Heinrich Heine Universität Düsseldorf

Structure and Dynamics of the N-terminal Domain of the SARS-CoV-2 Non-Structural Protein 3 (NSP3) Investigated by NMR Spectroscopy and Complementary Biophysical Methods

Inaugural Dissertation

zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen-Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Katharina Louise Vormann

aus Düsseldorf

Düsseldorf, November 2025

aus dem Institut für Physikalische Biologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

Erstgutachter: Dr. Nils-Alexander Lakomek, Institut für Physikalische Biologie,
Heinrich-Heine-Universität Düsseldorf

Zweitgutachter: Jun.-Prof. Dr. Wolfgang Hoyer, Institut für Physikalische Biologie,
Heinrich-Heine-Universität Düsseldorf

Tag der mündlichen Prüfung: **17.12.2025**

Eidesstattliche Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Es wurden keine früheren erfolglosen Promotionsversuche unternommen.

Düsseldorf, den 11.11.2025 

Ort, Datum | Unterschrift

„Research is to see what everybody else has seen,

and to think what nobody else has thought.“

Albert Szent-Györgyi

Nobel Prize in Physiology or Medicine (1937)

Acknowledgements

I would like to express my sincere gratitude to all those who have supported and guided me throughout the years of working on this project. First and foremost, I am deeply thankful to my supervisor, Dr. Nils Lakomek, for providing me with the opportunity to pursue this fascinating research topic at the IPB. I am especially grateful for his valuable time, insightful feedback on my thesis and publications, and for enabling me to attend numerous national and international conferences. I am deeply grateful for the opportunity to present my work at conferences and retreats, as well as engaging in enriching scientific discussions. It was truly an instructive and rewarding period. I would also like to express my sincere gratitude to Jun.-Prof. Wolfgang Hoyer for reviewing my thesis, and to Prof. Carsten Sachse for his continuous guidance and support throughout my PhD, particularly during our annual TAC meetings. Special thanks go to Sophia Werner, Justin Boecker, and Florian Tucholski for their constructive support. Furthermore, I am grateful to the faculty members of the department for their continuous assistance. My heartfelt thanks go to my group members Tobias Stief, Sophia Werner, and Mirko Kraus for their ongoing support and advice throughout my projects. Tobi was a great support during the NMR measurements. I would also like to thank my current master's student, Frederike Nith, for her assistance with my subproject. I am also grateful to Jessica Schmuck and Dr. Manuel Etzkorn for introducing me to the field of nucleic acids. I would like to thank all members of the IPB for creating a stimulating and enjoyable working environment that fosters not only scientific exchange but also camaraderie beyond research. I will fondly remember the conferences we attended and our shared car drives to Jülich, which made work more enjoyable. The collegial atmosphere at the IPB has been incredibly supportive. Moreover, I am very grateful that friendships developed over time with Lora Denson and Bianca Axinte, making the work experience even more enjoyable. Furthermore, the collaborative spirit during the COVIPA retreats and my exchange at the Bartenschlager lab in Heidelberg have been particularly memorable. I am especially thankful to Leon Hennecke for introducing me to new virus-related techniques. It has been a pleasure to be part of the COVIPA consortium. On a personal note, I would like to thank my parents for their unwavering encouragement and support. I am deeply grateful to my partner for standing by me throughout this journey, always providing an open ear, humor, and motivation. Lastly, I want to dedicate my thanks to my grandfather, who has always supported and motivated me throughout my career. Although he is no longer with us, I am determined to become your "Frau Doktor" - I hope you would be proud.

Thank you all,

Katharina

Table of Contents

List of Figures

List of Tables

List of Publications and Manuscripts

This work resulted in the following publications and manuscripts:

Publications

Stief, T., **Vormann, K.,** Lakomek, NA. (2024). Sensitivity-enhanced NMR ^{15}N R_1 and $R_{1\rho}$ relaxation experiments for the investigation of intrinsically disordered proteins at high magnetic fields. *Methods*, 223(10).

Stief, T., **Vormann, K.,** Lakomek, N.A (2024). NMR ^{15}N Relaxation Experiments for the Investigation of Picosecond to Nanoseconds Structural Dynamics of Proteins. *J. Vis. Exp.* (213), e67088.

Stief, T., Gremer, L., Pribicevic, S., Espinueva, D. F., **Vormann, K.,** Biehl, R., Jahn, R., Pérez-Lara, A., Lakomek, NA. (2023). Intrinsic Disorder of the Neuronal SNARE Protein SNAP25a in its Pre-fusion Conformation. *Journal of Molecular Biology*, 435, 168069.

Manuscripts

Vormann, K., Stief, T., Tucholski, FT., Cukkemane, A., Nagel-Steger, L., Bartenschlager, R., Biehl, R., Lakomek, NA. (2025). Integrative NMR pipeline: From an AlphaFold model to the dynamic validation of SARS-CoV-2 NSP3 N-terminus.
Manuscript close to submission, compare chapter 2

Vormann, K., Stief, T., Tucholski, FT., Nagel-Steger, L., Biehl, R., Lakomek, NA. (2025). Impact of ionic strengths on the self-association of the SARS-CoV-2 non-structural protein 3 N-terminal domain.
Manuscript in preparation, compare chapter 3

Tucholski, F. T., Wang, S.-Y., **Vormann, K.**, Janosch, A., Lakomek, N.-A., Nagel, L., Willbold, D., Haas, R., & Buell, A. K. (2025). Recombinant expression and characterization of a full-length pathogenic free light chain in Pichia pastoris.
Manuscript under review, Journal of Biological Chemistry.

Werner, S., Stief, T., Schulz, C. M., **Vormann, K.,** Backer, R., Schützmann, M., Gremer, L., Hoyer, W., Lakomek, N. A. (2025). Direct NMR evidence for weak binding of the vesicular SNARE protein synaptobrevin-2 / VAMP2 to the α-synuclein C-terminus.
Manuscript close to submission

Abbreviations

ACE2	Angiotensin-converting enzyme 2
AH1	Amphipathic helix 1
AI	After induction
ADP	Adenosine diphosphate
amu	Atomic mass unit
APS	Ammonium persulfate
ATP	Adenosine Triphosphate
ATR	Ataxia-telangiectasia mutated and Rad-3 related kinase
AUC	Analytical ultracentrifugation
BI	Before induction
bp	Base pairs
BSA	Bovine serum albumin
CCR	Cross-correlated relaxation
CD	Circular dichroism spectroscopy
CF	Correlation factor
CoV	Coronavirus
COVID-19	Coronavirus disease 2019
CSA	Chemical shift anisotropy
CSP	Chemical shift perturbation
CV	Column volume
Cy5	Cyanin 5
D	Diffusion coefficient
DC	Dipolar coupling
DLS	Dynamic light scattering
DMC	Double-membrane connector
DMEM	Dulbecco’s modified eagle medium
DMV	Double-membrane vesicle
DNA	Deoxyribonucleic acid
DPUP	Domain proceeding Ubl2
dsRNA	double-stranded RNA
E. coli	*Escherichia coli*

EM	Electron microscopy
ER	Endoplasmic reticulum
ERGIC	ER-Golgi intermediate compartment
EYFP	Enhanced yellow fluorescent protein
FT	Flowthrough
GFP	Green fluorescent protein
gRNA	genomic RNA
GSK-3	Glycogen synthase kinase 3
HA-tag	Human influenza hemagglutinin epitope tag
HEK cells	Human Embryonic Kidney cells
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
His-tag	Histidine (-tag)
HRP	Horseradish peroxidase
HSQC	Heteronuclear single quantum coherence
HuUb	Human ubiquitin
HVR	Hypervariable region
IB	Inclusion body
IDR	Intrinsically disordered region
IDP	Intrinsically disordered protein
IEC	Ion exchange chromatography
IF	Immunofluorescence
IMAC	Immobilized metal affinity chromatography
IP	Immunoprecipitation
IPTG	Isopropyl-ß-D-1-thiogalactopyranoside
K_{av}	Gel-phase distribution coefficient
kDa	Kilodalton
Kp	Kilobases
L	Lysate
LE	Labelling efficiency
LLPS	Liquid-like phase separation
Log	Logarithm
LP	Lambda phosphatase

LR	Linker region
NMR	Nuclear magnetic resonance
M	Marker
M-Protein	Membrane protein
MERS-CoV	Middle east respiratory syndrome coronavirus
MF	Model-free
MHV	Murine hepatitis virus
mRNA	messenger RNA
MS	Mass spectrometry
MST	Microscale thermophoresis
Munc13	Mammalian uncoordinated 13
Munc18	Mammalian uncoordinated 18
MW	Molecular weight
MWCO	Molecular weight cut-off
n	Number of repetitions
NAB	Nucleic acid binding domain
NaCl	Sodium chloride
NMR	Nuclear magnetic resonance
N-Protein	Nucleocapsid protein
ns	Number of scans
NSP	Non-structural protein
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
P	Pellet
PBS	Phosphate-buffered saline
PFA	Paraformaldehyde
pI	Isoelectric point
pH	Pondus hydrogenii
PTM	Posttranslational modification
RdRp	RNA-dependent RNA-polymerase
RNP	Ribonucleoprotein
RO	Replication Organelle

RTC	Replication-transcription complex
SAXS	Small angle X-ray scattering
S-Protein	Spike protein
SDS-PAGE	Sodium dodecylsulfate-polyacrylamide gel electrophoresis
sgRNA	sub genomic RNA
SM	SNARE motif
SNR	Signal-to-noise ratio
SV	Sedimental velocity
rcf	Relative centrifugal force
RDC	Residual dipolar coupling
RdRp	RNA-dependent RNA-polymerase
RNA	Ribonucleic acid
rpm	Revolutions per minute
R_1	Longitudinal relaxation
R_2	Transversal relaxation
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SEC	Size-exclusion chromatography
SL	Stem loop
SNARE	Soluble N-ethylmaleimide-sensitive factor attachment protein receptor
SNAP25	Synaptosomal-associated protein of 25 kDa
Syb-2	Synaptobrevin-2
Syx-1A	Syntaxin-1A
TBS-T	Tris-buffered saline (-Tween)
T_C	Tumbling time
TCEP	Tris-(2-carboxyethyl)-phosphine
TEMED	Tetramethyl ethylenediamine
TEV	*Tabacco etch virus*
TGS buffer	Tris-Glycine SDS buffer
Tm	Melting temperature
TMD	Transmembrane domain
TMPRSS2	Transmembrane serine protease 2
TMR	Transmembrane region

Tris	Tris(hydroxymethyl-)aminomethane
TROSY	Transverse relaxation-optimized spectroscopy
Ub	Ubiquitin
Ubl	Ubiquitin-like
Ubl1	Ubiquitin-like domain 1
ULK1	nc-51-like kinase 1
UTR	Untranslated region
VAMP2	Vesicle-associated membrane protein 2
V_c	Geometric column volume
V_e	Elution volume
VLP	Virus-like particles
V_o	Column void volume
WHO	World Health Organization
2xYT	2x Yeast extract and Triptone medium
3Ecto	Ectodomain 3

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused one of the largest pandemics in recent history. While rapidly developed vaccination schemes could help contain the pandemic, effective antiviral treatment, including for post-COVID symptoms, remains lacking. In this doctoral project, we aimed for deeper insights into the basic components of the viral replication machinery. This will eventually contribute to improved antiviral treatment strategies. SARS-CoV-2 relies on double-membrane vesicles (DMVs) for viral RNA replication. Non-structural protein 3 (NSP3) is a key component of the viral replication-transcription complex (RTC) and the DMV-spanning pore. This thesis investigates individual NSP3 domains - the ubiquitin-like domains 1 and 2 (Ubl1 and Ubl2) and the papain-like protease (PLpro) - to better understand viral replication. As a side project, the structure and dynamics of neuronal SNARE proteins (SNAP25 and Synaptobrevin-2), which are involved in synaptic vesicle fusion and affect vesicle plasticity, were also investigated. Biochemical expression and purification protocols were developed and established for all target proteins, enabling structural studies using nuclear magnetic resonance (NMR) spectroscopy, circular dichroism spectroscopy, small-angle X-ray scattering, and analytical ultracentrifugation. An integrative NMR-based pipeline was developed to validate AlphaFold3 structural models and analyze flexible regions. Using the N-terminal NSP3 domain Ubl1 as a model system, the predicted β-grasp fold was confirmed. Notably, NMR revealed a highly dynamic N-terminal segment (~15 residues) invisible in crystal or cryo-electron microscopy structures. NMR relaxation and secondary chemical shift data characterized its flexible nature. Further analyses showed that Ubl1 remains monomeric under physiological conditions but displays salt-dependent modulation of dynamics and weak reversible self-association, suggesting functional adaptation to crowded intracellular environments. Subcellular localization studies revealed cytosolic positioning of Ubl1 in HEK293T cells, aligning with its orientation at the cytosolic face of the DMV pore. Post-translational modifications were characterized using mass spectrometry and electrophoretic techniques. Ubl1 was phosphorylated in both bacterial and human systems, affecting its charge and possibly modulating RNA or protein interactions. NMR spectroscopy indicated random, condition-dependent binding of Ubl1 to viral ssRNA, without a defined binding site. The Ubl2 subdomain of PLpro was also studied by NMR, extending previous assignments and confirming its ubiquitin-like fold. For PLpro, optimized expression and refolding protocols yielded soluble protein suitable for high-field NMR studies. Overall, this work provides new structural and biophysical insights into key domains of SARS-CoV-2 and host proteins involved in vesicle dynamics. It establishes a framework for validating AI-generated models and studying dynamic protein regions, contributing to our understanding of coronavirus replication and viral-host interactions.

I. COVID-19 Pandemic: Outbreak, Origin, Genera, and Treatment

In December 2019, an outbreak of a novel and previously unidentified coronavirus disease (COVID-19), in Wuhan, China, captured global attention.[1] One month after the outbreak, the World Health Organization (WHO) declared COVID-19 a "public health emergency of international concern."[2] As of April 2024, the COVID-19 pandemic has resulted in over 700 million confirmed cases and approximately 7 million deaths globally (www.worldometers.info/coronavirus/). Because of the rapid spread of SARS-CoV-2, urgent global public health responses were required to control transmission, develop diagnostic tools, and accelerate vaccine development.

The causative agent of COVID-19 is severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the seventh known coronavirus capable of infecting humans.[1] Prior to the COVID-19 pandemic, several human-infecting coronaviruses were already known, including the human CoVs NL63 and 229E (members of the alpha-CoV genus), as well as SARS-CoV-1 (identified in 2003) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV; identified in 2012), both of which belong to the beta-CoV genus.[3,4,5,6]

Generally, coronaviruses (CoVs) are categorized into four genera: alpha, beta, gamma, and delta, with SARS-CoV-2 belonging to the beta coronaviruses.[7] Coronaviruses are classified within the family *Coronaviridae*, which falls under the order *Nidovirales*.[8,9] Coronaviruses are single-stranded RNA (ssRNA) viruses that rely on a host cell for replication and are capable of infecting both humans and animals.[8] Alpha-CoVs and beta-CoVs primarily infect mammals, while gamma-CoVs and delta-CoVs have a broader host range and are also capable of infecting birds.[10] Genomic analyses indicate that SARS-CoV-2 shares approximately 96% sequence identity with a bat CoV, suggesting that bats are likely the natural reservoir of the virus.[10,11] Although the exact mechanism and intermediate host involved in zoonotic transmission to humans remain unclear, the primary mode of subsequent human-to-human transmission is via respiratory droplets.[12] All known human CoVs have the potential to cause severe respiratory illnesses, including fatalities.[13] Clinical manifestations are highly variable and may include symptoms such as fever, fatigue, headache, and diarrhea.[14]

Furthermore, SARS-CoV-2 exhibits a high degree of sequence similarity with previously identified CoVs, including SARS-CoV-1 (79.6%) and MERS-CoV (50%).[15,16,17] This genomic and structural homology facilitated the early stages of vaccine development, as researchers were able to use existing data from these related viruses. To date, several vaccines and antiviral therapies have been developed to reduce the incidence of COVID-19 and to prevent further transmission of the virus.[18,19] However, the development of effective countermeasures has been challenged by the virus's high mutation rate.[20] Notably, mutations in the spike (S) protein have enhanced viral infectivity and enabled immune evasion, thereby complicating both vaccine efficacy and therapeutic targeting.[20] Based on their transmission rate, mutation rate and severity, SARS-CoV-2 are categorized into the alpha, beta, gamma, delta and omicron variant.[21]

Since SARS-CoV-2 evolves rapidly, a detailed understanding of its molecular biology is essential for developing robust and adaptable antiviral strategies. This knowledge is also crucial for effective pandemic preparedness. The recurring emergence of novel human CoVs further emphasizes the need to study existing CoVs in depth. Special attention should be given to the viral replication cycle to identify potential therapeutic and prophylactic targets.

II. The SARS-CoV-2 Genome

The SARS-CoV-2 genome is approximately 30 kb in length and features a 5' cap structure, a 3' polyadenylated tail, and a conserved viral leader sequence in the 5`untranslated region (UTR).[1,22] It encodes a variety of non-structural, structural, and accessory proteins essential for the viral life cycle.[11] The two open reading frames ORF1 and ORF2 encode for two viral polyproteins (pp1a and pp1ab).[11] Pp1a comprises for the non-structural proteins 1 through 11 (NSP1-NSP11), while pp1ab comprises NSPs 1-10 and 12-16.[11] These polyproteins are autoproteolytically processed by two viral proteases: the papain-like protease (PLpro) which cleaves NSP1 to NSP3, and the main protease (Mpro) which processes NSP4 to NSP16.[23,24,25] Among these, NSP3 is one of the largest and most multifunctional viral proteins.[26,27] It is a transmembrane protein consisting of 16 distinct domains (Figure 1).[28] The 3' region of the genome encodes the structural proteins spike (S), envelope (E), membrane (M), and nucleocapsid (N), as well as several accessory proteins.[29] These proteins are not translated

directly from the genomic RNA (gRNA), but from subgenomic RNA (sgRNA) species synthesized during replication (Figure 1). The open reading frames ORF3a, ORF3c, ORF6, ORF7a, ORF7b, ORF8, and ORF9b, encode for accessory proteins, which contribute to immune evasion, host interaction, and viral fitness.[30]

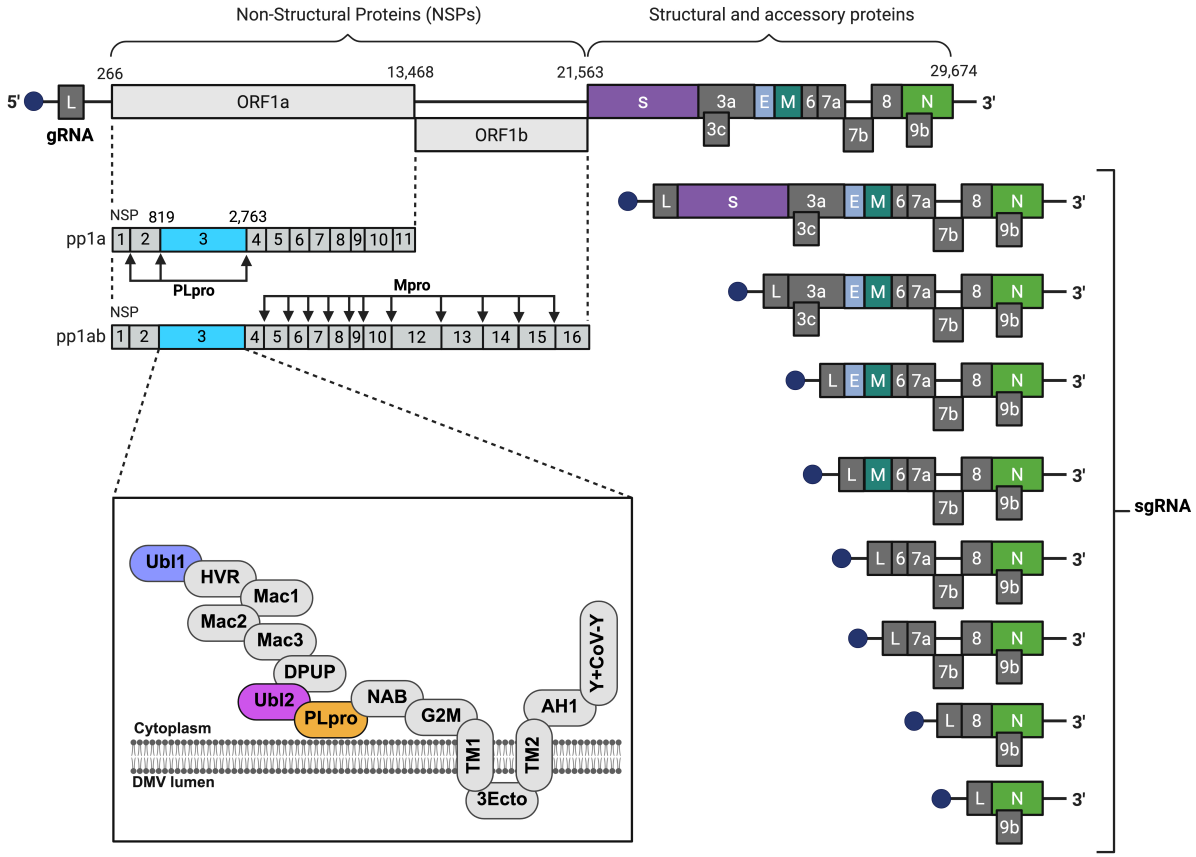


Figure 1: Genomic organization of SARS-CoV-2 highlighting NSP3. The approximately 30,000-nucleotide SARS-CoV-2 genome is capped at the 5' end, polyadenylated at the 3' end, and contains a viral leader sequence (L). Non-structural proteins (NSPs) are encoded by open reading frame 1a (ORF1a; 266-13,468 kb) and open reading frame 1b (ORF1b; 13,469-21,563 kb). ORF1a encodes NSP1-11 while ORF1b encodes NSP12 to NSP16. Two viral proteases process the translated polyproteins: the papain-like protease (PLpro) cleaves NSP1 to NSP3, and the main protease (Mpro) cleaves NSP4 to NSP16.[23,24,25] NSP3 (819-2,763 kb) is a large, multi-domain transmembrane protein comprising 16 distinct domains: ubiquitin-like domain 1 (Ubl1), a hypervariable region (HVR), three macrodomains (macrodomain 1 to 3), the domain preceding ubiquitin-like domain 2 (DPUP), the ubiquitin-like domain 2 (Ubl2), the papain-like protease (PLpro), the nucleic acid-binding domain (NAB), group 2-specific marker (G2M), two transmembrane domains (TM1 and TM2) separated by an ectodomain (3Ecto), followed by the amphipathic helix 1 (AH1) and the Y domain plus the coronavirus-specific Y domain (Y+CoV-Y) at the C-terminus.[31] The 3' end of the genome (nucleotides 21,563-29,674) encodes the structural proteins spike (S), envelope (E), membrane (M), and nucleocapsid (N), as well as multiple accessory proteins. These are translated from subgenomic RNA (sgRNA) species produced during replication. Identified accessory open reading frames include ORF3a, ORF3c, ORF6, ORF7a, ORF7b, ORF8, and ORF9b.

III. Structural Proteins of Coronaviruses

CoVs are spherical particles with an average diameter of approximately 100 nm.[32] Their envelope is composed of three capsid-forming structural proteins: the M-protein, the E-protein, and the S-protein (Figure 2).[33] The prominent spike proteins give the virus its crown-like appearance - resembling a solar corona - which is the origin of its name.[33] The S-protein is responsible for viral entry by binding to host cell receptors.[34] The M-protein is a multi-pass transmembrane protein that defines the shape of the virion and contributes to membrane curvature.[35,36] The E-protein supports viral assembly and is also associated with pathogenesis.[37] All three structural proteins are embedded in a phospholipid bilayer derived from the host cell ER-Golgi intermediate compartment (ERGIC).[38] Inside the envelope, the viral RNA genome is packaged by the N-protein, the fourth structural protein encoded by CoVs (Figure 2).[33]

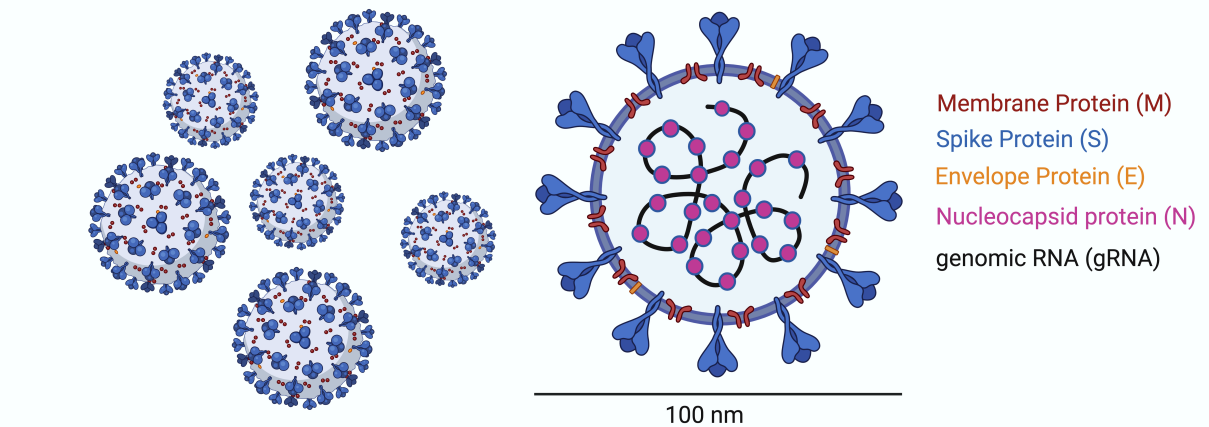


Figure 2: Molecular structure of SARS-CoV-2 with emphasis on its structural proteins. SARS-CoV-2 is an enveloped virus with an approximate diameter of 100 nm. Its viral envelope is composed of three structural proteins: the spike (S) protein, which facilitates host cell entry; the membrane (M) protein, which contributes to virion shape and assembly; and the envelope (E) protein, which is involved in viral budding and release. Encapsulated within the viral particle is a 30 kb positive-sense single-stranded RNA genome, which is packaged and organized by the nucleocapsid (N) protein into a ribonucleoprotein (RNP) complex. Figure was created in Biorender.com.

IV. SARS-CoV-2 Replication Cycle

SARS-CoV-2 is a positive-sense, ssRNA virus with a ~30 kb genome that encodes structural proteins, NSPs and accessory proteins. [11,29,39] To date, two major entry mechanisms into the host cell have been described.[40,41,42] The primary route involves direct membrane fusion, mediated by the interaction of the viral S-protein with the host angiotensin-converting

enzyme 2 (ACE2) receptor (Figure 3).[40] This process is facilitated by the host serine protease TMPRSS2, which cleaves the S-protein to promote viral-host membrane fusion.[40] An alternative entry pathway occurs via endocytosis, followed by proteolytic activation of the S-protein by cathepsins and subsequent fusion within the endosome.[41]

Upon entry and uncoating, the viral genome is released into the host cell cytoplasm (Figure 3). The positive-sense RNA genome is immediately accessible to the host ribosomal machinery and translated into two large polyproteins (pp1a and pp1ab).[43] These polyproteins are autoproteolytically processed PLpro[23] and Mpro[24], releasing 15 NSPs.[25] These NSPs assemble into the replication-transcription complex (RTC), a key driver of viral genome replication and sgRNA synthesis.[22,44]

A hallmark of CoV replication is the formation of virus-induced membrane compartments derived from the endoplasmic reticulum (ER). In SARS-CoV-2-infected cells, NSP3 and NSP4 trigger extensive remodeling of the rough ER, leading to the formation of double-membrane vesicles (DMVs).[45,46,47] These DMVs serve as replication organelles (ROs) that compartmentalize the RTC and protect viral RNA intermediates - particularly double-stranded RNA (dsRNA) - from host immune recognition.[47,48] The RNA-dependent RNA polymerase (RdRp) complex, synthesizes a negative-sense RNA intermediate from the gRNA template. This intermediate serves as the template for both full-length gRNA replication and the transcription of gRNA in sgRNAs.[49,50,51,52,53]

Within the DMV lumen, dsRNA intermediates are detectable,[54,55] but newly synthesized positive-sense RNAs must be exported into the cytoplasm for translation and packaging. Recent studies have revealed that this export is mediated by proteinaceous pores spanning the DMV membrane bilayer.[56] These pores are hypothesized to be formed by a complex involving NSP3 and NSP4.[46,56] Once in the cytoplasm, sgRNAs are translated into the structural proteins (E, M, N, S) along with accessory proteins.[52,57] The N-protein binds to newly synthesized gRNA to form the viral ribonucleoprotein (vRNP) complex, which is critical for genome encapsidation and virion assembly.[58,59,60,61] Structural proteins are co-translationally inserted into the ER membrane and trafficked to the ERGIC where they assemble with vRNPs into nascent virions.[52] Virions then mature during transit through the Golgi apparatus and the

trans-Golgi network. Finally, the fully assembled viral particles are released from the host cell via the secretory pathway.[62]

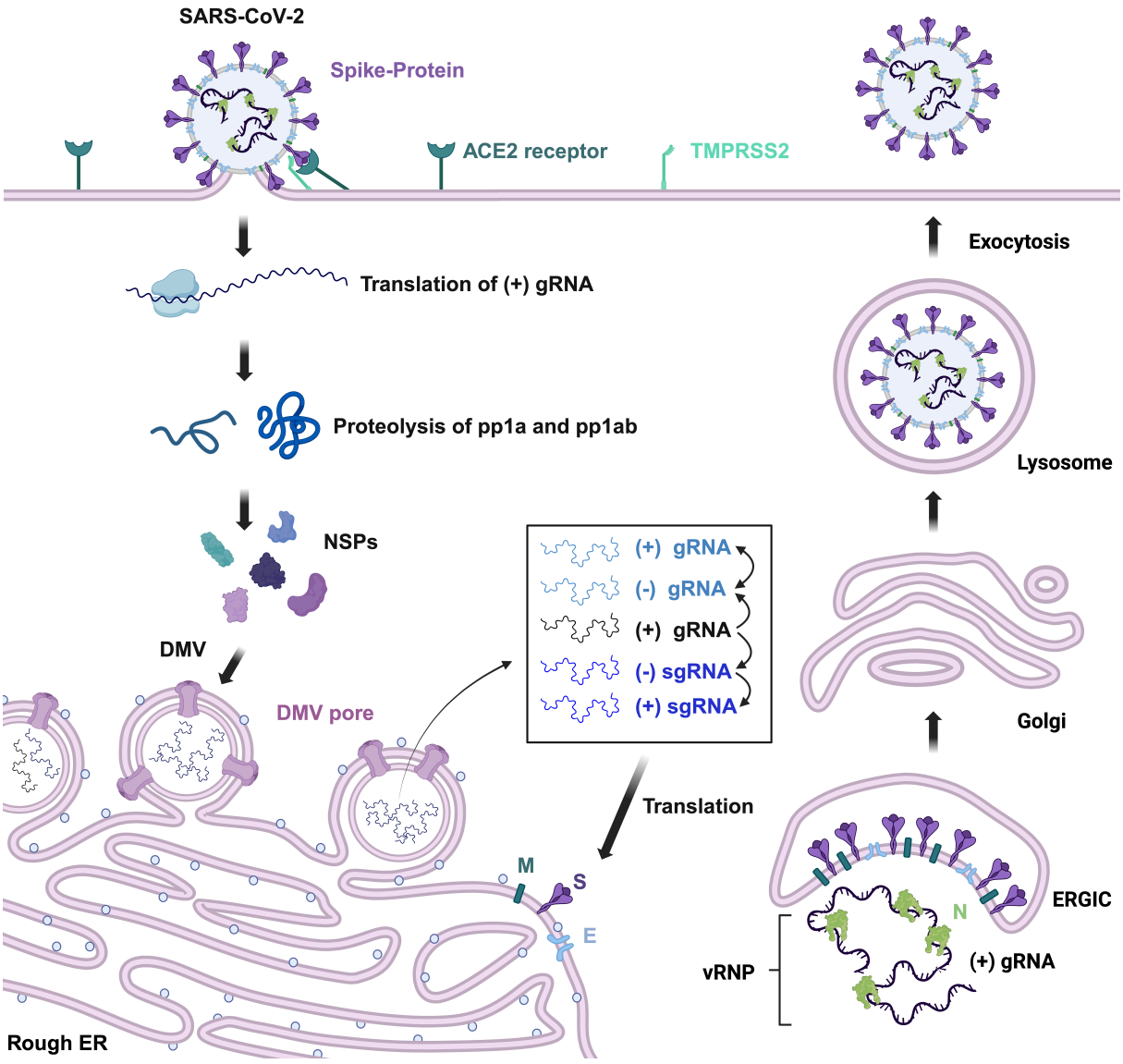


Figure 3: Schematic representation of the SARS-CoV-2 replication cycle. Viral entry is initiated by the binding of the S-protein to the ACE2 receptor on the host cell surface, facilitated by TMPRSS2, which cleaves the S-protein to enable membrane fusion. Following viral entry and uncoating, the viral RNA genome is released into the cytoplasm. Host ribosomes translate the two ORF1a and ORF1ab into polyproteins pp1a and pp1ab, which are subsequently cleaved by viral proteases into 15 NSPs. These NSPs assemble into the RTC. NSP3 and NSP4 mediate the formation of DMVs, which serve as ROs. Pore structures in the DMV membrane facilitate RNA export into the cytosol. gRNA is encapsidated by the N-protein to form vRNP complexes, while sgRNAs are translated into structural proteins (S, E, M, N) and accessory proteins. Structural proteins are inserted into the rough ER membrane and transported to the ERGIC, where they assemble with vRNPs to form nascent virions. These virions mature during transit through the Golgi and trans-Golgi network and are ultimately released from the host cell via lysosome-mediated exocytosis. Figure created with BioRender.com.

V. The Non-Structural Proteins of SARS-CoV-2

Non-structural proteins (NSPs) are viral proteins that do not contribute to the structural components of new virions but are instead primarily involved in viral replication and transcription.[18] They often function within DMVs, which are RO induced during infection.[54,63] Despite their name, "non-structural" does not imply that these proteins lack structure; rather, it refers to their functional role, indicating that they are not incorporated into the virion structure itself. NSPs perform a variety of functions, ranging from the disruption of host cellular processes to facilitating viral replication within DMVs.[18] Table 1 provides an overview of the SARS-CoV-2 NSPs and their associated functions.

Table 1: NSPs of SARS-CoV-2 and their functions.[18]

NSPs	Enzymatic activity	Function
1	None	Inhibition of host protein biosynthesis, degradation of host mRNA
2	None	Disrupting cell cycle progression, releasing NSP1, NSP2, NSP3, Inducing the
3	PLpro	Formation of DMVs, member of pore spanning DMVs, RNA-binding, N-protein interaction
4	None	Inducing the formation of DMVs, member of pore spanning DMVs
5	Mpro	Cleavage of viral polyprotein
6	None	Generating autophagosomes
7	None	Viral replication
8	None	Viral replication
9	None	Binding RNA, involved in viral replication
10	None	Cofactor of NSP14 and NSP16, stimulates viral exoribonuclease and methyltransferase
12	RNA-dependent RNA-polymerase (RdRp)	Transcription
13	Helicase	Viral replication and transcription: unwinding dsRNA
14	Exonuclease, Methyltransferase	Proof-reading and repairing viral RNA
15	Endoribonuclease	Interferes the innate immune response

VI. SARS-CoV-2 Induced Double-Membrane Vesicles

It has been demonstrated that all CoVs including SARS-CoV-2, induce extensive remodeling of intracellular membranes, particularly those of the rough ER.[47,55,56,63,64] These virus-induced membrane rearrangements give rise to ROs, which provide a protected microenvironment for viral RNA replication, thereby shielding viral RNA from host-mediated degradation.[47,55,56,63] Within these DMVs that constitute the core of these ROs, accumulations of gRNA, metabolites, viral cofactors, and NSPs have been observed. Here, NSP3, NSP4 and NSP6 play a crucial role in the formation of ROs, specifically DMVs (Figure 3).[45,56,64] It has been shown, that a co-expression of NSP3 and NSP4 is crucial for the induction of DMV formation and DMV pore formation in SARS-CoV-1 and SARS-CoV-2.[45,65,66] These vesicles originate from the ER and remain tethered to it, thereby creating a protected microenvironment that facilitates viral genome replication and transcription.[45,54,63] The functional role of NSP6 remains controversial. While some studies reported that NSP6 zippers the ER membrane, facilitates the formation of connectors between DMVs and the ER, and mediates contacts with lipid droplets via DFCP1 recruitment,[64] these findings were challenged by Zimmermann *et al.,* who demonstrated that DMVs induced by NSP3 and NSP4 remain connected to the ER even in the absence of NSP6 - suggesting that NSP6 is not essential for ER-DMV connectivity.[45] Riccardi *et al.* demonstrated that NSP6 interacts with lipid droplets via its amphipathic helix, facilitating the transfer and incorporation of lipids into NSP3/4-induced DMVs.[64] This identifies NSP6 as a central regulator of lipid-mediated membrane biogenesis within the SARS-CoV-2 replication organelle.[64] Furthermore, networks formed by the co-expression of NSP3, NSP4, and NSP6 exhibit a markedly higher vesicle density - approximately 15 DMVs per connector - and more compact, tightly apposed ER zippered domains compared to those formed by NSP3 and NSP4 alone, which display only about 3 DMVs per connector and a discernible intermembrane lumen.[64] Notably, treatment with the antiviral compound K22 diminished NSP6-dependent ER-zippering activity and significantly impaired viral replication, underscoring NSP6 as a potential therapeutic target for antiviral drug development.[64]

Cryo-electron tomography studies conducted by Wolff *et al.* in 2020 revealed the presence of molecular pores embedded in the DMV membranes of MHV infected cells (Figure 4 and 5).[56] In contrast, each DMV of SARS-CoV-2 infected cells has a diameter of approximately 336 nm and contains on average ten pores[45] exceeding the number previously reported for MHV virus

with 8 pores per DMV and a diameter of approximately 257 nm.[56] These pores are proposed to enable the bidirectional exchange of viral RNA and replication intermediates between the DMV lumen and the host cytoplasm.[46,56] The DMV-spanning pore is structurally organized into three distinct components: the hexameric prong, the upper base, and the lower base.[46] NSP3 contributes to the formation of both the prong and upper base, whereas NSP4 is responsible for assembling the lower base.[46] Consistent with these findings, truncation of the N-terminal NSP3 domains (ΔUbl1-Ubl2) resulted in a loss of curvature in the outer DMV membrane, while the inner membrane - associated with NSP4 - remained curved.[45] NSP3 and NSP4 are connected via their ectodomains, with the interaction mediated by both hydrophobic and hydrophilic residues.[45] The involved amino acids are highly conserved, underscoring the functional importance of this interface.[45]

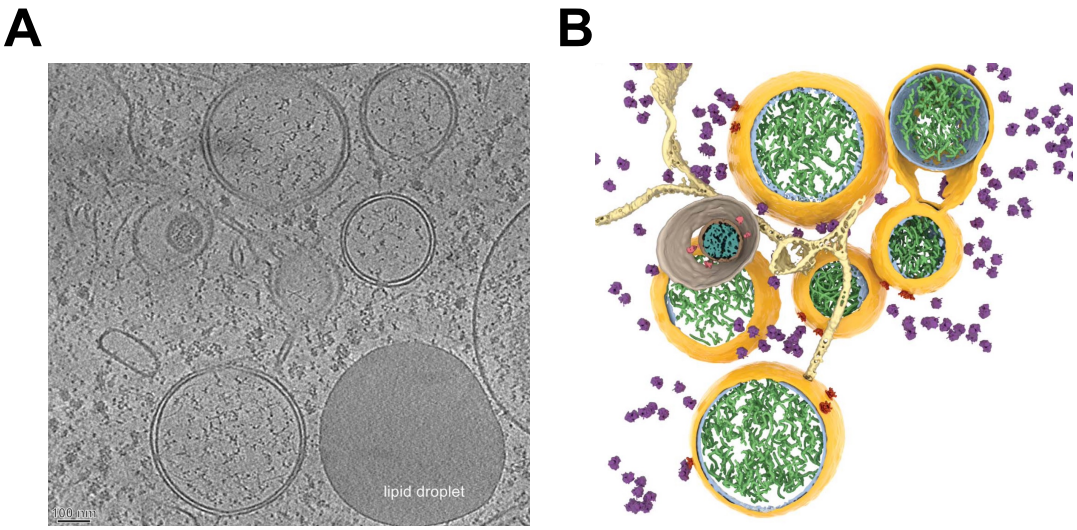


Figure 4: DMV formation in MHV infected cells, adapted and modified from Wolff *et al.*[56] (A) Tomographic slice of a MHV-infected cell. **(B)** 3D model of the corresponding tomogram showing ribosomes (violet), ER (beige), outer DMV membrane (yellow), inner DMV membrane (light blue), DMV luminal filaments (green), DMV pores (red), ERGIC (light brown), virus particle envelope (brown), S-protein (pink) and the vRNP (turquoise).

Fluorescence microscopy has shown that Ubl1 localizes to the prongs of the hexameric crown facing the cytosol, making it accessible to both host and viral proteins, including the viral N-protein.[56,67] Notably, deletion of the N-terminal domains of NSP3, specifically ΔUbl1-Ubl2, prevents the proper remodeling of the ER membrane into canonical DMVs.[45] Instead, an abnormal double-membrane network of interconnected tubules - termed double-membrane connectors (DMCs) - is formed.[45] This structural disruption is accompanied by the loss of the crown-like pore complex, reduced outer membrane curvature, and significant expansion of both the luminal space and the inter-pore distances.[45] Data indicated that domains located

between Ubl1 and Ubl2 are indispensable for NSP3 multimerization.[45] Huang *et al.* conducted a detailed investigation of the molecular organization of the DMV pore using cryogenic electron tomography (Figure 5).[46]

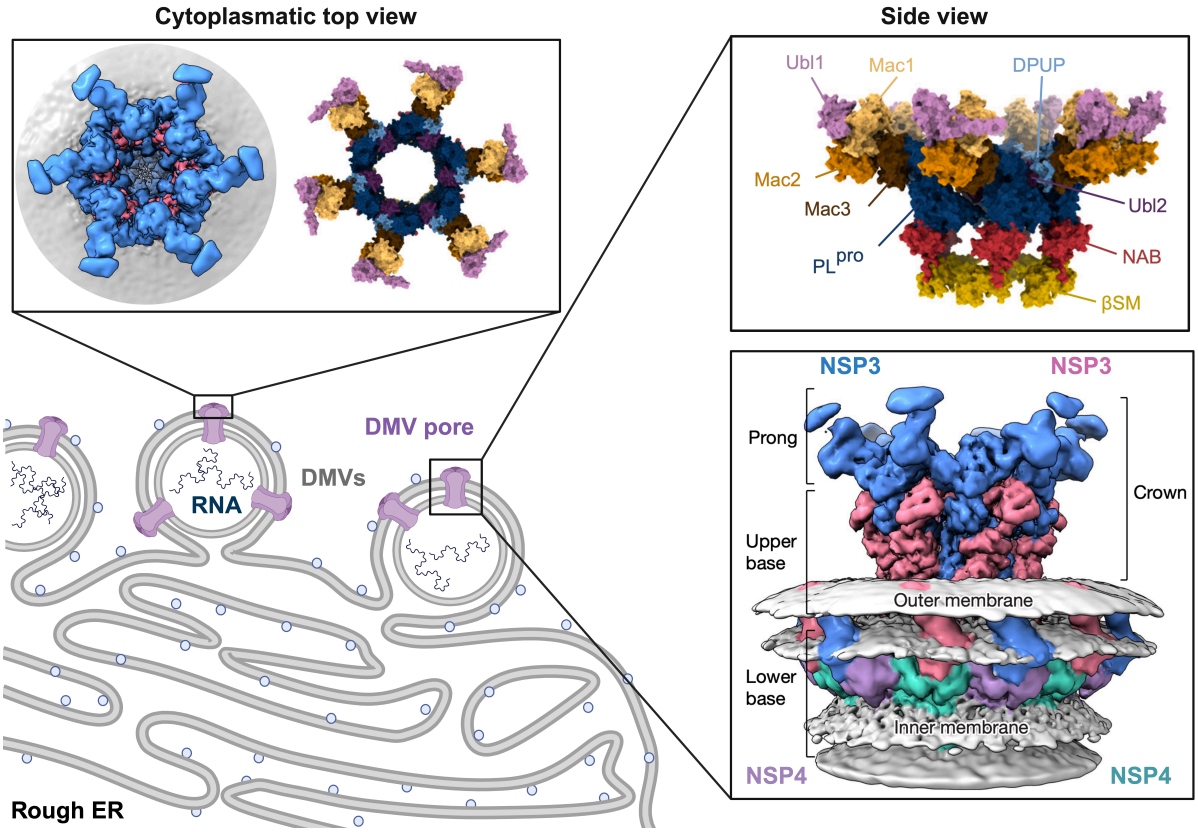


Figure 5: DMV pore formation induced by NSP3 and NSP4, adopted and modified from Huang *et al.* and Zimmermann *et al.*[45,46] NSP3 and NSP4 alter the morphology of the rough ER leading to the formation of double membrane vesicles (DMV) (cartoon, created with BioRender.com.).[45,54,56] NSP3 and NSP4 are transmembrane proteins, embedded in the DMV bilayer and responsible for the formation of DMV pores.[46] NSP3 forms the upper base and prongs of the pore, whereas NSP4 forms the lower base.[46] The transmembrane domains are incorporated in the DMV outer and inner membrane.[46] Zimmermann *et al.* predicted the domain organization of cytoplasmatic NSP3, with Ubl1 forming the tip of the crown, followed by the three macrodomains, DPUP, PLpro, NAB and the betaSM domain.[45] Cryo electron tomography images revealed a crown like structure of the pore, with a hexameric organization.[45,46]

Their analysis revealed the presence of 12 copies each of NSP3 and NSP4,[46] differing from earlier reports.[56] Both proteins are arranged into four hexameric rings.[46] Given that NSP3 and NSP4 are transmembrane proteins, they are embedded within the DMV bilayer, contributing to both membrane remodeling and pore formation.[46] Moreover, the DMV nuclear pore complex exhibits a positively charged ring within the channel that likely facilitates RNA translocation through the pore.[46] Although several studies have aimed to characterize DMVs, their pores, and the translocation of viral RNA, the precise mechanisms underlying the

formation of these replication organelles and RNA transport remain incompletely understood and require further investigation.

VII. SNARE-Dependent Virion Release after SARS-CoV-2 Infection

The formation of DMVs by SARS-CoV-2 is a hallmark of viral replication, as these structures create a shielded microenvironment that supports efficient viral RNA synthesis.[45,47] Following replication and assembly, newly formed virions must be transported through the host cell and released. This step involves the host's vesicular trafficking machinery, culminating in virion exocytosis (Figure 6). Vesicle formation and trafficking are essential cellular processes for the transport of macromolecules between membrane-enclosed compartments, including export (exocytosis) and import (endocytosis).[68]

Membrane fusion events, which are critical for vesicle trafficking and exocytosis, are mediated by SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) proteins.[68,69] The family of SNARE proteins mediate vesicle fusion with target membranes, for example during lysosomal exocytosis or neurotransmitter release in neuronal cells (Figure 6).[68,69] In the context of SARS-CoV-2, virion release occurs via the lysosomal exocytosis pathway, as demonstrated in beta-CoVs such as MHV and SARS-CoV-2.[70] This process is facilitated by ORF3a, a SARS-CoV-2 transmembrane protein embedded in lysosomal membranes, which acts as a channel and further promotes the interaction between the BORC-ARL8b complex and the SNARE proteins Synaptosomal Associated Protein of 23 kilodaltons (SNAP23), Syntaxin 4 (SYNTX4) and VAMP7.[71,72] The BORC-ARL8b complex guides lysosomes toward the plasma membrane along the cytoskeleton, where SNARE-mediated fusion allows for the release of virions into the extracellular space (Figure 6).[70,71,73] Membrane fusion in this context is dependent on intracellular Ca^{2+} levels mediated by TRPML3, and is orchestrated by the SNARE machinery interacting with ORF3a.[74,75] In parallel, ORF3a also disrupts the function of the HOPS complex, a critical component required for SNARE complex assembly during autolysosome formation.[76,77] Thereby, ORF3a inhibits the fusion of lysosomes with autophagosomes, disrupting autophagic flux and blocking autophagosome maturation in the host cell.[77] Recent findings have identified ORF3a as a water-permeable ion channel.[78] Its activity contributes to lysosomal volume expansion and promotes lysosomal deacidification and inactivation.[78] These changes are thought to facilitate viral egress.[78] Notably, it remains a matter of scientific debate whether ORF3a primarily functions as a water-permeable ion

channel, as a transmembrane protein involved in SNARE protein interactions, or potentially fulfills both roles depending on the cellular context. However, the precise role of ORF3a in lysosomal trafficking is still not fully understood. Further investigation is needed to clarify its specific functions and mechanisms in this context.

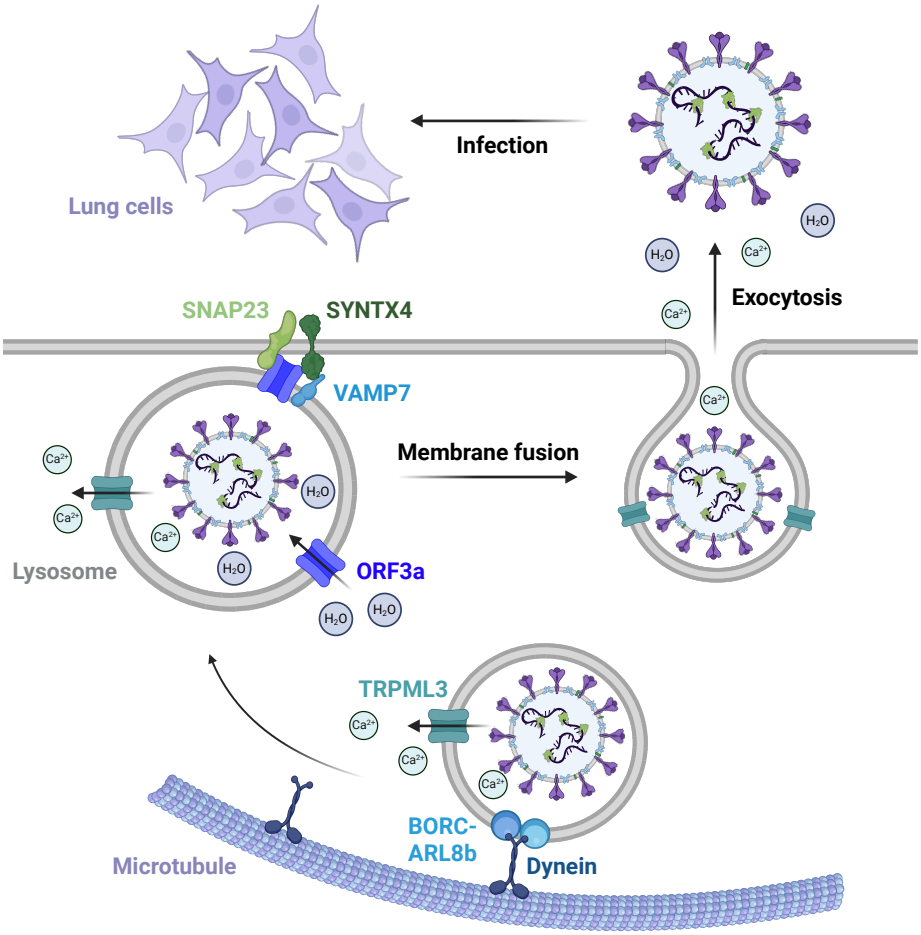


Figure 6: Schematic illustration of an ORF3a-mediated modulation of virion release in SARS-CoV-2 based on Chen *et al.*[71] and Michelucci *et al.*[78] Following virion assembly, SARS-CoV-2 particles are transported along microtubules through the host cell's secretory pathway toward lysosomal compartments. This trafficking is mediated by the interaction between the motor protein dynein and the BORC-ARL8b complex, connecting the cellular cytoskeleton with lysosomal vesicles. The viral protein ORF3a facilitates water influx into the lysosome, leading to an increase in lysosomal volume. Concurrently, calcium ions are released into the cytoplasm via the TRPML3 ion channel affecting the SNARE complex. In parallel, ORF3a interacts with components of the host SNARE complex, including VAMP7, SNAP23, and Syntaxin 4 (SYNTX4), promoting membrane fusion between lysosomes and the plasma membrane. Viral particles are subsequently released through lysosomal exocytosis, enabling infection of neighboring lung epithelial cells and propagation of the viral replication cycle.

In chapter 7 of this thesis, the expression and purification of two SNARE proteins is detailed described, which was a side-project of this PhD work. Instead of the SNARE proteins involved in virion exocytosis, the neuronal SNARE proteins were studied, which are the laboratory's main biological focus.

1. Chapter: Expression and Isolation of SARS-CoV-2 NSP3-Ubl1

1.1 Abstract

SARS-CoV-2 relies on the formation of DMVs to create a protected environment for viral RNA replication and transcription. While the structural components of these replication organelles have been partially elucidated, the molecular mechanisms underlying DMV biogenesis, DMV pore formation and RNA translocation remain incompletely understood. NSP3 is a central structural component of the DMV pore, with its N-terminal Ubl1 domain positioned at the cytosolic prongs of the pore complex. To facilitate structural and functional studies of Ubl1, we established an optimized purification protocol for recombinant 6xHis-tagged Ubl1. Small-scale expression screening across various *E. coli* strains and induction conditions identified BL21(DE3) at 25 °C as optimal for soluble Ubl1 production. For His-tag removal, an in-house-made TEV-protease was generated. Ubl1 and TEV protease were purified using different chromatography tools, including immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC). Final purification steps yielded monomeric Ubl1 of high purity, confirmed by elution behavior consistent with its expected molecular weight of around 12.5 kDa. This workflow enables the production of pure and monomeric Ubl1 samples suitable for downstream applications such as nuclear magnetic resonance (NMR) spectroscopy and further structural analysis. Future improvements may focus on enhancing TEV cleavage efficiency to increase the ratio of digested protein. The methodology provides a solid platform for in-depth structural and functional analysis of Ubl1 in the context of SARS-CoV-2 replication organelles.

1.2 Introduction

1.2.1 Structure of SARS-CoV-2 NSP3 and its Subdomain Ubl1

SARS-CoV-2 is a pathogenic beta-CoV responsible for the global outbreak of COVID-19.[22] As a positive-sense ssRNA virus, SARS-CoV-2 exploits the host cell's ribosomal machinery to translate its genomic RNA into large polyproteins.[27] Following translation, polyproteins pp1a and pp1ab undergo proteolytic processing by two viral proteases, yielding 15 NSPs, among which NSP3 is the largest.[26,27] Comprising over 1900 amino acid residues, NSP3 is a multifunctional protein made up of 16 distinct domains that contribute to viral replication and host modulation.[28] The N-terminal region of NSP3, referred to as NSP3a, consists of an ubiquitin-like domain (Ubl1; residues 1-111) and a highly flexible, glutamate-rich, intrinsically disordered hypervariable region (HVR) (Figure 7A).[67,79,80] As demonstrated by X-ray crystallography, Ubl1 is a globular and folded protein, comprising four α-helices and 3 β-sheets (PDB: 7KAG, Figure 7B).[80]

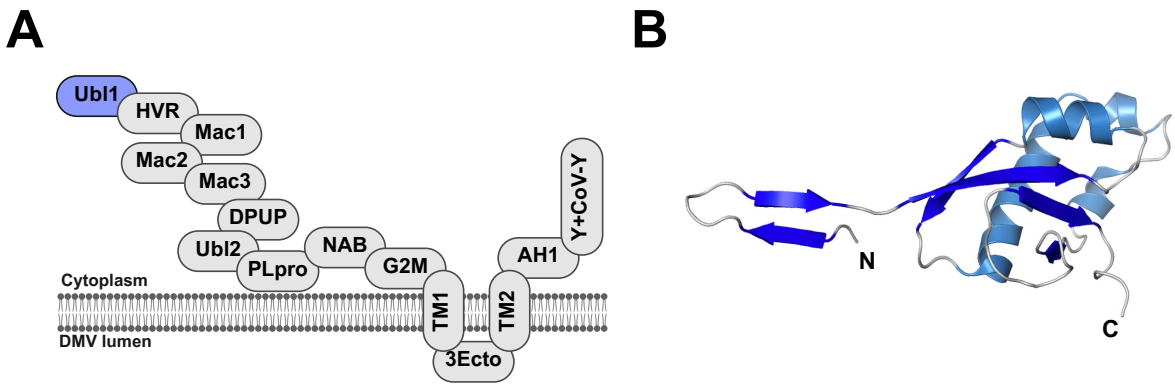


Figure 7: Structure of NSP3 with focus on Ubl1. (A) NSP3 embedded in the DMV bilayer. Schematic domain organization of NSP3, described as follows: The ubiquitin-like domain 1 (Ubl1), the hypervariable region (HVR), the three macrodomains (Mac1-3), the domain preceding the Ubiquitin-like domain 2 (DPUP), the ubiquitin-like domain 2 (Ubl2), the papain-like protease (PLpro), the nucleic-acid binding domain (NAB), the G2M domain, the two transmembrane domains (TM1 and TM2) separated by the ectodomain (3Ecto), and the C-terminal domains AH1 and Y+CoV-Y. Figure was created in BioRender.com. **(B)** Crystal structure of Ubl1 (PDB 7KAG).

1.2.2 The Origin of Ubiquitin-like Proteins

Ubiquitin (Ub) is a small, globular protein that is highly conserved across eukaryotes.[81] Its core structure adopts a β-grasp fold, which is also found in ubiquitin-like (Ubl) domains of SARS-CoV-1, SARS-CoV-2 and MERS-CoV such as Ubl1 and Ubl2 - hence the origin of their name.[79,82,83,84] While the amino acid sequences of Ubl domains can vary substantially, the

overall fold is conserved, even with minimal sequence similarity (Figure 8).[79] Notably, the proportion of α-helical content differs between human ubiquitin (HuUb), Ubl1, and Ubl2, reflecting structural adaptations within the conserved fold.[79] The proteins also vary in molecular weight: HuUb (8.8 kDa), Ubl1 (12.5 kDa), and Ubl2 (7.2 kDa). NSP3-Ubl1 has a maximum length of 64.9 Å (6.49 nm) and a width of 35.1 Å (3.51 nm) (Figure 8B).

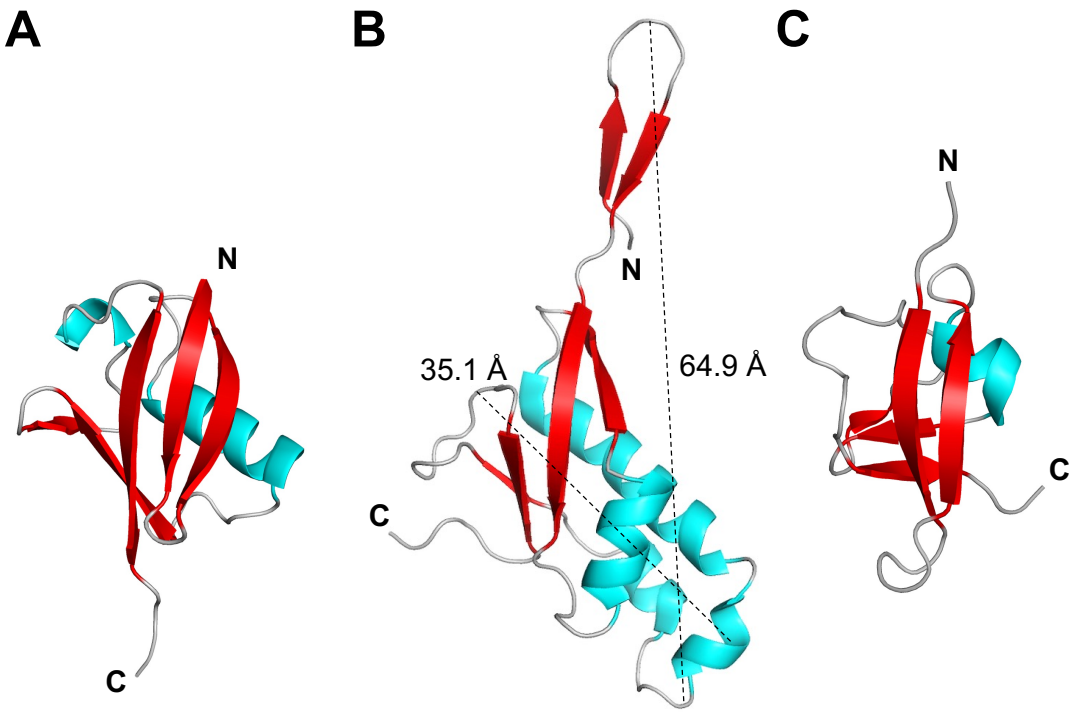


Figure 8: The ubiquitin-like fold. Crystal structures of **(A)** HuUb (PDB: 1UBQ), **(B)** SARS-CoV-2 NSP3 Ubl1 (PDB: 7KAG), and **(C)** SARS-CoV-2 NSP3 Ubl2 (PDB: 7CJD) are shown. α-helices are depicted in cyan and β-strands in red. The molecular size was calculated using PyMOL.[85]

Typically, ubiquitin is covalently attached to substrate proteins, marking them for degradation by the proteasome.[86] In addition to this canonical function, both ubiquitinated proteins and Ubl domains are key mediators of protein-protein interactions and play essential roles in the regulation of numerous intracellular signaling pathways, such as endocytosis and DNA repair.[86] In SARS-CoV, Ubl1 is implicated in RNA and nucleocapsid (N) protein binding, while Ubl2 is known to interfere with the host immune response. [67,87,88] However, the precise functional roles of Ubl1 and Ubl2 in the SARS-CoV-2 replication cycle remain poorly understood and require further investigation.

1.3 Aims

The objective of this study was to establish and optimize an efficient expression and purification protocol for recombinant, monomeric, and structurally homogeneous Ubl1 - the N-terminal ubiquitin-like domain of SARS-CoV-2 NSP3. Ubl1 is a key structural element of the DMV pore complex and is thought to mediate interactions with both viral and host proteins. To enable the production of high-quality Ubl1 suitable for downstream structural and functional studies, particularly NMR spectroscopy, we systematically evaluated various *E. coli* expression strains and induction conditions, followed by a multi-step purification strategy including IMAC, TEV protease cleavage, and SEC. The TEV protease required for site-specific cleavage of Ubl1 was produced in-house, and a dedicated purification protocol was likewise established to obtain catalytically active enzyme preparations. Together, these optimized workflows provide a robust foundation for future investigations into the role of Ubl1 in replication organelle formation and viral RNA translocation during SARS-CoV-2 infection.

1.4 Material and Methods

1.4.1 Buffer, Chemicals and Media

All standard reagents used for the respective analyses were obtained from AppliChem, Merck, Roche, and Sigma. Cell culture media were sourced from AppliChem. Details of the media, reagents, and buffers employed in the experimental procedures are provided in the corresponding sections. To prevent contamination during cellular experiments, all reagents were autoclaved and stored in sterile containers. All chemicals were of analytical grade and used without further purification. Buffers used for protein isolation were pre-filtered and degassed prior to use. Buffers and media were prepared using ultrapure water (Milli-Q grade). Buffers were filtered through 0.22 µm membranes prior to use.

1.4.2 Instruments

Instruments and materials, along with their respective manufacturers, used for the expression and purification of ^{1}H-^{15}N-labeled Ubl1 and the TEV protease are listed in Tables 2 and 3.

Table 2: List of instruments used for the expression and isolation of ^{1}H-^{15}N Ubl1 and TEV-protease.

Instrument	Manufacturer
Agarose-Gel chamber system	Bio-Rad Laboratories GmbH, München
ÄKTA Pure	GE Healthcare, Sweden
Autoclave VX-150	Systec GmbH, Linden
Bruker 700 MHz Avance III NMR spectrometer	Bruker, Billerica, MA, USA
Centrifuge 5430 R	Eppendorf GmbH, Hamburg
Centrifuge Avanti J-26S XP	Beckmann Coulter GmbH, Krefeld
Desk centrifuge 5415 R	Eppendorf GmbH, Hamburg
Dri-Block Heater 08-3	Techne, UK
Drying chamber T 5050	Heraeus, Hanau
Electrophoresis power supply 3501 XL	Amersham Pharmacia biotech, UK
Gel documentation system	Bio-Rad Laboratories GmbH, München
Incubation shaker Infors HT Multitron	Infors AG, Bottmingen
Incubation shaker New Brunswick innova 40	Eppendorf GmbH, Hamburg
Incubation shaker New Brunswick innova 44	Eppendorf GmbH, Hamburg
JASCO J-815 CD spectropolarimeter	Jasco, Gross-Umstadt
Magnetic stirrer C-MAG MS 7	IKA GmbH & CO. KG, Staufen
Milli-Q Direct 16	Merck KGaA, Darmstadt
Mini Trans-Blot module blotting chamber	Bio-Rad Laboratories GmbH, München
NanoDrop 2000	Thermo Fischer Scientific, Schwerte
pH meter	Satorius AG, Göttingen
Scale	Satorius AG, Göttingen
Sonication device Sonoplus	Bandelin electronic GmbH & Co KG, Berlin
Thermocycler compact	Eppendorf, Hamburg
T100 Thermal Cycler	Bio-Rad Laboratories GmbH, München
Vortexer	Heidolph Instruments, Schwabach
VS70T sonotrode	Bandelin electronic GmbH & Co KG, Berlin

Table 3: List of consumables for the expression and isolation of ^{1}H-^{15}N Ubl1 and TEV-protease.

Material	Manufacturer
Amicon MWCO 3 kDa	Merck Millipore, Darmstadt, Germany
Amicon MWCO 10 kDa	Merck Millipore, Darmstadt, Germany

cOmplete EDTA-free protease inhibitor tablet	Roche, Basel
Der Blaue Jonas - Single-step Coomassie Blue protein gel dye	GRP German Research, Haag a. d. Amper
Deuteriumoxid	Carl Roth, Karlsruhe
DNAse I, recombinant	Roche, Basel
E. coli BL21 (DE3) cells	New England Biolabs, Ipswich, USA
Goat anti-Mouse lgG (H+L) Secondary Antibody HRP conjugated	Thermo Fischer Scientific, Schwerte
HiLoad 16/60 Superdex 75 pg	Cytiva, Marlborough
HiLoad 16/600 Superdex 200 pg	Cytiva, Marlborough
Falcon tubes (15 mL and 50 mL)	Sarstedt, Nürnbrecht
MEM Vitamin Solution (100x)	Sigma-Aldrich, St. Louis, USA
Norell Select Series 3 mm NMR Tubes	Sigma-Aldrich, St. Louis, USA
PageRuler Prestained Protein Ladder	Thermo Fisher Scientific, Schwerte
Protino Ni-NTA 5 mL FPLC column	Macherey-Nagel, Düren
Plastic cuvettes	Brand GmbH & Co KG, Wertheim
Petri dishes	Hartenstein, Würzburg
Pipette tips	Sarsted, Nürnbrecht
PVDF-membrane 0.2 µm	GE Healthcare, Sweden
Quartz cuvettes – 1 mm path length	Hellma, Müllheim
Reaction tube (1.5 mL and 2 mL)	Sarstedt, Nürnbrecht
Regenerated Cellulose Membrane Filter 0.2 µm	Sartorius, Göttingen
Serological pipettes	Sarstedt, Nürnbrecht
SnakeSkin dialysis tubes (3 K MWCO)	Thermo Fischer Scientific, Schwerte
SnakeSkin dialysis tubes (10 K MWCO)	Thermo Fischer Scientific, Schwerte
Spectra Multicolor Low Range Protein Ladder	Thermo Fischer Scientific, Schwerte
Sterile Syringes for Single Use (3 mL, 20 mL) Fisherbrand	Thermo Fischer Scientific, Schwerte
Syringe filter, Filtropur S, PES, pore size 0.2 µm for sterile filtration	SARSted, Nürnbrecht
Sterican Standard canulla 0.45 x 25 mm	B. Braun, Melsungen
SuperSignal West Pico PLUS Chemiluminescent Substrate	Thermo Fischer Scientific, Schwerte

Whatman Cellulose Blotting Paper	Cytiva, Marlborough
2xYT medium	Thermo Fischer Scientific, Schwerte
6x-His Tag Monoclonal Antibody (HIS.H8)	Thermo Fischer Scientific, Schwerte
^{15}N-NH_4-Cl	99%, Cambridge Isotope Laboratories, USA

1.4.3 Buffers and Solutions

In Table 4, the composition of buffers used in this chapter is listed. Buffers and solutions were prepared using distilled Milli-Q water (18.2 MOhm). Buffers for isolation were degassed and filtered.

Table 4: Chemical solutions used for the expression and isolation of Ubl1 and the TEV-protease.

Solution	Ingredients and Concentrations
Acrylamide:Bisacrylamide (34.3%:1.1%)	34.3% (v/v) Acrylamide, 1.1% (v/v) Bisacrylamide
Acrylamide:Bisacrylamide (37.5%:1%)	37.5% (v/v) Acrylamide, 1% (v/v) Bisacrylamide,
Anode buffer (10x)	2 M Tris-HCl (pH 8.9)
Antibiotica stock	50 mg/mL Kanamycin
Blocking solution	5% milk (w/v) in (1x) TBS-T
Cathode buffer	1 M Tris-HCl (pH 8.25), 1 M Tricine, 1% (w/v) SDS
Column storage buffer	20% (v/v) EtOH
DNase I stock solution	20 mM Tris-HCl, 50 mM NaCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 50% (v/v) glycerol
Dialysis buffer Ubl1	50 mM Tris (pH 8), 250 mM NaCl, 2 mM DTT, 1 mM EDTA
Dialysis buffer TEV	50 mM Tris (pH 8), 25 mM NaCl, 50% (w/v) glycerine, 1 mM EDTA
Glycerol stock solution	31 g 2xYT powder per L, 50% (v/v) glycerol
Induction reagent	0.5 M IPTG
IMAC elution buffer Ubl1	50 mM Tris (pH 8), 250 mM NaCl, 500 mM imidazole

IMAC elution buffer TEV	25 mM Tris (pH 8), 100 mM NaCl, 500 mM imidazole, 10% (w/v) glycerine
IMAC lysis buffer TEV	25 mM Tris (pH 8), 500 mM NaCl, 10 mM imidazole, 10% (w/v) glycerine
Laemmli sample buffer (4x)	10% (w/v) SDS, 10 mM *β*-mercaptoethanol, 20% (v/v) glycerol, 0.05% (w/v) Bromphenol- Blue, 200 mM Tris- HCl (pH 6.8)
Lysis buffer Ubl1	50 mM Tris (pH 8), 250 mM NaCl
Lysis buffer TEV	25 mM Tris (pH 8), 500 mM NaCl, 0.4 mM PMSF
TBS-T buffer (20x)	0.5 M Tris (pH 7.6), 3M NaCl, 2% (v/v) Tween20
TEV-Pellet Washing Buffer	60 mM sodium phosphate (pH 7.4), 200 mM NaCl
TGS buffer (10x)	25 mM Tris-HCl (pH 8.3), 192 mM glycine, 0.1% (w/v) SDS
TG buffer (10x)	25 mM Tris-HCl (pH 8.3), 192 mM glycine
Towbin buffer (1x)	10% (v/v) TG buffer (10x), 20% (v/v) EtOH
Tris-Glycine Gel Buffer (Stacking Gel)	0.5 M Tris (pH 6.8)
Tris-Glycine Gel Buffer (Separating Gel)	1.5 M Tris (pH 8.8)
Tris-Tricine Gel Buffer (3x)	3 M Tris-HCl (pH 8.45), 0.3% (w/v) SDS
SEC buffer Ubl1	50 mM sodium phosphate (pH 6.5), 150 mM NaCl

1.4.4 Plasmids

The plasmid encoding the Ubl1 protein was rationally designed and subsequently synthesized and ordered through BioCat GmbH (Heidelberg, Germany). The construct encoding the TEV protease, along with its corresponding expression plasmid, was kindly provided by Dr. Bernd König from the Institute of Biological Information Processing 7 (IBI-7) at the Forschungszentrum Jülich. Detailed information regarding the specific plasmid backbones, inserted gene sequences, and target constructs used in this study is summarized in Table 5.

Table 5: Construct details of Ubl1 and TEV-protease.

Construct Information	**Ubl1**	**TEV-protease**
Original organism	*SARS-CoV-2*	*Tabacco etch virus (TEV)*
Plasmid	pET-28a(+)	pRK793
Codon-optimized	*E. coli*	*E. coli*
Cell strains	*E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) – short, *E. coli* BL21 (DE3)	*E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) – short, *E. coli* BL21 (DE3)
N-terminal Tag	6xHis	7xHis
Cleavage Sites	TEV & Thrombin	-
Antibiotic Resistance	Kanamycin	Ampicillin, Chloramphenicol
Restriction sites	NdeI, BamHI	-
Construct size with tag	15.55 kDa, 138 amino acids	28.69 kDa, 249 amino acids
Construct size	12.46 kDa, 111 amino acids	-
pI of His-tagged construct	4.70	9.58
pI of construct	4.08	-
Extinction coefficient (non-reduced conditions)	His-Ubl1: 16055 M^{-1} cm^{-1} Ubl1: 14565 M^{-1} cm^{-1}	His-TEV: 32220 M^{-1} cm^{-1}
Extinction coefficient (reduced conditions)	His-Ubl1: 15930 M^{-1} cm^{-1} Ubl1: 14440 M^{-1} cm^{-1}	His-TEV: 31970 M^{-1} cm^{-1}

In the following the amino acid sequences of tagged and digested constructs are listed. The His-tag is colored in green and the construct itself in red. The TEV cleavage side is displayed in cyan and the Thrombin cleavage side in purple. Stop codons are marked with "*".

6xHis-Ubl1:
MGSSHHHHHHSSGLVPRGSHMENLYFQAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYT
VELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEFKLASHMYCSFYPPDE*

Ubl1:
APTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYTVELGTEVNEFACVVADAVIKTLQPVSELL
TPLGIDLDEWSMATYYLFDESGEFKLASHMYCSFYPPDE*

6xHis-TEV:
MHHHHHHGESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNKHLFRRNNGTLLVQSLHGV

FKVKNTTTLQQHLIDGRDMIIIRMPKDFPPFPQKLKFREPQREERICLVTTNFQTKSMSSMVSDTSCTFPS
SDGIFWKHWIQTKDGQCGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSG
WRLNADSVLWGGHKVFMVKPEEPFQPVKEATQLMNRRRRR*

1.4.5 Sequencing of Ordered Gene Constructs

For sequencing, 12 µL of *E. coli* codon-optimized plasmids, at concentrations ranging from 50 to 100 ng/µL, were submitted to Microsynth Seqlab (Göttingen, Germany). Sequencing was performed using standard primers provided by Microsynth Seqlab. The specific primers used are listed in Table 6.

Table 6: Primer list used for sequencing of Ubl1 plasmid.

Construct	Plasmid concentration	Primer name	Primer sequence
pET-28a(+)-6xHis-Thrombin-TEV-Ubl1	109.5 ng/µL	T7	TAATACGACTCACTATAGGG

The online tool NucleoBlast[89] was used for the alignment of the resulting and theoretical sequences. Sequencing results of the constructs are listed in the supplementary data (Table SD1).

1.4.6 Transformation of Plasmids in Chemically Competent *E. coli* Cells

The various expression constructs were transformed into different *E.coli* strains, as detailed in Table 7. For protein production, *E. coli* BL21(DE3) cells were employed as the primary host for the expression of both Ubl1 and TEV protease constructs. To optimize Ubl1 expression, several *E.coli* strains were systematically evaluated. Among them, *E. coli* NiCo21 (DE3) - a derivative of BL21(DE3) - was selected due to its advantageous properties for downstream purification using immobilized metal affinity chromatography (IMAC).[90] This strain minimizes background binding during purification by reducing the expression of endogenous *E. coli* proteins that contain histidine-rich regions or metal-binding motifs.[90] Furthermore, *E. coli* Rosetta (DE3) cells were tested to improve the expression efficiency of Ubl1. The Rosetta strain supplies tRNAs for rare codons that are infrequently used in *E. coli*, thereby facilitating the expression of heterologous genes containing such codons and enhancing overall protein yield.[91]

Table 7: *E.coli* cell strains.

Construct	Cell strain	Manufacturer
pET-28a(+)-6xHis-Thrombin-TEV-Ubl1	*E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]*K-12(λS) – short, E. coli BL21 (DE3)*	New England Biolabs, Ipswich (USA)
pET-28a(+)-6xHis-Thrombin-TEV-Ubl1	*E. coli* str. *can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 – short, E.coli NiCo21(DE3)*	New England Biolabs, Ipswich (USA)
pET-28a(+)-6xHis-Thrombin-TEV-Ubl1	*E. coli str. F⁻ ompT hsdS$_B$(r_B^- m_B^-) gal dcm (DE3) pRARE (Cam^R) – short, E.coli Rosetta*	Merck Millipore, Massachusetts (USA)
pRK793-7xHis-TEV	*E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) – short, *E. coli* BL21 (DE3)	New England Biolabs, Ipswich (USA)

Initially, 50 µL of competent cells were thawed on ice for 20 minutes. Subsequently, 10 ng of plasmid DNA was added to the cell suspension and incubated on ice for an additional 30 minutes. Cells were then heat-shocked at 42 °C for 45 s, followed by immediate cooling on ice for 2 min. Afterward, 450 µL of pre-warmed 2xYT medium was added to the mixture, and cells were incubated at 37 °C with shaking at 800 rpm for 1 h. The transformation mixture was plated onto LB-agar plates containing the appropriate antibiotics and incubated overnight at 37 °C. On the following day, individual colonies were selected and cultured in 5 mL of 2xYT medium supplemented with the corresponding antibiotics at 37 °C and 160 rpm overnight. Cultures were centrifuged at 16,000 × g for 1 min, the supernatant was discarded, and the cell pellet was resuspended in glycerol stock solution (50% (v/v) 2xYT medium, 50% (v/v) glycerol). Glycerol stocks were flash-frozen in liquid nitrogen and stored at -80 °C for future use. These

stocks served as the inoculum for expression testing and large-scale protein expression in minimal medium.

1.4.7 Expression Tests of Ubl1

Expression tests were conducted to determine the optimal conditions for expression of the Ubl1 domain. Specifically, different post-induction temperatures and *E. coli* strains were evaluated to enhance recombinant protein yield. All expression tests were performed in 2xYT medium. To begin, 5 mL precultures of each construct were inoculated using the respective glycerol stocks and supplemented with 50 µg/mL kanamycin for selection. Precultures were incubated overnight at 37 °C and 160 rpm (16 h). The following day, 50 mL of fresh minimal medium was inoculated with 0.5 mL of the preculture and 50 µL of the appropriate antibiotic in a 250 mL baffled Erlenmeyer flask. Cultures were grown at 37 °C and 160 rpm until an optical density at 600 nm (OD_{600}) of 0.8 was reached. At this point, a 1 mL sample was collected to assess baseline (pre-induction) expression. Cells were pelleted by centrifugation at 16,000 × g for 1 min, the supernatant was discarded, and the pellet was resuspended in 45 µL of nuclease-free water and 15 µL of 6x Laemmli SDS loading buffer.[92] Protein expression was induced by the addition of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), which activates the T7 promoter system. Cultures were incubated overnight at the designated temperature. After induction, a second 1 mL culture sample was collected and processed as described above. Protein expression levels were analyzed by SDS-PAGE and Western blotting.

1.4.8 Large-Scale Expression of ^{1}H-^{15}N Ubl1 and ^{15}N-^{13}C Ubl1

For large-scale protein expression, a 50 mL preculture was prepared in 2xYT medium supplemented with 50 µg/mL kanamycin, using 250 mL baffled Erlenmeyer flasks. The culture was incubated overnight at 37 °C and 160 rpm. For the expression of uniformly ^{1}H-^{15}N Ubl1, two heat-sterilized 2 L Erlenmeyer flasks were each filled with 500 mL of minimal medium to ensure adequate aeration during cultivation (prepared as described in Table 8). Each flask was inoculated with 10 mL of the overnight preculture and supplemented with 50 µg/mL kanamycin. For the production of ^{15}N-^{13}C-labeled Ubl1, the same procedure was followed, with the exception that ^{13}C D-glucose was used in place of ^{12}C D-glucose as the single carbon source in the minimal medium.

Table 8: Protocol for 1 L minimal medium.

Chemical	Stock solution	Final concentration	Amount for 1 L main culture	Amount for 50 mL preculture
M9 salt solution	337 mM Na_2HPO_4, 220 mM KH_2PO_4, 85.5 mM NaCl (10x)	33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.55 mM NaCl (1x)	100 mL	5 mL
MEM vitamin solution	100x	1x	1 mL	50 µL
$MgSO_4$	1 M	1 mM	1 mL	50 µL
$CaCl_2$	0.1 M	0.3 mM	3 mL	150 µL
Biotine	1 mg/mL (w/v)	1 mg/L	1 mL	50 µL
Thiamine	1 mg/mL (w/v)	1 mg/L	1 mL	50 µL
Trace element solution	13.4 mM EDTA, 3.1 mM $FeCl_3$-$6H_2O$ 0.62 mM $ZnCl_2$ 76 µM $CuCl_2$-$2H_2O$ 42 µM $CoCl_2$-$2H_2O$ 162 µM H_3BO_3 8.1 µM $MnCl_2$-$4H_2O$ (100x)	134 µM EDTA, 31 µM $FeCl_3$-$6H_2O$ 6.2 µM $ZnCl_2$ 0.76 µM $CuCl_2$-$2H_2O$ 0.42 µM $CoCl_2$-$2H_2O$ 1.62 µM H_3BO_3 0.081 µM $MnCl_2$-$4H_2O$ (1x)	10 mL	500 µL
^{15}N NH_4Cl	99% (w/w)	0.1% (w/v)	1 g	50 mg
^{12}C D-Glucose-Monohydrate	20% (w/v)	0.4% (w/v)	20 mL	200 µL
^{13}C D-Glucose	99% (w/w)	0.4% (w/v)	4 g	200 mg

At OD_{600} of 0.8, 0.5 mM IPTG was added inducing the expression of target proteins. Samples before and after induction of expression were used to analyze the expression grade of both proteins. After another incubation period overnight at 25 °C and 120 rpm, cells were harvested at 4 °C via centrifugation at 6,000 x *g* for 15 min. After centrifugation, the medium was discarded. The pellet was weighted and stored at -20 °C for further use.

1.4.9 Large-Scale Expression of 7xHis-TEV Protease

For large-scale expression, a 50 mL preculture was prepared in 2xYT medium supplemented with 25 µg/mL chloramphenicol and 75 µg/mL ampicillin in 250 mL baffled Erlenmeyer flasks. The culture was incubated overnight at 37 °C and 160 rpm. For the main culture, five heat-sterilized 2 L Erlenmeyer flasks were each filled with 1 L of 2xYT medium, supplemented with the appropriate antibiotics, and inoculated with 5 mL of the overnight preculture. Cultures were incubated at 37 °C and 120 rpm until an optical density at 600 nm (OD_{600}) of 0.5-1.0 was reached. Protein expression was induced by the addition of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), followed by incubation at 25 °C and 120 rpm for 16 h. Cells were harvested the next day by centrifugation at 6,000 × g for 10 min at 4 °C. The resulting cell pellet was washed with TEV Pellet Washing Buffer (60 mM sodium phosphate (pH 7.4), 200 mM NaCl), followed by a second centrifugation step at 7,000 × g for 10 min. The washed pellet was weighed and stored at -20 °C for further use.

1.4.10 Disruption of *E.coli* Cells Harboring the Ubl1 Plasmid

The frozen cell pellet was thawed on ice and resuspended in Ubl1 lysis buffer at a ratio of 5 mL buffer per gram of cell pellet. For DNA degradation, DNase I was added at a concentration of 10 µg per gram of cell pellet, along with one EDTA-free protease inhibitor tablet per sample to minimize proteolytic activity. The suspension was incubated for 20 min at 4 °C on a roller mixer until homogenized. Cell lysis was performed by ultrasonication using a VS70T sonotrode for 20 min. The amplitude was set to 60%, with a pulse cycle of 1 s sonication followed by 3 s of rest. To prevent overheating during sonication, samples were kept in an ice-water bath throughout the process. Following lysis, the cell lysate was centrifuged at 40,000 × g for 45 min at 4 °C using a JA-20 rotor. The resulting supernatant and pellet fractions were separated, and samples were analyzed by SDS-PAGE to assess solubility and expression. The clarified supernatant was subjected to protein purification via IMAC using an ÄKTA start system.

1.4.11 Lysis of *E.coli* Cells Expressing the TEV-Protease

The frozen cell pellet was thawed on ice and resuspended in TEV lysis buffer at a ratio of 5 mL buffer per gram of cell pellet. For genomic DNA degradation, DNase I was added at a

concentration of 10 µg per gram of cell pellet, along with one EDTA-free protease inhibitor tablet to minimize proteolytic degradation. The cell suspension was incubated on a roller mixer at 4 °C for 20 min to ensure homogenization. Cell disruption was performed by ultrasonication using a VS70T sonotrode for 15 min at 55% amplitude, with 3.5-second sonication pulses followed by 11-second pauses. To prevent heat accumulation, samples were maintained in an ice-water bath throughout the sonication process. Following lysis, the cell suspension was centrifuged at 25,000 × g for 10 min at 4 °C using a JA-20 rotor. The supernatant and pellet were separated, and aliquots were analyzed by SDS-PAGE to evaluate solubility and expression levels. The supernatant was subsequently treated with 10.3% (v/v) Polymin-P in 25 mM Tris buffer (pH 8), added at a volume of 500 µL per 50 mL of supernatant. The mixture was incubated on ice for 15 min to precipitate nucleic acids and other contaminants, followed by a second centrifugation step at 48,400 x g for 75 min at 4 °C. A sample of the resulting clarified supernatant was taken for SDS-PAGE analysis. The remaining supernatant was subjected to protein purification by IMAC using an ÄKTA start system.

1.4.12 Sodium Dodecyl-Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Ubl1

SDS-PAGE was employed to assess and quantify protein expression and purification, enabling separation of proteins based on their molecular weight.[92,93] Due to the molecular weights of Ubl1 and TEV protease, both Tris-Tricine and Tris-Glycine SDS-PAGE systems were evaluated (see Tables 9 and 10 for gel compositions). For all analyses, samples were mixed with 15 µL of 4x Laemmli SDS loading buffer and denatured by boiling at 95 °C for 10 min. After denaturation, samples were centrifuged at 10,000 x g for 10 min to pellet any insoluble material. The resulting supernatants, along with a molecular weight protein marker, were loaded onto the appropriate SDS-PAGE gel and connected to an electrophoresis power supply unit (3501 XL). Electrophoresis was stopped once the tracking dye approached the bottom of the gel. Following electrophoresis, gels were washed four times with deionized water and subsequently stained with Coomassie Brilliant Blue (Blauer Jonas) for 1 h. No destaining step was required. Protein bands were visualized using a gel documentation system.

1.4.12.1 Tris-Tricine SDS-PAGE of Ubl1

A stacking and running SDS-PAGE gel was prepared according to the protocol outlined in Table 9.

Table 9: Protocol for 2x Tris-Tricine SDS-PAGES.

Component	20% Separating Gel	5.6% Stacking Gel
ddH_2O	-	3.2 mL
Gel buffer (3x)	5 mL	2.1 mL
Glycerol	1.6 mL	-
Acrylamide:Bisacrylamide (34.3%:1.1%)	8.5 mL	1 mL
APS (10% (w/v))	50 µL	50 µL
TEMED	25 µL	25 µL

For each gel, 3 µL of protein marker (Spectra Multicolor Low Range Ladder) and 10 µL of each sample were loaded. Electrophoresis was conducted at a constant current of 40 mA per gel using running buffer. The gel apparatus was assembled with the gel chamber filled with 1x cathode buffer and surrounded by 1x anode buffer.

1.4.12.2 Tris-Glycine SDS-PAGE of the TEV-protease

Table 10 summarizes the composition of the Tris-Glycine gels used.

Table 10: Protocol for 2x Tris-Glycine SDS-PAGES.

Component	15% Separating Gel	5% Stacking Gel
ddH_2O	-	3.2 mL
Tris-Glycine Gel Buffer (Separating Gel)	5 mL	2.1 mL
Tris-Glycine Gel Buffer (Stacking Gel)	1.6 mL	-
Acrylamide:Bisacrylamide (37.5%:1%)	8.5 mL	1 mL
SDS (10% (w/v))	50 µL	50 µL
APS (10%(w/v))		
TEMED	25 µL	25 µL

For electrophoresis, 3 µL of protein marker (PageRuler Prestained) and 10 µL of each sample were loaded onto the gel. The run was performed at a constant voltage of 120 V in running buffer (1x TGS buffer).

1.4.13 Western Blot Analysis of Ubl1

Western blot analysis was employed to detect the protein of interest following separation by SDS-PAGE as described in the previous section.[94] For membrane transfer, a PVDF membrane was activated by immersion in 100% (v/v) EtOH, and six Whatman papers were soaked in 1x Towbin transfer buffer. The SDS gel was carefully placed onto the activated PVDF membrane and sandwiched between three layers of Whatman paper on each side. Protein transfer was performed in a blotting chamber at 25 V and 1 A for 30 min. To verify successful protein transfer, the membrane was briefly stained with Ponceau S solution and subsequently washed four times with deionized water. The appearance of visible bands indicated effective transfer. The membrane was then blocked with blocking solution for 1 h at room temperature to saturate nonspecific binding sites. Following blocking, the solution was discarded, and the membrane was incubated overnight at 4 °C with primary antibody, 6x His-Tag Monoclonal Antibody (HIS.H8), diluted 1:1,000 in blocking solution under gentle agitation. The primary antibody solution was recovered the next day and stored at -20 °C for future use. The membrane was washed three times for 5 min each with 1x TBS-T buffer to remove excess blocking reagent. The membrane was then incubated with horseradish peroxidase (HRP)-conjugated Goat anti-Mouse IgG (H+L) secondary antibody diluted 1:10,000 in blocking solution for 1 h at room temperature with gentle shaking. Following incubation, the secondary antibody solution was discarded, and the membrane was washed three times for 5 min with 1x TBS-T. Chemiluminescent detection was carried out using a Bio-Rad Universal Hood III imaging system. For signal development, a 1:1 mixture of Super Signal Peroxidase Solution and Enhancer Solution was applied to the membrane and incubated for 1 min in a dark room. Imaging was performed in signal accumulation mode at high resolution, acquiring 40 frames over 200 s. The protein ladder was visualized separately using the colorimetric mode. Finally, images were merged and analyzed.

1.4.14 Immobilized Metal Affinity Chromatography (IMAC) of 6xHis-Ubl1 and 7xHis-TEV

IMAC was employed to selectively purify His-tagged proteins based on their affinity for nickel ions immobilized on the column matrix.[95] Purification was carried out using a Protino Ni-NTA 5 mL FPLC column connected to an ÄKTA Pure system (GE Healthcare, Sweden). Ubl1 purification was performed at room temperature, whereas TEV protease isolation was conducted at 4 °C to preserve enzymatic activity. All buffers were filtered and degassed prior to use. Two buffer systems were utilized during purification: IMAC lysis buffer for sample loading and washing, and an elution buffer supplemented with imidazole for protein elution. The exact compositions of these buffers are detailed in Table 3. Initially, the system was flushed with degassed, filtered double-distilled water (ddH_2O) until absorbance and conductivity values stabilized at 0 mAU. The column was then equilibrated with 10 column volumes (CVs) of lysis buffer at a flow rate of 2 mL/min and 0.3 MPa pressure. Following a reduction in flow rate to 0.5 mL/min, the column was assembled, and the system was washed again with 10 CVs of IMAC lysis buffer at 2 mL/min. The protein sample was subsequently loaded onto the column at 1 mL/min, and the flow-through (FT) was collected for SDS-PAGE analysis. The column was washed with IMAC lysis buffer until the absorbance at 280 nm returned to baseline. To remove nonspecifically bound proteins, the column was washed with elution buffer at 4% (v/v) imidazole for Ubl1 and 2% (v/v) for TEV protease. Target proteins were eluted using a linear gradient of elution buffer over 40 min at 1 mL/min, increasing the imidazole concentration to 100% (v/v). Fractions containing the protein of interest were pooled. Finally, the column and system were rinsed with ddH_2O and stored in 20% (v/v) EtOH.

1.4.15 Dialysis of 7xHis-TEV Protease

To reduce the imidazole concentration in the purified protein samples, IMAC fractions were pooled and subjected to dialysis using SnakeSkin dialysis tubing with a 10 kDa molecular weight cutoff (MWCO). The tubing was immersed in 1 L of the appropriate dialysis buffer and incubated overnight at 4 °C with gentle stirring. Following dialysis, the samples were concentrated using Amicon centrifugal filters with a 10 kDa MWCO. Concentrated samples were aliquoted, flash-frozen in liquid nitrogen, and stored at -80 °C until further use.

1.4.16 Digestion and Dialysis of Ubl1

The 6xHis affinity tag of Ubl1 was removed by proteolytic cleavage with an in-house-produced TEV protease. This protease specifically cleaves the polyhistidine tag at the N-terminus of the target protein. Digestion was carried out at a protease-to-substrate ratio of 1:100 (w/w). To simultaneously reduce the imidazole concentration and facilitate cleavage, the reaction mixture was subjected to dialysis overnight. For this purpose, the protein mixture was enclosed in a SnakeSkin dialysis tubing with a 3 kDa molecular weight cutoff (MWCO) and immersed in 1 L of the appropriate dialysis buffer. The sample was incubated at 4 °C under gentle stirring for the duration of the dialysis.

1.4.17 Size Exclusion Chromatography of Ubl1

The concentrated protein was further purified by size-exclusion chromatography (SEC). Ubl1 purification was conducted using a HiLoad 16/600 Superdex 75 pg column. To preserve column integrity, all buffers were sterile-filtered and degassed prior to use. Initially, the column was equilibrated with SEC buffer at a flow rate of 0.5 mL/min for 8 h, corresponding to approximately 2 CVs. The concentrated protein sample was then injected using a 3 mL syringe. Elution was performed at 1 mL/min with a delta column pressure of 0.3 MPa and a precolumn pressure of 0.5 MPa until 1 CV was completed. Fractions corresponding to the main peak were collected and analyzed by SDS-PAGE. The column was subsequently washed with degassed and filtered double-distilled water (ddH_2O) at 0.5 mL/min for 2 CVs, followed by an additional wash step with 20% (v/v) EtOH for 1 CV at 0.3 mL/min. Pooled fractions were concentrated using Amicon centrifugal filters with a 3 kDa molecular weight cutoff at 6,000 x g until a protein concentration of approximately 200 µM was achieved. Protein concentration was precisely determined using a NanoDrop 2000 spectrophotometer via UV-Vis absorbance. Finally, samples were flash-frozen in liquid nitrogen and stored at -80 °C until further use.

1.5 Results

The following sections provide a comprehensive overview of the expression and purification protocols for Ubl1. For structural investigations, such as NMR spectroscopy, the production of highly pure and homogeneous protein is paramount. Additionally, obtaining sufficient protein yield is critical for downstream applications. Consequently, initial efforts focused on optimizing high-level expression of the target protein. Since its establishment as an efficient and widely used host system in 1996, *E.coli* has remained a preferred platform for recombinant protein production.[96] However, gene expression in *E. coli* is influenced by multiple environmental factors, including growth medium composition and induction temperature. To maximize protein yield, various experimental parameters were systematically evaluated, particularly medium formulations, cell strains and post-induction temperatures. For subsequent purification steps, removal of the His-tag from Ubl1 was necessary to obtain the native protein. Given that the Ubl1 constructs contain a TEV protease cleavage site, recombinant TEV protease was produced and purified to facilitate tag cleavage.

1.5.1 Optimization of Ubl1 Expression in *E.coli* Cells

Small-scale expression trials of 6xHis-TEV-Ubl1 were conducted under varying temperature conditions across different *Escherichia coli* strains. BL21(DE3), Rosetta, and NiCo cells harboring the expression construct were initially grown overnight in 2xYT medium at 37 °C. Subsequently, main cultures were inoculated into fresh 2xYT medium and incubated post-induction at 20 °C, 25 °C, or 37 °C overnight. Optimal expression conditions were assessed by analyzing the intensity of protein bands corresponding to Ubl1 (15.5 kDa) via SDS-PAGE and Western blotting. Samples were collected prior to induction (BI) and after induction (AI), as depicted in Figures 9A and 9B. As illustrated in Figure 9, 6xHis-Ubl1 was successfully overexpressed in 2xYT medium at induction temperatures of 20 °C, 25 °C, and 37 °C, demonstrating efficient expression across a broad temperature range. Western blot analysis indicated that the highest protein yield was achieved at 25 °C, which was consequently selected as the optimal temperature for subsequent large-scale expression (Figure 9B). Furthermore, the target protein was expressed successfully in all three *E. coli* strains tested. Notably, expression in BL21(DE3) cells yielded slightly higher protein amounts compared to Rosetta and NiCo strains, and thus BL21(DE3) was chosen for further expression experiments.

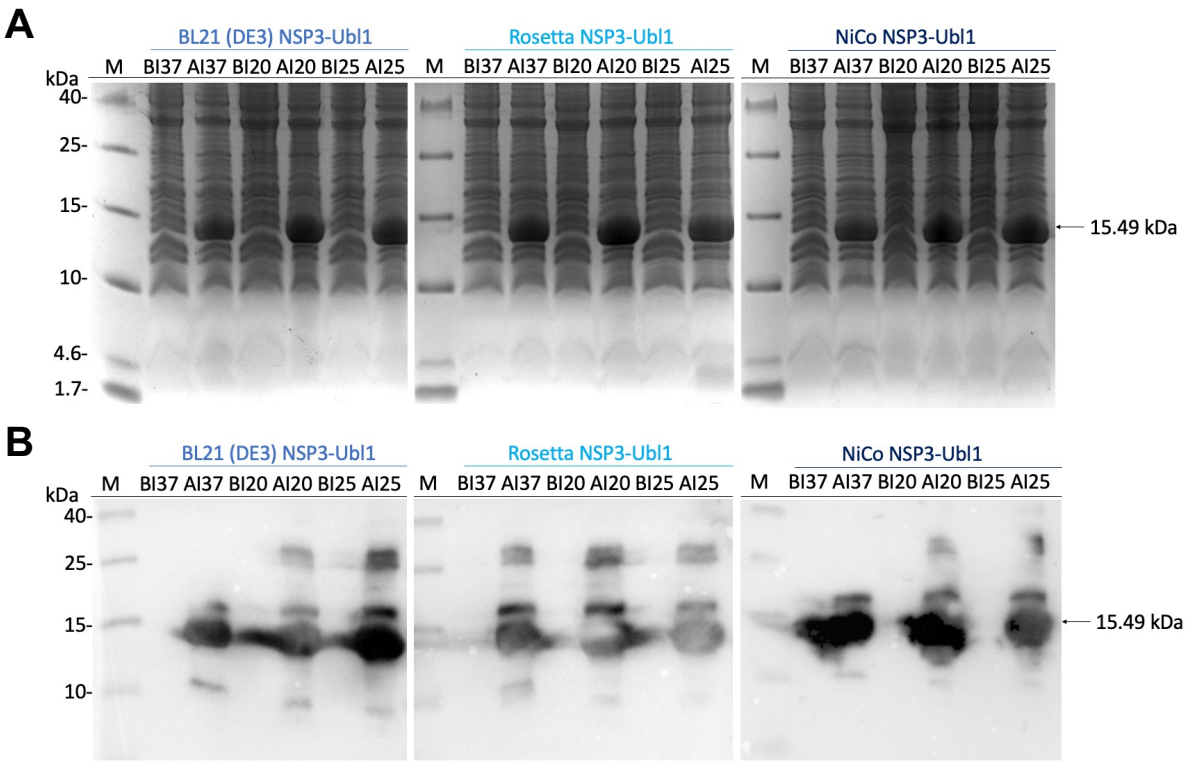


Figure 9: Expression tests of Ubl1 varying temperature and cell strain. *E. coli* BL21 (DE3) cells incorporating the pET-28a(+)-6xHis-Thrombin-TEV-Ubl1 vector were cultivated in 2xYT at 20 °C, 25 °C and 37 °C after induction of expression. Samples before (BI) and after induction (AI) of expression were taken and analyzed by a 20% (v/v) Tris-Tricine SDS-PAGE **(A).** Protein bands were compared to the marker (M) (Spectra Multicolor Low Range Protein Ladder). **(B)** Corresponding Western Blot membranes. The His-tagged protein was identified at the proposed size (15.55 kDa).

1.5.2 Isolation of 7xHis-TEV Protease and ^{1}H-^{15}N 6xHis-Ubl1 by IMAC

Following the confirmation of Ubl1 expression via Western blot analysis and the determination of optimal expression conditions, the subsequent aim was to produce large quantities of highly pure protein. IMAC was employed to selectively purify His-tagged Ubl1, effectively separating it from other histidine-rich *E. coli* proteins. Thereafter, TEV protease digestion was performed to remove the N-terminal His-tag, generating tag-free Ubl1 suitable for downstream applications. The TEV protease itself was previously expressed in *E. coli* and purified using IMAC, as shown in Figure 10.

Prior to purification, the successful expression of 7xHis-TEV protease in *E. coli* was confirmed by the presence of overexpression bands at approximately 28 kDa following IPTG induction (Figure 10A). Over time, the expression level of TEV protease increased, as indicated by more intense bands observed 18 hours post induction compared to the earlier time point at 2 hours post induction (Figure 10A). After cell lysis, the suspension was centrifuged, separating the

supernatant (S1) and pellet (P1), both of which were analyzed by SDS-PAGE. TEV protease was detected in both fractions. Additionally, the observation of a strong band at 45 kDa aligns with documented patterns of *E.coli* contaminations in our lab and is a common impurity in IMAC purified *E.coli* lysates. The supernatant (S1) was treated with 10.3% (v/v) Polymin-P in 25 mM Tris buffer (pH 8), followed by centrifugation to yield a second supernatant (S2) and pellet (P2). SDS-PAGE analysis showed that most protein remained soluble in S2, which was used for further purification (Figure 10A).

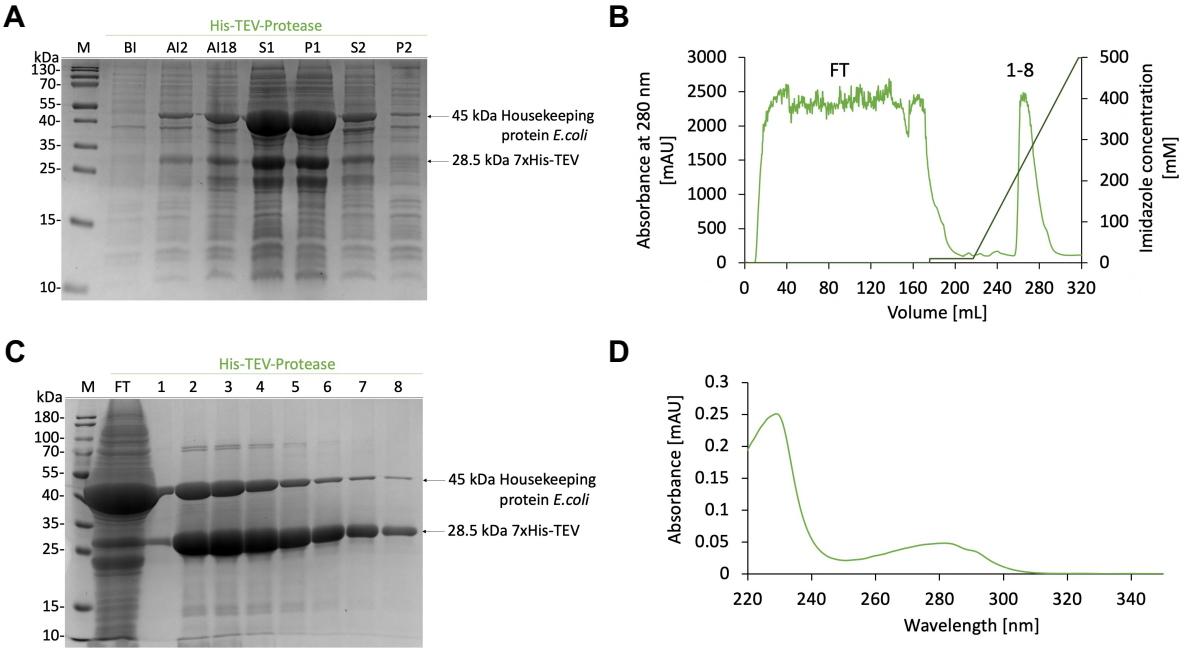


Figure 10: Isolation of 7xHis-TEV protease by IMAC. (A) 15% (v/v) Tris-Glycine SDS-PAGE of samples before induction (BI), after induction of expression (AI) taken after 2 h (AI2) and 18 h incubation (AI18), and after cell lysis with the supernatant (S) and pellet (P), each after the first and second centrifugation step (1/2). Samples were compared to a protein standard (Page-Ruler Prestained Protein Ladder). The protein sample was loaded on a Ni-NTA column equilibrated in 25 mM Tris (pH 8), 500 mM NaCl, 10 mM imidazole, 10% (w/v) glycerine. The corresponding chromatogram is shown in **B**. The absorbance (light green) measured at 280 nm was plotted against the volume in mL. His-tagged TEV-protease (28.5 kDa) was eluted using 25 mM Tris (pH 8), 500 mM NaCl, 500 mM imidazole, 10% (w/v) glycerine. **(C)** SDS-samples were taken for a 15% (v/v) Tris-Glycine SDS-PAGE. Samples of the flowthrough (FT) and the fractions (1-8) were taken and compared to the marker (M) (Page-Ruler prestained protein ladder). **(D)** UV-VIS spectrum of isolated 7xHis-TEV, 1:100 diluted. High absorbance levels at 280 nm were detected (0.0482).

The clarified lysate was loaded onto a Ni-NTA Sepharose 6 Fast Flow column equilibrated with TEV IMAC lysis buffer (25 mM Tris pH 8, 500 mM NaCl, 10 mM imidazole, 10% glycerol). Early fractions exhibited high absorbance, corresponding to elution of non-His-tagged *E. coli* proteins (Figure 10B). SDS-PAGE of the flow-through confirmed the presence of multiple *E. coli* proteins (Figure 10C). To remove nonspecifically bound proteins, the column

was washed with IMAC elution buffer containing 2% (v/v) imidazole (final 10 mM). Elution of 7xHis-TEV protease was achieved by applying a linear gradient of imidazole, with the protein eluting at approx. 265 mM imidazole. This was evident as a prominent absorbance peak between 256 and 296 mL, peaking at 266 mL (2432 mAU), indicating high yield (Figure 10B). Fractions (5 mL each) across the peak were analyzed by SDS-PAGE, showing strong bands at 28 kDa confirming the protease's presence (Figure 10C). However, an *E.coli* housekeeping protein was still detected, indicating some impurity and scope for further optimization. All protease-containing fractions were pooled for subsequent steps. Since imidazole can adversely affect protein stability and activity, the pooled sample was dialyzed overnight at 4 °C against TEV dialysis buffer (50 mM Tris (pH 8), 25 mM NaCl, 50% glycerol, 1 mM EDTA) to reduce imidazole concentration. After dialysis, the sample was concentrated and protein concentration was determined via UV-VIS spectroscopy at A_{280}, reaching 0.0482 (Figure 10D). Based on its extinction coefficient (32,220 M^{-1} cm^{-1}) and a 1:100 dilution, the protein concentration was determined to be 149.69 µM. Using the molecular weight of 28,560 g/mol, this corresponds to a concentration of 4.28 mg/mL. Given a total sample volume of 76 mL, the overall protein yield was 325.28 mg, obtained from a 5 L 2xYT culture, corresponding to a yield of 65.06 mg/L of culture.

Following TEV protease purification, ^{1}H-^{15}N-labeled Ubl1 was purified similarly but using different buffers. After sonication and centrifugation, the clarified lysate was loaded onto a Ni-NTA Sepharose 6 Fast Flow column equilibrated with Ubl1 lysis buffer (50 mM Tris (pH 8), 250 mM NaCl). Initial fractions contained non-His-tagged *E. coli* proteins, eluting with high absorbance (Figure 11A). An SDS-PAGE of the flow-through fraction confirmed multiple contaminant proteins (Figure 11B). To remove nonspecific binders, the column was washed with Ubl1 elution buffer containing 4% (v/v) imidazole (final 20 mM), which resulted in significant absorbance (max 855 mAU), indicating elution of loosely bound proteins and a low amount of target (Figure 11B). 6xHis-Ubl1 was eluted using an imidazole gradient, with the target protein eluting at approximately 387 mM imidazole, observed as a distinct absorbance peak between 99 mL and 111 mL, reaching a maximum of 3070 mAU (Figure 11A). Fractions (2 mL each) collected across this peak were analyzed by SDS-PAGE, confirming Ubl1 presence as strong bands around 15 kDa (Figure 11B). All target-containing fractions were pooled. To reduce imidazole concentration, the pooled Ubl1 sample was dialyzed overnight against Ubl1 dialysis buffer (50 mM Tris (pH 8), 250 mM NaCl, 2 mM DTT, 1 mM EDTA). Dialysis was

performed concurrently with TEV protease digestion to cleave the His-tag from Ubl1. SDS-PAGE analysis post-dialysis and digestion showed that the majority of protein remained His-tagged, indicating incomplete cleavage at this stage (Figure 11B).

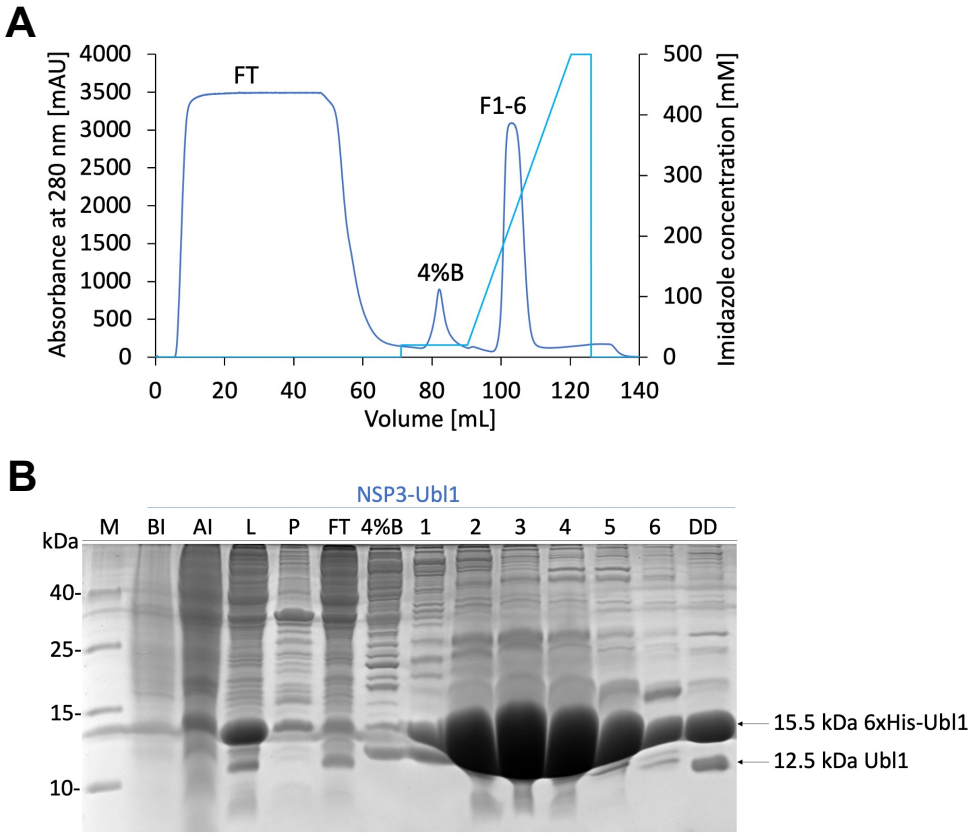


Figure 11: Purification of His-tagged ^{1}H-^{15}N Ubl1 by IMAC. The protein sample was loaded on a Ni-NTA column equilibrated in 50 mM Tris (pH 8), 250 mM NaCl. The corresponding chromatogram is shown in **A**. The absorbance (dark blue) measured at 280 nm was plotted against the volume in mL. The target protein was eluted using 50 mM Tris (pH 8), 250 mM NaCl, 500 mM Imidazole with a gradient starting at 4% (v/v) elution buffer and rising to 100% within 40 min. **(B)** SDS samples were taken for a 20% (v/v) Tris-Tricine SDS-PAGE. Samples before induction (BI), after induction (AI), lysate (L), pellet (P), flowthrough (FT), 4% elution buffer (4%B), fractions (1-6) and after dialysis and digestion (DD) were taken and compared to the marker (M) (Spectra Low Range Marker).

1.5.3 Isolation of ^{1}H-^{15}N Ubl1 by Reverse IMAC

To separate the cleaved, tag-free Ubl1 from undigested 6xHis-Ubl1, free His-tags, and His-tagged TEV protease, a second IMAC purification step was performed. As described previously, IMAC exploits the affinity of His-tags for nickel ions immobilized on the column.[95] Since TEV protease cleavage removes the His-tag from Ubl1, the digested, tag-free Ubl1 does not bind to the Ni-NTA resin and is thus collected in the flow-through fraction, effectively separating it from His-tagged contaminants.

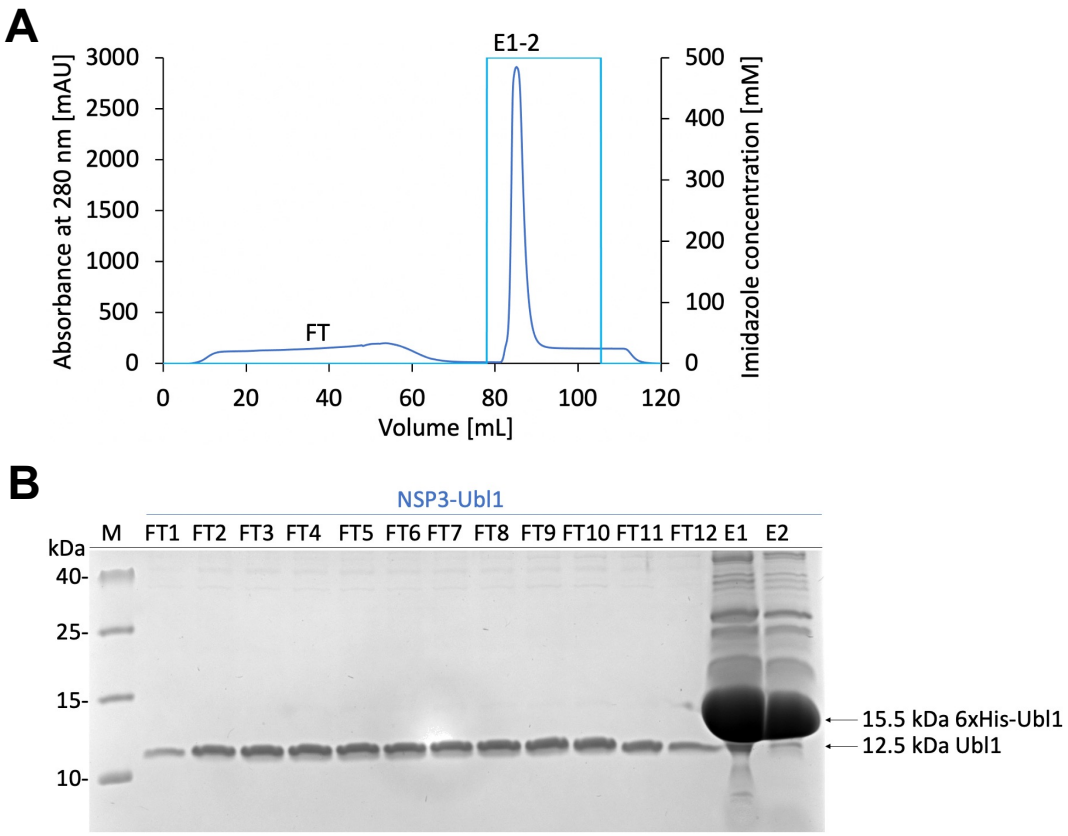


Figure 12: Purification of ^{1}H-^{15}N Ubl1 by IMAC. The protein sample was loaded on a Ni-NTA column equilibrated in 50 mM Tris (pH 8), 250 mM NaCl. The corresponding chromatogram is shown in **A**. The absorbance (dark blue) measured at 280 nm was plotted against the volume in mL. The target protein was expected in the flowthrough (FT). 6xHis-Ubl1 and 7xHis-TEV protease were eluted using 50 mM Tris (pH 8), 250 mM NaCl, 500 mM Imidazole. **(B)** SDS-samples were taken for a 20% (v/v) Tris-Tricine SDS-PAGE. Samples of the flowthrough (FT1-12) and eluate (E1, E2) were taken and compared to the marker (M) (Spectra Low Range Marker).

Following dialysis and TEV protease digestion, the sample was applied to a Ni-NTA Sepharose 6 Fast Flow column equilibrated with Ubl1 lysis buffer (50 mM Tris (pH 8), 250 mM NaCl). Due to successful cleavage, the target protein - now lacking the His-tag - was expected to be found in the flow-through fraction. A broad elution peak was observed between 10 mL and 70 mL, reaching a maximum absorbance of approximately 192 mAU (Figure 12A). Flow-through fractions (5 mL each) were collected and analyzed by SDS-PAGE (Figure 12B), confirming the presence of Ubl1 in all fractions and indicating efficient TEV protease cleavage.

To elute undigested His-tagged Ubl1, residual free His-tags, and His-tagged TEV protease, the imidazole concentration was increased to 500 mM, resulting in a strong elution peak with an absorbance of 2865 mAU (Figure 12A). SDS-PAGE analysis of the elution peak revealed a prominent band at approximately 15 kDa corresponding to His-Ubl1, and a band at 28 kDa representing the His-tagged TEV protease. Although the flow-through fractions contained

predominantly pure, digested Ubl1, the majority of the protein in the sample remained undigested.

1.5.4 SEC of ^{1}H-^{15}N Ubl1

With a sufficient amount of Ubl1 isolated, the next step aimed to further improve sample purity through SEC. This purification step also facilitated buffer exchange to optimize conditions for subsequent structural analyses of Ubl1. SEC separates proteins based on their hydrodynamic radius, eliminating the need for an elution buffer in this experiment.

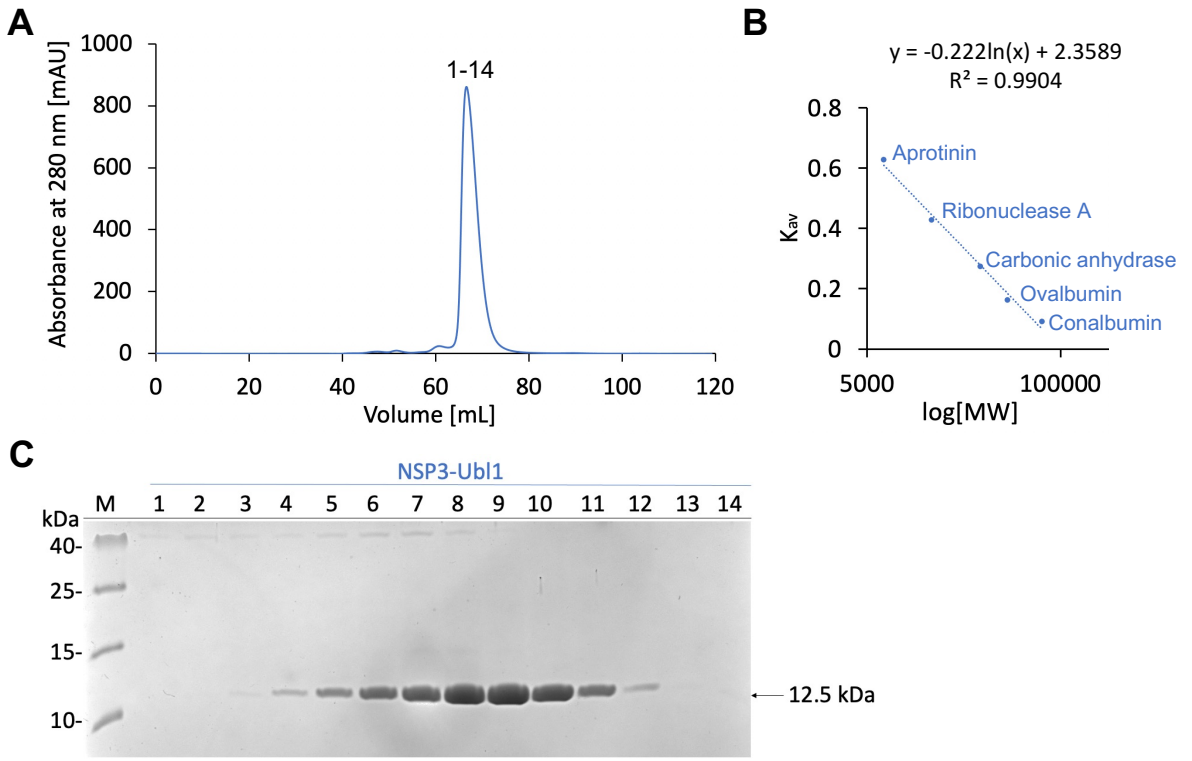


Figure 13: SEC of ^{1}H ^{15}N SARS-CoV-2 NSP3-Ubl1 (1-111). For SEC, a HighLoad 16/600 Superdex 75 pg column (Cytiva, Marlborough, Massachusetts, USA) equilibrated in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl was used. **(A)** Chromatogram of the isolation. The absorbance (mAU) is plotted against the volume (mL). **(B)** SEC calibration curve based on the following standard proteins: Aprotinin (6.5 kDa), Ribonuclease A (13.7 kDa), Carbonic Anhydrase (29 kDa), Ovalbumin (44 kDa) and Conalbumin (75 kDa). The calibration curve was plotted using the gel-phase distribution coefficient K_{av} vs. the logarithm of the molecular weight (MW) in g/mol. Based on the calibration curve, the calculated molecular weight of Ubl1 is 11.56 kDa (12.46 kDa expected) indicating monomeric Ubl1. **(C)** SDS-PAGE of collected SEC fractions (1-14). Bands correlating to monomeric Ubl1 (12.46 kDa) are indicated by an arrow.

Initially, the column was equilibrated with degassed and filtered 50 mM sodium phosphate buffer (pH 6.5) containing 150 mM NaCl. Once equilibrated, the protein sample was loaded onto a HiLoad Superdex 16/600 75 pg column for purification.During the SEC run, a prominent peak appeared in the chromatogram (Figure 13A). Based on the calibration curve (Figure 13B)

and the expected molecular weight of Ubl1, the target protein was anticipated to elute at approximately 65 mL. Indeed, the chromatogram revealed a prominent peak reaching 853 mAU at 66.8 mL, closely matching the predicted elution volume. Fractions of 2 mL each were collected across this peak and analyzed by SDS-PAGE (Figure 13C). Ubl1 was detected in fractions 3 through 12, with no additional impurities observed. Since SDS denatures proteins, the molecular weight of the eluted protein was determined from the SEC calibration curve (Figure 13B) rather than SDS-PAGE. The calibration curve was constructed by plotting the gel-phase distribution coefficient, K_{av}, against the logarithm of molecular weight (log[MW]) using the following protein standards: conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa), and aprotinin (6.5 kDa).

The K_{av} for Ubl1 was calculated using the equation:

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

where V_o is the column void volume (45.62 mL), V_c is the geometric column volume (120 mL), and V_e is the elution volume of Ubl1 (66.80 mL). Using this calibration, the molecular weight of Ubl1 was estimated as 11.56 kDa, which is in close agreement with the theoretical molecular weight of 12.49 kDa. These results confirm that the sample contains primarily monomeric Ubl1. Given the high purity of the pooled fractions (Figure 13C) and the monomeric state of Ubl1, the concentrated SEC samples (total volume: 4.2 mL; concentration: 2.85 mg/mL or 228.77 µM, corresponding to 11.97 mg of protein per liter of M9 medium) were considered suitable for subsequent structural analyses. Notably, half of the His-tagged Ubl1 sample was not subjected to the second IMAC and SEC purification steps and was stored for later use, indicating an expected total yield of approximately twice this amount per liter of M9 medium.

1.6 Discussion

Although significant progress has been made in characterizing the architecture and function of SARS-CoV-2 ROs - particularly DMVs and their associated pores - the precise molecular mechanisms governing their biogenesis and the translocation of viral RNA remain only partially understood. A central player in these processes is NSP3, the largest and one of the

most multifunctional components of the viral replication complex.[45,46,56] The N-terminal Ubl1 domain of NSP3 has emerged as a key structural element, localizing to the cytosolic prongs of the hexameric pore complex and potentially mediating interactions with host and viral proteins, including the N-protein.[56,67] Given its proposed structural and functional relevance in DMV pore formation and membrane remodeling, the Ubl1 domain represents an attractive target for further biochemical and structural studies. To enable such investigations, we established a purification protocol, in which we successfully expressed and purified the isolated Ubl1 domain.

The purification strategy for both 6xHis-TEV protease and 6xHis-Ubl1 was systematically optimized to achieve high yields of pure protein suitable for downstream structural analyses. Initial small-scale expression trials demonstrated robust overexpression of 6xHis-Ubl1 across various *E. coli* strains (BL21(DE3), Rosetta and NiCo) and induction temperatures (18 °C, 25 °C, 37 °C). In line with previous published purification protocols of Ubl1, BL21(DE3) cells yielded highest protein amounts at 25 °C.[80] This aligns well with previous studies suggesting that moderate induction temperatures often favor soluble expression of recombinant proteins, minimizing inclusion body formation and improving folding efficiency.[97,98,99]

Following expression, His-tagged Ubl1 and TEV protease were efficiently enriched from *E. coli* lysates using IMAC. Upon cell lysis, the majority of Ubl1 was recovered in the soluble fraction, while only a minor portion was found in the insoluble pellet. The formation of protein aggregates during sonication cannot be excluded, potentially due to local heat generation adversely affecting protein folding.

The use of a Polymin-P precipitation step prior to IMAC in the TEV protease purification reduced contaminating nucleic acids and precipitated unwanted proteins, thereby enhancing the clarity of the lysate and facilitating downstream purification. Despite this, SDS-PAGE revealed residual *E. coli* housekeeping proteins, particularly a 45 kDa contaminant co-eluting with TEV protease, highlighting an area for further optimization, such as altered wash conditions or additional chromatographic steps, such as SEC.

During IMAC purification, a portion of the His-tagged Ubl1 protein was lost, as evidenced by its presence in the flow-through fractions. This may be attributed to column overloading and could be mitigated in future experiments by using columns with larger bed volumes or by

connecting two Ni-NTA columns in tandem. Similarly, partial elution of His-tagged Ubl1 was observed during the wash step, likely for the same reason. Nevertheless, the majority of Ubl1 was successfully recovered during the elution phase, minimizing the impact of earlier losses.

The IMAC elution profiles of both Ubl1 and TEV protease exhibited sharp absorbance peaks, with TEV protease eluting at 265 mM imidazole and Ubl1 at 387 mM, emphasizing the necessity of a controlled elution gradient. This approach ensures minimal imidazole carryover in the final sample and facilitates downstream dialysis and digestion. Despite this, incomplete cleavage of Ubl1 by TEV protease during dialysis was observed, suggesting suboptimal digestion conditions, reduced enzymatic activity or insufficient incubation time. Optimization of the protease-to-substrate ratio (e.g., increasing to 1:10, protein:protease) may improve cleavage efficiency in future trials too.

Residual His-tagged Ubl1 following digestion necessitated a secondary IMAC purification step. This allowed for the separation of cleaved, tag-free Ubl1 - recovered in the flow-through - from undigested protein, free His-tags, and His-tagged TEV protease, exploiting the differential affinity for the Ni-NTA resin.

The final polishing step employing SEC not only improved sample purity but also facilitated buffer exchange, crucial for maintaining protein stability during subsequent NMR studies. The elution volume of Ubl1 in SEC corresponded closely with its monomeric molecular weight (calculated MW: 12.46 kDa, experimental MW: 11.56 kDa), as verified by the calibration curve using known standards. This confirmed that the purification protocol yielded monomeric Ubl1, a prerequisite for high-resolution structural characterization.

Overall, the integrated purification workflow successfully produced high-purity, monomeric Ubl1 and TEV protease. Nonetheless, future efforts could focus on optimizing TEV protease cleavage efficiency, possibly by fine-tuning enzyme-to-substrate ratios, digestion time, or buffer composition, to reduce the proportion of undigested target protein and simplify purification. Additionally, further refinement of wash conditions during IMAC might minimize co-purification of contaminant proteins, enhancing overall purity. Nevertheless, the current methodology provides a robust foundation for producing samples of sufficient quality and quantity for detailed structural and functional analyse

2. Chapter: An Integrative NMR Pipeline for Validating the Solution Structure and Dynamics of the N-terminal SARS-CoV-2 NSP3 domain

Journal: Nature communications (in preparation)

Authors: **Katharina Vormann**,
Tobias Stief,
Florian T. Tucholski,
Abishek Cukkemane,
Luitgard Nagel-Steger,
Ralf Bartenschlager,
Ralf Biehl,
Nils-Alexander Lakomek

Author contribution: Expression and isolation of ^{15}N-^{13}C labeled NSP3-Ubl1,
Expression and isolation of ^{1}H-^{15}N labeled NSP3-Ubl1,
Acquisition and analysis of 3D NMR experiments,
Assignment of NSP3-Ubl1,
Analysis of NMR relaxation data,
Acquisition and analysis of CD data,
Alpha-Fold prediction of NSP3-Ubl1,
Protein structure analysis,
Homology comparison,
Writing and proof-reading of manuscript

2.1 Abstract

The emergence of AI-based structure prediction tools, such as AlphaFold3 (AF3), has transformed structural biology. However, experimental validation remains essential. Especially intrinsically disordered and flexible regions frequently yield only low-confidence predictions or are lacking in X-ray or cryo-EM structures but can be accessed via Nuclear Magnetic Resonance (NMR) Spectroscopy. Here, we present an integrative NMR pipeline applied to the N-terminal ubiquitin-like domain 1 (Ubl1) of the multi-domain SARS-CoV-2 non-structural protein 3 (NSP3), a crucial component of the double-membrane vesicle (DMV) pore, which is part of the viral replication-transcription complex (RTC). NMR chemical shifts (CS), residual dipolar couplings (RDCs), and SAXS measurements are used to select and confirm the best average solution structure representation (X-ray structure vs AI prediction). NMR relaxation data provide residue-specific information on structural dynamics. We find a highly dynamic ~15-residue N-terminal segment that is undetectable or ambiguous in X-ray structures due to low electron density and has a low confidence level in the AF3 prediction. Thus, as we showcase, NMR provides information beyond static structures, including flexible regions and structural dynamics in a relatively fast manner.

2.2 Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a pathogenic β-coronavirus that caused the coronavirus disease 2019 (COVID-19) outbreak. SARS-CoV-2 is an enveloped virus with a single-stranded, positive-sense RNA genome. The viral genome encodes 16 non-structural proteins (NSPs) required for viral replication.[31] Amongst these, the non-structural protein 3 (NSP3)[26] is the largest one, comprising 15 different domains (Figure 14).

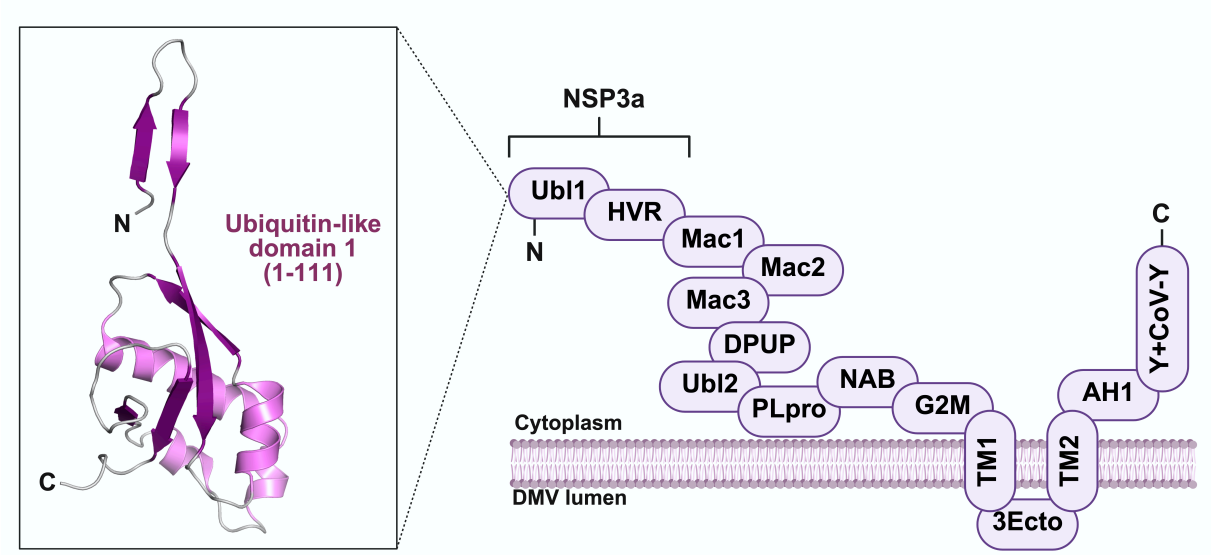


Figure 14: Domain organization of NSP3 highlighting Ubl1. Domain organization of NSP3 starting from the N-terminus. Ubl1 (PDB: 7KAG, 111 residues) is located at the N-terminus, followed by a hypervariable region (HVR), forming together NSP3a. The HVR is followed by three macrodomains (Mac1-3), the domain preceding the Ubiquitin-like domain 2 (DPUP), the ubiquitin-like domain 2 (Ubl2), the papain-like protease (PLpro), the nucleic-acid binding domain (NAB), the G2M domain, the two transmembrane domains (TM1 and TM2) separated by the ectodomain (3Ecto), and the C-terminal domains AH1 and Y+CoV-Y.

NSP3 and NSP4 are sufficient to induce the formation of replication organelles, comprising accumulations of double-membrane vesicles (DMVs)[45, 56,100]. Using cryo-EM tomography of MHV and SARS-CoV-2-infected cells, Wolff *et al.* were the first to report molecular pores embedded into DMVs[56]. These pores allow the exchange of metabolites and viral RNA between the DMV lumen and the cytoplasm[56]. The ~1.8 MDa pore complex presumably comprises both viral and host cell proteins and lipids, with NSP3 and NSP4 constituting the DMV pore[45,56,101]. The pore complex was recently described as encompassing 12 copies of NSP3 and NSP4, with six copies of NSP3 forming the hexameric crown-like structure[46]. The central DMV pore exhibits a diameter of 17 Å and is composed of positively charged arginine residues coordinating RNA translocation along the pore[46]. The most N-terminal NSP3 domain

contains an ubiquitin-like folded domain (Ubl1, residues 1-111), which is followed by an intrinsically disordered Glu-rich hypervariable acidic region[102]. Deletion of Ubl1 abrogates viral replication, while the disordered hypervariable region is dispensable for viral replication[58].

Using fluorescence microscopy, Ubl1 was observed in the prongs of the pore facing the cytosol, making it accessible to viral or host cell binding proteins, including the viral N-protein[56,67]. Notably, an NSP3 deletion comprising the N-terminal domains from Ubl1 up to Ubl2 rendered this protein incapable of remodeling the ER membrane into distinct DMVs[45]. Instead, a network of double-membrane connectors (DMC) was found. Further, no crown-like structure of the pore and a reduction of membrane curvature were observed, and the luminal space and the pore-to-pore distance were significantly increased[45].

Considering that ubiquitin-like domains regulate diverse biological processes,[86] the list of potential interaction partners is substantial. Ubl1 has been shown to interact with the intrinsically disordered region between the N-terminal domain (NTD) and the C-terminal domain (CTD) of the viral nucleocapsid protein (N-protein) via a bipartite interaction[67]. Additionally, fluorescently-labeled Ubl1 colocalizes with the N-protein in the cytosol of human hepatoma cells[61]. Serrano and colleagues identified single-stranded RNA (ssRNA) containing an AUA pattern as an interaction partner of SARS-CoV-1 NSP3a[87].

2.3 Results

Here, we establish an integrative NMR-based pipeline that delivers experimentally validated information on the Ubl1 average structure in solution and, beyond a static structural representation, also includes residue-specific information on structural dynamics (Figure 15). Most importantly, it can also access flexible regions, which are often missing from X-ray crystallography or cryo-EM structures. Also, AI-based predictions often yield only low confidence for those regions. In this integrative NMR-based pipeline, solution NMR chemical shifts (CS) and circular dichroism (CD) spectroscopy provide insights into both residue-specific and overall secondary structure. NMR residual dipolar couplings (RDCs) validate external structural information by X-ray crystallography or artificial intelligence (AI)-based structural predictions, such as e.g. AF3. Thus, a verified average solution structure can be obtained at physiological temperatures. Analytical ultracentrifugation (AUC) and small-angle X-ray

scattering (SAXS) examine the oligomerization state and overall shape of the protein. Beyond static structures, NMR ^{15}N relaxation experiments provide residue-specific information on structural dynamics and, together with secondary chemical shifts, identify and describe flexible and intrinsically disordered regions of the protein.

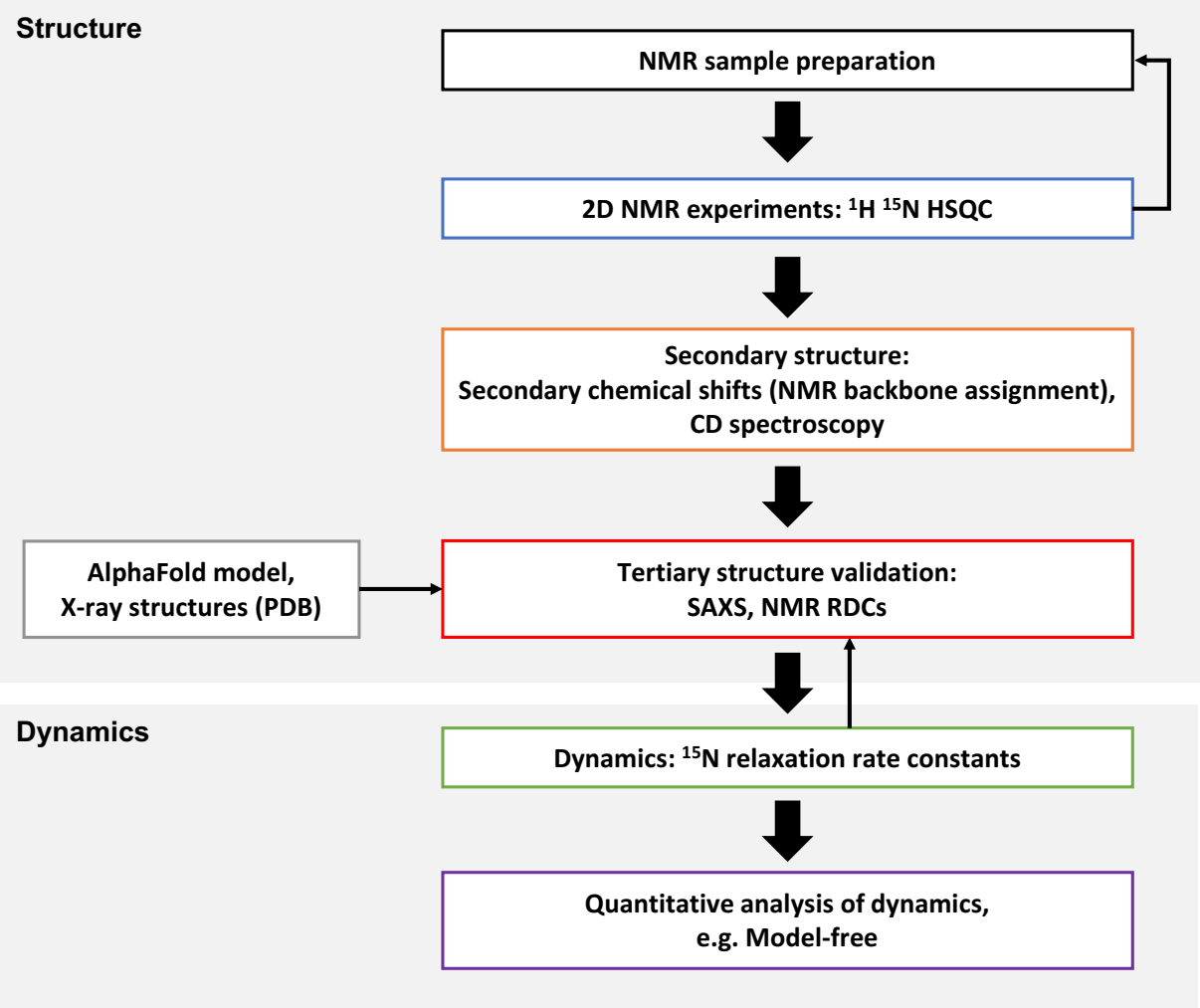


Figure 15: Integrative NMR-based pipeline for studying the structure and dynamics of small and soluble proteins.

2.3.1 Secondary Structure Assessment of Ubl1 by Circular Dichroism (CD) Spectroscopy and NMR Secondary Chemical Shifts (SCS) for Residue-Specific Information

We use the ^{15}N-^{13}C and ^{15}N isotopically labeled SARS-CoV-2 NSP3-Ubl1 domain (111 amino acids, 12.46 kDa, pI: 4.08, sequence: Figure S1) as a model system. NMR protein samples were expressed and purified as described in detail in the Materials and Methods section 2.5.1. The final size-exclusion chromatogram and SDS-PAGE gel are shown in Figure S2. CD data indicate that the secondary structure content of Ubl1 stays similar and stable across a temperature

range from 5 °C to 37 °C (Figure 16A, B). The secondary structure content was estimated using the web server BeStSel[103] (Figure 16D). The total α-helix content varied between 14.3% (37 °C) and 17.4% (25 °C). Antiparallel β-strand conformation fluctuates between 21.4% (15 °C) and 24.5% (25 °C). β-turns range from 13.5% (37 °C) to 15.8% (25 °C). Regions classified as "other" secondary structure elements, including loop regions, amounted to 42.3% (25 °C) up to 49.9% (37 °C). To obtain residue-specific information on the secondary structure elements, NMR secondary CS were measured (Figure 16C).

All NMR experiments were conducted at 25 °C. NMR spectral backbone assignment was obtained using a suite of three-dimensional (3D) backbone assignment experiments, as described in the Materials and Methods section 2.5.3. Of 105 visible resonances, 103 were assigned, in good agreement with the results of Salvi *et al.* and Bessa *et al.*[67,80]. However, slight differences were noted for residue D10, E14, N22, D32, and E111 compared to the assignment by Salvi *et al*.[80]. Further, we additionally assigned amino acids E92, L98, and F106 (for details on the NMR backbone assignment, see Table S6 in the Supporting Information). The assignment of the [^{1}H-^{15}N] resonances of the ^{1}H-^{15}N HSQC spectrum is shown in Figure S3. From the C_α and C_β CS (obtained from the NMR backbone assignment), residue-specific C_α and C_β secondary CS were calculated along the sequence, indicating the respective secondary structure propensities. The regions Y18 to I23, C39 to V44, L89 to D91 and H101 to Y107 comprise β-sheet secondary structure with negative C_α minus C_β SCS ($\Delta\delta C_\alpha$ - $\Delta\delta C_\beta$) up to -7.8 ppm. Quantitative analysis of NMR secondary CS revealed a helical content approximately twofold higher than that estimated from CD spectroscopy at 25 °C (CD: 17.4%, equivalent to 19 of 111 residues; NMR: 31.5%, equivalent to 35 of 111 residues). The regions E29 to N36, V50 to Q66, E70 to P74, and L79 to S83 are marked by two short and two longer α-helices with positive $\Delta\delta C_\alpha$ - $\Delta\delta C_\beta$ SCS shifts up to 8.2 ppm (Figure 16C). According to our analysis, the Ubl1 domain comprises four α-helices and four strands assembled into two β-sheets. This NMR-based information on the secondary structure was obtained without any prior external structural information needed.

Next, we compared the secondary structure content of Ubl1 in solution with the available structural information in the Protein Data Bank. The secondary structure is located in the same regions as those observed in the recent X-ray structure of Ubl1 (PDB code 7KAG; Figure 16D).

However, unlike the X-ray structure, we noted that the N-terminal region (residues T3 to V15) exhibited only weak β-sheet propensity, with a character more akin to a random coil. Notably, the Ubl1 structural model generated by AF3 predicted intrinsic disorder at the N-terminus (Figure 18E), albeit with low confidence. Our secondary chemical-shift analysis experimentally confirms that predictions (Figure 16C).

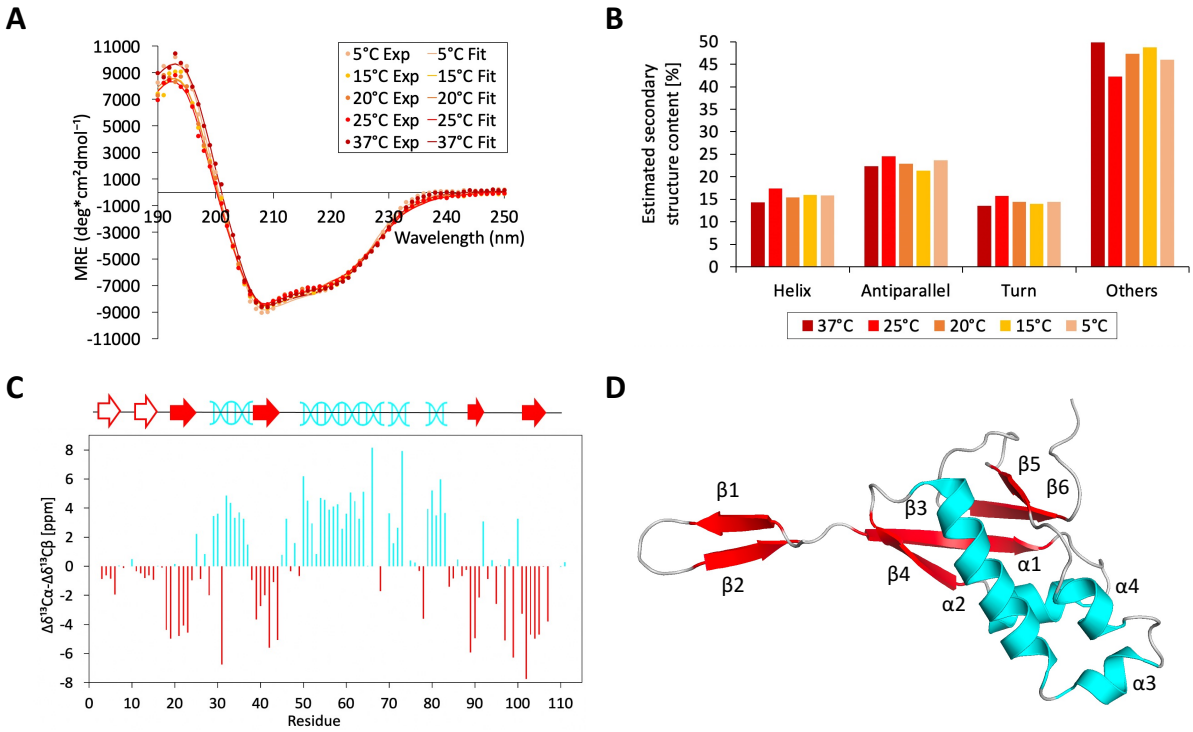


Figure 16: Secondary structure analysis of NSP3-Ubl1 by CD and NMR spectroscopy. (A) Far-UV CD spectrum of NSP3-Ubl1 (7.35 µM) in 50 mM sodium phosphate (pH 6.5) buffer. Data were recorded at different temperatures (5 °C, 15 °C, 20 °C, 25 °C, and 37 °C). The mean residue ellipticity (MRE) is plotted as a function of wavelength (nm). **(B)** Estimated secondary structure content (%) based on the fitted data. **(C)** $\Delta\delta C_\alpha$ - $\Delta\delta C_\beta$ secondary CS of ^{15}N-^{13}C-labelled Ubl1 derived from the backbone chemical shifts assignment (Figure S3) deliver secondary structure propensity. $\Delta\delta C_\alpha$ - $\Delta\delta C_\beta$ values > +4 ppm indicate 100% helical propensity, values < -4 ppm indicate 100% β-sheet propensity. Values between -1 ppm and +1 ppm report random coil characters. Structural elements are represented above the plot based on our data and the crystal structure shown in A. **(D)** Crystal structure of the Ubl1 domain (1-111) of SARS-CoV-2 (PDB: 7KAG) comprising four α-helices and six β-strands.

2.3.2 Assessment of X-ray Structures and AlphaFold3 (AF3) Model Validity for the Ubl1 Solution Structure Using NMR Residual Dipolar Couplings (RDCs)

We measured NMR residual dipolar couplings (RDC) and small-angle X-ray scattering (SAXS) data to discriminate between the X-ray structures of Ubl1 7KAG and 8XAB and the AF3

model[104,105,106]. From the measured NH RDCs (Figure 17), an alignment tensor was calculated using the program PALES[107], and the two respective X-ray structures as input.

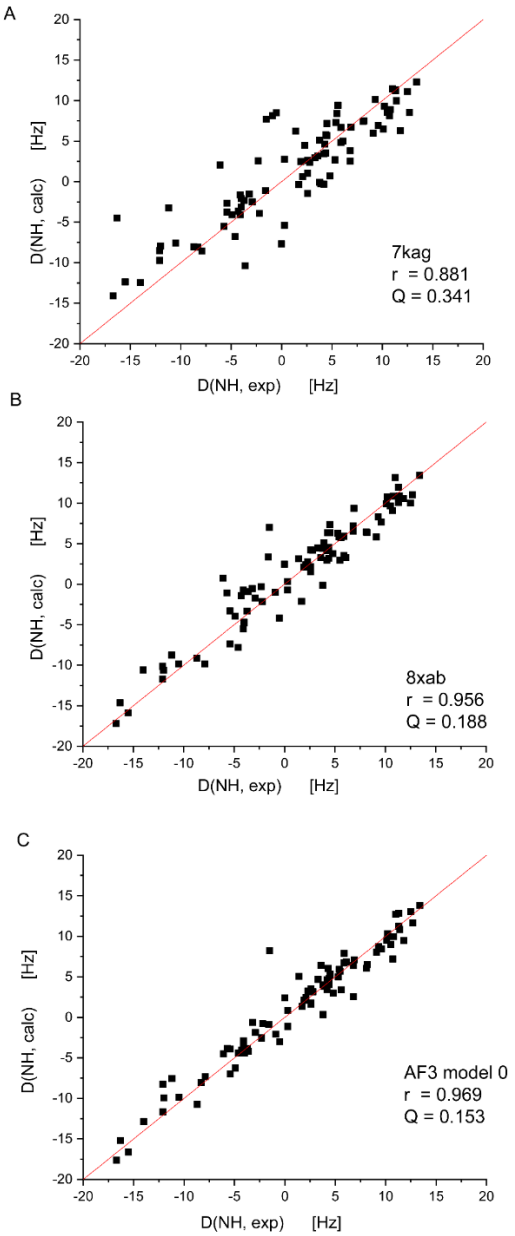


Figure 17: Experimental N-H RDC data. Experimental N-H RDC data are compared with back-calculated data using two X-ray structures and an AF3 model of SARS-CoV-2 Ubl1 for calculation of the alignment tensor: **(A)** X-ray 7KAG, **(B)** X-ray 8XAB and **(C)** AF3 model.

Only the structured regions of Ubl1 were included in the alignment tensor calculation; details are provided in section 2.5.5 of the Materials and Methods. Back-calculated N-H RDCs yield a Pearson correlation coefficient ρ = 0.881 and a Q-value of 0.341 to the experimental NH RDCs for the X-ray structure 7KAG, ρ = 0.956 and Q=0.188 for 8XAB, and ρ=0.969 and Q=0.153 for the AF3 model, favoring 8XAB over 7KAG and the AF3 model slightly over 8XAB as the best structural representation of the rigid part of the Ubl1 structure in solution (Figure 17). It

should be noted that the 8XAB X-ray structure does not include the N-terminus of Ubl1 (which was also excluded from the RDC-based structural validation anyway due to its identified mobility, which could skew the alignment tensor calculation).

2.3.3 Overall Shape Information from SAXS and AUC Data

As complementary information to the RDC-based structural analysis, which focuses on the local structure (and only the rigid part of the protein), SAXS data were recorded to obtain overall shape information on Ubl1 (Figure 18A).

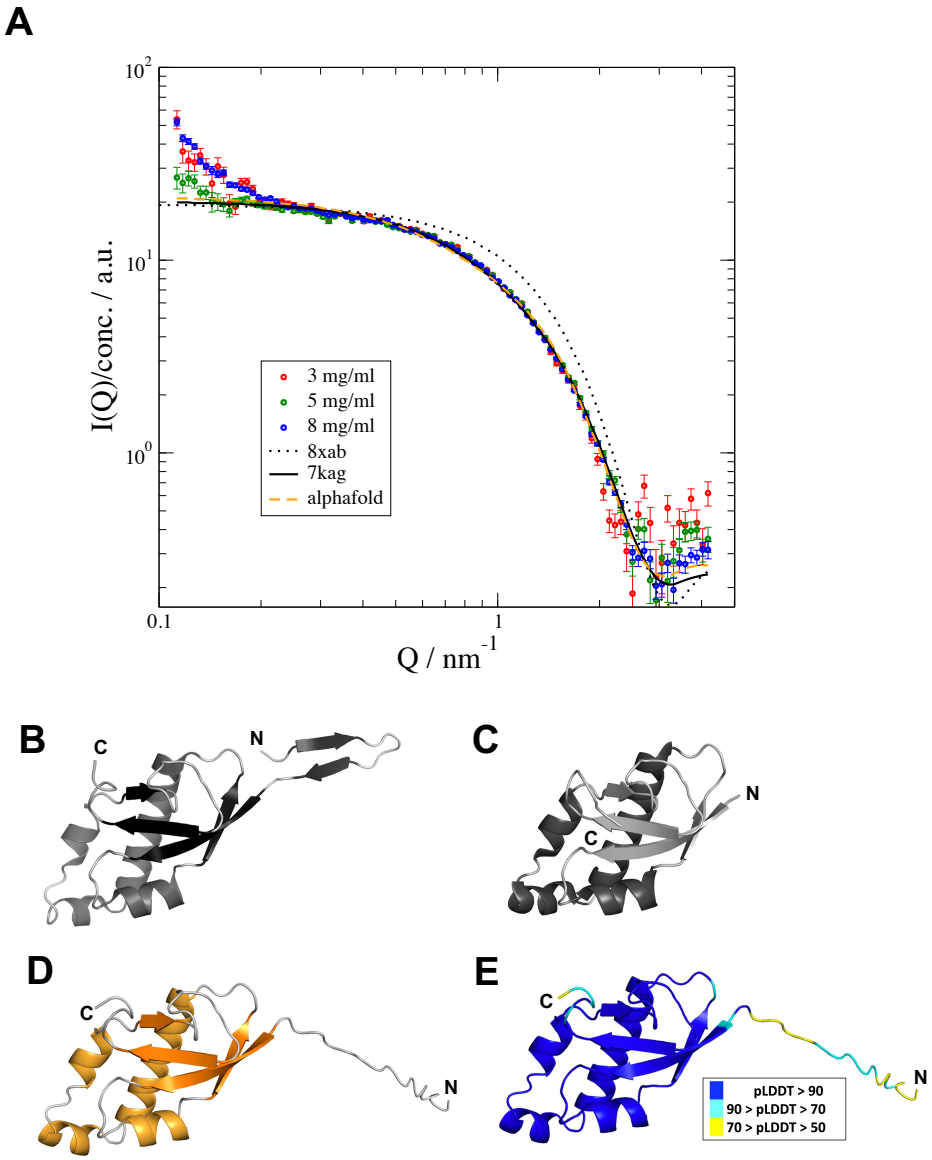


Figure 18: SAXS measurements of Ubl1 at different protein concentrations and structural representations of Ubl1. (A) SAXS form factor after concentration scaling. Lines indicate the calculated form factor from monomers present in PDB code 7KAG (black), 8XAB (dotted line) and the AF3

structure (orange). Corresponding crystal structures are demonstrated in panels B to E: **(B)** PDB 7KAG, **(C)** PDB 8XAB, **(D)** AF3 Ubl1, with **(E)** displaying the confidence level.

At the lowest scattering vectors, the slight increase indicates the presence of small amounts of aggregates. But the good overlap ($Q > 0.2$ nm^{-1}) suggests that no structure factor influences the data. To discriminate between the two PDB structures, the corresponding form factors were calculated from the atomic positions of the monomers. To confirm the absence of pre-existing oligomerization, analytical ultracentrifugation was performed, which revealed that Ubl1 remained monomeric (Figure S4). The measured frictional ratio (f/f_0) of 1.33 further supports a globular structural conformation. Examining the SAXS data (Figure 18A), we find that the 7KAG (Figure 18B) and the AF3 model (Figure 18D, E), both of which include the N-terminal 15 residues, can be fitted closely to the experimental data. In contrast, 8XAB (Figure 18C), lacking the Ubl1 N-terminus, cannot faithfully represent the experimental SAXS data for the full protein but yields a shift to lower Q in the best-fit model. This shift indicates a more compact structure present for 8XAB (Figure 18C). The minimum, around 2.5 nm$^{-1,}$ is due to the ellipsoidal shape of the central part of the protein structure. We note that the present data indicate only an extended domain, which may also be present in a disordered configuration or as an ensemble of configurations. Therefore, the SAXS data indicate that inclusion of the N-terminus is essential for a faithful fit. Still, the SAXS data cannot distinguish between the N-terminus being structured (as in 7KAG) or intrinsically disordered, as suggested by the AF3 model (with low confidence level, however), and experimentally established and verified by NMR secondary CS.

2.3.4 Structural Dynamics of Ubl1 Analyzed by NMR Relaxation Measurements

2.3.4.1 Fast ps-ns Dynamics

To characterize the residue-specific structural dynamics of Ubl1, NMR ^{15}N relaxation rate constants were measured (Figure 19). Data was recorded at three different magnetic field strengths: 14.1 T, 21.15 T, and 28.2 T, corresponding to 600 MHz, 900 MHz, and 1200 MHz ^{1}H Larmor frequencies, respectively, and a temperature of 25 °C. ^{15}N R_2 rate constants (derived from $R_{1\rho}$ experiments) characterize ps-ns structural dynamics of the protein, which are faster than the protein's rotational correlation time τc (Figure 19A). Higher R_2 ($R_{1\rho}$) rate constants indicate the presence of secondary structural elements, such as α-helices or β-sheets. Most residues in the region Q16-P109 have a R_2 rate constant ranging between 10 s^{-1} to 12 s^{-1} for

600 MHz, 13 s^{-1} to 16 s^{-1} for 900 MHz, and 18 s^{-1} to 22 s^{-1} (1.2 GHz), indicating a structured region, in agreement with the NMR chemical shift data (Figure 16C). (The systematic offset of the R_2 ($R_{1\rho}$) data is in line with the expected magnetic-field dependence of the data.) Different from the remaining protein, the N-terminus shows R_2 rate constants as low as 2.9 s^{-1} (600 MHz), indicating a highly flexible region (Figure 19C, D), in agreement with the close to random-coil character identified by the C_α minus C_β SCS (Figure 16C).

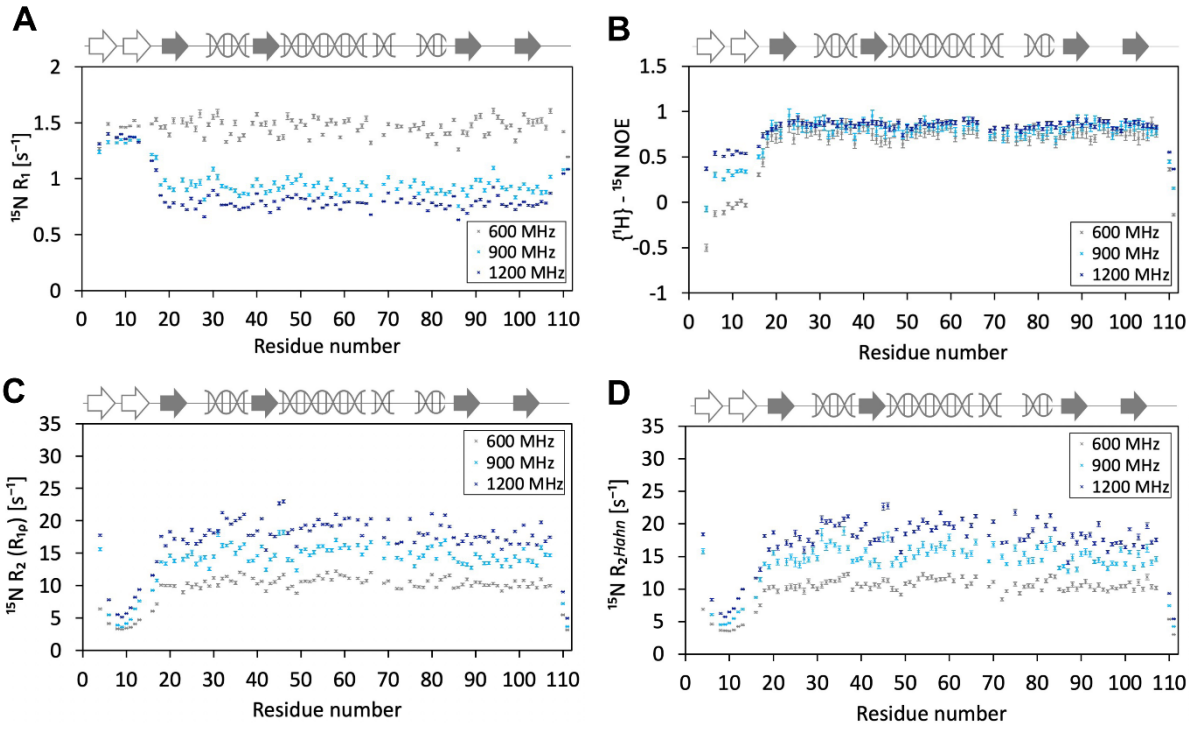


Figure 19: NMR relaxation data reveal a highly flexible N-terminus of Ubl1. NMR ^{15}N relaxation data of Ubl1 spectra were recorded at 600 MHz, 900 MHz, and 1.2 GHz. **(A)** High ^{15}N R_1 rate constants indicate a highly dynamic and flexible N-terminus of Ubl1, including the first 15 amino acids. **(B)** Low {^{1}H}-^{15}N NOE values indicate a high degree of internal dynamics on the fast picosecond timescale at the N-terminal region of Ubl1. **(C)** Low ^{15}N R_2 ($R_{1\rho}$) rate constants underline the finding of a high degree of intrinsic dynamics and flexibility at the N-terminus of Ubl1, while elevated ^{15}N R_2 rate constants for residues Q16-P109 point to rigid secondary structure. **(D)** ^{15}N R_2 rate constants derived from Hahn-echo-based measurements are sensitive to both ps-ns and µs-ms dynamics. Their close resemblance to the ^{15}N R_2 ($R_{1\rho}$) rate constants points to little additional (conformational) dynamics on the µs-ms timescale.

Consistent with the ^{15}N R_2 relaxation rate constants, low {^{1}H}-^{15}N heteronuclear NOE values further support a flexible N-terminus (Figure 19B). Thus, the observed high internal dynamics are also present on the fast picosecond timescale, as indicated by the heteronuclear NOE values. ^{15}N R_1 rate constants, sensitive to ps-ns dynamics, show only slight fluctuations at 600 MHz, but pronounced effects are observed at 900 MHz and even more so at 1200 MHz (Figure 19A). While the structured part of the protein shows systematically lower rate

constants at 900 MHz and even lower at 1200 MHz, as expected because of their field dependence, the ^{15}N R_1 rate constants for the N-terminus are substantially higher, underlining the increased local flexibility of the N-terminus (Figure 19A). Apparently, the ^{1}H and ^{15}N Larmor frequencies of 900 MHz and 1200 MHz are more sensitive for probing the underlying ps to low ns timescale than those at 600 MHz.

2.3.4.2 Extended Model-Free Analysis of Fast ps-ns Dynamics

To better visualize the underlying dynamics of the N-terminus relative to the remainder of the protein, we performed an extended model-free analysis of the NMR ^{15}N relaxation data using the *relax* software as described in the Materials and Methods Section[108,109,108]. The model-free analysis characterizes the amplitudes of motion by a residue-specific order parameter $S^2(i)$ and the timescale of motion by a residue-specific correlation time $\tau(i)$ for residue *i.* In the extended model-free analysis, both parameters are further discriminated by an order parameter plus associated correlation time characterizing fast-time-scale motions, $S_f^2(i)$ and $\tau_f(i)$, and a slow-time-scale order parameter, $S_s^2(i)$, and correlation time, $\tau_s(i)$. Both time scales are faster than the overall rotational correlation time τ_m. Order parameters describe the amplitudes of motion of the corresponding internuclear NH vector and usually range from 0 (highly dynamic) to 1 (rigid). The lower the order parameter, the larger the corresponding motional amplitudes of the internuclear NH vector.

For the implementation of the extended model-free analysis, we utilized ^{15}N relaxation data (R_1, R_2, hetNOE) measured at three different magnetic field strengths, 14.1 T, 21.15 T, and 28.2 T, corresponding to ^{1}H Larmor frequencies of 600, 900, and 1200 MHz. The AF3 structural model of Ubl1 (1-111) was employed for diffusion tensor calculation, yielding an axially symmetric (spheroidal) diffusion tensor with an overall tumbling correlation time of 8.36 ns. For comparison, based on the X-ray structures 7KAG and 8XAB, the diffusion of Ubl1 is characterized by an axially symmetric (spheroidal) diffusion tensor, with overall tumbling correlation times of 8.55 ns for 7KAG and 8.58 ns for 8XAB. A more detailed description of the diffusion tensor is provided in Supplementary Table S8.

The residue-specific parameters of the “model-free” analysis, using the AF3 structural model for diffusion tensor calculation, are visualized in Figure 20. Here, the order parameters S_f^2

(Figure 20A), characterizing the residue-specific amplitudes of fast internal motions, show values between 0.8 and 1.0 for the folded part of Ubl1, from Y18 to Y107. Order parameters at this level indicate restricted mobility of the corresponding NH internuclear vector on the fast timescale, as expected for structured regions in a folded protein and consistent with the secondary structure propensity derived from the secondary CS (Figure 16C). The timescale of the fast motion, described by the residue-specific correlation time τ_f (Figure 20B), varies between 5 and 20 ps for the structured part of the protein (Y18 to Y107) and is mainly associated with libration motions. The S_s^2 order parameters (Figure 20C), characterize the amplitudes of motions of backbone NH vectors in a time scale slower than τ_f but faster than the overall tumbling correlation time of the protein (here around 8.5 ns at 25 °C). Most residues in the structured region show S_s^2 order parameters higher than 0.9, pointing to only small amplitudes of motion on the slow timescale, with a correlation time τ_s between 1 ns and 2 ns for most residues (except for a few residues that show larger τ_s but with high uncertainty). This timescale has been connected to backbone dihedral angle sampling of the protein backbone, in the case of intrinsically disordered proteins[110]. We observe this timescale here also for a structured protein.

The N-terminal part of Ubl1 shows substantially lower S_f^2 , and therefore larger amplitudes of motion, than the rest of the Ubl1 backbone. The lowest S_f^2 is observed for residues D9 and D10 with 0.58, whereby both residues are less restricted than the remaining residues. The associated fast timescale of motion for the N-terminal residues lies in the range of τ_f =40 ps to 70 ps (Figure 20B).

Regarding the slower timescale, the order parameters S_s^2 of the N-terminal residues are much lower than those of the remaining structured part of Ubl1, as low as S_s^2=0.21 for amino acid D9. The associated slow timescale lies between τ_s =1 ns and 1.5 ns (Figure 20D). Thus, the unstructured N-terminus of Ubl1 has a high flexibility on both fast and slow timescales, while the remaining structured part of the protein is more rigid. This is in line with the secondary CS that indicate intrinsic disorder of the N-terminus (Figure 16C).

Back-calculation of the NMR relaxation input data reveals high agreement between the experimental NMR input ^{15}N relaxation data and the data back-calculated from the results of the extended model-free analysis (Figure S6).

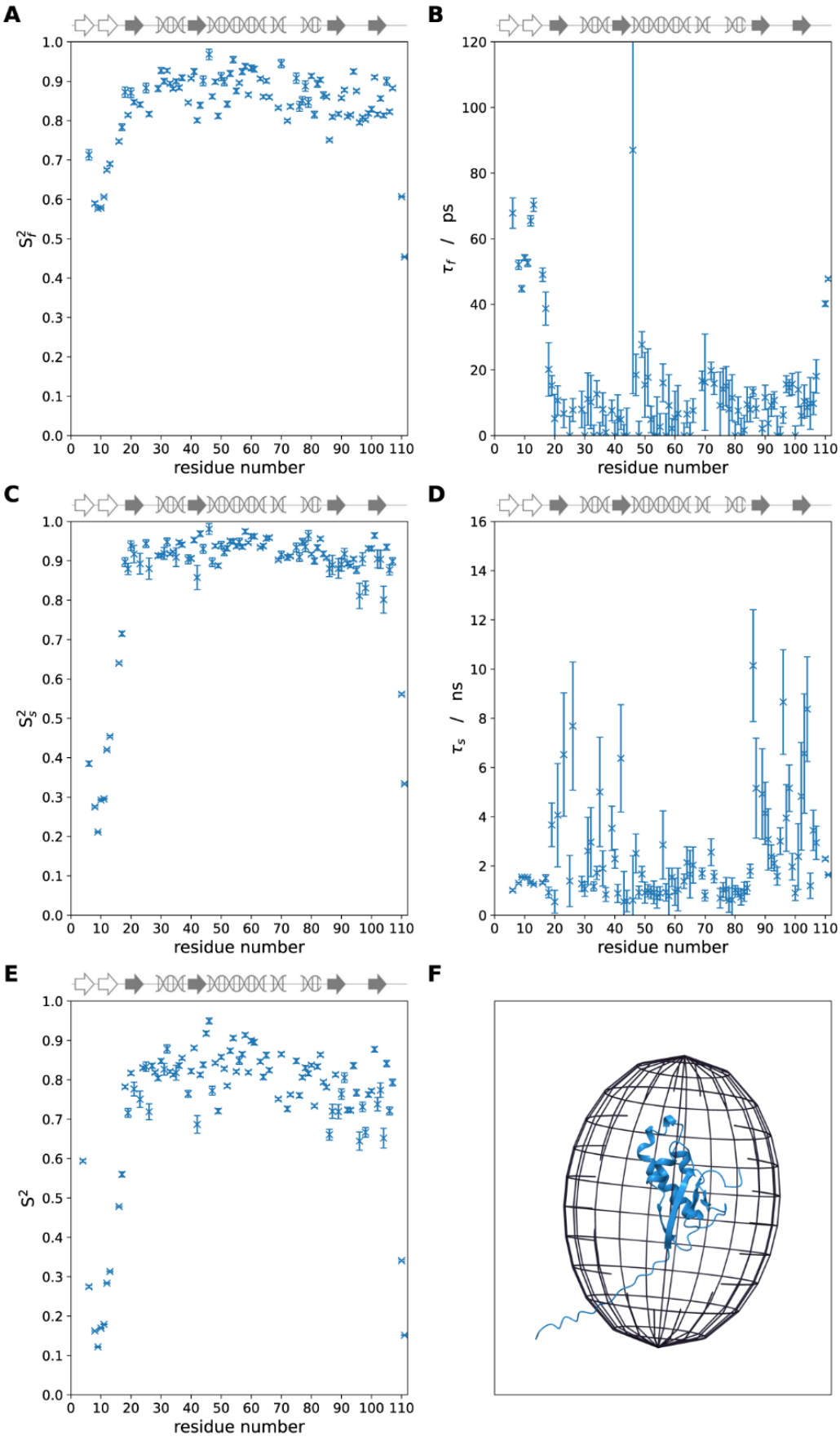


Figure 20: Extended model-free analysis of NMR ^{15}N relaxation data of Ubl1. Residue-specific order parameters and correlation times of the model-free analysis of Ubl1 using the AF3 structural model of

Ubl1 for diffusion tensor calculation. **(A)** Order parameter S_f^2 describing the residue-specific amplitudes of fast internal motions. **(B)** The associated correlation time τ_f defines the timescale of fast internal motions. **(C)** Order parameter S_s^2 describing slow internal motions. **(D)** Associated correlation time of slow internal motions τ_s **(E)** General order parameter $S^2 = S_s^2 * S_f^2$. **(F)** Spheroidal diffusion tensor representation relative to the AF3 structural model of Ubl1.

The model-free order parameters and correlation times were also calculated using the two X-ray structures 7KAG (Figure S7) and 8XAB (Figure S8) as input for the diffusion tensor calculation and extended model-free analysis using the relax software[108]. The obtained order parameters and correlation times show high agreement (Figures S7 and S8), proving the robustness of the analysis.

2.3.4.3 µs-ms Dynamics

Hahn-echo-based ^{15}N R_2 experiments were conducted at 600 MHz, 900 MHz, and 1200 MHz to probe for the presence of conformational exchange on a µs to ms timescale (Figure 19D). Unlike the $R_{1\rho}$ experiments that refocus conformational dynamics slower than the inverse of the spin-lock RF amplitude (here 2 kHz, thus slower than about 80 µs), the Hahn-echo-based ^{15}N R_2 experiments capture all ^{15}N R_{ex} exchange contributions. Those R_{ex} contributions are sensitive to µs-ms conformational dynamics (faster than about 10 ms) and scale approximately with the square of the magnetic field strengths. Therefore, at 1200 MHz, they are expected to be about four times larger than at 600 MHz. The Hahn-echo-based ^{15}N R_2 rate constants do not detect substantial R_{ex} contributions (and thus little µs-ms dynamics) for any residues of NSP3 Ubl1.

2.3.5 Structural conservation of Ubl1

Ubl1 adopts a ubiquitin-like fold composed of four α-helices and two β-sheets, with an intrinsically disordered, flexible N-terminal segment (Figures 16C, 19, and 20). The AF3-predicted structure (Figure 18D) aligned well with experimental RDCs, and SAXS/AUC analyses confirmed a monomeric, globular conformation in solution. Ubl1 is highly conserved between SARS-CoV-1 and SARS-CoV-2 (≈79% sequence identity; Figure S5), though minor structural differences occur at the N-terminus. The SARS-CoV-2 Ubl1 crystal structure (PDB 7KAG) revealed a homodimer comprising four α-helices and six β-strands per monomer (Figure 16D). In contrast, our data identified a similar fold but only four β-strands, consistent with the disordered N-terminal region and the AF3-predicted SARS-CoV-2 model (Table S10). The NMR structure of SARS-CoV-1 Ubl1 (PDB 2IDY)[87] showed three α-helices and four β-strands,

whereas the AF3 model predicted four α-helices and six β-strands, resembling the SARS-CoV-2 X-ray structure. Structural alignment of the SARS-CoV-1 and SARS-CoV-2 Ubl1 domains yielded RMSD values around 1.85 Å (Figure S9), indicating high overall similarity, with the largest differences localized to the flexible N-terminus. The greatest structural similarity was observed between the AF3-predicted Ubl1 domains of SARS-CoV-1 and SARS-CoV-2, despite differences at the N-terminus, which appeared flexible in SARS-CoV-2 and folded in SARS-CoV-1 (0.346 Å). These results support a conserved Ubl1 core architecture across both viruses.

2.4 Discussion

Using NMR, SAXS, CD, and AUC, we have studied the structure and structural dynamics of NSP3-Ubl1 from SARS-CoV-2 as an essential step in better understanding the functional role of this NSP3 N-terminal domain. We employed an integrative NMR pipeline, starting with CD spectroscopy, to obtain global information about the secondary structure of Ubl1. Based on NMR 3D backbone experiments, the secondary CS revealed additional residue-specific details about the secondary structure. Using AF3, we obtained a 3D structural model of Ubl1, which was validated by RDCs. SAXS measurements and AUC experiments investigated the overall shape and the oligomerization state of Ubl1. NMR relaxation data revealed residue-specific internal dynamics, which were further analyzed using a model-free approach, providing quantitative insights into the amplitude of motion and timescale of internal backbone motions.

Our data confirmed that Ubl1 is a globular and folded protein comprising four α-helices and two stable β-sheets (Figure 16A-C). NMR secondary chemical shift analysis indicated a helical content of 31.5%, approximately twice the value estimated by CD spectroscopy (17.4%). The helical content derived from NMR data is consistent with previously reported crystal structures and with available literature data (PDB: 7KAG, 8XAB). In contrast, the CD measurements appear to underestimate the total helical content, suggesting that not all helical elements contribute equally to the CD signal under the conditions used. However, a large part of the protein is well structured and relatively rigid on the ps-ns timescale, as expected for a protein with a ubiquitin-like fold[111,112]. Using NMR relaxation experiments followed by subsequent extended model-free analysis, we observed a high flexibility in the

N-terminal region of Ubl1 (residues T3-V15), consistent with an intrinsically disordered character and only a weak β-sheet propensity in this region, as indicated by secondary chemical shift data (Figure 16C). NMR relaxation data at high magnetic field strengths (900 MHz and 1200 MHz, Figure 19) provided new insights into the Ubl1 structural dynamics, particularly for the fast ps-ns timescale probed by the ^{1}H and ^{15}N Larmor frequencies.

Notably, the homologous Ubl1 protein from SARS-CoV-1 exhibits a structural organization similar to that of SARS-CoV-2 Ubl1 (Figure S9, Table S10) and shares a conserved amino acid sequence with approximately 79% identity (Figure S5), closely reflecting the 79.6% nucleotide sequence identity between the two viral genomes[113]. Interestingly, for SARS-CoV-1, {^{1}H}-^{15}N heteronuclear NOE data also indicated high internal dynamics[87]. Even though the domain organization of CoV genera is different, nine domains regarding structure and function are conserved, including Ubl1[31,114]. In SARS-CoV-1, SARS-CoV-2 and MHV, members of the beta-CoV genera, Ubl1 is known to interact with the viral N-protein[58,67,115,116] and RNA[87]. Deletion of the Ubl1 domains abrogated the viral replication cycle[58,45]. Equivalent to our results and published crystal structures of SARS-CoV-2, MHV, and SARS-CoV-1, Ubl1 exhibited a globular and folded structure[87,116]. Furthermore, the localization of Ubl1 in NSP3 is conserved along different CoV-genera, positioned at the very N-terminus of NSP3.[114] Double membrane structures induced by NSP3 have been observed in the human coronavirus NL63 (alpha CoV) speculating for a similar function of Ubl1[117]. Since both structure and function is conserved for the N-terminal NSP3 domain of alpha and beta CoVs, Ubl1 represents a potential drug target for future coronaviruses.[31] These structural and dynamic similarities might point to a similar functional role, such as interaction with other cellular and viral factors, including viral RNAs or the N-protein[67,87].

A critical interaction partner of Ubl1 is the SARS-CoV-2 N-protein[67]. The N-protein was found to form biomolecular condensates upon mixing with RNA[118,119,59,120,121,122]. The extent to which NSP3 participates in biomolecular condensate formation remains partially controversial. Carlson *et al.* found that GFP-tagged Ubl1 partitioned into N-protein droplets and filamentous structures[123]. Further, a colocalization of N-RNA condensates with NSP3-Ubl1 and NSP12 was observed [124]. Savastano *et al*. identified viral RNA polymerases in liquid N:RNA

droplets[119]. On the contrary, Ubl1 was also suggested to interrupt RNA-driven phase separation of N-protein by displacing RNA[61,84].

Without a doubt, Ubl1 plays a critical role in mediating the interaction between the DMV pores and the N-protein, which is required for the assembly of virus particles. As the Ubl1 domain is also indispensable for viral replication (at least in the case of MHV[116]), it represents a promising target for drug development. We speculate that the intrinsically disordered N-terminus may play a critical role in tuning the interaction of Ubl1 with viral RNA and the N-protein. The absence or presence of the Ubl1 N-terminus (A1 to V15), or its potential shielding by N-terminal Ubl1 fluorescence tags, may rationalize previously conflicting results. Further, intrinsically disordered proteins (IDP) or intrinsically disordered regions (IDR) are well known to participate in biomolecular condensate formation. Therefore, the intrinsically disordered character of the Ubl1 N-terminus may be critical in tuning the docking or undocking of the viral RNA to the N-protein condensates for viral RNA encapsidation vs RNA translation, respectively.

To summarize, solution NMR spectroscopy, supported by further biophysical methods, identifies a highly flexible and intrinsically disordered 15 amino acid residues long N-terminal region of the otherwise well-folded Ubl1 domain. This region constitutes the exposed N-terminus of SARS-CoV-2 NSP3, located at the DMV pore's exit site in the cytoplasm. Both the structure and dynamics of Ubl1 appear conserved across CoVs. We speculate that the disordered Ubl1 N-terminus connects the viral RTCs in the DMV interior to biomolecular condensates formed by the N-protein and viral RNA.

The recent advances in AI-based structure prediction, exemplified by AF3, have profoundly impacted structural biology by enabling high-confidence structural models, even in the absence of experimental data. For the globular core of SARS-CoV-2 NSP3-Ubl1, AF3 provided a reliable structural prediction that aligns well with both our experimental NMR data and available crystal structures. The predicted secondary structure elements and overall fold were confirmed by SCS, RDCs, and SAXS, supporting the robustness of AF3 in modeling stable and ordered protein domains.

However, our study also highlights the limitations of AF3 in predicting dynamic or intrinsically disordered regions. Specifically, the 15-residue N-terminal segment of Ubl1 (residues T3-V15), which exhibits high flexibility and an intrinsically disordered character in our NMR relaxation and secondary chemical shift analyses, was predicted with low confidence by AF3 and is not resolved in available X-ray or cryo-EM structures. This underscores the need for experimental approaches to accurately characterize flexible regions, which often play critical functional roles, particularly in protein-RNA or protein-protein interactions and phase separation processes.

Our integrative approach demonstrates that combining AF3 predictions with solution NMR and complementary biophysical techniques provides a powerful strategy for validating static structural models and extending them into the dynamic regime. This is especially relevant for multi-domain or partially disordered proteins such as NSP3, where functional interactions may rely on transient or flexible elements that current AI-based methods do not adequately capture. In this context, AF3 serves as a valuable starting point for hypothesis generation and model building, while experimental validation remains essential for functional interpretation, particularly at the dynamic and disordered interfaces.

2.5 Methods

Here, we briefly outline the integrative NMR-based pipeline for obtaining structural and dynamic information on a small, soluble protein (Figure 15). This particularly applies to small proteins (<20 kDa) that are difficult to study by cryo-EM-based methods but are well-suited for investigation by NMR spectroscopy.

Following the sample preparation of a ^{15}N-labeled sample (section 2.5.1), the NMR sample quality is evaluated by recording a ^{1}H-^{15}N HSQC spectrum (section 2.5.3). The sample preparation is optimized until a high-quality ^{1}H-^{15}N HSQC spectrum can be recorded. This might become an iterative process. After, or in parallel, the secondary structure content of the protein of interest is determined by CD spectroscopy for overall information (section 2.5.2). Then, 3D NMR spectra for backbone assignment are recorded on a ^{15}N-^{13}C-labeled sample, from which SCS are extracted for residue-specific information on the secondary structure of the protein (section 2.5.3).

To obtain information on the 3D structure of the protein in solution and avoiding the time-intensive step of determining an NMR-structure, the validity of available existing X-ray structures or structural models generated by AI-based structural prediction methods (here: AlphfaFold3; section 2.5.4) is evaluated by using NMR RDCs (section 2.5.5). Following that, the protein's overall shape and oligomerization state are examined by combining SAXS (section 2.5.7), AUC (section 2.5.6), and SEC chromatography (section 2.5.1). Residue-specific information on internal structural dynamics is obtained by NMR ^{15}N relaxation experiments (section 2.5.8). Information on amplitudes of motions and timescales is obtained by an NMR-based model-free analysis (section 2.5.9).

2.5.1 Sample Preparation of ^{1}H-^{15}N and ^{15}N-^{13}C NSP3-Ubl1

The gene base sequence encoding the SARS-CoV-2 NSP3-Ubl1 domain (111 amino acids, 12.46 kDa, pI: 4.08, Figure S1) was extracted from the NCBI genome entry (NC_045512.2) and codon-optimized for expression in *Escherichia coli* (*E. coli).* The construct was subcloned in a pET28a(+) vector with Kanamycin resistance. To facilitate protein isolation, a *hexa*-histidine tag was included at the N-terminus. Additionally, a *Tabacco Etch Virus* (TEV) - cleavage side (amino acid sequence: E-N-L-Y-F-Q-II-A, where II denotes the cleavage side) was inserted for tag removal. The pET28a(+)-6xHis-TEV-Ubl1 plasmid was transformed into chemically competent *E. coli* BL21 (DE3) cells. For precultures, cells were grown at 37 °C and 160 rpm in 2xYT medium (Thermo Fisher, Waltham, MA, USA) supplemented with 50 µg/mL Kanamycin overnight. For isotope labeling of ^{15}N-^{13}C Ubl1, cells were grown in M9 minimal medium supplemented with ^{15}N-NH_4-Cl (99%, Cambridge Isotope Laboratories, USA) and $^{13}C_6$-D-Glucose (U-$^{13}C6$, 99%, Cambridge Isotope Laboratories, Tewksbury, MA, USA) as single sources for protein synthesis. The main culture was incubated at 37 °C and 120 rpm, until an OD_{600} of 0.8 was reached; at this point, protein expression was induced with 0.5 mM IPTG. Cultures were incubated at 20°C after induction of expression for 18 h at 120 rpm. The next day, bacterial cells were harvested for 10 min at 6,000 × g (4 °C) using centrifugation. The cell pellet was resuspended in 50 mM Tris (pH 8), 250 mM NaCl buffer supplemented with one cOmplete EDTA-free protease inhibitor tablet (Roche, Basel, Switzerland) and DNAse I (10 ng per gram cell pellet) for nucleic acid digestion. Cells were lysed for 20 min by sonication (60% amplitude, 1 s pulse/ 3 s pause), followed by an additional centrifugation step for 30 min at 45,000 rpm

(4 °C). The lysate was purified by Ni-NTA affinity chromatography using a Protino 5 mL FPLC column (Macherey-Nagel, Düren, Germany). His-tagged Ubl1 was eluted with 50 mM Tris (pH 8), 250 mM NaCl, 500 mM Imidazole. TEV cleavage was combined with dialysis overnight at 4 °C in 50 mM Tris (pH 8), 250 mM NaCl, 1 mM DTT, 0.5 mM EDTA in a ratio of 1:100 (1 protease for 100 target proteins). TEV protease was removed from the digested protein via a second Ni-NTA affinity chromatography step, using the same buffer conditions as in the initial IMAC purification. This time, digested protein was expected in the flow-through. To increase the purity of the sample, a size exclusion chromatography was performed afterward using a HighLoad 16/600 Superdex 75 pg column (Cytiva, Marlborough, Massachusetts, USA) equilibrated in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl (Figure S2A, B). The quality of protein samples was analyzed by SDS-PAGE (Figure S2C). Lastly, the purified protein was concentrated using an Amicon with a cut-off of 3 kDa (Merck Millipore, Darmstadt, Germany).

2.5.2 Circular Dichroism (CD) Spectroscopy

Far-UV-CD spectroscopy of NSP3-Ubl1 protein samples was performed on a JASCO J-815 CD spectropolarimeter (Jasco, Gross-Umstadt, Germany) using 1 mm path length quartz cuvettes (Hellma, Müllheim, Germany). Spectra were recorded at 5 °C, 15 °C, 20 °C, 25 °C and 37 °C using 7.35 µM Ubl1 in 50 mM sodium phosphate (pH 6.5). The following instrument settings were used for the measurement: 0.5 nm step size, 50 nm/min scan speed, 1 nm bandwidth. For noise reduction 10 scans were accumulated each. The mean residue ellipticity $[\theta]_{MRW}$ in units of deg*cm^2*dmol^{-1} at wavelength is given by: $[\theta]_{MRW}$, λ = Ellipticity (mdeg)*10^6 / (Path length (mm) * Protein concentration (µM)*n). And n is the number of peptide bonds, which is the number of amino acids -1. Secondary structure analysis by BeStSel[103] (https://BeStSel.elte.hu/) was applied to all CD data recorded in the 190-250 nm range, using a scale factor of 1.0.

2.5.3 NMR Backbone Resonance Assignment

NMR spectral backbone resonance assignments were performed using a 114 µM ^{15}N-^{13}C-labeled Ubl1 wildtype sample in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl, 5% D_2O, 0.01% NaN_3 buffer, supplemented with 0.2 µM DSS for referencing (Figure S3). Standard three-dimensional HNCO, HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB NMR experiments were acquired at 25 °C on a Bruker 750 MHz Avance III NMR spectrometer (Bruker, Billerica,

MA, USA), equipped with a cryogenic triple resonance ^{1}H, ^{15}N, ^{13}C probe; details on the experimental setup can be found in the Supplementary Information Table S1 and S2. Spectra were processed using the Bruker Topspin 3.5 software. The software CCPN NMR Analysis Assign 3.0. was used to assist with manual backbone resonance assignments[125]. Residue-specific secondary chemical shift data were derived from the extracted backbone SC minus random-coil SC predicted by the POTENCI software[126], considering the Ubl1 wildtype amino acid sequence (Figure S1), experimental temperature, and buffer conditions (section 2.5.1). The NMR resonance assignment (^{1}H, ^{15}N, ^{13}C, and $^{13}C_{\alpha}$ and $^{13}C_{\beta}$) of Ubl1 (1-111) is provided as supplementary Table S6.

2.5.4 Generation of AF3 Model

In addition to the two available X-ray structures of the N-terminal Ubl1 domain of SARS-CoV-2 NSP3, with PDB codes 7KAG and 8XAB, a structural model of Ubl1 was predicted using the AF3 server (https://alphafoldserver.com/) and the Ubl1 amino acid sequence (Figure S1) as the only input.

2.5.5 N-H Residual Dipolar Coupling (RDC) Measurements

For N-H RDC measurements, 100 µM ^{1}H-^{15}N Ubl1 in 50 mM sodium phosphate (pH 6.5), 5% (v/v) D_2O, and 0.01% (v/v) NaN_3 buffer was mixed with Pf1-bacteriophages from strain LP11-92 (Asla-Biotech, Riga, Latvia; https://asla-biotech.com/products/pf1-phage-for-nmr-analysis) for alignment. The final concentration of Pf-1 phages was 20 mg/mL. Alignment was achieved following the procedure, similar to that described by Zweckstetter *et al.* [127]. Samples were recorded on a 1.2 GHz Bruker Avance Neo spectrometer (Bruker, Billerica, MA, USA) at 25 °C. Measurements were performed using a ^{1}H-^{15}N IPAP-HSQC experiment[128]. From that, NH *J+D* couplings were extracted, where *J* represents the *J*-coupling and *D* the residual dipolar coupling (RDC). A reference sample of 100 µM ^{1}H-^{15}N Ubl1 in the absence of phages was measured to determine the N-H *J*-couplings.

From the measured NH RDCs, an alignment tensor was calculated using the program PALES[107]. The two X-ray structures 7KAG, 8XAB, and the structural model of the Ubl1 generated by AF3 (see section 2.5.4) were used as input. Only the structured regions of Ubl1 were included in the calculation of the alignment tensor. Here, only residues with a 600 MHz ^{15}N R_2 relaxation

rate constant higher than $R_2 > 8\ s^{-1}$ were included, resulting in the exclusion of the dynamic residues of the N-terminus A1-C16 and C-terminus D110-E111. Hydrogen atoms were added to the X-ray structure for the calculation of the alignment tensor based on the N-H internuclear vector, using the NMRPipe Dynamo software tool[129].

2.5.6 Analytical Ultracentrifugation (AUC)

Sedimentation velocity (SV) experiments were performed using a Beckman Optima XL-A ultracentrifuge (Beckman Coulter, Brea, CA, USA) equipped with a UV/Vis absorbance detection system and an An-60Ti 4-hole rotor from Beckman Coulter. For the measurement, a concentration of 30 µM Ubl1 in 50 mM sodium phosphate buffer (pH 6.5) with 0.1 mM TCEP was prepared. A reference volume of 400 µL and a sample volume of 390 µL were used in standard double-sector measuring cells, featuring quartz glass windows and aluminum centerpieces with an optical path length of 12 mm. A wavelength of 276 nm was selected for detection, based on pre-measurement wavelength scans conducted at 3,000 rpm. The SV experiment was performed at 20 °C and 60,000 rpm in continuous mode with a radial resolution of 0.003 cm. Radial concentration profiles were measured at approximately 2-minute intervals until the protein was fully sedimented. Sedimentation data was analyzed using the continuous distribution *c*(*s*) Lamm equation model (*c*(*s*) model) of the software package SEDFIT Version 16.1c[130]. A resolution of 0.1 S and a confidence level (*F*-ratio) of 0.95 were selected for data analysis within the *s*-value range of 0 – 15 S. The RMSD value of the fit was 0.005, accounting for less than 1% of the total signal. The partial specific volume of Ubl1 (vbar = 0.73586 $cm^3\ g^{-1}$) was calculated based on the amino acid sequence (Figure S1) using the software SEDNTERP Version 2.0 Beta[131]. Additionally, solvent density (ρ = 1.0036 g/cm^3) and viscosity (η = 0.01022 P) were calculated from tabulated data sets of the same software. Standardization of measured *s*-values to water at 20 °C ($s_{20,w}$) was performed. Further analysis and graphical visualization of the sedimentation data were achieved using the software GUSSI[132] and DataGraph Version 5.3β, copyright 2020 (North Carolina, USA).

2.5.7 Small-Angle X-Ray Scattering (SAXS)

SAXS data were measured with the laboratory-based SAXS system “Ganesha-Air” from SAXSLAB/XENOCS at Forschungszentrum Jülich. The X-ray source is a D2-MetalJet (Excillum) with a liquid metal anode operating at 70 kV and 3.57 mA with Ga-Kα radiation (wavelength

λ = 0.1314 nm), providing a brilliant and small beam (< 200 µm). A PILATUS 300K (Dectris) was used to record the 2D scattering image. All samples were measured in sealed glass capillaries of 2 mm inner diameter. Images were circular averaged, normalized to incident intensity and transmission corrected. The covered Q range was 0.11-4.9 nm^{-1} measured at a detector distance of 187 cm. A dummy layer of solvent atoms represents the hydration layer of 0.3 nm thickness and a hydration layer density 5% larger than the bulk solvent. To exclude structure factor effects, concentrations of 3, 5, and 8 mg/mL were measured, and the corresponding buffer was subtracted as background. Protein form factors *P(Q)* were calculated based on protein PDB structures with atomic positions r_i and respective form factor $f_i'(Q)$ as

$$P(Q) = \langle \sum_{i,j} f_i'(Q) f_j'(Q) e^{-Q(r_i - r_j)} \rangle$$

In a solvent, according to Babinet's principle, we subtract from atomic form factors $f_{j,a}(Q)$ the excluded volume scattering $f_{j,d}(Q)$ of the solvent filled with dummy atoms yielding $f_j'(Q) = f_{j,a}(Q) - f_{j,d}(Q)$. All calculations are done with the program Jscatter.[133]

2.5.8 NMR Relaxation Data

NMR relaxation experiments were performed at 25 °C on 100 µM ^{1}H-^{15}N Ubl1 in 50 mM sodium-phosphate (pH 6.5), 0.1 mM TCEP, 5% (v/v) D_2O, and 0.01% (v/v) NaN_3 buffer, using a sensitivity-enhanced HSQC-detection scheme (https://bmrbig.bmrb.io/released/bmrbig102) [134,135,136]. Field-dependent NMR experiments were acquired on three different Bruker NMR spectrometers with ^{1}H Larmor frequencies of 600 MHz (Avance III), 900 MHz (Avance Neo), and 1.2 GHz (Avance Neo). Details on the experimental setup can be found in Supporting Information Table S3 (1.2 GHz), S4 (900 MHz) and S5 (600 MHz). NMR ^{15}N relaxation rate constants are shown in Table S7. The NMRPipe software was used to process the spectra[129].

2.5.9 Model-Free Analysis of Ubl1

An in-depths analysis of the Ubl1 structural dynamics was performed with a model-free analysis of the NMR relaxation data, using the *relax* 5.1.0 software[108,109]. Nine datapoints (R_1, R_2 and hetNOE values from 600 MHz, 900 MHz and 1200 MHz) were used as input for the analysis. *Relax* needs several optimization steps to decide on the best diffusion model and

residue-specific Model-free equation for fitting the data using an AIC criterion. The following diffusion models were tested: Sphere, Prolate Spheroid, Oblate Spheroid, and Ellipsoid. Two "Model-free" approaches were tested for each residue, the original Lipari and Szabo model[137] [S^2, τ_e] and the 1990 published extended model-free model[138] [S^2, τ_f, S_f^2, τ_s], both in the absence of chemical exchange relaxation to avoid miss-fitting of the anisotropic diffusion tensor[139]. The default parameters were used for the grid search increments (11), the Monte Carlo simulation number (500), and the maximum iterations (50). The chemical shift anisotropy was set to 172 ppm, and the N-H bond vector length was set to 1.02 Å. A fully automated GUI-based analysis was performed using NMRbox[140]: National Center for Biomolecular NMR Data Processing and Analysis, a Biomedical Technology Research Resource (BTRR), which is supported by NIH grant P41GM111135 (NIGMS). Three independent model-free analysis runs were performed, using either the Ubl1 (AF3) model or the X-ray structures 7KAG or 8XAB for diffusion tensor calculation.

2.5.10 Statistics

2.5.10.1 NMR Spectroscopy

The noise level of the various NMR spectra was taken as the error of the NMR spectral intensities. Gaussian error propagation was performed where needed. The error in the NMR relaxation rate constants was estimated using the NMRPipe software package[129] and dedicated scripts for calculating the NMR relaxation rate constants, available under https://bmrbig.bmrb.io/released/bmrbig102. Briefly, NMRPipe applies a Monte Carlo Simulation to generate a distribution of NMR relaxation rate constants, from which the standard deviation is taken as experimental error on the relaxation rates. For additional information, we refer to references [134,135,136]). The propagated error on the derived order Model-free order parameters and correlation times was calculated via the relax software, which also applies a Monte-Carlo algorithm for error estimation. For further details, we refer to [108].

2.6 Data Availability

NMR resonance assignment (^{1}H, ^{15}N, ^{13}C, and $^{13}C_{\alpha}$ and $^{13}C_{\beta}$) of Ubl1 (1-111) and NMR relaxation data are available in the supplementary information files. All data presented in this study are available from the corresponding author upon request.

2.7 Acknowledgments

We thank Filip Hasecke, Sophia Werner, Andrea Janosch, Robin Backer, Manuel Etzkorn, Wolfgang Hoyer, and Lothar Gremer for helpful discussions and Michelle Billig for initial help during the project. N.L. thanks the German Science Foundation for funding through the Heisenberg Program (DFG grant number 433700474) and the project “Virological and immunological determinants of COVID-19 pathogenesis – lessons to get prepared for future pandemics (KA1-Co-02 “COVIPA”), a grant from the Helmholtz Association's Initiative and Networking Fund. Work of R.Ba. was supported by the project “Virological and immunological determinants of COVID-19 pathogenesis – lessons to get prepared for future pandemics (KA1-Co-02 “COVIPA”), a grant from the Helmholtz Association's Initiative and Networking Fund. We acknowledge generous access to the Jülich-Düsseldorf Biomolecular NMR Center, jointly run by Forschungszentrum Jülich and Heinrich Heine University Düsseldorf (HHU).

2.8 Author Contributions

Conceptualization: N.L., R.Ba., R. Bi., L.N.S., Methodology: K.V., T.S., F.T., L.N.S., A.C., R.B., N.L., Resources (sample preparation): K.V., A.C., Investigation: K.V., T.S., F.T., L.N.S., R.B., N.L., Formal Analysis: K.V., T.S., F.T., L.N.S., R.B., N.L., Writing: K.V., T.S., F.T., L.N.S., R.Ba., R.B., N.L., Funding: N.L., R.Ba., Supervision: L.N.S., N.L.

2.9 Supplementary Information

```
          10        20        30        40        50        60
APTKVTFGDD TVIEVQGYKS VNITFELDER IDKVLNEKCS AYTVELGTEV NEFACVVADA
          70        80        90       100       110
VIKTLQPVSE LLTPLGIDLD EWSMATYYLF DESGEFKLAS HMYCSFYPPD E
```

Figure S1. The amino acid sequence of NSP3-Ubl1. Organism: *Sars-CoV-2,* NCBI genome entry NC_045512.2, 111 amino acids, 12.46 kDa. Amino acid residues with NMR backbone assignments are colored in green (103 out of 105 expected resonances).

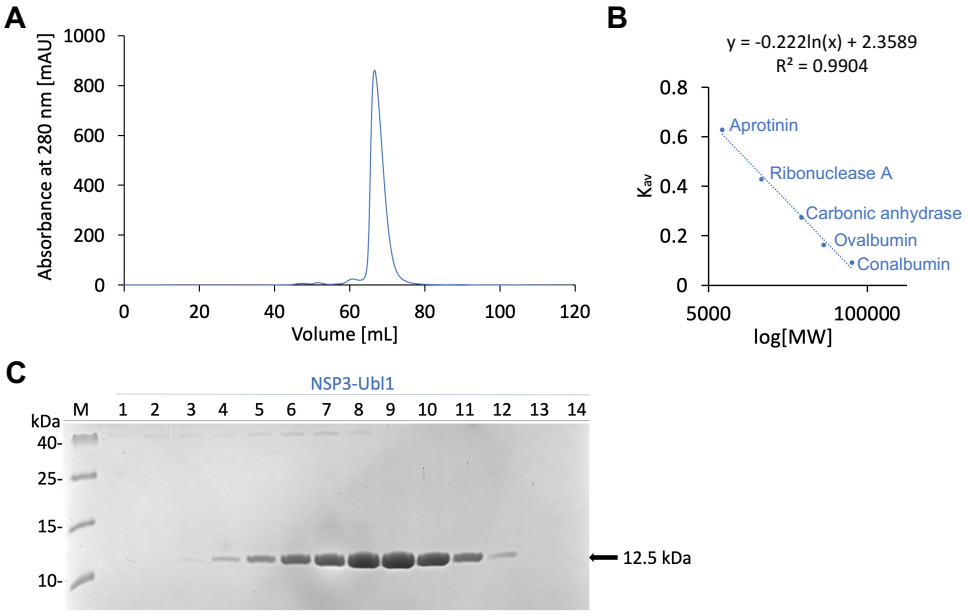


Figure S2: **Size Exclusion Chromatography (SEC) of ^{1}H ^{15}N Sars-CoV-2 NSP3a-Ubl1 (1-111). Size Exclusion Chromatography (SEC) of ^{1}H ^{15}N Sars-CoV-2 NSP3a-Ubl1 (1-111).** For SEC, a HighLoad 16/600 Superdex 75 pg column (Cytiva, Marlborough, Massachusetts, USA) equilibrated in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl was used. **(A)** Chromatogram of the isolation. The absorbance (mAU) is plotted against the volume (mL). **(B)** SEC calibration curve based on the following standard proteins: Aprotinin (6.5 kDa), Ribonuclease A (13.7 kDa), Carbonic Anhydrase (29 kDa), Ovalbumin (44 kDa) and Conalbumin (75 kDa). The calibration curve was plotted using the gel-phase distribution coefficient K_{av} vs. the logarithm of the molecular weight (MW) in g/mol. Based on the calibration curve, the calculated molecular weight of Ubl1 is 11.56 kDa (12.46 kDa expected), indicating monomeric Ubl1. **(C)** SDS-PAGE of collected SEC fractions (1-14). Bands correlating to monomeric Ubl1 (12.46 kDa) are indicated by an arrow.

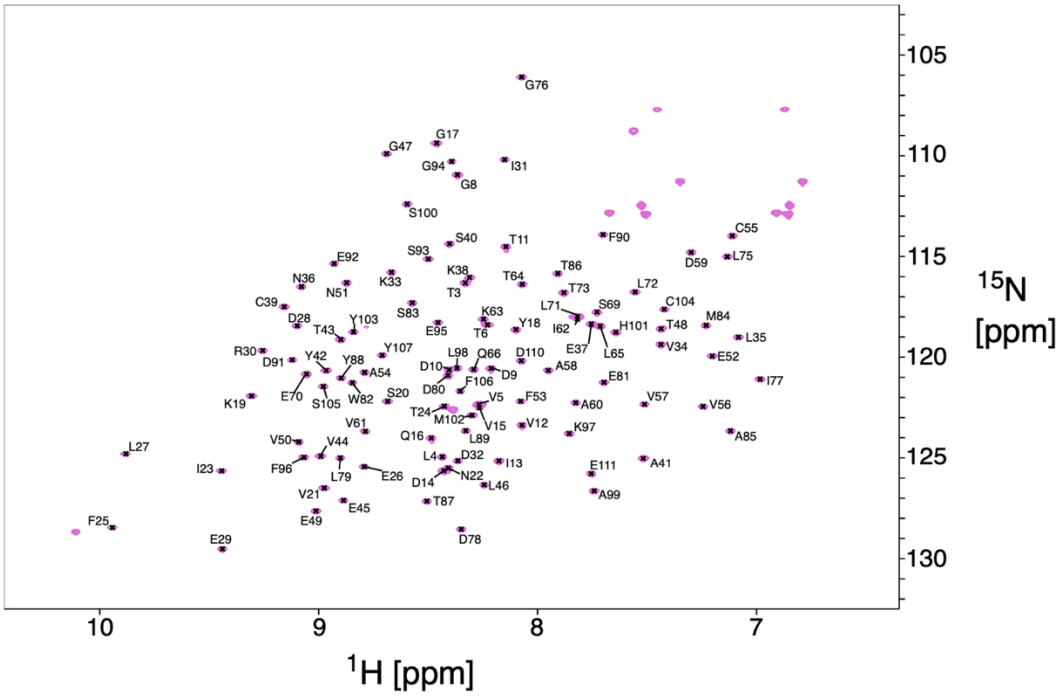


Figure S3: **NMR spectral backbone assignment of Sars-CoV-2 NSP3-Ubl1 (1-111).** The spectrum was acquired on a 900 MHz spectrometer at 25 °C and a concentration of 114 µM in 50 mM sodium

phosphate (pH 6.5), 150 mM NaCl. The sample was referenced to DSS. The assigned backbone amid groups are labelled.

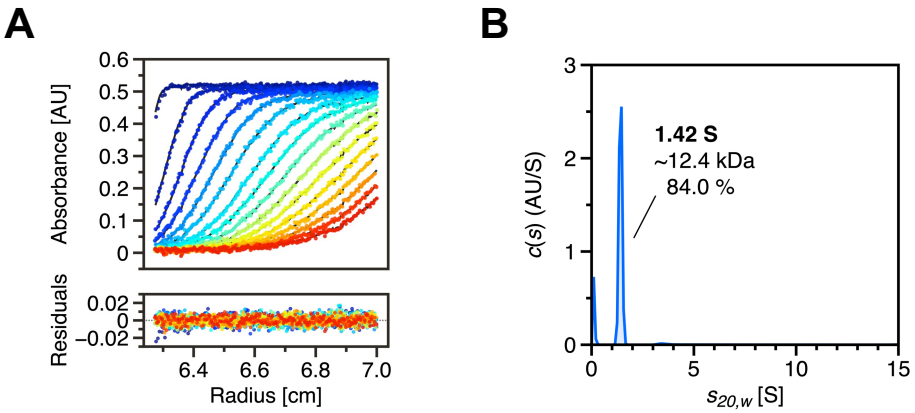


Figure S4: Sedimentation coefficient distribution of NSP3-Ubl1 determined by SV-AUC analysis. (A) Freshly prepared samples of 30 µM NSP3-Ubl1 in 50 mM sodium phosphate buffer, pH 6.5 with 0.1 M TCEP were sedimented at 60,000 rpm and 20 °C, with absorbance detected at 276 nm. Sedimentation profiles, i.e. raw data (dots), with overlaid fits (lines) are shown in the upper graph, while the deviations between fit and raw data are shown as residuals in the lower graph. **(B)** The continuous distribution *c*(*s*) Lamm equation model was applied to determine the distribution of sedimentation coefficients for NSP3-Ubl1. Calculated *s*-values were standardized to $s_{20,w}$ values.

Table S1: Experimental parameters of three-dimensional ^{13}C-^{15}N-Ubl1 NMR backbone assignment experiments.

	HNCO	HNCA	HN(CO)CA	HNCACB	HN(CO)CACB
Field [T]	17.6	17.6	17.6	17.6	17.6
^{1}H Larmor frequency [MHz]	750.196	750.196	750.196	750.196	750.196
Bruker pulse program	hncogp3d	hncagp3d	hncocagp3d	hncacbgp3d	hncocacbgp3d
Interscan recovery delay [s]	1	1	1	1	1
Number of scans	16	16	16	16	16
Total experimental time [h]	37.3	47.7	48.5	70.6	71.5
Direct dimension (^{1}H)					
^{1}H carrier [ppm]	4.70	4.70	4.70	4.70	4.70
Spectral width [ppm]	16.02	16.02	16.02	16.02	16.02
Data points (TD)	2048	2048	2048	2048	2048
Acquisition time [ms]	85.20	85.20	85.20	85.20	85.20
Indirect dimension (^{15}N)					
^{15}N carrier [ppm]	117.00	117.00	117.00	117.00	117.00
Spectral width [ppm]	35.00	35.00	35.00	35.08	35.00
Data points (TD)	100	100	100	100	100
Acquisition time [ms]	18.79	18.79	18.79	18.75	18.79
Indirect dimension (^{13}C)					
^{13}C carrier [ppm]	173.50	53.20	53.20	43.00	43.00
Spectral width [ppm]	12.05	31.18	31.18	70.68	70.68
Data points (TD)	70	90	90	130	130
Acquisition time [ms]	15.40	7.65	7.65	4.88	4.88

Assignment statistics					
# visible resonances	105	105	105	105	105
# assigned resonances	103	103	103	103	103

Table S2: Experimental parameters of two-dimensional ^{1}H-^{15}N Ubl1 HSQC experiment.

	HSQC
Field [T]	21.1
^{1}H Larmor frequency [MHz]	899.84
Bruker pulse program	fhsqcf3gpph
Interscan recovery delay [s]	1
Number of scans	8
Total experimental time [h]	0.9
Direct dimension (^{1}H)	
^{1}H carrier [ppm]	4.70
Spectral width [ppm]	16.34
Data points	2048
Acquisition time [ms]	69.63
Indirect dimension (^{15}N)	
^{15}N carrier [ppm]	117.00
Spectral width [ppm]	35.00
Data points	256
Acquisition time [ms]	40.11

Table S3: Experimental parameters of ^{15}N-Ubl1 NMR relaxation experiments at 1200 MHz.

	^{15}N R_1	**^{15}N $R_{1\rho}$**	**{^{1}H}-^{15}N NOE**	**^{15}N R_{2Hahn}**
Field [T]	28.2	28.2	28.2	28.2
1H Larmor frequency [MHz]	1199.92	1199.92	1199.92	1199.92
Pulse program	Stief *et al.*, Methods 2024	Stief *et al.*, Methods 2024	Lakomek *et al.*, Journal of Biomolecular NMR 2012	Stief *et al.*, JMB 2023
Interscan recovery delay [s]	2.00	2.00	2.00	2.00
Number of scans	8	8	32	8
Total experimental time [h]	23.8	20.5	66.3	21.0
Direct dimension (^{1}H)				
^{1}H carrier [ppm]	4.70	4.70	4.70	4.70
Spectral width [ppm]	16.03	16.03	16.03	16.03
Data points	4096	4096	4096	4096
Acquisition time [ms]	106.50	106.50	106.50	106.50
90° (^{1}H) pulse length [µs]	8.45	8.45	8.45	8.45
Indirect dimension (^{15}N)				
^{15}N carrier [ppm]	118.00	118.00	118.00	118.00
Spectral width [ppm]	35.00	35.00	35.00	35.00
Data points (total)	8	8	2	2
Acquisition time [ms]	60.16	60.16	60.16	60.16

90° (^{15}N) pulse length [µs]	34.8	34.8	34.8	34.8
Relaxation period				
# relaxation delays	8	8	2	2
Time points [ms]	0, 960, 160, 800, 80, 560, 240, 400	5, 65, 25, 45, 15, 55, 35, 10	5 s saturation on ^{1}H	1, 111

Table S4: Experimental parameters of ^{15}N-Ubl1 NMR relaxation experiments at 900 MHz.

	^{15}N R_1	**^{15}N $R_{1\rho}$**	**{^{1}H}-^{15}N NOE**	**^{15}N R_{2Hahn}**
Field [T]	21.1	21.1	21.1	21.1
1H Larmor frequency [MHz]	899.84	899.84	899.84	899.84
Pulse program	Stief *et al*., Methods 2024	Stief *et al*., Methods 2024	Lakomek *et al*., Journal of Biomolecular NMR 2012	Stief *et al*., JMB 2023
Interscan recovery delay [s]	3.00	3.00	2.00	3.00
Number of scans	8	8	32	32
Total experimental time [h]	32.3	29.4	66.1	30.5
Direct dimension (^{1}H)				
^{1}H carrier [ppm]	4.70	4.70	4.70	4.70
Spectral width [ppm]	16.34	16.34	16.34	16.34
Data points	2048	2048	2048	2048
Acquisition time [ms]	69.63	69.63	69.63	69.63
90° (^{1}H) pulse length [µs]	8.05	8.05	8.10	8.10
Indirect dimension (^{15}N)				
^{15}N carrier [ppm]	117.00	117.00	117.00	117.00
Spectral width [ppm]	36.55	36.55	36.55	36.55
Data points (total)	8	8	2	2
Acquisition time [ms]	76.8	76.8	76.8	76.8
90° (^{15}N) pulse length [µs]	42.50	42.50	42.50	42.50
Relaxation period				
# relaxation delays	8	8	2	2
Time points [ms]	0, 880, 80, 720, 160, 480, 240, 320	5, 55, 25, 20, 15, 45, 35, 10	5 s saturation on ^{1}H	1, 111

Table S5: Experimental parameters of ^{15}N-Ubl1 NMR relaxation experiments at 600 MHz.

	^{15}N R_1	**^{15}N $R_{1\rho}$**	**{^{1}H}-^{15}N NOE**	**^{15}N R_{2Hahn}**
Field [T]	14.1	14.1	14.1	14.1
1H Larmor frequency [MHz]	600.19	600.19	600.19	600.19
Pulse program	Stief *et al*., Methods 2024	Stief *et al.*, Methods 2024	Lakomek *et al.*, Journal of Biomolecular NMR 2012	Stief *et al*., JMB 2023
Interscan recovery delay [s]	2.00	2.00	2.00	2.00
Number of scans	8	8	32	8

Total experimental time [h]	1.1	20.8	66.6	5.6
Direct dimension (^{1}H)				
^{1}H carrier [ppm]	4.70	4.70	4.70	4.70
Spectral width [ppm]	16.02	16.02	16.02	16.02
Data points	2048	2048	2048	2048
Acquisition time [ms]	106.50	106.50	106.50	106.50
90° (^{1}H) pulse length [µs]	9.05	9.05	9.05	9.05
Indirect dimension (^{15}N)				
^{15}N carrier [ppm]	117.00	117.00	117.00	117.00
Spectral width [ppm]	34.98	34.98	34.98	34.98
Data points (total)	8	8	2	2
Acquisition time [ms]	120.32	120.32	120.32	120.32
90° (^{15}N) pulse length [µs]	42.20	42.20	42.20	42.20
Relaxation period				
# relaxation delays	8	8	2	2
Time points [ms]	0, 720, 160, 640, 80, 560, 320, 480	5, 65, 45, 15, 55, 35, 10	5 s saturation on ^{1}H	8.89, 133.32

Table S6: NMR backbone chemical shift assignment for Ubl1.
Data are provided in the following excel sheet: "Table_S6_Ubl1_assignment"

Table S7: NMR ^{15}N relaxation rate constants.
Data are provided in the following excel sheet: "Table_S7_Ubl1_relaxation"

10 20 30 40 50 60
APTKVTFGDD TV I EVQGYKS VNITFELDER IDKVLNEKCS AYTVELGTEV NEFACVVADA
APIKGVTFGED TVWEVQGYKN VRITFELDER VDKVLNEKCS VYTVESGTEV TEFACVVAEA

70 80 90 100 110
V I KTLQPVSE LLTP L GIDLD EWSMATYYLF DE SGEFKLAS HMYCSFYPPD E
VVKTLQPVSD LLTNMGIDLD EWS VAT FYLF DDAGEENFSS RMYCSFYPPD E

Figure S5: Sequence comparison of NSP3-Ubl1 from Sars-CoV-1 and Sars-CoV-2. Amino acid sequences are shown in the one-letter amino acid code. Non-conserved residues are highlighted. Sequences are shown in purple for Sars-CoV-1 Ubl1 and black for Sars-CoV-2 Ubl1. Sequence identity was analysed by using Protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Homologous Ubl1 proteins have shown a sequence identity of 79.28 % indicating that 88 of 111 amino acids are conserved.

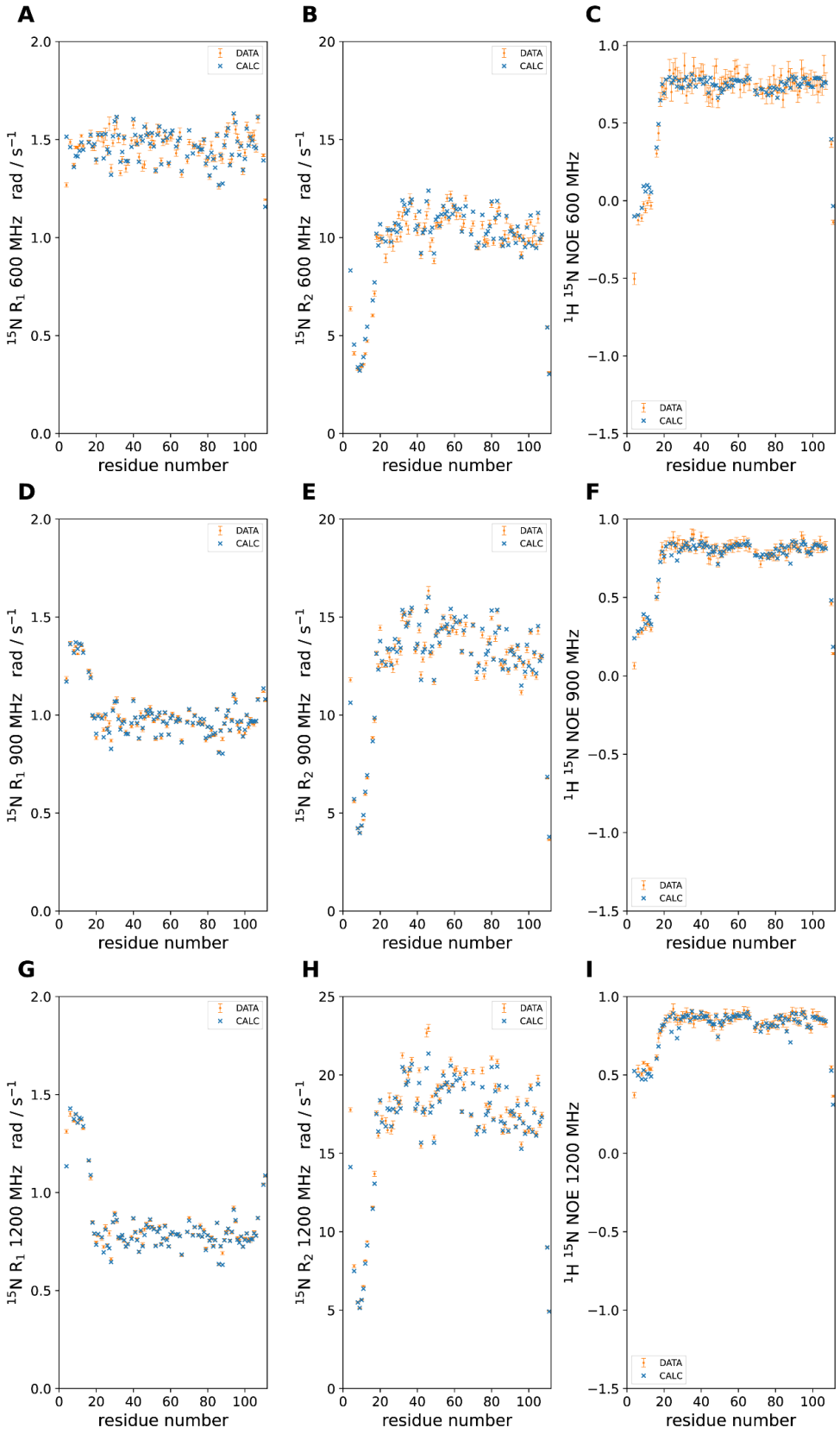


Figure S6: Back-calculated NMR ^{15}N relaxation rate constants based on the model free parameters of the AF3 structure. Experimental data are colored in orange, back-calculated data in blue color. **(A)** R_1 relaxation rates at 600 MHz. **(B)** R_2 relaxation rates at 600 MHz. **(C)** hetNOE values at 600 MHz. **(D)** R_1 relaxation rates at 900 MHz. **(E)** R_2 relaxation rates at 900 MHz. **(F)** hetNOE values at 900 MHz. **(G)** R_1 relaxation rates at 1200 MHz. **(H)**: R_2 relaxation rates at 1200 MHz. **(I)** hetNOE values at 1200 MHz.

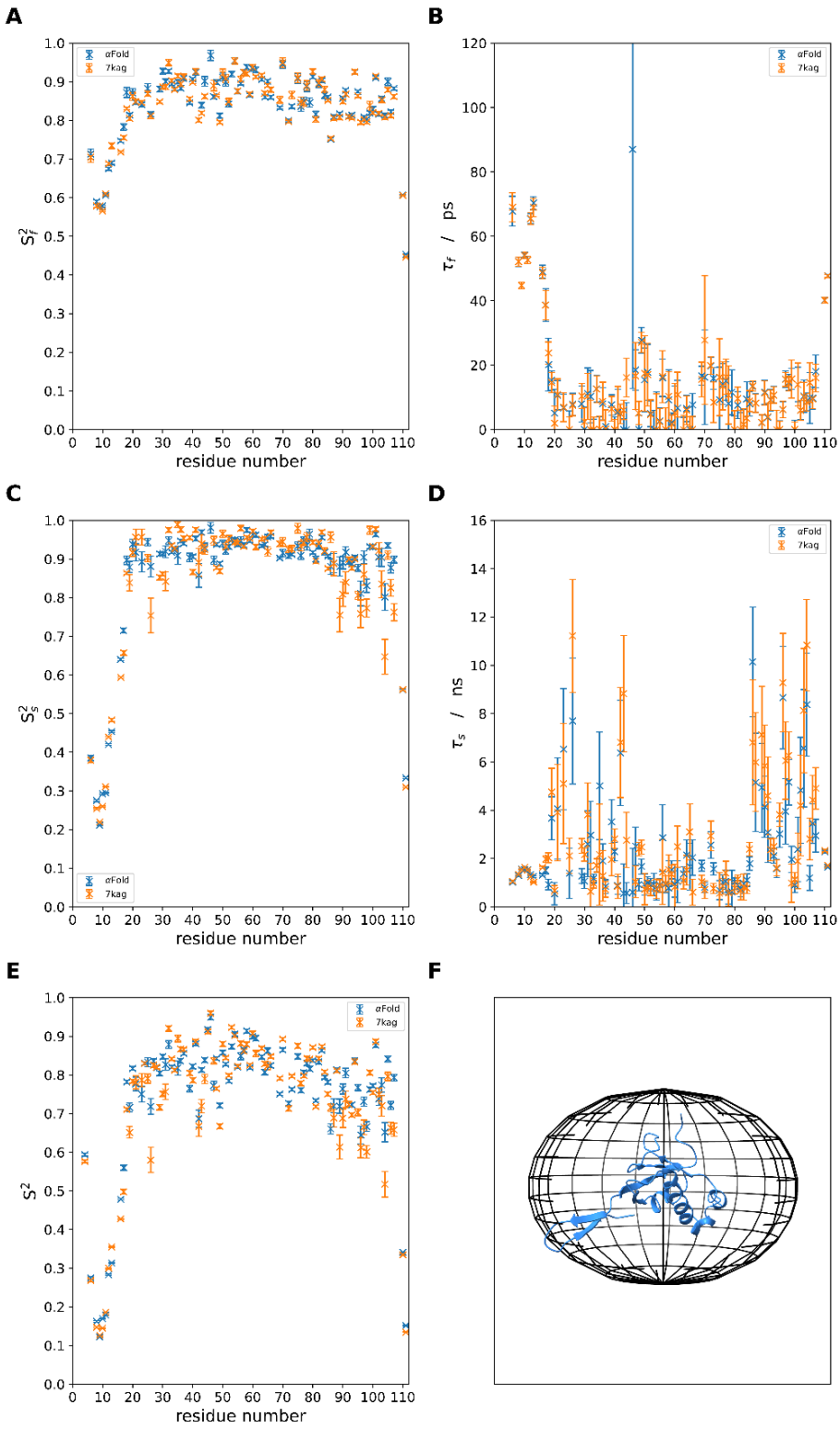


Figure S7: Comparison of the model-free analyses of Ubl1 using structure 7KAG or the AlphaFold3 structural model for diffusion tensor calculation. Data derived using the Ubl1 AlphaFold3 (AF3) structural model for diffusion tensor calculation are colored in blue, those using the X-ray structure 7KAG as input for diffusion tensor calculation are colored in orange. **(A)** Order parameters S_f^2 characterizing the amplitudes of fast internal motions. **(B)** Correlation times τ_f of fast internal motions. **(C)** Order parameters characterizing the amplitudes S_s^2 of slow internal motions. **(D)** Correlation time τ_s of slow internal motions. **(E)** General order parameter $S^2 = S_s^2 * S_f^2$. **(F)** Diffusion tensor versus structure alignment.

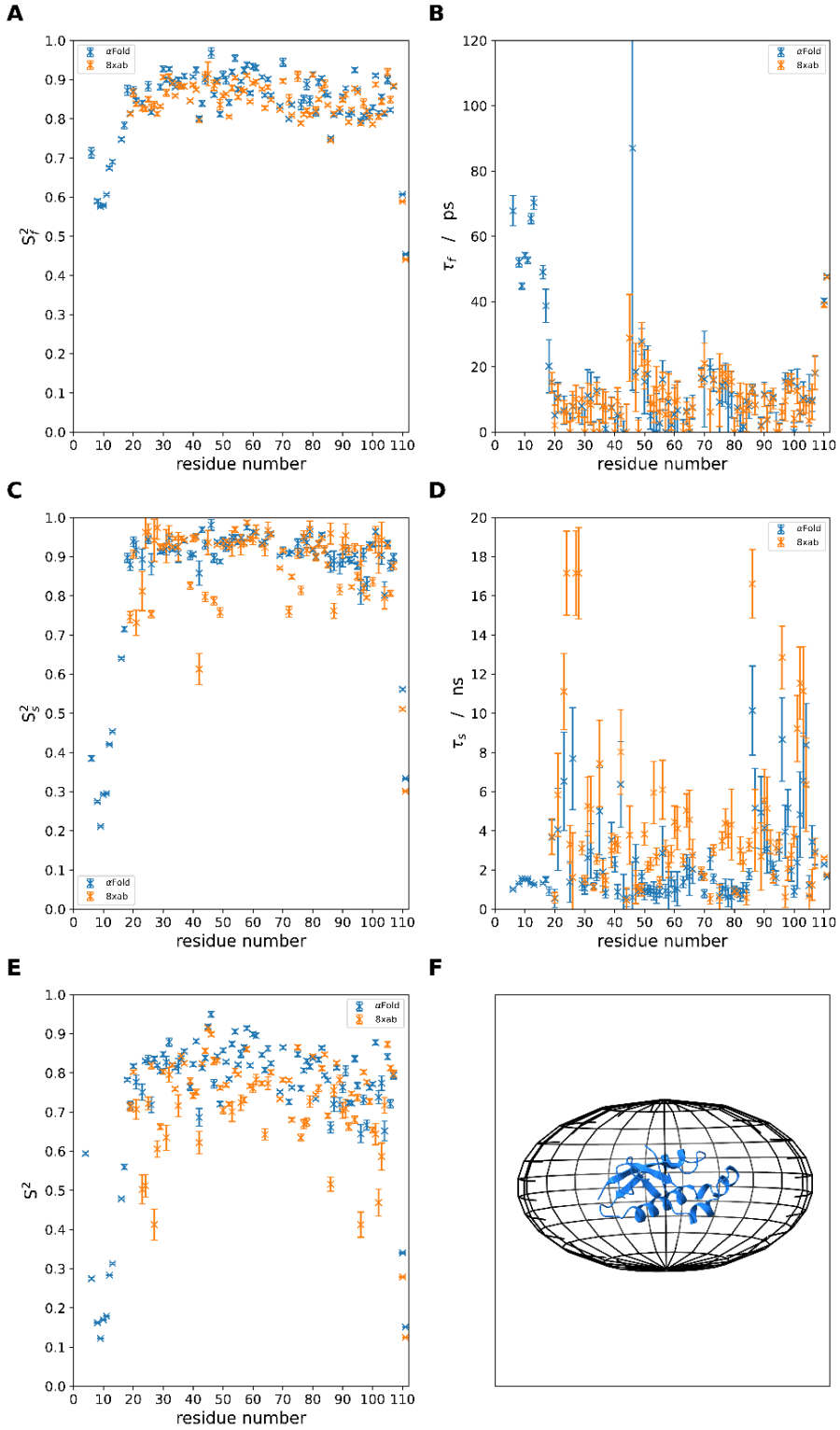


Figure S8: Comparison of the model-free analyses of Ubl1 using structure 8XAB or the AlphaFold3 structural model for diffusion tensor calculation. Data derived using the Ubl1 AlphaFold3 (AF3) structural model for diffusion tensor calculation are colored in blue, those using the X-ray structure 8XAB as input for diffusion tensor calculation are colored in orange. **(A)** Order parameters S_f^2 characterizing the amplitudes of fast internal motions. **(B)** Correlation times τ_f of fast internal motions. **(C)** Order parameters characterizing the amplitudes S_s^2 of slow internal motions. **(D)** Correlation time τ_s of slow internal motions. **(E)** General order parameter $S^2 = S_s^2 * S_f^2$. **(F)** Diffusion tensor versus structure alignment.

Table S8: Diffusion tensors fitted on the NMR ^{15}N relaxation data using the relax software nomenclature.

3D structure	**AF3 model**
Diffusion type	Spheroid (axially symmetric, prolate tensor)
τ_M s / rad	8.36e-09
D_{iso} rad / s	1.99e+07
D_a rad / s	6.49e+06
D_{par} rad / s	2.43e+07
D_{per} rad / s	1.78e+07
D_{ratio}	1.36
θ rad	1.547
ϕ rad	2.873
3D structure	**7kag**
Diffusion type	Spheroid (axially symmetric, oblate tensor)
τ_M s / rad	8.55-09
D_{iso} rad / s	1.95e+07
D_a rad / s	-5.72e+06
D_{par} rad / s	1.57e+07
D_{per} rad / s	2.14e+07
D_{ratio}	0.73
θ rad	2.152
ϕ rad	2.024
3D structure	**8xab**
Diffusion type	Spheroid (axially symmetric, oblate tensor)
τ_M s / rad	8.58-09
D_{iso} rad / s	1.94e+07
D_a rad / s	-9.30e+06
D_{par} rad / s	1.32e+07
D_{per} rad / s	2.25e+07
D_{ratio}	0.59
θ rad	0.82
ϕ rad	2.556

Table S9: Extended model-free analysis of NMR ^{15}N relaxation data of Ubl1. Residue-specific order parameters and correlation times of the model-free analysis of Ubl1 using the AlphaFold3 (AF3) structural model of Ubl1 for diffusion tensor calculation are provided in the EXEL file "Modelfree_results_Ubl1_AF3"

Table S10: Structural comparison of the SARS-CoV-1 and SARS-CoV-2 Ubl1 domain. Structural information derived from the Protein Data Bank (PDB) based on an X-Ray (7KAG, Stogios *et al.,* 2020) and NMR-structure (2IDY, Serrano *et al.,* 2007). For comparison, a structural model was predicted by AlphaFold3 (AF3).

	Ubl1 SARS-CoV-2			**Ubl1 SARS-CoV-1**	
Data	7KAG	This work	AF3	2IDY	AF3
α-helices	4	4	4	3	4
β-strands	6	4	4	4	6
N-terminus	folded	disordered	disordered	disordered	folded

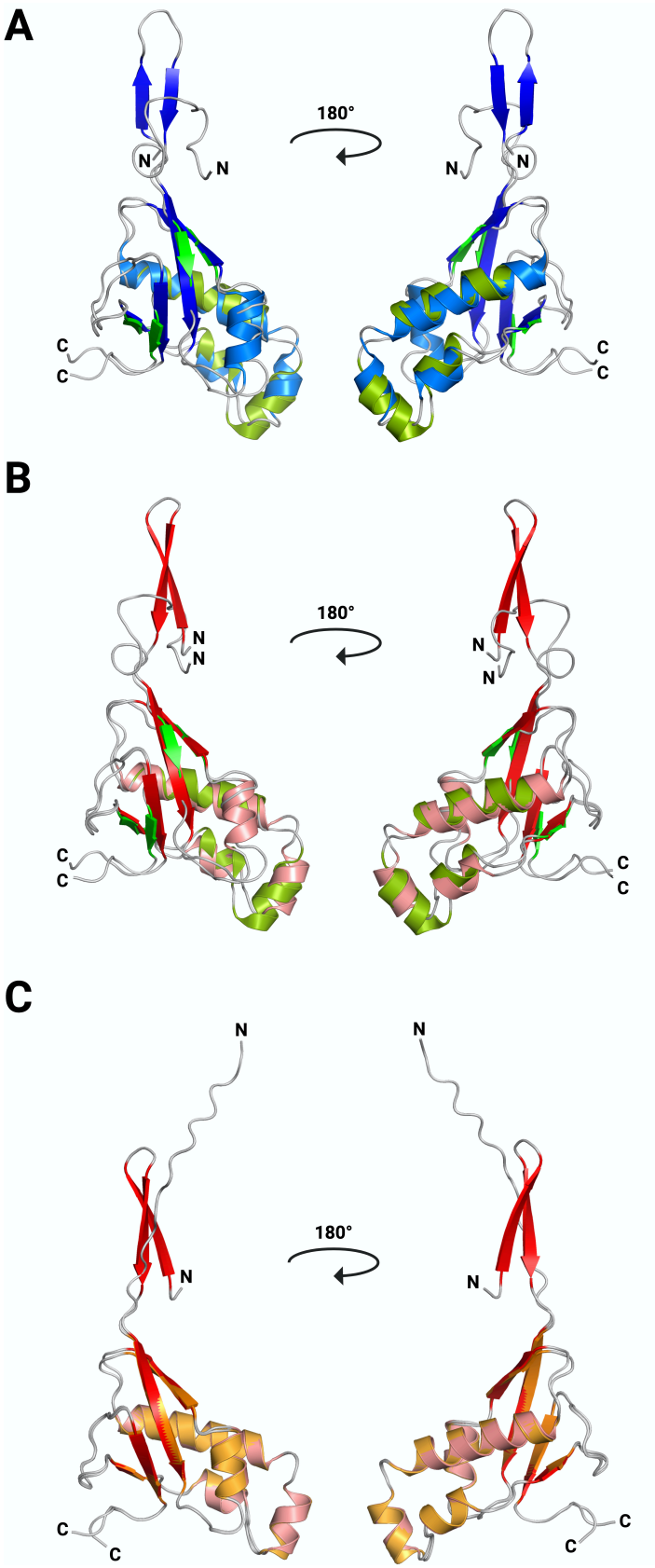


Figure S9: Overlay of Ubl1 domain structures from SARS-CoV-1 and SARS-CoV-2. (A) Structural alignment of 7KAG (SARS-CoV-2, blue) and 2IDY (SARS-CoV-1, green); RMSD = 1.854. **(B)** Alignment of 2IDY (green) with the corresponding AF3 model (red); RMSD = 1.854. **(C)** Comparison of AF3-predicted Ubl1 structures from SARS-CoV-2 (orange) and SARS-CoV-1 (red); RMSD = 0.346. All RMSD values were calculated using PyMOL[85].

3. Chapter: Impact of Ionic Strengths on the Self-Association of the SARS-CoV-2 Non-Structural Protein 3 N-terminal Domain

Journal: in preparation

Authors: **Katharina Vormann**,
Tobias Stief,
Florian T. Tucholski,
Luitgard Nagel-Steger,
Ralf Biehl,
Nils-Alexander Lakomek

Author contribution: Expression and isolation of ^{1}H-^{15}N labeled NSP3-Ubl1,
Analysis of 2D NMR data,
Analysis of NMR relaxation data,
Protein structure analysis,
Acquisition and analysis of DLS data,
Writing and proof-reading of manuscript

3.1 Abstract

The N-terminal Ubl1 domain of SARS-CoV-2 NSP3 plays a pivotal role in viral replication. Here, we investigated the structural dynamics and oligomerization behavior of Ubl1 under varying ionic conditions. ^{15}N NMR relaxation measurements revealed that in the absence of salt, Ubl1 exhibits increased mobility, characterized by low R_2 relaxation rates and minimal concentration dependence. In contrast, addition of 150 mM NaCl led to elevated R_2 rates and decreased molecular mobility, with a clear concentration-dependent effect, suggesting salt-induced molecular crowding. AUC and NMR data confirmed a monomeric state of Ubl1 regardless of redox conditions, excluding disulfide-mediated dimerization. However, small-angle X-ray scattering (SAXS) analysis indicated a minor dimeric population (~20%) under high-salt conditions. These findings support a model in which ionic strength modulates Ubl1 dynamics via charge shielding and weak, reversible self-association, contributing to its behavior in the crowded intracellular environment of SARS-CoV-2-infected cells.

3.2 Introduction

SARS-CoV-2 is a pathogenic beta coronavirus and the causative agent of the respiratory disease COVID-19. Due to its rapid global spread, there is an urgent need to identify and develop effective antiviral inhibitors. A key target in this context is the multi-domain NSP3, a crucial component of the viral replication machinery.[45,64,102] With more than 1900 amino acid residues, NSP3 is the largest protein encoded by the SARS-CoV-2 genome and consists of 16 distinct domains.[141] The N-terminal domain of NSP3, referred to as NSP3a, includes a ubiquitin-like fold and an intrinsically disordered, glutamate-rich acidic region.[102] NSP3 is essential for the formation of the SARS-CoV-2 replication organelles, a membranous structure supporting viral genome replication. It is also a major constituent of the molecular pores located in DMVs, proposed to facilitate RNA transport across the DMV bilayer.[45,56] Recent studies have demonstrated that NSP3-Ubl1 interrupts phase separation of the viral N-protein with RNA in a concentration-dependent manner.[61,84] The N-protein and viral RNA co-assemble into biomolecular condensates that are critical for viral assembly.[59,61,84,119,122,124,142] These condensates form via liquid–liquid phase separation (LLPS), creating dynamic, membrane-less compartments that locally concentrate proteins and nucleic acids.[143,144,145] LLPS is often driven by electrostatic interactions, a phenomenon known as complex coacervation, which governs the behavior of intracellular biomolecular condensates.[146] Electrostatic attraction between the negatively charged RNA phosphate backbone and positively charged residues in RNA-binding proteins is a key driver of LLPS.[146] Banerjee *et al.* demonstrated that increasing NaCl concentration disrupts RNA-peptide condensates, with complete dissolution observed at 225 mM NaCl, attributed to electrostatic screening effects.[143] Additional studies confirmed that salt concentration modulates SARS-CoV-2 N-protein phase behavior, with higher ionic strength reducing protein incorporation into liquid droplets.[118,124] This behavior is consistent with models in which charge screening and pH-dependent net charge influence LLPS.[143,147] Moreover, binding affinity between the N-protein (net charge +24 at pH 7.4) and nucleotides was enhanced at 50 mM NaCl, supporting a charge-regulated interaction mechanism.[118] Phase separation of intrinsically disordered regions (IDRs) is further promoted by RNA, molecular crowding agents, and moderate salt concentrations under *in vitro* conditions.[148] Despite the dense and viscous nature of these condensates, IDPs maintain nanosecond-scale internal dynamics, similar to those in dilute solutions.[149] However, reduced molecular mobility within condensates has been observed upon salt addition, emphasizing the critical role of the

surrounding aqueous environment in modulating LLPS behavior.[148] We hypothesize that condensates formed by N protein, RNA, and NSP3-Ubl1 promote increased viral replication and transcription efficiency within infected cells. However, structural and oligomerization data for NSP3 remain limited. In this context, our ongoing *in vitro* investigations aim to elucidate the structural and dynamic properties of NSP3, particularly the Ubl1 domain, using NMR spectroscopy, AUC, dynamic light scattering (DLS), and SAXS. These studies seek to provide deeper insight into the molecular mechanisms by which NSP3 contributes to the SARS-CoV-2 replication cycle.

3.3 Material and Methods

3.3.1 Sample Preparation of ^{1}H ^{15}N Ubl1

The gene base sequence encoding NSP3-Ubl1 (1–111), from *SARS-CoV-2* (111 amino acids, 12.46 kDa, pI: 4.08) was codon-optimized for expression in *E. coli* and subcloned in a pET28a(+) vector with Kanamycin resistance. To facilitate protein isolation, a *hexa*-histidine tag was included at the N-terminus. Additionally, a *Tabacco Etch Virus* (TEV) - cleavage side (amino acid sequence: E-N-L-Y-F-Q-II-A, where II denotes the cleavage side) was inserted for tag removal. The pET28a(+)-6xHis-TEV-Ubl1 plasmid was transformed into chemically competent *E. coli* BL21 (DE3) cells. For precultures, cells were grown at 37 °C and 160 rpm in 2xYT medium (Thermo Fisher, Waltham, MA, USA) supplemented with 50 µg/mL Kanamycin overnight. For isotope labeling of ^{1}H-^{15}N Ubl1, cells were grown in M9 minimal medium supplemented with ^{15}N-NH_4-Cl (99 %, Cambridge Isotope Laboratories, USA) as single sources for protein synthesis. The main culture was incubated at 37 °C and 120 rpm, until an OD_{600} of 0.8 was reached; at this point, protein expression was induced with 0.5 mM IPTG. Cultures were incubated at 20 °C after induction of expression for 18 h at 120 rpm. The next day, bacterial cells were harvested for 10 min at 6,000 x g (4 °C) using centrifugation. The cell pellet was resuspended in 50 mM Tris (pH 8), 250 mM NaCl buffer supplemented with one cOmplete EDTA-free protease inhibitor tablet (Roche, Basel, Switzerland) and DNAse I (10 ng per gram cell pellet) for nucleic acid digestion. Cells were lysed for 20 min by sonication (65% amplitude, 1 s pulse/ 3 s pause), followed by an additional centrifugation step for 30 min at 45,000 rpm (4 °C). The lysate was purified by Ni-NTA affinity chromatography using a Protino 5 mL FPLC

column (Macherey-Nagel, Düren, Germany). His-tagged Ubl1 was eluted with 50 mM Tris (pH 8), 250 mM NaCl, 500 mM Imidazole. TEV cleavage was combined with dialysis overnight at 4 °C in 50 mM Tris (pH 8), 250 mM NaCl, 1 mM DTT, 0.5 mM EDTA in a ratio of 1:100 (1 protease for 100 target proteins). TEV protease was separated from digested protein by an additional Ni-NTA affinity chromatography using the same buffer as before. However, this time digested protein was expected in the flowthrough. To increase the purity of the sample, a size exclusion chromatography was performed afterward using a HighLoad 16/600 Superdex 75 pg column (Cytiva, Marlborough, Massachusetts, USA) equilibrated in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl, 0.1 mM TCEP or 50 mM sodium phosphate (pH 6.5) and 0.1 mM TCEP, depending on the experiment. Quality of protein samples was analyzed by SDS-PAGE. Lastly, purified protein was concentrated using a Amicon with a cut-off of 3 kDa (Merck Millipore, Darmstadt, Germany).

3.3.2 Dynamic Light Scattering of Ubl1

Prior to DLS measurements, Ubl1 samples were concentrated up to 2 mg/mL, 4 mg/mL, 6 mg/mL, 8 mg/mL and 10 mg/mL in 50 mM sodium phosphate (pH 6.5) and 1 mg/mL, 2 mg/mL, 10 mg/mL and 12 mg/mL in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl using a Amicon with a cut-off of 3 kDa (Merck Millipore, Darmstadt, Germany). All the DLS measurements were performed with a Malvern Zetasizer Nano ZS apparatus with a backscattering setup (θ = 173°) and a He-Ne laser of the wavelength λ = 633 nm. Measurements were performed at 25 °C. DLS measures the collective translational diffusion coefficient D_c of particles [m^2/s] which is related under dilute conditions to the effective hydrodynamic radius R_h by the Stokes-Einstein equation:

$$D_0 = \frac{k_B T}{6\pi\eta R_H}$$

with k_B as the Boltzmann constant [m^2kg/Ks^2], T as the temperature [K], η as the viscosity [Pas], and R_H as the hydrodynamic radius [m]. For increasing concentration (c) the collective or mutual diffusion coefficient $D_m(c)$ is influenced by excluded volume effects, direct and hydrodynamic interactions between the particles leading for small concentrations to

$$D_m(c) = D_0(1 + k_m c)$$

with the interaction parameter k_m at low concentrations. Qualitatively, positive values of k_m indicate repulsive interaction while negative values indicate attractive interactions. In detail there is a small positive excluded volume contribution that appears for no direct interactions.

One finds $\mathrm{k_m} = 2B_{22}M - \nu - \xi_1$ with B_{22} as the second virial coefficient, the specific volume ν and the concentration coefficient of friction ξ_1. [150,151] On the other hand there are several predictions for k_m in the hard sphere case that deviate from each other dependent on the precision of the derivation.[152] A value of $\mathrm{k_m} = 1.56\nu \approx 1.1\ ml/g$ for $\nu = 0.71\ ml/g$ of Ubl1 can be estimated for a hard sphere case assuming no additional interactions except the excluded volume.

3.3.3 NMR Relaxation Data

NMR relaxation experiments were performed for 30 µM, 100 µM, 145 µM and 150 µM ^{1}H-^{15}N Ubl1 in 50 mM sodium phosphate (pH 6.5), 0.1 mM TCEP, 5% (v/v) D_2O, 0.01% (v/v) NaN_3 buffer at 25 °C. Samples were measured in the presence and absence of 150 mM NaCl. Experiments were acquired on different Bruker NMR spectrometers with ^{1}H Lamor frequencies of 600 MHz and 900 MHz. The NMRPipe software was used to process spectra.[129] Data were visualized using Thonny, a Python platform.[153] Experimental parameters of ^{15}N R_1 and ^{15}N $R_{1\rho}$ relaxation rate constants data measuring Ubl1 samples are listed in supplementary information (Table S11 and S12). ^{15}N R_2 rate constants were derived from $R_{1\rho}$ rate constants by correcting for R_1 contributions. Pulse sequences of ^{15}N R_1 and ^{15}N $R_{1\rho}$ relaxation rate experiments were used from previous publications.[134,135]

3.3.4 Analytical Ultracentrifugation

Sedimentation velocity (SV) experiments were performed using a Beckman Optima XL-A ultracentrifuge (Beckman Coulter, Brea, CA, USA) equipped with UV/Vis absorbance detection system and an An-60Ti 4-hole rotor from Beckman Coulter. For the measurement, a concentration of 145 µM Ubl1 in 50 mM sodium phosphate buffer (pH 6.5) with 150 mM NaCl in the presence and absence of 1 mM TCEP was prepared. A reference volume of 400 µl and a sample volume of 390 µl were used in standard double-sector measuring cells, with quartz glass windows and aluminum centerpieces with an optical path length of 12 mm. A wavelength of 290 nm was selected for detection, based on pre-measurement wavelength scans conducted at 3,000 rpm. The SV experiments were performed at 20 °C and 60,000 rpm in continuous mode with a radial resolution of 0.003 cm. Radial concentration profiles were measured in time intervals of approximately 2 min until the protein was fully sedimented. Sedimentation data was analyzed using the continuous distribution *c*(*s*) Lamm equation model

(*c*(*s*) model) of the software package SEDFIT Version 16.1c.[130] A resolution of 0.1 S and a confidence level (*F*-ratio) of 0.95 were selected for data analysis within the *s*-value range of 0-15 S. The partial specific volume of Ubl1 (vbar = 0,73586 $cm^3\ g^{-1}$) was calculated based on the amino acid sequence using the software SEDNTERP Version 2.0 Beta.[131] Additionally, solvent density (ρ = 1.0098 g/cm^3) and viscosity (η = 0.01036 P) were calculated from tabulated data sets of the same software. All rmsd values of the fits accounted for less than 1% of the total signal. Standardization of measured *s*-values to water at 20 °C ($s_{20,w}$) was performed. Further analysis and graphical visualization of the sedimentation data were achieved using the software GUSSI[132] and DataGraph Version 5.3β, copyright 2020 (North Carolina, USA).

3.3.5 Small Angle X-ray Scattering (SAXS) Analysis

For SAX measurements, Ubl1 samples were concentrated to 4 mg/mL, 10 mg/mL and 12 mg/mL in 50 mM sodium phosphate (pH 6.5) supplemented with 0.1 mM TCEP. SAXS data were measured with the laboratory-based SAXS system "Ganesha-Air" from (SAXSLAB/XENOCS) at Forschungszentrum Jülich. The X-ray source is a D2-MetalJet (Excillum) with a liquid metal anode operating at 70 kV and 3.57 mA with Ga-Kα radiation (wavelength λ = 0.1314 nm) providing a brilliant and small beam (<200 µm). A PILATUS 300K (Dectris) was used to record the 2D scattering image. All samples were measured in sealed glass capillaries of 2 mm inner diameter. Images were circular averaged, normalized to incident intensity and transmission corrected. The covered Q range was 0.11-4.17 nm^{-1} measured at a detector distance of 187 cm and 0.11-2.64 nm^{-1} measured at a detector distance of 117 cm. To exclude structure factor effects different concentrations were measured, and corresponding buffer was subtracted as background. Protein form factors *P(Q)* were calculated based on protein PDB structures with atomic positions r_i and respective form factor $f_i'(Q)$ as

$$P(Q) = \langle \sum_{i,j} f_i'(Q) f_j'(Q) e^{-Q(r_i - r_j)} \rangle$$

In a solvent, according to Babinet's principle, we subtract from the atomic form factors $f_{j,a}(Q)$ the excluded volume scattering $f_{j,d}(Q)$ of the solvent filled with dummy atoms yielding $f_j'(Q) = f_{j,a}(Q) - f_{j,d}(Q)$. A hydration layer at the surface can be fitted by adding a layer of dummy atoms around the protein structure. All calculations are done with the program Jscatter.[133]

3.4 Results

3.4.1 Translational Diffusion of Ubl1 Monitored by DLS

The translational diffusion of the Ubl1 domain was analyzed by dynamic light scattering (DLS), a technique commonly used to assess protein size and aggregation behavior in dilute solution.[154,155] Here, we investigated the influence of sodium chloride and varying protein concentrations on the diffusion properties of Ubl1 in aqueous solution (Figure 21). The DLS data were analyzed using the instrument software's NNLS fit algorithm, accounting for multimodal distributions, e.g., to include aggregates. We generally find large aggregates with sizes >400 nm and non-negligible intensities in the solution. As a scattering technique, the DLS signal is proportional to the particle/aggregate volume squared, V^2. Accounting for this, the relative volume fraction of aggregates is negligible (<<1%), as aggregates are ≈100 times larger than the monomer. Figure 21 presents the measured DLS data along with corresponding fits for $D_m(c)$ to evaluate the dilute limit D_0 and the interaction parameter k_m. Without salt we find $D_0 = 117 \pm 3$ µm²/s and $k_m = -7.9 \pm 3.9\, ml/g$. D_0 of a monomer can be estimated from a monomeric PDB structure using the software HYDROPRO[156] and evaluates to 107 µm²/s for a monomer of the PDB structure 7KAG with a β-sheet as N-terminus and to 130 µm²/s if the N-terminus is cut (residues A1-S20).

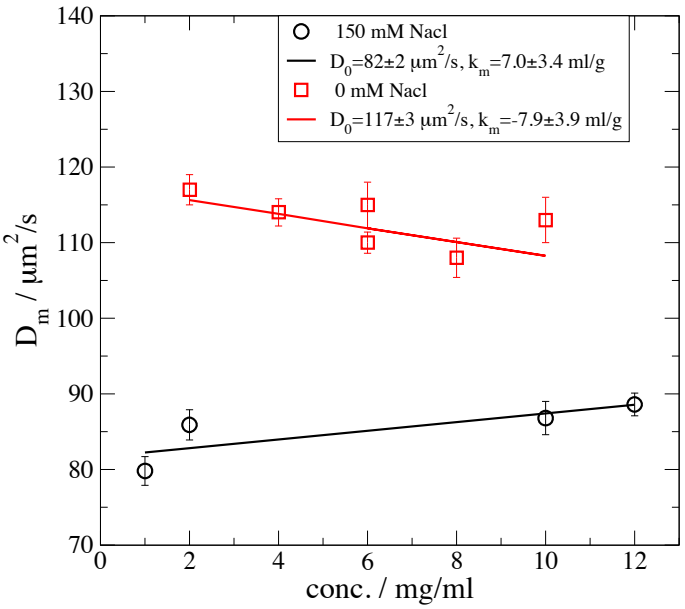


Figure 21: Salt affects the transversal diffusion of Ubl1. The diffusion coefficient D_m [µm²/ s] of Ubl1 samples were measured at increasing concentrations in 50 mM sodium phosphate (pH 6.5) (No Salt, red) and in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl (Salt, black). Measurements were performed at 25 °C. Lines indicate the fit according to $D_m(c) = D_0(1 + k_m c)$.

As we find experimentally by NMR that the N-terminus is flexible and disordered (comp. Chapter 2),[67] an intermediate value between rigid and cut N-terminus would be expected. This turns out to be in good agreement with the experimental value of $D_0 = 117 \pm 3$ µm²/s found in DLS measurement. Indeed, a monomeric structure with a flexible N-terminus is consistent with the data. The negative k_m indicates an attractive interaction between Ubl1 monomers, which seems to be the reason for the appearance of aggregates. At a salt concentration of 150 mM NaCl, we find a substantially smaller translational diffusion coefficient and weakly positive concentration dependence of the measured diffusion coefficients. The measured diffusion coefficient can be fitted by a dilute limit $\mathrm{D_0} = 82 \pm 2$ µm²/s and interactions parameter $\mathrm{k_m} = 7.0 \pm 3.4\ \mathrm{ml/g}$. A positive $\mathrm{k_m}$ indicates a repulsive stabilizing interaction in the presence of salt, while the smaller D_0 indicates a larger multimer. The dimeric structure in PDB 7KAG evaluates to 76 µm²/s that we expect to see at least a strong contribution of dimers.

3.4.2 NMR Relaxation Reveals Salt-Dependent Changes in Ubl1 Dynamics

NMR relaxation experiments provide valuable insights into both internal flexibility and overall tumbling of proteins in solution.[157] Our study builds upon the methodology established by Pielak *et al.* (2009), who introduced a strategy to discriminate between binding interactions and viscosity effects in crowded environments by evaluating the $R_1 \cdot R_2$ product, which remains independent of viscosity.[158] Here, we investigated the effect of sodium chloride on the dynamic properties of ^{15}N-labeled Ubl1. Salt is known to influence both the thermodynamics and kinetics of viral and cellular proteins.[159,160,161] To mimic physiologically relevant ionic conditions, ^{15}N-Ubl1 samples at 30 µM, 100 µM, and 150 µM were analyzed in the absence (0 mM) and presence (150 mM) of NaCl (Figure 22). The 0 mM condition represents a more repulsive environment, while 150 mM NaCl reflects an ionic strength comparable to intracellular conditions.[162,163,164] First, we assessed the effect of increasing Ubl1 concentration on its structural dynamics in the presence of 150 mM NaCl. As shown in Figure 22D, increasing concentrations resulted in slightly reduced ^{15}N R_1 relaxation rate constants, while ^{15}N R_2 ($R_{1\rho}$) rate constants increased (Figure 22C). This suggests a decrease in rotational mobility and an increase in the effective tumbling time (τ_c), potentially indicative of transient dimerization or crowding-induced compaction. In contrast, under low ionic strength conditions (0 mM NaCl), no significant concentration-dependent changes were observed in either ^{15}N R_1 or R_2 ($R_{1\rho}$)

values (Figure 22A, B), suggesting that Ubl1 remains monomeric and less influenced by intermolecular interactions. Notably, when comparing samples across ionic conditions, ^{15}N R_2 ($R_{1\rho}$) rates were consistently higher in the presence of NaCl (Figure 22C) than in its absence (Figure 22A), while ^{15}N R_1 differences remained modest.

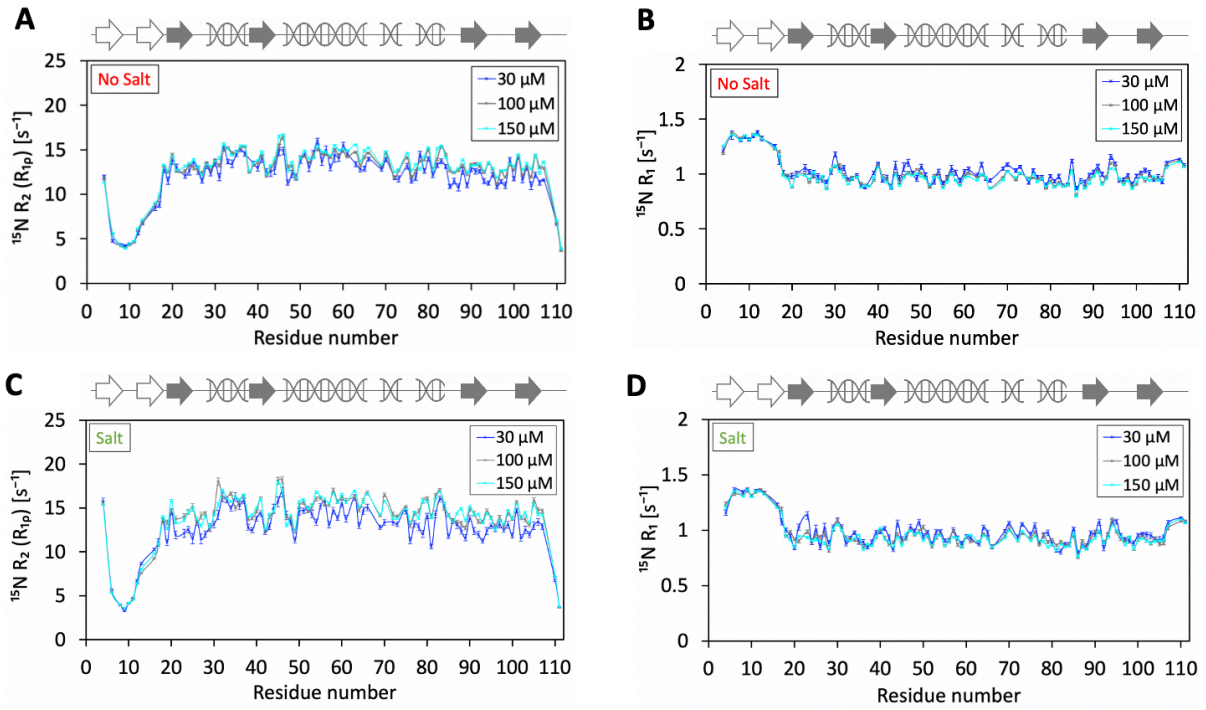


Figure 22: Concentration-dependent effect of Ubl1 in samples with salt. ^{15}N NMR relaxation of Ubl1 were measured in a concentration of 30 µM, 100 µM and 150 µM in 50 mM sodium phosphate (pH 6.5) and 0.1 mM TCEP (No Salt) or 50 mM sodium phosphate (pH 6.5), 150 mM NaCl and 0.1 mM TCEP (Salt). Spectra were recorded at 900 MHz at 25 °C. ^{15}N R_2 ($R_{1\rho}$) rate constants in absence **(A)** and presence of 150 mM NaCl **(C)**.^{15}N R_1 rate constants in absence **(B)** and presence of 150 mM NaCl **(D)**.

These observations indicate that NaCl enhances intermolecular interactions, likely by reducing electrostatic repulsion, thereby promoting crowding effects or weak self-association. In summary, our data suggest that Ubl1 exhibits salt-dependent dynamic behavior: under high-salt conditions, reduced electrostatic repulsion may facilitate transient interactions or compaction, whereas in low-salt environments, Ubl1 remains more dynamically free and unaffected by intermolecular interactions.

3.4.3 Oligomeric State and Cysteine Involvement in Ubl1 Dynamics

To further investigate the salt-induced decrease in Ubl1 mobility, we examined its oligomeric state and the potential involvement of cysteine-mediated disulfide bonding. AUC experiments were conducted in the presence and absence of the reducing agent TCEP (Figure S10). TCEP is

widely used in protein biochemistry to disrupt disulfide bonds within or between proteins by reducing cysteine residues.[165] AUC analysis confirmed a monomeric state of Ubl1 in 50 mM sodium phosphate buffer (pH 6.5), 150 mM NaCl at 25 °C, independent of the presence of TCEP (Figure S10). In both reduced and non-reduced conditions, a single sharp peak was observed in the sedimentation coefficient distribution at 1.50 S (reduced) and 1.53 S (non-reduced), respectively (Figure S10). The experimentally determined molecular weight was 12.7 kDa in both cases, which closely matches the theoretical molecular weight of Ubl1 (12.46 kDa), supporting the existence of a homogeneous monomeric species. Notably, in the absence of TCEP, the monomer accounted for 96.9% of the total signal. These findings were corroborated by NMR relaxation experiments performed on 145 µM ^{15}N-Ubl1 under both reducing (+1 mM TCEP) and non-reducing conditions (Figure S11, Table S12). Ubl1 contains three cysteine residues at positions C39, C55, and C103, which could potentially mediate disulfide-linked dimerization (Figure S12). However, the R_1 relaxation data of reduced and non-reduced samples showed a strong correlation (R = 0.766, rmsd = 0.064; Figure S11B), and box plot analysis revealed a narrow distribution with a mean difference close to zero (-0.768; Figure S11C), indicating minimal differences between conditions. Similarly, ^{15}N R_2 relaxation rates were highly correlated (R = 0.955, rmsd = 0.809; Figure S11E), with a box plot average of 0.353 (Figure S11F), further supporting the absence of structural or dynamic changes upon reduction. In conclusion, neither AUC nor NMR relaxation data provided evidence for disulfide-mediated dimerization of Ubl1. The observed dynamic behavior is therefore unlikely to be driven by covalent linkage via cysteines, and the protein remains monomeric under both reducing and non-reducing conditions.

3.4.4 SAXS Analysis of Ubl1

As complementary information, also SAXS data at different concentrations were recorded in the presence of 150 mM NaCl (Figure 23A) and without salt (Figure 23B). The data show in both cases larger aggregates visible at lower Q as an increase of the scattering. It was not possible to filter these aggregates as they were instantly reappearing as observed by DLS. At larger Q we observe in both cases reasonable scattering intensities that can be used to gain more information about the observed species by fitting a formfactor based on the PDB structure 7KAG. Without salt, we fitted a monomeric structure (Figure 23B black line) that even shows the minimum at Q=3nm^{-1} that appears as the protein has an ellipsoidal core

creating this minimum. With salt, the measured formfactor is in between the expected formfactor of a monomer and of the dimer present in PDB 7KAG. Fitting a number fraction of monomers and monomers present in dimers results in a monomer fraction of 0.69± 0.04 of monomers and the corresponding fraction of monomers in dimers. The corresponding fraction of scattering intensities at low Q as observed also by DLS, is about 50%, as the dimers contribute with this volume squared that in DLS an average value between monomer and dimer. From the present data a detailed information about the binding configuration of dimers cannot be obtained.

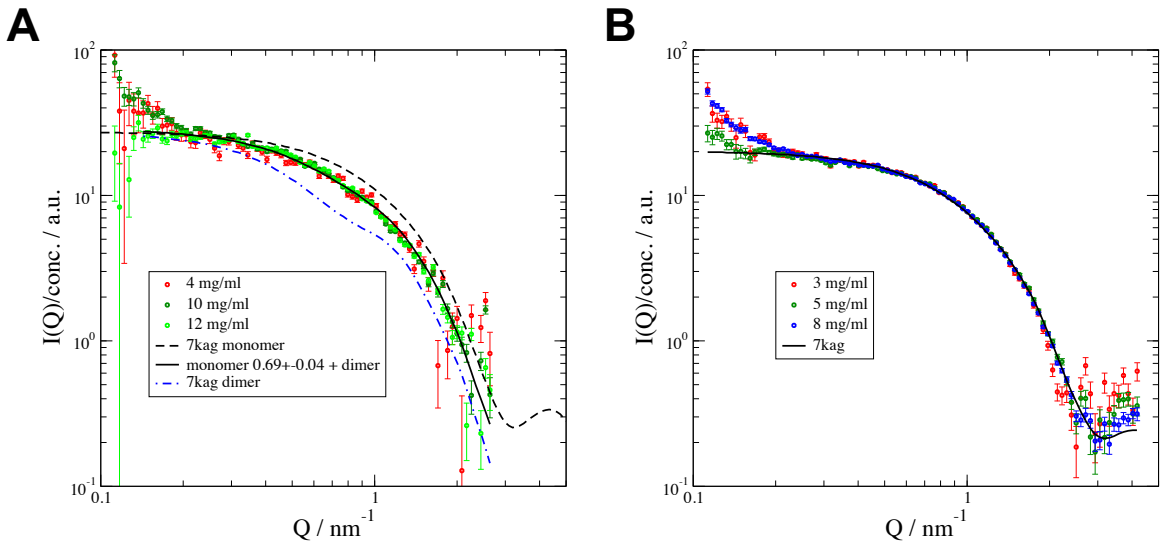


Figure 23: SAXS measurements of Ubl1 with 150 mM salt (A) and without salt (B). SAXS form factor after concentrations scaling. Lines indicate the calculated form factor for monomers and dimers present in PDB code 7KAG. A hydration layer of 0.3 nm thickness and 5% higher density than bulk water was fitted.

3.5 Discussion

The oligomerization state of SARS-CoV-2 NSP3-Ubl1 remains a subject under investigation. A recent study reported a heterotrimeric complex comprising two Ubl1 domains and the N-terminal domain of the N-protein.[61] Furthermore, two crystal structures of SARS-CoV-2 Ubl1 (PDB: 7KAG and 7TI9) revealed dimeric arrangements with open and closed V-like conformations, while a third structure lacking the N-terminal residues 1-17 (PDB: 8XAB) showed only a monomeric form, highlighting the N-terminal region as a key determinant of dimerization.[84] To gain further insights, we analyzed the oligomerization behavior of Ubl1 under varying redox and ionic conditions using a combination of ^{15}N NMR relaxation, AUC,

SAXS, and DLS. These complementary techniques provided a multifaceted view of Ubl1 association behavior in solution.

AUC measurements confirmed that Ubl1 predominantly exists as a monomer under both reducing and non-reducing conditions, with sedimentation coefficients of 1.50-1.53 S and an experimental molecular weight of ~12.7 kDa, closely matching the theoretical value of 12.46 kDa (Figure S10). ^{15}N relaxation data also showed no significant differences between reduced and non-reduced samples (Figure S11), effectively ruling out disulfide-mediated dimerization via the three cysteine residues (C39, C55, C103). However, despite the monomeric behavior detected by AUC, ^{15}N R_2 relaxation rates showed a concentration-dependence under physiological salt conditions (150 mM NaCl), which diminished in low-salt buffers, pointing to salt-induced molecular crowding or weak self-association rather than covalent oligomerization.

To probe the molecular assembly state further, SAXS measurements were performed at varying protein concentrations in both low-salt and physiological-salt buffers (Figures 23A, 23B). In both cases, an increase in scattering at low Q indicated the presence of larger aggregates, consistent with DLS results. These aggregates could not be removed by filtration, as they rapidly re-formed in solution. At higher Q-values, the scattering profiles were suitable for structural modeling using form factor fits based on PDB structure 7KAG. In salt-free buffer, the scattering matched closely to a monomeric model, including a characteristic minimum at $Q \approx 3\ nm^{-1}$, reflecting the ellipsoidal shape of Ubl1 (Figure 23B, black line). In contrast, under 150 mM NaCl, the experimental scattering profile lay between the theoretical curves for a monomer and the dimeric form of 7KAG. Quantitative fitting indicated a monomer fraction of 0.69 ± 0.04, with the remaining Ubl1 molecules existing as dimers. At low Q, this mixture results in disproportionately large scattering intensity due to the volume-squared dependence of SAXS, consistent with ~50% contribution from dimeric species. These data refine and correct the earlier assumption of exclusively monomeric Ubl1 under salt conditions and support a monomer-dimer equilibrium influenced by ionic strength and concentration.

To complement SAXS, DLS was used to analyze translational diffusion and aggregation behavior. At low ionic strength, Ubl1 showed a diffusion coefficient (D_0) of 117 ± 3 $\mu m^2/s$, in

line with a monomeric structure featuring a flexible N-terminal domain. This matches predictions from HYDROPRO calculations based on the 7KAG structure with (107 $\mu m^2/s$) or without (130 $\mu m^2/s$) the N-terminal β-hairpin. The negative interaction parameter (k_m = -7.9 ± 3.9 mL/g) indicates attractive interactions, likely driving aggregation in low-salt conditions, although the relative volume fraction of aggregates remains negligible due to their large size and low number fraction. In 150 mM NaCl, the diffusion coefficient decreased significantly to D_0 = 82 ± 2 $\mu m^2/s$, consistent with an increase in hydrodynamic size. This value closely approaches the calculated D_0 of 76 $\mu m^2/s$ for the dimeric structure in PDB 7KAG. Additionally, the interaction parameter became positive (k_m = 7.0 ± 3.4 mL/g), indicating repulsive interactions that help stabilize the solution by preventing further aggregation. These findings clearly demonstrate a salt- and concentration-dependent monomer-dimer equilibrium, with additional weak aggregation phenomena under low-salt conditions driven by attractive interactions, possibly mediated by hydrophobic contacts or charge clustering.

The N-terminal region of Ubl1 contains a β-hairpin with four charged residues - K4 (positively charged), and D9, D10, E14 (negatively charged). At pH 6.5 and 7.4, these residues are ionized, and under low ionic strength, electrostatic repulsion likely contributes to monomer stability and prevents close packing (Figure 24A). At 150 mM NaCl, charge screening reduces repulsion, allowing closer approach of Ubl1 molecules, facilitating weak dimerization and molecular crowding effects, reflected in elevated R_2 rates and reduced diffusion coefficients (Figures 21-24). In addition, the N-terminal β-hairpin contains eight hydrophobic residues (T3, V5, T6, F7, T11, V12, I13, V15) which may participate in hydrophobic interactions once electrostatic repulsion is mitigated. This supports the idea of transient, non-covalent dimers forming under physiological salt conditions.

Our findings provide a molecular basis for reconciling previous conflicting observations. For example, Ni *et al*. reported a heterotrimeric Ubl1-N-protein complex at high concentrations (40 mg/mL) in the presence of 150 mM NaCl with Ubl1 dimerization mediated by the N-terminus.[61] In contrast, SEC-MALS at lower concentrations (5 mg/mL) showed only monomeric Ubl1.[61] Similarly, Serrano *et al*. observed dimeric Ubl1 from SARS-CoV-1 at 1 mM protein concentration.[87] Our data support a concentration-dependent association modulated by ionic strength, protein flexibility, and hydrophobic exposure. Furthermore, Ubl1 may adopt

different oligomeric states *in vivo* due to local concentration effects. In the hexameric pore complex of NSP3, N-terminal Ubl1 domains are positioned at the tips of flexible arms,[56] likely reaching high local concentrations that could promote crowding-driven conformational restriction, analogous to crowding-induced binding in other viral systems such as NS3/4A of hepatitis C virus.[161]

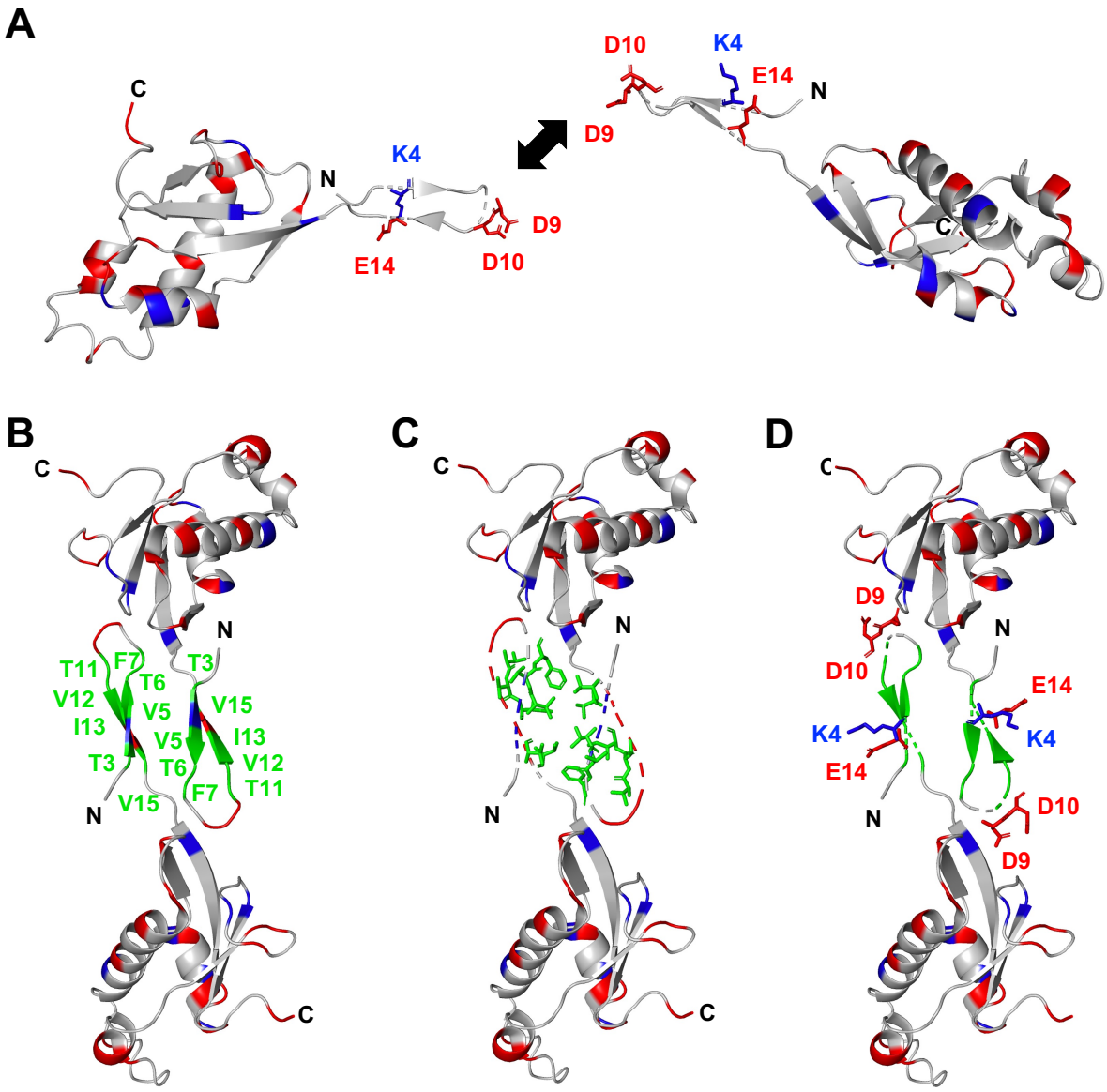


Figure 24: Schematic illustration of the impact of salt on Ubl1. (A) Crystal structure of Ubl1 (PDB: 7KGA) displayed as a monomer. Negatively charged residues (E, D) are highlighted in red and positively charged residues (K, R, H) in blue. Charged amino acids of the N-terminus (K4, D9, D10, E14) are displayed as sidechains. **(B)** Homodimer of Ubl1 (PDB: 7KAG) with focus on hydrophobic side chains (T3, V5, T6, F7, T11, V12, I13, V15) **(C)** and charged residues **(D).** Hydrophobic residues of the N-terminus are highlighted in green **(B, C, D).**

In summary, our combined AUC, NMR, SAXS, and DLS data establish a mechanistic framework in which ionic strength, charge shielding, hydrophobic contacts, and local

concentration govern the oligomerization behavior of Ubl1. While the protein is predominantly monomeric under dilute conditions, it exists in a dynamic monomer-dimer equilibrium under physiological salt concentrations, with additional weak aggregation driven by attractive interactions. These findings help explain previously contradictory *in vitro* observations and underscore the importance of biophysical context when interpreting Ubl1's oligomerization and function.

3.6 Supplementary Information

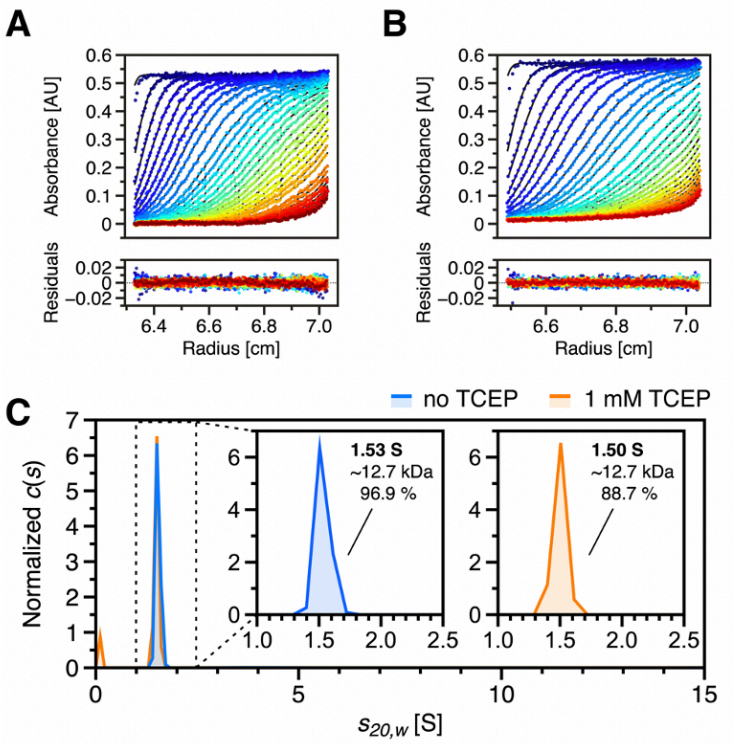


Figure S10: Sedimentation coefficient distribution of NSP3-Ubl1 determined by SV-AUC analysis. Sedimentation profiles of 145 µM NSP3-Ubl1 were analyzed in **(A)** 50 mM sodium phosphate buffer, pH 6.5 and **(B)** 50 mM sodium phosphate buffer, pH 6.5 with 1 mM TCEP. Freshly prepared samples were sedimented at 60,000 rpm and 20 °C, with absorbance detected at 290 nm. A total of 400 scans were recorded, with one in three of the first 300 scans utilized for sedimentation analysis. The continuous distribution c(*s*) Lamm equation model was applied to determine differences in sedimentation coefficients for NSP3-Ubl1 before (blue) and after (orange) TCEP addition (**C**). Data fitting was performed from 0 S to 15 S, and the identified species range is shown magnified. Measured *s*-values were standardized to $s_{20,w}$, reflecting water at 20 °C. The *c*(*s*) distributions were normalized to the area under the curve for consistent comparison.

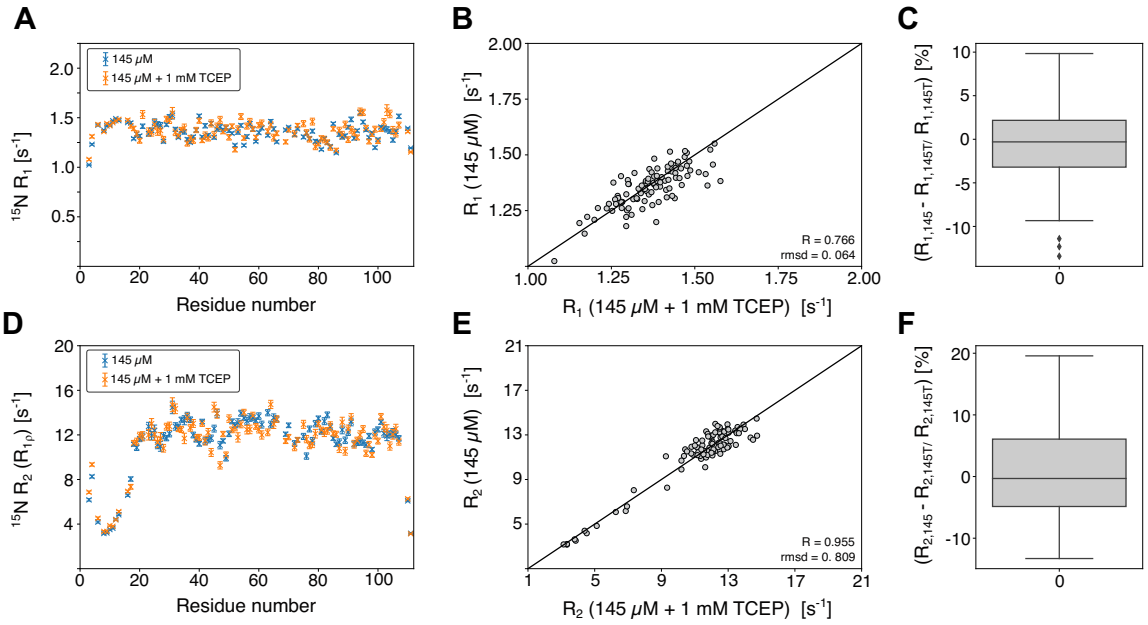


Figure S11: No effect of TCEP on the dynamics of Ubl1. ^{15}N NMR relaxation of Ubl1 was measured in a concentration of 145 µM in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl in the presence and absence of 1 mM TCEP at 25 °C. Spectra were recorded at 600 MHz. **(A)** Residue-specific ^{15}N R_1 relaxation rate constant comparison of 145 µM Ubl1 and 145 µM Ubl1 supplemented with 1 mM TCEP. **(B)** Correlation plot of the compared ^{15}N R_1 relaxation rate constants (145 µM Ubl1 and 145 µM Ubl1 + 1 mM TCEP). **(C)** Box plot of the residue-specific percentual differences of ^{15}N R_1 relaxation rate constants. **(D-F)** Same plots as in A-C, displaying the ^{15}N R_2 relaxation rate constants of ^{15}N Ubl1 samples.

Table S11: Experimental parameters of ^{1}H-^{15}N-Ubl1 (30 µM, 100 µM, 150 µM) NMR relaxation experiments at 900 MHz.

	^{15}N R_1	**^{15}N $R_{1\rho}$**
Field [T]	21.15	21.15
1H Larmor frequency [MHz]	900	900
Pulse program	Stief *et al.*, Methods 2024	Stief *et al.*, Methods 2024
Interscan recovery delay [s]	3.00	3.00
Number of scans	30 µM: 16 100 µM: 8 150 µM: 8	30 µM: 16 100 µM: 8 150 µM: 8
Total experimental time [h]	30 µM: 64.5 100 µM: 32.4 150 µM: 32.4	30 µM: 58.5 100 µM: 29.4 150 µM: 29.4
Direct dimension (^{1}H)		
^{1}H carrier [ppm]	4.70	4.70
Spectral width [ppm]	16.34	16.34
Data points	2048	2048
Acquisition time [ms]	69.63	69.63
90° (^{1}H) pulse length [µs]	8.05	8.05
Indirect dimension (^{15}N)		
^{15}N carrier [ppm]	117.00	117.00
Spectral width [ppm]	36.55	36.55

Data points	8 x 512	8 x 512
Acquisition time [ms]	76.8	76.8
90° (^{15}N) pulse length [µs]	42.50	42.50
Relaxation period		
# relaxation delays	8	8
Time points [ms]	0, 880, 80, 720, 160, 480, 240, 320	5, 55, 25, 20, 15, 45, 35, 10

Table S12: Experimental parameters of 145 µM ^{15}N-Ubl1 NMR relaxation experiments (presence and absence of 1 mM TCEP) at 600 MHz.

	^{15}N R_1	**^{15}N $R_{1\rho}$**
Field [T]	14.1	14.1
1H Larmor frequency [MHz]	600	600
Pulse program	Stief *et al.*, Methods 2024	Stief *et al.*, Methods 2024
Interscan recovery delay [s]	2.00	2.00
Number of scans	4	4
Total experimental time [h]	11.92	10.41
Direct dimension (^{1}H)		
^{1}H carrier [ppm]	4.70	4.70
Spectral width [ppm]	16.02	16.02
Data points	2048	2048
Acquisition time [ms]	106.50	106.50
90° (^{1}H) pulse length [µs]	8.90	8.90
Indirect dimension (^{15}N)		
^{15}N carrier [ppm]	117.00	117.00
Spectral width [ppm]	34.98	34.98
Data points (total)	8	8
Acquisition time [ms]	120.32	120.32
90° (^{15}N) pulse length [µs]	42.20	42.20
Relaxation period		
# relaxation delays	8	8
Time points [ms]	0, 720, 160, 640, 80, 560, 320, 480	5, 65, 25, 45, 15, 55, 35, 10

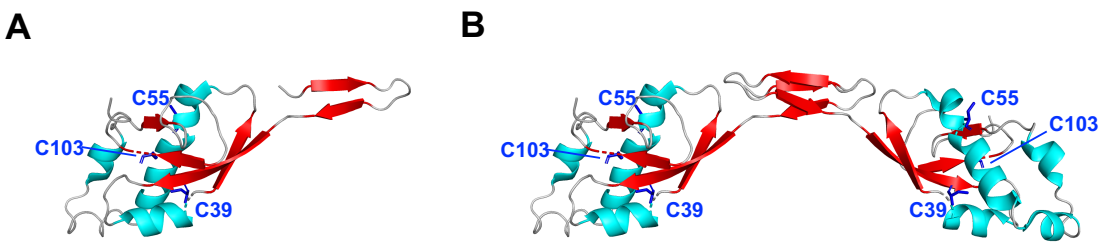


Figure S12: Cysteine residues of Ubl1. Crystal structure of a Ubl1 monomer **(A)** and a Ubl1 dimer **(B)** (PDB: 7kag). Cysteine residues of Ubl1 are highlighted in dark blue (C39, C55, C103).

4. Chapter: Posttranslational Modification of SARS-CoV-2 NSP3-Ubl1 Expressed in Prokaryotic and Eukaryotic Cells

4.1 Abstract

The Ubl1 domain of the SARS-CoV-2 NSP3 plays a pivotal role in the viral replication cycle and its interactions with host and viral components. In this study, we characterized the subcellular localization and posttranslational modifications (PTMs) of Ubl1 expressed in *E. coli* and HEK293T cells. Immunofluorescence microscopy revealed that HA-tagged Ubl1 localizes predominantly to the cytosol of HEK293T cells, consistent with its proposed orientation at the cytosolic surface of DMVs. Unexpectedly, Western blot and Native-PAGE analyses indicated heterogeneous phosphorylation of Ubl1 when expressed in *E. coli*, with mass spectrometry confirming the presence of up to six phosphate groups per monomer. These modifications significantly altered the protein's electrophoretic mobility. Although bacterial systems typically lack the enzymatic machinery for complex PTMs, prior reports of tyrosine kinase and phosphatase activity in *E. coli* BL21 may explain the observed phosphorylation pattern. In contrast, Ubl1 expressed in HEK cells displayed no detectable phosphorylation by Western blot, but mass spectrometry identified three phosphorylation events. Due to the presence of tyrosine residues in the HA-tag, tag-free constructs are required to conclusively identify modified residues. Functional implications of phosphorylation were considered in the context of known Ubl1 interactions with the viral N-protein and RNA. Given the negative charge of both phosphorylated proteins and nucleic acids, electrostatic effects may influence binding affinity.

Together, our findings suggest that Ubl1 is subject to phosphorylation in both bacterial and human expression systems, raising the possibility that such modifications regulate its role in the SARS-CoV-2 replication cycle. Further studies are needed to identify the specific phosphorylation sites and their functional impact on viral protein-RNA interactions.

4.2 Introduction

4.2.1 Posttranslational Modifications of SARS-CoV-2 Proteins

Posttranslational modifications (PTMs) such as glycosylation, ubiquitination, palmitoylation, and phosphorylation are common regulatory mechanisms that modulate the function, stability, and localization of viral proteins (Figure 25).[166,167,142] Among CoVs, these modifications play critical roles in facilitating viral replication, immune evasion, and host-pathogen interactions. For instance, the N-protein of SARS-CoV-1 and SARS-CoV-2 is phosphorylated by glycogen synthase kinase 3 (GSK-3) and the ataxia-telangiectasia mutated and Rad3-related (ATR) kinase, a modification that promotes the formation of biomolecular condensates essential for viral RNA transcription.[142,168,169] Phosphorylation of the N-protein is further known to enhance RNA-binding specificity, regulate subcellular localization, and influence antigenic properties, all of which are key to efficient viral assembly and immune modulation.[167]

Glycosylation is another widespread PTM in SARS-CoV-2, notably affecting the S-protein and NSP4.[170,171,172] The S-protein is glycosylated within the ER, where it acquires N-terminal mannose and complex glycan structures critical for proper folding, receptor binding, and immune recognition.[170] Similarly, NSP4 is glycosylated in a manner thought to contribute to the biogenesis of DMVs, which serve as replication organelles and are essential for the spatial coordination of viral RNA synthesis.[171,172] Comprehensive phosphor-proteomic analysis by Stukalov *et al*. revealed dynamic regulation of phosphorylation events in SARS-CoV-2-infected human lung cells, underscoring the functional relevance of PTMs throughout the viral life cycle.[173] Five viral proteins were identified as phosphorylated: the M-protein, the N-protein, the S-protein, ORF9b, and NSP3 - indicating that phosphorylation is not restricted to a single viral component but is a widespread regulatory mechanism.[173] While phosphorylation sites have been precisely mapped in some proteins, such as the C-terminal residues (S310, S311) of the N-protein, the specific phosphorylation sites on NSP3 remain undefined.[173] Phosphorylated N-protein expressed in infectious bronchitis virus has been shown to retain its ability to bind viral RNA, indicating that electrostatic repulsion between negatively charged phosphate groups and RNA does not necessarily inhibit this interaction.[174] A recent study demonstrated that the E-protein of SARS-CoV-2 is palmitoylated at three cysteine residues

(C40, C43, and C44) by palmitoyltransferases.[175] Mutation of these sites to serine significantly reduced the stability of the E-protein and disrupted its interactions with the S-, M- and N-protein.[175] As a result, virus-like particle (VLP) formation was impaired, highlighting the importance of palmitoylation for E-protein function and viral assembly.[175] Furthermore, ubiquitination is a central PTM observed in SARS-CoV-2 affecting several viral proteins (Figure 25).[166,167,173]

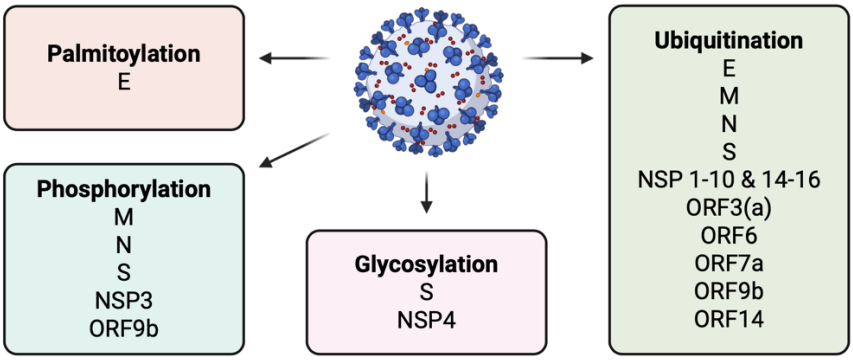


Figure 25: Posttranslational modifications of SARS-CoV-2 proteins. Palmitoylation, phosphorylation, glycosylation and ubiquitination are the four central PTM observed in SARS-CoV-2, concerning the envelope protein (E), the membrane protein (M), the nucleocapsid protein (N), the spike protein (S), multiple non-structural proteins (NSPs) and open-reading frames (ORF). Figure was created in BioRender.com.

As demonstrated, the prevalence and conservation of PTMs among SARS-CoV-2 proteins highlight their critical role in modulating viral replication, host cell manipulation, and immune evasion and needs to be investigated in more detail.

4.2.2 Posttranslational Modifications of SARS-CoV-2 NSP3

NSP3 has emerged as one of the most post-translationally modified non-structural proteins in both SARS-CoV-1 and SARS-CoV-2, underlining its potential regulatory complexity and functional versatility within the viral replication machinery.[173] Systematic proteomic studies have revealed that NSP3 undergoes extensive phosphorylation and ubiquitination, with several of these modifications occurring at conserved sequence motifs, suggesting evolutionary conservation of specific regulatory mechanisms.[173] Notably, phosphorylation sites were detected in both viral variants, with SARS-CoV-1 NSP3 containing three distinct phosphorylation events, while only one was identified in SARS-CoV-2.[173] Ubiquitination represents another major PTM type identified in NSP3. In an initial comparative analysis,

seven ubiquitination sites were found in SARS-CoV-1 NSP3, whereas twelve sites were mapped in SARS-CoV-2 NSP3, highlighting an expanded regulatory potential in the latter.[173] A follow-up study further extended this finding by identifying a total of 40 ubiquitination sites in SARS-CoV-2 NSP3 alone.[176] Within this broader dataset, domain-specific modification sites were reported, including a ubiquitination site at lysine 4 (K4) in the Ubl1 domain, lysine 798 (K798) in the Ubl2 domain, and 12 distinct sites located within PLpro, a region known for its role in viral polyprotein processing and host immune modulation.[176] These modifications may modulate domain-specific functions of NSP3 or contribute to its interactions with host proteins. In addition to phosphorylation and ubiquitination, other types of PTMs have been suggested based on NSP3's membrane localization. NSP3 is an integral membrane protein that spans the DMV bilayer twice, functioning as a crucial scaffold during RTC formation. In line with its transmembrane nature, studies in the related MHV have shown that NSP3 localizes to the ER and is subject to N-linked glycosylation, a modification that may contribute to protein folding, membrane topology, or interaction with host factors.[177]

Taken together, although the precise pattern and extent of PTMs can differ between CoV species, the consistent presence of multiple modifications on NSP3 - particularly in conserved functional domains - strongly suggests that PTMs play a critical regulatory role in modulating its activity, stability, and interactions. These findings underscore the importance of studying PTMs not only as markers of viral protein maturation but also as potential targets for therapeutic intervention.

4.3 Material and Methods

The following sections provide a detailed description of the experimental procedures used to investigate the potential phosphorylation of Ubl1. A detailed protocol of the expression and isolation of (His)Ubl1 is described in Chapter 1.

4.3.1 Buffer, Chemicals and Cell Media

Medium, chemicals and buffers used for the experimental procedures are listed in the respective paragraphs. To avoid contamination during the experiment, chemicals used for cellular experiments were autoclaved and stored in sterile vessels. All cell culture experiments were performed under strict sterile conditions.

4.3.2 Instruments

Instruments and materials used for the analysis of Ubl1 constructs are listed in Table 11 and 12.

Table 11: Instruments for the isolation and analysis of different Ubl1 constructs.

Instrument	Manufacturer
Agarose-Gel chamber system	Bio-Rad Laboratories GmbH, München
Autoclave VX-150	Systec GmbH, Linden
BZ-X810 fluorescence microscope	Keyence (Osaka, Japan)
CytoSMART Exact FL cell counter	Axion BioSystems, Atlanta (USA)
Desk centrifuge 5415 R	Eppendorf GmbH, Hamburg
Dri-Block Heater 08-3	Techne, UK
Drying chamber T 5050	Heraeus, Hanau
Electrophoresis power supply 3501 XL	Amersham Pharmacia biotech, UK
Mr. Frosty™ Freezing Container	Thermo Fischer Scientific, Schwerte
Gel documentation system	Bio-Rad Laboratories GmbH, München
Heracell™ 240i CO_2 Incubator	Thermo Fischer Scientific, Schwerte
Herasafe™ KS biological safety benches class II	Thermo Fischer Scientific, Schwerte
Magnetic stirrer C-MAG MS 7	IKA GmbH & CO. KG, Staufen
Milli-Q Direct 16	Merck KGaA, Darmstadt
Mini Trans-Blot module blotting chamber	Bio-Rad Laboratories GmbH, München
NanoDrop 2000	Thermo Fischer Scientific, Schwerte
Overhead Mixer	Carl Roth, Karlsruhe
pH meter	Satorius AG, Göttingen
Scale	Satorius AG, Göttingen
Thermocycler compact	Eppendorf, Hamburg
Thermostatic Water Bath WB 4	BIOBASE, China
T100 Thermal Cycler	Bio-Rad Laboratories GmbH, München
Vortexer	Heidolph Instruments, Schwabach

Table 12: List of consumables for the analysis of different Ubl1 constructs.

Material	Manufacturer
Anti-HA magnetic beads	Thermo Fischer Scientific, Schwerte

Assistant counter chamber according to Neubauer	Megro, Wesel
β-Actin Monoclonal Antibody (mouse, IgG1)	Thermo Fischer Scientific, Schwerte
Cell-culture flask (T75)	Thermo Fischer Scientific, Schwerte
cOmplete EDTA-free protease inhibitor tablet	Roche, Basel
Cover slips 22 x 22 mm (0.13 – 0.16 mm thickness)	Megro, Wesel
Cover slips 12 mm round (0.13 – 0.16 mm thickness)	Megro, Wesel
DAPI and Hoechst Nucleic Acid Stains	Thermo Fischer Scientific, Schwerte
Der Blaue Jonas - Single-step Coomassie Blue protein gel dye	GRP German Research, Haag a. d. Amper
Dulbecco's modified Eagle Medium (DMEM) (High Glucose-level, HEPES)	Thermo Fischer Scientific, Schwerte
DynaMag-2 magnetic rack (10 µL to 2 mL)	Thermo Fischer Scientific, Schwerte
FCS Gold	PAN-Biotech, Aidenbach
Goat anti-Mouse IgG (H+L) Secondary Antibody HRP conjugated	Thermo Fischer Scientific, Schwerte
Goat anti-Rabbit Secondary Antibody HRP conjugated	Thermo Fischer Scientific, Schwerte
HA-tag Monoclonal Antibody mouse IgG1	Thermo Fischer Scientific, Schwerte
Falcon tubes (15 mL and 50 mL)	Sarstedt, Nürnbrecht
Fluoromount-G Mounting Medium	Thermo Fischer Scientific, Schwerte
jetPRIME® transfection Kit	Sartorius Polyplus, France
Lambda Protein Phosphatase Kit	New England Biolabs, Ipswich, USA
Penicillin-Streptomycin antibiotics (10.000 U/mL)	Thermo Fischer Scientific, Schwerte
Phosphoserine/threonine/tyrosine Antibody (polyclonal, rabbit, IgG)	Thermo Fischer Scientific, Schwerte
PhosSTOP™	Roche, Basel

Plastic cuvettes	Brand GmbH & Co KG, Wertheim
Pipette tips	Sarstedt, Nürnbrecht
PIPETBOY acu2	Integra Biosciences GmbH, Biebertal
PVDF-membrane 0.2 µm	GE Healthcare, Sweden
Reaction tube (1.5 mL and 2 mL)	Sarstedt, Nürnbrecht
Regenerated Cellulose Membrane Filter 0.2 µm	Sartorius, Göttingen
Serological pipettes	Sarstedt, Nürnbrecht
Spectra Multicolor Low Range Protein Ladder	Thermo Fischer Scientific, Schwerte
Sterile Syringes for Single Use (3 mL, 20 mL) Fisherbrand	Thermo Fischer Scientific, Schwerte
Syringe filter, Filtropur S, PES, pore size 0.2 µm for sterile filtration	Sarstedt, Nürnbrecht
SuperSignal West Pico PLUS	Thermo Fischer Scientific, Schwerte
Trypan Blue Solution	BioChrom AG, Berlin
Whatman Cellulose Blotting Paper	Cytiva, Marlborough
10x Phosphate-Buffered Saline (PBS)	Sigma-Aldrich, St. Louis, USA
24-well plate	Sarstedt, Nürnbrecht
6x-His Tag Monoclonal Antibody (HIS.H8)	Thermo Fischer Scientific, Schwerte

4.3.3 Buffers and Solutions

In the following, the ingredients of buffers and solutions used for the expression, isolation and analysis of HA-Ubl1 (HEK), Ubl1 (*E. coli*) and His-Ubl1 (*E. coli*) are listed in Table 13. All buffers and solutions were prepared with Milli-Q water (18.2 MOhm). Buffers for cell culture and protein isolation were degassed and filtered.

Table 13: Buffer composition and chemical solutions.

Solution	**Ingredients and Concentrations**
Acrylamide:Bisacrylamide (30%:0.8%)	30% (v/v) Acrylamide, 0.8% (v/v) Bisacrylamide
Acrylamide:Bisacrylamide (34.3%:1.1%)	34.3% (v/v) Acrylamide, 1.1% (v/v) Bisacrylamide
BSA Blocking solution	3% BSA (w/v) in (1x) TBS-T

Cathode buffer	1 M Tris-HCl (pH 8.25), 1 M Tricine, 1% (w/v) SDS
Cryo Stock Buffer	10% (v/v) DMSO in DMEM medium
DMEM medium	500 mL DMEM stock, 10% (v/v) FCS Gold, 1% (v/v) Penicillin-Streptomycin antibiotics
IF washing buffer	0.02% PBS-T
IP acidic elution buffer	0.1 M Glycine (pH 2.5)
IP SDS elution buffer	50 mM sodium phosphate (pH 6.5), 150 mM NaCl, 4% (v/v) SDS
IP lysis buffer	50 mM Tris (pH 7.4), 150 mM NaCl, 0.5% (v/v) DDM, 5% (v/v) Glycerol, 2 mM EDTA + 1x protease inhibitor tablet + 1x phosphatase inhibitor tablet
IP wash buffer	50 mM Tris (pH 7.4), 150 mM NaCl, 0.5% (v/v) DDM, 5% (v/v) Glycerol, 2 mM EDTA
Laemmli sample buffer (4x)	10% (w/v) SDS, 10 mM *β*-mercaptoethanol, 20% (v/v) glycerol, 0.05% (w/v) Bromphenol-Blue, 200 mM Tris- HCl (pH 6.8)
Milk Blocking solution	1% milk (w/v) in (1x) TBS-T
Native-PAGE Gel Buffer	0.375 M Tris (pH 8.8)
Neutralization buffer	50 mM Tris (pH 8), 140 mM NaCl
Permeabilization Buffer	0.2% (v/v) Triton in 1x PBS
PFA solution	4% (v/v) paraformaldehyde in 1x PBS
Sample Buffer (2x)	62.5 mM Tris (pH 6.8), 25% (v/v) Glycerol, 1% (w/v) Bromphenol-Blue
TBS-T buffer (20x)	0.5 M Tris (pH 7.6), 3 M NaCl, 2% (v/v) Tween20
TGS buffer (10x)	25 mM Tris-HCl (pH 8.3), 192 mM glycine, 0.1% (w/v) SDS
TG buffer (10x)	25 mM Tris-HCl (pH 8.3), 192 mM glycine
Towbin buffer (1x)	10% (v/v) TG buffer (10x), 20% (v/v) EtOH
Tris-Tricine Gel Buffer (3x)	3 M Tris-HCl (pH 8.45), 0.3% (w/v) SDS

Acrylamide:Bisacrylamide (34.3%:1.1%)	34.3% (v/v) Acrylamide, 1.1% (v/v) Bisacrylamide

4.3.4 Plasmids and Cell Strains

The plasmid encoding for HA-Ubl1 (HEK) was obtained through a cooperation with the Bartenschlager lab (German Cancer Research Center, Institute for Virology, Heidelberg, Germany). The construct composition and further information are shown in Table 14. The plasmid encoding for (His-)Ubl1 is the same as described in Chapter 1.

Table 14: Construct details of HA-Ubl1 (HEK) and (His-)Ubl1 (*E. coli*).

Construct Information	**HA-Ubl1**	**(His-)Ubl1**
Original organism	*SARS-CoV-2*	*SARS-CoV-2*
Plasmid	pcDNA3.1(+)	pET-28a(+)
Codon-optimized	*Mammalian*	*E. coli*
Cell strain	*Human embryonic kidney cells 293T – short HEK293T*	*E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) – short, *E. coli* BL21 (DE3)
N-terminal Tag	HA (YPYDVPDYA)	6xHis
Cleavage Site	-	TEV & Thrombin
Antibiotic Resistance	-	Kanamycin
Restriction sites	EcoRI, BamHI	NdeI, BamHI
Construct size with tag	13.79 kDa, 123 amino acids	15.55 kDa, 138 amino acids
Construct size	-	12.46 kDa, 111 amino acids
pI of tagged construct	3.99	4.70
pI of construct	-	4.08
Extinction coefficient	19035 M^{-1} cm^{-1}	His-Ubl1: 16055 M^{-1} cm^{-1} Ubl1: 14565 M^{-1} cm^{-1}
Extinction coefficient (reduced conditions)	18910 M^{-1} cm^{-1}	His-Ubl1: 15930 M^{-1} cm^{-1} Ubl1: 14440 M^{-1} cm^{-1}

In the following the amino acid sequences of tagged and digested constructs are listed. The His-tag is colored in green, the HA-tag in blue and the construct itself in red. The TEV cleavage

side is displayed in cyan and the Thrombin cleavage side in purple. Stop codons are marked with "*".

6xHis-Ubl1:
MGSSHHHHHHSSGLVPRGSHMENLYFQAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYT
VELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEFKLASHMYCSFYPPDE*

Ubl1:
APTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYTVELGTEVNEFACVVADAVIKTLQPVSELL
TPLGIDLDEWSMATYYLFDESGEFKLASHMYCSFYPPDE*

HA-Ubl1:
MGYPYDVPDYAGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYTVELGTEVNEFACVVA
DAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEFKLASHMYCSFYPPDE*

4.3.5 Cultivation of HEK293T Cells

HEK293T cells were kindly offered by the Bartenschlager lab from the German Cancer Research Center (Institute for Molecular Virology, Heidelberg, Germany). Frozen cells were thawed in a prewarmed water bath (37 °C) for 10 min. The 1 mL cell suspension was mixed with 11 mL of prewarmed DMEM medium and incubated for 2-3 days in a T75 flask at 37 °C until a cell confluence of 80% was reached.

4.3.6 Splitting of HEK Cells

Before the experiments were started, the medium and buffers were prewarmed in a 37 °C water bath for at least 30 min. Starting cell splitting, the consumed DMEM medium of adherent cells was discarded, and cells were gently washed with 10 mL of prewarmed 1x PBS. The 1x PBS was discarded, and cells were harvested with 6 mL of fresh and prewarmed DMEM medium. The cell suspension was centrifuged for 2 min at 200 x *g*. The supernatant was discarded, and the cell pellet was resuspended in 4 mL fresh and prewarmed DMEM medium. For cell counting, 10 µL of the cell suspension were mixed with 10 µL of Tryptan blue (1:1) and added to a counting chamber according to Neubauer. The number of cells per mL was calculated by using the CytoSMART cell counter from Axion Biosystems. Based on that, the total volume required for $2x10^6$ cells/ T75 flask was calculated and filled up to 12 mL. The cell suspension was put into a fresh T75 flask and incubated for 2-3 days at 37 °C.

4.3.7 Cryo Stocks of HEK293T Cells

To generate cryo stocks of HEK293T cells, cell splitting was performed as described in 4.3.6. However, this time, the cell pellet was resuspended in 1 mL DMEM medium before counting. To generate cryo stocks, $6x10^6$ cells per cryo tube were required. After counting, cells were again centrifuged for 2 min at 200 x *g.* The 1 mL supernatant was discarded. The cell pellet was mixed with the corresponding volume of the Cryo Stock Buffer to get $6x10^6$ cells per cryo tube and stored in a freezing container filled with 100% (v/v) isopropanol at -80 °C for 2 days. Finally, cells were stored in liquid nitrogen.

4.3.8 Transfection of HEK293T Cells

Transfection of HEK293T cells was performed by using the jetPRIME® transfection kit from Sartorius Polyplus. For each reaction, 500 µL jetPRIME® buffer, 20 µL jetPRIME® reagent and 10 µg DNA were used. As two different DNA constructs were used (pcDNA3.1(+)-HA-Ubl1 (1.7 µg/µL) and pcDNA3.1(+)-EYFP (EYFP: Enhanced Yellow Fluorescent Protein) (3.6 µg/µL)), the double amount of transfection mix was assembled. Firstly, the DNA was diluted in the jetPRIME® buffer. The DNA sample was vortexed for 10 s and shortly centrifuged. Then the jetPRIME® reagent was added. Sample was again vortexed and centrifuged, followed by a 10 min incubation at room temperature. The final transfection mix (pcDNA3.1(+)-EYFP: 535.8 µL; pcDNA3.1(+)-HA-Ubl1: 532.8 µL) was added to each T75 flask. Transfected cells were incubated for 24 h at 37 °C and high humidity.

4.3.9 Immunoprecipitation (IP)

4.3.9.1 Harvesting of HEK293T Cells

The medium of transfected cells was discarded. Cells were washed with 5 mL 1x PBS (no prewarming necessary). Cells were scraped in 2 mL 1x PBS and transferred into an autoclaved 2 mL reaction tube. For harvesting, cells were centrifuged for 2 min at 1,400 x *g* (4 °C). The supernatant was discarded.

4.3.9.2 Cell Lysis of HEK293T Cells

The cell pellet was resuspended in 1 mL of IP Lysis Buffer. Cell lysis was continued by 30 min overhead-rotation at 4 °C. Samples were centrifuged for 15 min at 15,000 x *g* (4 °C). The cell lysate was transferred to a fresh tube and the protein concentration was determined by using

NanoDrop. The lysate amount containing 1 mg protein was calculated as a total input for IP. An SDS- sample of the cell lysate was taken. For that, 45 µL of the 1 mg lysate sample and 15 µL of 4x Reducing Laemmli Buffer were mixed.

4.3.9.3 Pulldown of HA-Ubl1

For the pulldown, autoclaved 1.5 mL reaction tube were prepared and filled with 50 µL of Anti-HA magnetic beads. Reaction tubes were put into a magnetic rack to separate beads from storage buffer. Storage buffer was removed, and beads were equilibrated with 1 mL IP wash buffer. Samples were incubated for 5 min at 4 °C while overhead-rotation. After equilibration, remaining IP wash buffer was discarded. The corresponding amount of cell lysate was added to the beads. Binding occurred in overhead rotation for 2 h at 4 °C. After incubation, magnetic beads binding the target protein and remaining cell lysate were separated by using a magnetic rack. Beads were washed 3 times with 1 mL IP wash buffer (2 min overhead rotation at 4 °C). To elute HA-tagged Ubl1, two different elution techniques were used based on later analysis methods.

4.3.9.4 Elution by SDS and Heat

After washing, HA-Ubl1 was eluted using 50 µL IP SDS elution buffer. SDS was used to provide all proteins with a negative charge for later SDS-PAGE analysis. Samples were incubated for 10 min at 95 °C for elution. Magnetic beads were separated from the eluate using the magnetic rack. The protein concentration of the eluate was determined by using NanoDrop. A gel sample of the eluate was taken (15 µL 4x Reduced Laemmli Buffer + 45 µL eluate).

4.3.9.5 Elution by Using an Acidic Buffer

As SDS and heat affects the structure and charge of proteins, another elution technique was used for mass spectrometry samples. Here, the target protein was eluted using 50 µL of IP acidic elution buffer using an acidic pH for elution. Beads were incubated for 10 min at room temperature. After incubation, the appropriate amount of neutralization buffer was added to get pH ≈ 6.5 (here 150 µL neutralization buffer was used). Beads were removed using the magnetic rack. The protein concentration of the eluate was determined by using NanoDrop. For later MS analysis, aliquots of 10 µM HA-Ubl1 were prepared.

4.3.10 SDS-PAGE of Different Ubl1 Constructs

SDS-PAGE was used to quantify successful expression and purification of samples allowing protein separation by mass.[92,93] Based on the MW of His-Ubl1 (15.55 kDa), Ubl1 (12.49 kDa) and HA-Ubl1 (13.79 kDa), a Tris- Tris-Tricine SDS-PAGE was performed (Table 15). Samples were diluted in 15 µl 6x Laemmli SDS-buffer and boiled at 95 °C for 10 min for protein denaturation. Samples were centrifugated for 10 min at 10,000 x *g*. Samples and the protein standard marker were loaded on the corresponding SDS gel and cable-connected to the electrophoresis power supply 3501 XL. When the blue running front was nearly at the bottom, the tank was taken off. After electrophoresis, the gel was washed four times with pure water prior to staining for 1 h with Coomassie (Blauer Jonas). Destaining was not required. Finally, the separated proteins were visualized by a gel documentation system.

4.3.10.1 Tris-Tricine SDS-PAGE of Ubl1 Constructs

A running and stacking SDS-gel were prepared (Table 15). 3 µL of marker (Spectra Multicolor Low Range Ladder) and 10 µL of each sample were loaded. The electrophoresis was performed with a constant current strength of 40 mA per gel in running buffer. For that, the gel chamber was filled with 1x cathode buffer and surrounded with 1x anode buffer.

Table 15: Tris-Tricine SDS-PAGES – protocol for 2 gels.

Component	20% Separating Gel	5.6% Stacking Gel
ddH_2O	-	3.2 mL
Gel buffer (3x)	5 mL	2.1 mL
Glycerol	1.6 mL	-
Acrylamide:Bisacrylamide (34.3%:1.1%)	8.5 mL	1 mL
APS (10% (w/v))	50 µL	50 µL
TEMED	25 µL	25 µL

4.3.11 Western Blot Analysis of Different Ubl1 constructs

Western blot analysis was used to identify NSP3-Ubl1 after protein separation by an SDS-gel.[94] Further, the house-keeping and cytoskeleton protein β-Actin with a molecular weight of 42 kDa was used as a loading reference.[178] Samples were separated using SDS-PAGE as described in section 4.3.10. For generating a Western-Blot-Sandwich, a PVDF-membrane was

activated by 100% (v/v) EtOH and six Watman papers were soaked in (1x) Towbin buffer. The SDS gel was put onto the PVDF-membrane and surrounded with three pieces of Watman papers from each side. The transfer was performed in a blotting chamber for 30 min at 25 V (1 A). To check if the transfer was successful, the membrane was incubated with Ponceau-S-solution shortly and washed four times with water. Visible bands gave a hint on a successful protein transfer. Then, the membrane was blocked in blocking solution for 1 h at room temperature to saturate all antibody sites. As milk contains the phosphoprotein casein, 3x (w/v) BSA in 1xTBS-T was used as a blocking solution to prevent cross-reaction with the phospho-specific antibody.[179] The blocking solution was discarded, and the primary antibody diluted in blocking solution was added to the membrane. Antibodies and the corresponding dilutions used in this experiment are listed in Table 16.

Table 16: List of antibodies used for Western Blot detection of Ubl1 constructs.

Construct	Primary Antibody	Secondary Antibody
HA-Ubl1 (HEK)	HA-tag Monoclonal Antibody mouse IgG1 (1:1000)	Goat anti-Mouse lgG (H+L) Secondary Antibody HRP conjugated (1:10000)
His-Ubl1 (*E. coli*)	6x-His Tag Monoclonal Antibody (HIS.H8) (1:1000)	Goat anti-Mouse lgG (H+L) Secondary Antibody HRP conjugated (1:10000)
HA-Ubl1 (HEK), His-Ubl1 & Ubl1 (*E. coli*)	Phosphoserine/threonine/tyrosine Antibody (polyclonal, rabbit, IgG) (1:250)	Goat anti-Rabbit Secondary Antibody HRP conjugated (1:1000)
β-Actin	β-Actin Monoclonal Antibody (mouse, IgG1) (1:5000)	Goat anti-Mouse lgG (H+L) Secondary Antibody HRP conjugated (1:10000)

Finally, the membrane was stored at 4 °C overnight under gentle shaking. The next day, the primary antibody solution was recycled and frozen at -20 °C for further use. The membrane was washed three times in 1x TBS-T for 5 min. The corresponding secondary antibody was diluted in blocking solution, added to the membrane, and incubated for 1 h at room temperature under gentle shaking. The antibody solution was discarded, and the membrane was washed again three times for 5 min in 1x TBS-T. The membrane was visualized using the

Bio-Rad Universal Hood III documentation system. For signal detection, a 1:1 dilution of Super Signal Peroxidase Solution and Enhancer Solution were added on the membrane and incubated for 1 min in a dark room. The signal accumulation mode was utilized in high resolution mode, with 40 frames in 200 s. The protein ladder was visualized separately by using the colorimetric mode. Finally, images were merged and analyzed.

4.3.12 Mass Spectrometry Analysis of Ubl1 Constructs

Following protein purification, samples were adjusted to a final concentration of 10 µM for mass spectrometry analysis. The measurements were carried out by Ian Gering at the Institute of Biological Information Processing 7 (IBI-7), Forschungszentrum Jülich. Intact protein mass spectrometry was employed to confirm the identity of the expressed proteins based on their molecular mass and to assess potential post-translational modifications present in constructs produced in *Escherichia coli* and HEK293 cells.

4.3.13 Phosphatase Assay

To dephosphorylate Ubl1 constructs, a Lambda Protein Phosphatase Kit from New England Biolabs (Ipswich, USA) was ordered including 10x PMP Buffer, 10 mM $MnCl_2$ and a Lambda phosphatase (LP) stock (400,000 units/mL). As a standard reaction mix, 40 µL of 10 µM Ubl1 were mixed with 5 µL of 10x PMP Buffer, 5 µL of 10 mM $MnCl_2$ and 1 µL LP stock (400,000 units/mL, New England Biolabs, Ipswich, USA). As a reference, an equal sample was prepared excluding the LP (40 µL 10 µM Ubl1, 5 µL 10x PMP Buffer, 5 µL 10 mM $MnCl_2$). Samples were incubated for 30 min at 30 °C. Dephosphorylation was stopped by adding 10 µL of 6xLaemmli-buffer and cooking for 10 min at 95 °C. Samples were analyzed by SDS-PAGE and western blotting.

4.3.14 Native-PAGE

Native-PAGE was used to separate polypeptides by their charge to mass ratio.[180] For that samples were prepared under non-denaturing conditions to maintain the protein's folding. Samples were mixed with 2x Sample Buffer (1:1). A running and stacking SDS-gel were prepared. 10 µL of Native Marker (Native Mark Unstained Protein Standard) and 10 µL of protein samples were loaded on the gel. The gel chamber was filled with 1xTG-buffer (without SDS). The exact gel composition is displayed in Table 17. To prevent protein denaturation

because of heating, the whole system was cooled on ice during the run. The system was cable-connected to the electrophoresis power supply 3501 XL. After around 1 h running at 120 mV, the blue running front was nearly at the bottom and the tank was taken off. After electrophoresis, the gel was washed four times with pure water prior to staining for 1 h with Coomassie (Blauer Jonas). Destaining was not required. Finally, the separated proteins were visualized by a gel documentation system.

Table 17: Native-PAGE protocol for 2 gels.

Component	15% Separating Gel	5.6% Stacking Gel
Native-PAGE Gel Buffer	4.89 mL	8.55 mL
Acrylamide:Bisacrylamide (30%:0.8%)	5 mL	1.34 mL
APS (10% (w/v))	100 µL	100 µL
TEMED	10 µL	10 µL

4.3.15 Immunofluorescence (IF) Staining of HA-Ubl1

4.3.15.1 Passaging of HEK293T Cells for IF

For later IF experiments, HEK293T cells were split into a 24-well plate, each well with a diameter of 1.6 cm. For that, cells were previously cultivated as described in 4.3.5. As a cell confluence of 80% was reached, cells were split into a 24-well plate. For that, each well was seeded with 50,000 HEK293T cells in 500 µL DMEM medium. Additionally, a round cover slip was added and fixed with a pipette tip on the bottom of the well. Cover slips were used for later fluorescence microscopy. Cells were incubated overnight in an incubator at 37 °C and high humidity.

4.3.15.2 Transfection of HEK293T Cells with Different pcDNA Constructs

Transfection of HEK293T cells was performed by using the jetPRIME® transfection kit from Sartorius as described in 4.3.8. Because of a lower well volume, the amount of transfection mix was adjusted. For each reaction, 50 µL jetPRIME® buffer, 1 µL jetPRIME® reagent and 0.5 µg DNA were used. As two different DNA constructs were used (pcDNA3.1(+)-HA-Ubl1 (1.7 µg/µL) and pcDNA3.1(+)-EYFP (3.6 µg/µL)), the double amount of transfection mix was prepared. The final transfection mix (pcDNA3.1(+)-EYFP: 51.14 µL; pcDNA3.1(+)-HA-Ubl1:

51.29 µL) was added to each well. Transfected cells were incubated for 24 h at 37 °C and high humidity.

4.3.15.3 Fixation of HEK293T Cells

24 h post-transfection, cells were fixed for further IF experiments. First, the used DMEM medium was aspirated. Then, 500 µL of PFA solution was added to each well of the 24-well-plate. Well plates were sealed and incubated for 15 min at room temperature on an elliptic shaker. The PFA solution was discarded, and cells were washed three times with 500 µL of 1x PBS. Lastly, 500 µL 1x PBS was added and cells were incubated for 24 h at 4 °C.

4.3.15.4 In-Cell Western Blot

This experiment was only performed with cells expressing HA-Ubl1 as this protein is not labelled with a fluorescent protein. The cells containing the EYFP control were not included here, as the plasmid already encoded for a fluorescent protein. Each incubation step was performed on an elliptic shaker.

Firstly, cells were washed with 500 µL 1x PBS. To allow the entrance of antibodies, the membrane was permeabilized. To do so, 500 µL permeabilization buffer was added per well, followed by an incubation time of 15 min. After the buffer was removed, cells were 3 times washed with 500 µL 1x PBS. Unspecific antibody binding was prevented by adding 500 µL of milk blocking solution per well. After 1 h incubation on an elliptic shaker, fixed cells were stained with 200 µL of an HA-specific antibody (1:1,000; Anti-HA mouse) and incubated for 1 h. The primary antibody was recycled, and cells three times washed with 500 µL of IF washing buffer (each time 10 min). After washing, the secondary antibody (Goat Anti-mouse IgG (H+L) from goat Alexa647; 1:5,000) was added and for another 60 min incubated. Lastly, cells were again washed with wash buffer as previously described.

4.3.15.5 Staining of Nucleic Acids

To visualize nucleic acids, HEK293T cells containing HA-Ubl1 and EYFP were stained with DAPI. For that 100 µL of diluted DAPI staining solution (1:5,000 in 1x PBS) was added to each well and incubated for 5 min. As the staining solution was removed, cells were washed with

500 µL of 1x PBS for 10 min on an elliptic shaker. Cells were stored in 500 µL fresh 1x PBS until imaging.

4.3.15.6 Mounting of Cells

To prepare the samples for mounting and later visualization, 8 µL of mounting medium was placed on a microscopy slide. The processed cover slips including adherent HEK cells were taken out of the well, gently washed in ddH_2O and shortly dried on a paper tissue. Finally, the cover slip was put onto the microscopy slide being fused by the mounting medium in between.

4.3.15.7 Fluorescence Microscopy of Stained HEK293T Cells

To visualize stained cells, the BZ-X810 fluorescence microscope from Keyence (Osaka, Japan) was used. Experimental parameters and the according dyes are listed in Table 18. Nucleic acids were colored using the blue-fluorescent dye DAPI, with an excitation peak monitored at 350 nm and an emission peak recorded at 465 nm. The target was marked with a fluorescent antibody (fluorophore: Alexa647). Alexa Fluor 647 is a red-fluorescent dye, with excitation suited at 650 nm and emission at 671 nm. The positive control (pcDNA3.1(+)-EYFP) already encoded for the enhanced yellow fluorescent protein (EYFP) and hence was not additionally stained. EYFP was excited at 513 nm and emission was detected at 527 nm.

Table 18: Experimental settings during fluorescence microscopy.

Construct	Dye/ Fluorophore	Gain	Exposure [s]	Filter
Nucleic acids	DAPI	x2 (6dB)	1/1.2	49000 ET - DAPI
HA-Ubl1	Alexa647	x2 (6dB)	4.5	49006 ET – Cy5
Positive Control	EYFP	x2 (6dB)	1/20	49011 ET – FTC / Alexa Fluor 488 / Fluo3 / Oregon Green

4.4 Results

The following section presents a comprehensive analysis of phosphorylation observed in Ubl1 following its heterologous expression in *E. coli* and subsequently in HEK cells. For advanced analytical approaches such as MS, the availability of highly pure and homogeneous protein is essential. Furthermore, sufficient protein yield represents a critical requirement for downstream biochemical and cell-based assays. Therefore, this chapter focuses on the expression of HA-tagged Ubl1 in human cells, followed by its purification through pulldown assays. Initial efforts were directed toward optimizing expression levels of the target protein in HEK cells. The successful production of pure Ubl1 protein in *E. coli* was described in Chapter 1. The results presented herein include data from Western blotting, SDS-PAGE, MS, phospho-specific assays, Native-PAGE, and immunofluorescence microscopy.

4.4.1 Phosphorylation Event of Recombinantly Expressed Ubl1 in *E. coli* Cells

Following the isolation of Ubl1 by SEC (for details see Chapter 1), the collected fractions were analyzed via SDS-PAGE (Figure 26A). The presence of double bands in each fraction indicated potential PTMs. Notably, these double bands became more distinct at lower protein concentrations, as observed in fractions 1 and 4.

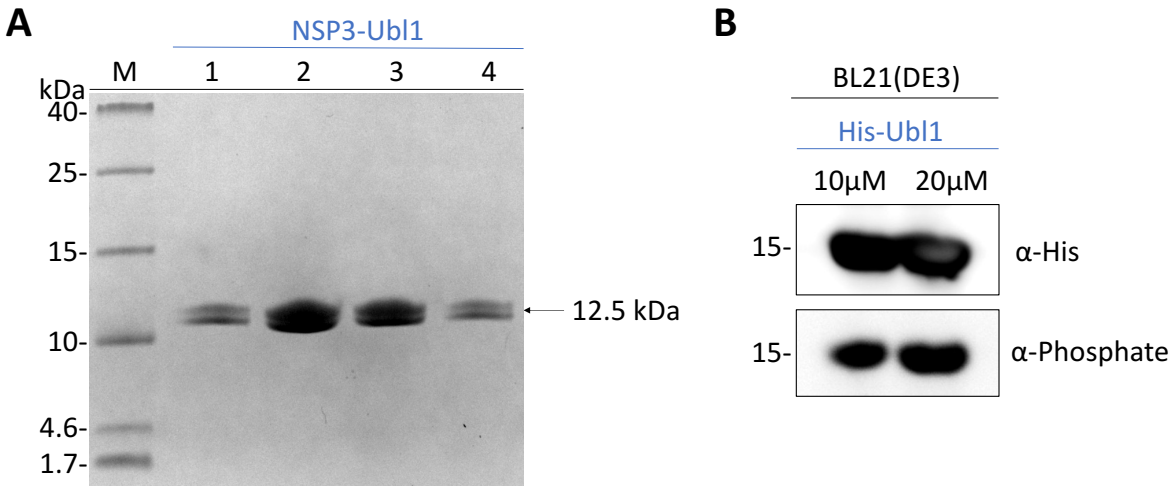
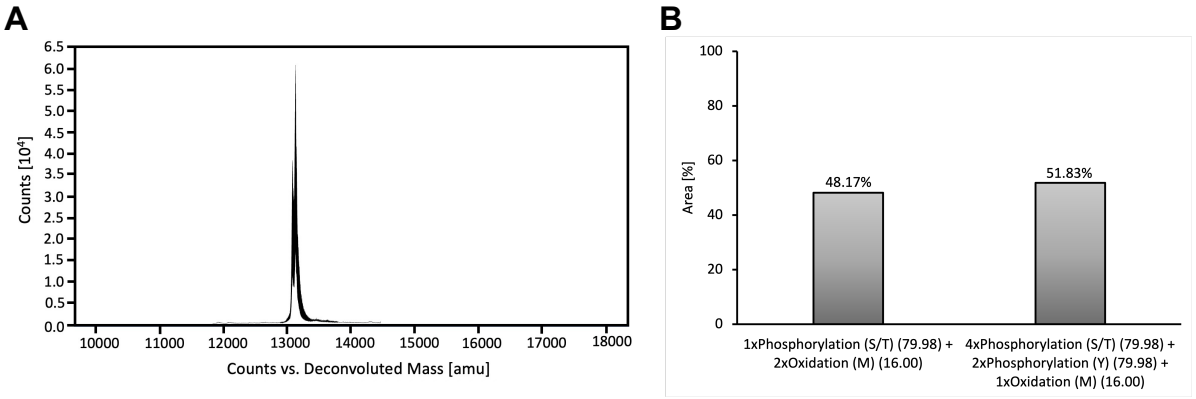


Figure 26: Ubl1 becomes phosphorylated during expression in *E. coli* cells. (A) Samples of fractions after SEC were taken and analyzed by a 20% (v/v) Tris-Tricine SDS-PAGE. Protein bands were compared to the marker (M) (Spectra Multicolor Low Range Protein Ladder). Double bands of low concentrated Ubl1 samples were detected in each fraction (12.5 kDa). **(B)** Western Blot analysis of recombinantly expressed His-tagged Ubl1 in *E. coli* cells followed by an anti-Phosphoserine/threonine/tyrosine immunodetection. Samples with concentrations of 10 µM and 20 µM were analyzed. His-Ubl1 was detected at around 15 kDa. Phosphorylation was confirmed by western blot analysis.

Previous studies have shown that PTMs can affect the electrophoretic mobility of cytosolic proteins during SDS-PAGE, a phenomenon known as gel shifting.[181] To further characterize these modifications, the samples were subjected to mass spectrometry analysis, performed by Ian Gehring at the Research Center Jülich (IBI-7). The corresponding mass spectrometry data are presented in Figure 27.

As illustrated in Figures 27A and 27B, intact protein mass spectrometry confirmed that Ubl1 expressed in *E. coli* undergoes posttranslational phosphorylation. The mass spectrum showed a prominent peak at approximately 13,200 amu, corresponding to a molecular mass of 13.2 kDa, which is consistent with the expected molecular weight of Ubl1 (12.5 kDa) (Figure 27A). The absence of additional peaks indicates that the sample is highly pure. Furthermore, approximately 48.17% of the protein was found to be singly phosphorylated, while 51.83% exhibited *hexa* phosphorylation, suggesting the concurrent activity of endogenous kinases and phosphatases within *E. coli* cells (Figure 27B). These two phosphorylation states were represented by the MS peak, which was split into two maxima (Figure 27A). Putative phosphorylation sites on SARS-CoV-2 Ubl1 are highlighted in the amino acid sequence shown in Figure 27C. While serine, threonine, and tyrosine residues are typical targets for phosphorylation, the exact modified residues have yet to be identified.



C

APTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYL
FDESGEFKLASHMYCSFYPPDE

Figure 27: Mass spectrometry analysis of Ubl1 expressed in *E. coli*. (A) Deconvoluted mass spectrum of isolated HA-Ubl1. Signal intensity (counts) is shown as a function of the molecular mass in amu (atomic mass units, equivalent to Daltons). **(B)** Bar chart presenting the amount of modified amino acids in Ubl1. Serine (S), threonine (T) or tyrosine (Y) residues are possible phosphorylation sites. Methionine (M) becomes oxidized. **(C)** Amino acid sequence of Ubl1 presented in the one-letter amino acid code. Possible phosphorylation sites are marked in grey.

Phosphorylation of His-tagged Ubl1 was further validated by Western blot analysis (Figure 26B). Using an anti-His primary antibody, His-Ubl1 was detected in samples at concentrations of 10 µM and 20 µM. Corresponding prominent bands at 15.5 kDa also reacted with phospho-specific antibodies, confirming the presence of phosphorylation. Collectively, MS and Western blot data provide clear evidence that Ubl1 expressed in *E. coli* undergoes PTM in the form of phosphorylation.

To investigate the reversibility of Ubl1 phosphorylation, a phosphatase assay was performed using the Lambda Phosphatase Kit from New England Biolabs. Two conditions were tested: in the absence (-LP) and presence (+LP) of LP. Samples were analyzed by SDS-PAGE (Figure 28A), and corresponding phosphorylation-specific detection was carried out by Western blot using anti-phosphoserine/threonine/tyrosine antibodies (Figure 28B). As shown in Figure 28, Ubl1 was detected at its expected molecular weight of approximately 12.46 kDa. Notably, in the +LP condition, two additional bands appeared around 22 kDa and 25 kDa on the SDS-PAGE gel (Figure 28A). These bands were not visible in the phosphorylation-specific Western blot (Figure 28B), indicating that they likely correspond to the LP enzyme rather than Ubl1 dimers. Crucially, Ubl1 remained detectable by phospho-specific antibodies even after LP treatment, suggesting that dephosphorylation was ineffective under the conditions used.

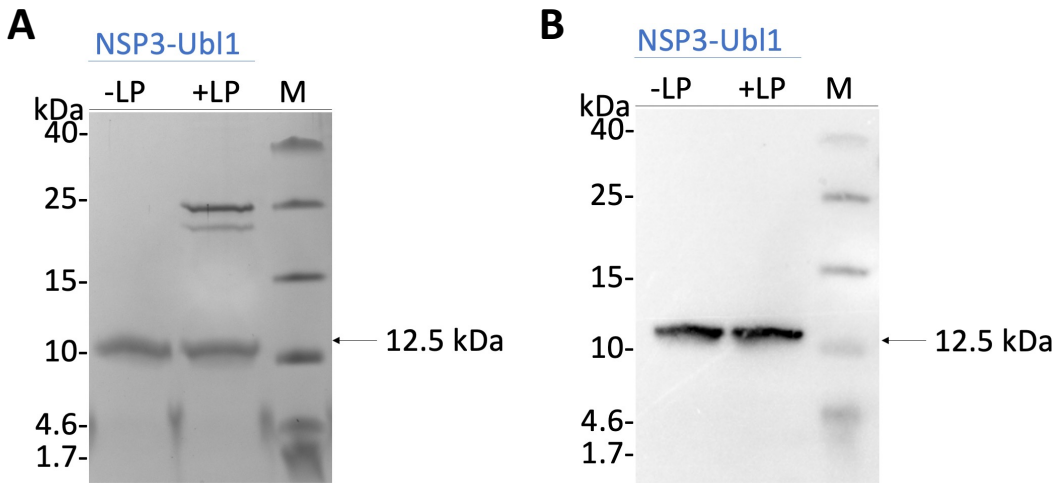


Figure 28: Western blot analysis of Ubl1 (*E. coli*) after dephosphorylation. (A) Ubl1 samples in absence (-) and presence (+) of Lambda Phosphatase (LP) were taken and analyzed by a 20% (v/v) Tris-Tricine SDS-PAGE. Protein bands were compared to the marker (M) (Spectra Multicolor Low Range Protein Ladder). **(B)** Phosphorylated protein was identified by Western Blot analysis at the proposed size (12.5 kDa) even in the presence of LP.

To determine whether the double bands observed in SDS-PAGE (Figure 26A) correspond to distinct protein populations with different PTMs, a Native-PAGE was conducted. In this

approach, proteins were prepared in a non-reducing, non-denaturing sample buffer, thereby preserving their secondary structure and native charge.[180] As shown in Figure 29, two distinct protein populations were observed at approximately 10 kDa and 21 kDa, independent of both the presence of 1 mM TCEP and the protein concentration (10 µM or 20 µM). Western blot analysis confirmed that both bands represent phosphorylated forms of Ubl1.

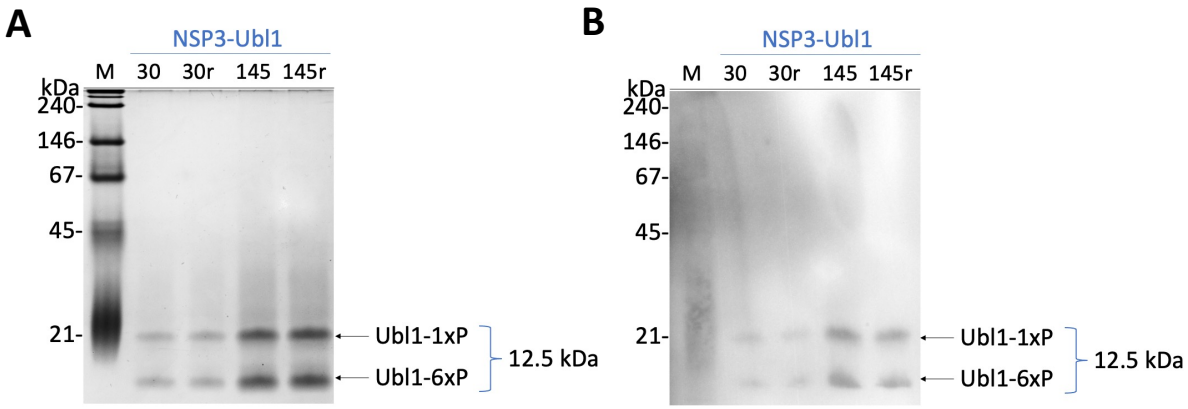


Figure 29: Native-PAGE of reduced and non-reduced Ubl1 samples. (A) Native-PAGE of different concentrated Ubl1 samples (30 mM and 145 µM). Samples were reduced (r) by using 1 mM TCEP. Ubl1 samples were analyzed by a 15% (v/v) Native-PAGE. Protein bands were compared to the marker (M) (Native Mark Unstained Protein Standard). Two different protein populations were detected. **(B)** Western blot analysis of Ubl1 samples after Native-PAGE. The two protein populations were identified as phosphorylated using the anti-phosphoserine/threonine/tyrosine antibody.

Taken together, these findings indicate that neither protein concentration nor reducing conditions significantly affect the electrophoretic mobility of Ubl1 under native conditions. The observed banding pattern is therefore most likely attributable to phosphorylation-dependent changes in conformation or charge.

4.4.2 Expression and Localization of HA-Ubl1 in Human Cells

To rule out *E. coli*-specific artifacts as the cause of Ubl1 phosphorylation and to investigate whether similar modifications occur in human cells, HA-tagged Ubl1 (HA-Ubl1) was expressed in HEK293T cells. These experiments were initially performed during a research exchange at the Bartenschlager Lab (German Cancer Research Center, Institute for Virology, Heidelberg) and subsequently replicated at the Institute for Physical Biology (IPB), Heinrich Heine University Düsseldorf. To evaluate transfection efficiency and subcellular localization of HA-Ubl1, immunofluorescence staining was conducted. Nuclei were stained with DAPI, HA-Ubl1 was detected using an Alexa Fluor 647-conjugated anti-HA antibody, and EYFP was used as a transfection and expression control. Transfection efficiency was calculated by determining the

percentage of HA-Ubl1-positive cells relative to the total number of DAPI-stained nuclei. EYFP demonstrated a transfection efficiency of 13.43%, while HA-Ubl1 reached 5.88%. As shown in Figure 30A, EYFP was predominantly localized to the cytosol of HEK293T cells. Similarly, HA-Ubl1 was mainly detected in the cytoplasm, indicating a comparable subcellular distribution.

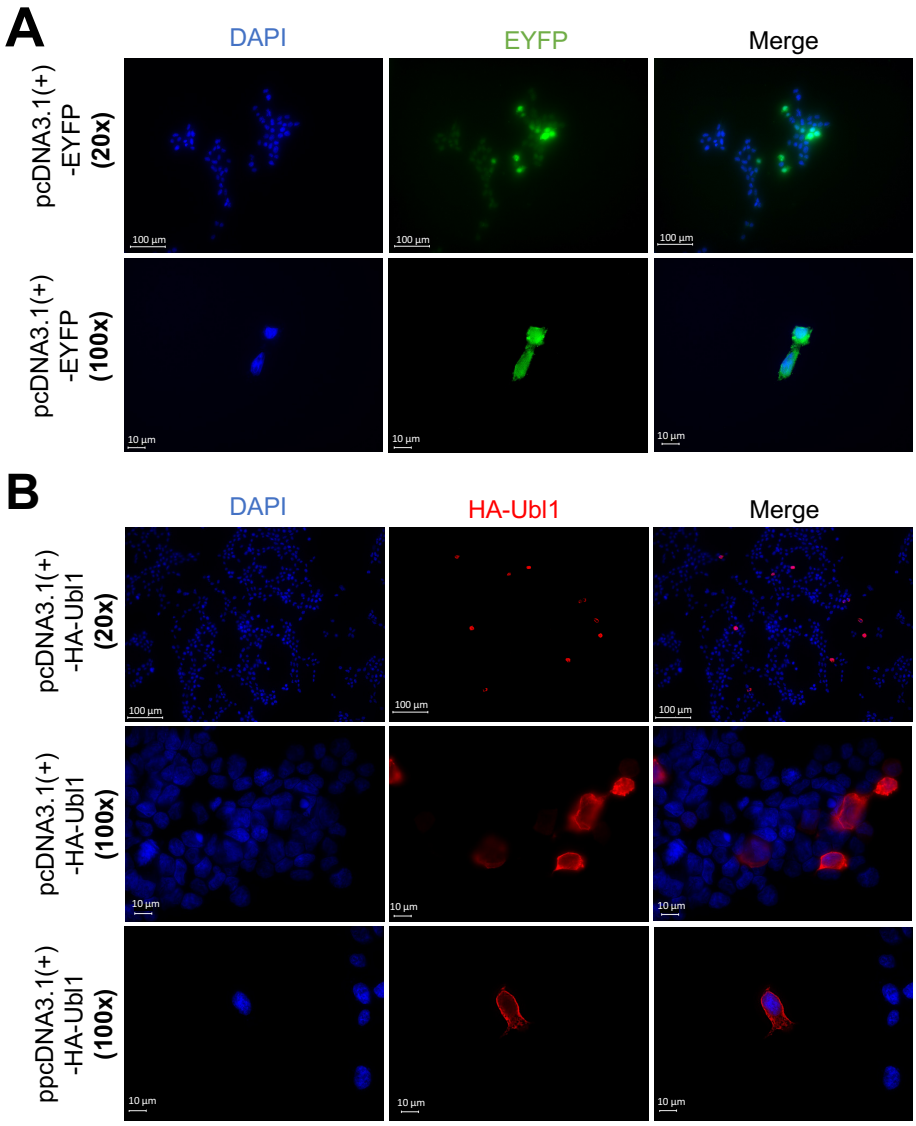


Figure 30: HA-Ubl1 is localized in the cytosol of HEK293T cells. Cells were fixed and stained with antibodies of the given species and viewed 48 h post-transfection using the Keyence BZ-X810 fluorescence microscope. Nuclear DNA was stained with DAPI. Cells were imaged at 20x and 100x magnifications. **(A)** As a positive control, cells expressing EYFP were used. **(B)** HA-Ubl1 was visualized using immunofluorescence staining and the Alexa467 fluorophore.

In the subsequent step, isolated EYFP and HA-Ubl1 samples were analyzed by Western blotting, using β-Actin as a cytoskeletal loading control. As shown in Figure 31A, HA-tagged Ubl1 was detected in both the cell lysate and the eluate, indicating efficient expression and successful purification. As expected, no EYFP signal was observed in either fraction, as the

EYFP construct lacks an HA tag. The β-Actin loading control was consistently detected across all cell lysate samples (Figures 31A and 31B). Importantly, phosphorylation-specific Western blot analysis revealed no detectable phosphorylation of HA-Ubl1 following expression in HEK293T cells (Figure 31B), suggesting that the PTM observed in *E. coli* is not replicated in the mammalian expression system.

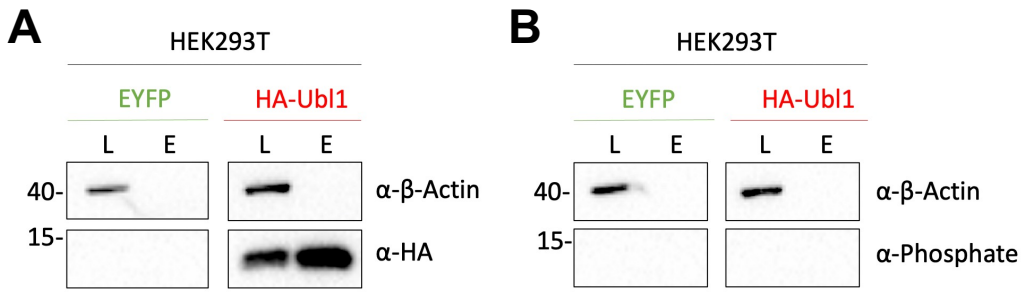
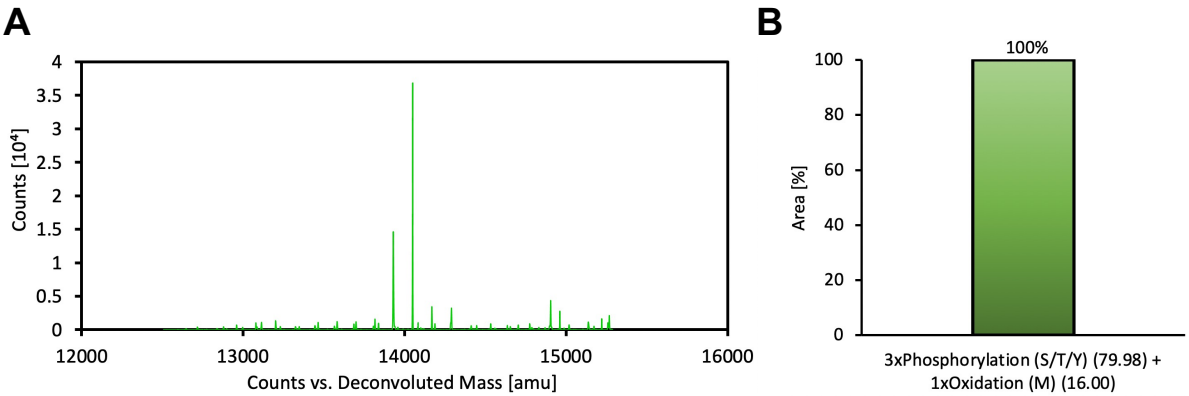


Figure 31: Western blot analysis of HA-Ubl1 after immunoprecipitation. Expression of HA-Ubl1 in HEK293T cells was confirmed by western blotting. EYFP served as a reference. Samples of the lysate (L) and eluate (E) were taken and analyzed by a western blot. Protein bands were compared to the marker (M) (Spectra Multicolor Low Range Protein Ladder). β-Actin served as a loading control. **(A)** HA-tagged Ubl1 was identified by Western Blot analysis at the proposed size (13.79 kDa). **(B)** Phospho-Blot of EYFP and HA-Ubl1 samples. Phosphorylation of HA-Ubl1 was not confirmed.

Although HA-Ubl1 was not detected as phosphorylated in the previous Western blot analysis (Figure 31B), a purified and isolated HA-Ubl1 sample was submitted for MS to definitively assess its phosphorylation status. The mass spectrum revealed a prominent peak at approximately 14,100 amu (14.1 kDa), which is in good agreement with the expected molecular weight of HA-Ubl1 (13.8 kDa) (Figure 32A). However, the presence of additional peaks suggests that the sample is less pure compared to the Ubl1-*E. coli* preparation (Figure 27A).

Unexpectedly, MS analysis revealed that HA-Ubl1 was phosphorylated, with each monomer carrying three phosphorylation events on serine, threonine, or tyrosine residues (Figure 32B), thereby contradicting the Western blot findings. Putative phosphorylation sites are highlighted in green within the amino acid sequence presented in Figure 32C; however, the exact residues remain to be identified. However, further investigation is required to elucidate the phosphorylation patterns of Ubl1 expressed in both *E. coli* and HEK293T cells, and to determine their functional significance in the context of the SARS-CoV-2 replication cycle. This research is ongoing in collaboration with Lara Rheinemann (AG Pichlmair, Technical University of Munich), a member of the COVIPA consortium.



C

MGYPYDVPDYAGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYTVELGTEVNEFACVVADAVIKT
LQPVSELLTPLGIDLDEWSMATYYLFDESGEFKLASHMYCSFYPPDE

Figure 32: Mass spectrometry analysis of HA-Ubl1 expressed in HEK cells. (A) Deconvoluted mass spectrum of isolated HA-Ubl1. Signal intensity (counts) is shown as a function of the molecular mass in amu (atomic mass units, equivalent to Daltons). **(B)** Bar chart presenting the amount of modified amino acids in Ubl1. Serine (S), threonine (T) or tyrosine (Y) residues are possible phosphorylation sites. Methionine (M) becomes oxidated. **(C)** Amino acid sequence of HA-Ubl1 presented in the one-letter amino acid code. The HA-tag is colored in grey. Possible phosphorylation sites are marked in green.

4.5 Discussion

The molecular mechanisms underlying the binding interactions of SARS-CoV-2 proteins remain poorly understood. Using IF, IP, MS and Western blot analyses, Ubl1 was localized to the cytosol of HEK293T cells. Furthermore, PTMs of NSP3-Ubl1 expressed in both *E. coli* and HEK cells were observed.

IF staining combined with fluorescence microscopy demonstrated that the expressed HA-tagged Ubl1 construct localizes to the cytosol of HEK293T cells, as evidenced by the distinct Alexa647 fluorescence signal of the HA-tag separate from the DAPI-stained nucleus (Figure 30B). Considering the position of the NSP3-Ubl1 domain at the pore of DMVs, the Ubl1 domain is oriented towards the cytosol, which is consistent with our findings.[46,56] Wolff and colleagues further confirmed, using cryo-electron tomography, that GFP-tagged Ubl1 localizes to the cytoplasm of human cells.[56] Additionally, colocalization of the N-protein and full-length NSP3 has been observed in the cytosol of Huh-7 cells.[61] In summary, the intracellular localization of single-domain Ubl1 expression in human cells aligns with that of full-length NSP3.

Surprisingly, Ubl1 became phosphorylated during expression in *E. coli* cells (Figure 26 and 27). It is well established that phosphorylation alters protein properties such as charge,

hydrophobicity, stability, and solubility.[181] Since phosphate groups carry a negative charge, the net charge of heavily phosphorylated proteins decreases after modification, resulting in increased electrophoretic mobility during SDS-PAGE. Consequently, strongly phosphorylated proteins migrate faster toward the positively charged anode, causing gel shifting and the appearance of double bands (Figure 26A).

A similar pattern was observed in Native-PAGE (Figure 29), where two distinct Ubl1 populations were detected at approximately 10 kDa and 21 kDa. This suggests the presence of two differently phosphorylated Ubl1 species, with the *hexa* phosphorylated form (6xP-Ubl1) exhibiting increased mobility. Ubl1 dimerization was excluded, as bands observed at 22 kDa and 25 kDa in SDS-PAGE (Figure 28A) disappeared in the anti-phospho Western blot (Figure 28B), indicating these bands likely correspond to the LP rather than Ubl1 dimers. Unfortunately, dephosphorylation of Ubl1 by LP was unsuccessful (Figure 28). Future experiments should consider optimizing parameters such as incubation time and the ratio of target protein to lambda phosphatase.

In research, bacterial cells are generally not the preferred system for generating PTMs on recombinantly expressed proteins, as *E. coli* lacks the necessary enzymatic machinery for most PTMs.[182,183] In contrast, insect cells are commonly used as expression hosts since they provide phosphorylation and glycosylation, which help preserve the native functionality of proteins.[184] Alternatively, plasmids encoding kinases can be introduced into *E. coli* cells to enable phosphorylation of target proteins.[182] However, in 1999, Vincent and colleagues reported the presence of protein-tyrosine kinase and phosphotyrosine-protein phosphatase activities in *E. coli* BL21 cells, suggesting reversible phosphorylation of tyrosine residues.[185] They identified the phosphatase Wzb as specifically dephosphorylating phosphotyrosine residues, without affecting phosphothreonine or phosphoserine residues.[185] This observation may explain the heterogeneous phosphorylation pattern observed for Ubl1 expressed in *E. coli*: while 48.17% of the 10 µM Ubl1 sample was singly phosphorylated, 51.83% was phosphorylated six times (Figure 27). Potential phosphorylation sites of Ubl1 expressed in *E. coli* or HEK cells are indicated in Figure 33A (PDB: 7KAG). Serine residues are marked in blue (S20, S40, S69, S83, S93, S100, S105), threonine residues in green (T3, T6, T11, T24, T43, T48, T64, T73, T86), and tyrosine residues in red (Y18, Y42, Y87, Y88, Y103, Y107). Interestingly, the total number of phosphorylation sites corresponds closely to the number of tyrosine residues

in Ubl1, supporting the hypothesis of tyrosine kinase and phosphatase activity (Figure 33B). According to this model, a protein kinase catalyzes the phosphorylation of Ubl1 in the cytoplasm using ATP hydrolysis, while dephosphorylation is mediated by a protein phosphatase that regenerates ATP (Figure 33B).

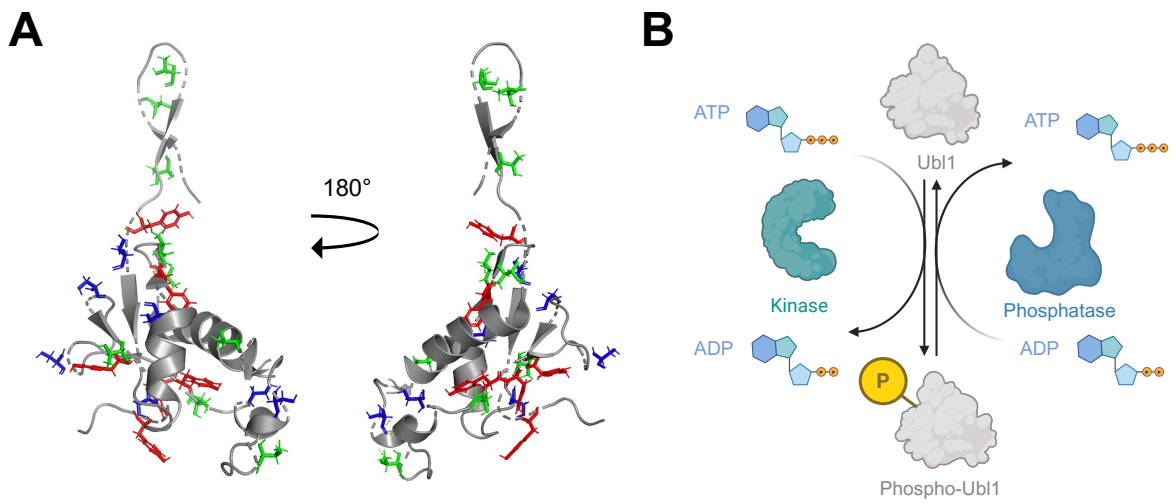


Figure 33: Phosphorylation of SARS-CoV-2 Ubl1. (A) Crystal structure of Ubl1 (PDB:7KAG). Possible phosphorylation sites are highlighted: serine residues (S20, S40, S69, S83, S93, S100, S105) are labelled in blue, threonine residues (T3, T6, T11, T24, T43, T48, T64, T73, T86) in green and tyrosine residues (Y18, Y42, Y87, Y88, Y103, Y107) in red. **(B)** Schematic illustration of Ubl1 phosphorylation mediated by a protein kinase. Phosphorylation is performed by using ATP hydrolysis. Dephosphorylation is mediated by a phosphatase resulting in ATP synthesis. This Figure was created in BioRender.com.

In this study, phosphorylation of SARS-CoV-2 Ubl1 expressed in HEK cells was detected by MS (Figure 32). The HA-tagged Ubl1 was found to be phosphorylated three times on serine, threonine, or tyrosine residues. However, since the HA-tag sequence itself contains three tyrosine residues (**Y**P**Y**DVPD**Y**A), it cannot be excluded that phosphorylation occurs on the tag rather than the Ubl1 protein. Therefore, further experiments using tag-less Ubl1 are required to definitively rule out phosphorylation of the HA-tag. Moreover, phosphorylation was not confirmed by Western Blot analysis using an anti-phosphoserine/threonine/tyrosine antibody (Figure 31B), indicating that these initial observations need to be investigated in greater detail for validation. Nevertheless, PTM of Ubl1 is plausible, as phosphorylation could regulate its binding interactions with the N-protein and viral RNA. Previous studies have demonstrated an interaction between Ubl1 and the N-protein by NMR spectroscopy.[67] Additionally, Ni and colleagues showed that SARS-CoV-2 NSP3-Ubl1 and RNA compete for binding to the N-protein in a dose-dependent manner.[61] Based on these findings, phosphorylation of Ubl1 might modulate its affinity for the N-protein and/or viral ssRNA.

Considering the composition of nucleic acids, RNA carries multiple negatively charged phosphate groups. This negative charge could lead to electrostatic repulsion with phosphorylated proteins, which also bear negatively charged phosphate groups. However, such repulsion does not necessarily preclude interaction, as exemplified by the viral N-protein: phosphorylated N-protein expressed in infectious bronchitis virus has been shown to bind viral RNA.[174]

To summarize, Ubl1 was localized to the cytosol of HEK293T cells and shown to undergo phosphorylation when expressed in both *E. coli* and HEK cells. Distinct phosphorylation states were detected, with MS indicating triply phosphorylated HA-tagged Ubl1, though modification of the tag itself cannot be excluded. These findings suggest that Ubl1 phosphorylation may influence its interactions with the viral N-protein and RNA. But our measurements need further validation to confirm or resolve some of the contradictory observations described above.

5. Chapter: Binding Interactions of NSP3-Ubl1 and Single-Stranded RNA studied by NMR Spectroscopy and MST

5.1 Abstract

The ubiquitin-like 1 (Ubl1) domain of SARS-CoV-1 NSP3 has been proposed to interact with single-stranded RNA (ssRNA) and may facilitate viral RNA translocation through the double-membrane vesicle (DMV) pore, potentially in cooperation with the viral N-protein. To investigate this, we examined the interaction between recombinant Ubl1 and ssRNA1 using NMR spectroscopy and microscale thermophoresis (MST) under varying experimental conditions. Thermodynamic predictions and 1D NMR analysis indicated that ssRNA1 remains largely unstructured and flexible in solution, with negligible formation of stable secondary structures. NMR chemical shift perturbation (CSP) analysis of Ubl1 revealed only weak, condition-dependent responses influenced by temperature, ionic strength, magnesium ions, and RNA-to-protein ratios. Residues showing minor perturbations (T24, L35, E37, G94) were distributed across different structural elements, suggesting non-specific, transient, or heterogeneous interactions rather than a defined RNA-binding site. MST measurements confirmed undetectable or very low-affinity binding, and no dissociation constant could be determined. Magnesium affected observed interactions in a non-systematic manner, and its direct influence on Ubl1 could not be isolated. Collectively, our results indicate that Ubl1 does not engage in specific or high-affinity binding to ssRNA1 under the tested conditions. While weak, transient, or context-dependent interactions cannot be excluded, further studies employing complementary biophysical and structural approaches are required to clarify the mechanistic and physiological role of Ubl1 in viral RNA recognition and genome translocation.

5.2 Introduction

5.2.1 Replication and Transcription of RNA in SARS-CoV-2 Induced Organelles

SARS-CoV-2 is a positive-sense single-stranded RNA virus that depends on the host cell's ribosomal machinery for translation of its genomic RNA.[52] The approximately 30 kb viral genome encodes both structural and NSPs essential for the viral life cycle.[52] Shortly after entry into the host cell, the viral RNA is directly translated into two large polyproteins, which are proteolytically cleaved into 15 NSPs Proteins.[22,44] Among these, the RNA-dependent RNA -polymerase (RdRp) complex - composed of NSP12 (catalytic subunit), NSP7, and NSP8 (cofactors) - is central to both genome replication and transcription of sub genomic RNA (sgRNA).[49,50,51,52,53] The replication of the viral genome and transcription of viral mRNAs are temporally and spatially coordinated processes, taking place in virus-induced membrane compartments.[47,56,63] As observed in other coronaviruses,[186] SARS-CoV-2 triggers a drastic reorganization of the ER, leading to the formation of DMVs, which serve as ROs.[55] NSP3 and NSP4 play a key role in this process by inducing double-membrane vesicle formation.[45,46] Within these DMVs, the replication-transcription complex (RTC) is assembled, providing a shielded microenvironment that protects viral RNA intermediates - particularly double-stranded RNA (dsRNA) - from host immune recognition.[47,48] Inside the DMVs, dsRNA was detected,[54,55] serving as an intermediate for both replication and transcription. The RdRp complex uses the genomic RNA (gRNA) as a template to synthesize a negative-sense intermediate, which in turn serves for the production of full-length genomic RNA and sgRNAs.[53] While dsRNA and replication machinery remain enclosed within the DMVs, newly synthesized positive-sense RNAs - both genomic and subgenomic - must be exported to the cytoplasm for translation and packaging.[187] Recent studies have identified proteinaceous pores spanning the DMV double membrane, through which nascent RNA transcripts are likely translocated into the cytosol.[56] These pores to involve NSP3, whose Ubl1 domain faces the cytoplasm, and NSP4, both of which anchor the pore complex within the membrane.[46,56] The DMV pore contains a positively charged ring within the central channel formed by NSP3 and NSP4, which is likely to facilitate the translocation of RNA through the pore.[46] Huang *et al.* provided a detailed structural visualization of the DMV, identifying residue R306 of NSP4 as a potential RNA-binding site due to its side chain orientation toward the DMV lumen.[46] Mutational analysis of R306 - through charge neutralization (R306A) and charge reversal

(R306Q, R306E) - disrupted the viral replication cycle, highlighting its functional significance.[46] Given the pore diameter of approximately 17 Å, it is sufficient to accommodate ssRNA and small metabolites essential for replication, while dsRNA, with a diameter of 24 Å, would be sterically excluded.[46] *In vivo* studies have shown that NSP3 co-purifies with dsRNA, suggesting a direct role in RNA handling and export.[187] Once in the cytoplasm, sgRNAs are translated into structural proteins (E, M, N, S).[52,57] The N-protein binds to newly synthesized gRNA to form the vRNP complex, which is essential for genome encapsidation and virion assembly.[58,59,60,61] The fate of the gRNA is multifaceted: it is used for further translation of NSPs, serve as a template for ongoing replication, or be packaged into progeny virions mediated by the N-protein.[52] Together, the formation of DMVs, the activity of the RTC, and the export of RNA through DMV pores represent highly coordinated steps in the SARS-CoV-2 replication cycle. Recently, Ke *et al*. demonstrated that the N-terminus of SARS-CoV-2 NSP3 disrupts RNA-induced phase separation of the N-protein by competitively displacing RNA.[84] However, the precise molecular mechanisms by which RNA and possibly replication factors are translocated along the DMV membrane are poorly understood and are the subject of ongoing investigation.

5.2.2 The Viral N-Ubl1-RNA Complex

The Ubl1 domain of SARS-CoV-1 and related CoVs has been implicated in RNA-binding and protein-protein interactions during viral replication (Figure 34). Bessa *et al.,* identified Ubl1 as a key target for the N-protein with high affinity in the nM range.[67] The intrinsically disordered and flexible N3 domain of N binds Ubl1 in a bipartite interaction, with N3 folding around sUbl1 (residues Q16-E111): the first N3 motif binds a hydrophobic groove of Ubl1 involving α-helix 1 and 3 (residues K33-C39 and L65-E70 of Ubl1); the second motif of N3 folds into an α-helix upon binding β-sheet 2 of Ubl1 (residues A41-E45 of Ubl1).[67] In MHV, Ubl1 has also been shown to interact with N.[116] Furthermore, binding interactions between the N-protein and viral RNA *in vitro* were identified,[61] with N phase-separating with RNA.[118] Interestingly, Bessa *et al.* identified an interaction between N234C (cysteine mutant of N3) and polymeric RNA (polyadenosine RNA) by 1D NMR spectroscopy, which is abrogated after the addition of sUbl1.[67] However, biochemical studies have reported low-affinity interactions between Ubl1 and ssRNA, with dissociation constants in the micromolar range, consistent across SARS-CoV-1 and SARS-CoV-2.[87,188] Serrano *et al.* identified two RNA species containing a repeating AUA motif that bind to SARS-CoV-1 Ubl1 with a dissociation constant of 20 µM.[87] Similarly, Lemak

et al. reported a comparable binding affinity for SARS-CoV-2 Ubl1, with a K_D of 31 µM for RNA binding.[188] Notably, no binding was observed with ssDNA or dsRNA containing an AUA motif, nor with homopolymeric RNA strands composed of A, G, C, or U, suggesting that Ubl1 engages in sequence-specific RNA interactions.

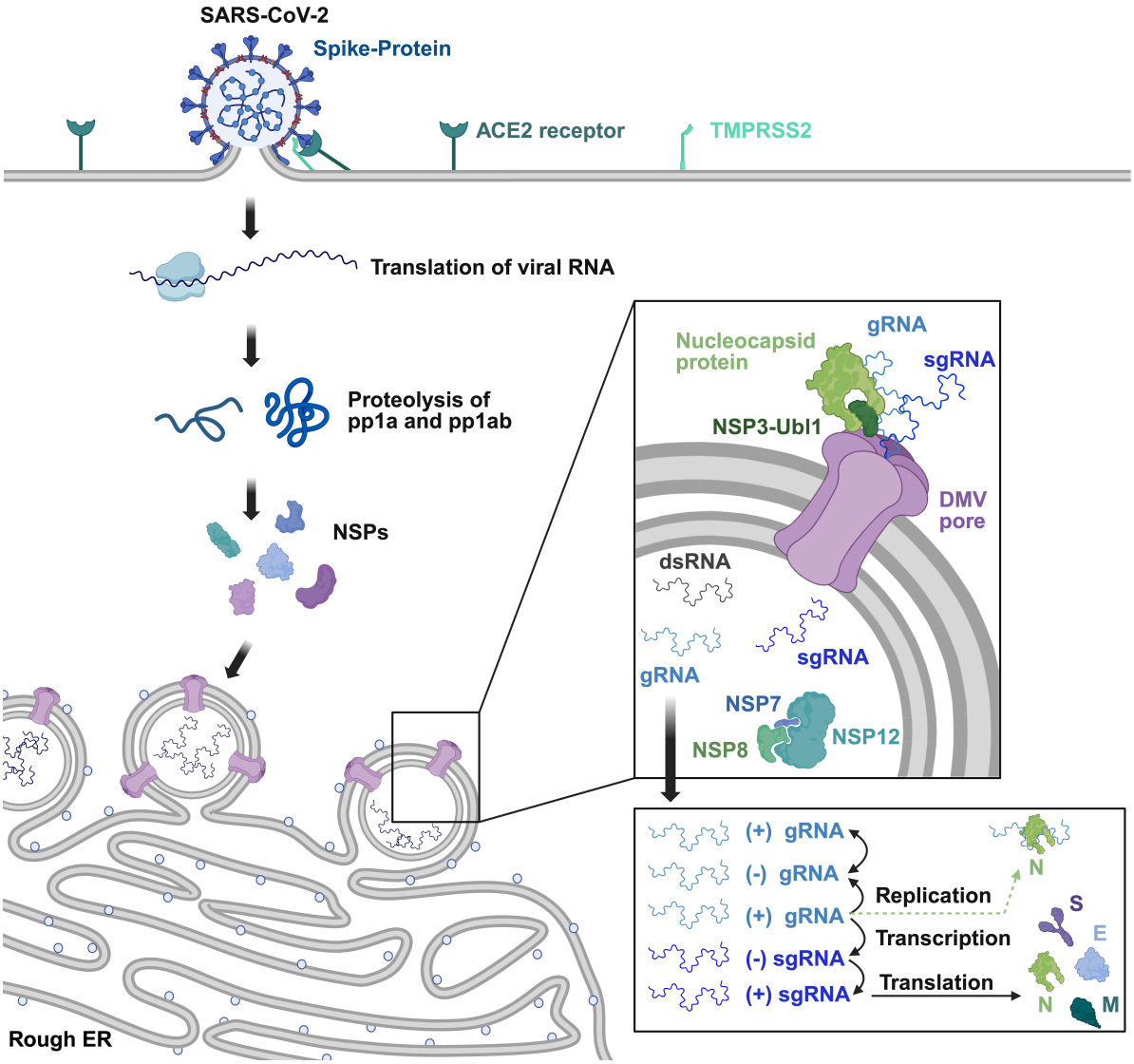


Figure 34: RNA species generated during the SARS-CoV-2 replication cycle. Viral entry is initiated by binding of the spike (S) protein to the host ACE2 receptor, facilitated by the transmembrane serine protease TMPRSS2, which cleaves the spike protein. Following uncoating, the viral genome is released into the cytoplasm, where two open reading frames (ORF1a and ORF1b) are translated by host ribosomes into polyproteins pp1a and pp1ab. These polyproteins are processed by viral proteases into 16 nonstructural proteins (NSPs). NSP3 and NSP4 induce the formation of double-membrane vesicles (DMVs) connected to the endoplasmic reticulum.[45] NSP3 assembles into a hexameric pore in the DMV membrane.[56] Each NSP3 monomer contains 16 domains, with the N-terminal ubiquitin-like domain 1 (Ubl1) exposed to the cytosol.[56] Ubl1 interacts with the nucleocapsid (N) protein,[67] which associates with genomic RNA (gRNA), promoting RNA packaging.[61] Within the DMV lumen, double-stranded RNA (dsRNA),[54] genomic RNA (gRNA), and subgenomic RNAs (sgRNAs) are present. RNA synthesis is mediated by the RNA-dependent RNA polymerase (RdRp) complex, composed of NSP12, NSP7, and

NSP8.[49,50,51,52,53] The RdRp uses positive-sense gRNA as a template to produce a negative-sense intermediate, which serves as a template for full-length gRNA and sgRNA synthesis.[53] Positive-sense sgRNAs are translated into the four structural proteins: nucleocapsid (N), membrane (M), envelope (E), and spike (S).[52] Concurrently, newly synthesized positive-sense gRNA is encapsidated by the N protein and directed to viral assembly sites.[52] Figure was created with BioRender.com.

Significant chemical shift perturbations were detected in two loop regions of SARS-CoV-1 Ubl1 - located between β-strands 3 and 4, and between β-strand 1 and α-helix 1 - as well as in α-helices 1 and 3_{10}.[87]Although RNA is negatively charged, Ubl1 contains several acidic residues that contribute to RNA binding, suggesting a mechanism based on long-range electrostatic interactions rather than direct molecular contact.[87,188] Nevertheless, current evidence suggests that Ubl1 exhibits sequence specificity and does not interact with unrelated RNA sequences.[87,188]

Given that both Ubl1 and the N protein are RNA-binding proteins, it has been proposed that Ubl1 facilitates viral genome translocation through the DMV pore and mediates RNA loading onto the N-protein (Figure 34). This is supported by the RNA-binding properties of multiple domains within NSP3, including Ubl1, the NAB domain, and the Y-domain.[188]Upon addition of Ubl1, the binding affinity of N towards the 5'UTR is dramatically increased.[188] . However, this finding contrasts with earlier observations suggesting that Ubl1 can reduce N-RNA binding in a dose-dependent manner.[61] The first 245 nucleotides of the SARS-CoV-2 5'UTR contain several AUA motifs located within conserved stem-loop structures (SL1-SL4, partially SL5),[189] pointing to a potential regulatory role of Ubl1 in viral RNA translation or replication.

5.3 Aims

This study aimed to investigate the RNA-binding properties of the Ubl1 domain of SARS-CoV-1 NSP3 a domain previously implicated in interactions with ssRNA particularly sequences containing an AUA motif.[87] Given Ubl1's established interaction with the N protein[67] and its localization at the pore of DMVs it is considered a potential mediator of viral RNA translocation during replication.[56] To explore this, we examined the interaction between recombinant Ubl1 and a model ssRNA1 ligand using NMR spectroscopy and MST under a range of biophysical conditions. Specifically, we varied salt concentration, temperature, the presence of magnesium ions, and the RNA-to-protein ratio. Prior to binding studies, the purity and structural properties of the RNA were assessed with particular focus on the absence of

stable secondary structure under physiological conditions. For interaction analysis, two-dimensional NMR spectra (^{1}H-^{15}N HSQC) were recorded for both the reference Ubl1 sample and Ubl1 in the presence of ssRNA1. The binding interface and dynamic effects were investigated by evaluating chemical shift perturbations (CSPs) as well as absolute and relative peak intensity changes. In parallel, MST was used to determine the affinity of the interaction and to assess the potential K_D under different conditions. Through this integrative biophysical approach, we aimed to gain a detailed understanding of the binding behavior of Ubl1 to RNA and to provide insight into its possible mechanistic role within the NSP3-N-RNA complex during the coronavirus replication cycle.

5.4 Material and Methods

The following sections provide a detailed description of the experimental procedures used to investigate the interaction between Ubl1 and RNA. Information regarding the isolation and purification of Ubl1 is presented in Chapter 1.

5.4.1 Buffer and Chemicals

All standard chemicals used for the respective analyses were obtained from commercial suppliers, including AppliChem, Merck, Roche, and Sigma. Detailed information on the chemicals and buffers used in the experimental procedures is provided in the corresponding sections. To minimize the risk of RNA degradation, all laboratory workspaces and consumables were thoroughly decontaminated prior to use. Buffers intended for RNA handling were sterilized by autoclaving, and glassware was heat-treated to eliminate potential contaminants, such as RNase.

5.4.2 Instruments

Instruments and materials used for the analysis of Ubl1-ssRNA binding interactions are listed in Table 19 and 20.

Table 19: Instruments for Ubl1 buffer and sample preparation.

Instrument	Manufacturer
Autoclave VX-150	Systec GmbH, Linden
Desk centrifuge 5415 R	Eppendorf GmbH, Hamburg

Dri-Block Heater 08-3	Techne, UK
Magnetic stirrer C-MAG MS 7	IKA GmbH & CO. KG, Staufen
Milli-Q Direct 16	Merck KGaA, Darmstadt
Monolith NT.115	NanoTemper Technologies GmbH, München
NanoDrop 2000	Thermo Fischer Scientific, Schwerte
Bruker 600 MHz Avance III NMR spectrometer	Bruker, Billerica, MA, USA
Bruker 900 MHz Avance III Neo spectrometer	Bruker, Billerica, MA, USA
Overhead Mixer	Carl Roth, Karlsruhe
pH meter	Satorius AG, Göttingen
Scale	Satorius AG, Göttingen

Table 20: List of consumables for the analysis of Ubl1-RNA interaction.

Material	**Manufacturer**
ALEXA488 Fluor NHS-Ester	Thermo Fischer Scientific, Schwerte
Amicon MWCO 3 kDa	Merck Millipore, Darmstadt, Germany
Deuteriumoxid	Carl Roth, Karlsruhe
Falcon tubes (15 mL and 50 mL)	Sarstedt, Nürnbrecht
Norell Select Series 3 mm NMR Tubes	Sigma-Aldrich, St. Louis, USA
Monolith NT. 115 Series Capillaries	NanoTemper Technologies GmbH, München
Monolith NT. 115 Series Premium Capillaries	NanoTemper Technologies GmbH, München
Plastic cuvettes	Brand GmbH & Co KG, Wertheim
Pipette tips	Sarsted, Nürnbrecht
PCR tube	Sarstedt, Nürnbrecht
PD-10 desalting column (Sephadex G-25 resin)	Cytiva, Marlborough
Reaction tube (1.5 mL and 2 mL)	Sarstedt, Nürnbrecht
Reaction tube (1.5 mL, black)	Sarstedt, Nürnbrecht
ssRNA1 (5 mg pellet)	BioSpring, Frankfurt am Main
Serological pipettes	Sarstedt, Nürnbrecht
SurPhob Low Binding Tips (10 µL)	Biozym Scientific GmbH, Hessisch Oldendorf
SurPhob Low Binding Tips (200 µL)	Biozym Scientific GmbH, Hessisch Oldendorf

5.4.3 Buffers and Solutions

In the following, the ingredients of buffers and solutions used for interaction studies are listed in Table 21. All buffers and solutions were prepared with Milli-Q water (18.2 MOhm). Buffers were degassed and filtered.

Table 21: Buffer composition and chemical solutions.

Solution	Ingredients and Concentrations
Labelling Buffer	50 mM sodium phosphate (pH 6.5), 150 mM NaCl
Low Salt Buffer	50 mM Tris (pH 7), 50 mM NaCl
High Salt Buffer	50 mM Tris (pH 7), 150 mM NaCl

Single-stranded RNA was ordered from BioSpring (Frankfurt am Main). The RNA-pellet (5 mg) was resuspended in filtered and autoclaved ddH_2O to get a final concentration of 2 mM. The RNA sample was aliquoted and stored at -20 °C for further use. Detailed information regarding the Ubl1 target protein construct is provided in Chapter 1, while information on the corresponding RNA construct is listed in Table 22.

Table 22: Construct details of ssRNA1.

Construct Information	ssRNA1
Sequence	5'AAAUACCUCUCAAAAAUAACACCACACCAUAUACCACAU3'
Size [bp]	39
MW [g/mol]	12,304
Tm [°C]	65
GC-content [%]	33
Extinction coefficient [cm^{-1} M^{-1}]	454,500
Weight [mg]	5.00
Manufacturer	BioSpring, Frankfurt am Main

5.4.4 NMR Interaction Studies of Ubl1 and ssRNA1

5.4.4.1 1D NMR Measurements of ssRNA1

To ensure high purity of ordered ssRNA1 (BioSpring, Frankfurt am Main), an 1D NMR spectrum of 50 µM ssRNA1 was acquired on a 600 MHz Bruker spectrometer diluted in 50 mM Tris-d11

(pH 7), 150 mM NaCl, 5% (v/v) D_2O, with 200 µL in volume. A temperature scan ranging from 5 °C to 41 °C in 2 °C steps was performed including breaks of 10 min between each step. The 90° pulse was determined using the average temperature of 22.5 °C (P1 = 8.7 µs). Spectra were recorded with 256 scans (ns). Data were analyzed by using TopSpin 4.0.3 (Bruker Biospin).

5.4.4.2 2D NMR Measurements of Ubl1 and ssRNA1

For starting measurements, purified ^{1}H-^{15}N Ubl1 samples in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl (see Chapter 1) were rebuffered in Low Salt Buffer (50 mM Tris (pH 7), 50 mM NaCl) or High Salt Buffer (50 mM Tris (pH 7), 150 mM NaCl) using an Amicon with a cut-off of 3 kDa (6,000 x *g* speed). Interaction studies were performed on a Bruker 900 MHz spectrometer at 5 °C or 25 °C. Starting with interaction measurements, ^{1}H-^{15}N Ubl1 HSQC reference spectra were recorded in a concentration of 50 µM in Low Salt Buffer or High Salt Buffer at different temperatures. Experiments were repeated by recycling the original sample and supplementing it with 2 mM $MgCl_2$. Interaction studies were performed in a ratio of 1:3 of 1:5 (Ubl1:ssRNA1). An overview of the performed experiment is provided in Table 23.

Table 23: Setup of experiments using BioSpring ssRNA1.

Exp. Nr.	Buffer	Ubl1 [µM]	Ubl1:ssRNA1 ratio	Temperature [°C]
1	High Salt	50	1:3	25
2	High Salt + 2 mM $MgCl_2$	50	1:3	25
3	Low Salt	50	1:5	25
4	Low Salt + 2 mM $MgCl_2$	50	1:5	25
5	Low Salt	50	1:5	5
6	Low Salt + 2 mM $MgCl_2$	50	1:5	5

For data analysis, differences in CSPs of Ubl1-ssRNA1 spectra and corresponding reference spectra were calculated. Relative peak intensity ratios were calculated using CSPs. Data analysis was supported by the software tool CCPN.[125]

5.4.5 Microscale Thermophoresis (MST)

MST is a solution-based fluorescence technique used to analyze interactions between a fluorescently labeled molecule - such as a protein or small compound - and its ligand in a dose-dependent manner.[190] In this study, MST was employed to investigate the binding interaction between fluorescence-labeled Ubl1 and ssRNA1 as the ligand. The following sections provide a detailed description of the experimental workflow and analysis related to MST measurements.

5.4.5.1 Protein-Labelling with ALEXA488

Starting with labelling Ubl1 with a fluorophore, ALEXA488 was resuspended in 1000 µL water-free DMSO to get a final concentration of 1 mg/mL. The green fluorophore was aliquoted in PCR tubes and stored at -20 °C for further use. To prevent bleaching, all experiments in the presence of ALEXA488 were performed in the absence of overhead light. For protein labelling, a pure and high-quality protein sample of Ubl1 was acquired as described in Chapter 1. As ALEXA488 requires amine-free buffer for labelling, Ubl1 was rebuffered in 50 mM sodium phosphate (pH 6.5), 150 mM NaCl (Labelling Buffer) using an Amicon (cutoff 3 kDa) and being centrifuged several times at 6,000 x *g*. Concentrated and rebuffered protein samples were mixed with 3x access of the fluorophore. Sample was incubated for 1 h overhead rotation. To prevent bleaching, sample was packaged in aluminum foil. To separate labelled protein from free fluorophore, a PD-10 desalting column was used. The column was previously washed with 8 CV degassed and filtered ddH_2O, followed by equilibration in 50 mM Tris (pH 8), 140 mM NaCl. Finally, 500 µL of the Ubl1-ALEXA mixture was loaded onto the column turning the column yellowish. The sample was eluted with 500 µL of buffer and fractions stored in black 1.5 mL reaction tubes. The column was washed with 8 CV water and 4 CV 20% (v/v) EtOH. The column was stored in 20% (v/v) EtOH at 4 °C.

5.4.5.2 Determination of Labelling Efficiency

The sample was centrifuged at 13,000 rpm for 5 min at 4 °C to separate aggregates from the remaining protein. The protein concentration at A_{280} was determined by using Nano-Drop. Labelling efficiency was determined by measuring absorbance at 494 nm based on the excitation of ALEXA488. The labelling efficiency [%] (LE) was finally calculated using the following formular:

$$LE = \frac{\varepsilon(Ubl1)}{\varepsilon(ALEXA488)}\ x\ \frac{1}{\left(\frac{A280}{A494}\right) - CF}\ x\ d$$

LE: Labelling efficiency [%]

Ɛ(Ubl1): Extinction coefficient of Ubl1 (14,565 cm^{-1} M^{-1})

Ɛ(ALEXA488): Extinction coefficient of ALEXA488 (73,000 cm^{-1} M^{-1})

CF: Correction factor (0.11 for ALEXA488 NHS ester)

d: pathlength (1 cm)

For optimal MST measurements, an LE of 50-100% is required. Samples were finally flash frozen in liquid nitrogen and stored at -80 °C until further measurements.

5.4.5.3 Optimization of MST measurements - Capillary Scans

To identify, which capillary sort (standard or premium) and protein concentration is suitable for later interaction measurements, 50 nM and 100 nM of labelled Ubl1 were analyzed in triplicates (Table 24) at 25 °C. For that, a 100 µL sample was put in a PCR tube. Because of capillary forces, the sample was soaked up into the capillary.

Table 24: Experimental setup for capillary scans.

Amount	Buffer	Ubl1 [nM]	Capillary	Temperature [°C]
3	50 mM Tris (pH 8), 140 mM NaCl	50	Standard	25
3	50 mM Tris (pH 8), 140 mM NaCl	50	Premium	25
3	50 mM Tris (pH 8), 140 mM NaCl	100	Standard	25
3	50 mM Tris (pH 8), 140 mM NaCl	100	Premium	25

For the measurement, the Monolith NT-115 device from NanoTemper was used with the following settings: LED-Power Level: 50%, LED-Color: Blue and MST-Power: 30%. The program “NT control” was used for measurements. Before starting the experiment, capillaries were

stored in the device for 15 min to warm up to 25 °C. Optimal conditions were identified by sharp, well-defined peaks reflecting non-sticking samples, and further, fluorescence counts between 600 and 1000, as the protein concentration is halved during interaction measurements.For later experiments and after buffer optimizations, capillary scans were performed before starting the final MST measurement to characterize the protein's behavior under given conditions.

5.4.5.4 Interaction Measurements with MST

MST measurements were performed as described for the capillary scans. However, here sample preparation differed as the target was mixed with ligand (ssRNA1). For better comprehension, a pipetting scheme is illustrated in Figure 35. Since the target protein and the ligand were diluted in a ration of 1:1, the double amount of target and ligand were prepared for starting the dilution series. Each sample was prepared in a PCR tube for better handling.

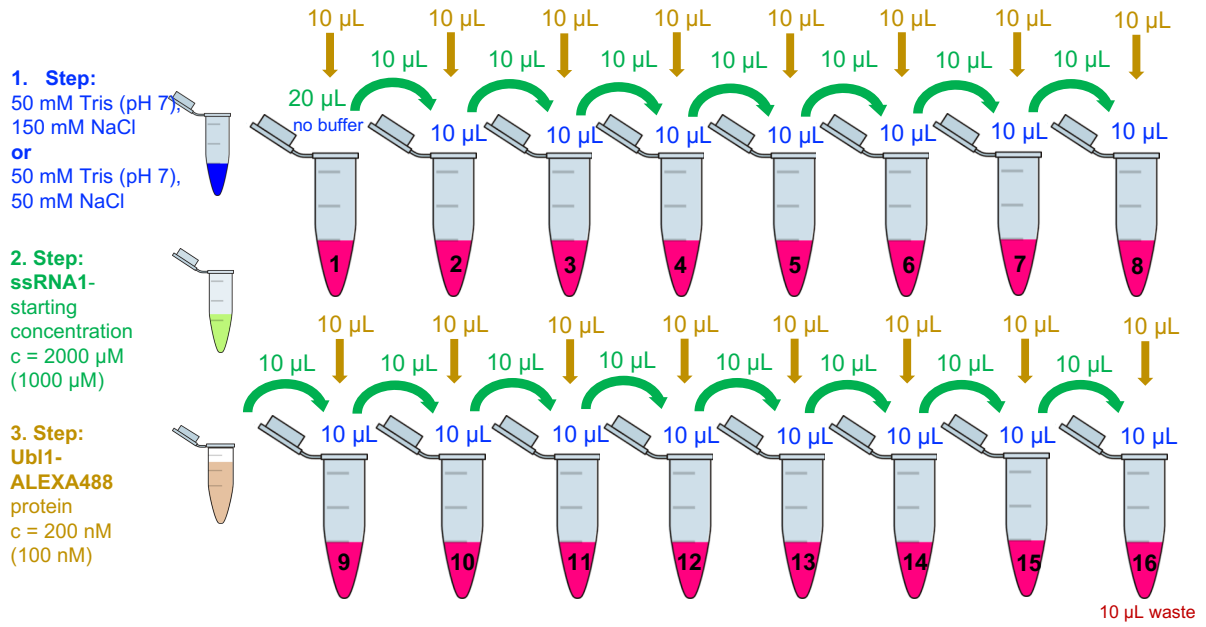


Figure 35: MST pipetting scheme. In total, 16 approaches were prepared based on the number of capillaries suitable for the Monolith NT.115 device. Buffer (High Salt or Low Salt Buffer), 10 µL in volume, was added to PCR tubes 2-16. Ligand (RNA) was added to reaction tube 1 (20 µL). Dilution series was started by mixing 10 µL of tube 1 with tube 2. After mixing, 10 µL of tube 2 were added to tube 3 etc. The final 10 µL of tube 16 were discarded. Lastly, 10 µL of labelled protein were added to each tube.

Finishing the dilution series, samples were soaked by the according capillary and stored for 15 min in the Monolith NT.115 device. Settings were copied from the capillary scans.

Capillaries including samples were discarded after measurement. Data were analyzed using the program “MO Analysis”.

5.5 Results

The following section provides a detailed analysis of the binding interactions between NSP3-Ubl1 and ssRNA1. The sequence of ssRNA1 used in this study was based on the work of Serrano *et al.* (2007), who identified an interaction between SARS-CoV-1 NSP3-Ubl1 and ssRNA containing an AUA motif.[87] To replicate these findings, the same RNA sequence was synthesized and obtained from BioSpring. The successful production of purified Ubl1 protein in *E. coli* has been described previously in Chapter 1. The results presented here are based on NMR spectroscopy and MST, which were employed to characterize the molecular interactions between Ubl1 and ssRNA with high sensitivity and resolution.

5.5.1 Analysis of ssRNA1

The sequence of ssRNA1 used in this study is shown in Figure 36A. As previously noted, the RNA contains a recurring AUA motif, with four AUA repeats highlighted in red. Prior to binding assays, the synthetic ssRNA1 (purchased from BioSpring) was assessed for structural integrity and purity. Secondary structure predictions were performed using the unaFOLD software (https://www.unafold.org/mfold/applications/rna-folding-form.php). As shown in Figures 36B and 36C, two potential folding conformations were predicted, each featuring three A-U base pairs, while the remainder of the sequence remained unpaired. The first predicted structure forms a loop comprising 26 nucleotides and has a calculated ΔG of -4.40 kcal/mol (Figure 36B). The second structure forms a smaller loop of 13 nucleotides with a ΔG of -3.40 kcal/mol (Figure 36C).

To assess the purity and structural integrity of the synthesized RNA, one-dimensional (1D) NMR spectra were acquired across a temperature range of 5 °C to 41 °C (Figure 37). As shown in Figure 37, the 1D spectra exhibit significant signal overlap and limited chemical shift dispersion. Notably, resonances corresponding to imino protons, typically observed in the downfield region between 12 and 14 ppm and indicative of stable base pairing, were absent.[191] Most signals were confined to the spectral window between 0 and 8 ppm, suggesting a lack of well-defined secondary structure.

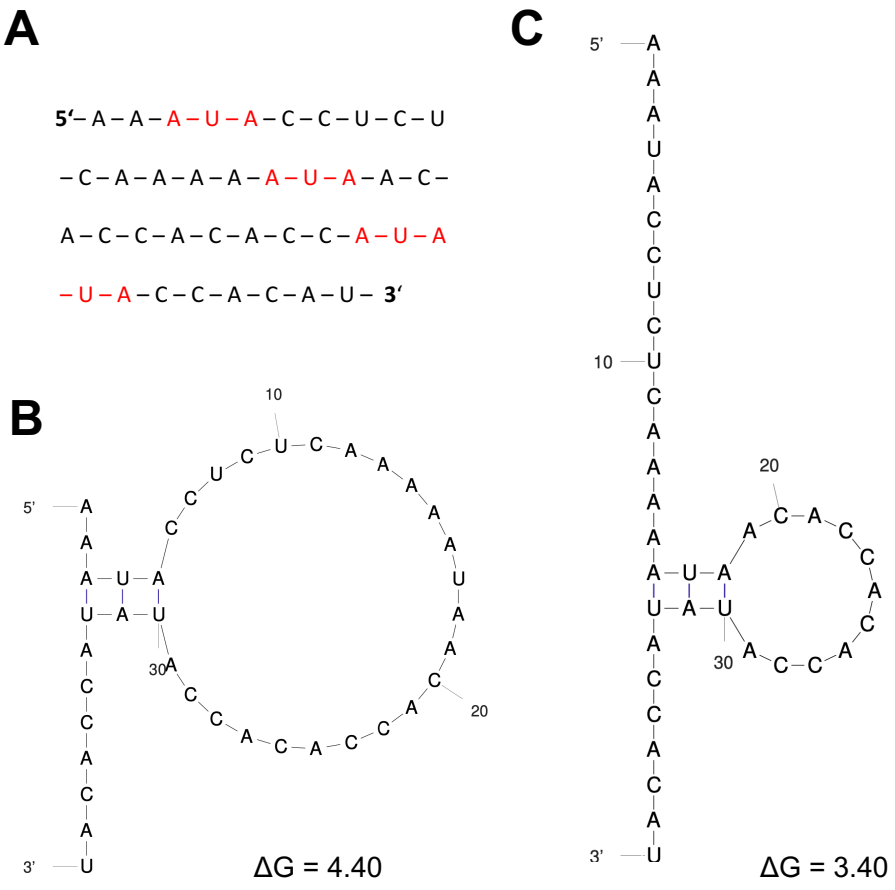


Figure 36: Folding prediction of ssRNA1. (A) Linear ssRNA sequence highlighting the AUA pattern (red). **(B)** and **(C):** Secondary structure of ssRNA predicted by unaFOLD with calculated ΔG values. (https://www.unafold.org/mfold/applications/rna-folding-form.php).

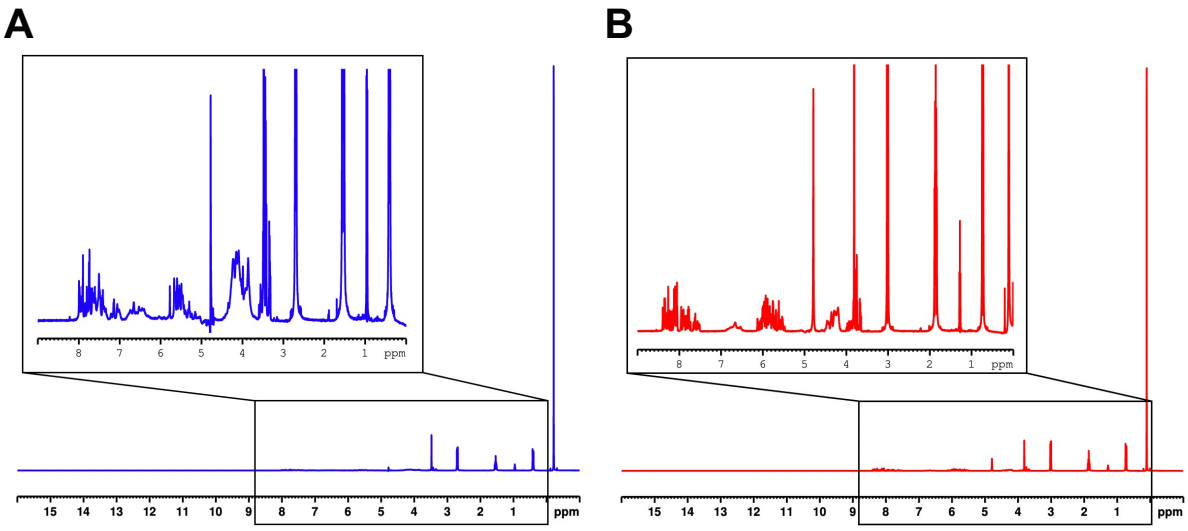


Figure 37: 1D NMR spectrum of ssRNA1. Spectra were acquired on a 600 MHz spectrometer using 50 µM ssRNA1 in 50 mM Tris-d11 (pH 7), 150 mM NaCl and 5% (v/v) D_2O. Spectra were recorded using a temperature gradient in 2 °C steps. Spectra recorded at 5 °C **(A)** and 41 °C **(B).**

Since the RNA construct is highly pure, the sample was used for further interaction studies with Ubl1. Interaction was studied by using NMR spectroscopy and MST.

5.5.2 Binding Interactions of ^{1}H-^{15}N Ubl1 and ssRNA1 Studied by NMR Spectroscopy

To investigate the interaction between ^{1}H-^{15}N Ubl1-labeled Ubl1 and ssRNA1, two-dimensional (2D) NMR spectroscopy was employed. A series of ^{1}H-^{15}N Ubl1 HSQC spectra was recorded at 5 °C and 25 °C using a 900 MHz spectrometer (Figure 38). CSPs were analyzed by comparing spectra of Ubl1 in the presence and absence of ssRNA1. To control for ionic strength-dependent effects previously observed (see Chapter 3), experiments were conducted in two buffer systems containing either 50 mM NaCl (Low Salt) or 150 mM NaCl (High Salt). Additionally, magnesium chloride was added after each measurement, as magnesium ions are known to influence the structural stability of nucleic acids.[192,193] NMR samples were prepared at two protein-to-RNA molar ratios: 1:3 (50 µM Ubl1 and 150 µM ssRNA1) and 1:5 (50 µM Ubl1 and 250 µM ssRNA1). The resulting spectra and CSP analyses are shown in Figure 38.

As shown in Figure 38, by NMR no consistent CSPs of ^{15}N Ubl1 were observed upon addition of ssRNA within the range of tested experimental conditions. In the absence of magnesium, no significant interaction was detected (Figure 38A, C), except under low-salt conditions at 5 °C (Figure 38E), where residues T24 and E37 exhibited notable chemical shift perturbations.

Under high-salt conditions (150 mM NaCl) and in the presence of magnesium, at 25 °C, only residue L35 showed significant CSP in the presence of RNA (Figure 38B). In contrast, under low-salt conditions (50 mM NaCl) and in the presence of magnesium, T24 was the only residue with significant CSPs (Figure 38D), therefore not yielding a clear trend.

In addition to CSPs, differences in peak intensities were analyzed to further characterize the interaction between Ubl1 and ssRNA1 (Figure 39). Intensity ratios were calculated by dividing the peak height of each Ubl1-RNA sample by the corresponding peak height of the reference (Ubl1 alone). Beforehand, each sample was normalized to its signal to noise ratio. Values below 1 indicate higher peak intensities in the reference sample compared to the RNA-bound condition whereas values above 1 reflect higher intensities in the presence of RNA (Intensity

rations of 1 indicate equal peak height intensities of the complex sample and its corresponding reference sample.

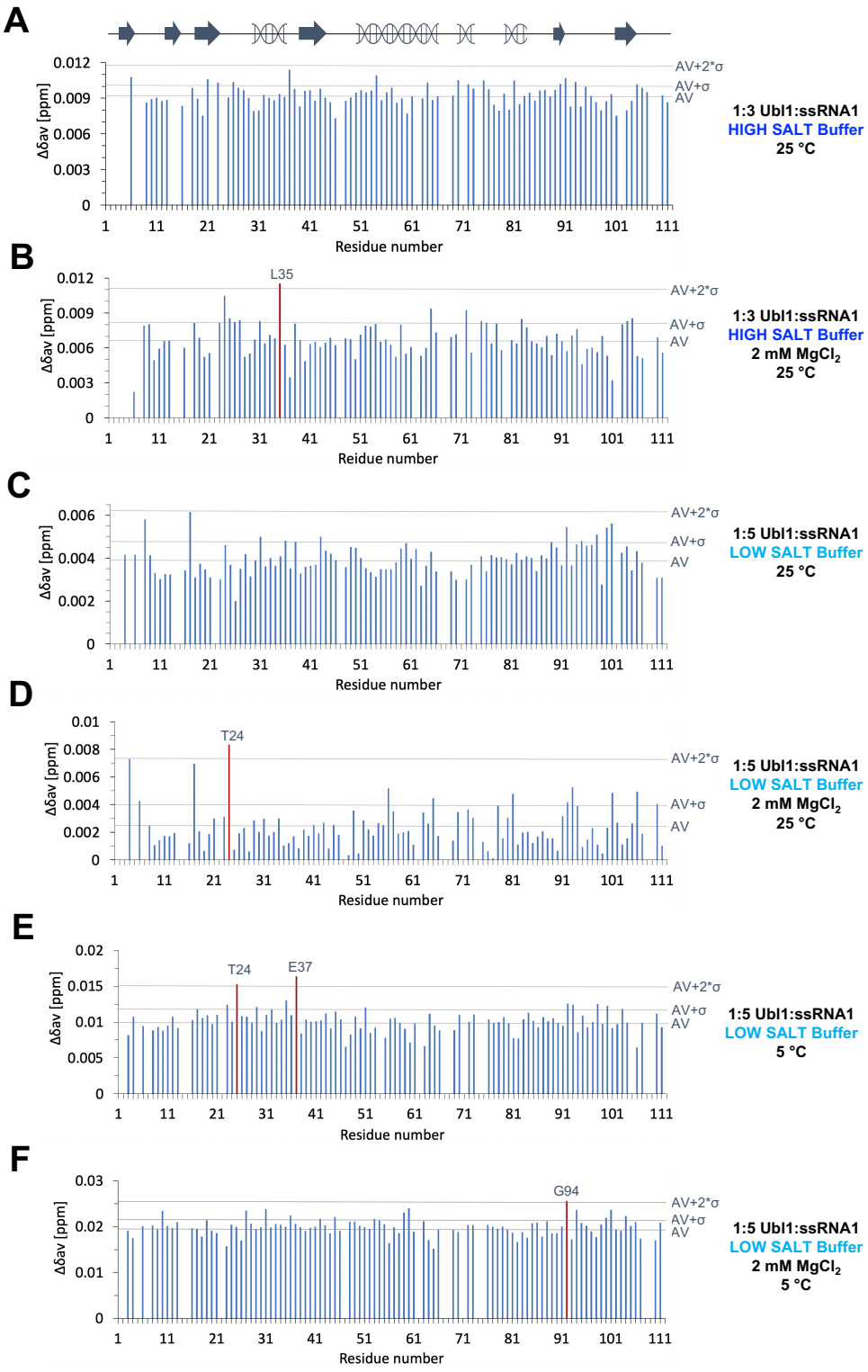


Figure 38: NMR CSP interaction analysis between ^{1}H-^{15}N Ubl1 and ssRNA1 under varying conditions. Displayed are CSPs of ^{1}H-^{15}N Ubl1 (residues 1-111) upon binding to ssRNA1. Residues exhibiting significant perturbations are marked as red columns. Experimental parameters and buffer

compositions for each condition are indicated alongside the corresponding plots. Each sample was supplemented with DSS for referencing.

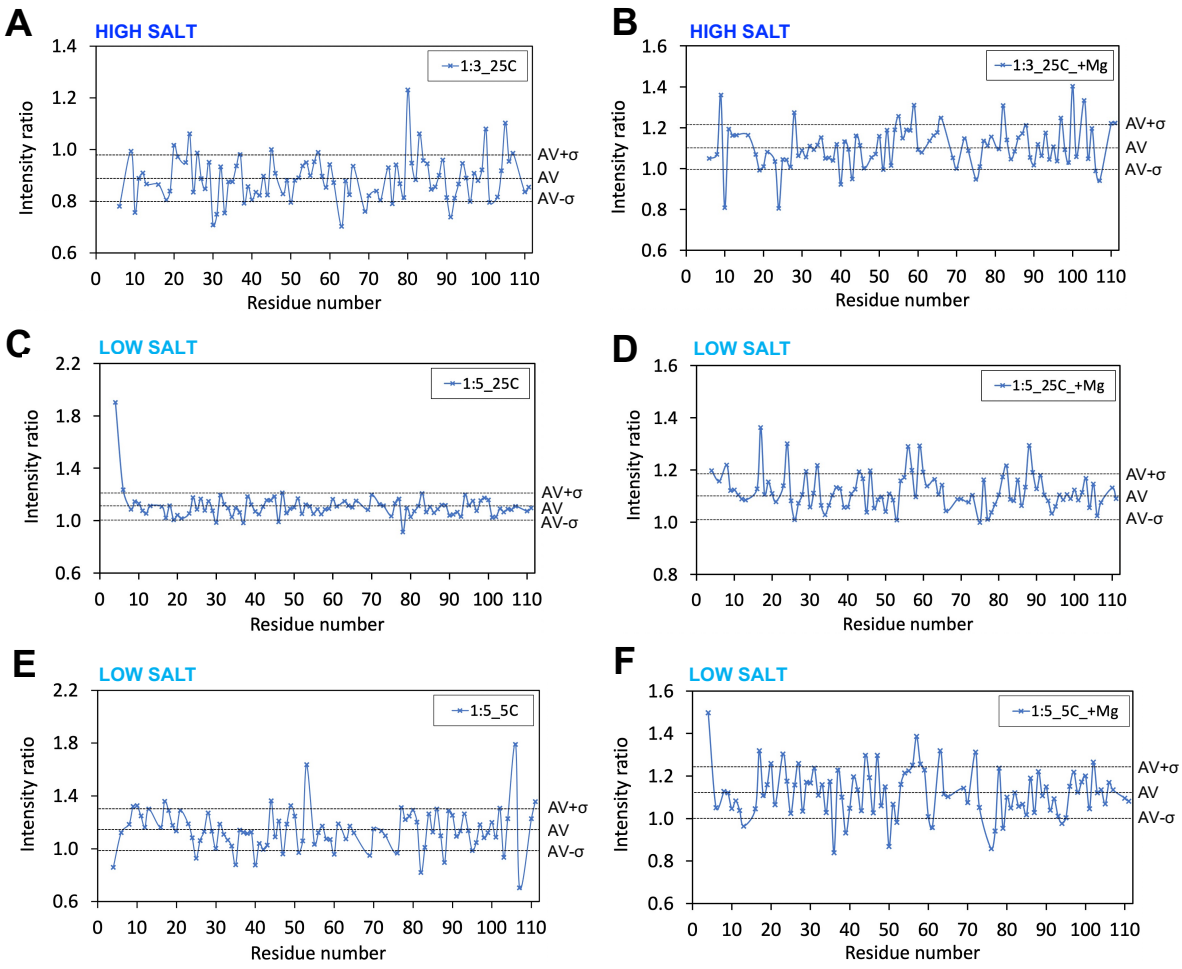


Figure 39: Relative peak intensities upon Ubl1-ssRNA1 interaction. Peak heights (absolute values) of Ubl1-ssRNA1 data points were normalized and divided by the peak height of the corresponding reference spectrum. Measurements were performed using High Salt (50 mM Tris (pH 7), 150 mM NaCl) or Low Salt (50 mM Tris (pH 7), 50 mM NaCl) buffer, at 25 °C (25C) or 5 °C (5C) in the absence (left panel) and presence (right panel) of 2 mM $MgCl_2$ (+Mg). Ubl1 to ssRNA1 was analyzed in a ratio of 1:3 and 1:5.

Since most measurements yielded intensity ratios of 1 or higher, no binding interactions were detected. The majority of datapoints was monitored around the average, with single values showing high significance (Figure 39).

To gain deeper insights into the effect of magnesium on binding, correlation plots were generated (Figure 40). Regardless of the experimental setup and conditions used, the data points were scattered along the linear regression line, indicating a lack of linear correlation. This observation is supported by consistently low R^2 values.

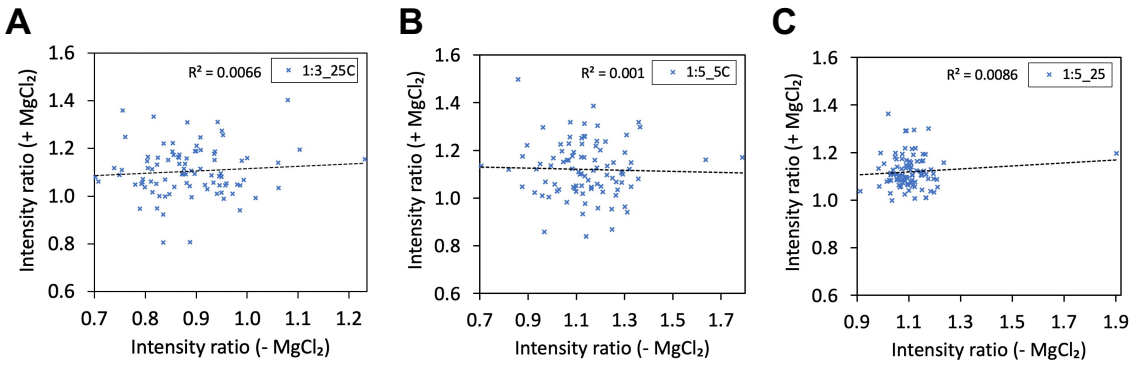


Figure 40: Correlation plots of intensity ratios. Previous analyzed peak intensity ratios were correlated between samples in the presence and absence of 2 mM $MgCl_2$. A linear regression line was fitted for each comparison, and the corresponding R^2 value was calculated. Lower R^2 values indicate weaker correlations. (A) Correlation plot of 1:3 Ubl1:RNA at 25 °C in High Salt buffer. (B) Correlation plot of 1:5 Ubl1:RNA at 5 °C in High Salt buffer. (C) Correlation plot of 1:5 Ubl1:RNA at 25 °C in High Salt buffer.

5.5.3 Binding Interactions of ^{1}H-^{15}N Ubl1 and ssRNA1 Studied by MST

MST was employed to further characterize the interaction between Ubl1 and ssRNA1, focusing on quantifying the binding affinity. Ubl1 was fluorescently labeled with ALEXA488, and a 1:1 serial dilution of ssRNA1 from 1 mM to 31 nM was prepared, yielding 16 samples containing 100 nM fluorescently labeled Ubl1 and the respective RNA concentrations.

Prior to the MST measurements (Figures 41C - F), capillary scans were performed under both buffer conditions - Low Salt and High Salt - to assess potential buffer effects on the fluorescence signal and capillary quality of Ubl1-ALEXA488 (Figures 41A and 41B). For that, Ubl1-ALEXA488 samples were prepared in the absence of RNA (see Material and Methods section). Fluorescence intensities obtained from capillary scans are shown in Figures 41A and 41B. Under low-salt conditions, fluorescence values ranged between 900 and 1000 counts (Figure 41A), while under high-salt conditions, values ranged from 1000 to 1100 counts (Figure 41B). Fluorescence levels remained consistent across all capillaries within each dilution series.

Figures 41C and 41D display MST binding curves based on normalized fluorescence, generated using a titration series ranging from 1 mM to 31 nM ssRNA1. The full sigmoidal binding curve could not be recorded within this concentration range. Corresponding MST traces are shown in Figures 41E and 41F, depicting relative fluorescence intensity over time for each ligand concentration. The MST traces showed minimal separation across the different ligand concentrations.

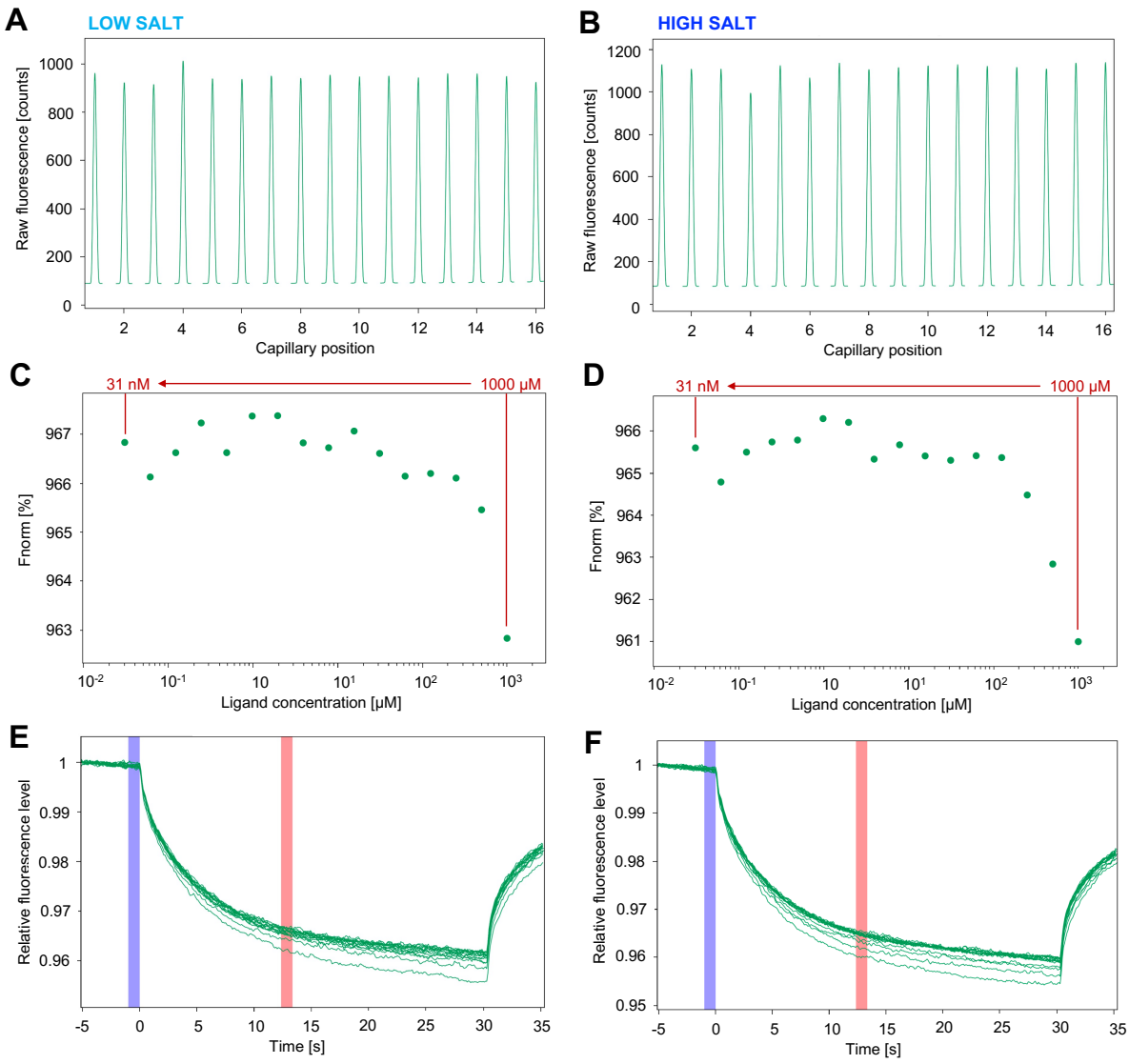


Figure 41: Interaction studies of Ubl1 and ssRNA1 using MST. MST experiments were performed using 100 nM Ubl1 fluorescently labeled with ALEXA488. The ssRNA1 ligand was titrated in a 1:1 serial dilution ranging from 1 mM to 31 nM. Measurements were carried out at 25 °C in either Low Salt buffer **(A, C, E)** or High Salt buffer **(B, D, F). (A, B)** Capillary scans showing the raw fluorescence intensity (counts) across capillaries 1-16 for quality control and uniformity of the samples. **(C, D)** MST single binding curve displaying normalized fluorescence (%) as a function of ligand concentration (µM), used to determine binding affinity. **(E, F)** MST traces representing relative fluorescence intensity over time (seconds), reflecting the thermophoretic behavior of Ubl1-ALEXA488 upon ssRNA1 binding.

5.6 Discussion

Previous studies have demonstrated that the Ubl1 domain of SARS-CoV-1 interacts with ssRNA1 containing an AUA motif.[87] Additionally, nucleic acid binding by the viral N-protein has been reported by Bessa *et al*., identifying it as a key RNA interaction partner.[67] Given that the N-protein plays a central role in RNA and genome packaging during virion assembly,[194] its interaction with RNA is well established and has been extensively characterized through methods such as NMR spectroscopy and electron tomography. [47,60,195, 196,197] Considering that the N-protein also interacts with Ubl1,[67] a direct interaction between Ubl1 and RNA is plausible. Furthermore, Ubl1 is localized at the tip of a crown-like pore embedded in the DMV bilayer,[56] positioning it as a potential interaction partner involved in the translocation of viral RNA through the pore. Therefore, we investigated the dose-dependent binding interactions between NSP3-Ubl1 and ssRNA under varying experimental conditions by NMR spectroscopy and MST. For this purpose, ssRNA1 with the identical sequence reported by Serrano *et al.* was synthesized and obtained from BioSpring, and its folding properties and purity were initially assessed.[87]

The thermodynamic analysis of the predicted ssRNA1 secondary structures indicates that spontaneous folding under physiological conditions is unlikely. Both predicted conformations exhibit positive ΔG values, signifying unfavorable thermodynamics for spontaneous formation.[198] Furthermore, the sequence's predominance of adenine-uracil (A-U) base pairs, which form only two hydrogen bonds compared to the three in guanine-cytosine (G-C) pairs, decreases the stability of potential folded structures.[199] Each predicted folding contains only three base pairs, leaving the majority of the RNA sequence unstructured and likely highly dynamic. Even under dynamic equilibrium, only a small fraction of the RNA population would be expected to adopt folded conformations. These factors collectively suggest that stable secondary structures of ssRNA1 are improbable in solution.

The 1D NMR analysis of the ssRNA1 across a temperature range of 5 °C to 41 °C provides insights into its structural properties and sample purity. The observed significant signal overlap and limited chemical shift dispersion are characteristic of highly flexible or disordered RNA conformations with no hydrogen bond formation.[200] Because of its geminal hydroxyl groups, the ribose part of the RNA shows signals mostly between 3.5 and 5 ppm, which makes

it hard to clearly separate the peaks.[191] This is further supported by the absence of imino proton resonances in the downfield region (12-14 ppm), where signals from hydrogen-bonded guanine and uracil imino protons typically appear in structured RNAs, but none in disordered regions.[191,201] The fact that most signals appear between 0 and 8 ppm supports the idea that the RNA is mostly unstructured and flexible. This behavior could result from an intrinsic sequence-related inability to form stable helices (see folding prediction) or from external factors such as ionic conditions or temperature influencing the folding equilibrium. Together, these findings indicate that the ordered RNA construct does not adopt a well-defined secondary structure in solution under the experimental conditions used. However, the spectra also demonstrate high sample purity, as no signals attributable to impurities were observed.

NMR spectroscopy showed condition dependent CSPs in Ubl1 upon addition of ssRNA1. The presence and magnitude of CSPs varied with experimental parameters such as temperature, salt concentration, magnesium ions, and RNA-to-protein ratio. In the absence of magnesium, CSPs were minimal, with only weak effects observed under low-salt conditions at 5 °C. Under physiological temperature and in the presence of magnesium, CSPs were detected for residues T24, L35, and E37, while higher RNA concentrations led to additional perturbations at G94. These observations suggest a heterogeneous and environment-sensitive response of Ubl1 to RNA addition. However, the residues involved in binding - T24, L35, E37, and G94 - possess distinct chemical properties, supporting different interaction types: E37 likely contributes to electrostatic interactions, while T24, L35, and G94 are more consistent with hydrophobic contacts.

In a very optimistic interpretation, looking for potential RNA binding epitopes of Ubl1, the residues T24, L35, and E37 would be located in conserved regions of Ubl1 along SARS-CoV-1 and SARS-CoV-2 regarding the amino acid sequence.[87,31] In SARS-CoV-2, T24 is located in the third β-strand, while L35 and E37 are positioned within the first α-helix (PDB: 7KAG), potentially indicating a partially overlapping RNA-binding interface. In contrast, in SARS-CoV-1, all three residues are situated within a loop region connecting the first two β-strands.[87] However, assuming a Gaussian distribution, 2.5% of all data points would show a deviation from the mean higher than 2σ, this corresponds to two residues out of one hundred. Therefore, the occurrence of the outliers, which differ across the various conditions, points to a pure statistical effect. Thus, based on the CSP analysis, no clear binding between Ubl1 and

RNA can be observed. The relative intensity analysis further supported this interpretation: since most measurements yielded intensity ratios of 1 or higher, no binding interactions were detected.

Using intensity ratios and correlation plots, we studied the effect of magnesium onto binding: magnesium affected binding, but not in a systematic way - the changes in peak intensities appear random. We could not display a global pattern that would suggest a specific or directed effect of magnesium. Hence, there is no evidence of a direct interaction and no consistent changes in intensity. The observed differences are not correlated, but rather random.

Furthermore, these data are in line with the MST measurements. Capillary scans demonstrated consistent fluorescence levels across dilution series under both low- and high-salt buffers, confirming sample homogeneity. However, MST binding curves did not reach saturation within the tested ligand concentrations even RNA concentrations up to 1 mM were used. Hence, MST traces showed minimal separation, indicating undetectable RNA-binding or weak binding in the low affinity range under these conditions.[190] As a result, the data could not be accurately fitted, and no dissociation constant (K_D) was determined. In contrast to the findings by Serrano *et al.*, who reported a low but measurable binding affinity (~20 µM) for the interaction between SARS-CoV-1 Ubl1 and ssRNA, we were unable to determine a dissociation constant under our experimental conditions, indicating potential differences in binding characteristics.[87] More recently, experiments utilizing ^{32}P-labeled ssRNA containing 5'UTR sequences demonstrated a comparable dissociation constant ($K_D \approx 31$ µM) for Ubl1, similar to that observed in SARS-CoV-1.[188] Furthermore, Ubl1 mutants exhibited altered RNA-binding properties.[188] Specifically, substituting negatively charged residues with positively charged ones led to a reduced RNA-binding affinity, while the purified Ubl1 constructs remained dimeric during SEC.[188] Deletion of the N-terminal 14 residues also diminished RNA-binding affinity, suggesting that both dimerization and the negatively charged surface of Ubl1 are critical for its interaction with RNA.[188] In the case of SARS-CoV-1 Ubl1, negatively charged and acidic residues were found to participate in nucleic acid binding, despite the fact that RNA itself carries a negative charge.[87] A putative binding pocket was identified between helices α1 and α2, as well as the loop regions connecting β1 to α1 and β3 to β4. These findings suggest that the observed binding effects may be mediated through long-range electrostatic

interactions rather than direct Ubl1:RNA contacts.[87] Nonetheless, sequence-specific recognition was demonstrated, as Ubl1 did not bind to unrelated RNA sequences.[87]

In the MHV, the Ubl1 domain has been shown to interact with the N-protein independently of RNA.[58,116] Since both Ubl1 and the N protein are RNA-binding proteins, it was proposed that Ubl1 may facilitate viral genome translocation through the DMV pore and mediate loading onto the N protein. This hypothesis is supported by evidence that several domains within NSP3, including Ubl1, the Y-domain, and the NAB domain, possess RNA-binding activity.[188] Moreover, Ubl1 significantly enhances the binding affinity of the N protein to the viral 5'UTR - by approximately 127-fold - compared to either protein alone.[188] This observation contrasts with earlier reports suggesting that Ubl1 reduces N-RNA binding in a dose-dependent manner.[61] Nevertheless, specific RNA-binding interactions remain plausible, particularly as the first 245 nucleotides of the SARS-CoV-2 5'UTR contain three AUA motifs, with at least one located within SLs 1-4 and partially SL5.[189] These findings point toward a potential role for Ubl1 in regulating viral translation or replication.[189]

In summary, our results indicate that no clear or specific interaction between Ubl1 and ssRNA1 could be observed under the tested conditions. While some weak signals were detected, they do not support the presence of a well-defined binding pocket or a specific binding site. Instead, potential interaction sites appear to be diffusely distributed across the protein, and the overall binding is weak and non-specific. The observed CSPs are likely the result of statistical or transient effects rather than a stable binding event. Similarly, intensity-based analyses and MST measurements did not reveal consistent or correlated changes upon ligand addition. Magnesium influenced RNA binding in a random, non-systematic manner, further supporting the absence of a defined binding interaction. Importantly, since Ubl1 was not analyzed alone (without RNA) in the presence and absence of Mg^{2+}, we cannot determine whether magnesium directly affects Ubl1 or only modulates RNA binding. Altogether, our data suggest that no specific or high-affinity interaction occurs between Ubl1 and ssRNA1 under the tested conditions. The possibility of weak, transient, or context-dependent interactions cannot be ruled out, but further studies using complementary biophysical and structural approaches will be necessary to clarify the mechanistic and physiological relevance of the Ubl1-RNA interaction.

6. Chapter: Expression and Isolation of NSP3 PLpro and its Subdomain Ubl2 followed by Structural Analysis using NMR and CD Spectroscopy

6.1 Abstract

The PLpro domain of SARS-CoV-2, including its subdomain Ubl2, play a vital role in viral replication by cleaving the viral polyprotein at NSP1/2, NSP2/3, and NSP3/4 junctions an essential step in the formation of DMVs. Although PLpro is a highly relevant drug target, there is still a lack of high-resolution structural data - particularly structures well-suited for drug screening and rational inhibitor design, such as NMR spectroscopy. In this study, we established optimized protocols for the high-yield expression and purification of recombinant PLpro and Ubl2, enabling their structural characterization using CD and NMR spectroscopy. Ubl2 was expressed in *E. coli* under optimized conditions and analyzed by CD and multidimensional NMR, confirming a ubiquitin-like β-grasp fold and revealing two helical regions. Near-complete NMR backbone resonance assignments were achieved, extending previous work by Shimada *et al.* and offering a more comprehensive structural map of Ubl2. Recombinant PLpro was also successfully produced, although inclusion body formation and aggregation were initially observed. Through a series of expression and purification optimizations - including modified lysis conditions, media selection, and buffer composition - soluble, folded PLpro was obtained. A well-dispersed ^{1}H-^{15}N TROSY-HSQC spectrum was recorded at 1.2 GHz, confirming the folded state of the protease and providing a reference spectrum for future interaction studies. These results offer valuable insights into the structural organization of PLpro and its Ubl2 subdomain and lay the groundwork for future studies aimed at elucidating their functional roles in the SARS-CoV-2 replication cycle and informing antiviral drug development.

6.2 Introduction

6.2.1 Structure and Localization of PLpro and Its Subdomain Ubl2

The viral genome of SARS-CoV-2 consists of two large ORFs, with ORF1 encoding two cysteine proteases: PLpro[23] and Mpro[24].[29,39] These proteases are embedded in NSP3 and NSP5, respectively - two of the 15 NSPs expressed from the around 30-kb SARS-CoV-2 genome.[22,44] PLpro constitutes one of the 16 domains within NSP3 and is centrally located within the full-length NSP3 polypeptide, also representing its largest domain (Figure 42).[25,79] Once NSP3 integrates into the DMV membrane, PLpro and Ubl2 are exposed to the cytosolic side (Figure 42).[202] Unlike certain other CoVs such as MHV, which encodes two PLpro domains, SARS-CoV-2 contains only a single PLpro domain.[203] The SARS-CoV-2 PLpro domain spans 315 amino acids, corresponding to residues 746-1060 of NSP3.[204] This domain is highly conserved across CoVs, sharing 83% sequence identity with the SARS-CoV-1 PLpro domain and exhibiting an almost identical active site architecture, underscoring its structural and functional conservation.[3,205]

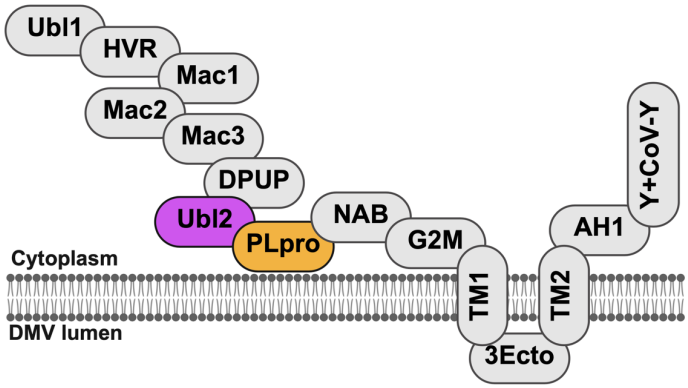


Figure 42: Domain organization of NSP3 highlighting PLpro and Ubl2. Ubl2 and PLpro are faced to the cytosol, being surrounded by the ubiquitin-like domain 1 (Ubl1), the hypervariable region (HVR), three macrodomains (Mac1-3), the domain preceding Ubl2 (DPUP), the nucleic acid binding domain (NAC), the G2M domain, two transmembrane domains (TM1, TM2), the ectodomain (3Ecto) and the very C-terminal domains AH1 and Y+CoV-Y. Figure was created in BioRender.com.

Structurally, PLpro is organized into four distinct domains. Three of these - termed the thumb, palm, and fingers - comprise the catalytic core (Figure 43).[206,207] A zinc-finger domain is located at the C-terminus and plays a critical structural role, coordinating a zinc ion via conserved cysteine residues (C189, C192, C224, and C226). Functional studies using SARS-CoV-1 PLpro knockout mutants demonstrated that zinc coordination is essential for protease activity.[208]

The thumb and palm domains are located internally and form the enzymatic core of the protease (Figure 43). The N-terminal region of PLpro is composed of the Ubl2 domain, which adopts a ubiquitin-like fold consisting of one α-helix and five β-strands, forming a compact, globular structure (Figure 43; PDB: 7CJD).

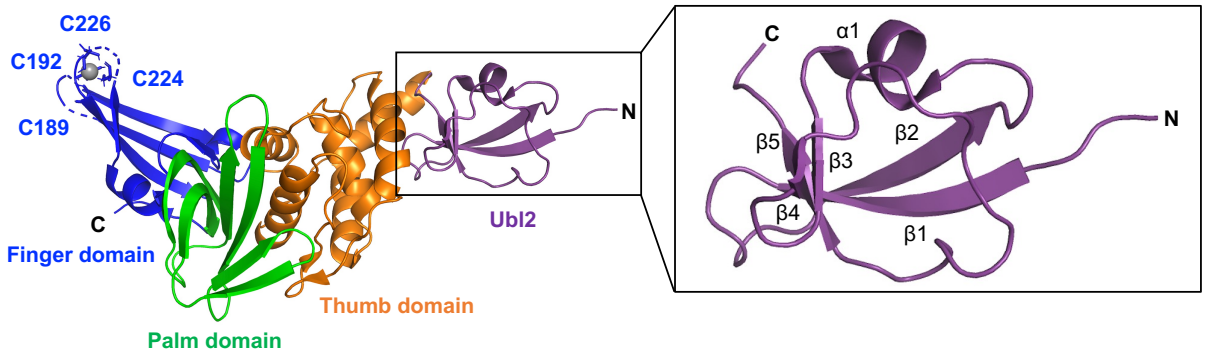


Figure 43: Structural organization of the PLpro domain. Subdomain organization of the SARS-CoV-2 papain-like protease (PLpro). Ribbon representation of the PLpro monomer (PDB: 7CJD) constituting the ubiquitin-like domain 2 (Ubl2, 1-60, purple), the thumb domain (61-178, orange), the zinc-binding finger domain (179-240 and 308-315, blue), and the palm domain (241-307, green). The zinc ion is stabilized by 4 cysteine residues within the finger domain (C189, C192, C224 and C226). Ubl2 is localized at the N-terminus of PLpro with the secondary structure elements β1-β2-α1-β3-β4-β5. Figure A and B were created in BioRender.com.

6.2.2 Function of PLpro and Ubl2

It has been shown that PLpro offers several functions in SARS-CoV-2 infected cells based on its enzymatic activity as a protease. At the very beginning of SARS-CoV-2 infection, PLpro mediates the proteolytic release of the first three NSPs, including NSP3;[208] an indispensable step for the formation of DMVs serving as replication organelles.[45,46,54,63,209] It has been shown that uncleaved NSP3-NSP4 is sufficient to induce membrane pairing but incapable of forming DMV pores, highlighting the essential role of PLpro proteolytic cleavage in DMV pore formation.[45] Recent findings by Huang *et al.* revealed an interaction between PLpro and DPUP, which contributes to the structural stabilization of the DMV pore.[46] Consistent with this observation, N-terminal truncation of NSP3 (ΔUbl1-Ubl2) including DPUP - affected pore integrity.[45]

In contrast to some other CoVs, such as MHV, which encode two PLpro domains, SARS-CoV-2 contains a single PLpro domain.[210] This functional role is complemented by Mpro, another viral cysteine protease responsible for cleaving downstream regions of the polyprotein.[24] PLpro specifically recognizes a conserved tetrapeptide motif (L-X-G-G↓X), where “X” denotes

residue variability and the cleavage occurs at the di-glycine site, enabling the release of NSP1, NSP2, NSP3, and NSP4 (Figure 44).[211,212] Each of these cleavage products exerts important functions during the viral replication cycle. For example, NSP1 suppresses host protein synthesis by binding to the 40S ribosomal subunit and obstructing the mRNA entry channel,[213,214] while NSP2 has been implicated in the regulation of cell cycle progression and mitochondrial biogenesis.[215,216]

ULK1 (unc-51-like kinase 1), a serine/threonine kinase, has been identified as a substrate of PLpro, resulting in the proteolytic cleavage between its N-terminal kinase domain and C-terminal substrate recognition region.[217] ULK1 plays a central role in the initiation of autophagy by phosphorylating multiple regulatory proteins involved in upstream autophagic signaling pathways.[218] Proteolytic cleavage of ULK1 impairs the initiation of the autophagy-related signaling cascade, thereby preventing autophagy-dependent cell death.[219] Notably, the recognition and cleavage side of PLpro as well as its structure is highly conserved along SARS-CoV-1 and SARS-CoV-2.[211] Consequently, PLpro has emerged as a promising antiviral target especially regarding preparedness for further pandemics. The PLpro domain of SARS-CoV-2 serves multiple essential functions in infected host cells, primarily through its cysteine protease activity. At the early stages of viral infection, PLpro mediates the proteolytic cleavage of the viral polyprotein, facilitating the release of NSP1-NSP3, including its own release from NSP3.[208] This processing step is indispensable for the formation of DMVs, which act as ROs during the viral life cycle.[45,46,54,63,209] Moreover, PLpro-mediated cleavage of NSP3-4 has been demonstrated to be essential for pore formation, while being dispensable for membrane pairing.[45]

Moreover, PLpro is capable of cleaving specific host proteins, thereby interfering with innate immune responses. One notable substrate is ULK1, a serine/threonine kinase involved in initiating autophagy. PLpro-mediated cleavage of ULK1 disrupts the connection between its N-terminal kinase domain and C-terminal substrate recognition domain,[217] effectively preventing the phosphorylation of downstream targets required for autophagy induction.[218,219] Consequently, this cleavage event inhibits autophagy-dependent cell death and may contribute to immune evasion. Notably, the overall structure of PLpro and its substrate recognition motif are highly conserved between SARS-CoV-1 and SARS-CoV-2,[211]

underlining its relevance as a strategic antiviral target. Due to its involvement in both viral polyprotein processing and the modulation of host defense mechanisms, PLpro is considered a promising candidate for the development of therapeutic inhibitors, with potential implications for pandemic preparedness and broad-spectrum antiviral strategies.

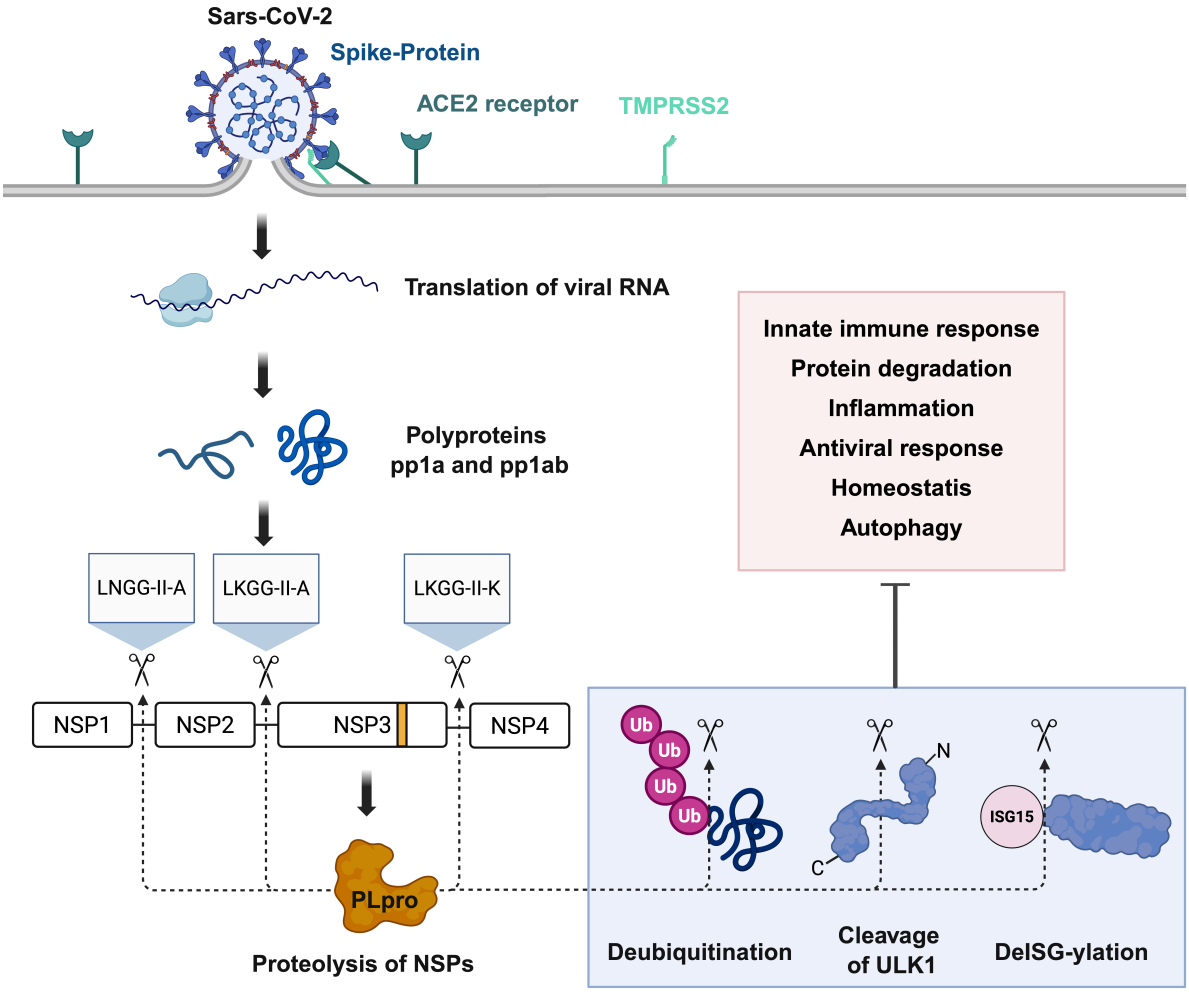


Figure 44: Effect of PLpro onto the viral replication cycle and the host cell's response. Upon viral entry, the viral genome is translated into two polyproteins (pp1a and pp1ab). The N-terminus of pp1a incorporates NSP1, NSP2, NSP3 and NSP4. PLpro is a catalytical active domain encoded in NSP3. Cleavage sites of PLpro are indicated by dotted arrows and scissors. Upon cleavage, NSP1, NSP2 and NSP3 are released, with NSP3 indispensable for the formation of DMVs.[56,45] PLpro is capable of deubiquitination and deISG-ylation, and further cleavage of ULK1, inhibiting the innate immune response, protein degradation, inflammation, antiviral response, homeostatis and autophagy.[220,217] Figure was created in BioRender.com.

In addition to its proteolytic role in viral polyprotein processing, PLpro plays a critical role in suppressing the host immune response by targeting PTMs of key immune-regulatory proteins.[208,221] Specifically, PLpro exhibits deubiquitinating and deISGylating activity, thereby interfering with the host's innate antiviral signaling pathways (Figure 44).[208,221] One of the key

targets of PLpro is the interferon-stimulating gene 15 (ISG15), a ubiquitin-like modifier that is structurally similar to HuUb and plays an essential role in antiviral defense.[222,223] Through its interaction with ISG15, PLpro reduces the level of ISG15-conjugated proteins in infected cells, leading to impaired type I interferon responses.[222] Structural studies have revealed that PLpro contains two distinct binding sites that accommodate the C-terminal tail of both ubiquitin and ISG15, recognizing a conserved L73-R74-G75-G76 motif.[220,222] Since PLpro consists of an ubiquitin-binding site,[205,221,222,224,225] an interaction with Ubl1 or Ubl2 was proposed. However, these domains do not exhibit the corresponding binding motif, excluding any interaction between Ubl1/2 and PLpro.

While the functional role of PLpro has been extensively studied, the function of Ubl2 remains controversial. Zimmermann *et al.* identified Ubl2 as one of four domains (Mac2, Mac3, DPUP) essential for NSP3 oligomerization and the structural integrity of the crown complex - both of which are critical for DMV pore formation.[45] Moreover, deletion of the ΔUbl1-Ubl2 region led to a reduction in NSP3-NSP4 cleavage products, highlighting a potential role for Ubl2 in facilitating PLpro-mediated proteolytic processing.[45] Since Ubl2 has been shown to modulate PLpro stability in delta coronaviruses,[226] it has been proposed that Ubl2, together with Mac2, Mac3, and DPUP, contributes to the overall structural stability of the crown.[45] Furthermore, deletion of the N-terminal region of NSP3 (ΔUbl1-Ubl2) failed to support proper formation of DMVs.[45] In contrast, a truncated construct extending to the first macrodomain was sufficient to induce characteristic ER membrane remodeling into distinct DMVs.[45] These findings suggest a potential role for Ubl2 - and possibly Mac2, Mac3 or DPUP - in mediating lipid-binding interactions., Frieman *et al*. reported that SARS-CoV-1 Ubl2 blocks the NF-κB and IRF3 pathway directly influencing the host innate immune response.[88]

Despite these advances, the precise function of the N-terminal Ubl2 domain of PLpro remains unclear and warrants further investigation.[205] Elucidating the role of the Ubl2 domain may provide new insights into how PLpro coordinates its enzymatic and regulatory activities, and could aid in the development of targeted inhibitors that disrupt its immunosuppressive functions.

6.2.3 PLpro as a Potential Drug Target

To date, several antiviral agents have been developed, most notably those targeting the RdRp complex, such as remdesivir and VV116.[227,228] However, the clinical efficacy of these compounds remains limited. Recently, increased attention has been directed toward the two viral cysteine proteases of SARS-CoV-2: Mpro and PLpro. Both enzymes represent attractive targets for antiviral drug development. While several inhibitors have already been developed against Mpro - most notably nirmatrelvir, which has shown considerable clinical success - efforts to target PLpro are still in earlier stages.[229] Unlike the S-protein, the PLpro domain exhibits a lower mutation rate, rendering it a promising target for antiviral drug development aimed at inhibiting viral replication and proliferation.[230,231] The high degree of sequence conservation within PLpro further enhances its potential as a candidate for therapeutic intervention and has already been studied in detail.[3,23,205,211,232] Although PLpro primes the release of NSP1, NSP2 and NSP3 and hence plays a critical role during the viral proliferation cycle, inhibition of PLpro got less attention.[233]

6.3 Aims

The two viral cysteine proteases of SARS-CoV-2, Mpro and PLpro, have recently gained increased attention as promising targets for antiviral therapy. While effective inhibitors such as nirmatrelvir have already been developed for Mpro,[229] drug development targeting PLpro remains in its early stages. This limited progress is mainly due to an incomplete understanding of PLpro's structural dynamics, particularly the role of its Ubl2 subdomain.[205]

Therefore, the objective of this subproject was to produce high-purity, high-yield protein samples of NSP3-Ubl2 and NSP3-PLpro for subsequent structural characterization. Expression optimization was conducted to improve recombinant production of Ubl2 and PLpro. Structural insights into the Ubl2 domain were obtained using CD and NMR spectroscopy. 3D NMR measurements of Ubl2 were conducted to complete the backbone resonance assignment of PLpro, which had been partially reported in previous work. Additionally, the secondary structure of Ubl2 was analyzed based on secondary structure propensity predictions and CD spectral data. A TROSY spectrum of PLpro was also recorded and compared with the chemical

shift assignments published by Shimada *et al.* (2021).[234] Finally, the acquired data were interpreted in the context of the SARS-CoV-2 replication cycle.

6.4 Material and Methods

6.4.1 Buffer, Chemicals and Media

All standard chemicals used for the respective analyses were purchased from AppliChem, Merck, Roche and Sigma. Cell culture media were obtained from AppliChem. Medium, chemicals and buffers used for the experimental procedures are listed in the respective paragraphs. To avoid contamination during the experiment, chemicals used for cellular experiments were autoclaved and stored in sterile vessels. Buffers used for protein isolation were previously filtered and degassed.

6.4.2 Instruments

Instruments and materials used for the expression and isolation of ^{1}H-^{15}N Ubl2, ^{15}N-^{13}C Ubl2 and ^{1}H-^{15}N PLpro are listed in Table 25 and 26.

Table 25: List of instruments used for the expression and isolation of ^{1}H-^{15}N Ubl2, ^{15}N-^{13}C Ubl2 and ^{1}H-^{15}N PLpro.

Instrument	Manufacturer
Agarose-Gel chamber system	Bio-Rad Laboratories GmbH, München
ÄKTA Pure	GE Healthcare, Sweden
Autoclave VX-150	Systec GmbH, Linden
Bruker 700 MHz Avance III NMR spectrometer	Bruker, Billerica, MA, USA
Centrifuge 5430 R	Eppendorf GmbH, Hamburg
Centrifuge Avanti J-26S XP	Beckmann Coulter GmbH, Krefeld
Desk centrifuge 5415 R	Eppendorf GmbH, Hamburg
Dri-Block Heater 08-3	Techne, UK
Drying chamber T 5050	Heraeus, Hanau
Electrophoresis power supply 3501 XL	Amersham Pharmacia biotech, UK
Gel documentation system	Bio-Rad Laboratories GmbH, München
Incubation shaker Infors HT Multitron	Infors AG, Bottmingen
Incubation shaker New Brunswick innova 40	Eppendorf GmbH, Hamburg
Incubation shaker New Brunswick innova 44	Eppendorf GmbH, Hamburg

JASCO J-815 CD spectropolarimeter	Jasco, Gross-Umstadt
Magnetic stirrer C-MAG MS 7	IKA GmbH & CO. KG, Staufen
Milli-Q Direct 16	Merck KGaA, Darmstadt
Mini Trans-Blot module blotting chamber	Bio-Rad Laboratories GmbH, München
NanoDrop 2000	Thermo Fischer Scientific, Schwerte
pH meter	Satorius AG, Göttingen
Scale	Satorius AG, Göttingen
Sonication device Sonoplus	Bandelin electronic GmbH & Co KG, Berlin
Thermocycler compact	Eppendorf, Hamburg
T100 Thermal Cycler	Bio-Rad Laboratories GmbH, München
Vortexer	Heidolph Instruments, Schwabach
VS70T sonotrode	Bandelin electronic GmbH & Co KG, Berlin

Table 26: List of consumables for the expression and isolation of ^{1}H-^{15}N Ubl2, ^{15}N-^{13}C Ubl2 and ^{1}H-^{15}N PLpro.

Material	**Manufacturer**
Amicon MWCO 3 kDa	Merck Millipore, Darmstadt, Germany
Amicon MWCO 10 kDa	Merck Millipore, Darmstadt, Germany
cOmplete EDTA-free protease inhibitor tablet	Roche, Basel
Der Blaue Jonas - Single-step Coomassie Blue protein gel dye	GRP German Research, Haag a. d. Amper
Deuteriumoxid	Carl Roth, Karlsruhe
DNAse I, recombinant	Roche, Basel
E. coli BL21 (DE3) cells	New England Biolabs, Ipswich, USA
Goat anti-Mouse lgG (H+L) Secondary Antibody HRP conjugated	Thermo Fischer Scientific, Schwerte
HiLoad 16/60 Superdex 75 pg	Cytiva, Marlborough
HiLoad 16/600 Superdex 200 pg	Cytiva, Marlborough
Falcon tubes (15 mL and 50 mL)	Sarstedt, Nürnbrecht
MEM Vitamin Solution (100x)	Sigma-Aldrich, St. Louis, USA
Norell Select Series 3 mm NMR Tubes	Sigma-Aldrich, St. Louis, USA
PageRuler Prestained Protein Ladder	Thermo Fisher Scientific, Schwerte

Protino Ni-NTA 5 mL FPLC column	Macherey-Nagel, Düren
Plastic cuvettes	Brand GmbH & Co KG, Wertheim
Petri dishes	Hartenstein, Würzburg
Pipette tips	Sarsted, Nürnbrecht
PVDF-membrane 0.2 µm	GE Healthcare, Sweden
Quartz cuvettes – 1 mm path length	Hellma, Müllheim
Reaction tube (1.5 mL and 2 mL)	Sarstedt, Nürnbrecht
Regenerated Cellulose Membrane Filter 0.2 µm	Sartorius, Göttingen
Serological pipettes	Sarstedt, Nürnbrecht
SnakeSkin dialysis tubes (3 K MWCO)	Thermo Fischer Scientific, Schwerte
SnakeSkin dialysis tubes (10 K MWCO)	Thermo Fischer Scientific, Schwerte
Spectra Multicolor Low Range Protein Ladder	Thermo Fischer Scientific, Schwerte
Sterile Syringes for Single Use (3 mL, 20 mL) Fisherbrand	Thermo Fischer Scientific, Schwerte
Syringe filter, Filtropur S, PES, pore size 0.2 µm for sterile filtration	SARSted, Nürnbrecht
Sterican Standard canulla 0.45 x 25 mm	B. Braun, Melsungen
SuperSignal West Pico PLUS Chemiluminescent Substrate	Thermo Fischer Scientific, Schwerte
Whatman Cellulose Blotting Paper	Cytiva, Marlborough
2xYT medium	Thermo Fischer Scientific, Schwerte
6x-His Tag Monoclonal Antibody (HIS.H8)	Thermo Fischer Scientific, Schwerte
^{15}N-NH_4-Cl	99%, Cambridge Isotope Laboratories, USA
^{13}C D-Glucose	99%, Cambridge Isotope Laboratories, USA

6.4.3 Buffers and Solutions

The composition of all buffers and solutions utilized in this study is summarized in Table 27. Unless otherwise specified, all buffers were prepared using distilled Milli-Q water with a resistivity of 18.2 MΩ·cm at 25 °C, ensuring high purity and minimal contamination. Buffers employed for protein isolation and purification procedures were degassed to remove dissolved gases, which can interfere with chromatographic performance, and subsequently filtered through 0.45 µm pore-size membranes to eliminate particulate contaminants and ensure sterility.

Table 27: Chemical solutions used for the expression and isolation of Ubl2 and PLpro.

Solution	Ingredients and Concentrations
Acrylamide:Bisacrylamide (34.3%:1.1%)	34.3% (v/v) Acrylamide, 1.1% (v/v) Bisacrylamide
Acrylamide:Bisacrylamide (37.5%:1%)	37.5% (v/v) Acrylamide, 1% (v/v) Bisacrylamide
Anode buffer (10x)	2 M Tris-HCl (pH 8.9)
Antibiotica stock	50 mg/mL Kanamycin
Blocking solution	5% milk (w/v) in (1x) TBS-T
Cathode buffer	1 M Tris-HCl (pH 8.25), 1 M Tricine, 1% (w/v) SDS
CD buffer Ubl2	50 mM sodium phosphate (pH 6.0)
Column storage buffer	20% (v/v) EtOH
DNase I stock solution	20 mM Tris-HCl, 50 mM NaCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 50% (v/v) glycerol
Dialysis buffer Ubl2	50 mM Tris (pH 6.0), 250 mM NaCl, 2 mM DTT, 1 mM EDTA
Dialysis buffer PLpro	20 mM Tris (pH 7.0), 250 mM NaCl, 0.3 mM TCEP
Glycerol stock solution	31 g 2xYT powder per L, 50% (v/v) glycerol
Induction reagent	0.5 M IPTG
IMAC elution buffer Ubl2	50 mM sodium phosphate (pH 6.0), 250 mM NaCl, 500 mM imidazole
IMAC elution buffer PLpro	20 mM Tris (pH 7.0), 300 mM NaCl, 1 mM TCEP, 300 mM imidazole
Laemmli sample buffer (4x)	10% (w/v) SDS, 10 mM *β*-mercaptoethanol, 20% (v/v) glycerol, 0.05% (w/v) Bromphenol-Blue, 200 mM Tris- HCl (pH 6.8)
Lysis buffer Ubl2	50 mM sodium phosphate (pH 6.0), 250 mM NaCl
Lysis buffer PLpro	20 mM Tris (pH 7.0), 300 mM NaCl, 1 mM TCEP, 1 µM $ZnSO_4$
TBS-T buffer (20x)	0.5 M Tris (pH 7.6), 3M NaCl, 2% (v/v) Tween20

TGS buffer (10x)	25 mM Tris-HCl (pH 8.3), 192 mM glycine, 0.1% (w/v) SDS
TG buffer (10x)	25 mM Tris-HCl (pH 8.3), 192 mM glycine
Towbin buffer (1x)	10% (v/v) TG buffer (10x), 20% (v/v) EtOH
Tris-Glycine Gel Buffer (Stacking Gel)	0.5 M Tris (pH 6.8)
Tris-Glycine Gel Buffer (Separating Gel)	1.5 M Tris (pH 8.8)
Tris-Tricine Gel Buffer (3x)	3 M Tris-HCl (pH 8.45), 0.3% (w/v) SDS
SEC buffer Ubl2	50 mM sodium phosphate (pH 6.0), 150 mM NaCl
SEC buffer PLpro	50 mM MES (pH 6.5), 100 mM NaCl, 1 mM TCEP

6.4.4 Plasmids and Bacterial Strains

The plasmids encoding for Ubl2 and PLpro were ordered at BioCat (Heidelberg, Germany). The construct composition and further information is shown in Table 28.

Table 28: Construct details of Ubl2 and PLpro.

Construct Information	**Ubl2**	**PLpro**
Original organism	SARS-CoV-2	SARS-CoV-2
Plasmid	pET-28a(+)	pET-28a(+)
Codon-optimized	*E. coli*	*E. coli*
Cell strains	*E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) – short, *E. coli* BL21 (DE3)	*E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) – short, *E. coli* BL21 (DE3)
N-terminal Tag	6xHis	6xHis
Cleavage Sites	Thrombin, TEV	Thrombin, TEV
Antibiotic Resistance	Kanamycin	Kanamycin
Restriction sites	Seamless cloning (CTCGAG), XhoI	Seamless cloning (CTCGAG), XhoI
Construct size with tag	9.98 kDa, 88 amino acids	38.81 kDa, 343 amino acids
Construct size	7.15 kDa, 63 amino acids	35.99 kDa, 318 amino acids
pI of His-tagged construct	7.14	8.18

pI of construct	6.79	8.17
Extinction coefficient	His-Ubl2: 5960 M^{-1} cm^{-1}	His-PLpro: 47385 M^{-1} cm^{-1}
(non-reduced conditions)	Ubl2: 4470 M^{-1} cm^{-1}	PLpro: 45895 M^{-1} cm^{-1}
Extinction coefficient	His-Ubl2: 5960 M^{-1} cm^{-1}	His-PLpro: 46760 M^{-1} cm^{-1}
(reduced conditions)	Ubl2: 4470 M^{-1} cm^{-1}	PLpro: 45270 M^{-1} cm^{-1}

In the following the amino acid sequences of His-tagged and digested constructs are listed. The His-tag is colored in green, the thrombin cleavage side in purple, the TEV cleavage side in blue and the construct itself in red. The stop codons are marked with "*". As Ubl2 is a subdomain of PLpro, the Ubl2 sequence is highlighted in the PLpro sequence by bold fonts.

6xHis-Thrombin-TEV-Ubl2:
MGSSHHHHHHSSGLVPRGSENLYFQSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGAD
VTKIKPHNSHEGKTFYVLPN*

Ubl2:
SLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPHNSHEGKTFYVLPN*

6xHis-Thrombin-TEV-PLpro:
MGSSHHHHHHSSGLVPRGSENLYFQ**SLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGA
DVTKIKPHNSHEGKTFYVLPN**DDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIKWAD
NNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELGDVRETMSYLFQHANL
DSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQIPCTCGKQATKYLVQQESPFVMMS
APPAQYELKHGTFTCASEYTGNYQCGHYKHITSKETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIK*

PLpro:
SLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPHNSHEGKTFYVLPNDDTLR
VEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIKWADNNCYLATALLTLQQIELKFNPPALQDA
YYRARAGEAANFCALILAYCNKTVGELGDVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEA
VMYMGTLSYEQFKKGVQIPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGH
YKHITSKETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIK*

6.4.5 Sequencing of Ordered Gene Constructs

For sequencing, 12 µL of *E. coli* codon-optimized plasmids were sent in a concentration between 50 and 100 ng/µL to Microsynth Seqlab (Göttingen). As a primer source, the standard primer list of Microsynth Seqlab was used. Primers used for sequencing are listed in Table 29. The online tool NucleoBlast[89] was used for the alignment of the resulting and theoretical sequences. Sequencing results of the constructs are listed in the supplementary data (Table SD1).

Table 29: Primer list used for sequencing of Ubl2 and PLpro plasmids.

Construct	Plasmid concentration	Primer name	Primer sequence
pET-28a(+)-6xHis-Thrombin-TEV-Ubl2	100.05 ng/µL	T7	TAATACGACTCACTATAGGG
pET-28a(+)-6xHis-Thrombin-TEV-Ubl2	100.05 ng/µL	T7term	TAATACGACTCACTATAG
pET-28a(+)-6xHis-Thrombin-TEV-PLpro	98.86 ng/µL	T7	TAATACGACTCACTATAGGG
pET-28a(+)-6xHis-Thrombin-TEV-PLpro	98.86 ng/µL	T7term	TAATACGACTCACTATAG

6.4.6 Transformation of Plasmids in Chemically Competent *E. coli* BL21 (DE3) Cells

Constructs were transformed in *E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) - short, *E. coli* BL21 (DE3) (New England Biolabs, Ipswich, USA). Firstly, 50 µL cells were thawed for 20 min on ice. Then, 10 ng plasmid was added to the cell suspension and incubated for 30 min on ice. Cells were heat-shocked for 45 s at 42 °C. After that, cells were cooled down for 2 min on ice. Then, 450 µL of preheated 2xYT medium were mixed with the cell suspension and incubated for 1 h at 37 °C (800 rpm). Cells were plated onto LB-agar plates containing the corresponding antibiotics and were incubated at 37 °C overnight. The next day, grown colonies were picked and cultured in 5 mL 2xYT medium supplemented with antibiotics at 37 °C and 160 rpm overnight. Cells were centrifuged for 1 min at 16,000 x *g*. The supernatant was discarded, and the cell pellet was resuspended in glycerol stock solution (50% (v/v) 2xYT medium, 50% (v/v) glycerol). Cells were flash frozen in liquid nitrogen and stored at -80 °C for further use. Glycerol stocks were used for expression tests and large-scale expression in minimal medium.

6.4.7 Expression Tests of Ubl2 and PLpro

Expression tests were used to identify optimal expression conditions for both NSP3 domains. For that, different temperature levels after induction of expression were quantified to optimize gene expression. As minimal medium was required for later protein labelling, small scale expression tests of Ubl2 and PLpro were performed in M9 medium and compared to expression in 2xYT medium. For that, 5 mL preculture of both constructs were inoculated with the corresponding glycerol stock and 50 mg/mL Kanamycin for selection. Precultures were incubated at 37 °C and 160 rpm overnight (16 h). The next day, 50 mL of fresh minimal medium was mixed with 0.5 mL preculture and 50 µL of the corresponding antibiotic in a 250 mL baffled Erlenmeyer flask. The main culture was incubated at 37 °C and 160 rpm. At an

OD_{600} of 0.8, a sample of 1 mL was taken to analyze the gene expression before induction. For that, the 1 mL culture was centrifuged for 1 min at 16,000 x *g*. The medium was discarded, and the pellet was diluted in 45 µl pure water and 15 µl 6x Laemmli SDS-buffer[92]. Gene expression was induced by 0.5 mM isopropyl-*ß*-D-1-thiogalactopyranoside (IPTG) causing the activation of the T7-promotor system. After another incubation time overnight at the corresponding temperature, a second 1 mL culture sample was taken after induction and mixed with Laemmli SDS-buffer[92]. Expression was analyzed by SDS-PAGE and Western Blot analysis.

6.4.8 Large-Scale Expression of labelled Ubl2 and PLpro in Minimal Medium

For large-scale expression, a preculture of 50 mL 2xYT medium was supplemented with 50 mg/mL Kanamycin in 250 mL baffled flasks. The culture was incubated at 37 °C and 160 rpm overnight. For the expression of ^{1}H-^{15}N Ubl2 and PLpro, two heat-sterilized two-liter flasks were filled with 500 mL minimal medium each (preparation of 1 L), each inoculated with 10 mL of preculture and supplemented with 50 mg/mL Kanamycin. The recipe of minimal medium is shown in Table 30.

Table 30: Protocol for the expression of Ubl2 and PLpro in 1 L minimal medium.

Chemical	**Stock solution**	**Final concentration**	**Amount for 1 L main culture**	**Amount for 50 mL preculture**
M9 salt solution	337 mM Na_2HPO_4, 220 mM KH_2PO_4, 85.5 mM NaCl (10x)	33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.55 mM NaCl (1x)	100 mL	5 mL
MEM vitamin solution	100x	1x	1 mL	50 µL
$MgSO_4$	1 M	1 mM	1 mL	50 µL
$CaCl_2$	0.1 M	0.3 mM	3 mL	150 µL
Biotine	1 mg/mL (w/v)	1 mg/L	1 mL	50 µL
Thiamine	1 mg/mL (w/v)	1 mg/L	1 mL	50 µL
Trace element solution	13.4 mM EDTA, 3.1 mM $FeCl_3$-$6H_2O$	134 µM EDTA, 31 µM $FeCl_3$-$6H_2O$	10 mL	500 µL

	0.62 mM $ZnCl_2$	6.2 µM $ZnCl_2$		
	76 µM $CuCl_2$-$2H_2O$	0.76 µM $CuCl_2$-$2H_2O$		
	42 µM $CoCl_2$-$2H_2O$	0.42 µM $CoCl_2$-$2H_2O$		
	162 µM H_3BO_3	1.62 µM H_3BO_3		
	8.1 µM $MnCl_2$-$4H_2O$	0.081 µM $MnCl_2$-		
	(100x)	$4H_2O$		
		(1x)		
^{15}N NH_4Cl	99% (w/w)	0.1% (w/v)	1 g	50 mg
^{12}C D-Glucose-Monohydrate	20% (w/v)	0.4% (w/v)	20 mL	200 µL
^{13}C D-Glucose	99% (w/w)	0.4% (w/v)	4 g	200 mg

For the expression of ^{15}N-^{13}C Ubl2, ^{13}C D-Glucose instead of ^{12}C D-Glucose was used. At OD_{600} of 0.8, 0.5 mM IPTG was added inducing the expression of target proteins. In the case of PLpro, 100 µM $ZnSO_4$ was supplemented to the medium to promote proper metal incorporation during the expression of the metalloprotein. Samples before and after induction of expression were used to analyze the expression grade of both proteins. After another incubation period overnight at 18 °C (PLpro) or 25 °C (Ubl2) and 120 rpm, cells were harvested at 4 °C via centrifugation at 6,000 x *g* for 15 min. After centrifugation, the medium was discarded. The pellet was weighted and stored at -20 °C for further use.

6.4.9 Cell Lysis by Sonication

The frozen cell pellet was thawed and resuspended in lysis buffer (5 mL lysis buffer/g cell pellet). For DNA degradation 10 µg per g cell pellet DNase I were added to the cell suspension followed by an EDTA-free protease inhibitor tablet to decrease protease activity. The cell suspension was incubated for 20 min on a roller mixer at 4 °C until the solution became homogenous. Samples were ultrasonicated using a VS70T sonotrode for 20 min. For that, the amplitude was set to 60%, with intervals of 1 s sonication and 3 s pause. As heat is generated during sonication, the sample was put onto an ice-water bath previously. After cell lysis, the suspension was centrifuged for 45 min at 40,000 x *g* at 4 °C using a JA20-rotor. Supernatant and cell pellet were separated, and samples analyzed by SDS-PAGE. The supernatant was used for protein purification by IMAC using an ÄKTA start system.

6.4.10 SDS-PAGE of Ubl2 and PLpro

SDS-PAGE was used to quantify successful expression and purification of samples allowing protein separation by mass.[92,93] Based on the MW of Ubl2 and PLpro, a Tris-Glycine and Tris-Tricine SDS-PAGE were assessed (Table 31 and 32). The exact protocols for Tricine and Glycine gels are shown in 6.4.10.1 and 6.4.10.2. Independent of the gel, samples were diluted in 15 µl 4x Laemmli SDS-buffer and boiled at 95 °C for 10 min for protein denaturation. Samples were centrifugated for 10 min at 10,000 x *g*. Samples and the protein standard marker were loaded on the corresponding SDS gel and cable-connected to the electrophoresis power supply 3501 XL. When the blue running front was nearly at the bottom, the tank was taken off. After electrophoresis, the gel was washed four times with pure water prior to staining for 1 h with Coomassie (Blauer Jonas). Destaining was not required. Finally, the separated proteins were visualized by a gel documentation system.

6.4.10.1 Tris-Tricine SDS-PAGE of Ubl2

A running and stacking SDS-gel were prepared (Table 31). 3 µL of marker (Spectra Multicolor Low Range Ladder) and 10 µL of each sample were added. The electrophoresis was performed with a constant current strength of 40 mA per gel in running buffer. For that, the gel chamber was filled with 1x cathode buffer and surrounded with 1x anode buffer.

Table 31: Tris-Tricine SDS-PAGE protocol for the preparation of 2 gels.

Component	20% Separating Gel	5.6% Stacking Gel
ddH_2O	-	3.2 mL
Gel buffer (3x)	5 mL	2.1 mL
Glycerol	1.6 mL	-
Acrylamide:Bisacrylamide (34.3%:1.1%)	8.5 mL	1 mL
APS (10% (w/v))	50 µL	50 µL
TEMED	25 µL	25 µL

6.4.10.2 Tris-Glycine SDS-PAGE of PLpro

The gel composition of glycine gels is shown in Table 32. For the run, 3 µL of marker (Page Ruler prestained) and 10 µL of each sample were added onto the gel. The electrophoresis was performed with a constant current strength of 120 mV in running buffer (1x TGS buffer).

Table 32: Tris-Glycine SDS-PAGE protocol for the preparation of 2 gels.

Component	15% Separating Gel	5% Stacking Gel
ddH_2O	-	3.2 mL
Tris-Glycine Gel Buffer (Separating Gel)	5 mL	2.1 mL
Tris-Glycine Gel Buffer (Stacking Gel)	1.6 mL	-
Acrylamide:Bisacrylamide (37.5%:1%)	8.5 mL	1 mL
SDS (10% (w/v))	50 µL	50 µL
APS (10%(w/v))		
TEMED	25 µL	25 µL

6.4.11 Western Blot Analysis

Western blot analysis was used to identify the protein of interest in a SDS-gel.[94] Therefore, samples were separated using SDS-PAGE as described in section 6.4.10. For generating a Western-Blot-Sandwich, a PVDF-membrane was activated by 100% (v/v) EtOH and six Watman papers were soaked in (1x) Towbin buffer. The SDS gel was put onto the PVDF-membrane and surrounded with three pieces of Watman papers from each side. The transfer was performed in a blotting chamber for 30 min at 25 V (1 A). To check if the transfer was successful, the membrane was incubated with Ponceau-S-solution shortly and washed four times with water. Visible bands gave a hint on a successful protein transfer. Then, the membrane was blocked in blocking solution for 1 h at room temperature to saturate all antibody sites. After that, the milk was discarded, and the 6x-His Tag Monoclonal Antibody (HIS.H8) was 1:1000 diluted in blocking solution. The membrane was stored at 4 °C overnight under gentle shaking. The next day, the primary antibody solution was recycled and frozen at -20 °C for further use. Excess milk was remoted by water. The membrane was washed three times in 1x TBS-T for 5 min. The Goat anti-Mouse lgG (H+L) Secondary Antibody HRP conjugated covalently linked to the horseradish peroxidase was 1:10.000 diluted in blocking solution, added to the membrane, and incubated for 1 h at room temperature under gentle shaking. The antibody solution was discarded, and the membrane was washed again three times for 5 min in 1x TBS-T. The membrane was documented using the Bio-Rad Universal

Hood III documentation system. For signal detection, a 1:1 dilution of Super Signal Peroxidase Solution and Enhancer Solution were added on the membrane and incubated for 1 min in a dark room. The signal accumulation mode was utilized in high resolution mode, with 40 frames in 200 s. The protein ladder was visualized separately by using the colorimetric mode. Finally, images were merged and analyzed.

6.4.12 IMAC of His-tagged Ubl2 and PLpro

IMAC was used to separate His-tagged proteins from others according to their affinity to nickel ions on a column.[95] For purification, a Protino Ni-NTA 5 mL FPLC column connected to an ÄKTA Pure purification system from GE Healthcare (Sweden) was used. The purification was performed at room temperature. Buffers were filtered and degassed for the isolation. For the purification process, two different buffers systems were used: lysis buffer and elution buffer supplemented with imidazole for protein elution. Firstly, the system was washed with degassed and filtered ddH_2O until the absorbance and conductivity reached 0 mAU. Then, the system was equilibrated with 10 CVs of lysis buffer at a flow rate of 2 mL/min and 0.3 mPa. As the flowrate was decreased to 0.5 mL/min, the column was assembled, and the system was washed again with 10 CVs of lysis buffer (2 ml/min). Then the sample was loaded onto the column (1 mL/min) and the flow through was collected for SDS-PAGE analysis. After that the column was washed again with lysis buffer until the absorbance at 280 nm reached the baseline. Then, the amount of elution buffer was increased to 4% to remove unspecifically bound proteins. After washing, target proteins were eluted using a gradient of elution buffer (40 min, 1 mL/min). As the percentage of elution buffer was increased to 100 %, the target protein was fully eluted. Fractions containing the target were pooled. The column and system were washed with ddH_2O and stored in 20% (v/v) EtOH.

6.4.13 Digestion and Dialysis of Target Proteins

The affinity tag of targets was removed by proteolytic cleavage. TEV-protease prepared in-house (see Chapter 1) was used for the digestion of 6xHis-tagged proteins removing the poly-histidine tag at the N-terminus of the protein. Constructs were digested in a ratio of 1:100 (1 mg TEV-protease added to 100 mg target protein). The approach was performed in combination with dialysis overnight to decrease the imidazole concentration in the sample. For that, the protein mix was put in a SnakeSkin dialysis tube (3 K MWCO for Ubl2 and

10 K MWCO for PLpro) surrounded by 1 L of the corresponding dialysis buffer (for buffer composition see Table 27). The sample was incubated overnight at 4 °C under gentle stirring.

6.4.14 Size Exclusion Chromatography of Ubl2 and PLpro

The concentrated protein was further purified using SEC. For the isolation, Ubl2 was purified using a HiLoad 16/600 Superdex 75 pg and PLpro using a HiLoad 16/600 Superdex 200 pg. Due to the sensitivity of the column, the buffer was sterile filtered and degassed previously. Firstly, the column was equilibrated with SEC buffer (see Table 27 for buffer composition) at a flowrate of 0.5 mL/min. After 8 h of equilibration (2 CVs), the concentrated sample was injected using a 3 mL syringe. The run was performed at 1 mL/min, 0.3 MPa delta column pressure and 0.5 MPa precolumn pressure until 1 CV was reached. Samples of the main peak were collected for SDS-PAGE. The column was washed with degassed and filtered ddH_2O at 0.5 mL/min (2 CVs), followed by an additional washing step with 20% (v/v) EtOH (1 CV) at 0.3 mL/min. Fractions were pooled and concentrated by centrifugation with an Amicon tube (MWCO 3 K for Ubl2 and MWCO 10 K for PLpro) at 6,000 x *g* until a sample concentration of around 200 µM was reached. The exact protein concentration was determined using NanoDrop and UV-VIS spectroscopy. Samples were flash-frozen in liquid nitrogen and stored at -80 °C.

6.4.15 NMR Measurements of Ubl2 and PLpro

6.4.15.1 NMR Sample Preparation of ^{1}H-^{15}N Ubl2 and PLpro

For NMR measurements, pure and high-quality protein samples were supplemented with 5% (v/v) D_2O, 0.1% (v/v) NaN_3 and 0.2 µM DSS (final concentration) and filled up to 200 µL with the corresponding SEC buffer. Samples were transferred into a Norell Select Series 3 mm NMR Tube and measured at different spectrometers at the Forschungszentrum Jülich (Germany). Exact protein concentrations and spectrometers are listed in the corresponding paragraphs.

6.4.15.2 TROSY Measurement of ^{1}H-^{15}N PLpro

Transverse relaxation-optimized spectroscopy (TROSY) experiments were performed using a 75 µM ^{1}H-^{15}N-labeled PLpro sample in 50 mM MES (pH 6.5), 100 mM NaCl, 1 mM TCEP, 5% (v/v) D_2O, 0.01% (v/v) NaN_3 buffer, supplemented with 0.2 µM DSS for referencing. Data

were recorded on a Bruker 1.2 GHz Avance Neo NMR spectrometer (Bruker, Billerica, MA, USA), equipped with a cryogenic triple resonance ^{1}H, ^{15}N, ^{13}C probe. Measurements were performed at 25 °C (298 K). Spectra were processed using the Bruker Topspin 3.5 software. The software CCPN NMR Analysis Assign 3.0. was used for data illustration.[125] This measurement has been performed in cooperation with the master student Frederike Nith under the supervision of Katharina Vormann and Tobias Stief.

6.4.15.3 3D Measurement of ^{15}N-^{13}C Ubl2

NMR spectral backbone resonance assignments were performed using a 170 µM ^{15}N-^{13}C-labeled Ubl2 wildtype sample in 50 mM sodium phosphate (pH 6.0), 150 mM NaCl, 5% (v/v) D_2O, 0.01% (v/v) NaN_3 buffer, supplemented with 0.2 µM DSS for referencing. Standard three-dimensional HNCO, HNCA, HN(CO)CA, HNCACB and HN(CO)CACB NMR experiments were acquired on a Bruker 700 MHz Avance III NMR spectrometer (Bruker, Billerica, MA, USA), equipped with a cryogenic triple resonance ^{1}H, ^{15}N, ^{13}C probe at 298 K. Spectra were processed using the Bruker Topspin 3.5 software. The software CCPN NMR Analysis Assign 3.0. was used to assist with manual backbone resonance assignments.[125] Residue-specific secondary chemical shift data were derived from the extracted backbone chemical shifts minus random-coil chemical shifts predicted by the POTENCI software[126] considering the Ubl2 amino acid sequence and buffer conditions. Experimental settings are listed in Table SD2 and SD3. The NMR resonance assignment (^{1}H, ^{15}N, ^{13}C, and $^{13}C_\alpha$ and $^{13}C_\beta$) of Ubl2 (1-63) has been performed in cooperation with the master student Frederike Nith under supervision of Katharina Vormann and Tobias Stief. Chemical shift values are displayed in Table SD4.

6.4.16 CD Spectroscopy of Ubl2

Far-UV-CD spectroscopy of an NSP3-Ubl2 protein sample was performed on a JASCO J-815 CD spectropolarimeter using 1 mm path length quartz cuvettes. Spectra were recorded at 5 °C, 15 °C, 20 °C, 25 °C and 37 °C using 300 µL 18.00 µM Ubl2 in 50 mM sodium phosphate (pH 6.0) (CD buffer Ubl2). As a blank, 300 µL of CD buffer was used. The following instrument settings were used for the measurement: 0.5 nm step size, 50 nm/min scan speed, 1 nm band with. The signal-to-noise ratio was improved by the accumulation of 10 scans per sample. The mean residue ellipticity $[\theta]_{MRW}$ in units of deg*cm^2*dmol^{-1} at wavelength is given by: $[\theta]_{MRW}$,

$\lambda = (\theta_{obs}*MRW)/(c*d*10)$, where θ_{obs} is the observed ellipticity at the wavelength λ (degrees); c is the concentration (M), and d is the cell path length (cm). The mean residue weight (MRW) is calculated by the molecular weight (MW: 7150 g/mol), divided by the number of peptide bonds (PB: 62) minus 1: MRW = MW/(PB-1). Secondary structure analysis by BeStSel[235] was applied to all CD data recorded in the 190 to 250 nm range with a scale factor of 1.0.

6.5 Results

The following sections provide a detailed account of the expression and purification procedures for PLpro and its subdomain Ubl2. For structural analyses such as those performed by NMR spectroscopy, the production of highly pure and homogeneous protein is essential. In addition, sufficient protein yield is a critical prerequisite for downstream applications. Therefore, the initial focus was placed on achieving high-level expression of the target proteins. To optimize protein yield, several experimental conditions were systematically tested, including variations in medium composition and post-induction temperature.

6.5.1 Optimization of Ubl2 Expression in *E.coli* Cells

Small-scale expression trials of 6xHis-TEV-Ubl2 were carried out under varying media and temperature conditions following induction. *E. coli* BL21 (DE3) cells harboring the expression construct were initially cultured overnight in 2xYT medium at 37 °C. Subsequently, main cultures were inoculated in 2xYT medium and incubated overnight at either 18 °C or 25 °C after induction. Optimal expression conditions were determined based on the presence of strong protein bands at the expected molecular weight of Ubl2 (~10 kDa) observed by SDS-PAGE and Western blot analysis. Samples were collected both before induction (BI) and after induction (AI), as shown in Figure 45A and 10B.

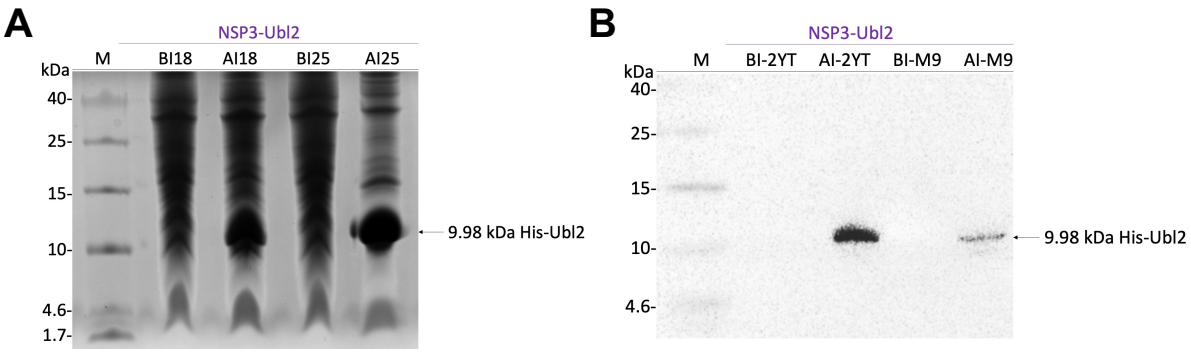


Figure 45: Expression tests of ^{1}H-^{15}N 6xHis-Ubl2. (A) *E. coli* BL21 (DE3) cells incorporating the pET-28a(+)-6xHis-Thrombin-TEV-6xHis-Ubl2 plasmid were cultivated in 2xYT at 18 °C (AI18) or 25 °C (AI25) overnight. Samples before (BI) and after induction (AI) of expression were taken and analyzed by a 20% (v/v) Tris-Tricine SDS-PAGE. Protein bands were compared to the marker (M) (Spectra Low Range Protein Ladder). 6xHis-tagged Ubl2 was expected at 9.98 kDa. **(B)** Expression test of Ubl2 expressed in 2xYT (BI-2YT, AI-2YT) or M9 medium (BI-M9, AI-M9) at 25 °C post induction of expression. The His-tagged protein was identified by Western Blot analysis at the proposed size, mainly expressed in 2xYT medium after induction of expression.

Given that minimal medium (M9) is essential for isotopic labeling of proteins, additional expression trials were conducted in M9 medium and directly compared to the results obtained in nutrient-rich 2xYT medium (Figure 45B). As shown in Figure 45A, 6xHis-Ubl2 was successfully overexpressed in 2xYT medium at both 18 °C and 25 °C following induction, indicating favorable expression conditions. SDS-PAGE analysis revealed a higher protein yield at 25 °C, which was therefore selected as the optimal temperature for subsequent large-scale expressions. In Figure 45B, His-Ubl2 was detected at approximately 10 kDa in samples from both 2xYT and minimal (M9) media, confirming successful expression of the target protein. However, expression in M9 medium resulted in a noticeably lower yield compared to 2xYT. Despite the reduced expression level, M9 medium was used for further protein production due to its suitability for isotopic labeling required for NMR spectroscopy.

6.5.2 Isolation of ^{1}H-^{15}N 6xHis-Ubl2 by IMAC

Following the confirmation of Ubl2 expression by Western blot analysis and the identification of optimal expression conditions, the next objective was to obtain large quantities of high-purity protein. To achieve this, IMAC was employed to efficiently isolate His-tagged Ubl2 from other histidine-rich *E. coli* proteins. Subsequent TEV protease digestion was carried out to remove the N-terminal His-tag, yielding tag-free Ubl2 for downstream applications.

After cell disruption by sonication and subsequent centrifugation, the clarified lysate was applied to a Ni-NTA Sepharose 6 Fast Flow column equilibrated with Ubl2 lysis buffer (50 mM sodium phosphate (pH 6.0), 250 mM NaCl). In the initial phase of the purification, non-His-tagged *E. coli* proteins were eluted, corresponding to high absorbance signals between 8 mL and 60 mL (Figure 46A). Consistent with this observation, multiple *E. coli* proteins were detected in the flow-through fraction by SDS-PAGE (Figure 46B). To remove nonspecifically bound proteins, the column was washed with IMAC elution buffer (50 mM sodium phosphate (pH 6.0), 250 mM NaCl, 500 mM imidazole) adjusted to 4% (v/v) imidazole, corresponding to a final concentration of 20 mM imidazole. Several *E. coli* proteins were still present during this washing step, as seen in the SDS-PAGE analysis of the 4%B fraction (Figure 46B). Elution of His-tagged Ubl2 was achieved using an imidazole gradient, with the target protein eluting at approximately 400 mM imidazole. This was reflected by a prominent absorbance peak between 108 mL and 120 mL, reaching a maximum at 115 mL (779 mAU) (Figure 46A). Two-

milliliter fractions were collected across the elution peak and analyzed by SDS-PAGE. Strong bands at around 10 kDa confirmed the presence of high amounts of His-Ubl2 (Figure 46B). Notably, fractions three to six contained the highest protein concentrations and were pooled for further processing. Since imidazole can interfere with protein stability and activity, the pooled sample was dialyzed overnight against dialysis buffer (50 mM Tris (pH 6.0), 250 mM NaCl, 2 mM DTT, 1 mM EDTA) to reduce the imidazole concentration. TEV protease was added simultaneously to cleave the N-terminal 6xHis-tag during dialysis.

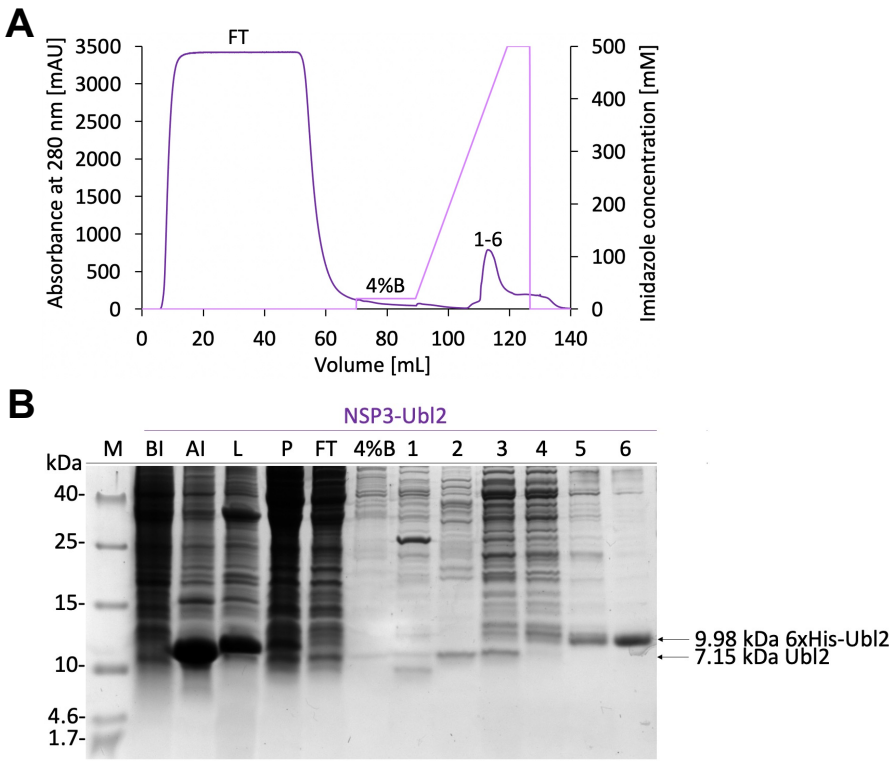


Figure 46: Isolation of His-tagged ^{1}H-^{15}N NSP3-Ubl2 by IMAC. The protein sample was loaded on a Ni-NTA column equilibrated in 50 mM sodium phosphate (pH 6.0), 250 mM NaCl. The corresponding chromatogram is shown in **A**. The absorbance (dark purple) measured at 280 nm was plotted against the volume in mL. Target protein was eluted using 50 mM sodium phosphate (pH 6.0), 250 mM NaCl, 500 mM imidazole. **(B)** SDS-samples were taken for a 20% (v/v) Tris-Tricine SDS-PAGE. Samples before induction (BI), after induction (AI), lysate (L), pellet (P), flowthrough (FT), 4% elution buffer (4%B) and fractions (1-6) were taken and compared to the marker (M) (Spectra Low Range Protein Ladder).

6.5.3 Isolation of ^{1}H-^{15}N Ubl2 by Reverse IMAC

To separate the cleaved (tag-free) Ubl2 from undigested 6xHis-Ubl2, free His-tags, and the His-tagged TEV protease, a second IMAC step was performed. As previously described, IMAC separates proteins based on their affinity for nickel ions via an N-terminal His-tag.[95] Due to the successful TEV protease cleavage, the digested (tag-free) Ubl2 was not expected to bind

to the column and was therefore collected in the flow-through. This purification strategy is commonly referred to as reverse IMAC.

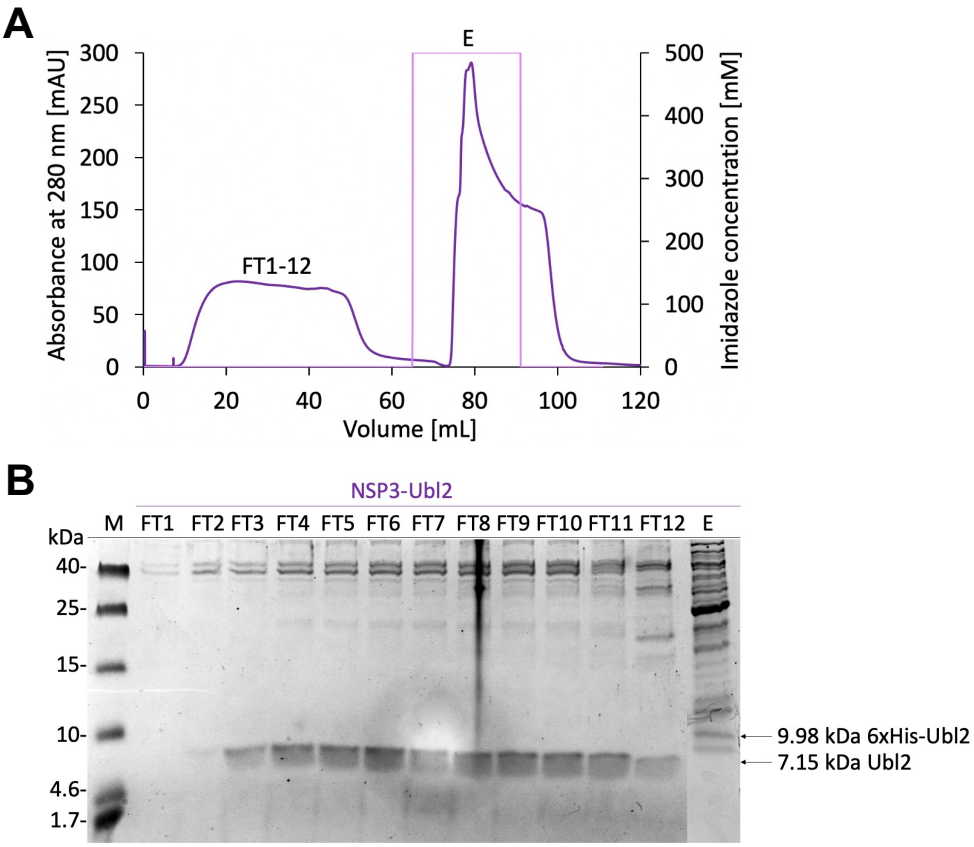


Figure 47: Isolation of ^{1}H-^{15}N NSP3-Ubl2 by reverse IMAC. The protein sample was loaded on a Ni-NTA column equilibrated in 50 mM sodium phosphate (pH 6.0), 250 mM NaCl. The corresponding chromatogram is shown in **A**. The absorbance (dark purple) measured at 280 nm was plotted against the volume in mL. Digested target protein was expected in the FT (7.15 kDa). His-tagged Ubl2 (9.98 kDa) and His-TEV-protease (25 kDa) were eluted using 50 mM sodium phosphate (pH 6.0), 250 mM NaCl, 500 mM imidazole. **(B)** SDS-samples were taken for a 20% (v/v) Tris-Tricine SDS-PAGE. Samples of the flowthrough (FT) and the eluate (E) were taken and compared to the marker (M) (Spectra Low Range Protein Ladder).

Following dialysis and TEV digestion, the sample was applied to a Ni-NTA Sepharose 6 Fast Flow column equilibrated with Ubl2 lysis buffer (50 mM sodium phosphate (pH 6.0), 250 mM NaCl). A broad elution peak was detected between 15 mL and 60 mL, with a maximum absorbance of approximately 80 mAU (Figure 47A). Flow-through fractions (4 mL each) were collected and analyzed by SDS-PAGE (Figure 47B). Ubl2 was identified in fractions 3 to 12, confirming efficient TEV protease cleavage. To elute non-digested His-tagged Ubl2, residual His-tags, and the His-tagged TEV protease, the imidazole concentration was raised to 500 mM (Figure 47A). A sample from the resulting elution peak was subjected to SDS-PAGE analysis (Figure 47B). A prominent band at approximately 25 kDa was observed, corresponding to the

His-tagged TEV protease (theoretical MW: 28,560 Da). A faint band at around 10 kDa was also visible, suggesting the presence of residual His-tagged Ubl2 and indicating incomplete digestion. To further assess the composition of the IMAC-purified sample, an anti-His Western blot was performed on selected fractions (Figure 48).

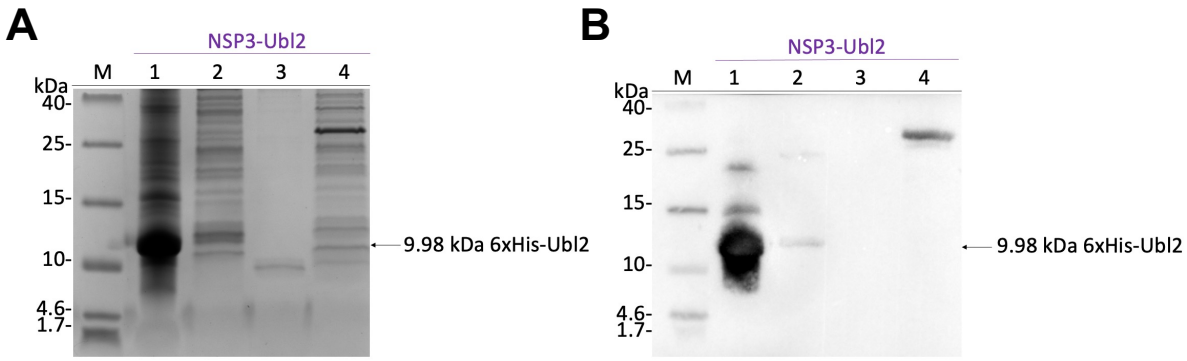


Figure 48: Sample analysis of ^{1}H-^{15}N 6xHis-Ubl2 by immunodetection. *E. coli* BL21 (DE3) cells incorporating the pET-28a(+)-6xHis-Ubl2 plasmid were cultivated in minimal medium at 22 °C overnight and Ubl2 protein was isolated. **(A)** Samples after induction of expression (1), IMAC fraction 3 (2), revIMAC FT fraction 3 (3) and of the IMAC Reverse eluate (4) were taken and analyzed by a 20% (v/v) Tris-Tricine SDS-PAGE. Protein bands were compared to the marker (M) (Spectra Low Range Protein Ladder). 6xHis-tagged Ubl2 was expected at 9.98 kDa. **(B)** The His-tagged Ubl2 was identified by Western Blot analysis at the proposed size.

In this experiment, SDS-PAGE and the corresponding Western blot analyses of Ubl2 samples following IMAC and reverse IMAC were performed to verify efficient protein isolation. As shown in fraction 1, a prominent overexpression band of Ubl2 at approximately 10 kDa was detected in the sample collected after induction (Figure 48). Fraction 2, representing an IMAC eluate with imidazole, contained both His-tagged Ubl2 and TEV protease. A faint Ubl2 band was observed in the flow-through fraction during reverse IMAC (fraction 3) on the SDS-PAGE, displaying low protein yields after digestion. However, no corresponding signal was detected in the Western blot, confirming effective digestion of Ubl2. The His-tagged TEV protease (28.56 kDa) was found exclusively in the reverse IMAC eluate, demonstrating successful separation of the protease from the cleaved protein (Figure 48B). In summary, Ubl2 was efficiently purified through the combined use of IMAC and reverse IMAC.

6.5.4 Isolation of ^{1}H-^{15}N Ubl2 by SEC

As a sufficient amount of Ubl2 was isolated, the next experimental step was to enhance the sample purity mediated by SEC. Sample purification was performed in combination with

buffer exchange to optimize later structural measurements of Ubl2. In general, SEC is used to separate proteins by their hydrodynamic radius. Therefore, no elution buffer was needed in this experiment. In a first step, the column was equilibrated with degassed and filtered 50 mM sodium phosphate (pH 6.0), 150 mM NaCl buffer. After equilibration, the protein was loaded on a HiLoad Superdex 16/600 75 pg column.

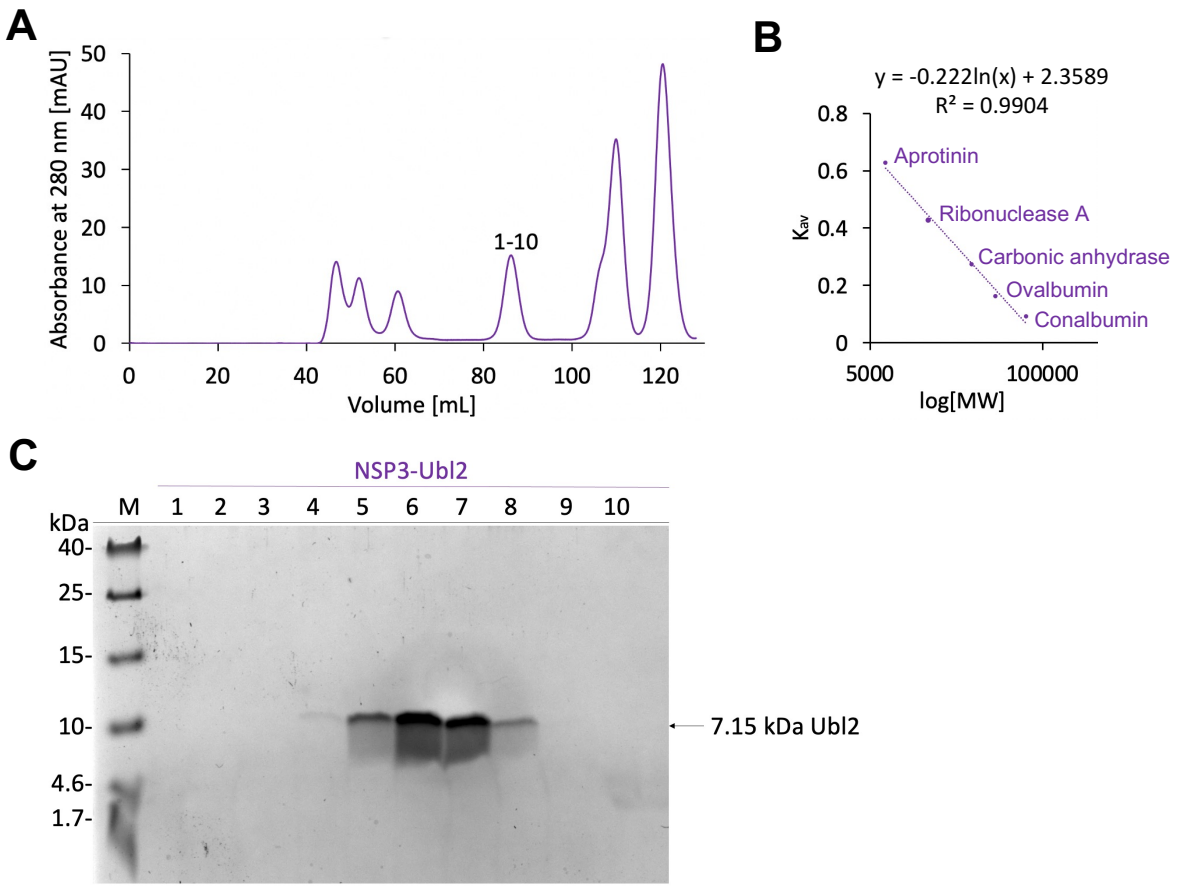


Figure 49: Size Exclusion Chromatography of ^{1}H-^{15}N NSP3-Ubl2. For SEC, a HighLoad 16/600 Superdex 75 pg column (Cytiva, Marlborough, Massachusetts, USA) equilibrated in 50 mM sodium phosphate (pH 6.0), 150 mM NaCl was used. **(A)** Chromatogram of the isolation. The absorbance (mAU) is plotted against the volume (mL). **(B)** SEC calibration curve based on the following standard proteins: Aprotinin (6.5 kDa), Ribonuclease A (13.7 kDa), Carbonic Anhydrase (29 kDa), Ovalbumin (44 kDa) and Conalbumin (75 kDa). The calibration curve was plotted using the gel-phase distribution coefficient K_{av} vs. the logarithm of the molecular weight (MW) in g/mol. Based on the calibration curve, the calculated molecular weight of Ubl2 is 3.46 kDa (7.15 kDa expected) indicating monomeric Ubl2. **(C)** 20% (v/v) Tris-Tricine SDS-PAGE of collected SEC fractions (1-10). Bands correlating to Ubl2 (7.15 kDa) are indicated by an arrow.

During the run, several elution peaks appeared in the SEC chromatogram (Figure 49A). Based on a calibration curve (Figure 49B) and the molecular weight of Ubl2, the target protein was expected to elute around 80 mL. As shown in Figure 49A, the chromatogram oft the SEC-run revealed a peak between 82 mL and 91 mL reaching 14.89 mAU. Fractions of the peak, 2 mL

each, were taken and analyzed via SDS-PAGE (Figure 49C). Ubl2 was identified in fraction 4 to 8, offering a low protein concentration in fraction 4. No further impurities were detected. However, as SDS causes the denaturation of proteins, the molecular weight of eluted protein was calculated based on a calibration curve (Figure 49B).

The calibration curve was plotted using the gel-phase distribution coefficient K_{av} versus the logarithm of the molecular weight (log [MW]) (Figure 49B). K_{av} was calculated using the column void volume (V_o = 45.62 mL), the geometric column volume (V_c = 120 mL), and the elution volume (V_e = 86.51 mL) of Ubl2. The following equation was used for calculation:

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

The SEC column was calibrated using conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa). The equation of the calibration curve based on the data from molecular weight standards was used to calculate the experimental molecular weights. Based on these data, a molecular weight of 3.46 kDa was calculated along with the proposed molecular weight of 7.15 kDa. These results indicate that only monomers were in the sample. As the purity of pooled fractions was high (Figure 49C) and only monomers were observed, the pooled and concentrated SEC samples were used for later NMR measurements.

6.5.5 Optimization of PLpro Expression in *E.coli* Cells

Expression tests of 6xHis-PLpro were performed at 18 °C and 25 °C after induction of expression to optimize the protein yield. For that, *E. coli* BL21 (DE3) cells containing the target construct, were cultivated in 2xYT medium at 37 °C overnight. Main cultures were inoculated in 2xYT medium and incubated at 18 °C and 25 °C overnight after induction of expression. Samples before (BI) and after induction of expression (AI) were taken (Figure 50A). Optimal expression conditions were identified by thick dark bands on the SDS-PAGE. Immunostaining was utilized to identify the target protein at the proposed size of 38.81 kDa (Figure 50B). His-tagged PLpro was overexpressed in 2xYT medium at 18 °C and 25 °C upon IPTG induction, with a slightly higher protein yield reached at 18°C (Figure 50A). Based on the Western Blot, a

higher protein yield was acquired during incubation at 25 °C (Figure 50B) as indicated by a slightly stronger band at 25 °C after induction (AI) compared to the band before induction (BI).

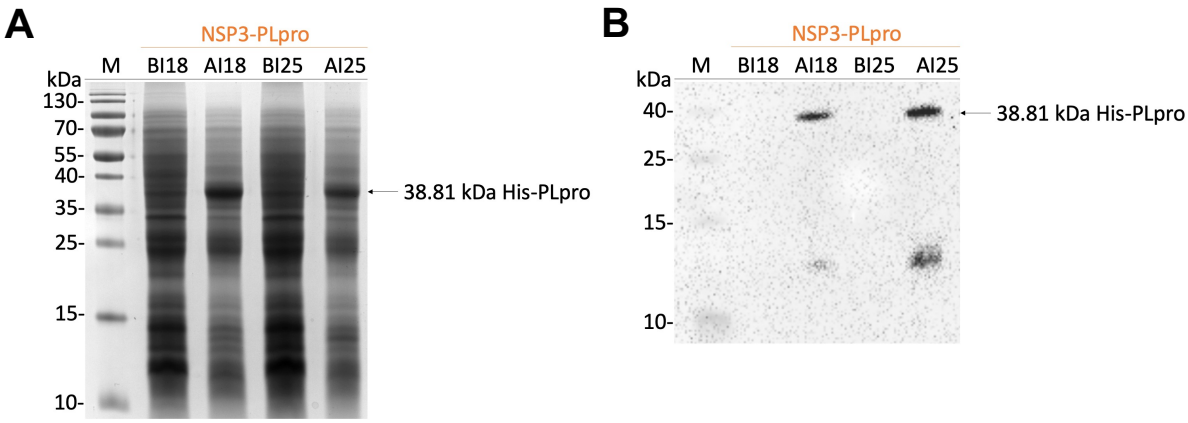


Figure 50: Expression tests of ^{1}H-^{15}N 6xHis-PLpro. (A) *E. coli* BL21 (DE3) cells incorporating the pET-28a(+)-6xHis-Thrombin-TEV-6xHis-PLpro plasmid were cultivated in 2xYT medium at 18 °C (AI18) or 25 °C (AI25) overnight. Samples before (BI) and after induction (AI) of expression were taken and analyzed by a 15% (v/v) Tris-Glycine SDS-PAGE. Protein bands were compared to the marker (M) (PageRuler Prestained Protein Ladder). 6xHis-tagged PLpro was expected at 38.81 kDa. **(B)** The His-tagged protein was identified by Western Blot analysis at the proposed size (38.81 kDa), previously using a 20% (v/v) Tris-Tricine SDS-PAGE (Spectra Low Range Protein Ladder).

Nevertheless, 18 °C post induction was used for further large-scale expressions, as lower temperatures reduce inclusion body formation and improve the solubility of recombinantly expressed proteins.[97]

6.5.6 Isolation of ^{1}H-^{15}N 6xHis-PLpro by IMAC

Following a 2-liter large-scale expression of ^{1}H-^{15}N 6xHis-PLpro, IMAC was carried out to separate 6xHis-PLpro from other histidine-rich *E. coli* proteins. Sonicated samples were loaded on a Ni-NTA sepharose 6 fast flow column pre-equilibrated with 20 mM Tris (pH 7.0), 300 mM NaCl, 1 mM TCEP, 1 µM $ZnSO_4$ (PLpro lysis buffer). PLpro is a metalloprotein that coordinates a zinc ion within its finger domain (Figure 43). Given that 0.35 mM zinc sulfate significantly impaired *E.coli* growth in M9 minimal medium,[236] zinc supplementation was limited to 100 µM in the expression medium and 1 µM in lysis buffer to balance metal incorporation and cell viability.

Before IMAC was performed, samples of the total cell extract (BI, AI) and cell lysate were analyzed by SDS-PAGE (Figure 51B). High protein yields were achieved during gene expression

at 18°C (AI). After sonication, samples of the lysate and pellet were analyzed, with the majority of protein detected in the pellet, suggesting inclusion body (IB) formation.

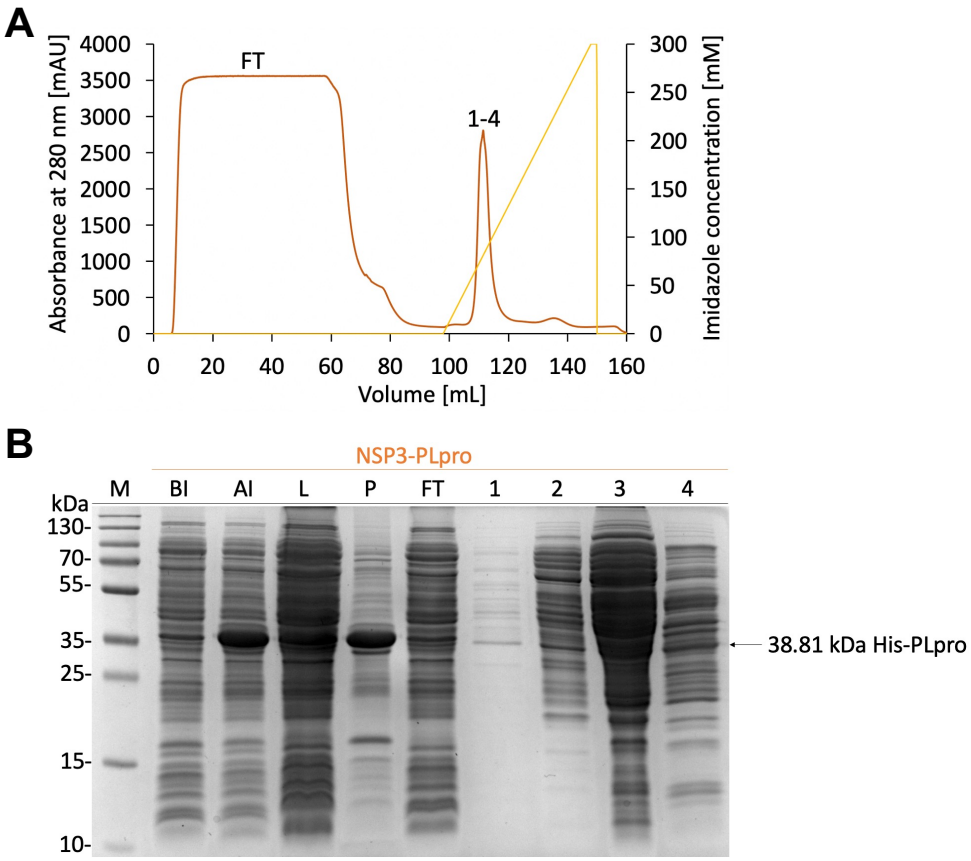


Figure 51: Isolation of His-tagged ^{1}H-^{15}N NSP3-PLpro by IMAC. Protein samples were loaded on a Ni-NTA column equilibrated in 20 mM Tris (pH 7.0), 300 mM NaCl, 1 mM TCEP, 1 µM $ZnSO_4$.The corresponding chromatogram is shown in **A**. The absorbance (dark orange) measured at 280 nm was plotted against the volume in mL. Target protein was eluted using 20 mM Tris (pH 7.0), 300 mM NaCl, 1 mM TCEP, 300 mM imidazole. **(B)** SDS-samples were taken for a 15% (v/v) Tris-Glycine SDS-PAGE. Samples before induction (BI), after induction (AI), lysate (L), pellet (P), flowthrough (FT) and fractions (1-4) were taken and compared to the marker (M) (Page Ruler Prestained protein Ladder).

As demonstrated by Figure 51A, *E. coli* proteins without His-tag were eluted, resulting in high absorbance levels between 6 and 85 mL. In line with these results, bands of several *E. coli* proteins were found in the flow through sample (Figure 51B). As PLpro already eluted during washing (4%B) in previous experiments (not shown), the washing step was skipped. His-tagged PLpro eluted at 210 mM imidazole as indicated by a peak appearing between 105 mL and 120 mL, with highest absorbance level at 111 mL (2767 mAU) (Figure 51A). Peak fractions (1 - 4), each 2 mL in volume, were collected and subjected to SDS-PAGE analysis. Most histidine-rich proteins including PLpro was found in fraction 3 (Figure 51B). Fraction 1 to 4 was pooled and used for subsequent experiments. Since imidazole can affect protein activity and stability, the

sample was dialyzed overnight in dialysis buffer (20 mM Tris (pH 7.0), 250 mM NaCl, 0.3 mM TCEP) to reduce the imidazole concentration. Dialysis was performed in parallel with TEV protease digestion to get rid of the N-terminal His-tag.

6.5.7 Isolation of ^{1}H-^{15}N PLpro by Reverse IMAC

To separate the digested target protein from undigested 6xHis-PLpro, cleaved His-tags, and His-tagged TEV protease, a second IMAC step was performed. As described earlier, IMAC separates proteins via interaction between polyhistidine tags and column-bound nickel ions.[95] Because the target protein lacks a His-tag after TEV-mediated cleavage, it was expected to appear in the flowthrough. This approach is therefore referred to as reverse IMAC.

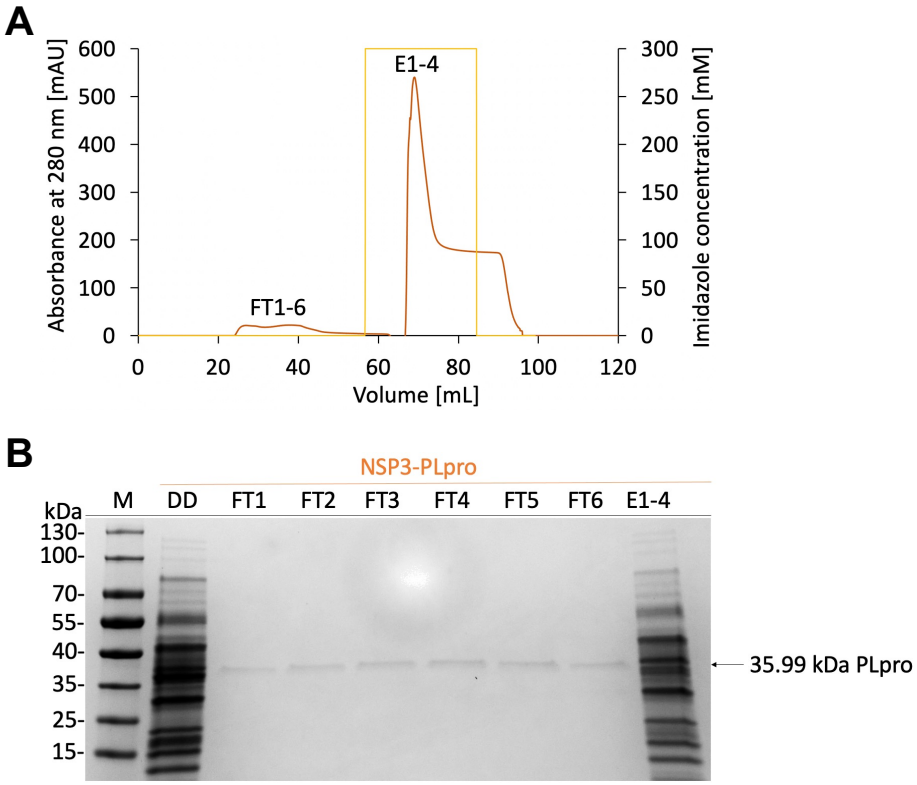


Figure 52: Isolation of ^{1}H-^{15}N NSP3-Ubl2 by reverse IMAC. Protein samples were loaded onto a Ni-NTA column pre-equilibrated with buffer containing 20 mM Tris (pH 7.0), 300 mM NaCl, 1 mM TCEP, and 1 µM $ZnSO_4$. **(A)** The corresponding chromatogram displays absorbance at 280 nm (dark orange) plotted against elution volume (mL). The target protein, lacking a His-tag after TEV cleavage, was collected in the flowthrough. Elution of His-tagged contaminants was achieved using 300 mM imidazole in the same buffer. **(B)** Samples were analyzed by 20% (v/v) Tris-Glycine SDS-PAGE. Lanes include before induction (BI), after induction (AI), lysate (L), pellet (P), flowthrough (FT), fractions 1-4, and molecular weight marker (M; PageRuler Prestained Protein Ladder).

After dialysis and TEV digestion, the sample was loaded onto a Ni-NTA Sepharose 6 Fast Flow column equilibrated with PLpro lysis buffer (20 mM Tris (pH 7.0), 300 mM NaCl, 1 mM TCEP,

and 1 µM ZnSO). Due to the cleavage by TEV protease, the target protein (PLpro) was expected to appear in the flowthrough. A broad peak was observed between 24 mL and 48 mL, reaching a maximum absorbance of 38 mAU (Figure 52A). Fractions of 4 mL were collected from the flowthrough and analyzed by SDS-PAGE (Figure 52B), confirming the presence of PLpro in each fraction and indicating efficient digestion. To elute residual His-tagged proteins, including undigested PLpro, the imidazole concentration was increased to 300 mM (Figure 52A). A large peak appeared at 70 mL, approximately 10-fold higher than the flowthrough signal. This was likely caused by a missed washing step during IMAC, resulting in the co-elution of unspecific bound *E. coli* proteins. This assumption was supported by SDS-PAGE analysis, which revealed multiple *E. coli*-derived protein bands in the eluate. For further confirmation and analysis of isolation steps, a western blot was performed (Figure 53).

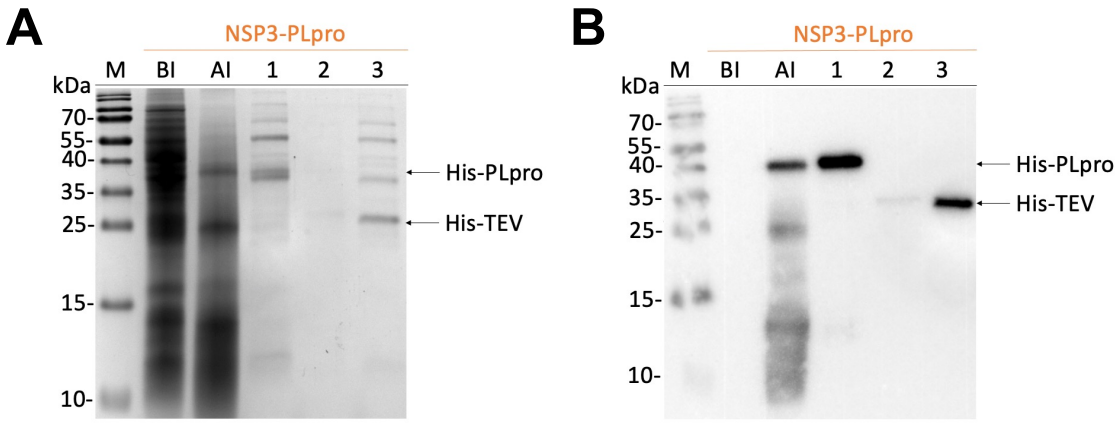


Figure 53: Immunodetection analysis of ^{1}H-^{15}N 6xHis-PLpro samples. (A) Samples taken before induction (BI), after induction (AI), IMAC fraction 3 (1), reverse IMAC flowthrough fraction 6 (2), and reverse IMAC eluate (4) were analyzed by 15% (v/v) Tris-Glycine SDS-PAGE. Protein bands were compared to the molecular weight marker (M; PageRuler Prestained Protein Ladder). The 6xHis-tagged PLpro is expected at approximately 38.8 kDa. **(B)** Western blot analysis confirmed the presence of 6xHis-PLpro at the expected molecular weight. Additionally, the His-tagged TEV protease was detected at around 28 kDa.

In this experiment, SDS-PAGE (Figure 53A) and the corresponding Western blot (Figure 53B) were performed to evaluate the isolation efficiency of PLpro after IMAC and reverse IMAC. His-tagged PLpro was detected in the sample after induction (AI) and in IMAC fraction 3 (1). A strong signal for the His-tagged TEV protease was observed in the reverse IMAC eluate (3), and a weaker signal was present in the reverse IMAC flowthrough (2), indicating incomplete separation of the protease from the target protein. Unfortunately, PLpro was only faintly visible in the reverse IMAC flowthrough by SDS-PAGE. In summary, PLpro was successfully

isolated through IMAC and reverse IMAC, although further optimization may be needed to improve separation from the TEV protease and the overall yield of PLpro.

6.5.8 Isolation of ^{1}H-^{15}N PLpro by SEC

As a sufficient amount of PLpro had been isolated, the next experimental step aimed to improve sample purity using SEC. In addition to purification, SEC was combined with buffer exchange to optimize conditions for subsequent structural analyses of PLpro by NMR spectroscopy. SEC separates proteins based on their hydrodynamic radius and does not require an elution buffer. Initially, the column was equilibrated with degassed and filtered SEC buffer (50 mM MES (pH 6.5), 100 mM NaCl, 1 mM TCEP). After equilibration, the protein sample was loaded onto a HiLoad Superdex 16/600 200 pg column.

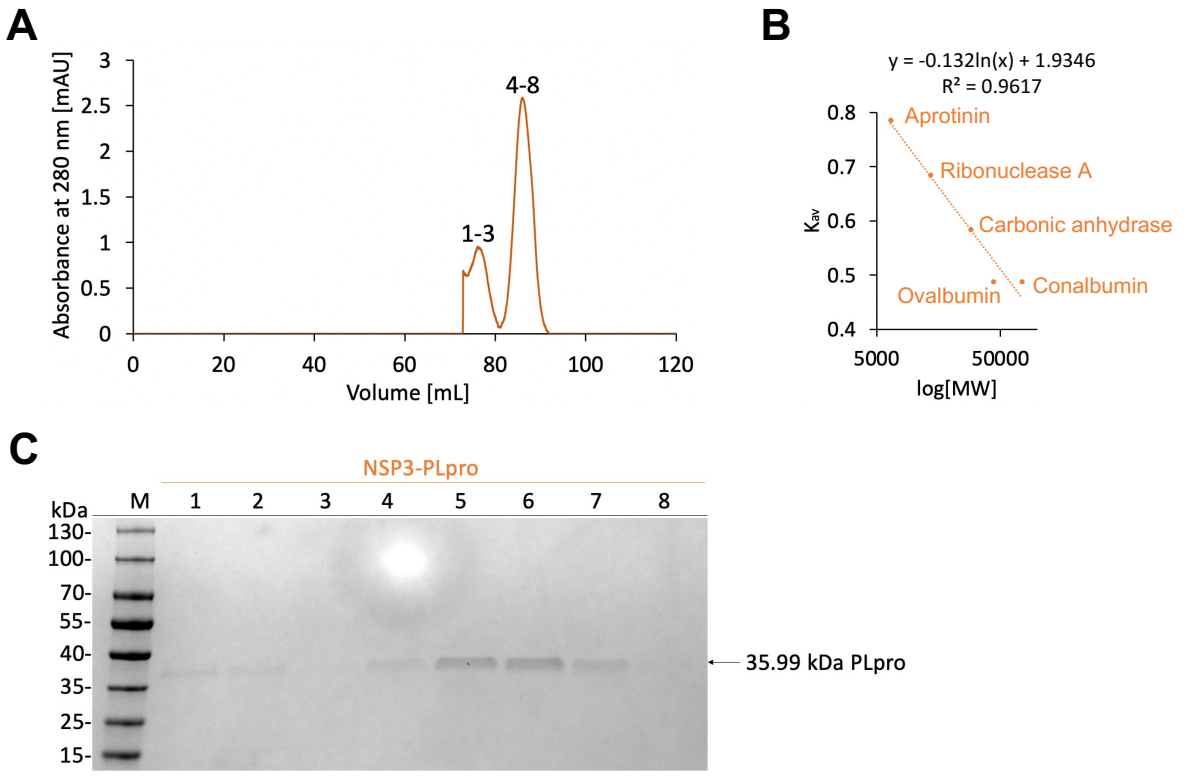


Figure 54: Size Exclusion Chromatography of ^{1}H-^{15}N NSP3-PLpro. For SEC, a HiLoad 16/600 Superdex 200 pg column (Cytiva, Marlborough, MA, USA) equilibrated with 50 mM MES (pH 6.5), 100 mM NaCl, and 1 mM TCEP was used. **(A)** Chromatogram of the SEC run. The absorbance (mAU) was plotted against elution volume (mL). **(B)** Calibration curve constructed using standard proteins: Aprotinin (6.5 kDa), Ribonuclease A (13.7 kDa), Carbonic Anhydrase (29 kDa), Ovalbumin (44 kDa), and Conalbumin (75 kDa). The gel-phase distribution coefficient (K_{av}) was plotted against the logarithm of the molecular weight (MW, g/mol). Based on the calibration curve, the molecular weight of PLpro was calculated as 30.85 kDa (theoretical MW: 35.99 kDa), consistent with a monomeric state of PLpro. **(C)** 15% (v/v) Tris-Glycine SDS-PAGE of collected SEC fractions (1-10). Bands corresponding to PLpro (35.99 kDa) are indicated by an arrow.

During the SEC run, two distinct elution peaks were observed in the chromatogram (Figure 54A). Based on the calibration curve (Figure 54B) and the known molecular weight of PLpro, the protein was expected to elute at approximately 85 mL. Consistent with this, a prominent peak with a maximum at 85.93 mL was detected, corresponding to the monomeric form of PLpro. In addition to this, a second, earlier peak appeared at 76.50 mL. Fractions from both peaks (2 mL each) were collected and analyzed by SDS-PAGE (Figure 54C). PLpro was identified in each fraction, while no additional impurities were detected. The presence of PLpro in both peaks suggests the occurrence of protein dimerization, with the earlier peak representing the dimeric form. To estimate the molecular weights of the eluted proteins, the gel-phase distribution coefficient K_{av} was calculated using the following equation:

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

The column was calibrated using a set of molecular weight standards: conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa), and aprotinin (6.5 kDa). Using the linear regression equation obtained from the calibration curve (Figure 19B), the molecular weights were calculated as 76.01 kDa for peak 1, indicating dimerization and 30.86 kDa for peak 2, what is in close agreement with the expected monomeric PLpro (35.99 kDa). The slightly lower experimental value for the monomer may be due to column calibration variability or protein conformation affecting elution behavior. PLpro was found to exist in both monomeric and dimeric forms under the conditions used. To ensure sample homogeneity for downstream NMR analysis, only fractions corresponding to Peak 2 (monomer) were pooled and concentrated. The presence of a dimeric peak indicates potential oligomerization via cysteine residues (see discussion, Figure 60), which could be further investigated through complementary biophysical methods.

6.5.9 Structural Analysis of Ubl2 and PLpro

6.5.9.1 CD Spectroscopy of Ubl2

CD experiments were performed across a temperature range from 5 °C to 37 °C (Figure 55) to get insights into the secondary structure and folding of Ubl2. Different secondary structural elements generally show characteristic minima and maxima in the CD spectra. The positive signal maximum near 195 nm and the more pronounced minima near 208 nm and 221 nm indicate α-helical structure.[237] β-sheet formation is described by a negative minimum near

195 nm and a weaker positive maximum near 218 nm, overlapping with the bands of helical structures.[237,238,239] The secondary structure content was estimated using the web server BeStSel (Figure 55B).[235] The total antiparallel β-sheet content varied between 38.0% (20 °C) and 38.3% (37 °C). β-turn formation fluctuates between 15.1% (37 °C) and 15.6% (37 °C). Further secondary structure elements, including loop regions, β-bridges and random coil structures, range from 46.3% (15 °C) up to 46.8% (20 °C). Interestingly, no α-helical content was detected, despite structural predictions indicating its presence in Ubl2 based on the crystal structure (PDB: 7THH, PDB: 7CJD).

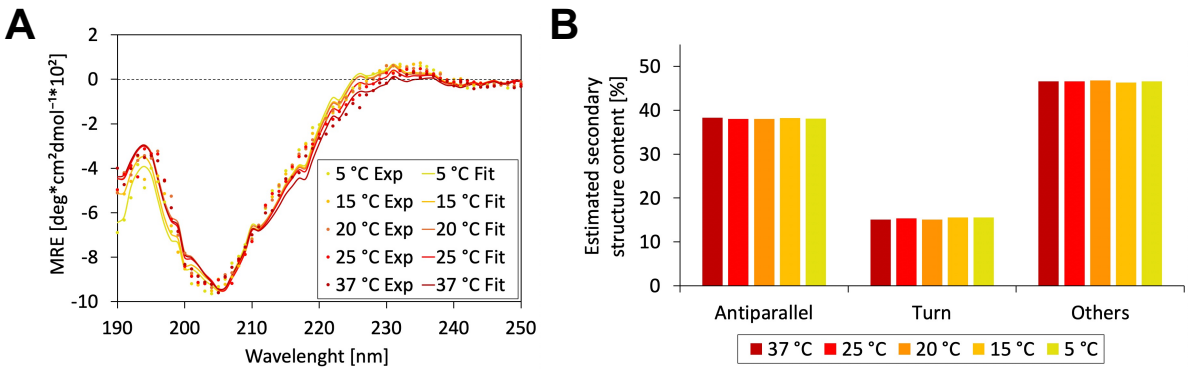


Figure 55: CD spectrum of Ubl2. (A) Far-UV CD spectrum of NSP3-Ubl2 (18.00 µM) in 50 mM sodium phosphate (pH 6.0) buffer. Data were recorded at different temperatures (5 °C, 15 °C, 20 °C, 25 °C and 37 °C). The mean residue ellipticity (MRE) is plotted against the wavelength in nm. Analysis of the data (dots, experimental data (Exp); lines, fitted data (Fit) between 190 and 250 nm using BeStSel[235]) indicates the presence of a structured and folded Ubl1 domain, as evident from a maximum at 193 nm. The presence of a double minimum at 208 nm and 221 nm reveals the retention of much secondary structure in Ubl2. **(B)** Estimated secondary structure content (%) based on the fitted data.

To summarize, CD data indicated that the secondary structure content of Ubl2 stays similar and stable between 5 °C and 37 °C (Figure 55). However, CD could not give residue-specific information. Hence, NMR spectroscopy was used for further investigations.

6.5.9.2 NMR Spectroscopy of Ubl2 - Assignment

The structure of Ubl2 was investigated by 3D NMR experiments. All NMR experiments were conducted at 25 °C using 170 µM Ubl2 in 50 mM sodium phosphate (pH 6.0), 150 mM NaCl buffer. Backbone resonance assignment was performed using a series of 3D NMR experiments for backbone correlation, as detailed in the Materials and Methods section. Due to the absence of detectable signals for the N-terminal residues S1 and L2, as well as for P36, P49, and P62, a total of 58 resonances was expected. Of these, 57 out of 58 detectable ^{1}H-^{15}N

resonances were successfully assigned, with the exception of M26. The assignment of N63 was guided by its anticipated C_α and C_β chemical shifts, consistent with its position at the C-terminus following a proline (P62).[240] The assigned ^{1}H-^{15}N resonances are presented in the ^{1}H-^{15}N HSQC spectrum shown in Figure 56. As expected for a folded, globular protein domain, backbone amide signals are distributed between 7 and 10 ppm. Side-chain resonances appear in the range of 6.5 to 7.5 ppm.

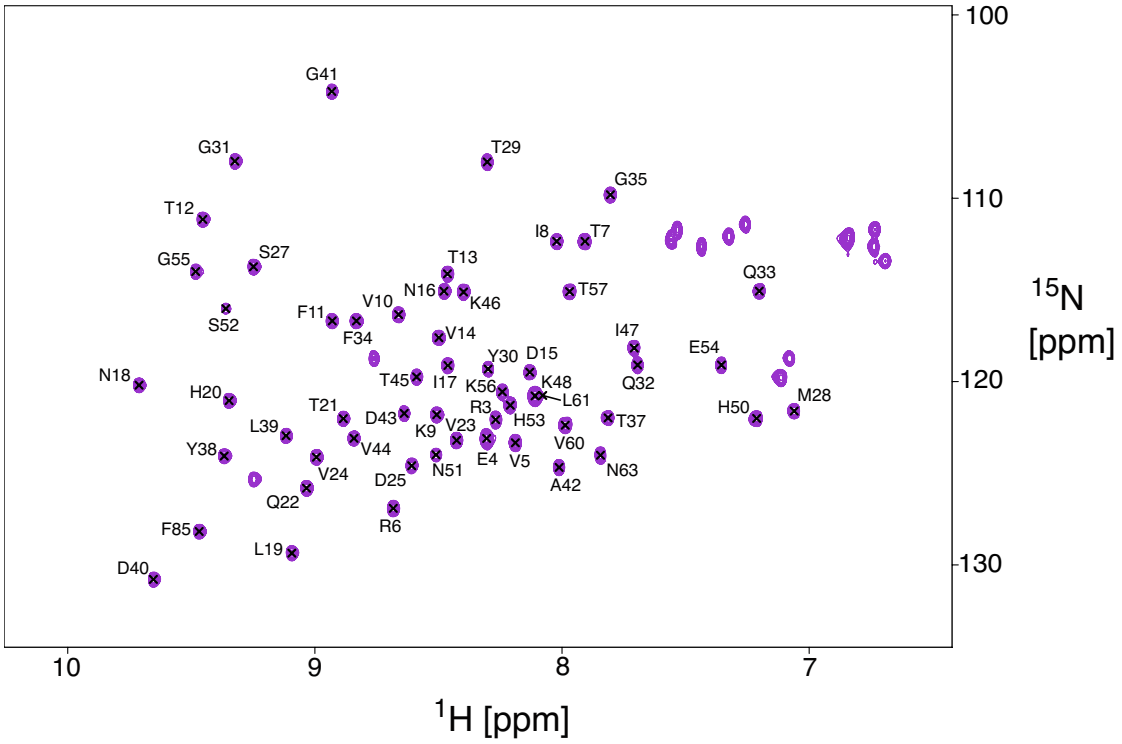


Figure 56: NMR spectral backbone assignment of SARS-CoV-2 ^{15}N-^{13}C NSP3-Ubl2 (1-63). The spectrum was acquired on a 700 MHz spectrometer at 25 °C and a concentration of 170 µM in 50 mM sodium phosphate (pH 6.0), 150 mM NaCl buffer. The sample was referenced to DSS and supplemented with 5% (v/v) D_2O and 0.01% (v/v) NaN_3. Assigned backbone amid groups are labelled. In total, 57 of 58 visible resonances were assigned, with S1, L2, and M26 missing.

6.5.9.3 Secondary Chemical Shift Perturbations of Ubl2

To get more insights into the secondary structure of Ubl2, 3D NMR experiments were used to calculate the secondary chemical shift perturbations (Figure 57C) and compare our results with literature. From the C_α and C_β chemical shifts (obtained from the NMR backbone assignment), residue-specific C_α and C_β secondary chemical shifts were calculated along the sequence, which indicate the respective secondary structure propensity. The regions R6 to V14, Q22 to M28, T37 to L39, I47 to H50 and Y59 comprise β-sheet secondary structure with negative C_α minus C_β secondary chemical shifts ($\Delta\delta C_\alpha$ - $\Delta\delta C_\beta$) up to -7.06 ppm. In total, the

Ubl2 domain comprises 5 β-strands assembled to 2 β-sheets, located in the same regions as displayed by the recently published X-ray structures of Ubl2 (PDB code 7CJD and 7THH, Figure 57A, B). The regions Y30 to Q32 and N51 to G55 are marked by two short α-helices with positive $\Delta\delta C_\alpha$ - $\Delta\delta C_\beta$ secondary chemical shifts up to 8.98 ppm (Figure 57B). Based on these NMR data, 12.7% (8 of 63 residues) of the total protein is composed of helices, whereas most of the domain comprises β-strands (38.1%; 24 of 63 residues). The sequence-specific secondary structure content of recorded data and crystal structures was compared in Figure 57D.

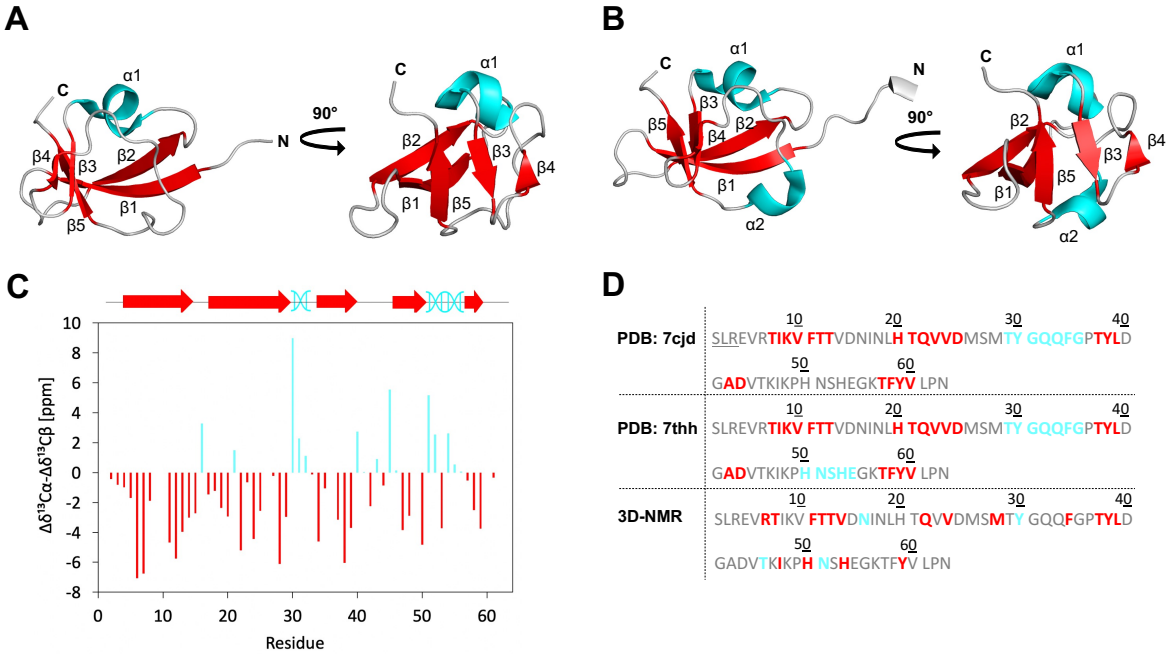


Figure 57: Secondary structure analysis of NSP3-Ubl2 by 3D NMR spectroscopy. (A) Crystal structure of the Ubl2 domain (4-63) of SARS-CoV-2 (PDB: 7CJD) comprising 1 helix and 5 β-strands (β1-β2-α1-β3-β4-β5). Notably, the first 3 amino acids are not shown by the crystal structure (residues S1, L2, R3). **(B)** Crystal structure of the Ubl2 domain (1-63) of SARS-CoV-2 (PDB: 7THH) comprising 2 helices and 5 β-strands (β1-β2-α1-β3-β4-α2-β5). **(C)** Secondary structure propensity of ^{15}N-^{13}C-labelled Ubl2 derived from backbone chemical shifts and on the basis of the current assignment (Figure 56). Values > +2 ppm are equivalent to 100% helical propensity, values < -2 ppm equivalent to 100% β-sheet propensity. Values between -1 ppm and +1 ppm report random coil characters. Structural elements are represented above the plot based on our data. Data were recorded on a 700 MHz spectrometer at 25 °C and a concentration of 170 µM in 50 mM sodium phosphate (pH 6.0), 150 mM NaCl. **(D)** Amino acid sequence of Ubl2 (PDB: 7CJD, 7THH; 3D NMR results) displayed by the one-letter amino acid code. Residues corresponding to secondary structure regions are highlighted in red for β-strands and cyan for α-helices.

6.5.10 Structural Analysis of PLpro

6.5.10.1 2D NMR Spectroscopy – TROSY Spectrum of ^{1}H-^{15}N PLpro

To gain deeper insights into the structural composition of PLpro, Transverse Relaxation-Optimized Spectroscopy (TROSY) was performed. This method enables the study of larger macromolecules, such as PLpro, which has a molecular weight of 35.99 kDa.[241,242] Pervushin and Wüthrich developed the TROSY technique capable of recording sharp signals from molecular complexes with sizes up to 200 kDa.[241] In combination with high-field NMR spectroscopy, a high-resolution TROSY spectrum of PLpro was obtained (Figure 58), serving as a molecular fingerprint for subsequent interaction studies.[243] Notably, progressive changes in the NMR spectrum over time suggested that the protein sample was only marginally stable under the given conditions (max. 48 h).

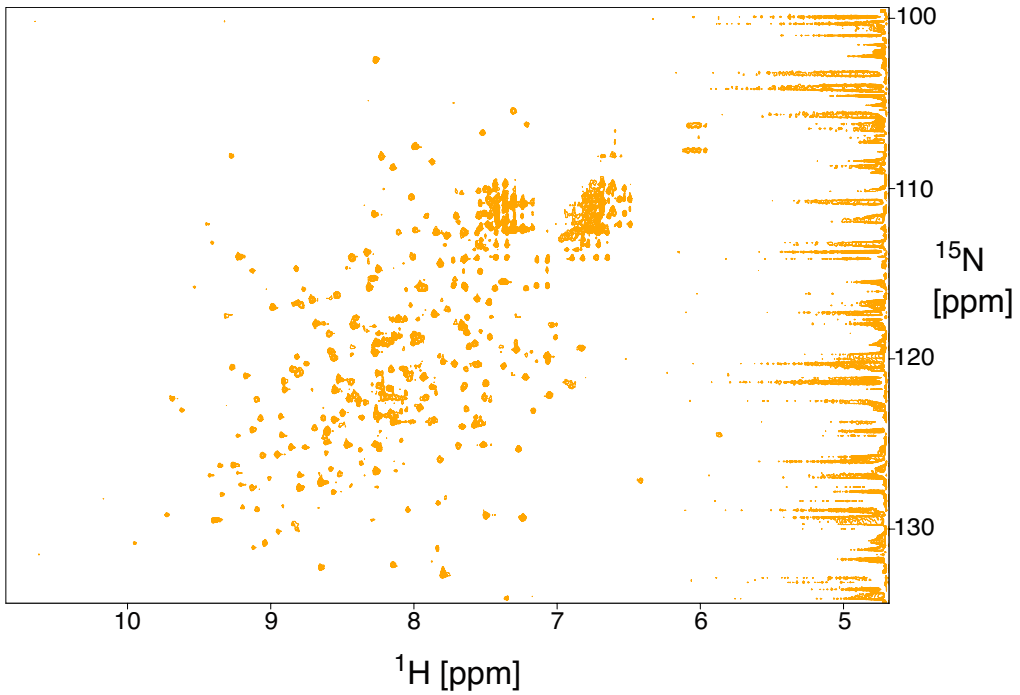


Figure 58: ^{1}H-^{15}N TROSY spectrum of ^{15}N PLpro. The spectrum was acquired on a 1.2 GHz spectrometer at 25 °C and a concentration of 75 µM in 50 mM MES (pH 6.0), 100 mM NaCl, 1 mM TCEP buffer. Sample was referenced to DSS and supplemented with 5% (v/v) D_2O and 0.01% (v/v) NaN_3.

6.6 Discussion

Over the past decades, PLpro including its Ubl2 subdomain, has been identified as a key component in the replication cycle of SARS-CoV-2, mediating the proteolytic release of Nsp1/2, Nsp2/3 and NSP3/4.[209] These cleavage events are critical for the formation of DMVs, which serve as ROs in infected cells.[46,45,54,63] Consequently, PLpro has emerged as a promising antiviral target to disrupt SARS-CoV-2 replication.[234] However, the development of effective inhibitors requires detailed structural information on PLpro, which remains limited to date. In this study, high-purity protein samples of PLpro and its Ubl2 subdomain were successfully produced and structurally characterized using NMR and CD spectroscopy to gain deeper insights into their folding, structural integrity, and potential functional roles within the replication cycle of SARS-CoV-2.

6.6.1 Expression and Isolation of Ubl2

Our initial approach focused on optimizing the expression and purification of the Ubl2 subdomain, corresponding to the N-terminal region of PLpro. Notably, to date, no dedicated protocols have been reported for the expression and isolation of Ubl2 alone; existing procedures primarily address the purification of full-length PLpro. Given the structural and folding similarities between Ubl1 and Ubl2, it was reasonable to assume that comparable conditions would be suitable for the expression and purification of Ubl2. Therefore, our strategy was adapted from a recently published protocol by Salvi *et al.* describing the expression and purification of NSP3-Ubl1. This protocol was further optimized in our laboratory to enhance protein yield.

As expected, and consistent with findings previously reported for Ubl1, optimal expression of Ubl2 was achieved at 25 °C in 2xYT medium using *E. coli* BL21 (DE3) cells (Figure 45). However, M9 medium was required for isotopic labeling and was therefore employed for subsequent large-scale protein production. Ubl2 was successfully expressed and remained soluble in the cell lysate following sonication (Figure 46B), indicating the absence of inclusion body formation or protein aggregation. Moreover, the target protein was not lost during the washing step with 4% (v/v) elution buffer (Figure 46), and high absorbance levels confirmed efficient recovery of His-tagged Ubl2 in the IMAC eluate. TEV protease digestion was confirmed to be successful based on immunodetection analysis: His-tagged Ubl2 was not

detected in either the reverse IMAC flow-through or the eluate, while the His-tagged TEV protease was efficiently separated from the target protein (Figure 47 and 48). SEC resulted in a sharp and well-defined elution peak, indicating effective separation of Ubl2 from remaining *E. coli* proteins (Figure 49). The final protein yields were sufficient for downstream biophysical analyses, including CD and NMR spectroscopy.

6.6.2 Structural Analysis of Ubl2

In this study, the secondary structure of Ubl2 was characterized using CD and NMR spectroscopy. As anticipated, Ubl2 adopts a globular and folded conformation, consistent with the ubiquitin-like fold, which features a β-grasp motif flanked by two α-helices.[244,245] NMR analysis identified five β-strands, in agreement with previously reported structures (PDB: 7CJD, 8THH). CD spectroscopy confirmed the overall folded nature of the protein and revealed a substantial content of antiparallel β-sheet structures (CD: 38.0%; NMR: 38.1%) (Figure 55B).

In contrast, analysis of α-helical content revealed discrepancies between the two methods. CD spectroscopy did not detect any helical content under the conditions used (Figure 55B), whereas NMR secondary chemical shift analysis indicated a helical content of 14.3% (Figure 57). These findings are consistent with crystallographic data, which report a short α-helix spanning residues T29 to G35 (PDB: 7CJD). NMR data further revealed two α-helical regions: one centered around Y30, aligning with the crystal structure, and a second near residue N51, located within the extended region between β-strands 4 and 5. The latter is not reported in PDB entry 7CJD but corresponds to a short helix (H50-F54) identified in an independent crystal structure by Osipiuk *et al.* (PDB: 7THH; Figure 17A, C).[23] These observations suggest that the formation of a second α-helix near N51 is structurally plausible and reflects conformational flexibility within this region of Ubl2, as commonly observed in ubiquitin-like proteins.[244,245]

Shimada *et al.* previously reported the NMR backbone assignment of SARS-CoV-2 NSP3-PLpro in which several residues corresponding to the Ubl2 domain were unassigned (Figure 59).[234] In total, 25 out of 63 residues corresponding to Ubl2 were assigned.[234] In this work, almost the whole Ubl2 was assigned except of the first two N-terminal residues (S1 and L2), the intermediate amino acid M26 and the three proline residues P36, P49 and P61. Notably, proline residues are not detectable by NMR spectroscopy because they lack the amide proton. Comparison of the current dataset with the peak list published by Shimada *et al.* revealed

substantial overlap in resonance positions (Figure 59C). However, slight deviations in chemical shifts were observed, which can be attributed to differences in experimental conditions, including buffer composition and magnetic field strength: Shimada *et al.* employed 20 mM sodium phosphate buffer (pH 7.0), 1 mM TCEP at 600 MHz, while in this study, 50 mM sodium phosphate buffer (pH 6.0), supplemented with 150 mM NaCl, was used at a field strength of 700 MHz. Moreover, this study successfully assigned additional 30 residues not previously reported by Shimada *et al.,* including: R3, E4, T12, T13, V14, D15, N16, I17, N18, L19, H20, T21, Q22, V23, G35, T37, Y38, L39, D40, G41, A42, D43, V44, T45, N51, S52, Y59, V60 and L61. For a detailed overview of the backbone resonance assignments, refer to Table SD4.

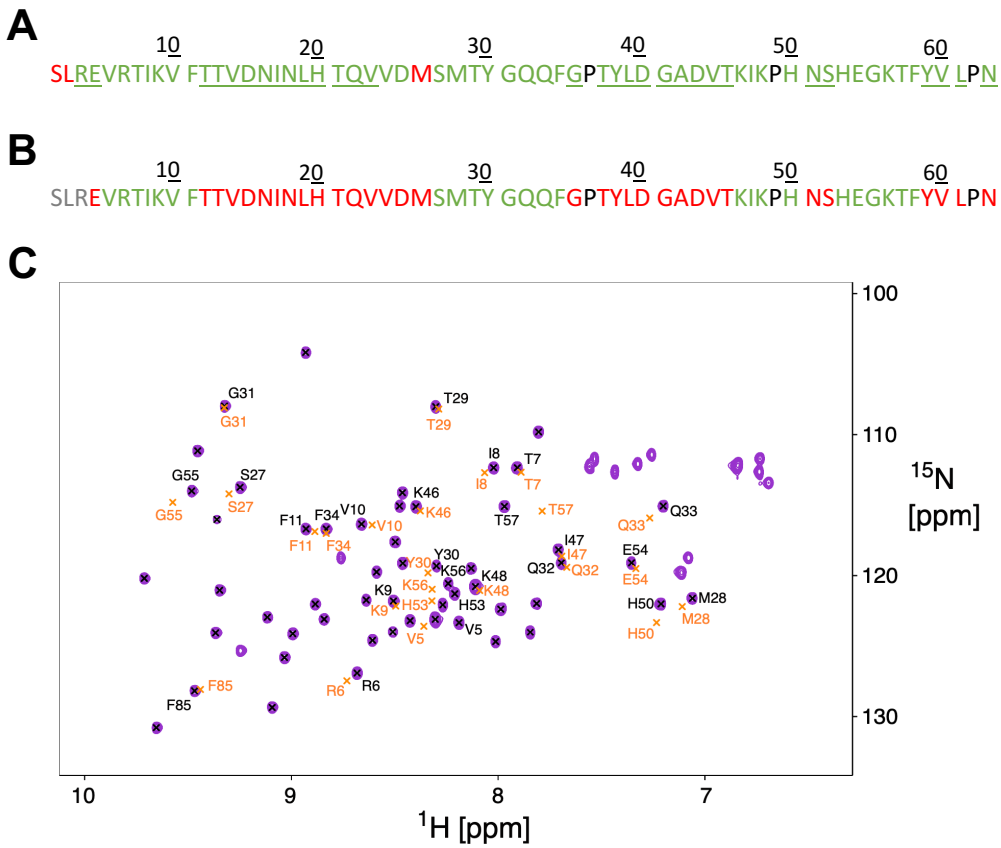


Figure 59: Ubl2 assignment comparison with literature. (A) Assigned amino acids of this work and **(B)** from Shimada *et al*. Notably, the first 3 amino acids are missing (greyish). Amino acid sequence of Ubl2 shown in the one-letter amino acid code. Assigned residues are highlighted in green. Additionally assigned residues by this work are greenish underlined. Non-assigned residues are annotated with red letters. Proline residues are labelled in black. **(C)** Overlay of Ubl2 assignments (orange: Shimada *et al.*, BMRB accession code 51992; purple: this work).

6.6.3 Expression and Isolation of PLpro

For the expression of PLpro in *E. coli*, post-induction temperatures of 18 °C and 25 °C were found to support optimal protein expression levels (Figure 50). However, 18 °C was selected

as the expression temperature to minimize inclusion body (IB) formation and protein aggregation.[97,98] At lower temperatures, the rate of protein biosynthesis is reduced, which slows the accumulation of nascent polypeptides and facilitates proper folding.[99] At low temperatures, translation and protein folding occur more slowly, including folding assisted by chaperones or disulfide bond formation. Since PLpro contains 11 cysteines, high temperatures may cause folding to happen too quickly, disrupting proper disulfide bridge formation. Moreover, it has been shown that the enzymatic activity of proteins sequestered in IBs is further diminished at elevated temperatures.[246] However, in this work, the enzymatic activity of PLpro was not tested and has to be confirmed previously. Furthermore, lower temperatures suppress the activity of endogenous proteases, thereby reducing proteolytic degradation of the target protein.[99] To preserve the catalytic activity of PLpro for downstream applications, expression was therefore conducted at the lowest effective temperature (18 °C).

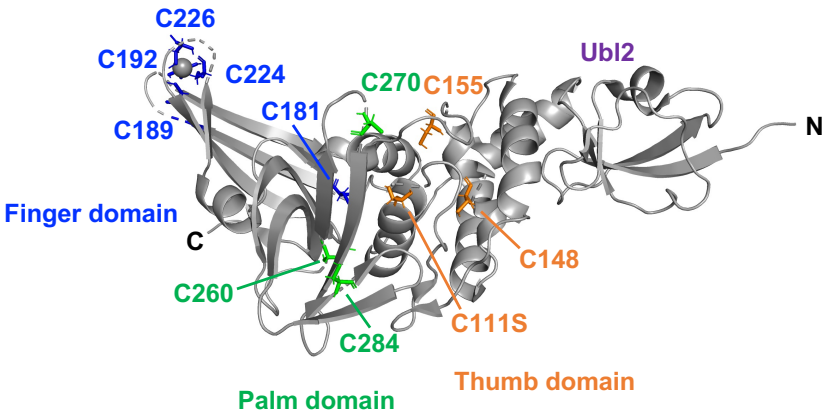


Figure 60: Crystal structure of PLpro highlighting its cysteine residues. PLpro contains 11 cysteine residues: C111, C148, and C155 are located in the thumb domain; C181, C189, C192, C224, and C226 in the finger domain; and C260, C270, and C284 in the palm domain. The Ubl2 subdomain is devoid of cysteine residues. Shimada *et al.* reported a C111S mutant to overcome protein precipitation (PDB:7CJD).[234]

Despite the use of a low expression temperature, IB formation was still observed upon cell lysis, as indicated by a substantial accumulation of PLpro in the insoluble fraction (cell pellet) following sonication (Figure 51B). One possible contributing factor may be the use of 2xYT medium during preculture growth, as demonstrated by the master thesis of Frederike Nith, which could lead to accelerated cell growth and increased protein synthesis rates, promoting aggregation and misfolding under overexpression conditions. As shown in the thesis by Frederike, minimal medium used for precultures reduced cell growth and enhanced the

soluble amount of PLpro in the lysate, while reducing the ribosomal activity and translation rate.

Another potential explanation for the observed aggregation is the use of *E. coli* as an expression system. Despite many advantages of using *E.coli*, it is well established that disulfide bond formation is inefficient in the cytoplasm of these cells due to its highly reducing environment.[247,248] The absence of proper disulfide bond formation (cysteinylation) can lead to misfolding of cysteine-rich proteins, ultimately resulting in aggregation, IB formation and loss of functional protein.[249] Given that PLpro contains 11 cysteine residues - corresponding to five disulfide bridges - correct folding is likely dependent on efficient disulfide bond formation. Shimada *et al.* observed that wild-type PLpro tends to precipitate and introduced a C111S mutation to improve its solubility.[234] This residue is located within an α-helix of the thumb domain (Figure 60) and is oriented toward the solvent. Due to its surface exposure, C111 has been proposed to contribute to oligomerization, potentially driving protein precipitation.

Alternatively, the periplasmatic space of *E.coli* is suitable to overcome the limitations of disulfide bond formation. The periplasm offers a more oxidizing environment, enriched with disulfide oxidoreductases such as DsbA and DsbC, which facilitate the formation and isomerization of disulfide bonds.[250] However, periplasmic expression often results in reduced overall protein yield, as a considerable portion of the recombinant protein may remain trapped in the cytoplasm and high protein concentrations are required for structural analysis.[251] An alternative strategy to overcome the limitations by *E.coli* is to use another expression system such as insect cells or the yeast *Pichia pastoris*, suitable for producing properly folded and enzymatic active PLpro.[184]

Following cell lysis, IMAC yielded a prominent elution peak, indicating high levels of His-tagged PLpro expression (Figure 51A). However, SDS-PAGE analysis of the IMAC eluate revealed the presence of several *E. coli* contaminants (Figure 51B), likely due to the omission of a washing step during purification. This step was deliberately excluded in subsequent preparations, as preliminary tests (data not shown; see Master thesis by Frederike Nith) showed that washing led to premature elution of the target protein, resulting in significant losses. Compared to the absorbance level of the IMAC eluate (Figure 52A), low absorbance levels of PLpro were

observed in the IMAC flow-through supporting the observation of high *E.coli* impurities in the IMAC eluate (Figure 51). A pronounced elution peak was also observed during reverse IMAC, likely attributed to the same residual contaminants and the absence of prior washing. Nevertheless, western blot analysis confirmed that soluble His-tagged PLpro was effectively isolated via IMAC and reverse IMAC, as evidenced by its detection in the IMAC eluate and successful separation from most *E.coli* proteins (Figure 53B). After TEV protease cleavage and subsequent reverse IMAC, PLpro was no longer detectable in the flow-through, confirming efficient tag removal (Figure 53B). This was further supported by the presence of a strong band corresponding to the His-tagged TEV protease in the reverse IMAC eluate, and the absence of His-tagged PLpro, indicating successful cleavage and purification.

Prior to SEC, the sample was concentrated to 3 mL using an Amicon centrifugal filter (cut-off 10 kDa). However, significant protein loss occurred due to non-specific adsorption of PLpro to the membrane (not shown), resulting in low absorbance during SEC (Figure 54). To address this, optimizations from the Master thesis by Frederike Nith were implemented. These included reducing the membrane surface area by using smaller concentrators (2 mL instead of 50 mL) and lowering the centrifugation speed (2,500 rpm instead of 6,000 x *g*), which substantially reduced protein loss and improved yields post-SEC. Additional optimizations involved shortening sonication time from 20 min to 6 min, which minimized mechanical stress and protein aggregation.

Moreover, IB formation was reduced by switching to minimal medium for preculture growth. These collective improvements led to the successful production of a high-concentration PLpro sample suitable for high-resolution NMR analysis. Furthermore, the addition of glycerol, as performed by Shimada *et al*.,[234] may help stabilize the protein in solution. However, since glycerol increases the viscosity of the buffer, it affects the rotational tumbling time of the protein. This leads to line broadening, reduced peak intensity, and increased peak overlap, which complicates peak differentiation and assignment. Additionally, glycerol may alter the dynamic properties of PLpro, potentially affecting its binding interactions. Hence, glycerol offers only limited benefit for NMR measurements.

6.6.4 Structural Analysis of PLpro

The two-dimensional TROSY-based ^{1}H-^{15}N HSQC NMR spectrum of PLpro displayed a dispersed pattern of cross-peaks across both the proton (^{1}H) and nitrogen (^{15}N) chemical shift dimensions, characteristic of a folded and structurally well-ordered protein (Figure 58). The combination of ultra-high field strength (1.2 GHz) and TROSY acquisition significantly minimized line broadening effects due to the effective transverse relaxation, facilitating the observation of numerous resonances despite the relatively large molecular weight of PLpro, with around 35 kDa in combination with protonation.[241,243]

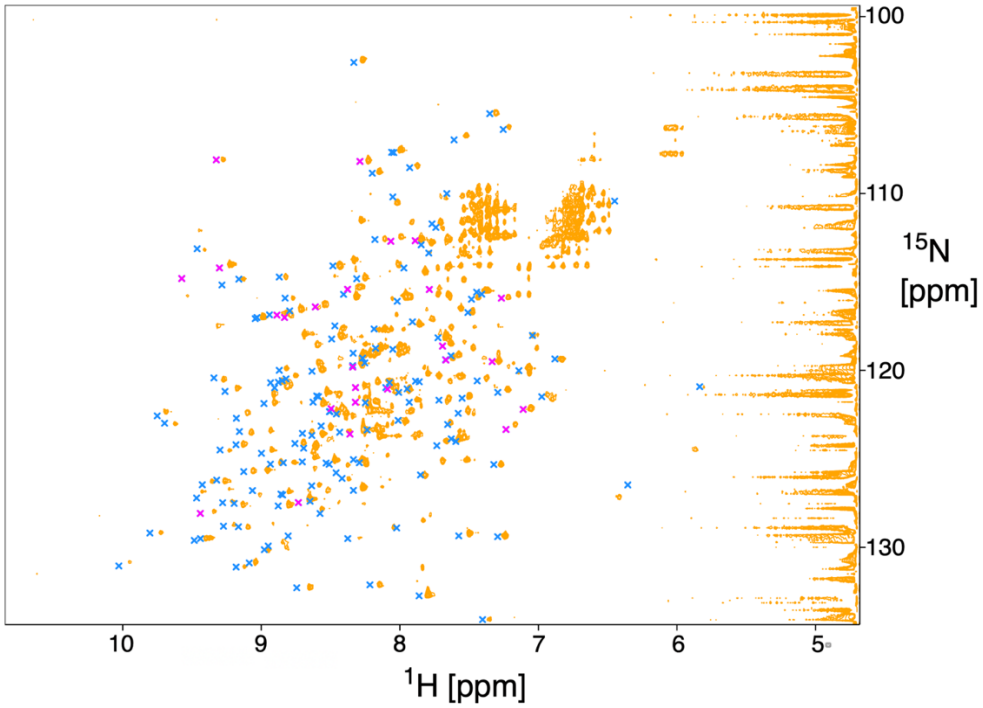


Figure 61: PLpro TROSY overlay with literature. Overlay of PLpro TROSY spectra (blue: Shimada *et al*., BMRB accession code 51992, with Ubl2 peaks highlighted in magenta; orange: PLpro TROSY spectrum acquired at this work).

In this experiment, we recorded the characteristic "fingerprint" HSQC spectrum of PLpro consistent with previously reported data (Figure 61).[234] The observed chemical shift differences can be attributed to variations in buffer conditions used during data acquisition. However, this spectrum serves as a reference for subsequent interaction studies and ligand-binding analyses.

7. Chapter: Expression and Isolation of the SNARE Proteins Synptobrevin-2 (Syb-2) and the Synaptosomal Associated Protein of 25 kilodaltons (SNAP25)

7.1 Abstract

High-quality structural characterization of neuronal SNARE proteins requires the production of isotopically labeled, soluble, and highly pure protein samples. In this study, we successfully optimized the expression and purification of SNAP25 and Syb-2 (residues 1-96) for subsequent NMR and CD spectroscopy analysis. Both proteins were recombinantly expressed in *E. coli* using minimal media for isotopic labeling and purified through multi-step protocols involving IMAC, IEC and SEC. For SNAP25, a cysteine-free variant was expressed at 20 °C, yielding soluble protein with high purity, as confirmed by SDS-PAGE and absorbance analysis. Thrombin digestion and IEC enabled separation from residual protease, and SEC further purified the sample to levels sufficient for high-resolution structural studies. Similarly, Syb-2 (1-96), lacking its transmembrane domain, was expressed and purified using an optimized workflow. Efficient removal of *E. coli* contaminants and successful tag removal were verified by SDS-PAGE and IEC. SEC yielded highly concentrated and structurally intact Syb-2 (1-96) suitable for biophysical characterization. These high-yield preparations enabled the detailed spectroscopic analyses presented in Chapter 8 and 9 and provide a robust platform for future studies on SNARE complex assembly and dynamics.

7.2 Introduction

7.2.1 SNARE Proteins

Throughout evolution, eukaryotic cells have developed vesicles to facilitate the transport of materials within and between cellular compartments, a process that is particularly crucial in neuronal cells for the release of neurotransmitters.[252] Membrane fusion of synaptic vesicles is a pivotal step during synaptic transmission.[253]

To ensure proper transport, the fusion of opposing lipid bilayers is necessary[254,255] and is generally triggered by changes in membrane potential, typically mediated by an action potential.[256] In response to electrically stimulation and Ca^{2+}-influx in the nerve terminal by voltage-gated channels, synaptic vesicles dock at the active zone, a specialized area of the presynaptic membrane organized by the SNARE complex (Figure 62).[255] Vesicle priming is mediated by proteins of the SNARE family in combination with calcium ions, facilitating membrane fusion.[255] Upon fusion of the synaptic vesicle with the presynaptic membrane, neurotransmitters are released via exocytosis in the synaptic cleft promoting sodium influx into the postsynaptic membrane and are further recycled by endocytosis.[257]

A small family of proteins mediating exocytosis and membrane fusion in neuronal cells are soluble NSF (*N*-ethylmaleimide-sensitive-factor) attachment receptor (SNARE) proteins. SNARE proteins are highly conserved in all species.[258] The secretion of neurotransmitters is regulated by the assembly of SNARE proteins including Synaptobrevin-2 (Syb-2), Synaptosomal Associated Protein of 25 kilodaltons (SNAP25) and Syntaxin-1A (Syx-1A) (Figure 63). Based on their localization in the presynaptic terminal, SNARE proteins are divided into two groups: v-SNAREs, including Syb-2 are located in the synaptic vesicle; Syx-1A and SNAP25 are referred as t-SNARE proteins and anchored in the presynaptic target membrane.[259,260] In their natural form, SNARE proteins are mainly intrinsically disordered and are further divided into 4 subtypes of SNARE motifs: the Qa, Qb, Qc and R-SNARE motif.[261,262] Vesicle trafficking is driven by the assembly of proteins into a stable four-helix bundle.[253,263] SNARE complex zippering is mediated by the transmembrane protein Syb-2 anchored in the lipid bilayer of synaptic vesicles. For zippering, a complex interaction of SNARE proteins and accessory proteins is required. To do so, the accessory protein Munc13 and Munc18 build a SNARE-acceptor complex in the plasma membrane.[253]

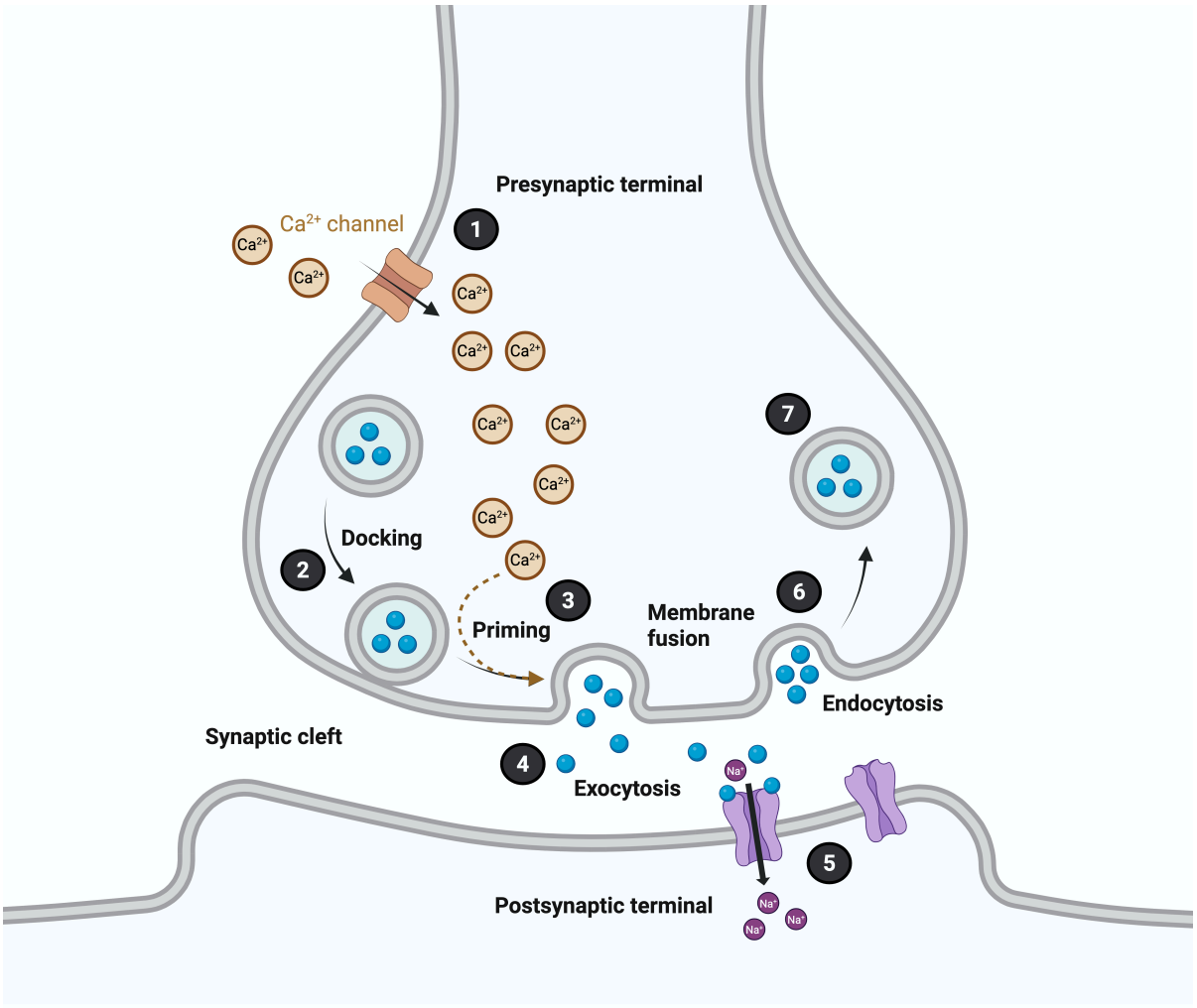


Figure 62: Schematic illustration of neurotransmitter release at the nerve terminal of a chemical synapse. Calcium ions trigger synaptic vesicle fusion with the presynaptic membrane. **(1)** An action potential induces membrane depolarization, leading to calcium influx at the presynaptic terminal via voltage-gated Ca^{2+} channels. **(2)** Synaptic vesicles dock at the active zone. **(3)** Vesicle priming is mediated by SNARE proteins and calcium ions, facilitating membrane fusion. **(4)** Upon fusion of the vesicle with the presynaptic membrane, neurotransmitters are released into the synaptic cleft via exocytosis. **(5)** Binding of neurotransmitters to sodium channels triggers conformational rearrangements of the channel protein, thereby promoting sodium (dark violet) influx into the postsynaptic neuron. **(6)** Neurotransmitters are subsequently recycled through endocytosis. **(7)** Newly formed vesicles are refilled with neurotransmitters and become available for the next cycle of release. Figure was created in BioRender.com.

By interacting with SNAP25 and Syntaxin-1A (Syx-1A) at the presynaptic membrane, Synaptobrevin-2 (Syb-2) initiates SNARE complex assembly (Figure 63). Syb-2 binds via its C-terminus to the C-terminal domain of Syx-1A. The resulting trans-SNARE complex adopts a coiled-coil structure composed of four heptad repeat motifs, with Syb-2 and Syx-1A each contributing one motif.[264] The remaining two helical motifs of the SNARE complex are contributed by SNAP25. This assembly brings the opposing membranes into close proximity, thereby enabling membrane fusion and neurotransmitter release (Figure 63). Notably, upon

Ca^{2+} influx, synaptotagmin and complexin - two further accessory proteins - are displaced, allowing completion of SNARE complex zippering and the transition from the trans-SNARE to the fully zippered cis-SNARE conformation, which drives membrane fusion and fusion pore formation (not shown).[265,266]

Following fusion, the SNARE complex adopts its cis configuration within the plasma membrane. It is subsequently disassembled by ATPase chaperones, allowing Syb-2 to be released and recycled for the next round of exocytosis. Although numerous studies have investigated the priming and triggering of SNARE complex formation, the prefusion conformations and the precise sequence of assembly steps remain incompletely understood and are still a matter of debate.[267] Accordingly, Figure 63 provides a simplified schematic representation of neurotransmitter release and the involvement of SNARE proteins.

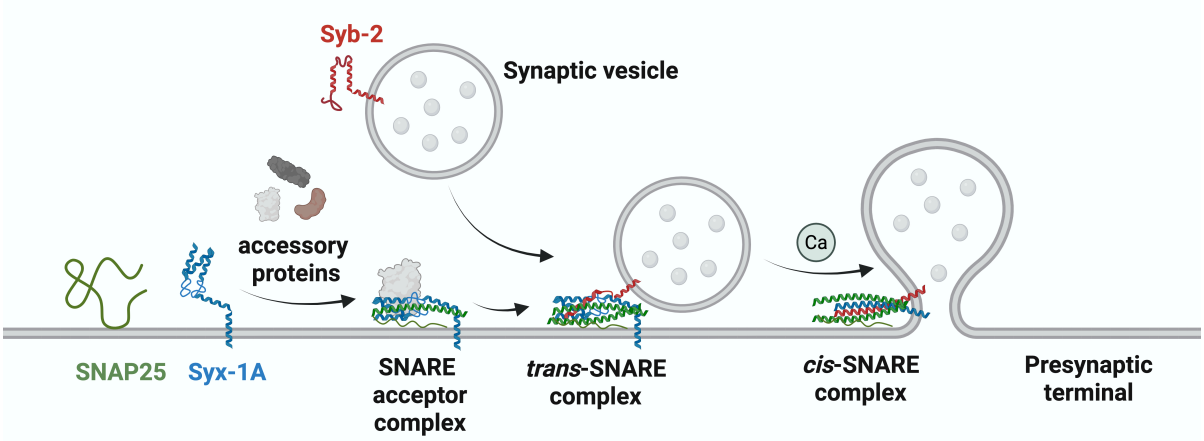


Figure 63: Scheme of calcium-dependent exocytosis of neurotransmitters mediated by SNARE proteins. SNAP25 (green) and Syntaxin-1A (Syx-1A, blue) are embedded in the plasma membrane of the presynaptic nerve terminal. Accessory proteins interact with SNAP25 and Syx-1A to initiate the formation of the SNARE acceptor complex. Synaptobrevin-2 (Syb-2, red), located in the membrane of the synaptic vesicle, then interacts with this acceptor complex to form the trans-SNARE complex, composed of Syb-2, SNAP25, and Syx-1A. The trans-SNARE complex is envisioned to be involved in both synaptic vesicle docking and "priming" (being ready for fusion upon Ca2+ influx). Progression of the N-to C-terminal zippering of the SNARE complex draws the synaptic vesicle toward the presynaptic membrane, ultimately leading to membrane fusion and neurotransmitter release. Note that this is a schematic illustration. Figure was created in BioRender.com.

7.2.2 SNAP25

With a molecular weight of approximately 25 kDa, SNAP25 is one of two t-SNARE proteins localized to the presynaptic plasma membrane. It is predominantly synthesized in endocrine and neuroendocrine cells and remains highly expressed throughout life.[268] Its mutated form

is correlated to neuronal disorders such as epileptic encephalopathies.[269] SNAP25 exists in two isoforms with SNAP25a (isoform 2) predominantly expressed in the embryonic brain and SNAP25b as its dominant isoform in central synapses during adulthood.[268] In its prefusion conformation, SNAP25 is largely an intrinsically disordered protein (IDP), with the exception of the N-terminal region of its first SNARE motif, which forms two α-helices (Figure 63). Unlike other SNARE proteins, SNAP25 lacks a transmembrane domain. Instead, it is anchored to the presynaptic plasma membrane via four palmitoylation sites located at cysteine residues C85, C88, C90, and C92.[270] It has been shown that the C-terminal part of the first SNARE motif (SN1), the connecting linker region (LR) and the majority of the second SNARE motif (SN2) are disordered and highly flexible (Figure 64).[271] In contrast to that, an α-helical propensity was determined at the C-terminal regions of SN1 and SN2 as well as inside the loop region (25%).[271] Interaction studies have demonstrated that SNAP25 binds to Syx-1A, nucleating the assembly of the SNARE complex.[272]

Figure 64: Domain organization of SNAP25 in the post-fusion *cis*-SNARE complex based on *Stief et al*.[271] SNAP25 is a 206 amino acids long protein, composed of 2 SNARE motifs (SN1 and SN2). SN1 raises from amino acid L11 to L81; SN2 from amino acid E145 to A199. In the post-fusion conformation, both SNs are of helical structure. The linker region (LR, residue G82 to N144) is incorporated between SN1 and SN2. Figure was created in BioRender.com.

7.2.3 Synaptobrevin-2

Synaptobrevin-2, short Syb-2 and also designated as vesicle-associated membrane protein 2 (VAMP-2), is a transmembrane protein composed of 116 amino acids, present on the surface of synaptic vesicles.[273] It has been shown that VAMP2 has a pivotal role in the regulation of neurotransmitter release.[274] The C-terminal region of Synaptobrevin-2 (Syb-2), consisting of approximately 20 hydrophobic residues, is embedded in the synaptic vesicle membrane (Figure 63), while the remaining 96 residues face the neuronal cytosol. In its monomeric prefusion state, Syb-2 is predominantly disordered and characterized by a highly flexible and dynamic N-terminal region, consistent with properties of an IDP.[275] Notably, rigidity within the protein increases progressively from the flexible N-terminus toward the hydrophobic C-terminus, correlating with enhanced lipid binding affinity and an accelerated rate of SNARE complex zippering.[275] Furthermore, the SNARE motif and adjacent linker region (LR) of Syb-2

exhibit pronounced Cα-Cβ secondary chemical shifts and α-helical propensity, indicating structural stiffness in both the linker and transmembrane regions (TMR).[275] Upon interaction with other SNARE proteins, Syb-2 undergoes partial folding and structural stabilization.[275] Its domain organization is illustrated in Figure 65.

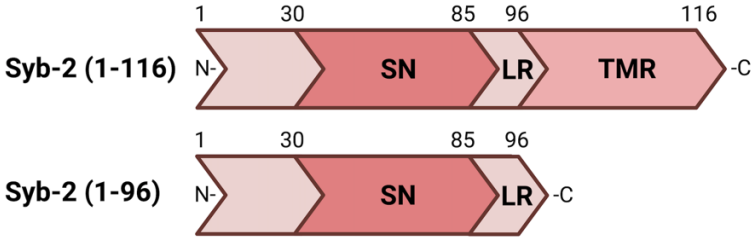


Figure 65: Domain organization of Syb-2 in the post-fusion SNARE-complex.[253,276,277] Syb-2 (1-116) is composed of four domains: a proline-enriched N-terminal domain (residues M1-R30), a SNARE motif (SN) raising from amino acid R31 to K85, a linker region (residues R86-M96) and a transmembrane domain at the C-terminus (residues I97-F114). The transmembrane region is embedded into the synaptic vesicle membrane. The remaining domains are faced to the cytoplasm. The truncated version of Syb-2 (1-96) leaking the TMR is shown below.

7.3 Aims

The aim of this work was to generate pure and high-quality protein samples of Syb-2 (1-96) and SNAP25 for later biophysical measurements. Samples were analyzed by CD spectroscopy, NMR spectroscopy and SAXS measurements. This chapter provides a detailed account of the expression and purification strategies employed to obtain homogeneous SNARE protein samples suitable for subsequent biophysical analyses. The following sections specifically describe the synthesis and preparation of uniformly labeled ^{15}N Syb-2 (1-96) and ^{15}N SNAP25 proteins.

7.4 Material and Methods

In the following sections, the protocols for the expression and purification of uniformly labeled ^{15}N SNAP25 and ^{15}N Syb-2 (1-96) are described in detail. Both SNARE proteins were successfully synthesized, and the resulting samples were used in the studies presented in Chapter 8 and 9. This chapter focuses exclusively on the preparation of ^{15}N SNAP25 and ^{15}N Syb-2 (1-96). Additional samples used in the respective publications were produced using the same expression and purification protocols; only the isotopic labeling schemes varied. For

further details, please refer to the Materials and Methods section. The isotopically labeled SNARE proteins used in the following chapters include: ^{1}H ^{15}N SNAP25, ^{15}N ^{13}C SNAP25 (Chapter 8); ^{2}H ^{15}N ^{13}C Syb-2 (1-96), ^{2}H ^{15}N Syb-2(1-96) and ^{1}H ^{15}N Syb-2 (1-96) (Chapter 9).

7.4.1 Buffer, Chemicals and Media

All standard chemicals used for the respective analyses were purchased from AppliChem, Merck, Roche and Sigma. Cell culture media were obtained from AppliChem. Medium, chemicals and buffers used for the experimental procedures are listed in the respective paragraphs. To avoid contamination during the experiment, chemicals used for cellular experiments were autoclaved and stored in sterile vessels.

7.4.2 Instruments

Instruments and materials used for the expression and isolation of ^{1}H-^{15}N SNAP25 and ^{2}H-^{15}N Syb-2 (1-96) are listed in Table 33 and 34.

Table 33: Instruments for the gene expression and protein purification of ^{1}H-^{15}N SNAP25 and ^{2}H-^{15}N Syb-2 (1-96).

Instrument	Manufacturer
Agarose-Gel chamber system	Bio-Rad Laboratories GmbH, München
ÄKTA Pure	GE Healthcare, Sweden
Autoclave VX-150	Systec GmbH, Linden
Centrifuge 5430 R	Eppendorf GmbH, Hamburg
Centrifuge Avanti J-26S XP	Beckmann Coulter GmbH, Krefeld
Desk centrifuge 5415 R	Eppendorf GmbH, Hamburg
Dri-Block Heater 08-3	Techne, UK
Drying chamber T 5050	Heraeus, Hanau
Electrophoresis power supply 3501 XL	Amersham Pharmacia biotech, UK
Gel documentation system	Bio-Rad Laboratories GmbH, München
Incubation shaker Infors HT Multitron	Infors AG, Bottmingen
Incubation shaker New Brunswick innova 40	Eppendorf GmbH, Hamburg
Incubation shaker New Brunswick innova 44	Eppendorf GmbH, Hamburg
Magnetic stirrer C-MAG MS 7	IKA GmbH & CO. KG, Staufen
Milli-Q Direct 16	Merck KGaA, Darmstadt
Mini Trans-Blot module blotting chamber	Bio-Rad Laboratories GmbH, München

NanoDrop 2000	Thermo Fischer Scientific, Schwerte
pH meter	Satorius AG, Göttingen
Scale	Satorius AG, Göttingen
Sonication device Sonoplus	Bandelin electronic GmbH & Co KG, Berlin
Thermocycler compact	Eppendorf, Hamburg
T100 Thermal Cycler	Bio-Rad Laboratories GmbH, München
Vortexer	Heidolph Instruments, Schwabach
VS70T sonotrode	Bandelin electronic GmbH & Co KG, Berlin

Table 34: List of consumables for the gene expression and protein purification of ^{1}H-^{15}N SNAP25 and ^{2}H-^{15}N Syb-2 (1-96).

Material	**Manufacturer**
Amicon MWCO 3 kDa	Merck Millipore, Darmstadt, Germany
Amicon MWCO 10 kDa	Merck Millipore, Darmstadt, Germany
cOmplete EDTA-free protease inhibitor tablet	Roche, Basel
Der Blaue Jonas - Single-step Coomassie Blue protein gel dye	GRP German Research, Haag a. d. Amper
Deuteriumoxid	Carl Roth, Karlsruhe
DNAse I, recombinant	Roche, Basel
E. coli BL21 (DE3) cells	New England Biolabs, Ipswich, USA
Goat anti-Mouse lgG (H+L) Secondary Antibody HRP conjugated	Thermo Fischer Scientific, Schwerte
HiLoad 16/60 Superdex 75 pg	Cytiva, Marlborough, USA
HiLoad 16/600 Superdex 200 pg	Cytiva, Marlborough, USA
HiPrep Q HP anion exchange chromatography column (20 mL)	Cytiva, Marlborough, USA
HiPrep SP FF cation exchange chromatography column (20 mL)	Cytiva, Marlborough, USA
Falcon tubes (15 mL and 50 mL)	Sarstedt, Nürnbrecht
MEM Vitamin Solution (100x)	Sigma-Aldrich, St. Louis, USA
Norell Select Series 3 mm NMR Tubes	Sigma-Aldrich, St. Louis, USA
PageRuler Prestained Protein Ladder	Thermo Fisher Scientific, Schwerte

Protino Ni-NTA 5 mL FPLC column	Macherey-Nagel, Düren
Plastic cuvettes	Brand GmbH & Co KG, Wertheim
Petri dishes	Hartenstein, Würzburg
Pipette tips	Sarsted, Nürnbrecht
PVDF-membrane 0.2 µm	GE Healthcare, Sweden
Reaction tube (1.5 mL and 2 mL)	Sarstedt, Nürnbrecht
Regenerated Cellulose Membrane Filter 0.2 µm	Sartorius, Göttingen
Serological pipettes	Sarstedt, Nürnbrecht
SnakeSkin dialysis tubes (3 K MWCO)	Thermo Fischer Scientific, Schwerte
SnakeSkin dialysis tubes (10 K MWCO)	Thermo Fischer Scientific, Schwerte
Spectra Multicolor Low Range Protein Ladder	Thermo Fischer Scientific, Schwerte
Sterile Syringes for Single Use (3 mL, 20 mL) Fisherbrand	Thermo Fischer Scientific, Schwerte
Syringe filter, Filtropur S, PES, pore size 0.2 µm for sterile filtration	Sarstedt, Nürnbrecht
Sterican Standard canulla 0.45 x 25 mm	B. Braun, Melsungen
SuperSignal West Pico PLUS Chemiluminescent Substrate	Thermo Fischer Scientific, Schwerte
Thrombin from bovine plasma (1000 units/mg protein)	Serva, Heidelberg
Whatman Cellulose Blotting Paper	Cytiva, Marlborough, USA
2xYT medium	Thermo Fischer Scientific, Schwerte
6x-His Tag Monoclonal Antibody (HIS.H8)	Thermo Fischer Scientific, Schwerte
^{15}N-NH_4-Cl	99%, Cambridge Isotope Laboratories, USA

7.4.3 Buffers and Solutions

In the following, the ingredients of buffers and solutions used for the expression and isolation of ^{1}H-^{15}N SNAP25 and ^{2}H-^{15}N Syb-2 (1-96) are listed in Table 35. All buffers and solutions were prepared with Milli-Q water (18.2 MOhm). Buffers for isolation were degassed and filtered.

Table 35: Buffer composition and chemical solutions.

Solution	Ingredients and Concentrations
Acrylamide:Bisacrylamide (34.3%:1.1%)	34.3% (v/v) Acrylamide, 1.1% (v/v) Bisacrylamide
Acrylamide:Bisacrylamide (37.5%:1%)	37.5% (v/v) Acrylamide, 1% (v/v) Bisacrylamide
Anode buffer (10x)	2 M Tris-HCl (pH 8.9)
Antibiotica stock	50 mg/mL Kanamycin
Blocking solution	5% milk (w/v) in (1x) TBS-T
Cathode buffer	1 M Tris-HCl (pH 8.25), 1 M Tricine, 1% (w/v) SDS
Column storage buffer	20% (v/v) EtOH
DNase I stock solution	20 mM Tris-HCl, 50 mM NaCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 50% (v/v) glycerol
Dialysis buffer SNAP25	20 mM HEPES (pH 7.4), 100 mM NaCl, 0.1 mM TCEP, 1 mM EDTA
Dialysis buffer Syb-2 (1-96)	20 mM HEPES (pH 7.4), 100 mM NaCl, 0.1 mM TCEP, 1 mM EDTA
Glycerol stock solution	31 g 2xYT powder per L 50% (v/v) glycerol
IEC buffer SNAP25	20 mM HEPES (pH 7.4), 0.1 mM TCEP, 1 mM EDTA
IEC elution buffer SNAP25	20 mM HEPES (pH 7.4), 1 M NaCl, 0.1 mM TCEP, 1 mM EDTA
IEC buffer Syb-2 (1-96)	20 mM HEPES (pH 7.4), 0.1 mM TCEP, 1 mM EDTA
IEC elution buffer Syb-2 (1-96)	20 mM HEPES (pH 7.4), 1000 mM NaCl, 0.1 mM TCEP, 1 mM EDTA
Induction reagent	0.5 M IPTG
IMAC elution buffer SNAP25	20 mM HEPES (pH 7.4), 500 mM NaCl, 400 mM imidazole
IMAC elution buffer Syb-2 (1-96)	20 mM HEPES (pH 7.4), 500 mM NaCl, 400 mM imidazole, 0.1 mM TCEP.

Laemmli sample buffer (4x)	10% (w/v) SDS, 10 mM *β*-mercaptoethanol, 20% (v/v) glycerol, 0.05% (w/v) Bromphenol-Blue, 200 mM Tris- HCl (pH 6.8)
Lysis buffer SNAP25	20 mM HEPES (pH 7.4), 500 mM NaCl, 8 mM imidazole
Lysis buffer Syb-2 (1-96)	20 mM HEPES (pH 7.4), 500 mM NaCl, 8 mM imidazole, 0.1 mM TCEP
TBS-T buffer (20x)	0.5 M Tris (pH 7.6), 3M NaCl, 2% (v/v) Tween20,
TGS buffer (10x)	25 mM Tris-HCl (pH 8.3), 192 mM glycine, 0.1% (w/v) SDS
TG buffer (10x)	25 mM Tris-HCl (pH 8.3), 192 mM glycine
Towbin buffer (1x)	10% (v/v) TG buffer (10x), 20% (v/v) EtOH
Tris-Glycine Gel Buffer (Stacking Gel)	0.5 M Tris (pH 6.8)
Tris-Glycine Gel Buffer (Separating Gel)	1.5 M Tris (pH 8.8)
Tris-Tricine Gel Buffer (3x)	3 M Tris-HCl (pH 8.45), 0.3% (w/v) SDS
SEC buffer SNAP25	20 mM sodium phosphate (pH 6.5), 150 mM NaCl, 0.1 mM TCEP
SEC buffer Syb-2 (1-96)	50 mM MES (pH 6.0), 150 mM NaCl, 0.1 mM TCEP, 1 mM EDTA

7.4.4 Plasmids and Bacterial Strains

The plasmids encoding for SNAP25 and Syb-2 (1-96) were thankfully received from our collaborators at the Jahn lab (Laboratory of Neurobiology, Max-Planck Institute for Multidisciplinary Sciences, Göttingen, Germany). The construct composition and further information are shown in Table 36.

Table 36: Construct details of SNAP25 and Syb-2 (1-96).

Construct Information	**SNAP25**	**Syb-2 (1-96)**
Original organism	*Rattus norvegicus*	*Rattus norvegicus*
Plasmid	pET-28a(+)	pET-28a(+)
Codon-optimized	*E. coli*	*E. coli*

Cell strains	*E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) – short, *E. coli* BL21 (DE3)	*E. coli* str. B F– *ompT gal dcm lon hsdSB*(*rB– mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) – short, *E. coli* BL21 (DE3)
N-terminal Tag	6xHis	6xHis
Cleavage Site	Thrombin	Thrombin
Antibiotic Resistance	Kanamycin	Kanamycin
Restriction sites	NdeI, XhoI	NdeI, XhoI
Uniprot accession number	P60881-2	P63045
Mutations	C84S, C85S, C90S, C92S	None
Construct size with tag	25.72 kDa, 229 amino acids	12.68 kDa, 116 amino acids
Construct size	23.84 kDa, 206 amino acids	10.80 kDa, 99 amino acids
pI of His-tagged construct	5.24	9.16
pI of construct	4.79	8.28
Extinction coefficient	6990 M^{-1} cm^{-1}	12490 M^{-1} cm^{-1}

In the following the amino acid sequences of His-tagged and digested constructs are listed. The His-tag is highlighted in green, the thrombin cleavage site in purple and the construct itself in red. The stop codons are marked with an asterisk ("*").

6xHis-Thrombin-SNAP25:
MGSSHHHHHHSSGLVPRGSHMASMAEDADMRNELEEMQRRADQLADESLESTRRMLQLVEESKDA
GIRTLVMLDEQGEQLDRVEEGMNHINQDMKEAEKNLKDLGKSSGLFISPSNKLKSSDAYKKAWGNNQD
GVVASQPARVVDEREQMAISGGFIRRVTNDARENEMDENLEQVSGIIGNLRHMALDMGNEIDTQNRQ
IDRIMEKADSNKTRIDEANQRATKMLGSG*

SNAP25:
GSHMASMAEDADMRNELEEMQRRADQLADESLESTRRMLQLVEESKDAGIRTLVMLDEQGEQLDRVE
EGMNHINQDMKEAEKNLKDLGKSSGLFISPSNKLKSSDAYKKAWGNNQDGVVASQPARVVDEREQMA
ISGGFIRRVTNDARENEMDENLEQVSGIIGNLRHMALDMGNEIDTQNRQIDRIMEKADSNKTRIDEANQ
RATKMLGSG*

6xHis-Thrombin-Syb-2(1-96):
MGSSHHHHHHSSGLVPRGSHMSATAATVPPAAPAGEGGPPAPPPNLTSNRRLQQTQAQVDEVVDIM
RVNVDKVLERDQKLSELDDRADALQAGASQFETSAAKLKRKYWWKNLKMM*

Syb-2 (1-96):
GSHMSATAATVPPAAPAGEGGPPAPPPNLTSNRRLQQTQAQVDEVVDIMRVNVDKVLERDQKLSELD
DRADALQAGASQFETSAAKLKRKYWWKNLKMM*

7.4.5 Sequencing of Ordered Gene Constructs

For sequencing, 12 µL of *E. coli* codon-optimized plasmids were sent in a concentration between 50 and 100 ng/µL to Microsynth Seqlab. As a primer source, the standard primer list of Microsynth Seqlab was used. Primers used for sequencing are listed in Table 37.

Table 37: Primer list used for sequencing of SNAP25 and Syb-2 (1-96) plasmids.

Construct	Plasmid concentration	Primer name	Primer sequence
pET-28a(+)-6xHis-Thrombin-SNAP25	95.3 ng/µL	T7	TAATACGACTCACTATAGGG
pET-28a(+)-6xHis-Thrombin-Syb-2(1-96)	85.0 ng/µL	T7	TAATACGACTCACTATAGGG

The online tool NucleoBlast[89] was used for the alignment of the resulting and theoretical sequences. Sequencing results of the constructs are listed in the supplementary information (Table S1).

7.4.6 Transformation of Plasmids in Chemically Competent *E. coli* BL21 (DE3) Cells

Constructs were transformed into *E. coli* strain B F– *ompT gal dcm lon hsdSB*(*rB*– *mB*–) λ (DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS) – short, *E. coli* BL21 (DE3) (New England Biolabs, Ipswich, USA). Firstly, 50 µL of cells were thawed for 20 min on ice. Then, 10 ng plasmid was added to the cell suspension and incubated for 30 min on ice. Cells were heat-shocked for 45 s at 42 °C. After that, cells were cooled down for 2 min on ice. Then, 450 µL of preheated 2xYT medium were mixed with the cell suspension and incubated for 1 h at 37 °C (800 rpm). Cells were plated onto LB-agar plates containing the corresponding antibiotics and were incubated at 37 °C overnight. The next day, grown colonies were picked and cultured in 5 mL 2xYT medium supplemented with antibiotics at 37°C and 160 rpm overnight. Cells were centrifuged for 1 min at 16,000 x *g*. The supernatant was discarded, and the cell pellet was resuspended in glycerol stock solution (50% (v/v) 2xYT medium, 50% (v/v) glycerol). Cells were flash frozen in liquid nitrogen and stored at -80 °C for further use. Glycerol stocks were used for expression tests and large-scale expression in minimal medium.

7.4.7 Expression Tests of SNAP25 and Syb-2 (1-96) in Minimal Medium

Expression tests were used to identify optimal expression conditions for both SNARE proteins. As minimal medium was required for protein labelling, small-scale expression tests of SNAP25 and Syb-2 (1-96) were directly performed in minimal medium instead of 2xYT medium. For that, 5 mL preculture of both constructs were inoculated with the corresponding glycerol stock and 50 mg/mL Kanamycin for selection. Precultures were incubated at 37 °C and 160 rpm overnight (16 h). The next day, 50 mL of fresh minimal medium was mixed with 0.5 mL preculture and 50 µL of the corresponding antibiotic in a 250 mL baffled Erlenmeyer flask. The main culture was incubated at 37 °C and 160 rpm. At an OD_{600} of 0.8, a sample of 1 mL was taken to analyze the gene expression before induction. For that, the 1 mL culture was centrifuged for 1 min at 16,000 x *g*. The medium was discarded, and the pellet was diluted in 45 µl Milli-Q water and 15 µl 6x Laemmli SDS-buffer[92]. Gene expression was induced by 0.5 mM IPTG causing the activation of the T7-promotor system. After another incubation time overnight at 20 °C, a second 1 mL culture sample was taken after induction and mixed with Laemmli SDS-buffer[92]. Expression was analyzed by SDS-PAGE and Western Blot analysis.

7.4.8 Large-Scale Expression of SNARE Proteins in Minimal Medium

For large-scale protein expression, an initial preculture was prepared by inoculating 50 mL of 2xYT medium supplemented with 50 µg/mL kanamycin in 250 mL baffled Erlenmeyer flasks. The preculture was incubated overnight at 37 °C with shaking at 160 rpm to ensure optimal aeration and bacterial growth. For the expression of uniformly ^{15}N-labeled SNAP25, two sterile 2-liter Erlenmeyer flasks, each containing 500 mL of minimal medium, were prepared by heat sterilization. Each flask was inoculated with 10 mL of the overnight preculture and supplemented with kanamycin at a final concentration of 50 µg/mL to maintain plasmid selection. The detailed composition of the minimal medium used for isotopic labeling is provided in Table 38.

In case of ^{2}H ^{15}N Syb-2 (1-96), another preculture was inserted to slowly get used to minimal medium and D_2O. After the first preculture, the second preculture was prepared simultaneously, replacing 2xYT medium with minimal medium (Table 39). For that, 50 mL minimal medium was supplemented with 50 mg/mL Kanamycin and 1 mL preculture one. The cells were incubated at 37 °C and 160 rpm overnight for 16 h. The next morning, a main culture

was prepared as described for ^{1}H-^{15}N SNAP25, replacing water by D_2O. For a better comprehension, details are displayed in Table 39.

Table 38: Minimal medium protocol.

Chemical	**Stock solution**	**Final concentration**	**Amount for 1 L main culture**	**Amount for 50 mL preculture**
M9 salt solution	337 mM Na_2HPO_4, 220 mM KH_2PO_4, 85.5 mM NaCl (10x)	33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.55 mM NaCl (1x)	100 mL	5 mL
MEM vitamin solution	100x	1x	1 mL	50 µL
$MgSO_4$	1 M	1 mM	1 mL	50 µL
$CaCl_2$	0.1 M	0.3 mM	3 mL	150 µL
Biotine	1 mg/mL (w/v)	1 mg/L	1 mL	50 µL
Thiamine	1 mg/mL (w/v)	1 mg/L	1 mL	50 µL
Trace element solution	13.4 mM EDTA, 3.1 mM $FeCl_3$-$6H_2O$ 0.62 mM $ZnCl_2$ 76 µM $CuCl_2$-$2H_2O$ 42 µM $CoCl_2$-$2H_2O$ 162 µM H_3BO_3 8.1 µM $MnCl_2$-$4H_2O$ (100x)	134 µM EDTA, 31 µM $FeCl_3$-$6H_2O$ 6.2 µM $ZnCl_2$ 0.76 µM $CuCl_2$-$2H_2O$ 0.42 µM $CoCl_2$-$2H_2O$ 1.62 µM H_3BO_3 0.081 µM $MnCl_2$-$4H_2O$ (1x)	10 mL	500 µL
^{15}N NH_4Cl	99% (w/w)	0.1% (w/v)	1 g	50 mg
^{12}C D-Glucose-Monohydrate	20% (w/v)	0.4% (w/v)	20 mL	200 µL

At OD_{600} of 0.8, 0.5 mM IPTG was added inducing the expression of target proteins. Also here, samples before and after induction of expression were used to analyze the expression grade of both proteins. After another incubation period overnight at 20 °C and 120 rpm, cells were

harvested at 4 °C via centrifugation at 6,000 x *g* for 15 min. After centrifugation, the medium was discarded. The pellet was weighted and stored at -20 °C for further use.

Table 39: Expression of ^{2}H ^{15}N Syb-2 (1-96) in deuterated minimal medium.

Day	Culture	Ingredients	Incubation time [h]	Temperature [°C]	Shaking [rpm]
1	Preculture 1	50 mL 2xYT medium, 50 µL 50 mg/mL Kanamycin, Glycerol stock	16	37	160
2	Preculture 2	50 mL minimal medium in ddH_2O, 50 µL 50 mg/mL Kanamycin, 1 mL preculture 1	16	37	160
3	Mainculture	500 mL minimal medium in D_2O, 500 µL 50 mg/mL Kanamycin, 10 mL preculture 2	16	37 (BI), 20 (AI)	120

7.4.9 Cell Lysis by Sonication

The frozen cell pellet was thawed and resuspended in lysis buffer (5 mL lysis buffer/g cell pellet). For DNA degradation, 10 µg per g cell pellet DNase I were added to the cell suspension followed by an EDTA-free protease inhibitor tablet to decrease protease activity. The cell suspension was incubated for 20 min on a roller mixer at 4 °C until the solution became homogenous. Samples were ultrasonicated using a VS70T sonotrode for 20 min. For that, the amplitude was set to 60%, with intervals of 1 s sonication and 3 s pause. As heat is generated during sonication, the sample was put onto an ice-water bath previously. After cell lysis, the suspension was centrifuged for 45 min at 40,000 x *g* at 4 °C using a JA20-rotor. Supernatant and cell pellet were separated, and samples analyzed by SDS-PAGE. The supernatant was used for protein purification by IMAC using an ÄKTA start system.

7.4.10 SDS-PAGE of SNARE Proteins

SDS-PAGE was used to quantify successful expression and purification of samples allowing protein separation by mass.[92,93] Based on the MW of Syb-2 (1-96) and SNAP25, a Tris-Glycine and Tris-Tricine SDS-PAGE were assessed (Table 40 and 41).

Independent of the gel, samples were diluted in 15 µl 4x Laemmli SDS-buffer and boiled at 95 °C for 10 min for protein denaturation. Samples were centrifugated for 10 min at 10,000 x *g*. Samples and the protein standard marker were loaded on the corresponding SDS gel and cable-connected to the electrophoresis power supply 3501 XL. When the blue running front was nearly at the bottom, the tank was taken off. After electrophoresis, the gel was washed four times with pure water prior to staining for 1 h with Coomassie (Blauer Jonas). Destaining was not required. Finally, the separated proteins were visualized by a gel documentation system.

7.4.10.1 Tris-Tricine SDS-PAGE of SNAP25

A running and stacking SDS-gel were prepared (Table 40). 3 µL of marker (Spectra Multicolor Low Range Ladder) and 10 µL of each sample were added. The electrophoresis was performed with a constant current strength of 40 mA per gel in running buffer. For that, the gel chamber was filled with 1x cathode buffer and surrounded by 1x anode buffer.

Table 40: Protocol for a Tris-Tricine SDS-PAGE (2x).

Component	20% Separating Gel	5.6% Stacking Gel
ddH_2O	-	3.2 mL
Gel buffer (3x)	5 mL	2.1 mL
Glycerol	1.6 mL	-
Acrylamide:Bisacrylamide (34.3%:1.1%)	8.5 mL	1 mL
APS (10% (w/v))	50 µL	50 µL
TEMED	25 µL	25 µL

7.4.10.2 Tris-Glycine SDS-PAGE of Syb-2 (1-96)

Table 41 demonstrates the components used for Tris-Glycine gels. For the run, 3 µL of marker (Page Ruler Prestained) and 10 µL of each sample were added onto the gel. The

electrophoresis was performed with a constant current strength of 120 mV in running buffer (1x TGS buffer).

Table 41: Protocol for a Tris-Glycine SDS-PAGE (2x).

Component	15% Separating Gel	5% Stacking Gel
ddH_2O	-	3.2 mL
Tris-Glycine Gel Buffer (Separating Gel)	5 mL	2.1 mL
Tris-Glycine Gel Buffer (Stacking Gel)	1.6 mL	-
Acrylamide:Bisacrylamide (37.5%:1%)	8.5 mL	1 mL
SDS (10% (w/v))	50 µL	50 µL
APS (10%(w/v))		
TEMED	25 µL	25 µL

7.4.11 Western Blot Analysis

Western blot analysis was used to identify the protein of interest in a SDS-gel.[94] Therefore, samples were separated using SDS-PAGE as described in section 7.4.10. For generating a Western-Blot-Sandwich, a PVDF-membrane was activated by 100% (v/v) EtOH and six Watman papers were soaked in (1x) Towbin buffer. The SDS gel was put onto the PVDF-membrane and surrounded with three pieces of Watman papers from each side. The transfer was performed in a blotting chamber for 30 min at 25 V (1 A). To check if the transfer was successful, the membrane was incubated with Ponceau-S-solution shortly and washed four times with water. Visible bands gave a hint on a successful protein transfer. Then, the membrane was blocked in blocking solution for 1 h at room temperature to saturate all antibody sites. After that, the milk was discarded, and the 6x-His Tag Monoclonal Antibody (HIS.H8) was 1:1,000 diluted in blocking solution. The membrane was stored at 4 °C overnight under gentle shaking. The next day, the primary antibody solution was recycled and frozen at -20 °C for further use. Excess milk was removed by water. The membrane was washed three times in 1x TBS-T for 5 min. The Goat anti-Mouse lgG (H+L) Secondary Antibody HRP conjugated covalently linked to the horseradish peroxidase was 1:10,000 diluted in blocking solution, added to the membrane, and incubated for 1 h at room temperature under gentle

shaking. The antibody solution was discarded, and the membrane was washed again three times for 5 min in 1x TBS-T. The membrane was documented using the Bio-Rad Universal Hood III documentation system. For signal detection, a 1:1 dilution of Super Signal Peroxidase Solution and Enhancer Solution were added on the membrane and incubated for 1 min in a dark room. The signal accumulation mode was utilized in high resolution mode, with 40 frames in 200 s. The protein ladder was visualized separately by using the colorimetric mode. Finally, images were merged and analyzed.

7.4.12 IMAC of His-tagged SNAP25 and Syb-2 (1-96)

For purification, a Protino Ni-NTA 5 mL FPLC column connected to an ÄKTA Pure purification system from GE Healthcare (Sweden) was used. Purification was conducted at room temperature. All buffers used for isolation were filtered and degassed prior to use. For the purification process, two different buffers systems were used: lysis buffer and elution buffer, supplemented with imidazole for protein elution. The system was washed with degassed and filtered ddH_2O until the absorbance and conductivity reached 0 mAU. Then, the system was equilibrated with 10 CVs of lysis buffer at a flow rate of 2 mL/min and 0.3 mPa. As the flowrate was decreased to 0.5 mL/min, the column was assembled, and the system was washed again with 10 CVs of lysis buffer (2 ml/min). Then the sample was loaded onto the column (1 mL/min) and the flow through was collected for SDS-PAGE analysis. After that the column was washed again with lysis buffer until the absorbance at 280 nm reached the baseline. Then, the amount of elution buffer was increased to 4% to remove unspecific bound proteins. After washing, target proteins were eluted using a gradient of elution buffer (40 min, 1 mL/min). As the percentage of elution buffer was increased to 100%, the target protein was fully eluted. Fractions containing the target were pooled. The column and system were washed with ddH_2O and stored in 20% (v/v) EtOH.

7.4.13 Digestion and Dialysis of Target Constructs

The affinity tag was removed from target proteins by proteolytic cleavage. Specifically, thrombin protease was employed to cleave 6×His-tags, thereby removing the N-terminal polyhistidine sequence. For that, 10 units of protease were mixed with 1 mg of protein. The approach was performed in combination with dialysis overnight to decrease the imidazole concentration in the sample. For that, the protein mix was put in a SnakeSkin dialysis tube

(3 K MWCO for Syb-2 and 10 K MWCO for SNAP25) surrounded by 1 L of the corresponding dialysis buffer (see Table 3). The sample was incubated overnight at 4 °C under gentle stirring.

7.4.14 Ion Exchange Chromatography (IEC) of SNAP25 and Syb-2 (1-96)

After digestion, target proteins were separated from the protease using IEC. Based on the pI of the constructs and the IEC buffers, a cation or anion exchanger column was used. In the case of SNAP25, a HiPrep Q HP anion exchange chromatography column was required; for Syb-2 (1-96) a HiPrep SP FF cation exchange chromatography column was used. Each column was equilibrated in IEC buffer (10 CVs). Prior to sample loading, the digested sample was diluted with IEC buffer to reduce the salt concentration to 10 mM. Diluted sample was loaded at a flowrate of 1 mL/min. The flowthrough was collected and analyzed by SDS-PAGE. Proteins were eluted using a salt gradient increasing the amount of IEC elution buffer within 40 min up to 100% (1 mL/min). Fractions of the main peak were collected and analyzed by SDS-PAGE. Fractions containing the target protein were pooled and concentrated up to 3 mL for SEC injection.

7.4.15 Size Exclusion Chromatography of SNAP25 and Syb-2 (1-96)

The concentrated protein was further purified using SEC. For the isolation, SNAP25 was purified using a HiLoad 16/600 Superdex 75 pg and Syb-2 (1-96) using a HiLoad 16/600 Superdex 200 pg. Firstly, the column was equilibrated with SEC buffer at a flowrate of 0.5 mL/min. Due to the sensitivity of the column, the buffer was sterile filtered and degassed previously. After 8 h of equilibration (2 CVs), the concentrated sample was injected using a 3 mL syringe. The run was performed at 1 mL/min, 0.3 MPa delta column pressure and 0.5 MPa precolumn pressure until 1 CV was reached. Samples of the main peak were collected for SDS-PAGE. The column was washed with degassed and filtered ddH_2O at 0.5 mL/min (2 CVs), followed by an additional washing step with 20% (v/v) EtOH (1 CV) at 0.3 mL/min. Fractions were pooled and concentrated by centrifugation with an Amicon tube (MWCO 3 K for Syb-2 (1-96) and MWCO 10 K for SNAP25) at 6,000 x *g* until a sample concentration of around 200 µM was reached. The final protein concentration was determined using NanoDrop and UV-VIS spectroscopy. Samples were flash-frozen in liquid nitrogen and stored at -80 °C.

7.4.16 NMR Sample Preparation of ^{1}H-^{15}N SNAP25 and ^{2}H-^{15}N Syb-2 (1-96)

For NMR measurements, pure and high-quality protein samples were supplemented with 5% (v/v) D_2O, 0.1% (v/v) NaN_3 and 0.8 µM DSS and filled up to 200 µL with the corresponding SEC buffer (SNAP25: 20 mM sodium phosphate (pH 6.5), 150 mM NaCl, 0.1 mM TCEP; Syb-2 (1-96): 50 mM MES (pH 6.0), 150 mM NaCl, 0.1 mM TCEP, 1 mM EDTA). Samples were transferred into a Norell Select Series 3 mm NMR Tube and measured at different spectrometers at the Forschungszentrum Jülich (Germany) according to the publications listed in Chapter 8 and 9.

7.5 Results

In general, structural analysis requires protein samples of high purity and homogeneity. The initial step in the purification process involves the high-yield production of the target protein. Following the transformation of plasmids into *E. coli*BL21 (DE3) cells, various expression conditions were systematically evaluated. Gene expression is influenced by multiple factors, including construct design, growth medium composition, and cultivation temperature. Consequently, these three parameters were examined in the subsequent experiments.

7.5.1 Expression of ^{1}H-^{15}N 6xHis-SNAP25 in Minimal Medium

Small-scale expression of 6xHis-Thrombin-SNAP25 was performed in minimal medium and 20 °C after induction of expression (AI), as minimal medium is required for protein labelling and indispensable for NMR measurements. For that, *E. coli* BL21 (DE3) cells containing the target construct, were cultivated in 2xYT medium at 37 °C overnight. Main cultures were cultivated in minimal medium and incubated at 20 °C overnight after induction of expression. Optimal expression conditions were identified by large bands on the SDS-PAGE and Western Blot membrane at the proposed size of around 25 kDa. Samples before (BI) and after induction (AI) were taken (Figure 66A, B).

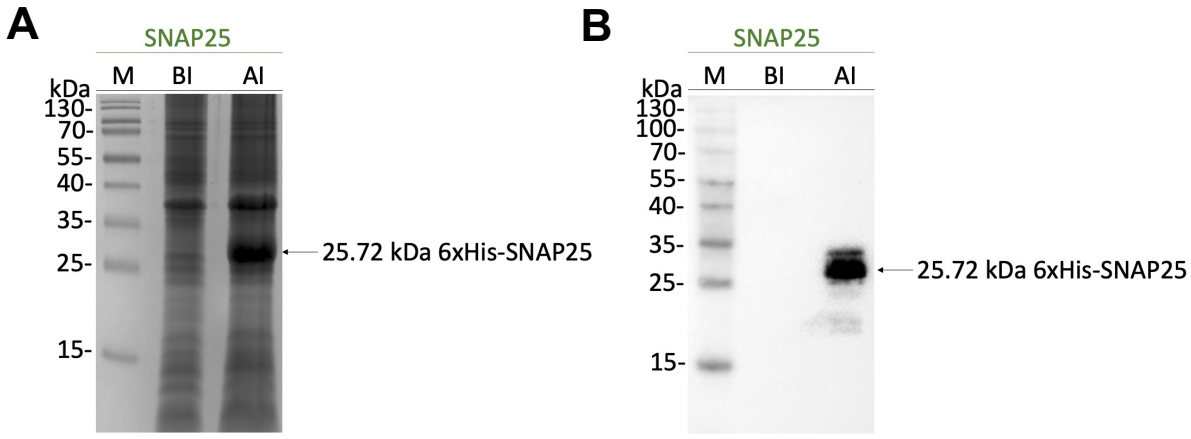


Figure 66: Expression of ^{1}H-^{15}N 6xHis-Thrombin-SNAP25 in minimal medium. *E. coli* BL21 (DE3) cells incorporating the pET-28a(+)-6xHis-Thrombin-SNAP25 plasmid were cultivated in minimal medium at 20 °C overnight. **(A)** Samples before (BI) and after induction (AI) of expression were taken and analyzed by a 15% Tris-Glycine SDS-PAGE. Protein bands were compared to the marker (M) (PageRuler Prestained Protein Ladder). 6xHis-tagged SNAP25 was expected at 25.72 kDa. **(B)** The His-tagged protein was identified by Western Blot analysis at the proposed size.

As shown in Figure 66A, overexpression of 6xHis-SNAP25 was confirmed by the appearance of a prominent band at approximately 25 kDa following expression in minimal medium at 20 °C overnight, indicating that these conditions are optimal for protein expression. In addition, a strong band at ~37 kDa was observed in both the BI and AI samples, likely corresponding to a housekeeping protein endogenously expressed by *E. coli*. In Figure 66B, the His-tagged SNAP25 protein was again detected at approximately 25 kDa. Furthermore, histidine-rich *E. coli* proteins with molecular weights slightly above and below that of SNAP25 were also visible on the Western blot membrane.

7.5.2 Isolation of ^{1}H-^{15}N 6xHis-SNAP25 by IMAC

After the presence of the target protein was determined by western blot analysis and optimal expression conditions were determined, the next aim of this work was to get large amounts of high-quality protein samples. For that, the 6xHis-tagged SNAP25 protein was rapidly separated from other histidine-rich *E. coli* proteins by IMAC followed by a Thrombin-cleavage to get rid of the tag. After the cells were harvested, the cell pellet was resuspended in lysis buffer (20 mM HEPES (pH 7.4), 500 mM NaCl, 8 mM imidazole). After sonication and a centrifugal step, the supernatant was loaded on an equilibrated column. Due to a hexa-histidine tag at the N-terminus of the construct, a Ni-NTA sepharose 6 fast flow column was used.

At the beginning of the isolation process, the clarified lysate was loaded onto a Ni-NTA affinity column. *E. coli* proteins lacking a His-tag did not bind to the resin and were eluted during the flowthrough, as indicated by the elevated absorbance values observed between 8 and 80 mL (Figure 67A). Consistent with these data, SDS-PAGE analysis revealed multiple *E. coli* proteins in the flowthrough fraction (Figure 67B). Upon a slight increase in imidazole concentration (to 20 mM, equivalent to 5% of 400 mM), nonspecifically bound proteins were washed off the column. These wash fractions, like the flowthrough, contained various *E. coli* proteins, as shown by SDS-PAGE. Interestingly, faint bands corresponding to SNAP25 were already visible in the 4% elution buffer fraction, suggesting that a small amount of the target protein may have been co-eluted at low imidazole concentrations. This could be due either to weak binding affinity to the resin or column overloading (Figure 67B). The majority of 6xHis-SNAP25, however, was eluted upon increasing the imidazole concentration to 180 mM (45% of the

elution buffer), as evidenced by a prominent peak between 106 and 120 mL in the chromatogram. Two-milliliter fractions across this peak were collected and analyzed by SDS-PAGE.

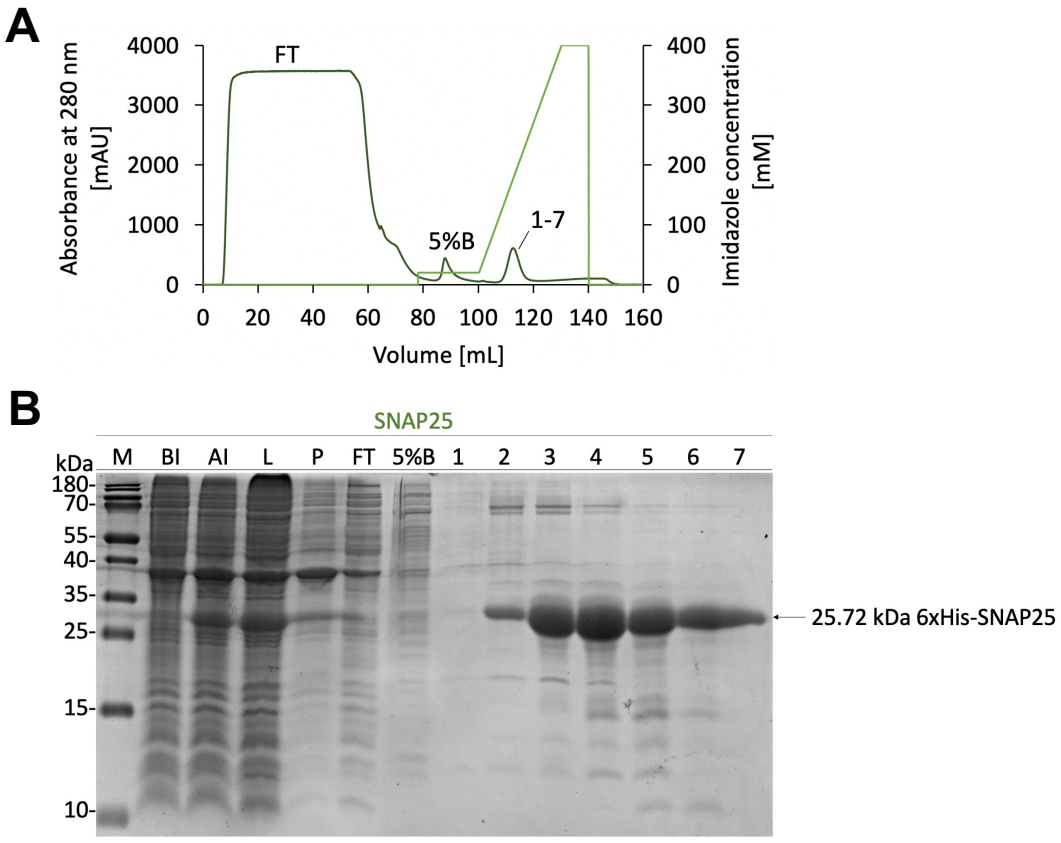


Figure 67: Purification of His-tagged ^{1}H-^{15}N SNAP25 by IMAC. The protein sample was loaded on a Ni-NTA column equilibrated in 20 mM HEPES (pH 7.4), 500 mM NaCl, 8 mM imidazole. The corresponding chromatogram is shown in **A**. The absorbance (dark green) measured at 280 nm was plotted against the volume in mL. Target protein was eluted using 20 mM HEPES (pH 7.4), 500 mM NaCl, 400 mM imidazole. **(B)** SDS-samples were taken for a 15% (v/v) Tris-Glycine SDS-PAGE. Samples before induction (BI), after induction (AI), lysate (L), pellet (P), flowthrough (FT), 5% elution buffer (5%B) and fractions (1-7) were taken and compared to the marker (M) (Page Ruler Prestained Protein Ladder).

As shown in Figure 67B, substantial amounts of 6xHis-SNAP25 were detected in fractions two through seven. These fractions were subsequently pooled for further experiments, as the target protein was present in each of these fractions. Due to the potential destabilizing effects of imidazole on protein activity and structure, the pooled sample was dialyzed overnight against a dialysis buffer containing 20 mM HEPES (pH 7.4), 100 mM NaCl, 0.1 mM TCEP, and 1 mM EDTA to reduce the imidazole concentration in the sample. Dialysis was carried out in parallel with overnight protease digestion to separate the His-tag from the target.

7.5.3 Isolation of ^{1}H-^{15}N SNAP25 by IEC

To separate the digested SNAP25 protein from undigested 6xHis-SNAP25, cleaved His-tags, and residual Thrombin, IEC was performed. IEC allows the separation of proteins based on their net charge. Given the isoelectric point (pI) of the digested SNAP25 (pI = 4.79) and the pH of the buffer used during dialysis (pH = 7.4), the target protein carries a net negative charge under these conditions. Therefore, an anion exchange column (Q-Sepharose) was employed to facilitate separation. Protein elution was achieved using a linear salt gradient, in which increasing concentrations of Na^{+} and Cl^{-} ions disrupt electrostatic interactions between the protein and the column matrix. Figure 68 illustrates the IEC purification process. Because the dialyzed sample initially contained 100 mM NaCl, it was diluted with a low-salt buffer (20 mM HEPES (pH 7.4), 0.1 mM TCEP, 1 mM EDTA) to reduce the NaCl concentration to approximately 10 mM, thereby enhancing the binding efficiency of SNAP25 to the Q-Sepharose resin.

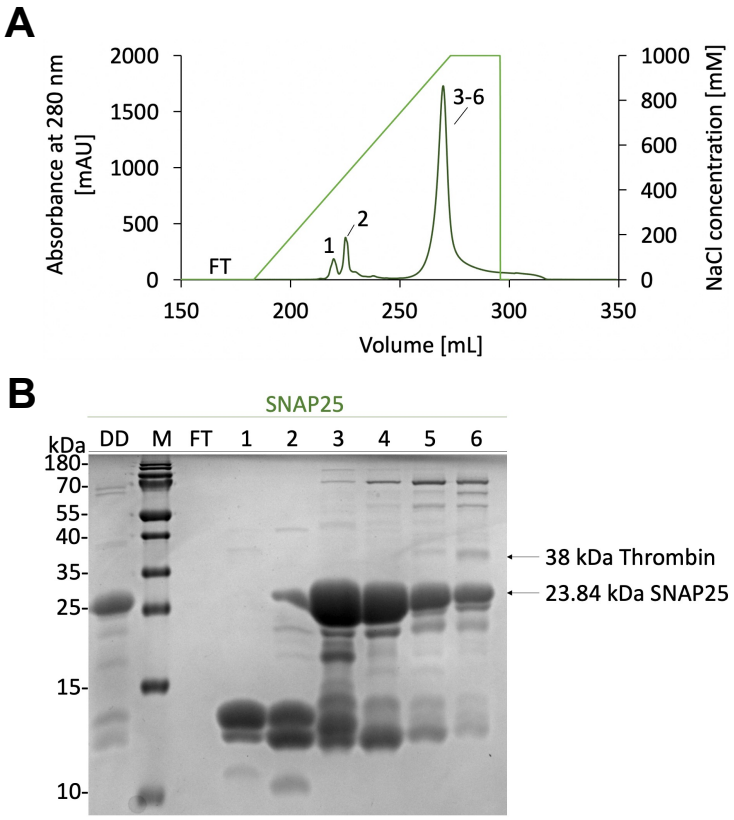


Figure 68: Isolation of ^{1}H-^{15}N SNAP25 by ion exchange chromatography. The protein sample was loaded on a Q-sepharose anion exchange chromatography column equilibrated in 20 mM HEPES (pH 7.4), 0.1 mM TCEP, 1 mM EDTA. The corresponding chromatogram is shown in **A**. The absorbance (dark green) measured at 280 nm was plotted against the volume in mL. Target protein was eluted by a sodium chloride gradient using 20 mM HEPES (pH 7.4), 1000 mM NaCl, 0.1 mM TCEP, 1 mM EDTA. **(B)** SDS-samples were taken for a 15% (v/v) Tris-Glycine SDS-PAGE. Samples after dialysis and digestion (DD), flowthrough (FT) and fractions (1-6) were taken and compared to the marker (M) (Page Ruler Prestained Protein Ladder). Digested SNAP25 was expected at 23.84 kDa.

The flowthrough, expected to contain no target protein, was collected separately during column loading. Elution of bound proteins was performed using a sodium chloride gradient (depicted in light green), with the gradient extending to a final concentration of 1000 mM NaCl. Two distinct elution peaks were observed at approximately 460 mM and 950 mM NaCl. Fractions corresponding to these peaks were collected and analyzed by SDS-PAGE. As shown in Figure 68B, substantial amounts of digested SNAP25 were detected in fractions three to six, correlating with the second elution peak at approximately 270 mL (950 mM NaCl). However, Thrombin contamination was detected in fractions five and six. To prevent undesired proteolysis in downstream applications, these fractions were excluded from further use. Due to the presence of remaining impurities, an additional purification step was deemed necessary.

7.5.4 Isolation of ^{1}H-^{15}N SNAP25 by SEC

Using SEC, the necessary purity of the sample was achieved for NMR measurements. In parallel, the buffer was exchanged generating optimal buffer conditions for structural and dynamic measurements. In general, SEC is used to separate proteins by their hydrodynamic radius. Therefore, no elution buffer was required in this experiment.

Initiating isolation, the column was equilibrated with degassed and filtered SEC buffer containing 20 mM sodium phosphate (pH 6.5), 150 mM NaCl, 0.1 mM TCEP. After equilibration, the protein was loaded on a HiLoad Superdex 16/600 75 pg column.

As shown in Figure 69A, the chromatogram from the SEC run revealed a major peak eluting between 46 mL and 80 mL, followed by a smaller peak between 99 mL and 106 mL. Two-milliliter fractions were collected across both peaks and analyzed by SDS-PAGE (Figure 69C). SNAP25 was detected in fractions two to four, corresponding to the main chromatographic peak (Figure 69A), which exhibited absorbance values of up to 170 mAU. However, due to minor residual impurities observed in fraction four (Figure 69C), only fractions two and three - both displaying high purity - were pooled for subsequent experiments. In contrast, fractions five to eight contained proteins of lower molecular weight, as indicated by SDS–PAGE, and were therefore excluded from further analyses (Figure 69C).

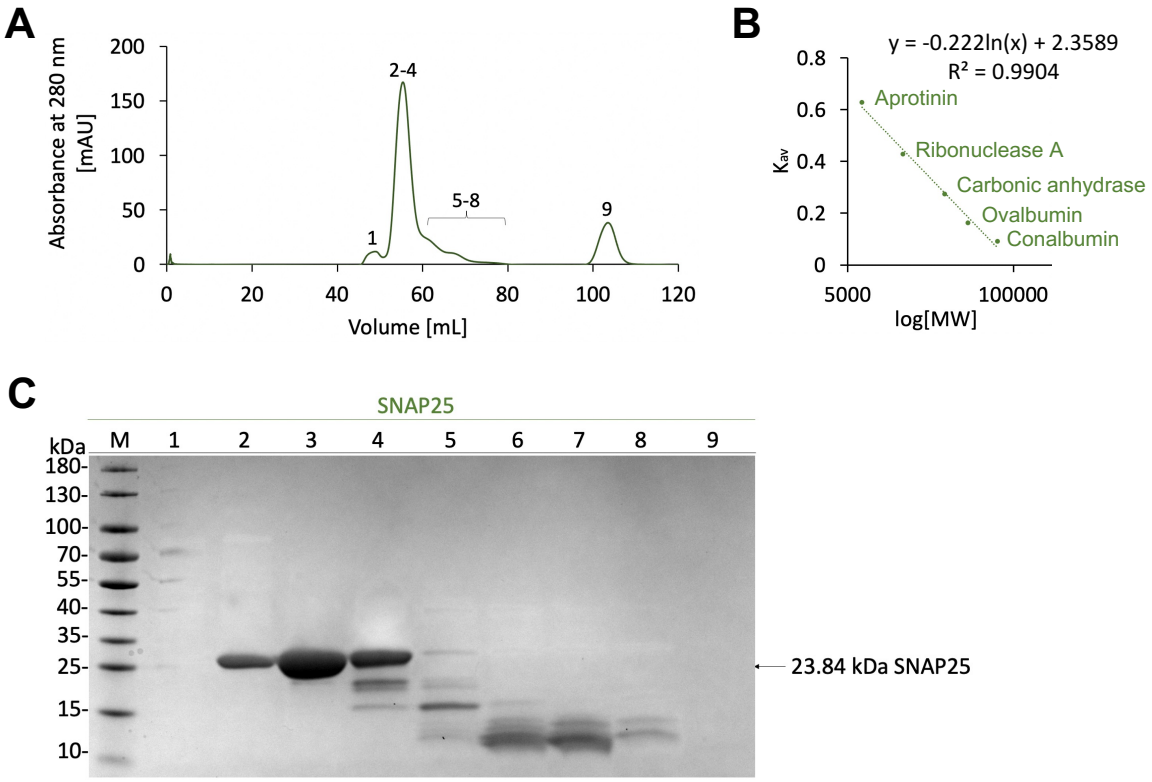


Figure 69: Size Exclusion Chromatography of ^{1}H-^{15}N SNAP25. For SEC, a HighLoad 16/600 Superdex 75 pg column (Cytiva, Marlborough, Massachusetts, USA) equilibrated in 20 mM sodium phosphate (pH 6.5), 150 mM NaCl, 0.1 mM TCEP was used. **(A)** Chromatogram of the isolation. The absorbance (mAU) is plotted against the volume (mL). **(B)** SEC calibration curve based on the following standard proteins: Aprotinin (6.5 kDa), Ribonuclease A (13.7 kDa), Carbonic Anhydrase (29 kDa), Ovalbumin (44 kDa) and Conalbumin (75 kDa). The calibration curve was plotted using the gel-phase distribution coefficient K_{av} vs. the logarithm of the molecular weight (MW) in g/mol. Based on the calibration curve, the calculated molecular weight of SNAP25 is 22.68 kDa (23.84 kDa expected) indicating monomeric SNAP25. **(C)** 15% (v/v) Tris-Glycine SDS-PAGE of collected SEC fractions (1-9). Bands correlating to SNAP25 (23.84 kDa) are indicated by an arrow.

However, as SDS causes the denaturation of proteins, the molecular weight of eluted protein was calculated based on a calibration curve (Figure 69B). The calibration curve was plotted using the gel-phase distribution coefficient K_{av} versus the logarithm of the molecular weight (log [MW]) (Figure 69B). K_{av} was calculated using the column void volume (V_o = 45.62 mL), the geometric column volume (V_c = 120 mL), and the elution volume (V_e = 55.34 mL) of SNAP25.

The following equation was used for calculation:

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

The SEC column was calibrated using conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa). The equation of the

calibration curve based on the data from molecular weight standards was used to calculate the experimental molecular weight. Based on these data, a molecular weight of 22.86 kDa was calculated along with the proposed molecular weight of 23.84 kDa. These results indicate that only monomers were in the sample. As the purity of pooled fractions was high (Figure 69C) and only monomers were observed, the pooled and concentrated SEC sample was used for later NMR measurements.

7.5.5 Expression of ^{2}H ^{15}N 6xHis-Syb-2 (1-96) in Minimal Medium

As demonstrated, a high protein yield and purity is required for high quality NMR measurements. Hence, the expression and isolation of Syb-2 (1-96) was optimized. For that, *E. coli* BL21 (DE3) cells containing pET-28a(+)-6xHis-Thrombin-Syb-2(1-96) were cultivated in 2xYT medium at 37 °C overnight. Main cultures were incubated at 37 °C until an OD_{600} of 0.8 was reached. After gene expression was induced, cells were incubated at 20 °C overnight. Samples before (BI) and after induction (AI) were taken (Figure 70A, B). Overexpression was confirmed by large bands on the Western Blot membrane at the proposed size of ~12 kDa.

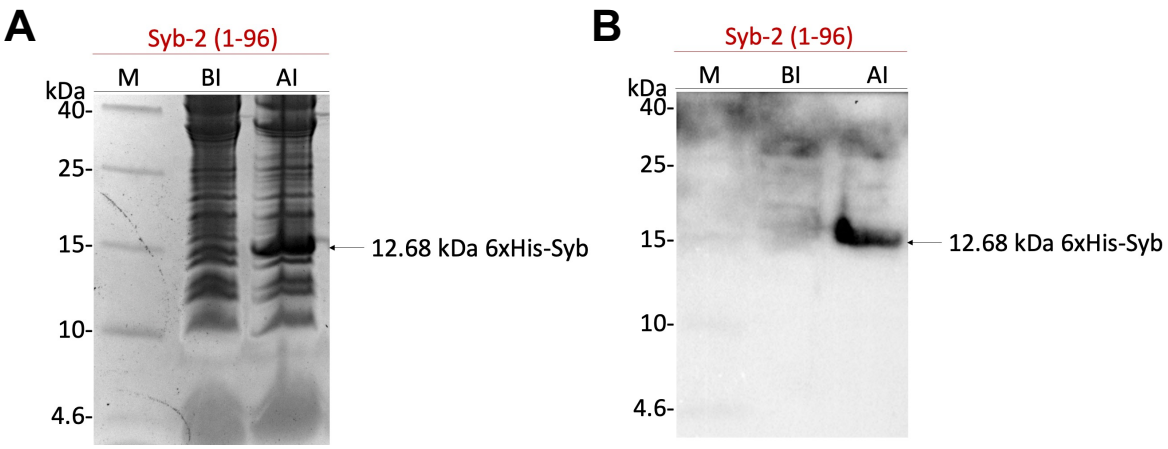


Figure 70: Expression of ^{2}H-^{15}N 6xHis-Thrombin-Syb-2 (1-96) in minimal medium. *E. coli* BL21 (DE3) cells incorporating the pET-28a(+)-6xHis-Thrombin-Syb-2(1-96) vector were cultivated in minimal medium at 20 °C overnight. **(A)** Samples before (BI) and after induction (AI) of expression were taken and analyzed by a 20% (v/v) Tris-Tricine SDS-PAGE. Protein bands were compared to the marker (M) (Spectra Multicolor Low Range Protein Ladder). **(B)** The His-tagged protein was identified by Western Blot analysis at the proposed size (12.68 kDa).

An overexpression band corresponding to 6xHis-Syb-2 (1–96) was observed at the expected molecular weight of approximately 12.68 kDa, as shown in Figure 70A. This result was further confirmed by Western blot analysis, where a prominent band appeared at ~15 kDa following immunodetection, consistent with the presence of His-tagged Syb-2 (1-96).

7.5.6 Isolation of ^{2}H ^{15}N 6xHis-Syb-2 (1-96) by IMAC

Following successful gene expression, labeled Syb-2 (1-96) was purified from *E. coli* host proteins using IMAC, followed by Thrombin cleavage combined with dialysis. After cell lysis via sonication, the clarified lysate was applied to a Ni-NTA column pre-equilibrated with lysis buffer containing 20 mM HEPES (pH 7.4), 500 mM NaCl, 8 mM imidazole, and 0.1 mM TCEP. The His-tagged Syb-2 (1-96) was expected to bind to the Ni-NTA resin, while *E. coli* proteins lacking a His-tag were eluted in the flowthrough, which showed elevated absorbance between 8 and 80 mL (Figure 71A). To remove nonspecifically bound proteins, the imidazole concentration was increased to 20 mM (4% of the elution buffer). Consistent with these conditions, SDS-PAGE analysis revealed several *E. coli* protein bands in the FT and the 4%B wash fraction (Figure 71B). The target protein was efficiently eluted at an imidazole concentration of approximately 230 mM (corresponding to 58% of the elution buffer), resulting in a prominent elution peak between 116 mL and 132 mL (Figure 71A).

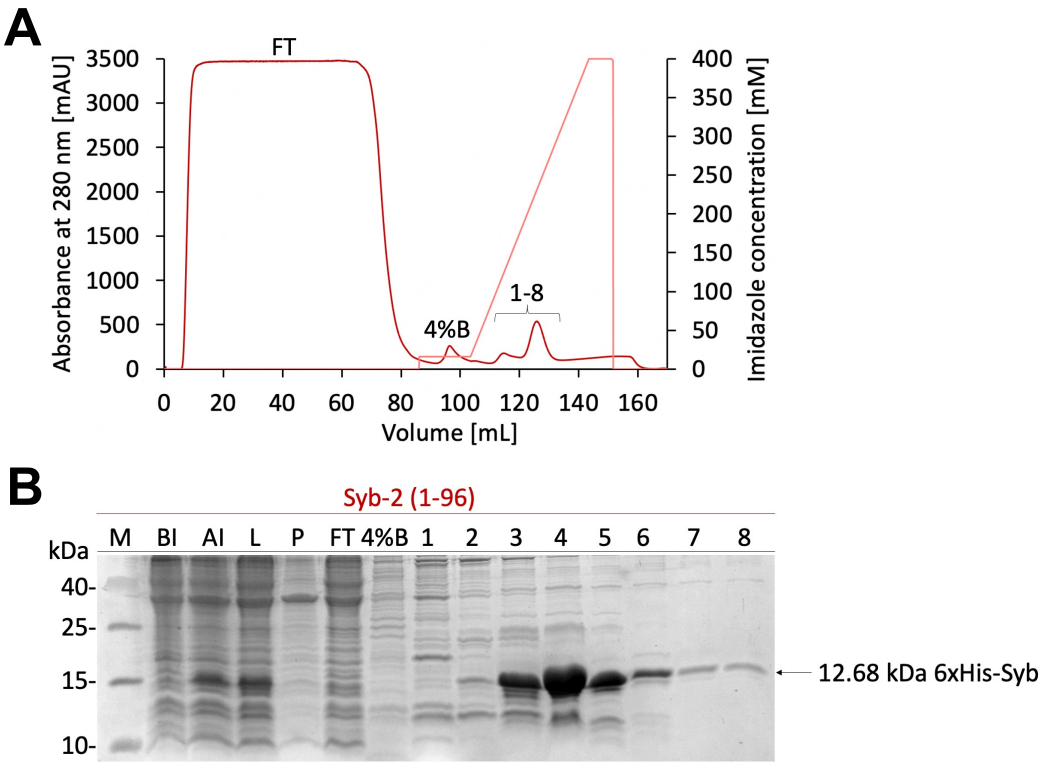


Figure 71: Purification of His-tagged ^{2}H-^{15}N Syb-2 (1-96) by immobilized metal affinity chromatography. The protein sample was loaded on a Ni-NTA column equilibrated in 20 mM HEPES (pH 7.4), 500 mM NaCl, 8 mM imidazole, 0.1 mM TCEP. The corresponding chromatogram is shown in **A**. The absorbance (dark red) measured at 280 nm was plotted against the volume in mL. Target protein was eluted using 20 mM HEPES (pH 7.4), 500 mM NaCl, 400 mM imidazole, 0.1 mM TCEP. **(B)** SDS-samples were taken for a 20% (v/v) Tris-Tricine SDS-PAGE. Samples before induction (BI), after induction (AI), lysate (L), pellet (P), flowthrough (FT), 4% elution buffer (4%B) and fractions (1-8) were taken and compared to the marker (M) (Spectra Multicolor Low Range Protein Ladder).

Two-milliliter fractions across this peak were collected and analyzed by Tris-Tricine SDS-PAGE. Fractions two through six contained Syb-2 (1-96); however, due to the presence of significant impurities in fraction two, only fractions three to six were pooled for subsequent steps. The pooled sample was dialyzed overnight in dialysis buffer containing 20 mM HEPES (pH 7.4), 100 mM NaCl, 0.1 mM TCEP, and 1 mM EDTA to gradually reduce the concentrations of sodium chloride and imidazole. Simultaneously, the His-tag was removed via overnight Thrombin digestion.

7.5.7 Isolation of ^{2}H ^{15}N Syb-2 (1-96) by IEC

IEC was employed to separate the digested Syb-2 (1–96) from undigested protein and residual protease. Based on pI of the cleaved Syb-2 (1-96) (pI = 8.28) and the pH of the dialysis buffer (pH = 7.4), the target protein is expected to carry a net positive charge under these conditions. Consequently, a cation exchange column (SP-Sepharose) was selected for purification.

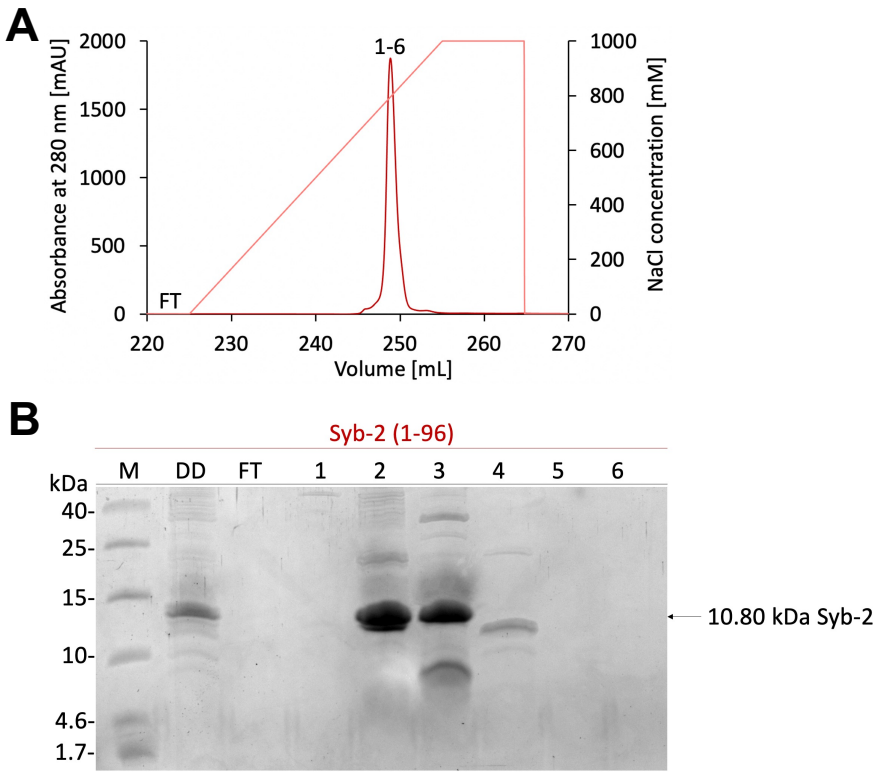


Figure 72: Isolation of ^{2}H-^{15}N Syb-2 (1-96) by ion exchange chromatography. The protein sample was loaded on a HiTrap SP-Sepharose cation exchange chromatography column equilibrated in 20 mM HEPES (pH 7.4), 0.1 mM TCEP, 1 mM EDTA. The corresponding chromatogram is shown in **A**. The absorbance (dark red) measured at 280 nm was plotted against the volume in mL. The target protein was eluted by a sodium chloride gradient using 20 mM HEPES (pH 7.4), 1000 mM NaCl, 0.1 mM TCEP, 1 mM EDTA. **(B)** SDS-samples were taken for a 20% (v/v) Tris-Tricine SDS-PAGE. Samples after dialysis and digestion (DD), flowthrough (FT) and fractions (1-6) were taken and compared to the marker (M) (Spectra Multicolor Low Range Protein Ladder). Digested Syb-2 (1-96) was expected at 10.80 kDa.

Protein elution was achieved using a linear sodium chloride gradient, with a final concentration of 1 M NaCl. Prior to IEC, the dialyzed sample was diluted in buffer without salt (20 mM HEPES, (pH 7.4), 0.1 mM TCEP, 1 mM EDTA) to reduce the NaCl concentration to approximately 10 mM, thereby enhancing protein binding to the column. Upon sample loading, the target protein was eluted using a sodium chloride gradient (depicted in light red, Figure 72A). A distinct elution peak was observed between 244 mL and 256 mL, corresponding to a salt concentration of approximately 790 mM (79% elution buffer). Two-milliliter fractions across this peak were collected and analyzed by SDS-PAGE (Figure 72B). As shown, high levels of purified Syb-2 (1-96) were detected in fractions two and three. However, due to remaining impurities, an additional purification step using SEC was necessary.

7.5.8 Isolation of ^{2}H ^{15}N Syb-2 (1-96) by SEC

This experiment was performed to achieve the final purity of protein samples for later NMR dynamic measurements. For that, the column was equilibrated with degassed and filtered 50 mM sodium phosphate (pH 6.0), 150 mM NaCl, 0.1 mM TCEP buffer. After equilibration, the protein was loaded on a HiLoad Superdex 16/600 200 pg column.

As displayed by Figure 73A, the chromatogram revealed a huge peak between 79 mL and 95 mL. Two-milliliter fractions of the peaks were taken and analyzed by a Tris-Tricine SDS-PAGE (Figure 73C). ^{2}H ^{15}N Syb-2 (1-96) was identified in fractions three to nine. However, because multiple impurities were present in the outer fractions (one to four and eight to nine), only fractions five to seven were combined to reduce the level of remaining impurities.

As described in previous sections, a calibration curve was used to estimate the molecular weight of proteins eluting at 90 mL. The calculation was based on the equation outlined in section 7.5.4. Specifically, the partition coefficient (K_{av}) was determined using the column's void volume (V_o = 40.75 mL), geometric column volume (V_c = 120 mL), and the elution volume of Syb-2 (V_e = 90.26 mL). From these parameters, a molecular weight of approximately 19.57 kDa was estimated. Although this value is higher than the theoretical molecular weight of cleaved Syb-2 (1-96) (10.80 kDa), the deviation is consistent with known effects of protein shape and potential dimerization or nonspecific interactions during SEC. The pooled and concentrated SEC fractions were subsequently used for NMR spectroscopy (Chapter 8 and 9).

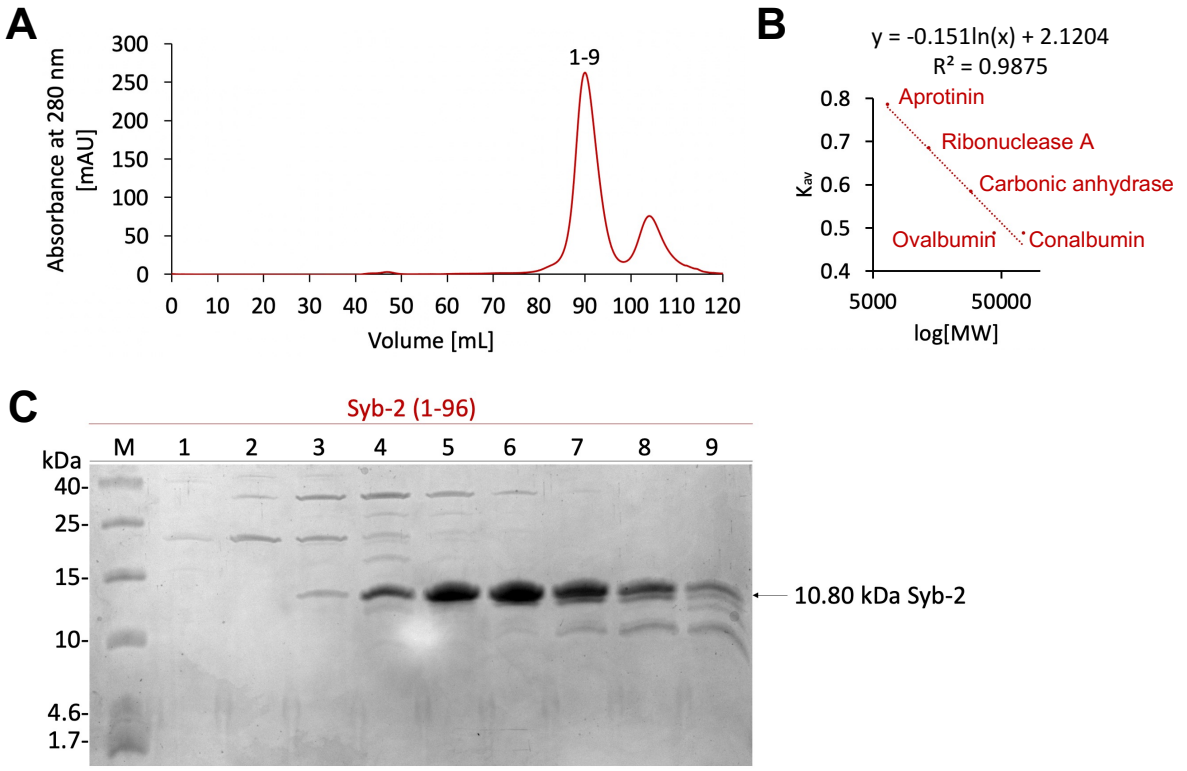


Figure 73: Size Exclusion Chromatography (SEC) of ^{2}H-^{15}N Syb-2 (1-96). For SEC, a HighLoad 16/600 Superdex 200 pg column (Cytiva, Marlborough, Massachusetts, USA) equilibrated in 50 mM MES (pH 6.0), 150 mM NaCl, 0.1 mM TCEP, 1 mM EDTA was used. **(A)** Chromatogram of the isolation. The absorbance (mAU) is plotted against the volume (mL). **(B)** SEC calibration curve based on the following standard proteins: Aprotinin (6.5 kDa), Ribonuclease A (13.7 kDa), Carbonic Anhydrase (29 kDa), Ovalbumin (44 kDa) and Conalbumin (75 kDa). The calibration curve was plotted using the gel-phase distribution coefficient K_{av} vs. the logarithm of the molecular weight (MW) in g/mol. Based on the calibration curve, the calculated molecular weight of Syb-2 (1-96) is 19.57 kDa (10.80 kDa expected) indicating monomeric Syb-2 (1-96). **(C)** 20% (v/v) Tris-Tricine SDS-PAGE of collected SEC fractions (1-9). Bands correlating to Syb-2 (1-96) (10.80 kDa) are indicated by an arrow.

7.6 Discussion

In this project, high-yield, high-purity Syb-2 (1–96) and SNAP25 samples were successfully produced and subsequently used for the NMR studies described in Chapters 8 and 9.

7.6.1 Expression and Isolation of the SNARE Protein SNAP25

Our initial approach focused on optimizing the expression and purification of SNAP25 following the expression and isolation of SNAP25 published by Pobatti *et al.* (2006) (further optimized in our lab).[278] Despite employing a cysteine-free SNAP25 mutant (all cysteines substituted with serines), 0.1 mM TCEP was included for NMR spectroscopy to ensure a reducing environment. Optimal expression of SNAP25 was achieved at 20 °C following induction with IPTG (Figure 66). As isotopic labeling required minimal medium, this medium

was subsequently used for large-scale protein production. SNAP25 was successfully expressed (sample AI) and remained soluble in the cell lysate (Figure 67B), as the target protein was not detected in the cell pellet after sonication. During IMAC, the protein was retained throughout the washing step with 5% (v/v) elution buffer, indicating minimal loss of the target protein at this stage (Figure 67B). The washing step effectively removed various *E. coli*-derived contaminants, thereby enhancing the purity of the His-tagged SNAP25 in the eluate. Strong absorbance signals and prominent gel bands at approximately 25 kDa confirmed the efficient recovery of His-tagged SNAP25 in the IMAC eluate (Figure 67). Subsequent purification via IEC successfully isolated digested SNAP25 (Figure 68). Fractions 5 and 6, which contained both the target protein and residual protease, were excluded from further experiments to avoid proteolytic degradation, despite the associated reduction in protein yield. SEC was then employed to separate SNAP25 from remaining *E. coli* proteins (Figure 69). High protein yields were obtained, as evidenced by the strong absorbance signals of the elution peak (Figure 33A). The final yields were sufficient to support downstream biophysical characterizations, including CD and NMR spectroscopy, as detailed in Chapters 8.

7.6.2 Expression and Isolation of the SNARE Protein Syb-2 (residues 1-96)

Although Syb-2 (1-96) lacks cysteine residues, a low concentration of the reducing agent TCEP (0.1 mM) was included during NMR measurements to maintain a reducing environment. Due to the absence of the transmembrane domain, Syb-2 (1-96) could be purified without the use of detergents such as CHAPS. The strategy was guided by a recent study published by Wittig *et al.*[279] Protocols were optimized in our lab. Consistent with previous reports on the expression of Syb-2 (1-96), high levels of the target protein were obtained following induction at 20 °C (Figure 70). The protein was effectively separated from *E. coli* contaminants using IMAC, as confirmed by the corresponding SDS-PAGE analysis (Figure 71B). The washing step removed non-specifically bound prokaryotic proteins, thereby increasing the purity of the sample prior to elution (Figure 71B). Successful thrombin cleavage was demonstrated by a shift in molecular weight observed in SDS-PAGE, from 12.68 kDa to 10.80 kDa, and corroborated by IEC data (Figure 72). During SEC, a sharp elution peak was observed, containing the target protein along with residual *E. coli* proteins (Figure 73). However, as the sample had already reached maximal achievable purity, only fractions predominantly containing the target protein were pooled for downstream biophysical analyses (Chapter 9).

8. Chapter: Intrinsic Disorder of the Neuronal SNARE Protein SNAP25a in its Pre-fusion Conformation

Journal: jmb

Volume: Volume 435, Issue 10, 15 May 2023, 168069

Authors: Tobias Stief,
Lothar Gremer,
Sonja Pribicevic,
Delane F. Espinueva,
Katharina Vormann,
Ralf Biehl, Reinhard Jahn,
Ángel Pérez-Lara,
Nils-Alexander Lakomek

DOI: 10.1016/j.jmb.2023.168069

Author contribution: Expression and purification of ^{15}N-^{13}C labeled SNAP25a,
Writing and proof-reading of manuscript

8.1 Abstract

The neuronal SNARE protein SNAP25a (isoform 2) forms part of the SNARE complex eliciting synaptic vesicle fusion during neuronal exocytosis. While the post-fusion cis-SNARE complex has been studied extensively, little is known about the pre-fusion conformation of SNAP25a. Here we analyze monomeric SNAP25a by NMR spectroscopy, further supported by small-angle X-ray scattering (SAXS) experiments. SAXS data indicate that monomeric SNAP25 is more compact than a Gaussian chain but still a random coil. We show that for monomeric SNAP25a, before SNAP25a interacts with its SNARE partners to drive membrane fusion, only the N-terminal part (region A5 to V36) of the first SNARE motif, SN1 (L11 - L81), is helical, comprising two α-helices (ranging from A5 to Q20 and S25 toV36). From E37 onwards, SNAP25a is mostly disordered and displays high internal flexibility, including the C-terminal part of SN1, almost the entire second SNARE motif (SN2, N144-A199), and the connecting loop region. Apart from the N-terminal helices, only the C-termini of both SN1 (E73 - K79) and SN2 (region T190 - A199), as well as two short regions in the connecting loop (D99 - K102 and E123 - M127) show a weak α-helical propensity (α-helical population < 25%). We speculate that the N-terminal helices (A5 to Q20 and S25 to V36) which constitute the N-terminus of SN1 act as a nucleation site for initiating SNARE zippering.

8.2 Introduction

SNARE proteins and their assembly into helical SNARE complexes drive membrane fusion during exocytosis and vesicle trafficking.[253,263] Because of their fundamental importance for brain function, neuronal SNARE complexes belong to the best-studied SNARE complexes[264,280] are the key engine eliciting synaptic vesicle membrane fusion with the plasma membrane, which propagates into fusion pore formation and subsequent neurotransmitter release into the synaptic cleft for signal transmission.[253,263]

The SNARE proteins SNAP25 (Synaptosomal associated protein of 25 kilodaltons) and syntaxin-1a reside at the plasma membrane, waiting for the vesicular SNARE protein, synaptobrevin-2, to initiate SNARE complex zippering. Formation of the SNARE complex will lead to the juxtaposition of the synaptic vesicle membrane and the plasma membrane (Figure 74), with synaptobrevin-2 anchored in the synaptic vesicle membrane and syntaxin-1a anchored in the plasma membrane, both via their C-terminal transmembrane domains.[253,263] The SNARE coiled-coil is formed by four heptad repeat motifs, called SNARE motifs, contributed by the individual SNARE proteins.4 Here, synaptobrevin-2 and syntaxin-1a contribute one SNARE motif each and SNAP25 two (Figure 74). The central layer of the helix bundle, also termed the “zero layer”[263,281] contains three Gln (Q) residues contributed by SNAP25 and syntaxin-1a, therefore dubbed Q-SNAREs, while synaptobrevin-2 provides an Arg (R) and is therefore named R-SNARE. It is widely accepted that the initial stages of the SNARE complex formation are templated by two accessory proteins, Munc13 and Munc18, interacting with syntaxin-1a.[253] However, the exact order of assembly is still under debate.[267]

SNAP25 exists in two isoforms, SNAP25b (isoform 1) and SNAP25a (isoform 2), with SNAP25a expressed in the embryonic brain while SNAP25b becomes the dominant form in the adult brain.[268,282] However, SNAP25a remains the predominant isoform in endocrine and neuroendocrine cells throughout life.[268] Mutations in the gene encoding SNAP25 are associated with epileptic encephalopathies[269] and neurodevelopmental and neurological disorders. [282,283]

Previous NMR studies on the monomeric forms of the SNAREs synaptobrevin-2 and syntaxin-1a provided structural insights into their pre-fusion conformation.[275,284,285,286,287] However, the

structure of SNAP25 in its monomeric pre-fusion conformation is still unclear. Early circular dichroism measurements detected intrinsic disorder and some helical content of SNAP25.[262] The intrinsic disorder of SNAP25 was later confirmed by EPR spectroscopy using several site-specific spin-labels.[288] However, detailed residue-specific insights into its secondary structure propensity have remained elusive.

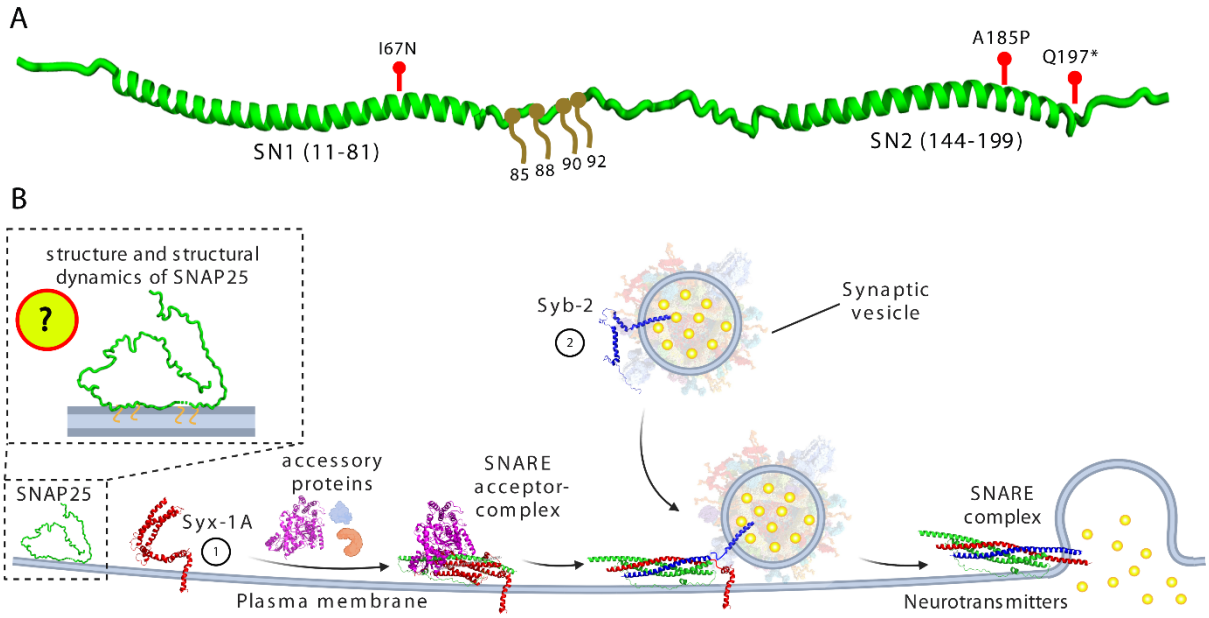


Figure 74: SNAP25. (A) SNAP25 domain organization showing the palmitoylation sites (in brown) and main disease mutations (in red) in humans. Its helical state is revealed in the SNARE complex. Mutations I67N and A185P lead to congenital myasthenic syndrome 18 (CMS18)[289], while Q197* leads to early infantile epileptic encephalopathy in humans (from UniProt database, https://www.uniprot.org/). **(B)** Scheme of SNARE assembly driving neuronal exocytosis. Note that the pre-fusion conformations of the SNARE proteins are not well understood. (Syx-1A, syntaxin-1A, modified from PDB ID: 1BR0 and 2M8R, and syb-2, synaptobrevin-2, modified from PDB ID: 2KOG. The accessory protein Munc18 appears in magenta, modified from PDB ID: 3C98, the SNARE acceptor complex is modified from the AlphaFold model AF-P60881-F1-model_v4 and PDB ID: 3C98, and the SNARE complex is modified from PDB ID: 3HD7).

Here, we investigate the secondary structure and structural dynamics of the monomeric pre-fusion conformation of SNAP25a (isoform 2 of SNAP25, from Rattus norvegicus, Uniprot accession number: P60881-2, see Figure S1 for sequence information) in a residue-specific manner by solution NMR spectroscopy, further supported by far-UV CD spectroscopy and SAXS. Importantly, our construct includes the SNAP25 loop region connecting the N-terminal SNARE motif, SN1 (region L11-L81), with the C-terminal SNARE motif, SN2 (region N144-A199). Wild-type SNAP25a comprises four cysteine residues, located in this loop region between SN1 and SN2 and adjacent to SN1. These four cysteines become posttranslational palmitoylated

and act as membrane anchor.[262,290] Although a structural study on fully palmitoylated SNAP25 would be highly desirable, it is hard to achieve yields compatible with structural investigations, particularly in the context of NMR isotope labeling. We, therefore, decided to use a variant in which the four cysteines are replaced by serines (C84S, C85S, C90S, C92S), thus lacking the palmitoyl residues.[288,289] For detailed sequence information on SNAP25a, see Figure S13. We find that SNAP25a is intrinsically disordered and highly dynamic, except of two adjacent N-terminal α-helices, which we interpret as a potential nucleation site for the initiation of SNARE zippering.

8.3 Results

8.3.1 NMR Chemical Shift Data Reveal Intrinsic Disorder of SNAP25a at a Residue-Specific Resolution

As a first inspection of the overall fold, we recorded far-UV-CD spectra of SNAP25a (Figure 75). CD spectra exhibit evident characteristics of a predominantly intrinsically disordered state through a prominent minimum at 203 nm. A second minimum is detected at 221 nm and can be interpreted as the presence of an α-helical portion to a lesser extent.

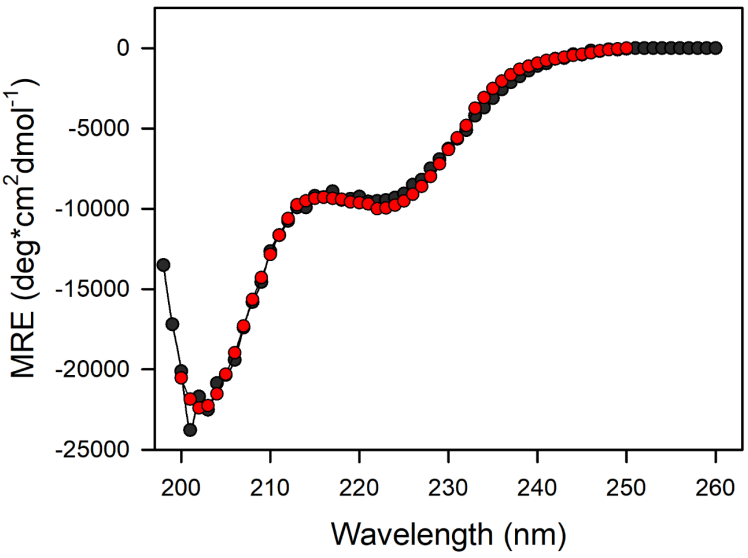


Figure 75: Far-UV CD spectrum of SNAP25a. Analysis of the data (black dots, experimental data; red dots, fitted data between 200 and 250 nm using BeStSel[103]) indicates the presence of a random-coil dominated secondary structure of SNAP25a, as evident from a strong minimum at 203 nm and the absence of any positive CD signal portions in the recorded data range down to 198 nm. MRE stands for mean residue ellipticity. The less pronounced minimum at 221 nm indicates the additional presence of a folded portion to a lesser extent. For the NMR analysis, first, we performed NMR spectral

backbone assignments of SNAP25a, employing three-dimensional NMR experiments described in the Materials and Methods section. Except for two proline residues (P91 and P117), complete backbone resonance assignments could be obtained (Table S16), which have been deposited to the Biological Magnetic Resonance Bank (BMRB ID 51598). A two-dimensional ^{1}H-^{15}N HSQC spectrum of SNAP25a, including assignments, is shown in Figures 3A and B. The narrow spectral dispersion in the proton dimension, with ^{1}H resonances ranging between 7.6 and 8.5 ppm, suggests an intrinsic disorder of SNAP25a. For a fully α-helical conformation, a slightly wider spectral dispersion (between about 7 and 9 ppm) would be expected. However, just from the ^{1}H dimension, it is difficult to distinguish between an α-helical or intrinsically disordered secondary structure. Therefore, we turned to Cα secondary chemical shifts, allowing for a much better distinction of secondary structure elements.

Although the buffer conditions at close to physiological salt concentrations (20 mM sodium phosphate buffer (pH 6.5), 150 mM NaCl, 0,1 mM TCEP) precluded CD data recording below 198 nm, and, therefore, did not allow unambiguous computational secondary structure calculations, analysis of the derived CD data (in the data range from 250 to 200 nm) fits 16.9% helical and at least 50.6% random-coil conformation, using the program BeStSel.[103] Those data agree with early CD data recorded on SNAP25[262], suggesting a large random-coil contribution and some helical content. Residue-specific insights into which parts are more structured and which are less have remained absent. We, therefore, turned to solution NMR spectroscopy which is well-suited to provide residue-specific insights into intrinsically disordered proteins.[291,292,293,294,295]

From the Cα chemical shifts, we calculated residue-specific secondary chemical shifts to obtain information on the secondary structure propensity of SNAP25a (see Materials and Methods section). Figure 76C shows the Cα secondary chemical shifts of the individual SNAP25a residues along the sequence (Table S17). The regions A5 to Q20 and S25 to V36 show strong α-helical propensity with positive secondary chemical shifts up to 2.5 ppm. They are connected by a short, slightly less structured, linker. From E37 to Q69, secondary chemical shifts indicate an intrinsic disorder. From E73 to K79, weak α-helical propensity marks the C-terminus of the N-terminal SNARE motif, SN1, according to the post-fusion structure of SNAP25 in the assembled SNARE complex. L11 marks the N-terminus of SN1 in the post-fusion structure. Therefore, the region A5 to V36 overlaps with the N-terminus of SN1. The α-helical propensity declines, however, well before the zero-layer positioned at Q53. Within the loop region (G82 to E143) that connects SN1 to the second SNARE motif, SN2, residues D99 to K102 and E123 to M127 show weak α-helical propensity. The adjacent second SNARE motif SN2 (N144 to A199) displays secondary chemical shifts close to zero ppm indicating intrinsic disorder. Only the C-terminus of SN2 (T190 to A199) shows weak α-helical propensity.

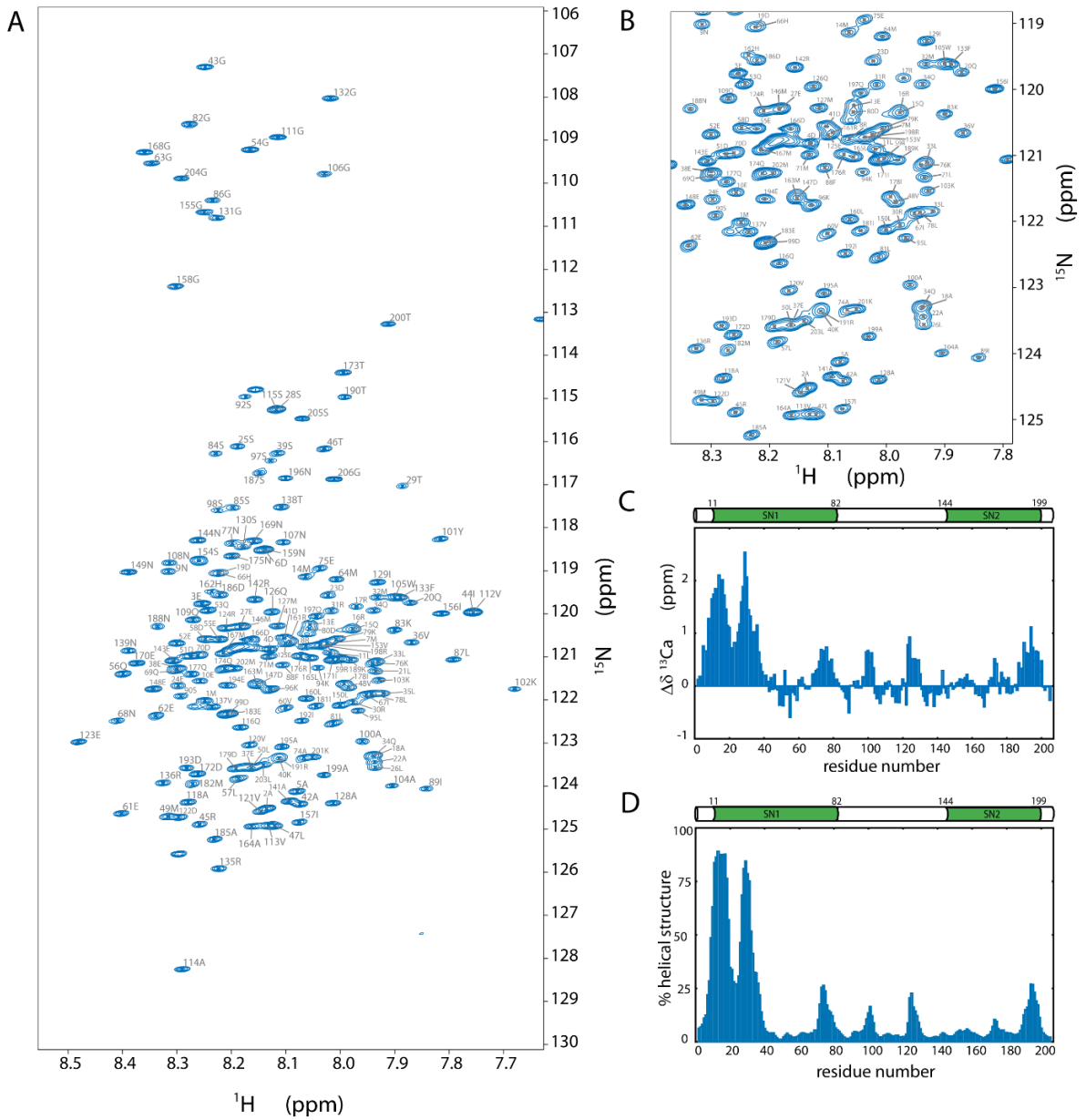


Figure 76: NMR spectroscopy of SNAP25. (A) ^{1}H-^{15}N HSQC spectrum of 150 µM ^{15}N^{13}C SNAP25a in 20 mM sodium phosphate buffer (pH 6.5), 150 mM NaCl, 0.1 mM TCEP, recorded at 900 MHz and 15 °C. **(B)** Zoom into the most crowded region of the ^{1}H-^{15}N HSQC spectrum of SNAP25a shown in panel a. **(C)** NMR Cα secondary chemical shift of SNAP25a plotted along the sequence. Positive values (> 0.5 ppm) indicate α-helical propensity (about 3-4 ppm stating full α-helical content). Negative values (< - 0.5 ppm) show β-sheet propensity. Values between -0.5 ppm and 0.5 ppm report random-coil characters, respectively. **(D)** Bar plot of population of local helical structure (in percent) of the SNAP25a (calculated by the Software CheSPI)[296]. Our data indicate that SNAP25a is primarily intrinsically disordered, apart from the N-terminal region A5 to V36, which shows a strong α-helical propensity and partially overlaps with the N-terminal SNARE motif SN1.

We also predicted the population of α-helical structure from the chemical shift values using the CheSPI software.[296] The α-helical population (in percent), shown in Figure 76D, follows the pattern of the Cα chemical shift values (see also Figure S14). We have tested different

computational disorder predictors (flDPnn[297], Metapredict[298], IDP-Seq2Seq[299], IUPred[300]) (Figure S15). While most of them correctly predict intrinsic disorder of SNAP25a, they fail to predict the N-terminal helices. One of them, Metapredict[298], predicts helical structure for both SN1 and SN2 and disorder for the loop region, therefore more similar to what is expected for SNAP25 in its post-fusion conformation. A similar model is predicted by AlphaFold2[301,302] (www.alphafold.ebi.ac.uk/entry/P60881). It has been discussed previously that caution is required when assessing AlphaFold2 predictions of IDP secondary structure.[303,304]

8.3.2 SAXS Analysis

SAXS data were measured for several concentrations of SNAP25a and corrected for background. Figure S16A shows SAXS data scaled by concentrations. Samples were checked for aggregation by dynamic light scattering (Zetasizer Nano, Malvern, U.S.A.), showing no significant aggregate formation. The hydrodynamic radius R_h extrapolated to zero concentration was R_h=4.46 nm. SAXS shows a typical pattern for Gaussian chains. The highest concentration was fitted by a wormlike chain model[305] showing a radius of gyration Rg=4.7 nm with a persistence length of 1.4 nm leading to a weaker power law decay at large Q (~ Q^{-d} with d<2). The compactness of the configuration can be examined looking at R_g/R_h=0.96. For an ideal Gaussian chain, we expect 0.66[306] while a sphere results in 1.29, respectively a globular protein like myoglobin in 1.26[307] The pair distance distribution, P(r), shows an end-to-end distance of ~12.5nm (Figure S16B). Consequently, the SNAP25 configuration is more compact than a Gaussian coil, but still a random chain as expected for intrinsically disordered proteins. The disordered myelin basic protein e.g., shows R_g/R_h=1.06.[308]

8.3.3 Structural Dynamics by NMR Relaxation Measurements

To characterize the structural dynamics of SNAP25, we turned to NMR relaxation measurements. Low heteronuclear $\{^1H\}$-^{15}N NOE data, ranging between 0.47 for M14 and -1.54 for G206, reveal high internal dynamics on the picosecond timescale (Figure 77A shows the $\{^1H\}$-^{15}N NOE values between -0.1 to 1, which are observed for the region M1-T200; for the full range including the highly dynamic C-terminus, see Figure S17A). These high internal dynamics are characteristic of an intrinsically disordered protein (IDP)[309] and are in line with the NMR secondary chemical shift data. Consistent with the $\{^1H\}$-^{15}N NOE values, ^{15}N R_1

relaxation rate constants further support the finding of high internal dynamics (Figure 77B). ^{15}N R_2 rate constants, derived from $R_{1\rho}$ measurements, show elevated rates between 8.0 s^{-1} and 10.5 s^{-1} for the region E10 to V36 (Figure 77C). In comparison, most remaining residues show an R_2 ($R_{1\rho}$) rate constant between 3 s^{-1} and 6 s^{-1}, apart from the C-terminus with values down to 0.8 s^{-1}. Low R_2 ($R_{1\rho}$) rates show high internal dynamics, consistent with the $\{^1H\}$-^{15}N NOE and ^{15}N R_1 relaxation rate constants, further supporting the finding of intrinsic disorder. The higher R_2 ($R_{1\rho}$) rates observed for the N-terminus of SN1 indicate the presence of a secondary structure, in line with the secondary chemical shifts indicating α-helical content. R_2 ($R_{1\rho}$) rate constants are sensitive to ps-ns dynamics, while NOE values are dominated by very fast ps dynamics (< 100 ps). To investigate the presence of conformational exchange, we recorded Hahn-echo-based ^{15}N R_2 rate constants at two magnetic fields, 14.1 T (corresponding to 600 MHz ^1H Larmor frequency) and 21.15 T (corresponding to 900 MHz ^1H Larmor frequency). While in the $R_{1\rho}$ experiments, with an applied spin-lock frequency of 2 kHz, all exchange processes slower than 80 µs have been refocused, and R_2($R_{1\rho}$) rate constants primarily report on fast ps-ns dynamics (faster than the rotational correlation time τ_c), these Hahn-echo-based R_2 measurements are sensitive to all conformational exchange processes slower than about 10 ms. The exchange processes scale with the square of the magnetic field and therefore have a more substantial impact on relaxation at higher fields. Figure 4D shows the difference between spin-echo-based R_2 rate constants measured at 900 MHz minus those measured at 600 MHz (for the individual rate constants, see Figure S17B and Table S18). For regions N9 to Q34, we observe differences higher than two s^{-1} for most resonances, while most residues for the remaining protein show differences lower than one s^{-1}. This suggests the presence of a conformational exchange process for the N-terminus of SN1, like, e.g., α-helix formation, which would be consistent with the high α-helical propensity revealed by the secondary chemical shifts data (Figure 77). SAXS data (see above) do not show indications of dimer formation of SNAP25a. We have further validated the monomeric character of SNAP25a by repeating the ^{15}N R_2 (R_{1rho}) measurements on a sample of lower concentration (70 µM vs. 150 µM). The ^{15}N R_2 (R_{1rho}) rate constants are sensitive to changes of the rotational correlation time τ_c, because of their $J(\omega_1) \approx J(0)$ dependence, with the spin-lock frequency ω_1 (which is 2 kHz in our case and small compared to the Larmor frequencies).

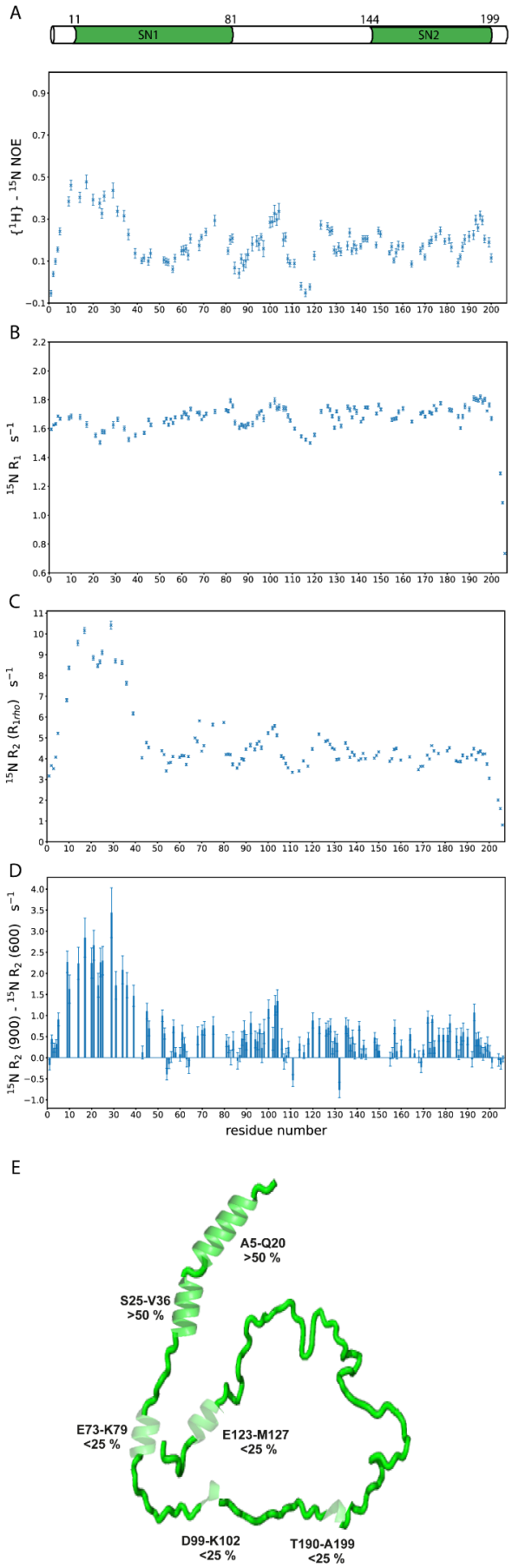


Figure 77: NMR relaxation data reveal high intrinsic dynamics of SNAP25a. (A) Low {^{1}H}-^{15}N NOE values indicate a high degree of internal dynamics on the fast ps timescale (for full range of data,

including the negative values of the C-terminus, see Figure S17A), which is also is in line with the **(B)** ^{15}N R_1 rate constants. **(C)** Low ^{15}N R_2 (R_{1rho}) rate constants underline the finding of a high degree of intrinsic dynamics; elevated ^{15}N R_2 rate constants for the N-terminal region A5 to V36 point to secondary structure. **(D)** Hahn-echo-based ^{15}N R_2 rate constants are sensitive to conformational exchange effects. Higher ^{15}N R_2 (900 MHz) minus R_2 (600 MHz) rate constants support a conformational exchange process for regions A5 to V36, such as, e.g., transient α-helix formation (for individual rate constants see Figure S17B). **(E)** Model of monomeric SNAP25a. The model was constructed manually (modified from PDB ID:3HD7), taking into account secondary structure information derived from secondary chemical shifts (see also Figures 76C and D) and using the program PyMOL (Schrödinger, Cambridge, MA). Figure 77E was created with BioRender.com.

^{15}N R_2 (R_{1rho}) rate constants of the 70 µM sample overlay well with the 150 µM sample and do not show a concentration dependence (Figure S18). We conclude that SNAP25 is monomeric at both concentrations.

8.3.4 Comparison with Post-Fusion Structure of SNAP25 in the Cis-SNARE Complex

Our Cα secondary chemical shift and NMR relaxation data reveal that SNAP25 is intrinsically disordered in its monomeric conformation, apart from two N-terminal helices A5-Q20 and S25-V36, with a-helical propensity > 75%. This region coincides with the N-terminus of SN1. We assign the N-terminal SNARE motif SN1 to residues L11 to L81. In the different X-ray structures of the post-fusion SNARE complex (PDB codes 1SFC, 1N7S, 3HD7)[264,280,310] the SN1 has fully α-helical conformation, with the N-terminus showing some variation regarding the beginning of the helix. The second SNARE motif SN2, which we assign to region N144-A199, is also fully helical in the post-fusion conformation of SNAP25 in the assembled cis-SNARE complex, but intrinsically disordered in the monomeric form; here only the C-terminus shows a weak α-helical propensity (< 25%). All X-ray structures (PDB codes 1SFC, 1N7S, 3HD7) were solved using two isolated constructs of SN1 and SN2, with the connecting loop region lacking. For a better visualization of the post-fusion conformation of SNAP25 its structure is highlighted in the context of the assembled (truncated) SNARE complex (Figure S19A, PDB code 1N7S). B-factors (or temperature factors, respectively) of the structure are color-coded onto the structure and plotted versus the residue number in Figure S19A. Towards the N- and C-termini of SN1 and SN2 B-factors increase, maybe not unexpectedly. For the first structure of the SNARE complex, 1SFC, which has been solved at lower resolution (2.4 Å compared to 1.45 Å for 1N7S), B-factors are overall higher, and also show an increase towards the N- and C-termini (Figure S19B). A similar effect is observed for 3HD7, the only structure of the SNARE complex that includes the syx-1a and syb-2 transmembrane regions and the only one which

has been solved in a lipid-mimicking detergent environment on a non-soluble construct (Figure S19B).

To summarize, we have studied the secondary structure and structural dynamics of SNAP25a using NMR chemical shifts and NMR relaxation data. Our data reveal that SNAP25a is intrinsically disordered in its monomeric pre-fusion form. Comparing NMR relaxation data to those obtained for synaptobrevin-2,[275] SNAP25a exhibits less intrinsic dynamics (and disorder) than synaptobrevin-2.

8.4 Discussion

NMR data on monomeric SNAP25a indicate that the second SNARE motif of SNAP25a, SN2, is intrinsically disordered, while the N-terminus of SN1 (up to residue V36) shows a strong α-helical propensity, with two α-helices A5-Q20 and S25-V36 connected by a short, less structured, linker. SAXS data indicate that SNAP25a is more compact than a Gaussian chain but still behaves like an intrinsically disordered protein. We speculate that the helical SNAP25a N-terminus may act as a nucleation site for N- to C-terminal zippering of the SNARE complex or upstream for the formation of an intermediate SNARE acceptor complex, presumably composed of Munc18:syntaxin-1a and SNAP25 (1:1:1).[311] Fasshauer and Margittai found previously, that N-terminal but not C-terminal truncations of SNAP25 prevented SNARE complex assembly.[272] Interestingly, while purified SNARE complexes containing a synaptobrevin-2 N-terminal (1-59) fragment are structured only N-terminally, purified SNARE complexes containing a synaptobrevin-2 C-terminal (60-96) fragment appear well-structured both for their N- and C-termini,[278] suggesting higher order of the syntaxin-1a and SNAP25 N-termini. This is in line with solution NMR ^{1}H-^{15}N HSQC spectra recorded on the homologous binary complex from yeast, composed of Sso1 (equivalent to Syntaxin-1a) and Sec9 (equivalent to SNAP25). Those spectra indicated that the C-terminus of Sso1 is less structured in the binary complex.[312] (As Sec9, similar to SNAP25, was not labeled with NMR-active isotopes, Sec9 was not directly observable in the NMR HSQC spectra.) Further, a SNARE complex containing a synaptobrevin-2 C-terminal fragment shows a more substantial decrease of the activation barrier for the assembly reaction than a SNARE complex containing a synaptobrevin-2 N-terminal fragment.[313] These results suggest that the N-termini of the SNAP25 and the syntaxin-1a SNARE motif exhibit a higher propensity to adopt an α-helical

structure than the SNAP25 C-termini, which would favor SNARE complex assembly via an N- to C-terminal zippering. This has been confirmed previously by NMR spectroscopy for the SNARE motif of syntaxin-1a in its pre-fusion conformation,[278] and we find this now for the N-terminus of SNAP25. Therefore, our data close the gap for SNAP25 and support a model of N-terminal nucleation and N- to C-terminal zippering for SNARE complex assembly.

8.5 Materials and Methods

8.5.1 NMR Sample Preparation

We studied the non-palmitoylated form of SNAP25a (isoform 2 of SNAP25 from *Rattus norvegicus,* Uniprot accession number: P60881-2, 206 amino acids, 23.3 kDa) with the four cysteine sites mutated to serines (C84S, C85S, C90S, C92S). The gene encoding SNAP25a from *Rattus norvegicus*, subcloned in a pET28a(+) vector with an N-terminal His_6-tag and Thrombin cleavage site, was heterologously expressed in BL21 (DE3) cells. For NMR sample preparation, cells were grown in M9 minimal media supplemented with ^{15}N enriched ammonium chloride (99%, Cambridge Isotope Laboratories, USA) and ^{13}C enriched Glucose (U-$^{13}C6$, 99%, Cambridge Isotope Laboratories, Tewksbury, MA, USA) as single sources for protein synthesis. Following an overnight preculture in 2xYT broth media (ThermoFisher, Waltham, MA, USA), cells were grown in M9 media for 4-5 h until an OD_{600} of 0.7 was reached, and protein expression was induced by the addition of 0.5 mM IPTG, and cells incubated over-night at 20 °C. SNAP25a was purified by Ni^{2+}-NTA affinity chromatography, followed by ion exchange chromatography (using a Resource Q column, Cytiva, Marlborough, MA, USA) and size exclusion chromatography (using a HiLoad 16/600 Superdex 75 pg column, Cytiva, USA), following the protocol described by Pobatti *et al*.[278] The final NMR sample contained 150 µM $^{15}N^{13}C$ SNAP25a in 20 mM sodium phosphate buffer at pH 6.5, 150 mM NaCl and 0.1 mM TCEP.

8.5.2 NMR Backbone Resonance Assignment

Spectral backbone resonance assignment was performed using a suite of three-dimensional HNCO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, and HN(CA)CO experiments. Spectra were referenced to DSS internally (0.2 µM), which was added to the NMR sample buffer. Measurements were conducted on a Bruker 750 MHz Avance III spectrometer equipped with a ^{1}H, ^{13}C, ^{15}N triple resonance Bruker TCI cryoprobe (Bruker, Billerica, MA, USA). The experimental temperature was 15 °C. Details on the experimental setup are summarized in

Table S13. Spectra were processed using the Bruker Topspin 3.5 and analyzed using the CCPN Analysis 3.0 software.[125] Secondary chemical shift data were derived from the backbone chemical shift assignments and random-coil chemical shifts predicted by the POTENCI software,[126] using the SNAP25 amino acid sequence and experimental buffer conditions as inputs. The secondary structure populations were predicted from chemical shift assignments using the CheSPI software.[296]

8.5.3 NMR Relaxation Data

^{15}N $R_{1,}$ ^{15}N $R_{1\rho}$ relaxation rate constants and heteronuclear {^{1}H}-^{15}N NOE data were recorded at 600 MHz using a Bruker Avance III HD spectrometer equipped with a ^{1}H, ^{13}C, ^{15}N triple resonance Bruker TCI cryoprobe, using the pulse programs described in [135]. Here, the pulse programs employing a sensitivity-enhanced HSQC read-out scheme were used. The experimental temperature was 15 °C. Rate constants and heteronuclear NOE data were evaluated using the NMRPipe software[129] and dedicated scripts as detailed in [135]; R_2 rate constants were derived from the $R_{1\rho}$ by correcting for R_1 contributions. Details on the experimental setup are summarized in Table S15. ^{15}N R_2 data, including potential R_{ex} contributions due to chemical or conformational exchange, were recorded using a Hahn-echo-based R_2 experiment; the pulse sequence scheme is shown in Figure S20. Conformational exchange processes result in a R_{ex} contribution to $R_{2,0}$ auto-relaxation, such that the Hahn-echo measured R_2 rate constants become $R_{2,app} = R_{2,0} + R_{ex}$. Data were recorded at 600 MHz and 900 MHz (using a Bruker 900 MHz Avance Neo Spectrometer equipped with a cryogenic ^{1}H, ^{13}C, ^{15}N TCI probe). As the R_{ex} contribution scales quadratically with the magnetic field, $R_{ex} \sim B^2$, the difference R_2 (900 MHz) minus R_2 (600 MHz) will indicate R_{ex} contributions due to conformational exchange.

8.5.4 CD Spectroscopy

Far-UV-CD spectroscopy of SNAP25a protein samples was performed on a JASCO J-815 CD spectropolarimeter (Jasco, Gross-Umstadt, Germany) using 1 mm path length quartz cuvettes (Helma, Müllheim, Germany). Typically, spectra were recorded at protein concentrations of ~6 to 12 µM SNAP25 in 20 mM sodium phosphate buffer, 150 mM NaCl, pH 6.5 at 20 °C with instrument settings as follows: 0.1 to 1 nm step size, 50 nm min^{-1} scan speed, 1 nm band with.

The signal-to-noise ratio was improved by the accumulation of 10 scans per sample. The mean residue ellipticity $[\theta]_{mrw}$ in deg·cm^2·dmol^{-1} was calculated from the equation:

$[\theta]_{\mathrm{mrw}} = (\theta_{\mathrm{obs}} \times \mathrm{MRW}) \,/\, (c \times d \times 10)$,

with θ_{obs}, observed ellipticity (in degrees); c, concentration (in g/ml); d, cell path length (in cm); MRW (mean residue weight), molecular weight divided by number of peptide bonds. Secondary structure analysis by BeStSel[103] (https://BeStSel.elte.hu/) was applied on CD data recorded at 11.55 µM SNAP25 in the 200 to 250 nm range with a scale factor of 1.0.

8.5.5 SAXS Analysis

SAXS experiments were performed the Jülich Center for Neutron Science (JCNS) at Forschungszentrum Jülich, Germany. The instrument "Ganesha-Air" from SAXSLAB/XENOCS was used. The X-ray source of the laboratory-based "Ganesha-Air" system is a D2-MetalJet (Excillum) with a liquid metal anode operating at 70 kV and 3.57 mA with Ga− Kα radiation (wavelength λ = 0.13414 nm). The data were acquired with a position-sensitive detector (PILATUS 300 K, Dectris). After calibration with silver behenate, the distance from the sample to the detector was set to cover a Q-range 0.1 – 6 nm^{-1}. Data analysis was done using the Python-based project Jscatter.[133] The pair distance distribution function, P(r) was calculated from the form factor with concentration 8 mg/ml, using the DATGNOM4 tool of the ATSAS software package.[314]

8.6 Accession Numbers

NMR backbone chemical shift assignments (^{1}H, ^{15}N, ^{13}C', and ^{13}Cα and ^{13}Cβ) for SNAP25a have been deposited to the BMRB with accession code 51598 and provided as supplementary Table S16.

8.7 CRediT Authorship Contribution Statement

Tobias Stief: Investigation, Methodology, Formal analysis, Resources, Writing. Lothar Gremer: Investigation, Formal analysis, Writing. Sonja Pribicevic: Resources, Writing. Delane F. Espinueva: Resources. **Katharina Vormann:** Resources. Ralf Biehl: Investigation, Formal Analysis, Writing. Reinhard Jahn: Conceptualization, Writing. Angel Perez-Lara:

Conceptualization, Writing. Nils-Alexander Lakomek: Conceptualization, Methodology, Writing, Supervision, Funding acquisition.

8.8 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8.9 Acknowledgments

A.P.-L. acknowledges funding by the Ramon y Cajal grant (RYC2018-023837-I) and R.J. financial support by the National Institutes of Health (NIH grant number P01 GM072694). N.L. acknowledges the German Science Foundation for funding through the Heisenberg Program (DFG grant number 433700474) as well as access to the Jülich-Düsseldorf Biomolecular NMR Center, jointly run by Forschungszentrum Jülich and Heinrich Heine University Düsseldorf (HHU).

8.10 Supplementary Information

Table S13: Experimental parameters of three-dimensional NMR backbone assignment experiments.

	HNCO	HNCA	HN(CO)CA	HNCACB	HN(CO)CACB	HN(CA)CO
Field / T	17.6	17.6	17.6	17.6	17.6	17.6
^{1}H Larmor frequency/ MHz	750	750	750	750	750	750
Bruker pulse program	hncogp3d	hncagp3d	hncocagp3d	hncacbgp3d	hncocacbgp3d	hncacogp3d
Interscan recovery delay / s	1	1	1	1	1	1
Number of scans	16	16	16	16	16	16
Total experimental time	2d	2d	2d	5d	5d	2d
Direct dimension (^{1}H)						
^{1}H carrier/ppm	4.70	4.70	4.70	4.70	4.70	4.70
Spectral width/ppm	16.02	16.02	16.02	16.02	16.02	16.02
Data points (TD)	2048	2048	2048	2048	2048	2048
Time/ms	85.2	85.2	85.2	85.2	85.2	85.2
Indirect dimension (^{15}N)						
^{15}N carrier/ppm	117.00	117.00	117.00	117.00	117.00	117.00

Spectral width/ppm	35.00	35.00	35.00	35.00	35.00	35.00
Data points (TD)	100	100	100	100	100	100
Acquisition time/ms	18.8	18.8	18.8	18.8	18.8	18.8
Indirect dimension (^{13}C)						
^{13}C carrier/ppm	173.50	53.20	53.20	43.00	43.00	173.50
Spectral width/ppm	10.00	31.18	31.18	70.68	70.68	10.00
Data points (TD)	90	90	90	220	220	90
Acquisition time/ms	23.9	7.7	7.7	8.3	8.3	23.9
Assignment statistics						
# visible resonances	204	204	204	204	204	204
# assigned resonances	204	204	204	204	204	204

Table S14: Experimental parameters of two-dimensional ^{1}H-^{15}N HSQC experiments.

	HSQC	**HSQC**	**HSQC**
Field / T	14.1	17.6	21.15 T
^{1}H Larmor frequency / MHz	600	750	900
Bruker pulse program	fhsqcf3gpph	fhsqcf3gpph	fhsqcf3gpph
Interscan recovery delay / s	1	1	2
Number of scans	8	8	8
Total experimental time	80 min	80min	145 min
Direct dimension (^{1}H)			
^{1}H carrier/ppm	4.70	4.694	4.70
Spectral width/ppm	16.03	16.02	16.34
Data points	2048	2048	2048
Acquisition time/ms	106.6	85.2	69.7
Indirect dimension (^{15}N)			
^{15}N carrier/ppm	117.00	117.00	117.00
Spectral width/ppm	35.00	35.00	35.00
Data points	512	512	512
Acquisition time /ms	120.3	96.0	80.1

Table S15: Experimental parameters of ^{15}N NMR relaxation experiments.

	R1	**R1rho**	**{1H} -15N NOE**	**R2 (Hahn-echo)**	**R2 (Hahn echo)**
Field / T	14.1	14.1	14.1	14.1	21.15
1H Larmor frequency / MHz	600	600	600	600	900
pulse program	Lakomek *et al.,* JBNMR 2012	Lakomek *et al.,* JBNMR 2012	Lakomek *et al.,* JBNMR 2012	see Figure S17	see Figure S17
Interscan recovery delay / s	2	2	2	2	3

Number of scans	8	8	64	16	16
Total experimental time	24h	21h	5d 13h	10.5h	15h
Direct dimension ^{1}H)					
^{1}H carrier/ppm	4.700	4.696	4.700	4.700	4.700
Spectral width/ppm	16.03	16.03	16.03	16.03	16.03
Data points	2048	2048	2048	2048	2048
Acquisition time	106.5	106.5	106.5	106.5	106.5
Indirect dimension (^{15}N)					
^{15}N carrier/ppm	117.00	117.00	117.00	117.00	117.00
Spectral width/ppm	29.91	35.00	35.00	29.91	30.00
Data points (total)	4096	4096	1024	1024	1024
Acquisition time/ms	140.8	120.3	120.3	140.8	93.6
Relaxation period					
# relaxation delays	8	8	2	2	2
Time points (ms)	0, 720, 320, 480, 80, 640, 160, 560	5,85, 15, 65, 35, 25, 45, 75	0, 5000	1, 111	1, 111

10 20 30 40 50 60
MAEDADMRNE LEEMQRRADQ LADESLESTR RMLQLVEESK DAGIRTLVML DEQGEQLDRV

70 80 90 100 110 120
EEGMNHINQD MKEAEKNLKD LGKCCGLFIC PCNKLKSSDA YKKAWGNNQD GVVASQPARV

130 140 150 160 170 180
VDEREQMAIS GGFIRRVTND ARENEMDENL EQVSGIIGNL RHMALDMGNE IDTQNRQIDR

190 200
IMEKADSNKT RIDEANQRAT KMLGSG

Figure S13: The sequence of SNAP25a (isoform 2). Organism: *Rattus norvegicus,* Uniprot accession number: P60881-2, 206 amino acids, 23.3 kDa. The two SNARE motifs, SN1 (L11 to L81) and SN2 (N1444 to A199), as observed in the post-fusion cis-SNARE complexe,[280,264] are color-coded in green. Backbone chemical shift assignments have been submitted to the Biological Magnetic Resonance Bank (BMRB ID 51598).

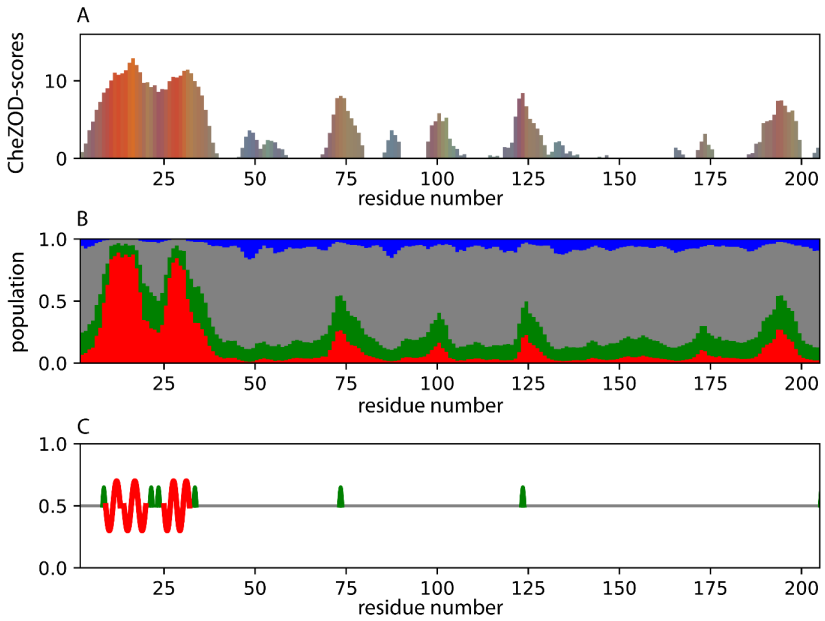


Figure S14: Output multi-panel plot generated by the CheSPI software[126] using the obtained NMR chemical shifts as input. (A) Bar plot with bar height corresponding to the so-called CheZOD Z-scores. These scores are translated to a color scheme in which blue colors stand for sheets, red for helices, and green colors for turns, while disordered states with principal components close to zero are represented in grey. **(B)** Stacked bar plot of populations of "extended" (blue), "helical" (red), "turn" (green), and "non-folded" (grey) local structures (corresponding to DSSP classes), as derived by the CheSPI software. **(C)** Representation of the most likely predictions for secondary structure separated into eight DSSP classes : α-helix (curved red lines), turns (green arcs), and coil (grey lines).[126]

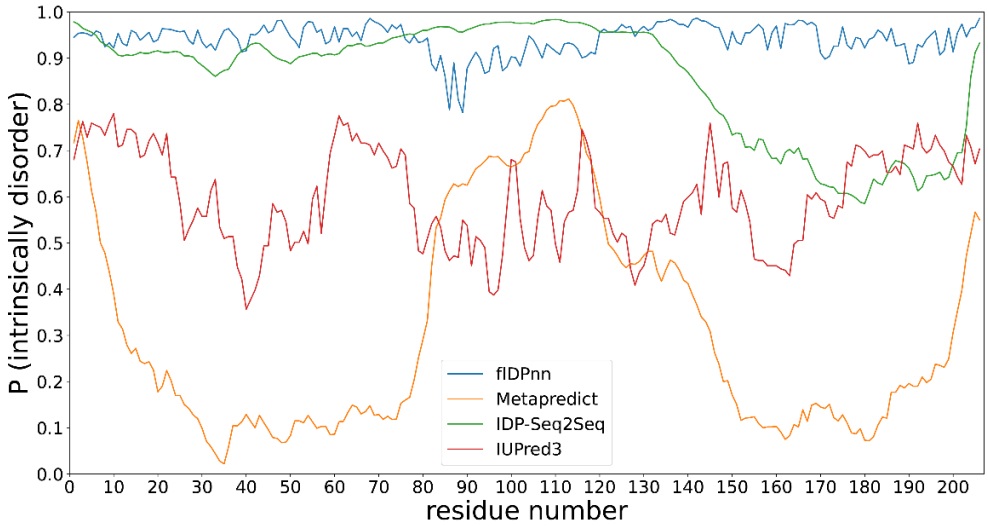


Figure S15: Computational disorder predictions of SNAP25a using different disorder predictors. A score of 1 means 100% probability to be intrinsically disordered, a score of 0 means 0%. Disorder prediction by four different disorder predictors flDPnn,[297] Metapredict,[298] IDP-Seq2Seq,[299] IUPred3[300] are shown.

Table S16: Backbone chemical shift assignments of SNAP25a. See the accompanied Excel table „SNAP25a_backbone_assignment". The columns contain the following information for each residue (rows): # (residue number), amino acid residue type (1-letter code), δH(N) chemical shift / ppm, δ(N) / ppm, δ(C') / ppm, δ(Cα) / ppm, δ(Cβ) / ppm.

Table S17: Secondary chemical shift data of SNAP25a. See the accompanied Excel table „SNAP25_secondary_chemical_shifts_POTENCI". Secondary chemical shift data have been calculated from the δ(Cα) and δ(Cβ) chemical shifts assignments minus random coil chemical shifts predicted from the amino acid sequence of SNAP25a and the experimental buffer conditions using the POTENCI software.[126] The columns contain the following information for each residue (rows): # (residue number), amino acid residue type (1-letter code), secondary chemical shifts ΔδCα / ppm, ΔδCβ / ppm, ΔδCα - ΔδCβ / ppm.

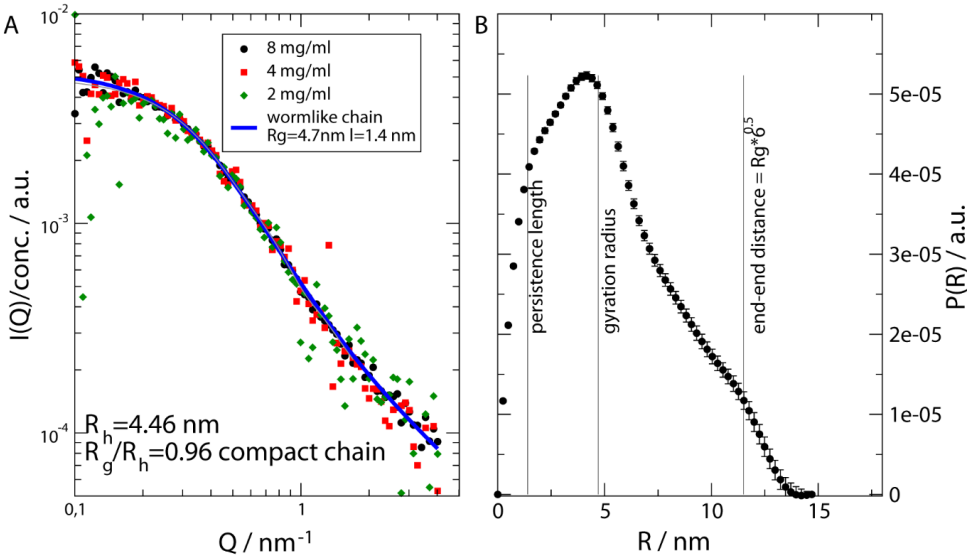


Figure S16: **SAXS analysis of SNAP25a. (A)** Scattering intensity I(Q) of SNAP25a as a function of the scattering vector Q. The scattering intensity is scaled by concentration after background subtraction. **(B)** Pair distance distribution function, P(r), of SNAP25a. The first shoulder indicates the persistence length, and the maximum corresponds to the radius of gyration. The end-to-end distance results in the shoulder around 12.5 nm as the maximum extension of the chain is a bit larger than $R_e=6^{0.5}R_g$ of a Gaussian chain.

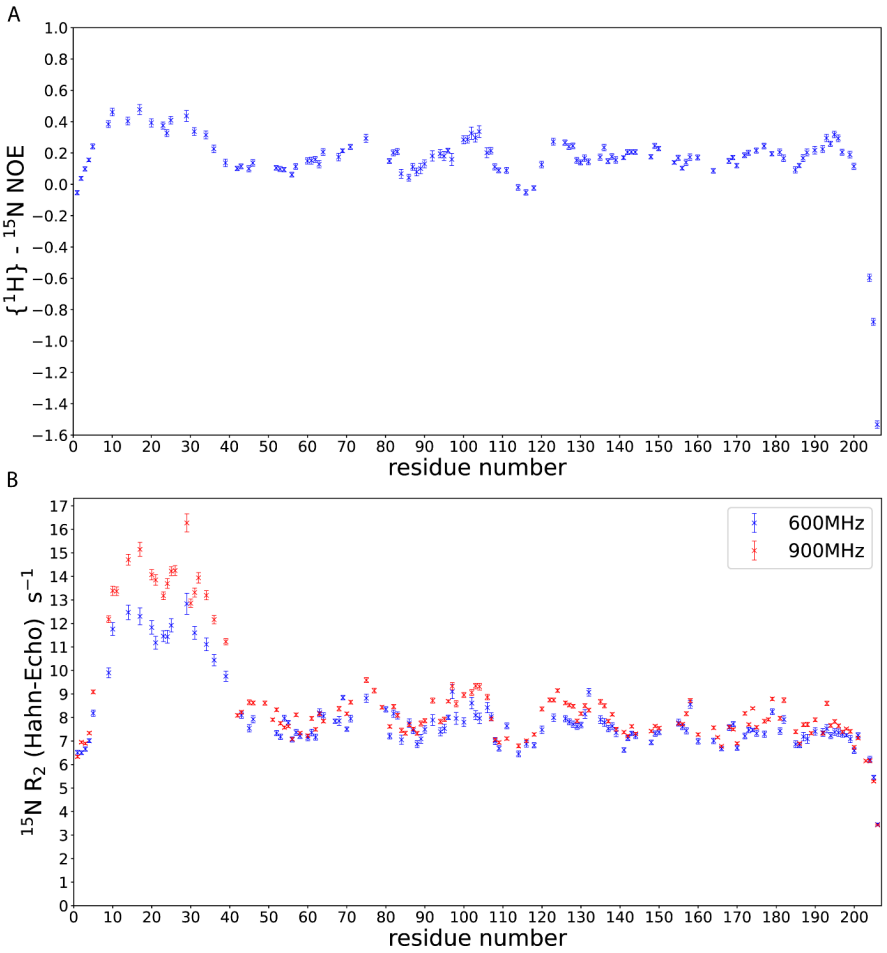


Figure S17: Additions on the NMR relaxation data on SNAP25 shown in Figure 4 of the main manuscript. (A) $\{^1H\}$-^{15}N NOE values indicate a high degree of internal dynamics on the fast ps timescale. **(B)** Individual ^{15}N R_2 (900 MHz) and R_2 (600 MHz) rate constants. In Figure 4B their difference is shown.

Table S18: NMR relaxation data. See the accompanied Excel table „SNAP25a_relaxation_data“. The columns contain the following information for each residue (rows): # (residue number), AA type (1-letter code), longitudinal relaxation rate constants ^{15}N R_1 / s^{-1} (600 MHz), experimental error $\delta(^{15}N\ R_1)$ / s^{-1} (600 MHz); transverse relaxation rate constants (derived from $R_{1\rho}$) ^{15}N $R_2(R_{1\rho})$ / s^{-1} (600 MHz), experimental error $\delta(^{15}N\ R_2(R_{1\rho}))$ / s^{-1} (600 MHz); heteronuclear NOE $\{^1H\}$-^{15}N values (I / I_0) (600 MHz), experimental error $\delta(I / I_0)$ (600 MHz); transverse relaxation rate constants (600 MHz) (derived from Hahn-echo experiment) ^{15}N R_2 / s^{-1} (600 MHz), experimental error $\delta(^{15}N\ R_2)$ / s^{-1} (600 MHz), ^{15}N R_2 / s^{-1} (900 MHz), $\delta(^{15}N\ R_2)$ / s^{-1} (900 MHz), difference ^{15}N R_2 / s^{-1} (900 MHz) minus ^{15}N R_2 / s^{-1} (600 MHz), propagated experimental error δ(R_2 / s^{-1} (900 MHz) - R_2 / s^{-1} (600 MHz)).

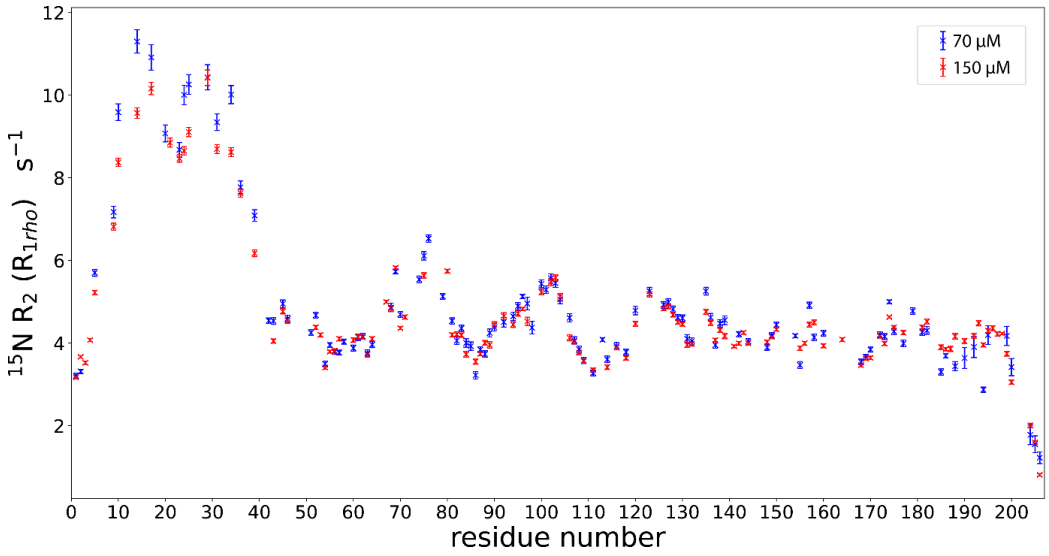


Figure S18: Comparison of ^{15}N R_2 ($R_{1\rho}$) rate constants of two SNAP25a samples at different concentrations of 150 µM and 70 µM.

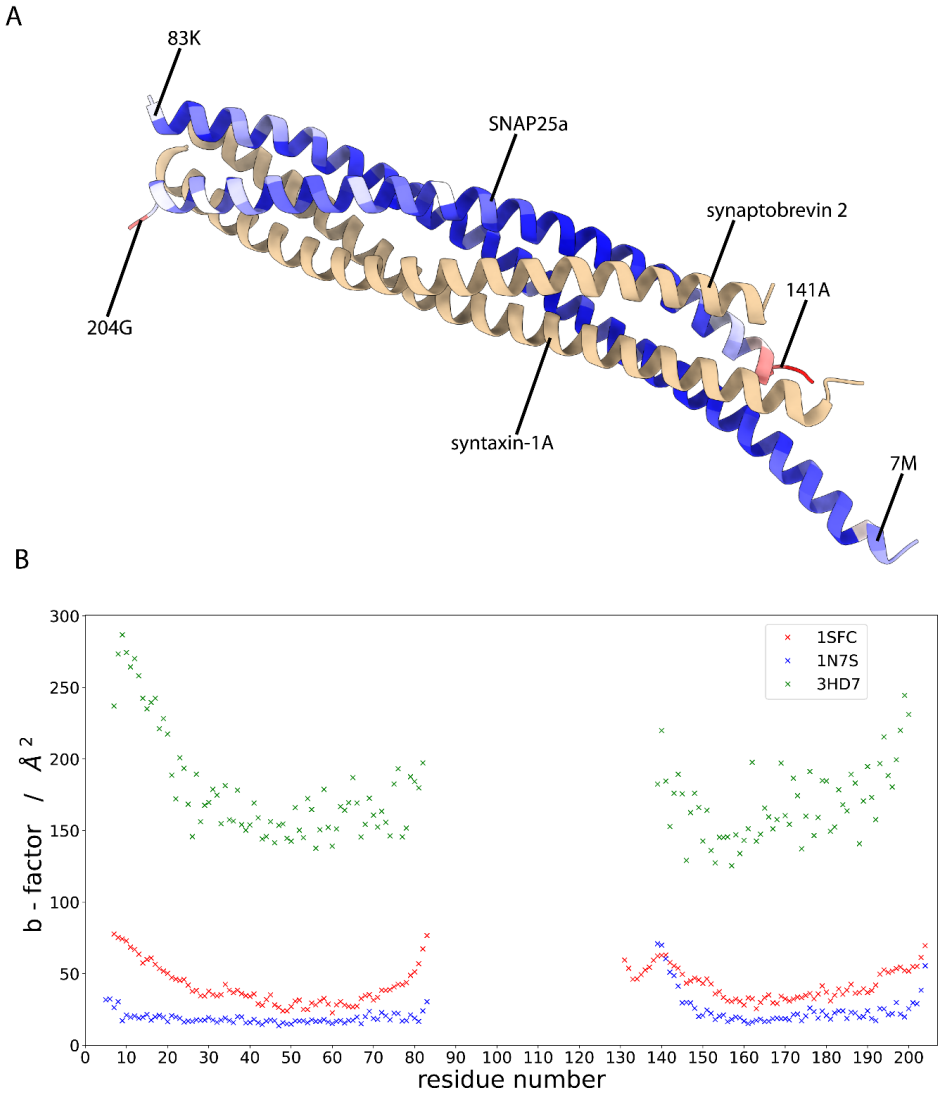


Figure S19: Post-fusion cis-SNARE complex. (A) Post-fusion structure of SNAP25a in the assembled cis-SNARE complex composed of synaptobrevin-2, syntaxin-1a and SNAP25 (PDB code 1N7S). Synataxin-1a and synaptobrevin-2 contribute one helix each and SNAP25a two helices (SN1) and (SN2).

B-factors/temperature factors are plotted onto the structure, blue indicates a low B-factor, white medium and red a high B-factor**. (B)** Comparison of B-factors of SNAP25a within three different X-ray structures (PDB codes: 1SFC, 1N7S, 3HD7) of the post-fusion cis-SNARE complex.

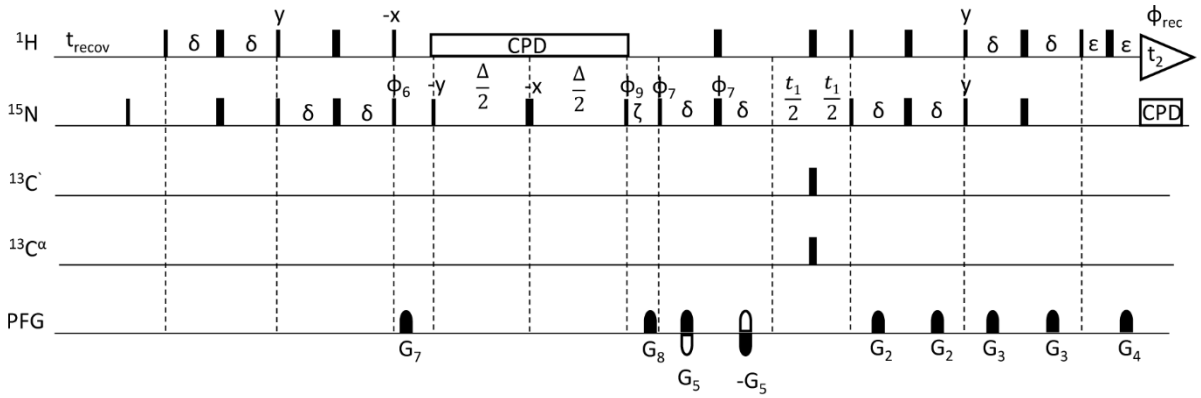


Figure S20: Pulse sequence scheme of the used Hahn-echo based ^{15}N R_2 experiment. 90° hard pulses with phase x if not marked otherwise are illustrated as narrow rectangular and 180° hard pulses as broad rectangular. The durations of the implemented delays are δ = 2.65ms and ϵ is equivalent to the duration of the decoding gradient G4 (201μs). With this offset it is possible to insert the decoding gradient without producing a linear phase error in proton dimension. Adjusting the delay ζ can improve the overall water-suppression; we used a delay of 20ms. The variable delay Δ corresponds to the relaxation period. The Gradients G2 (2.65 ms, -2.7 G/cm), G3 (2.65 ms, -0.3 G/cm) are rectangular shaped. Gradients G4 (201 μs, 21.4 G/cm), G5 (1 ms, 21.4 G/cm), G 7 (200 μs; -26.8 G/cm) and G8 (1 ms, 5.4 G/cm) are sine-bell shaped. Phase cycling: ϕ_6= y, y, -y, -y; ϕ_7= y, -y, ϕ_9= 4(-y), 4(y), ϕ_{rec}= y, -y, -y, y, -y, y, y, -y. The quadrature detection is implemented with the polarity inversion of the encoding gradient G5 and the phase cycling of ϕ_7.[315] For the decoupling in the proton dimension a waltz16 decoupling scheme was used. In the ^{15}N dimension a garp decoupling scheme was used.

9. Chapter: Sensitivity-Enhanced NMR ^{15}N R_1 and $R_{1\rho}$ Relaxation Experiments for the Investigation of Intrinsically Disordered Proteins at High Magnetic Fields

Journal: Methods

Volume: Volume 4233, March 2024, Pages 1-15

Authors: Tobias Stief,
Katharina Vormann,
Nils-Alexander Lakomek

DOI: 10.1016/j.ymeth.2024.01.008

Author contribution: Expression and purification of ^{2}H-^{15}N-^{13}C labeled Syb-2 (1-96),
Expression and purification of ^{2}H-^{15}N labeled Syb-2 (1-96),
Expression and purification of ^{1}H-^{15}N labeled Syb-2 (1-96),
Writing and proof-reading of manuscript

9.1 Abstract

NMR relaxation experiments provide residue-specific insights into the structural dynamics of proteins. Here, we present an optimized set of sensitivity-enhanced ^{15}N R_1 and $R_{1\rho}$ relaxation experiments applicable to fully protonated proteins. The NMR pulse sequences are conceptually similar to the set of TROSY-based sequences and their HSQC counterpart (Lakomek *et al*., J. Biomol. NMR 2012). Instead of the TROSY read-out scheme, a sensitivity-enhanced HSQC read-out scheme is used, with improved and easier optimized water suppression. The presented pulse sequences are applied on the cytoplasmic domain of the SNARE protein Synpatobrevin-2 (Syb-2), which is intrinsically disordered in its monomeric pre-fusion state. A two-fold increase in the obtained signal-to-noise ratio is observed for this intrinsically disordered protein, therefore offering a four-fold reduction of measurement time compared to the TROSY-detected version. The inter-scan recovery delay can be shortened to two seconds. Pulse sequences were tested at 600 MHz and 1200 MHz ^{1}H Larmor frequency, thus applicable over a wide magnetic field range. A comparison between protonated and deuterated protein samples reveals high agreement, indicating that reliable ^{15}N R_1 and $R_{1\rho}$ rate constants can be extracted for fully protonated and deuterated samples. The presented pulse sequences will benefit not only for IDPs but also for an entire range of low and medium-sized proteins.

9.2 Introduction

Intrinsically disordered proteins (IDPs), or proteins with intrinsically disordered regions (IDRs), compose about 30% of the human proteome and fulfill essential functions in cellular regulation and signaling.[295,309,316,317] They are characterized by high disorder and increased structural dynamics, resulting in high local mobility and flexibility, allowing them to fold and adapt to various binding partners.[292,318,319] Recently, IDPs have received increased attention as they are involved in organizing membrane-less organelles.[146,320,321,322,323,324] Misfolding of IDPs can lead to several diseases, including neurodegenerative diseases.[325,326,327,328] Characterizing the structural dynamics of intrinsically disordered proteins is critical to a deeper understanding of their function, conformational space, and binding interactions.[329,330] The high internal dynamics of IDPs pose challenges to state-of-the-art structural biology techniques such as X-ray crystallography or cryo-EM due to diffuse electron density. NMR spectroscopy is, however, well suited to studying IDPs and their conformational dynamics.[292,294,318, 331,332] ^{15}N relaxation experiments are the "working horse" for a protein dynamics investigation by NMR spectroscopy. The original "standard" NMR ^{15}N relaxation methods, developed in the early 1990s, cover the ps-ns time range faster than the overall tumbling rotational correlation time of the molecule.[315,333,334,335,336,337,338,339] Several sophisticated extensions were developed, e.g., *relaxation dispersion* experiments, which are sensitive to modulations of the isotropic part of the chemical shift tensor and can provide insights into the µs-ms time range.[340,341,342,343,344,345,346] Applications of NMR relaxation and relaxation dispersion experiments have led to numerous insights into protein structural dynamics, shaping our current understanding of the protein energy landscape[318,347,348,349,350,351] but also for RNA.[352,353,354] This manuscript focuses on the "standard" ^{15}N NMR relaxation methods addressing dynamics on the ps-ns time range, which can be considered as the flexibility of the protein. For these standard methods, several improvements have been added.[135,315,335,355,356,357,358,359,360,361] Systematic errors of used pulse sequences must be minimized to guarantee a reliable characterization of protein dynamics. This is particularly acute for IDPs as they are characterized by small relaxation rate constants and a high dynamic range. Due to the long history of NMR relaxation experiments, most pulse sequences have been developed at magnetic field strengths of 14.1 T, corresponding to a ^{1}H Larmor frequency of 600 MHz and lower. Frequently, these experiments have been designed for and tested on small, well-folded globular proteins. Advances in magnet technology have

led to the development of increasing magnetic field strengths of NMR spectrometers, with a 28 T magnet (1200 MHz ^{1}H Larmor frequency) recently becoming commercially available.[362] These high field strengths offer unprecedented resolution and new opportunities for investigating protein dynamics by NMR. They are particularly beneficial for studying IDPs (because of their narrow signal dispersion in the ^{1}H dimension). At the same time, high field strengths put new demands on underlying spectrometer electronics and probe design, as well as NMR pulse sequence design. Existing methods have to be tested and may have to be adapted, and novel NMR methods will be developed. In the following, we introduce two optimized ^{15}N R_1 and $R_{1\rho}$ pulse schemes with sensitivity-enhanced HSQC detection. The presented pulse schemes have been tested at 14.1 T and 28 T magnets, corresponding to ^{1}H Larmor frequencies of 600 MHz and 1200 MHz, respectively. Because research on IDPs represents a highly active and growing field within biomolecular solution NMR,[292, 294,318,331,332] we have tested the presented pulse sequences on an IDP, the vesicular SNARE protein Syb-2, which in its monomeric pre-fusion state behaves as an IDP.[275,284,363] Due to the small relaxation rate constants of IDPs, even small systematic errors of less than 1 s^{-1} will lead to substantial percentage-wise errors in rate constants. Therefore, IDPs present an ideal test case for NMR relaxation experiments and will identify even small systematic errors in the used pulse sequence. We have tested the optimized sensitivity-enhanced ^{15}N R_1 and $R_{1\rho}$ pulse sequences both at 600 MHz and 1200 MHz, using both a fully protonated as well as a deuterated sample, and compared the results to the original TROSY-based sequences, which had been tested on deuterated GB3 and 600 MHz.[135] For both fully protonated and deuterated samples, we find a high reproducibility and robustness at both field's strengths, suggesting that the presented sensitivity-enhanced sequences, as well as the original TROSY-based sequences,[135] will also be applicable at any magnetic field strength between 14.1 T and 28 T. High agreement of extracted relaxation rate constants, measured either in the sensitivity-enhanced HSQC-detected or the TROSY-detected sequence, is observed. Independent of HSQC- or TROSY-detection, we also find high agreement between the protonated and the deuterated samples.

9.3 Materials and Methods

9.3.1 NMR Sample Preparation

The gene base sequence encoding Synaptobrevin-2 (1-96), Syb-2 (1-96), the soluble part of Syb-2 (wt) without the transmembrane region, from *Rattus norvegicus* (UniProt accession

number: P63045, 96 amino acids, 10.52 kDa) was codon-optimized for expression in *E. coli* and subcloned in a pET28a(+) vector. To facilitate protein isolation, a hexa-histidine tag was included at the N-terminus. Additionally, a Thrombin-cleavage side (amino acid sequence: LVPR´GS) was inserted for tag removal. The pET28a(+)-His-Syb-2(1-96) plasmid was transformed into chemically competent *E. coli* BL21 (DE3) cells. Protein expression and purification followed protocols by Pobbati *et al*.[278] Briefly, for precultures, cells were grown at 37 °C in 2xYT medium (Thermo Fisher, Waltham, MA, USA) overnight. For isotope labeling of ^{15}N Syb-2 (1-96), cells were grown in M9 minimal medium supplemented with ^{15}N-NH_4-Cl (99%, Cambridge Isotope Laboratories, USA) as a single source for protein synthesis. The main culture was incubated at 37 °C until an OD_{600} of 0.8 was reached; at this point, protein expression was induced with 0.5 mM IPTG. Cultures were incubated at 20 °C after induction of expression for 18 hours. Bacterial cells were harvested for 10 min at 6,000 x *g* (4 °C) using centrifugation. The cell pellet was resuspended in 20 mM HEPES (pH 7.4), 500 mM NaCl, 8 mM Imidazole, and 0.1 mM TCEP supplemented with protease inhibitors (Complete, Roche) and DNAse I for nucleic acid digestion. Cells were lysed for 20 min by sonication, followed by an additional centrifugation step for 30 min at 45,000 rpm (4 °C). The lysate was purified by Ni^{2+}-NTA affinity chromatography. His-tagged Syb-2 (1-96) was eluted with 20 mM HEPES (pH 7.4), 500 mM NaCl, 400 mM Imidazole, and 0.1 mM TCEP. Thrombin cleavage was combined with dialysis overnight at 4 °C in 20 mM HEPES (pH 7.4), 150 mM NaCl, 0.1 mM TCEP, and 1 mM EDTA. Thrombin was separated from digested protein by ion exchange chromatography using 20 mM HEPES (pH 7.4), 0.1 mM TCEP, 1 mM EDTA, and a Resource Q column (Cytiva, Marlborough, MA, USA) for isolation. For elution, the same buffer composition was used with 1000 mM NaCl added. To increase the purity of the sample, a size exclusion chromatography was performed afterward using a HighLoad 16/600 Superdex 75 pg column (Cytiva, USA) equilibrated in 50 mM MES (pH 6.0), 150 mM NaCl, 0.1 mM TCEP and 1 mM EDTA. For the expression of deuterated ^{2}H^{15}N^{13}C Syb-2 (1-96) and ^{2}H^{15}N Syb-2 (1-96), commonly used filtered and deionized water (Milli-Q, Merck, USA) was replaced by D_2O. To reduce stressing conditions for cell growth, the transition from 2xYT medium into minimal medium with D_2O was performed slowly. First, cells were precultured in 2xYT medium and ddH_2O at 37 °C overnight. Then, cells were precultured in M9 minimal medium supplemented with ^{15}N-NH_4-Cl in ddH_2O at 37 °C for 8 hours. In the third overnight preculture, ddH_2O was replaced by D_2O. The main culture of deuterated ^{2}H^{15}N^{13}C Syb-2 (1-96) was supplemented

with ^{15}N-NH_4-Cl and ^{13}C-D-glucose (99% Cambridge Isotope Laboratories, USA) in D_2O. The main culture of deuterated $^{2}H^{15}N$ Syb-2 (1-96) was supplemented with ^{15}N-NH_4-Cl in D_2O. Induction of expression with IPTG and protein isolation was performed as described for ^{15}N Syb-2 (1-96); see above.

9.3.2 NMR Spectroscopy

9.3.2.1 Potential Sources of Artifacts in Standard NMR Relaxation Experiments

Well-known systematic errors are generated by cross-correlated relaxation (CCR) of the ^{1}H-^{15}N dipolar coupling (DC) and the ^{15}N chemical shift anisotropy (CSA),[315,355,357] which increases at higher magnetic fields.[360] Refocusing of the respective CCR during the NMR relaxation is usually achieved by 180° ^{1}H pulses.[315] For several ^{15}N R_1 experiments, frequently, a train of hard 180° ^{1}H pulses, spaced by a 5 ms delay, was used to suppress CCR. As discussed in the earlier work on ^{15}N NMR relaxation experiments with TROSY-detection,[135] this pulse train leads, however, to partial saturation of the water magnetization. The degree of water saturation will vary depending on the length of the relaxation period. In combination with the long T_1 times of the water in the order of 3-4 s, this can substantially attenuate the available amide proton magnetization via direct exchange with the water magnetization or through ^{1}H-^{1}H NOEs to nearby exchangeable protons,[135] as also pointed out by Chen and Tjandra independently.[359] The impact of progressive water saturation during the ^{15}N T_1 relaxation period increased the measured ^{15}N R_1 rate constant and was discussed in detail in [135]. This apparent (artificial) increase could be correlated to the solvent exposure of the respective amide group, with the most solvent-exposed amide groups displaying the strongest increase.[135] That systematic error can be circumvented by choosing very long interscan recovery delays (up to 10 s), leading to very long experimental times that are frequently not affordable. Another possibility is to saturate the water resonance entirely before the relaxation period starts, resulting in intensity losses for amide groups subject to solvent exchange. As discussed in the context of the TROSY-based sequences,[135] shaped I-BURP-2 180° ^{1}H pulses, being selective on the amide protons and spaced by a 40 ms delay, provide an alternative.[135] The 40 ms inter-pulse delay was sufficient to refocus CCR but, simultaneously (in combination with amide-selective shaped pulses), disturbs the water magnetization only minimally. We use the same strategy for the sensitivity-enhanced HSQC-detection scheme (Figure 78). As pointed out, the key is keeping or returning the water magnetization to the

z-axis. This is also achieved by the additional ^{1}H 90 (-x) pulse at the end of the refocused INEPT transfer, which fulfills two purposes: Returning the water to the z-axis and removing any residual antiphase coherence. Radiation damping of the water magnetization[364] is an additional but related concern, particularly acute for cryoprobes and high-field magnetics.[365,366,367] Very weak gradients can dephase the water magnetization and prevent radiation damping.[368] An alternative is to return the water magnetization to the z-axis using water-flip back pulses.[358,367,369] The TROSY-based- and HSQC-based experiments (Figure 78, see below) have been designed to return the water magnetization to the z-axis in the refocused ^{1}H-^{15}N INEPT transfer before the relaxation period. Therefore, the water magnetization is aligned along the z-axis at the start of the relaxation period. The R_1 relaxation delay element is designed to disturb the water magnetization minimally (see above). For the $R_{1\rho}$ relaxation period, we introduce amide-selective IBURP-2 pulses, which will be discussed further below. Therefore, systematic errors due to varying degrees of water saturation will be alleviated in the examined sequences. Any disturbance of the water magnetization after the relaxation period will reduce the achievable signal-to-noise ratio of the NMR experiment but will not introduce additional systematic errors. The TROSY detection element is advantageous regarding water suppression, as decoupling on the ^{1}H channel during t_1 evolution is avoided; for the sensitivity-enhanced HSQC detected relaxation experiments, a 180° (^{1}H) is used to refocus J-coupling evolution. A second 180° (^{1}H) pulse at the beginning of the t_1 evolution is used within the gradient-based Echo-/ Anti-Echo encoding element. The water magnetization will be returned to the z-axis after the second 180° (^{1}H) pulse. Radiation damping can occur in principle for longer t_1 times and the associated spacing of the two 180° (^{1}H) pulses. Although at the two spectrometers tested, we did not observe strong radiation damping, an optional weak rectangular gradient (< 2%) can be included in the pulse sequence (Figure 78, gradient G10, grey) to prevent radiation damping,[368] similarly as used in [135]. We further minimized the impact of the water signal caused by using hard 90° and 180° ^{1}H pulses by carefully choosing gradients (Figure 78, see below). Using amide-selective detection schemes, applied in SOFAST-HMQC or Band-selective Excitation Short-Transient (BEST) detection schemes, presents an alternative.[370,371] Very recently, during the revision of this manuscript, broad-band shaped pulses designed by optimum control theory were introduced to cover the large spectral width at very high magnetic fields.[372] Using those pulses, a significant SNR increase in a ^{1}H, ^{15}N TROSY-HSQC experiment could be obtained while requiring only moderate RF power levels.

9.3.2.2 General Setup of NMR Relaxation Experiments

We have recorded NMR relaxation experiments using a 171.1 µM ^{15}N Syb-2 (1-96) sample and, for comparison, a 171.7 µM ^{2}H^{15}N Syb-2 (1-96) sample in 50 mM MES (pH 6.0) buffer containing 150 mM NaCl, 0.1 mM TCEP and 1 mM EDTA. (The ^{15}N R_1 experiment, investigating the influence of the recovery delay, had been recorded on a 210 µM ^{2}H^{15}N^{13}C Syb-2 (1-96) sample.) All NMR experiments were performed at 278.15 K using a 250 µl volume filled in a 3 mm NMR sample tube. Experiments at 600 MHz ^{1}H Larmor frequency were conducted on a Bruker 600 MHz AVANCE III HD spectrometer equipped with a Bruker 5 mm QCI ^{1}H, ^{15}N, ^{13}C, ^{31}P quadruple resonance cryoprobe. Experiments at 1200 MHz ^{1}H Larmor frequency were recorded on a Bruker 1200 MHz AVANCE NEO spectrometer (Bruker, Billerica, MA, USA) equipped with a 3 mm TCI ^{1}H, ^{15}N, ^{13}C triple resonance cryoprobe. The optimized ^{15}N R_1 and $R_{1\rho}$ pulse sequences with sensitivity-enhanced HSQC detection presented here (Figure 78) employ a sensitivity-enhanced HSQC detection scheme[315,373] and can be applied to fully protonated proteins. The basic building blocks of those two sequences are conceptually similar to the previous TROSY-based sequences and their HSQC counterparts, initially designed in the Bax laboratory.[135] At that time, those pulse sequences had been tested on deuterated GB3 at 600 MHz. While the TROSY-detection[241,374] offers the best water suppression by using fewer ^{1}H pulses and avoiding ^{1}H decoupling, the sensitivity-enhanced HSQC scheme provides a higher signal-to-noise ratio (SNR) in principle (up to two-fold theoretically). In the sensitivity-enhanced read-out scheme, more ^{1}H pulses are used by design. While care was taken to return the water magnetization to the z-axis by the end of the pulse sequence, radiation damping of the water magnetization during the pulse sequence and its associated water trajectory is a potential pitfall. Previously, radiating damping was (partially) avoided by the use of soft rectangular gradients (of fixed duration) placed in the delay periods, requiring, however (in our hands), a very careful and frequently tedious gradient optimization. In turn, the potential gain in SNR of the HSQC-based sequence was often limited by a still strong water signal, limiting the affordable receiver gain setting.[135] The ^{15}N R_1 and $R_{1\rho}$ pulse sequences with sensitivity-enhanced HSQC detection presented here offer an improved water suppression (see Results section, Figure 80), and an easier to optimize water suppression (in our hands), by replacing those previous weak rectangular gradients by soft shaped gradients that can be adjusted in their gradient strengths and duration. Further, the ^{15}N $R_{1\rho}$ experiment uses two shaped ^{1}H pulses rather than two hard

180 (^{1}H) rectangular pulses spaced by a weak rectangular gradient in the delay period. To obtain the best water suppression in the sensitivity-enhanced HSQC experiments, we recommend optimization of the ^{1}H carrier frequency using a pre-saturation experiment (Bruker: zgpr sequence) and adjusting the ^{1}H carrier frequency (Bruker: o1) such that the water signal is minimized. Because of the easier-optimized water suppression combined with the high SNRs that can be achieved (see below), we anticipate these sequences to be of practical importance and valuable for various fully protonated small to medium-sized proteins. In the following, we briefly discuss the magnetization transfer pathway of the ^{15}N R_1 and $R_{1\rho}$ pulse sequences shown in Figure 1. After the recovery delay, a hard 90° (^{15}N) pulse destroys any residual non-Boltzmann magnetization that could be present on ^{15}N. A refocused INEPT element (two consecutive INEPT transfers[374]) transfers the initial ^{1}H Boltzmann magnetization to in-phase magnetization on the ^{15}N nucleus. During the following relaxation period, magnetization on the ^{15}N nucleus decays - depending on the length of the relaxation delay and resulting in decreased intensities as a function of the relaxation delay. The observed decay of intensities will be fitted by a mono-exponential decay function later to obtain the relaxation rate constants. Residue-specific intensities and rate constants will be acquired by the following t_1 evolution period of the transverse ^{15}N magnetization. For quadrature detection (phase sensitivity in the ^{15}N dimension), ^{15}N coherences are encoded by an Echo/ Anti-Echo encoding scheme with alternating gradients.[315,373] The ^{15}N evolution period is followed by a refocused INEPT element, which transfers the magnetization to the amide protons. During a final Hahn-Echo element, the decoding gradient of the Echo/ Anti-Echo scheme follows. The complete pulse scheme is illustrated in Figure 78. Gradients during the INEPT elements are implemented both for coherence selection and to prevent radiation damping of the water magnetization (here, only weak gradient strengths are used).[368,375]

9.3.2.3 ^{15}N R_1 Experiment

The central part of pulse schemes is the relaxation period, which will be discussed in more detail in the following: In the ^{15}N R_1 experiment, the relaxation period is elongated by increasing a loop counter and the associated number of repetitions (n) of the bracketed part of the pulse scheme (Figure 78A). The entire relaxation period is omitted in the first experiment, resulting in the $\tau_1 = 0$ reference point. Cross-correlated relaxation of the ^{15}N CSA and the ^{1}H-^{15}N dipolar coupling interactions during the relaxation period is refocused by a

central I-BURP-2 180° pulse,[376] selective to the amide proton resonances at $\Delta/2$, the center of the respective loop element. The relaxation loop is repeated an even number of times (n = 0, 2, …) to balance the evolution of longitudinal cross-correlated relaxation during the relaxation period Δ. These I-BURP-2 pulses invert the amide proton magnetization during the relaxation period. Here, it is essential to use selective pulses to avoid exciting the water resonance, as partial saturation of the water magnetization for more extended relaxation periods would impact the return to ^{1}H Boltzmann equilibrium adversely,[358] resulting in erroneous and too high ^{15}N R_1 rate constants for solvent-exposed residues.[135,359] Any amide proton will spend half of the relaxation period in the α-state and half the time in the β-state (assuming an ideal case, without any random ^{1}H-^{1}H spin-flips during the relaxation period). As a result, the longitudinal cross-correlated relaxation between the ^{1}H-^{15}N dipolar coupling and the ^{15}N chemical shift anisotropy will be averaged out. In the following, we briefly discuss the experimental parameters of the ^{15}N R_1 experiment: The duration of the delay in the INEPT transfer is δ = 2.65 ms; the delay ϵ corresponds to the decoding gradient G4 (201 µs). This delay makes inserting the decoding gradient without adding a linear phase error to the proton dimension possible. The variable delay corresponds to the delay between two shaped 180° I-BURP-2 pulses, selective on the amide proton region. For the duration of the individual I-BURP-2 pulses, we choose a length of 1000 µs at 1200 MHz to cover the amide ^{1}H chemical shift region of the IDP Syb-2 (1-96). We used a delay of 40 ms; the choice of the delay depends on the relaxation properties of the protein, which must be faster than the cross-correlated relaxation rate of the protein. The relaxation period is elongated by increasing the loop counter n and the associated repetition of the bracketed element of the pulse scheme. In the first experiment, the entire relaxation period will be jumped over. The relaxation loop is repeated an even number of times (n = 0, 2, …) to match the chosen relaxation delay. For the elimination of any cross-correlated relaxation between the ^{13}C and ^{15}N nuclei (applicable to ^{13}C labeled samples) on the ^{13}C channel, four 180° pulses, two selective rectangular 180° ^{13}C pulses are applied on ^{13}C' as well as on $^{13}C^{\alpha}$, at Δ*1/4 and Δ*3/4 of the relaxation period. The duration of the selective rectangular ^{13}C' pulse is chosen such that the $^{13}C^{\alpha}$ resonances are not excited and vice versa (see below).[369] The length of these selective rectangular 180° pulses is defined by $\frac{\sqrt{3}}{2\Omega}$.[369] Ω corresponds to the chemical shift difference between $^{13}C^{\alpha}$ and ^{13}C'; therefore, the pulse length of the 180° pulses was set to 23.7 µs at 1200 MHz. Gradients are sine-bell shaped, identical to the sine.20 gradient shape of the Bruker gradient library, apart from G7, which is

a sine.50 shaped. The gradient strengths are G_1 (200 µs, 1% (at our spectrometer setting), corresponding to 0.58 G/cm), and G_2 (200 µs, 2%, 1.16 G/cm), G_3 (800 µs, 31%, 17.98 G/cm),G_4 (200 µs, 11%, 6.38 G/cm), G_5 (201 µs, 50%, 29.00 G/cm), G_6 (1000 µs, 50%, 29.00 G/cm), G_7 (200 µs, 50%, 29.00 G/cm), G_8 (200 µs, 5%, 2.90 G/cm), and G_9 (200 µs, 14%, 8.12 G/cm). Further, we include an optional Gradient G10 (t1/4, 0.5%, 0.29 G/cm) during the first half of the t1 evolution period that can be switched on in case radiation damping is observed. In the ^{15}N dimension, a garp decoupling with an RF amplitude of 1.25 kHz, corresponding to a 90° pulse length of 200 µs, is applied. For the measurements conducted at 600 MHz we used the identical gradient setup. Therefore, this gradient setup should be widely applicable to various spectrometer frequencies between 600 MHz and 1200 MHz. At least the suggested gradient setup provides a good starting point for any further (and spectrometer dependent) optimization of the water suppression.

9.3.2.4 ^{15}N $R_{1\rho}$ Experiment

For the ^{15}N $R_{1\rho}$ experiment (Figure 78B), providing rate constants of the transverse relaxation, R_2 (see below), an adiabatic half-passage pulse turns the ^{15}N magnetization from the z-axis to the transverse plane, or the effective field direction for off-resonance nuclei, respectively.[377,378] The adiabatic half tanh/tan pulse74 has the same strength as the applied spin-lock RF amplitude (2 kHz). During the relaxation period, the magnetization is spin-locked in the rotating frame, and the intensity decay of the respective resonances is measured as a function of the lengths of the relaxation delay period Δ. In principle, the RF amplitude should be as high as affordable to reduce exchange contributions and off-resonance effects at the edges of the spectrum. The higher the spectrometer frequency, the more significant potential off-resonance effects will be. At 1200 MHz, for example, a 2 kHz spin-lock RF amplitude will be equivalent to 1 kHz at 600 MHz in terms of off-resonance effects. Particular care must be taken to properly align the magnetization along the effective field axis using an adiabatic pulse with the same RF amplitude as the following spin-lock. Confirming that the adiabatic pulse fulfills its adiabaticity condition for the given RF amplitude is important. Otherwise, magnetization components orthogonal to the effective field axis will be generated that will get dephased during the initial spin-lock period, leading to a sudden intensity drop in the decay curve and, consequently, a bi- or multi-exponential decay. During the relaxation period, two 180° I-BURP-2 pulses, selective to the amide proton resonances, are applied at Δ*1/4 and

Δ*3/4 of the relaxation period to average out contributions by transverse ^{15}N chemical shift anisotropy (CSA)/ ^{1}H-^{15}N dipolar coupling cross-correlated relaxation. In the case of ^{13}C labeled samples, a selective soft rectangular 180° pulse is applied, one at the ^{13}C' and the other at the $^{13}C^{\alpha}$ carrier, to prevent any errors potentially introduced by cross-correlated relaxation between ^{15}N and ^{13}C. In both $R_{1\rho}$ and R_1 experiments, a temperature-compensation element[135] is applied off-resonance on the ^{15}N channel. On modern NMR probes, that temperature compensation element will not necessarily be required and could be switched off to reduce the power deposition in the probe, at least in our experience. In the following, we briefly describe the setup of the ^{15}N $R_{1\rho}$ experiment: Apart from the bracketed relaxation period, the $R_{1\rho}$ experiment is similar to the R_1 experiment. The gradient setup is identical. The additional gradient G9 is a sine-bell-shaped (sine.20 Bruker gradient library), with a duration of 200 µs and a strength of 10%, corresponding to 5.8 G/cm. (This gradient will dephase residual orthogonal magnetization. A stronger gradient is advisable but has to be adjusted as a trade-off between dephasing efficiency and good overall water suppression.) The variable delay is equal to the length of the relaxation period. A spinlock with an RF amplitude of 2 kHz is applied during the relaxation period. Before the spinlock, an adiabatic half passage pulse (AHP) with a duration of 3000 µs aligns the magnetization on the effective field axis,[368] and a second AHP returns the magnetization to the z-axis after the spinlock period. Those AHPs are displayed as triangle pulses before and after the spinlock. They correspond to the first and second half of a tangent-hyperbolic tangent (tanh/tan) adiabatic inversion pulse, as defined in the Bruker pulse library (100 kHz total sweep width, ζ = 10, tan(κ) = 20, ωmax = 2 kHz). As for the R_1 experiment, 180° pulses on the ^{13}C' and the $^{13}C^{\alpha}$ are used to eliminate the cross-correlated effects resulting from the ^{13}C-^{15}N dipolar interactions in ^{13}C labeled samples. The two I-BURP-2 pulses stop the transverse cross-correlated relaxation contributions from the ^{15}N chemical shift anisotropy and the ^{1}H-^{15}N dipolar coupling interactions. The optimized NMR pulse sequences can be downloaded at www.ipb.hhu.de/en/teams/team-lakomek/pulsesequences.

9.3.2.5 Application to Intrinsically Disordered Proteins

To test the applicability of the pulse sequences on fully-protonated intrinsically disordered proteins and at high magnetic field strengths, we recorded ^{15}N R_1 and $R_{1\rho}$ relaxation experiments on monomeric Syb-2 (1-96), see above, which is intrinsically disordered in its

monomeric state.[275,284,379] NMR spectra were acquired at 1200 MHz and 5 °C, with 16.03 ppm spectral width in the ^{1}H direct dimension and 30.02 ppm spectral width in the ^{15}N indirect dimension. Altogether, eight different relaxation delays were recorded for the ^{15}N R_1 and ^{15}N $R_{1\rho}$ experiments. Each spectrum (= plane of the pseudo-3D) was acquired with 1024 complex data points in the ^{1}H dimension, corresponding to an acquisition time of 53.25 ms in the ^{1}H dimension. In the ^{15}N dimension, experiments were performed with an acquisition time of 70.14 ms and 256 complex data points per spectrum. Frequency carriers were placed for protons at 4.690 ppm, nitrogen at 117 ppm, and carbon at 176 ppm. For the R_1 experiment, the following relaxation delays were randomly shuffled: 0 ms, 960 ms, 240 ms, 800 ms, 160 ms, 640 ms, 320 ms, and 480 ms. In the $R_{1\rho}$ experiment, the following delays were used: 5 ms, 65 ms, 25 ms, 45 ms, 15 ms, 55 ms, 35 ms, and 10 ms. The first delay serves as the reference point. For the spinlock on the ^{15}N channel, a radio frequency (RF) amplitude of 2 kHz was applied. A four-step phase cycle was employed, limiting the applicable number of scans to multiples of four. As an interscan delay, 2.0 s were used unless specified otherwise. The experiments at 1200 MHz had a duration of 12 h 12 min total for the R_1 experiment and a duration of 10 h 19 min for the $R_{1\rho}$ experiment. A detailed overview of the chosen experimental parameters is shown in Table S19. Experimental NMR parameters for the relaxation experiments recorded at 600 MHz are shown in Table S20. For comparison, TROSY experiments were acquired as described in [135].

9.3.2.6 Evaluation of NMR Relaxation Data

To evaluate the experimental data, the dedicated NMR pipe[129] scripts originally published in the context of the TROSY-based sequences[135] were modified slightly. Scripts can be downloaded under the following link:

www.ipb.hhu.de/en/teams/team-lakomek/pulsesequences.

The R_2 rate constants were calculated by the following formula[380,381]:

$$R_2 = \frac{R_{1\rho}}{sin^2\,\theta} - \frac{R_1}{tan^2\,\theta}$$

The angle $\theta = arctan\left(\frac{\omega_1}{\Omega}\right)$ represents the angle between the axis of the effective magnetic field B_{eff} and the external magnetic field B_0, ω_1 is the RF amplitude of the spin-lock and Ω the chemical shift offset (the difference between the ^{15}N chemical shift of the respective residue and the ^{15}N carrier frequency).[380,381]

The experimental error was estimated using the NMR pipe-integrated Monte-Carlo-based error analysis.[129] The signal-to-noise ratio obtained for the various experiments (cf. Table S18) was calculated by taking the mean of the intensities of the three most intense resonances in the reference spectrum (first entry in the relaxation delay list). Those signals had to be non-overlapping signals. Their average intensity was divided by the noise estimated using nmrDraw / nmrPipe scripts.[379] Further, a "weak" SNR takes the average intensity of the three least intense signals and is divided by the noise calculated with nmrDraw. All SNR values were calculated for the reference spectrum acquired with the first entry of the vplist in the $R_{1\rho}$ experiment or with the first entry of the vclist in the R_1 experiment.

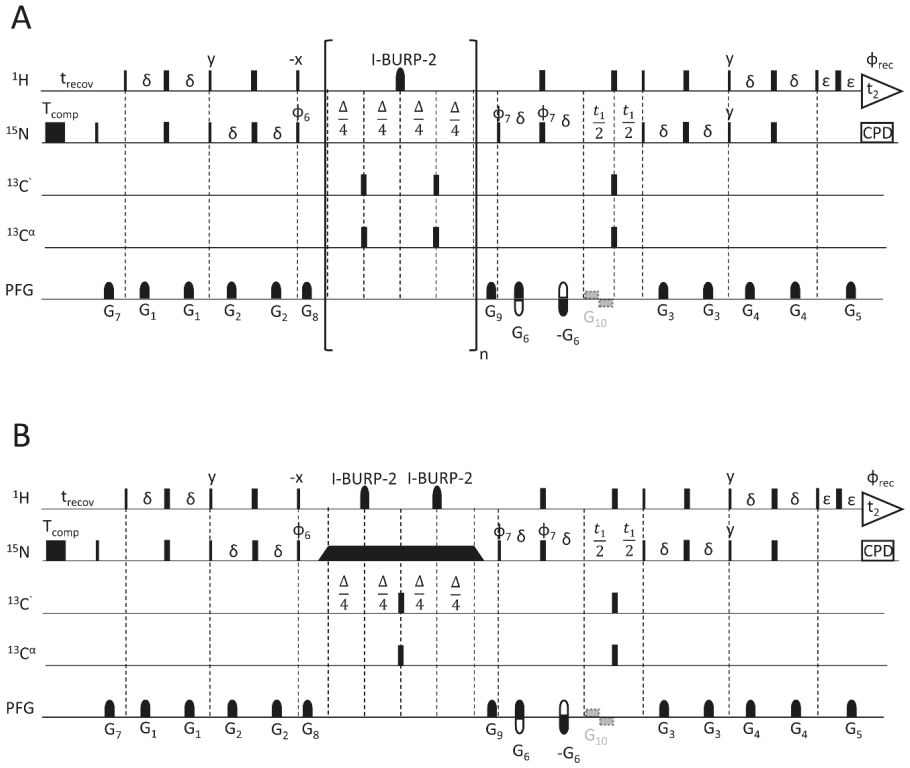


Figure 78: Pulse sequence schemes of NMR relaxation experiments using a sensitivity-enhanced HSQC read-out scheme. The following phase cycling is applied: ϕ_6 = y, y, -y, -y; ϕ_7 = y, -y, ϕ_{rec} = y, -y, -y, y. The quadrature detection is implemented with the polarity inversion of the encoding gradient G_5 and the phase cycling of ϕ_7 (Echo/ Anti-Echo detection[315]). **(A)** Pulse sequence scheme of the used ^{15}N R_1 relaxation experiment. Narrow rectangles correspond to 90° hard pulses with phase x, if not marked otherwise. 180° hard pulses are displayed as broad rectangles. **(B)** Pulse sequence scheme of the used ^{15}N $R_{1\rho}$ experiment. Apart from the bracketed relaxation period, the $R_{1\rho}$ experiment is similar to the R_1 experiment.

9.4 Results and Discussion

We have tested the optimized NMR relaxation experiments for their applicability to intrinsically disordered proteins. As a model system, we used the soluble cytoplasmic part of the vesicular SNARE protein Syb-2 (1-96), frequently referred to as VAMP-2. Syb-2 (1-96) is intrinsically disordered in its monomeric pre-fusion state. As a first assessment of the IDP character of Syb-2 (1-96), we recorded two-dimensional ^{1}H-^{15}N HSQC spectra on ^{15}N Syb-2 (1-96), both at 600 MHz and 1200 MHz.

9.4.1 Comparison of ^{1}H-^{15}N HSQC Spectra of ^{15}N Syb-2 (1-96) Recorded at 600 MHz and 1200 MHz

Figure 79 shows spectra of fully-protonated Syb-2 (1-96) measured at 600 MHz (Figure 79A) and 1200 MHz (Figure 79B). Both spectra have been recorded with an acquisition time of 106.5 ms in the direct ^{1}H dimension, corresponding to 1024 complex points in the direct ^{1}H dimension at 600 MHz and 2048 complex points at 1200 MHz.

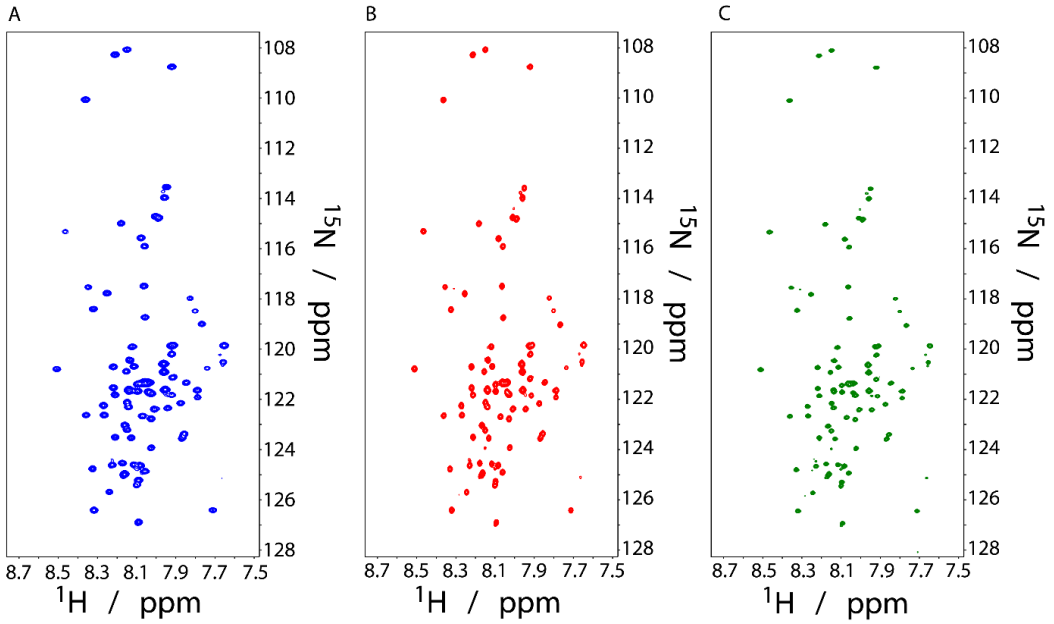


Figure 79: Two-dimensional ^{1}H-^{15}N HSQC spectra of fully protonated Syb-2 (1-96). All spectra were acquired at 5 °C with 8 scans, an interscan delay of 1.2 s, and a spectral width of 16 ppm in ^{1}H dimension and 30 ppm in ^{15}N dimension. **(A)** HSQC at 600 MHz with 1024 complex points in the ^{1}H dimension and 256 complex points recorded in the ^{15}N dimension. **(B)** HSQC at 1200 MHz with 2048 complex points in the ^{1}H dimension and 256 in the ^{15}N dimension. **(C)** HSQC at 1200 MHz with 2048 complex points in the ^{1}H dimension and 512 complex points in the ^{15}N dimension, resulting in the same acquisition time as 600 MHz (but doubling the experimental time).

In the indirect ^{15}N dimension, 256 complex points were recorded. The timewise increment is defined by 1/SWH (or 1/(2*SWH), respectively, depending on the employed quadrature detection and implementation of the pulse sequence), with the spectral width in Hz dubbed SWH. Therefore, at 1200 MHz, the length of the timewise increment, 1/SWH, will be only half as long as the corresponding one at 600 MHz. Thus, the double number of increments must be recorded at 1200 MHz relative to 600 MHz to achieve the same total acquisition time and resulting spectral resolution in Hz (which correlates inversely to the total acquisition time). While a longer acquisition time in the direct ^{1}H dimension will minimally increase the overall experimental time, the twice-as-long acquisition time in the indirect dimension due to the doubling of the recorded complex points will double the experimental time. However, when doing so, the resolution can be further increased (Figure 79C). As evident from Figure 79, higher magnetic field strengths will be beneficial for recording highly resolved NMR spectra of IDPs, even when fully protonated.

9.4.2 ^{15}N R_1, $R_{1\rho}$ with Sensitivity-Enhanced HSQC-Detection and Improved Water Suppression

9.4.2.1 Modifications in the ^{15}N R_1 and $R_{1\rho}$ Pulse Sequences

We briefly reiterate the main modifications for the ^{15}N R_1 and ^{15}N $R_{1\rho}$ relaxation experiments using a sensitivity-enhanced HSQC read-out scheme: starting from the original TROSY-based NMR relaxation experiments and their HSQC counterpart[135,318] we have implemented ^{15}N R_1 and $R_{1\rho}$ pulse sequences using a sensitivity-enhanced (Rance-Kay) HSQC read-out scheme with improved water suppression (Figure 78). We emphasize that for both read-out schemes, the TROSY and the HSQC-detected ones, the “standard” ^{15}N R_1 and $R_{1\rho}$ (auto-relaxation) rate constants are measured. Relative to the original implementation (compare SI Figure S1 in [135]), instead of weak rectangular gradients of fixed duration, we have used shorter shaped gradients that can be adjusted both in duration and time, allowing an easier to optimize water suppression. For the $R_{1\rho}$ experiment, we replaced the two hard 180° pulses (plus the soft gradient in between, preventing radiation damping) with two shaped I-BURP-2 pulses, which are selective on the amide protons. Figure 78 shows the improved set of NMR R_1 and $R_{1\rho}$ relaxation pulse sequences with sensitivity-enhanced (Rance-Kay) HSQC detection and improved water suppression, applicable at high-field magnets. We refer to the Materials and Methods section for details on the experimental setup.

9.4.2.2 Improved Water Suppression

Figure 80 displays the improved water suppression obtained using the presented ^{15}N R_1 and $R_{1\rho}$ NMR relaxation experiments with sensitivity-enhanced HSQC detection, tested both at 1200 MHz and 600 MHz.

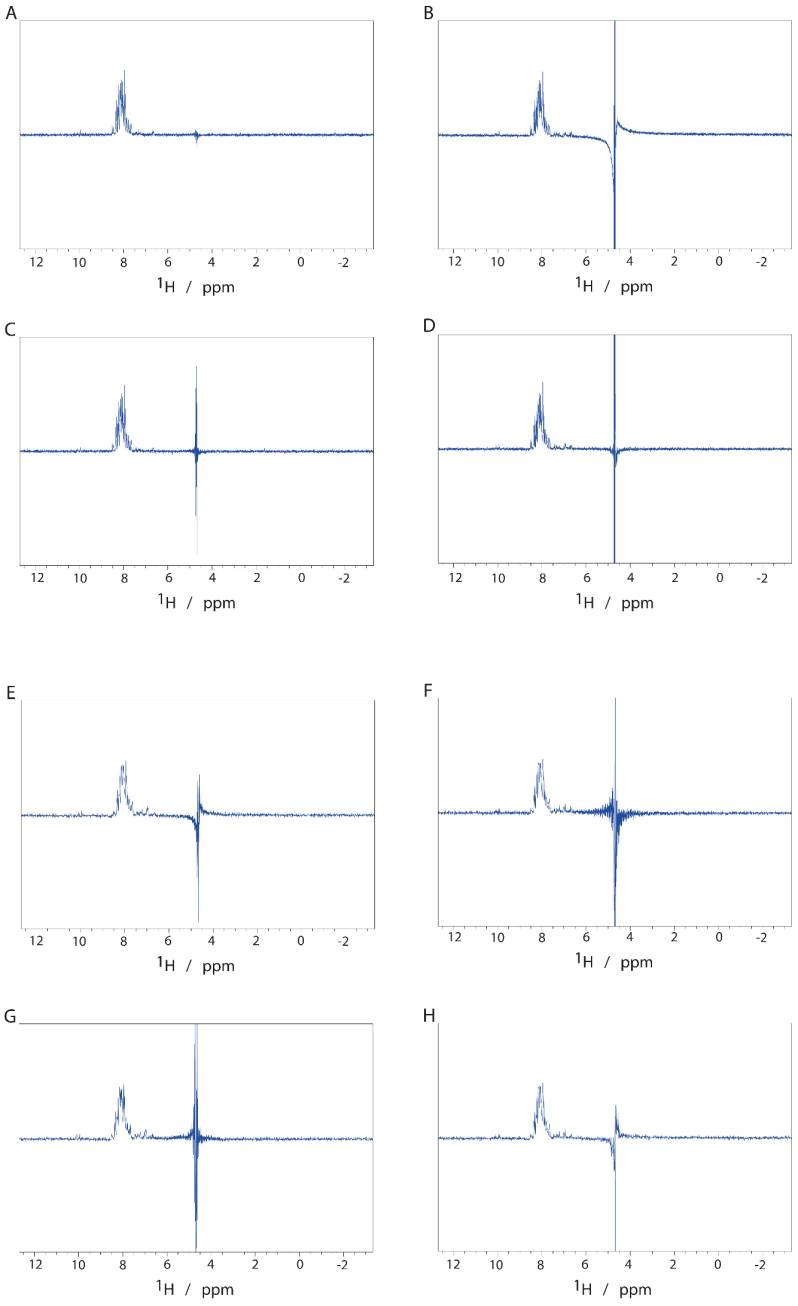


Figure 80: Improvement of water suppression in ^{15}N $R_{1\rho}$ and $R_{1\rho}$ NMR relaxation experiments. Sensitivity-enhanced HSQC-detection using the new implementation (left, see also Figure 1) and compared to the old implementation (right, SI Figure S1 in [135]). Shown is the 1D spectrum obtained from the first FID recorded on the protonated ^{15}N Syb-2 sample. **(A)** Optimized ^{15}N $R_{1\rho}$ experiment at 1200 MHz compared to **(B)** previous implementation. ^{15}N R_1 experiment at 1200 MHz: **(C)** New and **(D)** old. ^{15}N $R_{1\rho}$ experiment at 600 MHz: **(E)** new and **(F)** old. ^{15}N R_1 experiment at 600 MHz: **(G)** new and **(H)** old.

For comparison, the water-suppression using the previous implementation (SI Figure S1 in [135]) is shown.[135] The water-suppression obtained for the deuterated Syb-2 sample (see below) is shown in Figure S23. Apart from the ^{15}N R_1 experiment at 600 MHz, the level of water suppression of the new implementation is improved compared to the previous implementation. At 1200 MHz, using a 3 mm Bruker TCI cryoprobe, the level of improvement is higher than at 600 MHz using a 5 mm Bruker QCI probe. The same sample in a 3 mm NMR tube was used for both NMR spectrometers.

9.4.3 Protonated Sample: Comparison between TROSY- and HSQC-Detection Scheme

9.4.3.1 Application at 1200 MHz

The ^{15}N R_1 and $R_{1\rho}$ relaxation rate constants of the fully protonated ^{15}N Syb-2 (1-96) sample measured at 1200 MHz using the sensitivity-enhanced HSQC detection scheme displayed in Figure 78 are shown in Figures 81A and 81D (blue) and compared to the data obtained using TROSY-detected pulse sequences (orange).[135] The ^{15}N R_2 rate constants are derived from the measured ^{15}N R_1 and $R_{1\rho}$ rate constants using the formula (1).

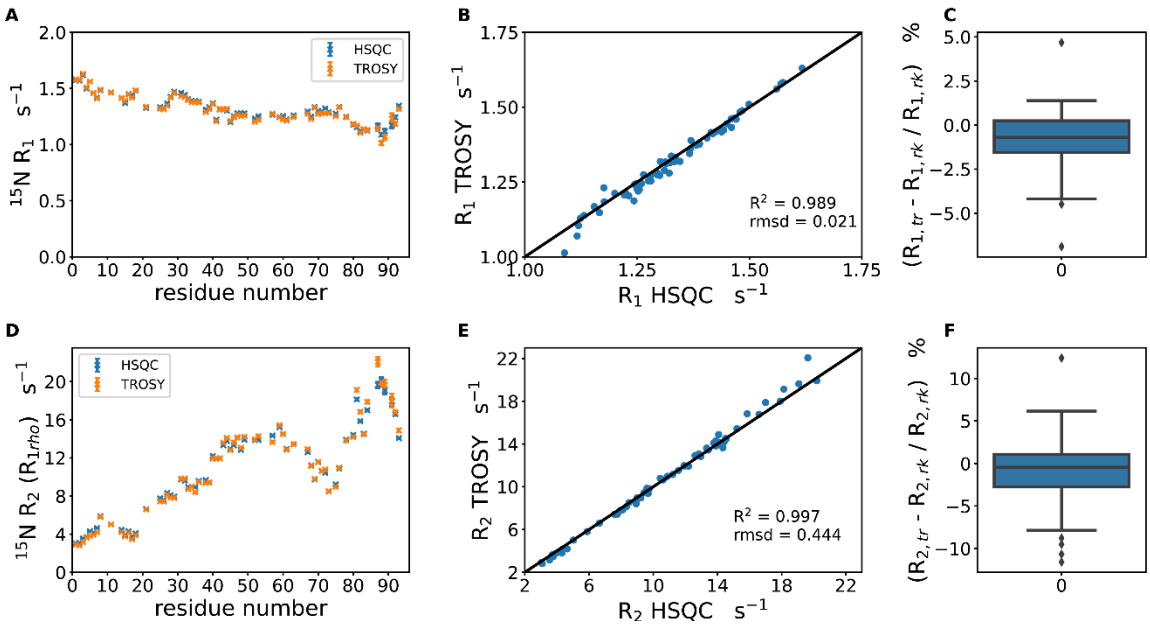


Figure 81: Comparison of relaxation rates of fully protonated ^{1}H-^{15}N Syb-2 (1-96) acquired with the pulse schemes in Figure 1 A, B at 1200 MHz and compared to data recorded on the same sample using the TROSY read-out pulse schemes. (A) Residue-specific R_1 relaxation rate comparison. **(B)** Correlation plot of the compared R_1 relaxation rate constants. **(C)** Box plot of the residue-specific percentual differences of R_1 relaxation rate constants. **(D-F)** Same plots as in A-C, illustrating the comparison of R_2 ($R_{1\rho}$) rate constants.

Low ^{15}N R_2 rate constants indicate high internal dynamics at the N-terminus of the Syb-2 (1-96), consistent with the intrinsically disordered state of monomeric Syb-2. Increasing relaxation rate constants along the protein backbone to the C-terminus point to decreasing internal dynamics. In the region of residue S80 up to W90 the R_2 relaxation rate constants reach a maximum before the rates drop due to the internal highly dynamic amino acids at the C-terminus. These observations correspond well to our earlier finding on $^{2}H^{15}N^{13}C$ Syb-2 (1-96).[275] The R_1 data only show small variations along the sequence and do not follow a clear trend. We re-emphasize that the HSQC-detected and the TROSY-detected schemes measure the "standard" R_1 and $R_{1\rho}$ auto-correlation rate constants. The ^{15}N R_1 rate constants, measured using either the HSQC or the TROSY-detection scheme, follow a very similar pattern and show a high correlation, with a Pearson correlation coefficient of R = 0.989 and a root mean square deviation (rmsd) of 0.021 (Figure 81B). A box plot analysis was performed to assess the agreement quality better. In Figure 81C, the difference of rate constants is measured using the HSQC or TROSY detection scheme. The agreement is best if the average deviation is zero and the width of the distribution, shown by the box plot bar (interquartile range), is small. We find that the average of the R_1 rate constants measured using the TROSY-detection scheme minus those measured using the HSQC-detection scheme is about -0.76%. This means that rate constants measured using the TROSY-detection scheme are underestimated compared to those obtained using the HSQC-detection scheme by about 0.76%. We observe an experimental error between 0.2% (most intense resonances) and 3% (least low resonances) for the HSQC-detected and between 0.3% and 4% for the TROSY experiment. The ^{15}N R_2 rate constants (obtained from the ^{15}N $R_{1\rho}$ rate constants by formula 1) are, on average, about -0.84% lower for the TROSY-detection scheme than for the HSQC-detection scheme (Figure 81F) but agree with a Pearson correlation coefficient R=0.997 and an rmsd=0.444. The experimental error is 0.3% and 3% for the HSQC-detected experiment and between 0.4% and 4% for the TROSY-detected experiments. Compared to the TROSY-based $R_{1\rho}$ experiment,[135] in the new pulse scheme presented in Figure 78B, we modified the way the evolution of cross-correlated relaxation is refocused. In the new experiment, two 180° I-BURP-2, selective to the amide protons and spaced at Δ*1/4 and Δ*3/4 duration of the relaxation period, refocus the cross-correlated relaxation. Previously, this was achieved by two hard 180° ^{1}H composite pulses. To avoid radiation damping of the water magnetization, a very weak gradient (that alternated its phase halfway) was introduced between those two 180° ^{1}H

composite pulses.[135] As evident from Figure S23, we find excellent agreement between relaxation rate constants measured using either of both implementations. In our hand, the new implementation using amide selective I-BURP-2 pulses shows slightly better overall water suppression. The 2 kHz spinlock RF amplitude used at 1200 MHz seems sufficient to avoid off-resonance effects. At least we do not observe any unexpected deviations of the ^{15}N R_2 rate constants of residues with the most ^{15}N upfield or downfield resonances (Figure S24).

9.4.3.2 Application at 600 MHz

To test the applicability of the pulse sequences with sensitivity-enhanced HSQC detection on a 600 MHz spectrometer, we recorded ^{15}N R_1 and ^{15}N $R_{1\rho}$ experiments on the protonated sample, using either the HSQC or the original TROSY-detection scheme. Also, at 600 MHz, the ^{15}N R_1 rate constants show good agreement (Figure 82A-C), with a Pearson correlation coefficient of R = 0.949 and a rmsd of 0.032 (Figure 82B). For the box plot analysis, the average deviation between both data sets is -0.45% and a relatively even distribution around that is observed (Figure 82C), pointing to a stochastic experimental error.

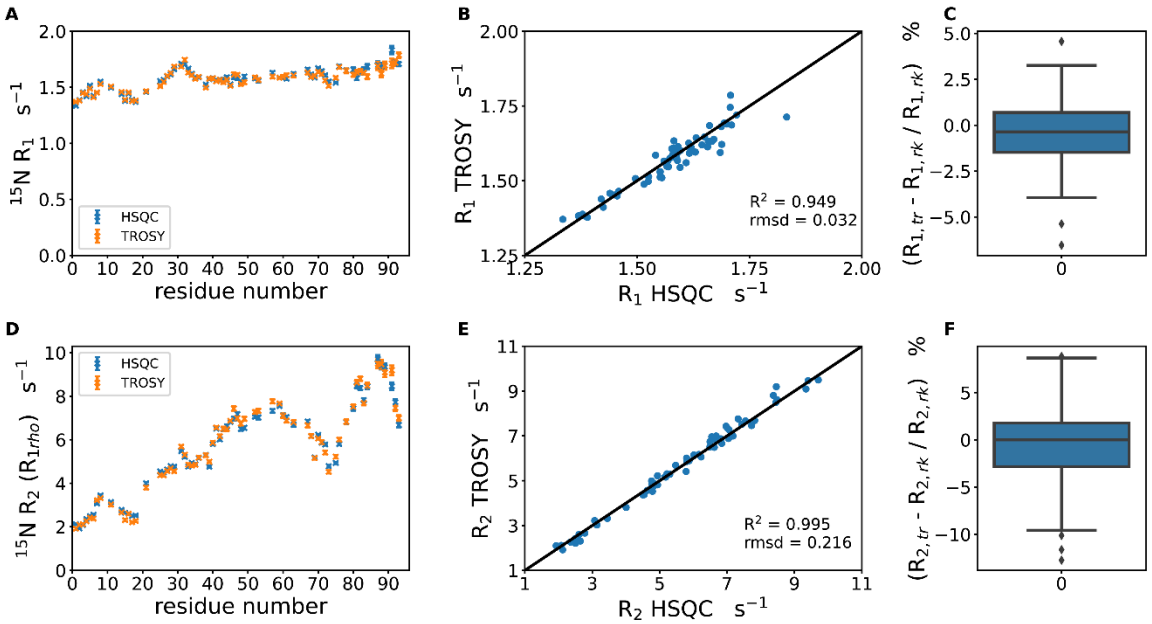


Figure 82: Comparison of relaxation rate constants of protonated ^{15}N Syb-2 (1-96) at 600 MHz, acquired with the pulse schemes in Figures 1 A, B, using sensitivity-enhanced HSQC detection and, for comparison, with the TROSY-detection. (A) Residue-specific R_1 relaxation rate constants comparison. **(B)** correlation plot of the R_1 relaxation rate constants. **(C)** Box plot of the residue-specific percentual differences of R_1 relaxation rates. **(D-F)** Same plots as in A-C, illustrating the comparison of R_2 ($R_{1\rho}$) rate constants.

The ^{15}N R_2 rate constants, derived from the ^{15}N $R_{1\rho}$ rate constants by formula 1, show good agreement with a Pearson correlation coefficient of R = 0.995 and a rmsd of 0.216 (Figure 82E). ^{15}N R_2 data obtained using the TROSY-detection scheme are, on average -0.43% lower than the ones obtained using the sensitivity-enhanced HSQC-detection scheme. For both experiments the average deviation lies within the experimental error margin (between 0.5% and 2% for HSQC and 0.5% and 3% for TROSY).

9.4.3.3 Signal-to-Noise Comparison

TROSY-detection schemes offer several advantages, such as excellent water suppression (see above) due to fewer pulses and the absence of proton decoupling. Also, in terms of line widths, TROSY-detection will yield smaller line widths and, therefore, better resolution compared to HSQC-detection because of the partial cancellation of the transverse auto-relaxation and cross-correlated relaxation, resulting in slower transverse relaxation and in turn reduced line widths. In addition, in principle, longer acquisition times in the direct dimension can be afforded due to the absence of composite pulse decoupling during the acquisition period, so resolution is not limited by the acquisition time.

For higher molecular weight proteins (> 20 kDa), TROSY will also win over HSQC regarding signal-to-noise. However, for small and medium-sized proteins, in particular, at low concentrations, the sensitivity-enhanced detection scheme offers a higher SNR. Table 42 shows a signal-to-noise comparison between the HSQC and TROSY-detection schemes. For fully protonated Syb-2, which is an IDP, we find an approximately two-fold increase in the SNR when using the sensitivity-enhanced HSQC scheme compared to the TROSY-detection scheme, which agrees with the maximum theoretically expected improvement of the HSQC vs TROSY-detection (which is two-fold), based on detection of the H^- operator of the sensitivity-enhanced HSQC detection compared to detection of the H^-N^β operator for (sensitivity-enhanced) TROSY detection. However, we also find slower ^1H T_1 relaxation of the H^-N^β TROSY line used for detection than the H^- operator in the HSQC scheme (see Figure S25). We estimated ^1H T_1 (TROSY) = 1.24 s, compared to ^1H T_1 (HSQC) = 0.86 s at 1200 MHz, which further enhances the SNR in the HSQC-detection scheme, given the inter-scan recovery delay of 2 s, used in both experiments.

Table 42: Signal-to-noise comparison for the ^{15}N R_1 and R_2 ($R_{1\rho}$) NMR relaxation experiments recorded on ^{1}H ^{15}N Syb-2, using the sensitivity-enhanced HSQC-detection or for comparison using the TROSY-detection scheme. (A) The SNR is calculated by taking the average of the three most intense resonances. For comparison, the "weak" SNR is also calculated by taking the average of the three least intense ("weakest") resonances **(B)** to get a better impression of the dynamic range.

A

Experiment	NS	read out	field / MHz	SNR (strong)	extrapolated SNR (at 8 NS)[1]
$R_{1\rho}$	4	HSQC	1200	688	972
$R_{1\rho}$	8	TROSY	1200	388	388
R_1	4	HSQC	1200	717	1014
R_1	8	TROSY	1200	404	404
$R_{1\rho}$	4	HSQC	600	332	469
$R_{1\rho}$	8	TROSY	600	237	237
R_1	4	HSQC	600	337	477
R_1	8	TROSY	600	228	228

B

Experiment	NS	read out	field / MHz	SNR (weak)	extrapolated SNR (at 8 NS)[1]
$R_{1\rho}$	4	HSQC	1200	47	66
$R_{1\rho}$	8	TROSY	1200	35	35
R_1	4	HSQC	1200	55	78
R_1	8	TROSY	1200	38	38
$R_{1\rho}$	4	HSQC	600	65	91
$R_{1\rho}$	8	TROSY	600	57	57
R_1	4	HSQC	600	71	100
R_1	8	TROSY	600	57	57

[1] Extrapolated SNR: As the HSQC-detected schemes were recorded with four scans only, their measured SNR was multiplied by a factor $\sqrt{2}$, to be comparable to the TROSY-detection schemes, which were recorded with eight scans.

9.4.4 Relaxation Rate Constants of Deuterated vs. Protonated ^{15}N Syb-2 (1-96)

9.4.4.1 ^{15}N R_1 Experiment

Figures 81 and 82 show the rate constants measured on a fully protonated ^{15}N Syb-2 (1-96) sample. In principle, the N-H amide group can be considered an isolated two-spin system in good approximation. However, we cannot exclude a residual interaction with the surrounding proton network. Therefore, measuring relaxation rate constants on a deuterated sample, with aliphatic side chains deuterated, and amide protons and exchangeable side-chain protons fully back-exchanged to protons (in a protonated buffer) is preferable in principle.[382,383,384] However, deuteration is cost-extensive, and frequently, deuteration puts an additional

burden on the biochemical sample preparation. We, therefore, tested how well the rate constants agree for a fully protonated versus a deuterated sample, with amide protons fully back-exchanged (Figure 83). Figure 83A shows a residue-wise comparison of the measured ^{15}N R_1 rate constants of the fully protonated ^{15}N Syb-2 (1-96) sample (compare Figure 81A) with those measured on the deuterated $^{2}H^{15}N$ Syb-2 (1-96) sample (of equal concentration).

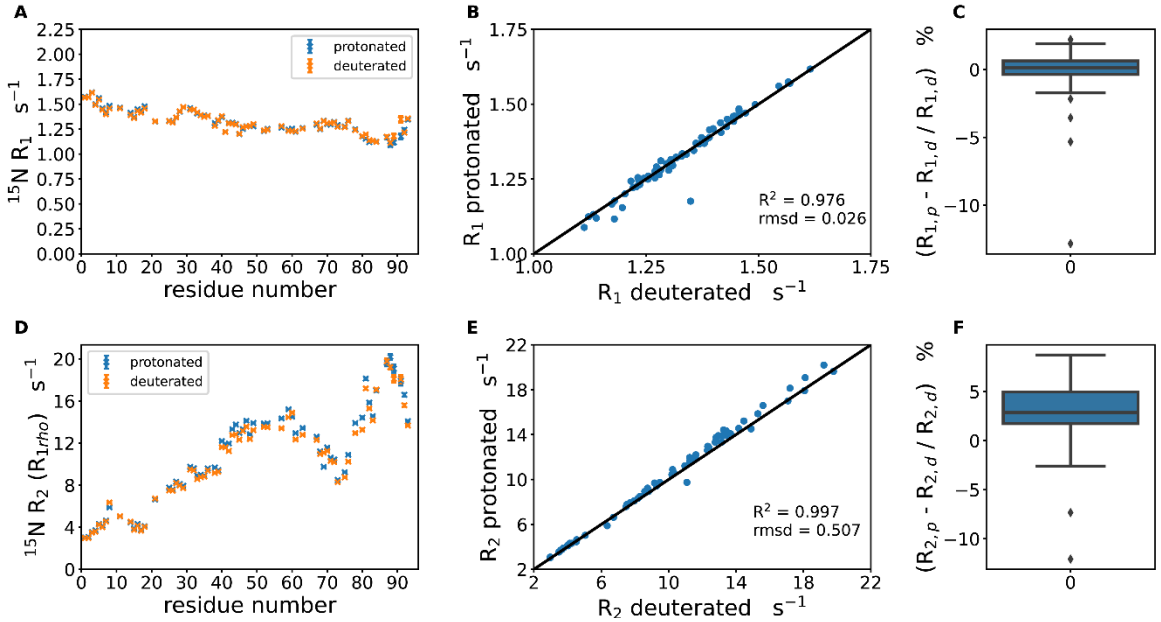


Figure 83: Comparison of relaxation rate constants of deuterated ^{2}H-^{15}N Syb-2 (1-96) and protonated ^{1}H-^{15}N Syb-2 (1-96) acquired with the pulse schemes in Figure 1 A, B at 1200 MHz. (A) Residue-specific comparison of ^{15}N R_1 relaxation rate constants. **(B)** Correlation plot of the compared R_1 relaxation rate constants. **(C)** Box plot of the residue-specific percentual differences of R_1 relaxation rate constants. **(D-F)** Same comparing plots as in A-C, but illustrating the comparison of R_2 ($R_{1\rho}$) rate constants.

Both samples were measured using the sensitivity-enhanced HSQC-detection scheme. On both samples, the ^{15}N R_1 rate constants follow a very similar pattern and show a high correlation, with a Pearson correlation coefficient of R = 0.976 and a root mean square deviation (rmsd) of 0.026 (Figure 83B). In Figure 83C, the difference in rate constants is measured on the protonated and deuterated samples. We find that the average deviation of the R_1 rate constants measured on the protonated sample minus those on the deuterated sample is close to zero on average. For deuterated samples, ^{1}H T_1 times increase because of the more dilute proton network. We, therefore, compared different interscan recovery delays (2s vs. 4s). Figure S26 compares the ^{15}N R_1 experiment (Figure 78A), recorded with a 2 s vs 4 s interscan recovery delay, both for the fully protonated (Figure S26 A-C) and the deuterated sample (Figure S26 D-F). The 2 s and the 4 s experiment correlate with R^2 = 0.992 (protonated) and R^2 = 0.985 (deuterated), and a rmsd of 0.016 and 0.02, respectively, indicating high

correspondence between both experiments (2 s vs 4 s interscan recovery delay). Also, data points scatter stochastically, and the average difference is close to zero. Thus, we concluded that a 2 s interscan delay is sufficient. A modified ^{15}N R_1 relaxation experiment has been suggested,[385] which maintains the transverse magnetization of aliphatic protons in a dephased state (saturated state) during the variable ^{15}N relaxation period T and uses a cosine modulated I-BURP-2 scheme that inverts both the amide and the aliphatic protons simultaneously. In our hands, ^{15}N R_1 data recorded on the fully protonated sample agree well with those on the deuterated sample, arguing against a strong impact of the magnetization state of the aliphatic protons.

9.4.4.2 ^{15}N $R_{1\rho}$ Experiment

Figure 83D compares the ^{15}N R_2 rate constants measured using the HSQC vs the TROSY detection scheme. While we find a Pearson correlation coefficient of R = 0.997 and rmsd of 0.507 (Figure 83E), indicating a close correlation, we also observe an offset of, on average, 2.79% higher values observed for the protonated sample (Figure 83F). More structure/rigid residues show a more substantial deviation. Deuteration of protein side chains may have a minimal impact on protein dynamics,[386] e.g., the C-D bond length is marginally shorter than the C-H bond lengths (approx. 0.005 Å).[387] However, this impact is subtle and less clear than dissolving the protein in a buffer of heavy water (D_2O).[386,388] Here, we work with a deuterated protein dissolved in H_2O buffer (with only 5% D_2O added for reference). The slightly reduced C-D bond lengths may lead to an overall marginally smaller hydrodynamic radius, affecting the overall tumbling correlation time. This may explain the slightly lower ^{15}N R_2 rate constants observed for the deuterated sample.

9.4.5 Deuterated Sample: Comparison between TROSY- and HSQC-Detection Scheme

To test whether, for the deuterated Syb-2 sample, the TROSY-detection scheme yields the same rate constants as when using a sensitivity-enhanced HSQC read-out scheme, we repeated the analysis described for the protonated sample also for the deuterated sample.

9.4.5.1 Application at 1200 MHz

Figure 84 compares the ^{15}N R_1 and $R_{1\rho}$ experiments recorded with the TROSY-detection scheme and the sensitivity-enhanced HSQC-detection scheme. Data were recorded at

1200 MHz. In Figure 84A, both R_1 data sets show only minor, non-systematic differences and follow a similar pattern. The correlation plot also shows a strong correlation between both data sets (Figure 84B). As demonstrated in the box plot in Figure 84C, ^{15}N R_1, and $R_{1\rho}$ rate constants are, on average, 0.59% and 0.80% higher for the HSQC-detected experiments vs. the TROSY-detected ones. An even stronger correlation between both pulse schemes is observed for the R_2 datasets (Figure 84 D-F), underlined by a correlation coefficient R = 0.997 and rmsd of 0.359.

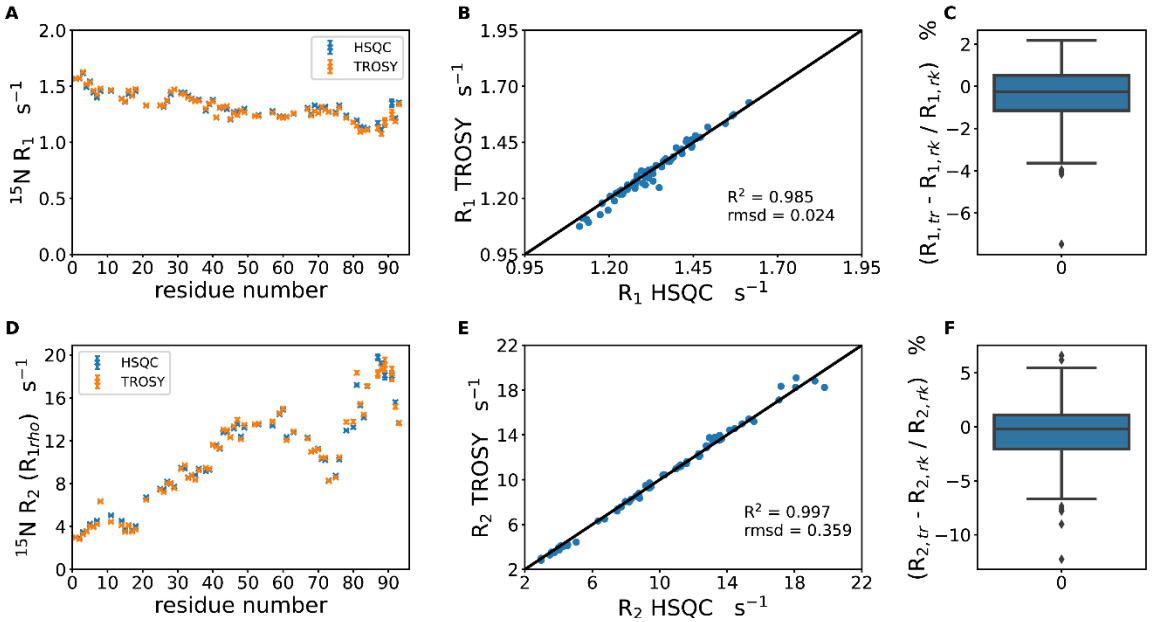


Figure 84: Comparison of relaxation rates of deuterated ^{2}H-^{15}N Syb-2 (1-96) acquired with the pulse schemes in Figure 1 A, B at 1200 MHz and of deuterated ^{2}H-^{15}N Syb-2 (1-96) obtained with the TROSY read-out schemes[135] at 1200 MHz. (A) Residue-specific R_1 relaxation rate comparison. **(B)** Correlation plot of the compared R_1 relaxation rate constants. **(C)** Box plot of the residue-specific percentual differences of R_1 relaxation rate constants. **(D-F)** Same plots as in A-C, but illustrating the R_2 ($R_{1\rho}$) rates comparison.

9.4.5.2 Application at 600 MHz

We also recorded the ^{15}N R_1 and ^{15}N $R_{1\rho}$ experiments with sensitivity-enhanced HSQC detection and, for comparison with TROSY-detection at 600 MHz spectrometer, using the deuterated sample. Also, at 600 MHz, the ^{15}N R_1 rate constants show good agreement (Figure 85A-C), with a Pearson correlation coefficient of R = 0.955 and a rmsd of 0.031 (Figure 85B). For the box plot analysis, a close to zero average deviation between both data sets and a relatively even distribution around that is observed (Figure 85C), pointing to a stochastic experimental error. The ^{15}N R_2 rate constants, derived from the ^{15}N $R_{1\rho}$ rate constants by

formula 1, show good agreement with a Pearson correlation coefficient of R = 0.994 and a rmsd of 0.214 (Figure 85E) and close to zero average deviation.

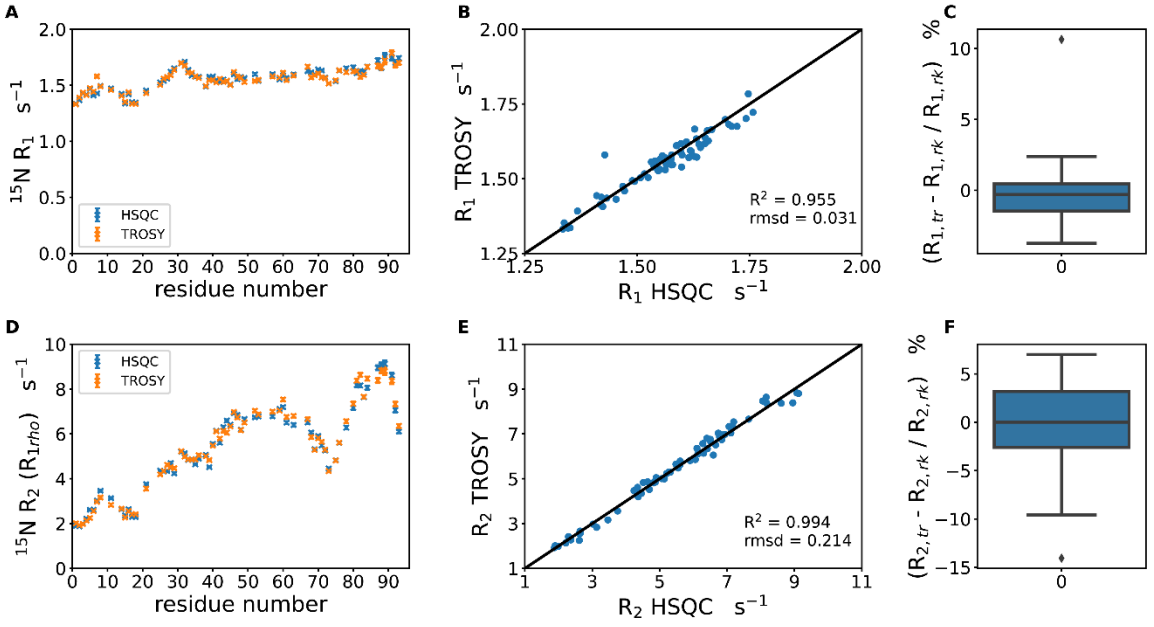


Figure 85: Comparison of relaxation rates of deuterated ^{2}H-^{15}N Syb-2 (1-96) acquired with the pulse schemes in Figure 1 A, B at 600 MHz and of deuterated ^{2}H-^{15}N Syb-2 (1-96) obtained with the TROSY read-out schemes[135] at 600 MHz. (A) Residue-specific R_1 relaxation rate comparison. **(B)** Correlation plot of the compared R_1 relaxation rate constants. **(C)** Box plot of the residue-specific percentual differences of R_1 relaxation rate constants. **(D-F)** Same plots as in A-C, but illustrating the R_2 ($R_{1\rho}$) rates comparison.

9.5 Conclusions

Advances in magnet technology have led to higher available field strengths, which have improved spectral resolution. Hence, new opportunities are offered, especially regarding the investigations of IDPs. Recently, NMR relaxation measurements could provide novel insights into fast internal protein dynamics of intrinsically disordered proteins.[110,294,330,389,390,391,392,393,394,395,396,397,398,399,400,401] Temperature-dependent NMR relaxation experiments have allowed for identifying different dynamics modes in an intrinsically disordered protein[390,391] or examining the effect of crowding on IDP dynamics [324,389,391] Detailed atomic-resolution insights on IDP dynamics can be obtained when NMR relaxation measurements are combined with molecular dynamics simulations[110,402] or further experimental probes, such as e.g. fluorescence or SAXS measurements, are included.[293,403,404] We established sensitivity-enhanced HSQC-based R_1 and $R_{1\rho}$ relaxation pulse sequences with improved water

suppression and tested them at 14.1 T (600 MHz ^{1}H Larmor frequency) and 28 T (1200 MHz ^{1}H Larmor frequency). We showed that relaxation rate constants obtained using these pulse sequences correspond well with those obtained using the TROSY-based pulse schemes introduced previously.[135] These pulse schemes reduce systematic errors from partial water saturation and cross-correlated relaxation. As we observe high robustness and reproducibility of the ^{15}N R_1 and $R_{1\rho}$ experiments both at 600 MHz and 1200 MHz, we conclude that both the sensitivity-enhanced HSQC-detected experiments as well as the TROSY-detected experiments can be applied for any magnetic field in between, ranging between 600 MHz and 1200 MHz. Further, we showed a high correlation and high agreement of relaxation rate constants between a deuterated and non-deuterated protein sample, using both sets of pulse sequences (sensitivity-enhanced HSQC and TROSY-based). Our data demonstrate that reliable relaxation rate constants can be obtained on fully protonated samples, and deuteration (for getting isolated ^{1}H-^{15}N spin pairs) is not required per se. Consequently, the subset of proteins accessible to a backbone structural dynamic investigation is enlarged, simultaneously reducing the costs of NMR sample preparation. The presented NMR relaxation sequences sensitivity-enhanced HSQC-detected are not limited to IDPs but will work equally well for small and medium-sized proteins.

9.5.1 Which Detection Scheme to Use? TROSY or HSQC?

The $H^{-}N^{\beta}$ TROSY-line detected in the TROSY-based read-out scheme shows slower relaxation (and sharper linewidth) than the H^{-} line detected in the sensitivity-enhanced HSQC experiments. (The H^{-} operator results from applying the H^{+} detection operator on the H_x or H_y operator, respectively.). The slower relaxation of the TROSY line originates from the partial cancellation of the R_2 auto-relaxation and the transverse cross-correlated relaxation of N-H dipolar coupling and the ^{15}N chemical shift anisotropy.[241] This partial cancellation becomes effective at ^{1}H Larmor frequencies above 500 MHz and reaches a maximum of around 1000 MHz.[242,405] The corresponding line widths will also be smaller because of the TROSY line's slower relaxation. Therefore, TROSY detection is always preferable in terms of line width. However, as only one doublet component is detected, the signal intensity will only be half if relaxation is neglected. Therefore, for a small and highly dynamic protein (which is characterized by small transverse relaxation rate constants), such as the IDP investigated here, the HSQC signal intensity will be twice, as both the $H^{-}N^{\beta}$ and $H^{-}N^{\alpha}$ line will add up and be

detected in the decoupled HSQC spectra. The resulting signal intensity, HSQC vs TROSY, will now be a trade-off between the double signal intensity of the HSQC spectra (in the absence of relaxation) and the TROSY line (H^-N^β in the direct ^1H dimension and N^-H^β in the indirect ^{15}N dimension, respectively) becoming gradually stronger for increased transverse relaxation. At some point, the intensity of the TROSY line will become more intense than the HSQC line (which is the average of the TROSY and anti-TROSY line because of $H^-N^\beta + H^-N^\alpha = 2\ H^-$). This is the breakeven that will be reached when TROSY outperforms HSQC, not only in terms of line width but also in terms of SNR. For a fully protonated rigid globular folded protein, the breakeven will be reached at a MW of about 20-30 kDa. For even higher MWs, deuteration of the protein is recommended to dilute the proton network and reduce transverse relaxation due to ^1H-^1H dipolar couplings with the surrounding proton network. This breakeven for proteins with high internal dynamics will be shifted towards higher MWs as fast internal dynamics will lead to lower transverse relaxation rates.

In the following, we will give a brief guideline on when to choose the TROSY and when to select the sensitivity-enhanced HSQC detected ^{15}N R_1 and ^{15}N $R_{1\rho}$ experiments:

1. When the main priority is to obtain the highest resolution (smallest line widths), the TROSY-detected experiments will always be the best option.

2. TROSY-detected experiments will also always give the best water suppression.

3. When a high signal-to-noise ratio (SNR) is the main criterion, it will be a judgment call, depending on the following points: (A) The molecular weight of the investigated biomolecule, (B) the (expected) internal dynamics of the studied biomolecule, (C) the magnetic field the sample is measured at, and (C) whether a fully protonated or a deuterated sample is used.

(A) Below 20 kDa, we recommend using the sensitivity-enhanced HSQC detection when SNR is the main criterion.

(B) Proteins with high internal dynamics (like IDPs) will give more intense resonances due to smaller transverse relaxation. Therefore, for IDPs, the HSQC detection may be beneficial up to 30 kDa.

(C) Below 800 MHz, the TROSY effect will be less effective. Therefore, the breakeven point for TROSY may be shifted towards higher MW.

(D) the TROSY effect will be more effective for deuterated proteins, shifting the breakeven points to lower MW.

This brief guideline can, however, only be over-simplified because all factors A to D contribute and will influence each other. Therefore, we strongly recommend to follow the following empirical approach:

TROSY will always win in terms of line width. If SNR is the main criterion, we suggest the following approach:

1. Set up both the TROSY-detected and the HSQC-detected NMR relaxation experiments. (With the easier-to-optimize HSQC-detected NMR experiments, this should be easy.)
2. Record the first FID of each sequence. The corresponding 1D spectrum (after the Fourier-Transformation of the first FID) will yield a good impression of the expected SNR. Choose the detection scheme that offers higher SNR.
3. To estimate the maximum duration that should be sampled, record, e.g., the second FID with the most extended relaxation delay period. If possible, the entire exponential decay curve down to 30% of the initial signal intensity should be sampled.
4. For the ^{15}N $R_{1\rho}$ experiment, be careful not to exceed the (probe and spectrometer-specific) maximum power and duration of the CW irradiation during the spinlock period of the relaxation delay.

9.6 Author Contributions

Tobias Stief: Investigation, Methodology, Formal Analysis, Writing. **Katharina Vormann: Resources (Sample preparation), Writing.** Nils-Alexander Lakomek: Conceptualization, Methodology, Writing, Supervision, Funding acquisition.

9.7 Acknowledgments

We thank Robin Backer, Lothar Gremer, Filip Hasecke, Nina Kirchgässler, Anne Pfitzer, Celina M. Schulz, Nasrollah Rezaei-Ghaleh, and Marie Schützmann for helpful discussions. N.L. thanks the German Science Foundation for funding through the Heisenberg Program (DFG grant number 433700474). This work is further supported by the project “Virological and immunological determinants of COVID-19 pathogenesis – lessons to get prepared for future pandemics (KA1-Co-02 “COVIPA”), a grant from the Helmholtz Association's Initiative and Networking Fund. We acknowledge generous access to the Jülich-Düsseldorf Biomolecular

NMR Center, jointly run by Forschungszentrum Jülich and Heinrich Heine University Düsseldorf (HHU).

9.8 Supplementary Information

Table S19: Experimental parameters of ^{15}N NMR relaxation experiments at 1200 MHz.

	R_1	**$R_{1\rho}$**
Field / T	28.2	28.2
1H Larmor frequency / MHz	1200	1200
Pulse program	Figure 1	Figure 1
Interscan recovery delay / s	2	2
Number of scans	4	4
Total experimental time	12 h 12 min	10 h 19 min
Direct dimension (^{1}H)		
^{1}H carrier / ppm	4.690	4.693
Spectral width / ppm	16.03	16.03
Data points	4096	4096
Acquisition time / ms	106.5	106.5
90° (^{1}H) pulse length / µs	8.0	8.0
Indirect dimension (^{15}N)		
^{15}N carrier / ppm	117.00	117.00
Spectral width / ppm	30.02	30.02
Data points (total)	4096	4096
Acquisition time / ms	70.14	70.14
90° (^{15}N) pulse length / µs	34.8	34.8
Relaxation period		
# relaxation delays	8	8
Time points / ms	0, 960, 240, 800, 160, 640, 320, 480	5, 65, 25, 45, 15, 55, 35, 10

Table S20: Experimental parameters of ^{15}N NMR relaxation experiments at 600 MHz.

	R_1	**$R_{1\rho}$**
Field / T	14.1	14.1
1H Larmor frequency / MHz	600	600

pulse program	Figure 1	Figure 1
Interscan recovery delay / s	2	2
Number of scans	4	4
Total experimental time	12 h 23 min	10 h 29 min
Direct dimension (^{1}H)		
^{1}H carrier / ppm	4.692	4.692
Spectral width / ppm	16.02	16.02
Data points	2048	2048
Acquisition time / ms	106.50	106.50
90° (^{1}H) pulse length / µs	8.75	8.75
Indirect dimension (^{15}N)		
^{15}N carrier/ppm	117.00	117.00
Spectral width / ppm	30.00	30.00
Data points (total)	4096	4096
Acquisition time / ms	140.29	140.29
90° (^{15}N) pulse length / µs	43	43
Relaxation period		
# relaxation delays	8	8
Time points / ms	0, 960, 240, 800, 160, 640, 320, 480	5, 65, 25, 45, 15, 55, 35, 10

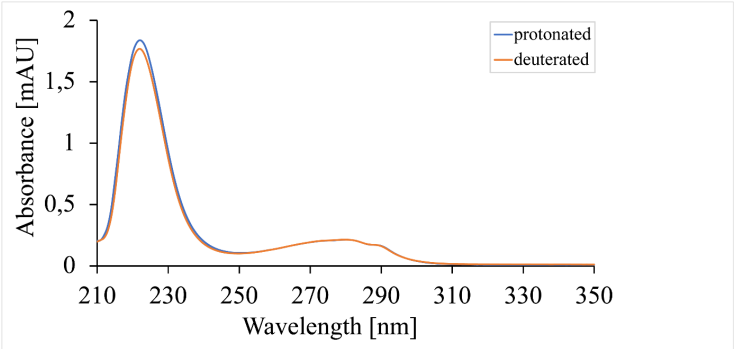


Figure S21: UV-vis spectrum of deuterated ^{2}H-^{15}N-^{13}C Syb-2 (1-96) and protonated ^{1}H^{15}N Syb-2 (1-96). Protein concentrations were calculated using the Beer-Lambert's Law (Molar extinction coefficient: 12490 $M^{-1}cm^{-1}$; path length: 1 cm; dilution factor: 10; A_{280} of deuterated ^{2}H-^{15}N Syb-2 (1-96):

0.21441475; A_{280} of protonated ^{1}H-^{15}N Syb-2 (1-96): 0.2137355). According to the equation, the final sample concentrations have been 171.7 µM of deuterated ^{2}H-^{15}N Syb-2 (1-96) and 171.1 µM of protonated ^{1}H-^{15}N Syb-2 (1-96) in 50 mM MES (pH 6.0), 150 mM NaCl, 1 mM EDTA and 0.1 mM TCEP.

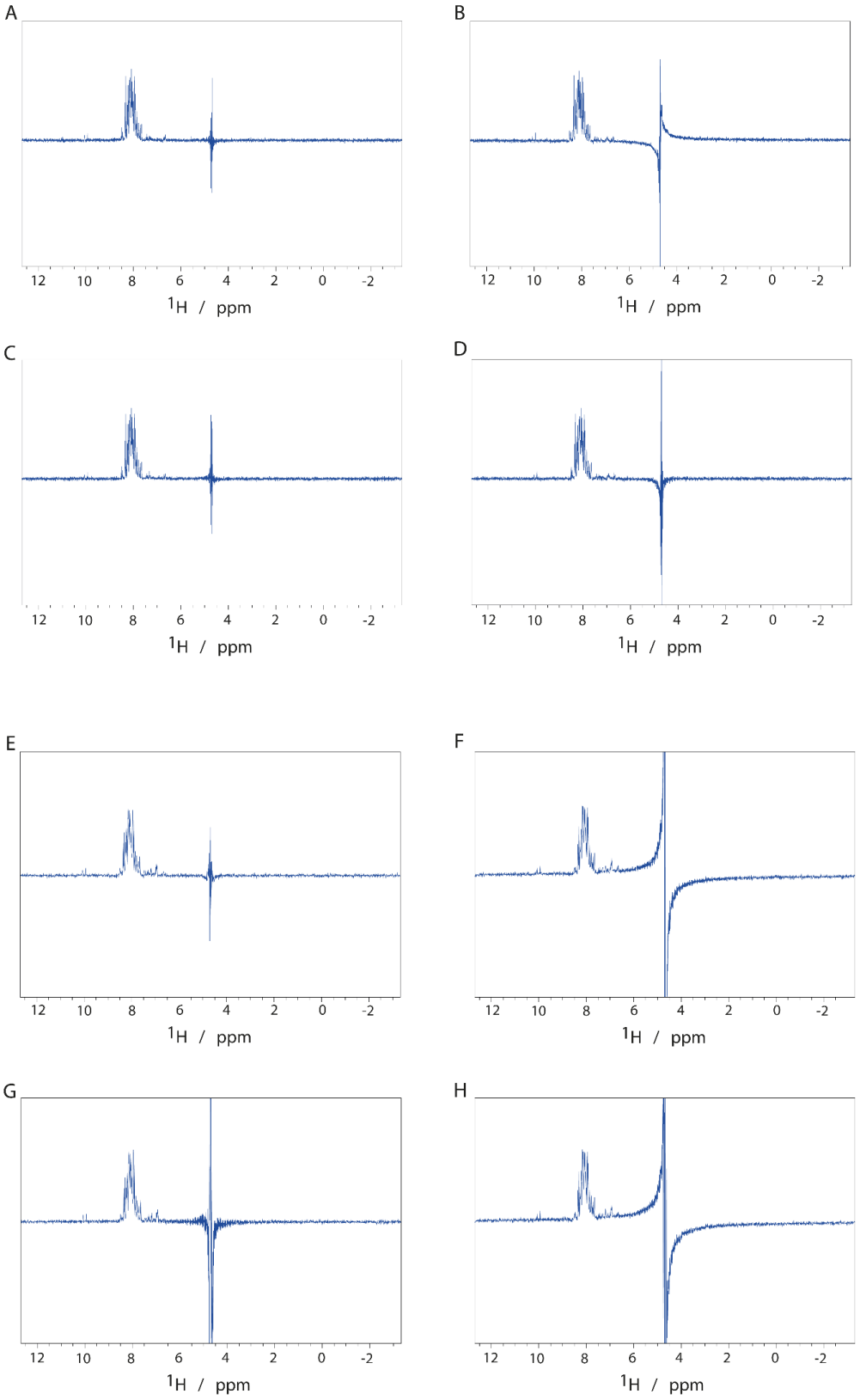


Figure S22: Improvement of water suppression in ^{15}N R_1 and ^{15}N $R_{1\rho}$ NMR relaxation experiments with sensitivity-enhanced HSQC-detection using the new implementation (left, see also Figure 1) and compared to the old implementation (right, SI Figure S1 in [135]). Shown is the 1D spectrum

obtained from the first FID recorded on the deuterated ^{15}N Syb-2 sample. **(A)** Optimized ^{15}N $R_{1\rho}$ experiment at 1200 MHz compared to **(B)** previous implementation. ^{15}N R_1 experiment at 1200 MHz: **(C)** New and **(D)** old. ^{15}N $R_{1\rho}$ experiment at 600 MHz: **(E)** new and **(F)** old. ^{15}N R_1 experiment at 600 MHz: **(G)** new and **(H)** old.

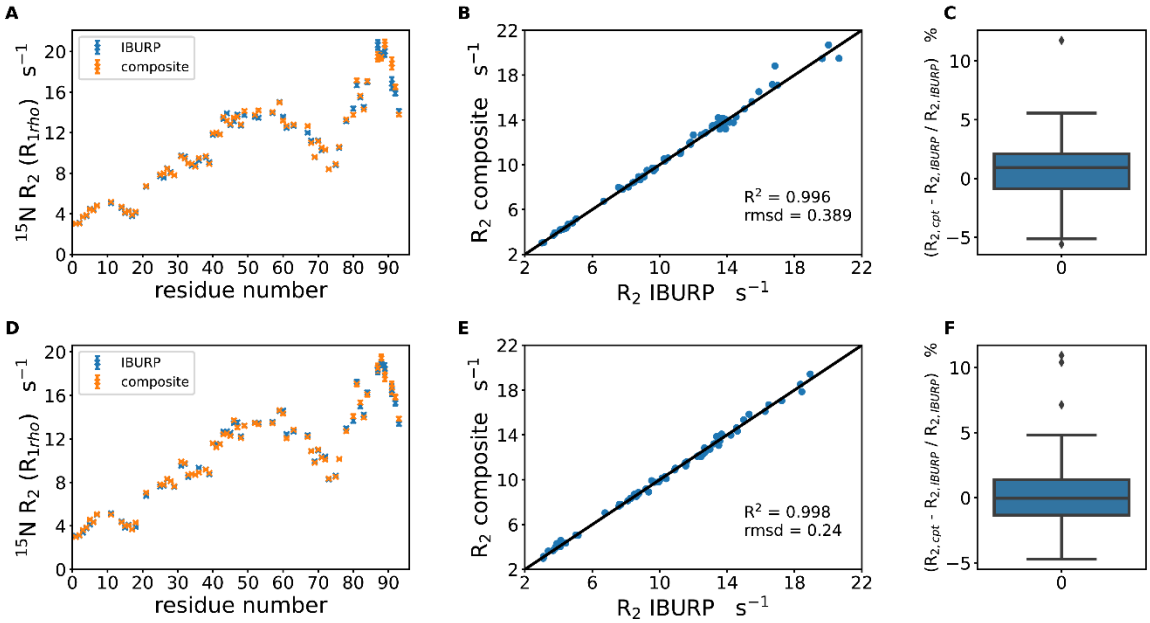


Figure S23: Comparison of R_2 relaxation rate constants of the protonated ^{1}H-^{15}N Syb-2 (1-96) at 1200 MHz, acquired with the pulse scheme in Figure 1B (IBURP) and the previous implementation to refocus cross-correlation relaxation during the $R_{1\rho}$ relaxation period, where composite pulses were used plus and additional weak gradient in between (that alternated its phase half-way), to prevent radiation damping, as described for the TROSY-based sequence.[135] **(A-C)** Protonated sample. **(D-F)** Deuterated sample.

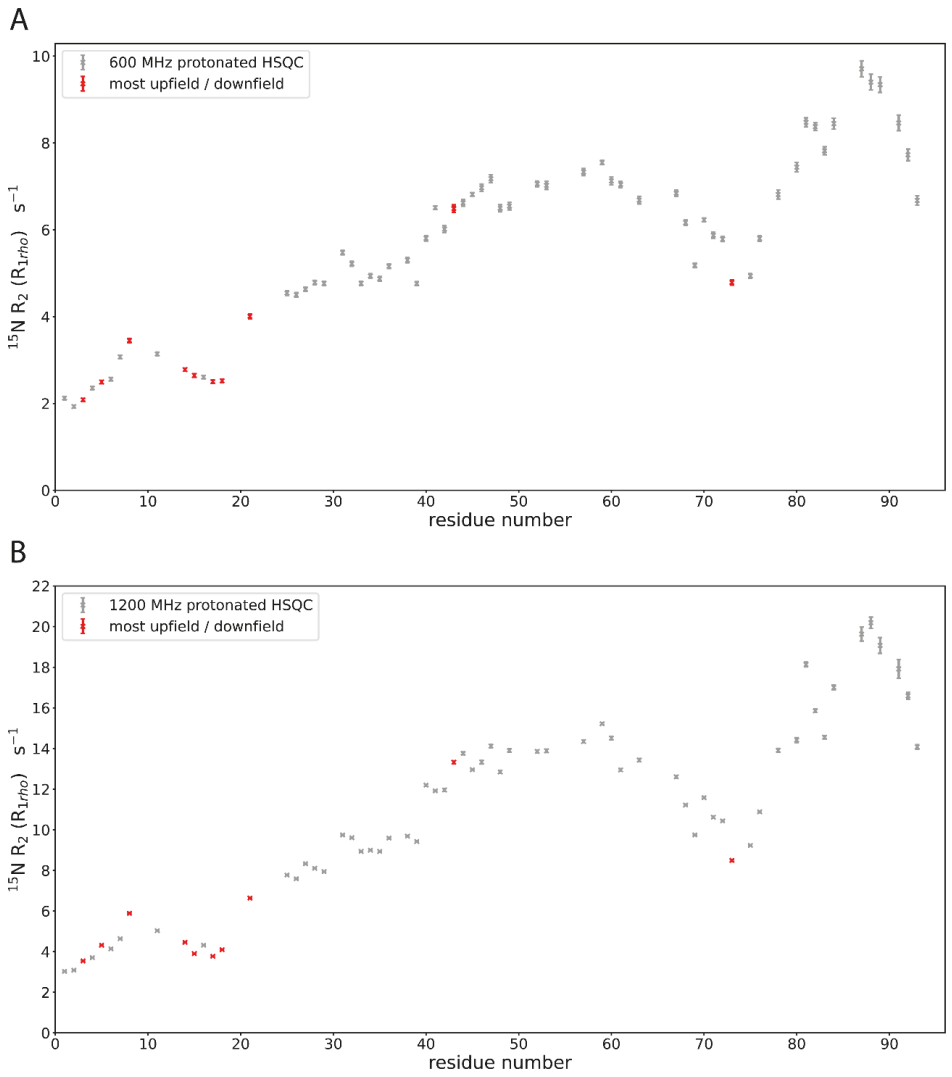


Figure S24: **^{15}N R$_2$ rate constants recorded at 600 MHz and 1200 MHz using the optimized sensitivity-enhanced HSQC detection scheme and a 2 kHz RF amplitude for the spinlock.** Rate constants for residues with the most ^{15}N upfield and downfield resonances are colored in red: **(A)** 600 MHz and **(B)** 1200 MHz for comparison.

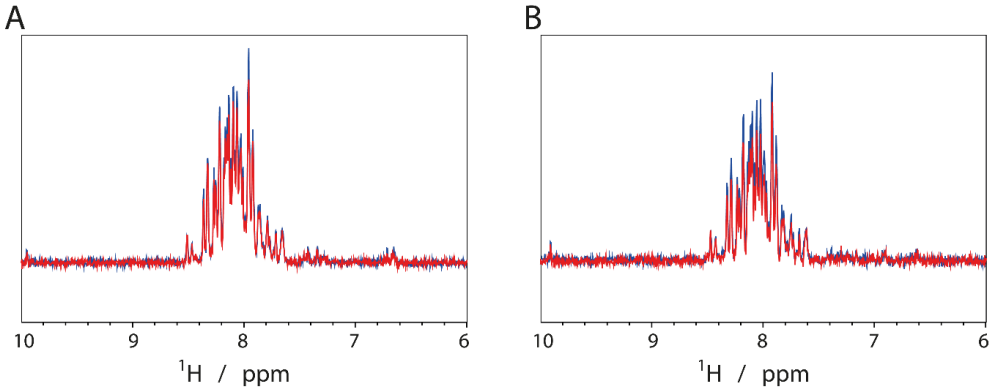


Figure S25: Intensity comparisons of the 1D spectrum obtained from the first FID of the ^{15}N R$_1$ experiments recorded on the protonated sample ^{15}N Syb-2 sample at 1200 MHz. (A) HSQC-detected experiment vs **(B)** TROSY detected experiment. The red spectrum indicates the intensity after d1=2s recovery delay, the blue spectrum after 4s recovery delay.

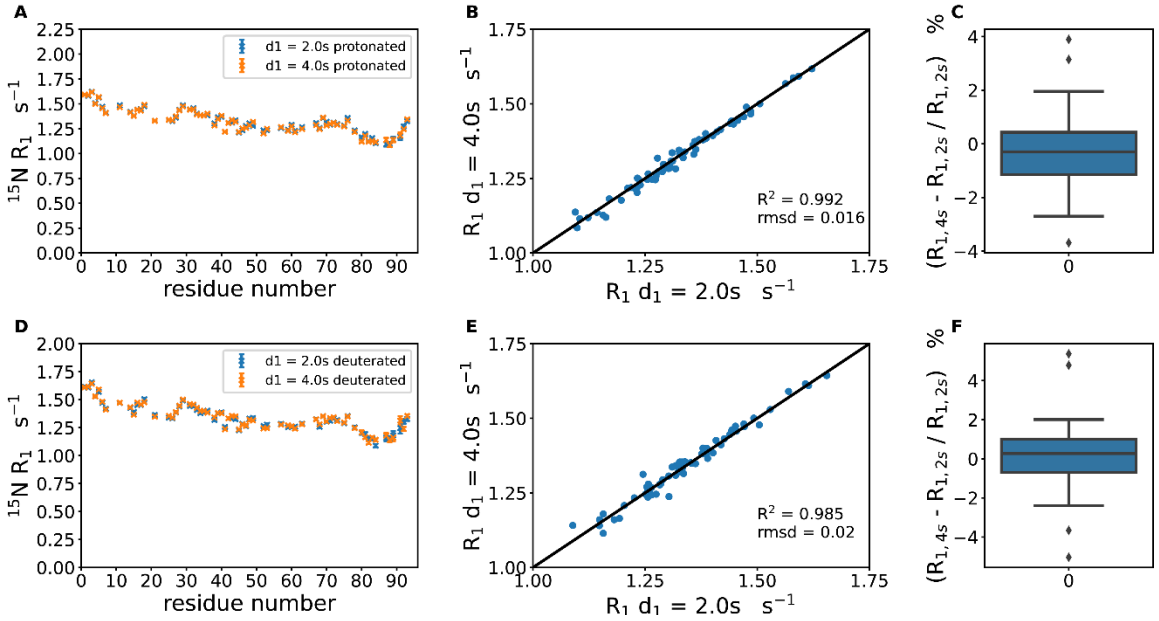


Figure S26: No impact of longer inter-scan recovery times on the derived R_1 relaxation rate constants. Experiments were conducted at 1200 MHz and 5°C, using the pulse scheme shown in Figure 1A with an interscan recovery delay of 2.0 s and 4.0 s. **(A)** Residue-specific R_1 relaxation rate comparison. **(B)** Correlation plot of the R_1 relaxation rate constants. **(C)** Box plot of the residue-specific percentual differences of R_1 relaxation rates. In **A-C**, the data for protonated ^{1}H-^{15}N Syb-2 (1-96) are shown, and in **D-F,** the data for deuterated ^{2}H-^{15}N-^{13}C Syb-2 (1-96).

Table S21. NMR relaxation data of ^{15}N Syb-2 measured at 1200 MHz and 5 °C. See the accompanied Excel table „TableS3_R1R2_1200“. The columns contain the following information for each residue (rows): # (residue number), AA type (1-letter code), transverse relaxation rate constants (derived from $R_{1\rho}$) ^{15}N $R_2(R_{1\rho})$/ s^{-1} (1200 MHz), experimental error $\delta(^{15}N\ R_2(R_{1\rho}))$/ s^{-1} (1200 MHz); longitudinal relaxation rate constants ^{15}N R_1 / s^{-1} (1200 MHz), experimental error $\delta(^{15}N\ R_1)$/ s^{-1} (1200 MHz).

Table S22. NMR relaxation data of ^{15}N Syb-2 measured at 600 MHz and 5 °C. See the accompanied Excel table „TableS3_R1R2_1200“. The columns contain the following information for each residue (rows): # (residue number), AA type (1-letter code), transverse relaxation rate constants (derived from $R_{1\rho}$) ^{15}N $R_2(R_{1\rho})$/ s^{-1} (1200 MHz), experimental error $\delta(^{15}N\ R_2(R_{1\rho}))$/ s^{-1} (1200 MHz); longitudinal relaxation rate constants ^{15}N R_1 / s^{-1} (1200 MHz), experimental error $\delta(^{15}N\ R_1)$/ s^{-1} (1200 MHz).

VIII. Discussion

Despite extensive global efforts, the molecular mechanisms underlying the structural dynamics and interaction profiles of SARS-CoV-2 proteins remain incompletely understood. Among these, NSP3 plays a central role in viral replication, particularly in the formation of DMVs and the establishment of the DMV pore complex. Although recent advances have illuminated aspects of DMV structure and function, key regions of NSP3 remain understudied. For example, structural investigations such as those by Huang *et al.* have omitted the N-terminal region of NSP3 - including the Ubl1 domain - thereby limiting insight into its potential regulatory or dynamic contributions to pore architecture and function.[46] Likewise, Bessa *et al.* examined Ubl1-N protein interactions using a truncated construct lacking the first 16 residues, precluding any conclusions about the role of Ubl1's flexible N-terminus in mediating N protein binding.[67]

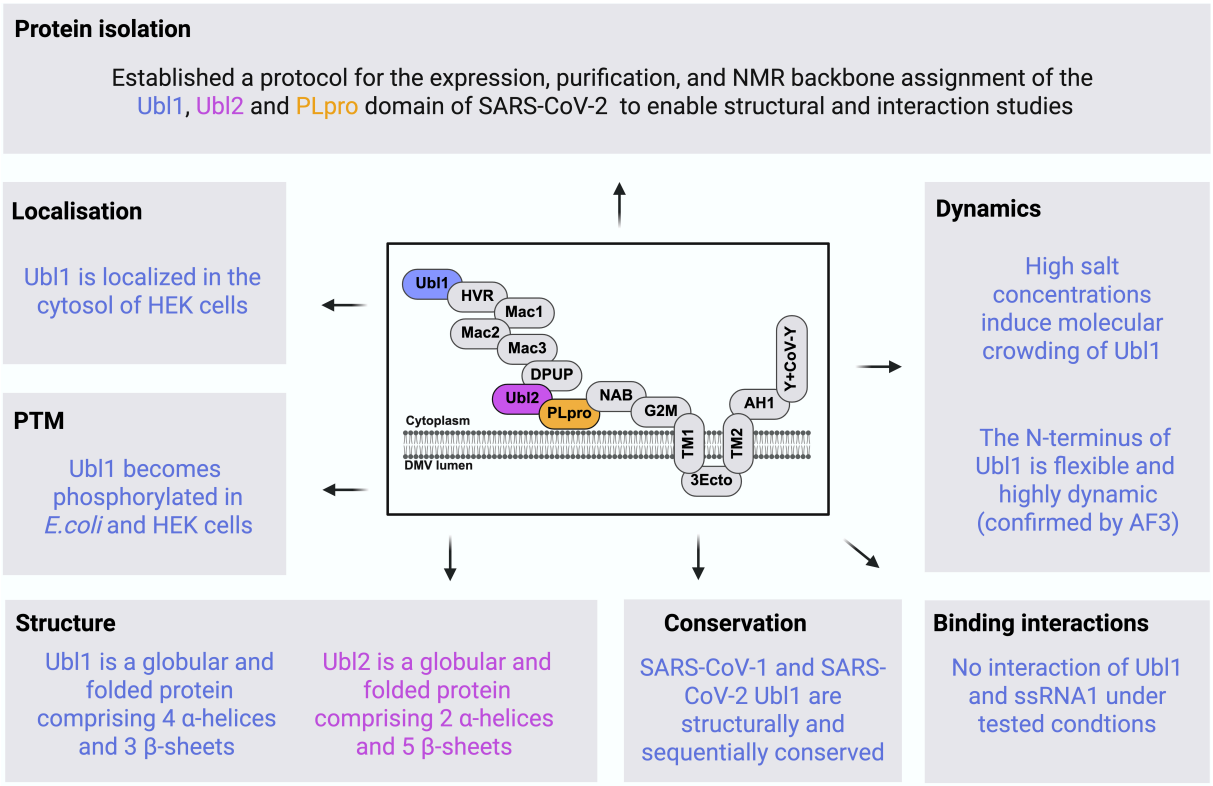


Figure 86: Summary of SARS-CoV-2 related findings in this work. Central overview of NSP3 domain organization with Ubl1 (blue), Ubl2 (purple), and PLpro (orange) highlighted; data and findings relate to protein isolation, localization, post-translational modifications (PTMs), structure, dynamics, and binding interactions, with construct-specific results color-coded accordingly. Figure was created in BioRender.com.

In the work described in this thesis, we conducted a comprehensive biophysical and cellular characterization of three key domains of NSP3: Ubl1, PLpro and Ubl2. Employing an integrative

structural biology approach - including NMR spectroscopy, SAXS, CD, MST, AUC, DLS, MS and in-cell fluorescence microscopy - we provide novel insights into the conformational behavior, dynamic properties, and interaction potentials of these domains under varying physiological and biochemical conditions. Figure 86 provides a summary of the SARS-CoV-2-related findings presented in this work.

Prior to structural analysis, I established a protocol for the isolation of pure, high-quality samples of Ubl1 (Chapter 1), Ubl2, and PLpro (Chapter 6). As PLpro, including its Ubl2 subdomain, is involved in the proteolytic cleavage of the viral polyprotein to release NSP1-NSP3 - a step essential for the formation of DMVs - both PLpro and Ubl2 represent promising drug targets for inhibiting the viral replication cycle.[45,63,54,209,406] Our results provide a platform for future interaction studies involving Ubl2 and PLpro, especially given that PLpro interferes with ULK1-mediated autophagy by dissociating its kinase and substrate-binding domains.[217,218]

Although no binding studies were performed in the present work, the performed sample production and optimization is a necessary first step and provides the basis for future NMR and biophysical studies aiming at deeper insights into functional roles of PLpro and Ubl2 in the SARS-CoV-2 replication cycle and their potential contributions to immune evasion. A structural foundation is provided by the near-complete NMR resonance assignment of Ubl2 achieved in this study (Chapter 6), which complements the previously published PLpro assignment by Shimada *et al.*[234]

Despite both Ubl1 and Ubl2 being classified as ubiquitin-like domains, our data reveal distinct differences in their secondary structures: Ubl1 consists of four α-helices and three β-sheets (Chapter 2), whereas Ubl2 is composed of two α-helices and five β-sheets and exhibits a lower molecular weight (Chapter 6). Nevertheless, both proteins were identified as globular and stably folded, exhibiting a conserved architecture across different CoVs.[79] However, the main focus of this work was on the structural, dynamic and functional characterization of Ubl1 (Chapter 1-5).

Our findings confirm that Ubl1 adopts a ubiquitin-like fold, composed of four α-helices and two β-sheets, with an intrinsically disordered N-terminal segment (residues T3-V15) (Chapter 2). This disordered region exhibited increased flexibility on the picosecond-nanosecond

timescale, as evidenced by relaxation experiments and model-free analysis. The AlphaFold3-predicted structural model of Ubl1 aligned well with experimental RDCs, while SAXS and AUC experiments verified its monomeric and globular state in solution.

Despite sharing approximately 79% sequence identity, the Ubl1 domains of SARS-CoV-1 and SARS-CoV-2 exhibit subtle but notable structural differences (Chapter 2). The crystal structure of SARS-CoV-2 Ubl1 (PDB: 7KAG) revealed a homodimeric arrangement stabilized by interactions within the folded N-terminal region, comprising four α-helices and six β-strands per monomer. In contrast, our structural analyses confirmed a conserved β-grasp fold but indicated four α-helices and only four β-strands, consistent with an intrinsically disordered N-terminal segment. This observation aligns with the AlphaFold3 (AF3) prediction for SARS-CoV-2 Ubl1, which likewise features a stable core and a flexible N-terminus. The NMR structure of SARS-CoV-1 Ubl1 (PDB: 2IDY)[87] also revealed a disordered N-terminal region, comprising three α-helices and four β-strands, while the AF3 model predicted a more folded N-terminus, resembling the SARS-CoV-2 crystal structure. Structural alignments between the experimental and predicted models yielded low RMSD values (1.85 Å), confirming high overall similarity between both viral Ubl1 domains (Chapter 2). The closest agreement was observed between the AF3-predicted Ubl1 models of SARS-CoV-1 and SARS-CoV-2 (RMSD: 0.35 Å). Together, these results indicate that Ubl1 adopts a structurally conserved core across both viruses, with variability confined mainly to the N-terminal region, which may influence its flexibility and interaction potential.

Our IF studies further revealed cytosolic localization of Ubl1 in HEK293T cells, consistent with the known cytosolic orientation of NSP3-Ubl1 at the DMV pore.[56,61] These findings align with previous work identifying interactions between Ubl1 and the N-protein both in eukaryotic cells (Huh-7) and *in vitro*, where Ubl1 binds the N-protein with high affinity in the nanomolar range.[61,67] The bipartite nature of this interaction - mediated by the intrinsically disordered N3 domain of the N-protein - suggests that Ubl1 may facilitate or modulate RNP complex assembly.[67] Bessa *et al.* described how the hydrophobic helix and polar strand of the N3 domain wrap around Ubl1, forming a stable complex.[67] Given Ubl1's position at the DMV pore, a role in RNA translocation is plausible.[56] Indeed, previous studies have demonstrated that Ubl1 from SARS-CoV-1 binds ssRNA containing an AUA motif.[87] However, we could not confirm RNA binding of Ubl1 (Chapter 5). Importantly, structural and dynamic features of Ubl1 appear

conserved between SARS-CoV-1 and SARS-CoV-2, consistent with their ~79% sequence identity, suggesting that key functional roles are maintained - particularly interactions with the N-protein.[61,67]

To investigate the potential interaction between SARS-CoV-2 Ubl1 and single-stranded RNA (ssRNA), we employed the same RNA motif previously used by Serrano *et al.* (2007) in their study of SARS-CoV-1 Ubl1.[87] NMR titration experiments revealed chemical shift perturbations at residues T24, L35, E37, and G94, suggesting possible interaction sites (Chapter 5). However, the observed binding pattern was dispersed and inconsistent along the protein, indicative of non-specific interactions within the range of statistical noise. The interaction was found to be condition-dependent, with enhanced signal perturbations observed at room temperature (25 °C), under low-salt conditions (50 mM NaCl), and in the presence of 2 mM Mg^{2+}. Correlation analysis suggested potential magnesium-related effects; however, these were non-systematic and likely stochastic in nature. Also, MST experiments did not show binding upon saturation, even at RNA concentrations up to 1 mM. We could not confirm previous reports showing low binding affinity (20 µM for SARS-CoV-1 and 31 µM for SARS-CoV-2).[87,188]

The oligomerization state of Ubl1 is another matter of debate. Therefore, we conducted concentration-dependent NMR studies (Chapter 3). Elevated R_2 relaxation rates suggested the presence of higher-order species. In contrast, AUC measurements confirmed a monomeric state under all tested redox conditions. This discrepancy is likely due to salt- and concentration-dependent molecular crowding effects, particularly involving the disordered N-terminal region. At physiological ionic strength (150 mM NaCl), electrostatic shielding reduces repulsion between charged N-terminal residues (e.g., K4, D9, D10, E14), thereby promoting transient self-association or crowding via hydrophobic residues in the β-hairpin (e.g., V5, F7, I13, V15). These effects increase local viscosity and steric hindrance, potentially impeding RNA access.[407]

In support of these findings, prior studies demonstrated condition-dependent oligomerization of Ubl1, with dimeric or trimeric assemblies observed only under high protein concentrations or when the N-terminal region was intact (e.g., PDB: 7KAG, 7TI9), while truncation of the N-terminus resulted in monomeric structures (e.g., PDB: 8XAB). These structural transitions may bear physiological relevance, as the hexameric arrangement of NSP3 in the DMV pore likely

leads to local enrichment of Ubl1 domains, creating an environment conducive to transient self-assembly.

Given that salt affects the oligomerization behavior of Ubl1 (Chapter 3), it is of particular interest that salt,[408] the protein concentration,[148] pH,[409] pressure,[410] temperature,[411] and phosphorylation are also known to regulate liquid-liquid phase separation (LLPS).[412] Such condensates create microenvironments enriched in proteins and nucleic acids.[143,144,145] Recent findings suggest that Ubl1 participates in biomolecular condensates formed via LLPS, in conjunction with viral RNA and the N-protein.[118,119,124,413] Notably, higher NaCl concentrations disrupt LLPS by screening electrostatic interactions, as shown for the SARS-CoV-2 N-protein.[118,124] In Table 43, the effects of salt onto Ubl1 based on our results and the literature are listed.

Table 43: The effect of salt onto LLPS and binding interactions based on literature and this work.

Category	High Salt	Low Salt
LLPS[148]	↓	↑
RNA-Protein interactions[143,414,147]	↓	↑
Molecular crowding of Ubl1	↑	↓
N-protein concentration in liquid droplets[118,124]	↓	↑
Kinase activity[415,416,417,418, 419,420,421]	↓	↑
Phosphatase activity[422,423,424]	↑	↓

A further regulatory layer of LLPS involves phosphorylation of Ubl1. Unexpectedly, recombinant Ubl1 expressed in *E. coli* was phosphorylated, as indicated by gel shifts and double bands (Chapter 4). The modification was resistant to λ-phosphatase treatment, suggesting stable phosphorylation or unsuitable dephosphorylation conditions. MS analysis of HA-tagged Ubl1 expressed in HEK cells revealed triple phosphorylation, although contributions from HA-tag tyrosine residues cannot be excluded. Phosphorylation likely introduces negative charges that modulate electrostatic interactions with RNA or the N-protein. The SARS-CoV-2 N-protein, phosphorylated by GSK-3, forms LLPS condensates essential for transcription,[413] implying that Ubl1 phosphorylation may similarly regulate RNA binding in a context-dependent manner. Phosphorylation of SARS-CoV-2 Ubl1 has been identified here for the first time (Chapter 4), and its effect on N-protein binding remains

unknown. Although Bessa *et al*.[67] characterized Ubl1–N interactions, the phosphorylation state of *E. coli*-expressed Ubl1 was likely unmodified. In MHV, mutation of phosphorylated serine residues in the N-protein reduced Ubl1 binding affinity threefold,[116] suggesting that phosphorylation enhances N–Ubl1 interactions. The reversible nature of this modification may fine-tune interaction strength; approximately half of purified Ubl1 appeared in a singly phosphorylated state. Predicted phosphorylation sites include clusters of serine, threonine, and tyrosine residues, and more than 500 host kinases could potentially target NSP3 domains.[412,425] Dynamic regulation of these PTMs may act as a molecular switch controlling Ubl1-mediated RNA or N-protein interactions during LLPS.

Given the known N-protein-Ubl1 interaction and N-protein-driven condensate formation,[118,119,124,413] we propose that Ubl1 also participates in LLPS. Such condensates concentrate macromolecules, accelerate reaction kinetics, and protect components from inhibitors.[143,144,145] Phase separation of RNA-binding proteins is typically driven by electrostatic interactions between positively charged residues and the RNA phosphate backbone, which can be disrupted by salt.[143] Increased ionic strength reduces SARS-CoV-2 N-protein condensation[124] and droplet protein content,[118] identifying salt as a critical regulator of LLPS.[408]

Intrinsically disordered regions (IDRs), known LLPS drivers.[122,412,426] We hypothesize, that this applies also to the disordered Ubl1 N-terminus. IDRs retain conformational flexibility within condensates and modulate phase behavior.[149] Decreased droplet dynamics upon salt addition highlight the importance of the surrounding aqueous environment.[148] Combined with phosphorylation, this dynamic property may enable Ubl1 to act as a molecular switch regulating RTC docking, RNA translocation, and N-mediated genome packaging. Phosphorylation and LLPS are further connected through kinase regulation. TBK1 senses ubiquitin condensates, and kinases often localize within such environments, promoting activation and substrate proximity.[427,428] Moreover, sodium chloride modulates kinase activity: phosphorylation increases under low-salt and decreases under high-salt conditions.[415,416,417,418, 419,420,421] For example, chloride binding to WNK1 inhibits its activity at ~200 mM NaCl,[417] and similar salt-sensing mechanisms exist in other WNK kinases.[418,419] Kinase activity is further enhanced by Mg^{2+} and ATP but inhibited by NaCl and KCl,[420,421] while alkaline phosphatase shows the opposite trend.[422,423,424] These findings suggest that

phosphorylation and dephosphorylation of Ubl1 may function as a salt-regulated molecular switch within liquid-like condensates.

However, previous structural studies have excluded the N-terminal region of Ubl1, precluding a complete understanding of its role in these interactions.[67] Our data suggest a critical role of the flexible N-terminus, not only in transient intermolecular contacts but also in promoting LLPS. Importantly, recent high-resolution cryo-EM reconstructions of the DMV pore complex omit this domain, leaving a significant gap in our understanding of early replication events.

Given that several domains within NSP3 - such as Ubl1, NAB, and the Y-domain - exhibit RNA-binding properties,[188] we propose an RNA translocation pathway through the DMV pore, terminating at the flexible N-terminal Ubl1 domain, which interacts with the viral N-protein (Figure 87). Along this path, the N-protein undergoes phase separation with viral RNA and Ubl1, resulting in a membraneless, shielded compartment (biomolecular condensate) that facilitates RNA packaging. This process appears to be regulated by phosphorylation events and salt modulating RNA-binding affinity and kinase/phosphatase activity, though the precise mechanisms require further investigation.

We hypothesize that these biomolecular condensates are composed of Ubl1, the N protein, viral RNA, and associated kinases, forming a complex multi-protein assembly. The formation and stability of these condensates are influenced by salt concentrations, indicating that the driving force for LLPS is dictated by both, the amino acid sequence and the environment.[408] Particularly NaCl affects molecular crowding of Ubl1, pore dynamics, the binding capacity of Ubl1 to RNA and the N-protein and the activity of kinases and phosphates. We propose a model in which RNA translocation and packaging mediated by Ubl1 are regulated in a salt-dependent manner (Figure 87). In this model, under low-salt conditions, RNP complex formation is initiated, facilitating the assembly of new virions. In this regime, Ubl1 monomers localized at the prongs of the DMV pore experience electrostatic repulsion due to their charged residues. This repulsion promotes interaction between Ubl1 and the viral N-protein, thereby supporting RNA translocation through the pore and subsequent RNA encapsidation. The process is further enhanced by phosphorylation of both Ubl1 and N, which is mediated by kinases, active under low-salt conditions. Additionally, LLPS contributes to the formation of a shielded, membrane-proximal environment that facilitates RNA packaging.

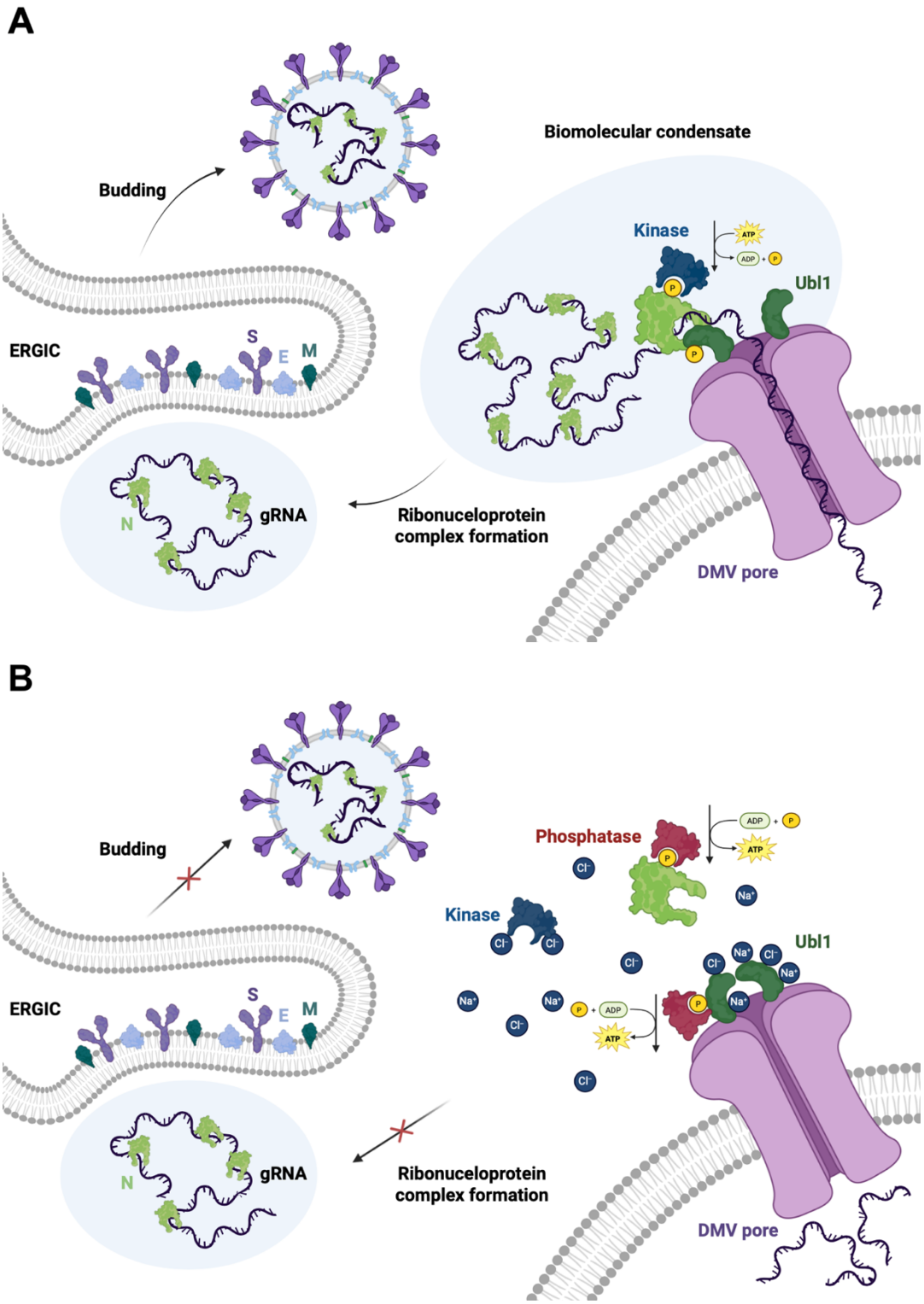


Figure 87: Proposed model of salt-dependent RNA translocation and packaging mediated by Ubl1 and the N-protein. (A) Low-salt conditions: In the absence of elevated salt concentrations, ribonucleoprotein complex formation is initiated, facilitating the assembly of new virions. Under these conditions, Ubl1 monomers located at the prongs of the double-membrane vesicle (DMV) pore experience electrostatic repulsion due to their charged residues. This promotes interaction between Ubl1 and the viral nucleocapsid (N) protein, thereby enabling RNA translocation through the pore and subsequent RNA packaging. These processes are further enhanced by the phosphorylation of Ubl1 and N, mediated by kinases, and are supported by the formation of a protective environment via liquid–liquid phase separation (LLPS). **(B)** High-salt conditions: Increased salt concentrations disrupt the

formation of biomolecular condensates, thereby inhibiting ribonucleoprotein assembly. Ubl1 becomes electrostatically shielded by sodium and chloride ions, which promotes molecular crowding. As a result, neighboring Ubl1 domains are brought into closer proximity, leading to pore closure - thus positioning Ubl1 as a gatekeeper. Phosphorylation is inhibited under these conditions, as kinases bind chloride ions, while phosphatase activity is enhanced, resulting in the dephosphorylation of both N and Ubl1. Consequently, the environment is no longer conducive to efficient RNA packaging.

In contrast, high-salt conditions disrupt the formation of biomolecular condensates and inhibit RNP assembly. In this context, in our model, Ubl1 would be electrostatically shielded by sodium and chloride ions, promoting molecular crowding and facilitating the close proximity of neighboring Ubl1 domains. This spatial arrangement leads to sealing of the pore, functionally identifying Ubl1 as a gate. At the same time, phosphorylation is suppressed due to chloride ion binding by kinases, while phosphatases become active and dephosphorylate both N and Ubl1. As a result, the conditions no longer favor efficient RNA packaging. However, how NaCl becomes enriched within these liquid compartments remains unclear and has to be investigated *in vivo*. The complexity of this supramolecular assembly exceeds initial expectations and warrants further detailed study.

Taken together, our results underscore the influence of environmental conditions - including ionic strength, phosphorylation, and molecular crowding - on Ubl1's interaction with the N-protein. These findings lead to the hypothesize of the discussed model that reconciles discrepancies in the literature and highlight the necessity of considering full-length constructs and physiological conditions in future studies. Since the structure of Ubl1 and Ubl2 and further the substrate specificity of PLpro are highly conserved in SARS-CoV-1 and SARS-CoV-2,[211] these domains are promising targets for broad-spectrum antivirals and pandemic preparedness. Although significant progress has been made, the specific function of the N-terminal Ubl2 of PLpro is still not fully understood and requires further study.[205] Clarifying the role of Ubl2 could enhance our understanding of how PLpro integrates its enzymatic and regulatory functions, potentially guiding the design of targeted inhibitors to counteract its immunomodulatory activity. Understanding the functional regulation of Ubl1 within ROs remains essential for elucidating the life cycle of SARS-CoV-2 and may inform future antiviral strategies.

IX. Supplementary Data

Table SD1: Sequencing results of pET28a(+)-constructs.

Construct	Primer	Sequence Result
pET28a(+)-6xHis-TEV-Thrombin-Ubl1	T7	CNNCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGC AGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATA TGGAAAACCTGTATTTCCAGGCCCCGACCAAAGTTACCTTTGGTGACGATACC GTTATTGAAGTTCAGGGTTATAAAAGCGTTAATATTACCTTTGAGCTGGATGA ACGCATTGATAAAGTTCTGAATGAAAAATGTAGCGCCTATACCGTTGAACTG GGTACCGAAGTGAATGAATTTGCATGTGTTGTTGCAGATGCCGTTATTAAGA CCCTGCAGCCGGTTAGTGAACTGCTGACCCCGCTGGGTATTGATCTGGATGA ATGGAGCATGGCCACCTATTATCTGTTTGATGAAAGTGGCGAATTCAAACTG GCCAGCCACATGTATTGCAGTTTTTATCCGCCGGATGAATAAGGATCCGAATT CGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCA CTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCC ACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGA GGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGC GCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGT GACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTC CTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCC CTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGAT TAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC CTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGA ACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCG ATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAA TTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGK GCSCGGAACCCCTATTTGTTTATTTTTCAAA
pET28a(+)-6xHis-Thrombin-SNAP25	T7	GGTAATTCCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCAT GGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGC AGCCATATGGCTAGCATGGCCGAAGACGCGGATATGCGCAATGAGCTGGAG GAGATGCAGAGGAGGGCTGACCAGCTGGCTGATGAGTCCCTGGAAAGCACC CGTCGCATGCTGCAGCTGGTTGAAGAGAGTAAAGATGCTGGCATCAGGACTT TGGTTATGTTGGATGAGCAAGGCGAACAACTCGATCGTGTCGAAGAAGGCAT GAACCATATCAACCAAGACATGAAGGAGGCCGAGAAAAATTTAAAAGATTTA GGCAAATCCTCTGGCCTTTTCATATCTCCTTCTAACAAGCTTAAATCCAGTGAT GCTTACAAAAAAGCCTGGGGCAATAATCAGGATGGAGTAGTGGCCAGCCAG CCCGCCCGTGTGGTGGATGAACGGGAGCAGATGGCCATCAGTGGTGGCTTC ATCCGCAGGGTAACAAATGATGCCCGGGAAAATGAGATGGATGAGAACCTG GAGCAGGTGAGCGGCATCATCGGAAACCTCCGCCATATGGCTCTAGACATGG GCAATGAGATTGACACCCAGAATCGCCAGATCGACAGGATCATGGAGAAGG CTGATTCCAACAAAACCAGAATTGATGAAGCCAACCAACGTGCAACAAAGAT GCTGGGCAGTGGGTAATAATAGTGACTCGAGCACCACCACCACCACCACTGA GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCG CTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGG TTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCT GTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCG CTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCT CGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAG GGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTG

		ATGGTTCMCGTAGTGGGCCATCGCCCTGATAGACGGTTTTCGCCCTTTGACG TTGGAGTCACGTTCTT
pET28a(+)-6xHis-Thrombin-Syb-2(1-96)	T7	CCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCA GCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GTCGGCTACCGCTGCCACCGTCCCGCCTGCCGCCCCGGCCGGCGAGGGTGGC CCCCCTGCACCTCCTCCAAATCTTACCAGTAACAGGAGACTGCAGCAGACCCA GGCCCAGGTGGATGAGGTGGTGGACATCATGAGGGTGAATGTGGACAAGG TCCTGGAGCGAGACCAGAAGCTATCGGAACTGGATGATCGCGCAGATGCCCT CCAGGCAGGGGCCTCCCAGTTTGAAACAAGTGCAGCCAAGCTCAAGCGCAA ATACTGGTGGAAAAACCTCAAGATGATGTAACTCGAGCACCACCACCACCAC CACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTG CCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTG AGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACG CGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTT CCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTC CCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGA TTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGC CCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGG AACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCC GATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGA ATTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATG TGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGC
pET28a(+)-6xHis-Thrombin-TEV-Ubl2	T7	ACATTTCCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGG GCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAG CGAAAACCTGTATTTCCAGAGCCTGCGTGAAGTGCGTACCATTAAGGTGTTTA CCACCGTGGATAATATTAATCTGCATACCCAGGTTGTGGATATGAGCATGACC TATGGTCAGCAGTTTGGTCCGACCTATCTGGATGGCGCCGATGTTACCAAAAT TAAGCCGCATAATAGCCATGAAGGTAAAACCTTTTATGTTCTGCCGAATTAAC TCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAA GGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTT GGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATAT CCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGG TGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCC GCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTC AAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCAC CTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC CCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGT GGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTT GATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGAT TTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAGG TGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAAT ACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGC ATCAAATGAAACT

pET28a(+)-6xHis-Thrombin-TEV-Ubl2	T7 term	TCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTC GAGTTAATTCGGCAGAACATAAAAGGTTTTACCTTCATGGCTATTATGCGGCT TAATTTTGGTAACATCGGCGCCATCCAGATAGGTCGGACCAAACTGCTGACC ATAGGTCATGCTCATATCCACAACCTGGGTATGCAGATTAATATTATCCACGG TGGTAAACACCTTAATGGTACGCACTTCACGCAGGCTCTGGAAATACAGGTTT TCGCTGCCGCGCGGCACCAGGCCGCTGCTGTGATGATGATGATGATGGCTGC TGCCCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGA ATTGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATC GAGATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCA CAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCG GGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCA GGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCC TTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATG CAGGAGTCGCATAAGGGAGAGCGTCGAGATCCCGGACACCATCGAATGGCG CAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGG GTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTG TCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCG AAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCC AACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTG CCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAA ATC
pET28a(+)-6xHis-Thrombin-TEV-PLpro	T7	CCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAG CCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCGAAAAC CTGTATTTCCAGAGTCTGCGTGAAGTGCGTACCATTAAGGTGTTTACCACCGT TGATAATATTAATCTGCATACCCAGGTTGTGGATATGAGCATGACCTATGGCC AGCAGTTTGGCCCGACCTATCTGGATGGTGCAGATGTGACCAAAATTAAGCC GCATAATAGCCATGAAGGCAAAACCTTTTATGTTCTGCCGAATGATGATACCC TGCGCGTTGAAGCATTTGAATATTATCATACCACCGATCCGAGCTTTCTGGGC CGCTATATGAGTGCCCTGAATCATACCAAAAAATGGAAATATCCGCAGGTTA ATGGCCTGACCAGTATTAAGTGGGCCGATAATAATTGTTATCTGGCCACCGCC CTGCTGACCCTGCAGCAGATTGAACTGAAATTCAATCCGCCGGCCCTGCAGG ATGCATATTATCGCGCCCGCGCAGGTGAAGCAGCCAATTTTTGCGCCCTGATT CTGGCCTATTGTAATAAGACCGTGGGTGAACTGGGTGACGTGCGTGAAACCA TGAGTTATCTGTTTCAGCATGCAAATCTGGATAGTTGTAAACGCGTTCTGAAT GTTGTTTGCAAAACCTGTGGCCAGCAGCAGACCACCCTGAAAGGTGTTGAAG CAGTGATGTATATGGGTACACTGAGTTATGAACAGTTTAAAAAAGGTGTGCA GATTCCGTGTACCTGTGGCAAACAGGCAACCAAATATCTGGTGCAGCAGGAA AGCCCGTTTGTGATGATGAGCGCACCGCCGGCACAGTATGAACTGAAACATG GCACCTTTACCTGCGCCAGCGAATATACCGGCAATTATCAGTGCGGCCATTAT AAACATATTACCAGCAAAGAAACCCTGTATTGCATTGATGGCGCCCTGCTGAC AAAAAGTAGCGAATATAAAGGCCCGATTACCGATGTTTTCTATAAAGAAAAT AGCTACACCACCACCATTAAGGTA

pET28a(+)-6xHis-Thrombin-TEV-PLpro	T7 term	TCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTC GAGTTACTTAATGGTGGTGGTGTAGCTATTTTCTTTATAGAAAACATCGGTAA TCGGGCCTTTATATTCGCTACTTTTTGTCAGCAGGGCGCCATCAATGCAATAC AGGGTTTCTTTGCTGGTAATATGTTTATAATGGCCGCACTGATAATTGCCGGT ATATTCGCTGGCGCAGGTAAAGGTGCCATGTTTCAGTTCATACTGTGCCGGC GGTGCGCTCATCATCACAAACGGGCTTTCCTGCTGCACCAGATATTTGGTTGC CTGTTTGCCACAGGTACACGGAATCTGCACACCTTTTTTAAACTGTTCATAACT CAGTGTACCCATATACATCACTGCTTCAACACCTTTCAGGGTGGTCTGCTGCT GGCCACAGGTTTTGCAAACAACATTCAGAACGCGTTTACAACTATCCAGATTT GCATGCTGAAACAGATAACTCATGGTTTCACGCACGTCACCCAGTTCACCCAC GGTCTTATTACAATAGGCCAGAATCAGGGCGCAAAAATTGGCTGCTTCACCT GCGCGGGCGCGATAATATGCATCCTGCAGGGCCGGCGGATTGAATTTCAGTT CAATCTGCTGCAGGGTCAGCAGGGCGGTGGCCAGATAACAATTATTATCGGC CCACTTAATACTGGTCAGGCCATTAACCTGCGGATATTTCCATTTTTTGGTATG ATTCAGGGCACTCATATAGCGGCCCAGAAAGCTCGGATCGGTGGTATGATAA TATTCAAATGCTTCAACGCGCAGGGTATCATCATTCGGCAGAACATAAAAGG TTTTGCCTTCATGGCTATTATGCGGCTTAATTTTGGTCACATCTGCACCATCCA GATAGGTCGGGCCAAACTGCTGGCCATAGGTCATGCTCATATCCACAACCTG GGTATGCAGATTAATATTATCAACGGTGGTAAACACCTTAATGGTACGCACTT CACGCAGACTCTGGAAATACAGGTTTTCGCTGCCGCGCGGCACCAGGCCG

Table SD2: Experimental parameters of three-dimensional ^{15}N-^{13}C Ubl2 NMR backbone assignment experiments.

	HNCO	HNCA	HN(CO)CA	HNCACB	HN(CO)CACB
Field [T]	16.45	16.45	16.45	16.45	16.45
^{1}H Larmor frequency [MHz]	700	700	700	700	700
Bruker pulse program	hncogp3d	hncagp3d	hncocagp3d	hncacbgp3d	hncocacbgp3d
Interscan recovery delay [s]	1	1	1	1	1
Number of scans	16	16	16	16	16
Total experimental time [h]	48.7	59.2	60.2	77.2	78.3
Direct dimension (^{1}H)					
^{1}H carrier [ppm]	4.699	4.699	4.699	4.699	4.699
Spectral width [ppm]	14.00	14.00	14.00	14.00	14.00
Data points (TD)	2048	2048	2048	2048	2048
Acquisition time [ms]	104.45	104.45	104.45	104.45	104.45
Indirect dimension (^{15}N)					

^{15}N carrier [ppm]	117.00	117.00	117.00	117.00	117.00
Spectral width [ppm]	35.05	35.05	35.05	35.05	35.05
Data points (TD)	100	100	100	100	100
Acquisition time [ms]	20.1	20.1	20.1	20.1	20.1
Indirect dimension (^{13}C)					
^{13}C carrier [ppm]	172.00	53.20	53.20	42.00	42.00
Spectral width [ppm]	14.02	35.05	35.05	60.41	60.41
Data points (TD)	90	110	110	140	140
Acquisition time [ms]	18.23	8.91	8.91	6.58	6.58
Assignment statistics					
# visible resonances	58	58	58	58	58
# assigned resonances	55	55	55	55	55

Table SD3: Experimental parameters of two-dimensional ^{1}H-^{15}N Ubl2 HSQC experiment.

	HSQC
Field [T]	16.45
^{1}H Larmor frequency [MHz]	700
Bruker pulse program	hsqcfpf3gpphwg
Interscan recovery delay [s]	1
Number of scans	4
Total experimental time [h]	0.2
Direct dimension (^{1}H)	
^{1}H carrier [ppm]	4.699
Spectral width [ppm]	15.94
Data points	2048
Acquisition time [ms]	91.75
Indirect dimension (^{15}N)	
^{15}N carrier [ppm]	117.00
Spectral width [ppm]	35.00
Data points	128
Acquisition time [ms]	25.77

Table SD4: NMR backbone assignment of ^{15}N-^{13}C Ubl2.

Residue	^{1}H [ppm]	^{15}N [ppm]	$^{13}C\alpha_i$ [ppm]	$^{13}C\beta_i$ [ppm]	$^{13}C\alpha_{i-1}$	$^{13}C\beta_{i-1}$
1S	n.	n.	n.	n.	n.	n.
2L	n.	n.	n.	n.	n.	n.
3R	8.3	121.9	53.4	28	52.9	39.7
4E	8.3	122.5	53.6	27.7	53.4	28
V5	8.2	124	59.6	30.1	53.6	27.6
R6	8.7	126.9	53	28.9	59.5	30.1
T7	7.9	112.3	56.5	70.3	53.1	28.9
I8	8	112.4	57.5	40.9	56.5	70.3
K9	8.5	121.6	53	31.1	57.5	40.9
V10	8.7	116.3	56	32.5	53.1	31.1
F11	8.9	116.4	53.6	39.3	56	32.4
T12	9.4	111.2	56.4	68.9	53.6	39.3
T13	8.5	114.1	57.3	68	56.4	68.9
V14	8.5	117.6	57.5	30	57.3	68
D15	8.1	119.5	50.8	40.3	57.6	30.1
N16	8.5	115.1	53.1	35.3	50.7	40.3
I17	8.5	119.1	60.3	38.4	53.1	35.3
N18	9.7	120.2	50.7	37.4	60.3	38.4
L19	9.1	129.4	51.5	41.1	50.7	37.4
H20	9.3	121	51.5	28.2	51.6	41.1
T21	8.9	122	61.1	66.4	51.5	28.2
Q22	9	125.8	51.6	29.8	61.1	66.4
V23	8.4	123.2	59.5	29.7	51.6	29.8
V24	9	124.1	57.6	31.6	59.5	29.7
D25	8.6	124.6	50.3	39.6	57.6	31.6
M26	n.	n.	n.	n.	n.	n.
S27	9.3	113.7	56.7	61.9	55.2	29.7
M28	7.1	121.6	51.2	34.6	56.7	61.9
T29	8.3	108	58.6	68.3	51.2	34.6
Y30	8.3	119.3	62.3	33.7	58.6	68.3
G31	9.3	108	44.6	44.6	62.3	33.7
Q32	7.7	118.9	55.4	27.2	44.7	44.7
Q33	7.2	115.1	54.8	27.9	55.4	27.2
F34	8.8	116.7	52.7	38.3	54.8	28
G35	7.8	109.8	41.3	41.3	52.7	38.3
P36	n.	n.	n.	n.	n.	n.
T37	7.8	121.9	60.4	70.3	61.6	28.5
Y38	9.4	124.1	53.5	39.9	60.3	70.2
L39	9.1	122.9	50.5	41.4	53.5	39.9
D40	9.7	130.8	52.8	36.8	50.5	41.4

G41	8.9	104.2	42.4	42.4	52.8	36.9
A42	8	124.7	48	17.2	42.4	42.4
D43	8.6	121.7	52	37.8	48	17.2
V44	8.8	123.1	57.7	28.1	52	37.9
T45	8.6	119.8	65.1	66.7	57.7	28.1
K46	8.4	115.1	53.2	29.3	65.5	66.7
I47	7.7	118.2	56.6	37.2	53.3	29.3
K48	8.1	120.9	49.8	28.9	56.5	37.1
P49	n.	n.	n.	n.	n.	n.
H50	7.2	122.2	51.6	30.1	60.6	30
N51	8.5	124	54.9	35.2	51.6	30.2
S52	9.4	116	57.9	60.3	54.9	35.2
H53	8.2	121.3	52.1	29.6	57.9	60.3
E54	7.4	119.1	55.9	26.6	52.1	29.6
G55	9.5	114	42.9	42.9	55.9	26.6
K56	8.2	120.6	53.6	29.8	42.9	42.9
T57	8	115.1	59.8	67.1	53.6	29.8
F58	9.5	128.2	54.5	38	59.8	67.1
Y59	9.3	125.3	53.7	37.8	54.5	38
V60	8	122.4	56.4	31.8	53.7	37.4
L61	8.1	120.8	51.4	38.9	56.5	31.8
P62	n.	n.	n.	n.	n.	n.
N63	7.8	124.0	50.2	39.2	59.0	30.3

X. References

1. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
2. Li, X. *et al.* Transmission dynamics and evolutionary history of 2019-nCoV. *J. Med. Virol.* **92**, 501–511 (2020).
3. Rota, P. A. *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394–1399 (2003).
4. Drosten, C. *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003).
5. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
6. Wu, D., Wu, T., Liu, Q. & Yang, Z. The SARS-CoV-2 outbreak: What we know. *Int. J. Infect. Dis.* **94**, 44–48 (2020).
7. Yang, D. & Leibowitz, J. L. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* **206**, 120–133 (2015).
8. Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
9. Hu, B., Guo, H., Zhou, P. & Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
10. Woo, P. C. Y. *et al.* Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavi. *J. Virol.* **86**, 3995–4008 (2012).
11. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
12. Jin, Y.-H. *et al.* A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). *Mil. Med. Res.* **7**, 4 (2020).
13. Weiss, S. R. & Leibowitz, J. L. Coronavirus pathogenesis. *Adv. Virus Res.* **81**, 85–164 (2011).
14. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, England)* **395**, 497–506 (2020).
15. Grifoni, A. *et al.* A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **27**, 671-680.e2 (2020).
16. Abdolmaleki, G. *et al.* A comparison between SARS-CoV-1 and SARS-CoV2: an update on current COVID-19 vaccines. *Daru* **30**, 379–406 (2022).
17. Liu, J. *et al.* A comparative overview of COVID-19, MERS and SARS: Review article. *Int. J. Surg.* **81**, 1–8 (2020).
18. Jahirul Islam, M., Nawal Islam, N., Siddik Alom, M., Kabir, M. & Halim, M. A. A review on structural, non-structural, and accessory proteins of SARS-CoV-2: Highlighting drug target sites. *Immunobiology* **228**, 152302 (2023).
19. Kyriakidis, N. C., López-Cortés, A., González, E. V., Grimaldos, A. B. & Prado, E. O. SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *npj Vaccines* **6**, (2021).
20. Van Egeren, D. *et al.* Risk of rapid evolutionary escape from biomedical interventions

targeting SARS-CoV-2 spike protein. *PLoS One* **16**, e0250780 (2021).

21. Farhud, D. D. & Mojahed, N. SARS-COV-2 Notable Mutations and Variants: A Review Article. *Iran. J. Public Health* **51**, 1494–1501 (2022).

22. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

23. Osipiuk, J. *et al.* Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.* **12**, 743 (2021).

24. Jin, Z. *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).

25. Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**, 853–879 (2000).

26. Snijder, E. J., Decroly, E. & Ziebuhr, J. The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv. Virus Res.* **96**, 59–126 (2016).

27. Brierley, I., Digard, P. & Inglis, S. C. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**, 537–547 (1989).

28. Neuman, B. W. Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles. *Antiviral Res.* **135**, 97–107 (2016).

29. Naqvi, A. A. T. *et al.* Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim. Biophys. acta. Mol. basis Dis.* **1866**, 165878 (2020).

30. Zandi, M. *et al.* The role of SARS-CoV-2 accessory proteins in immune evasion. *Biomed. Pharmacother.* **156**, 113889 (2022).

31. Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res.* **149**, 58–74 (2018).

32. Neuman, B. W. *et al.* Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J. Virol.* **80**, 7918–7928 (2006).

33. Fehr, A. R. & Perman, S. Coronaviruses: An Overview of Their Replication and Pathogenesis. in *Coronaviruses: Methods and Protocols* vol. 1282 1–23 (2015).

34. Collins, A. R., Knobler, R. L., Powell, H. & Buchmeier, M. J. Monoclonal antibodies to murine hepatitis virus-4 (strain JHM) define the viral glycoprotein responsible for attachment and cell--cell fusion. *Virology* **119**, 358–371 (1982).

35. Armstrong, J., Niemann, H., Smeekens, S., Rottier, P. & Warren, G. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature* **308**, 751–752 (1984).

36. Neuman, B. W. *et al.* A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* **174**, 11–22 (2011).

37. Zhou, S. *et al.* SARS-CoV-2 E protein: Pathogenesis and potential therapeutic development. *Biomed. Pharmacother.* **159**, 114242 (2023).

38. Tobler, K., Ackermann, M. & Fraefel, C. *Allgemeine Virologie*. (2016).

39. Brian, D. A. & Baric, R. S. Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* **287**, 1–30 (2005).

40. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).

41. Zhao, M.-M. *et al.* Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Signal Transduct. Target. Ther.* **6**, 134 (2021).

42. Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* **23**, 3–20 (2022).

43. Bhatt, P. R. *et al.* Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science* **372**, 1306–1313 (2021).

44. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

45. Zimmermann, A. L., Zhao, X., Makroczyova, J. & Wachsmuth-melm, M. SARS-CoV-2 nsp3 and nsp4 are minimal constituents of a pore spanning replication organelle. *Nat. Commun.* **14**, 1–12 (2023).

46. Huang, Y. *et al.* Molecular architecture of coronavirus double-membrane vesicle pore complex. *Nature* **633**, 224–231 (2024).

47. Klein, S. *et al.* SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat. Commun.* **11**, 5885 (2020).

48. Knoops, K. *et al.* SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol.* **6**, e226 (2008).

49. Yan, L. *et al.* Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell* **184**, 184-193.e10 (2021).

50. Chen, J. *et al.* Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **182**, 1560-1573.e13 (2020).

51. Hillen, H. S. *et al.* Structure of replicating SARS-CoV-2 polymerase. *Nature* **584**, 154–156 (2020).

52. Malone, B., Urakova, N., Snijder, E. J. & Campbell, E. A. Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat. Rev. Mol. Cell Biol.* **23**, 21–39 (2022).

53. Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **10**, 1–9 (2019).

54. Klein, S. *et al.* SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat. Commun.* **11**, 1–10 (2020).

55. Snijder, E. J. *et al.* A unifying structural and functional model of the coronavirus replication organelle: Tracking down RNA synthesis. *PLoS Biol.* **18**, e3000715 (2020).

56. Wolff, G. *et al.* A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* **369**, 1395–1398 (2020).

57. Siu, Y. L. *et al.* The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *J. Virol.* **82**, 11318–11330 (2008).

58. Hurst, K. R., Koetzner, C. A. & Masters, P. S. Characterization of a Critical Interaction between the Coronavirus Nucleocapsid Protein and Nonstructural Protein 3 of the Viral Replicase-Transcriptase Complex. *J. Virol.* **87**, 9159–9172 (2013).

59. Iserman, C. *et al.* Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol. Cell* **80**, 1078-1091.e6 (2020).

60. Huang, Q. *et al.* Structure of the N-terminal RNA-binding domain of the SARS CoV

nucleocapsid protein. *Biochemistry* **43**, 6059–6063 (2004).

61. Ni, X., Han, Y., Zhou, R., Zhou, Y. & Lei, J. Structural insights into ribonucleoprotein dissociation by nucleocapsid protein interacting with non-structural protein 3 in SARS-CoV-2. *Commun. Biol.* **6**, (2023).

62. Perlman, S. & Masters, P. S. *Coronaviridae: The viruses and their replication*. vols 1, 7th edi (2020).

63. Cortese, M. *et al.* Integrative Imaging Reveals SARS-CoV-2-Induced Reshaping of Subcellular Morphologies. *Cell Host Microbe* **28**, 853-866.e5 (2020).

64. Ricciardi, S. *et al.* The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature* **606**, 761–768 (2022).

65. Angelini, M. M., Akhlaghpour, M., Neuman, B. W. & Buchmeier, M. J. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *MBio* **4**, (2013).

66. Oudshoorn, D. *et al.* Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with Coronaviral RNA Replication. *MBio* **8**, (2017).

67. Bessa, L. M. *et al.* The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a. *Sci. Adv.* **8**, 1–12 (2022).

68. Vardjan, N., Jorgacevski, J. & Zorec, R. Fusion pores, SNAREs, and exocytosis. *Neuroscientist* **19**, 160–174 (2012).

69. Rizo, J. & Südhof, T. C. Snares and munc18 in synaptic vesicle fusion. *Nat. Rev. Neurosci.* **3**, 641–653 (2002).

70. Ghosh, S. *et al.* β-Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell* **183**, 1520-1535.e14 (2020).

71. Chen, D. *et al.* ORF3a of SARS-CoV-2 promotes lysosomal exocytosis-mediated viral egress. *Dev. Cell* **56**, 3250-3263.e5 (2021).

72. Walia, K. *et al.* SARS-CoV-2 virulence factor ORF3a blocks lysosome function by modulating TBC1D5-dependent Rab7 GTPase cycle. *Nat. Commun.* **15**, (2024).

73. Pu, J., Guardia, C. M., Keren-Kaplan, T. & Bonifacino, J. S. Mechanisms and functions of lysosome positioning. *J. Cell Sci.* **129**, 4329–4339 (2016).

74. Rodríguez, A., Webster, P., Ortego, J. & Andrews, N. W. Lysosomes Behave as Ca2+-regulated Exocytic Vesicles in Fibroblasts and Epithelial Cells. *J Cell Biol* **137**, 93–104 (1997).

75. Saftig, P. & Klumperman, J. Lysosome biogenesis and lysosomal membrane proteins: Trafficking meets function. *Nat. Rev. Mol. Cell Biol.* **10**, 623–635 (2009).

76. Qu, Y. *et al.* ORF3a-Mediated Incomplete Autophagy Facilitates Severe Acute Respiratory Syndrome Coronavirus-2 Replication. *Front. cell Dev. Biol.* **9**, 716208 (2021).

77. Miao, G. *et al.* ORF3a of the COVID-19 virus SARS-CoV-2 blocks HOPS complex-mediated assembly of the SNARE complex required for autolysosome formation. *Dev. Cell* **56**, 427-442.e5 (2021).

78. Michelucci, A. *et al.* SARS-CoV-2 ORF3a accessory protein is a water-permeable channel that induces lysosome swelling. *Commun. Biol.* **8**, (2025).

79. Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res.* **149**, 58–74 (2018).

80. Salvi, N. *et al.* 1H, 13C and 15N backbone chemical shift assignments of SARS-CoV-2 nsp3a.

Biomol. NMR Assign. **15**, 173–176 (2021).

81. Sharp, P. M. & Li, W. H. Molecular evolution of ubiquitin genes. *Trends Ecol. Evol.* **2**, 328–332 (1987).

82. Lei, J. *et al.* Crystal structure of the papain-like protease of MERS coronavirus reveals unusual, potentially druggable active-site features. *Antiviral Res.* **109**, 72–82 (2014).

83. Yang, X. *et al.* Proteolytic processing, deubiquitinase and interferon antagonist activities of Middle East respiratory syndrome coronavirus papain-like protease. *J. Gen. Virol.* **95**, 614–626 (2014).

84. Ke, Z. *et al.* N terminus of SARS-CoV-2 nonstructural protein 3 interrupts RNA-driven phase separation of N protein by displacing RNA. *J. Biol. Chem.* **300**, 107828 (2024).

85. Rigsby, R. E. & Parker, A. B. Using the PyMOL application to reinforce visual understanding of protein structure. *Biochem. Mol. Biol. Educ. a Bimon. Publ. Int. Union Biochem. Mol. Biol.* **44**, 433–437 (2016).

86. Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **458**, 422–429 (2009).

87. Serrano, P. *et al.* Nuclear Magnetic Resonance Structure of the N-Terminal Domain of Nonstructural Protein 3 from the Severe Acute Respiratory Syndrome Coronavirus. *J. Virol.* **81**, 12049–12060 (2007).

88. Frieman, M., Ratia, K., Johnston, R. E., Mesecar, A. D. & Baric, R. S. Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF-kappaB signaling. *J. Virol.* **83**, 6689–6705 (2009).

89. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

90. Robichon, C., Luo, J., Causey, T. B., Benner, J. S. & Samuelson, J. C. Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. *Appl. Environ. Microbiol.* **77**, 4634–4646 (2011).

91. Tegel, H., Tourle, S., Ottosson, J. & Persson, A. Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta(DE3). *Protein Expr. Purif.* **69**, 159–167 (2010).

92. Laemmli, U. K. 227680a0. *Nature* **227**, 680–685 (1970).

93. Weber, K. & Osborn, M. The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J. Biol. Chem.* **244**, 4406–4412 (1969).

94. Burnette, W. N. Western blotting. *Clin. Chem.* **57**, 132–133 (1981).

95. Porath, J., Carlsson, J., Olsson, I. & Greta Belfrage. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598–599 (1975).

96. Makrides, S. C. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol. Rev.* **60**, 512–538 (1996).

97. Shirano, Y. and Shibata, D. Low temperature cultivation of Escherichia coli carrying a rice lipoxygenase L-2 cDNA produces a soluble and active enzyme at a high level. *FEBS Lett.* **271**, 128-130 (1990).

98. Vasina, J. A. & Baneyx, F. Expression of aggregation-prone recombinant proteins at low temperatures: A comparative study of the Escherichia coli cspA and tac promoter systems. *Protein Expr. Purif.* **9**, 211–218 (1997).

99. C.P., C. Engineering cell physiology to enhance recombinant protein production in Escherichia coli. *Appl. Microbiol. Biotechnol.* **76**, (2007).

100. Tabata, K. *et al.* Convergent use of phosphatidic acid for hepatitis C virus and SARS-CoV-2

replication organelle formation. *Nat. Commun.* **12**, (2021).

101. Botova, M. *et al.* A specific phosphorylation-dependent conformational switch in SARS-CoV-2 nucleocapsid protein inhibits RNA binding. *Sci. Adv.* **10**, 1–15 (2024).

102. Neuman, B. W. Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles. *Antiviral Res.* **135**, 97–107 (2016).

103. Micsonai, A. *et al.* BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **46**, W315–W322 (2018).

104. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science (80-.).* **278**, 1111–1114 (1997).

105. Zweckstetter, M. NMR hawk-eyed view of AlphaFold2 structures. *Protein Sci.* **30**, 23333–2337 (2021).

106. Robertson, A. J., Courtney, J. M., Shen, Y., Ying, J. & Bax, A. Concordance of X-ray and AlphaFold2 Models of SARS-CoV-2 Main Protease with Residual Dipolar Couplings Measured in Solution. *J. Am. Chem. Soc.* **143**, 19306–19310 (2021).

107. Zweckstetter, M. NMR: Prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* **3**, 679–690 (2008).

108. d'Auvergne, E. J. & Gooley, P. R. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR* **40**, 107–119 (2008).

109. Bieri, M., d'Auvergne, E. J. & Gooley, P. R. relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and µs motion of proteins. *J. Biomol. NMR* **50**, 147–155 (2011).

110. Salvi, N., Abyzov, A. & Blackledge, M. Solvent-dependent segmental dynamics in intrinsically disordered proteins. *Sci. Adv.* **5**, eaax2348 (2019).

111. Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation. *J Am Chem Soc.* **117**, 11341–11352 (1995).

112. Wirmer, J., Wolfgang, P. & Schwalbe, H. Motional properties of unfolded ubiquitin: A model for a random coil protein. *J. Biomol. NMR* **35**, 175–186 (2006).

113. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).

114. Almasy, K. M., Davies, J. P. & Plate, L. Comparative host interactomes of the SARS-CoV-2 nonstructural protein 3 and human coronavirus homologs. *Mol. Cell. Proteomics* **20**, 100120 (2021).

115. Hurst, K. R., Ye, R., Goebel, S. J., Jayaraman, P. & Masters, P. S. An Interaction between the Nucleocapsid Protein and a Component of the Replicase-Transcriptase Complex Is Crucial for the Infectivity of Coronavirus Genomic RNA. *J. Virol.* **84**, 10276–10288 (2010).

116. Keane, S. C. & Giedroc, D. P. Solution Structure of Mouse Hepatitis Virus (MHV) nsp3a and Determinants of the Interaction with MHV Nucleocapsid (N) Protein. *J. Virol.* **87**, 3502–3515 (2013).

117. Castillo, G. *et al.* Molecular mechanisms of human coronavirus NL63 infection and replication. *Virus Res.* **327**, 199078 (2023).

118. Cubuk, J. *et al.* The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **12**, 1–17 (2021).

119. Savastano, A., Ibáñez de Opakua, A., Rankovic, M. & Zweckstetter, M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat. Commun.* **11**, (2020).

120. Jack, A. *et al.* SARS-CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. *PLoS Biol.* **19**, (2021).

121. Chen, H. *et al.* Liquid–liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res.* **30**, 1143–1145 (2020).

122. Perdikari, T. M. *et al.* SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J.* **39**, e106478 (2020).

123. Carlson, C. R. *et al.* Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. *Mol. Cell* **80**, 1092-1103.e4 (2020).

124. Zhao, D. *et al.* Understanding the phase separation characteristics of nucleocapsid protein provides a new therapeutic opportunity against SARS-CoV-2. *Protein Cell* **12**, 734–740 (2021).

125. Skinner, S. P. *et al.* CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* **66**, 111–124 (2016).

126. Nielsen, J. T. & Mulder, F. A. A. Potenci: Prediction of temperature, neighbor and ph-corrected chemical shifts for intrinsically disordered proteins. *J. Biomol. NMR* **70**, 141–165 (2018).

127. Zweckstetter, M. & Bax, A. Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *J. Biomol. NMR* **20**, 365–377 (2001).

128. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J. Magn. Reson.* **131**, 373–378 (1998).

129. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

130. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).

131. Laue, T. M. Computer-aided interpretation of analytical sedimentation data for proteins. in *Analytical ultracentrifugation in biochemistry and polymer science* 90–125 (1992).

132. Brautigam, C. A. Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Methods Enzymol.* **562**, 109–133 (2015).

133. Biehl, R. Jscatter, a program for evaluation and analysis of experimental data. *PLoS One* **14**, 1–18 (2019).

134. Stief, T., Vormann, K. & Lakomek, N. A. Sensitivity-enhanced NMR 15N R1 and R1ρ relaxation experiments for the investigation of intrinsically disordered proteins at high magnetic fields. *Methods* **223**, 1–15 (2024).

135. Lakomek, N. A., Ying, J. & Bax, A. Measurement of 15N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR* **53**, 209–221 (2012).

136. Stief, T., Vormann, K. & Lakomek, N. NMR 15 N Relaxation Experiments for the Investigation of Picosecond to Nanoseconds Structural Dynamics of Proteins. *J. Vis. Exp.* **213**, 1–31 (2024).

137. Lipari, G. & Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. *J. Am. Chem. Soc.* **104**, 4559–4570 (1982).

138. Clore, G. M. *et al.* Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *Am. Chem. Soc.* 4989–4991 (1990).

139. Osborne, M. J. & Wright, P. E. Anisotropic rotational diffusion in model-free analysis for a

ternary DHFR complex. *J. Biomol. NMR* **19**, 209–230 (2001).

140. Maciejewski, M. W. *et al.* NMRbox: A Resource for Biomolecular NMR Computation. *Biophys. J.* **112**, 1529–1534 (2017).

141. Snijder, E. J. *et al.* Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **331**, 991–1004 (2003).

142. Carlson, C. R. *et al.* Phosphorylation modulates liquid-liquid phase separation of the SARS-CoV-2 N protein. *bioRxiv* 2020.06.28.176248 (2020).

143. Banerjee, P. R., Milin, A. N., Moosa, M., Onuchic, P. L. & Deniz, A. A. Reentrant phase transition drives dynamic substructure formation in ribonucleoprotein droplets Vacuolated Ribonucleoprotein Droplets RNA controls the reentrant phase transition of ribonucleoproteins (RNP) to assemble and dissolve RNP droplets. During diss. *Angew Chem Int Ed Engl* **56**, 11354–11359 (2017).

144. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

145. Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651–663 (2016).

146. Pak, C. W. *et al.* Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol Cell.* **63**, 72–85 (2016).

147. Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).

148. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **60**, 208–219 (2015).

149. Galvanetto, N. *et al.* Extreme dynamics in a biomolecular condensate. *Nature* **619**, 876–883 (2023).

150. Kenrick, S. & Some, D. The Diffusion Interaction Parameter (k D) as an Indicator of Colloidal and Thermal Stability. *Appl. Note Wyatt Technol. Corp.* 1–6 (2014).

151. Kuehner, D. E. *et al.* Interactions of lysozyme in concentrated electrolyte solutions from dynamic light-scattering measurements. *Biophys. J.* **73**, 3211–3224 (1997).

152. Meechai, N., Jamieson, A. M. & Blackwell, J. Translational Diffusion Coefficients of Bovine Serum Albumin in Aqueous Solution at High Ionic Strength. *J. Colloid Interface Sci.* **218**, 167–175 (1999).

153. Annamaa, A. Introducing Thonny, a Python IDE for learning programming. 183 (2015).

154. Lin, M. & Larive, C. K. Detection of Insulin Aggregates with Pulsed-Field Gradient Nuclear Magnetic Resonance Spectroscopy. *Anal. Biochem.* **229**, 214–220 (1995).

155. Nesmelova, I. V. & Fedotov, V. D. Self-diffusion and self-association of lysozyme molecules in solution. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1383**, 311–316 (1998).

156. García De La Torre, J., Huertas, M. L. & Carrasco, B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**, 719–730 (2000).

157. Korzhnev, D. M. & Kay, L. E. Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc Chem Res.* **41**, 442–451 (2008).

158. Li, C. & Pielak, G. J. Protein Binding under Crowded Conditions. *J. Am. Chem. Soc.* **131**, 1368–1369 (2010).

159. Maity, H., Muttathukattil, A. N. & Reddy, G. Salt Effects on Protein Folding Thermodynamics. *J. Phys. Chem. Lett.* **9**, (2018).

160. Ganji, M., Docter, M., Le Grice, S. F. J. & Abbondanzieri, E. A. DNA binding proteins explore multiple local configurations during docking via rapid rebinding. *Nucleic Acids Res.* **44**, 8376–8384 (2016).

161. Ostrowska, N., Feig, M. & Trylska, J. Crowding affects structural dynamics and contributes to membrane association of the NS3/4A complex. *Biophys. J.* **120**, 3795–3806 (2021).

162. Fulton, A. B. How crowded is the cytoplasm? *Cell* **30**, 345–347 (1982).

163. Spitzer, J. From water and ions to crowded biomacromolecules: In vivo structuring of a prokaryotic cell. *Microbiol. Mol. Biol. Rev.* **75**, 491–506 (2011).

164. Briggs, J.A.G., Simon, M.N., Gross, I., Krausslich, H.-G., Fuller, S. D. & Vogt, V.M. and Johnson, M. C. The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.* **11**, 672–675 (2004).

165. Burns, J.A., Butler, J.C., Moran, J., Whitesides, G. M. Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *J. Org. Ch.* **56**, 2648–2650 (1991).

166. Fung, T. S. & Liu, D. X. Post-translational modifications of coronavirus proteins: Roles and function. *Future Virol.* **13**, 405–430 (2018).

167. Fung, T. S. & Liu, D. X. PTM of CoVs 2018.pdf. **13**, 405–430 (2018).

168. Wu, C. H. *et al.* Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus mucleocapsid protein and viral replication. *J. Biol. Chem.* **284**, 5229–5239 (2009).

169. Fang, S., Xu, L., Huang, M., Qisheng Li, F. & Liu, D. X. Identification of two ATR-dependent phosphorylation sites on coronavirus nucleocapsid protein with nonessential functions in viral replication and infectivity in cultured cells. *Virology* **444**, 225–232 (2013).

170. Ritchie, G. *et al.* Identification of N-linked carbohydrates from severe acute respiratory syndrome (SARS) spike glycoprotein. *Virology* **399**, 257–269 (2010).

171. Gadlage, M. J. *et al.* Murine Hepatitis Virus Nonstructural Protein 4 Regulates Virus-Induced Membrane Modifications and Replication Complex Function. *J. Virol.* **84**, 280–290 (2010).

172. Beachboard, D. C., Anderson-Daniels, J. M. & Denison, M. R. Mutations across Murine Hepatitis Virus nsp4 Alter Virus Fitness and Membrane Modifications. *J. Virol.* **89**, 2080–2089 (2015).

173. Stukalov, A. *et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV*. *Nature* vol. 594 (Springer US, 2021).

174. Chen, H. *et al.* Mass Spectroscopic Characterization of the Coronavirus Infectious Bronchitis Virus Nucleoprotein and Elucidation of the Role of Phosphorylation in RNA Binding by Using Surface Plasmon Resonance. *J. Virol.* **79**, 1164–1179 (2005).

175. Wang, Z. *et al.* Palmitoylation of SARS-CoV-2 Envelope protein is central to virus particle formation. *J. Virol.* **98**, e0107224 (2024).

176. Xu, G. *et al.* Multiomics approach reveals the ubiquitination-specific processes hijacked by SARS-CoV-2. *Signal Transduct. Target. Ther.* **7**, (2022).

177. Kanjanahaluethai, A., Chen, Z., Jukneliene, D. & Baker, S. C. Membrane topology of murine coronavirus replicase nonstructural protein 3. *Virology* **361**, 391–401 (2007).

178. Zhang, R. *et al.* β-Actin as a loading control for plasma-based Western blot analysis of major depressive disorder patients. *Anal. Biochem.* **427**, 116–120 (2012).

179. Mahmood, T. & Yang, P. C. Western blot: Technique, theory, and trouble shooting. *N. Am. J.*

Med. Sci. **4**, 429–434 (2012).

180. Arndt C, Koristka S, Bartsch H, B. M. Native polyacrylamide gels. *Methods Mol Biol.* **869**, 49–53 (2012).

181. Shi, Y. *et al.* Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. *Protein Sci.* **21**, 1197–1209 (2012).

182. Kenji Sugase, Mindy A. Landes, Peter E. Wright, M. M.-Y. Overexpression of post-translationally modified peptides in Escherichia coli by co-expression with modifying enzymes. *Protein Expr Purif.* **57**, 108–115 (2008).

183. Sahdev, S., Khattar, S. K. & Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. *Mol. Cell. Biochem.* **307**, 249–264 (2008).

184. Devendra Shivhare, Ting Yan Aw, M. N. Production of Recombinant Viral Antigens Using the Baculovirus-Insect Cell Expression System. in *Springer Nature* 185–194 (2024).

185. Vincent, C. *et al.* Cells of Escherichia coli Contain wzb and wzc. *J. Bacteriol.* **181**, 3472–3477 (1999).

186. Romero-Brey, I. & Bartenschlager, R. Membranous replication factories induced by plus-strand RNA viruses. *Viruses* **6**, 2826–2857 (2014).

187. Chen, A. *et al.* A coronaviral pore-replicase complex links RNA synthesis and export from double-membrane vesicles. *Sci. Adv.* **10**, 1–12 (2024).

188. Lemak, S., Skarina, T., Patel, D. T., Stogios, P. J. & Savchenko, A. specific interactions with the 5 ’ UTR of the viral genome. 1–17.

189. Lan, T. C. T. *et al.* Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat. Commun.* **13**, 1128 (2022).

190. Seidel, S. A. I. *et al.* Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* **59**, 301–315 (2013).

191. Borggräfe, J. & Etzkorn, M. Solution NMR Spectroscopy as a Tool to Study DNAzyme Structure and Function. *Methods Mol. Biol.* **2439**, 131–151 (2022).

192. Draper, D. E. A guide to ions and RNA structure. *RNA* **10**, 335–343 (2004).

193. Kankia, B. I. Binding of Mg2+ to single-stranded polynucleotides: hydration and optical studies. *Biophys. Chem.* **104**, 643–654 (2003).

194. Masters, P. S. Coronavirus genomic RNA packaging. *Virology* **537**, 198–207 (2019).

195. Redzic, J. S. *et al.* The Inherent Dynamics and Interaction Sites of the SARS-CoV-2 Nucleocapsid N-Terminal Region. *J. Mol. Biol.* **433**, (2021).

196. Dinesh, D. C. *et al.* Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.* **16**, e1009100 (2020).

197. Yao, H. *et al.* Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **183**, 730-738.e13 (2020).

198. Reis, M. Chapter 4 - Thermodynamics. in *Fundamentals of magnetism* 45–52 (2013).

199. Alberts B, Johnson A, Lewis J, et al. The Structure and Function of DNA. in *Molecular biology of the cell* (2002).

200. Schnieders, R., Keyhani, S., Schwalbe, H. & Fürtig, B. More than Proton Detection—New Avenues for NMR Spectroscopy of RNA. *Chem. - A Eur. J.* **26**, 102–113 (2020).

201. Scott, L. G. & Hennig, M. RNA Structure Determination by NMR. in *Bioinformatics: Data, Sequence Analysis and Evolution* (ed. Keith, J. M.) 29–61 (Humana Press, 2008).

doi:10.1007/978-1-60327-159-2_2.

202. Sokolinskaya, E. L., Putlyaeva, L. V, Polinovskaya, V. S. & Lukyanov, K. A. Genetically Encoded Fluorescent Sensors for SARS-CoV-2 Papain-like Protease PLpro. *Int. J. Mol. Sci.* **23**, (2022).

203. Mielech, A. M., Chen, Y., Mesecar, A. D. & Baker, S. C. Nidovirus papain-like proteases: multifunctional enzymes with protease, deubiquitinating and deISGylating activities. *Virus Res.* **194**, 184–190 (2014).

204. Mahmoudvand, S. & Shokri, S. Interactions between SARS coronavirus 2 papain-like protease and immune system: A potential drug target for the treatment of COVID-19. *Scand. J. Immunol.* **94**, 1–12 (2021).

205. Bosken, Y. K., Cholko, T., Lou, Y.-C., Wu, K.-P. & Chang, C. A. Insights Into Dynamics of Inhibitor and Ubiquitin-Like Protein Binding in SARS-CoV-2 Papain-Like Protease. *Front. Mol. Biosci.* **7**, 1–14 (2020).

206. Shin, D. *et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657–662 (2020).

207. Yang, H. & Rao, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nat. Rev. Microbiol.* **19**, 685–700 (2021).

208. Barretto, N. *et al.* The Papain-Like Protease of Severe Acute Respiratory Syndrome Coronavirus Has Deubiquitinating Activity. *J. Virol.* **79**, 15189–15198 (2005).

209. Ullrich, S. & Nitsche, C. SARS-CoV-2 Papain-Like Protease: Structure, Function and Inhibition. *ChemBioChem* **23**, 1–11 (2022).

210. Wu, A. *et al.* Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* **27**, 325–328 (2020).

211. Rut, W. *et al.* Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Sci. Adv.* **6**, 1–12 (2020).

212. Harcourt, B. H. *et al.* Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-Like Protease Activity. *J. Virol.* **78**, 13600–13612 (2004).

213. Schubert, K. *et al.* SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat. Struct. Mol. Biol.* **27**, 959–966 (2020).

214. Thoms, M. *et al.* Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science (80-.).* **369**, 1249–1256 (2020).

215. Cornillez-Ty, C. T., Liao, L., Yates, J. R., Kuhn, P. & Buchmeier, M. J. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling. *J. Virol.* **83**, 10314–10318 (2009).

216. Angeletti, S. *et al.* COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J. Med. Virol.* **92**, 584–588 (2020).

217. Mohamud, Y. *et al.* The papain-like protease of coronaviruses cleaves ULK1 to disrupt host autophagy. *Biochem. Biophys. Res. Commun.* **540**, 75–82 (2021).

218. Russell, R. C. *et al.* ULK1 induces autophagy by phosphorylating Beclin-1 and activating VPS34 lipid kinase. *Nat. Cell Biol.* **15**, 741–750 (2013).

219. Chan, E. Y. W., Kir, S. & Tooze, S. A. siRNA screening of the kinome identifies ULK1 as a multidomain modulator of autophagy. *J. Biol. Chem.* **282**, 25464–25474 (2007).

220. Snyder, N. A. & Silva, G. M. Deubiquitinating enzymes (DUBs): Regulation, homeostasis, and

oxidative stress response. *J. Biol. Chem.* **297**, 101077 (2021).

221. Lindner, H. A. *et al.* Selectivity in ISG15 and ubiquitin recognition by the SARS coronavirus papain-like protease. *Arch. Biochem. Biophys.* **466**, 8–14 (2007).

222. Wydorski, P. M. *et al.* Dual domain recognition determines SARS-CoV-2 PLpro selectivity for human ISG15 and K48-linked di-ubiquitin. *Nat. Commun.* **14**, 1–18 (2023).

223. Dos Santos, P. F. & Mansur, D. S. Beyond ISGlylation: Functions of Free Intracellular and Extracellular ISG15. *J. Interf. cytokine Res. Off. J. Int. Soc. Interf. Cytokine Res.* **37**, 246–253 (2017).

224. Ratia, K., Kilianski, A., Baez-Santos, Y. M., Baker, S. C. & Mesecar, A. Structural Basis for the Ubiquitin-Linkage Specificity and deISGylating Activity of SARS-CoV Papain-Like Protease. *PLoS Pathog.* **10**, (2014).

225. Békés, M. *et al.* Recognition of Lys48-Linked Di-ubiquitin and Deubiquitinating Activities of the SARS Coronavirus Papain-like Protease. *Mol. Cell* **62**, 572–585 (2016).

226. Li, M. *et al.* Structure of the multiple functional domains from coronavirus nonstructural protein 3. *Emerg. Microbes Infect.* **10**, 66–80 (2021).

227. Kokic, G. *et al.* Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* **12**, 279 (2021).

228. Cao, Z. *et al.* VV116 versus Nirmatrelvir-Ritonavir for Oral Treatment of Covid-19. *N. Engl. J. Med.* **388**, 406–417 (2023).

229. Lamb, Y. N. Correction to: Nirmatrelvir Plus Ritonavir: First Approval. *Drugs* vol. 82 1025 (2022).

230. Wang, H., Xue, S., Yang, H. & Chen, C. Recent progress in the discovery of inhibitors targeting coronavirus proteases. *Virol. Sin.* **31**, 24–30 (2016).

231. Li, X. & Song, Y. Targeting SARS-CoV-2 nonstructural protein 3: Function, structure, inhibition, and perspective in drug discovery. *Drug Discov. Today* **29**, 103832 (2024).

232. Freitas, B. T. *et al.* Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infect. Dis.* **6**, 2099–2109 (2020).

233. Hartenian, E. *et al.* The molecular virology of coronaviruses. *J. Biol. Chem.* **295**, 12910–12934 (2020).

234. Shiraishi, Y. & Shimada, I. NMR Characterization of the Papain-like Protease from SARS-CoV-2 Identifies the Conformational Heterogeneity in Its Inhibitor-Binding Site. *J. Am. Chem. Soc.* **145**, 16669–16677 (2023).

235. Wien, F. *et al.* BeStSel : webserver for secondary structure and fold Andr as ´ Moln ar. **50**, 90–98 (2022).

236. Lee, L. J., Barrett, J. A. & Poole, R. K. Genome-wide transcriptional response of chemostat-cultured Escherichia coli to zinc. *J. Bacteriol.* **187**, 1124–1134 (2005).

237. Woody, R. W. Structural composition of βI - and βII-proteins. *Protein Sci.* **12**, 384–388 (2003).

238. Bondesen, B. A. & Schuh, M. D. Advanced Chemistry Classroom and Laboratory Circular Dichroism of Globular Proteins. *J. Chem. Educ.* **78**, (2001).

239. Brahms, S., Brahms, J., Spach, G. & Brack, A. Identification of β,β-turns and unordered conformations in polypeptide chains by vacuum ultraviolet circular dichroism. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3208–3212 (1977).

240. Zhang, H., Neal, S. & Wishart, D. S. A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **25**, 173–195 (2003).

241. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12366–12371 (1997).

242. Tzakos, A. G., Grace, C. R. R., Lukavsky, P. J. & Riek, R. NMR techniques for very large proteins and rnas in solution. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 319—342 (2006).

243. Takeuchi, K., Arthanari, H., Shimada, I. & Wagner, G. Nitrogen detected TROSY at high field yields high resolution and sensitivity for protein NMR. *J. Biomol. NMR* **63**, 323–331 (2015).

244. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of Ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531–544 (1987).

245. Stockman, B. J., Euvrard, A. & Scahill, T. A. Heteronuclear three-dimensional NMR spectroscopy of a partially denatured protein: The A-state of human ubiquitin. *J. Biomol. NMR* **3**, 285–296 (1993).

246. de Groot, N. S. & Ventura, S. Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Lett.* **580**, 6471–6476 (2006).

247. Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M. & Panda, A. K. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microb Cell Fact.* **14**, (2015).

248. Stewart, E. J., Åslund, F. & Beckwith, J. Disulfide bond formation in the Escherichia coli cytoplasm: An in vivo role reversal for the thioredoxins. *EMBO J.* **17**, 5543–5550 (1998).

249. Palmer, I. & Wingfield, P. T. Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from Escherichia coli. *CP Protein Sci.* (2012).

250. Denoncin, K. & Collet, J. F. Disulfide bond formation in the bacterial periplasm: Major achievements and challenges ahead. *Antioxidants Redox Signal.* **19**, 63–71 (2013).

251. Schütz, A. *et al.* A concise guide to choosing suitable gene expression systems for recombinant protein production. *STAR Protoc.* **4**, 1–16 (2023).

252. Wickner, W. & Schekman, R. Membrane fusion. *Nat. Struct. Mol. Biol.* **15**, 658–664 (2008).

253. Rizo, J. & Südhof, T. C. The membrane fusion enigma: SNAREs, Sec1/Munc18 proteins, and their accomplices--guilty as charged? *Annu. Rev. Cell Dev. Biol.* **28**, 279–308 (2012).

254. Rothman, J. E. Mechanisms of Intracellular Protein Transport. *Biol. Chem. Hoppe. Seyler.* **377**, 407–410 (1996).

255. Jahn, R. & Südhof, T. C. Membrane fusion and exocytosis. *Annu. Rev. Biochem.* **68**, 863–911 (1999).

256. Diao, J. *et al.* Synaptic proteins promote calcium-triggered fast transition from point contact to full fusion. *Elife* **1**, e00109 (2012).

257. Südhof, T. C. The synaptic vesicle cycle: a cascade of protein-protein interactions. *Nature* **375**, 645–653 (1995).

258. Fasshauer, D., Sutton, R. B., Brunger, A. T. & Jahn, R. Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q-and R-SNAREs (membrane fusion neurotransmission clostridial neurotoxins). *Neurobiol. Commun. by Peter B. Moore* **95**, 15781–15786 (1998).

259. Hong, W. SNAREs and traffic. *Biochim. Biophys. Acta* **1744**, 120–144 (2005).

260. Tian, Z. *et al.* Biochemical studies of membrane fusion at the single-particle level. *Prog. Lipid Res.* **73**, 92–100 (2019).

261. Kloepper, T. H., Kienle, C. N. & Fasshauer, D. An elaborate classification of SNARE proteins sheds light on the conservation of the eukaryotic endomembrane system. *Mol. Biol. Cell* **18**, 3463–3471 (2007).

262. Fasshauer, D., Otto, H., Eliason, W. K., Jahn, R. & Brünger, A. T. Structural changes are associated with soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor complex formation. *J. Biol. Chem.* **272**, 28036–28041 (1997).

263. Jahn, R. & Fasshauer, D. Molecular machines governing exocytosis of synaptic vesicles. *Nature* **490**, 201–207 (2012).

264. Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature* **395**, 347–353 (1998).

265. Lai, Y., Lou, X., Wang, C., Xia, T. & Tong, J. Synaptotagmin 1 and Ca2+ drive trans SNARE zippering. *Sci. Rep.* **4**, 4575 (2014).

266. Xu, J., Brewer, K. D., Perez-Castillejos, R. & Rizo, J. Subtle Interplay between synaptotagmin and complexin binding to the SNARE complex. *J. Mol. Biol.* **425**, 3461–3475 (2013).

267. Zhang, Y. & Hughson, F. M. Chaperoning SNARE Folding and Assembly. *Annu. Rev. Biochem.* **90**, 581–603 (2021).

268. Bark, I. C., Hahn, K. M., Ryabinin, A. E. & Wilson, M. C. Differential expression of SNAP-25 protein isoforms during divergent vesicle fusion events of neural development. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1510–1514 (1995).

269. Alten, B. *et al.* Role of Aberrant Spontaneous Neurotransmission in SNAP25-Associated Encephalopathies. *Neuron* **109**, 59-72.e5 (2021).

270. Greaves, J. *et al.* The Hydrophobic Cysteine-rich Domain of SNAP25 Couples with Downstream Residues to Mediate Membrane Interactions and Recognition by DHHC Palmitoyl Transferases. *Mol. Biol. Cell* **20**, 1845–1854 (2009).

271. Stief, T. *et al.* Intrinsic Disorder of the Neuronal SNARE Protein SNAP25a in its Pre-fusion Conformation. *J. Mol. Biol.* **435**, 1–10 (2023).

272. Fasshauer, D. & Margittai, M. A Transient N-terminal Interaction of SNAP-25 and Syntaxin Nucleates SNARE Assembly. *J. Biol. Chem.* **279**, 7613–7621 (2004).

273. Raptis, A., Torrejón-Escribano, B., Gómez De Aranda, I. & Blasi, J. Distribution of synaptobrevin/VAMP 1 and 2 in rat brain. *J. Chem. Neuroanat.* **30**, 201–211 (2005).

274. Kesavan, J., Borisovska, M. & Bruns, D. v-SNARE Actions during Ca2+-Triggered Exocytosis. *Cell* **131**, 351–363 (2007).

275. Lakomek, N. A., Yavuz, H., Jahn, R. & Pérez-Lara, Á. Structural dynamics and transient lipid binding of synaptobrevin-2 tune SNARE assembly and membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8699–8708 (2019).

276. Deák, F., Shin, O.-H., Kavalali, E. T. & Südhof, T. C. Structural determinants of synaptobrevin 2 function in synaptic vesicle fusion. *J. Neurosci. Off. J. Soc. Neurosci.* **26**, 6668–6676 (2006).

277. Fang, Q., Zhao, Y. & Lindau, M. Juxtamembrane tryptophans of synaptobrevin 2 control the process of membrane fusion. *FEBS Lett.* **587**, 67–72 (2013).

278. Pobbati, A. V, Stein, A. & Fasshauer, D. N- to C-terminal SNARE complex assembly promotes rapid membrane fusion. *Science* **313**, 673–676 (2006).

279. Wittig, S. *et al.* Oligomerisation of Synaptobrevin-2 Studied by Native Mass Spectrometry and Chemical Cross-Linking. *J. Am. Soc. Mass Spectrom.* **30**, 149–160 (2019).

280. Stein, A., Weber, G., Wahl, M. C. & Jahn, R. Helical extension of the neuronal SNARE complex

into the membrane. *Nature* **460**, 525–528 (2009).

281. Hernandez, J. M. *et al.* Membrane fusion intermediates via directional and full assembly of the SNARE complex. *Science* **336**, 1581–1584 (2012).

282. Irfan, M., Daraio, T. & Bark, C. SNAP-25 Puts SNAREs at Center Stage in Metabolic Disease. *Neuroscience* **420**, 86–96 (2019).

283. Antonucci, F. *et al.* SNAP-25, a Known Presynaptic Protein with Emerging Postsynaptic Functions. *Front. Synaptic Neurosci.* **8**, 7 (2016).

284. Ellena, J. F. *et al.* Dynamic structure of lipid-bound synaptobrevin suggests a nucleation-propagation mechanism for trans-SNARE complex formation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 20306–20311 (2009).

285. Liang, B., Dawidowski, D., Ellena, J. F., Tamm, L. K. & Cafiso, D. S. The SNARE motif of synaptobrevin exhibits an aqueous-interfacial partitioning that is modulated by membrane curvature. *Biochemistry* **53**, 1485–1494 (2014).

286. Liang, B., Kiessling, V. & Tamm, L. K. Prefusion structure of syntaxin-1A suggests pathway for folding into neuronal trans-SNARE complex fusion intermediate. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19384–19389 (2013).

287. Brewer, K. D., Li, W., Horne, B. E. & Rizo, J. Reluctance to membrane binding enables accessibility of the synaptobrevin SNARE motif for SNARE complex formation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12723–12728 (2011).

288. Margittai, M., Fasshauer, D., Pabst, S., Jahn, R. & Langen, R. Homo- and heterooligomeric SNARE complexes studied by site-directed spin labeling. *J. Biol. Chem.* **276**, 13169–13177 (2001).

289. Shen, X.-M., Selcen, D., Brengman, J. & Engel, A. G. Mutant SNAP25B causes myasthenia, cortical hyperexcitability, ataxia, and intellectual disability. *Neurology* **83**, 2247–2255 (2014).

290. Veit, M., Söllner, T. H. & Rothman, J. E. Multiple palmitoylation of synaptotagmin and the t-SNARE SNAP-25. *FEBS Lett.* **385**, 119–123 (1996).

291. Arai, M., Sugase, K., Dyson, H. J. & Wright, P. E. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9614–9619 (2015).

292. Bah, A. *et al.* Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106–109 (2015).

293. Borgia, A. *et al.* Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).

294. Schneider, R., Blackledge, M. & Jensen, M. R. Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. *Curr. Opin. Struct. Biol.* **54**, 10–18 (2019).

295. Dyson, H. J. & Wright, P. E. NMR illuminates intrinsic disorder. *Curr. Opin. Struct. Biol.* **70**, 44–52 (2021).

296. Nielsen, J. T. & Mulder, F. A. A. CheSPI: chemical shift secondary structure population inference. *J. Biomol. NMR* **75**, 273–291 (2021).

297. Hu, G. *et al.* flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nat. Commun.* **12**, 4438 (2021).

298. Emenecker, R. J., Griffith, D. & Holehouse, A. S. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys. J.* **120**, 4312–4319 (2021).

299. Tang, Y.-J., Pang, Y.-H. & Liu, B. IDP-Seq2Seq: identification of intrinsically disordered regions based on sequence to sequence learning. *Bioinformatics* **36**, 5177–5186 (2021).

300. Erdős, G., Pajkos, M. & Dosztányi, Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* **49**, W297–W303 (2021).

301. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

302. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

303. Ruff, K. M. & Pappu, R. V. AlphaFold and Implications for Intrinsically Disordered Proteins. *J. Mol. Biol.* **433**, 167208 (2021).

304. Wilson, C. J., Choy, W.-Y. & Karttunen, M. AlphaFold2: A Role for Disordered Protein/Region Prediction? *Int. J. Mol. Sci.* **23**, (2022).

305. Kholodenko, A. Analytical calculation of the scattering function for polymers of arbitrary flexibility using the Dirac propagator. *Macromolecules* **26**, (1993).

306. Doi, M. & Edwards, S. The Theory of Polymer Dynamics. *Oxford University Press* (1988).

307. Stadler, A. M., Pellegrini, E., Johnson, M., Fitter, J. & Zaccai, G. Dynamics-stability relationships in apo- and holomyoglobin: a combined neutron scattering and molecular dynamics simulations study. *Biophys. J.* **102**, 351–359 (2012).

308. Stadler, A. M. *et al.* Internal nanosecond dynamics in the intrinsically disordered myelin basic protein. *J. Am. Chem. Soc.* **136**, 6987–6994 (2014).

309. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).

310. Ernst, J. A. & Brunger, A. T. High resolution structure, stability, and synaptotagmin binding of a truncated neuronal SNARE complex. *J. Biol. Chem.* **278**, 8630–8636 (2003).

311. Jakhanwal, S., Lee, C.-T., Urlaub, H. & Jahn, R. An activated Q-SNARE/SM protein complex as a possible intermediate in SNARE assembly. *EMBO J.* **36**, 1788–1802 (2017).

312. Fiebig, K. M., Rice, L. M., Pollock, E. & Brunger, A. T. Folding intermediates of SNARE complex assembly. *Nat. Struct. Biol.* **6**, 117–123 (1999).

313. Li, F., Tiwari, N., Rothman, J. E. & Pincet, F. Kinetic barriers to SNAREpin assembly in the regulation of membrane docking/priming and fusion. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10536–10541 (2016).

314. Petoukhov, M. V, Konarev, P. V, Kikhney, G. & Dmitri, I. ATSAS 2.1 - Supported Small-Angle Scattering Data Analysis. *Appl. Crystallogr.* **40**, s223–s228 (2007).

315. Kay, L. E., Keifer, P. & Saarinen, T. Pure Absorption Gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665 (1992).

316. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533 (2002).

317. Uversky, V. N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739–56 (2002).

318. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025 (2007).

319. Mittag, T. & Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles.

Curr. Opin. Struct. Biol. **17**, 3–14 (2007).

320. Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* **60**, 231–241 (2015).

321. Nott, T. J. *et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).

322. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **28**, 420–435 (2018).

323. Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).

324. Guseva, S. *et al.* Liquid-Liquid Phase Separation Modifies the Dynamic Properties of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **145**, 10548–10563 (2023).

325. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).

326. Bertoncini, C. W. *et al.* Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1430–1435 (2005).

327. Buell, A. K. *et al.* Solution conditions determine the relative importance of nucleation and growth processes in α-synuclein aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7671–7676 (2014).

328. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).

329. Jensen, M. R., Salmon, L., Nodet, G. & Blackledge, M. Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J Am Chem Soc.* **132**, 1270–2 (2010).

330. Camacho-Zarco, A. R. *et al.* NMR Provides Unique Insight into the Functional Dynamics and Interactions of Intrinsically Disordered Proteins. *Chem. Rev.* **122**, 9331–9356 (2022).

331. Rezaei-Ghaleh, N. *et al.* Local and Global Dynamics in Intrinsically Disordered Synuclein. *Angew. Chem. Int. Ed. Engl.* **57**, 15262–15266 (2018).

332. Alderson, T. R. & Kay, L. E. NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. *Cell* **184**, 577–595 (2021).

333. Kay, L. E., Torchia, D. A. & Bax, A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979 (1989).

334. Peng, J. W. & Wagner, G. Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments. *Biochemistry* **31**, 8571–8586 (1992).

335. Farrow, N. A. *et al.* Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).

336. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry* **31**, 5269–5278 (1992).

337. Kay, L. E. Protein dynamics from NMR. *Nat. Struct. Biol.* **5 Suppl**, 513–517 (1998).

338. Ishima, R. & Torchia, D. A. Protein dynamics from NMR. *Nat. Struct. Biol.* **7**, 740–743 (2000).

339. Lienin, S. F., Bremi, T., Brutscher, B., Brüschweiler, R. & Ernst, R. R. Anisotropic intramolecular backbone dynamics of ubiquitin characterized by NMR relaxation and MD computer simulation. *J. Am. Chem. Soc.* **120**, 9870–9879 (1998).

340. Akke, M. & Palmer, A. G. Monitoring macromolecular motions on microsecond to millisecond time scales by R(1)rho-R(1) constant relaxation time NMR spectroscopy. *J. Am. Chem. Soc.* **118**, 911–912 (1996).

341. Loria, J. P., Rance, M. & Palmer, A. G. A Relaxation-Compensated Carr−Purcell−Meiboom−Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *J. Am. Chem. Soc.* **121**, 2331–2332 (1999).

342. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001).

343. Mulder, F. A., Skrynnikov, N. R., Hon, B., Dahlquist, F. W. & Kay, L. E. Measurement of slow (micros-ms) time scale dynamics in protein side chains by (15)N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J. Am. Chem. Soc.* **123**, 967–975 (2001).

344. Hansen, D. F., Vallurupalli, P. & Kay, L. E. An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J. Phys. Chem. B* **112**, 5898–5904 (2008).

345. Palmer, A. G. 3rd. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **104**, 3623–3640 (2004).

346. Mittermaier, A. & Kay, L. E. New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224–228 (2006).

347. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638–1642 (2006).

348. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).

349. Neudecker, P. *et al.* Structure of an intermediate state in protein folding and aggregation. *Science* **336**, 362–366 (2012).

350. Morrison, E. A. *et al.* Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* **481**, 45–50 (2011).

351. Stiller, J. B. *et al.* Structure determination of high-energy states in a dynamic protein ensemble. *Nature* **603**, 528–535 (2022).

352. Zhang, Q., Sun, X., Watt, E. D. & Al-Hashimi, H. M. Resolving the motional modes that code for RNA adaptation. *Science* **311**, 653–656 (2006).

353. Kimsey, I. J., Petzold, K., Sathyamoorthy, B., Stein, Z. W. & Al-Hashimi, H. M. Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature* **519**, 315–320 (2015).

354. Rinnenthal, J. *et al.* Mapping the landscape of RNA dynamics with NMR spectroscopy. *Acc. Chem. Res.* **44**, 1292–1301 (2011).

355. Jonathan Boyd Ulrich Hommel, I. D. C. Influence of cross-correlation between dipolar and anisotropic chemical shift relaxation mechanisms upon longitudinal relaxation rates of 15N in macromolecules. *Elsevier ScienceDirect Journals* vol. 175 477 (1990).

356. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A. & D.A., T. Pulse sequences for removal of the effects of cross-correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T1 and T2 values in proteins. *J. Magn. Reson.* **92**, 359–375 (1992).

357. Palmer, A. G., Skelton, N. J., Chazin, W. J., Wright, P. E. & Rance, M. SUPPRESSION OF THE EFFECTS OF CROSS-CORRELATION BETWEEN DIPOLAR AND ANISOTROPIC CHEMICAL-SHIFT RELAXATION MECHANISMS IN THE MEASUREMENT OF SPIN SPIN RELAXATION RATES. *Mol.*

Phys. **75**, 699–711 (1992).

358. Grzesiek, S. & Bax, A. The importance of not saturating H2O in protein NMR - Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594 (1993).

359. Chen, K. & Tjandra, N. Water proton spin saturation affects measured protein backbone 15N spin relaxation rates. *J. Magn. Reson.* **213**, 151–157 (2011).

360. Ishima, R. A probe to monitor performance of ^{15}N longitudinal relaxation experiments for proteins in solution. *J. Biomol. NMR* **58**, 113–122 (2014).

361. Yuwen, T. & Skrynnikov, N. R. Proton-decoupled CPMG: a better experiment for measuring (15)N R2 relaxation in disordered proteins. *J. Magn. Reson.* **241**, 155–169 (2014).

362. Wikus, P., Frantz, W., Kummerle, R. & Vonlanthen, P. Commercial gigahertz-class NMR magnets. *Supercond Sci Tech* **35**, (2022).

363. Hazzard, J., Südhof, T. C. & Rizo, J. NMR analysis of the structure of synaptobrevin and of its interaction with syntaxin. *J. Biomol. NMR* **14**, 203–207 (1999).

364. Bloembergen, N. & Pound, R. V. Radiation Damping in Magnetic Resonance Experiments. *Phys. Rev.* **95**, 8–12 (1954).

365. Ishima, R. Effects of radiation damping for biomolecular NMR experiments in solution: a hemisphere concept for water suppression. *Concepts Magn. Reson. Part A. Bridg. Educ. Res.* **44A**, 252–262 (2015).

366. Shishmarev, D. & Otting, G. Radiation damping on cryoprobes. *J. Magn. Reson.* **213**, 76–81 (2011).

367. Loria, J. P., Rance, M. & Palmer, A. G. 3rd. Transverse-relaxation-optimized (TROSY) gradient-enhanced triple-resonance NMR spectroscopy. *J. Magn. Reson.* **141**, 180–184 (1999).

368. Sklenar, V. Suppression of Radiation Damping in Multidimensional NMR Experiments Using Magnetic Field Gradients. *J. Magn. Reson. Ser. A* **114**, 132–135 (1995).

369. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).

370. Schanda, P., Van Melckebeke, H. & Brutscher, B. Speeding up three-dimensional protein NMR experiments to a few minutes. *J. Am. Chem. Soc.* **128**, 9042–9043 (2006).

371. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127**, 8014–8015 (2005).

372. Joseph, D. & Griesinger, C. Optimal control pulses for the 1.2-GHz (28.2-T) NMR spectrometers. *Sci. Adv.* **9**, eadj1133 (2023).

373. Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson.* **93**, 151–170 (1991).

374. Salzmann, M., Wider, G., Pervushin, K. & Wüthrich, K. Improved sensitivity and coherence selection for [15N,1H]-TROSY elements in triple resonance experiments. *J. Biomol. NMR* **15**, 181–184 (1999).

375. Wider, G. Technical aspects of NMR Spectroscopy with biological macromolecules and studies of hydration in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **32**, 193–275 (1998).

376. Geen, H. & Freeman, R. Band-selective radiofrequency pulses. *J. Magn. Reson.* **93**, 93–141

(1991).

377. Mulder, F. A. A., RA, de G., Kaptein, R. & Boelens, R. An Off-resonance Rotating Frame Relaxation Experiment for the Investigation of Macromolecular Dynamics Using Adiabatic Rotations. *J. Magn. Reson.* **131**, 351–357 (1998).

378. Garwood, M. & Ke, Y. Symmetrical Pulses to Induce Arbitrary Flip Angles with Compensation for Rf Inhomogeneity and Resonance Offsets. *J. Magn. Reson.* **94**, 511–525 (1991).

379. Shao, X., Sudhof, T. C. & Rizo, J. Assignment of the 1H, 15N and 13C resonances of the calcium-free and calcium-bound forms of the first C2-domain of synaptotagmin I. *Biomol. NMR Assign.* **10**, 307–8 (1997).

380. Desvaux, H. & Berthault, P. Study of dynamic processes in liquids using off-resonance rf irradiation. *Prog. Nucl. Magn. Reson. Spectrosc.* **35**, 295–340 (1999).

381. Palmer, A. G. 3rd & Massi, F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem. Rev.* **106**, 1700–1719 (2006).

382. Lemaster, D. M. Isotope Labeling in Solution Protein Assignment and Structural-Analysis. *Prog Nucl Mag Res Sp* **26**, 371–419 (1994).

383. Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J. D. & Bax, A. Four-Dimensional 15N-Separated NOESY of Slowly Tumbling Perdeuterated 15N-Enriched Proteins. Application to HIV-1 Nef. *J. Am. Chem. Soc.* **117**, 9594–9595 (1995).

384. Eletsky, A., Kienhöfer, A. & Pervushin, K. TROSY NMR with partially deuterated proteins. *J. Biomol. NMR* **20**, 177–180 (2001).

385. Gairí, M. *et al.* An optimized method for (15)N R(1) relaxation rate measurements in non-deuterated proteins. *J. Biomol. NMR* **62**, 209–220 (2015).

386. Nichols, P. J. *et al.* Deuteration of nonexchangeable protons on proteins affects their thermal stability, side-chain dynamics, and hydrophobicity. *Protein Sci.* **29**, 1641–1654 (2020).

387. Mugridge, J. S., Bergman, R. G. & Raymond, K. N. Does size really matter? The steric isotope effect in a supramolecular host-guest exchange reaction. *Angew. Chem. Int. Ed. Engl.* **49**, 3635–3637 (2010).

388. Cioni, P. & Strambini, G. B. Effect of heavy water on protein flexibility. *Biophys. J.* **82**, 3246–3253 (2002).

389. Adamski, W. *et al.* A Unified Description of Intrinsically Disordered Protein Dynamics under Physiological Conditions Using NMR Spectroscopy. *J. Am. Chem. Soc.* **141**, 17817–17829 (2019).

390. Salvi, N., Abyzov, A. & Blackledge, M. Multi-Timescale Dynamics in Intrinsically Disordered Proteins from NMR Relaxation and Molecular Simulation. *J. Phys. Chem. Lett.* **7**, 2483–2489 (2016).

391. Abyzov, A., Mandelkow, E., Zweckstetter, M. & Rezaei-Ghaleh, N. Fast Motions Dominate Dynamics of Intrinsically Disordered Tau Protein at High Temperatures. *Chemistry* **29**, e202203493 (2023).

392. Rezaei-Ghaleh, N., Parigi, G. & Zweckstetter, M. Reorientational Dynamics of Amyloid-β from NMR Spin Relaxation and Molecular Simulation. *J. Phys. Chem. Lett.* **10**, 3369–3375 (2019).

393. Kurzbach, D., Kontaxis, G., Coudevylle, N. & Konrat, R. NMR Spectroscopic Studies of the Conformational Ensembles of Intrinsically Disordered Proteins. *Adv. Exp. Med. Biol.* **870**, 149–185 (2015).

394. Kim, S., Wu, K.-P. & Baum, J. Fast hydrogen exchange affects ^{15}N relaxation measurements in intrinsically disordered proteins. *J. Biomol. NMR* **55**, 249–256 (2013).

395. Gill, M. L., Byrd, R. A. & Palmer, A. G. I. I. I. Dynamics of GCN4 facilitate DNA interaction: a model-free analysis of an intrinsically disordered region. *Phys. Chem. Chem. Phys.* **18**, 5839–5849 (2016).

396. Brady, J. P. *et al.* Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E8194–E8203 (2017).

397. Libich, D. S., Tugarinov, V. & Clore, G. M. Intrinsic unfoldase/foldase activity of the chaperonin GroEL directly demonstrated using multinuclear relaxation-based NMR. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8817–8823 (2015).

398. Sólyom, Z. *et al.* The Disordered Region of the HCV Protein NS5A: Conformational Dynamics, SH3 Binding, and Phosphorylation. *Biophys. J.* **109**, 1483–1496 (2015).

399. Chiliveri, S. C. *et al.* Experimental NOE, Chemical Shift, and Proline Isomerization Data Provide Detailed Insights into Amelotin Oligomerization. *J. Am. Chem. Soc.* **145**, 18063–18074 (2023).

400. Gruber, T. *et al.* Macromolecular Crowding Induces a Binding Competent Transient Structure in Intrinsically Disordered Gab1. *J. Mol. Biol.* **434**, 167407 (2022).

401. Kämpf, K. *et al.* What Drives (15)N Spin Relaxation in Disordered Proteins? Combined NMR/MD Study of the H4 Histone Tail. *Biophys. J.* **115**, 2348–2367 (2018).

402. Salvi, N. *et al.* Convergent views on disordered protein dynamics from NMR and computational approaches. *Biophys. J.* **121**, 3785–3794 (2022).

403. Milles, S. *et al.* Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell* **163**, 734–745 (2015).

404. Naudi-Fabra, S., Tengo, M., Jensen, M. R., Blackledge, M. & Milles, S. Quantitative Description of Intrinsically Disordered Proteins Using Single-Molecule FRET, NMR, and SAXS. *J. Am. Chem. Soc.* **143**, 20109–20121 (2021).

405. Fernández, C. & Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr. Opin. Struct. Biol.* **13**, 570–580 (2003).

406. Barretto, N. *et al.* The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J. Virol.* **79**, 15189–15198 (2005).

407. Minton, A. P. & Wilf, J. Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry* **20**, (1981).

408. Krainer, G. *et al.* Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. *Nat. Commun.* **12**, 1–14 (2021).

409. Adame-Arana, O., Weber, C. A., Zaburdaev, V., Prost, J. & Jülicher, F. Liquid Phase Separation Controlled by pH. *Biophys. J.* **119**, 1590–1605 (2020).

410. Cinar, H. *et al.* Temperature, Hydrostatic Pressure, and Osmolyte Effects on Liquid-Liquid Phase Separation in Protein Condensates: Physical Chemistry and Biological Implications. *Chemistry* **25**, 13049–13069 (2019).

411. Dignon, G. L., Zheng, W., Kim, Y. C. & Mittal, J. Temperature-Controlled Liquid-Liquid Phase Separation of Disordered Proteins. *ACS Cent. Sci.* **5**, 821–830 (2019).

412. Meca, E., Fritsch, A. W., Iglesias-Artola, J. M., Reber, S. & Wagner, B. Predicting disordered regions driving phase separation of proteins under variable salt concentration. *Front. Phys.* **11**, 1–13 (2023).

413. Carlson, C. R. *et al.* Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. *Mol. Cell* **80**, 1092-1103.e4 (2020).

414. OVERBEEK, J. T. & VOORN, M. J. Phase separation in polyelectrolyte solutions; theory of complex coacervation. *J Cell Physiol Suppl.* **49**, 7–22 (1957).

415. Chiga, M. *et al.* Dietary salt regulates the phosphorylation of OSR1/SPAK kinases and the sodium chloride cotransporter through aldosterone. *Kidney Int.* **74**, 1403–1409 (2008).

416. Xu, Y. *et al.* Effects of Different Na+ Concentrations on cAMP-Dependent Protein Kinase Activity in Postmortem Meat. *Foods* **13**, (2024).

417. Piala, A. T. *et al.* Chloride sensing by WNK1 involves inhibition of autophosphorylation. *Sci. Signal.* **7**, ra41 (2014).

418. Huang, C.-L. & Cheng, C.-J. A unifying mechanism for WNK kinase regulation of sodium-chloride cotransporter. *Pflugers Arch.* **467**, 2235–2241 (2015).

419. Carlson, G. M. & Graves, D. J. Site of Action and Biphasic Effect of Neutral Salts in the Phosphorylase Kinase Reaction. *Biochemistry* **15**, 4476–4481 (1976).

420. Krebs, E. G. *et al.* Purification and properties of rabbit skeletal muscle phosphorylase B kinase. *Biochemistry* **3**, 1022–1033 (1964).

421. Sacktor, B., Wu, N. C., Lescure, O. & Reed, W. D. Regulation of muscle phosphorylase b kinase activity by inorganic phosphate and calcium ions. *Biochem. J.* **137**, 535–542 (1974).

422. Utida, S. Effect of Sodium Chloride on Alkaline Phosphatase Activity in Intestinal Mucosa of the Rainbow Trout. *Proc. Jpn. Acad.* **43**, 783–788 (1967).

423. Poe, R. W., Sangadala, V. S. & Brewer, J. M. Effects of various salts on the steady-state enzymatic activity of E. coli alkaline phosphatase. *J. Inorg. Biochem.* **50**, 173–180 (1993).

424. Beaugé, L. & Campos, M. A. Calcium inhibition of the ATPase and phosphatase activities of (Na+ + K+)-ATPase. *Biochim. Biophys. Acta* **729**, 137–149 (1983).

425. Hou, Z. & Liu, H. Mapping the Protein Kinome: Current Strategy and Future Direction. *Cells* **12**, (2023).

426. Kamps, J. *et al.* The N-terminal domain of the prion protein is required and sufficient for liquid-liquid phase separation: A crucial role of the Aβ-binding domain. *J. Biol. Chem.* **297**, 100860 (2021).

427. Schlütermann, D. *et al.* FIP200 controls the TBK1 activation threshold at SQSTM1/p62-positive condensates. *Sci. Rep.* **11**, 1–18 (2021).

428. López-Palacios, T. P. & Andersen, J. L. Kinase regulation by liquid–liquid phase separation. *Trends Cell Biol.* **33**, 649–666 (2023).