where the same P_{CTD} sample was injected into buffer was performed, and the measurement was subtracted from each experimental data set before fitting to a single binding site model using *MicroCal ORIGIN* software. Association binding constant (K in M^{-1}), number of binding sites (N), enthalpy (ΔH in kcal/mol) and entropy (ΔS in kcal/mol/deg) were obtained from the fit. Dissociation constants (K_D) were obtained as the inverse of the association constant (K), while Gibbs free energy (ΔG in kcal/mol) was calculated using the following equation:

$$\Delta G = \Delta H - T\Delta S \qquad \text{Equation 3}$$

where T is absolute temperature in Kelvin.

3.3 Results

3.3.1 The C-terminal loop region of N-protein is intrinsically disordered

In spite of its implicated important role in N-RNA/P interaction, the structural properties of the C-terminal loop region of N protein are unknown due to flexibility in the crystal structure. The structural properties of N-pep were investigated using NMR. The ^{1}H-^{15}N HSQC spectrum was recorded at 25 °C. In the ^{1}H-^{15}N HSQC spectra, the resonances were dispersed over a narrow range in the $^{1}H^{N}$-dimension, indicating that N-pep is predominantly unstructured. Backbone ^{1}H, ^{15}N and ^{13}C resonances of N-pep (residues 363-414) were assigned using standard three-dimensional heteronuclear NMR experiments (Figure 3-1A). The $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts depend on local backbone geometry and are exquisitely sensitive to the presence of secondary structure, which provides a means to identify regions of regular

secondary structure [154]. Consistent with the data indicating the N-pep as unstructured, no patterns could be discerned from the N-pep chemical shifts (Figure 3-1B).

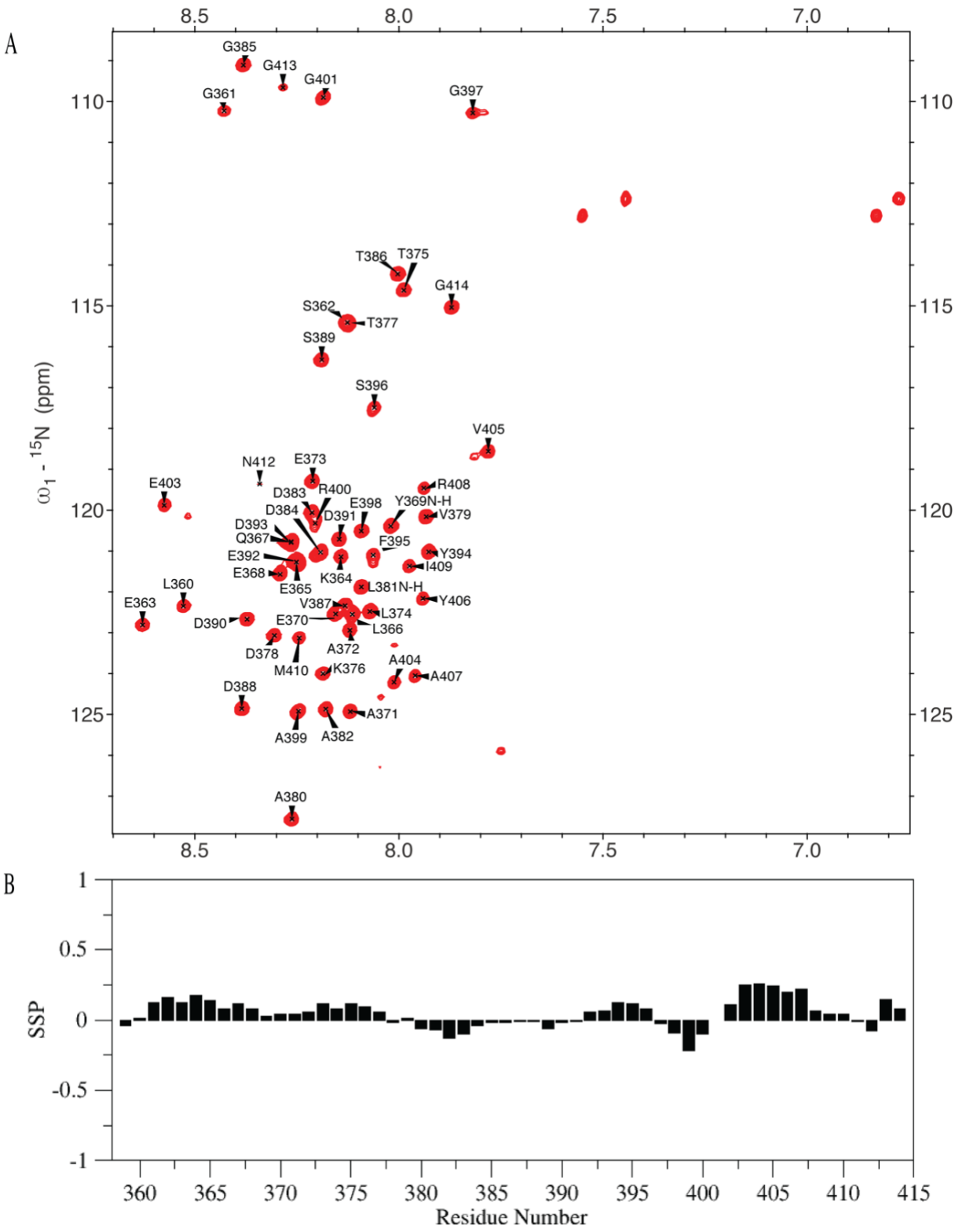


Figure 3-1. N-pep backbone assignment and secondary structure propensity analysis.

A. 700 MHz ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled N-pep at pH 6.8 and 25 °C. Assignment of N-pep backbone amide resonances are indicted by residue and number. Resonances from Q367 and N412 sidechain on the upper right spectra are not assigned.

B. Secondary structure propensity analysis of N-pep. Secondary structure was analysed by *SSP* [131] using ^{13}Cα and ^{13}Cβ chemical shift data. No patterns could be discerned from the N-pep chemical shifts, indicating the N-pep was predominantly unstructured.

3.3.2 Atomic level structural insights into the N-pep/P_{CTD} interaction

Data on interaction of N-RNA/P is available only from unvalidated modelling [83] and random mutagenesis [103], and from targeted mutagenesis of the P_{CTD} positive patch [96], which could involve indirect off-target effects. To directly assess binding sites in both N and P, as well as to gain insight into the binding kinetics, NMR chemical shift titration experiments were performed. Addition of an increasing amount of unlabelled N-pep into ^{15}N-labelled P_{CTD} resulted in a discrete perturbation (Figure 3-2). Most of these perturbed peaks shifted smoothly from the free state to the bound state (Figure 3-2A), suggesting that they are experiencing a fast exchange on the chemical shift timescale. Some of the signals which moved the most (e.g. F212, F215) broadened significantly with the addition of N-pep and then started to sharpen up again near the saturation point. In addition, some other resonances (e.g. Y213) broadened out and remained unobservable during the titration, indicating their limiting shifts are larger compared with the continuous shifts of other resonances and that they are in the moderately fast~intermediate exchange regime. The largest chemical shift perturbations in P_{CTD} were observed to be located within or close to a positive patch (K211, K212, K214, L224 and R260, figure 3-2C and 3-2E), and the most significant resonance broadening is also observed in this region, providing validation of random mutagenesis that identified predominant change in this region as affecting P_{CTD}/N binding [103], and similar specific site-directed mutagenic data

using MOKV [96]. In contrast, the W-hole (C261, W265 and M287) on the opposite face of P_{CTD} which was predicted based on random mutagenesis to be the second N-binding site [83, 103, 111] did not show any significant perturbation (CSP < 0.16 ppm, which lies within one standard deviation of the CSP of the remaining peaks, calculated at the last titration point of P_{CTD} : N-pep = 1 : 9 according to *Equation 1*). Importantly, the $N^{\varepsilon 1}/H^{\varepsilon 1}$ of the indole ring of W265 was not perturbed throughout the titration (Figure 3-2B). All observed shifts were in a linear manner, indicative of a single binding site with a single dissociation constant. Fitting the chemical shift differences of a number of well-resolved resonances showing obvious shifts throughout the titration to a single binding site model (Figure 3-2D), shows the affinity of N-pep for P_{CTD} is 234 $\pm$ 19 μM.

N-pep binding onto P_{CTD} has been directly assessed. To map the interaction site of P_{CTD} on N-pep, a complementary titration monitoring chemical shift changes in ^{15}N-labelled N-pep upon addition of unlabelled P_{CTD} was performed. Upon the addition of P_{CTD}, about 20 resonances within the flexible loop region (residue 375-395) were distinctly shifted (e.g. D388) and/or broadened (e.g. A380, A382) while the remaining 30 resonances were not affected (Figure 3-3A, 3-3B). Certain resonances showed small but significant shifts typical of the fast exchange regime (e.g. D390). In the presence of 2-fold molar excess of P_{CTD}, resonances from D378, A380, A382 and D383 broadened out and remained unobservable throughout the remainder of the titration. The signal intensity within the binding interface on N-pep decreased much faster compared to that of P_{CTD}, suggesting that N-pep may be undergoing a conformational transition to a folded form upon binding to P_{CTD}, and that the possible folding event occurs on the micro-millisecond time scale, based on the line broadening observed in the binding site of N-pep. The binding affinity obtained from the chemical shift perturbation is 224 $\pm$ 9 μM (Figure 3-3C), consistent with the previous titration result.

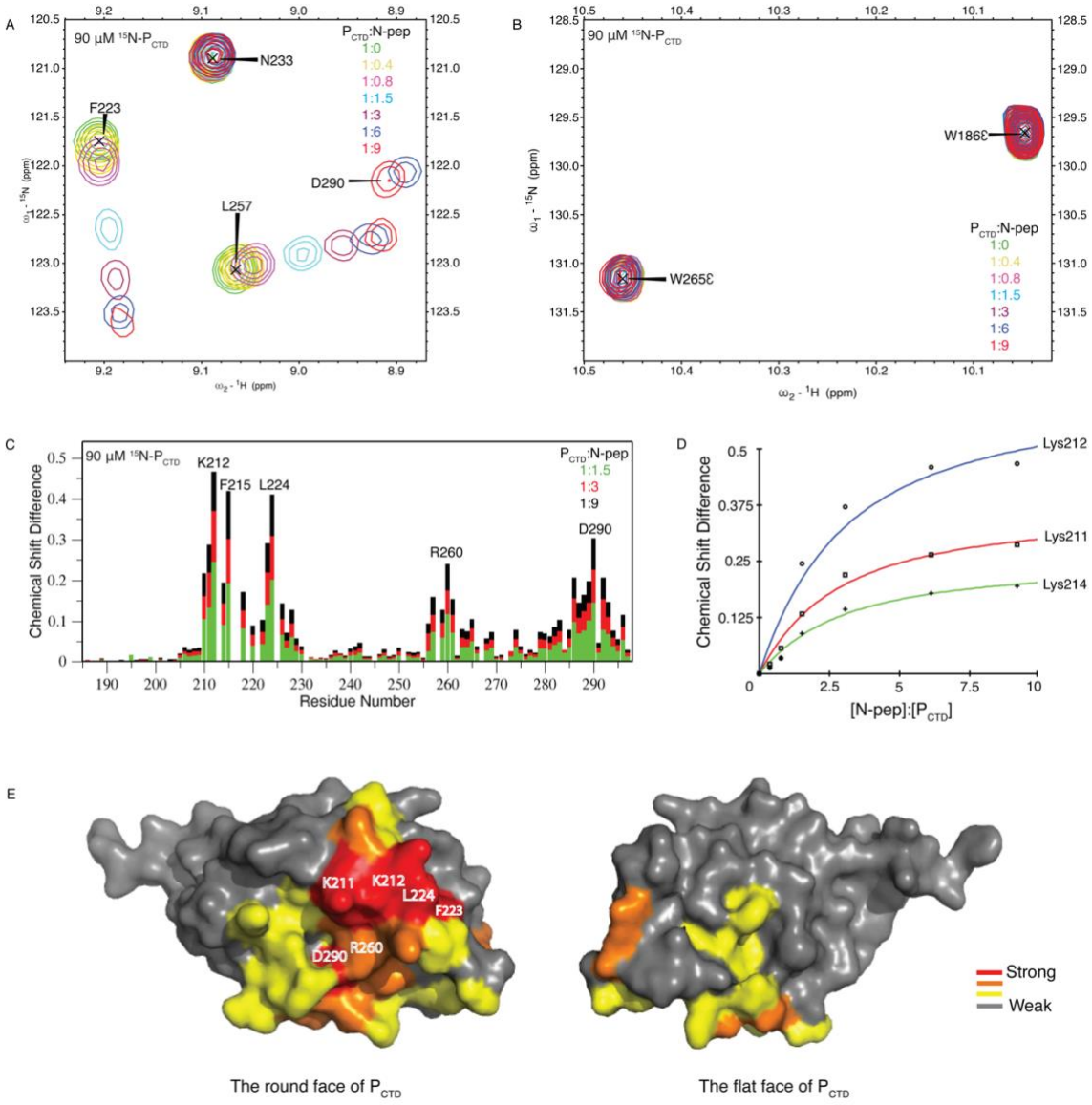


Figure 3-2. Titration of ^{15}N-labelled P_{CTD} with N-pep.

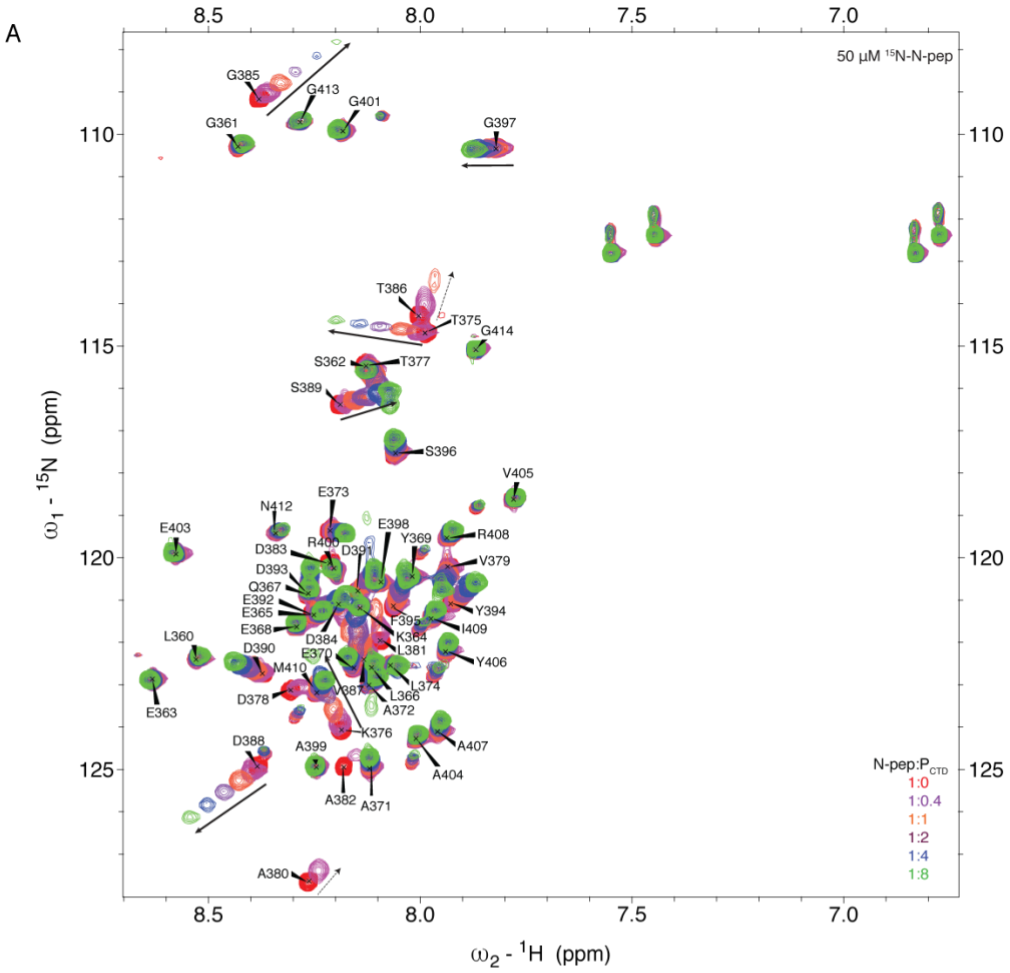
A. Segment from ^{15}N-P_{CTD}, ^{1}H-^{15}N HSQC titration experiment. 7 spectra were recorded at P_{CTD} : N-pep molar ratios of 1:0, 1:0.4, 1:0.8, 1:1.5, 1:3, 1:6 and 1:9. Experiments were conducted at pH 6.8 and 25 °C. Green colour indicated apo form of 90 μM of ^{15}N-P_{CTD}. Adding N-pep resulted in the chemical shift perturbation and reduction of intensity of F223 and L257, while N233 was not perturbed.

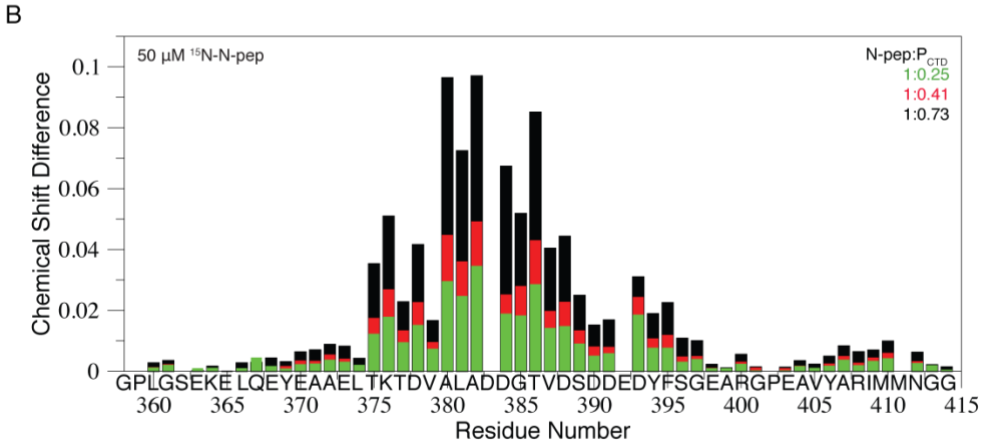
B. Segment from ^{15}N-P_{CTD}, ^{1}H-^{15}N HSQC titration experiment showing that the $N^{\varepsilon 1}/H^{\varepsilon 1}$ of W265 indole sidechain in the W-hole was not perturbed by the addition of N-pep throughout the titration.

C. Plot of P_{CTD} amide chemical shift changes. The green, red and black colour corresponds to the chemical shift changes at P_{CTD} : N-pep molar ratios of 1:1.5, 1:3, 1:9 respectively.

D. Single-site saturation binding curves (K_D = 234 $\pm$ 19 μM) for the residues K211, K212 and K214, which showed the largest chemical shift changes of all residues and remained resolved throughout the titration.

E. The round face (left) and the flat face (right) of P_{CTD}. The residues are coloured according to the chemical shift differences of the last titration point (red, two standard deviation; orange, one standard deviation; yellow, mean). Mapping the chemical shift changes onto the crystal structure of P_{CTD} indicated the positive patch on the round face of P_{CTD} forms the key binding site for N-pep.


A
50 μM ¹⁵N-N-pep
ω1 - ¹⁵N (ppm)
ω2 - ¹H (ppm)
N-pep:PCTD
1:0
1:0.4
1:1
1:2
1:4
1:8


B
50 μM ¹⁵N-N-pep
N-pep:PCTD
1:0.25
1:0.41
1:0.73
Chemical Shift Difference
Residue Number

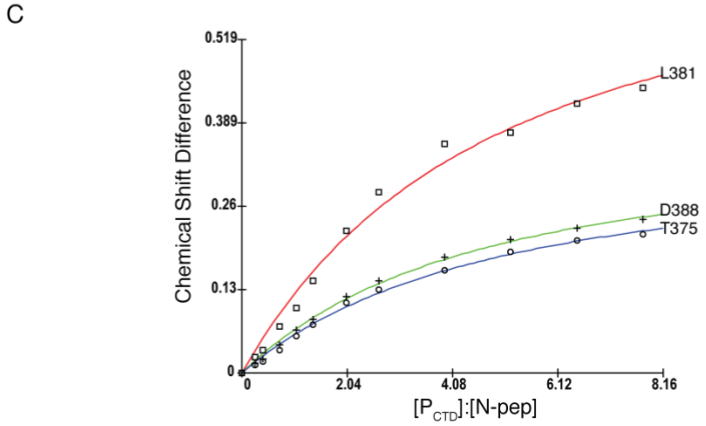

C
Chemical Shift Difference
L381
D388
T375
[PCTD]:[N-pep]

Figure 3-3. Titration of ^{15}N-labelled N-pep at 25 °C and pH 6.8 with unlabelled P_{CTD} up to 8-fold molar excess.

A. ^{1}H-^{15}N HSQC spectrum of ^{15}N-labelled N-pep showing chemical shift dependence on P_{CTD}. Arrows indicate the shift of resonances. Signals also showing significant broadening are marked with broken arrows.

B. Plot of the change in average $^{1}H^{N}$ and ^{15}N chemical shift. The green, red and black colour corresponds to the chemical shift changes at N-pep : P_{CTD} molar ratios of 1:0.25, 1:0.41, 1:0.73 respectively. The N-pep sequence is shown under the X-axis.

C. Single-site saturation binding curves ($K_D = 224 \pm 9$ μM) for the residues T375, L381 and D388.

3.3.3 Effect of cyclization of N-pep

The micromolar affinity observed in our N-pep/P_{CTD} titration experiments (see 3.3.2) is much weaker compared to the reported nanomolar affinity of N-RNA/P_{CTD} measured by surface plasmon resonance (SPR) [83], which could be partially due to lack of constraint at the termini of the flexible loop region in the N-pep construct. Models of the N-pep binding site for P_{CTD} suggest that it appears as a loop in the full-length N protein (Supplementary Figure 3-2), with residues T375 close to S396, which naturally proposes the idea of building in a disulphide bond (two Cysteine insertions, one between E370 and A371, and the other one between G401 and P402) to cyclize the N-pep, possibly mimicking a loop conformation of the P_{CTD} binding site in the N protein, and explore how cyclization would affect the interaction. A ^{1}H-^{15}N HSQC titration of ^{15}N-labelled cyclized N-pep with unlabelled P_{CTD} was performed as above, showing the perturbed peaks are experiencing two-state exchange mostly in an intermediate times-scale regime as indicated by the linearity of the peak shifts and linewidth changes (broadening and narrowing again, Supplementary Figure 3-3) during the titration. The dissociation constant determined from the saturation binding curves is 51 ± 6 μM (Table 3-1), which is almost

4.5-fold tighter than linear N-pep (224 ± 9 μM), implying that the conformation of the N_{CTD} loop in the full-length N protein is critical for its interaction with P_{CTD}. By cyclizing the N-pep through a disulphide bond, we perhaps are mimicking the conformation of the N_{CTD} loop region within the full-length N protein and enhancing its interaction with P_{CTD}.

Table 3-1. [&]Kinetics data of the N-pep variants binding to P_{CTD} determined by saturation binding curves (*Xcrvfit* [151]) and two-dimensional lineshape analysis (*TITAN* [153]).

	K_D (μM)	K_D (μM)	k_{off} (s^{-1})	k_{on} (s^{-1} μM^{-1})
Mutant	**Xcrvfit*		*#TITAN*	
N-pep (WT)	224 ± 9	215 ± 6	3200 ± 200	15
K376A	35 ± 3	32± 1	1800 ± 100	56
D378A	309 ± 10	298 ± 9	4000 ± 400	13
A380G	332 ± 12	325 ± 8	9000 ± 900	28
L381A	770 ± 34	785 ± 37	9000 ± 1600	11
A382G	300 ± 9	272 ± 8	7400 ± 1000	27
D383A	608 ± 46	603 ± 26	8300 ± 1400	14
D384A	546 ± 28	554 ± 22	7300 ± 1200	13
T386A	24 ± 1	16 ± 1	634 ± 15	40
D388A	323 ± 23	340 ± 10	3800 ± 200	11
S389E	88 ± 4	79 ± 3	2000 ± 100	25
S389A	98± 2	91 ± 2	3200 ± 200	35
D390A	100 ± 6	95± 1	4200 ± 200	44
D391A	160 ± 5	153 ± 3	2900 ± 200	19
E392A	180 ± 27	166 ± 3	3600 ± 200	22
D393A	88± 7	81 ± 4	2400 ± 400	30
T386A/S389E	11± 1	9.9 ± 0.3	640 ± 30	64
K376A/T386A/S389E	9.7 ± 0.9	8.9 ± 0.4	310 ± 20	35
Cyclized N-pep (WT)	51 ± 6	48 ± 2	720 ± 50	15
Cyclized S389E	19 ± 2	16 ± 1	550 ± 20	34
Cyclized-Phospho-N-pep	9.0 ± 0.4	7.4 ± 0.4	420 ± 20	57
Cyclized T386A/S389E	2.3 ± 0.2	2.4 ± 0.2	130 ± 10	54
Cyclized K376A/T386A/S389E	1.7 ± 0.1	2.4 ± 0.3	110 ± 10	46

*_Xcrvfit_: dissociation constant (K_D in μM) derived from _Xcrvfit_ software by fitting the chemical shift perturbation into a single binding site model. The chemical shift changes of at least 4 residues from one titration were used to calculate the overall K_D and standard deviation.

#_TITAN_: dissociation constant (K_D in μM) and off-rate (k_{off} in s^{-1}) were determined by simultaneously fitting the lineshape to a two-state single binding site model using _TITAN_ software. The on-rate (k_{on} in s^{-1} μM^{-1}) was determined by dividing the off-rate by the dissociation constant. The lineshape changes of at least 4 residues from one titration were used in _TITAN_ to calculate the overall K_D and off-rate.

&Kinetics data: The binding kinetics between each N-pep variant and P_{CTD} were calculated from a single set of ^{1}H-^{15}N HSQC titration experiments. Considering the errors in the protein concentration estimation using extinction coefficients, the actual uncertainty of the binding kinetics estimate may be larger than the table listed here.

NMR titration experiments contain rich information about the kinetic, structural and mechanistic aspects of the binding events. Apart from the chemical shift information used in saturation binding curves to extract the dissociation constant, the extensive linewidth changes caused upon complex formation can be analysed in _TITAN_ to extract information about the binding mode and detailed kinetics [153]. Using two-dimensional lineshape analysis of the wild type ^{15}N-labelled N-pep and P_{CTD} NMR titration data, the lineshape changes of residues G385, D388, D390, S396 and G397 of N-pep fitted well to a two-state binding model (Supplementary Figure 3-2), and yielded a dissociation constant of 215 ± 6 μM and an off-rate of 3200 ± 200 s^{-1}. Similarly, the lineshape analysis was performed on ^{15}N-labelled cyclized N-pep and P_{CTD} titration data (Table 3-1) and the K_D, k_{off} and k_{on} were determined. Compared with the linear wild type N-pep, cyclization improved the binding affinity 4.5-fold to 48 ± 2 μM by extending the complex lifetime (k_{off} = 720 ± 50 s^{-1}), while the association rate (k_{on}) remained the same. The dissociation constants calculated from the two-dimensional lineshape analysis for both the linear and cyclized N-pep are in good agreement with the

saturation binding curve analysis. Quality of the lineshape fitting of the experimental titration data to the two-state model was demonstrated in Supplementary Figure 3-4.

3.3.4 Identification of N-pep residues important for binding P_{CTD}

While mutagenic analysis of P_{CTD} from MOKV [96, 103] and RABV [102] has indicated potential residues on the N-RNA/P binding interface, similar targeted mutagenic analysis of MOKV N [96] failed to identify any critical residues – mutating individually or in combination the two highly conversed acidic regions (residues $_{373}$EE$_{374}$ and $_{390}$DEED$_{393}$) in MOKV N did not affect the N/P_{CTD} interaction [96]. To identify the specific residues of N that make direct contact with P_{CTD}, an alanine scan of the flexible loop region of N-pep was performed. Mutations included residues that shifted the most during the ^{15}N-labelled N-pep and P_{CTD} titration, the previously reported phosphorylation site S389 [79] and the negatively charged residues within the flexible loop region that could be making ionic interactions with the positive patch on P_{CTD}. Of special interest are D378, D383 and D384 which were predicted to be forming ionic bonds with K211, K212 and K214 of P_{CTD} respectively [83] but have not been tested empirically. The variant N-pep proteins were ^{15}N-labelled and purified, and their interactions with P_{CTD} studied using ^{1}H-^{15}N HSQC titration experiments, with k_{on}, k_{off} and K_D calculated using two-dimensional lineshape analysis as described above (Table 3-1). K_D values derived from saturation binding curves using *Xcrvfit* software were also included here to indicate the validity of the lineshape analysis.

Mutating L381, D383 or D384 to alanine caused significant reductions in binding with P_{CTD}. The dissociation constant was at least 2.5-fold weaker compared with the wild type N-pep, due to the much more rapid dissociation. Very limited changes in peak intensity or chemical shifts were observed during the titration (Supplementary Figure 3-3), suggesting L381 makes crucial

hydrophobic interactions and D383 and D384 are essential for electrostatic interactions with the positive patch on the P_{CTD}. L381 is first identified here, and indicates that in addition to the electrostatic interactions [96, 103], there are hydrophobic interactions between N-pep/P_{CTD}. D383 and D384 were predicted to be involved in P_{CTD} binding in the SAXS model [83], but have not been experimentally investigated. Mutating D378, A380, A382 and D388 had little effect suggesting that these residues were not making essential contacts with P_{CTD}. Notably, while the importance of L381 has not been predicted previously, the non-critical role of D378 identified here contradicted the predictions made in the model [83].

The phosphorylation of N protein has been reported to modulate viral transcription and replication [80] by enhancing the interaction between P protein and the viral nucleocapsid [81]. The phosphorylation site has been mapped to S389, with phosphorylation catalysed by cellular casein kinase II (CK-II) [79]. To investigate the effect of phosphorylation on N/P interaction, a phosphomimetic mutant (S389E), phosphorylation-deficient mutant (S389A) and *in vitro* CK-II phosphorylated N-pep (Supplementary Figure 3-5) were produced and ^{15}N-labelled, before titration with unlabelled P_{CTD}. Phosphorylation of N-pep improved the binding affinity almost 7-fold compared with the unphosphorylated N-pep through more rapid association and extended complex lifetime, indicated by a faster on-rate and a slower off-rate. Possibly the long-range electrostatic steering between the phosphate group and P_{CTD} accelerated the complex formation after a diffusive encounter [155]. Also due to the extra negative charge introduced onto N-pep, more ionic interactions would be expected to be formed, making the complex more stable with a slower off-rate. These interactions are likely to be formed between the phosphate group of S389 of N-pep and K211, K256, R249 and R293 of P_{CTD}, according to predictions of the SAXS model [83]. Phosphomimetic N-pep（S389E） bound to P_{CTD} 3-fold more strongly than wild type N-pep, but more weakly than phosphorylated N-pep. The kinetic

differences between unphosphorylated N-pep and phosphorylated or phosphomimetic mutant of N-pep agreed well with previous mammalian cell-based studies performed with wild type N, and mutated phosphorylation-deficient (S389A or S389G, mutants lacking the phosphorylation site) or phosphorylation-mimetic (S389E, S389D) N, and phosphatase-treated N, where immunoprecipitation assays, minigenome system and full-length infectious virus [80, 81] indicated that S389 phosphorylation of N is important to the formation of the viral replication complex, enabling transcription and replication.

Unexpectedly, the phosphorylation-deficient N-pep (S389A) also bound to P_{CTD} 2-fold more tightly than unphosphorylated N-pep, with a much more rapid association rate but similar complex lifetime. Several other alanine substitutions were also found to bind to P_{CTD} with an accelerated association rate and/or extended complex lifetime and, therefore, a tighter affinity (e.g. K376A, T386A, D390A, D393A). The reason behind the faster on-rate effected by these mutations is unclear, but may be due to removal of certain steric hindrance during the complex formation process, or prevention of non-specific interactions that could slow down the diffusion and, consequently, the on-rate [156]. Alternatively, the alanine substitutions might selectively promote or stabilize a favourable transient conformation of free N-pep that enables formation of the "folded" final complex [112, 157-159]. Apart from the faster on-rate shown by these alanine substitutions, the alanine substitution at T386 also extended complex lifetime as indicated by the slower off-rate, probably due to more productive hydrophobic interactions with P_{CTD}.

3.3.5 Improving the affinity of N-pep for P_{CTD}

To solve the N-pep/P_{CTD} complex structure by X-ray crystallography, we need to obtain a stable complex. Crystallography has solved structures of even weak binding complexes [160,

161], although ideally, strong affinity is preferred. With the object of improving N-pep/P_{CTD} affinity, we combined cyclization and mutations of N-pep that showed stronger affinity for P_{CTD}, generating linear and cyclized double mutant (T386A/S389E) and triple mutant (K376A/T386A/S389E) N-pep. Interaction of these variants with P_{CTD} was characterized using NMR titration experiments and isothermal titration calorimetry (ITC). ITC showed wild type linear N-pep bound P_{CTD} with a K_D of about 267 μM (K = 3.7E3 ± 2E2 M^{-1}), and the affinity increased to 44 μM (association constant K = 2.3E4 ± 2.0E3 M^{-1}) after cyclization (Table 3-2, Supplementary Figure 3-6), in agreement with the K_D derived from saturation binding curves and lineshape analysis from NMR titration experiments (Table 3-1). Combining the strong affinity N-pep mutations yielded linear N-pep variants with even stronger affinity for P_{CTD}, with a K_D of 13 μM (K = 7.5E4 ± 1.1E4 M^{-1}) for the double mutant (T386A/S389E) N-pep and 7.1 μM (K = 1.4E5 ± 7.9E3 M^{-1}) for the triple mutant (K376A/T386A/S389E) N-pep. Cyclization of the mutated N-pep further improved the affinity to 3.0 μM (K = 3.3E5 ± 3.2E4 M^{-1}) and 1.2 μM (K = 8.2E5 ± 3.0E5 M^{-1}) for the double and triple mutants respectively, determined by NMR titrations (Table 3-1) and ITC experiments (Table 3-2). The estimates for the model parameters from ITC confirmed a 1:1 stoichiometry for all of the P_{CTD}/N-pep variant interactions, in accordance with NMR saturation binding curve results.

ITC has the added benefit of directly measuring the change in enthalpy (ΔH) of binding and, from this, the entropy ($T\Delta S$) and free energy (ΔG) can be calculated (Table 3-2). The enthalpy change is the overall result of building and interrupting interactions between the binding partners and the solvent, and could make favourable or unfavourable contribution to the binding free energy. The entropy change in a binding event could arise from three aspects: the unfavourable loss of the overall translational and rotational entropy because of complex formation, favourable entropy contribution from the solvent at the binding interface being

released upon binding, and the uncertain conformational entropy changes of the binding partners. Overall, the binding of all N-pep variants and P_{CTD} was solely driven by favorable enthalpy contributions, which overcompensated the small unfavourable entropic ($T\Delta S$) contribution likely associated with the disorder-to-order transition of N-pep upon binding, therefore dominating the coupled folding and binding mechanism [162].

Compared with the corresponding linear N-pep variants, cyclization of N-pep resulted in enthalpy gains on binding (Table 3-2) ($\Delta H_{cyclized} - \Delta H_{linear}$, WT: -1.2, double mutant: -2.8, triple mutant: -4.6 kcal/mol) but were accompanied by increasing entropic penalties ($T\Delta S_{cyclized} - T\Delta S_{linear}$, WT: -0.2, double mutant: -2.0, triple mutant: -3.6 kcal/mol). Nevertheless, lower free energy of binding is observed for all cyclized variants ($\Delta G_{cyclized} - \Delta G_{linear}$, WT: -1.0, double mutant: -0.8, triple mutant: -1.0 kcal/mol) and enhanced N-pep/P_{CTD} interaction.

The thermodynamic reasons behind the stronger affinities of the mutants are more complicated and varies between the different mutations introduced. T386A/S389E was enthalpically less favourable, suggesting a loss of binding interactions ($\Delta H_{T386A/S389E} - \Delta H_{WT}$, linear version: 2.2, cyclized: 0.6 kcal/mol). However, this loss was compensated by more favourable entropic contributions ($T\Delta S_{T386A/S389E} - T\Delta S_{WT}$, linear version: 4.0, cyclized: 2.2 kcal/mol) and thus yielded a lower binding free energy ($\Delta G_{T386A/S389E} - \Delta G_{WT}$, linear version: -1.8, cyclized: -1.6 kcal/mol). While showing stronger affinity than all other N-pep variants, the introduction of the three mutations K376A/T386A/S389E resulted in slightly poorer enthalpy compared to wild type for the linear N-pep ($\Delta H_{K376A/T386A/S389E} - \Delta H_{WT}$ = 1.0 kcal/mol) which was well compensated for by favourable entropy ($T\Delta S_{K376A/T386A/S389E} - T\Delta S_{WT}$ = 3.2 kcal/mol). There was a marked improvement in enthalpy on cyclization ($\Delta H_{K376A/T386A/S389E,\ cyclized} - \Delta H_{WT,\ cyclized}$ = -2.4 kcal/mol) and only a small change in entropy ($T\Delta S_{K376A/T386A/S389E,\ cyclized}$ - $T\Delta S_{WT,}$

$_{cyclized}$ = 0.2 kcal/mol). The triple mutant, in comparison to the double mutant showed improved enthalpy ($\Delta H_{K376A/T386A/S389E} - \Delta H_{T386A/S389E,}$ linear version: -1.2, cyclized: -3.0 kcal/mol) which was large enough to overcompensate for the unfavourable entropic penalty ($T\Delta S_{K376A/T386A/S389E} - T\Delta S_{T386A/S389E}$, linear version: -0.8, cyclized: -2.4 kcal/mol).

Interestingly, the enthalpy-entropy compensation between these N-pep variants show a linear relationship (Supplementary Figure 3-7), which has been observed and reported for other protein interactions, but the mechanism behind this physical phenomenon is still under discussion [163, 164]. Nevertheless, by combining mutations and cyclization, the N-pep/P_{CTD} binding affinity has been improved 200-fold to 1.2 μM, and the affinity determined by ITC remained strong between pH 5~7 (Table 3-2).

Table 3-2. The kinetics and energetics of the interaction between N-pep variants and P_{CTD} derived from ITC studies.

	pH	Stoichiometry N	K (M^{-1})	K_D (μM)	ΔH (kcal/mol)	ΔS (cal/mol/deg)	$T\Delta S$ (kcal/mol)	ΔG (kcal/mol)
N-pep (WT)	6.8	1	3.7E3 ± 2E2	267	-12.2 ± 0.5	-24.6	-7.3	-4.9
Cyclized WT	6.8	1	2.3E4 ± 2.0E3	44	-13.4 ± 0.5	-25.0	-7.5	-5.9
T386A/S389E	6.8	1.32 ± 0.03	7.5E4 ± 1.1E4	13	-10.0 ± 0.4	-11.2	-3.3	-6.7
Cyclized T386A/S389E	6.8	0.98 ± 0.01	3.3E5 ± 3.2E4	3.0	-12.8 ± 0.2	-17.7	-5.3	-7.5
K376A/T386A/S389E	6.8	1.13 ± 0.01	1.4E5 ± 7.9E3	7.1	-11.2 ± 0.1	-13.9	-4.1	-7.1
Cyclized K376A/T386A/S389E	6.8	0.93 ± 0.04	8.2E5 ± 3.0E5	1.2	-15. 8 ± 1.0	-25.9	-7.7	-8.1
	6	0.94 ± 0.02	6.6E5 ± 8.9E4	1.5	-16.1 ± 0.4	-27.4	-8.2	-7.9
	5	0.92 ± 0.02	5.7E5 ± 7.0E4	1.8	-14.6 ± 0.5	-22.7	-6.8	-7.8

3.3.6 Evidence that the highest affinity binder for P_{CTD} adopts a transient secondary structure

To gain more insight into the structural property of the highest affinity binder for P_{CTD} and the possible reasons behind the 200-fold affinity improvement, the backbone resonances of the cyclized triple mutant (K376A/T386A/S389E) N-pep variant were assigned using triple

resonance experiments and secondary structure propensity analysed using the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts by *SSP* [131].

In contrast to the predominately unstructured wild type N-pep (Figure 3-4B), the cyclized triple mutant showed a tendency to form secondary structure from residues A381-E392 and P404-G415 (corresponds to A380-E391 and P402-G413 in the wild type N, as there are two cysteine insertions, one between E370 and A371, and the other one between G401 and P402). What we observed was, as the primary binding site for P_{CTD} identified in this study (see 3.3.2), the flexible loop region (residues A381-E392) showed a β-strand propensity, as indicated by the continuous negative values up to -0.3 in this region (Figure 3-4A). The terminal part P404-G415 exhibited a strong helical propensity up to 43%, which resembles the actual helical structure of this region (α14, P402-M411) in the context of full-length N [64, 72].

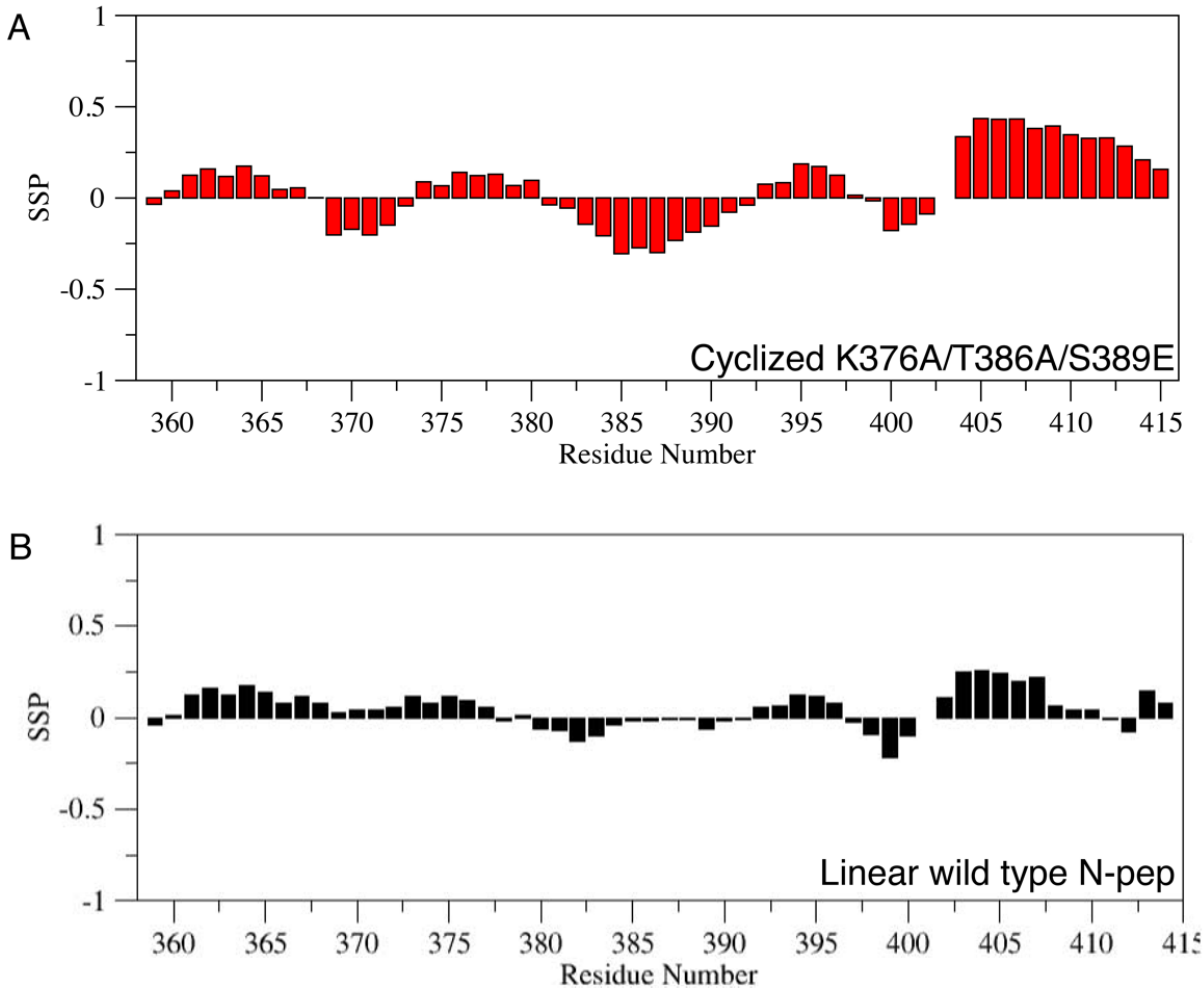


Figure 3-4. Secondary structure propensity analysis of the cyclized triple mutant K376A/T386A/S389E.

The wild type N-pep were shown in (B) in black for comparison. The cyclized triple mutant (A) might adopt a transient β-strand structure from residue A381 to E392 which corresponds to A380 to E391 in the wild type N, and a transient α helical structure from P404 to G415 (P402-G413 in the wild type N). The artificially introduced loop was formed by a disulphide bond between the two Cysteine insertions, one between E370 and A371, and the other one between G401 and P402.

The mechanism behind the 200-fold affinity improvement upon cyclization and introduction of mutation is unclear. However, the higher secondary structure propensity shown within the cyclized triple mutant may be an important contributor. In the N-pep/P_{CTD} interaction, the

extensive resonance broadening observed within the flexible loop region of N-pep during the titration and the coupled folding and binding mechanism suggested by ITC both indicate that the flexible loop region of N might go through a conformational change upon binding to P_{CTD}. Notably, the N-pep-binding interface on P_{CTD}, the positive patch ($_{211}$KKYK$_{214}$, L224, R260), is mostly clustered on or around an anti-parallel β-sheet formed by two β-strands. Possibly upon interaction with P_{CTD} the transient β-strand structure in the N_{CTD} loop is involved in formation and stabilization of the complex via a mechanism of β-strand addition [165].

The strong helical propensity of the terminal part of N-pep (α14, resides P402-M411 in wild type N) may also contribute to the enhanced affinity by allowing the formation of salt bridges between α14 of N-pep and α6 of P_{CTD}. In the N-RNA/P_{CTD} model, the α14 of N was predicted to be packed against the α6 (residues 279-297) of P_{CTD}, with E403 and R408 of N making ionic interactions with K285 and D289 of P_{CTD} [83], although this prediction has not been experimentally assessed. In our data of ^{1}H-^{15}N HSQC titration of wild type ^{15}N-labelled N-pep and P_{CTD}, E403, R408 and the neighboring region were clearly not perturbed (Figure 3-3B, see 3.3.2). In the reciprocal titration (Figure 3-2C, 3-2E), only D289 of P_{CTD} was moderately perturbed (CSP = 0.20 ppm, lies beyond one standard deviation but within two standard deviations, calculated at the last titration point of P_{CTD} : N-pep = 1 : 9 according to *Equation 1*), which could be due to the influence of D290 as its intramolecular hydrogen bond with R260 of P_{CTD} was interrupted by R260 interacting with N-pep. Although we found no strong evidence of the interaction between E403 and R408 of N-pep and K285 and D289 of P_{CTD}, these salt bridges might be present in the N-RNA/P_{CTD} complex when the α14 of N is properly folded and a stable binding interface is formed. In the linear wild type N-pep segment, the lack of a stable helical structure within this region (residues P402-M413, Figure 3-1B) could hinder this potential ionic interaction. Here, the greater helical propensity of this region in the cyclized

triple mutant N-pep might enhance the N-pep/P_{CTD} interaction by allowing the formation of the two salt bridges between α14 of N and α6 of P_{CTD}.

3.4 Discussion

From the previously published studies, the P protein binding interface on rabies virus nucleocapsid has been narrowed down to the C-terminal region of N protein (residues 377–450) using limited proteolysis [73]. Based on SAXS and other biochemical data, a N-RNA/P_{CTD} model was proposed with the N_{CTD} loop binding to the positive patch and W-hole of P_{CTD}, and detailed interactions were predicted [83]. However, the predictions in this SAXS model may be flawed, as the two binding sites used in the SAXS modelling process came from a random mutagenic study [103]. In this latter study, P_{CTD} constructs with defective N interaction containing W-hole mutations were identified, but these were rare, and were always in proteins containing multiple substitutions, including mutations in the positive patch [103] and mutations disabling the core packing [96], providing little genuine support for the roles of the W-hole region. By contrast, mutations in the positive patch were enriched in proteins defective for N interaction, supporting a role in N-binding. Importantly, targeted W-hole mutations were subsequently found to have little impact on N-binding or viral replication in mammalian cell-based assays [102], whereas positive patch mutations indicated critical roles [96, 102]. Nevertheless, these assays did not discount possible off-target effects of positive patch mutations. In addition, this model lacked predictions of roles for several additional residues within the positive patch and N_{CTD} loop subsequently implicated by targeted mutagenesis [96] and our study.

3.4.1 N-pep/P_{CTD} interaction is governed by electrostatic and hydrophobic forces

Here, by combining NMR titration and mutagenesis data, we mapped the binding interfaces on N and P proteins and successfully defined the specific N residues involved in the interaction for the first time. As predicted in the SAXS model [83], RABV P_{CTD} structure [111], targeted mutagenesis of MOKV P [96] and RABV P [102], and random mutagenic study of MOKV P [103], residues of the positive patch (K211, K212, K214, L224 and R260) on P_{CTD} formed the key binding site for the N_{CTD} loop. Importantly, our data indicated an interaction directly by monitoring the involvement of the backbones in the interaction, discounting possible off-target effects in the mutagenic studies. D383, D384 and phosphorylated S389 of N made electrostatic interactions with the positively charged residues on $P_{CTD,}$ which validated the prediction in the SAXS model [83] and is consistent with the previous literature reporting that the phosphorylation state of N is important for P-binding and replication [79-81], while two other highly conserved acidic regions in the vicinity of S389 (residues $_{373}$EE$_{374}$ and $_{390}$DEED$_{393}$ of MOKV N) are not involved, consistent with mutagenic data [96]. Notably, L381 of N also appeared to be vital and likely makes a crucial hydrophobic interaction with L224 of P_{CTD}, the importance of which has not previously been shown. On the other hand, in contrast to the prediction that D378 of N interacts with K214 of P_{CTD} [83], we found that D378 was dispensable for the interaction. Consistent with critical roles in replication, all the residues identified within the N-pep/P_{CTD} interface are conserved among lyssaviruses (Supplementary Figure 3-9). Experiments monitoring the sidechain resonances such as $^{1}H^{-13}C$ HSQC titrations could be used to further validate the involvement of specific residues in the interaction.

E403 and R408 of N were also predicted to form salt bridges with K285 and D289 of P_{CTD} with the α14 (residues 402-413) of N packed against the α6 (residues 279-297) of P_{CTD} in the SAXS model, but this prediction has not been experimentally assessed [83]. In our study, we found no

strong evidence of the formation of such salt bridges, as in the ^{1}H-^{15}N HSQC titration of wild type N-pep with P_{CTD}, E403, R408 and the neighboring region of N were clearly not perturbed (Figure 3-3B), and the moderate perturbation of D289 of P_{CTD} may be an indirect influence from the perturbation in the positive patch (see 3.3.2 and 3.3.6). Nevertheless, these salt bridges might be present in the N-RNA/P_{CTD} complex and in the cyclized triple mutant N-pep/P_{CTD} complex when the α14 of N is properly or at least partially folded and a stable binding interface is formed. In the wild type linear N-pep segment, the lack of a stable helical structure within this region (residues P402-M413, Figure 3-1B) could hinder this potential ionic interaction.

Apart from the positive patch, the W-hole (C261, W265 and M287) on P_{CTD} was predicted to be the second binding site for N-RNA complex [83, 103, 111] based on a random mutagenic study of P_{CTD} [103], although there is no direct mutagenic or structural data supporting this. The idea that W-hole is important is based on a random mutagenic study which results in multiple mutations including enriched changes at the positive patch [103], and could have off-target effects. In the NMR titration of ^{15}N-labelled P_{CTD} with N-pep, only residues close to the positive patch on the round face of P_{CTD} were perturbed and the W-hole residues on the flat face were clearly not affected. Importantly, the $N^{\varepsilon 1}/H^{\varepsilon 1}$ peak of W265 indole side chain which serves as a useful probe for the W-hole was not affected in the titration with N-pep. Throughout our study of the interaction between P_{CTD} and all the N-pep variants, only one binding site (the positive patch) was discovered from NMR titrations and ITC experiments. In agreement with lack of major role in N-RNA/P binding, mammalian cell-based assays previously showed that W-hole mutations did not impair the N-binding or the genome replication functions of P protein [102]. And the W-hole residues, especially W265, appear to be poorly conserved among lyssaviruses (Supplementary Figure 3-9B). A possible scenario is that the affinity of the

W-hole as the second binding interface is extremely weak (e.g. K_D > 1 mM), which may not be distinguishable in the presence of the μM affinity with the positive patch. Or the effects of mutations used to implicate W-hole in RABV P_{CTD} structure [111] and the SAXS model [83] were actually due to simultaneous mutations at other sites, including the positive patch. Hence, no clear direct data supporting the W-hole in N-binding have been found so far, which is consistent with our findings. Taken together, the positive patch of P_{CTD} is the primary N-pep binding site, and the W-hole does not play an important role in the interaction.

3.4.2 N-pep/P_{CTD} form a relatively weak-affinity complex

P protein attaches the polymerase L to the N-RNA template and the interaction requires the phosphorylation of S389 on N protein [79-81]. We found phosphorylation of S389 did enhance the N-pep/P_{CTD} interaction about 7-fold through more rapid association and a longer complex lifetime. However, the low micromolar affinity between the CK-II phosphorylated cyclized N-pep and P_{CTD} (7.4 ± 0.4 μM) obtained from our study differs to the nanomolar affinity between N-RNA ring and P_{CTD} (160 ± 20 nM) reported from the Jamin group [83]. This discrepancy might arise from the major difference in the length of the reactant and the technique applied. Here, instead of the 550 kDa N_{10}-RNA ring used in the SPR study in the Jamin group, a 57 amino acid segment from the N_{CTD} flexible loop region was used in our solution NMR and ITC study, which might partially account for the almost 50-fold difference in affinity. Also, if the W-hole or the K285 and D289 of α6 of P_{CTD} do play a role in N-RNA/P interaction as predicted [83] even with a very weak affinity in the context of full-length N protein, the avidity would lead to a much tighter interaction as well. We also tried to express and purify N-terminal His-tagged full-length N protein in *E. coli* BL21(DE3) bacteria but failed. Similar bacterial expression of N has been attempted and failed previously probably due to the lack of phosphorylation in *E. coli* such that N is unable to bind RNA or form an ordered

N-RNA ring [72, 82]. Currently the full-length N proteins used in the published studies are mostly derived from insect cells [64, 72, 73, 83].

P protein self-assembles both *in vitro* and *in vivo* through its central dimerization domain (residues 95-131, see supplementary figure 3-1) to form a stable parallel dimer P_2 [87]. If both P_{CTD} within P_2 interact with N-RNA simultaneously with a low micromolar affinity, the overall affinity between P_2 and N-RNA would be expected to be enhanced dramatically. However, this appears not to be the case. Previous SEC/MALLS/RI (size-exclusion chromatography/multi-angel laser light scattering/refractometry) titration experiments [83] showed that a circular N-RNA ring with 9-12 N protomers can only accommodate two P_2 possibly due to steric hindrance, with each P_2 binding via a single P_{CTD}. The viral helical nucleocapsid (N-RNA) has ~53 N protomers per turn when packed with other viral proteins in the virion, or ~15 N protomers per turn when it is released [71, 73, 166] as the active template for replication, which provides a wider angle to possibly accommodate more P_{CTD}. Nevertheless, it has been shown that P mutant without the dimerization domain remains to be highly functional in transcription [103], indicating that such high affinity is not necessary for transcription, and that the transcription/replication complex is based on a lower affinity interaction, more consistent with that derived from our study. For example, N and P proteins have been shown to accumulate within viral replication factories, forming cytoplasmic inclusions (Negri bodies-like structures) with properties of liquid organelles [67]; the latter suggests a weak and dynamic interaction, and is not compatible with the strong affinity reported by SPR (160 ± 20 nM) [83] or the possibly even stronger affinity when P_2 interacts with N-RNA via both P_{CTD} simultaneously.

Considering the diverse binding partners of P_{CTD} and multiple roles played by P in the RABV life cycle, in contrast to the relatively low abundance of P in the host cells compared to N and most likely other binding partners, it would be unlikely that P and N-RNA form a very strong and stable complex. The fact that we were able to find multiple variants of N-pep with higher affinity to P_{CTD} also suggested that too tight an interaction with N-RNA may not be favoured in the viral life cycle, and might even impose constraints in the viral replication process.

3.4.3 Evidence that N_{CTD} flexible loop may fold upon binding to P_{CTD}

The affinity between N-pep and P_{CTD} was improved 200-fold through cyclization and introduction of mutations that further improved affinity. Cyclized wild type N-pep exhibited 4.5-fold stronger affinity towards P_{CTD} compared with when linear, due to the slower off-rate and consequently extended complex lifetime, which could be explained by the likely contribution of pre-organizing and constraining N-pep into a P_{CTD} binding conformation via cyclization. Interestingly, ITC data indicated that cyclization resulted in mostly enthalpy gains on binding compared to linear N-pep while the entropy change is minor, which also suggests that more productive interactions were made so that the N-pep/P_{CTD} complex was stabilized.

The three single mutants K376A, T386A and S389E all exhibited faster on-rate and slower off-rate compared to the linear N-pep (Table 3-1), although K376A influenced on-rate more, whereas T386A contributed more to extending complex lifetime. The combined effect could be complicated, and showed overall enhanced affinity through more rapid association and slower dissociation. Unexpectedly, ITC analysis of T386A/S389E and K376A/T386A/S389E indicated that the combined double or triple mutant was enthalpically less favourable for N-pep/P_{CTD} affinity, suggesting a loss of binding interactions; the enhanced affinity of the double or triple mutant was achieved by more favourable entropic contributions.

Conformational changes within the binding site could weaken the affinity since a portion of the intrinsic binding free energy is consumed in converting the protein to the conformation which is competent for binding [167]. The ITC study revealed that the N-pep/P_{CTD} interaction was solely driven by favorable enthalpy contributions, overcompensating the unfavourable entropic penalty ($T\Delta S$), which in part can be rationalized by the decrease in conformational entropy of N-pep on obtaining a level of order within the complex [168]. This disordered-to-ordered conformational transition of N-pep is also indicated by the extensive broadening of N-pep peaks in the NMR titration and more dispersed N-pep spectra in the bound state (Supplementary Figure 3-8). If we take the ITC result to revisit the possible reasons behind the enhanced affinity between certain N-pep variants and P_{CTD} (see 3.3.4), the most likely scenario would be that the strong affinity mutations selectively induce or stabilize a transient conformation of free N-pep which could be involved in complex formation and then trapped as the bound state; thus the entropic penalty during the binding process is reduced, resulting in a significant affinity increase [168].

The formation of a more ordered complex results in a favorable entropy gain but also accompanies a loss of conformational entropy of the binding partners, or at least the N-pep, which lays the basis of the enthalpy-entropy compensation relationship [168]. As the binding event can be viewed as replacing the water molecules in or around the active sites with the ligand, the water network also contributes to and complicates the enthalpy and entropy change in the binding events [163, 169]. The enthalpy-entropy compensation between the N-pep variants shows a linear relationship except for the cyclized N-pep, which is an outlier (Supplementary Figure 3-7). Due to the limitation of sample concentration and the weak binding affinity, a complete binding curve for the linear and cyclized form of wild type N-pep

could not be obtained (Supplementary Figure 3-6), which may affect the accuracy of the estimation of enthalpic and entropic contributions to the interaction.

3.4.4 Previous findings supporting the possible folding event of the N_{CTD} loop

As discussed above, NMR and ITC studies suggested that the flexible loop region of N-pep might go through conformational change and become folded upon binding with P_{CTD}. The secondary structure propensity analysis of the highest affinity binder for P_{CTD} (see 3.3.6) further indicated that the N_{CTD} loop might adopt a β-strand conformation and bind to the anti-parallel β-sheet of P_{CTD} through the β-addition mechanism [165]. A conformational change of the N_{CTD} loop region of the N-RNA complex upon P binding was also suggested by analysis using a monoclonal antibody [170]. The antibody against N was found to exclusively bind to N in the context of an N-RNA complex with P, with the epitope mapped to the N_{CTD} loop region (residues 388-407). This linear epitope within N was not recognized by the antibody when P was removed from the N-RNA complex by chemical treatment, or when N existed in the RNA-free form as N^0 protein in complex with P (N^0/P complex) [170]. Thus, P might be necessary in keeping the N_{CTD} loop within N-RNA in the correct conformation for binding by this antibody.

Although there is very little sequence similarity among Mononegaviruses and the nature of the N-RNA/P interaction varies from electrostatic forces-dominant [157] to hydrophobic forces-dominant [171], certain levels of structural conservation of P_{CTD} are observed [95], and a similar folding event for the N proteins has been reported. The C-terminal disordered region of N protein of the Paramyxoviruses, Measles virus [172], Hendra virus [158] and Sendai virus [157] adopt an α-helical structure upon binding to the helices of the C-terminal domain of P protein. N_{CTD} disordered loop of Vesicular stomatitis virus [112] of the *Rhabdoviridae* family

became ordered, although not folded, upon binding to P protein. P binds to N-RNA complex as a polymerase cofactor to enable access of the polymerase to the viral genome RNA encapsidated by N. The conformational change of the N_{CTD} flexible loop region within N-RNA complex upon binding with P could be a signal that is further transferred from N_{CTD} to N_{NTD}, leading to the opening of the N_{NTD}/N_{CTD} clamp and the release of the encapsidated RNA, enabling access of the polymerase L to the viral genome [64].

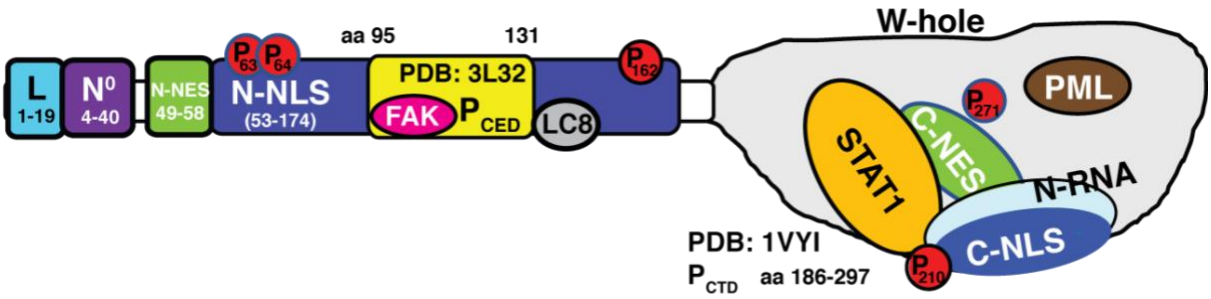
Definite demonstration of this conformational change of the N_{CTD} loop in RABV will ideally require more direct structural and dynamic evidence of the N-pep/P_{CTD} complex. The N-pep variants which bind to P_{CTD} with low micromolar affinity identified in this study could enable elucidation of the complex structure and how it forms by X-ray crystallography and NMR, which would also provide valuable information for structure-based drug discovery and anti-viral drug design. Unfortunately, our attempt to crystallize several N-pep variants with P_{CTD} by either co-crystallisation or soaking N-pep into P_{CTD} crystals failed (Supplementary Figure 3-10). NMR structural studies to solve the P_{CTD}/N-pep (linear triple mutant) complex structure have also failed because of significant broadening of a large portion of the C^{α}/H^{α} peaks of $^{13}C,^{15}N$-labelled P_{CTD} saturated with unlabelled triple mutant N-pep in aliphatic ^{1}H-^{13}C HSQC spectra, and a lack of resonances of unlabelled triple mutant N-pep from the 2D $^{13}C,^{15}N$-filtered NOESY experiment due to line-broadening. Given the various N-pep/P_{CTD} complexes with different affinities and off-rates engineered in this study, it would still be possible to collect NOE data to determine the complex structure using the suitable N-pep/P_{CTD} complex. Alternatively, with the information of the N-pep/ P_{CTD} interface and binding residues identified in this study, a P_{CTD}/N complex model could be predicted using HADDOCK [160].

3.4.5 The replication machinery

The replication/transcription mechanism within the *Mononegavirales* order has been widely studied and a "cartwheel" model has been proposed for Sendai virus [138, 157, 173] from the family *Paramyxoviridae* and Vesicular stomatitis virus [112] from the family *Rhabdoviridae.* In this model, P protein binds to the N-RNA template with a weak affinity, and thus can continuously bind and release the N-RNA template, carrying the polymerase (L) along the helical nucleocapsid. Consistent with our findings, the dissociation constant of the C-terminal domains of P and N proteins of Sendai virus was in the low micromolar range determined by saturation binding curve analysis of NMR titrations. In contrast, a "jumping" model has been proposed for Measles virus [174] and Rabies virus [83] based on a high nanomolar affinity of the C-terminal domains of P and N proteins for Measles virus and P_{CTD}/N_m-RNA complex for Rabies virus in which P remains bound to the N-RNA template and the progressive movement is driven by the dynamic association and dissociation of L from P/N-RNA complex.

As discussed, a tight N-RNA/P complex is not compatible with its liquid behaviour as cytoplasmic inclusion bodies (Negri bodies) [67], and may not be favoured in the rabies virus life cycle considering the multiple roles played by limited amount of P. In fact, many cellular proteins are bound by P protein and involved in viral life cycle. The positive patch on P_{CTD} is not only the N-RNA binding site, but also the identified nuclear localization sequence which binds to importins and confers predominantly nuclear accumulation of P_{CTD} when expressed in cells [97, 98]. The dynamic nature of the N-RNA/P_{CTD} interaction may be enhanced by competitive binding of N-RNA with importins because of the overlapped binding interface. Also, heat shock protein 70 (HSP70) has been identified to regulate rabies virus infection positively [68] and has been found in both purified nucleocapsids [175] and the liquid organelles of Negri body-like structure [66] where viral replication/transcription take place, by

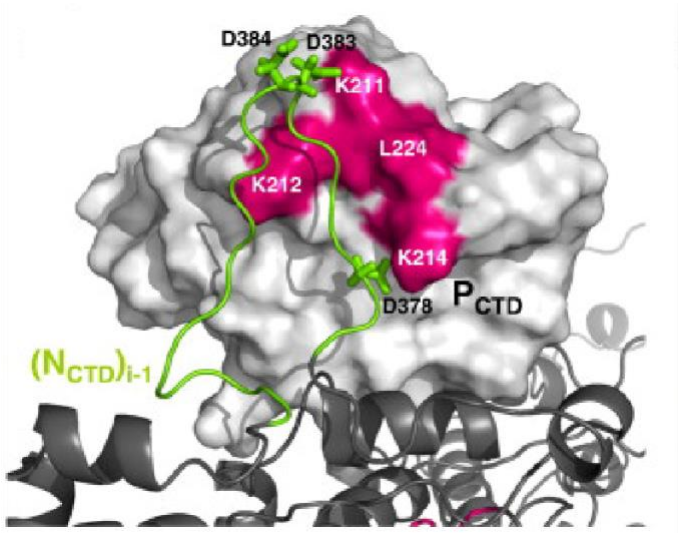
directly or indirectly interacting with N-RNA. HSP70 also performs a pro-viral function in several other Mononegaviruses including respiratory syncytial virus [176], canine distemper virus [177, 178] and measles virus [179-181], possibly via an interaction with N-RNA. Similar to importins competing for P with N-RNA, HSP70 could be competing for N-RNA with P, thus destabilizing the P/N-RNA complex, modulating the P/N-RNA interaction strength and facilitating the cartwheeling. But validation of this hypothesis would need to be further investigated.



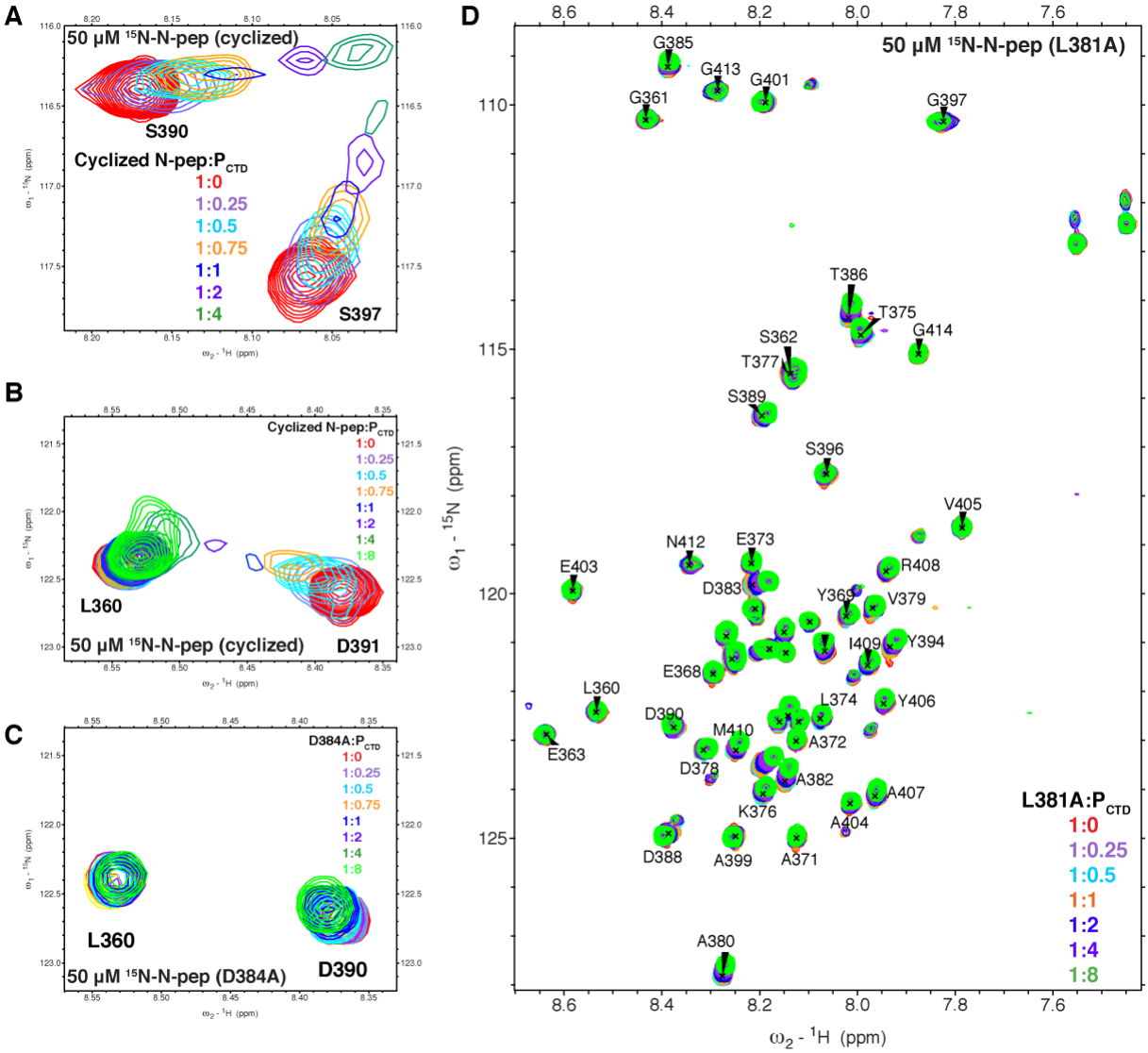
Supplementary Figure 3-1. Schematic diagram of RABV P protein.

The NESs, NLSs, phosphorylation sites (P) and binding sites for protein partners are labelled out.

L: large subunit of the polymerase complex; N^0: nascent RNA-free N protein; FAK: focal adhesion kinase; P_{CED}: central domain, and dimerization domain of RABV P (residues 95-131); LC8: dynein light chain 8; P_{CTD}: C-terminal domain (residues 186-297); STAT1: signal transducer and activator of transcription 1; N-RNA: nucleocapsid (genome encapsidated by N); PML: Promyelocytic leukemia protein; N-NES: N-terminal nuclear export sequence (NES); N-NLS: N-terminal nuclear localization sequence (NLS); C-NES: C-terminal nuclear export sequence (NES); C-NLS: C-terminal nuclear localization sequence (NLS); Serine phosphorylation sites: P63, P64, P162, P210 and P271.



Supplementary Figure 3-2. A model of the N-pep (green) binding site for P_{CTD} (light grey) suggests that N-pep appears as a loop in the full-length N protein. Adapted from [83].



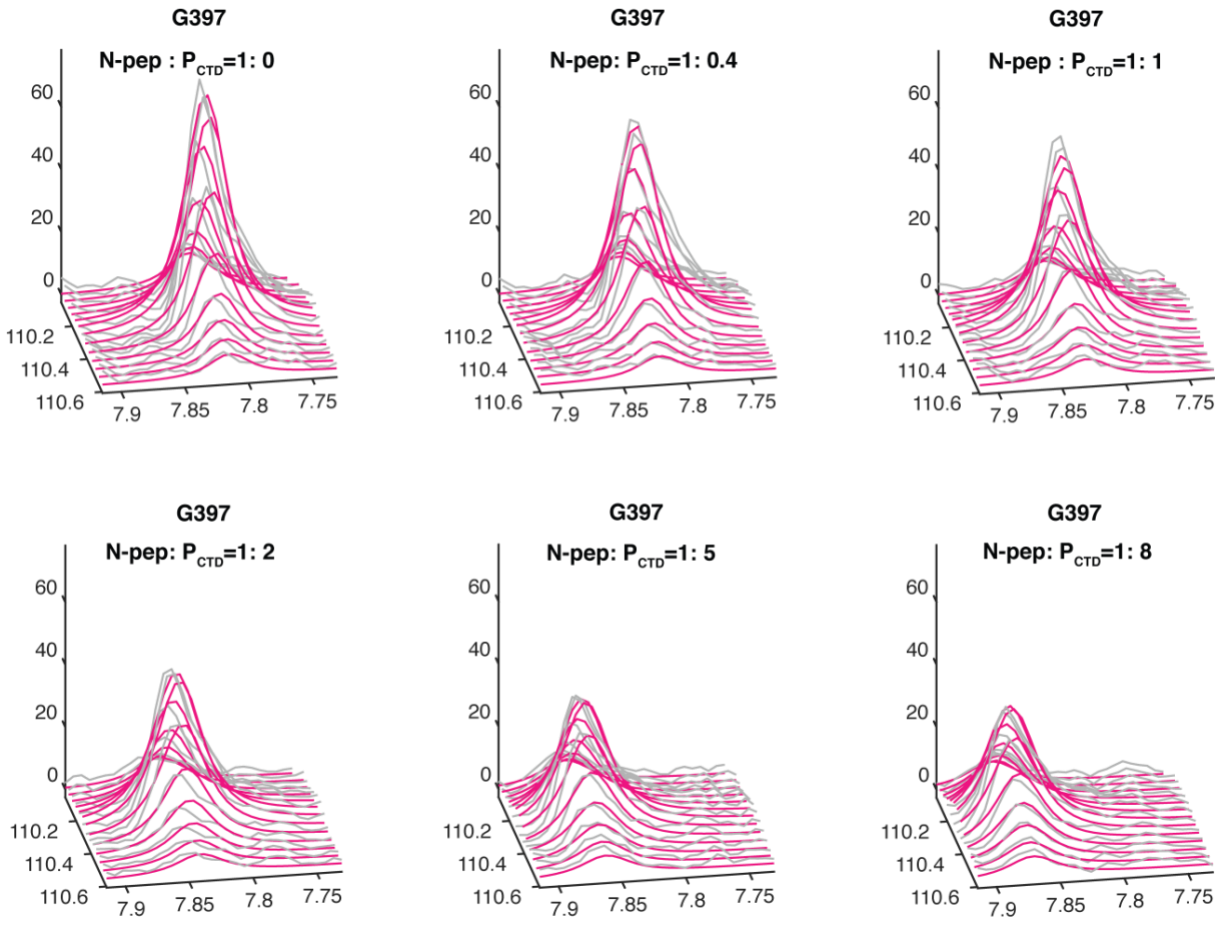
Supplementary Figure 3-3. Titrations of ^{15}N-N-pep variants with unlabelled P_{CTD}.

A. Segment from ^{15}N-labelled cyclized N-pep, ^{1}H-^{15}N HSQC titration experiment. The resonance of residue S390 (residue S389 in the wild type N) first broadened then sharpened up during the titration, suggesting intermediate exchange.

B. Residue D391 of ^{15}N-labelled cyclized N-pep (D390 in the wild type N) was also experiencing two-state exchanges in an intermediate timescale regime.

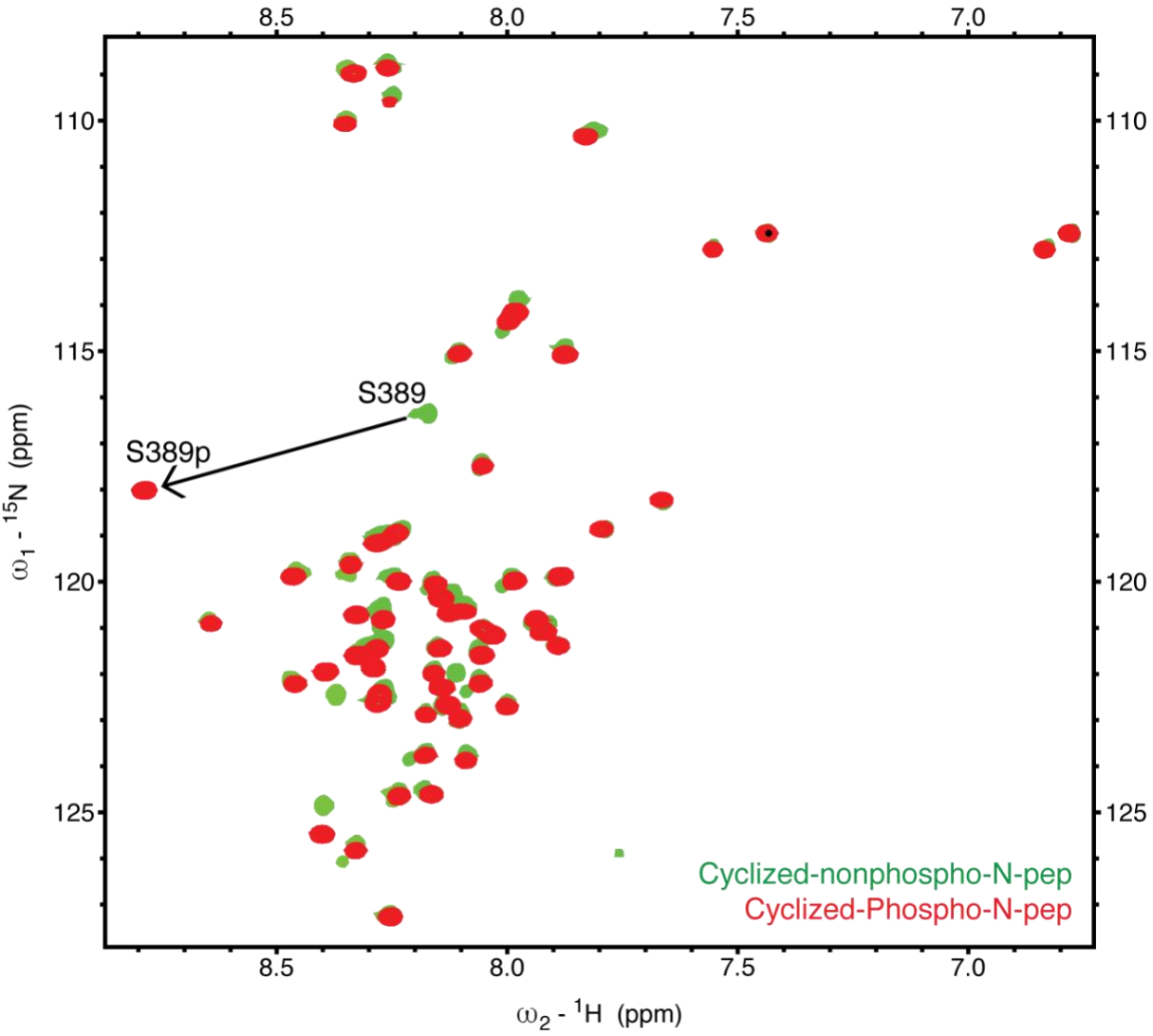
C. Mutating D3484 of N-pep to Ala causes significant reductions in binding with P_{CTD}. Very few changes were observed in the titration of ^{15}N-labelled D384 with P_{CTD}.

D. Mutating L381 to Ala causes significant reductions in binding with P_{CTD}.



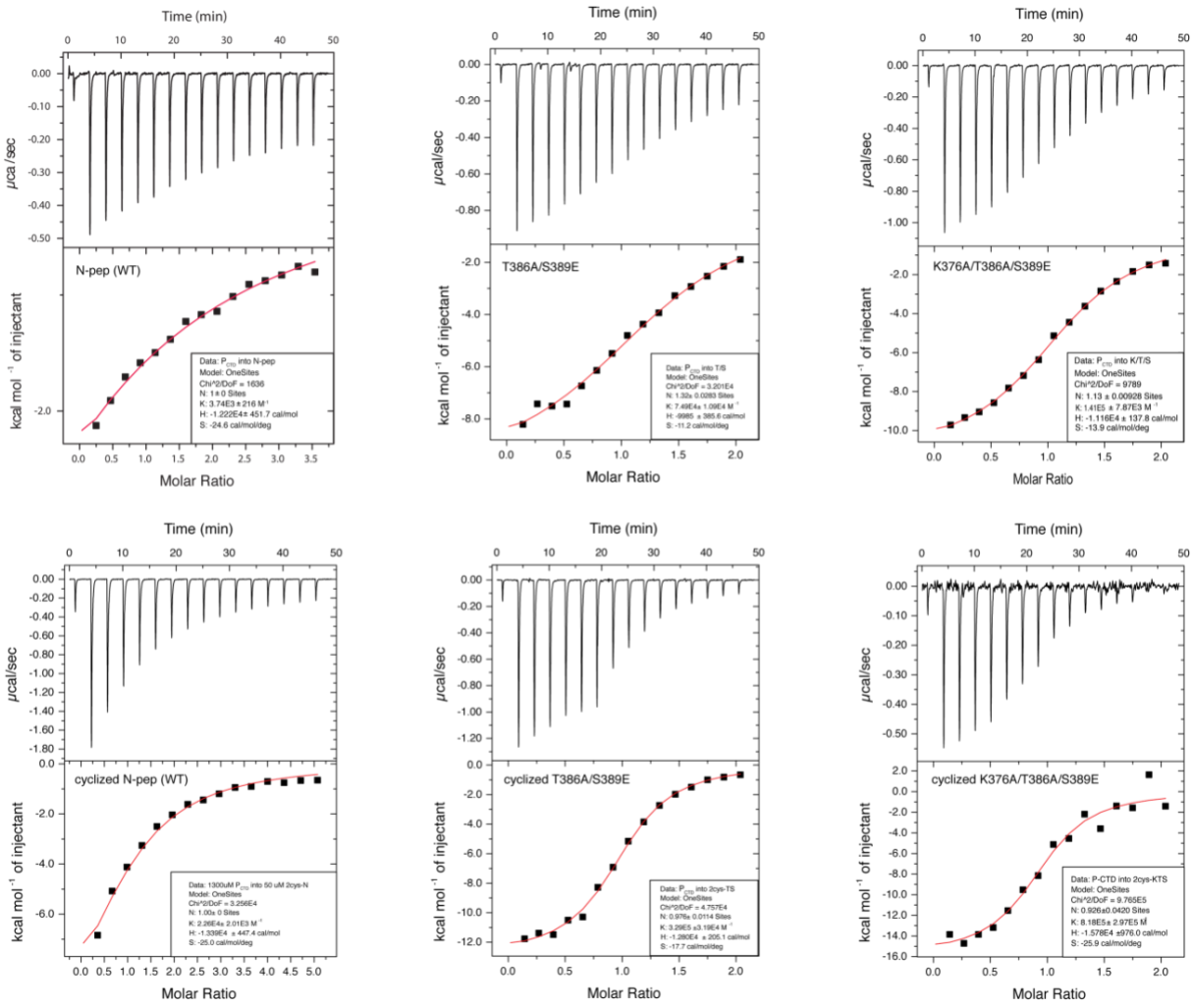
Supplementary Figure 3-4. Lineshape analysis performed on wild type ^{15}N-labelled N-pep with P_{CTD} titration data.

The lineshape of G397 (grey) of N-pep at apo state and 11 different titration points (5 shown here) were fitted into a two-state binding model (magenta) with a K_D of 215 $\pm$ 6 μM and k_{off} of 3164 ± 168 s^{-1}.



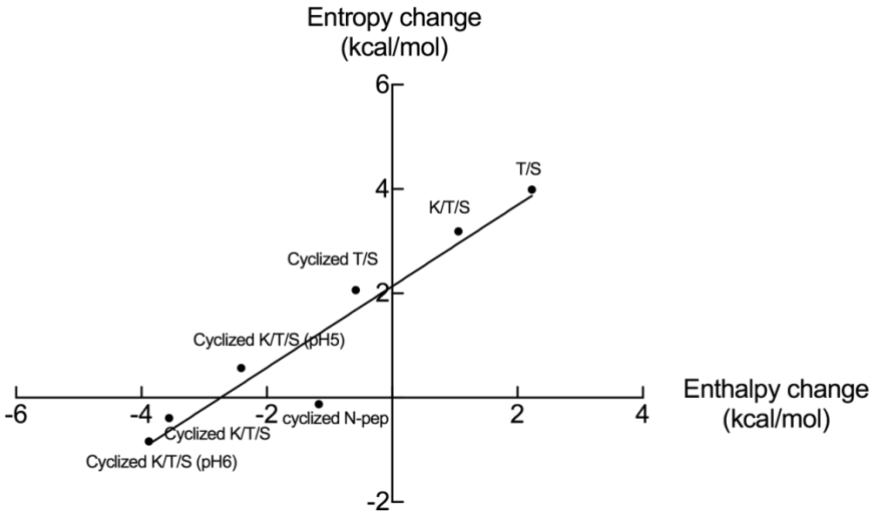
Supplementary Figure 3-5. An overlay of ^{1}H-^{15}N HSQC spectra of Ser389-phosphorylated cyclized N-pep (red) and non-phosphorylated form (green).

Spectra were recorded at 25 ℃ and pH 6.8. Ser389 and phosphorylated S389 (S389p) were labelled in the spectra.



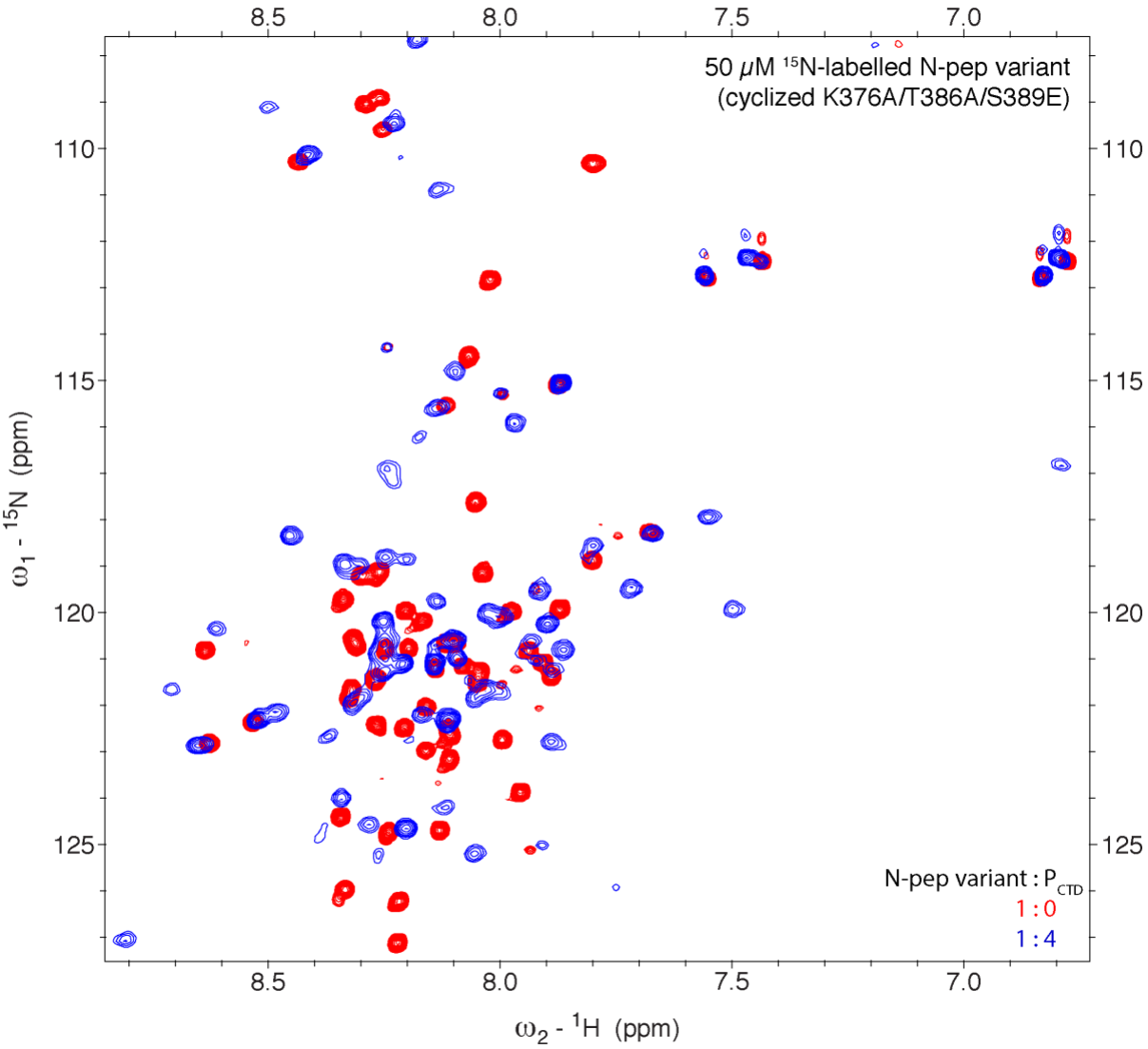
Supplementary Figure 3-6. Isothermal titration calorimetric analysis of interactions between N-pep variants and P_{CTD}.

Upper panel is the raw data representing the heat flow traces caused by injections of P_{CTD}. After integration and correction for the dilution heat, the data was fit to a one binding site model as is shown in the bottom panel.



	pH	*K* (M^{-1})	Δ*H* (kcal/mol)	**Δ*H* change (kcal/mol)**	TΔ*S* (kcal/mol)	**T*ΔS* change (kcal/mol)**	Δ*G* (kcal/mol)
N-pep (WT)	6.8	3.7E3 ± 2E2	-12.2±0.5	**0**	-7.3	**0**	-4.9
Cyclized WT	6.8	2.3E4 ± 2.0E3	-13.4±0. 5	**-1.2**	-7.5	**-0.2**	-5.9
T386A/S389E	6.8	7.5E4 ± 1.1E4	-10.0±0.4	**2.2**	-3.3	**4.0**	-6.7
Cyclized T386A/S389E	6.8	3.3E5 ± 3.2E4	-12.8±0.2	**-0.6**	-5.3	**2.0**	-7.5
K376A/T386A/S389E	6.8	1.4E5 ± 7.9E3	-11.2±0.1	**1.0**	-4.1	**3.2**	-7.1
Cyclized K376A/T386A/S389E	6.8	8.2E5 ± 3.0E5	-15.8±1.0	**-3.6**	-7.7	**-0.4**	-8.1
	6	6.6E5 ± 8.9E4	-16.1±0.4	**-3.9**	-8.2	**-0.9**	-7.9
	5	5.7E5 ± 7.0E4	-14.6±0.5	**-2.4**	-6.8	**0.5**	-7.8

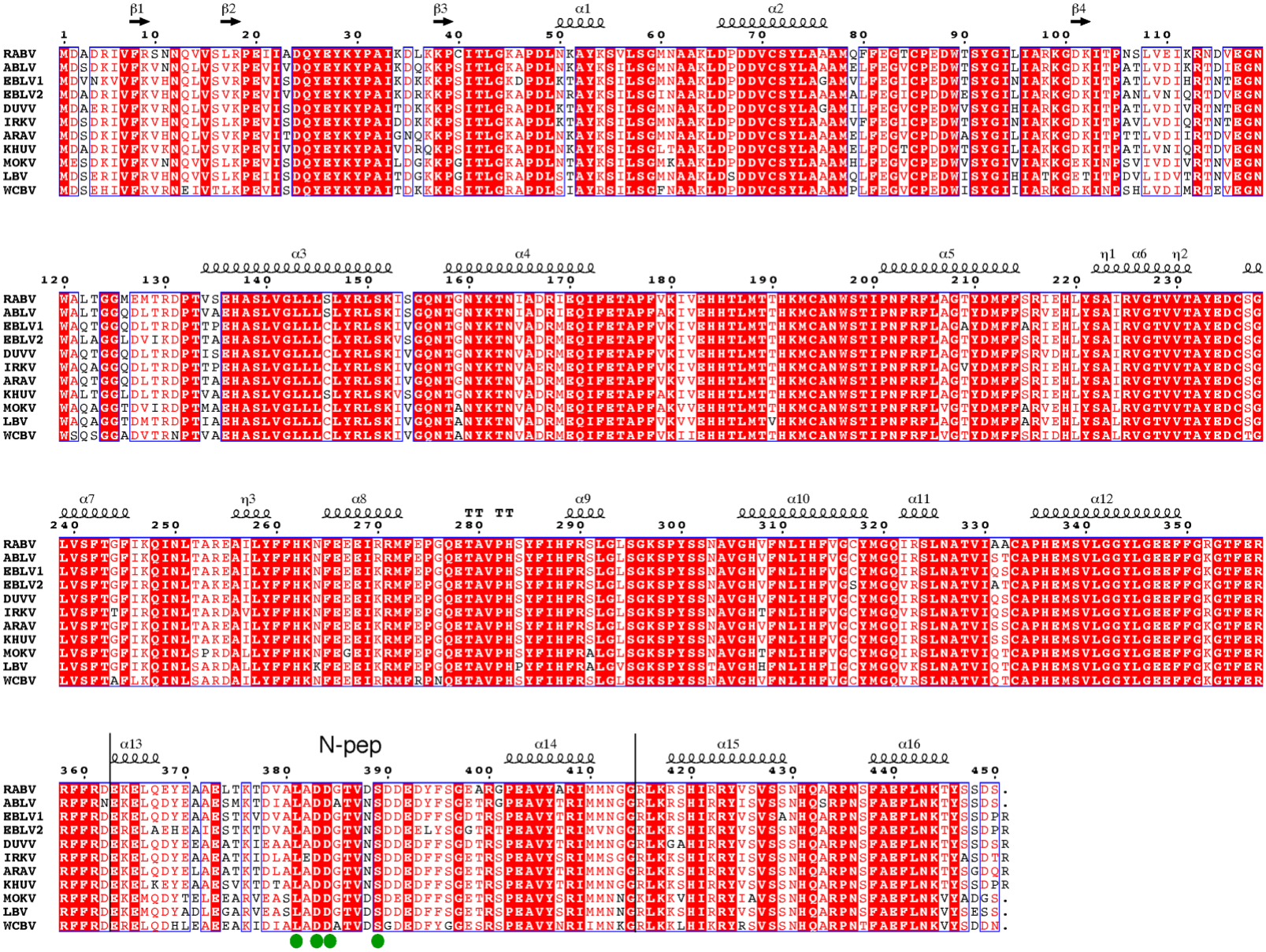
Supplementary Figure 3-7. The enthalpy-entropy compensation between the indicated N-pep variants shows a linear relationship ($R^2 = 0.90$).



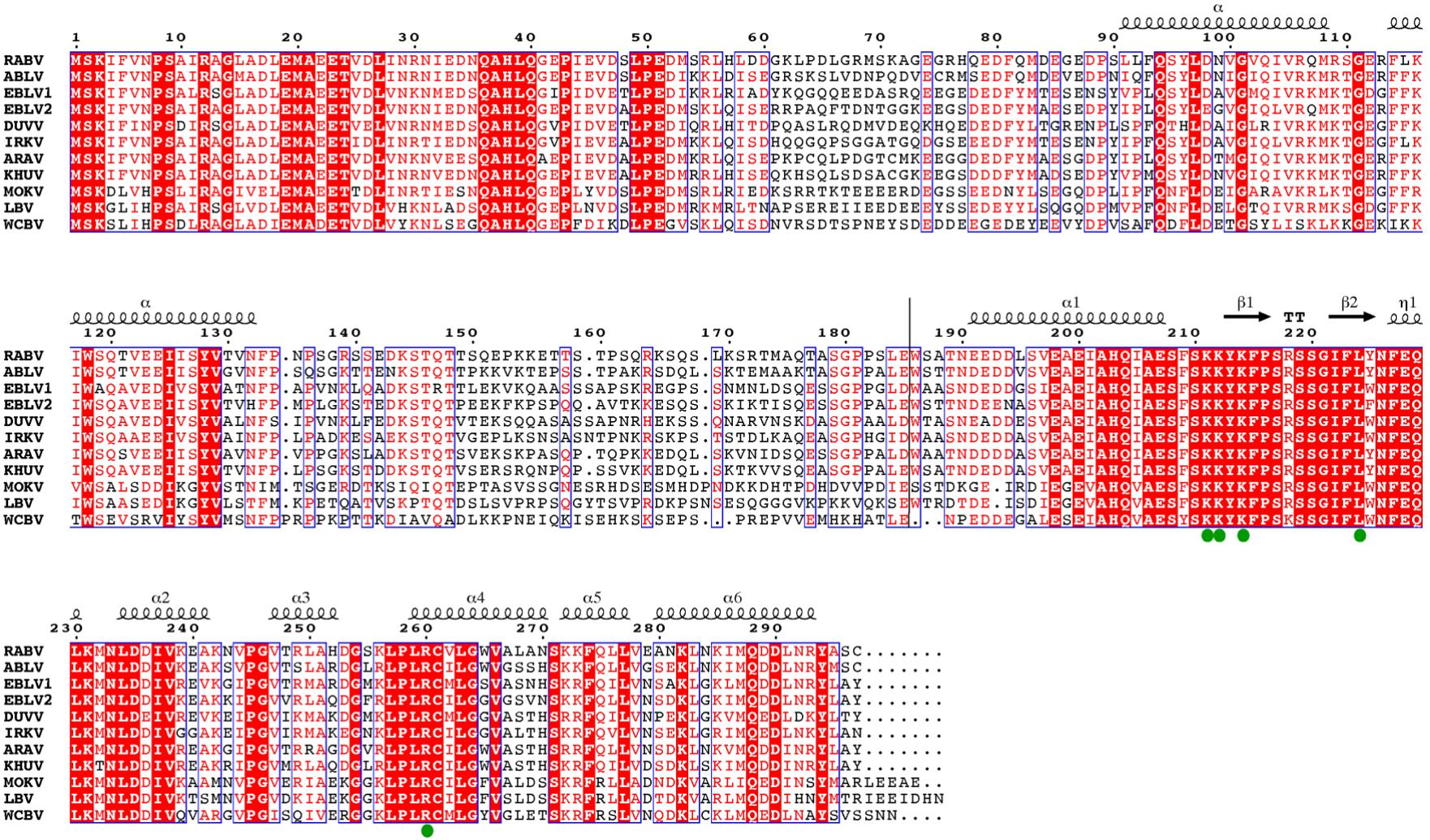
Supplementary Figure 3-8. An overlay of ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled cyclized K376A/T386A/S389E in the free from (red) and bound form (blue).

The HSQC spectra of ^{15}N-labelled cyclized K376A/T386A/S389E recorded at 4-fold molar excess of P_{CTD} showed more dispersed peaks compared to the apo form. The spectra were recorded at 25 °C and pH 6.8.

A

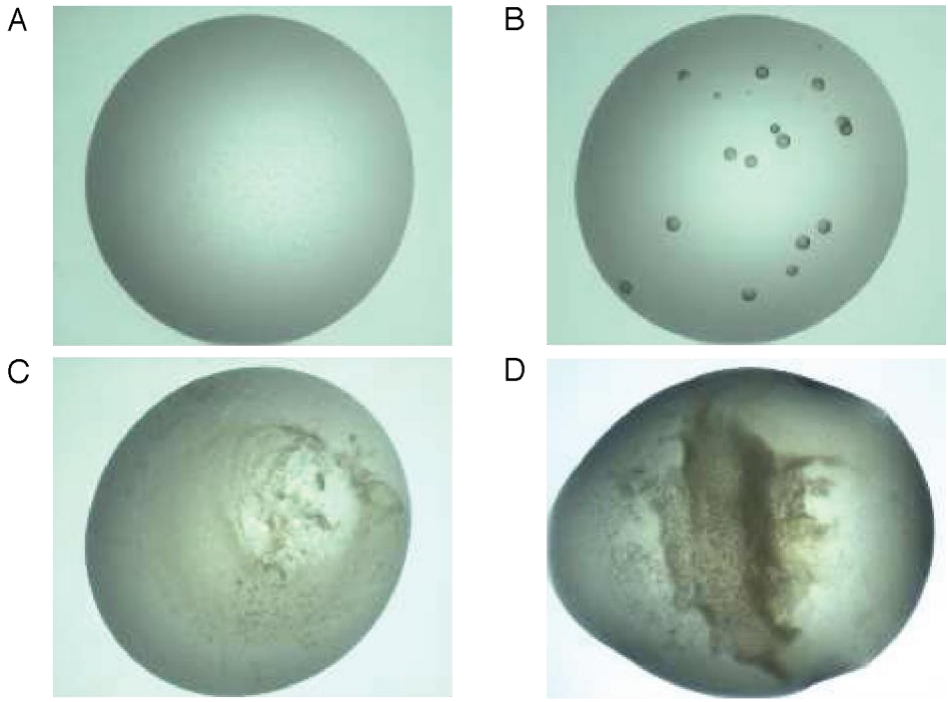

β1
β2
β3
α1
α2
β4
α3
α4
α5
η1
α6
η2
α7
η3
α8
TT TT
α9
α10
α11
α12
α13
N-pep
α14
α15
α16
RABV
ABLV
EBLV1
EBLV2
DUVV
IRKV
ARAV
KHUV
MOKV
LBV
WCBV

B


RABV
ABLV
EBLV1
EBLV2
DUVV
IRKV
ARAV
KHUV
MOKV
LBV
WCBV
α
α1
β1
TT
β2
η1
α2
α3
α4
α5
α6

Supplementary Figure 3-9. Multiple sequence alignment of *Lyssavirus* N (A) and P proteins (B).

The residues critical for N-pep/P_{CTD} interaction are marked with a green dot. Although P protein is much less conserved between lyssaviruses than N, the critical residues identified within the N/P interface appear to be highly conserved among the lyssaviruses. Protein sequences were retrieved from the GenBank: RABV (N: O55611, P: Q9IPJ8), ABLV (N: Q8JTH3, P: Q8JTH2), EBLV1 (N: A4UHP8, P: A4UHP9), EBLV2 (N: A4UHQ3, P: A4UHQ4), DUVV (N: Q66453, P: O56774), IRKV (N: Q5VKP6, P: Q5VKP5), ARAV (N: Q6X1D8, P: Q6X1D7), KHUV (N: Q6X1D4, P: Q6X1D3), MOKV (N: P0C570, P: P0C569), LBV (N: Q82994, P: O56773), WBCV (N: Q5VKP2, P: Q5VKP1); accession numbers are shown in parentheses



Supplementary Figure 3-10. Failed attempts to co-crystallize N-pep with P_{CTD}.

A. Clear drop.
B. Spherulites.
C. Light precipitate.
D. Heavy brown precipitate.

Chapter 4:

Structural and dynamic analysis of the phosphomimetics of the C-terminal domain of rabies virus P protein

Structural and dynamic analysis of the phosphomimetics of the C-terminal domain of rabies virus P protein

Jingyu Zhan[1], Riley Metcalfe[1], Michael D.W. Griffin[1], Gregory W. Moseley[1,2], Paul R. Gooley[1]

[1] Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3010, Australia and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia.

[2] Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton Campus, Victoria 3800, Australia.

* Correspondence to:

Paul R. Gooley
The Department of Biochemistry and Molecular Biology and Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Victoria 3010, Australia
Ph: +61 3 9034 2273
Email: prg@unimelb.edu.au.

4.1 Introduction

Members of the *Lyssavirus* genus cause lethal encephalitic disease which remains a significant global public health threat with no effective treatment available. The virulence of these viruses results from their ability to hijack cellular machinery for viral replication and strategies to evade the host immune system. Due to the limited coding capacity of the single negative-stranded RNA genome, viruses express multi-functional proteins in different isoforms comprised of intrinsically disordered regions and overlapping functional sequences, which can be further regulated by post-translational modifications (PTMs) to carry out a diversity of functions.

Rabies virus (RABV) is a prototypical member of the *Lyssavirus* genus of the *Rhabdoviridae* family. Its 12 kb genome encodes five viral proteins, among which P, L and N together with the RNA genome (N-RNA complex) form liquid organelles (Negri bodies) of viral replication factories within the host cell's cytoplasm [67]. Phosphoprotein P is identified as a multifunctional protein for its ability not only to work as the polymerase factor but also to antagonize the innate antiviral responses. The N-terminal region of P bears the binding sites for the viral polymerase L and the nascent RNA-free nucleoprotein N^0 and is proposed to be disordered. Following this region is a structured dimerization domain (PDB: 3L32, from China/MRV strain), then another disordered region which is linked to a well-structured globular domain P_{CTD} (PDB: 1VYI, from CVS strain) at the C-terminus. *In vivo* experiments showed that P_{CTD} contains multiple functional important sites, including the viral N-RNA binding site [73, 83], nuclear import and export sequences [97, 98], and binding sites for host factors including nucleolin [100], microtubules [29], Promyelocytic leukemia (PML) protein [101] and signal transducers and activators of transcription (STAT) proteins [84, 102].

However, how these juxtapositioned binding interfaces and diverse functions are coordinated on this 13 kDa P_{CTD} domain remains unclear.

Phosphorylation is a major reversible post-translational modification which may play important roles in regulating viral protein functions. CKII-catalyzed phosphorylation of rabies virus N protein at S_{389} [79] is necessary for viral replication by enhancing the interaction between the N-RNA complex and P_{CTD} [81]. *In vitro* experiments showed that P protein can be phosphorylated by a putative rabies virus protein kinase (RVPK) at non-conserved S_{63} and S_{64}, and by protein kinase C (PKC) at conserved S_{162}, S_{210} and S_{271} [90]. Among the two phosphorylation sites identified within P_{CTD}, PKC phosphorylation of S_{210} and its phosphomimetic mutation were found to shift the nucleo-cytoplasmic distribution of P_{CTD} towards nuclear export [98] possibly by inducing conformational change to P_{CTD} and unmasking the originally buried NES, thus simultaneously regulating NLS and NES functions. Since the P_{CTD} interacts with multiple viral and host cellular proteins, changes in its phosphorylation state may impact on these interactions, and ultimately on viral growth and pathogenicity. The possible conformational change means investigation into the different phosphorylated forms of P_{CTD} is necessary for understanding their biological significance. In this study, we aimed to investigate the potential differences between unphosphorylated and phosphorylated P_{CTD} from Nishigahara strain through structural, dynamic and functional aspects using recombinantly expressed and purified P_{CTD} and its phosphomimetics $S_{210}E$ and $S_{271}E$.

4.2 Materials and Methods

4.2.1 Site-directed mutagenesis

The bacterial expression construct of P_{CTD} was the same as previously described [143]. Phosphomimetic mutations of P_{CTD} were introduced into the original P_{CTD} construct using PrimeSTAR Max DNA Polymerase (Takara) according to the manufacturer's instructions, and using mutagenic primers designed according to Zheng et al. [141]. The amplified material was digested with DpnI for 1.5 hours at 37 °C to remove wild type bacterially expressed plasmid before transformation into Top10 *Escherichia coli* cells. Mutation of the plasmid was confirmed by sequencing at the Micromon sequencing facility (Monash University).

4.2.2 Protein expression and purification

The expression and purification of unlabelled, ^{15}N-labelled or ^{13}C,^{15}N-labelled phosphomimetic mutations of P_{CTD} were the same as the wild type P_{CTD} as described before [143]. Briefly, protein was expressed in *E. coli* BL21(DE3) in 2YT auto-induction media (N-5052) where the expression of the target protein is induced upon the change of glucose-lactose metabolic state during the culture growth [124]. After OD_{600} reached 0.8 at 37 °C, the bacterial culture was transferred to 16 °C for overnight expression. NMR samples were uniformly ^{15}N-labelled by growing cells in N-5052 supplemented with 1 g/L of $^{15}NH_4Cl$ as a sole source of nitrogen [124], or ^{13}C,^{15}N-labelled by growing cells in N-5052 [124] supplemented with 1 g/L of $^{15}NH_4Cl$ (Sigma-Aldrich) and 3 g/L of D-[^{13}C] glucose (Sigma-Aldrich) as sole sources of nitrogen and carbon.

Recombinant P_{CTD} or its mutants were purified by TALON metal affinity chromatography (Clontech) and the His-tag was removed by overnight TEV treatment (0.5 ml of 1.8 mg/ml purified TEV per 50 ml of protein sample). Protein samples were further purified by

size-exclusion chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare) in 50 mM Na_2HPO_4/NaH_2PO_4 (pH 6.8) and 100 mM NaCl. The crystallography samples were exchanged into 20 mM Tris (pH 7.5) and 100 mM NaCl by overnight dialysis using the mini dialysis kit with 1 kDa cut-off (Amersham Biosciences).

The expression and purification of N-pep (residues 363-414 of N protein, with $S_{389}E$ mutation) were the same as previously described in Chapter 3. N-pep was produced in a pGEX-6P-3 vector, with an N-terminal GST-tag followed by a PreScission protease cleavage site, expressed in the same way using auto-induction as P_{CTD}. For purification of N-pep, the supernatant of the cell lysate was applied to glutathione-Sepharose 4B resin (GE Healthcare) before washing with Phosphate-buffered saline (PBS) and incubation with Tris-buffered saline (TBS) solution containing PreScission protease (200 μL of 3 mg/ml purified PreScission protease in 25 mL of TBS) for 2 hours at 4 ℃ to remove the GST-affinity tag while bound to the Glutathione Sepharose matrix. The cleaved N-pep was washed out of the column with TBS and further purified by reversed-phase high performance liquid chromatography (RP-HPLC) with a gradient from 0 to 60% acetonitrile (0.1% trifluoroacetic acid) in 60 min at a flow rate of 5 ml/min using a C18 column (Agilent ZORBAX 300SB-C18, 5 μm, 9.4×250 mm). The fractions collected from RP-HPLC were analysed by mass spectrometry to confirm the molecular weight before freeze drying.

4.2.3 Crystallization and data collection

An initial crystallization condition screen was performed at the Collaborative Crystallisation Centre (C3 Australia) using a nano dispensing robot (Art Robins Phenix) and a sitting-drop vapour-diffusion method where 384 conditions were tested. Conditions that produced crystals were optimised manually in sitting drops by mixing 1.5 μL of protein solution at 6 mg/ml in 20

mM Tris (pH 7.5), 100 mM sodium chloride with 1.5 μL of the reservoir solution. Diffraction-quality rod-shaped crystals of $S_{210}E$ and wild type P_{CTD} were grown between 2-4 weeks at 8 ℃ in 1.9 M ammonium sulfate, 0.2 M potassium sodium tartrate and 0.1 M sodium citrate (pH 4.75). Three-dimensional crystals of wild type P_{CTD} grown at 20 ℃ and appeared within 1-2 weeks in 25% (w/v) PEG 3350, 0.2 M sodium acetate and 0.1 M HEPES (pH 8.5) were used for data collection as well as serving as a seed-stock preparation to produce $S_{271}E$ crystals following Luft et al. [182]. Similar three-dimensional diffraction-quality crystals of $S_{271}E$ mutant were obtained at 20 ℃ within a week by seeding with wild type three-dimensional P_{CTD} crystals in 20% PEG 3350 and 0.2 M sodium acetate. Crystals were flash-cooled in liquid nitrogen from the crystallization drop prior to data collection. X-ray diffraction data were collected at 100 K at the MX2 beamline [183].

4.2.4 X-ray structure determination and refinement

Diffraction data were indexed, integrated and scaled with *XDS* [184], analysed by *Pointless* [185] and then merged using *Aimless* [186] from the *CCP4* suite [187]. Initial phases were solved by molecular replacement using the P_{CTD} structure from CVS strain (PDB: 1VYI) using *phenix.phaser* [188]. *Phenix.autobuild* [189] was used for automatic model building with simulated annealing to reduce model bias. Restrained refinement was carried out using *phenix.refine* [190] with iterative manual real space model building and corrections using *Coot* [191]. Translation-Libration-Screw-rotation (TLS) refinement was performed in the later stage of refinement with each protein molecule defined as an individual TLS group and in the final rounds using automatically defined TLS groups by *Phenix*. Solvent molecules were not added until the end stages of refinement. Summary of data processing and refinement statistics are listed in Table 4-1.

4.2.5 NMR assignment and solution structure calculation

All NMR experiments were performed on Bruker Avance III 600 MHz, Bruker Avance II 800 MHZ or Bruker Avance IIIHD 700 MHz spectrometers, all equipped with 5-mm HCN (TCI) cryogenically cooled probes to maximize sensitivity. All data were collected at 25 °C under the standard condition: 0.5-0.6 mM $S_{210}E$ dissolved in 50 mM phosphate buffer (pH 6.8), 100 mM sodium chloride. The nearly complete ^{1}H, ^{13}C and ^{15}N assignments were accomplished as previously described [143]. To obtain distance restraints, 2D NOESY, 3D ^{15}N-edited NOESY and two 3D ^{13}C-edited NOESY (with ^{13}C offsets at 40 ppm for aliphatic carbons and 123 ppm for aromatic carbons) were acquired. The 2D NOESY was acquired conventionally whereas the 3D ^{13}C and ^{15}N-edited NOESY spectra were recorded with 25% NUS Poisson gap sampling [125]. The 2D NOESY was processed in *NMRPipe* whereas the data collected using NUS were reconstructed using *qMDD* [126] with the compressed sensing algorithm and processed using *NMRPipe* [127]. NMR data were usually processed with a Lorentz–Gaussian function in the direct dimension and cosine-squared bells in the indirect dimensions and zero-filling once. Dihedral backbone angle restraints were predicted by *TALOS+* [132] on the basis of backbone assignments (HN, Hα, ^{15}N, $^{13}Cα$, $^{13}Cβ$, $^{13}C'$). Distance restraints were obtained by auto-assigning NOE data using the *CANDID* algorithm in *CYANA* [192-194] (version 2.1) where seven iterative cycles of NOE assignment and structure calculation were performed in which initial NOE assignments based on chemical shift agreement were assessed with respect to network anchoring and spatial proximity in the intermediary structure, and only unambiguous NOE assignments were passed down to the succeeding cycle to guide further NOE assignment and structure refinement. 100 trial structures were generated in each cycle and the 20 best structures with the lowest energy calculated in the last cycle were selected to represent the solution structure. The experimental restraints and structure statistics can be found in Table 4-2.

4.2.6 Relaxation experiments

^{15}N transverse relaxation T_2 experiments for wild type P_{CTD} and $S_{210}E$ were acquired at 15 °C on a Bruker Avance III 600 MHz spectrometer and at 25 °C on a Bruker Avance II 800 MHz spectrometer in 50 mM sodium phosphate (pH 6.8) and 100 mM sodium chloride. The data were collected with 1024 t_2 and 150 t_1 points and 8 different relaxation delays (2×0.017, 0.034, 0.068, 2×0.102, 0.136, 2×0.1696, 0.204, 0.237 seconds), processed in *NMRPipe* [127] and analysed in *relax* [130, 195].

^{15}N Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments [196] were performed at 3 °C, 15 °C and 25 °C for wild type P_{CTD} and $S_{210}E$ on a Bruker Avance II 800 MHz spectrometer. Experiments were performed with a constant delay (T_{cpmg}) of 80 ms and 13 different CPMG pulse frequencies (0, 50, 2×75, 100, 2×150, 200, 300, 500, 2×700, 800, 900, 1000, 1500 Hz). Data were collected with 2048 t_2 and 256 t_1 points and 16 scans for each t_1 point on ^{15}N-labelled $S_{210}E$ or wild type P_{CTD} samples. Spectra were processed using *NMRPipe* [127] and peak intensities were measured with *NMRFAM-Sparky* [129]. The automatic relaxation dispersion analysis software *NESSY* [197] was used to analyse data on a per residue basis to look for any residues with chemical exchange.

4.2.7 NMR titration experiments

The binding of N-pep ($S_{389}E$) to P_{CTD} and its phosphomimetic mutants were monitored using 2D ^{1}H-^{15}N HSQC titration experiments. Concentrated unlabeled P_{CTD} or its phosphomimetic mutants was gradually added into the ^{15}N-labelled N-pep ($S_{389}E$) sample until reaching 8-fold molar excess. Seven HSQC spectra (2048×256 data points) were recorded at P_{CTD} : N-pep molar ratios of 1:0, 1:0.25, 1:0.5, 1:1, 1:2, 1:4 and 1:8.

Resonances showing obvious chemical shift changes yet remaining well-resolved were picked to calculate binding affinity by fitting the titration curve into a model for a protein with only one ligand binding site using *Xcrvfit* [151]. Dilution through the titration was taken into account. The chemical shift perturbation (CSP) of each residue was calculated by the formula [152]:

$$CSP = \sqrt{\Delta\delta_H^2 + 0.154^2 \cdot \Delta\delta_N^2}$$

For all three sets of NMR titration data between ^{15}N-labelled N-pep ($S_{389}E$) and P_{CTD} variants (wild type, $S_{210}E$ and $S_{271}E$), two-dimensional lineshape analysis were performed using the software *TITAN* [153]. The off-rate (k_{off}) and dissociation constant (K_D) were determined by simultaneously fitting the lineshape of some shifted yet resolved peaks to a two-state single binding site model. The on-rate (k_{on}) was determined by the equation:

$$k_{on} = \frac{k_{off}}{K_D}$$

4.3 Results

4.3.1 Structure of $S_{210}E$, $S_{271}E$ compared with wild type P_{CTD} by X-ray crystallography

To investigate possible conformational changes that are induced upon phosphorylation, the structure of the phosphomimetic mutants of P_{CTD}, $S_{210}E$ and $S_{271}E$, were solved using X-ray crystallography. For a complete comparison, wild type P_{CTD} was crystallised in the same condition as $S_{210}E$ or $S_{271}E$ separately and the corresponding crystal structures were solved.

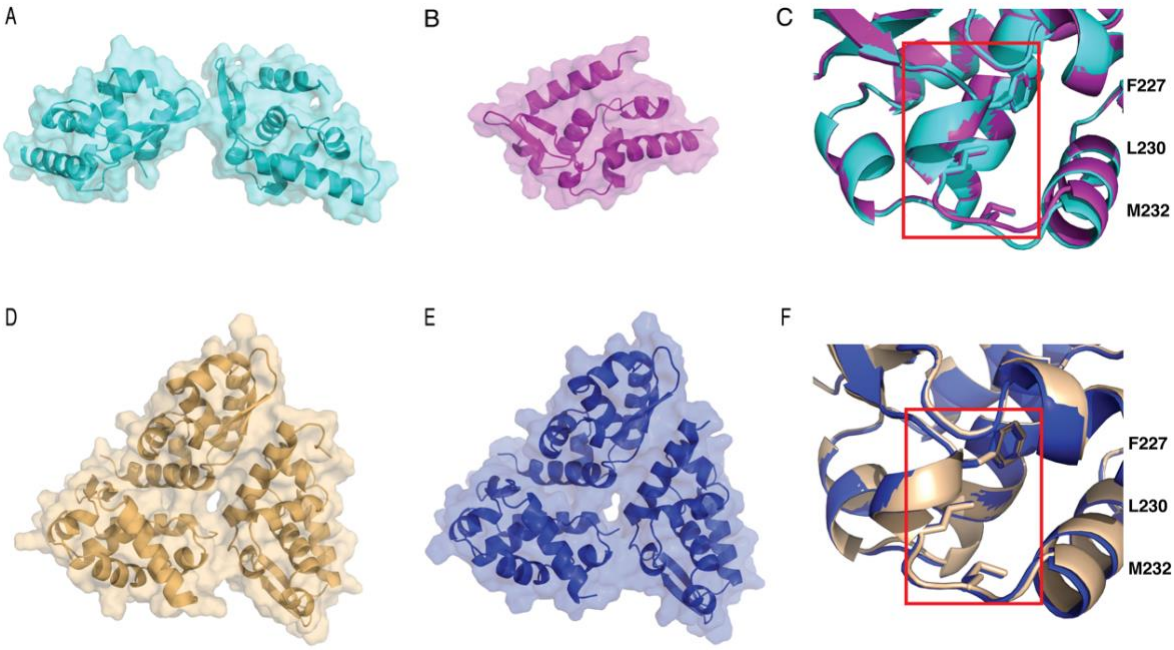


Figure 4-1. X-ray crystal structure of Nishigahara P_{CTD}, $S_{210}E$ and $S_{271}E$.

A. The crystal structure of $S_{210}E$ belonged to the orthorhombic space group $P2_12_12$ with two monomers in the asymmetric unit.

B. Wild type P_{CTD} crystallized in the same condition as $S_{210}E$. Crystal belonged to the orthorhombic space group $P2_12_12_1$ with one monomer in the asymmetric unit.

C. Comparison of the side chain position of the critical residues F_{227}, L_{230} and M_{232} in the putative NES between $S_{210}E$ (cyan) and wild type (magenta).

D. $S_{271}E$ crystalized in space group $P1$ with three monomers per asymmetric unit.

E. Wild type P_{CTD} crystallized in the same condition and same space group as $S_{271}E$ with almost identical unit cell dimensions.

F. Comparison of the side chains of the critical residues F_{227}, L_{230} and M_{232} in the putative NES between $S_{271}E$ (wheat) and wild type (blue).

Table 4-1. X-ray Data collection and model refinement statistics.

Statistics for the highest-resolution shell are given in parentheses.

	$S_{271}E$	Wild type	$S_{210}E$	Wild type
Data collection				
Space group	*P*1	*P*1	$P2_12_12$	$P2_12_12_1$
Wavelength (Å)	0.9537	0.9537	0.9537	0.9537
Number of images	3600	3600	3600	3600
Oscillation range per image (°)	0.1	0.1	0.1	0.1
Detector	EIGER X 16M	EIGER X 16M	EIGER X 16M	EIGER X 16M
Cell dimensions				
a, b, c (Å)	39.0, 47.8, 48.8	38.3, 47.9, 48.9	44.7, 69.1, 75.7	42.3, 44.0, 61.3
α, β, γ (°)	66.3, 88.6, 90.0	66.2, 88.5, 89.8	90, 90, 90	90, 90, 90
Resolution range (Å)	43.73-1.85 (1.92-1.85)	38.26-1.60 (1.66-1.60)	37.83-2.00 (2.07-2.00)	34.83-1.50 (1.55-1.50)
R_{merge}	0.075 (0.777)	0.050 (0.880)	0.112 (2.724)	0.084 (1.375)
R_{meas}	0.088 (0.911)	0.059 (1.038)	0.117 (2.826)	0.088 (1.429)
$R_{p.i.m}$	0.046 (0.473)	0.031 (0.545)	0.032 (0.749)	0.025 (0.388)
CC(1/2)	0.997 (0.672)	0.998 (0.557)	0.999 (0.718)	0.999 (0.813)
I/σ(I)	9.4 (1.4)	11.8 (1.3)	12.8 (1.3)	16.5 (1.6)
Total reflections	95941 (9731)	143403 (14546)	214001 (21800)	240989 (25038)
Unique reflections	26755 (2642)	40367 (4030)	16314 (1562)	18933 (1860)
Completeness (%)	97.3 (96.4)	96.1 (94.9)	98.9 (98.4)	99.9 (99.9)
Multiplicity	3.6 (3.7)	3.6 (3.6)	13.1 (13.9)	12.7 (13.5)
Wilson B-factor (Å^2)	30.07	29.87	42.81	23.00
Matthews coefficient, V_M (Å^{3}Da^{-1})	2.14	2.12	2.26	2.21
Solvent content (%)	42.6	41.9	45.6	44.3
Refinement				
Resolution range (Å)	43.73-1.85 (1.92-1.85)	38.26-1.60 (1.66-1.60)	37.83-2.00 (2.07-2.00)	34.83-1.50 (1.55-1.50)
Reflections used in refinement	26750 (2642)	40356 (4029)	16255 (1562)	18916 (1858)
R_{free} reflections	1258 (108)	2053 (195)	782 (74)	949 (105)
R_{work}	0.178 (0.308)	0.171 (0.320)	0.230 (0.349)	0.174 (0.278)
R_{free}	0.201 (0.377)	0.188 (0.322)	0.247 (0.350)	0.194 (0.301)
Protein molecules in asymmetric unit	3	3	2	1
Protein residues	325	318	218	109
Total atoms	5434	5304	3560	1835
Protein molecules	2599	2536	1747	868
Ligands	None	None	15	15
Water	227	205	48	79
Average B-factor (Å^2)	40.88	44.54	76.59	36.67
Protein	40.75	44.36	76.17	35.24
Ligands	None	None	123.16	91.25
Water	42.43	46.81	77.26	42.08
RMSD				
Bond length (Å)	0.011	0.006	0.005	0.010

Bond angles (°)	1.05	0.92	0.97	1.28
Ramachandran statistics				
Favoured /allowed/outliers (%)	96.9/2.8/0.3	98.1/1.9/0.0	97.7/2.3/0.0	99.1/0.9/0.0

As an autonomous folding domain isolated from the full-length P protein, wild type Nishigahara P_{CTD} crystalized in two different crystallisation conditions with different crystal morphology and packing. Rod-shaped Nishigahara P_{CTD} crystals were obtained at 8 °C (1.9 M ammonium sulfate, 0.2 M potassium sodium tartrate and 0.1 M sodium citrate, pH 4.75) that diffracted to a 1.5 Å resolution and belonged to the orthorhombic space group $P2_12_12_1$ with one monomer in the asymmetric unit (Figure 4-1B). Three-dimensional P_{CTD} crystals were grown at 20 °C in 25% (w/v) PEG 3350, 0.2 M sodium acetate and 0.1M HEPES (pH 8.5) and the structure was solved to 1.6 Å resolution with 3 monomers per asymmetric unit crystalized in space group *P*1 (Figure 4-1E). Despite the differences in crystallisation condition and crystal morphology, the structures solved were essentially the same (RMSD = 0.47 Å). Overall the structure can be described as the shape of a sliced pear with a round face and a flat face, composed of seven helices (six α-helices and one 3_{10} helix) and an antiparallel β-sheet formed by two short β-strands, in good agreement with the crystal structure of P_{CTD} from CVS strain of rabies virus (PDB: 1VYI) (RMSD = 0.55 Å) which differs by four residues ($I_{257}L$, $D_{281}N$, $S_{284}N$ and $T_{295}A$). The only major difference is a shortened N-terminus of Nishigahara P_{CTD} compared to the published CVS P_{CTD} crystal structure, which could be due to the different crystal packing and crystal contacts, a result of different crystal forms. In the CVS P_{CTD} (PDB: 1VYI) crystal structure which belonged to $P3_121$ space group and was refined at 1.5 Å, the N-terminus was stabilized by multiple crystal contacts which were not present in the crystal lattice of the Nishigahara P_{CTD} (Supplementary Figure 4-1). The N-terminus of the Nishigahara P_{CTD} in two different crystal forms are both exposed and flexible, thus residues 186-188 were not defined in the electron density and were not included in the models.

Rod-shaped crystals of $S_{210}E$ were obtained from the same condition as wild type rod-shaped crystals, but with a slightly different crystal packing. Crystals of $S_{210}E$ belonged to the orthorhombic space group $P2_12_12$ and the structure was solved to a resolution of 2 Å with 2 monomers in the asymmetric unit (Figure 4-1A). Three-dimensional $S_{271}E$ crystals were grown by seeding with wild type three-dimensional crystals in a similar condition as the seed. $S_{271}E$ crystalized in space group *P*1 with 3 monomers per asymmetric unit with almost identical unit cell dimensions as the wild type three-dimensional crystals, and the structure was solved to 1.85 Å resolution (Figure 4-1D). Overall, $S_{210}E$ and $S_{271}E$ structures have nearly identical conformations to wild type P_{CTD} ($S_{210}E$ and the corresponding wild type: RMSD=0.28 Å, $S_{271}E$ and the corresponding wild type: RMSD=0.27 Å) except for the minor difference in the loop region (residues S_{219}-G_{221}) and in both termini. Importantly, no conformational change was observed for the amphipathic 3_{10} helix (residues 227-230) within the proposed NES motif (residues 221-238) [98], and the side chains of the key hydrophobic residues ($F_{227}xxL_{230}xM_{232}$) of the NES were still buried (Figure 4-1C, F), and thus, presumably, not readily accessible to CRM-1.

4.3.2 Solution structure of P_{CTD} and $S_{210}E$ by NMR spectroscopy

Although the crystal structures of $S_{210}E$, $S_{271}E$ and wild type P_{CTD} were well-defined and the phosphomimetics showed no sign of conformational change compared to wild type P_{CTD}, it may be possible that crystallization traps the lowest energy conformation of a protein where several functionally important conformational states may exist. As NMR provides a solution based complementary approach to study protein structure and dynamics under near physiological conditions, we incorporated NMR into the study to solve the solution structure of P_{CTD} and the phosphomimetics.

Table 4-2. Experimental restraints and structure statistics of the ensemble of 20 best Nishigahara P_{CTD} solution structures.

P_{CTD} structural restraints (residue 193-296)	
Total number of NOE restraints used &	1560
Intra-residual	388
Sequential (|i-j|=1)	390
Short range (1<|i-j|<5)	316
Long range (|i-j|≥5)	466
NOE constraints per constrained residue	15
Dihedral angel restrains	
TALOS+	191
Structural statistics	
Average maximum upper distance violation	0.13
Average maximum angle violation (°)	3.26
Average *CYANA* target function (Å^2)	0.53
Ramachandran statistics #	
Most favored	87.5%
Additionally allowed	12.5%
Generously allowed	0.0%
Disallowed	0.0%
RMSD to mean structure	
Average backbone	0.46 ± 0.07
All heavy (non-hydrogen) atoms	0.99 ± 0.08

& Number of peaks from the final.upl file of an automated calculation in *CYANA*.

Statistics derived from the rama.ps file of an automated calculation in *CYANA*.

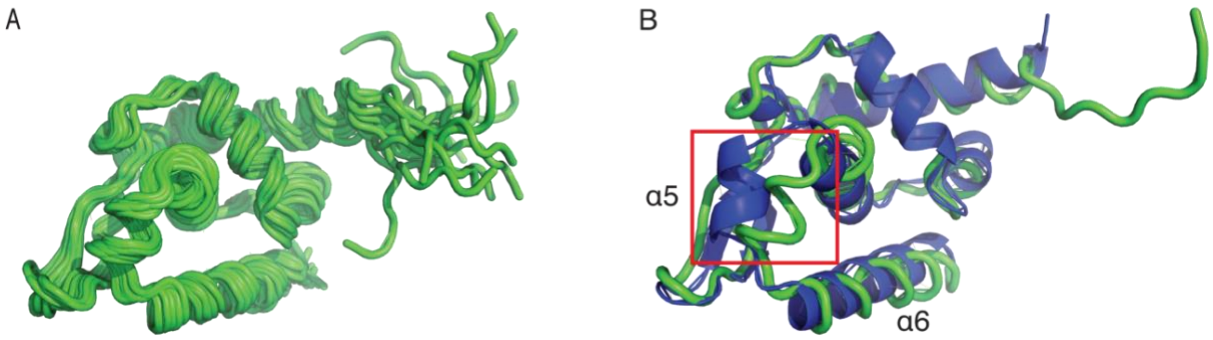


Figure 4-2. NMR solution structure of Nishigahara P_{CTD}.

A. An ensemble of 20 best structures of Nishigahara P_{CTD}. The conformation of the first 10 residues at the N-terminus (residues 183-192) is not well defined by the NMR data. Average RMSD for the ordered backbone atoms from residue 193 to 296 is 0.46 Å.

B. A superposition based on the alignment of the Nishigahara P_{CTD} structures solved from NMR (green) and X-ray crystallography (blue) with an RMSD of 1.5 Å. The major difference lies in the position of α5.

The nearly complete ^{1}H, ^{15}N and ^{13}C resonances assignment of Nishigahara P_{CTD} was reported (BMRB: 27498) with overall 98.7% backbone atoms and 92.9% proton assigned [143]. The solution structure of P_{CTD} was determined from 388 intra-residual, 390 sequential, 316 short range ($1<|i-j|<5$) and 466 long range ($|i-j|\geq5$) non-redundant distance restraints auto-assigned from 2D ^{1}H-^{1}H NOESY, 3D ^{15}N-edited NOESY, aromatic and aliphatic ^{13}C-edited NOESY, and 191 dihedral angle restraints predicted by *TALOS+* [132] using the backbone assignments, which corresponds to an average of nearly 17 restraints per residue (Table 4-2). Regions with fewer than average NOEs were found to be mostly located in the N- and C-terminus and the loop region (residues R_{218}-G_{221}) between the two β-strands (Supplementary Figure 4-2B). The best 20 low-energy structures from the final *CYANA* calculation are shown in Figure 4-2A to represent the solution structure of P_{CTD} with an average backbone RMSD of 0.46 Å and target function of 0.53 $Å^2$ over residues 193-296. The solution structure only slightly differs from

X-ray crystal structure (RMSD=1.5 Å) in the position of α5 (residues 272-276), which was packed more closely to the neighbouring α4 and α6 in the solution structure while being more extended towards the β-sheet in the crystal structure (Figure 4-2B). The similarity of crystal structures and the NMR solution structure supports that the NMR assignments are correct.

To assess the effect of phosphorylation on protein backbone, ^{1}H-^{15}N HSQC spectra were acquired on ^{15}N-labelled $S_{210}E$ and $S_{271}E$ and their chemical shift differences compared to the wild type P_{CTD} were plotted out in Figure 4-3A and Figure 4-3B respectively. $S_{210}E$ showed clear localized perturbation in the putative NES motif (residues 221-238) while $S_{271}E$ exhibited a relatively wide range effect on the P_{CTD} backbone. These chemical shift differences suggest conformational perturbations. Considering S_{210} was also the identified phosphorylation site regulating NES function [98], we chose to focus on $S_{210}E$ in the following NMR structural and dynamic study.

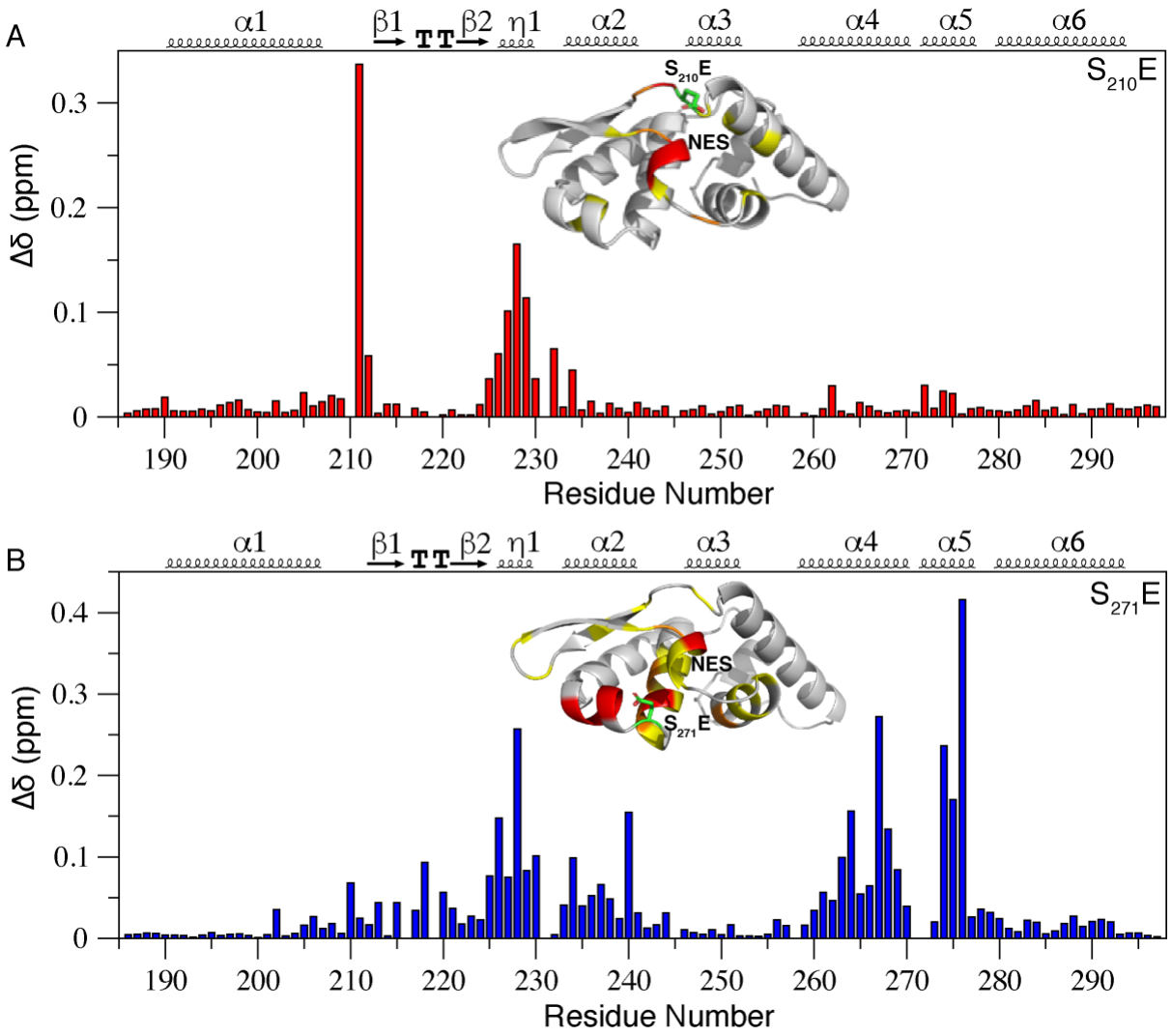


Figure 4-3. NMR characterization of the effect of phosphomimetic mutations $S_{210}E$ and $S_{271}E$ on P_{CTD} backbone.

A. Backbone amide chemical shift difference of $S_{210}E$ and wild type. $S_{210}E$ showed localized perturbation in the putative NES region (residues 221-238). The $S_{210}E$ structure is coloured according to the chemical shift differences of the last titration point (red, two standard deviation; orange, one standard deviation; yellow, mean), and the side chain of residue $S_{210}E$ is shown as sticks.

B. Backbone amide chemical shift difference of $S_{271}E$ and wild type. $S_{271}E$ exhibited stronger and wider range effect on P_{CTD} backbone than $S_{210}E$. The $S_{271}E$ structure is coloured according to the chemical shift differences of the last titration point as above, and the side chain of residue $S_{271}E$ is shown as sticks.

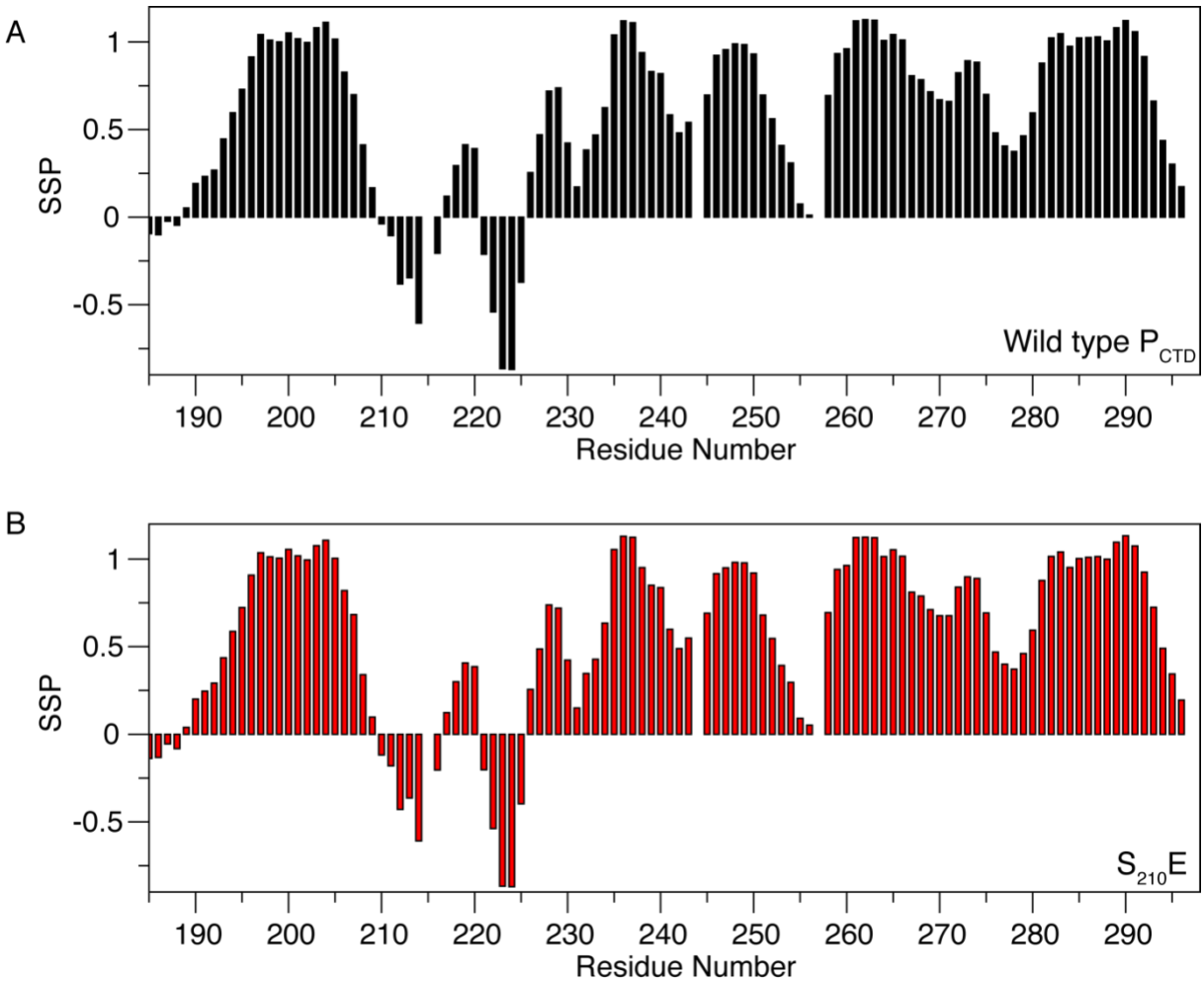


Figure 4-4. Comparison of secondary structure propensity of wild type P_{CTD} (A) and $S_{210}E$ (B).

A. Secondary structure of P_{CTD} analyzed using $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts by *SSP*.

B. Secondary structure of $S_{210}E$ analyzed in the same way as wild type, showing $S_{210}E$ shared the same secondary structure arrangement as wild type P_{CTD}.

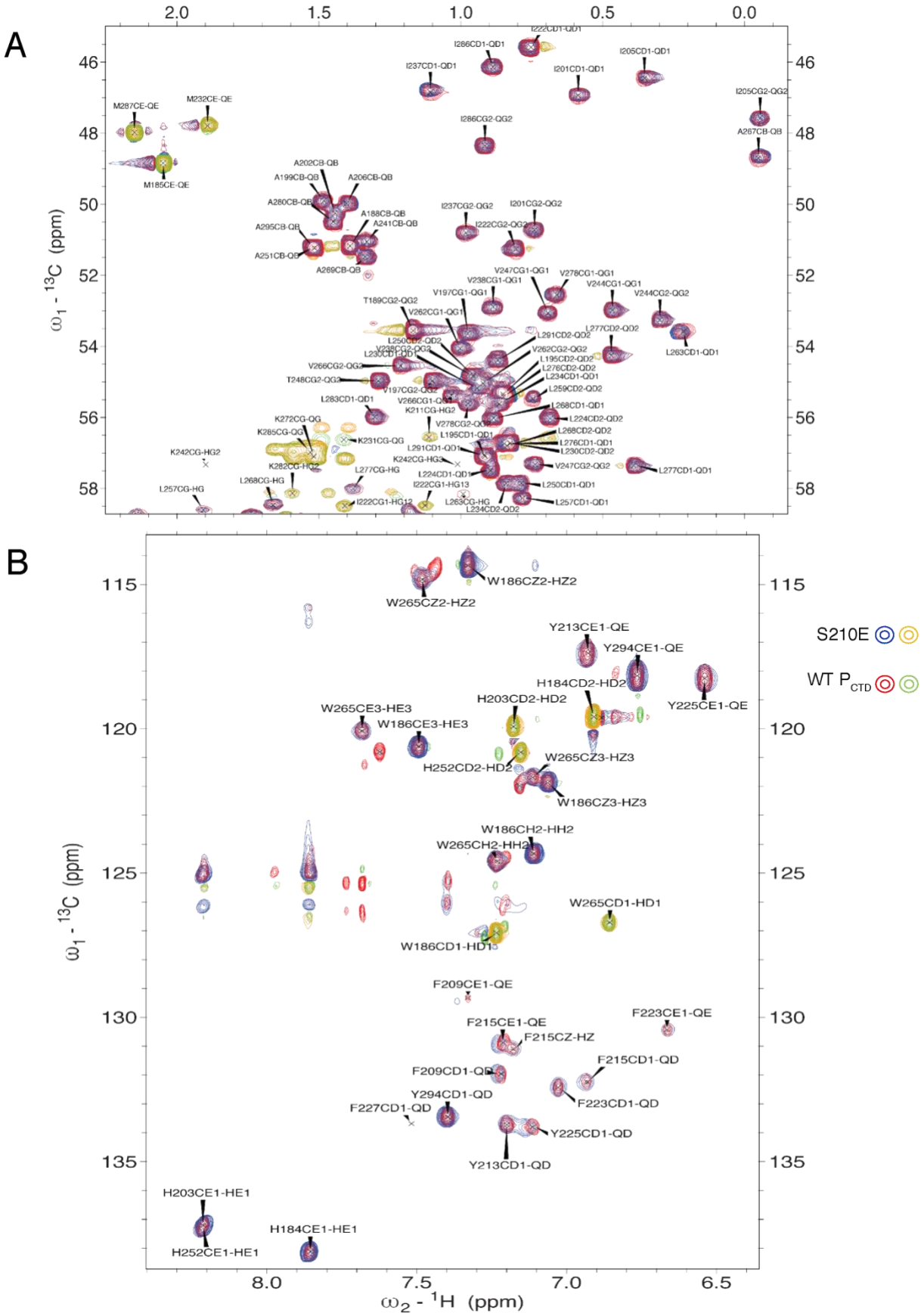

A
B
S210E
WT P$_{CTD}$
ω_1 - ^{13}C (ppm)
ω_2 - ^{1}H (ppm)
M287CE-QE
M232CE-QE
M185CE-QE
A202CB-QB
A199CB-QB
A206CB-QB
A280CB-QB
A188CB-QB
A241CB-QB
A295CB-QB
A251CB-QB
A269CB-QB
I286CD1-QD1
I237CD1-QD1
I201CD1-QD1
I205CD1-QD1
I286CG2-QG2
I237CG2-QG2
I201CG2-QG2
I222CG2-QG2
I205CG2-QG2
A267CB-QB
V247CG1-QG1
V278CG1-QG1
V238CG1-QG1
V244CG1-QG1
V197CG1-QG1
V244CG2-QG2
T189CG2-QG2
V262CG1-QG1
L277CD2-QD2
L263CD1-QD1
L291CD2-QD2
V262CG2-QG2
L195CD2-QD2
L276CD2-QD2
L234CD1-QD1
T248CG2-QG2
L259CD2-QD2
L283CD1-QD1
K211CG-HG2
L268CD1-QD1
L224CD2-QD2
K272CG-QG
K285CG-QG
K231CG-QG
L195CD1-QD1
L268CD2-QD2
L291CD1-QD1
L276CD1-QD1
L230CD2-QD2
K242CG-HG2
K242CG-HG3
K282CG-HG2
V247CG2-QG2
L277CD1-QD1
L277CG-HG
L224CD1-QD1
L268CG-HG
I222CG1-HG13
L250CD1-QD1
L257CG-HG
I222CG1-HG12
L263CG-HG
L257CD1-QD1
L234CD2-QD2
W186CZ2-HZ2
W265CZ2-HZ2
Y213CE1-QE
Y294CE1-QE
H184CD2-HD2
H203CD2-HD2
Y225CE1-QE
W265CE3-HE3
W186CE3-HE3
W265CZ3-HZ3
H252CD2-HD2
W186CZ3-HZ3
W186CH2-HH2
W265CH2-HH2
W265CD1-HD1
W186CD1-HD1
F209CE1-QE
F223CE1-QE
F215CE1-QE
F215CZ-HZ
F215CD1-QD
F209CD1-QD
F223CD1-QD
Y294CD1-QD
F227CD1-QD
Y225CD1-QD
Y213CD1-QD
H203CE1-HE1
H184CE1-HE1
H252CE1-HE1

Figure 4-5. Spectra overlay of the selected methyl region (A) and aromatic region (B) of the assigned wild type P_{CTD} and $S_{210}E$.

A. Selected methyl region of the aliphatic constant time ^{1}H-^{13}C HSQC spectra of $S_{210}E$ overlaid with wild type P_{CTD} showing nearly identical chemical shifts. The spectra were recorded with ^{13}C offset at 40 ppm, spectra width of 32 ppm and constant time period of 13.3 ms. The ^{13}C nuclei coupled to odd number of carbons (e.g. $C^{\delta 1}/H^{\delta 1}$ of Leu) appear with opposite sign with those coupled to even number of carbons (e.g. $C^{\varepsilon}/H^{\varepsilon}$ of Met). The ^{13}C dimension of the methyl region showing here is folded and a spectra width of 32 ppm should be subtracted to get the accurate ^{13}C chemical shifts of the methyl groups.

B. Aromatic constant time ^{1}H-^{13}C HSQC spectra of $S_{210}E$ overlaid with wild type P_{CTD} showing nearly identical chemical shifts of the aromatic rings. The spectra were recorded with ^{13}C offset at 123 ppm, spectral width of 32 ppm and constant time period of 8.3 ms. The ^{13}C nuclei coupled to odd number of carbons (e.g. $C^{\delta 2}/H^{\delta 2}$ of His) appear with opposite sign with those coupled to even number of carbons (e.g. $C^{\varepsilon 1}/H^{\varepsilon 1}$ of His).

^{1}H, ^{15}N and ^{13}C resonances of $S_{210}E$ were assigned using the same experiments under the same conditions as for wild type P_{CTD}. Secondary structure analysis based on $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts indicated that $S_{210}E$ shared the same secondary structure arrangement as wild type P_{CTD} (Figure 4-4). As the hydrophobic side chains of the NES are buried, their chemical shifts will be exquisitely sensitive to the local chemical environment, we further compared the chemical shifts of the methyl groups and aromatic side chains of $S_{210}E$ and wild type P_{CTD}. Within the putative NES motif (residues 221-238), there are 7 residues with 13 methyl groups (I_{222}, L_{224}, L_{230}, M_{232}, L_{234}, I_{237} and V_{238}) and 3 aromatic residues (F_{223}, Y_{225} and F_{227}) which make extensive NOE connections and were critical for P_{CTD} structure. For example, the methyl protons H^{ε} of M_{230} were in close contact with the H^{δ} of the phenyl ring of F_{227} (4.7 Å) and the two methyl groups of L_{230} are closely facing the aromatic ring of Y_{225} (4.3 Å, 3.4 Å). If phosphorylation of S_{210} does expose the NES, the chemical environment of these aromatic

rings and especially the methyl groups within the NES would change significantly because of the ring current effect, thus shifts of their peak positions on the spectra would be expected. Figure 4-5A shows the methyl region of the aliphatic ^{1}H-^{13}C HSQC spectra of $S_{210}E$ overlaid with wild type P_{CTD} and Figure 4-5B shows the overlaid aromatic ^{1}H-^{13}C HSQC spectra. Both overlays show nearly identical chemical shifts, suggesting no structural difference between $S_{210}E$ and wild type P_{CTD}.

4.3.3 Dynamic study of $S_{210}E$ and wild type P_{CTD}

Protein function is determined by the interplay between structure and dynamics. Up until now our structural study using X-ray crystallography and NMR has focused on the ground-state, low-energy conformers. In many cases these ground-state low-energy conformers are in equilibrium with weakly populated, excited-state higher energy conformers, which only transiently exist, thus being invisible to structural study techniques, but are known to play critical roles in biological functions [198]. Typically, ligand binding or PTMs shift the equilibria such that these high-energy states are visible to protein NMR dynamics techniques.

Conformational exchange between different protein states on micro-millisecond timescales may produce an additional transverse relaxation term (R_{ex}) to the transverse relaxation characteristic of the tumbling motion of the protein. Referred to as chemical exchange, this additional component to transverse relaxation arises from residues going through conformational exchange processes within the molecular frame of the protein, and may exhibit higher transverse relaxation rates (R_2) compared to the neighbouring residues. These effects are most often observed in ^{15}N R_2 experiments [199, 200]. Figure 4-6 shows the plots of the transverse relaxation rate R_2 of P_{CTD} and $S_{210}E$ measured at 15 °C (A) and 25 °C (B). The flexible N- and C-terminus and the loop region (S_{220} and G_{221}) exhibited lower than average R_2

values indicative of a rapid pico-nanosecond internal motion. The relatively uniform high R_2 values of the middle region suggested this well-structured globular domain is experiencing overall molecular tumbling. There was no sign of chemical exchange for either wild type P_{CTD} or the mutant $S_{210}E$ at the two different temperatures.

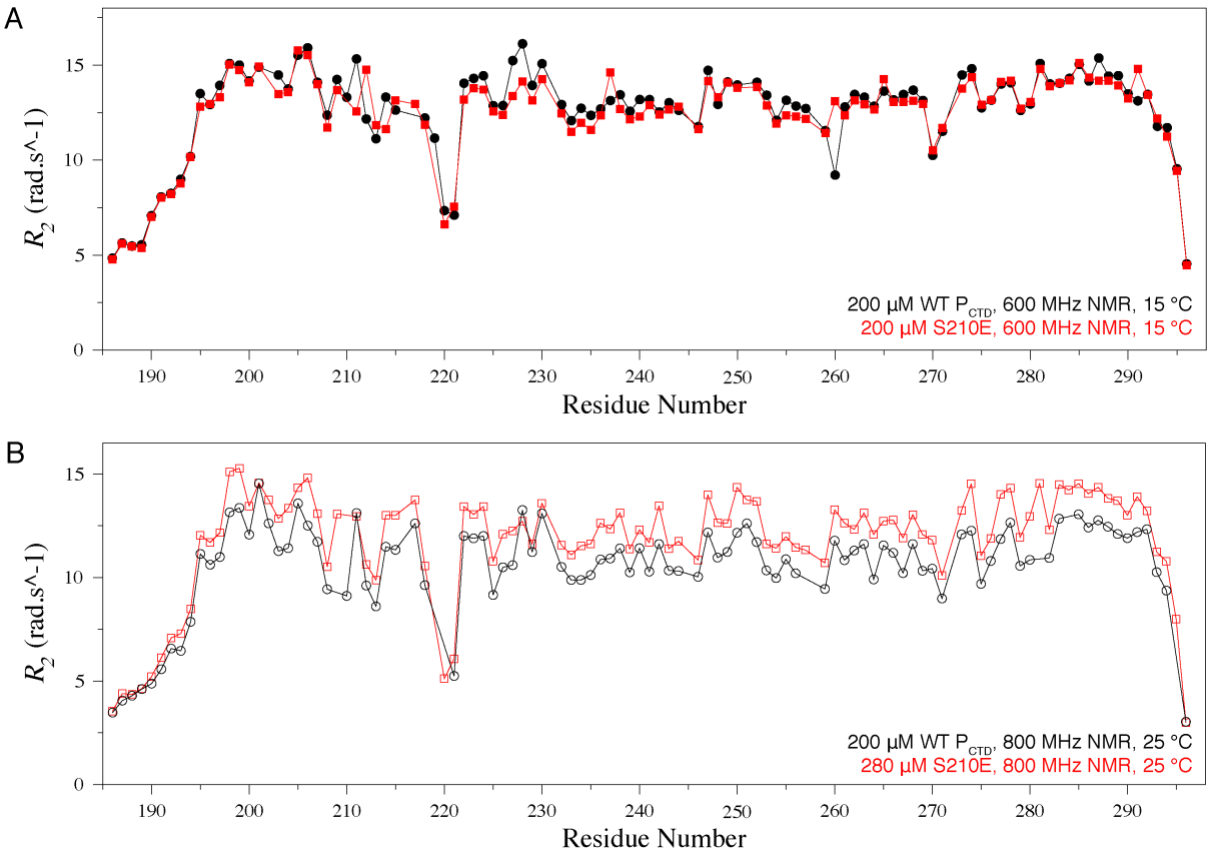


Figure 4-6. Transverse relaxation rate R_2 of wild type P_{CTD} and $S_{210}E$ measured at 15 °C (A) and 25 °C (B).

The R_2 profile of wild type P_{CTD} is shown by black circles and $S_{210}E$ is shown by red squares. No sign of chemical exchange was observed for either P_{CTD} or $S_{210}E$ at two different temperatures.

In an effort to confirm there is no conformational exchange on the long micro- to millisecond timescales within P_{CTD} or $S_{210}E$, we also performed ^{15}N CPMG relaxation dispersion experiments on ^{15}N-labelled P_{CTD} and $S_{210}E$ at 3, 15 and 25 °C. CPMG relaxation dispersion experiments are widely used for probing invisible states in exchange particularly on these timescales [198, 201, 202]. The magnitude of the chemical exchange contribution R_{ex} to R_2 can be systemically modulated via the CPMG pulse train comprising refocusing CPMG pulses applied with different repetition rates (ν_{CPMG}): by increasing the frequency of the CPMG pulse (ν_{CPMG}), the contribution of R_{ex} is reduced therefore a lower effective transverse relaxation rate $R_{2,eff}$ is observed. When $R_{2,eff}$ is plotted against ν_{CPMG} as the relaxation dispersion profile, it can be fitted into a conformational exchange model to extract information about the exchange rate or chemical shift difference of a nuclei in two different states. Here, the relaxation dispersion profiles of all the residues from wild type P_{CTD} and the mutant $S_{210}E$ measured at the three different temperatures are all flat (Figure 4-7), show consistent $R_{2,eff}$ values at 13 different CPMG repetition rates, indicating no chemical exchange contribution to R_2, which supports that there is no conformational change in wild type P_{CTD} or the mutant $S_{210}E$.

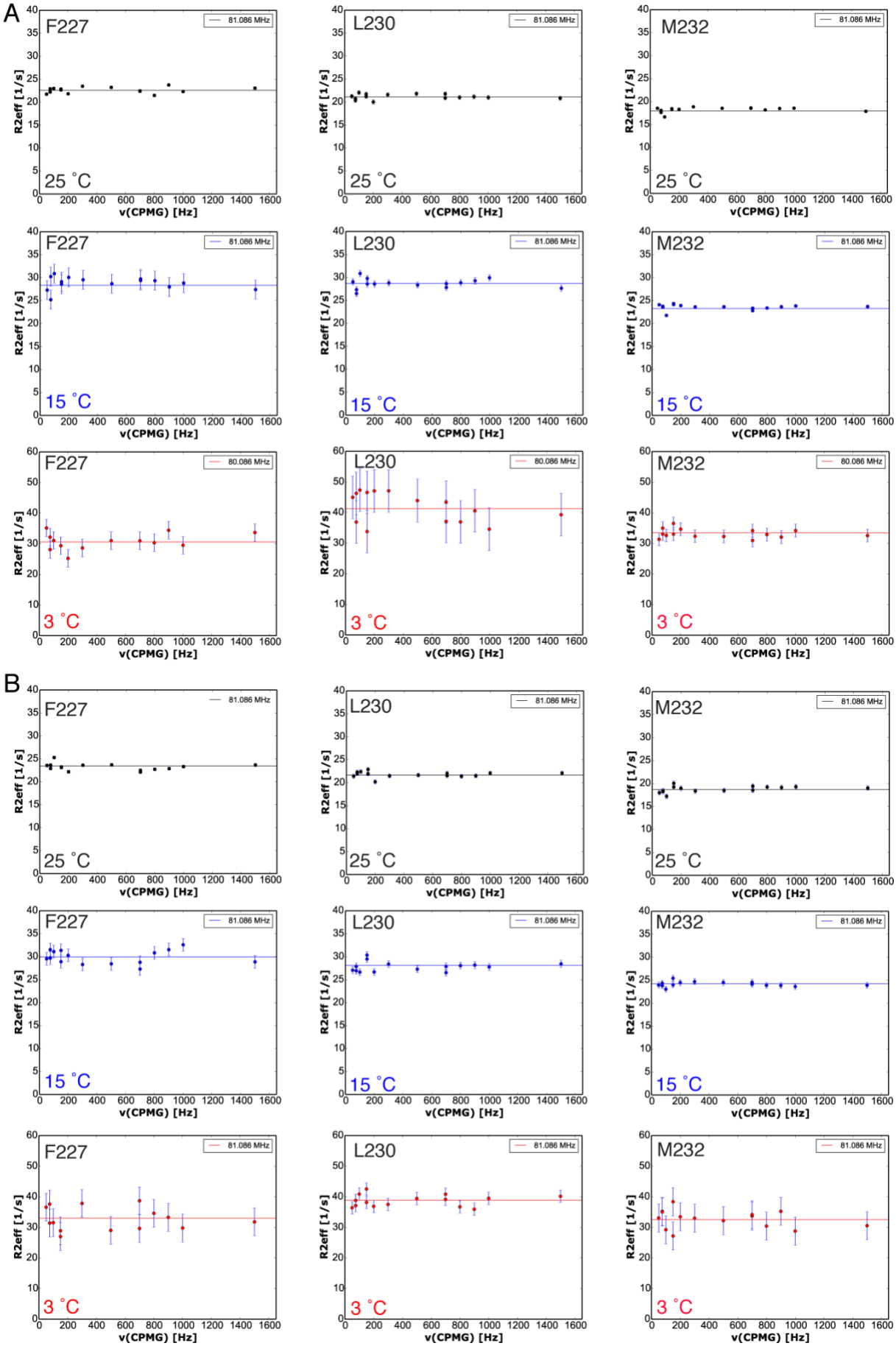

A
F227
L230
M232
81.086 MHz
80.086 MHz
25 °C
15 °C
3 °C
R2eff [1/s]
v(CPMG) [Hz]
B
F227
L230
M232
81.086 MHz
25 °C
15 °C
3 °C
R2eff [1/s]
v(CPMG) [Hz]

Figure 4-7. CPMG profile of wild type P_{CTD} (A) and $S_{210}E$ (B) measured at 3 °C (red), 15 °C (blue) and 25 °C (black).

Critical residues within the NES motif F_{227}, L_{230} and M_{232} were shown in the first, middle and last lane respectively. The relaxation dispersion profiles of the three residues from P_{CTD} and $S_{210}E$ are all flat, showing consistent $R_{2,eff}$ values at 13 different CPMG repetition rates, indicating no chemical exchange contribution to R_2.

4.3.4 Assessment of replication function: P_{CTD}/N-pep interaction

One of the many important functions of P protein is aiding the binding of the polymerase L protein to the N-RNA template. The process is through the interaction between the P_{CTD} and the C-terminal flexible loop of the N protein (N-pep). Despite the fact that no conformational change was found in the structural and dynamic study of P_{CTD} and its phosphomimetics, we wondered if phosphorylation of P_{CTD} has any impact on its role as the polymerase cofactor and binding with N-RNA. This question is important as even if no structural change occurs, the introduction of a negative charge near the positive patch that is critical for binding to N protein may disrupt an interaction.

The interaction between N-pep and the phosphomimetics of P_{CTD} was studied using NMR chemical shift perturbation experiments and the binding kinetics were obtained from saturation binding curve analysis and two-dimensional lineshape analysis performed on the titration data (as described in chapter 3). $S_{210}E$ and $S_{271}E$ bound to N-pep with a similar affinity, less than two-fold weaker compared to the wild type P_{CTD}, due to a slightly slower on-rate and faster off-rate than wild type (Table 4-3). The similar binding affinity shown among wild type and the phosphomimetics of P_{CTD} indicated that phosphorylation states of P_{CTD} may not have an impact or be regulatory of its interaction with N protein.

Table 4-3. Kinetics data of wild type P_{CTD}, $S_{210}E$ and $S_{271}E$ binding to N-pep.

K_D values were determined by saturation binding curve analysis using *Xcrvfit* [151] and lineshape analysis using *TITAN* [153]; on-rate and off-rate were determined from lineshape analysis using *TITAN*.

P_{CTD}	K_D (μM)	K_D (μM)	k_{off} (s^{-1})	k_{on} (s^{-1} μM^{-1})
	Xcrvfit		*TITAN*	
Wild type	88 ± 4	79 ± 3	2000 ± 100	25
$S_{210}E$	140 ± 6	123 ± 3	2200 ± 100	19
$S_{271}E$	149 ± 22	123 ± 4	2300 ± 200	19

4.4 Discussion

We studied Nishigahara P_{CTD} and its phosphomimetics $S_{210}E$ and $S_{271}E$ using NMR and X-ray crystallography, and found no structural, dynamic or functional importance that phosphorylation may have on P_{CTD} in this study. The crystal structure of Nishigahara P_{CTD}, $S_{210}E$ and $S_{271}E$ were solved to 1.5 to 2 Å resolution and showed no difference between the wild type and the phosphomimetics. The solution structural study of P_{CTD} and $S_{210}E$ confirmed the crystal structures solved by X-ray crystallography, indicating that there is no significant conformational change upon phosphorylation. The dynamics of wild type P_{CTD} and the mutant $S_{210}E$ complemented the previous structural study and contributed to a more comprehensive understanding of P_{CTD} properties, suggesting that P_{CTD} is a well-structured globular domain with no internal micro-millisecond motions, which ruled out the possibility of the presence of sparsely-populated high-energy excited state conformers in either unphosphorylated or phosphorylated P_{CTD}. The result of these studies strongly argues against the hypothesis of unmasking NES by phosphorylation-induced conformational change, and calls for further investigations into the mechanism behind the different nucleo-cytoplasmic distribution of

phosphorylated and unphosphorylated P_{CTD} [98]. Our experiments with titrating $P_{CTD,}$ $S_{210}E$ and $S_{271}E$ with the N-pep showed a similar binding affinity to the C-terminal loop of N protein (N-pep) suggesting that the single binding site model of P_{CTD}/N-pep is not affected by the phosphorylation state of P_{CTD}.

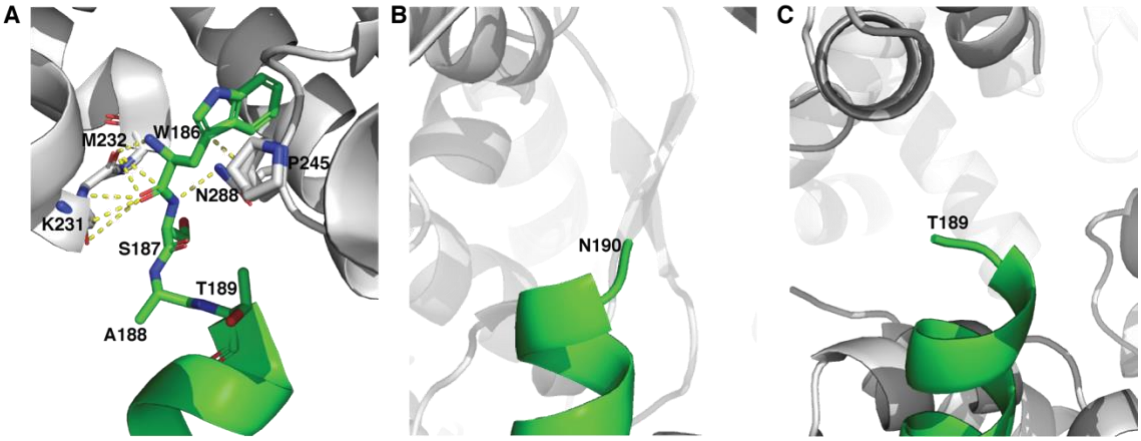
Thus, our conclusions raise the issue to the significance of phosphorylation of P_{CTD} in the viral life cycle. Firstly, does P_{CTD} get phosphorylated in cells at all? Earlier studies reported that infecting insect cells with recombinant baculovirus yielded unphosphorylated P protein [203], while RABV N protein was phosphorylated at S_{389} not only in virus-infected cells [204] but also in insect [205] and mammalian cells [80, 206] when expressed alone. Nonetheless, physiological phosphorylation is dynamic and is often transient in response to various stimuli or cellular events. The identified isomers of PKC [90] responsible for P_{CTD} phosphorylation are staurosporine-sensitive and are regulated under two independent mechanisms: by maturation through multisite phosphorylation, and by activation of phorbol ester or diacylglycerol which removes the pseudosubstrate from the catalytic site [207, 208]. Because of such a complex dual regulation system, PKC has a weak basal activity, but it can be activated by type I [209] and type II interferons [210] and mediate STAT1 phosphorylation [209, 210], which are the key components of host immune response to viral infection [211]. PKC was found to become activated in the case of herpesviruses infection [212] and was implicated in the entry process of some enveloped viruses from *Togaviridae*, *Poxviridae*, *Herpesviridae* and *Rhabdoviridae* families [212-214]. Hence, the unphosphorylated form of P yielded from insect cells may simply be the result of lacking the suitable phosphorylation condition or being dephosphorylated by active cellular phosphatases.

Secondly, if phosphorylation of P_{CTD} does happen in infected cells, under what circumstances does it happen; how many different phosphorylation states exist; and what are the functional roles for each of them? The only functional description reported so far is that phosphorylation at S_{210} changes the nucleo-cytoplasmic distribution of P_{CTD} possibly by unmasking NES and/or causing structural changes [98]. Our structural and dynamic studies using both X-ray crystallography and NMR strongly argues against this structural change exposing the NES as we found no alternative state in wild type or phosphomimetics of P_{CTD} in any condition, the NES remains buried with no sign of chemical exchange, and the mechanism behind the more cytoplasmic distribution of phosphorylated P_{CTD} remains unclear. However, in the previous study investigating the relationship between P_{CTD} phosphorylation and its nucleo-cytoplasmic distribution [98], only the phosphorylation state of S_{210} is carefully controlled and examined using phospho-deficient mutant $S_{210}A$ and phospho-mimetic mutant $S_{210}D$ while the phosphorylation state of S_{271} is unknown. Considering the spatial proximity of both S_{210} and S_{271} to the putative NES motif and the expectation that they are phosphorylated by the same category of PKC isomers, it is possible that the biologically relevant phosphorylation form of P_{CTD} is the S_{210}/S_{271} double phosphorylated version in which the critical hydrophobic NES residues might be exposed and able to bind to CRM-1 to be exported out of the nucleus. Further, one could also question the effectiveness of using Glu as a phosphomimetic in our study. The more definite answer to which phosphorylation states are biologically relevant and whether conformational change exists in any of them could be found from a structural and dynamic study using PKC-phosphorylated P_{CTD}.

As an alternative to phosphorylation exposing the NES, phosphorylation may enable an initial interaction with CRM-1 that results in the exposure of NES. The negative charge acquired after P_{CTD} phosphorylation might be vital in initiating and orientating CRM-1 in the interaction, as it

was implied that in the typical amphipathic NES helix, the negatively charged residues on the opposite side of the hydrophobic residues are important for creating the initial contact with CRM-1 [215], leading to induced local unfolding upon interaction with CRM-1 which allows the hydrophobic NES residues to be accessible. Another possibility to consider is that S_{210} phosphorylation could inactivate or weaken the NLS that is immediately after S_{210}. The NLS, comprises the positive patch residues KKYK [216], and the negative charge positioned adjacent to the NLS region after S_{210} phosphorylation might cause charge repulsion and interrupt NLS/importin interaction. This possibility could be investigated by experiments measuring the affinity between importin and P_{CTD} with and without PKC phosphorylation.

In considering the impact of phosphorylation on viral replication, we found that phosphorylation at either S_{210} or S_{271} has little impact on the N-pep/P_{CTD} interaction (less than 2-fold weakening of K_D). This agrees well with the *in vivo* study published recently [67] in which the three PKC phosphorylation sites S_{162}, S_{210} and S_{271} on full-length P were mutated to either Ala or Asp and none of these mutations affected the ability of P protein to bind to N-RNA complex and form the liquid droplets of viral replication factories in the cytoplasm [67]. It also validated our previous P_{CTD}/N-pep interaction study, suggesting that the micromolar-affinity single binding site model of P_{CTD}/N-pep is not affected by the phosphorylation states of P_{CTD}, and that the replication process of rabies virus does not require P_{CTD} phosphorylation. The questions with regard to the actual phosphorylation states of P_{CTD} in virus infected cells need to be answered and their functional roles remain to be firmly established.

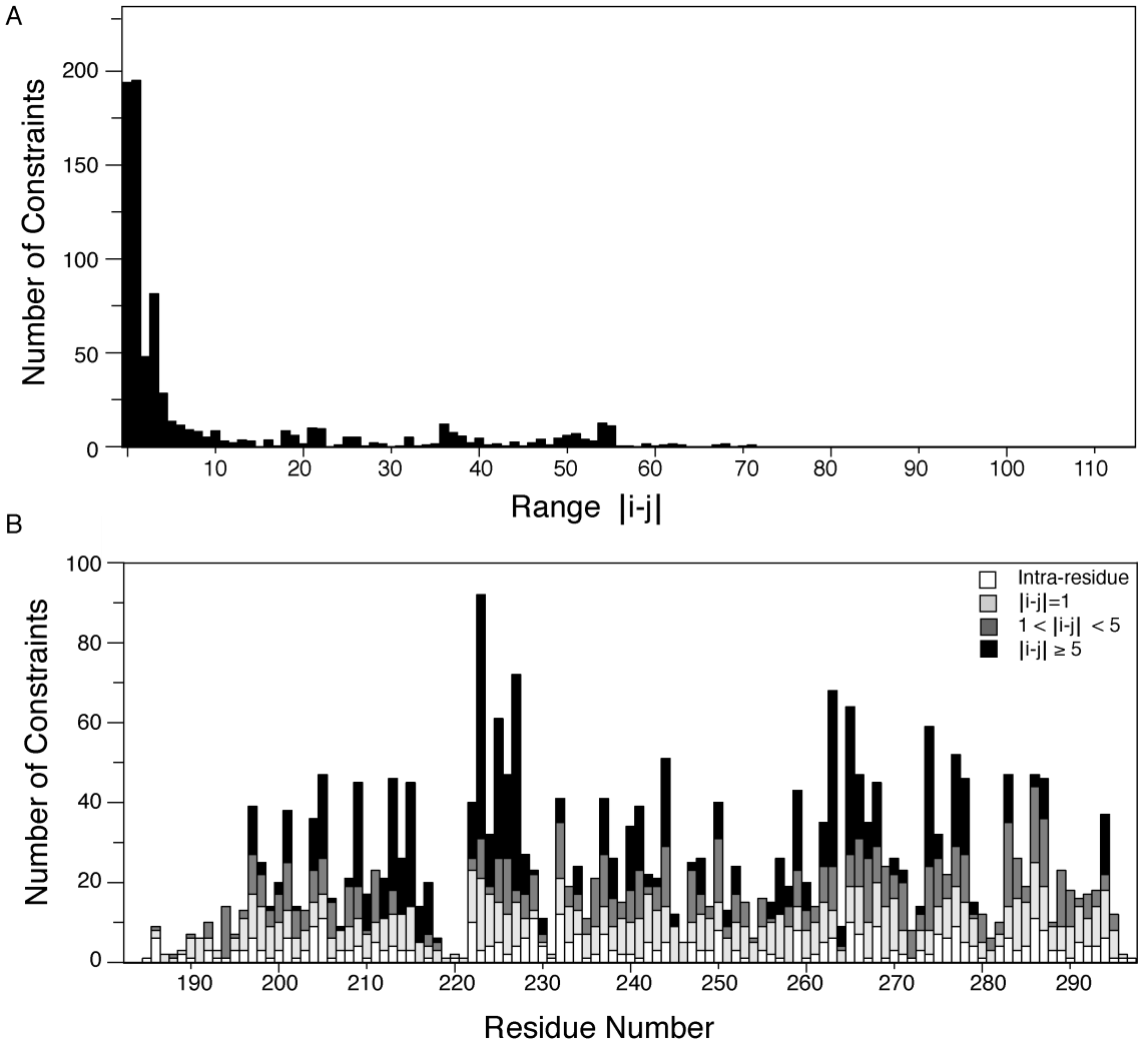


Supplementary Figure 4-1. The N-terminus of P_{CTD} from CVS strain is stabilized by multiple crystal contacts.

A. P_{CTD} from CVS strain crystalized in space group $P3_121$ and the structure was refined at 1.5 Å (PDB: 1VYI). The N-terminus (coloured in green with residues $_{186}$WSAT$_{189}$ labelled) is stabilized by multiple crystal contacts.

B. P_{CTD} from Nishigahara strain crystalized in space group $P1$. The N-terminus (green) is exposed and equivalent crystal contacts are not present.

C. P_{CTD} from Nishigahara strain crystalized in space group $P2_12_12_1$. The N-terminus (green) is exposed.

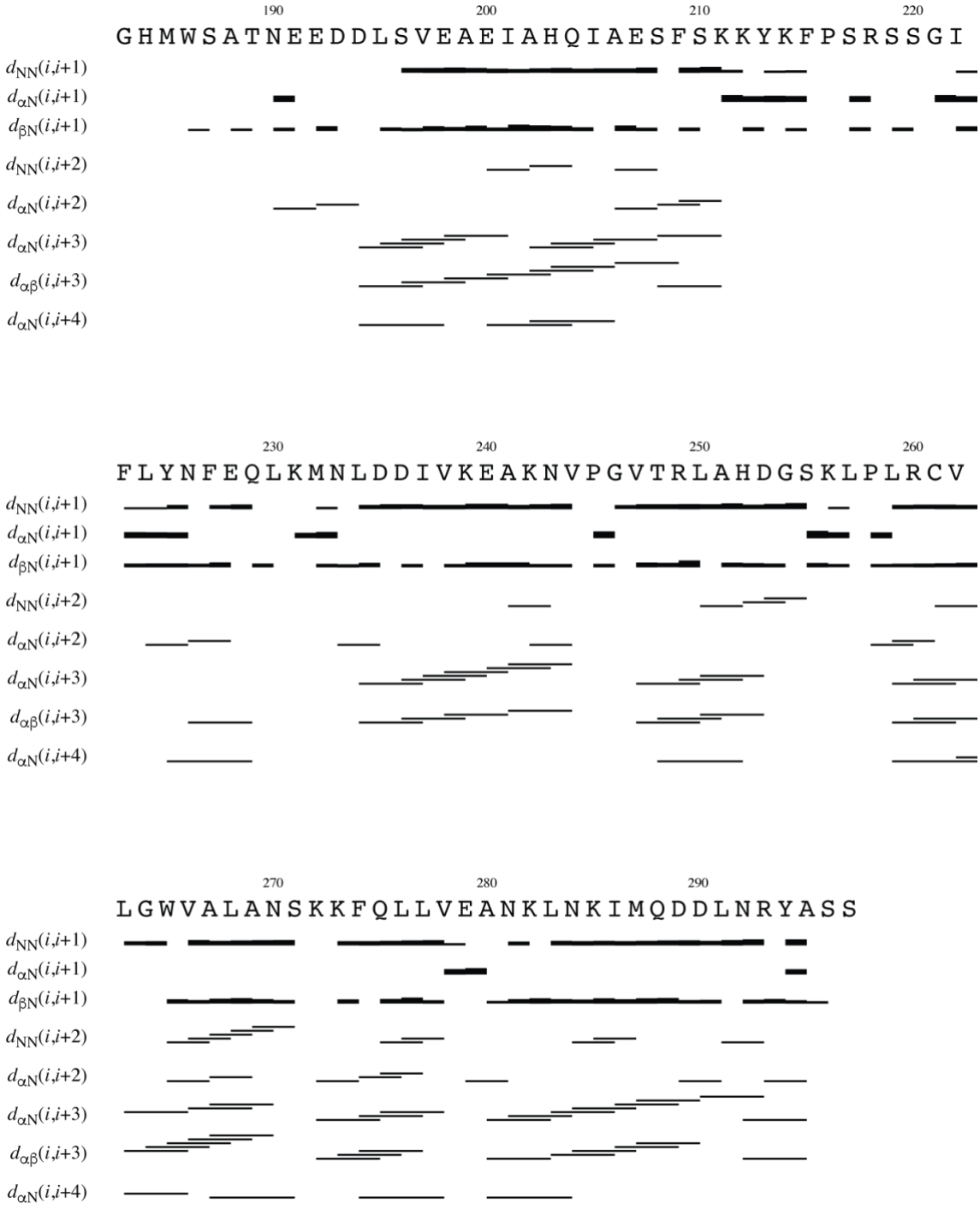


Supplementary Figure 4-2. Details of the distance restraints used in the Nishigahara P_{CTD} NMR structure calculation.

A. Distribution of the NOE restraints for residue range of |i-j|.

B. Histogram of the distribution of NOE restraints per residue. The NOE restraints are colored and categorized as follow: ☐ intra-residual, ☐ sequential (|i-j|=1), ▒ short range (1<|i-j|<5), ■ long range (|i-j|≥5).

The figure was generated using the macro "docstat" within *CYANA* (version 2.1).



Supplementary Figure 4-3. Connectivity of the NOE restrains per residue.

The figure was generated using the macro “seqplot” within *CYANA* (version 2.1).

Chapter 5:

General Discussion

General discussion

Rabies virus (RABV) is a non-segmented negative-sense RNA virus causing fatal encephalitis with no treatment. With the limited coding capacity of its 12 kb genome, RABV expresses the multifunctional phosphoprotein (P) to achieve functional diversity. P protein has well-structured domains separated by disordered regions, with overlapping functional sequences mapped onto them. The C-terminal domain of P protein (P_{CTD}) itself is an interaction hub with nuclear import and export sequences, PTMs, binding sites for both viral and host cellular proteins, and is involved in viral replication and innate immunity evasion. Using NMR and other biophysical techniques, we studied the structure and dynamics of $P_{CTD,}$ and its interaction with N protein to understand its role in viral replication.

The X-ray crystal structure of P_{CTD} from the pathogenetic Nishigahara strain was solved to 1.5 Å resolution, and was in good agreement with an available crystal structure of P_{CTD} from CVS strain (PDB: 1VYI) [111]. Since reasonably comprehensive NMR resonance assignments are a prerequisite for further NMR structural, dynamic and functional studies, the ^{1}H, ^{15}N and ^{13}C resonances of Nishigahara P_{CTD} were assigned and deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number 27498 [143]. Overall, 98.7% of the backbone resonances and 92.9% of the proton resonances were unambiguously assigned. Around 87.9% of the aromatic side chain ^{1}H and ^{13}C resonances and 98.4% of the methyl groups were identified, among which all the 39 identified methyl groups from Val and Leu were stereo-specifically assigned. The secondary structure propensity analysis based on the backbone assignments agreed well with its X-ray crystal structure.

The resonance assignments of P_{CTD} enabled solution structural, dynamic studies and the effects of post-translational modifications. An ensemble of 20 best P_{CTD} structures with an average

backbone RMSD of 0.46 Å and target function of 0.53 $Å^2$ was determined from an automated *Cyana* calculation [192, 194], in which a total of 1560 non-redundant distance restraints and 191 dihedral angle restraints were used, corresponding to an average of nearly 17 restraints per residue. The solution structure of Nishigahara P_{CTD} is consistent with its X-ray crystal structure, which in turn corroborates the chemical shift assignments. A preliminary NMR dynamics analysis suggested that P_{CTD} is a well-structured globular domain with no significant internal micro-millisecond motions or alternative sparsely-populated high-energy conformers.

Phosphorylation is a major reversible post-translational modification which may induce conformational transitions and/or regulate protein functions. It is reported that P_{CTD} can be phosphorylated by PKC at conserved S210 and S271 [90], and that phosphorylation at S210 or its phosphomimetic mutation could shift the nucleo-cytoplasmic distribution of P_{CTD} towards nuclear export [98] possibly by inducing conformational change to P_{CTD} and unmasking the originally buried NES, thus simultaneously regulating NLS and NES functions. However, by comparing the X-ray crystal structures, NMR chemical shifts of the methyl groups and aromatic rings, and internal motions on the long micro-millisecond timescales, we found no evidence suggesting that the phosphomimetic mutation affected the structure or the backbone dynamics of P_{CTD}. Questions regarding if P_{CTD} ever gets phosphorylated, and how phosphorylation regulates its nucleo-cytoplasmic distribution and overall functions need to be further investigated.

The chemical shift assignments of P_{CTD} also enabled characterization of the protein-binding interface. Critical to viral replication and transcription, the P_{CTD} bears the N-RNA binding site. This interdomain interface has been studied previously by low resolution negative staining [73], yeast two-hybrid screens [96, 103] and mammalian cell-based assays [102], from which

the binding sites have been localized to a positively charged patch ($_{211}$KKYK$_{214}$) [96, 102, 103] and a hydrophobic pocket ('W-hole', which includes C261, W265, M287) [96, 103] of P_{CTD} and the C-terminal loop region of N protein [73]. Based on the SAXS data, and the information of W-hole and positive patch being N-binding sites, a N-RNA/P_{CTD} complex model was generated, which also remains to be the only data related to the structure of this complex so far. In this model, the P_{CTD} was proposed to lie on the C-terminal top of the circular N-RNA ring, with the N_{CTD} loops from two neighbouring N protomers contacting the positive patch and W-hole of P_{CTD}. Specifically, K211, K212 and K214 of P_{CTD} were predicted to be forming ionic bonds with D378, D383 and D384 of N, and the affinity was reported to be 160 ± 20 nM measured by SPR.

However, this model is controversial as there is little genuine data supporting the role of the W-hole in N protein binding. While the importance of the positive patch is consistently implicated by random or targeted mutagenic studies performed with MOKV and RABV P_{CTD} [96, 102, 103], there are contradicting findings regarding the role of W-hole in N-binding. W-hole was only identified [111] in a random mutagenic study of MOKV P_{CTD} [103] where P_{CTD} constructs with defective N-binding containing W-hole mutations were identified, but these were rare and always appeared with positive patch mutations or destabilizing mutations. Importantly, targeted W-hole mutations were later found to have little impact on N-binding or replication [102]. Therefore, the exact nature of the N-RNA/P_{CTD} interaction needs to be defined.

The N-RNA/P_{CTD} interaction was studied on an atomic level in this study. NMR chemical shift mapping of the binding interfaces suggested a single binding mode for N-pep at the positive patch of P_{CTD} and the flexible loop region of N-pep, which provides the first direct

confirmation of binding sites. The indole side chain of the W-hole residue W265, which is exposed on the P_{CTD} surface and is readily accessible, is not perturbed by the presence of N-pep, suggesting that the W-hole does not have a major role in N-binding. This is consistent with the previous mammalian cell-based assays [96] mentioned before, and thus strongly questions the accuracy of the SAXS model. Phosphomimetic mutants of P_{CTD} (S210E, S271E) showed a similar interaction with N-pep, suggesting PKC phosphorylation of P_{CTD} is not required for viral replication, which also agrees with the recent finding that phosphorylation-deficient mutations in the PKC phosphorylation sites did not impair the ability of P protein to bind to N-RNA and form viral replication factories (Negri bodies) in the cytoplasm [67].

The interaction interface was further examined by an alanine scan of the flexible loop residues of N-pep to determine the specific residues contributing to P_{CTD}-binding, and the results were not completely in line with the predictions of the SAXS model either. Among the three aspartate residues of N that are implicated in P_{CTD}-binding by the SAXS model [83], D378 did not appear essential for the interaction, while D383 and D384 were confirmed to be important. Furthermore, L381 was also found to be critical in this study, the function of which has not been predicted previously and is first identified here. CK-II phosphorylation of S389 of N brought a 7-fold improvement to the N-pep/P_{CTD} binding affinity, which is consistent with the previous literature reporting that phosphorylation of N protein enhanced P-binding [81, 83] and viral transcription/replication [206]. The key residues of N identified in this study are in good agreement with a previous study where two highly conserved acidic regions in the vicinity of S389 (residues $_{373}$EE$_{374}$ and $_{390}$DEED$_{393}$ of MOKV N) were found to be dispensable for the interaction [96]. From this mutagenesis data, it would be expected that D383, D384 and phosphorylated S389 of N make electrostatic interactions with the P_{CTD}

positive patch residues, and L381 of N makes a hydrophobic interaction with L224 of P_{CTD}. As a previous targeted-mutagenic study of MOKV N protein failed to identify any critical N residues for P_{CTD}-binding [96], this would be the first time that the key N residues in the N-RNA/P interface have been experimentally identified. All the residues identified within the N-pep/P_{CTD} interface, for both proteins, are conserved among lyssaviruses, which is consistent with their critical roles in the viral replication process. In theory, the N-pep binding interface identified here can be used as a potential target for anti-viral drug design. For instance, in silico fragment screening and SPR-based fragment screening combined with NMR can be used to identify small novel compounds that bind to P_{CTD} with a stronger affinity than N-RNA complex, thus block the RABV replication process. However, the positive patch of P_{CTD} is a flat site which is not an ideal target for fragment screening, and in fact a preliminary SPR-based fragment screening combined with NMR performed by Siqiong Zheng (unpublished data) did not find any fragment that binds to the P_{CTD} positive patch specifically. The N_{CTD} loop, on the other hand, is disordered and thus may be difficult to target as well.

Using saturation binding [151], line-shape analysis [153] and ITC [162], the kinetic and thermodynamic properties of the interactions of the P_{CTD} and N-pep variants were analysed. The interactions were all solely driven by favourable enthalpy contributions, which overcompensated the unfavourable entropic penalty likely associated with the disorder-to-order transition of N-pep upon binding. The conformational transition of N-pep, suggesting coupled folding, is also indicated by the extensive broadening of N-pep resonances in the titration, although definite demonstration of this conformational change will ideally require more direct structural and dynamic evidence. While the analysis of the on-rate, off-rate and the enthalpy-entropy compensation relationship enhanced the understanding of the

N-RNA/P_{CTD} interaction, it could also guide the development of high affinity binders for the positive patch of P_{CTD} as anti-viral drugs.

The affinity between the P_{CTD} and cyclized phosphorylated N-pep was 7.4 ± 0.4 μM measured by NMR titration, which suggests a weak interaction. This is consistent with the reported liquid behaviour of the viral factories (Negri bodies) formed by N and P in the cytosol [67], and the observation that P_{CTD} bound to the N-RNA ring in a flexible manner using negative stain EM [73]. However, such a weak and potentially dynamic interaction differs to the high nanomolar affinity between P_{CTD} and N-RNA ring measured by SPR reported together with the SAXS model [83]. This discrepancy might arise from the major difference in the length of the reactant and the technique applied. As we were unable to produce the full-length N protein in *E. coli*, a 57-residue segment from the N_{CTD} flexible loop region was used instead, which might not have all the components of the N protein in the P_{CTD}-binding interface in the properly folded form. It could also be that artefacts may arise in the SPR experiment, for example matrix interactions or mass effects [217]. Importantly, sensorgrams were not reported in the publication where SPR was used to determine the K_D [83]. Ideally, a solution-based measurement of P_{CTD}/N-RNA binding affinity, e.g. ITC, is needed to confirm the result of the SPR study. Furthermore, cellular proteins such as importins and HSP70, which could bind to the P_{CTD} or N protein at the P_{CTD}/N-RNA interface, may also enhance the dynamic nature of the N-RNA/P_{CTD} interaction by competitive binding because of the overlapped binding interface.

By combining cyclization and mutations of N-pep that showed stronger affinity for P_{CTD}, a cyclized N-pep variant with mutations K376A/T386A/S389E was generated which bound to P_{CTD} 200-fold stronger than the wild type linear N-pep. This high affinity N-pep variant that was serendipitously discovered in this study was used to form the N-pep/P_{CTD} complex with

the goal of elucidating its structure by X-ray crystallography or NMR. We have tried to crystallize the complex by co-crystallization and soaking N-pep into P_{CTD} crystals using P_{CTD} and the triple mutant N-pep (K376A/T386A/S389E) in both linear and cyclized forms. Considering that the disordered regions could hamper crystallization, two more cyclized triple mutant N-pep with the flexible N- and C-terminus trimmed to different lengths were also generated and used in the crystallization trial. For each P_{CTD}/N-pep variant combination, 384 co-crystallization conditions have been tested and none of them produced crystals. Soaking was also attempted where P_{CTD} crystals in three different space groups [111] (Chapter 4) were soaked with four N-pep variants mentioned above, and was also unsuccessful. As P_{CTD} can crystallize in three different conditions and space groups, seeding the complex droplets with P_{CTD} crystal seeds could be another way to generate the N-pep/P_{CTD} complex crystals. If successful, this could provide valuable information for structure-based drug discovery. The highest affinity binder for P_{CTD}, the cyclized triple mutant N-pep, could be optimized and developed into peptidomimetics to serve as anti-viral drugs.

Apart from its role in viral replication, P_{CTD} also plays diverse roles in the virus-host interface. The nearly complete assignment of P_{CTD} provides a basis to use NMR to study the interactions formed between P_{CTD} and host factors including nucleolin [100], STAT proteins [84, 102], PML protein [101] and microtubules [99]. Furthermore, since RABV P is a multi-domain protein with well-structured domains (P_{CED} and P_{CTD}) and disordered regions (IDR_{NT} and IDR_{CT}, see 1.5.4), NMR could be a useful tool in investigating the other regions of P, especially the disordered regions. IDR_{NT} is implicated in interactions with L protein [88] and the nascent RNA-free N protein (N^0) [89] which are also important for viral replication, and is predicted to marginally populate a helical structure [86]; IDR_{CT} interacts with dynein light chain (LC8) [93] and is critical for the formation of Negri Bodies in cells [67]. As a prominent method to study

the IDPs, NMR structural, dynamic and functional studies of the other regions of P could help to understand the replication machinery, the strategies exploited by RABV to evade innate immune system, and the broad virus-host interface.

References

1. Adamson, P.B., *The Spread of Rabies into Europe and the Probable Origin of This Disease in Antiquity.* Journal of the Royal Asiatic Society of Great Britain and Ireland, 1977(2): p. 140-144.
2. World Health, O., *WHO Expert Consultation on Rabies. Second report.* World Health Organ Tech Rep Ser, 2013(982): p. 1-139, back cover.
3. Organization, W.H., *WHO expert consultation on rabies: third report.* 2018.
4. Manning, S.E., et al., *Human rabies prevention--United States, 2008: recommendations of the Advisory Committee on Immunization Practices.* MMWR Recomm Rep, 2008. **57**(RR-3): p. 1-28.
5. Hampson, K., et al., *Estimating the global burden of endemic canine rabies.* PLoS Negl Trop Dis, 2015. **9**(4): p. e0003709.
6. Srinivasan, A., et al., *Transmission of rabies virus from an organ donor to four transplant recipients.* New England Journal of Medicine, 2005. **352**(11): p. 1103-1111.
7. Houff, S.A., et al., *Human-to-Human Transmission of Rabies Virus by Corneal Transplant.* New England Journal of Medicine, 1979. **300**(11): p. 603-604.
8. Control, C.f.D. and Prevention, *Investigation of rabies infections in organ donor and transplant recipients--Alabama, Arkansas, Oklahoma, and Texas, 2004.* MMWR. Morbidity and mortality weekly report, 2004. **53**(26): p. 586.
9. Fooks, A.R., et al., *Current status of rabies and prospects for elimination.* Lancet, 2014. **384**(9951): p. 1389-1399.
10. Charlton, K.M., et al., *The long incubation period in rabies: Delayed progression of infection in muscle at the site of exposure.* Acta Neuropathologica, 1997. **94**(1): p. 73-77.
11. Israsena, N., A. Mahavihakanont, and T. Hemachudha, *Rabies virus infection and microRNAs.* Adv Virus Res, 2011. **79**: p. 329-44.
12. Jackson, A.C., *Chapter 7 - Human Disease*, in *Rabies (Third Edition)*, A.C. Jackson, Editor. 2013, Academic Press: Boston. p. 269-298.
13. Ugolini, G., et al., *Horizontal eye movement networks in primates as revealed by retrograde transneuronal transfer of rabies virus: Differences in monosynaptic input to "slow" and "fast" abducens motoneurons.* Journal of Comparative Neurology, 2006. **498**(6): p. 762-785.
14. Ugolini, G., *Rabies Virus as a Transneuronal Tracer of Neuronal Connections.* Advances in Virus Research: Research Advances in Rabies, Vol 79, 2011. **79**: p. 165-202.
15. Shankar, V., B. Dietzschold, and H. Koprowski, *Direct entry of rabies virus into the central nervous system without prior local replication.* J Virol, 1991. **65**(5): p. 2736-8.
16. Johnson, N., et al., *Lyssavirus infection activates interferon gene expression in the brain.* Journal of General Virology, 2006. **87**: p. 2663-2667.
17. Johnson, N., et al., *Human rabies due to lyssavirus infection of bat origin.* Veterinary Microbiology, 2010. **142**(3-4): p. 151-159.
18. Badrane, H., et al., *Evidence of two Lyssavirus phylogroups with distinct pathogenicity and immunogenicity.* Journal of Virology, 2001. **75**(7): p. 3268-3276.
19. Morimoto, K., et al., *Characterization of a unique variant of bat rabies virus responsible for newly emerging human cases in North America.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(11): p. 5653-5658.
20. Tsiang, H., *Pathophysiology of rabies virus infection of the nervous system.* Adv Virus Res, 1993. **42**: p. 375-412.
21. Hemachudha, T., J. Laothamatas, and C.E. Rupprecht, *Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges.* Lancet Neurology, 2002. **1**(2): p. 101-109.
22. Mitrabhakdi, E., et al., *Difference in neuropathogenetic mechanisms in human furious and paralytic rabies.* Journal of the Neurological Sciences, 2005. **238**(1-2): p. 3-10.
23. Hemachudha, T. and P. Phuapradit, *Rabies.* Curr Opin Neurol, 1997. **10**(3): p. 260-7.
24. Hemachudha, T., et al., *Pathophysiology of human paralytic rabies.* J Neurovirol, 2005. **11**(1): p. 93-100.

25. Li, Z., Z. Feng, and H. Ye, *Rabies viral antigen in human tongues and salivary glands.* J Trop Med Hyg, 1995. **98**(5): p. 330-2.
26. Dacheux, L., et al., *A reliable diagnosis of human rabies based on analysis of skin biopsy specimens.* Clin Infect Dis, 2008. **47**(11): p. 1410-7.
27. Lanska, D.J., *Rabies virus, paralytic and classical.* Lancet, 1992. **339**(8796): p. 809.
28. Johnson, N., A.F. Cunningham, and A.R. Fooks, *The immune response to rabies virus infection and vaccination.* Vaccine, 2010. **28**(23): p. 3896-901.
29. Lieu, K.G., et al., *The rabies virus interferon antagonist P protein interacts with activated STAT3 and inhibits Gp130 receptor signaling.* J Virol, 2013. **87**(14): p. 8261-5.
30. Crick, J., *The vaccination of man and other animals against rabies.* Postgrad Med J, 1973. **49**(574): p. 551-64.
31. Pasteur, L., *Methode pour prevenir la rage apres morsure*. 1885.
32. Wu, X., T.G. Smith, and C.E. Rupprecht, *From brain passage to cell adaptation: the road of human rabies vaccine development.* Expert Rev Vaccines, 2011. **10**(11): p. 1597-608.
33. Fuenzalida, E., R. Palacios, and J. Borgono, *Antirabies antibody response in man to vaccine made from infected suckling-mouse brains.* Bulletin of the World Health Organization, 1964. **30**(3): p. 431.
34. Hicks, D.J., A.R. Fooks, and N. Johnson, *Developments in rabies vaccines.* Clin Exp Immunol, 2012. **169**(3): p. 199-204.
35. Maki, J., et al., *Oral vaccination of wildlife using a vaccinia-rabies-glycoprotein recombinant virus vaccine (RABORAL V-RG (R)): a global review.* Veterinary Research, 2017. **48**.
36. Osinubi, M.O.V., et al., *Enhancing comparative rabies DNA vaccine effectiveness through glycoprotein gene modifications.* Vaccine, 2009. **27**(51): p. 7214-7218.
37. Gupta, P.K., et al., *Sindbis virus replicon-based DNA vaccine encoding Rabies virus glycoprotein elicits specific humoral and cellular immune response in dogs.* Acta Virologica, 2009. **53**(2): p. 83-88.
38. Kaur, M., A. Rai, and R. Bhatnagar, *Rabies DNA vaccine: No impact of MHC Class I and Class II targeting sequences on immune response and protection against lethal challenge.* Vaccine, 2009. **27**(15): p. 2128-2137.
39. Chen, H., et al., *Adenovirus-Based Vaccines: Comparison of Vectors from Three Species of Adenoviridae.* Journal of Virology, 2010. **84**(20): p. 10522-10532.
40. Wiktor, T.J., et al., *Protection from Rabies by a Vaccinia Virus Recombinant Containing the Rabies Virus Glycoprotein Gene.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1984. **81**(22): p. 7194-7198.
41. Cadoz, M., et al., *Immunization with Canarypox Virus Expressing Rabies Glycoprotein.* Lancet, 1992. **339**(8807): p. 1429-1432.
42. Yarosh, O.K., et al., *Human adenovirus type 5 vectors expressing rabies glycoprotein.* Vaccine, 1996. **14**(13): p. 1257-64.
43. Jallet, C., et al., *Chimeric lyssavirus glycoproteins with increased immunological potential.* Journal of Virology, 1999. **73**(1): p. 225-233.
44. Fooks, A.R., et al., *Rabies.* Nature Reviews Disease Primers, 2017. **3**.
45. Rupprecht, C., I. Kuzmin, and F. Meslin, *Lyssaviruses and rabies: current conundrums, concerns, contradictions and controversies.* F1000Res, 2017. **6**: p. 184.
46. Gunawardena, P.S., et al., *Lyssavirus in Indian Flying Foxes, Sri Lanka.* Emerg Infect Dis, 2016. **22**(8): p. 1456-9.
47. Shinwari, M.W., et al., *Australian bat lyssavirus infection in two horses.* Vet Microbiol, 2014. **173**(3-4): p. 224-31.
48. Kgaladi, J., et al., *Diversity and epidemiology of Mokola virus.* PLoS Negl Trop Dis, 2013. **7**(10): p. e2511.
49. Horton, D.L., et al., *Antigenic and genetic characterization of a divergent African virus, Ikoma lyssavirus.* J Gen Virol, 2014. **95**(Pt 5): p. 1025-32.
50. Tordo, N., et al., *Walking along the rabies genome: is the large G-L intergenic region a remnant gene?* Proc Natl Acad Sci U S A, 1986. **83**(11): p. 3914-8.
51. Conzelmann, K.K., et al., *Molecular cloning and complete nucleotide sequence of the attenuated rabies virus SAD B19.* Virology, 1990. **175**(2): p. 485-99.
52. Arnheiter, H., et al., *Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis.* Cell, 1985. **41**(1): p. 259-67.

53. Schnell, M.J., et al., *The cell biology of rabies virus: using stealth to reach the brain.* Nature Reviews Microbiology, 2010. **8**(1): p. 51-61.
54. Leyrat, C., et al., *Structure, interactions with host cell and functions of rhabdovirus phosphoprotein.* Future virology, 2011. **6**(4): p. 465-481.
55. Lafon, M., *Rabies virus receptors.* J Neurovirol, 2005. **11**(1): p. 82-7.
56. Tordo, N. and A. Kouknetzoff, *The rabies virus genome: an overview.* Onderstepoort J Vet Res, 1993. **60**(4): p. 263-9.
57. Abraham, G. and A.K. Banerjee, *Sequential transcription of the genes of vesicular stomatitis virus.* Proc Natl Acad Sci U S A, 1976. **73**(5): p. 1504-8.
58. Shakin-Eshleman, S.H., et al., *N-linked glycosylation of rabies virus glycoprotein. Individual sequons differ in their glycosylation efficiencies and influence on cell surface expression.* J Biol Chem, 1992. **267**(15): p. 10690-8.
59. Wojczyk, B.S., et al., *The role of site-specific N-glycosylation in secretion of soluble forms of rabies virus glycoprotein.* Glycobiology, 1998. **8**(2): p. 121-30.
60. Masters, P.S. and A.K. Banerjee, *Resolution of multiple complexes of phosphoprotein NS with nucleocapsid protein N of vesicular stomatitis virus.* J Virol, 1988. **62**(8): p. 2651-7.
61. Howard, M. and G. Wertz, *Vesicular stomatitis virus RNA replication: a role for the NS protein.* J Gen Virol, 1989. **70 (Pt 10)**: p. 2683-94.
62. Finke, S. and K.M. Conzelmann, *Dissociation of rabies virus matrix protein functions in regulation of viral RNA synthesis and virus assembly.* Journal of Virology, 2003. **77**(22): p. 12074-12082.
63. Finke, S., R. Mueller-Waldeck, and K.K. Conzelmann, *Rabies virus matrix protein regulates the balance of virus transcription and replication.* Journal of General Virology, 2003. **84**: p. 1613-1621.
64. Albertini, A.A., et al., *Crystal structure of the rabies virus nucleoprotein-RNA complex.* Science, 2006. **313**(5785): p. 360-3.
65. Kristensson, K., et al., *Rabies: Interactions between neurons and viruses. A review of the history of Negri inclusion bodies.* Neuropathology and Applied Neurobiology, 1996. **22**(3): p. 179-187.
66. Lahaye, X., et al., *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): p. 7948-58.
67. Nikolic, J., et al., *Negri bodies are viral factories with properties of liquid organelles.* Nat Commun, 2017. **8**(1): p. 58.
68. Lahaye, X., et al., *Hsp70 protein positively regulates rabies virus infection.* J Virol, 2012. **86**(9): p. 4743-51.
69. Fouquet, B., et al., *Focal Adhesion Kinase Is Involved in Rabies Virus Infection through Its Interaction with Viral Phosphoprotein P.* Journal of Virology, 2015. **89**(3): p. 1640-1651.
70. Finke, S. and K.K. Conzelmann, *Replication strategies of rabies virus.* Virus Res, 2005. **111**(2): p. 120-31.
71. Iseni, F., et al., *Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures.* J Gen Virol, 1998. **79 (Pt 12)**: p. 2909-19.
72. Albertini, A.A., et al., *Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings.* J Struct Biol, 2007. **158**(1): p. 129-33.
73. Schoehn, G., et al., *Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site.* J Virol, 2001. **75**(1): p. 490-8.
74. Tawar, R.G., et al., *Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus.* Science, 2009. **326**(5957): p. 1279-83.
75. Green, T.J., et al., *Structure of the vesicular stomatitis virus nucleoprotein-RNA complex.* Science, 2006. **313**(5785): p. 357-360.
76. Mountcastle, W.E., et al., *Proteolytic cleavage of subunits of the nucleocapsid of the paramyxovirus simian virus 5.* J Virol, 1974. **14**(5): p. 1253-61.
77. Mountcastle, W.E., et al., *Nucleocapsid protein subunits of simian virus 5, Newcastle disease virus, and Sendai virus.* J Virol, 1970. **6**(5): p. 677-84.
78. Heggeness, M.H., A. Scheid, and P.W. Choppin, *the relationship of conformational changes in the Sendai virus nucleocapsid to proteolytic cleavage of the NP polypeptide.* Virology, 1981. **114**(2): p. 555-62.
79. Wu, X., X. Lei, and Z.F. Fu, *Rabies virus nucleoprotein is phosphorylated by cellular casein kinase II.* Biochemical and Biophysical Research Communications, 2003. **304**(2): p. 333-338.

80. Wu, X., et al., *Both Viral Transcription and Replication Are Reduced when the Rabies Virus Nucleoprotein Is Not Phosphorylated.* Journal of Virology, 2002. **76**(9): p. 4153-4161.
81. Toriumi, H. and A. Kawai, *Association of rabies virus nominal phosphoprotein (P) with viral nucleocapsid (NC) is enhanced by phosphorylation of the viral nucleoprotein (N).* Microbiol Immunol, 2004. **48**(5): p. 399-409.
82. Kawai, A., et al., *Nucleocapsid formation and/or subsequent conformational change of rabies virus nucleoprotein (N) is a prerequisite step for acquiring the phosphatase-sensitive epitope of monoclonal antibody 5-2-26.* Virology, 1999. **263**(2): p. 395-407.
83. Ribeiro Ede, A., Jr., et al., *Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes.* J Mol Biol, 2009. **394**(3): p. 558-75.
84. Vidy, A., M. Chelbi-Alix, and D. Blondel, *Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways.* J Virol, 2005. **79**(22): p. 14411-20.
85. Vidy, A., et al., *The nucleocytoplasmic rabies virus P protein counteracts interferon signaling by inhibiting both nuclear accumulation and DNA binding of STAT1.* J Virol, 2007. **81**(8): p. 4255-63.
86. Ivanov, I., et al., *Structural insights into the rhabdovirus transcription/replication complex.* Virus Res, 2011. **162**(1-2): p. 126-37.
87. Ivanov, I., et al., *Structure of the dimerization domain of the rabies virus phosphoprotein.* J Virol, 2010. **84**(7): p. 3707-10.
88. Chenik, M., et al., *Mapping the interacting domains between the rabies virus polymerase and phosphoprotein.* J Virol, 1998. **72**(3): p. 1925-30.
89. Mavrakis, M., et al., *Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA.* Virology, 2006. **349**(2): p. 422-9.
90. Gupta, A.K., et al., *The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C.* J Virol, 2000. **74**(1): p. 91-8.
91. Mebatsion, T., *Extensive attenuation of rabies virus by simultaneously modifying the dynein light chain binding site in the P protein and replacing Arg333 in the G protein.* Journal of Virology, 2001. **75**(23): p. 11496-11502.
92. MacGibeny, M.A., et al., *Retrograde axonal transport of rabies virus is unaffected by interferon treatment but blocked by emetine locally in axons.* Plos Pathogens, 2018. **14**(7).
93. Moseley, G.W., et al., *Dynein light chain association sequences can facilitate nuclear protein import.* Molecular Biology of the Cell, 2007. **18**(8): p. 3204-3213.
94. Oksayan, S., et al., *A novel nuclear trafficking module regulates the nucleocytoplasmic localization of the rabies virus interferon antagonist, P protein.* J Biol Chem, 2012. **287**(33): p. 28112-21.
95. Delmas, O., et al., *The structure of the nucleoprotein binding domain of lyssavirus phosphoprotein reveals a structural relationship between the N-RNA binding domains of Rhabdoviridae and Paramyxoviridae.* RNA Biol, 2010. **7**(3): p. 322-7.
96. Assenberg, R., et al., *Structure of the nucleoprotein binding domain of Mokola virus phosphoprotein.* J Virol, 2010. **84**(2): p. 1089-96.
97. Pasdeloup, D., et al., *Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1-dependent nuclear export signal.* Virology, 2005. **334**(2): p. 284-93.
98. Moseley, G.W., et al., *Nucleocytoplasmic distribution of rabies virus P-protein is regulated by phosphorylation adjacent to C-terminal nuclear import and export signals.* Biochemistry, 2007. **46**(43): p. 12053-61.
99. Brice, A., et al., *Quantitative Analysis of the Microtubule Interaction of Rabies Virus P3 Protein: Roles in Immune Evasion and Pathogenesis.* Sci Rep, 2016. **6**: p. 33493.
100. Oksayan, S., et al., *Identification of a role for nucleolin in rabies virus infection.* J Virol, 2015. **89**(3): p. 1939-43.
101. Blondel, D., et al., *Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies.* Oncogene, 2002. **21**(52): p. 7957-70.
102. Wiltzer, L., et al., *Interaction of rabies virus P-protein with STAT proteins is critical to lethal rabies disease.* J Infect Dis, 2014. **209**(11): p. 1744-53.
103. Jacob, Y., E. Real, and N. Tordo, *Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains.* J Virol, 2001. **75**(20): p. 9613-22.
104. Chang, C.E., W. Chen, and M.K. Gilson, *Ligand configurational entropy and protein binding.* Proc Natl Acad Sci U S A, 2007. **104**(5): p. 1534-9.

105. Gerard, F.C., et al., *Modular organization of rabies virus phosphoprotein.* J Mol Biol, 2009. **388**(5): p. 978-96.
106. Barr, J.N. and G.W. Wertz, *Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination.* J Virol, 2001. **75**(15): p. 6901-13.
107. Hercyk, N., S.M. Horikami, and S.A. Moyer, *The vesicular stomatitis virus L protein possesses the mRNA methyltransferase activities.* Virology, 1988. **163**(1): p. 222-5.
108. Mellon, M.G. and S.U. Emerson, *Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus.* J Virol, 1978. **27**(3): p. 560-7.
109. Emerson, S.U. and M. Schubert, *Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus.* Proc Natl Acad Sci U S A, 1987. **84**(16): p. 5655-9.
110. Chenik, M., K. Chebli, and D. Blondel, *Translation Initiation at Alternate in-Frame Aug Codons in the Rabies Virus Phosphoprotein Messenger-Rna Is Mediated by a Ribosomal Leaky Scanning Mechanism.* Journal of Virology, 1995. **69**(2): p. 707-712.
111. Mavrakis, M., et al., *Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus.* J Mol Biol, 2004. **343**(4): p. 819-31.
112. Green, T.J. and M. Luo, *Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11713-8.
113. Ribeiro, E.A., Jr., et al., *Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein.* J Mol Biol, 2008. **382**(2): p. 525-38.
114. Kouznetzoff, A., M. Buckle, and N. Tordo, *Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA.* J Gen Virol, 1998. **79 (Pt 5)**: p. 1005-13.
115. Amarasinghe, G.K., et al., *Taxonomy of the order Mononegavirales: update 2018.* Arch Virol, 2018.
116. Ito, N., G.W. Moseley, and M. Sugiyama, *The importance of immune evasion in the pathogenesis of rabies virus.* J Vet Med Sci, 2016. **78**(7): p. 1089-98.
117. Oksayan, S., et al., *Subcellular trafficking in rhabdovirus infection and immune evasion: a novel target for therapeutics.* Infect Disord Drug Targets, 2012. **12**(1): p. 38-58.
118. Crow, M.S., et al., *Diverse mechanisms evolved by DNA viruses to inhibit early host defenses.* Crit Rev Biochem Mol Biol, 2016. **51**(6): p. 452-481.
119. Leyrat, C., et al., *Structural disorder in proteins of the rhabdoviridae replication complex.* Protein Pept Lett, 2010. **17**(8): p. 979-87.
120. Albertini, A.A., R.W. Ruigrok, and D. Blondel, *Rabies virus transcription and replication.* Adv Virus Res, 2011. **79**: p. 1-22.
121. Rowe, C.L., et al., *Nuclear Trafficking of the Rabies Virus Interferon Antagonist P-Protein Is Regulated by an Importin-Binding Nuclear Localization Sequence in the C-Terminal Domain.* PLoS One, 2016. **11**(3): p. e0150477.
122. Moseley, G.W., et al., *Dual modes of rabies P-protein association with microtubules: a novel strategy to suppress the antiviral response.* J Cell Sci, 2009. **122**(Pt 20): p. 3652-62.
123. Ito, N., et al., *Role of interferon antagonist activity of rabies virus phosphoprotein in viral pathogenicity.* J Virol, 2010. **84**(13): p. 6699-710.
124. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-34.
125. Hyberts, S.G., K. Takeuchi, and G. Wagner, *Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data.* J Am Chem Soc, 2010. **132**(7): p. 2145-7.
126. Kazimierczuk, K. and V.Y. Orekhov, *Accelerated NMR spectroscopy by using compressed sensing.* Angew Chem Int Ed Engl, 2011. **50**(24): p. 5556-9.
127. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.
128. Bahrami, A., et al., *Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy.* PLoS Comput Biol, 2009. **5**(3): p. e1000307.
129. Lee, W., M. Tonelli, and J.L. Markley, *NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy.* Bioinformatics, 2015. **31**(8): p. 1325-7.

130. d'Auvergne, E.J. and P.R. Gooley, *Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor.* J Biomol NMR, 2008. **40**(2): p. 121-33.
131. Marsh, J.A., et al., *Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation.* Protein Sci, 2006. **15**(12): p. 2795-804.
132. Shen, Y., et al., *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* J Biomol NMR, 2009. **44**(4): p. 213-23.
133. Güntert, P., et al., *Automated stereospecific 1 H NMR assignments and their impact on the precision of protein structure determinations in solution*, in *Nmr In Structural Biology: A Collection of Papers by Kurt Wüthrich*. 1995, World Scientific. p. 419-426.
134. Wuthrich, K., M. Billeter, and W. Braun, *Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance.* J Mol Biol, 1983. **169**(4): p. 949-61.
135. Neri, D., et al., *Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling.* Biochemistry, 1989. **28**(19): p. 7510-6.
136. Amarasinghe, G.K., et al., *Taxonomy of the order Mononegavirales: update 2018.* Arch Virol, 2018. **163**(8): p. 2283-2294.
137. *Human rabies: 2016 updates and call for data.* Wkly Epidemiol Rec, 2017. **92**(7): p. 77-86.
138. Kolakofsky, D., et al., *Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis.* Virology, 2004. **318**(2): p. 463-73.
139. Stillman, E.A. and M.A. Whitt, *Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription.* J Virol, 1999. **73**(9): p. 7199-209.
140. Wiltzer, L., et al., *Conservation of a unique mechanism of immune evasion across the Lyssavirus genus.* J Virol, 2012. **86**(18): p. 10194-9.
141. Zheng, L., U. Baumann, and J.L. Reymond, *An efficient one-step site-directed and site-saturation mutagenesis protocol.* Nucleic Acids Res, 2004. **32**(14): p. e115.
142. Marley, J., M. Lu, and C. Bracken, *A method for efficient isotopic labeling of recombinant proteins.* J Biomol NMR, 2001. **20**(1): p. 71-5.
143. Zhan, J., et al., *(1)H, (15)N and (13)C resonance assignments of the C-terminal domain of the P protein of the Nishigahara strain of rabies virus.* Biomol NMR Assign, 2018.
144. Mayzel, M., et al., *Time-resolved multidimensional NMR with non-uniform sampling.* J Biomol NMR, 2014. **58**(2): p. 129-39.
145. Kazimierczuk, K. and V.Y. Orekhov, *Accelerated NMR spectroscopy by using compressed sensing.* Angewandte Chemie, 2011. **123**(24): p. 5670-5673.
146. Zhang, H., S. Neal, and D.S. Wishart, *RefDB: a database of uniformly referenced protein chemical shifts.* J Biomol NMR, 2003. **25**(3): p. 173-95.
147. Schrodinger, LLC, *The PyMOL Molecular Graphics System, Version 1.8*. 2015.
148. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W320-4.
149. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.
150. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-637.
151. Boyko, R. and B.D. Sykes. *Xcrvfit: A graphical x-windows program for binding curve studies and NMR spectroscopic analysis*. 2012; Available from: http://www.bionmr.ualberta.ca/bds/software/xcrvfit.
152. Ayed, A., et al., *Latent and active p53 are identical in conformation.* Nat Struct Biol, 2001. **8**(9): p. 756-60.
153. Waudby, C.A., et al., *Two-Dimensional NMR Lineshape Analysis.* Sci Rep, 2016. **6**: p. 24826.
154. Wishart, D.S., B.D. Sykes, and F.M. Richards, *Relationship between nuclear magnetic resonance chemical shift and protein secondary structure.* J Mol Biol, 1991. **222**(2): p. 311-33.
155. Bongrand, P., *Ligand-receptor interactions.* Reports on Progress in Physics, 1999. **62**(6): p. 921-968.
156. Etoc, F., et al., *Non-specific interactions govern cytosolic diffusion of nanosized objects in mammalian cells.* Nat Mater, 2018. **17**(8): p. 740-746.
157. Houben, K., et al., *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**(13): p. 6807-16.

158. Communie, G., et al., *Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus.* PLoS Pathog, 2013. **9**(9): p. e1003631.
159. Fenley, A.T., H.S. Muddana, and M.K. Gilson, *Entropy-enthalpy transduction caused by conformational shifts can obscure the forces driving protein-ligand binding.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(49): p. 20006-20011.
160. van Zundert, G.C.P., et al., *The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes.* J Mol Biol, 2016. **428**(4): p. 720-725.
161. Schiebel, J., et al., *High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits.* Structure, 2016. **24**(8): p. 1398-1409.
162. Du, X., et al., *Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods.* Int J Mol Sci, 2016. **17**(2).
163. Lumry, R. and S. Rajender, *Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water.* Biopolymers, 1970. **9**(10): p. 1125-227.
164. Liu, L. and Q.X. Guo, *Isokinetic relationship, isoequilibrium relationship, and enthalpy-entropy compensation.* Chem Rev, 2001. **101**(3): p. 673-95.
165. Remaut, H. and G. Waksman, *Protein-protein interaction through beta-strand addition.* Trends in Biochemical Sciences, 2006. **31**(8): p. 436-444.
166. Albertini, A.A., et al., *Structural aspects of rabies virus replication.* Cell Mol Life Sci, 2008. **65**(2): p. 282-94.
167. Kuriyan, J., B. Konforti, and D. Wemmer, *The molecules of life: Physical and chemical principles*. 2012: Garland Science.
168. Geschwindner, S., J. Ulander, and P. Johansson, *Ligand Binding Thermodynamics in Drug Discovery: Still a Hot Tip?* Journal of Medicinal Chemistry, 2015. **58**(16): p. 6321-6335.
169. Breiten, B., et al., *Water Networks Contribute to Enthalpy/Entropy Compensation in Protein-Ligand Binding.* Journal of the American Chemical Society, 2013. **135**(41): p. 15579-15584.
170. Toriumi, H. and A. Kawai, *Structural difference recognized by a monoclonal antibody #404-11 between the rabies virus nucleocapsid (NC) produced in virus infected cells and the NC-like structures produced in the nucleoprotein (N) cDNA-transfected cells.* Microbiol Immunol, 2005. **49**(8): p. 757-70.
171. Kingston, R.L., et al., *Structural basis for the attachment of a paramyxoviral polymerase to its template.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(22): p. 8301-8306.
172. Longhi, S., et al., *The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein.* Journal of Biological Chemistry, 2003. **278**(20): p. 18638-18648.
173. Curran, J., *A role for the Sendai virus P protein trimer in RNA synthesis.* J Virol, 1998. **72**(5): p. 4274-80.
174. Blocquel, D., et al., *Interaction between the C-terminal domains of measles virus nucleoprotein and phosphoprotein: a tight complex implying one binding site.* Protein Sci, 2012. **21**(10): p. 1577-85.
175. Sagara, J. and A. Kawai, *Identification of heat shock protein 70 in the rabies virion.* Virology, 1992. **190**(2): p. 845-8.
176. Brown, G., et al., *Evidence for an association between heat shock protein 70 and the respiratory syncytial virus polymerase complex within lipid-raft membranes during virus infection.* Virology, 2005. **338**(1): p. 69-80.
177. Oglesbee, M., S. Ringler, and S. Krakowka, *Interaction of canine distemper virus nucleocapsid variants with 70K heat-shock proteins.* J Gen Virol, 1990. **71 (Pt 7)**: p. 1585-90.
178. Oglesbee, M.J., et al., *The highly inducible member of the 70 kDa family of heat shock proteins increases canine distemper virus polymerase activity.* J Gen Virol, 1996. **77 (Pt 9)**: p. 2125-35.
179. Zhang, X., et al., *Identification and characterization of a regulatory domain on the carboxyl terminus of the measles virus nucleocapsid protein.* J Virol, 2002. **76**(17): p. 8737-46.
180. Zhang, X., et al., *Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus.* Virology, 2005. **337**(1): p. 162-74.
181. Carsillo, T., et al., *A single codon in the nucleocapsid protein C terminus contributes to in vitro and in vivo fitness of Edmonston measles virus.* J Virol, 2006. **80**(6): p. 2904-12.
182. Luft, J.R. and G.T. DeTitta, *A method to produce microseed stock for use in the crystallization of biological macromolecules.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 5): p. 988-93.

183. Aragao, D., et al., *MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron.* J Synchrotron Radiat, 2018. **25**(Pt 3): p. 885-891.
184. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
185. Evans, P.R., *An introduction to data reduction: space-group determination, scaling and intensity statistics.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 282-92.
186. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1204-14.
187. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.
188. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
189. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 61-9.
190. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
191. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
192. Guntert, P. and L. Buchner, *Combined automated NOE assignment and structure calculation with CYANA.* J Biomol NMR, 2015. **62**(4): p. 453-71.
193. Herrmann, T., P. Guntert, and K. Wuthrich, *Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA.* J Mol Biol, 2002. **319**(1): p. 209-27.
194. Guntert, P., *Automated NMR structure calculation with CYANA.* Methods Mol Biol, 2004. **278**: p. 353-78.
195. d'Auvergne, E.J. and P.R. Gooley, *Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces.* J Biomol NMR, 2008. **40**(2): p. 107-19.
196. Loria, J.P., M. Rance, and A.G. Palmer, *A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy.* Journal of the American Chemical Society, 1999. **121**(10): p. 2331-2332.
197. Bieri, M. and P.R. Gooley, *Automated NMR relaxation dispersion data analysis using NESSY.* Bmc Bioinformatics, 2011. **12**.
198. Baldwin, A.J. and L.E. Kay, *NMR spectroscopy brings invisible protein states into focus.* Nat Chem Biol, 2009. **5**(11): p. 808-14.
199. Ishima, R. and D.A. Torchia, *Protein dynamics from NMR.* Nat Struct Biol, 2000. **7**(9): p. 740-3.
200. Ishima, R. and D.A. Torchia, *Estimating the time scale of chemical exchange of proteins from measurements of transverse relaxation rates in solution.* J Biomol NMR, 1999. **14**(4): p. 369-72.
201. Sanchez-Medina, C., et al., *Probing the free energy landscape of the fast-folding gpW protein by relaxation dispersion NMR.* J Am Chem Soc, 2014. **136**(20): p. 7444-51.
202. Mittermaier, A. and L.E. Kay, *New tools provide new insights in NMR studies of protein dynamics.* Science, 2006. **312**(5771): p. 224-8.
203. Prehaud, C., K. Nel, and D.H. Bishop, *Baculovirus-expressed rabies virus M1 protein is not phosphorylated: it forms multiple complexes with expressed rabies N protein.* Virology, 1992. **189**(2): p. 766-70.
204. Sokol, F. and H.F. Clark, *Phosphoproteins, structural components of rhabdoviruses.* Virology, 1973. **52**(1): p. 246-63.
205. Prehaud, C., et al., *Expression, characterization, and purification of a phosphorylated rabies nucleoprotein synthesized in insect cells by baculovirus vectors.* Virology, 1990. **178**(2): p. 486-97.
206. Yang, J., et al., *Phosphorylation of rabies virus nucleoprotein regulates viral RNA transcription and replication by modulating leader RNA encapsidation.* J Virol, 1999. **73**(2): p. 1661-4.
207. Nishizuka, Y., *Protein kinase C and lipid signaling for sustained cellular responses.* FASEB J, 1995. **9**(7): p. 484-96.
208. Newton, A.C., *Protein kinase C: structure, function, and regulation.* J Biol Chem, 1995. **270**(48): p. 28495-8.
209. Uddin, S., et al., *Protein kinase C-delta (PKC-delta) is activated by type I interferons and mediates phosphorylation of Stat1 on serine 727.* J Biol Chem, 2002. **277**(17): p. 14408-16.

210. Deb, D.K., et al., *Activation of protein kinase C delta by IFN-gamma.* J Immunol, 2003. **171**(1): p. 267-73.
211. Randall, R.E. and S. Goodbourn, *Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures.* J Gen Virol, 2008. **89**(Pt 1): p. 1-47.
212. Park, R. and J.D. Baines, *Herpes simplex virus type 1 infection induces activation and recruitment of protein kinase C to the nuclear membrane and increased phosphorylation of lamin B.* J Virol, 2006. **80**(1): p. 494-504.
213. Constantinescu, S.N., C.D. Cernescu, and L.M. Popescu, *Effects of protein kinase C inhibitors on viral entry and infectivity.* FEBS Lett, 1991. **292**(1-2): p. 31-3.
214. Sieczkarski, S.B., H.A. Brown, and G.R. Whittaker, *Role of protein kinase C betaII in influenza virus entry via late endosomes.* J Virol, 2003. **77**(1): p. 460-9.
215. la Cour, T., et al., *Analysis and prediction of leucine-rich nuclear export signals.* Protein Eng Des Sel, 2004. **17**(6): p. 527-36.
216. Marfori, M., et al., *Molecular basis for specificity of nuclear import and prediction of nuclear localization.* Biochim Biophys Acta, 2011. **1813**(9): p. 1562-77.
217. Schuck, P. and H. Zhao, *The role of mass transport limitation and surface heterogeneity in the biophysical characterization of macromolecular binding processes by SPR biosensing.* Methods Mol Biol, 2010. **627**: p. 15-54.

Appendix A:

Publication Zhan et al.

Biomolecular NMR Assignments
https://doi.org/10.1007/s12104-018-9841-4

ARTICLE

^{1}H, ^{15}N and ^{13}C resonance assignments of the C-terminal domain of the P protein of the *Nishigahara* strain of rabies virus

Jingyu Zhan[1,2] · Md. Alamgir Hossain[1,2] · Ashish Sethi[1,2] · Toyoyuki Ose[3] · Gregory W. Moseley[1,2,4] · Paul R. Gooley[1,2]

Received: 4 June 2018 / Accepted: 14 September 2018

Abstract
The C-terminal domain of the P protein of rabies virus is a multifunctional domain that interacts with both viral and host cell proteins. Here we report the ^{1}H, ^{13}C and ^{15}N chemical shift assignments of this domain from P protein of the *Nishigahara* strain of rabies virus, a pathogenic laboratory strain well established for studies of virulence functions of rabies virus proteins, including P protein. The data and secondary structure analysis are in good agreement with the reported predominantly helical structure of the same domain from the CVS strain of rabies solved by crystallography. These assignments will enable future solution studies of the interactions of the P protein with viral and host proteins, and the effects of post-translational modifications.

Keywords Rabies · Lyssavirus · P protein · Phosphoprotein · Chemical shift assignments

Biological context

Rabies virus is a species of the *Lyssavirus* genus of the *Rhabdoviridae* family in the order *Mononegavirales* (Amarasinghe et al. 2018; Ito et al. 2016). These viruses have a single-stranded negative-sense RNA genome that is limited in size (c. 12 kb) and, consequently, coding capacity. Viral pathogenicity depends on the ability of virus to enter cells, replicate, assemble progeny, and escape cells, for which lyssaviruses have a 'skeleton crew' of only five genes, all essential in the basic life cycle (Ito et al. 2016; Ivanov et al. 2011). However, as viruses are obligate intracellular parasites, their capacity to infect, spread and cause disease is also critically dependent on their ability to remodel host cells into efficient virus factories, by controlling elements such as innate antiviral responses, gene expression, cell survival and metabolism. To achieve such functional diversity, viruses express multifunctional proteins, which comprise overlapping functional sequences and domains (Ito et al. 2016; Oksayan et al. 2012). Mechanisms that enable and regulate this functional diversity include the expression of multiple isoforms of the protein from one gene, post-translational modification and intrinsically disordered regions (Crow et al. 2016; Ito et al. 2016; Leyrat et al. 2010; Oksayan et al. 2012). The rabies phosphoprotein (P protein) is an archetypal example of this multifunctionality.

The P proteins of rabies virus and of other lyssavirus mediate multiple essential roles in replication and host cell manipulation (Ito et al. 2016; Oksayan et al. 2012). There are five isoforms, full-length P1 and P2–P5, which differ only through deletions at the N-terminal end (Gerard et al. 2009). The N-terminal region, which is likely to be disordered, is followed by a dimerization domain, another disordered region of about 30 residues and a globular 13-kDa C-terminal domain (CTD) (Albertini et al. 2011). The latter domain contains multiple interaction sites, including nuclear import and export sequences (Moseley et al. 2007; Rowe et al. 2016); two potential phosphorylation sites (Gupta et al. 2000); a basic-rich site that binds to the viral RNA-N protein complex (Ribeiro Ede et al. 2009); and sites for

✉ Paul R. Gooley
prg@unimelb.edu.au

1 Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, VIC 3010, Australia

2 Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, VIC 3010, Australia

3 Faculty of Advanced Life Science, Hokkaido University, Sapporo 060-0810, Japan

4 Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton Campus, Clayton, VIC 3800, Australia

Published online: 20 September 2018

interacting with host-cell proteins including the transcription factor STAT1 (Wiltzer et al. 2014), nucleolin (Oksayan et al. 2015), PML protein (Blondel et al. 2002) and microtubule proteins (Brice et al. 2016; Moseley et al. 2009). Importantly, several studies indicate that a number of key interactions are dynamic and regulated by processes such as dimerization and phosphorylation (Moseley et al. 2007, 2009), while other experiments using recombinant viruses derived from the pathogenic *Nishigahara* rabies virus strain have indicated critical roles in virulence for interactions of P protein with STATs (Wiltzer et al. 2014), microtubules (Brice et al. 2016), and with nucleo-cytoplasmic trafficking machinery (Ito et al. 2010). While an X-ray crystal structure of the P protein CTD from the rabies CVS strain is available (Mavrakis et al. 2004) we have initiated NMR studies of the CTD from the *Nishigahara* strain (NiP-CTD) to define the sites of interaction and conformational changes that occur on post-translational modification and protein interaction. Here we report the ^{1}H, ^{13}C and ^{15}N chemical shift assignments of NiP-CTD which can be used for future protein-interactions and dynamics-based studies.

Methods and experiments

Protein expression and purification

To express NiP-CTD, the gene corresponding to residues 186–297 were PCR amplified and inserted into pET28a vector using *Nde*I–*Eco*RI restriction sites. The terminal cysteine (Cys297) was mutated to serine to avoid formation of intermolecular disulphide bonds. This pET28a vector was designed to have a His_{10}-tag on the N-terminal region of NiP-CTD. A TEV protease cleavage site was inserted between the His_{10}-tag and the NiP-CTD gene.

To express unlabelled NiP-CTD, vectors were transformed into BL21(DE3) using heat shock transformation protocols. 5 ml of overnight culture, prepared from a single colony in LB media supplemented with 0.5 ml of 30 mg/ml Kanamycin, was used to inoculate 500 ml of 2YT auto-induction media (N-5052), where the gene of interest is induced upon glucose–lactose metabolism (Studier 2005). At first the culture was grown at 37 °C for 4 h until the OD_{600} reached around 0.7–0.8. The culture was then transferred to a 16 °C incubator for overnight expression. Cells were harvested by centrifugation at 3000×*g* for 10 min at 4 °C. Cell pellets were collected and stored at − 20 °C until required for purification.

Expression of uniformly ^{15}N-labelled NiP-CTD was auto-induced by growing BL21 (DE3) in N-5052 containing 1 g/l of $^{15}NH_4Cl$ as a sole source of nitrogen (Studier 2005). Cells were grown and harvested similar to unlabelled NiP-CTD and stored at − 20 °C until use. To label with ^{13}C and ^{15}N isotopes cells were grown in N-5052 (Studier 2005) supplemented with 1 g/l of $^{15}NH_4Cl$ (Sigma-Aldrich) and 3 g/l of D-[^{13}C] glucose (Sigma-Aldrich) as sole sources of nitrogen and carbon. Cells were grown at 37 °C until an OD_{600} of 0.6–0.7, transferred to 16 °C and induced with 0.4 mM IPTG and protein expressed overnight.

To purify NiP-CTD, pellets from a 500 ml culture were resuspended in 50 ml of NaPi resuspension buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 7.4) with one complete-EDTA free protease inhibitor tablet (Sigma-Aldrich) dissolved. A hand-held homogenizer was used to completely resuspend the cells before applying to an Avestin EmulsiFlex C3 cell crusher. Cells were lysed with three cycles to ensure complete disruption. Insoluble debris was separated from the supernatant by centrifugation (13,000×*g* for 30 min at 4 °C). The clear supernatant was passed through 0.22 µm filters and applied to 5 ml Talon® metal affinity resin (Clontech, TAKARA) packed in a gravity flow column and pre-equilibrated in resuspension buffer. The resin was rocked for 90 min at 4 °C, the column was washed twice (approximately 100 ml) with resuspension buffer and bound NiP-CTD was then eluted from the resin using ~ 80 ml of elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, pH 7.4). To remove imidazole, the eluted sample was concentrated down to 5 ml using an Amicon Ultra-15 (MWCO 3 kDa) (Sigma-Aldrich) at 3000×*g* at 4 °C and exchanged with 50 ml of buffer (50 mM Na_2HPO_4, 300 mM NaCl, pH 7.4). Following buffer exchange, and to cleave the His_{10}-tag, the protein sample was incubated overnight with TEV protease (0.5 ml of 1.8 mg/ml of purified TEV to 50 ml of protein sample) in the presence of 1 mM DTT. After cleavage, the protein was further concentrated to 1.5 ml using an Amicon Ultra-15 centrifugal unit (MWCO 3 kDa). NiP-CTD was further purified using a HiLoad™ 16/60 Superdex™ 75 prep grade column (GE Healthcare) pre-equilibrated with SEC buffer (50 mM Na_2HPO_4, 100 mM NaCl, 1 mM DTT, pH 6.8) and the flow rate was kept at 1 ml/min. Eluted 1 ml fractions with high purity were collected, pooled and concentrated in Amicon Ultra-4 centrifugal units (MWCO 3 kDa).

NMR spectroscopy

All NMR experiments were conducted at 298 K on Bruker Avance II 800 MHZ or Avance IIIHD 700 MHz spectrometers equipped with 5-mm HCN (TCI) cryogenic probes. Assignments were accomplished with the following experiments: 2D ^{1}H–^{1}H TOCSY and NOESY recorded on unlabelled NiP-CTD dissolved in 100% $^{2}H_2O$ and 50 mM phosphate buffer at pH 6.8; 2D ^{1}H–^{15}N HSQC, 3D HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, HBHA(CO)NH, CC(CO)NH, H(CCO)NH and ^{15}N-edited NOESY recorded on ^{15}N or ^{13}C,^{15}N NiP-CTD dissolved in

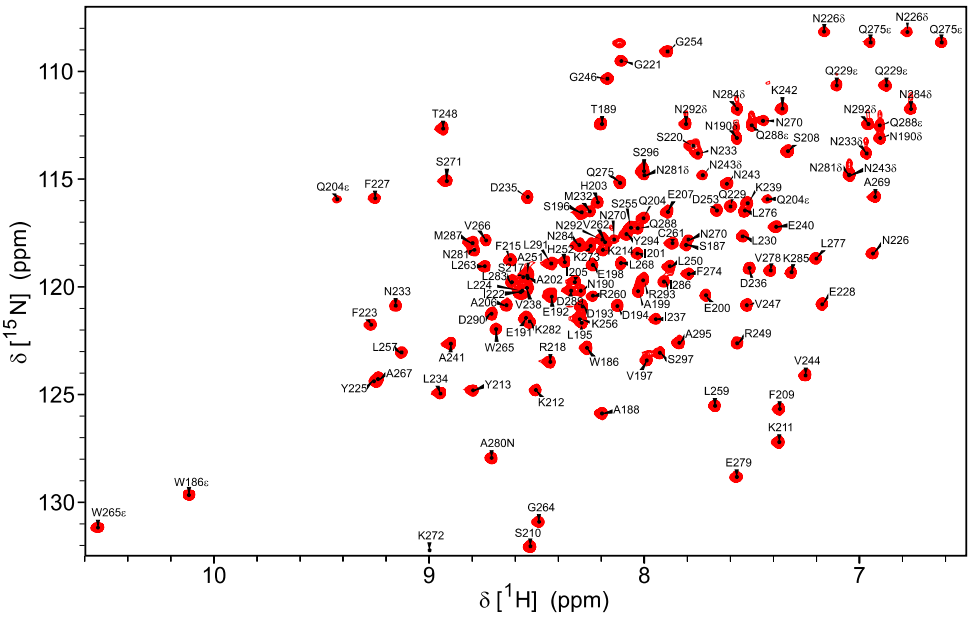


Fig. 1 700 MHz ^{1}H–^{15}N HSQC spectra of ^{15}N-labelled NiP-CTD at pH 6.8 and 298 K. Assignments of backbone amide resonances are indicted by residue and number; side-chain amides for Gln and Asn residues and the indole of the two Trp residues are indicated by residue, number and atom

10%/90% $^{2}H_2O/H_2O$ and 50 mM phosphate buffer at pH 6.8; two 2D constant-time ^{1}H–^{13}C HSQC [constant time periods of 13.3 ms (aliphatic) and 8.3 ms (aromatic)], 3D HCCH-TOCSY and two ^{13}C-edited NOESY [with ^{13}C offsets at 40 ppm (aliphatic) and 123 ppm (aromatic)] recorded on ^{13}C,^{15}N NiP-CTD dissolved in 100% $^{2}H_2O$ and 50 mM phosphate buffer at pH 6.8. A single high-resolution ^{1}H-^{13}C HSQC was recorded on a 10% ^{13}C-labelled sample of NiP-CTD dissolved in 100% $^{2}H_2O$ and 50 mM phosphate buffer at pH 6.8. All 3D spectra were recorded using 10% non-uniform sampling (NUS) and Poisson gap sampler (Hyberts et al. 2010), except the 3D ^{13}C and ^{15}N-edited NOESY spectra which were recorded with 25% NUS. Spectra were reconstructed with the compressed sensing algorithm using qMDD (Kazimierczuk and Orekhov 2011) and processed using NMRPipe (Delaglio et al. 1995). All other non-NUS data were processed in NMRPipe. NMR data were usually processed with a Lorentz–Gaussian function in the direct dimension and cosine-squared bells in the indirect dimensions and zero-filling once. Preliminary auto-assignments of backbone atoms (HN, ^{15}N, ^{13}Cα, ^{13}Cβ, ^{13}C′) were obtained from the PINE server (Bahrami et al. 2009), and manually verified and extended to include Hα and sidechain atoms using NMRFAM-Sparky (Lee et al. 2015). Preliminary ^{15}N transverse relaxation T_2 experiments were acquired at 800 MHz, processed in NMRPipe and analysed in relax (d'Auvergne and Gooley 2008). Secondary structure analysis was performed using ^{13}Cα and ^{13}Cβ chemical shifts with SSP (Marsh et al. 2006).

Extent of assignments and data deposition

The 2D ^{1}H–^{15}N HSQC spectrum of NiP-CTD showed well-dispersed resonances of a folded domain (Fig. 1). As the first three residues (Gly-His-Met) represent non-native sequence of the rabies virus P-protein, introduced from cloning, they are not included in the following statistics. The overall backbone assignment completeness is 98.73% among which 107 backbone NH resonances were assigned for the 109 non-proline residues, except for Ser219 and Lys231. All 24 of the amide side-chain resonances of Asn and Gln were present in the spectra and assigned (Fig. 1). A total of 63 out of 64 methyl groups of Ala, Leu, Ile, Val, Met and Thr were assigned except for the $C^{\delta2}/H^{\delta2}$ of Leu257. All the 39 identified Leu and Val methyl groups were stereo-specifically assigned using biosynthetic fractional (10%) ^{13}C labelling. For the aromatic residues, the side-chain resonances of Trp186, Trp265, His203, His252, Tyr213, Tyr225 and Tyr294 were unambiguously assigned. The $C^{\delta2}/H^{\delta2}$, $C^{\varepsilon1}/H^{\varepsilon1}$ and C^{ζ}/H^{ζ} pairs of Phe215 and Phe223 were fully assigned while Phe209, Phe227 and Phe274 were partially assigned. Secondary structure analysis (Fig. 2a) reveals seven α-helices and two β-strands: residues 194–208, α1; 227–230, α2; 233–243, α3; 245–252, α4; 258–271 α5; 272–276 α6; 280–293 α7; 212–214 β1 and 222–225 β2. These secondary structures are in good agreement with the crystal structure (pdb ident: 1vyi) of the CTD from the CVS strain of rabies (Mavrakis et al. 2004). The only significant difference is that the first residue of helix α1 is 189 in the crystal structure. However, analysis of ^{15}N-R_2 suggests that the N-terminal of

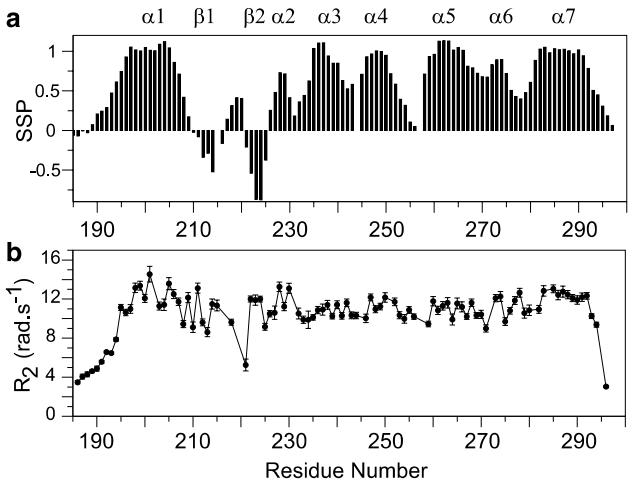


Fig. 2 Secondary structure (**a**) and ^{15}N-R_2 analysis (**b**) of NiP-CTD. Secondary structure was assessed by SSP (Marsh et al. 2006) for the ^{13}Cα and ^{13}Cβ chemical shift data. Seven α-helices and two β-strands are identified in agreement with the crystal structure for the same domain from the CVS rabies strain (Mavrakis et al. 2004). ^{15}N-R_2 analysis shows mobility in the N-terminus up to residue 195

NiP-CTD is dynamic up to residue 195 (Fig. 2b). The chemical shift assignments for NiP-CTD have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number 27498.

Acknowledgements This work was supported by National Health and Medical Research Council project Grant 1125704 (to PRG and GWM); and Australian Research Council equipment Grant LE120100022.

References

Albertini AA, Ruigrok RW, Blondel D (2011) Rabies virus transcription and replication. Adv Virus Res 79:1–22

Amarasinghe GK, Arechiga Ceballos NG, Banyard AC, Basler CF et al (2018) Taxonomy of the order Mononegavirales: update 2018. Arch Virol 163:2283–2294

Bahrami A, Assadi AH, Markley JL, Eghbalnia HR (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. PLoS Comput Biol 5:e1000307

Blondel D, Regad T, Poisson N, Pavie B, Harper F, Pandolfi PP, De The H, Chelbi-Alix MK (2002) Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies. Oncogene 21:7957–7970

Brice A, Whelan DR, Ito N, Shimizu K, Wiltzer-Bach L, Lo CY, Blondel D, Jans DA, Bell TD, Moseley GW (2016) Quantitative analysis of the microtubule interaction of rabies virus P3 protein: roles in immune evasion and pathogenesis. Sci Rep 6:33493

Crow MS, Lum KK, Sheng X, Song B, Cristea IM (2016) Diverse mechanisms evolved by DNA viruses to inhibit early host defenses. Crit Rev Biochem Mol Biol 51:452–481

d'Auvergne EJ, Gooley PR (2008) Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. J Biomol NMR 40:121–133

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Gerard FC, Ribeiro Ede A Jr, Leyrat C, Ivanov I, Blondel D, Longhi S, Ruigrok RW, Jamin M (2009) Modular organization of rabies virus phosphoprotein. J Mol Biol 388:978–996

Gupta AK, Blondel D, Choudhary S, Banerjee AK (2000) The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C. J Virol 74:91–98

Hyberts SG, Takeuchi K, Wagner G (2010) Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. J Am Chem Soc 132:2145–2147

Ito N, Moseley GW, Blondel D, Shimizu K, Rowe CL, Ito Y, Masatani T, Nakagawa K, Jans DA, Sugiyama M (2010) Role of interferon antagonist activity of rabies virus phosphoprotein in viral pathogenicity. J Virol 84:6699–6710

Ito N, Moseley GW, Sugiyama M (2016) The importance of immune evasion in the pathogenesis of rabies virus. J Vet Med Sci 78:1089–1098

Ivanov I, Yabukarski F, Ruigrok RW, Jamin M (2011) Structural insights into the rhabdovirus transcription/replication complex. Virus Res 162:126–137

Kazimierczuk K, Orekhov VY (2011) Accelerated NMR spectroscopy by using compressed sensing. Angew Chem Int Ed Engl 50:5556–5559

Lee W, Tonelli M, Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 31:1325–1327

Leyrat C, Gerard FC, de Almeida Ribeiro E Jr, Ivanov I, Ruigrok RW, Jamin M (2010) Structural disorder in proteins of the rhabdoviridae replication complex. Protein Pept Lett 17:979–987

Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. Protein Sci 15:2795–2804

Mavrakis M, McCarthy AA, Roche S, Blondel D, Ruigrok RW (2004) Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J Mol Biol 343:819–831

Moseley GW, Filmer RP, DeJesus MA, Jans DA (2007) Nucleocytoplasmic distribution of rabies virus P-protein is regulated by phosphorylation adjacent to C-terminal nuclear import and export signals. Biochemistry 46:12053–12061

Moseley GW, Lahaye X, Roth DM, Oksayan S, Filmer RP, Rowe CL, Blondel D, Jans DA (2009) Dual modes of rabies P-protein association with microtubules: a novel strategy to suppress the antiviral response. J Cell Sci 122:3652–3662

Oksayan S, Ito N, Moseley G, Blondel D (2012) Subcellular trafficking in rhabdovirus infection and immune evasion: a novel target for therapeutics. Infect Disord Drug Targets 12:38–58

Oksayan S, Nikolic J, David CT, Blondel D, Jans DA, Moseley GW (2015) Identification of a role for nucleolin in rabies virus infection. J Virol 89:1939–1943

Ribeiro Ede A Jr, Leyrat C, Gerard FC, Albertini AA, Falk C, Ruigrok RW, Jamin M (2009) Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes. J Mol Biol 394:558–575

Rowe CL, Wagstaff KM, Oksayan S, Glover DJ, Jans DA, Moseley GW (2016) Nuclear trafficking of the rabies virus interferon antagonist P-protein is regulated by an importin-binding nuclear localization sequence in the C-terminal domain. PLoS ONE 11:e0150477

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207–234

Wiltzer L, Okada K, Yamaoka S, Larrous F, Kuusisto HV, Sugiyama M, Blondel D, Bourhy H, Jans DA, Ito N, Moseley GW (2014) Interaction of rabies virus P-protein with STAT proteins is critical to lethal rabies disease. J Infect Dis 209:1744–1753

Appendix B:

^{1}H, ^{15}N and ^{13}C resonance assignments of the C-terminal domain of the P protein of the Nishigahara strain of rabies virus (BMRB: 27498)

```
#####################################################################
#       Chemical Shift Ambiguity Index Value Definitions            #
#                                                                   #
# The values other than 1 are used for those atoms with different   #
# chemical shifts that cannot be assigned to stereospecific atoms   #
# or to specific residues or chains.                                #
#                                                                   #
#   Index Value            Definition                               #
#                                                                   #
#      1             Unique (including isolated methyl protons,    #
#                         geminal atoms, and geminal methyl         #
#                         groups with identical chemical shifts)    #
#                         (e.g. ILE HD11, HD12, HD13 protons)       #
#      2             Ambiguity of geminal atoms or geminal methyl  #
#                         proton groups (e.g. ASP HB2 and HB3       #
#                         protons, LEU CD1 and CD2 carbons, or      #
#                         LEU HD11, HD12, HD13 and HD21, HD22,      #
#                         HD23 methyl protons)                      #
#      3             Aromatic atoms on opposite sides of           #
#                         symmetrical rings (e.g. TYR HE1 and HE2   #
#                         protons)                                  #
#      4             Intraresidue ambiguities (e.g. LYS HG and     #
#                         HD protons or TRP HZ2 and HZ3 protons)    #
#      5             Interresidue ambiguities (LYS 12 vs. LYS 27)  #
#      6             Intermolecular ambiguities (e.g. ASP 31 CA    #
#                         in monomer 1 and ASP 31 CA in monomer 2   #
#                         of an asymmetrical homodimer, duplex      #
#                         DNA assignments, or other assignments     #
#                         that may apply to atoms in one or more    #
#                         molecule in the molecular assembly)       #
#      9             Ambiguous, specific ambiguity not defined     #
#                                                                   #
#####################################################################
loop_
    _Atom_chem_shift.ID
    _Atom_chem_shift.Comp_index_ID
    _Atom_chem_shift.Comp_ID
    _Atom_chem_shift.Atom_ID
    _Atom_chem_shift.Atom_type
    _Atom_chem_shift.Atom_isotope_number
    _Atom_chem_shift.Val
    _Atom_chem_shift.Val_err
    _Atom_chem_shift.Auth_seq_ID

1             2             H             CD2           C             13            119.592       0             184
2             2             H             CE1           C             13            138.119       0             184
3             2             H             HD2           H             1             6.901         0.006         184
4             2             H             HE1           H             1             7.843         0.007         184
5             3             M             C             C             13            175.582       0.001         185
6             3             M             CA            C             13            55.521        0.075         185
7             3             M             CB            C             13            32.826        0.041         185
8             3             M             CE            C             13            16.858        0             185
9             3             M             CG            C             13            31.783        0.025         185
10            3             M             HA            H             1             4.433         0.01          185
11            3             M             HB2           H             1             1.942         0.011         185
12            3             M             HB3           H             1             1.91          0.011         185
13            3             M             HG2           H             1             2.451         0.002         185
14            3             M             HG3           H             1             2.4           0.011         185
15            3             M             HE            H             1             2.049         0             185
16            4             W             C             C             13            175.797       0.002         186
17            4             W             CA            C             13            57.781        0.047         186
18            4             W             CB            C             13            29.627        0.059         186
19            4             W             CD1           C             13            127.08        0.006         186
20            4             W             CE3           C             13            120.631       0.017         186
21            4             W             CH2           C             13            124.394       0.06          186
22            4             W             CZ2           C             13            114.378       0.033         186
23            4             W             CZ3           C             13            121.874       0             186
24            4             W             H             H             1             8.264         0.002         186
25            4             W             HA            H             1             4.592         0.006         186
26            4             W             HD1           H             1             7.231         0.007         186
27            4             W             HE1           H             1             10.117        0.001         186
28            4             W             HE3           H             1             7.484         0.007         186
29            4             W             HH2           H             1             7.11          0             186
30            4             W             HZ2           H             1             7.326         0.005         186
31            4             W             HZ3           H             1             7.057         0.004         186
32            4             W             N             N             15            122.776       0.054         186
33            4             W             NE1           N             15            129.635       0.017         186
34            4             W             HB2           H             1             3.235         0.006         186
35            4             W             HB3           H             1             3.235         0.006         186
36            5             S             C             C             13            181.329       0.01          187
37            5             S             CA            C             13            57.715        0.091         187
38            5             S             CB            C             13            64.215        0.059         187
39            5             S             H             H             1             7.804         0.002         187
40            5             S             HA            H             1             4.265         0.006         187
41            5             S             HB2           H             1             3.65          0.012         187
42            5             S             HB3           H             1             3.696         0.004         187
43            5             S             N             N             15            118.019       0.039         187
44            6             A             C             C             13            177.992       0.001         188
45            6             A             CA            C             13            52.665        0.088         188
46            6             A             CB            C             13            19.243        0.036         188
47            6             A             H             H             1             8.196         0.003         188
48            6             A             HA            H             1             4.176         0.004         188
49            6             A             N             N             15            125.867       0.033         188
```

50	6	A	HB	H	1	1.379	0.012	188
51	7	T	C	C	13	174.328	0.002	189
52	7	T	CA	C	13	61.811	0.084	189
53	7	T	CB	C	13	69.771	0.053	189
54	7	T	CG2	C	13	21.599	0.048	189
55	7	T	H	H	1	8.197	0.002	189
56	7	T	HA	H	1	4.313	0.008	189
57	7	T	HB	H	1	4.252	0.005	189
58	7	T	N	N	15	112.442	0.039	189
59	7	T	HG2	H	1	1.168	0.004	189
60	8	N	C	C	13	175.284	0.004	190
61	8	N	CA	C	13	53.26	0.062	190
62	8	N	CB	C	13	39.106	0.033	190
63	8	N	H	H	1	8.295	0.004	190
64	8	N	HA	H	1	4.724	0.009	190
65	8	N	HB2	H	1	2.881	0.01	190
66	8	N	HB3	H	1	2.79	0.014	190
67	8	N	HD21	H	1	7.571	0.001	190
68	8	N	HD22	H	1	6.901	0.002	190
69	8	N	N	N	15	120.16	0.041	190
70	8	N	ND2	N	15	113.023	0.069	190
71	9	E	C	C	13	177.077	0.002	191
72	9	E	CA	C	13	57.792	0.054	191
73	9	E	CB	C	13	30.05	0.053	191
74	9	E	CG	C	13	36.414	0	191
75	9	E	H	H	1	8.552	0.006	191
76	9	E	HA	H	1	4.215	0.01	191
77	9	E	HB2	H	1	2.073	0.01	191
78	9	E	HB3	H	1	1.977	0.015	191
79	9	E	N	N	15	121.442	0.05	191
80	9	E	HG2	H	1	2.268	0	191
81	9	E	HG3	H	1	2.268	0	191
82	10	E	C	C	13	177.197	0.009	192
83	10	E	CA	C	13	57.777	0.054	192
84	10	E	CB	C	13	29.884	0.041	192
85	10	E	CG	C	13	36.457	0.047	192
86	10	E	H	H	1	8.432	0.002	192
87	10	E	HA	H	1	4.17	0.01	192
88	10	E	HB2	H	1	2.056	0.012	192
89	10	E	HB3	H	1	1.99	0.013	192
90	10	E	N	N	15	120.413	0.043	192
91	10	E	HG2	H	1	2.266	0.005	192
92	10	E	HG3	H	1	2.266	0.005	192
93	11	D	C	C	13	176.676	0.006	193
94	11	D	CA	C	13	55.2	0.05	193
95	11	D	CB	C	13	41.236	0.027	193
96	11	D	H	H	1	8.286	0.002	193
97	11	D	HA	H	1	4.625	0.009	193
98	11	D	N	N	15	120.932	0.062	193
99	11	D	HB2	H	1	2.722	0.01	193
100	11	D	HB3	H	1	2.722	0.01	193
101	12	D	C	C	13	177.483	0.001	194
102	12	D	CA	C	13	55.804	0.084	194
103	12	D	CB	C	13	41.255	0.007	194
104	12	D	H	H	1	8.126	0.003	194
105	12	D	HA	H	1	4.562	0.01	194
106	12	D	N	N	15	120.863	0.043	194
107	12	D	HB2	H	1	2.712	0.011	194
108	12	D	HB3	H	1	2.712	0.011	194
109	13	L	C	C	13	179.542	0.011	195
110	13	L	CA	C	13	57.291	0.101	195
111	13	L	CB	C	13	41.504	0.078	195
112	13	L	CD1	C	13	25.111	0.064	195
113	13	L	CD2	C	13	23.302	0.086	195
114	13	L	CG	C	13	27.196	0.105	195
115	13	L	H	H	1	8.287	0.004	195
116	13	L	HA	H	1	4.192	0.011	195
117	13	L	HB2	H	1	1.803	0.008	195
118	13	L	HB3	H	1	1.609	0.007	195
119	13	L	HG	H	1	1.73	0.008	195
120	13	L	N	N	15	121.588	0.072	195
121	13	L	HD1	H	1	0.914	0.006	195
122	13	L	HD2	H	1	0.855	0.005	195
123	14	S	C	C	13	177.19	0	196
124	14	S	CA	C	13	61.126	0.036	196
125	14	S	CB	C	13	62.718	0.026	196
126	14	S	H	H	1	8.289	0.009	196
127	14	S	HA	H	1	4.3	0.006	196
128	14	S	HB2	H	1	4.041	0.006	196
129	14	S	HB3	H	1	4.017	0.005	196
130	14	S	N	N	15	116.535	0.051	196
131	15	V	C	C	13	177.456	0.013	197
132	15	V	CA	C	13	66.288	0.041	197
133	15	V	CB	C	13	31.843	0.044	197
134	15	V	CG1	C	13	21.739	0.082	197
135	15	V	CG2	C	13	23.02	0.053	197
136	15	V	H	H	1	7.994	0.007	197
137	15	V	HA	H	1	3.655	0.009	197
138	15	V	HB	H	1	2.285	0.01	197
139	15	V	N	N	15	123.4	0.032	197
140	15	V	HG1	H	1	0.969	0.012	197
141	15	V	HG2	H	1	1.094	0.009	197
142	16	E	C	C	13	178.011	0.003	198
143	16	E	CA	C	13	60.895	0.068	198

144	16	E	CB	C	13	28.835	0.051	198
145	16	E	CG	C	13	38.079	0.062	198
146	16	E	H	H	1	8.242	0.007	198
147	16	E	HA	H	1	3.578	0.01	198
148	16	E	HB2	H	1	2.075	0.016	198
149	16	E	HB3	H	1	2.005	0.02	198
150	16	E	HG2	H	1	2.384	0.004	198
151	16	E	HG3	H	1	2.035	0.011	198
152	16	E	N	N	15	118.948	0.054	198
153	17	A	C	C	13	180.137	0.004	199
154	17	A	CA	C	13	54.88	0.028	199
155	17	A	CB	C	13	17.981	0.038	199
156	17	A	H	H	1	8.026	0.005	199
157	17	A	HA	H	1	4.095	0.013	199
158	17	A	N	N	15	120.186	0.082	199
159	17	A	HB	H	1	1.473	0.006	199
160	18	E	C	C	13	178.326	0.004	200
161	18	E	CA	C	13	58.795	0.062	200
162	18	E	CB	C	13	29.624	0.036	200
163	18	E	CG	C	13	35.097	0	200
164	18	E	H	H	1	7.713	0.002	200
165	18	E	HA	H	1	4.247	0.013	200
166	18	E	HB2	H	1	2.298	0.016	200
167	18	E	HB3	H	1	2.047	0.01	200
168	18	E	N	N	15	120.356	0.042	200
169	19	I	C	C	13	177.41	0	201
170	19	I	CA	C	13	66.086	0.047	201
171	19	I	CB	C	13	37.911	0.02	201
172	19	I	CD1	C	13	14.971	0.017	201
173	19	I	CG2	C	13	18.773	0.037	201
174	19	I	H	H	1	8.032	0.003	201
175	19	I	HA	H	1	3.38	0.012	201
176	19	I	HB	H	1	1.807	0.014	201
177	19	I	N	N	15	118.448	0.061	201
178	19	I	HD1	H	1	0.578	0.013	201
179	19	I	HG2	H	1	0.73	0.015	201
180	20	A	C	C	13	179.178	0.019	202
181	20	A	CA	C	13	55.66	0.068	202
182	20	A	CB	C	13	18.32	0.063	202
183	20	A	H	H	1	8.541	0.004	202
184	20	A	HA	H	1	3.598	0.012	202
185	20	A	N	N	15	119.522	0.13	202
186	20	A	HB	H	1	1.441	0.014	202
187	21	H	C	C	13	177.407	0.005	203
188	21	H	CA	C	13	59.13	0.14	203
189	21	H	CB	C	13	29.194	0.067	203
190	21	H	CD2	C	13	119.987	0.053	203
191	21	H	CE1	C	13	137.215	0	203
192	21	H	H	H	1	8.215	0.003	203
193	21	H	HA	H	1	4.308	0.013	203
194	21	H	HD2	H	1	7.159	0.012	203
195	21	H	HE1	H	1	8.185	0.016	203
196	21	H	N	N	15	116.09	0.046	203
197	21	H	HB2	H	1	3.341	0.008	203
198	21	H	HB3	H	1	3.341	0.008	203
199	22	Q	C	C	13	179.646	0.008	204
200	22	Q	CA	C	13	60.759	0.083	204
201	22	Q	CB	C	13	28.973	0	204
202	22	Q	CG	C	13	35.292	0.073	204
203	22	Q	H	H	1	8.002	0.005	204
204	22	Q	HA	H	1	3.957	0.007	204
205	22	Q	HE21	H	1	9.423	0.004	204
206	22	Q	HE22	H	1	7.428	0.002	204
207	22	Q	HG2	H	1	2.985	0.009	204
208	22	Q	HG3	H	1	2.28	0.007	204
209	22	Q	N	N	15	116.767	0.053	204
210	22	Q	NE2	N	15	115.861	0.082	204
211	22	Q	HB2	H	1	1.86	0.017	204
212	22	Q	HB3	H	1	1.86	0.017	204
213	23	I	C	C	13	177.251	0.008	205
214	23	I	CA	C	13	66.202	0.05	205
215	23	I	CB	C	13	37.42	0.049	205
216	23	I	CD1	C	13	14.476	0.023	205
217	23	I	CG1	C	13	29.087	0.052	205
218	23	I	CG2	C	13	15.661	0.095	205
219	23	I	H	H	1	8.321	0.003	205
220	23	I	HA	H	1	3.261	0.011	205
221	23	I	HB	H	1	1.467	0.016	205
222	23	I	HG12	H	1	1.652	0.008	205
223	23	I	HG13	H	1	0.287	0.019	205
224	23	I	N	N	15	119.767	0.082	205
225	23	I	HD1	H	1	0.353	0.006	205
226	23	I	HG2	H	1	-0.063	0.007	205
227	24	A	C	C	13	173.771	0	206
228	24	A	CA	C	13	55.916	0.068	206
229	24	A	CB	C	13	18.04	0.088	206
230	24	A	H	H	1	8.639	0.008	206
231	24	A	HA	H	1	3.862	0.009	206
232	24	A	N	N	15	120.849	0.072	206
233	24	A	HB	H	1	1.396	0.014	206
234	25	E	C	C	13	177.524	0.009	207
235	25	E	CA	C	13	58.803	0.062	207
236	25	E	CB	C	13	29.969	0.036	207
237	25	E	CG	C	13	36.611	0.012	207

238	25	E	H	H	1	7.892	0.003	207
239	25	E	HA	H	1	4.123	0.01	207
240	25	E	HB2	H	1	2.081	0.016	207
241	25	E	HB3	H	1	2.036	0.014	207
242	25	E	N	N	15	116.512	0.028	207
243	25	E	HG2	H	1	2.24	0.006	207
244	25	E	HG3	H	1	2.24	0.006	207
245	26	S	C	C	13	176.375	0.005	208
246	26	S	CA	C	13	57.338	0.104	208
247	26	S	CB	C	13	64.42	0.058	208
248	26	S	H	H	1	7.33	0.003	208
249	26	S	HA	H	1	5.017	0.008	208
250	26	S	HB2	H	1	4.621	0.01	208
251	26	S	HB3	H	1	3.947	0.008	208
252	26	S	N	N	15	113.715	0.038	208
253	27	F	C	C	13	175.263	0.012	209
254	27	F	CA	C	13	63.011	0.069	209
255	27	F	CB	C	13	40.862	0.042	209
256	27	F	CD1	C	13	132.045	0.133	209
257	27	F	CE1	C	13	129.333	0	209
258	27	F	H	H	1	7.37	0.002	209
259	27	F	HA	H	1	4.349	0.009	209
260	27	F	HB2	H	1	3.751	0.013	209
261	27	F	HB3	H	1	3.079	0.009	209
262	27	F	HZ	H	1	7.104	0.004	209
263	27	F	N	N	15	125.661	0.059	209
264	27	F	HD1	H	1	7.21	0.007	209
265	27	F	HD2	H	1	7.21	0.007	209
266	27	F	HE1	H	1	7.318	0.005	209
267	27	F	HE2	H	1	7.318	0.005	209
268	28	S	C	C	13	175.537	0.003	210
269	28	S	CA	C	13	57.173	0.071	210
270	28	S	CB	C	13	64.132	0.062	210
271	28	S	H	H	1	8.531	0.003	210
272	28	S	HA	H	1	4.709	0.003	210
273	28	S	HB2	H	1	3.924	0.008	210
274	28	S	HB3	H	1	3.963	0.009	210
275	28	S	N	N	15	106.035	0.05	210
276	29	K	C	C	13	174.795	0.003	211
277	29	K	CA	C	13	57.101	0.112	211
278	29	K	CB	C	13	32.569	0.14	211
279	29	K	CG	C	13	24.591	0.02	211
280	29	K	H	H	1	7.37	0.003	211
281	29	K	HA	H	1	3.93	0.007	211
282	29	K	HG2	H	1	1.122	0.009	211
283	29	K	HG3	H	1	0.679	0.007	211
284	29	K	N	N	15	127.193	0.054	211
285	29	K	HB2	H	1	1.781	0.008	211
286	29	K	HB3	H	1	1.781	0.008	211
287	30	K	C	C	13	175.625	0.011	212
288	30	K	CA	C	13	56.379	0.069	212
289	30	K	CB	C	13	32.727	0.023	212
290	30	K	CD	C	13	29.063	0	212
291	30	K	CG	C	13	26.295	0	212
292	30	K	H	H	1	8.503	0.002	212
293	30	K	HA	H	1	4.255	0.007	212
294	30	K	HB2	H	1	1.815	0.007	212
295	30	K	HB3	H	1	1.581	0.014	212
296	30	K	N	N	15	124.776	0.052	212
297	31	Y	C	C	13	181.549	0.013	213
298	31	Y	CA	C	13	57.484	0.087	213
299	31	Y	CB	C	13	41.404	0.148	213
300	31	Y	CD1	C	13	133.764	0.061	213
301	31	Y	CE1	C	13	117.524	0.146	213
302	31	Y	H	H	1	8.792	0.003	213
303	31	Y	HA	H	1	4.531	0.008	213
304	31	Y	HB2	H	1	2.662	0.016	213
305	31	Y	HB3	H	1	2.559	0.016	213
306	31	Y	N	N	15	124.759	0.074	213
307	31	Y	HD1	H	1	7.186	0.015	213
308	31	Y	HD2	H	1	7.186	0.015	213
309	31	Y	HE1	H	1	6.919	0.007	213
310	31	Y	HE2	H	1	6.919	0.007	213
311	32	K	C	C	13	175.192	0.005	214
312	32	K	CA	C	13	56.119	0.137	214
313	32	K	CB	C	13	34.459	0.031	214
314	32	K	CD	C	13	29.832	0	214
315	32	K	CG	C	13	25.456	0	214
316	32	K	H	H	1	8.189	0.003	214
317	32	K	HA	H	1	4.432	0.009	214
318	32	K	HB2	H	1	1.556	0.015	214
319	32	K	HB3	H	1	1.381	0.009	214
320	32	K	N	N	15	118.238	0.054	214
321	33	F	C	C	13	180.266	0	215
322	33	F	CA	C	13	55.444	0.042	215
323	33	F	CB	C	13	42.207	0.07	215
324	33	F	CD1	C	13	132.189	0.065	215
325	33	F	CE1	C	13	130.886	0	215
326	33	F	CZ	C	13	131.102	0	215
327	33	F	H	H	1	8.622	0.006	215
328	33	F	HA	H	1	4.603	0.008	215
329	33	F	HB2	H	1	1.094	0.013	215
330	33	F	HB3	H	1	2.323	0.011	215
331	33	F	HZ	H	1	7.178	0	215

332	33	F	N	N	15	118.702	0.043	215
333	33	F	HD1	H	1	6.927	0.016	215
334	33	F	HD2	H	1	6.927	0.016	215
335	33	F	HE1	H	1	7.212	0	215
336	33	F	HE2	H	1	7.212	0	215
337	34	P	C	C	13	176.604	0	216
338	34	P	CA	C	13	62.346	0.089	216
339	34	P	CB	C	13	31.881	0.043	216
340	34	P	CD	C	13	49.61	0.066	216
341	34	P	HA	H	1	4.783	0.01	216
342	34	P	HB2	H	1	2.287	0.005	216
343	34	P	HB3	H	1	2.098	0.009	216
344	34	P	HD2	H	1	3.672	0.009	216
345	34	P	HD3	H	1	3.672	0.009	216
346	35	S	C	C	13	176.314	0.01	217
347	35	S	CA	C	13	56.43	0.069	217
348	35	S	CB	C	13	64.891	0.08	217
349	35	S	H	H	1	8.557	0.006	217
350	35	S	HA	H	1	5.06	0.012	217
351	35	S	HB2	H	1	4.007	0.007	217
352	35	S	HB3	H	1	3.638	0.009	217
353	35	S	N	N	15	119.522	0.056	217
354	36	R	C	C	13	176.318	0	218
355	36	R	CA	C	13	58.901	0	218
356	36	R	CB	C	13	29.947	0	218
357	36	R	H	H	1	8.436	0.003	218
358	36	R	N	N	15	123.469	0.055	218
359	37	S	C	C	13	174.267	0	219
360	37	S	CA	C	13	58.035	0.089	219
361	37	S	CB	C	13	62.849	0.07	219
362	37	S	HA	H	1	4.584	0.007	219
363	37	S	HB2	H	1	3.952	0.007	219
364	37	S	HB3	H	1	3.865	0.005	219
365	38	S	C	C	13	173.949	0	220
366	38	S	CA	C	13	59.646	0.053	220
367	38	S	CB	C	13	62.735	0.061	220
368	38	S	H	H	1	7.772	0.003	220
369	38	S	HA	H	1	4.38	0.009	220
370	38	S	N	N	15	113.401	0.072	220
371	38	S	HB2	H	1	4.06	0.003	220
372	38	S	HB3	H	1	4.06	0.003	220
373	39	G	C	C	13	174.581	0.004	221
374	39	G	CA	C	13	44.848	0.106	221
375	39	G	H	H	1	8.102	0.005	221
376	39	G	HA2	H	1	4.186	0.012	221
377	39	G	HA3	H	1	4.075	0.01	221
378	39	G	N	N	15	109.501	0.042	221
379	40	I	C	C	13	176.137	0.002	222
380	40	I	CA	C	13	58.811	0.067	222
381	40	I	CB	C	13	42.898	0.048	222
382	40	I	CD1	C	13	13.647	0.034	222
383	40	I	CG1	C	13	26.558	0.099	222
384	40	I	CG2	C	13	19.3	0.055	222
385	40	I	H	H	1	8.577	0.004	222
386	40	I	HA	H	1	5.391	0.01	222
387	40	I	HB	H	1	1.817	0.009	222
388	40	I	HG12	H	1	1.41	0.015	222
389	40	I	HG13	H	1	1.126	0.013	222
390	40	I	N	N	15	120.191	0.059	222
391	40	I	HD1	H	1	0.751	0.009	222
392	40	I	HG2	H	1	0.797	0.009	222
393	41	F	C	C	13	179.798	0.007	223
394	41	F	CA	C	13	55.779	0.081	223
395	41	F	CB	C	13	41.774	0.099	223
396	41	F	CD1	C	13	132.545	0.115	223
397	41	F	CE1	C	13	130.35	0.079	223
398	41	F	CZ	C	13	129.358	0.071	223
399	41	F	H	H	1	9.263	0.005	223
400	41	F	HA	H	1	5.082	0.012	223
401	41	F	HB2	H	1	3.246	0.004	223
402	41	F	HB3	H	1	3.017	0.012	223
403	41	F	HZ	H	1	5.574	0.013	223
404	41	F	N	N	15	121.734	0.046	223
405	41	F	HD1	H	1	7.019	0.01	223
406	41	F	HD2	H	1	7.019	0.01	223
407	41	F	HE1	H	1	6.658	0.01	223
408	41	F	HE2	H	1	6.658	0.01	223
409	42	L	C	C	13	177.192	0.001	224
410	42	L	CA	C	13	52.91	0.086	224
411	42	L	CB	C	13	44.359	0.086	224
412	42	L	CD1	C	13	25.501	0.064	224
413	42	L	CD2	C	13	24.12	0.084	224
414	42	L	CG	C	13	27.33	0.06	224
415	42	L	H	H	1	8.572	0.006	224
416	42	L	HA	H	1	5.646	0.012	224
417	42	L	HB2	H	1	1.888	0.013	224
418	42	L	HB3	H	1	1.276	0.016	224
419	42	L	HG	H	1	1.683	0.012	224
420	42	L	N	N	15	120.219	0.062	224
421	42	L	HD1	H	1	0.893	0.008	224
422	42	L	HD2	H	1	0.679	0.01	224
423	43	Y	C	C	13	180.754	0.015	225
424	43	Y	CA	C	13	58.694	0.094	225
425	43	Y	CB	C	13	45.681	0.059	225

426	43	Y	CD1	C	13	133.822	0.022	225
427	43	Y	CE1	C	13	118.321	0.109	225
428	43	Y	H	H	1	9.243	0.006	225
429	43	Y	HA	H	1	4.917	0.014	225
430	43	Y	HB2	H	1	3.042	0.014	225
431	43	Y	HB3	H	1	2.939	0.017	225
432	43	Y	N	N	15	124.342	0.062	225
433	43	Y	HD1	H	1	7.113	0.007	225
434	43	Y	HD2	H	1	7.113	0.007	225
435	43	Y	HE1	H	1	6.537	0.007	225
436	43	Y	HE2	H	1	6.537	0.007	225
437	44	N	C	C	13	173.844	0	226
438	44	N	CA	C	13	51.115	0.08	226
439	44	N	CB	C	13	41.816	0.084	226
440	44	N	H	H	1	6.933	0.007	226
441	44	N	HA	H	1	5.058	0.011	226
442	44	N	HB2	H	1	3.201	0.01	226
443	44	N	HB3	H	1	2.632	0.008	226
444	44	N	HD21	H	1	7.162	0.006	226
445	44	N	HD22	H	1	6.777	0.001	226
446	44	N	N	N	15	118.446	0.047	226
447	44	N	ND2	N	15	108.081	0.05	226
448	45	F	C	C	13	179.608	0.005	227
449	45	F	CA	C	13	62.04	0.143	227
450	45	F	CB	C	13	39.072	0.139	227
451	45	F	CD1	C	13	133.687	0	227
452	45	F	H	H	1	9.249	0.002	227
453	45	F	HA	H	1	3.932	0.01	227
454	45	F	HB2	H	1	2.919	0.014	227
455	45	F	HB3	H	1	2.344	0.015	227
456	45	F	HZ	H	1	6.529	0	227
457	45	F	N	N	15	115.878	0.076	227
458	45	F	HD1	H	1	7.505	0.007	227
459	45	F	HD2	H	1	7.505	0.007	227
460	45	F	HE1	H	1	7.107	0.01	227
461	45	F	HE2	H	1	7.107	0.01	227
462	46	E	C	C	13	178.234	0.014	228
463	46	E	CA	C	13	59.331	0.063	228
464	46	E	CB	C	13	29.246	0.041	228
465	46	E	CG	C	13	36.241	0.079	228
466	46	E	H	H	1	7.175	0.007	228
467	46	E	HA	H	1	3.501	0.007	228
468	46	E	HB2	H	1	1.675	0.008	228
469	46	E	HB3	H	1	1.873	0.013	228
470	46	E	HG2	H	1	2.185	0.012	228
471	46	E	HG3	H	1	2.313	0.012	228
472	46	E	N	N	15	120.78	0.044	228
473	47	Q	C	C	13	177.091	0.003	229
474	47	Q	CA	C	13	57.917	0.06	229
475	47	Q	CB	C	13	29.678	0.073	229
476	47	Q	CG	C	13	35.778	0.058	229
477	47	Q	H	H	1	7.599	0.004	229
478	47	Q	HA	H	1	3.863	0.009	229
479	47	Q	HB2	H	1	1.378	0.014	229
480	47	Q	HB3	H	1	1.784	0.009	229
481	47	Q	HE21	H	1	7.099	0.004	229
482	47	Q	HE22	H	1	6.872	0.002	229
483	47	Q	HG2	H	1	2.438	0.013	229
484	47	Q	HG3	H	1	2.23	0.008	229
485	47	Q	N	N	15	116.252	0.064	229
486	47	Q	NE2	N	15	110.534	0.07	229
487	48	L	C	C	13	176.906	0	230
488	48	L	CA	C	13	56.648	0.082	230
489	48	L	CB	C	13	41.909	0.052	230
490	48	L	CD1	C	13	23.231	0.076	230
491	48	L	CD2	C	13	24.706	0.073	230
492	48	L	CG	C	13	26.771	0.018	230
493	48	L	H	H	1	7.545	0.004	230
494	48	L	HA	H	1	4.091	0.013	230
495	48	L	HG	H	1	1.74	0.008	230
496	48	L	N	N	15	117.637	0.045	230
497	48	L	HB2	H	1	1.377	0.009	230
498	48	L	HB3	H	1	1.377	0.009	230
499	48	L	HD1	H	1	0.939	0.014	230
500	48	L	HD2	H	1	0.839	0.01	230
501	49	K	C	C	13	175.504	0	231
502	49	K	CA	C	13	56.705	0.07	231
503	49	K	CB	C	13	28.994	0.063	231
504	49	K	CG	C	13	24.608	0.025	231
505	49	K	HA	H	1	3.976	0.015	231
506	49	K	HB2	H	1	2.07	0.015	231
507	49	K	HB3	H	1	1.939	0.007	231
508	49	K	HG2	H	1	1.403	0.016	231
509	49	K	HG3	H	1	1.403	0.016	231
510	50	M	C	C	13	176.004	0.006	232
511	50	M	CA	C	13	54.209	0.077	232
512	50	M	CB	C	13	38.987	0.027	232
513	50	M	CE	C	13	15.825	0.028	232
514	50	M	CG	C	13	31.415	0.066	232
515	50	M	H	H	1	8.251	0.01	232
516	50	M	HA	H	1	4.646	0.01	232
517	50	M	HB2	H	1	1.442	0.014	232
518	50	M	HB3	H	1	2.302	0.016	232
519	50	M	HG2	H	1	2.434	0.01	232

520	50	M	HG3	H	1	2.622	0.015	232
521	50	M	N	N	15	116.492	0.097	232
522	50	M	HE	H	1	1.887	0.005	232
523	51	N	C	C	13	176.638	0.005	233
524	51	N	CA	C	13	52.922	0.1	233
525	51	N	CB	C	13	38.667	0.073	233
526	51	N	H	H	1	9.156	0.003	233
527	51	N	HA	H	1	4.656	0.012	233
528	51	N	HD21	H	1	6.965	0.003	233
529	51	N	HD22	H	1	7.748	0.006	233
530	51	N	N	N	15	120.886	0.057	233
531	51	N	ND2	N	15	113.756	0.055	233
532	51	N	HB2	H	1	2.939	0.006	233
533	51	N	HB3	H	1	2.939	0.006	233
534	52	L	C	C	13	177.726	0.007	234
535	52	L	CA	C	13	58.828	0.089	234
536	52	L	CB	C	13	41.816	0.077	234
537	52	L	CD1	C	13	23.658	0.066	234
538	52	L	CD2	C	13	25.834	0.062	234
539	52	L	CG	C	13	26.837	0.082	234
540	52	L	H	H	1	8.949	0.003	234
541	52	L	HA	H	1	3.996	0.011	234
542	52	L	HB2	H	1	1.76	0.012	234
543	52	L	HB3	H	1	1.432	0.012	234
544	52	L	HG	H	1	1.563	0.013	234
545	52	L	N	N	15	124.914	0.042	234
546	52	L	HD1	H	1	0.854	0.007	234
547	52	L	HD2	H	1	0.83	0.01	234
548	53	D	C	C	13	178.738	0.007	235
549	53	D	CA	C	13	57.578	0.087	235
550	53	D	CB	C	13	40.019	0.054	235
551	53	D	H	H	1	8.547	0.003	235
552	53	D	HA	H	1	4.196	0.007	235
553	53	D	HB2	H	1	2.626	0.008	235
554	53	D	HB3	H	1	2.55	0.013	235
555	53	D	N	N	15	115.794	0.073	235
556	54	D	C	C	13	178.88	0.006	236
557	54	D	CA	C	13	57.237	0.098	236
558	54	D	CB	C	13	40.532	0.03	236
559	54	D	H	H	1	7.51	0.005	236
560	54	D	HA	H	1	4.524	0.012	236
561	54	D	HB2	H	1	2.806	0.014	236
562	54	D	HB3	H	1	2.673	0.014	236
563	54	D	N	N	15	119.125	0.057	236
564	55	I	C	C	13	177.512	0.008	237
565	55	I	CA	C	13	66.32	0.025	237
566	55	I	CB	C	13	38.111	0.053	237
567	55	I	CD1	C	13	14.882	0.03	237
568	55	I	CG2	C	13	18.833	0.049	237
569	55	I	H	H	1	7.946	0.004	237
570	55	I	HA	H	1	3.49	0.013	237
571	55	I	HB	H	1	2.115	0.007	237
572	55	I	N	N	15	121.481	0.051	237
573	55	I	HD1	H	1	1.105	0.005	237
574	55	I	HG12	H	1	2.06	0.014	237
575	55	I	HG13	H	1	2.06	0.014	237
576	55	I	HG2	H	1	0.978	0.009	237
577	56	V	C	C	13	176.611	0.019	238
578	56	V	CA	C	13	67.029	0.074	238
579	56	V	CB	C	13	31.393	0.06	238
580	56	V	CG1	C	13	20.944	0.033	238
581	56	V	CG2	C	13	22.943	0.053	238
582	56	V	H	H	1	8.544	0.005	238
583	56	V	HA	H	1	3.197	0.006	238
584	56	V	HB	H	1	2.22	0.01	238
585	56	V	N	N	15	119.955	0.092	238
586	56	V	HG1	H	1	0.887	0.011	238
587	56	V	HG2	H	1	0.941	0.017	238
588	57	K	C	C	13	179.214	0	239
589	57	K	CA	C	13	59.82	0.022	239
590	57	K	CB	C	13	32.495	0.091	239
591	57	K	CG	C	13	25.223	0	239
592	57	K	H	H	1	7.521	0.003	239
593	57	K	HA	H	1	3.795	0.007	239
594	57	K	N	N	15	116.126	0.05	239
595	57	K	HB2	H	1	1.934	0.007	239
596	57	K	HB3	H	1	1.934	0.007	239
597	57	K	HG2	H	1	1.497	0	239
598	57	K	HG3	H	1	1.497	0	239
599	58	E	C	C	13	179.044	0.018	240
600	58	E	CA	C	13	57.948	0.083	240
601	58	E	CB	C	13	30.226	0.016	240
602	58	E	CG	C	13	37.084	0.062	240
603	58	E	H	H	1	7.386	0.005	240
604	58	E	HA	H	1	4.197	0.006	240
605	58	E	N	N	15	117.179	0.065	240
606	58	E	HB2	H	1	2.168	0.01	240
607	58	E	HB3	H	1	2.168	0.01	240
608	58	E	HG2	H	1	2.359	0.007	240
609	58	E	HG3	H	1	2.359	0.007	240
610	59	A	C	C	13	178.288	0.001	241
611	59	A	CA	C	13	54.467	0.096	241
612	59	A	CB	C	13	19.094	0.035	241
613	59	A	H	H	1	8.899	0.004	241

614	59	A	HA	H	1	3.79	0.013	241
615	59	A	N	N	15	122.639	0.029	241
616	59	A	HB	H	1	1.321	0.009	241
617	60	K	C	C	13	176.963	0.009	242
618	60	K	CA	C	13	58.836	0.098	242
619	60	K	CB	C	13	32.276	0.068	242
620	60	K	CG	C	13	25.325	0.019	242
621	60	K	H	H	1	7.357	0.005	242
622	60	K	HA	H	1	3.635	0.015	242
623	60	K	HB2	H	1	1.811	0.013	242
624	60	K	HB3	H	1	1.724	0.009	242
625	60	K	HG2	H	1	1.898	0	242
626	60	K	HG3	H	1	1.005	0.006	242
627	60	K	N	N	15	111.729	0.053	242
628	61	N	C	C	13	174.948	0.001	243
629	61	N	CA	C	13	53.311	0.056	243
630	61	N	CB	C	13	38.84	0.114	243
631	61	N	H	H	1	7.616	0.006	243
632	61	N	HA	H	1	4.59	0.012	243
633	61	N	HB2	H	1	3.031	0.013	243
634	61	N	HB3	H	1	2.952	0.015	243
635	61	N	HD21	H	1	7.728	0.001	243
636	61	N	HD22	H	1	7.047	0.001	243
637	61	N	N	N	15	115.22	0.064	243
638	61	N	ND2	N	15	114.73	0.083	243
639	62	V	C	C	13	174.324	0	244
640	62	V	CA	C	13	60.484	0.096	244
641	62	V	CB	C	13	30.692	0.045	244
642	62	V	CG1	C	13	21.027	0.054	244
643	62	V	CG2	C	13	21.313	0.032	244
644	62	V	H	H	1	7.249	0.003	244
645	62	V	HA	H	1	2.705	0.012	244
646	62	V	HB	H	1	1.968	0.007	244
647	62	V	N	N	15	124.123	0.049	244
648	62	V	HG1	H	1	0.463	0.007	244
649	62	V	HG2	H	1	0.298	0.009	244
650	63	P	C	C	13	176.106	0.002	245
651	63	P	CA	C	13	64.69	0.057	245
652	63	P	CB	C	13	31.854	0.027	245
653	63	P	CD	C	13	51.352	0.063	245
654	63	P	CG	C	13	27.55	0	245
655	63	P	HA	H	1	4.115	0.009	245
656	63	P	HB2	H	1	2.208	0.014	245
657	63	P	HB3	H	1	1.885	0.013	245
658	63	P	HD2	H	1	3.578	0.013	245
659	63	P	HD3	H	1	2.882	0.009	245
660	64	G	C	C	13	175.591	0	246
661	64	G	CA	C	13	45.743	0.063	246
662	64	G	H	H	1	8.169	0.002	246
663	64	G	HA2	H	1	4.239	0.01	246
664	64	G	HA3	H	1	3.635	0.007	246
665	64	G	N	N	15	110.328	0.089	246
666	65	V	C	C	13	177.344	0.001	247
667	65	V	CA	C	13	67.754	0.028	247
668	65	V	CB	C	13	30.688	0.038	247
669	65	V	CG1	C	13	21.128	0.057	247
670	65	V	CG2	C	13	25.341	0.029	247
671	65	V	H	H	1	7.526	0.005	247
672	65	V	HA	H	1	3.43	0.012	247
673	65	V	HB	H	1	2.245	0.016	247
674	65	V	N	N	15	120.836	0.051	247
675	65	V	HG1	H	1	0.697	0.01	247
676	65	V	HG2	H	1	0.729	0.008	247
677	66	T	C	C	13	176.312	0	248
678	66	T	CA	C	13	66.579	0.096	248
679	66	T	CB	C	13	67.451	0.074	248
680	66	T	CG2	C	13	23.011	0.055	248
681	66	T	H	H	1	8.932	0.002	248
682	66	T	HA	H	1	3.742	0.01	248
683	66	T	HB	H	1	4.267	0.006	248
684	66	T	N	N	15	112.642	0.057	248
685	66	T	HG2	H	1	1.282	0.01	248
686	67	R	C	C	13	178.448	0.005	249
687	67	R	CA	C	13	59.22	0.081	249
688	67	R	CB	C	13	29.441	0.075	249
689	67	R	CD	C	13	42.934	0.052	249
690	67	R	CG	C	13	26.978	0.005	249
691	67	R	H	H	1	7.566	0.005	249
692	67	R	HA	H	1	3.958	0.009	249
693	67	R	HD2	H	1	3.117	0.003	249
694	67	R	HD3	H	1	3.207	0.007	249
695	67	R	N	N	15	122.608	0.034	249
696	67	R	HB2	H	1	1.827	0.015	249
697	67	R	HB3	H	1	1.827	0.015	249
698	67	R	HG2	H	1	1.605	0.011	249
699	67	R	HG3	H	1	1.605	0.011	249
700	68	L	C	C	13	179.118	0.004	250
701	68	L	CA	C	13	57.77	0.074	250
702	68	L	CB	C	13	42.841	0.027	250
703	68	L	CD1	C	13	25.835	0.075	250
704	68	L	CD2	C	13	22.931	0.093	250
705	68	L	CG	C	13	27.252	0.029	250
706	68	L	H	H	1	7.878	0.004	250
707	68	L	HA	H	1	4.07	0.009	250

708	68	L	HG	H	1	1.813	0.013	250
709	68	L	N	N	15	119.033	0.055	250
710	68	L	HB2	H	1	1.118	0.015	250
711	68	L	HB3	H	1	1.118	0.015	250
712	68	L	HD1	H	1	0.789	0.014	250
713	68	L	HD2	H	1	0.953	0.008	250
714	69	A	C	C	13	180.471	0.007	251
715	69	A	CA	C	13	54.782	0.034	251
716	69	A	CB	C	13	19.245	0.024	251
717	69	A	H	H	1	8.542	0.004	251
718	69	A	HA	H	1	4.015	0.013	251
719	69	A	N	N	15	119.447	0.045	251
720	69	A	HB	H	1	1.51	0.009	251
721	70	H	C	C	13	175.939	0.007	252
722	70	H	CA	C	13	59.357	0.079	252
723	70	H	CB	C	13	28.851	0.056	252
724	70	H	CD2	C	13	120.813	0	252
725	70	H	CE1	C	13	137.215	0	252
726	70	H	H	H	1	8.367	0.003	252
727	70	H	HA	H	1	4.104	0.008	252
728	70	H	HB2	H	1	3.332	0.006	252
729	70	H	HB3	H	1	3.192	0.004	252
730	70	H	HD2	H	1	7.146	0.005	252
731	70	H	HE1	H	1	8.186	0.014	252
732	70	H	N	N	15	118.878	0.049	252
733	71	D	C	C	13	176.64	0.009	253
734	71	D	CA	C	13	54.562	0.082	253
735	71	D	CB	C	13	41.051	0.027	253
736	71	D	H	H	1	7.663	0.003	253
737	71	D	HA	H	1	4.576	0.004	253
738	71	D	N	N	15	116.429	0.047	253
739	71	D	HB2	H	1	2.807	0.007	253
740	71	D	HB3	H	1	2.807	0.007	253
741	72	G	C	C	13	174.728	0.003	254
742	72	G	CA	C	13	45.768	0.051	254
743	72	G	H	H	1	7.891	0.003	254
744	72	G	HA2	H	1	4.03	0.01	254
745	72	G	HA3	H	1	3.504	0.011	254
746	72	G	N	N	15	109.053	0.047	254
747	73	S	C	C	13	180.943	0.006	255
748	73	S	CA	C	13	58.314	0.062	255
749	73	S	CB	C	13	64.398	0.08	255
750	73	S	H	H	1	8.066	0.003	255
751	73	S	HA	H	1	4.429	0.008	255
752	73	S	HB2	H	1	3.99	0.012	255
753	73	S	HB3	H	1	3.684	0.018	255
754	73	S	N	N	15	117.239	0.07	255
755	74	K	C	C	13	177.254	0.01	256
756	74	K	CA	C	13	56.828	0.088	256
757	74	K	CB	C	13	32.436	0.037	256
758	74	K	CG	C	13	24.8	0	256
759	74	K	H	H	1	8.298	0.004	256
760	74	K	HA	H	1	4.232	0.011	256
761	74	K	N	N	15	121.406	0.065	256
762	74	K	HB2	H	1	1.884	0.011	256
763	74	K	HB3	H	1	1.884	0.011	256
764	74	K	HG2	H	1	1.569	0	256
765	74	K	HG3	H	1	1.569	0	256
766	75	L	C	C	13	175.917	0	257
767	75	L	CA	C	13	52.428	0.06	257
768	75	L	CB	C	13	42.59	0.099	257
769	75	L	CD1	C	13	26.236	0.09	257
770	75	L	CG	C	13	26.729	0.074	257
771	75	L	H	H	1	9.131	0.004	257
772	75	L	HA	H	1	4.385	0.01	257
773	75	L	HB2	H	1	1.621	0.009	257
774	75	L	HB3	H	1	0.814	0.017	257
775	75	L	HG	H	1	1.898	0.013	257
776	75	L	N	N	15	123.042	0.034	257
777	75	L	HD1	H	1	0.784	0.013	257
778	76	P	C	C	13	175.712	0.002	258
779	76	P	CA	C	13	61.192	0.018	258
780	76	P	CB	C	13	27.811	0.02	258
781	76	P	CD	C	13	49.519	0.093	258
782	76	P	CG	C	13	27.18	0.134	258
783	76	P	HA	H	1	4.354	0.013	258
784	76	P	HD2	H	1	3.346	0.01	258
785	76	P	HD3	H	1	3.088	0.013	258
786	76	P	HB2	H	1	1.269	0.009	258
787	76	P	HB3	H	1	1.269	0.009	258
788	76	P	HG2	H	1	1.744	0.013	258
789	76	P	HG3	H	1	1.744	0.013	258
790	77	L	C	C	13	177.104	0.005	259
791	77	L	CA	C	13	59.026	0.211	259
792	77	L	CB	C	13	41.755	0.066	259
793	77	L	CD1	C	13	26.835	0.073	259
794	77	L	CD2	C	13	23.553	0.044	259
795	77	L	CG	C	13	26.785	0.032	259
796	77	L	H	H	1	7.672	0.003	259
797	77	L	HA	H	1	3.754	0.009	259
798	77	L	HB2	H	1	2.076	0.011	259
799	77	L	HB3	H	1	1.463	0.008	259
800	77	L	HG	H	1	1.165	0.009	259
801	77	L	N	N	15	125.513	0.029	259

802	77	L	HD1	H	1	1.148	0.014	259
803	77	L	HD2	H	1	0.744	0.011	259
804	78	R	C	C	13	178.195	0.005	260
805	78	R	CA	C	13	60.488	0.096	260
806	78	R	CB	C	13	28.155	0.043	260
807	78	R	CG	C	13	27.247	0	260
808	78	R	H	H	1	8.236	0.007	260
809	78	R	HA	H	1	3.07	0.009	260
810	78	R	N	N	15	120.371	0.075	260
811	78	R	HB2	H	1	1.528	0.01	260
812	78	R	HB3	H	1	1.528	0.01	260
813	79	C	C	C	13	175.639	0.008	261
814	79	C	CA	C	13	62.283	0.095	261
815	79	C	CB	C	13	25.732	0.084	261
816	79	C	H	H	1	7.865	0.006	261
817	79	C	HA	H	1	4.084	0.01	261
818	79	C	HB2	H	1	3.305	0.007	261
819	79	C	HB3	H	1	3.162	0.013	261
820	79	C	N	N	15	117.946	0.058	261
821	80	V	C	C	13	176.471	0.002	262
822	80	V	CA	C	13	67.064	0.092	262
823	80	V	CB	C	13	31.855	0.038	262
824	80	V	CG1	C	13	22.073	0.005	262
825	80	V	CG2	C	13	23.078	0.064	262
826	80	V	H	H	1	8.178	0.007	262
827	80	V	HA	H	1	3.601	0.013	262
828	80	V	HB	H	1	2.238	0.019	262
829	80	V	N	N	15	117.954	0.044	262
830	80	V	HG1	H	1	0.994	0.013	262
831	80	V	HG2	H	1	0.92	0.015	262
832	81	L	C	C	13	177.368	0.004	263
833	81	L	CA	C	13	57.958	0.107	263
834	81	L	CB	C	13	39.709	0.075	263
835	81	L	CD1	C	13	21.667	0.04	263
836	81	L	CD2	C	13	27.644	0.084	263
837	81	L	CG	C	13	26.235	0.052	263
838	81	L	H	H	1	8.737	0.004	263
839	81	L	HA	H	1	3.586	0.011	263
840	81	L	HB2	H	1	1.689	0.014	263
841	81	L	HB3	H	1	0.593	0.01	263
842	81	L	HG	H	1	0.995	0.012	263
843	81	L	N	N	15	119.046	0.045	263
844	81	L	HD1	H	1	0.209	0.011	263
845	81	L	HD2	H	1	0.798	0.011	263
846	82	G	C	C	13	181.414	0.001	264
847	82	G	CA	C	13	47.866	0.062	264
848	82	G	H	H	1	8.488	0.004	264
849	82	G	HA2	H	1	3.598	0.011	264
850	82	G	HA3	H	1	3.105	0.011	264
851	82	G	N	N	15	104.879	0.068	264
852	83	W	C	C	13	179.52	0.005	265
853	83	W	CA	C	13	61.034	0.042	265
854	83	W	CB	C	13	29.055	0.066	265
855	83	W	CD1	C	13	126.803	0.067	265
856	83	W	CE3	C	13	120.166	0.064	265
857	83	W	CH2	C	13	124.679	0.053	265
858	83	W	CZ2	C	13	114.875	0.039	265
859	83	W	CZ3	C	13	121.693	0	265
860	83	W	H	H	1	8.693	0.009	265
861	83	W	HA	H	1	3.838	0.006	265
862	83	W	HB2	H	1	3.528	0.014	265
863	83	W	HB3	H	1	3.398	0.007	265
864	83	W	HD1	H	1	6.848	0.006	265
865	83	W	HE1	H	1	10.528	0.003	265
866	83	W	HE3	H	1	7.666	0.007	265
867	83	W	HH2	H	1	7.234	0.009	265
868	83	W	HZ2	H	1	7.475	0.007	265
869	83	W	HZ3	H	1	7.099	0.011	265
870	83	W	N	N	15	121.944	0.054	265
871	83	W	NE1	N	15	131.176	0.027	265
872	84	V	C	C	13	178.973	0	266
873	84	V	CA	C	13	66.513	0.097	266
874	84	V	CB	C	13	32.368	0.037	266
875	84	V	CG1	C	13	23.398	0.034	266
876	84	V	CG2	C	13	22.623	0.051	266
877	84	V	H	H	1	8.733	0.003	266
878	84	V	HA	H	1	3.633	0.013	266
879	84	V	HB	H	1	2.15	0.014	266
880	84	V	N	N	15	117.832	0.053	266
881	84	V	HG1	H	1	1.028	0.014	266
882	84	V	HG2	H	1	1.207	0.01	266
883	85	A	C	C	13	179.201	0.006	267
884	85	A	CA	C	13	55.161	0.087	267
885	85	A	CB	C	13	16.717	0.037	267
886	85	A	H	H	1	9.235	0.003	267
887	85	A	HA	H	1	3.581	0.012	267
888	85	A	N	N	15	124.245	0.041	267
889	85	A	HB	H	1	-0.06	0.009	267
890	86	L	C	C	13	178.388	0.003	268
891	86	L	CA	C	13	57.858	0.063	268
892	86	L	CB	C	13	41.856	0.097	268
893	86	L	CD1	C	13	24.043	0.041	268
894	86	L	CD2	C	13	24.755	0.035	268
895	86	L	CG	C	13	26.492	0.027	268

896	86	L	H	H	1	8.111	0.005	268
897	86	L	HA	H	1	3.88	0.015	268
898	86	L	HB2	H	1	1.65	0.008	268
899	86	L	HB3	H	1	1.618	0.007	268
900	86	L	HG	H	1	1.662	0.007	268
901	86	L	N	N	15	118.907	0.029	268
902	86	L	HD1	H	1	0.875	0.013	268
903	86	L	HD2	H	1	0.835	0.018	268
904	87	A	C	C	13	179.246	0.004	269
905	87	A	CA	C	13	53.347	0.04	269
906	87	A	CB	C	13	19.535	0.035	269
907	87	A	H	H	1	6.922	0.005	269
908	87	A	HA	H	1	4.169	0.005	269
909	87	A	N	N	15	115.797	0.053	269
910	87	A	HB	H	1	1.326	0.009	269
911	88	N	C	C	13	174.608	0	270
912	88	N	CA	C	13	55.439	0.071	270
913	88	N	CB	C	13	44.28	0.034	270
914	88	N	H	H	1	7.447	0.004	270
915	88	N	HA	H	1	4.916	0.012	270
916	88	N	HB2	H	1	2.777	0.007	270
917	88	N	HB3	H	1	2.47	0.015	270
918	88	N	HD21	H	1	7.79	0.008	270
919	88	N	HD22	H	1	8.139	0.003	270
920	88	N	N	N	15	112.295	0.055	270
921	88	N	ND2	N	15	117.776	0.071	270
922	89	S	C	C	13	174.432	0.035	271
923	89	S	CA	C	13	55.967	0.094	271
924	89	S	CB	C	13	62.688	0.029	271
925	89	S	H	H	1	8.915	0.003	271
926	89	S	HA	H	1	4.967	0.014	271
927	89	S	HB2	H	1	4.03	0.013	271
928	89	S	HB3	H	1	3.839	0.01	271
929	89	S	N	N	15	115.055	0.046	271
930	90	K	C	C	13	178.617	0.011	272
931	90	K	CA	C	13	59.036	0.175	272
932	90	K	CB	C	13	31.738	0.025	272
933	90	K	CD	C	13	28.632	0	272
934	90	K	CG	C	13	25.012	0.074	272
935	90	K	H	H	1	9	0.005	272
936	90	K	HA	H	1	3.999	0.013	272
937	90	K	HB2	H	1	1.859	0.016	272
938	90	K	HB3	H	1	1.835	0.011	272
939	90	K	N	N	15	132.231	0.027	272
940	90	K	HG2	H	1	1.512	0.008	272
941	90	K	HG3	H	1	1.512	0.008	272
942	91	K	C	C	13	178.142	0.001	273
943	91	K	CA	C	13	59.352	0.069	273
944	91	K	CB	C	13	32.365	0.074	273
945	91	K	CD	C	13	29.248	0	273
946	91	K	CG	C	13	25.201	0	273
947	91	K	H	H	1	8.247	0.004	273
948	91	K	HA	H	1	3.919	0.014	273
949	91	K	HB2	H	1	1.892	0.011	273
950	91	K	HB3	H	1	1.767	0.01	273
951	91	K	N	N	15	118.083	0.041	273
952	91	K	HD2	H	1	1.38	0.008	273
953	91	K	HD3	H	1	1.38	0.008	273
954	91	K	HE2	H	1	3.244	0	273
955	91	K	HE3	H	1	3.244	0	273
956	91	K	HG2	H	1	1.485	0.008	273
957	91	K	HG3	H	1	1.485	0.008	273
958	92	F	C	C	13	176.707	0.004	274
959	92	F	CA	C	13	61.535	0.097	274
960	92	F	CB	C	13	39.56	0.07	274
961	92	F	H	H	1	7.787	0.004	274
962	92	F	HA	H	1	3.416	0.013	274
963	92	F	HB2	H	1	2.305	0.019	274
964	92	F	HB3	H	1	2.917	0.009	274
965	92	F	N	N	15	119.369	0.052	274
966	92	F	HD1	H	1	6.15	0	274
967	92	F	HD2	H	1	6.15	0	274
968	93	Q	C	C	13	178.011	0.003	275
969	93	Q	CA	C	13	59.126	0.186	275
970	93	Q	CB	C	13	28.061	0.062	275
971	93	Q	CG	C	13	33.919	0.078	275
972	93	Q	H	H	1	8.111	0.003	275
973	93	Q	HA	H	1	3.671	0.008	275
974	93	Q	HB2	H	1	2.347	0.012	275
975	93	Q	HB3	H	1	1.986	0.009	275
976	93	Q	HE21	H	1	6.947	0.003	275
977	93	Q	HE22	H	1	6.62	0.004	275
978	93	Q	HG2	H	1	2.797	0.011	275
979	93	Q	HG3	H	1	2.542	0.007	275
980	93	Q	N	N	15	115.18	0.056	275
981	93	Q	NE2	N	15	108.556	0.076	275
982	94	L	C	C	13	178.363	0.014	276
983	94	L	CA	C	13	56.693	0.043	276
984	94	L	CB	C	13	42.404	0.047	276
985	94	L	CD1	C	13	24.739	0.11	276
986	94	L	CD2	C	13	23.406	0.036	276
987	94	L	CG	C	13	26.809	0	276
988	94	L	H	H	1	7.535	0.005	276
989	94	L	HA	H	1	4.093	0.013	276

990	94	L	HB2	H	1	1.747	0.01	276
991	94	L	HB3	H	1	1.531	0.009	276
992	94	L	HG	H	1	1.927	0	276
993	94	L	N	N	15	116.457	0.072	276
994	94	L	HD1	H	1	0.846	0.004	276
995	94	L	HD2	H	1	0.847	0.006	276
996	95	L	C	C	13	177.929	0.002	277
997	95	L	CA	C	13	55.935	0.087	277
998	95	L	CB	C	13	43.003	0.071	277
999	95	L	CD1	C	13	25.406	0.094	277
1000	95	L	CD2	C	13	22.306	0.04	277
1001	95	L	CG	C	13	26.029	0.046	277
1002	95	L	H	H	1	7.201	0.003	277
1003	95	L	HA	H	1	3.846	0.009	277
1004	95	L	HB2	H	1	0.695	0.012	277
1005	95	L	HB3	H	1	-0.163	0.009	277
1006	95	L	HG	H	1	1.378	0.013	277
1007	95	L	N	N	15	118.659	0.048	277
1008	95	L	HD1	H	1	0.391	0.01	277
1009	95	L	HD2	H	1	0.469	0.01	277
1010	96	V	C	C	13	180.638	0.006	278
1011	96	V	CA	C	13	60.965	0.082	278
1012	96	V	CB	C	13	33.006	0.11	278
1013	96	V	CG1	C	13	20.609	0.053	278
1014	96	V	CG2	C	13	23.69	0.081	278
1015	96	V	H	H	1	7.413	0.003	278
1016	96	V	HA	H	1	4.475	0.012	278
1017	96	V	HB	H	1	1.232	0.009	278
1018	96	V	N	N	15	119.221	0.072	278
1019	96	V	HG1	H	1	0.651	0.013	278
1020	96	V	HG2	H	1	0.968	0.007	278
1021	97	E	C	C	13	176.527	0.007	279
1022	97	E	CA	C	13	55.695	0.066	279
1023	97	E	CB	C	13	29.136	0.116	279
1024	97	E	CG	C	13	38.619	0.092	279
1025	97	E	H	H	1	7.571	0.005	279
1026	97	E	HA	H	1	4.411	0.006	279
1027	97	E	HB2	H	1	2.326	0.006	279
1028	97	E	HB3	H	1	1.339	0.007	279
1029	97	E	HG2	H	1	2.468	0.008	279
1030	97	E	HG3	H	1	2.397	0.006	279
1031	97	E	N	N	15	128.804	0.052	279
1032	98	A	C	C	13	179.444	0.001	280
1033	98	A	CA	C	13	56.038	0.087	280
1034	98	A	CB	C	13	18.536	0.036	280
1035	98	A	H	H	1	8.708	0.003	280
1036	98	A	HA	H	1	3.856	0.01	280
1037	98	A	N	N	15	127.944	0.033	280
1038	98	A	HB	H	1	1.437	0.011	280
1039	99	N	C	C	13	178.883	0.011	281
1040	99	N	CA	C	13	56.637	0.063	281
1041	99	N	CB	C	13	37.968	0.048	281
1042	99	N	H	H	1	8.786	0.003	281
1043	99	N	HA	H	1	4.548	0.014	281
1044	99	N	HB2	H	1	3.181	0.008	281
1045	99	N	HB3	H	1	2.885	0.009	281
1046	99	N	HD21	H	1	7.999	0.002	281
1047	99	N	HD22	H	1	7.045	0.001	281
1048	99	N	N	N	15	118.263	0.068	281
1049	99	N	ND2	N	15	114.717	0.08	281
1050	100	K	C	C	13	179.733	0.01	282
1051	100	K	CA	C	13	59.312	0.048	282
1052	100	K	CB	C	13	32.928	0.037	282
1053	100	K	CG	C	13	26.139	0.013	282
1054	100	K	H	H	1	8.532	0.003	282
1055	100	K	HA	H	1	4.132	0.013	282
1056	100	K	HG2	H	1	1.581	0.014	282
1057	100	K	HG3	H	1	1.53	0	282
1058	100	K	N	N	15	121.616	0.054	282
1059	100	K	HB2	H	1	1.879	0.009	282
1060	100	K	HB3	H	1	1.879	0.009	282
1061	101	L	C	C	13	177.531	0.001	283
1062	101	L	CA	C	13	57.852	0.107	283
1063	101	L	CB	C	13	41.487	0.105	283
1064	101	L	CD1	C	13	24.036	0.07	283
1065	101	L	CD2	C	13	27.34	0.083	283
1066	101	L	CG	C	13	26.796	0.066	283
1067	101	L	H	H	1	8.615	0.002	283
1068	101	L	HA	H	1	4.158	0.013	283
1069	101	L	HB2	H	1	2.13	0.009	283
1070	101	L	HB3	H	1	1.5	0.017	283
1071	101	L	HG	H	1	1.864	0.008	283
1072	101	L	N	N	15	119.768	0.046	283
1073	101	L	HD1	H	1	1.299	0.011	283
1074	101	L	HD2	H	1	1.185	0.009	283
1075	102	N	C	C	13	177.391	0.008	284
1076	102	N	CA	C	13	57.203	0.108	284
1077	102	N	CB	C	13	38.425	0.058	284
1078	102	N	H	H	1	8.305	0.007	284
1079	102	N	HA	H	1	4.447	0.008	284
1080	102	N	HB2	H	1	2.966	0.012	284
1081	102	N	HB3	H	1	2.858	0.007	284
1082	102	N	HD21	H	1	7.566	0.001	284
1083	102	N	HD22	H	1	6.756	0.002	284

1084	102	N	N	N	15	118.035	0.096	284
1085	102	N	ND2	N	15	111.684	0.034	284
1086	103	K	C	C	13	178.304	0.006	285
1087	103	K	CA	C	13	59.378	0.094	285
1088	103	K	CB	C	13	32.258	0.04	285
1089	103	K	CD	C	13	28.706	0	285
1090	103	K	CG	C	13	24.96	0.035	285
1091	103	K	H	H	1	7.315	0.003	285
1092	103	K	HA	H	1	4.136	0.006	285
1093	103	K	N	N	15	119.326	0.025	285
1094	103	K	HB2	H	1	2.009	0.013	285
1095	103	K	HB3	H	1	2.009	0.013	285
1096	103	K	HG2	H	1	1.544	0.012	285
1097	103	K	HG3	H	1	1.544	0.012	285
1098	104	I	C	C	13	178.239	0.022	286
1099	104	I	CA	C	13	64.872	0.066	286
1100	104	I	CB	C	13	38.74	0.072	286
1101	104	I	CD1	C	13	14.193	0.056	286
1102	104	I	CG1	C	13	29.232	0.066	286
1103	104	I	CG2	C	13	16.379	0.033	286
1104	104	I	H	H	1	7.908	0.003	286
1105	104	I	HA	H	1	3.918	0.012	286
1106	104	I	HB	H	1	2.042	0.01	286
1107	104	I	HG12	H	1	1.813	0.01	286
1108	104	I	HG13	H	1	1.281	0.008	286
1109	104	I	N	N	15	119.743	0.037	286
1110	104	I	HD1	H	1	0.888	0.011	286
1111	104	I	HG2	H	1	0.909	0.009	286
1112	105	M	C	C	13	177.888	0.015	287
1113	105	M	CA	C	13	60.43	0.092	287
1114	105	M	CB	C	13	33.224	0.042	287
1115	105	M	CE	C	13	16.032	0.021	287
1116	105	M	CG	C	13	31.816	0.107	287
1117	105	M	H	H	1	8.795	0.004	287
1118	105	M	HA	H	1	4.182	0.013	287
1119	105	M	HB2	H	1	2.5	0.017	287
1120	105	M	HB3	H	1	2.292	0.018	287
1121	105	M	HG2	H	1	2.593	0.012	287
1122	105	M	HG3	H	1	3.101	0.008	287
1123	105	M	N	N	15	117.919	0.097	287
1124	105	M	HE	H	1	2.14	0.005	287
1125	106	Q	C	C	13	178.565	0.004	288
1126	106	Q	CA	C	13	59.132	0.17	288
1127	106	Q	CB	C	13	28.527	0.065	288
1128	106	Q	CG	C	13	34.027	0.009	288
1129	106	Q	H	H	1	8.027	0.009	288
1130	106	Q	HA	H	1	4.065	0.011	288
1131	106	Q	HB2	H	1	2.251	0.004	288
1132	106	Q	HB3	H	1	2.293	0.001	288
1133	106	Q	HE21	H	1	7.499	0.001	288
1134	106	Q	HE22	H	1	6.905	0	288
1135	106	Q	HG2	H	1	2.536	0.005	288
1136	106	Q	HG3	H	1	2.558	0.004	288
1137	106	Q	N	N	15	117.269	0.098	288
1138	106	Q	NE2	N	15	112.481	0.027	288
1139	107	D	C	C	13	179.226	0.009	289
1140	107	D	CA	C	13	57.573	0.112	289
1141	107	D	CB	C	13	40.377	0.083	289
1142	107	D	H	H	1	8.336	0.005	289
1143	107	D	HA	H	1	4.49	0.011	289
1144	107	D	HB2	H	1	2.918	0.011	289
1145	107	D	HB3	H	1	2.752	0.012	289
1146	107	D	N	N	15	120.127	0.097	289
1147	108	D	C	C	13	178.939	0.008	290
1148	108	D	CA	C	13	57.552	0.091	290
1149	108	D	CB	C	13	40.869	0.092	290
1150	108	D	H	H	1	8.707	0.006	290
1151	108	D	HA	H	1	4.524	0.01	290
1152	108	D	HB2	H	1	3.039	0.01	290
1153	108	D	HB3	H	1	2.576	0.011	290
1154	108	D	N	N	15	121.218	0.057	290
1155	109	L	C	C	13	179.875	0.003	291
1156	109	L	CA	C	13	57.827	0.085	291
1157	109	L	CB	C	13	40.986	0.072	291
1158	109	L	CD1	C	13	25.208	0.087	291
1159	109	L	CD2	C	13	22.395	0.058	291
1160	109	L	CG	C	13	26.752	0.013	291
1161	109	L	H	H	1	8.432	0.005	291
1162	109	L	HA	H	1	4.18	0.013	291
1163	109	L	HB2	H	1	1.566	0.008	291
1164	109	L	HB3	H	1	1.999	0.013	291
1165	109	L	HG	H	1	2.131	0.008	291
1166	109	L	N	N	15	118.926	0.022	291
1167	109	L	HD1	H	1	0.902	0.01	291
1168	109	L	HD2	H	1	0.869	0.006	291
1169	110	N	C	C	13	177.418	0.014	292
1170	110	N	CA	C	13	55.623	0.093	292
1171	110	N	CB	C	13	38.451	0.08	292
1172	110	N	H	H	1	8.193	0.004	292
1173	110	N	HA	H	1	4.585	0.011	292
1174	110	N	HB2	H	1	3.027	0.006	292
1175	110	N	HB3	H	1	2.928	0.02	292
1176	110	N	HD21	H	1	7.804	0.001	292
1177	110	N	HD22	H	1	6.958	0	292

1178	110	N	N	N	15	117.836	0.096	292
1179	110	N	ND2	N	15	112.4	0.052	292
1180	111	R	C	C	13	177.774	0.011	293
1181	111	R	CA	C	13	58.495	0.039	293
1182	111	R	CB	C	13	30.18	0.034	293
1183	111	R	CD	C	13	43.338	0.044	293
1184	111	R	CG	C	13	27.372	0.067	293
1185	111	R	H	H	1	8.007	0.006	293
1186	111	R	HA	H	1	4.147	0.009	293
1187	111	R	HB2	H	1	1.907	0.014	293
1188	111	R	HB3	H	1	1.847	0.014	293
1189	111	R	HG2	H	1	1.623	0.011	293
1190	111	R	HG3	H	1	1.468	0.007	293
1191	111	R	N	N	15	119.665	0.053	293
1192	111	R	HD2	H	1	3.177	0.01	293
1193	111	R	HD3	H	1	3.177	0.01	293
1194	112	Y	C	C	13	175.955	0.004	294
1195	112	Y	CA	C	13	59.314	0.073	294
1196	112	Y	CB	C	13	38.442	0.069	294
1197	112	Y	CD1	C	13	133.439	0.031	294
1198	112	Y	CE1	C	13	118.251	0.056	294
1199	112	Y	H	H	1	8.083	0.006	294
1200	112	Y	HA	H	1	4.42	0.006	294
1201	112	Y	HB2	H	1	3.143	0.006	294
1202	112	Y	HB3	H	1	2.883	0.01	294
1203	112	Y	N	N	15	117.508	0.068	294
1204	112	Y	HD1	H	1	7.394	0.009	294
1205	112	Y	HD2	H	1	7.394	0.009	294
1206	112	Y	HE1	H	1	6.756	0.005	294
1207	112	Y	HE2	H	1	6.756	0.005	294
1208	113	A	C	C	13	177.483	0.008	295
1209	113	A	CA	C	13	52.877	0.08	295
1210	113	A	CB	C	13	19.31	0.034	295
1211	113	A	H	H	1	7.836	0.003	295
1212	113	A	HA	H	1	4.443	0.009	295
1213	113	A	N	N	15	122.607	0.017	295
1214	113	A	HB	H	1	1.506	0.008	295
1215	114	S	C	C	13	181.708	0	296
1216	114	S	CA	C	13	58.326	0.063	296
1217	114	S	CB	C	13	64.114	0.084	296
1218	114	S	H	H	1	8.006	0.004	296
1219	114	S	HA	H	1	4.542	0.015	296
1220	114	S	N	N	15	114.598	0.047	296
1221	114	S	HB2	H	1	3.932	0.006	296
1222	114	S	HB3	H	1	3.932	0.006	296
1223	115	S	C	C	13	178.644	0	297
1224	115	S	CA	C	13	60.078	0.057	297
1225	115	S	CB	C	13	64.854	0.051	297
1226	115	S	H	H	1	7.929	0.003	297
1227	115	S	HA	H	1	4.293	0.006	297
1228	115	S	N	N	15	123.054	0.037	297
1229	115	S	HB2	H	1	3.877	0.007	297
1230	115	S	HB3	H	1	3.877	0.007	297

stop_