

**Structural insights into the activity regulation of  
full-length non-structural protein 1  
from SARS-CoV-2  
and  
substrate recruitment by the hexameric  
MecA–ClpC complex**

Von der Naturwissenschaftlichen Fakultät der  
Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Ying Wang

Seoksa (Master), Sookmyung Women's University Seoul

2023

Referentin: Prof. Dr. ric.Teresa Carlomagno

Korreferent: Prof. Dr. rer. nat. Mark Brönstrup

Tag der Promotion: 14.12.2022

## The following publications contributed to this thesis:

1. **Wang Y**, Kirkpatrick J, Zur Lage S, Carlomagno T. (Submitted) Structural insights into the activity regulation of full-length Nonstructural Protein 1 from SARS-CoV-2.
2. **Wang, Y.**, Kirkpatrick, J., Zur Lage, S., Korn, S.M., Neissner, K., Schwalbe, H., Schlundt, A., and Carlomagno, T. (2021).  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  backbone chemical-shift assignments of SARS-CoV-2 non-structural protein 1 (leader protein). *Biomol NMR Assign* 15, 287-295.
3. **Wang Y**, Kirkpatrick J, Hinse P, Carlomagno T. (To be submitted) Substrate recruitment by the hexameric MecA–ClpC complex.

# Table of Contents

<b>Abstract .....</b>	<b>4</b>
<b>Zusammenfassung.....</b>	<b>5</b>
<b>Abbreviations and Symbols.....</b>	<b>6</b>
<b>Chapter 1: Introduction.....</b>	<b>8</b>
<b>1.1 Nonstructural protein 1 (Nsp1) from the severe acute respiratory syndrome– coronavirus 2 (SARS-CoV-2) .....</b>	<b>8</b>
1.1.1 Identification of severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2).....	8
1.1.2 The coronavirus life cycle .....	9
1.1.3 SARS-CoV-2 genome structure, gene expression and structural proteins .....	11
1.1.4 Nonstructural proteins (Nsps) involved in replication and transcription.....	13
1.1.5 Nonstructural protein 1 (Nsp1) mediated translation inhibition, viral evasion, and lower immune responses .....	18
<b>1.2 Substrate protein (ComK) recruitment by the hexameric MecA–ClpC complex.....</b>	<b>24</b>
1.2.1 AAA+ protease complexes.....	24
1.2.2 Hexameric MecA–ClpCP complex .....	26
1.2.3 The substrate ComK.....	31
1.2.4 The questions we ask and what we find.....	33
<b>1.3 Methodological Background .....</b>	<b>35</b>
1.3.1 Nuclear Magnetic Resonance (NMR) spectroscopy.....	35
1.3.1.1 Genesis of the NMR signal .....	35
1.3.1.2 Spin relaxation .....	39
1.3.1.3 NMR assignment of protein resonances .....	41
1.3.1.4 TROSY-NMR for large biological macromolecules in solution .....	45
1.3.1.5 Structure calculation by NMR.....	49
1.3.2 Small-angle X-ray scattering (SAXS) .....	54
1.3.3 Size-exclusion Chromatography and Multiangle Light Scattering (SEC-MALS).....	57
<b>1.4 References.....</b>	<b>59</b>
<b>Chapter 2: Publications and Manuscripts .....</b>	<b>70</b>
2.1 Manuscript 1 .....	70
Structural insights into the activity regulation of full-length non-structural protein 1 from SARS- CoV-2.....	70
2.2 Publication 1 .....	105
<sup>1</sup> H, <sup>13</sup> C, and <sup>15</sup> N backbone chemical-shift assignments of SARS-CoV-2 non-structural protein 1 (leader protein).....	105
2.3 Manuscript 2 .....	115

Substrate recruitment by the hexameric MecA–ClpC complex .....	115
<b>Chapter 3: Appendix .....</b>	<b>146</b>
Appendix 1: Cloning Primers .....	146
Appendix 2: SAXS Data .....	148
Appendix 3: Chemical-shift assignments .....	149
<b>Acknowledgements .....</b>	<b>162</b>
<b>Curriculum Vitae and List of publications.....</b>	<b>163</b>

## Abstract

Since the outbreak of the COVID-19 pandemic in December 2019, SARS-CoV-2 has generated awareness for the requirement of novel antiviral drugs that target new proteins. Previous studies have pointed to the pathogenic significance of the non-structural protein 1 (Nsp1), which was proposed to be a major virulence factor due to its dual role in host translation inhibition and viral replication of SARS-CoV-2. The precise mechanisms of these two functions of Nsp1 are still uncovered. Here, I report the backbone chemical-shift assignments and the atomic-resolution NMR structure of full-length Nsp1 from SARS-CoV-2 solved by NMR. I found that Cov-2 Nsp1 consists of a folded N-terminal domain and an unfolded C-terminal region. Previous studies have identified a surface of the folded N-terminal domain of Nsp1 that associates with both host mRNA, in its function as inhibitor of the host protein translation, and the viral 5'-UTR RNA, in its function as promoter of viral protein translation. I found that the acidic C-terminal tail of Nsp1 folds back on this surface and masks the RNA binding site. A recent Cryo-EM study has shown that the end of C-terminal region of Nsp1 interacts with the mRNA entry tunnel on the 40S subunit of the ribosome, thus inhibiting host mRNA entry and promoting its degradation. I propose that the RNA binding site on the Nsp1 N-terminal domain is protected by its C-terminal tail before Nsp1 contacts the ribosome. Upon ribosome binding, the C-terminal tail is displaced, and the RNA binding site is exposed to recruit either host mRNA or the viral 5' UTR. My findings have consequences for the design of drugs targeting the RNA binding surface of Nsp1, as it demonstrates the need to develop a molecule that can not only recognize this surface with high affinity, but also displace the C-terminal tail to render this surface accessible.

In bacteria, the unfoldase ClpC is a member of the conserved Hsp100/Clp family of AAA+ ATPases and is involved in various cellular processes. The functional form of ClpC is an hexameric assembly, which is responsible for controlled unfolding of substrate proteins. The ClpC hexamer can further associate with the protease ClpP to form a complete protein degradation machine. MecA functions as an adaptor protein of ClpC and is necessary to both promote the formation of the functional ClpC hexamer and to recruit specific substrate proteins, such as the transcription factor ComK. Degradation of ComK via MecA-mediated recruitment to the ClpCP complex is part of the regulatory mechanisms of the development of cell competence. The mechanisms of selective substrate recruitment are still unknown, due to the conspicuous conformational dynamics and heterogeneity that characterize this step. To understand how the substrate ComK is recruited to the MecA–ClpC complex, I reconstituted the ComK–MecA–ClpC complex in vitro and applied nuclear magnetic resonance (NMR) spectroscopy and other biophysical techniques, such as multi-angle light scattering. I found that addition of ComK stabilizes the MecA–ClpC complex by forming a homogeneous ternary protein complex, which contains a ClpC hexamer, four MecA and two ComK molecules in the presence of ATP. The structural differences between the MecA–ClpC and the ComK–MecA–ClpC complexes were also monitored by small-angle X-ray scattering datasets, which furthered confirmed the presence of the interaction between ComK and MecA.

**Keywords:** SARS-CoV-2; non-structural protein 1; ComK–MecA–ClpC complex; NMR spectroscopy

## Zusammenfassung

Seit dem Ausbruch der COVID-19-Pandemie im Dezember 2019 hat SARS-CoV-2 das Bewusstsein für den Bedarf an neuartigen antiviralen Medikamenten geweckt, die neue Proteine abzielen. Frühere Studien haben auf die pathogene Bedeutung des Unstrukturiertenproteins 1 (Nsp1) hingewiesen, das aufgrund seiner doppel Rolle bei der Inhibierung von Translationsprozessen im Wirt, sowie der viralen Replikation von SARS-CoV-2 als wichtiger Virulenzfaktor betrachtet wird. Die genauen Mechanismen dieser beiden Funktionen von Nsp1 sind noch weitestgehend unerforscht. Ich mittels NMR erhobener Daten die Zuordnungen der chemischen Verschiebung des Rückgrats sowie die atomare NMR-Struktur von Nsp1 in voller Länge aus SARS-CoV-2. Ich fand heraus, dass Cov-2 Nsp1 aus einer gefalteten N-terminalen Domäne und einer ungefalteten C-terminalen Region besteht. Frühere Studien haben einen Oberflächenbereich der gefalteten N-terminalen Domäne von Nsp1 identifiziert, die sowohl mit Wirts-mRNA in ihrer Funktion als Inhibitor der Translation assoziiert, als auch mit der viralen 5'-UTR-RNA in ihrer Funktion als Promotor des viralen Proteins. Ich fand heraus, dass sich der saure C-terminale Ende von Nsp1 auf dieser Oberfläche zurückfaltet und die RNA-Bindungsstelle maskiert. Eine kürzlich durchgeführte Cryo-EM-Studie hat ergeben, dass das Ende der C-terminalen Region von Nsp1 mit dem mRNA-Eintrittstunnel auf der 40S-Untereinheit des Ribosoms interagiert, wodurch der mRNA-Eintritt des Wirts gehemmt und dessen Abbau gefördert wird. Ich schlage vor, dass die RNA-Bindungsstelle auf der N-terminalen Domäne von Nsp1 durch ihr C-terminalen Ende geschützt wird, bevor Nsp1 das Ribosom kontaktiert. Bei der Ribosomenbindung wird der C-terminale Bereich verschoben und die RNA-Bindungsstelle wird freigelegt, um entweder Wirts-mRNA oder die virale 5'-UTR zu rekrutieren. Meine Ergebnisse haben Konsequenzen für das Design von Medikamenten, die auf die RNA-bindende Oberfläche von Nsp1 abzielen, da sie die Notwendigkeit aufzeigen ein Molekül zu entwickeln, welches diese Oberfläche nicht nur mit hoher Affinität erkennen kann, sondern auch das flexible C-terminale Ende verdrängt, um diese Oberfläche zugänglich zu machen.

In Bakterien ist die *Unfoldase* ClpC, ein Mitglied der Hsp100/Clp-Familie von AAA+ ATPasen, und ist an verschiedenen zellulären Prozessen beteiligt. Die funktionelle Form von ClpC folgt einer hexameren Anordnung, die für die kontrollierte Entfaltung von Substratproteinen verantwortlich ist. Das ClpC-Hexamer kann ferner mit der Protease ClpP assoziieren, um eine vollständige Proteinabbaumaschine zu bilden. MecA fungiert als Adapterprotein von ClpC und ist notwendig um sowohl die Bildung des funktionellen ClpC-Hexamers zu fördern, als auch spezifische Substratproteine wie unter anderem den Transkriptionsfaktor ComK zu rekrutieren. Der Abbau von ComK über MecA-vermittelte Rekrutierung zum ClpCP-Komplex ist Teil der Regulationsmechanismen der Entwicklung von Zellkompetenz. Dies liegt hauptsächlich an der auffällige Konformationsdynamik und Heterogenität des Komplexes, die diesen Schritt charakterisieren. Um zu verstehen, wie das Substrat ComK zum MecA-ClpC-Komplex rekrutiert wird, wurde der ComK-MecA-ClpC-Komplex in vitro rekonstituiert und Kernspinresonanz(NMR)-Spektroskopie sowie weitere biophysikalische Techniken wie zum Beispiel Mehrwinkel-Lichtstreuung angewendet. Ich fand heraus, dass die Zugabe von ComK den MecA-ClpC-Komplex durch Bildung eines homogenen ternären Proteinkomplexes stabilisiert, der in Gegenwart von ATP ein ClpC-Hexamer, vier MecA- und zwei ComK-Moleküle enthält. Die strukturellen Unterschiede zwischen den MecA-ClpC- und den ComK-MecA-ClpC-Komplexen wurden auch durch Kleinwinkel-Röntgenstreuungsdatensätze überwacht, die das Vorhandensein der Wechselwirkung zwischen ComK und MecA weiter bestätigten.

**Schlüsselwörter:** SARS-CoV-2; Nichtstrukturprotein 1; ComK-MecA-ClpC-Komplex; NMR-Spektroskopie

## Abbreviations and Symbols

AAA+	ATPases associated with various cellular activities
ATP	Adenosintriphosphate
aa	Amino acid
bp	Base pairs
COVID-19	<i>Coronavirus</i> disease 2019
CTD	C-terminal domain
Cryo-EM	Cryogenic-electron microscopy
DWB	Double walker B mutations
DNAse	Deoxyribonuclease
D <sub>max</sub>	Maximum dimension of the particle
FL	Full-length
EDTA	Ethylenediaminetetraacetic acid
kDa	Kilodalton
kb	Kilobasepair
HMQC	Heteronuclear multiple quantum coherence
HSQC	Heteronuclear single quantum coherence
RDC	Residual dipolar coupling
RMSD	Root-mean-square deviation
ILV	Isoleucine, leucine, and valine
IFN-β	Interferon-beta
mRNA	Messenger RNA
M <sub>w</sub>	Molecular weight
NMR	Nuclear magnet resonance
Nsp	Non-structural protein
NTD	N-terminal domain
NTDL	N-terminal domain with linker
NaN <sub>3</sub>	Sodium azide
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect spectroscopy
NBD	Nucleotide-binding domain
LB	Lysogeny broth
IPTG	Isopropyl β-d-1-thiogalactopyranoside
MALS	Multi Angle Light Scattering
rpm	Revolutions per minute
PCR	Polymerase chain reaction



SARS-CoV-2	Severe acute respiratory syndrome–coronavirus 2
SCoV	Severe acute respiratory syndrome coronaviruses
SAXS	Small-angle X-ray scattering
SEC	Size exclusion chromatography
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SL	Stem-loop
TROSY	Transverse relaxation optimized spectroscopy
TRIS	Tris(hydroxymethyl)aminomethane
TCEP	Tris-(2-carboxyethyl)-phosphin
TEV	Tobacco Etch Virus
UTR	Untranslated region
WT	Wild type
ORF	Open-reading-frames

## Chapter 1: Introduction

Chapter 1 is divided in three sections. The first section frames my structural studies of the non-structural protein 1 (Nsp1) from SARS-CoV-2 in a biological context. The second section provides an overview on the bacterial AAA+ protease complexes and sets the questions I ask in this work. The third section provides the methodological background of my work.

### 1.1 Nonstructural protein 1 (Nsp1) from the severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2)

#### 1.1.1 Identification of severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2)

The tremendous disruption to public health and the global economy caused by the ongoing COVID-19 pandemic has initiated massive scientific research into severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2).

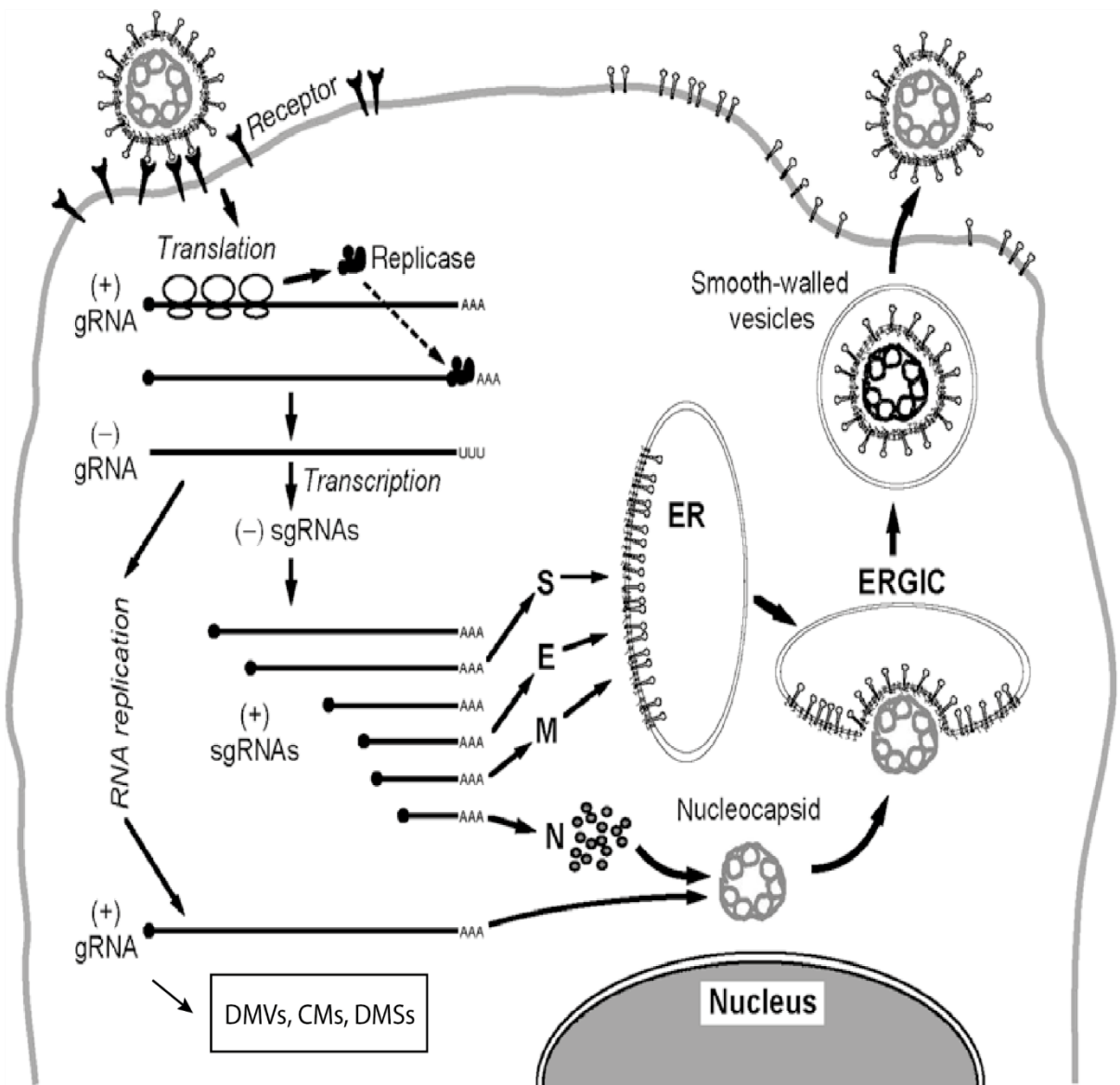
In March of 2020, the World Health Organization declared the novel coronavirus disease 2019 (COVID-19) a global pandemic. Since the first patient was identified in Wuhan, Hubei province, China, at least 6,090,000,000 cases of infection have been reported globally and approximately 6,520,000 deaths ascribed to this pandemic. The etiological agent of COVID-19 was found to be a novel coronavirus named severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2), a sister virus of the severe acute respiratory syndrome–coronavirus 1 (SARS-CoV-1) that was responsible for the SARS epidemic in 2003 (Gorbalenya *et al.*, 2020; Zhou *et al.*, 2020).

Coronaviruses (CoVs) belong to the family of Coronaviridae. These are enveloped positive-sense single-stranded RNA viruses that cause respiratory and enteric disease among mammals and avian species. CoVs can be divided in four genera: *alpha-coronavirus*, *beta-coronavirus*, *gamma-coronavirus* and *delta-coronavirus*. Alpha-coronaviruses and beta-coronaviruses more commonly infect mammalian species, whereas gamma-coronaviruses and delta-coronaviruses prevalently infect birds and fish (Gorbalenya *et al.*, 2020; Kneller *et al.*, 2020; Mariano *et al.*, 2020). Up to now, seven coronaviruses that belong to alpha- and beta-coronaviruses genera are known to infect mammalian species (Hartenian *et al.*, 2020). Human alpha-coronaviruses such as HCoV-229E and HCoV-OC43, together with the more recently identified HCoV-NL63 and HCoV-HKU1, generate seasonal and mild respiratory tract infections with common cold-like symptoms (Paules *et al.*, 2020). In strong contrast, SARS-CoV-2, together with its homolog SARS-CoV-1 and Middle East respiratory syndrome

coronavirus (MERS-CoV), which have emerged in human population within the past two decades, are famed agents of life-threatening acute respiratory pathologies and severe lung injury diseases (Skowronski *et al.*, 2005; Tang *et al.*, 2015).

### **1.1.2 The coronavirus life cycle**

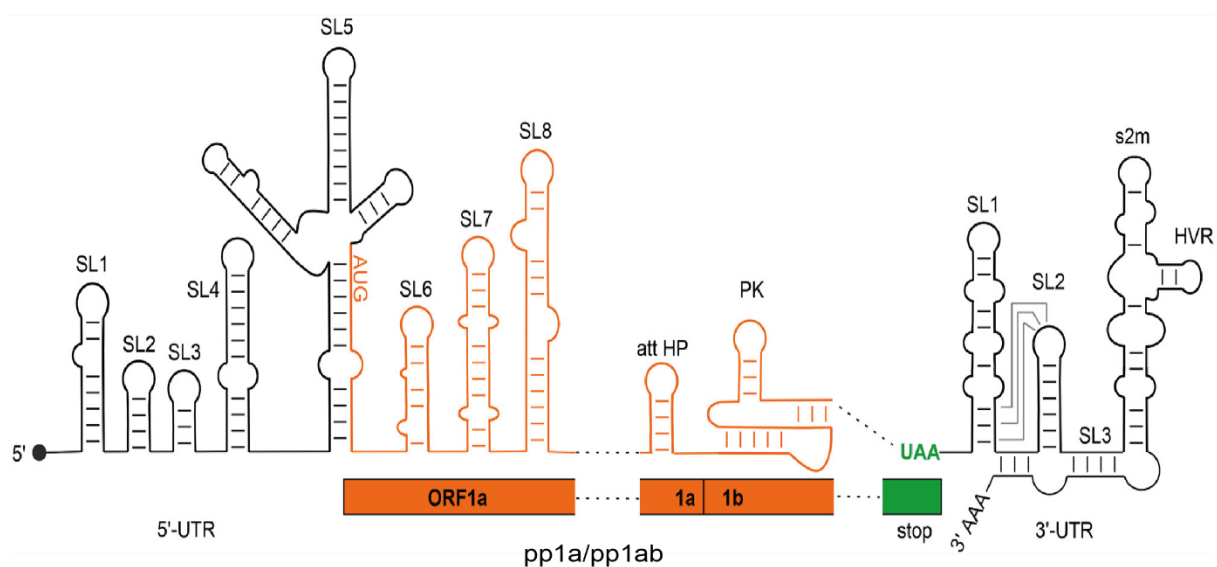
The life cycle of a coronavirus is initiated by the binding of virions to host receptors. Coronavirus virion is composed of four structural proteins, namely spike (S), envelope (E), membrane (M), and nucleocapsid (N). The enveloped, single-stranded positive-sense RNA genome of the virus is coated with N protein and enclosed in a lipid bilayer which includes the other three membrane proteins S, M, and E. The latter two proteins guarantee the incorporation of viral components into new virions during the viral assembly process, whereas the S protein is necessary for viral entry, as it binds to the target cell and provides specificity for cellular entry receptors (such as angiotensin-converting enzyme 2 (ACE2)), together with host factors (such as the cell surface serine protease TMPRSS2). Upon cell entry, RNA release and uncoating, the incoming genomic RNA is directly translated into the replicase polyprotein (pp1a and pp1ab), which is proteolytically cleaved by viral cysteine proteases and subsequently yields the 16 non-structural proteins (Nsps) that come together to form the viral replication and transcription complex (RTC). These proteins then use the genomic RNA as the template for viral replication, which involves the synthesis of the complementary full-length negative-strand intermediates, of both progeny genomes and a nested set of subgenomic mRNAs (sg mRNAs). Viral genome replication is thought to occur on the double-membrane vesicles (DMVs), convoluted membranes (CMs) and small open double-membrane spherules (DMSs), which create a protective environment for viral genomic RNA replication and transcription of sg mRNAs. This in turn leads to the accommodation of the translated membrane-bound structural proteins, M, S, and E into the endoplasmic reticulum (ER), together with the helical nucleocapsids that are assembled by the nucleocapsid (N) protein and genomic RNA in the cytoplasm, forming virions by budding into the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). Finally, progeny virions are exported from the infected cells by exocytosis. During infection by certain coronaviruses (not all), the S protein is cleaved into S1 and S2 domains. A portion of S protein that has not been incorporated into the virions is eventually transported to the plasma cell surface, which results in the fusion of virion-containing cells with their neighbor, uninfected cells. This mechanism empowers the spread of infection independent of the action of extracellular virus, thus providing a way of viral evasion from immune surveillance (Stadler *et al.*, 2003; Masters, 2006; Perlman & Netland, 2009; Hartenian *et al.*, 2020; V'Kovski *et al.*, 2021) (Fig. 1.1.1).



**Fig 1.1.1. Life cycle of the coronavirus.** This figure is adapted from (Masters, 2006).

### 1.1.3 SARS-CoV-2 genome structure, gene expression and structural proteins

SARS-CoV-2 is an enveloped, positive-sensed, single-stranded RNA genome of approximately 30kb in length. The genome is 5'-capped and 3' polyadenylate. At the 5' end of the genome, there is a leader sequence of ~ 60-95 nucleotides, which is followed by a highly structured untranslated region (5'-UTR) and the initiation codon for open reading frame (ORF). At the 3' end, there is another untranslated region (3'-UTR) following the end of the last ORF. Secondary structures such as stem loop and pseudoknot have been characterized recently in both UTR regions (Fig. 1.1.2). The function of these putative *cis*-acting RNA elements has been predicted in different beta-coronaviruses and are associated with regulation of replication, subgenomic mRNA (sg mRNA) production and translation (Madhugiri *et al.*, 2016; Wacker *et al.*, 2020).

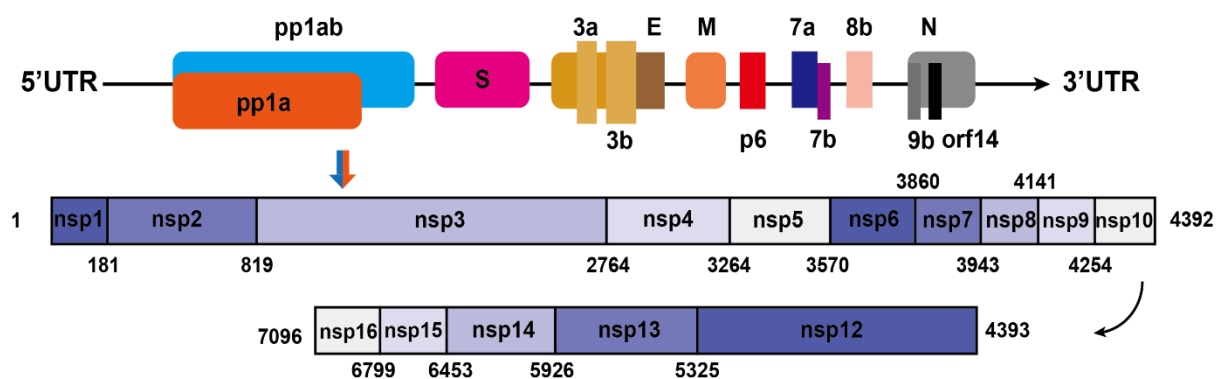


**Figure 1.1.2 Overview of *cis*-acting RNA elements of the SARS-CoV-2 genome.** Black: untranslated regions; orange: coding regions; green: stop codon at the end of the last ORF. This figure is adapted from (Wacker *et al.*, 2020).

Approximately two-thirds of the genome, located in the 5' genomic region, encodes the replicase gene, which contains two open reading frames, ORF1a and ORF1b. Translation of ORF1a yields polyprotein 1a (pp1a), and -1 ribosomal frameshifting allows translation of both ORF1a and ORF1b to yield pp1ab (Sola *et al.*, 2015). Sixteen non-structural proteins are co-translationally and post-translationally released from pp1a (nsp1–11) and pp1ab (nsp1–10, nsp12–16) upon proteolytic cleavage by two cysteine proteases that are located within Nsp3 (papain-like protease; PLpro) and Nsp5 (chymotrypsin-like protease) (V'Kovski *et al.*, 2021). The 3' third of the genome contains four structural proteins [Spike (S)-Envelope (E)-Membrane

(M)-Nucleocapsid (N)], and eight small accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, and orf14) that exist in addition to the structural proteins (Fig. 1.1.3).

Among the four structural proteins, the spike (S) surface glycoprotein is important for invasion of host cells through binding to the host cell receptor, namely the angiotensin converting enzyme 2 (ACE2) (Hoffmann *et al.*, 2020). The nucleocapsid protein N plays a multifaceted role in the viral infection cycle. It binds to, and packages the viral RNA against the internal surface of the envelope protein (E) (Chang *et al.*, 2014; McBride *et al.*, 2014). The membrane protein (M) is embedded in the viral membrane and was reported to drive the assembly of new virions into host cells (Masters, 2006; Neuman *et al.*, 2011). The E protein was described to self-interact, through its transmembrane domain, to form an ion channel (Pervushin *et al.*, 2009). Interaction between the C-terminus of the E protein and the M protein initiate the virus assembly and budding into the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) of host cells (Schoeman & Fielding, 2019).



**Figure 1.1.3 Genome composition of SARS-CoV-2.** Schematic diagram of the genome organization and the encoded proteins for SARS-CoV-2. This figure is modified from (Wu *et al.*, 2020).

### 1.1.4 Nonstructural proteins (Nsps) involved in replication and transcription

Upon cell entry, the virus exploits the host translational machinery to produce the polypeptides corresponding to ORF1a and ORF1ab. The polypeptides are then proteolytically cleaved to yield 16 so-called non-structural proteins. Recently, possible functions of nonstructural proteins (Nsps) have been predicted based on their structural characterizations by canonical structural biology approaches, such as X-ray crystallography, NMR spectroscopy and cryo-electron microscopy (cryo-EM). Most of Nsps have been shown, or are predicted, to have enzymatic functions and are involved in assembling host-viral replication and transcription complexes (Stadler *et al.*, 2003; Khailany *et al.*, 2020) (Table 1.1). Identification and characterization of non-structural proteins within the SARS-CoV-2 genome is critically

important for us to understand how the virus operates and to devise effective approaches to mitigate the impact of the pandemic.

**Table 1.1 Functions and available PDB structures of Nonstructural proteins (Nsp 1-16).** Table is from (Mariano *et al.*, 2020), modified.

<b>Protein name</b>	<b>Functions</b>	<b>PDB-IDs</b>	<b>Ligand</b>	<b>References</b>
Nsp1	Host mRNA degradation / Translation Inhibition / attenuation of innate Immune response to the virus	6ZLW	Ribosomal complex	(Thoms <i>et al.</i> , 2020)
		6ZOJ	Ribosomal complex	(Schubert <i>et al.</i> , 2020)
		7K3N	–	(Semper <i>et al.</i> , 2021)
		7K7P	–	(Clark <i>et al.</i> , 2021)
		7EQ4	–	(Zhao <i>et al.</i> , 2021)
Nsp2	Putative role in apoptosis induction	–	–	–
Nsp3	Polyprotein processing / Viral replication Nsp3-4-6 complex for domain endoplasmic (ER) modification / Formation of double membrane vesicles (DMVs) / Papain-like protease domain / interferon antagonist	6YWK	HEPES	–
		6YWL	ADP-ribose	–
		6YWM	MES	–
		6W6Y	AMP	–
		6WUU	VIR250 ubiquitin propargylamide	(Rut <i>et al.</i> , 2020)
		6XAA	ISG15 C-terminal Domain propargylamide	(Fu <i>et al.</i> , 2020)
		6XA9	ISG15	(Fu <i>et al.</i> , 2020)
			ADP ribose	–
			MES	–
6W02	ADP ribose Peptide inhibitor VIR251	–		

Nsp4	Nsp3-4-6 complex for endoplasmic (ER) modification / Formation of double membrane vesicles (DMVs)	–	–	–
Nsp5	Polyprotein processing / Inhibition of interferon signalling / 3C-like cysteine proteinase	6W63 6WTT 6XCH 6WQF 7C8R 7C8T 6WNP 6WTK 7BRR 7BRP 7BUY 6Y2G 6Y2F 6LZE 6M0K 6YVA 6YT8 6YNQ	Inhibitor X77 Inhibitor GC-376 Leupeptin – TG-0203770 TG-0205221 Boceprevir Feline coronavirus drug GC376 – Carmofur Aalpha-ketoamide 13b Aalpha-ketoamide 13b Inhibitor 11b Inhibitor 11b mISG15 Pyrithione zinc AZD6482	– – (Ma <i>et al.</i> , 2020)  (Kneller <i>et al.</i> , 2020) – – – – – (Jin <i>et al.</i> , 2020) (Jin <i>et al.</i> , 2020) (Jin <i>et al.</i> , 2020) (Zhang <i>et al.</i> , 2020) (Zhang <i>et al.</i> , 2020) (Zhang <i>et al.</i> , 2020) (Dai <i>et al.</i> , 2020) (Dai <i>et al.</i> , 2020)  
Nsp6	Nsp3-4-6 complex for endoplasmic (ER) modification / Formation of double membrane vesicles (DMVs)	–	–	–
Nsp7	Nsp7-8 replication-transcription complex, involved in virus replication / Cofactor for RNA-dependent RNA polymerase	7BTF 7BV1 7BV2 7BW4	Nsp8, Nsp12 Nsp8, Nsp12 Nsp8, Nsp12, Remdesivir Nsp8, Nsp12	(Chen <i>et al.</i> , 2020) (Yin <i>et al.</i> , 2020) (Yin <i>et al.</i> , 2020) (Peng <i>et al.</i> , 2020)



		6YYT	Nsp8	(Hillen <i>et al.</i> , 2020)
		6M5I	Nsp8 C-term	–
		7C2K	Nsp8, Nsp12	(Wang <i>et al.</i> , 2020)
		6XEZ	Nsp8, Nsp12,	(Chen <i>et al.</i> , 2020)
		6XQB	Nsp13	(Shi <i>et al.</i> , 2020b)
		6XIP	Nsp8, Nsp12	–
		6YHU	Nsp8, Nsp12	–
			Nsp8	
Nsp8	Nsp7-8 replication-transcription involved in virus replication / RNA-primase of the virus / cofactor for RNA-dependent RNA polymerase	7BTF	Nsp7, Nsp12	(Chen <i>et al.</i> , 2020)
		7BV1	Nsp7, Nsp12	(Yin <i>et al.</i> , 2020)
		7BV2	Nsp7, Nsp12	(Yin <i>et al.</i> , 2020)
			Remdesivir	(Peng <i>et al.</i> , 2020)
		7BW4	Nsp7, Nsp12	(Hillen <i>et al.</i> , 2020)
		6YYT	Nsp7, Nsp12	–
		6M5I	Nsp7	(Wang <i>et al.</i> , 2020)
		7C2K	Nsp7, Nsp12	(Chen <i>et al.</i> , 2020)
		6XEZ	Nsp7, Nsp12,	(Shi <i>et al.</i> , 2020b)
		6XQB	Nsp13	
		6XIP	Nsp7, Nsp12	–
		6YHU	Nsp7, Nsp12	–
			Nsp7	
Nsp9	RNA replicase / Binding of single-stranded RNA	6W9Q	Main protease	–
		6W4B	derived peptide	–
		6WXD	–	–
			–	
Nsp10	Play multiple roles in RNA capping/ Modulates (cofactor) Nsp14 and Nsp16 activities/ Nsp10-Nsp11-Nsp14-Nsp16-complex/ Nucleoside-2'O-methyl transferase (2'O-MTase) cofactor	7C2I	Nsp16, SAM	–
		7C2J	Nsp16, SAM	–
		6W61	Nsp16, SAM	–
		6XKM	Nsp16, SAM	–
		6WJT		–

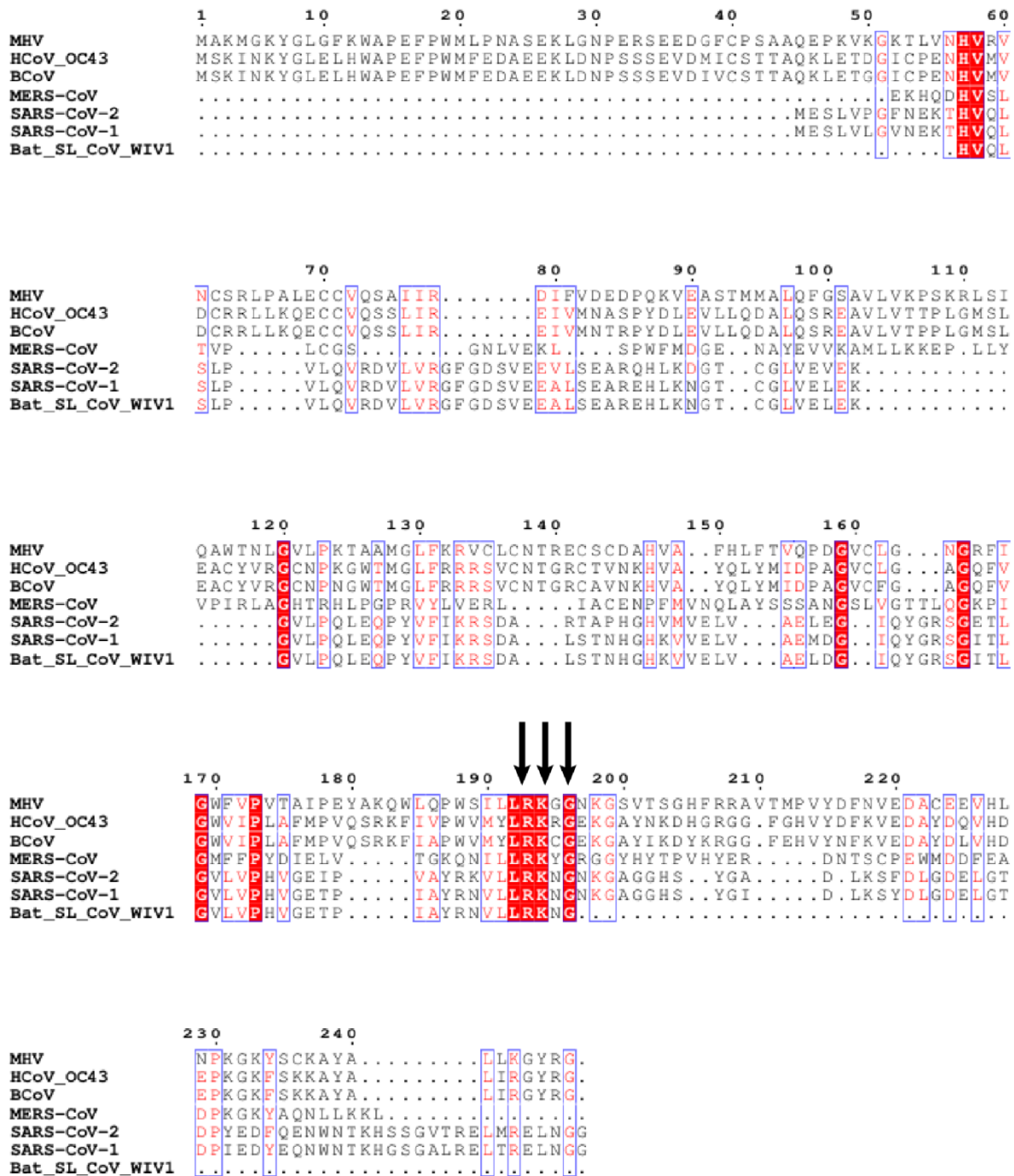
		6WVN	Nsp16, S-Adenosyl-L-Homocysteine	–
		6WQ3	Nsp16, S-Adenosylmethionine	–
		6WKS	Nsp16, S-Adenosylmethionine, 7-methyl-GpppA	–
		6ZCT	Nsp16, S-Adenosylmethionine, 7-methyl-GpppA	–
		6YZ1	Nsp16, S-Adenosylmethionine, 7-methyl-GpppA	–
		7BQ7	Nsp16, Sinefungin	–
		6W75	Nsp16, Sinefungin	–
		6W4H	Nsp16, S-Adenosyl-L-Homocysteine	(Rosas-Lemus <i>et al.</i> , 2020)
			Nsp16	
			Nsp16, S-Adenosyl-L-Homocysteine	
Nsp11	–	–	–	–
Nsp12	RNA-dependent RNA polymerase (RdRp) / part of the replication-transcription complex / nucleotidyltransferase	7BTF	Nsp7, Nsp8	(Gao <i>et al.</i> , 2020)
		6M71	Nsp7, Nsp8	(Gao <i>et al.</i> , 2020)
		7BV1	Nsp7, Nsp8	(Yin <i>et al.</i> , 2020)
		7BV2	Nsp7, Nsp8	(Yin <i>et al.</i> , 2020)
		7BW4	Nsp7, Nsp8	–
		6YYT	Nsp7, Nsp8	(Peng <i>et al.</i> , 2020)
		6X2G	Nsp7, Nsp8	–
		7C2K	Nsp7, Nsp8	(Wang <i>et al.</i> , 2020)
		7BZF	Nsp7, Nsp8	(Wang <i>et al.</i> , 2020)
		6XEZ	Nsp7, Nsp8, Nsp13	(Chen <i>et al.</i> , 2020)

Nsp13	5'-3' Helicase, part of the replication-transcription complex (RTC).	6XEZ	Nsp7, Nsp8, Nsp12	(Chen <i>et al.</i> , 2020)
Nsp14	RNA cap N7-methyltransferase activity and exonuclease activity /Nsp10-Nsp11-Nsp14-Nsp16 complex/ 3'-5'exonuclease	–	–	–
Nsp15	RNA uridylate-specific endoribonuclease (NendoU)	6WWW	–	(Kim <i>et al.</i> , 2020)
		6WLC	Uridine-5'	(Kim <i>et al.</i> , 2020)
		6W01	Monophosphate	(Kim <i>et al.</i> , 2020)
		6WXC	Citrate	–
		6XDH	Tripiracil	–
		6X41	–	–
		6X1B	3'-uridinemono phosphate	–
		6X1B	GpU	–
Nsp16	S-adenosyl-L-methionine-dependent (SAM)-dependent, nucleoside-2'O-methyl transferase (2'O-Mtase) /Nsp10-Nsp11-Nsp14-Nsp16-complex / RNA cap formation	6W4H	S-adenosylmethionine	(Rosas-Lemus <i>et al.</i> , 2020)
		6W75	adenosylmethionine	–
		6WJT	Nsp10	–
		6WVN	Nsp10, S-Adenosyl-L-Homocysteine	–
		6WQ3	Nsp10, S-Adenosylmethionine	–
		6WRZ	Nsp10, S-Adenosylmethionine	–
		6ZCT	Adenosylmethionine, 7-methyl-GpppA	–
		6YZ1	Nsp10, Sinefungin	–
		7C2I	Nsp10, Sinefungin	–
		7C2J	Nsp10, SAM	–
		6W61	Nsp10, SAM	–
		6W61	Nsp10, SAM	–
		6W61	Nsp10, SAM	–

### **1.1.5 Nonstructural protein 1 (Nsp1) mediated translation inhibition, viral evasion, and lower immune responses**

SARS-CoV-2 Nsp1, the first N-terminal cleavage product released from the replicase polyprotein pp1a/1ab, is 180 residues in length, and shares ~84% amino-acid sequence identity with the homologous protein from SARS-CoV-1.

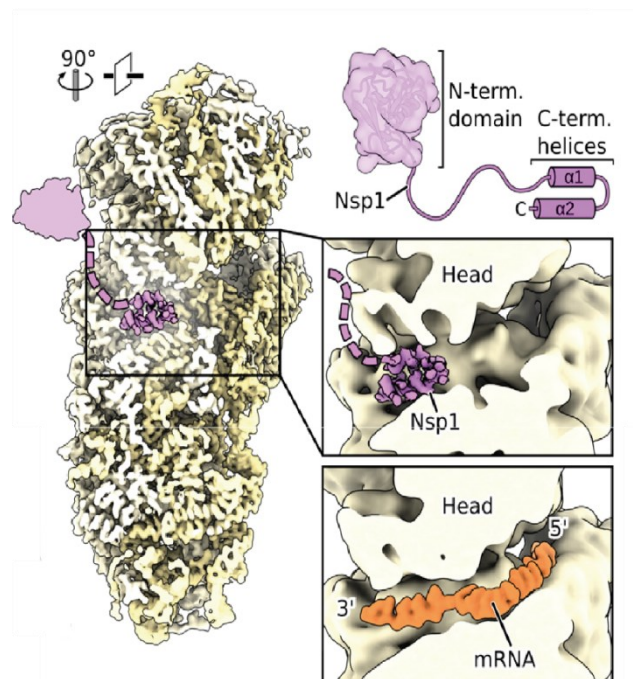
The protein is predicted to consist of a globular domain (residues 10–125), which is flanked by short N-terminal and long C-terminal tails (residues 1–11 and 126–180, respectively), which exhibit a high — but not complete — degree of intrinsic disorder (Wang *et al.*, 2021). A positive charged region, which is composed of eight amino acid residues (LRKxGxKKG) that conserved among Nsp1 proteins of beta-coronaviruses has been predicted to participate in the RNA-binding activity and function of viral replication (Fig. 1.1.4) (Tanaka *et al.*, 2012; Terada *et al.*, 2017).



**Fig 1.1.4. Sequence alignment of SARS-CoV-2 Nsp1 with other Nsp1 proteins that derived from beta-coronaviruses.** Alignment was achieved by using Clustal Omega and the colored figures were generated by ESPrpt 3.0 (Goujon *et al.*, 2010; Sievers *et al.*, 2011; Robert & Gouet, 2014). Residues boxed in red are completely identical. The conserved motif (LRKxGxKG) is indicated in black arrows.

Nsp1 has an important role in inhibition of the translation of the host proteome. This inhibition is mediated by the interaction of the unfolded C-terminal region of Nsp1 with the entrance of the mRNA tunnel in the 40S ribosomal subunit (Fig. 1.1.5). Upon binding to the ribosome, the Nsp1 C-terminus forms two  $\alpha$  helices. This interaction blocks the host mRNA translation during

the initiation stage, leading to a template-dependent endonucleolytic cleavage of the mRNA transcripts, eventually causing host mRNAs degradation of infected cells (Kamitani *et al.*, 2006; Kamitani *et al.*, 2009; Lokugamage *et al.*, 2012; Schubert *et al.*, 2020; Thoms *et al.*, 2020). Deletion or mutations of Nsp1 in its N-terminal folded domain (aa 79-89) or C-terminal unfolded region (K164A, H165A) substantially reduce the impact of Nsp1 viral load, signal transduction and gene expression, underlining the importance of Nsp1 as a major virulence factor for viral pathogenicity (Wathelet *et al.*, 2007; Benedetti *et al.*, 2020). However, it is still unclear how Nsp1 distinguishes between host mRNA and viral RNA to block translation of the first and promoting translation of the second. Due to its dual role in inhibition of host-protein translation and stimulation of viral-protein translation, Nsp1 has been proposed as an attractive drug target for the treatment of COVID-19.

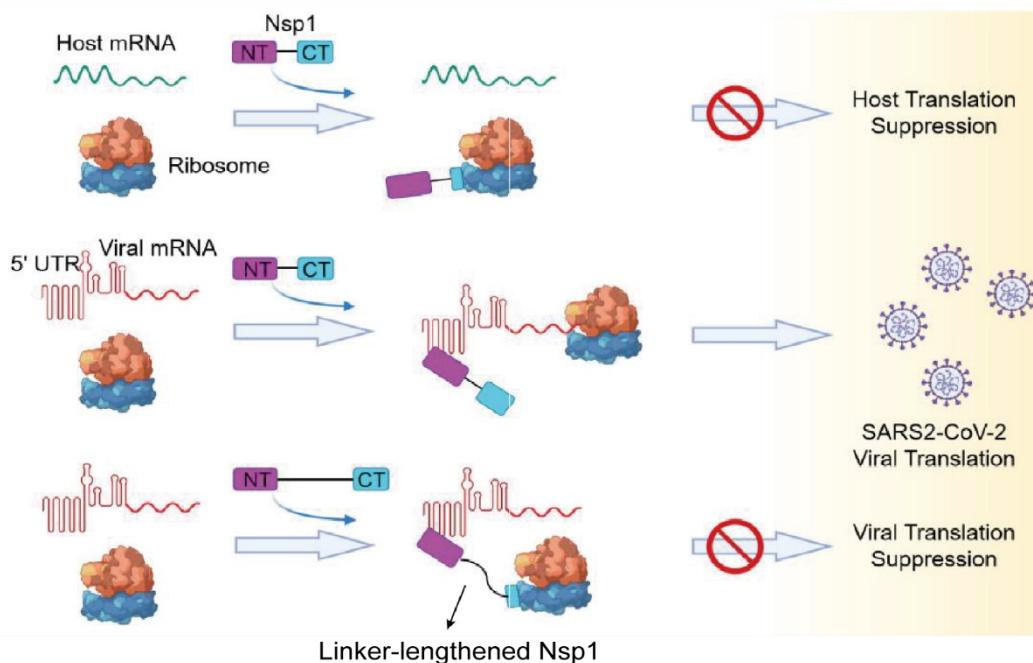


**Figure 1.1.5 Cryo-EM structure of Nsp1-bound ribosomal complex.** Cross-section of the reconstituted Nsp1-40S complex with highlighted central position of Nsp1 within the mRNA tunnel. The putative position of the N-terminal domain of Nsp1 is modelled based on PDB-2HSX and PDB-6Y0G. This figure was used by (Thoms *et al.*, 2020)..

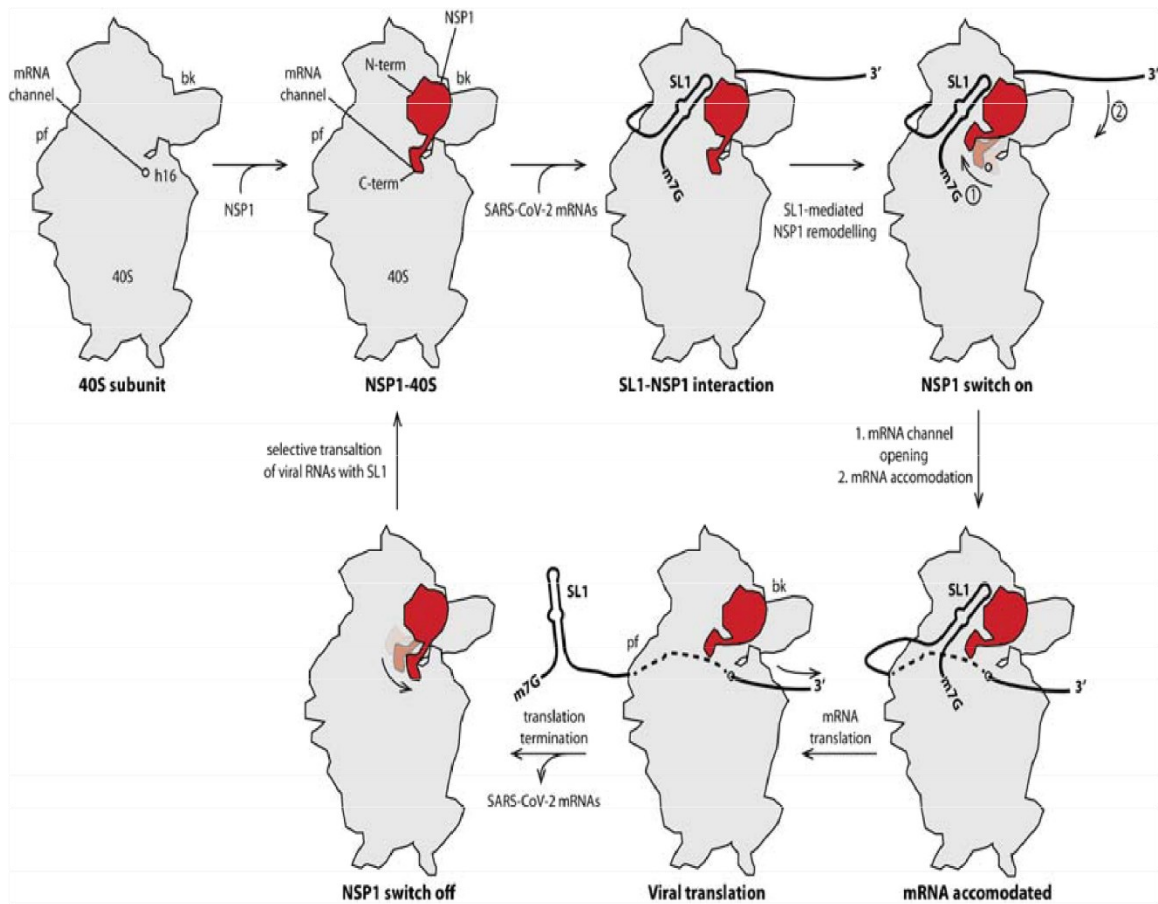
An interaction between the Nsp1 N-terminal domain and the 5' UTR of the viral RNA has been found to help the viral transcripts evade Nsp1-induced translation inhibition in infected cells (Huang *et al.*, 2011; Shi *et al.*, 2020a). However, the 5' UTR has no affinity for free Nsp1 or any component of the ribosome on its own (Tidu *et al.*, 2021). Also, the linker-length between the N-terminal domain of Nsp1 and the region interacting with the ribosome entry tunnel has been found to be critical for supporting translation of the viral genome (Shi *et al.*, 2020a).The

Nsp1-induced translational inhibition suppresses the ability of human lung epithelial cells to respond to IFN- $\beta$  stimulation, thus inactivating virus- and IFN-dependent signaling pathways and resulting in the attenuation of the innate immune responses to the virus (Banerjee *et al.*, 2020). Recently, it has been shown that the Nsp1 globular domain associates directly with the catalytic subunit of DNA polymerase  $\alpha$  of primosome, which is possibly involved in regulating the antiviral response of the host cells (Kilkenny *et al.*, 2022).

The viral RNA has been proposed to escape the Nsp1-mediated translation inhibition through the interaction of its 5' UTR with the N-terminal domain of Nsp1. This interaction should promote an allosteric conformational change that displaces the C-terminal region from the ribosome, thus liberating the mRNA entry site and reactivating translation (Shi *et al.*, 2020a) (Fig. 1.1.6). Tidu *et al.* proposed that the *cis-acting* stem-loop 1 (SL1) of 5'UTR executes this task. They further propose that the 5'UTR SL1 is released from the 40S-Nsp1 complex once the translation terminated thus allowing the Nsp1 C-terminal region to bind to the ribosome again and impede cellular translation. Nsp1 (regulated by SL1) should then act as a gatekeeper of the ribosomal mRNA channel, uniquely switching host gene translation to viral protein production (Tidu *et al.*, 2021) (Fig. 1.1.7). This mechanism is shadowed by the fact that neither free Nsp1 nor the ribosomal 40S subunit show any affinity to the 5'UTR RNA (Tidu *et al.*, 2021). Furthermore, it remains unclear how Nsp1 should remain bound to the ribosome once its interaction with the C-terminal region is released.



**Figure 1.1.6 Model mechanism for Nsp1-mediated translation inhibition and evasion by SARS-CoV-2 5' UTR.** This figure is adapted from (Shi *et al.*, 2020a).



**Figure 1.1.7 NSP1, as a gatekeeper to guarantee the viral evasion by associating with SARS-CoV-2 5'UTR.** Figure is adapted from (Tidu *et al.*, 2021).

Structure-function studies, especially those addressing the regulation of Nsp1 activity, are still limited due to the lack of structural information on intact Nsp1. Here, as part of this thesis, I report the nuclear magnetic resonance (NMR) structure of full-length Nsp1 from SARS-CoV-2. My data reveals that Nsp1 consists of an  $\alpha/\beta$  fold formed by a six-stranded, capped  $\beta$ -barrel motif comprising residues 10-127 and of intrinsically disordered short N-terminal and long C-terminal tails (residues 1–9 and 128–180, respectively). 3D NOESY experiments indicate that the intrinsically disordered C-terminal tail transiently interacts with the globular domain of Nsp1 via long-range interactions. In particular, the positively charged RNA-binding surface, positioned at the end of the  $\beta_6$  strand of the folded domain of Nsp1, is masked by the negatively charged C-terminal tail. In view of the cryo-EM structure, which demonstrated an interaction of the C-terminal tail with the mRNA entry tunnel, we postulate that the unfolded C-terminal tail regulates access to the RNA binding surface, by masking the surface to any potential binders until Nsp1 has reached the ribosome. Comparative analysis of SARS-CoV Nsp1 sequences unveils how evolution has reshaped the loops and the charged surface patches. A comparison of our NMR structure with the three available crystal structures of

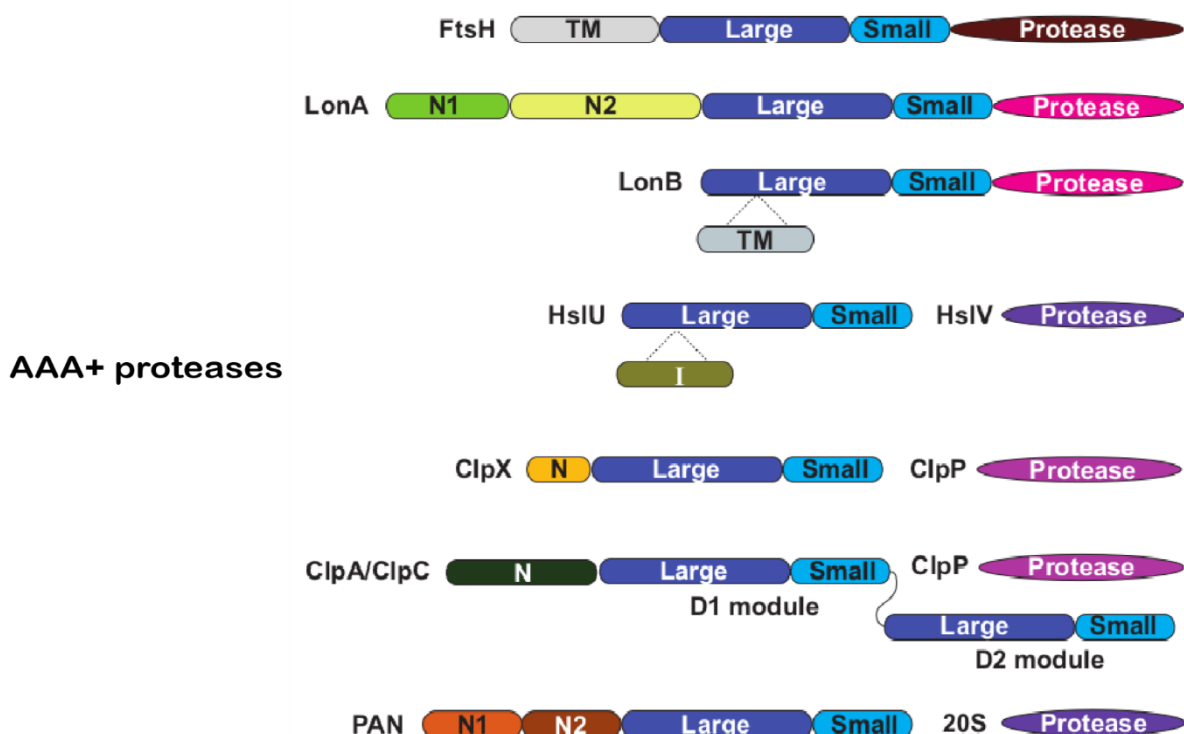


folded Nsp1 in SARS-CoV-2 demonstrates only minor differences among the structures. Our results provide the groundwork for future studies focusing on the structure-function relations of Nsp1, providing an understanding into the structural features of this attractive target for anti-COVID-19 therapeutics.

## 1.2 Substrate protein (ComK) recruitment by the hexameric MecA–ClpC complex

### 1.2.1 AAA+ protease complexes

Regulated proteolysis by members of the ATP-dependent AAA+ (ATPases associated with various cellular activities) protein family is a crucial component of the protein quality-control system that has evolved in all domains of life to maintain cellular homeostasis and to respond to external stimuli and environmental changes. The unifying function of the proteins of this family is to utilize the energy of ATP hydrolysis to remodel or unfold target proteins through specific loops in the pore formed by the AAA+ hexameric ring structures (Elsholz *et al.*, 2017). To date, several major families of AAA+ proteases have been found to associate with a large number of substrate proteins and participate in protein unfolding and degradation, including the Lon (LonA and LonB) and FtsH families, the PAN/20S protease family, the Hsp100/Clp (such as ClpXP, ClpAP, ClpCP) and HslUV (also known as ClpYQ) families (Ogura & Wilkinson, 2001; Sauer & Baker, 2011) (Fig. 1.2.1)



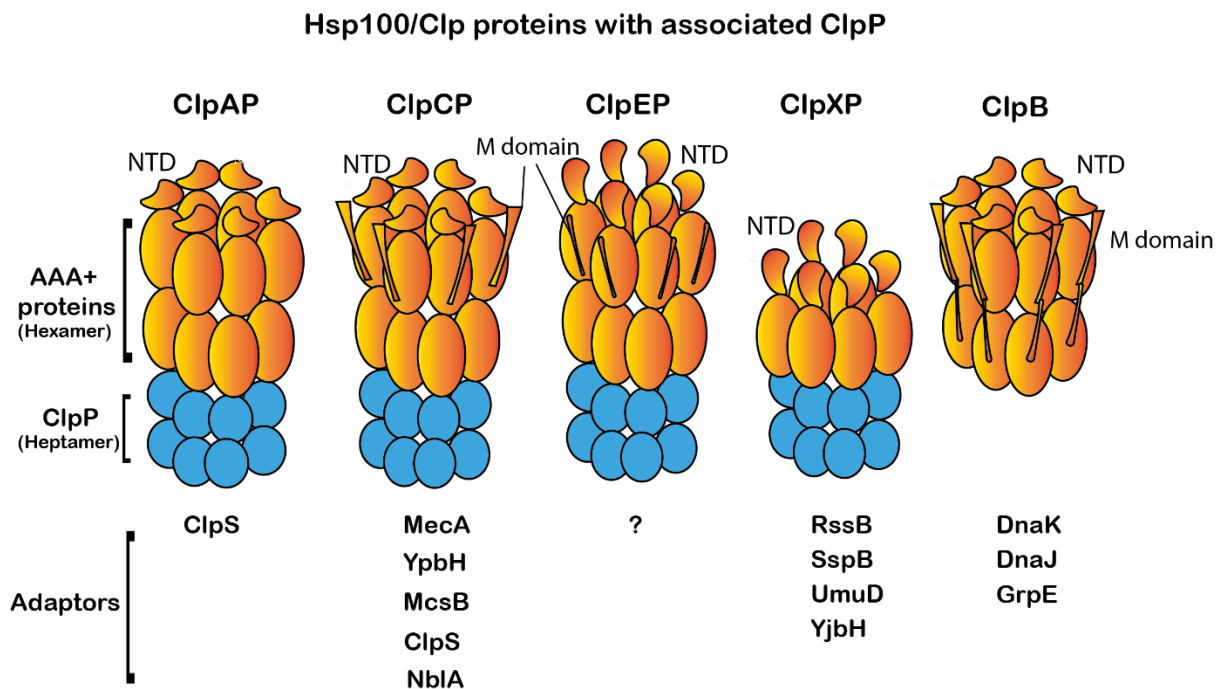
**Fig 1.2.1 Domain structures of AAA+ proteases.** Each protease contains one or two AAA+ subunits (in the figure indicated by D1 and D2 domains), a protease subunit and additional family-specific domains. The protease domain and AAA+ module are present either on the same polypeptide (FtsH, LonA, and LonB) or on different polypeptide chains (HslUV, ClpXP, ClpAP, ClpCP, and PAN/20S). This figure is adapted from (Sauer & Baker, 2011).

In bacteria, the cellular protein quality control system comprises the conserved Hsp100/Clp family of AAA+ protease complexes. Each of these complexes consists of an ATP-dependent, protein-unfolding unit, such as ClpA, ClpC, ClpE, ClpX and a compartmentalized protease unit, such as ClpP. ClpA, ClpC, ClpE and ClpX are active as homo-hexameric assemblies and recognize, unfold and translocate specific protein substrates to ClpP for degradation (Horwich *et al.*, 1999). However, molecular chaperones like ClpB and Hsp104 (the orthologue of ClpB in yeast), which also belong to Hsp100/Clp protein family, lack the sequence motif that recognizes the ClpP protease and can thus only function as protein remodellers, i.e., they use the energy of ATP binding and hydrolysis to rescue misfolded or aggregated proteins (Weibezahn *et al.*, 2004; Doyle & Wickner, 2009; Haslberger *et al.*, 2010; Mertens & Svergun, 2010). Furthermore, AAA+ unfoldases, such as ClpA, ClpC, ClpX can also function alone to disaggregate and remodel unfolded polypeptides, when they are not in complex with ClpP (Andersson *et al.*, 2006). For classification, ClpA, ClpB, ClpE and ClpC belong to class 1 of the Hsp100/Clp family, bearing two nucleotide-binding domains (NBD), known as D1 and D2, while ClpX belongs to class 2, containing only one nucleotide-binding domain (Schirmer *et al.*, 1996).

Hsp100/Clp protease complexes consist of six identical multidomain subunits to form a hexameric complex with a barrel-like structure, whereas the compartmentalized ClpP protease consists of a double-ring structure with 14 subunits (Fig. 1.2.2). Inside the hexameric complex, the diameter of the pore is controlled by multiple layers of axial loops. These loops, located at different positions of D1 and D2 domains, are essential for substrate binding, unfolding, and translocation to ClpP for degradation. In addition to nucleotide-binding domains, Hsp100/Clp proteins also possess loops or domains inserted at different positions of the AAA+ domain. For example, the C-terminal recognition loop (called VGF loop) is necessary for the interaction with ClpP, the middle domain (also referred to as linker domain) and the N-terminal domain (NTD) are often necessary for the interaction with adaptor proteins (Kirstein *et al.*, 2006).

Several adaptor proteins have been discovered for the proteins of the bacterial Hsp100/Clp family (Fig. 1.2.2). They modulate substrate specificity, proteolytic or chaperone activity of their respective partner AAA+ proteins and assist in recognition and subsequent degradation of specific substrate proteins. For example, the adaptor protein ClpS significantly affects the intrinsic substrate specificity of ClpAP by altering ClpA activity from degradation of SsrA-tagged proteins to degradation of defined aggregated or oligomeric protein substrates (Dougan *et al.*, 2002). Adaptor protein SspB targets and delivers ssrA-tagged substrates to ClpXP through the interaction of the ssrA tag with the N-domain of ClpX. Interestingly, ClpAP also degrades ssrA-tagged substrates but not when SspB is bound to them, implying that SspB hinders ClpA recognition of the substrates and at the same time facilitates recognition

by ClpX (Dougan *et al.*, 2003; Sauer & Baker, 2011). However, all these adaptor proteins are only regulators and not essential for the basic functions of their partner proteins, which are also fully active in the absence of their adaptors (Kirstein *et al.*, 2006). In strong contrast, all the functional activities of ClpC, including chaperone and degradation of general substrates, rely on the presence of adaptor proteins, such as MecA. In addition, MecA also appears to be critical for the development of competence in *Bacillus subtilis* by delivering the transcription factor ComK to ClpCP for degradation (Turgay *et al.*, 1998).

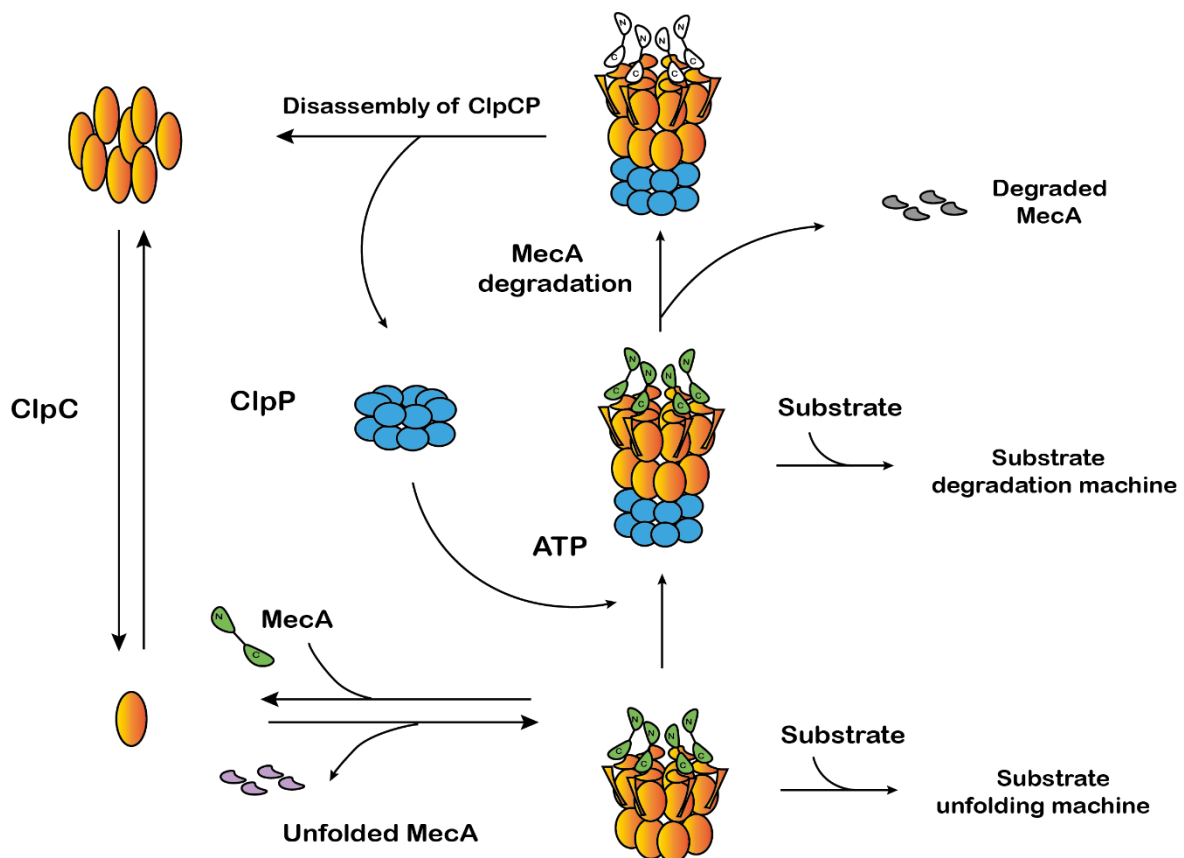


**Fig 1.2.2 Composition and structural architecture of Hsp100/Clp protease complexes.** The identified adaptor and/or regulator proteins are indicated under the respective protease complex. The AAA+ and associated domains are shown in orange, and the protease is shown in blue. This figure is modified from (Kirstein *et al.*, 2006).

### 1.2.2 Hexameric MecA–ClpCP complex

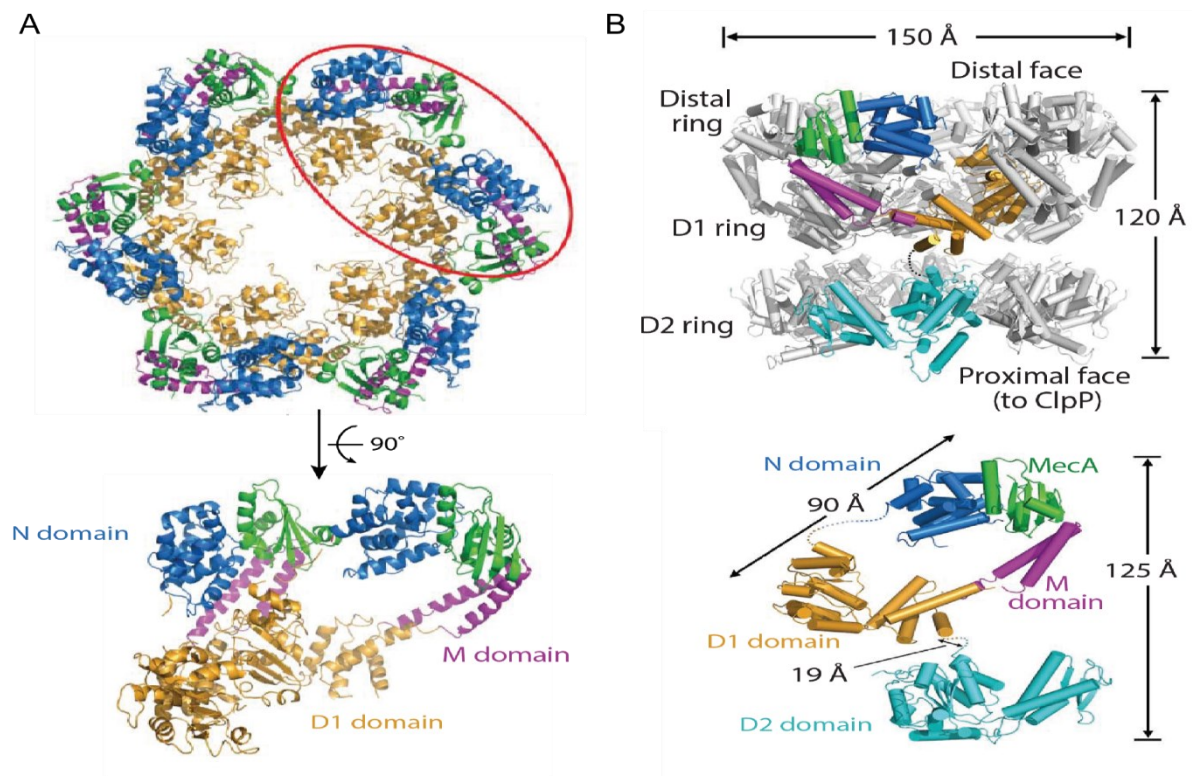
MecA is a ClpC-specific adaptor protein that is necessary for the recognition and targeting of substrate proteins as well as for the formation of the functional ClpC hexamer and assembly of the ClpCP degradation machinery. MecA itself consists of two functional domains. The N-terminal domain (NTD, residues 1-77) is required for associating with substrate proteins, whereas the C-terminal domain (CTD, residues 121-218) is critical for the interaction with ClpC and the induction of the activity of both the ClpC hexamer and the ClpCP protease complex. The NTD and CTD of MecA are separated by a flexible linker (residues 78-120), which may play an important role in substrate delivery (Kirstein *et al.*, 2006).

Assembly of the ClpCP protease is strictly determined by the presence of MecA, and the assembled MecA–ClpCP protease complex is highly dynamic. ATP binding to ClpC and hydrolysis are needed for unfolding and translocating substrate proteins. In the presence of ATP, MecA binds to ClpC to facilitate the formation of the functional MecA–ClpC hexamer, which acts as an unfoldase. When the hexameric MecA–ClpC complex interacts with ClpP, the MecA–ClpCP complex acts as a substrate degradation machine. MecA recognizes and delivers specific substrate proteins to the ClpC hexamer; interestingly, in the absence of substrates, MecA itself can be degraded by ClpCP, which leads to the disassembly of ClpCP. This regulatory mechanism impedes the activity of the ClpCP degradation machine when substrate proteins are not available (Kirstein *et al.*, 2006; Mei *et al.*, 2009; Elsholz *et al.*, 2017) (Fig. 1.2.3).



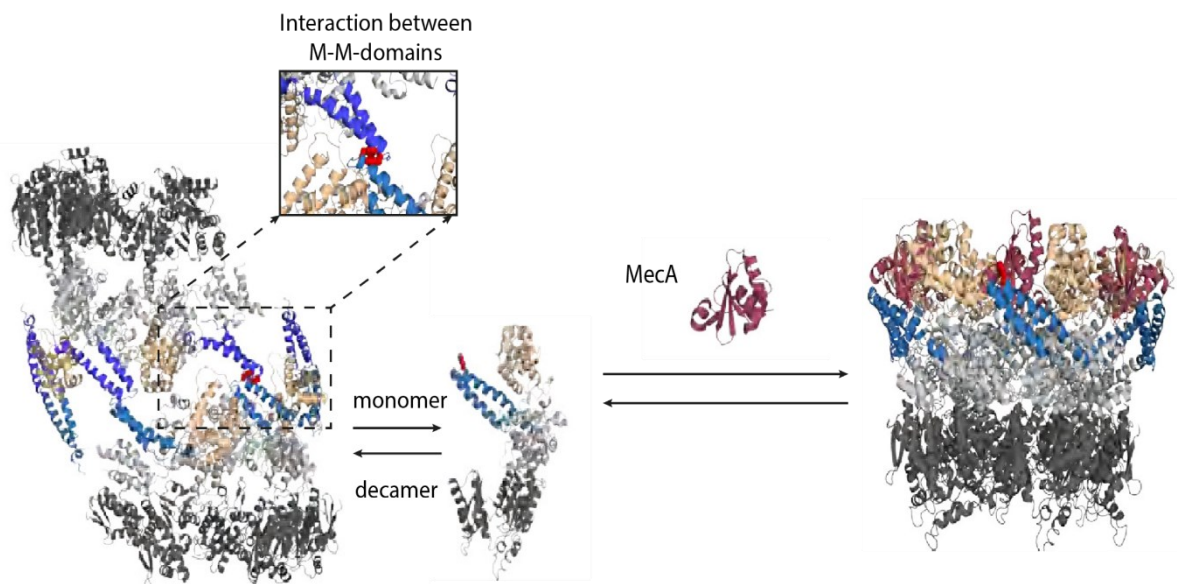
**Fig 1.2.3 Model of the MecA–ClpCP degradation system.** Assembly of the ClpCP protease depends on the presence of the adaptor protein MecA, which targets and translocates specific substrate proteins to ClpCP for degradation. In the absence of substrates, MecA itself is degraded by the ClpCP protease. This in turn results in the disassembly of the ClpCP complex. The exact molar ratio between MecA and ClpC, although shown here as 1:1, remains to be determined. This figure is modified from (Mei *et al.*, 2009).

To understand how MecA facilitates the assembly and activation of the ClpC complex, the Shi group determined two crystal structures of constructs from the MecA–ClpC complex (Wang *et al.*, 2011). The first construct comprised the CTD of MecA (residues 121–218) and a truncated ClpC lacking the D2 domain and two flexible loops (ClpC- $\Delta$ D2; residues 1–485, Glu280 to Ala mutant,  $\Delta$ 247–251 and  $\Delta$ 281–292). The mutation E280A in the walker B motif of the D1 domain does not affect ATP binding but abolishes the activity of ATP hydrolysis, thus stabilizing the complex. In this structure, six molecules of MecA CTD and six molecules of ClpC- $\Delta$ D2 form a barrel-shaped complex, featuring two stacked rings. The wide ring of the barrel is composed of six MecA CTD and six ClpC N-terminal domains, whereas the narrow ring consists of six ClpC D1 domains. Six M domains, connecting the N-terminal and the D1 domains of ClpC, are positioned at the exterior of the barrel (Fig. 1.2.4A). The six MecA CTD–ClpC N-domain heterodimers associate with each other through polar contacts, while the M domain binds to MecA CTD through a mixture of van der Waals contacts and polar interactions. The second construct comprised MecA108 (MecA residues 108–218) and a full-length ClpC protein with Glu to Ala mutations in the walker B Motifs of both D1 and D2 without four flexible internal loops (E280A, E618A,  $\Delta$ 247–251,  $\Delta$ 281–292,  $\Delta$ 587–596 and  $\Delta$ 665–685). This complex has dimensions of 125 Å (height) and 150 Å (diameter) and contains three rings. The overall size of the D2 ring, which is responsible for binding ClpP, is greater than that of the D1 ring, and the six D2 domains were found to be asymmetric. Each ClpC protomer appears to be extended and with discrete N-terminal, D1 and D2 domains (Fig. 1.2.4B).



**Fig 1.2.4 Crystal structures of the MecA–ClpC hexameric complex.** (A) Cartoon representation of hexameric MecA121–ClpC- $\Delta$ D2 complex (PDB:3PXG). the M domains connect the two stacked rings. MecA121 is colored in green, while the ClpC N-terminal, M and D1 domains are represented in blue, magenta and orange, respectively. (B) Overall structure of the hexameric complex between MecA108 and full-length ClpC (PDB: 3PXI). One ClpC molecule bound to one MecA108 is colour-coded; all other molecules are depicted in grey. The ClpC protomer bound to MecA108 exhibits an extended architecture and discrete functional domains (N-terminal, D1 and D2). The D2 domain of ClpC is shown in cyan. Figures were reproduced from (Wang *et al.*, 2011).

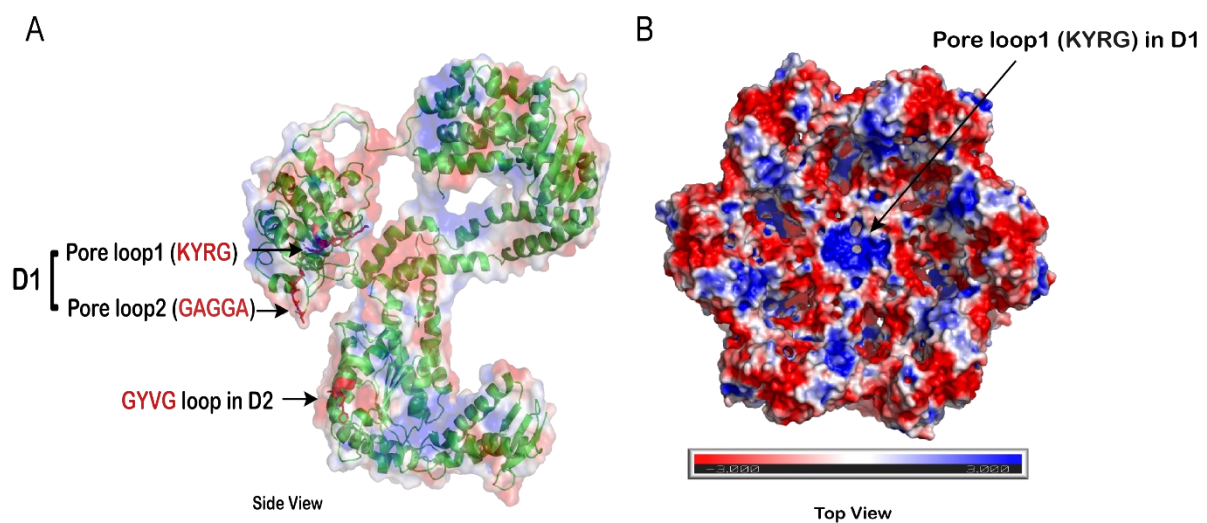
Structural characterization of full-length ClpC-DWB (ClpC with mutations in both walker B motifs) by cryo-EM unveiled the mechanism of activity control of ClpC (Carroni *et al.*, 2017). In the absence of MecA, the activity of ClpC is repressed by formation of an inactive resting state consisting of a decamer with helical arrangement of the monomers, where the coiled-coil M domains are associated head-to-head (Fig. 1.2.5). In this inhibited state the binding of adaptor molecules, substrates and ClpP is hindered, and the efficiency of ATP hydrolysis is low. Binding of MecA to the N-terminal and M domains of ClpC breaks this resting state and generates an active hexameric complex. Thus, the ClpC M domain serves as a molecular switch, ensuring the suppression of ClpC activity in the resting state and promoting hexameric ClpC activation when MecA is present. Since the MecA binding sites in the ClpC N-terminal and M domains are not accessible in the inactive state, occasional disassembly of ClpC into monomers is a prior condition for MecA binding (Fig. 1.2.5).



**Fig 1.2.5 Dual function of ClpC M domains in both the resting and active states.** The ClpC inactive state consists of two layers of ClpC molecules arranged in a helical configuration. Pairs of protomers share contacts of their middle M domains in a head-to-head arrangement. The activity of hexameric ClpC is stimulated by MecA through binding to the N-terminal and M domain of ClpC, an event which

breaks the interaction between the ClpC M domains and lifts the repression of ClpC activity. This figure is adapted from (Carroni *et al.*, 2017).

Further, the structure of the full-length MecA–ClpC complex by cryo-EM provided evidence for the presence of an allosteric conformational coupling of the two nucleotide-binding domains (D1 and D2) during a functional cycle, with the information being transferred by the D1–D2 interface. In this work, it was shown that the substrate protein, translocated through the central pore, successively contacts the axial loops [pore loop1 (KYRG) and pore loop2 (GAGGA) in D1, and GYVG loop in D2] (Wang *et al.*, 2011; Liu *et al.*, 2013) (Fig. 1.2.6A and B). The pore loops within D1 and D2 rings likely drag the substrate proteins in a spiral trajectory, which facilitates the efficiency of substrates unfolding and translocation during cycles of ATP hydrolysis. Moreover, nucleotide binding and hydrolysis also regulate the compactness of D2 ring and the spacing between the six ClpP-interacting VGF loops, suggesting that the nucleotide state may also affect the interaction of MecA–ClpC complex with ClpP. Mutational experiments indicated that increasing or decreasing the linker length at the D1-D2 interface of ClpC remarkably impairs the ATPase and unfolding activities of the hexameric complex, further validating the contribution of inter-ring communication to the overall activity of MecA–ClpC complex (Liu *et al.*, 2013).



**Figure 1.2.6 Electrostatic surface representation of MecA CTD–ClpC–DWB complex (PDB: 3J3S).**

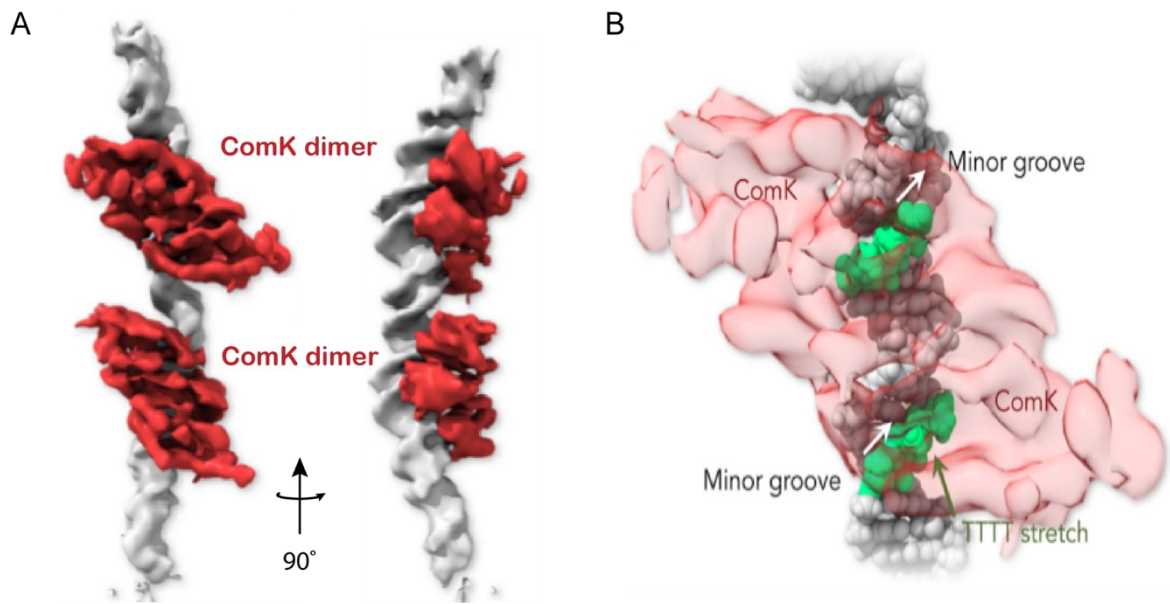
(A) Surface representation of the protomer of that MecA CTD–ClpC–DWB complex. Pore loops located in D1 and D2 domains are indicated in red. (B) Top surface view of the hexameric complex. The surface of pore loop1 is highly positively charged. The figure is reproduced from (Liu *et al.*, 2013) (PDB 3J3S).



### 1.2.3 The substrate ComK

In *Bacillus subtilis*, the development of natural competence is a highly regulated global response to drastic environmental changes, like heat or oxidative stress. Development of cellular competence necessitates the expression of genes that are subject to a complicated competence-signal-transduction network involving sensing or processing diverse competence-inducing signals. Numerous proteins encoded by competence genes exist uniquely at the onset of late exponential growth under harsh nutritional conditions. They are necessary for recognizing and transporting exogenous DNAs into the cell, and exhibit properties as growth inhibition, which are characteristics of persister-like cell states that can provide a survival advantage for bacteria upon various environmental stresses. The transcription factor ComK is essential for the transcription of competence genes and is expressed only in the cells that are destined to become competent. ComK acts positively both on its own transcription as well as on the transcription of the late competence genes encoding the DNA uptake machinery and of genes needed for homologous recombination and DNA repair (comC, comE, comF and comG) (van Sinderen *et al.*, 1995; Haijema *et al.*, 1996a; Haijema *et al.*, 1996b; Dubnau, 1997; Hamoen *et al.*, 1998; Turgay *et al.*, 1998). Moreover, ComK is negatively regulated by MecA (MecB) or ClpC in *Bacillus subtilis*. MecA normally binds to ComK for protein degradation; overproduction of MecA results in the downregulation of ComK synthesis (Dubnau & Roggiani, 1990; Roggiani *et al.*, 1990; Hahn *et al.*, 1994).

ComK activates transcription of competence genes by binding to the minor groove of specific A/T-rich DNA sequences, so called K-boxes (one A/T-rich motif is one box), which are positioned upstream of ComK-dependent genes. Previous *in vitro* experiments demonstrated that ComK acts as a tetramer with two dimers bound to each K-box of its own promoter. Recently, 3D density maps obtained by cryo-EM further indicated that the two ComK molecules in the two boxes of each DNA promoter face the same side, and protein density corresponding to two ComK molecules is located in each box (Hamoen *et al.*, 1998; Susanna *et al.*, 2006; Rosenblum *et al.*, 2021) (Fig. 1.2.7).

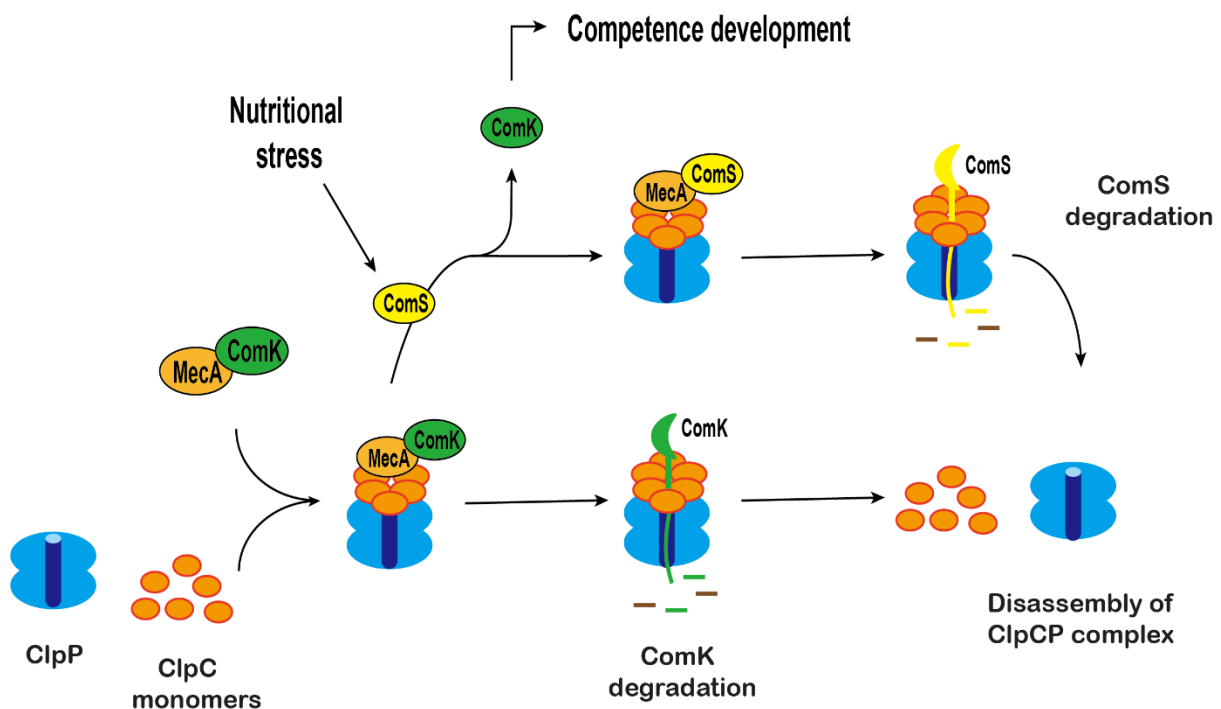


**Fig 1.2.7 Structure of ComK-DNA complex obtained by cryo-EM.** (A) 3D density map of ComK (red) bound to DNA promoter (grey) with an 8 bp spacer. (B) Zoom view of ComK dimer (transparent red) relative to a single box of the promoter DNA (8 bp spacer). ComK faces the thymine bases (green) in the minor groove. This figure is adapted from (Rosenblum *et al.*, 2021).

ComK is involved in both DNA–protein and protein–protein interactions. Furthermore, ComK is a specific substrate protein for the MecA–ClpC complex [When ComK is sequestered by MecA binding, the synthesis of ComK is suspended (Turgay *et al.*, 1998)]. When ComK is sequestered during exponential cell growth state by forming a ternary complex with MecA and ClpC, ComK is incapable of activating the transcription of competence genes. Strains carrying mutations in either MecA or ClpC generate substantially high levels of active ComK and develop competence under conditions in which competence is normally suppressed. This regulation is crucial for the cells as improper activation of ComK and competence development can inhibit cell growth (Lazazzera & Grossman, 1997; Turgay *et al.*, 1998).

Higher cell density in post-exponential or stationary cell growth state results in the synthesis of a small anti-adaptor, so called ComS, which disrupts the ComK–MecA–ClpC ternary complex and restores the ability of ComK as a transcription factor. ComS is a 46 amino-acid polypeptide which is essential for competence development as well. As genetic competence is regulated by cell density signals, production of ComS is stimulated by two quorum-sensing pheromones (CSF and ComX) that accumulate in culture medium as cells grow to high density. These pheromones are also necessary for the development of competence and can induce the development of competence when added to cells. Through ComS, CSF and ComX pheromones serve as signals for high cell density and trigger the activation of ComK (Magnuson *et al.*, 1994; Solomon *et al.*, 1996; Lazazzera & Grossman, 1997).

The synthesis of ComS also interrupts the degradation of ComK by the MecA–ClpC complex. Similarities in the primary sequences of ComS and ComK allow ComS to compete with ComK for MecA binding. Thus, ComS is able to disassociate the ComK–MecA–ClpCP complex and compete with ComK for degradation. When ComK or ComS are not available, MecA is degraded by ClpCP, which eventually induces the disassembly of MecA–ClpCP complex. Thus, ClpC, MecA and ComK, together with anti-adaptor ComS, form a regulatory model which controls bacterial competence development through a regulated degradation of transcription factor ComK (Turgay *et al.*, 1998; Battesti & Gottesman, 2013; Elsholz *et al.*, 2017) (Fig. 1.2.8).



**Fig 1.2.8 Regulated proteolysis of transcription factor ComK by MecA–ClpCP protease complex.** ClpC oligomerization and interaction with ClpP depends on the presence of its adaptor protein MecA, which targets transcription factor ComK for ClpCP degradation under normal cell growth conditions. In the case of nutritional stress, anti-adaptor protein ComS, promotes the development of competence by releasing ComK from the MecA–ClpCP complex. MecA degradation leads to ClpCP disassembly. This figure is adapted from (Battesti & Gottesman, 2013).

#### 1.2.4 The questions we ask and what we find

Despite several structure-function based studies of Hsp100/Clp proteins, our understanding of the mechanistic cycle of the hexameric Hsp100/Clp proteins is far from complete. Until now, little is known about the mechanisms that ensure substrate specificity and the recruitment of substrate proteins to the MecA–ClpC complex. Due to its strong tendency to aggregation and

its low stability, no structural information is available for ComK, which increases the difficulty of investigating the mechanism of ComK recruitment by the MecA–ClpC complex. Here, as part of this thesis, we report the first *in-vitro* assembly of the ComK–MecA–ClpC ternary complex for both biochemical and structural analyses. Using a combination of NMR spectroscopy and biochemical methods, we find that homogenous binding of ComK to MecA depends on the presence of ClpC. Binding of ClpC to MecA CTD may induce a conformational change that enables the MecA NTD to target ComK with improved specificity. Furthermore, we have undertaken investigations of both the MecA–ClpC and ComK–MecA–ClpC complexes by solution-state NMR spectroscopy, using methyl-TROSY spectra of MecA in the context of these complexes. We find that the MecA NTD shows a reduced mobility when in complex with ClpC, indicating that this domain is confined upon complex formation, possibly due to its electrostatic interaction with the positive charged central pore loop inside ClpC hexamer. Moreover, NMR titration experiments of ComK onto the MecA–ClpC complex demonstrate that MecA recruits a ComK dimer to ClpC. Combining the NMR data with SEC-MALS and SAXS measurements, we conclude that addition of ComK stabilizes the MecA–ClpC complex yielding a homogeneous ternary complex, involving a ClpC hexamer, four MecA and two ComK molecules in the presence of ATP. Therefore, we postulate that four instead of six MecA molecules are sufficient to accomplish substrate delivery of ComK to ClpC or ClpCP. Our results represent a starting point for further NMR-based structural and functional studies of this system to understand the mechanism by which MecA targets ComK for degradation.

## 1.3 Methodological Background

### 1.3.1 Nuclear Magnetic Resonance (NMR) spectroscopy

#### 1.3.1.1 Genesis of the NMR signal

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique to determine biomolecular structure and dynamics at the atomic level. Biomolecular NMR in the solution-state was initially limited to proteins or nucleic acids up to ~10 kDa in size, but is now applicable to much larger macromolecules by virtue of developments in instrumental hardware, pulse sequences and isotopic labeling schemes, together with the advent of transverse-relaxation-optimized spectroscopy (TROSY) for the study of large complexes (Pervushin *et al.*, 1997).

The fundamental phenomenon behind NMR spectroscopy is the existence of nuclear spin angular momentum and nuclear magnetism. Many atomic nuclei possess intrinsic angular momentum, known as *spin*, which is a quantum-mechanical property with no classical analogue. The nuclear spin angular momentum,  $I$ , is a vector quantity; its magnitude is characterized by the nuclear spin quantum number  $I$ :

$$|I| = \sqrt{\hbar[I(I + 1)]} \quad (1.1)$$

$\hbar$  equals to the Planck's constant  $h$  divided by  $2\pi$ . The nuclear spin quantum number  $I$  of a nucleus may have either positive integer ( $I = 1, 2, 3, \dots$ ) or positive half-integer values ( $I = 1/2, 3/2, 5/2, \dots$ ), or can also be zero ( $I = 0$ ). Quantum numbers greater than 4 are rather rare. Different isotopes of the same element may have different nuclear spin quantum numbers and the most abundant isotopes of some common elements, such as  $^{12}\text{C}$  or  $^{16}\text{O}$ , have  $I = 0$ , and hence have no angular momentum or magnetic moment and cannot be detected by NMR. Only nuclei with non-zero spin quantum numbers are detectable by NMR, and among them those with  $I = 1/2$  ( $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$  &  $^{31}\text{P}$ ) are particularly useful for biomolecular solution NMR.

The component of the vectorial spin angular momentum along any particular axis is said to be quantized; for a spin- $I$  nucleus, there are  $2I+1$  projections onto the chosen axis (in NMR, the axis of quantization is defined by the direction of the static magnetic field,  $B_0$ ). By convention, this axis is labelled as the z-axis in the Cartesian frame, so the quantized component of the spin angular momentum is denoted as  $I_z$  and is given by:

$$I_z = m\hbar \quad (1.2)$$

where  $m$  is the magnetic quantum number and has values in integral steps between  $+I$  and  $-I$  ( $I, I - 1, I - 2, \dots, -I + 1, -I$ ). Therefore, the z-component of the angular momentum of a

spin- $\frac{1}{2}$  system can take two values, with  $m = \pm\frac{1}{2}$ . Thus,  $I_z = \pm\frac{1}{2} \hbar$ . For a nucleus with  $I = 1$ , the z-component of the angular momentum of has three possible values:  $I_z = \pm \hbar$  or  $I_z = 0$ . In the absence of external magnetic field, the axis of quantization is arbitrary and the spin angular momentum has no energetically preferred directions; in other words, the different spin states of a nucleus — characterized by the values of the magnetic quantum number — have the same energy.

The magnetic moment of nuclei is closely linked to their nuclear spin angular momenta, and nuclei with no spin angular momentum are non-magnetic. The nuclear magnetic moment  $\mu$  (also a vector quantity) is given by:

$$\mathbf{M} = \gamma \mathbf{I} \quad (1.3)$$

in which  $\gamma$  is the ratio of the magnetic moment to the nuclear spin angular momentum (known as the gyromagnetic or magnetogyric ratio). Consequently, the z component of the magnetic momentum is given by:

$$\mu_z = \gamma I_z = \gamma m \hbar \quad (1.4)$$

When an external static magnetic field  $B_0$  is applied, the magnetic interaction between the nuclear magnetic moment and the external field means that magnetic energy associated with the nucleus becomes dependent on the orientation of the nuclear magnetic moment relative to the external magnetic field. If the z-axis is taken as the axis of the external magnetic field, then the magnetic energy of the nucleus depends on the component of the nuclear magnetic moment along the z-axis, and is given by:

$$E = -\mu_z B_0 \quad (1.5)$$

From equations of (1.4) and (1.5), this becomes:

$$E = -\mu_z B_0 = -\gamma I_z B_0 = -m \hbar \gamma B_0 \quad (1.6)$$

When the nuclei are within the external magnetic field, they can be considered to be distributed amongst the  $2I+1$  available energy levels. The  $2I+1$  states for a spin- $I$  nucleus are equally spaced, with an energy separation between adjacent energy levels of  $\hbar \gamma B_0$ .

The two energy levels of a spin- $\frac{1}{2}$  nucleus correspond to the nuclear magnetic moment being aligned parallel and anti-parallel to the external magnetic field, the parallel alignment is the lower in energy. By convention, the nuclear spin-state corresponding to the lower energy level is labelled as the  $\alpha$  state, and that corresponding to the higher energy level as the  $\beta$  state. At thermal equilibrium, the nuclei are distributed between the two states according to the Boltzmann distribution, which means that there are more nuclei in the  $\alpha$  state than the  $\beta$  state,

and hence more nuclei with magnetic moments parallel to field than nuclei with magnetic moments anti-parallel to the field. This translates into a non-zero macroscopic magnetic moment for the sample as a whole, which is usually referred to as the bulk magnetization. This bulk magnetization is a vector quantity, denoted  $\mathbf{M}$ , having both magnitude and direction; at equilibrium it is perfectly aligned with the external field, having zero component in the transverse ( $xy$ ) plane.

Application of electromagnetic radiation (time-dependent magnetic field) whose frequency corresponds to the energy separation of the energy levels is able to stimulate transitions. The selection rule for NMR is  $\Delta m = \pm 1$ , so that the energy separation of two energy levels between which transitions can occur is given by:

$$\Delta E = \hbar\gamma B_0 \quad (1.7)$$

If the time-dependent magnetic field is applied to the equilibrium state, the effect is to rotate the bulk magnetization vector  $\mathbf{M}$  away from the  $z$ -axis; the rotation axis corresponds to the direction of the time-dependent field, which is always in the transverse ( $xy$ ) plane. This motion of the bulk magnetization vector about the time-dependent field is called *nutation* and is the means of generating transverse magnetization. The typical magnitude of nuclear gyromagnetic ratios and the strengths of magnets suitable for NMR are such that electromagnetic radiation required to stimulate transitions between the nuclear spin states (and equivalently rotate the bulk magnetization into the transverse plane) usually falls in the radio-frequency (RF) region of the electromagnetic spectrum. The short bursts of this radiation applied during an NMR experiment are therefore referred to as RF pulses.

The time-dependent magnetic field is turned off (ending the RF pulse) once the bulk magnetization vector has been rotated to the transverse plane. The bulk magnetization vector then rotates instead about the axis of the static magnetic field (i.e. around the  $z$ -axis); this rotation is called *precession*. The precessional frequency, also known as the *Larmor frequency*  $\omega$ , is given by:

$$\omega = \gamma B_0 \text{ in rad S}^{-1} \quad (1.8)$$

Or 
$$\nu = \gamma B_0 / 2\pi \text{ in Hz} \quad (1.9)$$

The Larmor frequency of a nucleus is directly related to its gyromagnetic ratio  $\gamma$  and the strength of the applied magnetic field  $B_0$ . The precession of the magnetization in the transverse plane generates a current in a detection coil oriented in the transverse plane; this electrical response is the detected signal for all pulsed NMR experiments, and is called the *free induction decay* (FID). The FID corresponds to the time-domain NMR signal; the NMR spectrum is generated by Fourier transformation of the FID.

For a given external magnetic field, the Larmor frequency for a particular isotope is not fixed at a single frequency, but actually depends sensitively on the molecular structure; the Larmor frequencies for any isotope span a small range. This is due to the fact that the effective magnetic field at any nucleus is not exactly equal to the applied external magnetic field but is modulated by the surrounding electronic environment. Due to differences in the local electron distributions, individual nuclei inside a molecule will experience different effective magnetic fields and consequently precess at different frequencies. This effect is called the *chemical shift*, and allows different types of the same isotope in different chemical environments inside a molecule to be distinguished.

The effective magnetic field  $B$  experienced by a nucleus in an atom or molecule is usually slightly smaller than the external magnetic field  $B_0$ , which induces the electrons to circulate within their atomic orbitals, generating a local magnetic field  $B'$  that points in the opposite direction to  $B_0$ . A nucleus in a molecule is said to be shielded from the external field  $B_0$  by its surrounding electrons; the effective magnetic field is given by:

$$B = B_0 - B' = B_0(1 - \sigma) \quad (1.10)$$

Here,  $\sigma$  is the constant of proportionality between  $B_0$  and  $B'$  and is called the *shielding constant*.

The NMR frequency of a specific nucleus in a molecule becomes:

$$\nu = \gamma B_0(1 - \sigma) / 2\pi \quad (1.11)$$

This means that the absolute value of the nucleus frequency depends not only on the external field  $B_0$ , but also on the shielding constant and hence its local electronic environment.

Since the absolute chemical shifts are rarely needed, it is conventional to define a chemical-shift scale that is based on the difference in frequencies between the nucleus of interest  $\nu$  and reference signal  $\nu_{ref}$  by means of a dimensionless parameter  $\delta$  (typically expressed in parts per million, ppm):

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} \times 10^6 \quad (1.12)$$

$\delta$  is an intrinsic molecular property and independent of the magnetic field.

Conventionally, NMR spectra are plotted with  $\delta$  increasing from right to left. The more shielded nuclei have smaller absolute Larmor frequencies and appear on the right-hand side of the spectrum, with the less shielded nuclei having larger absolute Larmor frequencies and appearing towards the left-hand side of the spectrum.



### 1.3.1.2 Spin relaxation

Relaxation is the process by which the bulk magnetization returns to its equilibrium state after some perturbation. In NMR, application of an RF pulse to equilibrium magnetization will produce deviations in the spin-state populations. The bulk magnetization vector is rotated towards the transverse plane, which reduces its z-component and generates a component in the transverse plane, which is then allowed to precess about the external magnetic field, generating the signal that is detected in the form of the FID. As soon as the RF field is switched off, the system evolves back to the equilibrium state through a process called nuclear spin relaxation. There are two fundamental types of relaxation: *longitudinal relaxation* (*spin-lattice relaxation* or  $T_1$  relaxation) and *transverse relaxation* (*spin-spin relaxation* or  $T_2$  relaxation).

The process by which the z-magnetization returns to its thermal equilibrium value is called *longitudinal relaxation*. It is characterized by the recovery of the thermal equilibrium spin populations, where the population of the lower-energy  $\alpha$  state is greater than that of higher-energy  $\beta$  state. The term of spin-lattice relaxation arises from the understanding of the flow of energy between the spins and molecular motion. This relaxation process is described by the following equation:

$$M_z = M_0 \cdot (1 - e^{-\frac{t}{T_1}}) \quad (1.13)$$

where  $M_0$  is the magnitude of the magnetization vector along the z-axis at equilibrium, and  $M_z$  represents the value of the z-component of the magnetization at time t. Occasionally, instead of the relaxation time-constant  $T_1$ , the equation is expressed in terms of the respective spin-lattice relaxation rate constant,  $R_1$  ( $R_1 = 1/T_1$ ). It is clear from this equation that the  $T_1$  relaxation process results in the longitudinal magnetization follows an exponential trajectory in returning to equilibrium.

The process by which transverse magnetization returns to its equilibrium value is called *transverse relaxation*. At thermal equilibrium, there is no transverse magnetization, so transverse relaxation always corresponds to non-zero transverse magnetization decaying to zero. In contrast to the  $T_1$  relaxation which requires spin-spin transitions in order to restore the equilibrium population distribution,  $T_2$  relaxation does not necessarily involve transitions, and is the result of the loss of transverse phase coherence amongst the individual magnetic moments that constitute the bulk magnetization; this loss of phase coherence leads to a decay of the transverse component of the bulk magnetization. As for  $T_1$  relaxation, transverse relaxation also results in the magnetization following an exponential trajectory, here with the characteristic relaxation time-constant  $T_2$ :

$$M_{xy} = M_0 \cdot e^{-\frac{t}{T_2}} \quad (1.14)$$

Similarly, instead of the relaxation time-constant  $T_2$ , the above equation can be instead be written in terms of the respective relaxation rate constant  $R_2 = 1/T_2$ .

The effects of  $T_1$  and  $T_2$  relaxation have different implications for NMR experiments. Since  $T_1$  determines the time taken for the bulk magnetization to be regenerated along the z-axis — which is the implied starting-point for most NMR experiments — slower  $T_1$ -relaxation typically results in longer overall experiment times, as more time is spent in between consecutive scans waiting for the magnetization to reappear along the z-axis (a *scan* is a single repetition of an experiment; most experiments are acquired with multiple scans, with the detected signals from all the scans being co-added to increase the signal-to-noise in the final spectrum). Conversely, faster  $T_2$  relaxation leads to broad linewidths of NMR signals and thus decreased sensitivity and poor spectral resolution.

Both types of relaxation are driven by the fluctuations of local magnetic fields around the nuclear spins due to the thermal motion of the molecules. Two types of local magnetic fields tend to be the dominant sources of relaxation for spin- $1/2$  systems: dipole-dipole (DD) interactions and chemical-shift anisotropy (CSA).

The dipole-dipole mechanism is due to the mutual interaction between the magnetic moment of one nuclear spin and the magnetic fields associated with the magnetic moments of neighboring spins. The instantaneous strength of the dipole-dipole interaction depends on several factors. First, the distance,  $r$ , between the two spins; the interaction is relatively short-range, with a distance-dependence of  $1/r^3$ . Second, the orientation of the inter-nuclear vector with respect to the applied magnetic field  $B_0$ . Third, the gyromagnetic ratio  $\gamma$  of the spins. As described earlier, the strength (energy) of the interaction of a magnetic moment with a surrounding magnetic field is proportional to both the magnitude of the magnetic moment and the strength of the field. In the context of the dipole-dipole relaxation mechanism, the field in question is that due to another nuclear magnetic moment, and therefore the strength of the dipole-dipole interaction depends on the gyromagnetic ratios of both nuclei. Because of molecular tumbling — rotational diffusion due to Brownian motion in solution — the orientation of the inter-spin vector with respect to the external magnetic field  $B_0$  is continually changing over time, which leads to a time-dependent fluctuation of the local magnetic fields at each of the nuclei. As the fluctuation frequency of the local magnetic fields becomes slower, the  $T_2$  relaxation caused by dipole-dipole interaction is more efficient.

The chemical shift anisotropy (CSA) is the anisotropy of the local magnetic field at a given nucleus that arises due to the effect of the external applied field on the anisotropic electron

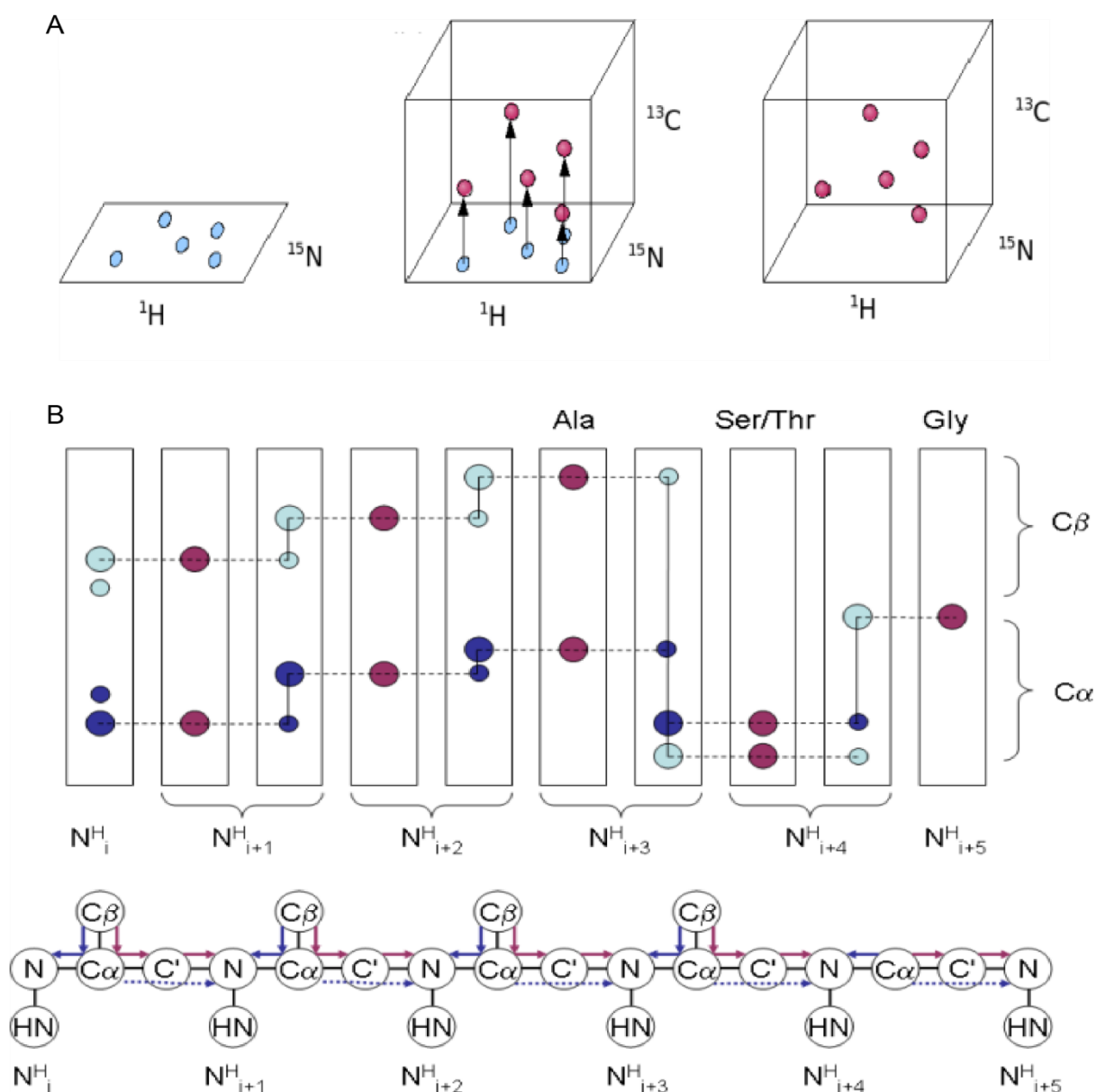
density surrounding the nucleus. In general, the local magnetic field will have components along all three Cartesian axes, and crucially (in most cases) depends on the orientation of the molecule relative to the applied field. It follows that the chemical shift is a tensorial interaction, and the associated tensor is usually anisotropic. It is this anisotropy that can represent a relaxation mechanism, whereby the continually changing orientation of the local electron density relative to the external field leads to a continually changing local field due to that electron density. Importantly — and unlike the dipole-dipole mechanism — the magnitude of the local fields associated with the CSA depends on the strength of the external magnetic field, and the corresponding contribution to the transverse relaxation rate is then dependent on the square of the external magnetic field strength.

### 1.3.1.3 NMR assignment of protein resonances

Assignment of the resonances — the NMR signals — of the proton, nitrogen and carbon nuclei that exist in proteins is an essential step for any structural or dynamical investigation by solution NMR. In the beginning of any protein resonance assignment, the first spectrum acquired is the two dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC (heteronuclear single quantum coherence) spectrum, which shows peaks for all the N–H pairs from the backbone amide groups within a protein. The exact positions of the peaks — dictated by the chemical-shifts of the  $^1\text{H}$  and  $^{15}\text{N}$  resonances of the NH groups — are highly sensitive to the 3D structure of the protein, and hence the complete  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum is often thought of as being the “fingerprint” of the protein. The  $^{15}\text{N}$ -HSQC pulse-sequence exploit the INEPT (“insensitive nuclei enhanced by polarization transfer”) element to transfer magnetization from  $^1\text{H}$  to  $^{15}\text{N}$  and back again. Peaks arising from N–H pairs in the side-chains of some amino acids, such as the amide groups of asparagine and glutamine, and the indole group of tryptophan, can also be observed in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra, though they are usually easy distinguished from the peaks of the backbone amides. However, prolines are not visible as they do not have an amide proton.

The assignment of the backbone amide resonances is performed using a series of experiments that allow for the connection of sequential backbone amides. These standard triple-resonance backbone assignment spectra record carbon chemical shifts on a third dimension in addition to the two dimensions of the  $^{15}\text{N}$ -HSQC (Fig. 1.3.1A). For example, the HNCACB spectrum contains peaks that correlate the resonances of the backbone NH group of each residue with the  $\text{C}^\alpha$  and  $\text{C}^\beta$  resonances of the same residue (“intra-type” peaks, strong intensity) and with those of the preceding residue (“inter-type” peaks, weak intensity). Is it the intra-type peaks that are of most interest, and hence this is denoted as an intra-type spectrum. In turn, the HN(CO)CACB spectrum only contains peaks that correlate the NH resonances of each residue with the  $\text{C}^\alpha$  and  $\text{C}^\beta$  resonances of the preceding residue, and is denoted an inter-

type spectrum. Matching the  $C^\alpha$  and  $C^\beta$  chemical shifts of the peaks in intra- and inter-type spectra establishes sequential connectivities between the respective backbone amides (Fig. 1.3.1B). The HNCA/HN(CO)CA pair of experiments only show peaks that correlate the  $C^\alpha$  nuclei with the backbone amides, and hence contain less information than the HNCACB/HN(CO)CACB pair, but the advantage of the HNCA/HN(CO)CA pair is that the experiments are more sensitive, and hence may be useful to supply information that is missing from the HNCACB/HN(CO)CACB spectra due to insufficient signal-to-noise. The HNCO/HN(CA)CO pair represent additional information, correlating the amide resonances with those of the carbonyl ( $C'$ ) nuclei; here the HNCO spectrum is the inter-type spectrum, and the HN(CA)CO is the intra-type. (Ikura *et al.*, 1990; Kay *et al.*, 1990; Bax & Ikura, 1991; Clubb *et al.*, 1992; Grzesiek & Bax, 1992; Constantine *et al.*, 1993; Yamazaki *et al.*, 1994; Zhang *et al.*, 1994).



**Figure 1.3.1 NMR assignment strategy of protein backbone resonance assignment.** (A) Representation of the introduction of the third dimension  $^{13}\text{C}$  in addition to the 2D  $^{15}\text{N}$ -HSQC spectrum. These  $^{13}\text{C}$  resonances have a certain correlation with the corresponding amide resonance ( $^{15}\text{N}$ ), depending on the spectrum that recorded. (B) Assignment of protein backbone resonances is commonly performed by using combinations of inter- and intra-type 3D spectra pairs. For a certain amide resonance, HNCACB correlates the backbone amide with four carbon resonances ( $\text{C}\alpha$  and  $\text{C}\beta$  of the current and preceding residues), whereas the HN(CO)CACB yields strips with two carbon resonances ( $\text{C}\alpha$  and  $\text{C}\beta$  of the preceding residue). The  $\text{C}\alpha$  are colored in dark blue, the  $\text{C}\beta$  are in light blue; red circles indicate corresponding resonances in the HN(CO)CACB. By matching the identical  $\text{C}\alpha$ - $\text{C}\beta$  pair in these spectra pairs, the sequential connectivity can be established and the  $^1\text{H}$ ,  $^{15}\text{N}$  shifts of the preceding residue can be identified. Some amino acids (Alanine, Threonine, Serine and Glycine) with special  $\text{C}\alpha$  and  $\text{C}\beta$  chemical shifts are particularly conducive as they allow for the direct identification via their unique chemical shifts. Figures are adapted from <https://www.protein-nmr.org.uk>.

The next step is the assignment of the resonances of the nuclei of the amino-acid side-chains. The assignment is typically performed using 3D H(CCCO)NH-TOCSY and H(C)CH-TOCSY spectra, with a 2D constant-time  $^{13}\text{C}$ -HSQC spectrum serving the role of book-keeper for the resonance assignments (Bax *et al.*, 1990; Montelione *et al.*, 1992; Olejniczak *et al.*, 1992; Bax & Grzesiek, 1993).

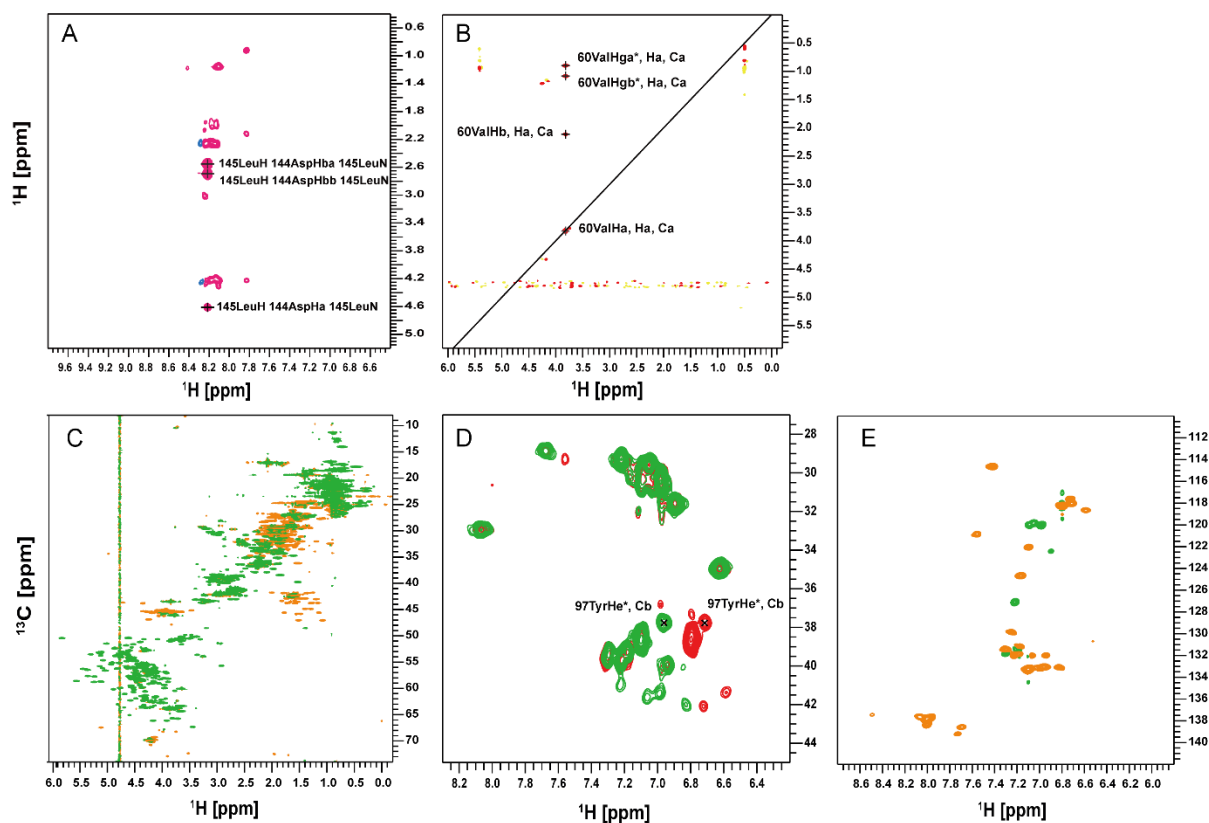
The H(CCCO)NH-TOCSY experiment correlates the NH amide resonances with the resonances of the side-chain protons of the preceding residue. Magnetization is first transferred from the side-chain protons to their directly attached  $^{13}\text{C}$  nuclei, and then to the  $\text{C}\alpha$  nucleus by TOCSY mixing. From there the magnetization is transferred to the  $\text{C}'$  nucleus of the same residue and then the amide group of the following residue, finally being detected on the amide proton. There are chemical-shift evolution periods for the side-chain protons, the amide nitrogen, and its attached proton, resulting in a 3D spectrum with one nitrogen and two proton dimensions (Fig. 1.3.2A).

The H(C)CH-TOCSY spectrum correlates the  $^{13}\text{C}$  and  $^1\text{H}$  resonances of each CH group with all the  $^1\text{H}$  resonances in the side-chain of the same residue. The magnetization transfer pathway of the pulse sequence is approximately symmetric; it starts with proton magnetization, which is then transferred to the directly attached respective  $^{13}\text{C}$  nuclei, before being spread over all the other  $^{13}\text{C}$  nuclei of the side-chain by TOCSY mixing. Finally, the magnetization is transferred back to the directly attached protons for detection. Two chemical shift evolution periods are incorporated while the magnetization is on the protons (at the start and end of the magnetization transfer pathway), and a third chemical shift evolution is implemented while the magnetization is on the  $^{13}\text{C}$  nuclei after the TOCSY mixing element. The 3D spectrum is typically viewed as 2D H-H planes, which show “peak-strips” at the chemical-shift position of each side-chain  $^{13}\text{C}$  nucleus that contain peaks correlating all the side-chain protons with the

proton directly attached to the respective  $^{13}\text{C}$  nucleus (Fig. 1.3.2B). The constant-time  $^{13}\text{C}$ -HSQC spectrum shows peaks for all  $^1\text{H}$ - $^{13}\text{C}$  moieties in a protein, regardless of chemical type. The magnetization is transferred from  $^1\text{H}$  to  $^{13}\text{C}$  via J-coupling and is then transferred back to the hydrogen for detection. The constant-time variant of this experiment involves implementing the  $^{13}\text{C}$  chemical-shift evolution period in a constant-time fashion, which allows the chemical-shift to evolve without concomitant evolution of the homonuclear  $^{13}\text{C}$ - $^{13}\text{C}$  J-coupling. The resolution in the  $^{13}\text{C}$  dimension is much improved compared to a plain  $^{13}\text{C}$ -HSQC spectrum, where the peaks are broadened due to the unresolved peak-splitting arising from evolution of the  $^{13}\text{C}$ - $^{13}\text{C}$  couplings (Fig. 1.3.2C).

The assignment of the protein aromatic side-chain resonances was facilitated by 2D (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE spectra (Yamazaki *et al.*, 1993). The former correlates the  $\text{C}^\beta$  resonances of Phe, Tyr, Trp and His residues with the  $\text{H}^\delta$  resonances of the aromatic ring, while the latter provides the corresponding correlation with the  $\text{H}^\epsilon$  resonances for Phe and Tyr residues (Fig. 1.3.2D). In addition, because the homonuclear  $^{13}\text{C}$ - $^{13}\text{C}$  J-coupling is larger for aromatic than for aliphatic carbon nuclei, a specific constant-time  $^{13}\text{C}$ -HSQC spectrum tailored for aromatic residues is also needed for the assignment (Fig. 1.3.2E). Assignment of the methionine  $\text{C}^\epsilon$  group that is separated by the sulfur atom from the remainder of the methionine side-chain is achieved by inspection of the  $^{13}\text{C}$ -edited NOESY spectrum (described in 1.3.1.5).

The spectra mentioned above are critical for completing the list of chemical shifts of proton, nitrogen, and carbon resonances, so that further studies can be performed. However, assignment of side-chain resonances in large proteins or complexes is more challenging due to the line-broadening that becomes ever more severe with increasing molecular-weight and the resulting lack of resolution and low signal-to-noise ratios of the acquired NMR spectra. In these cases, the development of TROSY (transverse relaxation-optimized spectroscopy) techniques, combined with uniform or selective deuterium-labelling (replacing protons with deuterons), have been instrumental in extending the application of solution NMR to much larger biomolecular systems than was previously possible (chapter 1.3.1.4).



**Figure 1.3.2 NMR assignment of side-chain resonances of proteins.** (A) A (NH) strip that centered on residue Leu145 of the H(CCCO)NH-TOCSY spectrum is shown. In each strip, the hydrogen chemical shifts from the whole side chain of the preceding residue are detectable, although the signal to noise ratio of this experiment is sometimes not good enough. (B) A typical example of a HCCH-TOCSY strip, which is centered on the  $\text{H}\alpha$ - $\text{C}\alpha$  pair of amino acid Val60 (diagonal peak), other H-C groups that from this residue are also presented. Identical strips in which each of these peaks will be diagonal can be found somewhere else in the spectrum. Thus, assignment of all side-chain aliphatic groups can be performed by matching their identical chemical shifts. (C) A constant-time  $^{13}\text{C}$ -HSQC spectrum shows the peaks of each  $^1\text{H}$ - $^{13}\text{C}$  group inside a protein, regardless of chemical type. In this spectrum, positive peaks are shown in green and negative ones are in orange. (D) 2D NMR spectra of (HB)CB(CGCD)HD (green) and (HB)CB(CGCDCE)HE (red) depicts the correlation of the  $\text{C}\beta$  resonance of Try97 with its aromatic protons  $\text{H}\delta$  and  $\text{H}\epsilon$ , as an example of the side chain assignment of aromatic residues. (E) A particular  $^{13}\text{C}$ -HSQC spectrum displays the peaks that corresponding to the H-C groups of aromatic amino acids. Notice that these aromatic groups show higher carbon chemical shifts than that of the aliphatic ones.

### 1.3.1.4 TROSY-NMR for large biological macromolecules in solution

In the early days of biomolecular solution NMR, structure determination was limited to relatively small biomolecules, with molecular weights below  $\sim 10$  kDa. For larger proteins beyond this size, the lengthening of the rotational correlation time has unfavourable

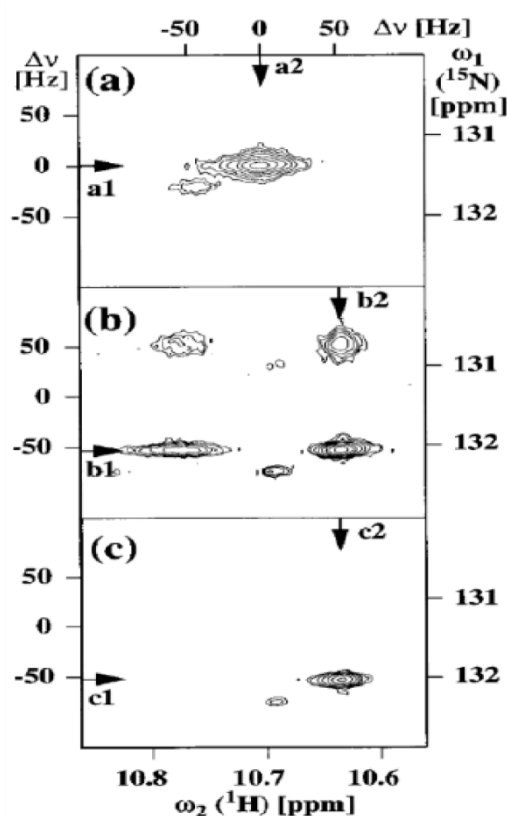
consequences for the relaxation properties of the system; in short, the transverse relaxation become ever faster, which results in very broad linewidths and hence poor resolution and reduced signal-to-noise. In addition, the larger the protein becomes, the greater the number of signals in its NMR spectra, which leads to more peak-overlap and makes analysis and assignment increasingly difficult. Thus, it gets ever more challenging to acquire high-quality NMR spectra as the system under investigation becomes larger. The development of methods for the uniform isotopic labeling of proteins with  $^{15}\text{N}$  and  $^{13}\text{C}$ , together with triple-resonance pulse schemes, led to the possibility of structure determination for proteins up to approximately 30 kDa by solution NMR. For even bigger proteins, obtaining high-resolution spectra requires the partial replacement of protons by deuterons ( $^2\text{H}$ ). This alleviates the problems described above in two ways. Firstly, it reduces the overlap of the remaining  $^1\text{H}$  signals simply because there are fewer signals in the  $^1\text{H}$  spectrum. Secondly, the lower proton density leads to improvements in the relaxation properties of all other magnetic nuclei by strongly reducing the contribution to the transverse relaxation that arises from dipolar interactions with protons (the gyromagnetic ratio of  $^2\text{H}$  — and hence the magnitude of its magnetic moment — is six times smaller than that of  $^1\text{H}$ ).  $^{15}\text{N}$  and  $^{13}\text{C}$  nuclei therefore relax more slowly when attached to and surrounded by deuterons rather than protons, and the remaining protons will also relax more slowly when there are fewer other protons in their vicinity. The slower transverse relaxation leads to narrower linewidths and hence better resolution and improved signal-to-noise (the height of an NMR peak is inversely proportional to its linewidth, so that narrower peaks are also taller peaks).

The development TROSY techniques (Transverse Relaxation-Optimized Spectroscopy) has also been pivotal in extending the size-limit of biomolecular systems that can be studied by solution NMR. The TROSY approach is usually combined with selective protonation of certain types of chemical-groups (e.g. backbone amide groups or the terminal methyl-groups of Ile, Leu, Val, Met, Thr and Ala) in an otherwise perdeuterated background (Pervushin *et al.*, 1997; Sprangers *et al.*, 2007).

All TROSY approaches rely on the phenomenon of cross-correlated cross-relaxation, exploiting the cross-correlation of different types of relaxation mechanism that in favourable cases results in line-narrowing of particular components of the NMR signal. The first implementation of TROSY was that developed for N–H groups of backbone amides. Here, it is the cross-correlation between the dipole-dipole coupling (DD) and chemical shift anisotropy (CSA) (see chapter 1.3.1.2) that results in the two lines of both the  $^1\text{H}$  and  $^{15}\text{N}$  doublets having different linewidths. For example, in a 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum in which decoupling has not been applied in either dimension, the 2D peaks are each split into four component lines due to the signals in each dimension appearing as doublets (the doublet splitting is caused by the



N–H J-coupling). The four lines arise from components of the signal that have different transverse relaxation times during the two chemical-shift evolution periods (due to correlation of the DD and CSA relaxation mechanisms), and thus different linewidths and intensities (peak-heights). The TROSY experiment is designed to select the slowest relaxing component, for which the interaction of the DD and CSA relaxation mechanisms is such that the resulting net local magnetic field is reduced to almost zero, leading to a single, sharp peak (Fig. 1.3.3). As local fields due to chemical shift anisotropy (CSA) scale with the main magnetic field  $B_0$  but those due to dipole-dipole (DD) interactions are independent of the external field, the degree of destructive interference between the two mechanisms depends on the strength of the external magnetic field; it turns out that the optimum destructive interference occurs at high magnetic fields (approximately 900 MHz), at which DD interactions and CSA are of comparable magnitude.



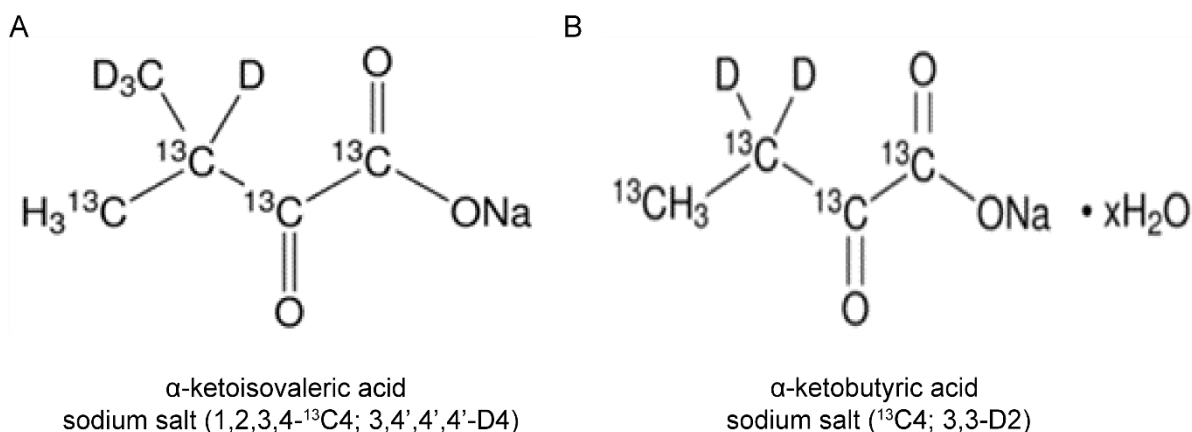
**Figure 1.3.3 TROSY-NMR for large biological systems.** (A) Peak in a conventional  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiment with decoupling. (B) A non-decoupled  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum shows four multiplet components of one amide appear with different line-shapes and intensities. (C) The line shape improvement achieved by  $^1\text{H}$ - $^{15}\text{N}$  TROSY spectrum of same amides shown in A and B. Only the slowest relaxation multiplet component from B was selected. This figure was used by (Pervushin *et al.*, 1997).

Following the invention of TROSY for backbone amide groups, another type of TROSY approach was developed for the  $^{13}\text{C}$ - $^1\text{H}_3$  spin-systems of methyl groups. In the context of

biomolecular protein NMR, suitable methyl groups are located in the side-chains of several hydrophobic amino acids, including isoleucine, leucine, valine, alanine, methionine and threonine. This approach has become known as methyl-TROSY, and has been applied to systems well over 100 kDa in molecular weight. Here, the TROSY effect is generated from the cross-correlation between the intra-methyl  $^1\text{H}$ - $^1\text{H}$  and  $^1\text{H}$ - $^{13}\text{C}$  dipole-dipole interactions (Ollerenshaw *et al.*, 2003); as with  $^1\text{H}$ - $^{15}\text{N}$  TROSY, deuteration of the rest of the protein is strongly recommended to eliminate the dipolar interactions of the methyl-group protons with all other protons, which would otherwise significantly reduce the TROSY effect. Thus, successful application of methyl-TROSY NMR experiment is highly dependent on the preparation of samples containing  $^1\text{H}$  and  $^{13}\text{C}$  labeled methyl groups in a perdeuterated background. In order to achieve high levels of deuteration, bacterial cultures are grown in a  $\text{D}_2\text{O}$ -based culture-medium that contains either  $^2\text{H}$ -glucose or  $^2\text{H}$ -glycerol as the sole carbon. Commercially available precursors (added to the culture shortly before induction of protein expression) allow the selective labeling of the methyl groups of Ile, Leu, Val, Met, Thr and Ala amino acids. However, the yield of perdeuterated protein sample obtained from cultures grown in  $\text{D}_2\text{O}$ -based media is usually less than half the amount that can be obtained from  $\text{H}_2\text{O}$ -based media.

Spectroscopically, the pulse-sequence used to record methyl-TROSY spectra is a simple 2D  $^1\text{H}$ - $^{13}\text{C}$  HMQC (heteronuclear multiple-quantum coherence) correlation experiment that had been developed for 2D H-C correlation spectra many years before the advent of TROSY techniques.

The standard assignment strategy for methyl resonances in large biological systems is based on the “divide and conquer” approach, by which large biomolecules are divided into smaller fragments (sub-complexes or sub-domains) that can be accessed by more traditional NMR methods. For the assignment of methyl  $^1\text{H}$  and  $^{13}\text{C}$  resonances, the backbone resonances and the terminal labeled methyl groups can be connected by preparing a perdeuterated,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled protein sample with addition of  $\alpha$ -keto-acid amino-acid precursors for selective methyl protonation of Ile, Leu and Val during bacterial over-expression (Ollerenshaw *et al.*, 2003). The precursors are deuterated, selectively  $^1\text{H}$ ,  $^{13}\text{C}$ -labelled at one of the terminal methyl groups and  $^{13}\text{C}$ -labeled along the rest of the side-chain to allow transfer of magnetization from the backbone amide nuclei to the methyl groups (Fig. 1.3.4). After assignment of methyl resonances in smaller fragments, assignments are transferred by inspection to the parent complex for further investigation by methyl-TROSY. This strategy has been applied to several biological systems (Sprangers & Kay, 2007; Lapinaite *et al.*, 2013; Danilenko *et al.*, 2019).



**Figure 1.3.4 Methyl group labeling scheme for TROSY-NMR assignment.** Selectively methyl protonated and <sup>13</sup>C-labeled amino acid precursors of Leu and Val (A) and Ile (B) are required for the assignment of methyl <sup>1</sup>H and <sup>13</sup>C resonances of perdeuterated, <sup>13</sup>C, <sup>15</sup>N labeled protein samples.

### 1.3.1.5 Structure calculation by NMR

High-resolution protein structures can be obtained using solution NMR by means of a large collection of inter-proton distance restraints derived from nuclear Overhauser effect Spectroscopy (NOESY). NOESY experiments are based on the nuclear Overhauser effect (NOE), which is the transfer of longitudinal magnetization between two spins that arises due to dipolar-mediated cross-relaxation. The phenomenological definition of the NOE is the change in the longitudinal magnetization of one spin when the magnetization of another spin is perturbed from its equilibrium state. The mutual interaction of the dipolar fields of the two spins leads to relaxation-induced transitions in which the spin-states of both spins are flipped simultaneously. If we consider two spins, spin 1 and spin 2, the differential equation that dictates the rate of change of the *z*-magnetization is:

$$\frac{dI_{1z}}{dt} = -R_z^{(1)}(I_{1z} - I_{1z}^0) - \sigma_{12}(I_{2z} - I_{2z}^0) \quad (1.15)$$

If spin 2 is not in its equilibrium state ( $I_{2z} - I_{2z}^0 \neq 0$ ), the rate of change of the *z*-magnetization on spin 1 will have a contribution that is proportional to the cross-relaxation rate constant  $\sigma_{12}$ , which dictates the time-evolution of the magnetization of spin 1 resulting from a non-equilibrium magnetization on spin 2. This rate constant will be non-zero if there is an appreciable dipolar interaction between the two spins. Here,  $I_{jz}$  is the quantum mechanical operator that represents the *z*-component of the nuclear spin angular momentum for spin *j*.  $R_z^{(1)}$  is the self-relaxation rate constant for spin 1, which represents the recovery of the longitudinal magnetization of spin 1 after perturbation and is independent of the magnetization of spin 2. The superscript 0 represents the equilibrium value of the corresponding quantity.

Since the cross-relaxation rate falls off rapidly with distance ( $\sigma_{12} \propto 1/r^6$ ), the spins need to be reasonably close in space if the cross-relaxation effect between them is to be observable. The presence of cross-relaxation between two spins is responsible for the NOE. Detection of cross relaxation (and hence the NOE) is achieved with experiments that monitor the  $z$ -magnetization of one spin after perturbation of the magnetization of the other.

For the 1D *transient NOE experiment*, two scans with slightly different pulse sequences are acquired to obtain the transient NOE spectrum. In the first scan, the  $z$ -magnetization of spin 2 is inverted by a selective  $180^\circ$  pulse, followed by a delay  $\tau$  during which the cross relaxation occurs, and finally  $z$ -magnetization is made observable by a non-selective  $90^\circ$  pulse. This scan corresponds to what is called the *irradiated spectrum*. The second scan of the experiment is a simple pulse-acquire sequence, with no inversion of spin 2; this corresponds to the *reference spectrum*. By subtracting the reference spectrum from the irradiated spectrum, one obtains the NOE difference spectrum, which reveals the small perturbation — if there is cross-relaxation between the two spins — of the  $z$ -magnetization on spin 1 due to inversion of spin 2.

By solving the pair of coupled differential equations (also known as the *Solomon equations*), of which equation 1.15 is the first, one can compute the  $z$ -magnetization for each spin as a function of the delay  $\tau$  in the irradiated spectrum. For the reference spectrum, the evolution of the magnetization is simple as both spins are in their equilibrium states. The results of both experiments are summarized in the following table:

Table 1.3.1

Experiment	Spin-1 magnetization	Spin-2 magnetization
Irradiate	$I_{1z}(\tau) = 2\sigma_{12}\tau I_{2z}^0 + I_{1z}^0$	$I_{2z}(\tau) = 2R_z^{(2)}\tau I_{2z}^0 - I_{2z}^0$
Reference	$I_{1z}^0$	$I_{2z}^0$

The  $90^\circ$  pulse in both irradiated and reference experiments rotates  $z$ -magnetization into the transverse plane so that the FID can be detected. Fourier transformation of the FID will give a spectrum in which there are two peaks at the respective offsets of spin 1 and spin 2.

The height of the peak for spin 1 and spin 2 is proportional to the  $z$ -magnetization of the corresponding spins  $I_{1z}(\tau)$  and  $I_{2z}(\tau)$ . If both spins are of the same type, their equilibrium  $z$ -

magnetizations will be equal. The data in the table 1.3.1 can be converted into peak heights when observe at time  $\tau$ :

Table 1.3.2

Spectrum	Spin-1 peak height	Spin-2 peak height
Irradiated	$c(1 + 2\sigma_{12}\tau)$	$c(2R_z^{(2)}\tau - 1)$
Reference	$c$	$c$
NOE difference	$c2\sigma_{12}\tau$	$c(2R_z^{(2)}\tau - 2)$

In this table,  $c$  is a constant of proportionality between the magnitude of z-magnetization and peak height of the corresponding signal in the spectrum. The NOE difference spectrum shows the difference when the reference spectrum is subtracted from the irradiated spectrum. Note that the results in the tables above are only valid if the *initial rate limit assumption* can be applied:  $2\sigma_{12}\tau \ll 1$  and  $R_z^{(l)}\tau \ll 1$ . In words, this means that magnetization of neither spin varies significantly from its initial value over the duration of the delay  $\tau$ . If the cross-relaxation rate-constant is positive, then in the irradiated spectrum, the peak for spin 1 will be more intense than that in the reference spectrum; the peak is said to have received an *NOE enhancement*.

The magnitude of the NOE enhancement,  $\eta$ , is described as:

$$\eta = \frac{\text{peak height in irradiated spectrum} - \text{peak height in reference spectrum}}{\text{peak height in reference spectrum}} \quad (1.16)$$

According to table 1.3.2, the enhancement is computed as:

$$\eta = \frac{c(1 + \sigma_{12}\tau) - c}{c} = 2\sigma_{12}\tau \quad (1.17)$$

Thus, the larger the cross-relaxation rate  $\sigma_{12}$ , the greater the enhancement  $\eta$ . Since  $\sigma_{12} \propto 1/r^6$ , spins that are closer in space will have faster cross-relaxation and therefore exhibit larger NOE enhancements. Consequently, the size of  $\eta$  can be used as an indicator of internuclear spatial distance.

The 1D *steady-state NOE experiment* is also recorded as a difference spectrum. Here, in the first scan (the irradiated spectrum), spin 2 is irradiated continuously with a weak RF field,

which leads to its saturation, reducing its z-magnetization to zero. The reference spectrum, as with the transient NOE experiment, is just a standard pulse-acquire experiment. As before, the difference spectrum is obtained by subtracting the reference spectrum from the irradiated spectrum.

By following a similar approach to that shown above, the steady-state NOE enhancement  $\eta_{ss}$  found to be given by:

$$\eta_{ss} = \frac{c \left( 1 + \frac{\sigma_{12}}{R_z^{(1)}} \right) - c}{c} = \frac{\sigma_{12}}{R_z^{(1)}} \quad (1.18)$$

In contrast to the transient NOE experiment in which the NOE enhancement only depends on the cross-relaxation rate  $\sigma_{12}$ , the steady-state NOE enhancement depends on the ratio of the cross-relaxation rate constant  $\sigma_{12}$  to the self-relaxation constant of the observed spin. Since the magnitude of steady-state NOE reflects a balance of between cross and self-relaxation, such NOE measurements tend to only be used qualitatively.

An important application of the steady-state NOE is to enhance the intensity of signals recorded from heteronuclei (e.g.,  $^{13}\text{C}$ ). By irradiating the protons, the z-magnetization of the  $^{13}\text{C}$  heteronucleus will be enhanced due to cross-relaxation. When the z-magnetization is rotated into the transverse plane, stronger  $^{13}\text{C}$  signals can be detected because of the NOE enhancement.

The 1D experiments described above are useful for detecting and measuring NOEs between specific protons and are typically only applied to small molecules. The power of NOEs for structure determination of proteins relies on the measurement of a set of multi-dimensional NOESY spectra, which are all based on a 2D NOESY experiment. In a 2D NOESY experiment, it is the NOE that provides the mechanism by which magnetization is transferred between the nuclei (during the mixing-time that is sandwiched between the indirect chemical-shift evolution period and the directly-detected FID). The presence of a cross-peak connecting two resonances in a NOESY spectrum indicates that the corresponding protons share an NOE, and are therefore close in space. Furthermore, the intensity of the cross-peak depends on the rate of cross-relaxation between the protons, showing an approximate  $1/r^6$ -dependence; the closer the two protons, the stronger the corresponding cross-peak in the NOESY spectrum. In principle, a NOESY spectrum will contain cross-peaks for every pair of protons that are separated by less than about 6 Å. The complete set of cross-peaks therefore represents a dense mesh of inter-atomic distances that accurately define the overall 3D molecular structure. To calculate this structure from the information in the NOESY spectrum requires conversion

of each cross-peak into a corresponding distance-restraint between the two respective protons; the first step in this process is to assign each and every cross-peak to the proton-pair from which it arises. This NOESY assignment is additional to the process of resonance assignment described earlier, and relies on the resonance assignment having been completed beforehand.

Since a typical small-to-medium-sized protein will contain at least 1000 protons, the 2D NOESY spectra of proteins are exceptionally crowded, and extensive peak-overlap means that it is difficult to resolve and assign the individual cross-peaks. Therefore, the 2D NOESY spectrum is extended into a 3D NOESY–HSQC spectrum, in which the additional dimension corresponds to the chemical shift of the heteronucleus that is bound to the directly detected proton. The expansion of the 2D NOESY spectrum into a 3D spectrum significantly reduces — but does not completely eliminate — the problem of overlapping cross-peaks. Since the protein is usually available in  $^{13}\text{C},^{15}\text{N}$ -labelled form, two separate 3D NOESY-HSQC spectra can be recorded: the 3D NOESY– $^{15}\text{N}$ -HSQC spectrum and the 3D NOESY– $^{13}\text{C}$ -HSQC spectrum (Kay *et al.*, 1989; Marion *et al.*, 1989; Zuiderweg & Fesik, 1989).

Structure determination by NMR spectroscopy is the combination of two types of potential-energy functions, chemical information and NMR-derived information, which give a multi-dimensional potential-energy landscape representing the energy of the protein as a function of its atomic coordinates. The chemical information involves information such as bond lengths, bond angles and dihedral angles that are often very precisely defined. The task of the structure-calculation protocol is to convert the information generated by NOE-derived experimental restraints into a structure that can satisfy them while retaining good covalent and non-covalent geometry. Finding the global minimum of the potential energy landscape modeled by these two functions corresponds to determining the set of atomic coordinates that represent the native structure of a protein in solution. The most commonly used algorithms are based on molecular dynamics combined with simulated annealing (Brünger *et al.*, 1997; Stein *et al.*, 1997).

The generation of potential-energy functions that incorporate the NOE-derived information from the NOESY spectra involves first converting the intensities (volumes) of the NOE cross-peaks ( $V$ ) into distances ( $d$ ). This conversion is termed calibration, with the peak intensities being related to the distances by a calibration-factor,  $C$ . The calibration-factor can be determined from the cross-peak intensity,  $v_{ref}$ , that is measured for a proton-pair whose internuclear separation,  $d_{ref}$ , is known and fixed by the chemical structure:

$$V = Cd^{-6} \Rightarrow d = \left(\frac{V}{C}\right)^{-\frac{1}{6}} = \left(\frac{d_{ref}^{-6}}{v_{ref}} V\right)^{-1/6} \quad (1.19)$$

This distance is then converted into upper ( $U$ ) and lower ( $L$ ) distance restraints by introduction of an error estimate,  $\Delta$ :

$$U = d + \Delta \quad (1.20)$$

$$L = d - \Delta \quad (1.21)$$

(1.20) and (1.21) are then incorporated into the potential-energy function. In its simplest form, this function is a quadratic-well type:

$$E_{NOE} = \begin{cases} (L - D)^2 & \text{if } D < L \\ 0 & \text{if } L \leq D \leq U \\ (D - U)^2 & \text{if } D > U \end{cases} \quad (1.22)$$

Here,  $D$  is the inter-nuclear separation in the structure; in other words, the energy of the NOE component of the potential is minimized when the separation of the nuclei in the calculated structure is close to that of the distance estimated from the intensity of the corresponding cross-peak in the NOESY spectrum.

Most of the steps of the molecular dynamics/simulated annealing simulation are performed in a vacuum, with the final step of the structure calculation being a refinement in a thin layer of water, and for this step, van der Waals interactions are modeled using the full Lennard-Jones potential and electrostatic interactions are included as standard Coulomb potentials

### 1.3.2 Small-angle X-ray scattering (SAXS)

Small angle X-ray scattering (SAXS) is a powerful biophysical technique for the structural characterization of both ordered and disordered proteins in solution. It can provide low-resolution information on the overall shape, conformation, and quaternary structure of biological molecules. By incorporating with complementary techniques, such as NMR, crystallography, electron microscopy (EM) or bioinformatics, SAXS is also able to provide unique information on the structural properties of flexible macromolecules that show various structural states under different physiological conditions.

A typical SAXS experiment is performed by synchrotron X-ray radiation, which is then scattered by isotropic diluted protein solution in different concentration range (mg/ml). A 2D scattering pattern that collected by a 2D detector is radially averaged to obtain the scattered intensity, which is a function of the momentum transfer  $s$ , with units of nm or Å (Fig. 1.3.5A).



$$s = \frac{4\pi \sin \theta}{\lambda} \quad (1.23)$$

Here,  $\theta$  is half the angle between the incident and the scattered beam and  $\lambda$  is the wavelength of the X-ray beam.

In a SAXS experiment, the signal of the protein solution must be carefully determined and subtracted from the buffer signal to make sure only get the scattering pattern of the solute.

The Guinier approximation is widely used to determine the basic molecular parameters (Guinier, 1939; Feigin & Svergun, 1987). It describes that in the initial region of the curve, scattering intensity is dependent on the radius of gyration  $R_g$  and the forward scattering intensity  $I(0)$ . For a monodisperse solution, the Guinier equation is given by:

$$I(s) = I(0) \exp \left[ \frac{-s^2 R_g^2}{3} \right] \quad (1.24)$$

In practice, the value of  $R_g$  and  $I(0)$  are estimated from the best possible linear fit of  $\ln[I(s)]$  versus  $s^2$  (Guinier plot), which is valid for sufficiently small  $s$  values, typically in the range up to  $sR_g < 1.3$ . The plot should be linear for monodisperse particles, thence the  $R_g$  is determined as the slope of the line and  $I(0)$  is the intersection with the y-axis.

In fact, the linearity region of the Guinier plot can indicate whether the sample system behaves good or not. Deviation from the initial region of the curve may reflect strong interference effects, polydispersity of the sample or improper background subtraction. Intermolecular attraction or nonspecific aggregation lead to an increase at low angles and overestimation of both  $R_g$  and  $I(0)$ . In contrast, intermolecular repulsion or inter-particle effects result in a decrease at low angles and an underestimation of  $R_g$  and  $I(0)$  (Mertens & Svergun, 2010; Blanchet & Svergun, 2013) (Fig. 1.3.5B).

The scattering curve can be described by the distance distribution function  $P(r)$  which reports the distances between all possible pairs of atoms within the scattering particle. The scattering pattern of a particle  $I(s)$  is a Fourier transform of its  $P(r)$  function; the relation is given by equation:

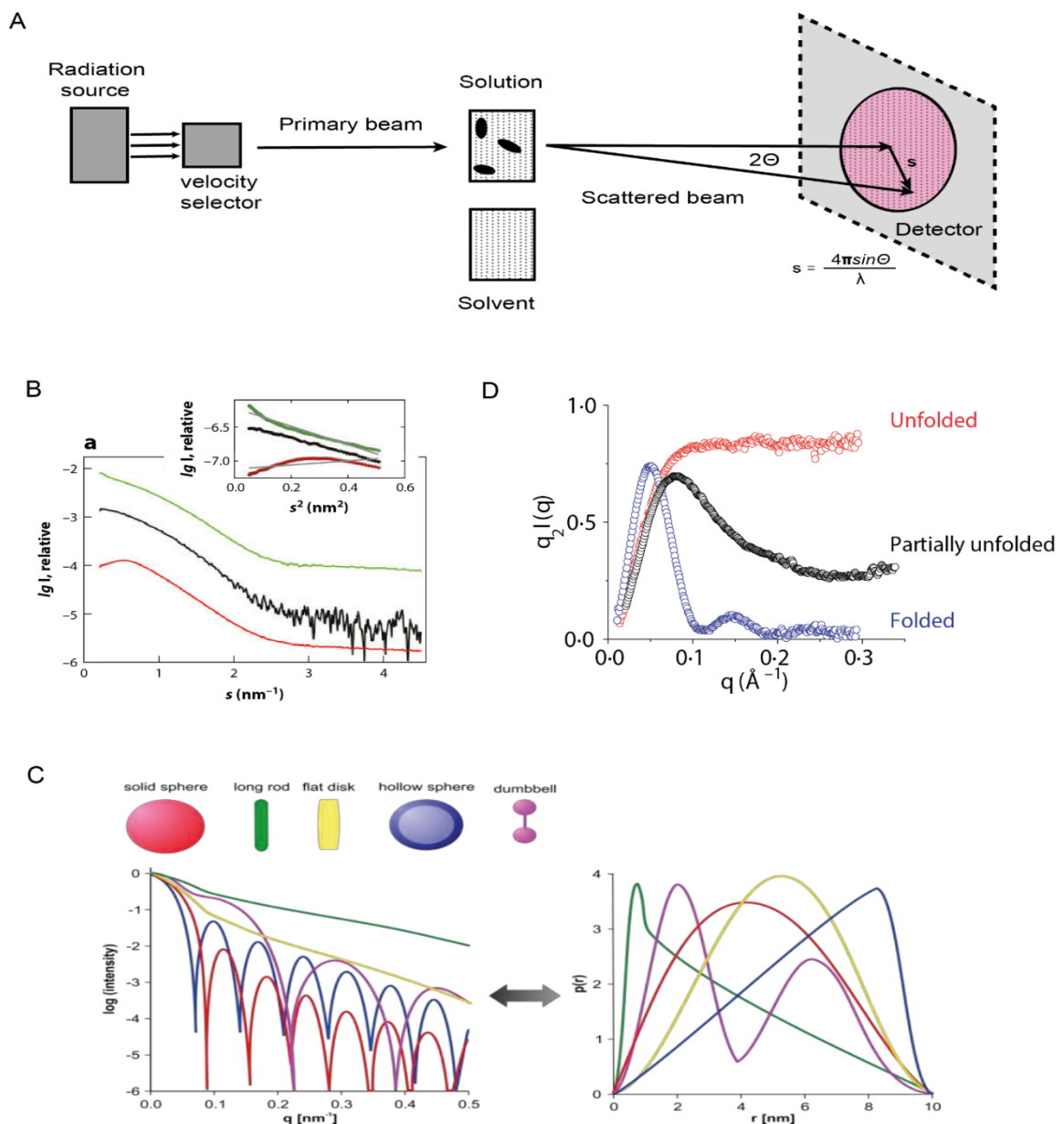
$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin(sr)}{sr} ds \quad (1.25)$$

As the experimental scattering data do not reach from zero to infinite, the  $P(r)$  is calculated by an indirect Fourier transformation, typically using the GNOM program (Mylonas & Svergun, 2007). This program computes  $P(r)$  from the SAXS scattering curve by giving a user defined  $D_{\max}$  and a representation of the scattering curve as a sum of spline functions (smoothed curve). Importantly,  $P(r)$  reflects the distance distribution within the particle. Particles of

different overall shapes theoretically correspond to different scattering curves (Svergun & Koch, 2003) (Fig. 1.3.5C).

Another important application from SAXS data is assessing the folding status of the protein sample by using the Kratky plot [ $s^2 I(s)$  versus  $s$ ]. Well-folded globular proteins typically show a peak at low angles in the plot, whereas unfolded proteins or random coils show no peak, but rather a continuous increase at high  $s$  (Putnam *et al.*, 2007) (Fig. 1.3.5D).

The scattering curves obtained from SAXS data can be used to generate an *ab initio* reconstruction of low-resolution structural envelopes using the program DAMMIF (Svergun, 1999). However, full structural characterisation of a system of interest still depends on high-resolution structural information derived from other techniques such as NMR or crystallography.



**Figure 1.3.5 Small-angle X-ray scattering (SAXS)** (A) Schematic representation of a typical set up of a SAXS experiment. The X-ray beam hits the sample, and the scattered pattern is collected on a 2D detector. (B) An example of Guinier plots dictates the presence of inter-particle interactions (inset). Black: ideal protein solution; Green: attractive interactions; Red: repulsive interactions. Structural parameters of  $R_g$  and  $I(0)$  also could be extracted from the Guinier plot. Figure is adapted from (Blanchet & Svergun, 2013). (C) The distance-distribution functions  $P(r)$  show the particles with different shapes. This figure is adapted from (Svergun & Koch, 2003). (D) Kratky plot of folded (blue), partially folded (black) and unfolded (red) proteins. This figure is adapted from (Putnam *et al.*, 2007).

### 1.3.3 Size-exclusion Chromatography and Multiangle Light Scattering (SEC-MALS)

Size-exclusion chromatography (SEC) is normally used to separate and analyze proteins or other biomolecules, such as nucleic acids (DNA and RNA) or bionanoparticles. It has the advantage of preserving the structure and biological activity of the system of interest. However, it also has limited applicability in the determination of absolute mass and size distribution of biological macromolecules. In order to reliably characterize the oligomeric property and molar mass of proteins or biomolecular complexes, installation of multiangle light scattering (MALS) downstream of the SEC is required for obtaining the molecular size of biomolecules, independent of their shape and SEC retention time:

$$M_w = \frac{R(0)}{Kc \left(\frac{dn}{dc}\right)^2} \quad (1.26)$$

Where  $M_w$  is the molar mass of a given molecule,  $R(0)$  is the reduced Rayleigh ratio that extrapolated to angle zero;  $c$  is the concentration of the molecule;  $dn/dc$  is the polarizability or refractive index increment (the difference between the refractive index of the analyte and solvent);  $K$  is a constant defined in equation 1.26.

In SEC-MALS, the SEC is responsible for separating different species in the solution sample so that they can enter the MALS and their molecular size can be determined respectively. In a typical MALS experiment, the sample solution is irradiated by a laser beam and the scattering intensity is measured at several fixed angles by a few fixed detectors or a movable detector. The fundamental principle of MALS is the proportionality of the resulting scattered light intensity at angle  $\Theta$ ,  $I(\Theta)$  to the molecular weight  $M_w$ , the molecular concentration  $c$  and the refractive index increment  $dn/dc$ .

$$I(\Theta)_{scattered} \propto M_w \left(\frac{dn}{dc}\right)^2 \quad (1.27)$$

Furthermore, the reduced Zimm equation (Zimm, 1948) describes the angular dependence of a certain molecule's Rayleigh scattering intensity, which relates  $I(\Theta)_{\text{scattered}}$  and  $P(\Theta)$ :

$$\left(\frac{Kc}{I(\Theta)_{\text{scattered}}}\right) = \left(\frac{1}{MP(\Theta)}\right) + 2A_2c \quad (1.28)$$

Here,  $P(\Theta)$  is form factor or particle scattering function. Large particles that show an angular variation in the intensity of scattered light is called anisotropic scatterers. The scattered light intensity can be measured at different angles when the molecule is an anisotropic scatterer (>10 nm at  $\lambda = 660$  nm).  $I(\Theta)$  is the intensity of scattered light at angle  $\Theta$ ,  $K$  is a constant defined in equation 1.26, and  $A_2$  is the second virial coefficient as a measure for non-ideal solution behavior of the analyte.

The constant  $K$  is described as:

$$K = \frac{4\pi^2 n_0^2}{N_A \lambda_0^4} \left(\frac{dn}{dc}\right)^2 \quad (1.29)$$

where  $n_0$  is the refractive index of the solvent,  $N_A$  is the Avogadro number as  $6.022 \times 10^{23}$  and  $\lambda_0$  is the vacuum wavelength of the incident laser beam.

Moreover,  $I(\Theta)_{\text{scattered}}$  in 1.28 can be replaced by the excess Rayleigh ratio  $R(\Theta, c)$ , which dictates the difference of the scattered light from the sample solution over the scattered light from solvent.

The refractive index increment value  $dn/dc$  can be measured with the differential refractive index (dRI) detectors. While its value is quite similar for most biomolecules, thence the default value 0.185 ml/g is commonly used for the calculation of the molecular weight (Zhao *et al.*, 2011). To determine the molecular size of biomolecules or biomolecular complexes in solution, the analyte needs to be an anisotropic scatterer detected at different angles under the given incident beam wavelength. The angular dependence  $P(\Theta)$  for smaller molecules (also called isotropic scatterers, <10 nm) is often undetectable.

## 1.4 References

- Andersson, F.I., Blakytyn, R., Kirstein, J., Turgay, K., Bukau, B., Mogk, A. & Clarke, A.K. (2006) Cyanobacterial ClpC/HSP100 protein displays intrinsic chaperone activity. *J Biol Chem*, **281**, 5468-5475.
- Banerjee, A.K., Blanco, M.R., Bruce, E.A., Honson, D.D., Chen, L.M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S.A., Loney, C., Thai, J., Miller, Z.D., Lin, A.E., Schmidt, M.M., Stewart, D.G., Goldfarb, D., De Lorenzo, G., Rihn, S.J., Voorhees, R.M., Botten, J.W., Majumdar, D. & Guttman, M. (2020) SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell*, **183**, 1325-1339 e1321.
- Battesti, A. & Gottesman, S. (2013) Roles of adaptor proteins in regulation of bacterial proteolysis. *Curr Opin Microbiol*, **16**, 140-147.
- Bax, A., Clore, G.M. & Gronenborn, A.M. (1990)  $^1\text{H}$  -  $^1\text{H}$  correlation via isotropic mixing of  $^{13}\text{C}$  magnetization, a new three-dimensional approach for assigning  $^1\text{H}$  and  $^{13}\text{C}$  spectra of  $^{13}\text{C}$ -enriched proteins. *J. Magn. Reson.*, **88**, 425-431.
- Bax, A. & Grzesiek, S. (1993) Methodological advances in protein NMR. *Acc. Chem. Res.*, **26**, 131-138.
- Bax, A. & Ikura, M. (1991) An efficient 3D NMR technique for correlating the proton and  $^{15}\text{N}$  backbone amide resonances with the  $\alpha$ -carbon of the preceding residue in uniformly  $^{15}\text{N}/^{13}\text{C}$  enriched proteins. *J. Biomol. NMR*, **1**, 99-104.
- Benedetti, F., Snyder, G.A., Giovanetti, M., Angeletti, S., Gallo, R.C., Ciccozzi, M. & Zella, D. (2020) Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1. *J Transl Med*, **18**, 329.
- Blanchet, C.E. & Svergun, D.I. (2013) Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annu Rev Phys Chem*, **64**, 37-54.
- Brünger, A.T., Adams, P.D. & Rice, L.M. (1997) New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure*, **5**, 325-336.
- Carroni, M., Franke, K.B., Maurer, M., Jager, J., Hantke, I., Gloge, F., Linder, D., Gremer, S., Turgay, K., Bukau, B. & Mogk, A. (2017) Regulatory coiled-coil domains promote head-to-head assemblies of AAA+ chaperones essential for tunable activity control. *Elife*, **6**.
- Chang, C.K., Hou, M.H., Chang, C.F., Hsiao, C.D. & Huang, T.H. (2014) The SARS coronavirus nucleocapsid protein--forms and functions. *Antiviral Res*, **103**, 39-50.

- Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Maruthi, K., Eng, E.T., Vatandaslar, H., Chait, B.T., Kapoor, T.M., Darst, S.A. & Campbell, E.A. (2020) Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*, **182**, 1560-1573 e1513.
- Clark, L.K., Green, T.J. & Petit, C.M. (2021) Structure of nonstructural protein 1 from SARS-CoV-2. *J Virol*, **95**, e02019-02020.
- Clubb, R.T., Thanabal, V. & Wagner, G. (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue <sup>1</sup>HN, <sup>15</sup>N, and <sup>13</sup>C chemical shifts in <sup>15</sup>N-<sup>13</sup>C-labelled proteins. *J. Magn. Reson.*
- Constantine, K.L., Goldfarb, V., Wittekind, M., Friedrichs, M.S., Anthony, J., Ng, S.-C. & Mueller, L. (1993) Aliphatic <sup>1</sup>H and <sup>13</sup>C resonance assignments for the 26-10 antibody VL domain derived from heteronuclear multidimensional NMR spectroscopy. *J. Biomol. NMR*, **3**, 41-54.
- Dai, W., Zhang, B., Jiang, X.-M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J. & Liu, F. (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science*, **368**, 1331-1335.
- Danilenko, N., Lercher, L., Kirkpatrick, J., Gabel, F., Codutti, L. & Carlomagno, T. (2019) Histone chaperone exploits intrinsic disorder to switch acetylation specificity. *Nat Commun*, **10**, 1-11.
- Dougan, D.A., Reid, B.G., Horwich, A.L. & Bukau, B. (2002) ClpS, a Substrate Modulator of the ClpAP Machine. *Molecular Cell*, **9**, 673-683.
- Dougan, D.A., Weber-Ban, E. & Bukau, B. (2003) Targeted delivery of an ssrA-tagged substrate by the adaptor protein SspB to its cognate AAA+ protein ClpX. *Mol Cell*, **12**, 373-380.
- Doyle, S.M. & Wickner, S. (2009) Hsp104 and ClpB: protein disaggregating machines. *Trends Biochem Sci*, **34**, 40-48.
- Dubnau, D. (1997) Binding and transport of transforming DNA by *Bacillus subtilis*: the role of type-IV pilin-like proteins—a review. *Gene*, **192**, 191-198.
- Dubnau, D. & Roggiani, M. (1990) Growth medium-independent genetic competence mutants of *Bacillus subtilis*. *J. Bacteriol.*, **172**, 4048-4055.
- Elsholz, A.K.W., Birk, M.S., Charpentier, E. & Turgay, K. (2017) Functional Diversity of AAA+ Protease Complexes in *Bacillus subtilis*. *Front Mol Biosci*, **4**, 44.

- Feigin, L. & Svergun, D.I. (1987) *Structure analysis by small-angle X-ray and neutron scattering*. Springer Science and Business Media.
- Fu, Z., Huang, B., Tang, J., Liu, S., Liu, M., Ye, Y., Liu, Z., Xiong, Y., Cao, D. & Li, J. (2020) Structural basis for the inhibition of the papain-like protease of SARS-CoV-2 by small molecules. *Biorxiv*.
- Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z. & Zhang, L. (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, **368**, 779-782.
- Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M. & Neuman, B.W. (2020) Severe acute respiratory syndrome-related coronavirus: The species and its viruses—a statement of the Coronavirus Study Group. *BioRxiv*.
- Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.*, **38**, W695-W699.
- Grzesiek, S. & Bax, A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.*, **114**, 6291-6293.
- Guinier, A. (1939) La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques. Vol. 11, Annales de physique. City. p. 161-237.
- Hahn, J., Kong, L. & Dubnau, D. (1994) The regulation of competence transcription factor synthesis constitutes a critical control point in the regulation of competence in *Bacillus subtilis*. *J. Bacteriol.*, **176**, 5753-5761.
- Hajjema, B.J., Van Sinderen, D., Winterling, K., Kooistra, J., Venema, G. & Hamoen, L.W. (1996) Regulated expression of the *dinR* and *recA* genes during competence development and SOS induction in *Bacillus subtilis*. *Mol. Microbiol.*, **22**, 75-85.
- Hamoen, L.W., Van Werkhoven, A.F., Bijlsma, J.J., Dubnau, D. & Venema, G. (1998) The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev*, **12**, 1539-1550.
- Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J.M. & Glaunsinger, B.A. (2020) The molecular virology of coronaviruses. *J Biol Chem*, **295**, 12910-12934.

- Haslberger, T., Bukau, B. & Mogk, A. (2010) Towards a unifying mechanism for ClpB/Hsp104-mediated protein disaggregation and prion propagation. *Biochem Cell Biol*, **88**, 63-75.
- Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D. & Cramer, P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature*, **584**, 154-156.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., Muller, M.A., Drosten, C. & Pohlmann, S. (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, **181**, 271-280 e278.
- Horwich, A.L., Weber-Ban, E.U. & Finley, D. (1999) Chaperone rings in protein folding and degradation. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 11033-11040.
- Huang, C., Lokugamage, K.G., Rozovics, J.M., Narayanan, K., Semler, B.L. & Makino, S. (2011) Alphacoronavirus transmissible gastroenteritis virus nsp1 protein suppresses protein translation in mammalian cells and in cell-free HeLa cell extracts but not in rabbit reticulocyte lysate. *J Virol*, **85**, 638-643.
- Ikura, M., Kay, L.E. & Bax, A. (1990) A novel approach for sequential assignment of <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N spectra of larger proteins: Heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*, **29**, 4659-4667.
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L.W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z. & Yang, H. (2020) Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature*, **582**, 289-293.
- Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K.G. & Makino, S. (2009) A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat Struct Mol Biol*, **16**, 1134-1140.
- Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., Kubo, H. & Makino, S. (2006) Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc Natl Acad Sci U S A*, **103**, 12885-12890.
- Kay, L.E., Ikura, M., Tschudin, R. & Bax, A. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.*, **89**, 496-514.
- Kay, L.E., Marion, D. & Bax, A. (1989) Practical aspects of 3D heteronuclear NMR of proteins. *J. Magn. Reson.*, **84**, 72-84.



- Khailany, R.A., Safdar, M. & Ozaslan, M. (2020) Genomic characterization of a novel SARS-CoV-2. *Gene Rep*, **19**, 100682.
- Kilkenny, M.L., Veale, C.E., Guppy, A., Hardwick, S.W., Chirgadze, D.Y., Rzechorzek, N.J., Maman, J.D. & Pellegrini, L. (2022) Structural basis for the interaction of SARS-CoV-2 virulence factor nsp1 with DNA polymerase  $\alpha$  - primase. *Protein Sci*, **31**, 333-344.
- Kim, Y., Jedrzejczak, R., Maltseva, N.I., Wilamowski, M., Endres, M., Godzik, A., Michalska, K. & Joachimiak, A. (2020) Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci*, **29**, 1596-1605.
- Kirstein, J., Schlothauer, T., Dougan, D.A., Lilie, H., Tischendorf, G., Mogk, A., Bukau, B. & Turgay, K. (2006) Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. *EMBO J*, **25**, 1481-1491.
- Kneller, D.W., Phillips, G., O'Neill, H.M., Jedrzejczak, R., Stols, L., Langan, P., Joachimiak, A., Coates, L. & Kovalevsky, A. (2020) Structural plasticity of SARS-CoV-2 3CL M(pro) active site cavity revealed by room temperature X-ray crystallography. *Nat Commun*, **11**, 3202.
- Lapinaite, A., Simon, B., Skjaerven, L., Rakwalska-Bange, M., Gabel, F. & Carlomagno, T. (2013) The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature*, **502**, 519-523.
- Lazazzera, B.A. & Grossman, A.D. (1997) A regulatory switch involving a Clp ATPase. *Bioessays*, **19**, 455-458.
- Liu, J., Mei, Z., Li, N., Qi, Y., Xu, Y., Shi, Y., Wang, F., Lei, J. & Gao, N. (2013) Structural Dynamics of the MecA-ClpC Complex: A TYPE II AAA+ PROTEIN UNFOLDING MACHINE\* $\diamond$ . *J. Biol. Chem.*, **288**, 17597-17608.
- Lokugamage, K.G., Narayanan, K., Huang, C. & Makino, S. (2012) Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation. *J Virol*, **86**, 13598-13608.
- Ma, C., Sacco, M.D., Hurst, B., Townsend, J.A., Hu, Y., Szeto, T., Zhang, X., Tarbet, B., Marty, M.T., Chen, Y. & Wang, J. (2020) Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res*, **30**, 678-692.
- Madhugiri, R., Fricke, M., Marz, M. & Ziebuhr, J. (2016) Coronavirus cis-Acting RNA Elements. *Adv Virus Res*, **96**, 127-163.

- Magnuson, R., Solomon, J. & Grossman, A.D. (1994) Biochemical and genetic characterization of a competence pheromone from *B. subtilis*. *Cell*, **77**, 207-216.
- Mariano, G., Farthing, R.J., Lale-Farjat, S.L.M. & Bergeron, J.R.C. (2020) Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. *Front Mol Biosci*, **7**, 605236.
- Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. & Clore, G.M. (1989) Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1. beta. *Biochemistry*, **28**, 6150-6156.
- Masters, P.S. (2006) The Molecular Biology of Coronaviruses. *Adv Virus Res*, 193-292.
- McBride, R., van Zyl, M. & Fielding, B.C. (2014) The coronavirus nucleocapsid is a multifunctional protein. *Viruses*, **6**, 2991-3018.
- Mei, Z., Wang, F., Qi, Y., Zhou, Z., Hu, Q., Li, H., Wu, J. & Shi, Y. (2009) Molecular determinants of MecA as a degradation tag for the ClpCP protease. *J Biol Chem*, **284**, 34366-34375.
- Mertens, H.D. & Svergun, D.I. (2010) Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol*, **172**, 128-141.
- Montelione, G.T., Lyons, B.A., Emerson, S.D. & Tashiro, M. (1992) An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J Am Chem Soc*, **114**, 10974-10975.
- Mylonas, E. & Svergun, D.I. (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.*, **40**, s245-s249.
- Neuman, B.W., Kiss, G., Kunding, A.H., Bhella, D., Baksh, M.F., Connelly, S., Droese, B., Klaus, J.P., Makino, S., Sawicki, S.G., Siddell, S.G., Stamou, D.G., Wilson, I.A., Kuhn, P. & Buchmeier, M.J. (2011) A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol*, **174**, 11-22.
- Ogura, T. & Wilkinson, A. (2001) AAA+ superfamily ATPases: common structure—diverse function. *Genes Cells*, **6**, 575-597.
- Olejniczak, E.T., Xu, R.X. & Fesik, S.W. (1992) A 4D HCCH-TOCSY experiment for assigning the side chain <sup>1</sup>H and <sup>13</sup>C resonances of proteins. *J. Biomol. NMR*, **2**, 655-659.

- Ollerenshaw, J.E., Tugarinov, V. & Kay, L.E. (2003) Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry*, **41**, 843-852.
- Paules, C.I., Marston, H.D. & Fauci, A.S. (2020) Coronavirus Infections-More Than Just the Common Cold. *JAMA*, **323**, 707-708.
- Peng, Q., Peng, R., Yuan, B., Zhao, J., Wang, M., Wang, X., Wang, Q., Sun, Y., Fan, Z., Qi, J., Gao, G.F. & Shi, Y. (2020) Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. *Cell Rep*, **31**, 107774.
- Perlman, S. & Netland, J. (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol*, **7**, 439-450.
- Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. (1997) Attenuated T<sub>2</sub> 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.
- Pervushin, K., Tan, E., Parthasarathy, K., Lin, X., Jiang, F.L., Yu, D., Vararattanavech, A., Soong, T.W., Liu, D.X. & Torres, J. (2009) Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoS Pathog*, **5**, e1000511
- Putnam, C.D., Hammel, M., Hura, G.L. & Tainer, J.A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys*, **40**, 191-285.
- Robert, X. & Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, **42**, W320-W324.
- Roggiani, M., Hahn, J. & Dubnau, D. (1990) Suppression of early competence mutations in *Bacillus subtilis* by mec mutations. *J. Bacteriol.*, **172**, 4056-4063.
- Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N.L., Kiryukhina, O., Wiersum, G., Kim, Y., Jedrzejczak, R., Maltseva, N.I., Endres, M., Jaroszewski, L., Godzik, A., Joachimiak, A. & Satchell, K.J.F. (2020) The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv*.
- Rosenblum, G., Elad, N., Rozenberg, H., Wiggers, F., Jungwirth, J. & Hofmann, H. (2021) Allostery through DNA drives phenotype switching. *Nat Commun*, **12**, 2967.
- Rut, W., Lv, Z., Zmudzinski, M., Patchett, S., Nayak, D., Snipas, S.J., El Oualid, F., Huang, T.T., Bekes, M., Drag, M. & Olsen, S.K. (2020) Activity profiling and structures of inhibitor-bound SARS-CoV-2-PLpro protease provides a framework for anti-COVID-19 drug design. *bioRxiv*.

- Sauer, R.T. & Baker, T.A. (2011) AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem*, **80**, 587-612.
- Schirmer, E.C., Glover, J.R., Singer, M.A. & Lindquist, S. (1996) HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem. Sci.*, **21**, 289-296.
- Schoeman, D. & Fielding, B.C. (2019) Coronavirus envelope protein: current knowledge. *Virology*, **16**, 69.
- Schubert, K., Karousis, E.D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L.A., Leibundgut, M., Thiel, V., Muhlemann, O. & Ban, N. (2020) SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol*, **27**, 959-966.
- Semper, C., Watanabe, N. & Savchenko, A. (2021) Structural characterization of nonstructural protein 1 from SARS-CoV-2. *iScience*, **24**, 101903.
- Shi, M., Wang, L., Fontana, P., Vora, S., Zhang, Y., Fu, T.M., Lieberman, J. & Wu, H. (2020a) SARS-CoV-2 Nsp1 suppresses host but not viral translation through a bipartite mechanism. *BioRxiv*.
- Shi, W., Chen, M., Yang, Y., Zhou, W., Chen, S., Yang, Y., Hu, Y. & Liu, B. (2020b) A dynamic regulatory interface on SARS-CoV-2 RNA polymerase. *bioRxiv*.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. & Söding, J. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
- Skowronski, D.M., Astell, C., Brunham, R.C., Low, D.E., Petric, M., Roper, R.L., Talbot, P.J., Tam, T. & Babiuk, L. (2005) Severe acute respiratory syndrome (SARS): a year in review. *Annu Rev Med*, **56**, 357-381.
- Sola, I., Almazan, F., Zuniga, S. & Enjuanes, L. (2015) Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol*, **2**, 265-288.
- Solomon, J.M., Lazazzera, B.A. & Grossman, A.D. (1996) Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes Dev.*, **10**, 2014-2024.
- Sprangers, R. & Kay, L.E.J.N. (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. **445**, 618-622.
- Sprangers, R., Velyvis, A. & Kay, L.E. (2007) Solution NMR of supramolecular complexes: providing new insights into function. *Nat. Methods*, **4**, 697-703.

- Stadler, K., Massignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H.D. & Rappuoli, R. (2003) SARS--beginning to understand a new virus. *Nat Rev Microbiol*, **1**, 209-218.
- Stein, E.G., Rice, L.M. & Brünger, A.T. (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J. Magn. Reson.*, **124**, 154-164.
- Susanna, K.A., Fusetti, F., Thunnissen, A.W.H., Hamoen, L.W. & Kuipers, O.P. (2006) Functional analysis of the competence transcription factor ComK of *Bacillus subtilis* by characterization of truncation variants. *Microbiology (Reading)*, **152**, 473-483.
- Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.*, **76**, 2879-2886.
- Svergun, D.I. & Koch, M.H. (2003) Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.*, **66**, 1735.
- Tanaka, T., Kamitani, W., DeDiego, M.L., Enjuanes, L. & Matsuura, Y. (2012) Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA. *J Virol*, **86**, 11128-11137.
- Tang, Q., Song, Y., Shi, M., Cheng, Y., Zhang, W. & Xia, X.Q. (2015) Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. *Sci Rep*, **5**, 17155.
- Terada, Y., Kawachi, K., Matsuura, Y. & Kamitani, W. (2017) MERS coronavirus nsp1 participates in an efficient propagation through a specific interaction with viral RNA. *Virology*, **511**, 95-105.
- Thoms, M., Buschauer, R., Ameisemeier, M., Koepke, L., Denk, T., Hirschenberger, M., Kratzat, H., Hayn, M., Mackens-Kiani, T. & Cheng, J. (2020) Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science*, **369**, 1249-1255.
- Tidu, A., Janvier, A., Schaeffer, L., Sosnowski, P., Kuhn, L., Hammann, P., Westhof, E., Eriani, G. & Martin, F. (2021) The viral protein NSP1 acts as a ribosome gatekeeper for shutting down host translation and fostering SARS-CoV-2 translation. *Rna*, **27**, 253-264.
- Turgay, K., Hahn, J., Burghoorn, J. & Dubnau, D. (1998) Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J*, **17**, 6730-6738.
- V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. (2021) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol*, **19**, 155-170.

- van Sinderen, D., Luttinger, A., Kong, L., Dubnau, D., Venema, G. & Hamoen, L. (1995) comK encodes the competence transcription factor, the key regulatory protein for competence development in *Bacillus subtilis*. *Mol. Microbiol.*, **15**, 455-462.
- Wacker, A., Weigand, J.E., Akabayov, S.R., Altincekic, N., Bains, J.K., Banijamali, E., Binas, O., Castillo-Martinez, J., Cetiner, E., Ceylan, B., Chiu, L.Y., Davila-Calderon, J., Dhamotharan, K., Duchardt-Ferner, E., Ferner, J., Frydman, L., Furtig, B., Gallego, J., Grun, J.T., Hacker, C., Haddad, C., Hahnke, M., Hengesbach, M., Hiller, F., Hohmann, K.F., Hyman, D., de Jesus, V., Jonker, H., Keller, H., Knezic, B., Landgraf, T., Lohr, F., Luo, L., Mertinkus, K.R., Muhs, C., Novakovic, M., Oxenfarth, A., Palomino-Schatzlein, M., Petzold, K., Peter, S.A., Pyper, D.J., Qureshi, N.S., Riad, M., Richter, C., Saxena, K., Schamber, T., Scherf, T., Schlagnitweit, J., Schlundt, A., Schnieders, R., Schwalbe, H., Simba-Lahuasi, A., Sreeramulu, S., Stirnal, E., Sudakov, A., Tants, J.N., Tolbert, B.S., Vogeles, J., Weiss, L., Wirmer-Bartoschek, J., Wirtz Martin, M.A., Wohnert, J. & Zetzsche, H. (2020) Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res*, **48**, 12415-12435.
- Wang, F., Mei, Z., Qi, Y., Yan, C., Hu, Q., Wang, J. & Shi, Y. (2011) Structure and mechanism of the hexameric MecA-ClpC molecular machine. *Nature*, **471**, 331-335.
- Wang, Q., Wu, J., Wang, H., Gao, Y., Liu, Q., Mu, A., Ji, W., Yan, L., Zhu, Y., Zhu, C., Fang, X., Yang, X., Huang, Y., Gao, H., Liu, F., Ge, J., Sun, Q., Yang, X., Xu, W., Liu, Z., Yang, H., Lou, Z., Jiang, B., Guddat, L.W., Gong, P. & Rao, Z. (2020) Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell*, **182**, 417-428 e413.
- Wang, Y., Kirkpatrick, J., Zur Lage, S., Korn, S.M., Neissner, K., Schwalbe, H., Schlundt, A. & Carlomagno, T. (2021) (1)H, (13)C, and (15)N backbone chemical-shift assignments of SARS-CoV-2 non-structural protein 1 (leader protein). *Biomol NMR Assign*, **15**, 287-295.
- Wathelet, M.G., Orr, M., Frieman, M.B. & Baric, R.S. (2007) Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. *J Virol*, **81**, 11620-11633.
- Weibezahn, J., Tessarz, P., Schlieker, C., Zahn, R., Maglica, Z., Lee, S., Zentgraf, H., Weber-Ban, E.U., Dougan, D.A. & Tsai, F.T. (2004) Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell*, **119**, 653-665.
- Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G. & Jiang, T. (2020) Genome

- Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe*, **27**, 325-328.
- Yamazaki, T., Forman-Kay, J. & Kay, L. (1993) Two-Dimensional NMR Experiments for Correlating  $^1\text{H}$  / Chemical Shifts of Aromatic Residues in  $^1\text{H}$  /  $^{13}\text{C}$ -Labeled Proteins via Scalar Couplings. *J Am Chem Soc*, **115**, 11054-11055.
- Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D. & Kay, L.E. (1994) A suite of triple resonance NMR experiments for the backbone assignment of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  labeled proteins with high sensitivity. *J. Am. Chem. Soc.*, **116**, 11655-11666.
- Yin, W., Mao, C., Luan, X., Shen, D.-D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W. & Gao, M.J.S. (2020) Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*, **368**, 1499-1504.
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K. & Hilgenfeld, R.J.S. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science*, **368**, 409-412.
- Zhang, O., Kay, L.E., Olivier, J.P. & Forman-Kay, J.D. (1994) Backbone  $^1\text{H}$  and  $^{15}\text{N}$  resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J Journal of biomolecular NMR*, **4**, 845-858.
- Zhao, H., Brown, P.H. & Schuck, P. (2011) On the distribution of protein refractive index increments. *Biophys J*, **100**, 2309-2317.
- Zhao, K., Ke, Z., Hu, H., Liu, Y., Li, A., Hua, R., Guo, F., Xiao, J., Zhang, Y., Duan, L., Yan, X.F., Gao, Y.G., Liu, B., Xia, Y. & Li, Y. (2021) Structural Basis and Function of the N Terminus of SARS-CoV-2 Nonstructural Protein 1. *Microbiol Spectr*, **9**, e0016921.
- Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F. & Shi, Z.L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, **579**, 270-273.
- Zimm, B.H (1948) The scattering of light and the radial distribution function of high polymer solutions. *J. Chem. Phys*, **16**, 1093-1099.
- Zuiderweg, E.R. & Fesik, S.W. (1989) Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *J. Biochem.*, **28**, 2387-2391.

## Chapter 2: Publications and Manuscripts

### 2.1 Manuscript 1

#### **Structural insights into the activity regulation of full-length non-structural protein 1 from SARS-CoV-2**

Ying Wang<sup>1</sup>, John Kirkpatrick<sup>1,2,3</sup>, Susanne zur Lage<sup>2</sup>, Teresa Carlomagno<sup>1,2,3,4\*</sup>

<sup>1</sup> Centre of Biomolecular Drug Research (BMWZ), Leibniz University Hannover, Schneiderberg 38, D-30167 Hannover, Germany

<sup>2</sup> NMR-based Structural Chemistry, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig, Germany

<sup>3</sup> School of Biosciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK

<sup>4</sup> Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK

Type of authorship:	First author
Type of article:	Research article
Share of the work:	75 %
Contribution to the publication:	Planned and performed all experiments, analyzed data, prepared all figures, and wrote the paper
Journal:	Structure
Impact factor:	5.006
Status of publication:	Submitted



# Structural insights into the activity regulation of full-length non-structural protein 1 from SARS-CoV-2

Ying Wang<sup>1</sup>, John Kirkpatrick<sup>1,2,3</sup>, Susanne zur Lage<sup>2</sup>, Teresa Carlomagno<sup>1,2,3,4\*</sup>

<sup>1</sup> Centre of Biomolecular Drug Research (BMWZ), Leibniz University Hannover, Schneiderberg 38, D-30167 Hannover, Germany

<sup>2</sup> NMR-based Structural Chemistry, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig, Germany

<sup>3</sup> School of Biosciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK

<sup>4</sup> Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK

\*To whom correspondence should be addressed: t.carlomagno@bham.ac.uk. Correspondence can also be addressed to: covid19-ffm@dlist.server.uni-frankfurt.de.

## Summary

Non-structural protein 1 (Nsp1) of SARS-CoV-2 is a major virulence factor and thus an attractive drug target. The last 33 amino acids of Nsp1 have been shown to bind within the mRNA entry tunnel of the 40S ribosomal subunit, shutting off host gene expression. Here, we report the first solution-state structure of full-length Nsp1, which features an  $\alpha/\beta$  fold formed by a six-stranded, capped  $\beta$ -barrel-like globular domain (N-terminal domain, NTD), flanked by short N-terminal and long C-terminal flexible tails. The NTD has been found to be critical for 40S-mediated viral mRNA recognition and promotion of viral gene expression. We find that in free Nsp1 the NTD mRNA-binding surface is occluded by interactions with the acidic C-terminal tail, suggesting a mechanism of activity regulation based on the interplay between the folded NTD and the disordered C-terminal region, which will guide drug-design approaches for this important target.

## Keywords

SARS-CoV-2; Nsp1; non-structural proteins; drug targets; NMR spectroscopy

## Introduction

The COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has considerably increased the awareness of coronaviruses as a threat to human health worldwide. Seven coronaviruses — all belonging to the alpha- and beta-coronavirus genera — are known to infect mammalian species (Hartenian et al., 2020). Human alpha-

coronaviruses such as HCoV-229E, HCoV-NL63, HCoV-OC43 and HCoV-HKU1 cause seasonal and mild respiratory-tract infections with common-cold-like symptoms (Paules et al., 2020), while SARS-CoV-2, together with its homolog SARS-CoV-1 and the Middle East respiratory syndrome coronavirus (MERS-CoV) are established agents of life-threatening acute respiratory pathologies and severe lung disease (Skowronski et al., 2005; Tang et al., 2015).

SARS-CoV-2 has an enveloped, positive-sense, single-stranded RNA genome, approximately 30 kb in length. The 5'-capped and 3'-poly-adenylated genome comprises two terminal untranslated regions (the 5' and 3' UTRs), a positive-sense single-strand genomic RNA and a nested set of 9 sub-genomic mRNAs (Chan et al., 2020; Masters, 2006). Two open-reading-frames (ORF1a and ORF1b) span the first two-thirds of the SARS-CoV-2 genome and encode two polyproteins, which are then proteolytically cleaved by two viral proteases to yield 16 so-called non-structural proteins (Khailany et al., 2020; Stadler et al., 2003). Among these, non-structural protein 1 (Nsp1), also known as the leader protein, plays a key role in suppressing host-cell protein translation, and thus the host immune response to the virus.

Nsp1 inhibits host translational activity by inserting the second half of its C-terminal tail (residues 148–180; termed the C-terminal domain (CTD), Figure 1A) into the entrance of the mRNA tunnel of the 40S ribosomal subunit. This blocks translation of host transcripts at the initiation stage, and may lead to a template-dependent endonucleolytic cleavage of the mRNA transcript, and eventually host mRNA degradation in infected cells (Kamitani et al., 2009; Kamitani et al., 2006; Lokugamage et al., 2012; Schubert et al., 2020; Thoms et al., 2020). Nsp1-dependent translational inhibition represses the ability of human lung epithelial cells to respond to IFN- $\beta$  stimulation, thus inactivating virus- and IFN-dependent signaling pathways and resulting in the attenuation of the innate immune response (Banerjee et al., 2020; Wathelet et al., 2007).

The viral mRNA evades Nsp1-mediated translation inhibition through as-yet unknown mechanisms that has been proposed to rely on the association between the N-terminal domain (NTD) of Nsp1 (defined here as residues 10–127, Figure 1A) and the SARS-CoV-2 5'-UTR (Banerjee et al., 2020; Huang et al., 2011; Lapointe et al., 2021; Schubert et al., 2020; Shi et al., 2020; Tanaka et al., 2012; Tidu et al., 2021). However, the viral 5' UTR appears to have no affinity for either free Nsp1 or any component of the ribosome (Tidu et al., 2021). Deletion of either the NTD or the CTD or mutations in both regions substantially reduce the impact of Nsp1 on viral load and suppression of host gene expression, underlining the significance of both the Nsp1-NTD and the C-terminal tail in sustaining the activity of Nsp1 as a major virulence factor (Benedetti et al., 2020; Lin et al., 2021; Mendez et al., 2021; Schubert et al., 2020; Tanaka et al., 2012; Thoms et al., 2020; Wathelet et al., 2007).

There have been several structural studies on Nsp1 from SARS-CoV-1 (PDB code: 2GDT/2HSX (Almeida et al., 2007)) and SARS-CoV-2 (PDB codes: 7K7P, 7K3N, 7EQ4 (Clark et al., 2021; Semper et al., 2021; Zhao et al., 2021)), but all of the corresponding structures represent only the globular core and end at residue 126 or 127. Cryo-EM studies on Nsp1 from SARS-CoV-2 in complex with the 40S ribosomal subunit have shown that the CTD forms two  $\alpha$ -helices stably bound inside the ribosomal mRNA entry channel (Schubert et al., 2020; Shi et al., 2020; Thoms et al., 2020; Yuan et al., 2020), but the structure does not extend further towards the N-terminus of the protein. Recently, the NMR-derived secondary-structure determination of full-length Nsp1 (FL-Nsp1) has shown that the C-terminal tail (residues 128–180, linker + CTD, Figure 1A) remains largely disordered in solution, although helix-forming propensity was detected in the region that forms the helical hairpin in the ribosome mRNA tunnel (Wang et al., 2021). Up to now, there has been no complete atomic-resolution structure of FL-Nsp1, from either SARS-CoV-1 or SARS-CoV-2.

Here, we report the high-resolution solution structure of FL-Nsp1 from SARS-CoV-2. The full-length protein retains its globular core domain formed by the various truncated forms of the protein (NTD, residues 10–127, Figure 1A): the  $\alpha/\beta$  fold comprising a six-stranded, capped  $\beta$ -barrel motif (Almeida et al., 2007). Both the short N-terminal and long C-terminal tails (residues 1–9 and 128–180, respectively) are mostly unstructured and flexible, but the NMR spectra reveal the presence of interactions between the tails and the globular core domain. The structure reveals that — in the full-length protein — the positively charged patch comprising the end of the  $\beta_6$  strand, which has been demonstrated to be critical for RNA binding (Almeida et al., 2007; Lokugamage et al., 2012; Mendez et al., 2021; Nakagawa et al., 2016; Tanaka et al., 2012; Terada et al., 2017), interacts with the highly acidic C-terminal tail, effectively occluding the RNA-binding surface from intermolecular interactions. From the perspective of drug design, our study indicates that the RNA-binding surface of Nsp1 is not completely accessible in the free protein and that competition with the long C-terminal tail for this binding site needs to be taken into account in the design of novel therapeutics targeting free Nsp1.

## **Results and Discussion**

### **Structure determination of FL-Nsp1 from SARS-CoV-2**

We used NMR spectroscopy to determine the structure of FL-Nsp1 in solution. The protein was expressed with  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labelling and purified as described previously (Wang et al., 2021). The rotational correlation-time estimated from backbone  $^{15}\text{N}$  relaxation experiments (Dayie and Wagner, 1994; Farrow et al., 1994) indicated that Nsp1 is a monomer in solution. Nearly complete backbone and side-chain resonance assignments were determined using standard triple-resonance NMR methods. The backbone assignment has been reported previously

(Wang et al., 2021). In brief, we were able to obtain  $^1\text{H}$  and  $^{15}\text{N}$  assignments for 99.4% and 93.9% backbone amide groups, and  $^{13}\text{C}$  assignments for 96.7%, 100% and 100% of the  $\text{C}'$ ,  $\text{C}^\alpha$  and  $\text{C}^\beta$  nuclei, respectively. Amide assignments are missing for residue G127. The sidechain  $^1\text{H}$  resonances were assigned to a completeness of 98.4%. NOEs were derived from standard 3D NOESY- $^{15}\text{N}$ -HSQC and NOESY- $^{13}\text{C}$ -HSQC spectra. The final data-collection statistics and structural quality scores are shown in Tables 1 and 2. The structure closest-to-the-mean was chosen as the representative structure of the ensemble and used in all figures, unless stated otherwise.

**Table 1. Assignment and restraint statistics.** NOE distance restraints are classified by sequence-separation ( $\Delta i$ ): intra-residue (intra-res.;  $\Delta i = 0$ ), sequential ( $\Delta i = 1$ ), short-range ( $2 \leq \Delta i \leq 3$ ), medium-range (med.-range;  $4 \leq \Delta i \leq 5$ ) and long-range ( $\Delta i > 5$ ).

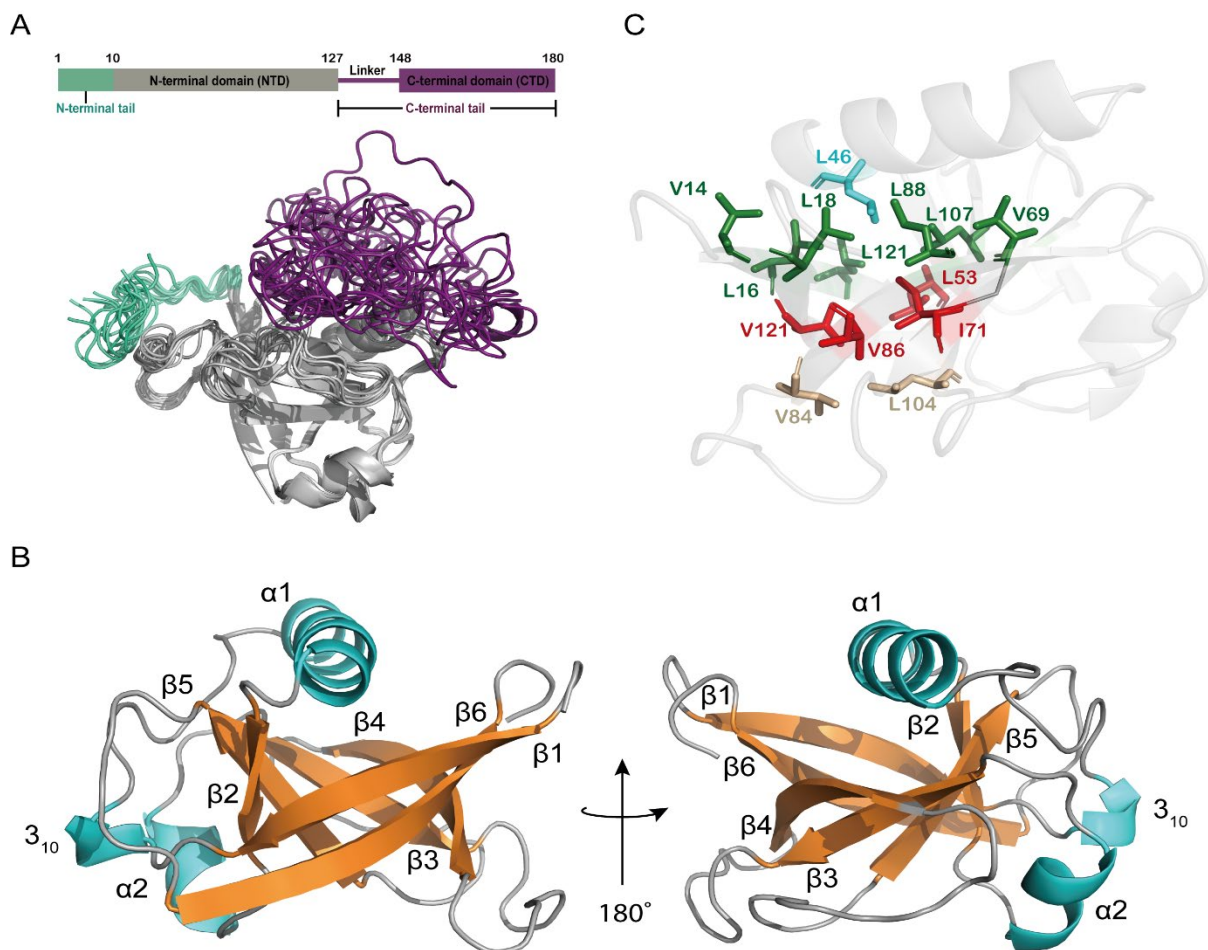
<b>Assignment coverage</b>						
<b>Overall, by element:</b>						
	H	C	N			
	98.9 %	91.6 %	75.4 %			
<b>Backbone, by nucleus-type:</b>						
	$\text{H}^{\text{N}}$	N	$\text{C}'$	$\text{C}^\alpha/\text{C}^\beta$	$\text{H}^\alpha$	Overall
	99.4 %	93.9 %	96.7 %	100.0 %	100.0 %	<b>98.3 %</b>
<b>Sidechain, by element:</b>						
	H	C/N				
	98.4 %	79.3 %				
<b>Restraint counts</b>						
<b>NOE distance restraints:</b>						
	Intra-res.	Sequential	Short-range	Med.-range	Long-range	Total
Unambig.	1484	813	283	74	783	<b>3437</b>
Ambig.	528.0	472.4	309.6	93.8	797.2	<b>2201</b>
Overall	2012.0	1285.4	592.6	167.8	1580.2	<b>5638</b>
<b>Backbone dihedral restraints:</b>						
	$\phi$	$\psi$	Total			
	94	94	<b>188</b>			
<b>Hydrogen-bond restraints:</b>						
	H-to-O	N-to-O	Total			
	51	38	<b>89</b>			
<b>RDC restraints (backbone N-H):</b>						
	Tight	Loose	Total			
	73	16	<b>89</b>			

**Table 2. Structural quality parameters.** The first column lists average values for the ensemble (with standard deviations in parentheses), while the second column lists the corresponding values for the representative structure (closest-to-the-mean). Structural RMSDs were calculated over residues 12–126.

	Ensemble average ( $\pm$ S.D.)	Representative structure
<b>Structural RMSDs (fitting to mean):</b>		
Backbone ( $\text{\AA}$ )	0.57 ( $\pm$ 0.11)	0.43
Heavy atoms ( $\text{\AA}$ )	0.99 ( $\pm$ 0.09)	0.92
<b>RMSDs from experimental restraints:</b>		
NOE distances ( $\text{\AA}$ )	0.02897 ( $\pm$ 0.00091)	0.02773
H-bond distances ( $\text{\AA}$ )	0.0142 ( $\pm$ 0.0029)	0.0129
Dihedral angles (deg)	0.977 ( $\pm$ 0.087)	1.012
RDCs, tight (Hz)	0.894 ( $\pm$ 0.066)	0.881
RDCs, loose (Hz)	2.92 ( $\pm$ 0.19)	2.63
<b>Violation-counts:</b>		
NOE distances > 0.5 $\text{\AA}$	0.10 ( $\pm$ 0.30)	0
NOE distances > 0.3 $\text{\AA}$	10.9 ( $\pm$ 2.3)	7
NOE distances > 0.1 $\text{\AA}$	103.5 ( $\pm$ 4.4)	104
H-bond distances > 0.5 $\text{\AA}$	0 ( $\pm$ 0)	0
H-bond distances > 0.5 $\text{\AA}$	0 ( $\pm$ 0)	0
H-bond distances > 0.5 $\text{\AA}$	0.6 ( $\pm$ 0.5)	1
Dihedral angles > 10 $^\circ$	0 ( $\pm$ 0)	0
Dihedral angles > 5 $^\circ$	2.00 ( $\pm$ 0.63)	2
<b>RMSDs from idealized geometry:</b>		
Bond-lengths ( $\text{\AA}$ )	0.004177 ( $\pm$ 0.000084)	0.004083
Bond-angles (deg)	0.567 ( $\pm$ 0.013)	0.558
Improper-angles (deg)	1.726 ( $\pm$ 0.056)	1.760
<b>Ramachandran plot summary (PROCHECK):</b>		
Most favoured	84.5 %	71.3 %
Additionally allowed	13.9 %	24.0 %
Generously allowed	1.1 %	1.3 %
Disallowed	0.5 %	3.3 %

The solution NMR structure of FL-Nsp1 includes a short unstructured N-terminal region (N-terminal tail) (residues 1–9), followed by a folded globular domain (NTD, residues 10–127) and finally a long and largely disordered C-terminal tail (linker + CTD, residues 128–180) (Fig. 1A, PDB: 8AOU). Canonical secondary-structure elements of the folded core of Nsp1 are labeled sequentially as  $\beta$ 1,  $3_{10}$ ,  $\alpha$ 1,  $\beta$ 2,  $\alpha$ 2,  $\beta$ 3,  $\beta$ 4,  $\beta$ 5 and  $\beta$ 6 (Fig. 1B). The globular domain

is defined by the NOEs and the RDCs to a precision of 0.8 Å (backbone RMSD calculated over residues 10–127 for the 10 lowest-energy structures). This domain features an  $\alpha/\beta$  fold formed by three helices and a six-stranded  $\beta$ -barrel motif, which is composed of three pairs of antiparallel  $\beta$ -strands:  $\beta 2$  (51–54) and  $\beta 5$  (104–109);  $\beta 3$  (68–73) and  $\beta 4$  (84–91); and  $\beta 1$  (13–20) and  $\beta 6$  (116–124).  $\alpha 1$  (residues 35–48) acts as a ‘cap’, covering one opening of the barrel, while the  $3_{10}$ -helix (residues 23–25) and  $\alpha 2$  (residues 58–63) are oriented parallel to and alongside the  $\beta$ -barrel. The residues that form the core of the  $\beta$ -barrel are predominantly hydrophobic. The side-chains of residues L16, L18, V69, L88, L107 and L123 form one of the two edging layers of the  $\beta$ -barrel, which is also stabilized by L46 in  $\alpha 1$  donating its side-chain to the center of this layer. Residues L53, I71, V86 and V121 contribute their side-chains to form the central layer of  $\beta$ -barrel, while the other edging layer is comprised of residues V84 and L104 (Fig. 1C).



**Figure 1. Solution NMR structure of Nsp1 from SARS-CoV-2.** (A) Ribbon representation of the final water-refined ensemble of 10 NMR structures of full-length Nsp1. The short N-terminal and long C-terminal tails are largely disordered (colored in turquoise and purple, respectively). The globular NTD, comprising residues 10–127, is colored in gray. Coordinates

have been deposited in the PDB with accession code 8AOU. **(B)** Ribbon representation of the closest-to-the-mean structure for the NTD alone. The  $\beta$ -strands are colored in orange, the three helices are in light-blue, and segments not adopting canonical secondary-structure are in gray. The secondary-structure elements are labeled sequentially. **(C)** Ribbon representation of Nsp1 with the hydrophobic residues that comprise the core of the  $\beta$ -barrel shown as sticks. The three layers are colored in green, red and light brown. L46, which stabilizes the opening layer of  $\beta$ -barrel, is shown in light-blue.

### **Comparison with solution NMR structure of Nsp1<sup>13–127</sup> from SARS-CoV-1 and crystal structures of the globular core-domain of Nsp1 from SARS-CoV-2**

SARS-CoV-2 Nsp1 shows approximately 84% amino-acid sequence identity with its homologous protein from SARS-CoV-1 (Fig. S1A). Tertiary structural alignment of the two globular domains reveals a few differences, for example in the conformations of the major loops (Fig. S1B). However, the conformations of loop 1 and loop 2 are variable in the NMR ensemble of SARS-CoV-2 Nsp1 (Figure 1A and Figure S1C) and also differ between the representative structure of the NMR ensemble and the available crystal structures of SARS-CoV-2 Nsp1 (PDB codes: 7K7P, 7K3N and 7EQ4; Figure S1D and S1E). Thus, we conclude that the differences observed between the conformations of the SARS-CoV-2 and SARS-CoV-1 loops is likely due to the internal flexibility of these loops rather than to differences in the primary sequence.

In the crystal structures of SARS-CoV-2 Nsp1-NTD a one-residue extension of strand  $\beta$ 4 is associated with the appearance of an additional  $\beta$ -strand in loop 2, named  $\beta$ 5 (residues 95 to 97), which together with  $\beta$ 4 forms a short  $\beta$ -hairpin (Clark et al., 2021) (Fig. S1E). The emergence of the extra  $\beta$  sheet may be due to the influence of intermolecular contacts in the crystal lattice; in solution, this loop does not appear to form a canonical  $\beta$ -hairpin.

Comparing the electrostatic surface of SARS-CoV-2 Nsp1-NTD with that of SARS-CoV-1 Nsp1<sup>13–127</sup> shows that the two proteins have similar surface potentials, even in regions where the primary sequences differ the most, such as loop 1 (Figure S2A–C). However, some deviations between the electrostatic surfaces are brought about by the different conformations of the loops 1 and 2. One striking difference caused by the flexibility of loop 2 is the emergence of a negatively charged pocket in SARS-CoV-2 Nsp1, formed by residues S74, E87, G98 and S100 (Figure S2D–F). The presence of this negative pocket is highly dependent on the position of S100 in loop 2. In conformers where the S100 side-chain is far from the side-chains of S74 and E87, this negatively charged pocket disappears (Figure S1C and Figure S2G). In the crystal structure of SARS-CoV-2 Nsp1-NTD, the negative patch is still present but with reduced size.

## Interactions between the Nsp1-NTD and the disordered terminal tails

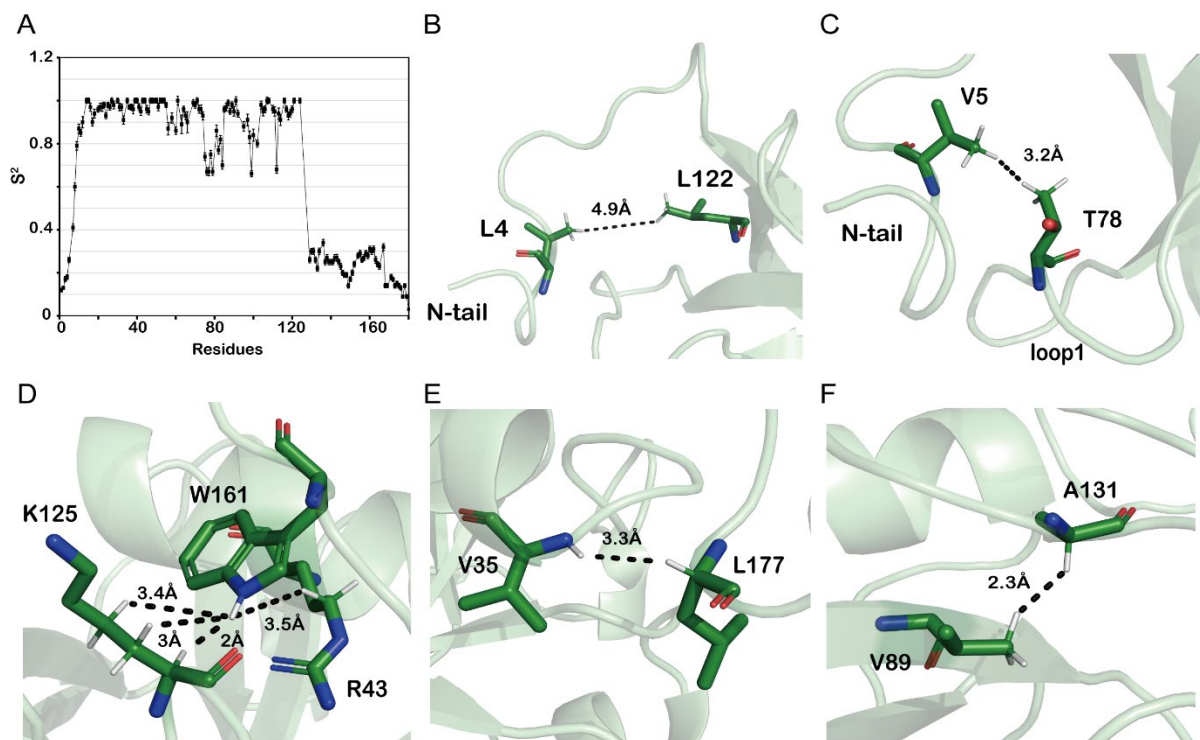
Our initial NMR studies of SARS-CoV-2 Nsp1 yielded spectra whose appearance were characteristic of a protein consisting of a well-folded globular domain followed by an extended, largely disordered C-terminal tail (Wang et al., 2021). To further analyze the dynamics of the terminal regions, we performed the standard set of backbone  $^{15}\text{N}$  relaxation experiments ( $^{15}\text{N}$   $R_1$ ,  $^{15}\text{N}$   $R_2$  ( $R_{1\rho}$ ) and  $\{^1\text{H}\}^{15}\text{N}$  heteronuclear NOE) and analysed the relaxation rates within the Lipari-Szabo Model-Free framework (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b). Figure 2A shows the resulting backbone N–H order parameters ( $S^2$ ), which confirm that the globular domain exhibits only limited internal motion, with the exception of two loops (loop1, between  $\beta_3$  and  $\beta_4$ , and loop2, between  $\beta_4$  and  $\beta_5$ ), whose  $S^2$  values of 0.6–0.7 reveal the presence of enhanced local flexibility relative to the remainder of the core domain. The N-terminal tail and C-terminal tails (residues 1–9 and 128–180) are highly dynamic ( $S^2$  values in the range 0.2–0.3), but some regions of the C-terminal tail (130–144 and 154–163) appear to experience partial restrictions on their local flexibility, which could arise from transient secondary-structure formation or weak interactions with the NTD, or a combination of the two. Interestingly, the backbone chemical shifts indicate a clear helical propensity only for residues 171–176, which correspond to the segment of the CTD that forms the one of the two helices upon interaction with the ribosome (Wang et al., 2021). However, in free Nsp1 in solution, these residues have very low order parameters. The second helix of the CTD formed upon ribosome binding comprises residues 154–163, but the chemical shift analysis does not indicate a clear propensity to form a secondary structure in this region in free Nsp1 (Wang et al., 2021). Thus, the higher order parameters shown in stretches 130–144 and 154–163 are likely due to intramolecular interactions.

Interactions between the tails and the globular domain were revealed in the NOESY spectra. Long-range NOEs were detected from the short N-terminal tail to the globular core-domain, between the side-chains of L4 and L122 (Figure 2B), and between V5 and T78 and the stretch 78–80 of loop1; Figure 2C). Similarly, we could also detect three unambiguous NOEs between the long flexible C-terminal tail and the globular NTD. For example, residue W161 experiences contacts with residues R43 and K125, the latter of which belongs to the conserved eight-amino-acid sequence (LRKxGxKG) that was shown to be important for the RNA-binding activity of Nsp1 proteins (Figure 2D) (Almeida et al., 2007; Lokugamage et al., 2012; Tanaka et al., 2012; Terada et al., 2017). Residue L177, close to the very end of the CTD, contacts residue V35 of the core domain (Figure 2E). Similarly, another long-range NOE was observed between the side-chains of the NTD V89 and A131 in the linker region (Figure 2F). In addition to these unambiguous NOEs, we could detect numerous ambiguous NOEs with contributions from proton-pairs that correspond to long-range contacts between the tails and the globular



core. All these NOEs together indicate that the mostly disordered C-terminal tail of Nsp1 tends to fall back onto the surface of the globular NTD.

Despite the presence of these contacts between the NTD and the N- and C-terminal tails of Nsp1, the structural ensemble clearly shows that neither the secondary nor the tertiary structures of the short N-terminal and long C-terminal tail are well-defined. This is in agreement with the  $^{15}\text{N}$  relaxation data, which indicate that the two tails are flexible. The overall picture that emerges is of a C-terminal tail with significant local flexibility but some propensity to engage in interactions with the globular NTD. Interestingly, the two helices at the very end of the C-terminal tail that are observed in the cryo-EM structure of Nsp1 interacting with 40S ribosomal subunit (CTD) appear to be mostly disordered in free Nsp1 and tend to fall back onto the surface of the Nsp1-NTD.



**Figure 2. The C-terminal tail is dynamic but interacts with the NTD. (A)** Order parameters ( $S^2$ ) for backbone amide N–H groups derived from  $^{15}\text{N}$  relaxation data. An N–H group that is totally rigid in the molecular frame has an  $S^2$  value of 1, while a group with completely unrestricted motion has an  $S^2$  value of 0. **(B–C)** Long-range NOE contacts (black dashed lines) between the short N-terminal tail and the NTD. **(D–F)** Long-range NOE contacts between the NTD and the long C-terminal tail. The dashed lines are labeled with the corresponding distances in the closest-to-the-mean structure.

**Free Nsp1 has no affinity for RNAs derived from the 5'UTR region of the SARS-CoV-2 genome**

The 40S ribosomal subunit has been suggested to be critical for mediating association of Nsp1 with the 5'-UTR of viral mRNA (Mendez et al., 2021; Tidu et al., 2021). However, Shi et al have also reported a direct interaction of the SARS-CoV-2 5'-UTR RNA with the Nsp1-NTD (Shi et al., 2020). To test whether Nsp1 is able to bind SARS-CoV-2 5'-UTR RNA or any of its stem-loop elements, we measured  $^1\text{H},^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labeled Nsp1 (200  $\mu\text{M}$ ) in the presence and in the absence of equimolar concentrations of two RNA constructs comprising either stem-loop 1 (SL1) or stem-loops 1-to-4 (SL1–4). No significant changes were observed in the  $^1\text{H},^{15}\text{N}$ -HSQC spectra after addition of the RNAs, suggesting that neither of these RNAs bind Nsp1 in the respective binary mixtures (Figure S3A). We next tested the binding between Nsp1 and the complete 5'-UTR RNA via an electrophoretic mobility shift assay (EMSA). The migration of 0.3, 0.6, 1.25, 2.5  $\mu\text{M}$  of the 5'-UTR RNA through a native gel was monitored after incubation with increasing amounts of Nsp1 (0-, 5-, 10-, 20-, 40-fold excess). No significant binding was observed from this EMSA (Figure S3B), confirming that Nsp1 has no direct binary affinity for the RNAs SL1, SL1–4 or even the full-length 5'-UTR RNA of SARS-CoV-2.

### **Importance of charged surface regions for the potential functional role of Nsp1**

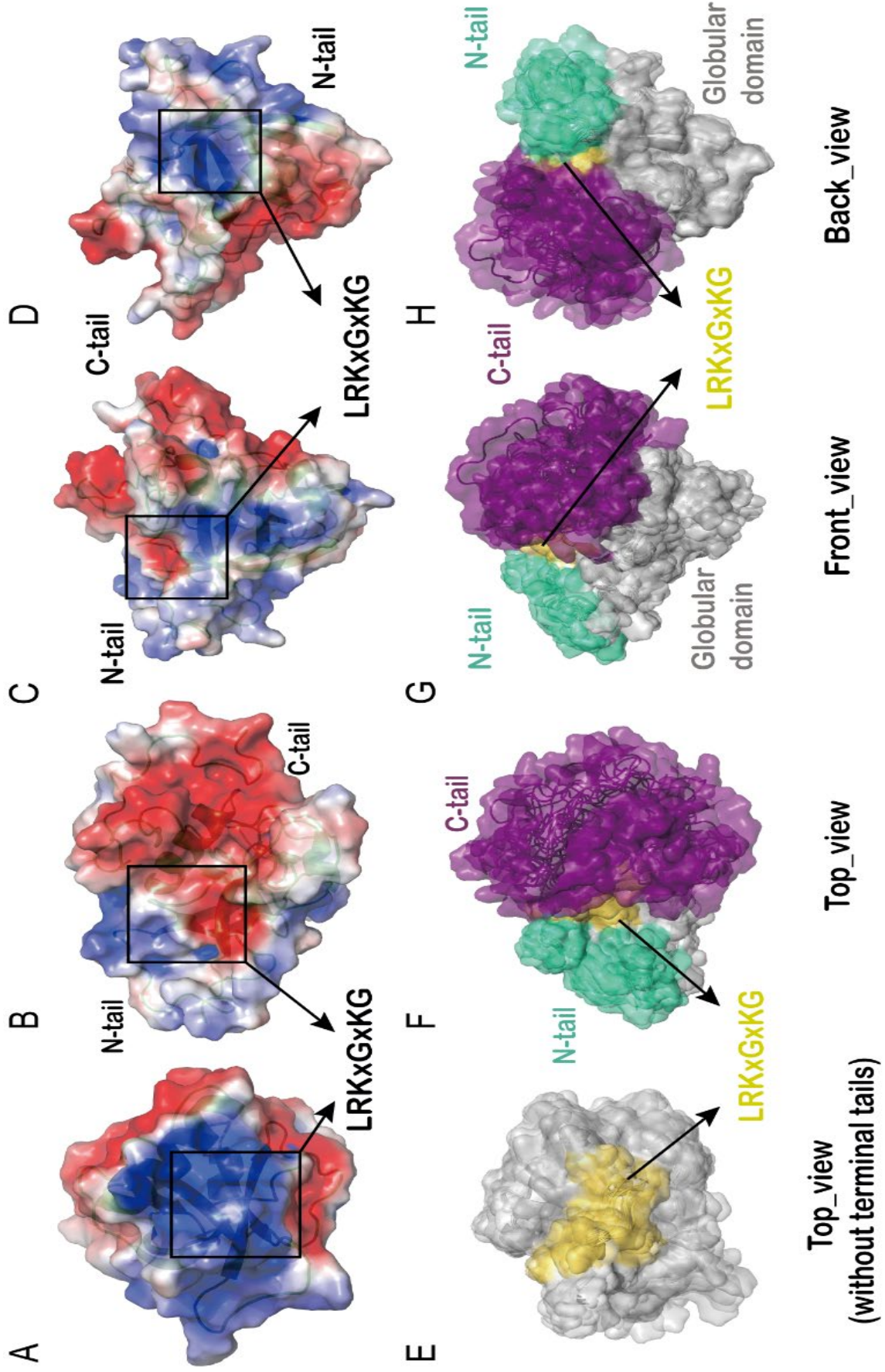
Several studies have indicated that Nsp1 is a major virulence factor for viral pathogenicity of SARS-CoV-2, inhibiting host gene expression by binding to the 40S ribosomal subunit through its CTD and inducing cellular mRNA degradation (Huang et al., 2011; Kamitani et al., 2009; Kamitani et al., 2006; Nakagawa et al., 2016; Tanaka et al., 2012). The host mRNA associates with Nsp1 in a ribosome-dependent manner (Graziadei et al., 2022; Mendez et al., 2021) and this association requires both the NTD and CTD of Nsp1. Recent cryo-electron microscopy (cryo-EM) studies have demonstrated that inhibition of host protein synthesis by Nsp1 is mediated by insertion of the end of the C-terminal tail into the entrance of the mRNA tunnel in the 40S ribosomal subunit (Schubert et al., 2020; Shi et al., 2020; Thoms et al., 2020; Yuan et al., 2020). It therefore appears that two distinct functional domains of Nsp1 are required for SARS-CoV-2 to selectively inhibit host protein production: the CTD, as a general inhibitor of mRNA translation, is necessary for translational shutoff, while the NTD is required for host mRNA binding and induction of host mRNA degradation. The mode of interaction of the host mRNA with the Nsp1 NTD in the Nsp1–40S–mRNA complex is still unknown. In addition, the CTD domain is not the sole region responsible for the recruitment of Nsp1 to the 40S ribosomal subunit, as both the Nsp1-NTD and the linker domain have been shown to reinforce binding (Graziadei et al., 2022; Mendez et al., 2021).

SARS-CoV-2 mRNAs containing the 5'-untranslated region (5'-UTR) are resistant to Nsp1-induced mRNA cleavage as well as to repression of translation (Huang et al., 2011; Lapointe et al., 2021; Schubert et al., 2020; Shi et al., 2020; Tanaka et al., 2012; Tidu et al., 2021). This function of the 5'-UTR of viral mRNA has been shown to depend on the ribosome-mediated

association of the Nsp1 core-domain with stem-loop 1 (SL1) of the 5'-UTR (Banerjee et al., 2020; Shi et al., 2020; Tanaka et al., 2012; Tidu et al., 2021). In addition, a mutant Nsp1, where the linker domain was extended beyond 20 residues failed to support viral translation, emphasizing the importance of this spacer in the regulation of Nsp1 activity (Shi et al., 2020). Despite all these data, the precise mechanism by which SARS-CoV-2 eludes the Nsp1-mediated translation inhibition — thereby switching the host translation machinery from host to viral protein synthesis — and how the association between the Nsp1-NTD and the SARS-CoV-2 5'-UTR supports this mechanism are still unknown.

The consensus amino-acid sequence (LRKxGxKG), located at the end of  $\beta 6$  of the Nsp1-NTD, is conserved among beta-coronaviruses (Fig. S4). Its positively charged amino acids R124 and K125 have been shown to be crucial for the association of Nsp1 with viral 5'-UTR RNA, as well as with host mRNA (Almeida et al., 2007; Lokugamage et al., 2012; Mendez et al., 2021; Nakagawa et al., 2016; Tanaka et al., 2012; Terada et al., 2017). In our structure, the positively charged surface formed by this conserved sequence (Fig. 3A and E) is partially masked by both the N-terminal and C-terminal tails (the solvent-accessible surface-area drops from 515 Å<sup>2</sup> in Nsp1<sup>10-130</sup> to 262 Å<sup>2</sup> in FL-Nsp1, Fig. 3). While the N-terminal tail is predominantly hydrophobic, the long C-terminal tail is highly acidic, so that the electrostatic potential of the accessible surface of this region in FL-Nsp1 is substantially different to that of the masked patch beneath.

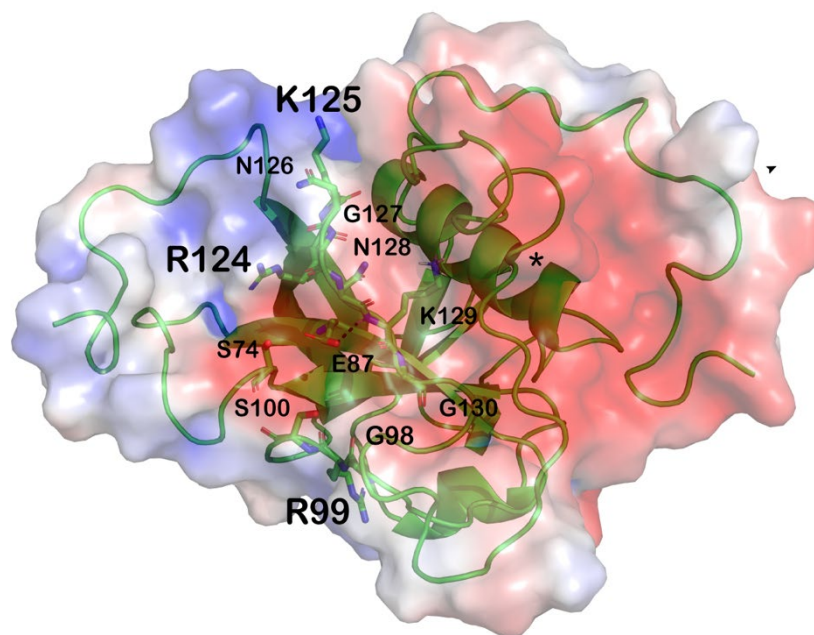
The interaction between the Nsp1-NTD and the C-terminal tail may have functional relevance in two ways. First, it may serve to protect the long disordered region from either degradation or non-specific binding before it encounters the 40S ribosomal complex; the intra-molecular interaction is incompatible with binding of residues 148–180 to the ribosome mRNA entry tunnel and thus needs to be disrupted upon recruitment of Nsp1 to the ribosome. Second, the occlusion of positively charged surface patches on the NTD by the C-terminal tail may function to prevent non-specific RNA recruitment to Nsp1 away from the ribosome; this would explain why free full-length Nsp1 has no binding affinity to RNAs corresponding to SL1, SL1–4 or full-length 5'-UTR of SARS-CoV-2 (Fig. S3), while Shi et al found that the Nsp1-NTD can directly bind to SARS-CoV-2 5'-UTR RNA (Shi et al., 2020). It follows that the engagement of the C-terminal tail by the ribosome would make the RNA-binding surface of Nsp1 more accessible, supporting the model proposed by Mendez et al in which recruitment of Nsp1 by the ribosome is coupled to its recognition of RNA (Mendez et al., 2021).



**Figure 3. Electrostatic surface potential and functional surface regions of full-length SARS CoV-2 Nsp1.** (A) Electrostatic surface potential of the closest-to-the-mean structure of the Nsp1<sup>10–130</sup>, i.e. the surface potential that would be observed in the absence of the N- and C-terminal tails. The positively charged putative RNA-binding site made up of the conserved amino-acid sequence (LRKxGxKG) is indicated by the black square. (B–D) Electrostatic surface potential of the closest-to-the-mean structure of FL-Nsp1 shown from three different viewing-angles. The position of the conserved amino-acid sequence (LRKxGxKG) is indicated as before. (E) Surface/ribbon representation of the structural ensemble of Nsp1<sup>10–130</sup>, with the surface patch corresponding to the eight-residue consensus sequence colored in yellow. (F–H) Surface/ribbon representations of the structural ensemble of FL-Nsp1 shown from three different views. The NTD is colored gray, and the N- and C-terminal tails are colored turquoise and purple, respectively. Here, the yellow surface patch of the consensus sequence visible in (E) is largely covered by the C-terminal tail.

The recognition modes of host mRNA and viral RNA by the Nsp1-NTD remain unknown, as does the explanation of why binding of the viral 5'-UTR RNA to the Nsp1-NTD promotes translation of the viral RNA, while binding of the host mRNA to the Nsp1-NTD instead results in mRNA cleavage. It has been proposed that binding of the viral 5'-UTR to the Nsp1-NTD induces a large conformational change in the linker, which ultimately leads to weakening of the interaction of the CTD domain with the mRNA entry site of the 40S particle (Mendez et al., 2021; Tidu et al., 2021). Our structure supports this hypothesis. Besides R124 and K125, R99 of loop 2 was also found to be important for binding both types of RNA (Mendez et al., 2021), thus identifying the surface formed by the N-terminal end of  $\beta$ 1, the C-terminal end of  $\beta$ 6 and  $\beta$ 3, as well as loop 2 as a potential RNA binding platform (Figure 4). Even if this surface contains a negative patch in free Nsp1 (Figure S2D and Figure 4), the rearrangement of loop 2 expected upon RNA binding, is likely to cause a conformational change that eliminates the negative patch, as readily seen in other conformers of the ensemble (Figure S2G). In free Nsp1, this surface is flanked on one side by the linker region, as confirmed by the unambiguous NOEs seen between V89 and A131 (Figure 2F). In our structure the linker region makes many fuzzy contacts with the Nsp1-NTD and shows reduced flexibility with respect, for example, to the residues at the very end of the C-terminal tail. Depending on the secondary and tertiary structure of the RNA binding to this surface, and the exact mode of the interaction, the linker may be either displaced from its location on the Nsp1-NTD, or, at the minimum, undergo a conformational change. The nature of this displacement and/or conformational change may determine the different functional consequences of host mRNA or viral 5'-UTR RNA binding to the Nsp1-NTD. It is worth noting that, when measuring the weakened binding of mRNA to Nsp1 mutants, deletion of the NTD had similar effects on the association of Nsp1

with both host and viral mRNAs (15-fold and 17-fold reductions in binding, respectively) while the mutations R124A/K125A and R99A had a stronger effect on the association of Nsp1 with viral mRNA (13-fold and 7-fold reductions for the R124A/K125A and R99A mutants, respectively) than with the host mRNA (5-fold and 2-fold reductions for the R124A/K125A and R99A mutants, respectively) (Mendez et al., 2021). This data supports the hypothesis that the binding mode of the RNA to Nsp1, and the consequences of that binding event on the conformation of the linker region, is sequence- and thus structure-dependent. This hypothesis is also in agreement with the observation that not all host mRNAs are subject to Nsp1-induced cleavage (Huang et al., 2011).



**Figure 4. The putative RNA binding surface of SARS-CoV-2 Nsp1 is flanked by the linker region.** Electrostatic potential of the RNA-binding surface defined by residues R99, R124 and R125 (shown in sticks and labeled in large font). Also shown in sticks and labeled in small font are the residues that can form a negative patch, depending on the conformation of loop 2 (S74, E87, G98 and S100) and the residues of the conserved LRKxGxKG motif. An asterisk marks the end of the linker region and the start of the CTD domain.

## Conclusion

We have presented the first structure of full-length Nsp1, which reveals that the two highly disordered terminal tails are prone to fall back onto the surface of the globular core-domain, largely due to long-range and fuzzy electrostatic interactions. As a result, a large portion of the positively charged surface around the key RNA-binding residues R124 and K125 is not completely accessible in free Nsp1. This provides an explanation for the proposed coupling of RNA-binding by the NTD with ribosome-binding of the CTD, which necessarily disrupts the

interaction of the NTD with the C-terminal tail. Furthermore, the linker region between the NTD and CTD delineates the RNA-binding surface, so that binding of RNA is likely to significantly perturb the linker region, which may in turn influence the CTD and its interaction with the ribosome. The precise nature of the perturbation would depend on what type of RNA binds to Nsp1, rationalizing the different fates for host mRNA and viral RNA. As ribosome-dependent binding of RNA to Nsp1 is essential for viral replication, the surface region of Nsp1 containing R124 and K125 (and also R99) is an attractive drug target. Our study shows that in free full-length Nsp1, and probably also in the ribosome-bound state, the nature of the accessible surface in this key region is different from the corresponding surface of the isolated globular domain. The structure determined here provides important insights to guide drug-design efforts that target the RNA-binding function of Nsp1.

## Methods

### Construct design and protein preparation

The construct design, protein expression and purification were carried out as described previously (Wang et al., 2021).

### RNA transcription and purification

All RNA sequences were derived from the 5'-UTR region of the SARS-CoV-2 genome (Wacker et al., 2020) and comprised either the entire 5'-UTR or stem-loops 1 to 4 (SL1–4) or stem-loop 1 (SL1). The exact sequences are given below. All RNAs were generated by *in-vitro* transcription.

RNAs	# nt	Sequence 5' to 3'
SL1	29	GGUUUUAUACCUUCCCAGGUAACAAACCC
SL1–4	119	GGUUUUAUACCUUCCCAGGUAACAAACCAACCAACUUCGAUCUC UUGUAGAUCUGUUCUCUAAACGAACUUUAAAUCUGUGGGCU GUCACUCGGCUGCAUGCUUAGUGCACUCACGC
5'-UTR	472	AUUAAAGGUUUUAUACCUUCCCAGGUAACAAACCAACCAACUUC GAUCUCUUGUAGAUCUGUUCUCUAAACGAACUUUAAAUCUGU GUGGCUGUCACUCGGCUGCAUGCUUAGUGCACUCACGCAGUAU AAUUAUAACUAAUUACUGUCGUUGACAGGACACGAGUAACUCG UCUAUCUUCUGCAGGCUGCUUACGGUUUCGUCCGUGUUGCAG CCGAUCAUCAGCACAUUAGGUUUCGUCCGGGUGUGACCGAAA GGUAAGAUGGAGAGCCUUGUCCCUGGUUUAACGAGAAAACAC ACGUCCAACUCAGUUUGCCUGUUUUACAGGUUCGCGACGUGCU CGUACGUGGCUUUGGAGACUCCGUGGAGGAGGUCUUAUCAGA GGCACGUCAACAUCUUAAGAUGGCACUUGUGGCUUAGUAGAA GUUGAAAAGGCGUUUUGCCUCAACUUGAACAGCCCUAUGUG

Plasmids containing DNA templates were transformed into *E. coli* Top10 cells; transformed cells were grown in LB medium at 37 °C overnight and harvested by centrifugation at 6000 rpm and 4 °C for 20 minutes. Plasmids were extracted using the Qiagen Plasmid Mega Kit (Qiagen) according to the manufacturer's instructions and linearized with HindIII (NEB) before *in-vitro* transcription by T7 RNA polymerase (prepared in-house). Plasmid DNAs were then purified by phenol/chloroform/isoamyl alcohol and chloroform/isoamyl alcohol (Carl Roth) extraction and precipitated by pure ethanol and 5 M NaCl.

For the transcription of each RNA construct, the concentrations of DNA, nucleoside triphosphates (NTPs, Carl Roth), MgCl<sub>2</sub> and T7 polymerase were optimized to maximize the



yield. Large-scale transcription reactions were run for 4 to 6 hours at 37 °C. RNAs were purified using preparative, denaturing polyacrylamide gels containing 7 M urea, except for the 5'-UTR, which was purified directly after transcription by size-exclusion chromatography (Superdex S200 16/60) to preserve its fold. Purity was verified using analytical denaturing polyacrylamide gels.

### Electrophoretic mobility shift assays

Samples for EMSA were prepared by diluting the appropriate volumes of concentrated stock solutions (table below) of 5'-UTR (7.7  $\mu$ M) and Nsp1 (38 and 152  $\mu$ M) in EMSA buffer (50 mM sodium phosphate (pH 6.5), 50 mM NaCl, 2 mM EDTA, 2 mM DTT; prepared with nuclease-free water) to a total volume of 10  $\mu$ L, followed by incubation for 30 mins at room-temperature. Four different concentrations of 5'-UTR were used (0.3, 0.6, 1.25 & 2.5  $\mu$ M), with five samples prepared at each 5'-UTR concentration (corresponding to 0-, 5-, 10-, 20- and 40-fold excess Nsp1), giving 20 samples in total. 2.5  $\mu$ L 5x loading-dye (50 mM Tris-HCl (pH 7.5), 0.25% xylene cyanol, 0.25% bromophenol blue, 30% glycerol) was added to each sample just prior to loading onto a 5% polyacrylamide gel, which had been pre-run for 1 h with TBE running-buffer (100 mM Tris-HCl (pH 8.3), 100 mM boric acid, 2 mM EDTA). After sample-loading, the gel was run overnight (with the same running-buffer) at 4 °C and a current of 2 mA. The gel was stained in a solution of ethidium bromide (0.25  $\mu$ g/ml, 2 mins incubation), washed three times in water and then visualized using a Gel Doc XR+ gel documentation system (Bio-Rad).

5'-UTR (7.7 $\mu$ M stock)		Nsp1 added ( $\mu$ L)					Nsp1 stock ( $\mu$ M)
$\mu$ M	Add ( $\mu$ L)	0	5-fold	10-fold	20-fold	40-fold	
0.3	0.4	0	0.4	0.8	1.6	3.2	38
0.6	0.8	0	0.8	1.6	3.2	6.4	38
1.25	1.65	0	0.4	0.8	1.6	3.2	152
2.5	3.3	0	0.8	1.6	3.2	6.4	152

### NMR data acquisition

**Sample conditions.** Except where specified otherwise, NMR spectra were collected on uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labelled Nsp1 samples at concentrations of 500–700  $\mu$ M, dissolved in NMR buffer (50 mM sodium phosphate (pH 6.5), 200 mM sodium chloride, 2 mM dithiothreitol, 2

mM ethylene-diamine-tetra-acetic acid, 0.01% w/v sodium azide, 0.001% w/v 3-(trimethylsilyl)propane-1-sulfonate, 10% v/v D<sub>2</sub>O), and loaded into 3-mm NMR tubes (sample volume ~180  $\mu$ L). All NMR spectra were measured at a temperature of 298 K.

**Instrumentation.** All spectra were measured on Bruker AVIII-HD spectrometers running Bruker Topspin software (v3.2), operating at <sup>1</sup>H field-strengths of 600 MHz and 850 MHz, and equipped with inverse HCN CP-TCI (helium-cooled) and CPP-TCI (nitrogen-cooled) cryogenic probeheads, respectively.

**Assignment experiments.** 2D <sup>15</sup>N-HSQC spectra were recorded using States-TPPI for frequency discrimination, with water suppression achieved via a combination of WATERGATE and water flip-back pulses to preserve the water magnetization (Bodenhausen and Ruben, 1980; Piotto et al., 1992). Backbone resonance assignments were obtained as described previously (Wang et al., 2021). 2D <sup>1</sup>H-<sup>13</sup>C correlation spectra for sidechain assignment purposes were recorded as <sup>13</sup>C-HSQCs with both real-time and constant-time <sup>13</sup>C chemical-shift evolution periods (Santoro and King, 1992; Vuister and Bax, 1992). Separate constant-time <sup>13</sup>C-HSQC spectra were recorded for aliphatic and aromatic CH groups, with the length of the constant-time period optimized for the respective <sup>1</sup>J<sub>CC</sub> coupling constants. Sidechain assignments were obtained from the following set of experiments: 3D aliphatic HC(C)H-TOCSY (Bax et al., 1990; Kay et al., 1993), 3D aromatic HC(C)H-TOCSY, 3D aromatic (H)CCH-TOCSY, 3D H(CCCO)NH (Logan et al., 1992; Montelione et al., 1992), 3D HA(CACO)NH and 3D HBHA(CBCACO)NH, 2D (HB)CB(CD)HD and 2D (HB)CB(CDCE)HDHE (Yamazaki et al., 1993) (all recorded at 600 MHz, with the exception of the aliphatic HC(C)H-TOCSY, which was recorded at 850 MHz). Stereospecific assignments of Leu and Val methyl groups were obtained from a constant-time <sup>13</sup>C-HSQC spectrum (850 MHz) recorded on a sample prepared from cells grown in minimal medium containing a mixture of 10% <sup>13</sup>C-glucose and 90% <sup>12</sup>C-glucose (Neri et al., 1989).

**Structural restraint data.** Distance restraints for structure calculation were extracted from 3D NOESY-<sup>15</sup>N-HSQC (Fesik and Zuiderweg, 1988; Marion et al., 1989) and 3D NOESY-<sup>13</sup>C-HSQC spectra (Ikura et al., 1990; Muhandiram et al., 1993), both recorded at 850 MHz. The NOESY mixing-time was 100 ms. Hydrogen-bond restraints were partly inferred from a CLEANEX experiment (Hwang et al., 1997), recorded at 850 MHz as a pseudo-3D spectrum with incrementation of the CLEANEX mixing-time. Backbone amide N-H residual-dipolar-coupling (RDC) restraints were extracted from two pairs of spectra, both recorded under isotropic (unaligned) and anisotropic (aligned) conditions at 850 MHz using the same sample of <sup>15</sup>N-labelled Nsp1: (1) <sup>15</sup>N-HSQC and <sup>15</sup>N-TROSY-HSQC spectra; and (2) the upfield/downfield sub-spectra generated from an IPAP-<sup>15</sup>N-HSQC experiment (Ottiger et al.,

1998). Anisotropic conditions were achieved by addition of *Pf1* filamentous bacteriophage (ASLA Biotech, Latvia) to a final concentration of 13.5 mg/ml. The formation and homogeneity of the anisotropic phase was confirmed by inspection of the D<sub>2</sub>O <sup>2</sup>H spectrum, which showed a well-resolved doublet with a splitting of 13.0 Hz. The <sup>15</sup>N-HSQC spectrum was recorded using semi-constant-time <sup>15</sup>N chemical-shift-evolution together with <sup>1</sup>H CPD (WALTZ-16) (Ottiger and Bax, 1998). The <sup>15</sup>N-TROSY-HSQC and IPAP-<sup>15</sup>N-HSQC experiments were recorded with H<sup>α</sup>/H<sup>β</sup> band-selective decoupling for <sup>15</sup>N chemical-shift evolution (Yao et al., 2009).

**Relaxation experiments.** Backbone <sup>15</sup>N relaxation rates ( $R_1$  and  $R_{1\rho}$ ) were measured at 600 MHz using established proton-detected experiments based on a gradient-selected, sensitivity-enhanced, refocused <sup>15</sup>N-HSQC sequence (Dayie and Wagner, 1994; Farrow et al., 1994). The water signal was preserved using selective water pulses and weak bipolar gradients during the indirect chemical shift evolution time to maintain the water magnetization along the +z axis. The relaxation delays were varied between 10 ms and 1.2 s for the  $R_1$  experiment, and between 3 ms and 120 ms for the  $R_{1\rho}$  experiment. In the  $R_1$  sequence, N–H cross-relaxation pathways were suppressed by application of <sup>1</sup>H amide-selective IBURP-1 inversion pulses (Geen and Freeman, 1991) at intervals of 5 ms during the relaxation delay. In the  $R_{1\rho}$  sequence, cross-relaxation was suppressed by application of between one and four <sup>1</sup>H amide-selective inversion pulses during the <sup>15</sup>N spin-lock relaxation period (Korzhnev et al., 2002), and the <sup>15</sup>N magnetization was explicitly aligned with the spin-lock field (Hansen and Kay, 2007), which was applied at a <sup>15</sup>N field-strength of 2.5 kHz. Backbone {<sup>1</sup>H}<sup>15</sup>N steady-state heteronuclear NOEs were measured using the standard method (LI and Montelione, 1993). The reference and saturated spectra were recorded in an interleaved fashion, with an inter-scan recycle delay of 6 s for both spectra. Water magnetization was preserved in the reference spectrum as described above. Saturation of the amide proton magnetization was achieved using a train of high-power 180° pulses applied at 10.9-ms intervals for the duration of the inter-scan recycle delay (Ferrage et al., 2009). An inter-increment delay of 18 s was added to ensure full recovery of the water magnetization at the start of each increment of the reference experiment.

### **NMR data processing and analysis**

All NMR spectra were processed/visualized with a combination of the software packages NMRPipe v.10.1 (Delaglio et al., 1995) and CcpNmr Analysis v2.4 (Vranken et al., 2005).

**CLEANEX spectrum.** The final CLEANEX spectrum that was inspected in the process of deriving hydrogen-bond restraints was generated as a 2D projection over the subset of six 2D sub-spectra with mixing-times of 30.41, 40.55, 50.69, 60.82, 81.10 & 101.38 ms.

**Extraction of RDCs.** Two sets of backbone amide N–H RDCs,  $D_{\text{NH}}^{\text{HSQC/TROSY}}$  and  $D_{\text{NH}}^{\text{IPAP}}$ , were calculated from the HSQC/TROSY and IPAP spectra, respectively, according to:

$$D_{\text{NH}}^{\text{HSQC/TROSY}} = 2\nu_{\text{N}}[(\delta_{\text{N}}^{\text{HSQC,anisotropic}} - \delta_{\text{N}}^{\text{TROSY,anisotropic}}) - (\delta_{\text{N}}^{\text{HSQC,isotropic}} - \delta_{\text{N}}^{\text{TROSY,isotropic}})]$$

$$D_{\text{NH}}^{\text{IPAP}} = \nu_{\text{N}}[(\delta_{\text{N}}^{\text{upfield,anisotropic}} - \delta_{\text{N}}^{\text{downfield,anisotropic}}) - (\delta_{\text{N}}^{\text{upfield,isotropic}} - \delta_{\text{N}}^{\text{downfield,isotropic}})]$$

where the various  $\delta_{\text{N}}$  are the  $^{15}\text{N}$  chemical-shift positions of the peaks in the four sub-spectra indicated in the respective superscript labels ( $\delta_{\text{N}}^{\text{upfield}} < \delta_{\text{N}}^{\text{downfield}}$ ) and  $\nu_{\text{N}}$  is the  $^{15}\text{N}$  Larmor frequency (–86.151 MHz). Estimated errors in the individual RDCs were calculated based on the following empirical relationship for the estimated error in the frequency position of any individual peak:

$$\sigma[\nu_{\text{N}}\delta_{\text{N}}] = 0.5 \times \text{FWHH}/\text{SINO}$$

where FWHH and SINO are the full-width at half-height (in Hz) and the signal-to-noise of the peak, respectively. RDCs were not calculated for amides where any of the peaks in the four sub-spectra were heavily overlapped, and the estimated error was doubled for amides where any of the peaks were partially overlapped. Final raw RDCs were calculated as averages of the  $D_{\text{NH}}^{\text{HSQC/TROSY}}$  and  $D_{\text{NH}}^{\text{IPAP}}$  values where both were available or taken as one or the other when only one value was available. Where the final raw RDC was calculated as the average of the two values, the corresponding error was set either as the standard-deviation of the two values if this was larger than both of the estimated errors for the individual values, or otherwise as the average of the estimated errors for the individual values. Scaled RDCs used for fitting of the alignment tensor and as restraints for structure calculation were generated by scaling the raw values by the inverse of the backbone order-parameter,  $S_{\text{NH}}$ , as derived from analysis of the relaxation data (errors were scaled accordingly, and a lower-limit of 0.25 Hz was imposed on the final set of errors for the scaled RDCs). Alignment-tensor parameters were obtained by SVD fitting (Losonczi et al., 1999) of the scaled RDCs with the software package PALES (Zweckstetter and Bax, 2000); RDCs from amides with order parameters below 0.95 were excluded from the fit.

**Relaxation data analysis.** Relaxation spectra were processed with partial Lorentzian-to-Gaussian apodization in both frequency dimensions and limited linear-prediction in the  $^{15}\text{N}$  dimension, and the corresponding peak intensities quantified by lineshape-fitting with the software package FuDA (D. Flemming Hansen; <https://www.ucl.ac.uk/hansen-lab/fuda/>).  $R_1$  and  $R_{1\rho}$  rate-constants were calculated by fitting the intensity profiles to mono-exponential decay functions, also with FuDA, and  $\{^1\text{H}\}^{15}\text{N}$  heteronuclear NOEs (hNOEs) were calculated as the ratios of the peak-intensities in the saturated and reference sub-spectra. For

subsequent analysis, the pure transverse relaxation rate-constants,  $R_2$ , were calculated from the following equation:

$$R_{1\rho} = R_1 \cos^2 \theta + R_2 \sin^2 \theta$$

$\theta$  is the tip-angle of the magnetization-vector, given by:

$$\theta = \arctan(v_1 / \Delta\nu)$$

where  $v_1$  is the field-strength of the spin-lock and  $\Delta\nu$  is the offset relative to the frequency of the transmitter (both in Hz).

The three relaxation parameters —  $R_1$ ,  $R_2$  &  $\{^1\text{H}\}^{15}\text{N}$  hNOE — were analysed within the Model-Free framework, as implemented in the software package TENSOR2 (Dosset et al., 2000). Briefly, in the first step, the relaxation parameters for a subset of backbone amides with restricted internal mobility and negligible residual exchange-contributions to  $R_2$  are used to determine a global anisotropic rotational diffusion tensor. The second step yields the parameters describing the internal mobility for all backbone amides by fitting the relaxation parameters to the Model-Free (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b) and Extended Model-Free (Clare et al., 1990) spectral-density functions, using the diffusion-tensor parameters determined in the first step. For the first step, amides with restricted internal mobility were defined according to the hNOE; only amides with hNOE values higher than 0.62 were retained. Amides with residual exchange-contributions to  $R_2$  were identified from within the subset of those with restricted internal mobility by reference to the exchange-indicator statistic,  $Ex^{\text{Ind}}$ , calculated as:

$$Ex^{\text{Ind}} = \frac{\langle T_2 \rangle - T_2}{\langle T_2 \rangle} - \frac{T_1 - \langle T_1 \rangle}{\langle T_1 \rangle}$$

where  $T_1 = 1/R_1$  and  $T_2 = 1/R_2$ . Amides for which  $Ex^{\text{Ind}} < \sigma[Ex^{\text{Ind}}]$ , where  $\sigma[Ex^{\text{Ind}}]$  is the standard deviation of the exchange-indicator statistic over the considered subset, were considered to have negligible exchange-contributions to  $R_2$  and retained for diffusion-tensor fitting. The backbone order-parameter used for filtering the RDC data,  $S_{\text{NH}}$ , was calculated according to:

$$S_{\text{NH}} = \sqrt{S_{\text{NH,fast}}^2 S_{\text{NH,slow}}^2}$$

where  $S_{\text{NH,fast}}^2$  and  $S_{\text{NH,slow}}^2$  are the squares of the order-parameters for fast and slow internal motions derived from the analysis of the relaxation data.

## Structure calculation

Structures were calculated with the software package ARIA v2.3 (Rieping et al., 2007), which uses CNS v1.21 (Brunger, 2007) for simulated-annealing molecular-dynamics. All restraint tables and NOESY peak-lists have been included in the BMRB submission (accession number 34748).

**Distance restraints.** Distance restraints were generated from NOE peak-heights using the ARIA internal relaxation-matrix calibration routine to account for the effects of spin-diffusion. The majority of NOE cross-peaks in the NOESY-<sup>15</sup>N-HSQC and NOESY-<sup>13</sup>C-HSQC spectra were manually assigned in the two HSQC dimensions; the NOESY dimension was left unassigned for automatic assignment by ARIA. Some heavily overlapped peaks were left unassigned in all dimensions. For each NOESY spectrum, the cross-peaks were divided into three lists: (1) peaks for which the HSQC assignments were to residues in the globular domain; (2) peaks for which the HSQC assignments were to residues in the flexible tails; and (3) peaks that were unassigned in all dimensions. For this purpose, the globular domain was taken to comprise residues 8–127, with the remaining residues (0–7 and 128–180) constituting the flexible tails. Each peak-list was converted to distance restraints using a separate calibration; the rotational correlation times required for the calibrations were set to 13.5, 4.5 and 9.0 ns for peak-lists (1), (2) and (3), respectively. The peak-heights in peak-lists (1) and (2) from the NOESY-<sup>15</sup>N-HSQC spectrum were pre-scaled prior to the ARIA run according to the peak-heights of the respective peaks in the 2D <sup>15</sup>N-HSQC spectrum:

$$h_{\text{NOE}}^{\text{scaled}} = h_{\text{NOE}} \times \frac{\langle h_{2\text{D-HSQC}} \rangle}{h_{2\text{D-HSQC}}}$$

where  $h_{\text{NOE}}$  and  $h_{\text{NOE}}^{\text{scaled}}$  are the unscaled and scaled NOE peak-heights, respectively,  $h_{2\text{D-HSQC}}$  is the peak-height of the corresponding amide peak in the 2D <sup>15</sup>N-HSQC spectrum and  $\langle h_{2\text{D-HSQC}} \rangle$  is the average peak-height in that spectrum. Peak-heights in peak-list (3) from the NOESY-<sup>15</sup>N-HSQC spectrum and in all three peak-lists from the NOESY-<sup>13</sup>C-HSQC spectrum were unscaled.

**Dihedral-angle restraints.** The backbone dihedral angles  $\phi$  and  $\psi$  were predicted by TALOS-N (Shen and Bax, 2013) using the chemical shifts for the nuclei H<sup>N</sup>, N<sup>H</sup>, H <sup>$\alpha$</sup> , C <sup>$\alpha$</sup> , C <sup>$\beta$</sup>  and C'. Dihedral-angle predictions described as “strong” or “generous” according to the TALOS-N classification system were used as dihedral-angle restraints for the structure calculation. The error-bounds for the restraints were calculated by doubling and tripling the TALOS-N estimated errors, with lower limits of 20° and 30°, for “strong” and “generous” predictions, respectively.

**RDC restraints.** Scaled N–H RDCs were incorporated into the structure calculation as SANI-format restraints. Only RDCs from amides in the globular domain (here defined as residues

7–128) and with backbone order-parameters,  $S_{\text{NH}}$ , greater than 0.80 were used as restraints. The restraints were divided into two classes depending on the respective values of  $S_{\text{NH}}$ : (1)  $S_{\text{NH}} \geq 0.95$ ; and (2)  $0.80 \leq S_{\text{NH}} \leq 0.95$ . The SANI force-constants  $k_{\text{cool1}}/k_{\text{cool2}}$  were set to 0.1/0.5 and 0.05/0.25 for classes (1) and (2), respectively (half and quarter as strong as the respective default ARIA settings). The tensor parameters (magnitude,  $D_a$ , and rhombicity,  $R$ ) for each structure-calculation run were calculated by SVD fitting of the scaled RDCs to the closest-to-the-mean structure of the previous run; the tensor parameters for the final run were:  $D_a = -8.8$  Hz &  $R = 0.20$ .

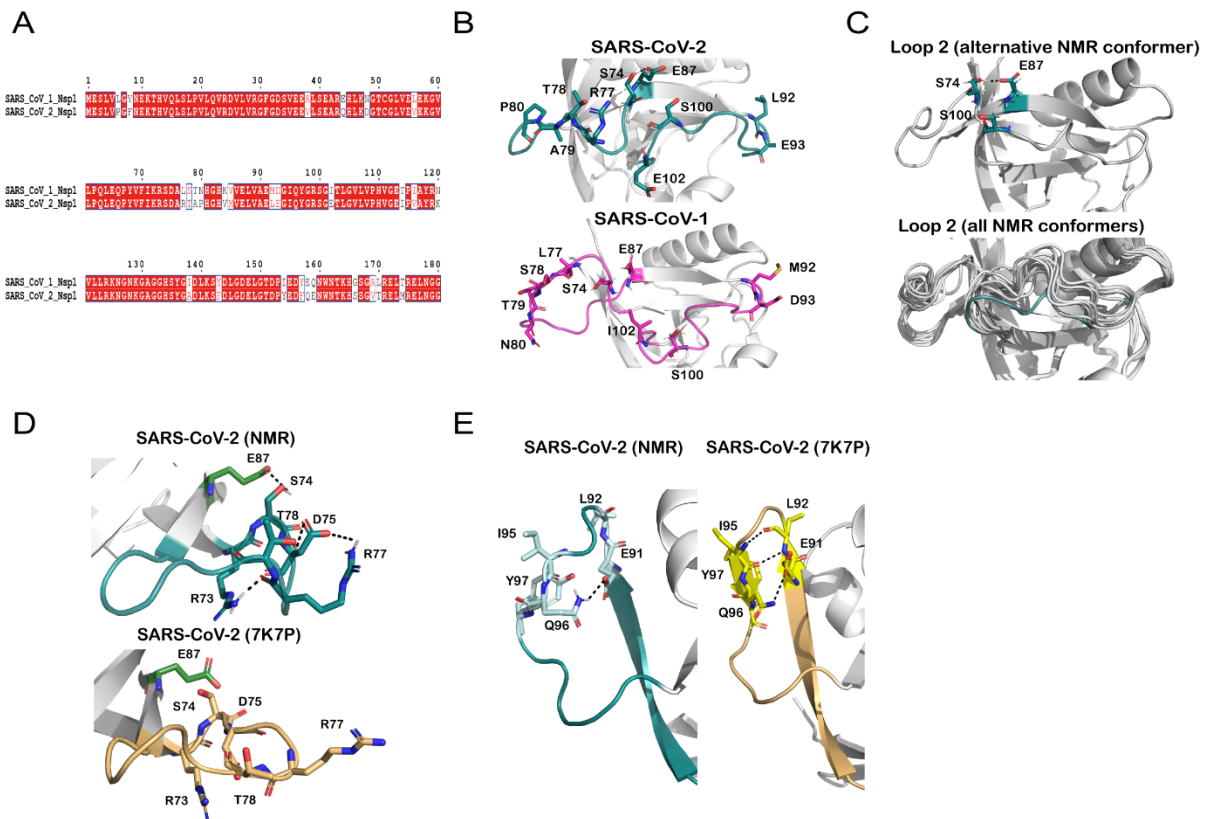
**Hydrogen-bond restraints.** Where appropriate, hydrogen-bond restraints were imposed between backbone amide and backbone carbonyl groups. The final set of hydrogen-bond restraints was derived in an iterative fashion during the structure-calculation process, by careful inspection for the presence of recognisable hydrogen-bonds in interim structures (involving amide and carbonyl groups between which no hydrogen-bond restraints had been imposed in the generation of the interim structure under inspection) and adhering to the criterion that there was no peak in the 2D CLEANEX spectrum for the corresponding amide-group. Linearity of the restraints was imposed for the subset of identified hydrogen-bonds located in regions of canonical secondary-structure.

**Cis-prolines.** The  $^{13}\text{C}$  chemical-shifts of 67P- $\text{C}^\beta$  (downfield-shifted) and 67P- $\text{C}^\gamma$  (upfield-shifted), and the pattern of NOEs between 67P- $\text{H}^\alpha/\text{H}^\delta$  and 66Q- $\text{H}^\beta/\text{H}^\alpha$  were indicative of *cis* geometry for the amide bond between residues 66Q and 67P. Therefore, a *cis*-geometry 66Q–67P amide-bond was imposed during the structure calculation via the standard patch to the force-field parameter-set.

**Protocol settings.** The following adjustments were made to the default ARIA protocol settings:

- The number of calculated structures per iteration were: 50 structures for it0–it5 and 100 structures for it6–it8. Per-iteration violation analysis and calibration were done using the 10 lowest-energy structures from the previous iteration. The 10 lowest-energy structures from it8 were water-refined to generate the final structural ensemble.
- The number of MD steps for each stage of the simulated-annealing were: steps\_high = 30000; steps\_cool1 = 15000; and steps\_cool2 = 12000.
- The maximum numbers of allowed contributions to an ambiguous distance restraint were increased slightly in the early iterations only: from 20 to 30 in it0 & it1, and from 20 to 25 in it2 & it3.
- Chemical-shift matching tolerances for automatic assignment were set to:  $^1\text{H}$ [direct/HSQC dimension] = 0.02 ppm;  $^1\text{H}$ [indirect/NOESY dimension] = 0.04 ppm;  $^{13}\text{C}/^{15}\text{N}$  = 0.2 ppm.

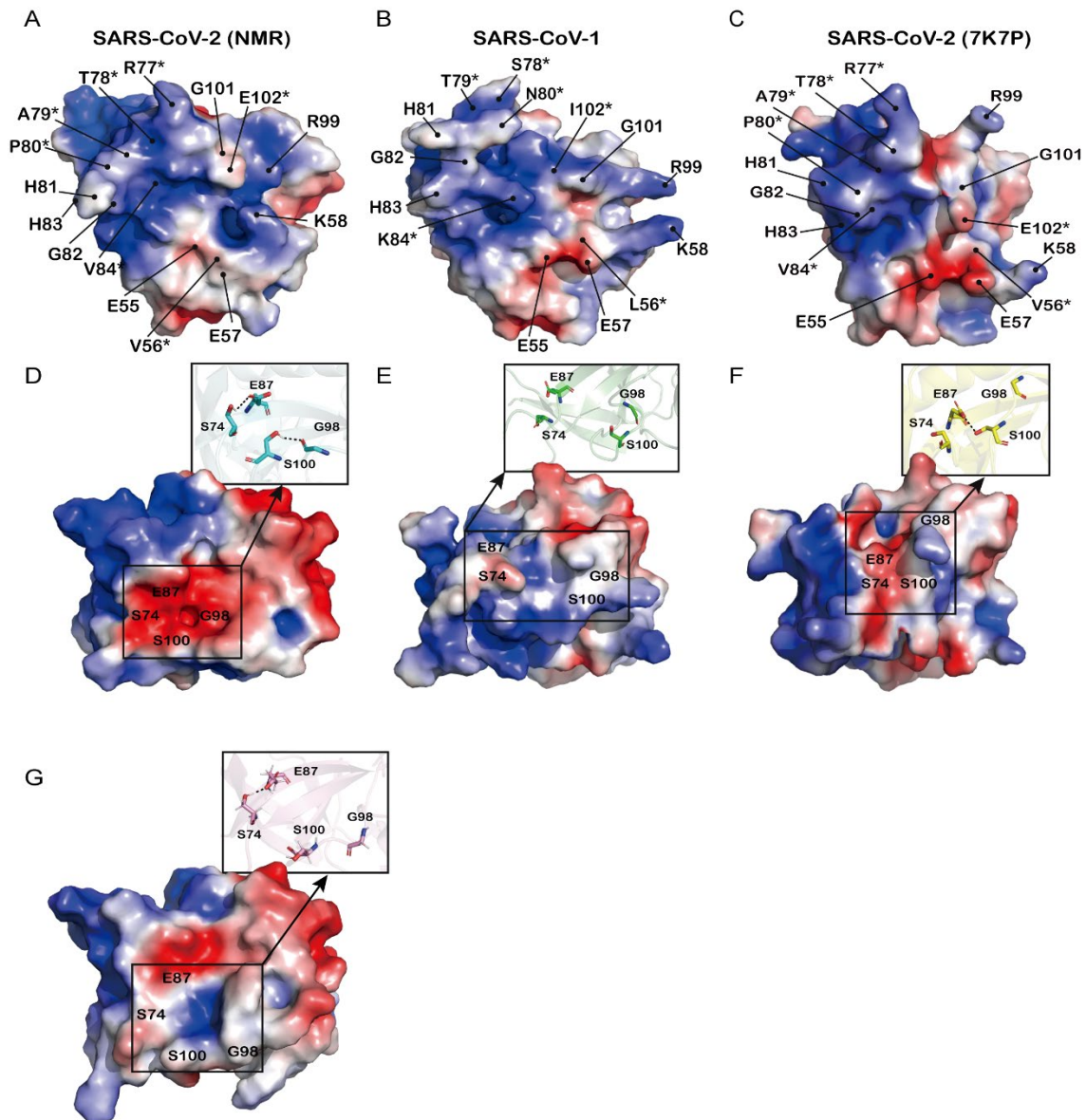
## Supplemental Figures



**Figure S1. Comparison of the solution structure of the NTD of SARS-CoV-2 FL-Nsp1 with the solution structure of SARS-CoV-1 Nsp1-NTD and the crystal structure of SARS-CoV-2 Nsp1-NTD. (A)** Sequence alignment of Nsp1 from SARS-CoV-1 and SARS-CoV-2. The alignment was done with Clustal Omega and the colored figures were generated by ESPript 3.0 (Robert and Gouet, 2014; Sievers et al., 2011). Residues boxed in red are identical. **(B)** Conformational differences in loop1 and loop2 between Nsp1 from SARS-CoV-2 (turquoise, upper panel) and SARS-CoV-1 (magenta, lower panel). Residues that are different between the two species are labeled and shown as sticks. **(C)** Variability of loop 2 conformations in the NMR structure of SARS-CoV-2 FL-Nsp1. The upper panel shows the conformation of loop 2 of one member of the NMR ensemble, where it deviates substantially from that of representative conformer shown in (B). The lower panel shows an overlay of the conformations of loop 2 for all members of the NMR ensemble. **(D)** Conformational differences of loop 1 between the solution NMR structure of SARS-CoV-2 FL-Nsp1 (turquoise, top) and the crystal structure of PDB 7K7P (light-orange, bottom). PDB 7K7P is shown as representative of all available crystal structures of SARS-CoV-2 Nsp1-NTD (PDB codes: 7K7P, 7K3N and 7EQ4). Residue E87 is colored in green in both structures and residues that are involved in the formation of hydrogen-bonds are shown as sticks. Hydrogen-bonds are shown

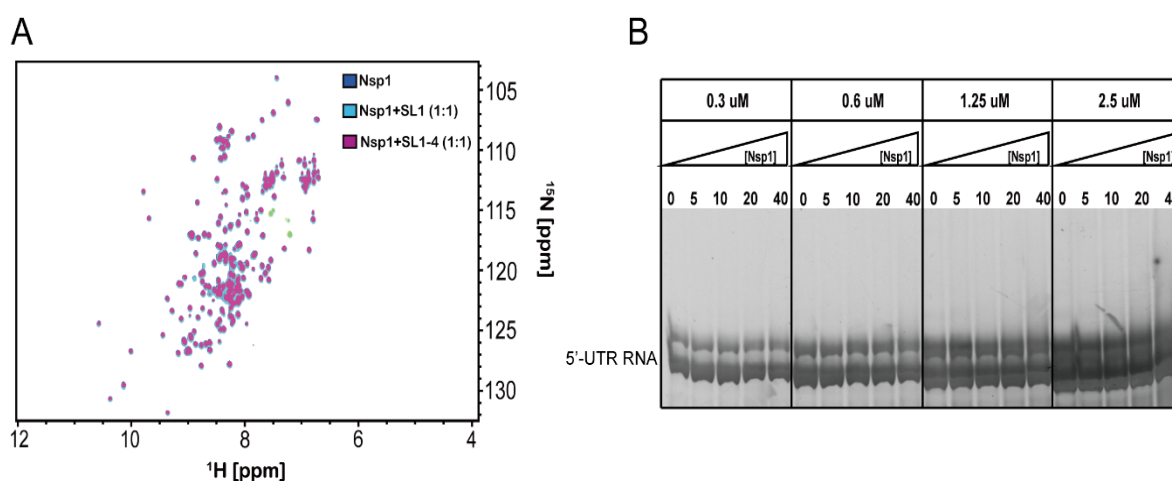


as black dashed lines. (E) The  $\beta$ 4 strand and loop 2 in the NMR structure of SARS-CoV-2 FL-Nsp1 (turquoise, left) and in the crystal structure of PDB code 7K7P (light-orange, right). Residues that are implicated in the formation of a small  $\beta$ -sheet in the crystal structures are shown as sticks in both structures (light-blue and yellow, respectively). Polar contacts and hydrogen bonds are represented as black dashed lines.

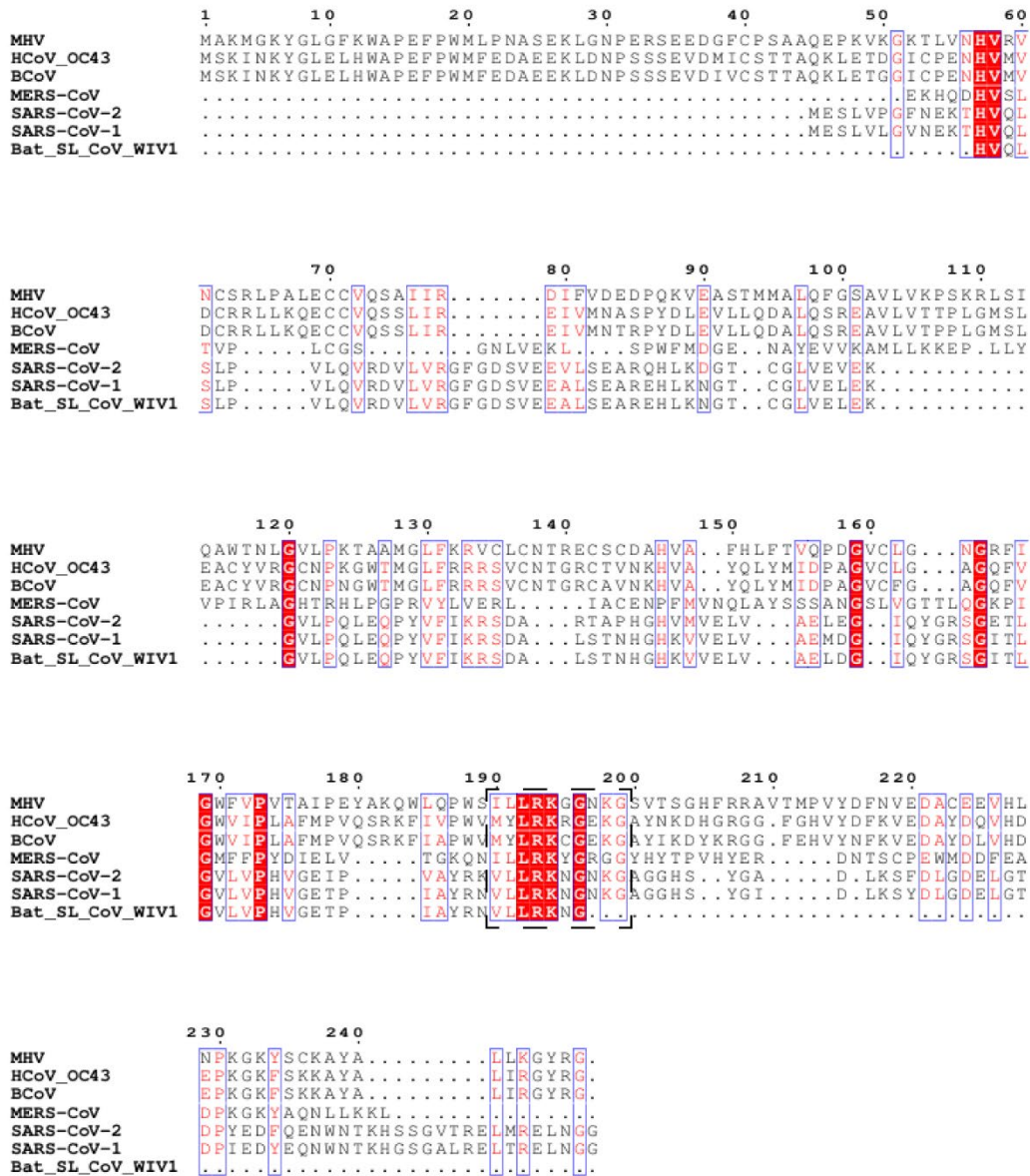


**Figure S2. Comparison of electrostatic surfaces of the solution structure of SARS-CoV-2 Nsp1-NTD (tails removed *in silico*), the solution structure of SARS-CoV-1 Nsp1-NTD and the crystal structure of SARS-CoV-2 Nsp1-NTD. (A–C)** The electrostatic surface is broadly similar for all three structures, despite differences in the amino-acid sequence between the SARS-CoV-1 and SARS-CoV-2 proteins (labeled with asterisks). PDB 7K7P is

shown as representative of all available crystal structures of SARS-CoV-2 Nsp1-NTD. **(D–F)** Emergence of a negatively charged pocket in the solution structure of SARS-CoV-2 Nsp1-NTD. The residues forming the pocket (S74, E87E, G98 and S100) are colored in cyan and green for the solution structures of SARS-CoV-2 and SARS-CoV-1 Nsp1-NTD, respectively, and yellow for the crystal structure of SARS-CoV-2 Nsp1-NTD. Hydrogen bonds are represented as black dashed lines. **(G)** The negative pocket is absent from the NMR conformers of SARS-CoV-2 Nsp1-NTD where the conformation of loop 2 differs from that of the representative structure of the ensemble. The conformer shown here is the same as in Fig. S1C. The sequence boundaries of the NTD domains as shown are as follows: SARS-CoV-2 Nsp1-NTD (NMR), residues 10–128; SARS-CoV-1 Nsp1-NTD, residues 13–127; SARS-CoV-2 Nsp1-NTD (7K7P), residues 10–126.



**Figure S3. SARS-CoV-2 FL-Nsp1 does not show a binary interaction with the 5'-UTR. (A)** Overlay of  $^{15}\text{N}$ -HSQC spectra of 200  $\mu\text{M}$  of  $^{15}\text{N}$ -labelled FL-Nsp1 dissolved in NMR buffer in isolation (dark blue) and upon addition of 200  $\mu\text{M}$  of either SL1 (light-blue) or SL1–4 (purple) RNAs. **(B)** EMSA assays of the SARS-CoV-2 5'-UTR RNA in the presence of increasing stoichiometric ratios of FL-Nsp1. Four different RNA concentrations were tested (0.3, 0.6, 1.2 and 2.4  $\mu\text{M}$ ) in the presence of Nsp1 at 0-, 5-, 10-, 20- and 40-fold excess.



**Figure S4. Sequence alignment of SARS-CoV-2 Nsp1 with other Nsp1 proteins derived from beta-coronaviruses.** Alignments were done with Clustal Omega and the colored figures were generated using ESPrnt 3.0 (Robert and Gouet, 2014; Sievers et al., 2011). Residues boxed in red are identical. The conserved LRKxGxKG motif is indicated with a black dashed box.

## References

- Almeida, M.S., Johnson, M.A., Herrmann, T., Geralt, M., and Wuthrich, K. (2007). Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. *J Virol* *81*, 3151-3161. <https://doi.org/10.1128/JVI.01939-06>.
- Banerjee, A.K., Blanco, M.R., Bruce, E.A., Honson, D.D., Chen, L.M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S.A., Loney, C., et al. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell* *183*, 1325-1339 e1321. <https://doi.org/10.1016/j.cell.2020.10.004>.
- Bax, A., Clore, G.M., and Gronenborn, A.M. (1990).  $^1\text{H}$ – $^1\text{H}$  correlation via isotropic mixing of  $^{13}\text{C}$  magnetization, a new three-dimensional approach for assigning  $^1\text{H}$  and  $^{13}\text{C}$  spectra of  $^{13}\text{C}$ -enriched proteins. *J. Magn. Reson.* *88*, 425-431. [https://doi.org/10.1016/0022-2364\(90\)90202-K](https://doi.org/10.1016/0022-2364(90)90202-K).
- Benedetti, F., Snyder, G.A., Giovanetti, M., Angeletti, S., Gallo, R.C., Ciccozzi, M., and Zella, D. (2020). Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1. *J Transl Med* *18*, 329. <https://doi.org/10.1186/s12967-020-02507-5>.
- Bodenhausen, G., and Ruben, D.J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* *69*, 185-189. [https://doi.org/10.1016/0009-2614\(80\)80041-8](https://doi.org/10.1016/0009-2614(80)80041-8).
- Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* *2*, 2728-2733. <https://doi.org/10.1038/nprot.2007.406>.
- Chan, J.F., Kok, K.H., Zhu, Z., Chu, H., To, K.K., Yuan, S., and Yuen, K.Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* *9*, 221-236. <https://doi.org/10.1080/22221751.2020.1719902>.
- Clark, L.K., Green, T.J., and Petit, C.M. (2021). Structure of nonstructural protein 1 from SARS-CoV-2. *J Virol* *95*, e02019-02020. <https://doi.org/10.1128/JVI.02019-20>.
- Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M. (1990). Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* *112*, 4989-4991. <https://doi.org/10.1021/ja00168a070>.
- Dayie, K., and Wagner, G. (1994). Relaxation-rate measurements for  $^{15}\text{N}$ – $^1\text{H}$  groups with pulsed-field gradients and preservation of coherence pathways. *J. magn. reson., Ser. A* *111*, 121-126. <https://doi.org/10.1006/jmra.1994.1236>.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* *6*, 277-293. <https://doi.org/10.1007/BF00197809>.

- Dosset, P., Hus, J.-C., Blackledge, M., and Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* 16, 23-28. <https://doi.org/10.1023/A:1008305808620>.
- Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994). Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation. *J. Biochem.* 33, 5984-6003. <https://doi.org/10.1021/bi00185a040>.
- Ferrage, F., Cowburn, D., and Ghose, R. (2009). Accurate Sampling of High-Frequency Motions in Proteins by Steady-State <sup>15</sup>N-<sup>1</sup>H Nuclear Overhauser Effect Measurements in the Presence of Cross-Correlated Relaxation. *J Am Chem Soc* 131, 6048-6049. <https://doi.org/10.1021/ja809526q>.
- Fesik, S.W., and Zuiderweg, E.R. (1988). Heteronuclear three-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson.* 78, 588-593. [https://doi.org/10.1016/0022-2364\(88\)90144-8](https://doi.org/10.1016/0022-2364(88)90144-8).
- Geen, H., and Freeman, R. (1991). Band-selective radiofrequency pulses. *J. Magn. Reson.* 93, 93-141. [https://doi.org/10.1016/0022-2364\(91\)90034-Q](https://doi.org/10.1016/0022-2364(91)90034-Q).
- Graziadei, A., Schildhauer, F., Spahn, C., Kraushar, M.L., and Rappilber, J. (2022). SARS-CoV-2 Nsp1 N-terminal and linker regions as a platform for host translational shutoff. *BioRxiv*. <https://doi.org/10.1101/2022.02.10.479924>.
- Hansen, D.F., and Kay, L.E. (2007). Improved magnetization alignment schemes for spin-lock relaxation experiments. *J. Biomol. NMR* 37, 245-255. <https://doi.org/10.1007/s10858-006-9126-6>.
- Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J.M., and Glaunsinger, B.A. (2020). The molecular virology of coronaviruses. *J Biol Chem* 295, 12910-12934. <https://doi.org/10.1074/jbc.REV120.013930>.
- Huang, C., Lokugamage, K.G., Rozovics, J.M., Narayanan, K., Semler, B.L., and Makino, S. (2011). SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. *PLoS Pathog* 7, e1002433. <https://doi.org/10.1371/journal.ppat.1002433>.
- Hwang, T.-L., Mori, S., Shaka, A., and van Zijl, P.C. (1997). Application of phase-modulated CLEAN chemical EXchange spectroscopy (CLEANEX-PM) to detect water-protein proton exchange and intermolecular NOEs. *J Am Chem Soc* 119, 6203-6204. <https://doi.org/10.1021/ja970160j>.
- Ikura, M., Kay, L.E., Tschudin, R., and Bax, A. (1990). Three-dimensional NOESY-HMQC spectroscopy of a <sup>13</sup>C-labeled protein. *J. Magn. Reson.* 86, 204-209. [https://doi.org/10.1016/0022-2364\(90\)90227-Z](https://doi.org/10.1016/0022-2364(90)90227-Z).
- Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K.G., and Makino, S. (2009). A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat Struct Mol Biol* 16, 1134-1140. <https://doi.org/10.1038/nsmb.1680>.

Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., Kubo, H., and Makino, S. (2006). Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc Natl Acad Sci U S A* *103*, 12885-12890. <https://doi.org/10.1073/pnas.0603144103>.

Kay, L.E., Xu, G.-Y., Singer, A.U., Muhandiram, D.R., and FORMAN-KAY, J.D. (1993). A gradient-enhanced HCCH-TOCSY experiment for recording side-chain <sup>1</sup>H and <sup>13</sup>C correlations in H<sub>2</sub>O samples for proteins. *J. magn. reson., Ser. B* *101*, 333-337. <https://doi.org/10.1006/jmrb.1993.1053>.

Khailany, R.A., Safdar, M., and Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2. *Gene Rep* *19*, 100682. <https://doi.org/10.1016/j.genrep.2020.100682>.

Korzhev, D.M., Skrynnikov, N.R., Millet, O., Torchia, D.A., and Kay, L.E. (2002). An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J Am Chem Soc* *124*, 10743-10753. <https://doi.org/10.1021/ja0204776>.

Lapointe, C.P., Grosely, R., Johnson, A.G., Wang, J., Fernandez, I.S., and Puglisi, J.D. (2021). Dynamic competition between SARS-CoV-2 NSP1 and mRNA on the human ribosome inhibits translation initiation. *Proc Natl Acad Sci U S A* *118*. <https://doi.org/10.1073/pnas.2017715118>.

Li, Y.-C., and Montelione, G. (1993). Solvent saturation-transfer effects in pulsed-field-gradient heteronuclear single-quantum-coherence (PFG-HSQC) spectra of polypeptides and proteins. *J. magn. reson., Ser. B* *101*, 315-319. <https://doi.org/10.1006/jmrb.1993.1049>.

Lin, J.W., Tang, C., Wei, H.C., Du, B., Chen, C., Wang, M., Zhou, Y., Yu, M.X., Cheng, L., Kuivanen, S., et al. (2021). Genomic monitoring of SARS-CoV-2 uncovers an Nsp1 deletion variant that modulates type I interferon response. *Cell Host Microbe* *29*, 489-502 e488. <https://doi.org/10.1016/j.chom.2021.01.015>.

Lipari, G., and Szabo, A. (1982a). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* *104*, 4546-4559. <https://doi.org/10.1021/ja00381a009>.

Lipari, G., and Szabo, A. (1982b). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc* *104*, 4559-4570. <https://doi.org/10.1021/ja00381a010>.

Logan, T.M., Olejniczak, E.T., Xu, R.X., and Fesik, S.W. (1992). Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS Lett* *314*, 413-418. [https://doi.org/10.1016/0014-5793\(92\)81517-P](https://doi.org/10.1016/0014-5793(92)81517-P).

Lokugamage, K.G., Narayanan, K., Huang, C., and Makino, S. (2012). Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation. *J Virol* *86*, 13598-13608. <https://doi.org/10.1128/JVI.01958-12>.

- Losonczi, J.A., Andrec, M., Fischer, M.W., and Prestegard, J.H. (1999). Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* *138*, 334-342. <https://doi.org/10.1006/jmre.1999.1754>.
- Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., and Bax, A. (1989). Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J Am Chem Soc* *111*, 1515-1517. <https://doi.org/10.1021/ja00186a066>.
- Masters, P.S. (2006). The Molecular Biology of Coronaviruses. *Adv Virus Res*, 193-292. [https://doi.org/10.1016/s0065-3527\(06\)66005-3](https://doi.org/10.1016/s0065-3527(06)66005-3).
- Mendez, A.S., Ly, M., Gonzalez-Sanchez, A.M., Hartenian, E., Ingolia, N.T., Cate, J.H., and Glaunsinger, B.A. (2021). The N-terminal domain of SARS-CoV-2 nsp1 plays key roles in suppression of cellular gene expression and preservation of viral gene expression. *Cell Rep* *37*, 109841. <https://doi.org/10.1016/j.celrep.2021.109841>.
- Montelione, G.T., Lyons, B.A., Emerson, S.D., and Tashiro, M. (1992). An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J Am Chem Soc* *114*, 10974-10975. <https://doi.org/10.1021/ja00053a051>.
- Muhandiram, D., Farrow, N., XU, G.-Y., Smallcombe, S., and Kay, L. (1993). A gradient 13C NOESY-HSQC experiment for recording NOESY spectra of 13C-labeled proteins dissolved in H2O. *J. magn. reson., Ser. B* *102*, 317-321. <https://doi.org/10.1006/jmrb.1993.1102>.
- Nakagawa, K., Lokugamage, K.G., and Makino, S. (2016). Viral and Cellular mRNA Translation in Coronavirus-Infected Cells. *Adv Virus Res* *96*, 165-192. <https://doi.org/10.1016/bs.aivir.2016.08.001>.
- Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuethrich, K. (1989). Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional carbon-13 labeling. *J. Biochem.* *28*, 7510-7516. <https://doi.org/10.1021/bi00445a003>.
- Ottiger, M., and Bax, A. (1998). Determination of Relative N-HN, N-C', C $\alpha$ -C', and C $\alpha$ -H $\alpha$  Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase. *J Am Chem Soc* *120*, 12334-12341. <https://doi.org/10.1021/ja9826791>.
- Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* *131*, 373-378. <https://doi.org/10.1006/jmre.1998.1361>.
- Paules, C.I., Marston, H.D., and Fauci, A.S. (2020). Coronavirus Infections-More Than Just the Common Cold. *JAMA* *323*, 707-708. <https://doi.org/10.1001/jama.2020.0757>.
- Piotto, M., Saudek, V., and Sklenář, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* *2*, 661-665. <https://doi.org/10.1007/BF02192855>.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., and Nilges, M. (2007). ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23, 381-382. <https://doi.org/10.1093/bioinformatics/btl589>.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42, W320-W324. <https://doi.org/10.1093/nar/gku316>.

Santoro, J., and King, G.C. (1992). A constant-time 2D overbodenhausen experiment for inverse correlation of isotopically enriched species. *J. Magn. Reson.* 97, 202-207. [https://doi.org/10.1016/0022-2364\(92\)90250-B](https://doi.org/10.1016/0022-2364(92)90250-B).

Schubert, K., Karousis, E.D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L.A., Leibundgut, M., Thiel, V., Muhlemann, O., and Ban, N. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol* 27, 959-966. <https://doi.org/10.1038/s41594-020-0511-8>.

Semper, C., Watanabe, N., and Savchenko, A. (2021). Structural characterization of nonstructural protein 1 from SARS-CoV-2. *iScience* 24, 101903. <https://doi.org/10.1016/j.isci.2020.101903>.

Shen, Y., and Bax, A. (2013). Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* 56, 227-241. <https://doi.org/10.1007/s10858-013-9741-y>.

Shi, M., Wang, L., Fontana, P., Vora, S., Zhang, Y., Fu, T.M., Lieberman, J., and Wu, H. (2020). SARS-CoV-2 Nsp1 suppresses host but not viral translation through a bipartite mechanism. *BioRxiv*. <https://doi.org/10.1101/2020.09.18.302901>.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., and Söding, J. (2011). Fast, scalable generation of high - quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. <https://doi.org/10.1038/msb.2011.75>.

Skowronski, D.M., Astell, C., Brunham, R.C., Low, D.E., Petric, M., Roper, R.L., Talbot, P.J., Tam, T., and Babiuk, L. (2005). Severe acute respiratory syndrome (SARS): a year in review. *Annu Rev Med* 56, 357-381. <https://doi.org/10.1146/annurev.med.56.091103.134135>.

Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H.D., and Rappuoli, R. (2003). SARS--beginning to understand a new virus. *Nat Rev Microbiol* 1, 209-218. <https://doi.org/10.1038/nrmicro775>.

Tanaka, T., Kamitani, W., DeDiego, M.L., Enjuanes, L., and Matsuura, Y. (2012). Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA. *J Virol* 86, 11128-11137. <https://doi.org/10.1128/JVI.01700-12>.

Tang, Q., Song, Y., Shi, M., Cheng, Y., Zhang, W., and Xia, X.Q. (2015). Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. *Sci Rep* 5, 17155. <https://doi.org/10.1038/srep17155>.

Terada, Y., Kawachi, K., Matsuura, Y., and Kamitani, W. (2017). MERS coronavirus nsp1 participates in an efficient propagation through a specific interaction with viral RNA. *Virology* 511, 95-105. <https://doi.org/10.1016/j.virol.2017.08.026>.



Thoms, M., Buschauer, R., Ameisemeier, M., Koepke, L., Denk, T., Hirschenberger, M., Kratzat, H., Hayn, M., Mackens-Kiani, T., and Cheng, J. (2020). Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* 369, 1249-1255. <https://doi.org/10.1126/science.abc8665>.

Tidu, A., Janvier, A., Schaeffer, L., Sosnowski, P., Kuhn, L., Hammann, P., Westhof, E., Eriani, G., and Martin, F. (2021). The viral protein NSP1 acts as a ribosome gatekeeper for shutting down host translation and fostering SARS-CoV-2 translation. *Rna* 27, 253-264. <https://doi.org/10.1261/rna.078121.120>.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Struct. Funct. Genet* 59, 687-696. <https://doi.org/10.1002/prot.20449>.

Vuister, G.W., and Bax, A. (1992). Resolution enhancement and spectral editing of uniformly <sup>13</sup>C-enriched proteins by homonuclear broadband <sup>13</sup>C decoupling. *J. Magn. Reson.* 98, 428-435. [https://doi.org/10.1016/0022-2364\(92\)90144-V](https://doi.org/10.1016/0022-2364(92)90144-V).

Wacker, A., Weigand, J.E., Akabayov, S.R., Altincekic, N., Bains, J.K., Banijamali, E., Binas, O., Castillo-Martinez, J., Cetiner, E., Ceylan, B., et al. (2020). Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res* 48, 12415-12435. <https://doi.org/10.1093/nar/gkaa1013>.

Wang, Y., Kirkpatrick, J., Zur Lage, S., Korn, S.M., Neissner, K., Schwalbe, H., Schlundt, A., and Carlomagno, T. (2021). (<sup>1</sup>H, (<sup>13</sup>C, and (<sup>15</sup>N backbone chemical-shift assignments of SARS-CoV-2 non-structural protein 1 (leader protein). *Biomol NMR Assign* 15, 287-295. <https://doi.org/10.1007/s12104-021-10019-6>.

Wathelet, M.G., Orr, M., Frieman, M.B., and Baric, R.S. (2007). Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. *J Virol* 81, 11620-11633. <https://doi.org/10.1128/JVI.00702-07>.

Yamazaki, T., Forman-Kay, J., and Kay, L. (1993). Two-Dimensional NMR Experiments for Correlating <sup>1</sup>H and <sup>13</sup>C / Chemical Shifts of Aromatic Residues in <sup>13</sup>C-Labeled Proteins via Scalar Couplings. *J Am Chem Soc* 115, 11054-11055. <https://doi.org/10.1021/ja00076a099>.

Yao, L., Ying, J., and Bax, A. (2009). Improved accuracy of <sup>15</sup>N-<sup>1</sup>H scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins. *J. Biomol. NMR* 43, 161-170. <https://doi.org/10.1007/s10858-009-9299-x>.

Yuan, S., Peng, L., Park, J.J., Hu, Y., Devarkar, S.C., Dong, M.B., Shen, Q., Wu, S., Chen, S., Lomakin, I.B., and Xiong, Y. (2020). Nonstructural Protein 1 of SARS-CoV-2 Is a Potent Pathogenicity Factor Redirecting Host Protein Synthesis Machinery toward Viral RNA. *Mol Cell* 80, 1055-1066 e1056. <https://doi.org/10.1016/j.molcel.2020.10.034>.

Zhao, K., Ke, Z., Hu, H., Liu, Y., Li, A., Hua, R., Guo, F., Xiao, J., Zhang, Y., Duan, L., et al. (2021). Structural Basis and Function of the N Terminus of SARS-CoV-2 Nonstructural Protein 1. *Microbiol Spectr* 9, e0016921. <https://doi.org/10.1128/Spectrum.00169-21>.

Zweckstetter, M., and Bax, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J Am Chem Soc* 122, 3791-3792. <https://doi.org/10.1021/ja0000908>.

## 2.2 Publication 1

### **<sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N backbone chemical-shift assignments of SARS-CoV-2 non-structural protein 1 (leader protein)**

Ying Wang<sup>1</sup>, John Kirkpatrick<sup>1,2</sup>, Susanne zur Lage<sup>2</sup>, Sophie M. Korn<sup>3,5</sup>, Konstantin Neißner<sup>3,5</sup>, Harald Schwalbe<sup>4,5</sup>, Andreas Schlundt<sup>3,5</sup>, Teresa Carlomagno<sup>1,2\*</sup>

<sup>1</sup>Center of Biomolecular Drug Research (BMWZ), Leibniz University Hannover, Schneiderberg 38, D-30167, Hannover, Germany

<sup>2</sup>Group of NMR-based Structural Chemistry, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig

<sup>3</sup>Institute for Molecular Biosciences, St Lucia, QLD 4072, Australia

<sup>4</sup>Institute for Organic Chemistry and Chemical Biology, 60438 Frankfurt, Germany

<sup>5</sup>Center for Biomolecular Magnetic Resonance (BMRZ), Johann Wolfgang Goethe-University Frankfurt, Max-von-Laue-Str. 7, 60438 Frankfurt, Germany

Type of authorship:	First author
Type of article:	Research article
Share of the work:	70 %
Contribution to the publication:	Performed all experiments, analyzed data, wrote parts of the paper
Journal:	Biomolecular NMR Assignments
5-year impact factor:	0.583
Date of publication:	March 26, 2021
Number of citations:	3
DOI:	10.1007/s12104-021-10019-6
PubMed-ID:	33770349



# $^1\text{H}$ , $^{13}\text{C}$ , and $^{15}\text{N}$ backbone chemical-shift assignments of SARS-CoV-2 non-structural protein 1 (leader protein)

Ying Wang<sup>1</sup> · John Kirkpatrick<sup>1,2</sup> · Susanne zur Lage<sup>2</sup> · Sophie M. Korn<sup>3,5</sup> · Konstantin Neißner<sup>3,5</sup> · Harald Schwalbe<sup>4,5</sup> · Andreas Schlundt<sup>3,5</sup> · Teresa Carlomagno<sup>1,2</sup>

Received: 29 January 2021 / Accepted: 13 March 2021  
© The Author(s) 2021

## Abstract

The current COVID-19 pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has become a worldwide health crisis, necessitating coordinated scientific research and urgent identification of new drug targets for treatment of COVID-19 lung disease. The *covid19-nmr* consortium seeks to support drug development by providing publicly accessible NMR data on the viral RNA elements and proteins. The SARS-CoV-2 genome comprises a single RNA of about 30 kb in length, in which 14 open reading frames (ORFs) have been annotated, and encodes approximately 30 proteins. The first two-thirds of the SARS-CoV-2 genome is made up of two large overlapping open-reading-frames (ORF1a and ORF1b) encoding a replicase polyprotein, which is subsequently cleaved to yield 16 so-called non-structural proteins. The non-structural protein 1 (Nsp1), which is considered to be a major virulence factor, suppresses host immune functions by associating with host ribosomal complexes at the very end of its C-terminus. Furthermore, Nsp1 facilitates initiation of viral RNA translation via an interaction of its N-terminal domain with the 5' untranslated region (UTR) of the viral RNA. Here, we report the near-complete backbone chemical-shift assignments of full-length SARS-CoV-2 Nsp1 (19.8 kDa), which reveal the domain organization, secondary structure and backbone dynamics of Nsp1, and which will be of value to further NMR-based investigations of both the biochemical and physiological functions of Nsp1.

**Keywords** SARS-CoV-2 · Nsp1 · Non-structural proteins · New drug targets · 5' untranslated region · NMR spectroscopy

## Biological context

The ongoing COVID-19 pandemic has initiated intense scientific research into the causative pathogen, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Previous studies have demonstrated that SARS-CoV-2 shows close sequence homology to SARS-CoV-1, which was responsible for the SARS epidemic in 2003, and also to MERS-CoV, which causes Middle-Eastern Respiratory Syndrome (Benedetti et al. 2020; A. Wu et al. 2020a, b). With a high transmissibility and potential to induce life-threatening acute respiratory distress syndrome, SARS-CoV-2 represents a severe threat to human health worldwide. Rapid identification and characterization of druggable protein targets within the SARS-CoV-2 genome is the first step towards development of targeted therapeutic treatments for COVID-19, which together with vaccination approaches, will be critically important in mitigating the impact of the pandemic.

SARS-CoV-2 belongs to the family of beta-coronaviruses, with an enveloped, positive-sense single-stranded

✉ Teresa Carlomagno  
teresa.carlomagno@oci.uni-hannover.de;  
covid19-nmr@dlist.server.uni-frankfurt.de

<sup>1</sup> Centre of Biomolecular Drug Research (BMWZ), Leibniz University Hannover, Schneiderberg 38, 30167 Hannover, Germany

<sup>2</sup> Group of NMR-Based Structural Chemistry, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany

<sup>3</sup> Institute for Molecular Biosciences, St Lucia, QLD 4072, Australia

<sup>4</sup> Institute for Organic Chemistry and Chemical Biology, 60438 Frankfurt, Germany

<sup>5</sup> Center for Biomolecular Magnetic Resonance (BMRZ), Johann Wolfgang Goethe-University Frankfurt, Max-von-Laue-Str. 7, 60438 Frankfurt, Germany

RNA genome (Gorbalenya et al. 2020), which encodes two large overlapping open reading frames (ORF1a and ORF1b) at the 5'-end, as well as four structural proteins and eight accessory proteins at the 3'-end (Lim et al. 2016; Zhou et al. 2020). After cell entry, the virus exploits the host translational machinery to produce the polypeptides corresponding to ORF1a and ORF1b. The polypeptides are then proteolytically cleaved into 16 functional non-structural proteins (Khailany et al. 2020), most of which are involved in assembling the host–viral replication/transcription complex (Masters 2006). Among these proteins, non-structural protein 1 (Nsp1), also known as the leader protein, plays a key role in hampering host gene expression.

SARS-CoV-2 Nsp1, the first N-terminal cleavage protein from the replicase polyprotein, is 180 residues in length, and shares ~85% amino-acid sequence identity with the homologous protein from SARS-CoV-1. SARS-CoV-1 Nsp1 has been structurally characterized using solution-state NMR methods, leading to an atomic-resolution structure of the globular domain comprising residues 13–128 and revealing short N-terminal and long C-terminal tails (residues 1–12 and 129–179, respectively) that are flexibly disordered (Almeida et al. 2007). Recent cryo-electron microscopy (cryo-EM) studies on Nsp1 from SARS-CoV-2 have demonstrated that host protein translation inhibition by Nsp1 is mediated by insertion of its C-terminus into the entrance of the mRNA tunnel in the small ribosomal subunit (Schubert et al. 2020; Thoms et al. 2020). In addition, it was shown biochemically that the 5' UTR of SARS-CoV-2 is capable of promoting translation initiation of viral mRNAs through binding to the N-terminal domain (NTD) of Nsp1 (Schubert et al. 2020). However, the precise mechanism by which SARS-CoV-2 escapes from the Nsp1-mediated translation inhibition—thereby switching the host translation machinery from host to viral protein synthesis—and which specific region of SARS-CoV-2 5' UTR binds to the NTD of Nsp1 are still unclear. Due to its dual role in inhibition of host-protein translation and stimulation of viral-protein translation, Nsp1 has been proposed as an attractive drug target for the treatment of COVID-19.

The research consortium *covid19-nmr* seeks to rapidly and publicly support the search for anti-viral COVID-19 drugs using an NMR-based screening approach that in the initial stage involves the production of all druggable proteins and RNAs from SARS-CoV-2 and their subsequent NMR chemical-shift assignment, followed by a second stage of solution-structure determination, drug-library screening and rational structure-based drug-design. Here we report the near-complete backbone assignment of full-length SARS-CoV-2 Nsp1 that represents the first step towards its structural characterization and provides a basis for residue-resolved drug-screening and protein–RNA interaction studies.

## Methods and experiments

### Construct design

The Nsp1 protein sequence studied here derives from the SARS-CoV-2 genomic sequence corresponding to NCBI GenBank entry NC\_045512.2, which is identical to GenBank entry MN908947.3 (F. Wu et al. 2020a, b). A codon-optimized expression construct of SARS-CoV-2 full-length Nsp1 was inserted into the vector pETM-11 (EMBL bacterial expression vector database), containing an N-terminal His<sub>6</sub>-tag, and a tobacco etch virus (TEV) cleavage site. Due to the nature of the TEV cleavage site, two artificial N-terminal residues (Gly[−1] and Ala[0]) are included in the cleaved protein, before the native Nsp1 sequence starts at Met[1] (corresponding to NCBI GenPept entry YP\_009725297.1).

For the NTD of SARS-CoV-2 Nsp1, domain boundaries were defined in analogy to the available NMR structure (PDB codes 2hsx/2gdt (Almeida et al. 2007)) of the SARS-CoV-1 homologue Nsp1 (85% sequence identity). The expression construct contained residues 13–127 of the Nsp1 sequence. The construct was amplified from the full-length Nsp1 gene, obtained as an *E. coli*-codon-optimized DNA construct from GenScript Biotech (Netherlands). It was inserted into the pET3b-based vector pKM263, containing an N-terminal His<sub>6</sub>-tag, a GST-tag and a tobacco etch virus (TEV) cleavage site. Four artificial N-terminal residues (Gly[−3], Ala[−2], Met[−1] and Gly[0]) remained after proteolytic TEV cleavage, resulting in a 12.9 kDa protein.

### Protein expression and purification

Large-scale expressions of uniformly <sup>13</sup>C,<sup>15</sup>N-labelled full-length Nsp1 in *E. coli* BL21 (DE3) cells were carried out in M9 minimal medium, containing 1 g/L <sup>15</sup>NH<sub>4</sub>Cl (Cambridge Isotope Laboratories), 2.5 g/L <sup>13</sup>C<sub>6</sub>-D-glucose (Cambridge Isotope Laboratories) and 50 µg/mL kanamycin. The cells were cultured at 37 °C until the OD<sub>600</sub> reached 0.7, at which point protein expression was induced with 0.6 mM isopropyl-beta-thiogalactopyranoside (IPTG). The temperature was reduced to 16 °C and expression continued for 22 h. After harvesting (4 °C, 4000 rpm for 40 min), cell pellets were stored at −20 °C. For protein purification, the cells were resuspended in buffer A (50 mM Tris–HCl, 500 mM sodium chloride, 100 mM sodium sulfate, 5% v/v glycerol, 5 mM imidazole, 1 mM Tris-(2-carboxyethyl)-phosphine (TCEP), pH 7.5) with one tablet of EDTA-free protease inhibitor cocktail (Roche), 100 µg of lysozyme (Carl Roth), and 50 µg of

deoxyribonuclease (DNase) (New England Biolabs) per 1 L of original culture volume. Cell lysis was performed by sonication (50% power; 5 s pulse/10 s rest duty-cycle, for 20 min). The cell lysate was cleared by centrifugation (4 °C, 18000 rpm for 1 h). The supernatant was filtered and loaded onto a HisTrap HP column (GE Healthcare), washed first with buffer A and then with buffer A containing additional 2 M LiCl, before elution with a linear gradient of buffer A to buffer B (composition same as buffer A but with 300 mM imidazole). Nsp1-containing fractions of the eluate were exchanged back into buffer A using a desalting column, and then incubated with TEV protease (0.5 mg per 1 L of original culture volume) overnight at 4 °C. The cleaved Nsp1 was separated from the TEV protease and residual uncleaved protein by passing the dialysis mixture over the HisTrap column and washing with buffer A. The flow-through was collected, concentrated and loaded onto a HiLoad Superdex 75 16/600 size-exclusion column (GE Healthcare), previously equilibrated in buffer C (50 mM Tris-HCl, 500 mM sodium chloride, 100 mM sodium sulfate, 1 mM EDTA, 1 mM TCEP, pH 7.5). Pure Nsp1-containing fractions were identified by SDS-PAGE, pooled and exchanged into NMR buffer (50 mM sodium phosphate (pH 6.5), 200 mM sodium chloride, 2 mM dithiothreitol, 2 mM ethylene diamine tetra-acetic acid, 0.01% w/v sodium azide, 0.001% w/v 3-(trimethylsilyl)propane-1-sulfonate) by means of repeated dilution/concentration using Amicon centrifugal concentrators (10-kDa molecular-weight cutoff).

Uniformly <sup>15</sup>N-labelled Nsp1-NTD (residues 13–127) was expressed in *E. coli* strain BL21 (DE3) in M9 minimal medium containing 1 g/L <sup>15</sup>NH<sub>4</sub>Cl (Cambridge Isotope Laboratories) and 100 µg/mL ampicillin. Expression was induced at an OD<sub>600</sub> of 0.7 with 1 mM IPTG for 18 h at 16 °C. Cell pellets were resuspended in Buffer D (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 4 mM DTT and 100 µL protease inhibitor mix (SERVA), pH 8.0). The supernatant was cleared by centrifugation (40 min, 10000 g, 4 °C) and subsequently loaded onto a Ni<sup>2+</sup>-NTA gravity-flow column (Sigma Aldrich). Cleavage of the His<sub>6</sub>-GST-tag was achieved overnight at 4 °C with 0.5 mg of TEV protease per 1L of culture while dialyzing into Buffer D. The TEV protease and the cleaved tag were removed via a second Ni<sup>2+</sup>-NTA gravity-flow column. Further purification of Nsp1-NTD by size exclusion chromatography (HiLoad 16/600 SD 75, GE Healthcare) was carried out in Buffer E (25 mM NaPi, 250 mM NaCl, 2 mM TCEP, 0.02% w/v NaN<sub>3</sub>, pH 7.0). Fractions containing pure Nsp1-NTD were determined by SDS-PAGE, pooled and concentrated using Amicon centrifugal concentrators (3-kDa cutoff). Final NMR samples were prepared in Buffer E, containing 5% v/v D<sub>2</sub>O at Nsp1-NTD concentrations of 15 µM.

## NMR experiments

NMR samples (~550 µL total volume in 5-mm-diameter NMR tubes) were prepared with Nsp1 at a concentration of 300–400 µM, dissolved in NMR buffer. The protein appeared relatively stable over a period of several days, except for the gradual appearance of a set of small, sharp peaks characteristic of the dipeptides resulting from proteolysis. Appearance of these peaks was also accompanied by small shift-changes in the main set of peaks.

NMR experiments on full-length Nsp1 were recorded at 298 K on two Bruker Avance III-HD spectrometers running Topspin 3.2 software, with <sup>1</sup>H field-strengths of 850 MHz and 600 MHz, and equipped with inverse HCN CP-TCI (helium-cooled) and CPP-TCI (nitrogen-cooled) cryogenic probeheads, respectively.

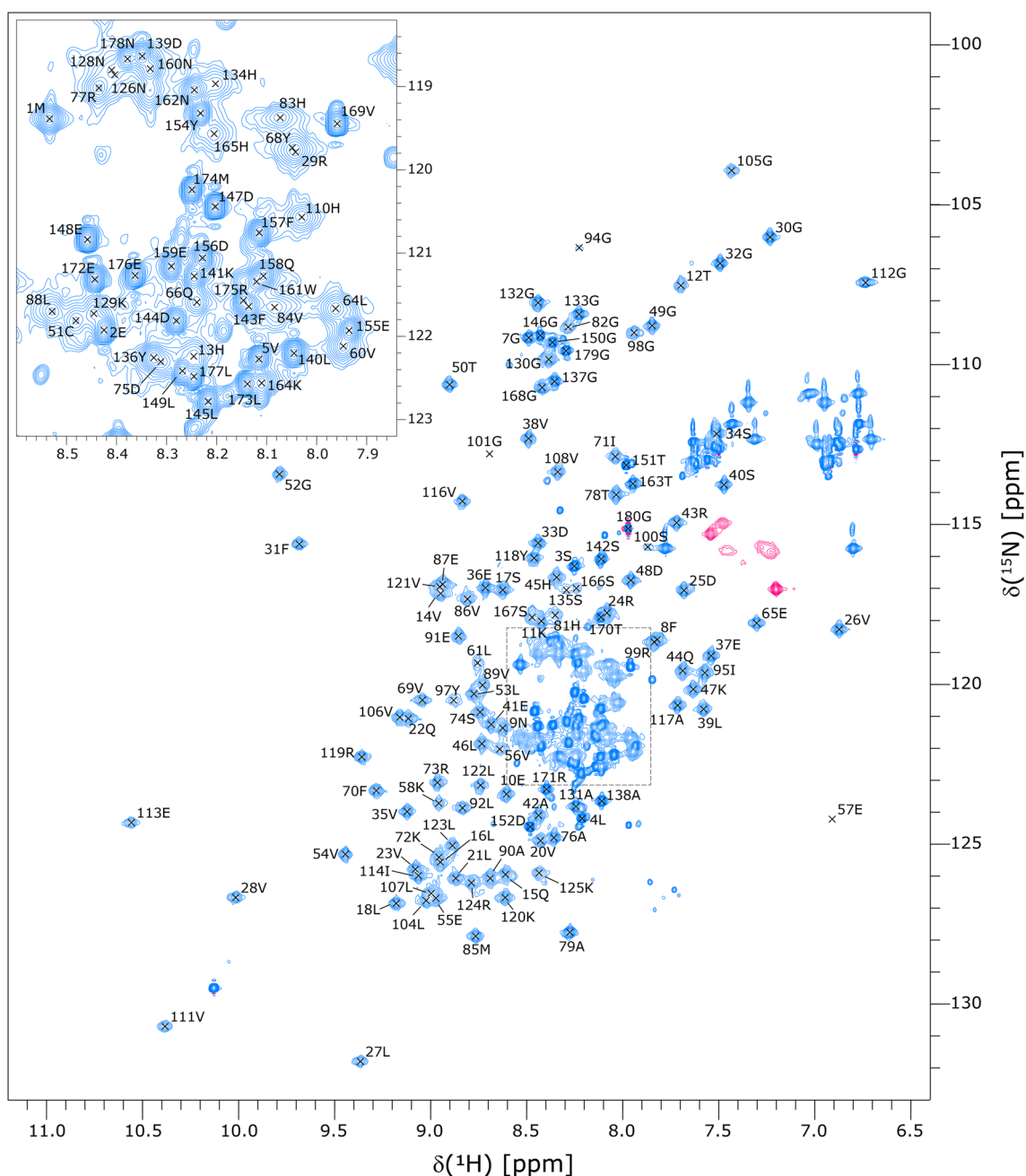
2D <sup>15</sup>N-HSQC spectra were recorded using States-TPPI for frequency discrimination, with water suppression achieved via a combination of WATERGATE and water flip-back pulses to preserve the water magnetization (Bodenhausen & Ruben 1980; Piotto et al. 1992). Backbone resonance assignments were obtained from a standard-suite of 3D triple-resonance out-and-back-type spectra, comprising HNCO (Ikura et al. 1990; Kay et al. 1990), HN(CA)CO (Clubb et al. 1992; Kay et al. 1994), HNCA, HNCACB (Grzesiek & Bax 1992; Wittekind & Mueller 1993), HN(CO)CA and HN(CO)CACB (Bax & Ikura 1991; Yamazaki et al. 1994) spectra. With the exception of the HNCO experiment, which was recorded at both field-strengths, the through-carbonyl and the HN(CA)CO spectra were recorded at 600 MHz, while the remaining spectra were recorded at 850 MHz. Water suppression and frequency discrimination in the triple-resonance spectra were achieved as for the <sup>15</sup>N-HSQC spectra.

For the Nsp1 NTD, 2D <sup>15</sup>N-HSQC-TROSY experiments were measured at 950 MHz with acceleration of longitudinal <sup>1</sup>H relaxation between scans via the Band-Selective Excitation Short-Transient (BEST) approach (Favier & Brutscher 2011; Lescop et al. 2007; Solyom et al. 2013) using exclusively shaped-pulses on <sup>1</sup>H (bandwidth and offset of 4.5 and 8.5 ppm, respectively) and an inter-scan delay of 0.3 s.

NMR data were processed with NMRPipe v10.1 (Delaglio et al. 1995) and analysed in CcpNmr Analysis v2.4 (Vranken et al. 2005).

## Assignments and data deposition

The <sup>15</sup>N-HSQC spectrum of SARS-CoV-2 full-length Nsp1 has the appearance expected for a protein comprising a well-folded globular domain together with an extended and largely disordered tail (Fig. 1). The amide peaks from the residues in the globular domain are widely dispersed with a <sup>1</sup>H chemical-shift dispersion of ~4 ppm, while the peaks



**Fig. 1**  $^{15}\text{N}$ -HSQC spectrum of SARS-CoV-2 Nsp1 with backbone amide assignments. The excerpt in the top-left corner is an expansion of the crowded central region (corresponding to the dashed box in the main spectrum). The unlabeled peak at  $\delta(^1\text{H})=10.15$  ppm,  $\delta(^{15}\text{N})=129.8$  ppm corresponds to the side-chain  $\text{N}^{\epsilon 1}\text{-H}^{\epsilon 1}$  group of

residue 161W. The unlabeled pink peaks correspond to arginine side-chain  $\text{N}^{\epsilon}\text{-H}^{\epsilon}$  groups, and are aliased from their true  $^{15}\text{N}$  chemical-shift positions at  $\sim 84\text{--}86$  ppm. The peaks of the asparagine and glutamine side-chain amide groups are also unlabeled. The spectrum was recorded at a  $^1\text{H}$  field-strength of 850 MHz

from the C-terminal region are clustered more closely together with  $^1\text{H}$  chemical-shifts in the range 7.9–8.5 ppm. In general, the peaks from the C-terminal region are much sharper and more intense than the more widely dispersed peaks from the globular domain, as would be expected for a disordered tail. Differential linewidths and intensities characteristic of exchange-mediated line-broadening

effects were observed within both the C-terminal region and the globular domain. Some residues in the C-terminal region and towards the C-terminal end of the globular domain appeared to give rise to multiple peaks, with one or two weak-intensity peaks in addition to a strong-intensity peak. The weak-intensity peaks are probably due to

low-populated alternate conformers associated with cis-geometry amide bonds at proline residues.

We assigned the backbone resonances to a high-degree of completeness, obtaining <sup>1</sup>H and <sup>15</sup>N assignments for 165 of the 172 assignable backbone amide groups (96.0%), and <sup>13</sup>C assignments for 96.1%, 97.2% and 97.5% of the C', C<sup>α</sup> and C<sup>β</sup> nuclei, respectively. Amide assignments are missing for residues 59G, 63Q, 93E, 96Q, 102E, 103 T and 127G. These residues are located in four stretches for which the assigned amide peaks are significantly broader and weaker. The amide signals for the missing residues are probably so broadened that the corresponding peaks simply do not appear above the noise in the <sup>15</sup>N-HSQC spectrum, and certainly would not yield detectable peaks in the 3D triple-resonance spectra.

The assigned backbone chemical shifts were used to predict the secondary-structure with DANGLE (Cheung et al. 2010) and TALOS-N (Shen & Bax 2013), and also to calculate random-coil-index-derived order-parameters (RCI-S<sup>2</sup>) (Berjanskii & Wishart 2005). Panel A of Fig. 2 shows the RCI-S<sup>2</sup> values and TALOS-N-derived helix/strand probabilities. Panel B shows the TALOS-N- and DANGLE-predicted secondary structures of SARS-CoV-2 Nsp1 together with the secondary-structure assignments of the two crystal structures of the folded globular domain (PDB codes 7k7p and 7k3n), the two cryo-EM structures of the C-terminus bound to the small ribosomal subunit (PDB codes 6zn5 (Thoms et al. 2020) and 7k5i), and the solution-NMR structure of Nsp1 from SARS-CoV-1 (PDB code 2hsx).

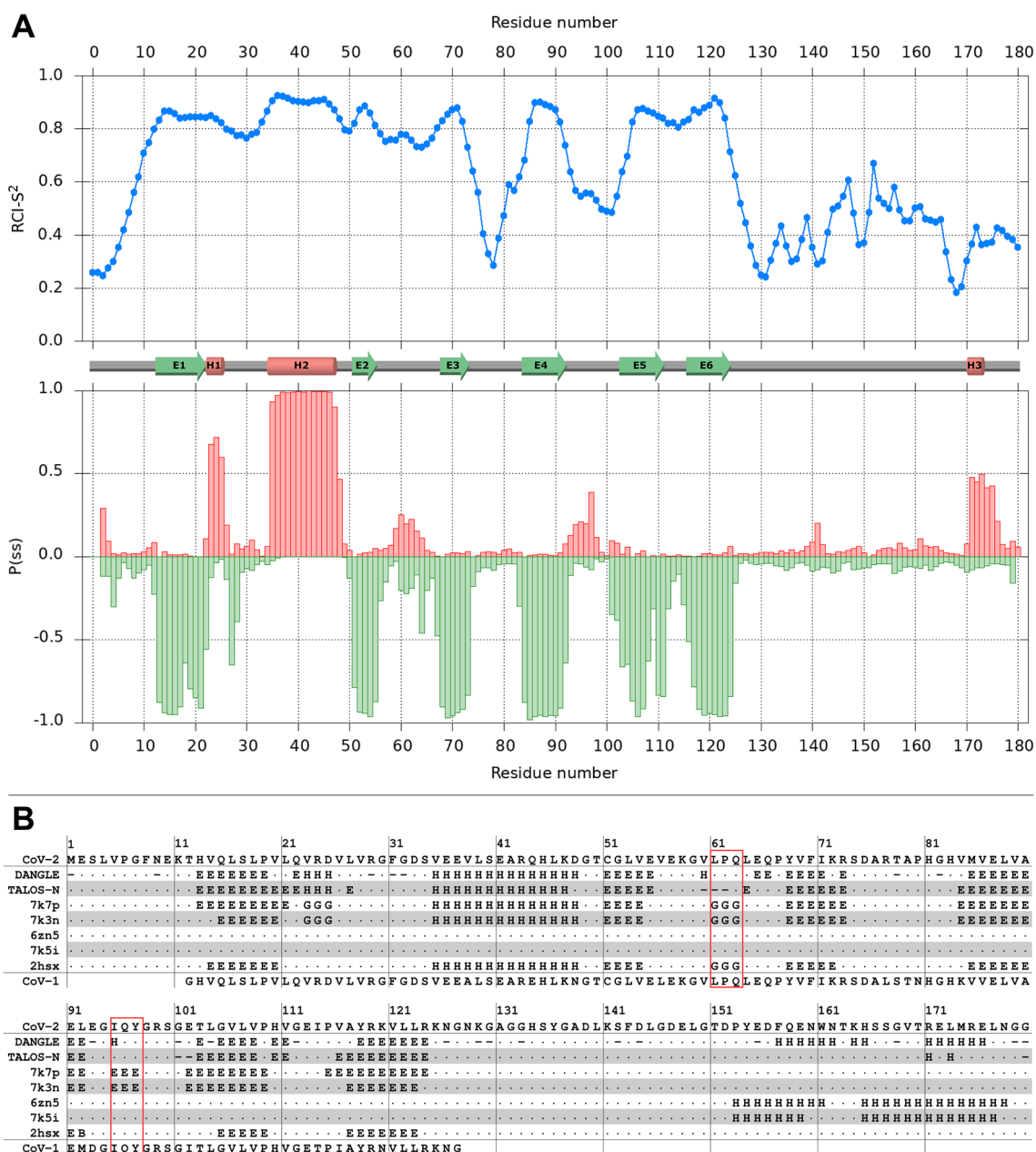
The well-folded globular domain of Nsp1 extends approximately from residue 10 to residue 125, and consists of six β-strands and two helices (labelled E1–E6 and H1–H2 in Fig. 2, respectively). The two long loops between strands E3 and E4, and between E4 and E5 show relatively low order-parameters, indicative of significant internal flexibility. Overall, the chemical-shift-derived secondary-structure predictions for the globular domain are very similar to the secondary structures of the two crystal structures and the solution-structure of Nsp1 from SARS-CoV-1. The short proline-containing 3<sub>10</sub> helix (residues 61–63) and the short

β-sheet formed by residues 95–97 observed in the crystal structures are not predicted from the backbone chemical shifts, but some assignments were missing for residues in these two regions, compromising the ability to make accurate chemical-shift-based secondary-structure predictions.

In full-length Nsp1, the C-terminal region from residue 126 onwards is partially but not completely disordered, with RCI-S<sup>2</sup> values falling mostly in the range 0.3–0.6, and very little canonical secondary-structure propensity for residues 125–155. DANGLE predicts two additional helices at the very C-terminus (residues 157–162 and 171–176). While the TALOS-N helical assignments in the same sequence-stretch are limited to just two residues (171 & 173; labelled H3 in Fig. 2), the corresponding probabilities indicate a clear helical propensity for residues 171–175. Interestingly, the region of helical propensity at the very end of the C-terminus corresponds approximately to the two helices observed in the cryo-EM structure of Nsp1 interacting with the small ribosomal subunit, suggesting that these helices, which are partially formed in free Nsp1, become stabilized upon insertion into the mRNA tunnel of the small ribosomal subunit.

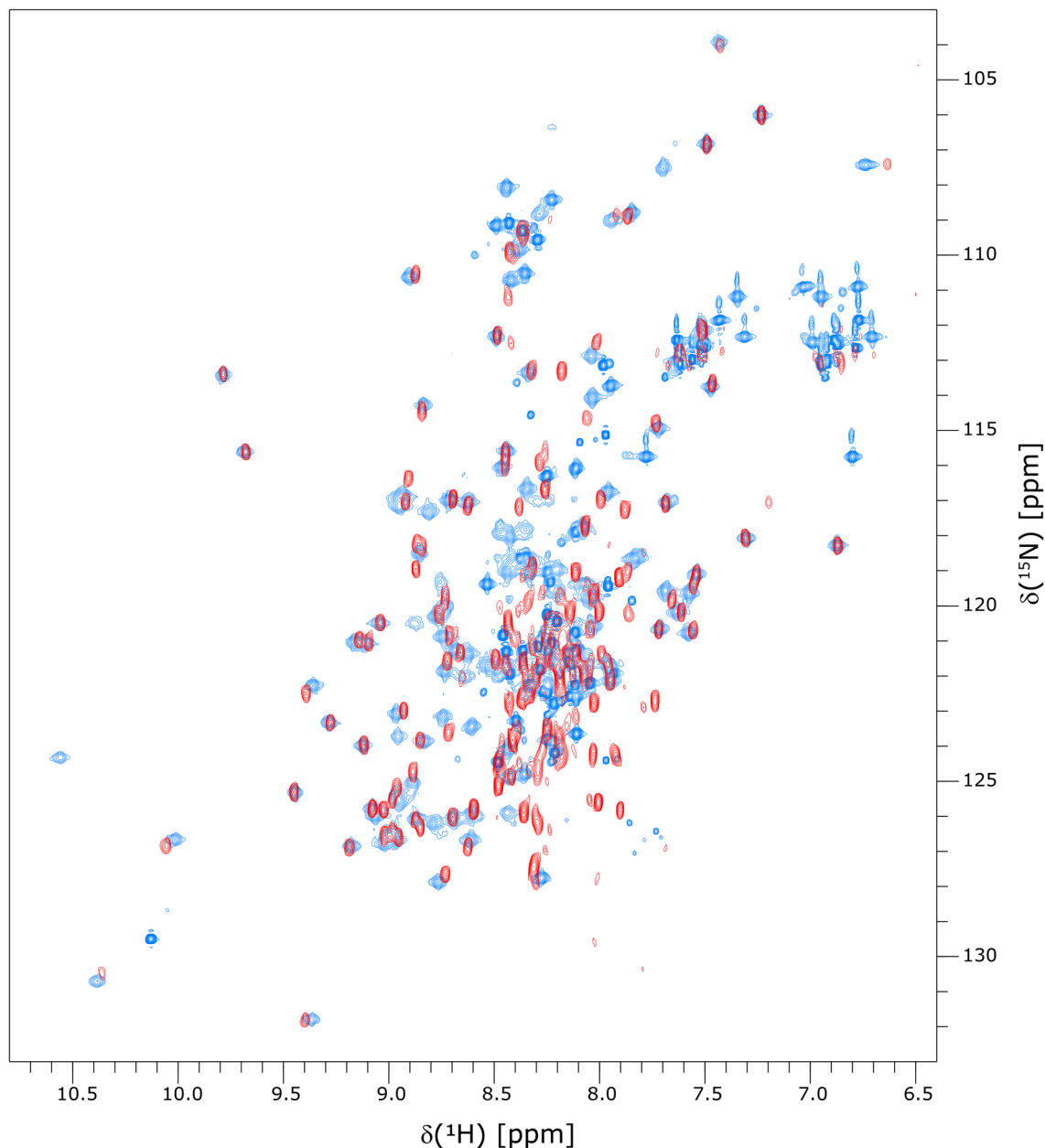
Based on the strong sequence-similarity with SARS-CoV-1 Nsp1, we also investigated the isolated Nsp1 NTD (residues 13–127). Unexpectedly, the protein showed little solubility and we were not able to achieve high concentrations in NMR-compatible buffers, although the domain boundaries are in agreement with the available crystal and NMR structures of SARS-CoV-1 Nsp1 (Fig. 2b). Nevertheless, we were able to record 2D <sup>1</sup>H, <sup>15</sup>N correlation spectra. An overlay of the <sup>15</sup>N-HSQC-TROSY spectrum of Nsp1 NTD with the <sup>15</sup>N-HSQC spectrum of full-length Nsp1 (Fig. 3) shows reasonable agreement for the residues of the NTD, although there are some chemical shift differences and also instances of relative line-broadening for the Nsp1 NTD peaks. It is conceivable that in the full-length protein, the disordered C-terminal domain interacts with the NTD, thereby stabilizing its fold and improving solubility. We are currently conducting further experiments to investigate this hypothesis.





**Fig. 2** Secondary-structure of SARS-CoV-2 Nsp1. **a.** Results from TALOS-N analysis of the assigned backbone chemical shifts. Top: random-coil-index-derived squared-order-parameters ( $RCI-S^2$ ) plotted by residue number. Bottom: probabilities of helical (positive, red) and strand (negative, green) secondary-structure plotted by residue number. The schematic secondary-structure shown between the two plots is based on the TALOS-N prediction. **b.** Comparison of the DANGLE- and TALOS-N-predicted secondary-structures for SARS-CoV-2 Nsp1 with those of the crystal structures for SARS-CoV-2 Nsp1 with those of the crystal structures of the globular domain (PDB codes 7k7p & 7k3n), the cryo-EM structures of the C-terminus in complex with the small ribosomal subunit (PDB codes 6zn5 & 7k5i) and the NMR structure of the globular domain

of SARS-CoV-1 Nsp1 (PDB code 2hsx). The two regions where secondary-structure elements present in the crystal structures are missing from the chemical-shift-based predictions are highlighted in red boxes. The secondary structures of the PDBs were calculated with STRIDE (Frishman & Argos 1995). The Nsp1 sequences from SARS-CoV-2 and SARS-CoV-1 (shown above and below the secondary-structure annotations, respectively) were aligned with ClustalW (Sievers et al. 2011). The codes for the secondary-structure annotations are as follows: ‘E’ denotes extended  $\beta$ -strand; ‘H’ denotes either generic helix (DANGLE & TALOS-N) or specific  $\alpha$ -helix (STRIDE); ‘G’ denotes  $3_{10}$  helix (STRIDE only)



**Fig. 3** Overlay of  $^1\text{H}$ , $^{15}\text{N}$  correlation spectra of full-length Nsp1 and the Nsp1 NTD (residues 13–127). Overlay of the  $^{15}\text{N}$ -HSQC-TROSY spectrum of Nsp1 NTD (red; recorded at 950 MHz) with the  $^{15}\text{N}$ -HSQC spectrum of full-length Nsp1 (blue; recorded at

850 MHz). The signal-to-noise ratios of the two spectra are not quantitatively comparable due to differences in concentrations and acquisition parameters

**Acknowledgements** We acknowledge Dr Andrea Graziadei (TU Berlin) for the kind gift of the Nsp1 expression plasmid.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This work was funded by Goethe Corona Funds, the Deutsche Forschungsgemeinschaft through grant SFB902/B18 (to *covid19-nmr*), DFG large scale equipment grant 452632086, EU commission access programme iNEXT-discovery (grant agreement no. 871037), and grant CA294/16-1 to T.C.

**Data availability** The  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  backbone chemical-shift assignments of SARS-CoV-2 Nsp1 have been deposited at the BioMagRes-Bank (<https://www.bmrb.wisc.edu>) under accession number 50620.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Almeida MS, Johnson MA, Herrmann T, Geralt M, Wüthrich K (2007) Novel  $\beta$ -Barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. *J Virol* 81(7):3151–3161. <https://doi.org/10.1128/JVI.01939-06>
- Bax A, Ikura M (1991) An efficient 3D NMR technique for correlating the proton and  $^{15}\text{N}$  backbone amide resonances with the  $\alpha$ -carbon of the preceding residue in uniformly  $^{15}\text{N}/^{13}\text{C}$  enriched proteins. *J Biomol NMR* 1(1):99–104. <https://doi.org/10.1007/BF01874573>
- Benedetti F, Snyder GA, Giovanetti M, Angeletti S, Gallo RC, Ciccozzi M, Zella D (2020) Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1. *J Transl Med* 18(1):4–9. <https://doi.org/10.1186/s12967-020-02507-5>
- Berjanskii MV, Wishart DS (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* 127(43):14970–14971. <https://doi.org/10.1021/ja054842f>
- Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* 69(1):185–189. [https://doi.org/10.1016/0009-2614\(80\)80041-8](https://doi.org/10.1016/0009-2614(80)80041-8)
- Cheung MS, Maguire ML, Stevens TJ, Broadhurst RW (2010) DAN-GLUE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J Magn Reson* 202(2):223–233. <https://doi.org/10.1016/j.jmr.2009.11.008>
- Clubb RT, Thanabal V, Wagner G (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intrasite  $^1\text{H}$ ,  $^{15}\text{N}$ , and  $^{13}\text{C}$  chemical shifts in  $^{15}\text{N}$ - $^{13}\text{C}$ -labelled proteins. *J Magn Reson*. [https://doi.org/10.1016/0022-2364\(92\)90252-3](https://doi.org/10.1016/0022-2364(92)90252-3)
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. <https://doi.org/10.1007/BF00197809>
- Favier A, Brutscher B (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J Biomol NMR* 49(1):9–15. <https://doi.org/10.1007/s10858-010-9461-5>
- Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. *Proteins Struct Funct Genet* 23(4):566–579. <https://doi.org/10.1002/prot.340230412>
- Gorbalenya A, Baker S, Baric R, de Groot R, Drosten C, Gulyaeva A, Haagmans B, Lauber C, Leontovich A, Neuman B, Penzar D, Perlman S, Poon L, Samborskiy D, Sidorov I, Sola I, Ziebuhr J (2020) Severe acute respiratory syndrome-related coronavirus: the species and its viruses – a statement of the Coronavirus Study Group. *Nat Microbiol*. <https://doi.org/10.1101/2020.02.07.937862>
- Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson* 96(2):432–440. [https://doi.org/10.1016/0022-2364\(92\)90099-5](https://doi.org/10.1016/0022-2364(92)90099-5)
- Ikura M, Kay LE, Bax A (1990) A novel approach for sequential assignment of  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy Application to Calmodulin. *Biochemistry* 29(19):4659–4667. <https://doi.org/10.1021/bi00471a022>
- Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson* 89(3):496–514. [https://doi.org/10.1016/0022-2364\(90\)90333-5](https://doi.org/10.1016/0022-2364(90)90333-5)
- Kay LE, Xu GY, Yamazaki T (1994) Enhanced-sensitivity triple-resonance spectroscopy with minimal  $\text{H}_2\text{O}$  saturation. *J Magn Reson Ser A*. <https://doi.org/10.1006/jmra.1994.1145>
- Khailany RA, Safdar M, Ozaslan M (2020) Genomic characterization of a novel SARS-CoV-2. *Gene Rep* 19(January):100682. <https://doi.org/10.1016/j.genrep.2020.100682>
- Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* 187(1):163–169. <https://doi.org/10.1016/j.jmr.2007.04.002>
- Lim Y, Ng Y, Tam J, Liu D (2016) Human coronaviruses: a review of virus-host interactions. *Diseases* 4(4):26. <https://doi.org/10.3390/diseases4030026>
- Masters PS (2006) The molecular biology of coronaviruses. *Adv Virus Res* 65(January):193–292. [https://doi.org/10.1016/S0065-3527\(06\)66005-3](https://doi.org/10.1016/S0065-3527(06)66005-3)
- Piotto M, Saudek V, Sklenář V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR*. <https://doi.org/10.1007/BF02192855>
- Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler LA, Leibundgut M, Thiel V, Mühlemann O, Ban N (2020) SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol* 27(10):959–966. <https://doi.org/10.1038/s41594-020-0511-8>
- Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56(3):227–241. <https://doi.org/10.1007/s10858-013-9741-y>
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. <https://doi.org/10.1038/msb.2011.75>
- Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR* 55(4):311–321. <https://doi.org/10.1007/s10858-013-9715-0>
- Thoms M, Buschauer R, Ameismeier M, Koepke L, Denk T, Hirschenberger M, Kratzat H, Hayn M, MacKens-Kiani T, Cheng J, Straub JH, Stürzel CM, Fröhlich T, Berninghausen O, Becker T, Kirchhoff F, Sparrer KMJ, Beckmann R (2020) Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* 369(6508):1249–1256. <https://doi.org/10.1126/SCIENCE.ABC8665>
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct Funct Genet*. <https://doi.org/10.1002/prot.20449>
- Wittekind M, Mueller L (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the  $\alpha$ - and  $\beta$ -carbon resonances in proteins. *J Magn Reson Ser B* 101(2):201–205. <https://doi.org/10.1006/jmrb.1993.1033>
- Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T (2020a) Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* 27(3):325–328. <https://doi.org/10.1016/j.chom.2020.02.001>

Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, Yuan M-L, Zhang Y-L, Dai F-H, Liu Y, Wang Q-M, Zheng J-J, Xu L, Holmes EC, Zhang Y-Z (2020b) A new coronavirus associated with human respiratory disease in China. *Nature* 579(7798):265–269. <https://doi.org/10.1038/s41586-020-2008-3>

Yamazaki T, Muhandiranv DR, Kay LE, Lee W, Arrowsmith CH (1994) A suite of triple resonance NMR experiments for the backbone assignment of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  labeled proteins with high sensitivity. *J Am Chem Soc* 116(26):11655–11666. <https://doi.org/10.1021/ja00105a005>

Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang

R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Xiao GF, Shi Z-L (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579(7798):270–273. <https://doi.org/10.1038/s41586-020-2012-7>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## 2.3 Manuscript 2

### Substrate recruitment by the hexameric MecA–ClpC complex

Ying Wang<sup>1</sup>, John Kirkpatrick<sup>1,2,3</sup>, Teresa Carlomagno<sup>1,2,3,4</sup> \*

<sup>1</sup>Centre of Biomolecular Drug Research (BMWZ), Leibniz University Hannover, Schneiderberg 38, D-30167 Hannover

<sup>2</sup>Group of NMR-based Structural Chemistry, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig

<sup>3</sup>School of Biosciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK

<sup>4</sup>Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK

Type of authorship:	First author
Type of article:	Research article
Share of the work:	80 %
Contribution to the publication:	Planned and performed all experiments, prepared all figures, and wrote the paper
Journal:	Journal of Structural Biology
5-year impact factor:	3.121
Status of publication:	To be submitted

## Substrate recruitment by the hexameric MecA–ClpC complex

Ying Wang<sup>1</sup>, John Kirkpatrick<sup>1,2,3</sup>, Teresa Carlomagno<sup>1,2,3,4</sup> \*

<sup>1</sup>Centre of Biomolecular Drug Research (BMWZ), Leibniz University Hannover, Schneiderberg 38, D-30167 Hannover

<sup>2</sup>Group of NMR-based Structural Chemistry, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig

<sup>3</sup>School of Biosciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK

<sup>4</sup>Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK

\*To whom correspondence should be addressed: [teresa.carlomagno@oci.uni-hannover.de](mailto:teresa.carlomagno@oci.uni-hannover.de)

### Abstract

In bacteria, the unfoldase ClpC interacts with its adaptor protein MecA to form a hexameric molecular machine that is responsible for controlled unfolding of substrate proteins. The ATP-dependent MecA–ClpC complex can further associate with the protease ClpP to form a complete protein degradation machine. The protein ComK, which is a key player in the development of bacterial competence, is one of the most important substrates for the ClpCP degradation system. Here, we report for the first time the principles of the three-dimensional architecture of the ComK–MecA–ClpC complex, shedding light on the process by which ComK is recruited for unfolding and subsequent degradation. NMR experiments confirmed that ComK is recruited to the MecA–ClpC complex by the MecA N-terminal domain (NTD). Contrary to most previous studies, we show that the MecA–ClpC binary complex in solution does not have a stable and well-defined MecA:ClpC stoichiometry of 6:6, but that the number of MecA molecules attached to the hexameric ClpC particle is variable, with a prevailing stoichiometry of 4:6. This “4:6” MecA–ClpC state appears to be stabilized by addition of ComK, which results in a ternary complex with a distinct 2:4:6 stoichiometry (ComK:MecA:ClpC). Furthermore, we demonstrate that the MecA-NTD in the MecA–ClpC complex has restricted mobility, suggesting an interaction of the MecA-NTD with ClpC that could be at the basis of MecA auto-degradation by the MecA–ClpC complex in the absence of other substrate proteins.

### Keywords

ATP-dependent MecA–ClpC complex; ComK; Protein degradation; NMR spectroscopy

### Introduction

Regulated proteolysis by ATP-dependent AAA+ proteins (ATPases associated with various cellular activities) is a critical component of the protein quality-control system that has evolved in all domains of life to both maintain cellular homeostasis and allow response to external stimuli and environmental changes. To date, several major families of AAA+ proteins have been shown to associate with large numbers of substrate proteins and engage in protein remodeling, unfolding and degradation, including the Lon and FtsH families, the PAN/20S protease family, the Hsp100/Clp protein family and the HslUV (also known as ClpYQ) family. (Ogura and Wilkinson, 2001; Sauer and Baker, 2011).

In bacteria, the most extensively studied AAA+ proteins are from the conserved Hsp100/Clp protein family, which includes the ClpAP, ClpCP and ClpXP protease complexes. Each of these protein-degradation complexes consists of an ATP-dependent, protein-unfolding unit — such as ClpA, ClpC and ClpX — and an associated protease unit, such as ClpP. In the presence of ATP, the unfoldase proteins normally form a hexameric barrel-like assembly with a central pore, through which substrate proteins are unfolded and translocated to the protease unit for proteolytic degradation (Horwich et al., 1999; Wang et al., 2011). However, some unfoldase members of the Hsp100/Clp protein family, such as ClpB and Hsp104, are incapable of binding to ClpP and serve solely as molecular chaperones, acting to disaggregate and subsequently refold protein aggregates during cycles of ATP hydrolysis. Importantly, even those unfoldases that usually associate with ClpP, such as ClpA, ClpC, and ClpX, can function independently of ClpP to disaggregate or remodel misfolded polypeptides (Andersson et al., 2006; Elsholz et al., 2017). Members of the Hsp100/Clp protein family can be divided into two distinct classes based on the number of nucleotide-binding domains (NBDs) they contain. ClpA, ClpB, ClpC and Hsp104 belong to class 1, bearing two NBDs (known as D1 and D2), while ClpX belongs to class 2, having only a single NBD (Fig. S1B) (Schirmer et al., 1996).

Several adaptor proteins for bacterial Hsp100/Clp proteins have been identified and characterized. These adaptor proteins impart substrate specificity, modulate the proteolytic or chaperoning activity of their respective partner AAA+ proteins, and assist in recognition and subsequent degradation of specific substrate proteins. While most adaptor proteins do not fundamentally alter the basic functions of the AAA+ proteins they bind, some of them, such as MecA, trigger a complete switch in the activity of the AAA+ partner proteins. The cellular functions of ClpC, which include chaperoning and degradation of specific substrates, are strictly dependent on its association with MecA to form a binary MecA–ClpC complex (Fig. S1A).

MecA itself consists of two well-structured, globular domains (Fig. 1A). The N-terminal domain (NTD) is necessary for recruitment of substrate proteins (Persuh et al., 1999; Schlothauer et al., 2003), while the C-terminal domain (CTD) interacts with ClpC, promoting its self-

association into the functional hexameric state (Kirstein et al., 2006). The MecA-NTD and CTD are separated by a flexible linker (~40 residues), whose role in substrate binding and delivery is unknown.

In *Bacillus subtilis*, bacterial competence development is controlled by a regulated proteolysis of transcription factor ComK, which activates the expression of competence genes that encode proteins responsible for binding and internalization of exogenous DNA (Dubnau, 1997). The controlled degradation of ComK takes place via the ClpCP machinery; during escape-of-competence, ComK is recruited to the MecA–ClpCP complex via interaction with MecA, where it is unfolded and degraded (Fig. S1A). Thereby, the cellular concentration of ComK is reduced to a level that is insufficient for transcriptional activation of competence genes. In post-exponential or stationary growth-phases, two quorum-sensing pheromones (CSF and ComX), that serve as signals for high cell density, stimulate the synthesis of a small anti-adaptor protein called ComS (Magnuson et al., 1994; Solomon et al., 1996). ComS and ComK bind competitively to MecA: increased concentrations of ComS result in ComK being displacement from the MecA–ClpCP complex, effectively acting to inhibit the ClpCP-mediated degradation of ComK. Since ComK is a positive auto-regulatory transcription factor, the increased levels of ComK resulting from the slower rate of degradation act to accelerate the rate of ComK synthesis, ultimately leading to the development of competence. However, when neither ComK nor ComS are present in significant quantities, MecA itself is degraded by ClpCP, which eventually induces the disassembly of the obsolete ClpCP complex. Thus, ClpCP, MecA and ComK, together with the anti-adaptor ComS, form a regulatory system for the careful control of bacterial competence development (Battesti and Gottesman, 2013; Elsholz et al., 2017; Lazazzera and Grossman, 1997; Turgay et al., 1998).

ClpC comprises two nucleotide-binding domains (D1 and D2), a “middle” domain (M-domain, also referred to as a linker domain) and an N-terminal domain (Fig. S1B). Together, the D1 and N-terminal domains interact with MecA, while the D2 domain interacts with ClpP to form the ClpCP degradation machine. In addition to the well-structured cores of these domains, ClpC has a series of pore-loops within the D1 and D2 domains (pore-loop-1 (KYRG) and pore-loop-2 (GAGGA) in D1, and pore-loop-3 (GYVG) in D2), which are critical for substrate unfolding and degradation (Fig. S1C) (Dougan et al., 2002; Kirstein et al., 2006; Wang et al., 2011).

Despite a number of structural and functional studies of Hsp100/Clp proteins, our understanding of the mechanistic details of substrate recruitment by adaptor proteins remains limited. Here, we report the first *in-vitro* assembly of the ComK–MecA–ClpC ternary complex for both biochemical and structural analysis. Using methyl-TROSY spectra, we study MecA within both the MecA–ClpC and ComK–MecA–ClpC complexes by solution NMR spectroscopy.



We find that in MecA–ClpC, the MecA-NTD shows reduced mobility, indicating that this domain interacts with ClpC also in the absence of substrate, possibly providing a mechanism for the auto-degradation of MecA by the MecA–ClpC complex. NMR titration experiments of ComK onto the MecA–ClpC complex in the presence of ATP, combined with SEC-MALS and SAXS measurements, demonstrate that the MecA–ClpC complex recruits two copies of ComK, to yield a particle consisting of six copies of ClpC, four copies of MecA and two copies of ComK. This, together with MALS data showing that the MecA–ClpC complex is not a homogeneous 6:6 complex, suggest plasticity in the number of adaptor proteins necessary to activate the ClpC complex, which could represent a means to fine-tune ClpC activity through MecA concentration in the presence and absence of substrates.

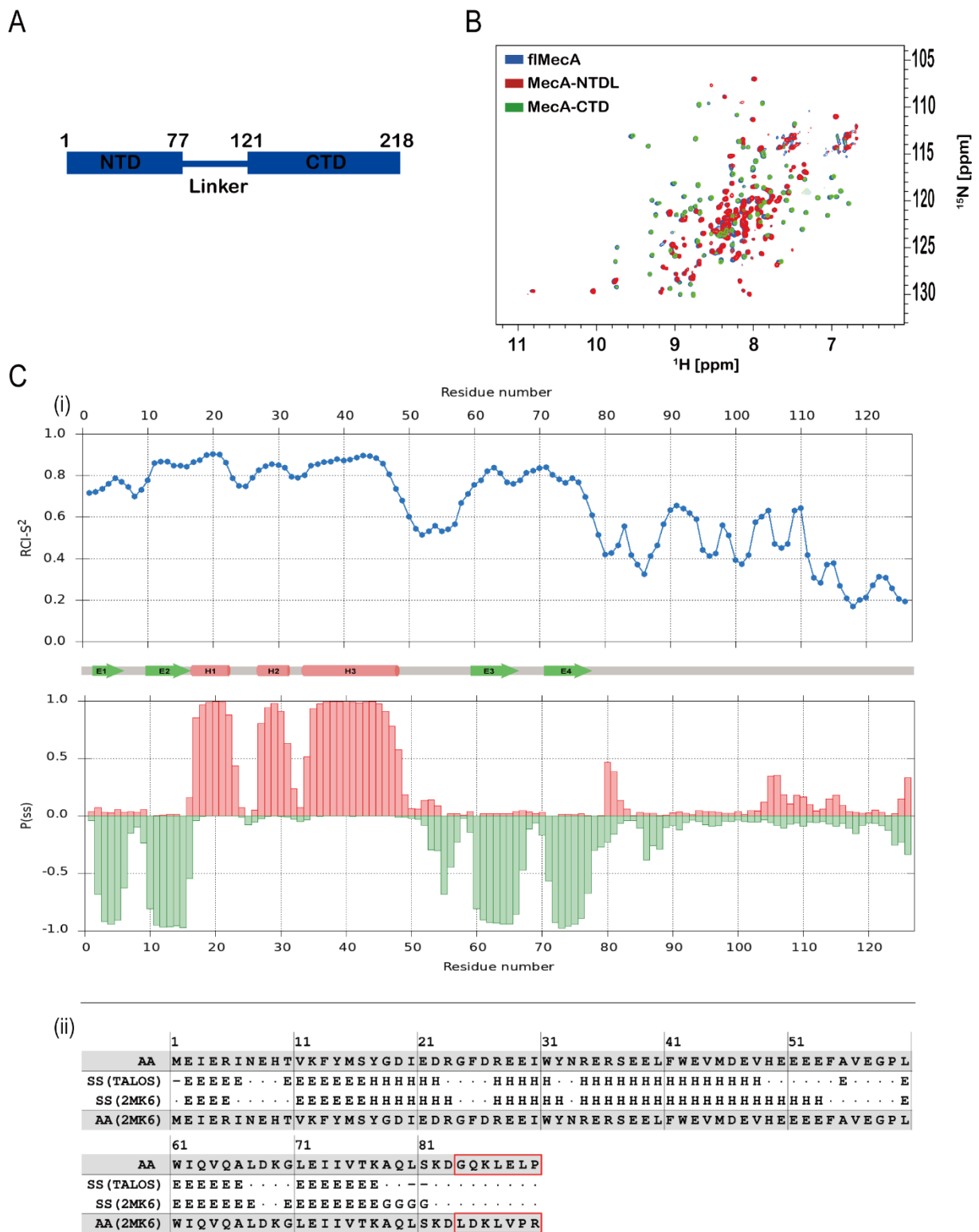
## Results

### NMR analysis of full-length and truncated MecA constructs

The  $^{15}\text{N}$ -HSQC spectrum of full-length (fl) MecA (26.5 kDa) at 650  $\mu\text{M}$  showed broad line widths indicative of the formation of oligomers in solution. The NMR signals sharpened up upon dilution (Figure S2A–C) and plateaued at a concentration of 30  $\mu\text{M}$ . At this concentration MALS experiments also confirmed that the protein is monomeric in solution. To discover which domain of MecA is responsible for the oligomerization, we designed four truncated constructs corresponding to the N-terminal domain of the protein (NTD, MecA<sup>1–77</sup>), the C-terminal domain (CTD, MecA<sup>121–218</sup>), the NTD and the linker region (NTDL, MecA<sup>1–120</sup>), the linker region and the CTD (CTDL, MecA<sup>78–218</sup>). The NTD eluted from size exclusion chromatography as two peaks, of which only the one eluting later delivered a  $^{15}\text{N}$ -HSQC spectrum with sharp signals (Fig. S3A–B). The NTDL, which eluted from SEC as one individual peak, yielded a high quality  $^{15}\text{N}$ -HSQC spectrum, indicating that neither the NTD nor the linker region are responsible for the oligomerization of MecA. Conversely, the  $^{15}\text{N}$ -HSQC spectrum of the CTD showed a similar behavior as that of flMecA, where sharp signals could be obtained only at low protein concentration (Fig. S3D–E). The tendency of the CTD to aggregate was not relieved by addition of the linker region in the CTDL construct (Fig. S3F). Thus, we concluded that the CTD is responsible for MecA oligomerization and used the NTDL construct to assign the backbone and ILV methyl-groups resonances of this portion of MecA to a high-degree of completeness with the exception of residues M1, E8, E50, E53, E54, E57, I62, K69K, S81 and K82, for which the amide peaks were also missing in the  $^{15}\text{N}$ -HSQC spectrum. In addition, most residues showed two peaks for both the backbone amide and the methyl groups in both MecA-NTDL and flMecA constructs (Fig. S4). The presence of a second peak for most residues of the core  $\beta$ -sheet and helix  $\alpha$ 3 let us hypothesize that the tertiary packing of the

secondary structural elements of the MecA-NTD may adopt two different conformations (conformers A and B).

The  $^{15}\text{N}$  HSQC spectrum of free flMecA closely resembles a simple superposition of the respective spectra of the isolated NTD and CTD (Fig. 1B), differing significantly only with respect to the presence of the peaks arising from the inter-domain linker, indicating that any interactions between the two globular domains in the full-length protein are non-existent or negligible. In agreement, the narrow and intense peaks due to residues in the linker are characteristic of a region with limited propensity for secondary-structure formation and increased local flexibility relative to residues in the NTD and CTD. To confirm this hypothesis, the program TALOS-N (Shen and Bax, 2013) was used to predict the secondary-structure for MecA-NTDL from the assigned chemical shifts of the backbone nuclei ( $\text{H}^{\text{N}}$ ,  $\text{N}$ ,  $\text{C}'$ ,  $\text{C}^{\alpha}$  &  $\text{C}^{\beta}$ ), and also to calculate random-coil-index-derived order-parameters ( $\text{RCI-S}^2$ ) (Berjanskii and Wishart, 2005). Panel (i) of Fig. 1C shows the  $\text{RCI-S}^2$  values (upper plot) and the TALOS-N-predicted helix/strand probabilities (lower plot). The corresponding TALOS-N secondary-structure classification is shown in between the two plots. As can be clearly seen, MecA-NTDL comprises a well-folded domain extending from residue 1 to residue 77 — corresponding to the globular NTD — followed by a long C-terminal tail that appears to be largely unstructured. The globular NTD consists of four  $\beta$ -strands and three  $\alpha$ -helices (labelled E1–E4 and H1–H3 in Fig. 1C, respectively). The  $\text{RCI-S}^2$  order parameters are mostly clustered around 0.8, as is expected for regions of defined secondary-structure within a globular domain, but there is a noticeable dip for the loop between helix H3 and strand E3, with order parameters dropping close to 0.5, revealing that this loop possesses significant local flexibility. Panel (ii) of Fig. 1C compares the secondary-structure of MecA-NTDL predicted by TALOS-N from the backbone chemical shifts with the secondary-structure of the solution NMR structure of MecA-NTD (PDB code 2mk6). The two sets secondary-structure classifications are very similar, indicating the overall structure of the NTD in MecA-NTDL is largely unchanged from that of the NTD in isolation. The remainder of MecA-NTDL, comprising the  $\sim 40$  residues C-terminal to the NTD that correspond to the inter-domain linker in flMecA, does not adopt any canonical secondary-structure, with low TALOS-predicted probabilities for both helix and strand. The  $\text{RCI-S}^2$  order parameters are also lower than for the globular NTD, with values in the range 0.2–0.6, confirming that this part of MecA-NTDL has a relatively high degree of flexibility. In summary, the secondary-structure of MecA-NTDL predicted from the backbone chemical shifts confirms its proposed description as a well-folded globular domain followed by a flexible and largely disordered tail.



**Figure 1. NMR studies of MecaA in the absence and in the presence of interaction partners.** (A) Domain architecture of MecaA. (B) Overlap of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of flMecA, MecA-NTDL and MecA-CTD. The spectra were recorded at 25 °C in a buffer consisting of 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 100 mM  $\text{Na}_2\text{SO}_4$ , 1 mM TCEP and 1 mM EDTA, 0.01%  $\text{NaN}_3$ , using a spectrometer operating at 600 MHz  $^1\text{H}$  frequency equipped with a cryogenic

probehead. (C) Secondary-structure prediction of the MecA-NTDL. (i) Results from TALOS-N analysis of the backbone chemical shift assignments. Upper plot: random-coil-index-derived squared-order-parameters (RCI-S<sup>2</sup>) plotted by residue number. Lower plot: probabilities of helical (positive, pinkred) and strand (negative, green) secondary structure plotted by residue number. The schematic secondary-structure shown between the two plots is based on the TALOS-N prediction. (ii) Comparison of the TALOS-N-predicted secondary-structure of MecA-NTDL with the solution NMR structure of the MecA-NTD (PDB code 2mk6).

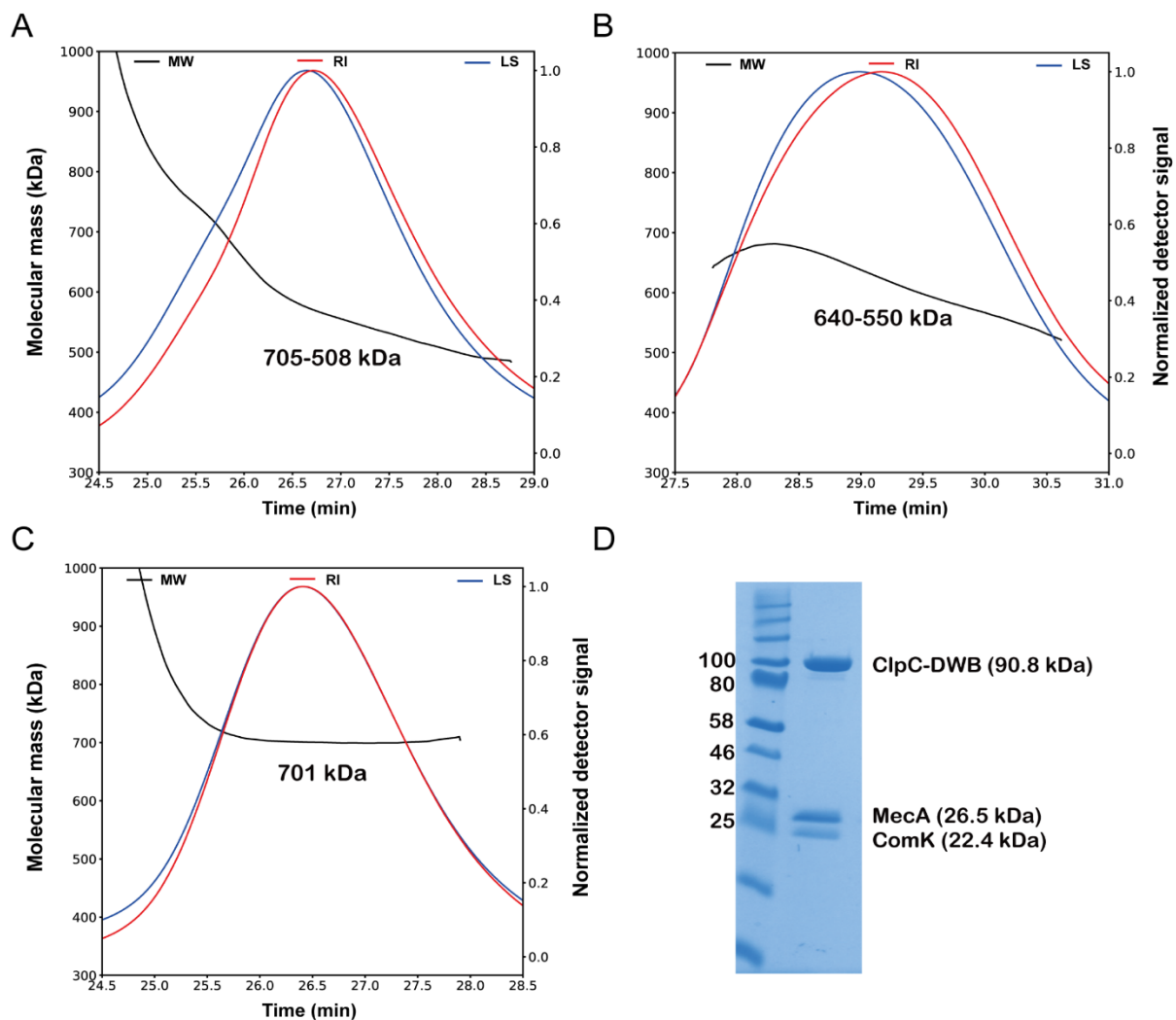
### **The MecA–ClpC complex**

MecA, as a ClpC-specific adaptor protein, performs a dual function in the ClpC-mediated unfolding processes. Firstly, the NTD of MecA has been shown to be responsible for substrate recruitment (Persuh et al., 1999; Schlothauer et al., 2003), and secondly the CTD interacts with ClpC to facilitate its assembly in the functional hexameric state (Kirstein et al., 2006). X-ray crystallography and electron microscopy structures of both truncated and full-length ClpC constructs in complex with MecA indicate a 6:6 stoichiometry ratio with MecA-CTD bridging the N-terminal domains of two ClpC monomers to form an heterododecameric complex (Carroni et al., 2017; Liu et al., 2013; Wang et al., 2011). In the electron microscopy structures, the MecA-NTD is found to be flexible above the entrance of the central chamber; however, electron density measured next to the ClpC-D1 loops has been attributed to the MecA-NTD transiently entering the unfolding chamber for auto-unfolding in cis.

Here we set out to study MecA in complex with ClpC in solution with the goal of discovering the role of its NTD in the absence of substrates. We first assembled a complex consisting of ClpC and MecA in the presence of ATP. To avoid ATP hydrolysis while retaining binding, we used a mutant of ClpC, named ClpC-DWB, that contains one mutation of the Walker B motif in each of the D1 and D2 domains (E280A/E618A). This mutant has been shown to bind ATP and assemble in a stable hexameric complex in the presence of both MecA and ATP, but to be defective in ATP hydrolysis (Carroni et al., 2017; Liu et al., 2013; Wang et al., 2011). Herein we refer to the mutant ClpC-DWB as ClpC for simplicity, but we used the ClpC-DWB construct in all our experiments.

We assembled the MecA–ClpC complex by mixing stoichiometric concentrations of ClpC and MecA and subjected the complex to size-exclusion chromatography followed by multiple angle light scattering (SEC-MALS). In the presence of ATP, ClpC alone yielded a peak of molecular weight (MW) of ~ 980 kDa corresponding to an assembly of 10–11 protomers (ClpC MW = 90.8) and consistent with previous studies (Fig. S5) (Carroni et al., 2017). The SEC-MALS data indicate that the assembly is heterogeneous. Upon addition of an equimolar concentration of MecA (MW = 26.5 kDa), the MALS profile showed the formation of a smaller

but highly heterogeneous complex with MW ranging from 508 to 705 kDa (Fig. 2A). As a 6:6 MecA:ClpC ratio would correspond to a MW of 704 kDa, we conclude that the 6:6 stoichiometry is only one of the many forms adopted by the MecA–ClpC complex, which is instead highly dynamic. To test whether the MecA-NTD is responsible for the heterogeneity of the complex, we mixed ClpC with the MecA-CTD in the presence of ATP. In this case, we obtained a heterogeneous complex with MW in the 550–640 range (Fig. 2B). We concluded that, while the presence of MecA-NTD seems to increase the heterogeneity of the MecA–ClpC particle, as demonstrated by the increase in the spread of MW when adding flMecA rather than MecA-CTD, the formation of complexes with variable stoichiometry is an intrinsic characteristic of flClpC and may explain why it has been so far impossible to obtain a crystal structure of flClpC in complex with MecA.



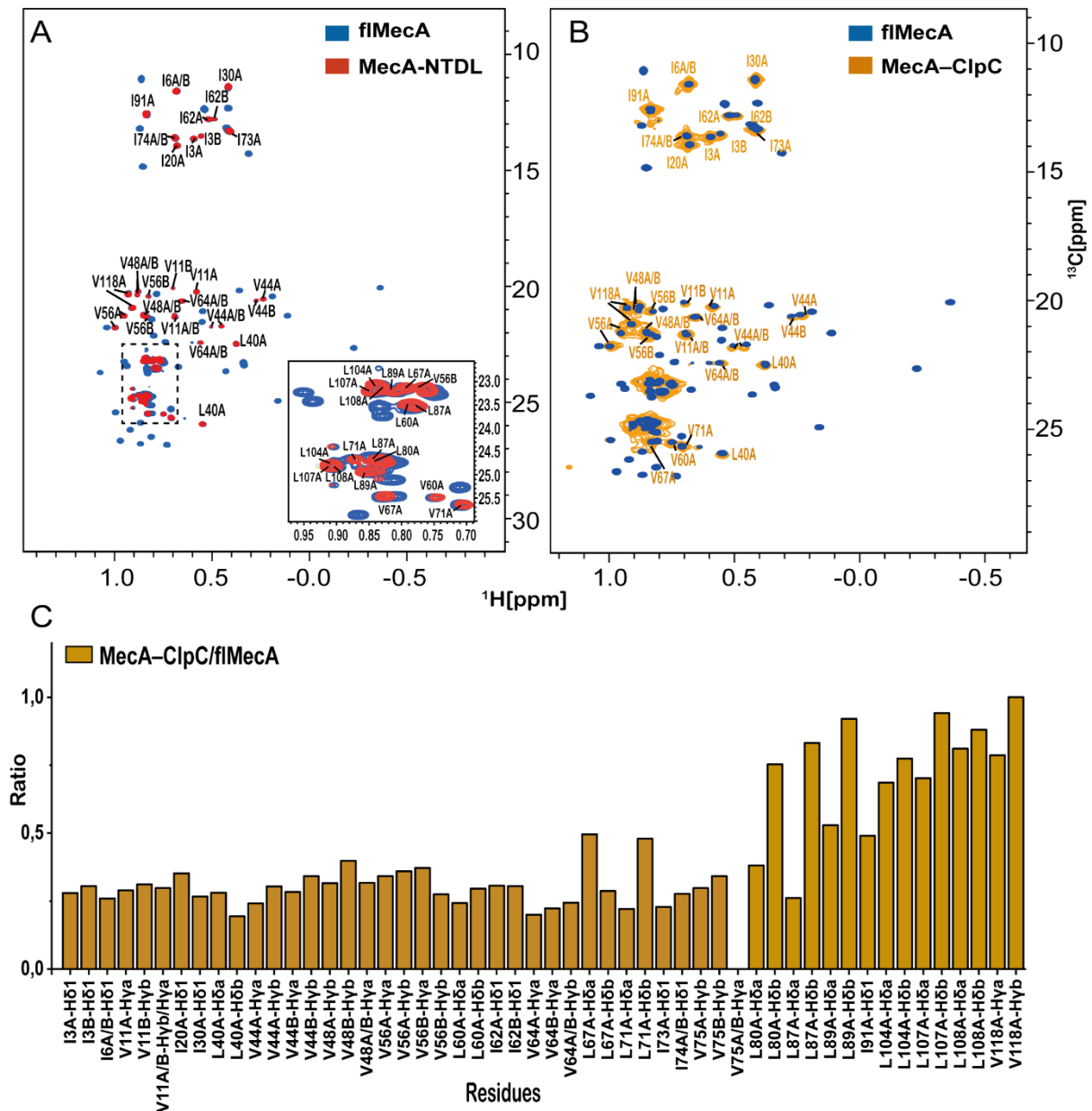
**Figure 2 SEC-MALS analysis of ClpC complexes.** (A) SEC-MALS profile of the MecA–ClpC complex obtained mixing roughly equimolar quantities of MecA and ClpC (MecA in slight excess) to reach a concentration of 35  $\mu$ M of each protein in a buffer containing 20 mM Tris-HCl, pH 7.5 100 mM NaCl, 1 mM TCEP, 1mM EDTA, 5 mM ATP and 0.01% v/v NaN<sub>3</sub>. The

offset of the light scattering-(LS) and refractive index-(RI) indicates heterogeneity in the size of the particle. The complex has molecular weights (MW) ranging from 508 kDa to 705 kDa, which corresponds to a 6:6 MecA:ClpC stoichiometric ratio. (B) SEC-MALS profile of the MecA CTD–ClpC complex obtained by mixing roughly equimolar quantities of MecA-CTD and ClpC (MecA-CTD in slight excess) to reach a concentration of 35  $\mu$ M of each protein in the same buffer as in A. The SEC-MALS profile reveals that the complex formed is heterogeneous with molecular weights (MW) ranging from 550 kDa to 640 kDa, which would correspond to a 6:6 MecA-CTD:ClpC stoichiometric ratio. (C) SEC-MALS profile of the ComK–MecA–ClpC complex obtained by mixing roughly equimolar quantities of ComK, MecA and ClpC (ComK and MecA in slight excess) to reach a concentration of 35  $\mu$ M of each protein in the same buffer as in A. The ternary complex has a homogeneous MW of 701 kDa. The presence of all three proteins in the complex has been confirmed by Tricine-SDS-PAGE, as shown in (D).

Next, we monitored MecA in the MecA–ClpC complex by methyl-group NMR spectroscopy in  $^1\text{H}$ - $^{13}\text{C}$ -HMQC spectra (Fig. 3). Firstly, we could easily transfer the assignments of ILV methyl-groups from free MecA-NTDL to free flMecA, as the NMR peaks of the NTDL amino acids in the MecA-NTDL and flMecA constructs appeared at identical positions (Fig. 3A). The methyl-group peaks not present in the MecA-NTDL  $^1\text{H}$ , $^{13}\text{C}$ -HMQC spectrum but present in the flMecA  $^1\text{H}$ - $^{13}\text{C}$ -HMQC spectrum were attributed to the MecA-CTD. Then, we assembled the MecA–ClpC complex using unlabelled ClpC and  $^1\text{H}$ , $^{13}\text{C}$ -ILV-methyl labelled, perdeuterated MecA. In this sample, we sought to selectively observe the  $^1\text{H}$  and  $^{13}\text{C}$  labelled methyl groups of MecA in a  $^1\text{H}$ - $^{13}\text{C}$ -HMQC spectrum of the complex particles.

The  $^1\text{H}$ , $^{13}\text{C}$ -HMQC spectra of free flMecA apo and MecA in complex with ClpC exhibited significant differences (Fig. 3B). When MecA bound to ClpC in the presence of ATP, many of the NMR signals corresponding to the MecA-CTD disappeared completely, indicating that the MecA-CTD adopted the rotational correlation of the complex and its NMR lines broadened beyond detection. This is in agreement with the X-ray crystallography and electron microscopy structures showing that MecA-CTD is stably sandwiched between the N-terminal domains of two ClpC protomers and thus rigidly attached to the high-MW particle formed by ClpC. In contrast, the methyl-group peaks corresponding to NTD residues broadened upon complex formation but remained visible, and those corresponding to linker residues nearly did not change. When we quantified the broadening of the NTDL methyl-group signals by calculating the intensity ratios of all MecA signals in the MecA–ClpC complex ( $I_{\text{complex}}$ ) and in free MecA ( $I_{\text{free}}$ , Fig. 3C), we observed that both sets of NTD peaks (conformers A and B) displayed a notable decrease in intensity (average of  $I_{\text{complex}}/I_{\text{free}} \sim 0.3$ ), while peaks of linker residues had values of  $I_{\text{complex}}/I_{\text{free}}$  close to 1. This observation suggests that, while the linker remains highly flexible in the complex, the MecA-NTD is in conformational exchange between a ClpC-

interacting state and a second flexible state where it wobbles with the linker. This is consistent with that observed in electron microscopy structure (Liu et al., 2013) and confirms that the MecA-NTD functions as a substrate for ClpC in the absence of any other substrate. The recruitment of MecA-NTD to the ClpC unfolding chamber occurs in *cis*, as, in our experiments we measure a purified MecA–ClpC complex and the lack of any peaks corresponding to the MecA-CTD demonstrates that all MecA is bound to ClpC.



**Figure 3 Methyl-TROSY spectra of  $^1\text{H}$ ,  $^{13}\text{C}$ -ILV-methyl/ $^2\text{H}$ -labeled MecA in the MecA–ClpC complex.** (A) Superimposition of  $^1\text{H}$ ,  $^{13}\text{C}$ -HMQC spectra of free flMecA (blue) and free MecA-NTDL (red) with assignment transferred from MecA-NTDL to flMecA. (B) Overlay of  $^1\text{H}$ ,  $^{13}\text{C}$ -HMQC spectra of free flMecA (blue) and flMecA in complex with ClpC (orange) with

assignments transferred from A. (C) Ratios of peaks' intensities in the spectra of flMecA in the MecA–ClpC complex and in the apo form. Only assigned peaks corresponding to the NTDL region are shown. A and B refer to the two conformers of Figure S4.

### **The MecA–ComK interaction**

While it is known that MecA-NTD is responsible for the recruitment of ComK to the MecA–ClpC complex (Kirstein et al., 2006), the interaction between ComK and MecA-NTD has not been characterized in detail. Here, we were able to express and purify ComK as a stable tetramer in high salt buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 100 mM Na<sub>2</sub>SO<sub>4</sub>, 1 mM EDTA, 1 mM TCEP, 0.01% NaN<sub>3</sub>, Fig. S6A). To verify direct binding of MecA-NTD to ComK, we recorded <sup>15</sup>N-TROSY spectra of 50 μM <sup>15</sup>N,<sup>2</sup>H-labeled ComK in the presence of one equivalent of either flMecA or MecA-CTD. Many ComK <sup>1</sup>H-<sup>15</sup>N peaks disappeared after addition of flMecA, while they were still visible in the presence of MecA-CTD, confirming that the interaction is mediated by the MecA-NTDL (Fig. S6B and C).

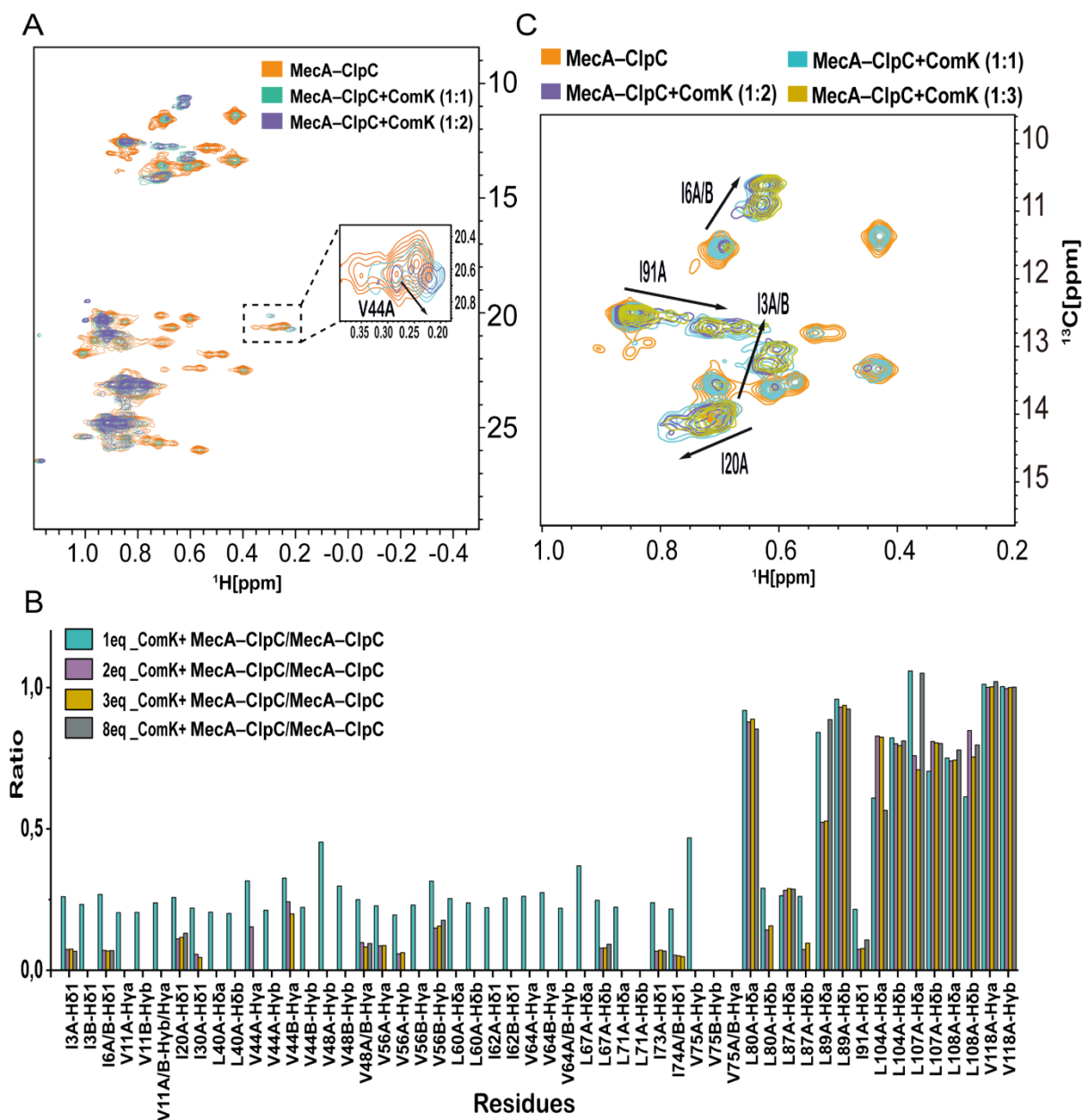
### **The MecA–ClpC–ComK complex**

To understand how ComK interacts with MecA-NTD in the MecA–ClpC complex and in particular to find out how many copies of ComK are recruited for degradation, we titrated <sup>2</sup>H,<sup>15</sup>N-labelled ComK to the MecA–ClpC complex assembled using unlabelled ClpC and <sup>1</sup>H,<sup>13</sup>C-ILV-methyl labelled, perdeuterated MecA. The purified MecA–ClpC complex was divided in seven aliquots to which 0, 1, 2, 3, 4, 6 and 8 equivalents of ComK were added. The equivalents of ComK were calculated with respect to the molarity of the MecA–ClpC complex, thus, one equivalent of ComK corresponds to one copy of ComK and one copy of the average MecA–ClpC complex. The concentration of the complex was calculated assuming a MW of ~650 kDa. Because the peak is inhomogeneous, this assumption is only an approximation, and the concentration of the complex could have been miscalculated by ~10–15%. However, this error is unlikely to impact the results presented below.

We then monitored the methyl-group peaks of MecA in <sup>1</sup>H,<sup>13</sup>C-HMQC spectra (Fig. 4A). As expected, the signals of the methyl groups of MecA-NTD notably broadened upon addition of ComK. When we quantified this line broadening by calculating the intensity ratios of the MecA methyl-groups in the presence and in the absence of ComK ( $I_{\text{complex,ComK}}$  and  $I_{\text{complex}}$ , respectively), we observed that addition of one equivalent of ComK led to a drop in intensity of ~70% and that the intensity loss plateaued at two equivalents of ComK around a value of 0–10% (Fig. 4B). Thus, our NMR titrations indicated that a maximum of two ComK copies binds one MecA–ClpC complex particle. A few methyl groups did not disappear completely, but became weaker, while new peaks appeared in the vicinity (Fig. 4C). Thus, for a few residues multiple peaks were present in the spectra, beyond the two sets corresponding to



the two conformations of MecA-NTDL (conformers A and B). This process, which is also saturated after addition of 2 equivalents of ComK, may indicate that not all copies of MecA present in the complex experience the same chemical environment. Opposite to the MecA-NTD methyl groups, the methyl groups corresponding to the linker region, in particular after residue 91, did not lose their intensity upon addition of ComK, suggesting that the linker does not participate in the interaction and moreover remains flexible in the MecA–ClpC–ComK complex (Fig. 4B).

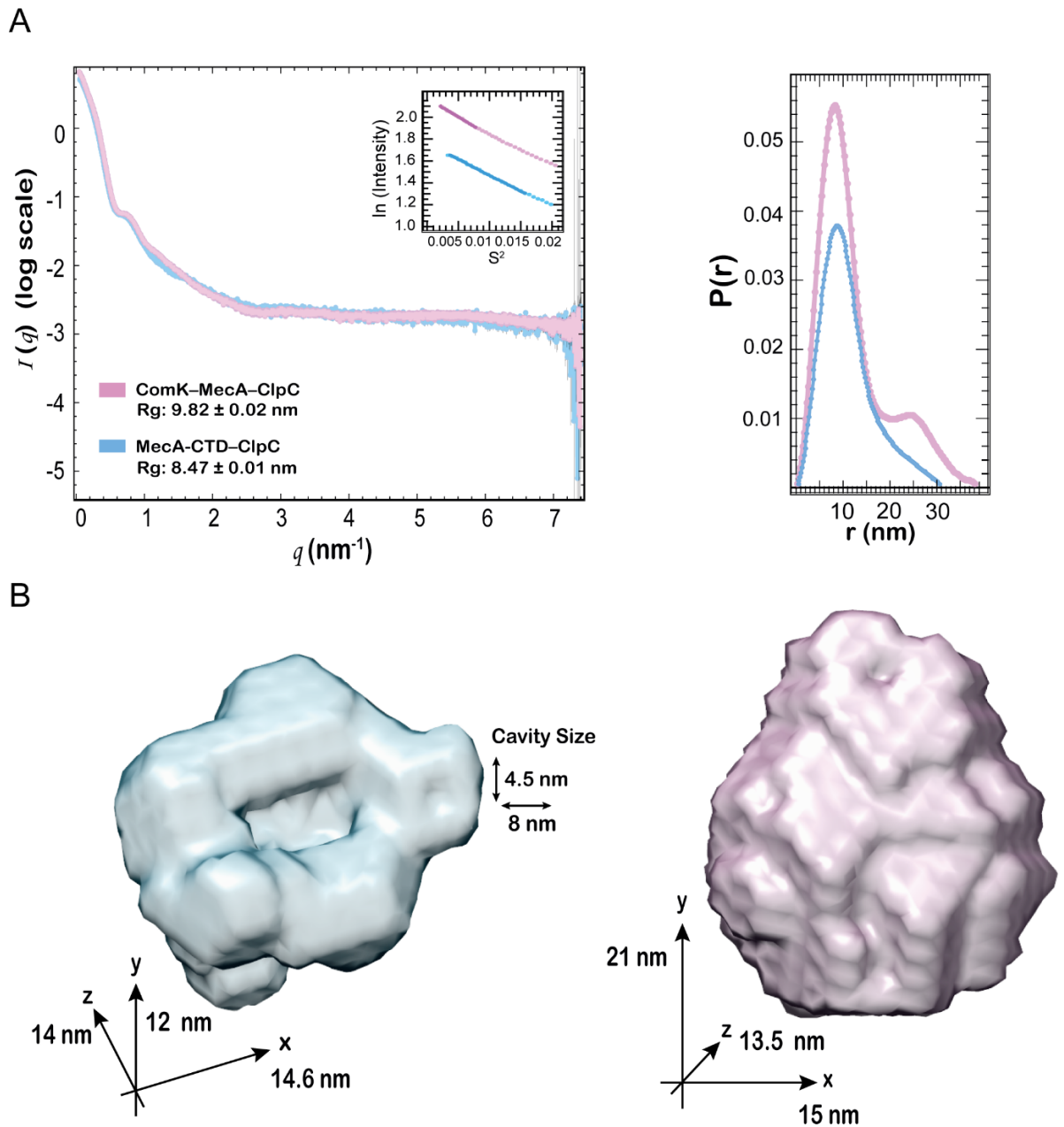


**Figure 4. NMR titrations of ComK to the MecA–ClpC complex.** (A) Overlay of  $^1\text{H}$ ,  $^{13}\text{C}$ -HMQC spectra of  $^1\text{H}$ ,  $^{13}\text{C}$ -ILV-methyl/ $^2\text{H}$ -labeled MecA in 48  $\mu\text{M}$  MecA–ClpC complex (orange) and upon addition of 1 (green) and 2 (purple) molar equivalents of ComK. The molar

equivalents refer to the molarity of one copy of the average MecA–ClpC complex and not to the individual MecA and ClpC proteins. (B) Ratios of peaks' intensities in the spectra of flMecA in the ComK–MecA–ClpC complex and in the MecA–ClpC complex. Upon addition of more than 2 molar equivalents of ComK, no further reduction of peak intensities is observed. (C) A few peaks of the MecA-NTD region of fl-MecA remain visible with ~10% of intensity, while new peaks appear nearby.

To further investigate the stoichiometry of the MecA–ClpC–ComK complex, we run a SEC–MALS experiment of a complex assembled with equimolar concentrations of MecA, ClpC and ComK in the presence of 5 mM ATP (Fig. 2C). We obtained a homogeneous complex of 701 kDa that contained all three proteins (Fig. 2D). This molecular weight is incompatible with six copies of MecA being present in the MecA–ClpC–ComK complex, as the MW of a 6:6 MecA–ClpC complex equals ~704 kDa. Pulling together the NMR data, which indicate recruitment of two copies of ComK by the MecA–ClpC complex, and the SEC-MALS data, which measure a MW of 701 kDa for the MecA–ClpC–ComK complex, we conclude that the trimeric MecA–ClpC–ComK complex can contain a maximum of four copies of MecA. (Note that molecular weights of ClpC, MecA and ComK are 90.8, 26.5 and 22.4 kDa, respectively).

Finally, we measured Small-Angle-X-Ray-Scattering curves of the MecA-CTD–ClpC and ComK–MecA–ClpC complexes. As expected, the radius of gyration ( $R_g$ ) of the ComK–MecA–ClpC complex is larger than that of the MecA-CTD–ClpC complex and the distance distribution function  $P(r)$  shows that the trimeric complex adopts an elongated structure (Fig. 5A). Despite the heterogeneity of the composition of the MecA-CTD–ClpC complex, the *ab initio* shape reconstruction yielded a cylinder with a squeezed internal cavity of dimensions 4.5x8x12 nm (Fig. 5B). We think that the squeeze of the cavity is not real but is a consequence of the heterogeneous nature of the complex. The homogeneous ComK–MecA–ClpC has a well-defined half conical shape, with the basis of the cone having dimension of ~14x14 nm, similar to the MecA-CTD–ClpC complex and to the crystal structure of a 6:6 MecA-CTD–ClpC complex derived using a truncated ClpC construct (Wang et al., 2011). On top of the barrel a conical cap closes the pore and brings the overall vertical dimension to 21 nm (Fig. 5B). This low-resolution envelope demonstrates that in the ComK–MecA–ClpC complex, the MecA-NTD and ComK subunits are not disordered but interact with the ClpC barrel from above the pore.



**Figure 5. Small-Angle X-Ray Scattering (SAXS) analysis of MecA-CTD-ClpC and ComK-MecA-ClpC complexes.** (A) Left, comparison of SAXS profiles of MecA-CTD-ClpC- and ComK-MecA-ClpC complexes. The ternary complex exhibits a higher  $R_g$  than the MecA-CTD-ClpC complex. The Guinier plot and fit from these scattering curves are shown as an inset. Right, the increased maximum distance of the  $P(r)$  function together with the shape of the function suggest the presence of a small domain next to a larger domain in the ComK-MecA-ClpC complex. (B) SAXS-derived *ab initio* envelopes of MecA-CTD-ClpC (left) and in ComK-MecA-ClpC (right) complexes with dimensions.

## Discussion

The assembly architecture of ClpC has been object of several studies. Early crystal structures of a truncated ClpC construct with D2 deleted (ClpC- $\Delta$ D2; residues 1–485, Glu 280 Ala,  $\Delta$ 247–251 and  $\Delta$ 281–292) in complex with MecA-CTD revealed a 6:6 ClpC:MecA stoichiometry (Wang et al., 2011) in the absence of both nucleotide and magnesium. In the same work, a low-resolution structure of a non-truncated version of ClpC (Glu 280 Ala, Glu 618 Ala,  $\Delta$ 247–251,  $\Delta$ 281–292,  $\Delta$ 587–596 and  $\Delta$ 665–685) bound to MecA<sup>108-208</sup> was obtained in the presence of AMP-PNP, which also revealed a 6:6 stoichiometry. Finally, a low-resolution electron-microscopy structure of wild-type ClpC in complex with flMecA in the presence of ATP and magnesium (Carroni et al., 2017) revealed a 6:6 ClpC:MecA stoichiometry as well, which was confirmed by SEC-MALS and quantitative gel analysis. As opposed to other studies that used SEC-MALS to establish the MW of the MecA–ClpC complex (Carroni et al., 2017; Wang et al., 2011), we do not see inhomogeneity at MW higher than that corresponding to a 6:6 MecA–ClpC complex, but rather at lower MWs (Figure 2). Except for a small percentage of complex containing less than six copies of ClpC, in our study most of the complex seems to contain a range of MecA copies spanning from one (~570 kDa) to six (~700 kDa). This inhomogeneity appears an intrinsic characteristic of the complex and may be the reason why the electron microscopy structure of the MecA–ClpC particle failed to reach a resolution better than 11 Å. However, when we assemble the complex in the presence of magnesium (50 mM sodium phosphate, pH 7.0, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 2mM TCEP, 5 mM ATP and 0.01% NaN<sub>3</sub>), we obtain a SEC-MALS profile that is not only more homogeneous but also indicative of larger particles (Fig. S7, MW range = 655–725 kDa). Thus, we conclude that magnesium is needed to assemble a stable complex of MecA–ClpC with a well-defined stoichiometry, besides being needed for ATP hydrolysis.

Our NMR analysis of the MecA–ClpC complex suggests that the N-terminal part of MecA is in transient contact with ClpC. This observation is consistent with that observed in electron microscopy structures (Kirstein et al., 2006). The interaction of the MecA-NTD with ClpC in *cis* may represent a mechanism through which MecA becomes a substrate for ClpC in the absence of any other substrate. Previously, it has been proposed that the MecA–ClpC complex may recruit MecA in *trans* for degradation (Persuh et al., 1999). However, our NMR studies fail to detect any interactions between MecA-NTD and any other MecA domain. Only MecA-CTD shows a tendency to oligomerize; however, when in complex with ClpC, MecA-CTD is involved in other interactions and is unable to recruit further MecA copies to the complex. Thus, our NMR analysis favors the model by which the MecA-NTD–ClpC interactions in *cis* initiate MecA unfolding and degradation.

As opposed to the MecA–ClpC complex, the ComK–MecA–ClpC complex has a well-defined size of 701 kDa also in the absence of magnesium, suggesting that the interaction with substrate stabilizes the active conformation of ClpC. Our NMR titrations indicate that a maximum of two copies of ComK bind the MecA–ClpC complex. In the absence of magnesium, binding of ComK to form the ComK–MecA–ClpC seems to be incompatible with the presence of six copies of MecA, as the complex appears too small to accommodate six copies of MecA and two copies of ComK. Whether this incompatibility is released by magnesium binding will be revealed by future experiments.

In conclusion, we have demonstrated that the MecA–ClpC complex can exist in a multitude of compositions in the presence of ATP and in the absence of magnesium, but that this inhomogeneity is released by ComK binding to yield a stable complex. This plasticity of the MecA–ClpC complex is likely relevant for its function and turnover. Furthermore, we have shown that MecA-NTD does not interact with other MecA structural elements, suggesting that MecA functions as a substrate to ClpC in *cis*.

## Methods

**Protein expression and purification.** Plasmids of ClpC (UniProtKB accession code P37571) double-walker B mutant (ClpC-DWB, E280A/E618A) and full-length (fl) MecA (P37958) were cloned into pET28a(+), containing a Tobacco Etch Virus (TEV) protease cleavage site, followed by a C-terminal hexa-histidine tag. The DNA sequence of ComK (P40396) was amplified and cloned into pETM11-Sumo3 expression vector (European Molecular Biology Laboratory collection). Four truncated MecA constructs (MecA-NTD, MecA<sup>1-77</sup>; MecA-NTDL, MecA<sup>1-120</sup>; MecA-CTD, MecA<sup>121-218</sup>; MecA-CTDL, MecA<sup>78-218</sup>) were generated with the Quik-change II Site-Directed Mutagenesis protocol (Agilent Technologies). All constructs were confirmed by DNA sequencing.

The recombinant ComK and MecA variants were transformed into *Escherichia coli* BL21 (DE3) and expression was induced by addition of 0.6 mM isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) at OD 600 = 0.6–0.8 and continued for 18–20 h at 16 °C. For ClpC-DWB, expression was also performed in BL21 (DE3) *E. coli* strains, induced with 1 mM IPTG for ~18 hours at 25 °C. After harvesting (4 °C, 4000 rpm for 40 min), cell pellets were stored at –20 °C.

For protein purification, cell pellets of ClpC-DWB and MecA variants were resuspended in ice-cold buffer A (50 mM Tris-HCl, 500 mM sodium chloride, 100 mM sodium sulfate, 5% (v/v) glycerol, 10 mM imidazole, 1 mM Tris-(2-carboxyethyl)-phosphine (TCEP), pH 7.5) with one tablet of EDTA-free protease inhibitor cocktail (Roche), 100  $\mu$ g of lysozyme (Carl Roth), and 50  $\mu$ g of deoxyribonuclease (DNAse) (New England Biolabs) per 1 L of original culture volume. Cells were lysed via sonication (50% power; 5 sec pulse/10 sec rest duty-cycle, for 30 mins). Total cell lysate was cleared at 18000 rpm for 1 h at 4 °C. The supernatant was filtered and loaded onto a HisTrap HP column (GE Healthcare), which was equilibrated and washed with buffer A, and then eluted with a linear gradient from buffer A to buffer B (50 mM Tris-HCl, 500 mM sodium chloride, 100 mM sodium sulfate, 5% (v/v) glycerol, 300 mM imidazole, 1 mM Tris-(2-carboxyethyl)-phosphine (TCEP), pH 7.5). The C-terminal histidine tag was then cleaved using TEV protease (0.5 mg per 1 L of original culture volume) overnight at 4 °C in buffer A. To isolate the cleaved proteins, the mixture was injected onto a HisTrap HP column equilibrated in buffer A and the flow-through was collected. Pure fractions were pooled, concentrated, and purified further on HiLoad Superdex 200 16/60 or HiLoad Superdex 75 16/60 size-exclusion chromatography columns (GE Healthcare) equilibrated in buffer C (50 mM Tris-HCl, 500 mM NaCl, 100 mM Na<sub>2</sub>SO<sub>4</sub>, 1 mM TCEP, 1mM EDTA, 0.01% v/v NaN<sub>3</sub>, pH 7.5). Pure protein fractions were identified by SDS-PAGE. Purification of ComK was performed using the same protocol above, except for an extra washing step with buffer A containing additional 2 M LiCl, before elution with a linear gradient from buffer A to buffer B. In addition, the SUMO-tagged

ComK was cleaved by Sentrin-specific protease 2 (SEN2; 1:1000 protease: protein ratio) overnight in buffer A. All steps of the ComK purification were done at 4 °C.

Production of <sup>15</sup>N-labeled MecA and its four truncated constructs was performed in H<sub>2</sub>O M9 minimal medium supplemented with <sup>15</sup>NH<sub>4</sub>Cl (Cambridge Isotope Laboratories); perdeuterated, <sup>15</sup>N-labeled MecA, MecA-NTDL and MecA-CTD were expressed in 100% D<sub>2</sub>O M9 minimal medium supplemented with <sup>2</sup>H-glucose (Cambridge Isotope Laboratories) and <sup>15</sup>NH<sub>4</sub>Cl. Isoleucine, leucine, and valine (ILV) methyl-labeled MecA protein was expressed in 100% D<sub>2</sub>O M9 minimal medium supplemented with d<sub>7</sub>-D-glucose (Cambridge Isotope Laboratories), α-ketobutyric acid sodium salt (60 mg/L, Sigma, cat. no. 589276) and α-ketoisovaleric acid sodium salt (120 mg/mL, Sigma, cat. no. 691887). Sparse deuterated, <sup>15</sup>N labeled ComK was produced in 99% D<sub>2</sub>O M9 minimal medium with <sup>15</sup>NH<sub>4</sub>Cl. <sup>13</sup>C, <sup>15</sup>N-labeled MecA-NTDL was produced in H<sub>2</sub>O M9 minimal medium containing 1 g/L <sup>15</sup>NH<sub>4</sub>Cl, 2.5 g/L <sup>13</sup>C<sub>6</sub>-D-glucose (Cambridge Isotope Laboratories).

Labeled and truncated proteins were purified identically to the unlabeled wild-type proteins.

**Complex assembly.** For MALS/SAXS experiments: MecA and ClpC-DWB (herewith abbreviated to ClpC) were mixed in a 1.1:1 molar ratio in buffer D (20 mM Tris-HCl, pH 7.5 100 mM NaCl, 1 mM TCEP, 1mM EDTA, 5 mM ATP, 0.01% v/v NaN<sub>3</sub>) before injection onto a size-exclusion chromatography (SEC) column. The MecA–ClpC complex was separated by SEC in the same buffer. The ComK–MecA–ClpC ternary complex was assembled by mixing the proteins in a 1.1:1.1:1 molar ratio in buffer D and subsequently purified by SEC. Pure fractions were identified by SDS-PAGE, pooled, and concentrated for MALS/SAXS measurement. All SEC runs were performed with a Superose 6 10/300 column (GE Healthcare).

For NMR experiments: the MecA–ClpC complex was separated by SEC in buffer D. Titrations were performed with 50-200 μM ComK added to the MecA–ClpC complex in 1, 2, 3, 4, 6, 8 molar ratios. The ternary ComK–MecA–ClpC complex was buffer exchanged into NMR buffer (100% D<sub>2</sub>O based buffer D) by means of repeated dilution/concentration steps using Amicon centrifugal concentrators (10-kDa molecular-weight cutoff).

**NMR spectroscopy.** NMR experiments were recorded at a temperature of 298 K on two Bruker Avance III-HD spectrometers running Topspin 3.2 software, with <sup>1</sup>H field-strengths of 850 MHz and 600 MHz, equipped with inverse HCN CP-TCI (helium-cooled) and CPP-TCI (nitrogen-cooled) cryogenic probeheads, respectively. NMR data were processed with the NMRPipe (Delaglio, F. et al. 1995) or Topspin (Bruker) software packages and analyzed within CcpNmr Analysis v2.4 (Vranken et al., 2005).

$^{15}\text{N}$ -HSQC spectra of flMecA and its four truncated constructs were recorded at various protein concentrations in 50mM sodium phosphate, pH 7.0, 200 mM NaCl, 10 mM  $\text{MgCl}_2$ , 1 mM TCEP, 0.01%  $\text{NaN}_3$ . Titrations of unlabeled flMecA and MecA-CTD onto  $^2\text{H}$ ,  $^{15}\text{N}$ -labeled ComK were performed in 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 100 mM  $\text{Na}_2\text{SO}_4$ , 1 mM TCEP, 1 mM EDTA, and 0.01%  $\text{NaN}_3$ . Perdeuterated,  $^1\text{H}$ ,  $^{13}\text{C}$ -ILV methyl-labeled free flMecA was measured at a concentration of  $\sim 80$   $\mu\text{M}$  in 100%  $\text{D}_2\text{O}$  based NMR buffer (20 mM Tris-HCl, pH 7.5 100 mM NaCl, 1 mM TCEP, 1 mM EDTA, 5 mM ATP). The  $^1\text{H}$ ,  $^{13}\text{C}$ -ILV methyl-labeled MecA in complex with non-labelled ClpC was first separated by SEC, then prepared at a concentration of  $\sim 0.2$  mM in the same NMR buffer.  $^2\text{H}$ ,  $^{15}\text{N}$ -labeled ComK was titrated into the MecA–ClpC complex in six different molar ratios (1:0, 1:1, 1:2, 1:3, 1:4, 1:6, 1:8) to record the methyl groups of the  $^1\text{H}$ ,  $^{13}\text{C}$ -ILV methyl-labeled flMecA in the context of ComK–MecA–ClpC complex.  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled MecA-NTDL was prepared at 0.5 mM concentration in 20 mM Tris-HCl, pH 7.5 100 mM NaCl, 1 mM TCEP, 1 mM EDTA.

2D  $^1\text{H}$ – $^{15}\text{N}$  and  $^1\text{H}$ – $^{13}\text{C}$  (ILV-methyl) correlation spectra were acquired using  $^{15}\text{N}$ -HSQC ( $^{15}\text{N}$  heteronuclear single-quantum coherence),  $^{15}\text{N}$ -TROSY ( $^{15}\text{N}$  Transverse Relaxation-Optimized Spectroscopy) and  $^{13}\text{C}$ -HMQC ( $^{13}\text{C}$  heteronuclear multiple-quantum coherence) pulse sequences (Bodenhausen and Ruben, 1980; Ollerenshaw et al., 2003; Pervushin et al., 2009; Piotto et al., 1992; Sklenar et al., 1993). Protein backbone resonance assignments of the  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled MecA-NTDL were obtained from a standard-suite of 3D triple-resonance experiments comprising HNCO (Ikura et al., 1990; Kay et al., 1990), HN(CA)CO (Clubb et al., 1992; Kay et al., 1994), HNCA, HNCACB (Grzesiek and Bax, 1992; Wittekind and Mueller, 1993), HN(CO)CA and HN(CO)CACB spectra (Bax and Ikura, 1991; Yamazaki et al., 1994). ILV-methyl resonances were assigned using 3D HC(C)H–total correlation spectroscopy (TOCSY) (Bax et al., 1990; Kay et al., 1993), 3D H(CCCO)NH (Logan et al., 1992; Montelione et al., 1992). 2D  $^1\text{H}$ – $^{13}\text{C}$  correlation spectra for sidechain assignment purposes were recorded as  $^{13}\text{C}$ -HSQCs with both real-time and constant-time  $^{13}\text{C}$  chemical-shift evolution periods (Santoro and King, 1992; Vuister and Bax, 1992).

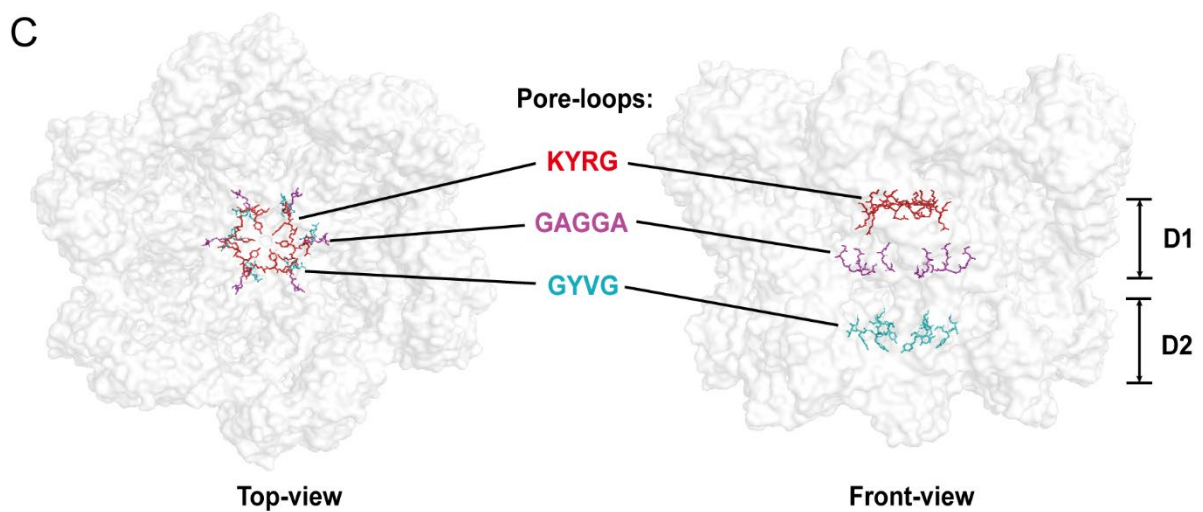
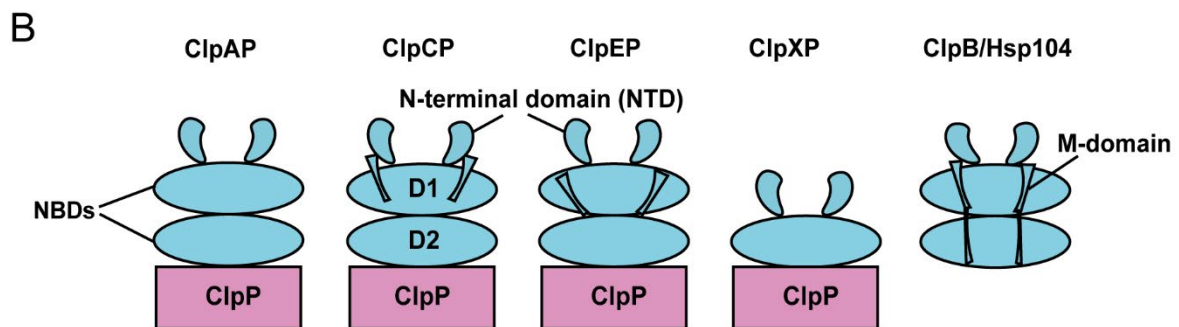
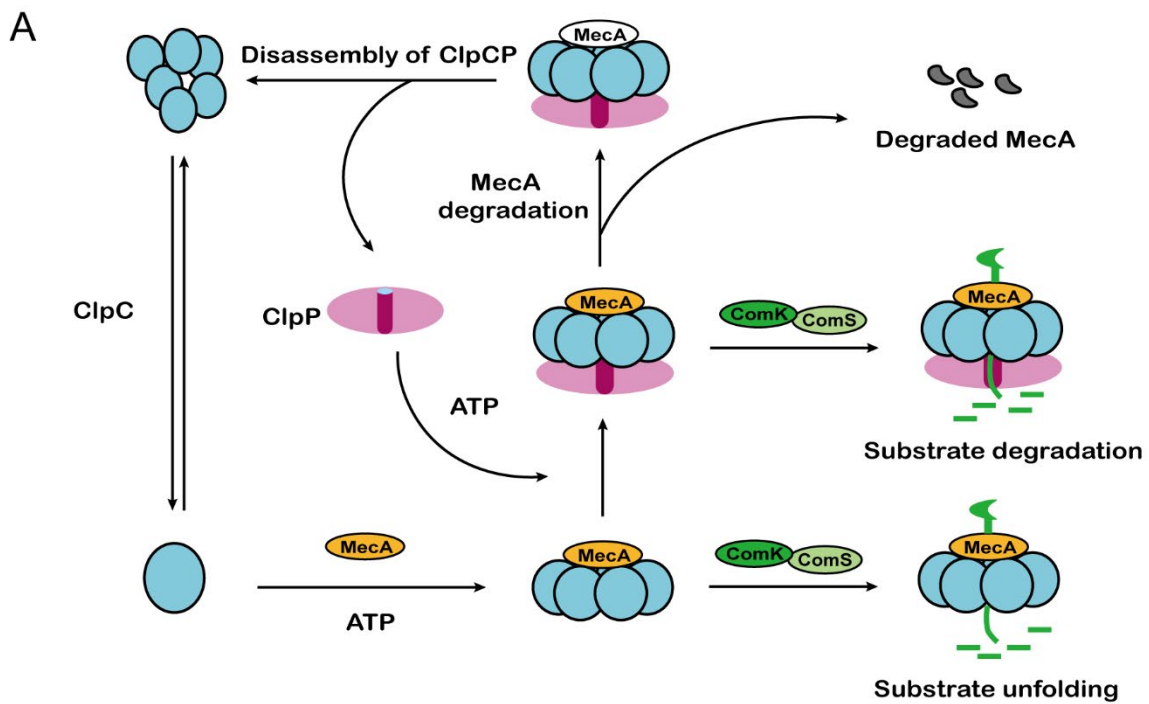
Peak intensities were obtained as peak heights and the spectral noise was taken as the experimental error.

**Multi-angle-light scattering (MALS) experiments.** The measurements were performed using an in-line SEC-MALS set-up, including a Superose 6 10/300 (GE Healthcare) column on an Åkta Pure system coupled with a Optilab T-rEX detector and mini-DAWN TREOS system (Wyatt). All samples were prepared in 20 mM Tris-HCl, pH 7.5 100 mM NaCl, 1 mM TCEP, 1 mM EDTA, 5 mM ATP, 0.01% (v/v)  $\text{NaN}_3$ . Data were analyzed using the ASTRA 7.0 software package (Wyatt Technologies).

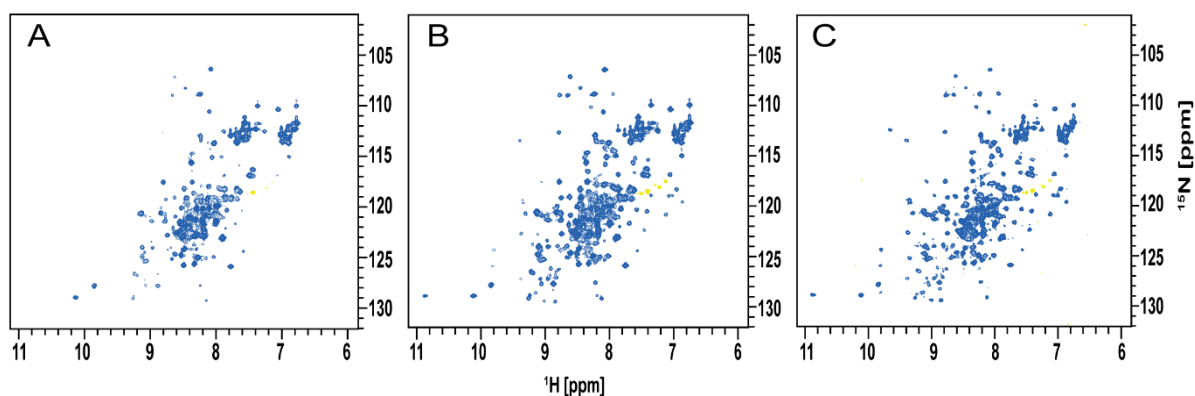


**SAXS data collection and analysis.** Before measurement, concentrated samples were prepared in fresh buffer D (20 mM Tris-HCl, pH 7.5 100 mM NaCl, 1 mM TCEP, 1 mM EDTA, 5 mM ATP, 0.01% (v/v) NaN<sub>3</sub>). Samples of MecA–ClpC and ComK–MecA–ClpC were diluted to 5, 10, 18 and 25 mg/ml. SAXS curves were measured at the European Molecular Biology Laboratory in Hamburg, Germany, on the beamline P12.

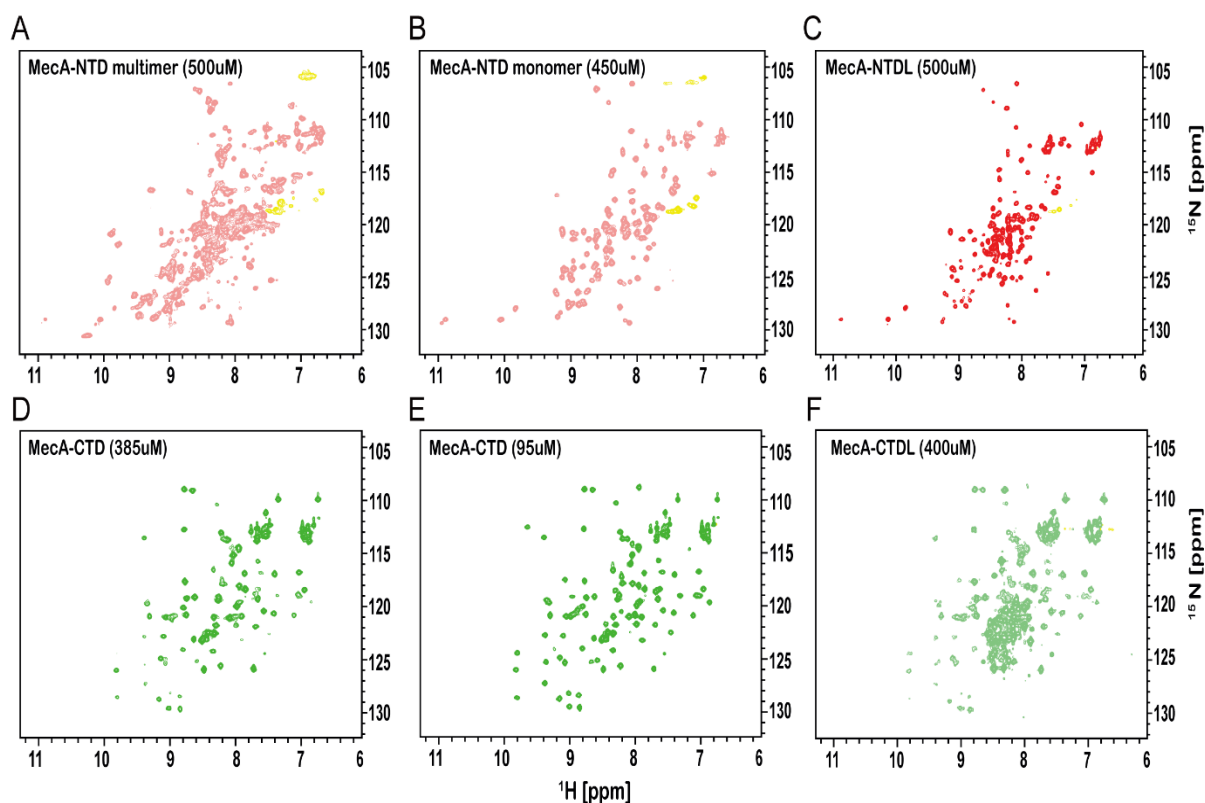
## Supplemental Figures



**Figure S1. Schematics of ClpC activity and architecture.** **A.** ClpC can act either in complex with ClpP to unfold and degrade proteins or by itself, as unfolding and chaperoning machinery. The activation of ClpC requires binding to the co-factor protein MecA, which also functions to bind and recruit the substrate. ComK and ComS are competitive substrates of the MecA–ClpC complex. In the absence of other protein substrates, MecA can become a substrate to ClpC. **B.** Domain architecture of ClpC and other proteins of the Hsp100/Clp protein family. **C.** Electron microscopy structure of 3J3S showing the position of pore-loops 1–3 in the ClpC barrel (Liu et al., 2013).



**Figure S2. Concentration-dependent oligomeric states of MecA.** <sup>1</sup>H, <sup>15</sup>N-HSQC spectra of MecA measured at three different protein concentrations: (A) 690 μM; (B) 315 μM and (C) 30 μM. The peaks sharpen up with decreasing protein concentrations demonstrating transition to a lower molecular-weight state. The spectra were recorded at 25 °C in a buffer consisting of 50mM sodium phosphate, pH 7.0, 200 mM NaCl, 10 mM MgCl<sub>2</sub>, 1 mM TCEP, 0.01% NaN<sub>3</sub>, using a spectrometer operating at 850 MHz <sup>1</sup>H frequency equipped with a cryogenic probehead.



**Figure S3.  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -labeled truncated MecA constructs.** (A-B) SEC purification of the MecA-NTD yields two peaks, one corresponding to a multimeric species, the  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum of which is shown in (A) and one corresponding to monomeric MecA-NTD, the  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum of which is shown in B. (C)  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum of MecA-NTDL. (D-E)  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum of MecA-CTD at two different concentrations. (F)  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum of MecA-CTDL. All spectra were recorded at 25 °C in a buffer consisting of 50mM sodium phosphate, pH 7.0, 200 mM NaCl, 10 mM  $\text{MgCl}_2$ , 1 mM TCEP, 0.01%  $\text{NaN}_3$ , using a spectrometer operating at 600 MHz  $^1\text{H}$  frequency equipped with a cryogenic probehead.

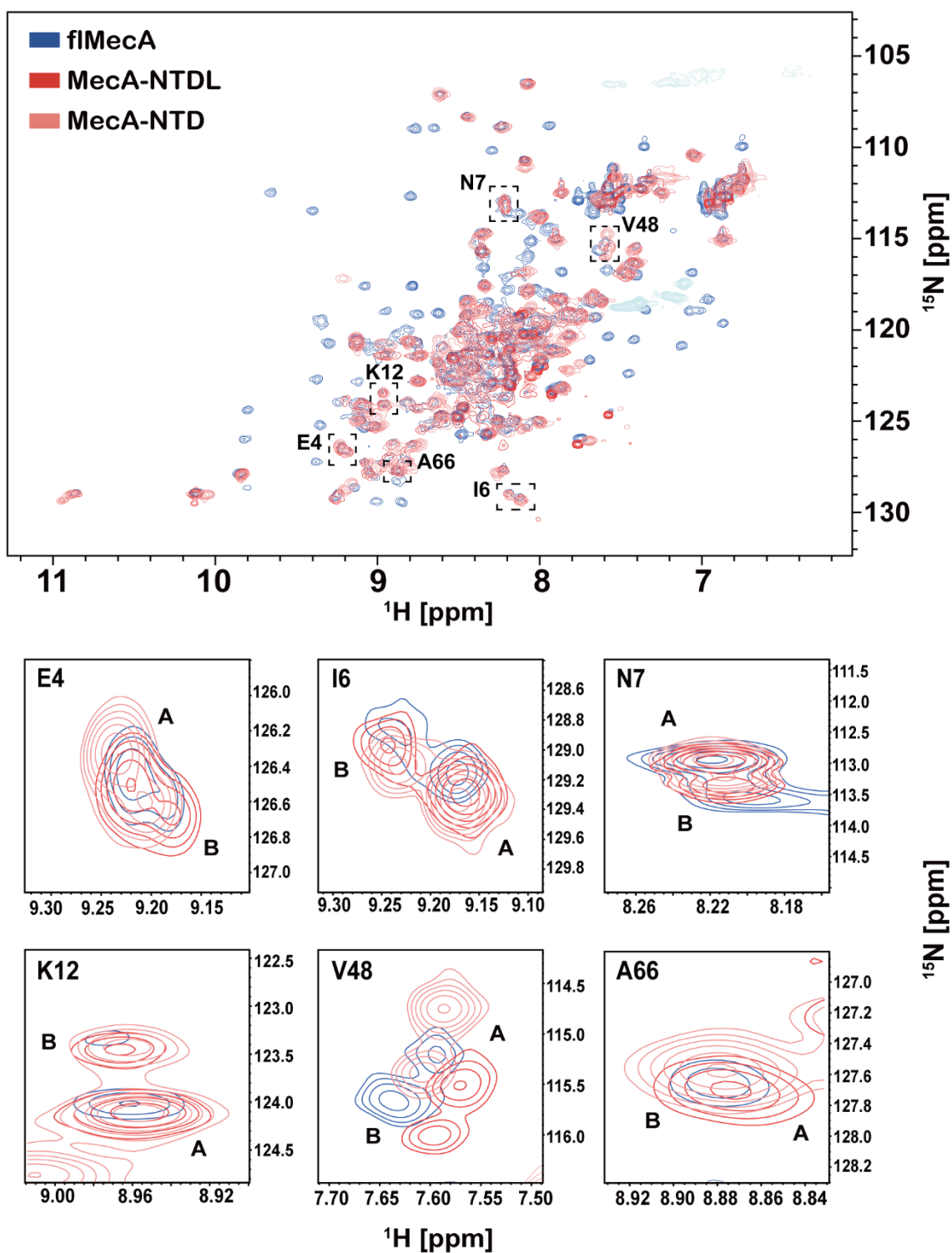
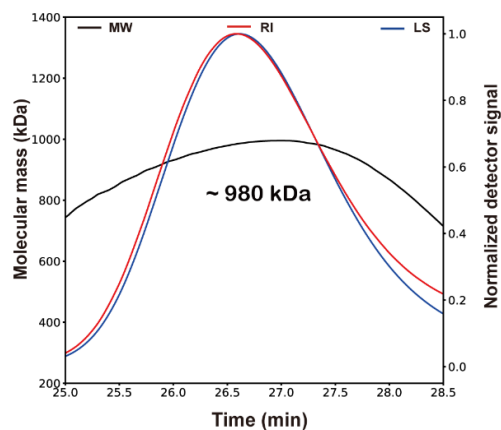
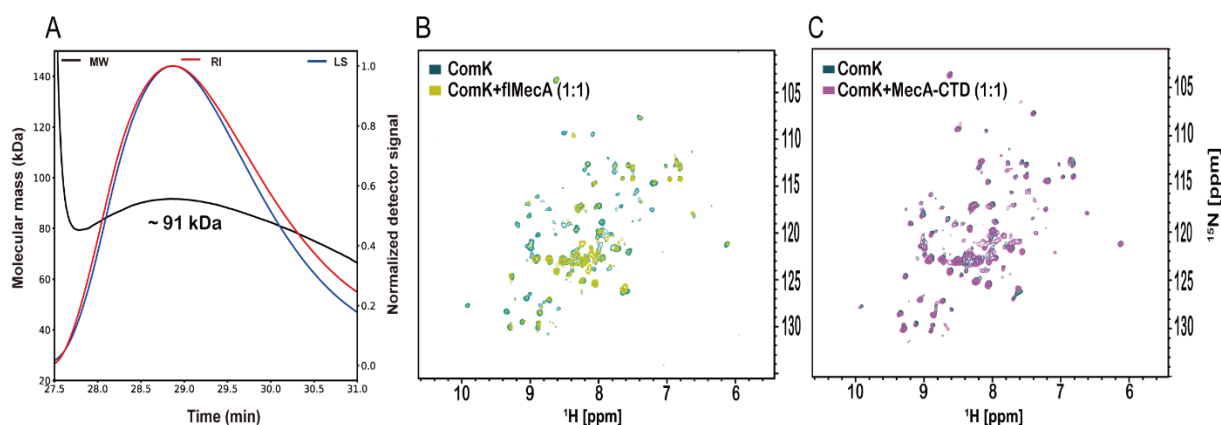


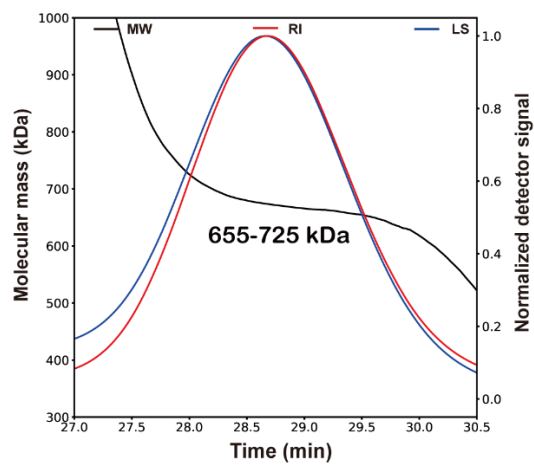
Figure S4. Figure shows the two conformations (A and B) of MecA-NTD in NTD, NTDL and fIMecA constructs.



**Figure S5. SEC-MALS analysis of ClpC-DWB in the absence of adaptor proteins.** SEC-MALS profile of 40  $\mu$ M ClpC-DWB in a buffer containing 20 mM Tris-HCl, pH 7.5 100 mM NaCl, 1 mM TCEP, 1mM EDTA, 0.01% v/v  $\text{NaN}_3$  and 5mM ATP.



**Figure S6. Interaction of ComK with MecA in the absence of ClpC.** (A) SEC-MALS profile of 50  $\mu$ M ComK in the buffer containing 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 100 mM  $\text{Na}_2\text{SO}_4$ , 1 mM TCEP, 1 mM EDTA and 0.01%  $\text{NaN}_3$ . In this buffer, ComK is a homogeneous tetramer, as demonstrated by SEC-MALS. (B-C) Overlay of  $^1\text{H}$ ,  $^{15}\text{N}$ -TROSY-HSQC spectra of 50  $\mu$ M  $^2\text{H}$ ,  $^{15}\text{N}$ -labeled ComK (teal) upon addition of equimolar quantities of either filMecA (olive, B) or MecA-CTD (magenta, C). Both samples were measured in the buffer given above. All spectra were recorded at 25  $^\circ\text{C}$  using a spectrometer operating at 600 MHz  $^1\text{H}$  frequency equipped with a cryogenic probehead.



**Figure S7. SEC-MALS profile of the MecA-ClpC complex in the presence of magnesium.**

## References

- Andersson, F.I., Blakytyn, R., Kirstein, J., Turgay, K., Bukau, B., Mogk, A., Clarke, A.K., 2006. Cyanobacterial ClpC/HSP100 protein displays intrinsic chaperone activity. *J Biol Chem* 281, 5468-5475.
- Battesti, A., Gottesman, S., 2013. Roles of adaptor proteins in regulation of bacterial proteolysis. *Curr Opin Microbiol* 16, 140-147.
- Bax, A., Ikura, M., 1991. An efficient 3D NMR technique for correlating the proton and  $^{15}\text{N}$  backbone amide resonances with the  $\alpha$ -carbon of the preceding residue in uniformly  $^{15}\text{N}/^{13}\text{C}$  enriched proteins. *J. Biomol. NMR* 1, 99-104.
- Bax, A., Clore, G.M., Gronenborn, A.M., 1990.  $^1\text{H}$   $^1\text{H}$  correlation via isotropic mixing of  $^{13}\text{C}$  magnetization, a new three-dimensional approach for assigning  $^1\text{H}$  and  $^{13}\text{C}$  spectra of  $^{13}\text{C}$ -enriched proteins. *J. Magn. Reson.* 88, 425-431.
- Berjanskii, M.V., Wishart, D.S., 2005. A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* 127, 14970-14971.
- Bodenhausen, G., Ruben, D.J., 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* 69, 185-189.
- Carroni, M., Franke, K.B., Maurer, M., Jager, J., Hantke, I., Gloge, F., Linder, D., Gremer, S., Turgay, K., Bukau, B., Mogk, A., 2017. Regulatory coiled-coil domains promote head-to-head assemblies of AAA+ chaperones essential for tunable activity control. *Elife* 6.
- Clubb, R.T., Thanabal, V., Wagner, G., 1992. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue  $^1\text{HN}$ ,  $^{15}\text{N}$ , and  $^{13}\text{C}$  chemical shifts in  $^{15}\text{N}$ --- $^{13}\text{C}$ -labelled proteins. *J. Magn. Reson.*
- Dougan, D.A., Mogk, A., Zeth, K., Turgay, K., Bukau, B., 2002. AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett* 529, 6-10.
- Dubnau, D., 1997. Binding and transport of transforming DNA by *Bacillus subtilis*: the role of type-IV pilin-like proteins—a review. *Gene* 192, 191-198.
- Elsholz, A.K.W., Birk, M.S., Charpentier, E., Turgay, K., 2017. Functional Diversity of AAA+ Protease Complexes in *Bacillus subtilis*. *Front Mol Biosci* 4, 44.
- Grzesiek, S., Bax, A., 1992. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* 114, 6291-6293.
- Horwich, A.L., Weber-Ban, E.U., Finley, D., 1999. Chaperone rings in protein folding and degradation. *Proc. Natl. Acad. Sci. U.S.A.* 96, 11033-11040.



- Ikura, M., Kay, L.E., Bax, A., 1990. A novel approach for sequential assignment of  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  spectra of larger proteins: Heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* 29, 4659-4667.
- Kay, L.E., Xu, G.Y., Yamazaki, T., 1994. Enhanced-sensitivity triple-resonance spectroscopy with minimal  $\text{H}_2\text{O}$  saturation. *J. magn. reson., Ser. A.* 109, 129-133.
- Kay, L.E., Ikura, M., Tschudin, R., Bax, A., 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* 89, 496-514.
- Kay, L.E., Xu, G.-Y., Singer, A.U., Muhandiram, D.R., FORMAN-KAY, J.D., 1993. A gradient-enhanced HCCH-TOCSY experiment for recording side-chain  $^1\text{H}$  and  $^{13}\text{C}$  correlations in  $\text{H}_2\text{O}$  samples for proteins. *J. magn. reson., Ser. B* 101, 333-337.
- Kirstein, J., Schlothauer, T., Dougan, D.A., Lilie, H., Tischendorf, G., Mogk, A., Bukau, B., Turgay, K., 2006. Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. *EMBO J* 25, 1481-1491.
- Lazazzera, B.A., Grossman, A.D., 1997. A regulatory switch involving a Clp ATPase. *Bioessays* 19, 455-458.
- Liu, J., Mei, Z., Li, N., Qi, Y., Xu, Y., Shi, Y., Wang, F., Lei, J., Gao, N., 2013. Structural Dynamics of the MecA-ClpC Complex: A TYPE II AAA+ PROTEIN UNFOLDING MACHINE\* $\diamond$ . *J. Biol. Chem.* 288, 17597-17608.
- Logan, T.M., Olejniczak, E.T., Xu, R.X., Fesik, S.W., 1992. Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS Lett* 314, 413-418.
- Magnuson, R., Solomon, J., Grossman, A.D., 1994. Biochemical and genetic characterization of a competence pheromone from *B. subtilis*. *Cell* 77, 207-216.
- Montelione, G.T., Lyons, B.A., Emerson, S.D., Tashiro, M., 1992. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J Am Chem Soc* 114, 10974-10975.
- Ogura, T., Wilkinson, A., 2001. AAA+ superfamily ATPases: common structure—diverse function. *Genes Cells* 6, 575-597.
- Ollerenshaw, J.E., Tugarinov, V., Kay, L.E., 2003. Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry* 41, 843-852.

- Persuh, M., Turgay, K., Mandic-Mulec, I., Dubnau, D., 1999. The N-and C-terminal domains of MecA recognize different partners in the competence molecular switch. *Mol. Microbiol.* 33, 886-894.
- Pervushin, K., Tan, E., Parthasarathy, K., Lin, X., Jiang, F.L., Yu, D., Vararattanavech, A., Soong, T.W., Liu, D.X., Torres, J., 2009. Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoS Pathog* 5, e1000511.
- Piotto, M., Saudek, V., Sklenář, V., 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2, 661-665.
- Santoro, J., King, G.C., 1992. A constant-time 2D overboderhausen experiment for inverse correlation of isotopically enriched species. *J. Magn. Reson.* 97, 202-207.
- Sauer, R.T., Baker, T.A., 2011. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* 80, 587-612.
- Schirmer, E.C., Glover, J.R., Singer, M.A., Lindquist, S., 1996. HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem. Sci.* 21, 289-296.
- Schlothauer, T., Mogk, A., Dougan, D.A., Bukau, B., Turgay, K., 2003. MecA, an adaptor protein necessary for ClpC chaperone activity. *Proc Natl Acad Sci U S A* 100, 2306-2311.
- Shen, Y., Bax, A., 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* 56, 227-241.
- Sklenar, V., Piotto, M., Leppik, R., Saudek, V., 1993. Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity. *J. magn. reson., Ser. A.* 102, 241-245.
- Solomon, J.M., Lazazzera, B.A., Grossman, A.D., 1996. Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes Dev.* 10, 2014-2024.
- Turgay, K., Hahn, J., Burghoorn, J., Dubnau, D., 1998. Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J* 17, 6730-6738.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D., 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Struct. Funct. Genet* 59, 687-696.

- Vuister, G.W., Bax, A., 1992. Resolution enhancement and spectral editing of uniformly  $^{13}\text{C}$ -enriched proteins by homonuclear broadband  $^{13}\text{C}$  decoupling. *J. Magn. Reson.* 98, 428-435.
- Wang, F., Mei, Z., Qi, Y., Yan, C., Hu, Q., Wang, J., Shi, Y., 2011. Structure and mechanism of the hexameric MecA-ClpC molecular machine. *Nature* 471, 331-335.
- Wittekind, M., Mueller, L., 1993. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-and beta-carbon resonances in proteins. *J. magn. reson., Ser. B.* 101, 201-205.
- Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D., Kay, L.E., 1994. A suite of triple resonance NMR experiments for the backbone assignment of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  labeled proteins with high sensitivity. *J. Am. Chem. Soc.* 116, 11655-11666.

## Chapter 3: Appendix

### Appendix 1: Cloning Primers

We acknowledge Prof. Kürşad Turgay (MPUSP Berlin) for the generous gift of the ClpC-DWB, MecA and ComK template plasmids. We also want to thank Dr Andrea Graziadei (TU Berlin) for the gift of Nsp1 expression plasmid, and Prof. Dr. Harald Schwalbe for the kind gift of the template plasmids of SARS-CoV-2 5'-UTRs.

Primers used to generate MecA-TEV construct:

Forward: 5'CCTCGAGCCCCTGAAAGTATAAATTTTCTGATGCAAAGTGTTTTTTATCGT

Reverse: 5'CGAATTCCTCGAGCCCCTGAAAGTATAAATTTTCTGATGCAAAGTGTTTTTTT  
ATCGT

Primers used to generate ClpC-DWB-TEV construct:

Forward: 5'GGAATTCCTCATGGGGTTTGGAAGATTTACAGAACGAGCT

Reverse: 5'GGAATTCCTCGAGGCCCTGAAAGTATAAATTTTCATTCGTTTTAGCAGT  
CGTTTTTAC

Primers used to generate ComK-SUMO construct:

Forward: 5'GGAATTCGGATCCATGAGCCAGAAAACCGATGC

Reverse: 5'CGAATTCCTCGAGTCATTAATAGCGTTCGCCAGTTCA

Primers used to generate mutants for MecA four truncated constructs (-TEV):

MecA NTD-For: 5'GAATTCCTCATGGAAATTGAAAGAATTAA

MecANTD-Rev: 5'GAATTCCTCGAGTTATTAATGATGATGATGATGATGGCCCTGAA  
AGTATAAATTTTCTTTTGTGACGATGATTTCTAA

MecA NTDL-For: 5'GAATTCCTCATGGAAATTGAAAGAATTAA

MecANTDL-Rev: 5'GAATTCCTCGAGTTATTAATGATGATGATGATGATGGCCCTGAAAG  
TATAAATTTTCTGATTGACGGCTTGCTCTT

MecA CTDL-For: 5'GAATTCCCATGGCCCAGCTTTCCAAAGACGG

MecACTDL-Rev: 5'GAATTCCTCGAGTTATTAATGATGATGATGATGATGGCCCTGAAAG  
TATAAATTTTCTGATGCAAAGTGTTTTTTTATC

MecA CTD-For: 5'GAATTCCCATGGAAGAGAAGGAGCAAAAAGCT

MecACTD-Rev: 5'GAATTCCTCGAGTTATTAATGATGATGATGATGATGGCCCTG  
AAAGTATAAATTTTCTGATGCAAAGTGTTTTTTTATC

Primers used to generate mutants for ClpC-DWB truncated constructs (-TEV):

ClpC-DWB- $\Delta$ D<sub>2</sub>-For: 5'GAATTCCCATGGGGTTTGGAAGATTTACA

ClpC-DWB- $\Delta$ D<sub>2</sub>-Rev: 5'GAATTCCTCGAGTTATTAATGATGATGATGATGATGGCCC  
TGAAAGTATAAATTTTCACCGGTCCAGCTGGATAC

ClpC-DWB- $\Delta$ D<sub>2</sub>- $\Delta$ 281-292- $\Delta$ 247-251-For: 5' ctcaaattctccgcggtattttgtcccatgtctaattgcat

ClpC-DWB- $\Delta$ D<sub>2</sub>- $\Delta$ 281-292- $\Delta$ 247-251-Rev: 5' atgacattagacatgggaacaaaataccgcggaatttgag

ClpC-N-For: 5'GAATTCCCATGGGGTTTGGAAGATTTACAGAA

ClpC-N-Rev:5'GAATTCCTCGAGTTATTAATGATGATGATGATGATGGCCCTGAAAGTATAA  
ATTTTCTGATGATCCTGTTTCATTACT

## Appendix 2: SAXS Data

Sample	MecA CTD–ClpC-DWB	ComK–MecA–ClpC-DWB
Source	<i>E. coli</i> BL21(DE3)	<i>E. coli</i> BL21(DE3)
Concentration (mg/ml)	5, 10, 17, 24	5, 10, 18, 25
Solvent	20 mM Tris-HCl, pH 7.5 100 mM NaCl, 1 mM TCEP, 1mM EDTA, 5mM ATP, 0.01% (v/v) NaN <sub>3</sub>	20 mM Tris-HCl, pH 7.5 100 mM NaCl, 1 mM TCEP, 1mM EDTA, 5mM ATP, 0.01% (v/v) NaN <sub>3</sub>
Source, beamline	P12 (Petra III U29 undulator)	P12 (Petra III U29 undulator)
Beam size at the detector	0.2 x 0.05 mm <sup>2</sup>	0.2 x 0.05 mm <sup>2</sup>
Wavelength	0.124 nm	0.124 nm
Monochromator	Double crystal Si (111)	Double crystal Si (111)
Detector model	Pilatus 6M	Pilatus 6M
Data analytical software	Primus UCSF Chimera DAMMIF	Primus UCSF Chimera DAMMIF
Absolute scaling method	Comparison with buffer scattering	Comparison with buffer scattering
Sample temperature	298K	298K
R <sub>g</sub> Guinier (nm)	8.47 ± 0.01 nm	9.82 ± 0.02 nm
sR <sub>g</sub>	0.50	0.47
Exp. I(0) (cm <sup>-1</sup> )	5.59 ± 0.0042	8.72 ± 0.0075
P(r) GNOM (nm)	31.44	39.84

### Appendix 3: Chemical-shift assignments

Table 3.1: Assignment of backbone resonances of MecA NTDL, residues 1-45

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	C <sub>α</sub> (ppm)	C <sub>β</sub> (ppm)
2	Glu	8, 6	123, 81	55, 59	32, 18
3	I1	8, 6	121, 68	59, 94	40, 54
4	Glu	9, 2	126, 33	55, 07	32, 65
5	Arg	9, 05	127, 76	54, 79	29, 46
6	I1	8, 09	129, 18	62, 76	37, 99
7	Asn	8, 19	112, 81	52, 68	38, 17
8	Glu	8, 81	121, 09	-	-
9	His	7, 97	113, 74	55, 53	32, 82
10	Thr	7, 4	116, 22	62, 6	71, 38
11	Val	9, 09	124, 91	59, 76	35, 81
12	Lys	8, 94	124, 12	54, 03	37, 32
13	Phe	9, 11	120, 51	55, 82	41, 76
14	Tyr	8, 71	121, 29	55, 74	41, 38
15	Met	7, 97	124, 97	54, 19	37, 32
16	Ser	9, 11	120, 72	57, 45	64, 81
17	Tyr	8, 47	120, 01	61, 8	36, 85
18	Gly	8, 59	107	47, 28	36, 71
19	Asp	7, 83	123, 01	57, 32	40, 58
20	I1	7, 81	119, 21	66, 1	38, 61
21	Glu	8, 16	120, 02	58, 98	29, 09
22	Asp	8, 74	122, 7	57, 14	40, 62
23	Arg	7, 44	116, 88	56, 49	30, 67
24	Gly	8, 06	106, 36	45, 35	30, 69
25	Phe	7, 6	118, 33	56, 74	41, 65
26	Asp	8, 94	121, 33	52, 91	43, 83
27	Arg	9, 08	124, 16	59, 65	29, 62
28	Glu	8, 42	118, 52	59, 08	29, 36
29	Glu	7, 86	118, 37	58, 74	-
30	I1	7, 41	115, 51	61, 85	35, 65
31	Trp	7, 46	116, 73	59, 38	-
32	Tyr	7, 62	112, 82	58, 59	40, 07
33	Asn	8, 42	120, 54	52, 13	39, 05
34	Arg	8, 88	126, 12	-	-
35	Glu	8, 43	119, 11	59, 99	29, 25
36	Arg	7, 74	119, 36	57, 12	29, 41
37	Ser	8, 34	114, 65	62, 22	-
38	Glu	7, 89	121, 14	60, 07	28, 96
39	Glu	7, 73	119, 12	60, 01	29, 85
40	Leu	7, 8	120, 24	57, 99	40, 27
41	Phe	7, 88	118, 19	61, 11	37, 86
42	Trp	8, 31	117, 49	59, 87	29, 23
43	Glu	7, 96	120, 46	59, 95	29, 41
44	Val	7, 94	119, 35	66, 12	31, 01
45	Met	8, 04	117, 77	56, 45	30, 89

Table 3.1: Assignment of backbone resonances of MecA NTDL, residues 46-90

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	C α (ppm)	C β (ppm)
46	Asp	7, 83	119, 01	56, 67	41, 24
47	Glu	7, 65	117, 93	57, 76	30, 27
48	Val	7, 55	115, 28	-	-
48	Val	7, 55	115, 32	62, 25	-
49	His	8, 14	121, 22	-	-
49	His	8, 16	121, 32	56, 91	29, 43
50	Glu	-	-	-	-
51	Glu	8, 18,	119, 67	56, 28	30, 54
52	Glu	8, 41	122, 48	55, 82	29, 63
52	Glu	8, 46	122, 41	-	-
53	Glu	8, 41	121, 41	-	-
54	Phe	8, 19	121, 13	-	-
55	Ala	8, 21	127, 59	52, 16	19, 24
56	Val	8, 18	120, 6	62, 14	33, 54
57	Glu	8, 46	124, 59	-	-
58	Gly	8, 07	110, 75	44, 65	-
59	Pro	-	-	-	-
60	Leu	8, 41	120, 61	-	-
60	Leu	8, 39	120, 48	53, 67	44, 95
61	Trp	9, 05	124, 05	56, 14	-
62	Il	7, 91	121, 31	-	-
63	Gln	8, 94	127, 34	54, 34	32, 75
64	Val	8, 38	123, 38	60, 89	34, 25
65	Gln	8, 98	125, 22	54, 34	31, 76
66	Ala	8, 87	127, 58	52, 59	19, 11
67	Leu	8, 59	124, 67	53, 54	43, 64
68	Asp	8, 58	120, 81	58, 04	40, 78
69	Lys	-	-	-	-
70	Gly	7, 03	110, 39	45, 11	-
71	Leu	8, 07	116, 74	52, 86	45, 57
72	Glu	8, 79	120, 52	54, 79	33, 26
73	Il	8, 8	127, 24	-	39, 01
74	Il	8, 75	126, 14	-	40, 15
75	Val	9, 25	129, 28	61, 3	33, 6
76	Thr	8, 91	121, 29	59, 88	-
77	Lys	8, 76	124, 3	55, 74	34, 08
78	Ala	8, 14	124, 87	52, 02	19, 7
79	Gln	8, 02	119, 35	55, 22	29, 74
80	Leu	8, 22	123, 72	-	-
81	Ser	8, 26	116, 58	-	-
82	Lys	8, 3	122, 97	-	-
83	Asp	8, 21	119, 41	54, 41	41, 01
84	Gly	8, 22	108, 84	45, 63	-
85	Gln	8, 1	119, 38	55, 74	29, 75
86	Lys	8, 37	122, 66	56, 22	33, 11
87	Leu	8, 32	124, 58	54, 81	42, 7
88	Glu	8, 46	122, 75	55, 83	30, 61
89	Leu	8, 25	125, 02	52, 79	41, 86
90	Pro	-	-	-	-



Table 3.1: Assignment of backbone resonances of MecA NTDL, residues 91-126

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	C α (ppm)	C β (ppm)
91	I1	8, 19	122, 84	58, 62	38, 7
93	Glu	8, 51	121, 49	56, 86	30, 51
94	Asp	8, 4	121, 35	54, 29	41, 02
95	Lys	8, 17	121, 88	56, 02	32, 92
96	Lys	8, 29	122, 98	56, 36	32, 91
97	Gln	8, 4	121, 93	55, 49	29, 64
98	Glu	8, 46	124, 28	54, 37	29, 73
100	Ala	8, 48	124, 74	52, 28	19, 33
101	Ser	8, 34	115, 69	58, 52	63, 85
102	Glu	8, 54	122, 61	56, 77	30, 38
103	Asp	8, 31	121, 32	54, 28	40, 94
104	Leu	8, 17	123, 03	55, 79	42, 14
105	Asp	8, 28	120, 49	54, 97	40, 94
106	Ala	7, 91	123, 49	52, 82	19, 14
107	Leu	8, 04	120, 28	55, 25	42, 03
108	Leu	8, 01	122, 28	55, 16	42, 37
109	Asp	8, 18	121, 06	54, 48	41, 42
110	Asp	8, 2	120, 35	54, 68	40, 99
111	Phe	8, 08	120, 24	58, 34	39, 22
112	Gln	8, 21	121, 46	56, 05	29, 22
113	Lys	8, 15	122, 29	56, 77	33, 1
114	Glu	8, 38	122, 02	56, 8	30, 24
115	Glu	8, 42	122, 25	56, 91	30, 32
116	Gln	8, 35	121, 15	55, 84	29, 47
117	Ala	8, 29	125, 55	52, 64	19, 16
118	Val	8, 1	119, 13	62, 39	32, 81
119	Asn	8, 46	122, 21	53, 31	38, 88
120	Gln	8, 42	121, 32	56, 21	29, 38
121	Glu	8, 44	121, 24	57, 09	30, 13
122	Asn	8, 34	118, 79	53, 26	38, 75
123	Leu	7, 97	121, 96	55, 33	42, 36
124	Tyr	7, 95	119, 61	57, 58	38, 78
125	Phe	7, 98	121, 5	57, 54	39, 63
126	Gln	7, 74	126, 19	57, 35	30, 62

Table 3.2 Assignments of all resonances of MecA NTDL, residues 1-71

Residue index	Amino acid	H <sup>1</sup> (ppm)	N (ppm)	NO (ppm)	NC (ppm)	Hβ (ppm)	Hγ (ppm)	Hδ (ppm)	Hε (ppm)	Ca (ppm)	Cβ (ppm)	Cγ (ppm)	Cδ (ppm)	Cε (ppm)	Cl (ppm)
1	Met														17.24
2	Glu	8.6	123.81			2.042.11	2.322.44		1.9	54.95	33.55	30.97			
3	Glu	9	126.9			2.111.99	0.850.85	0.641.33		56.99	32.18	36.1			
4	Glu	9.5	126.95			1.941.88	0.850.85	0.641.33		56.99	32.18	36.1	13.92		
5	Asp	9.05	127.76			1.971.71	1.541.57			54.79	29.46	35.56	43.32		
6	Ile	9.05	129.18			1.51	0.871.33	1.191.134		62.76	38.02	27.33	11.81		
7	Asn	8.19	112.81	114.94		3.163.31	1.972.04			52.68	38.17	35.62			
8	Glu	8.81	121.09			1.741.90	1.06			59.56	29.21	35.62			
9	His	7.97	113.74			3.343.12	0.690.69	0.710.88		62.5	32.82	21.2			
10	Thr	7.4	116.22			3.56	1.221.16			59.77	33.53	20.48	21.51	21.51	21.51
11	Val	9.09	124.91			1.73	1.521.48		2.762.68	54.03	37.32	24.53			41.76
12	Lys	8.94	124.13			1.661.52				55.92	41.36				
13	Ala	8.1	121.5			2.683.04				55.92	41.36				
14	Thr	8.11	121.59			1.661.52				55.92	41.36				
15	Met	7.97	124.97			1.661.81	2.412.56		1.82	54.19	37.32	32.87			17.3
16	Ser	9.11	120.72			4.074.35				57.45	64.81				
17	Tyr	8.47	120.01			3.142.81				61.8	36.85				
18	Gly	8.59	107							47.28	36.71				
19	Asp	7.83	123.01			2.712.88				57.32	40.58				
20	Ile	7.81	119.21			2.11	0.811.76	0.961.77		66.09	38.67	28.53	18.15		
21	Glu	8.16	120.03			2.282.13	2.592.54		0.7	59.96	29.09	36.84	14.18		
22	Asp	7.83	119.21			2.722.22				46.72	30.62				
23	Ala	7.44	116.58			2.461.91	1.771.89	2.843.12		55.46	30.62	29.09	43.41		
24	Gly	8.06	105.38			3.704.03				46.35	30.69				
25	Phe	7.6	118.33			4.63				56.74	41.65				
26	Asp	8.94	121.33			2.782.72				52.91	43.83				
27	Glu	9.08	124.16			4.9	1.621.35			59.65	29.62	25.35	43.31		
28	Glu	8.42	118.52			2.192.12	2.372.28	2.862.52		58.08	29.36	37.15			
29	Glu	7.86	118.37			2.372.09	2.432.51			58.74	30.32	37.17			
30	Ile	7.41	115.51			3.64	1.111.46	1.110.64	1.46	61.84	35.7	18.52	27.44		11.71
31	Tyr	8.1	119.21			2.683.04			0.43	59.56	40.07				
32	Tyr	7.92	112.82			2.953.16				59.56	40.07				
33	Asn	8.42	120.54	112.38		2.733.42				52.13	44.98	39.05			
34	Asp	8.88	126.12			5.06				59.25	26.86	26.86	42.89		
35	Glu	8.43	119.11			4.19	1.711.74			59.99	29.25	36.56			
36	Asp	7.74	119.36			2.132.24	2.382.27	3.203.31		57.12	29.41	26.47	41.78		
37	Ser	8.34	114.65			3.91	1.811.77	3.363.42		62.22	29.41	26.47			
38	Glu	7.89	121.14			3.77	2.532.43			60.07	29.86	37.53			
39	Glu	7.73	119.12			4.24	2.722.43			59.01	29.86	36.7			
40	Glu	7.89	119.36			3.81	1.24			59.01	29.86	36.7			
41	Phe	7.87	118.19			3.81	0.390.55	0.39		61.11	37.86	20.93			22.71
42	Tyr	8.31	117.49			3.383.17			10.85	59.87	29.22				22.71
43	Glu	7.96	120.46			4.17	2.522.24			59.95	29.39	36.61			
44	Val	7.94	119.35			3.46	0.250.47	0.47	0.24	66.12	31.05	20.82	20.82	21.91	21.96
45	Met	8.04	117.77			4.13	1.42		1.69	56.45	30.89	32.76			17.22
46	Asp	7.83	119.01			4.55	2.442.22			57.75	30.28	36.24			
47	Glu	7.65	117.93			4.18	0.990.60	0.60	0.83	57.24	30.28	36.24			
48	Val	7.55	115.32			4.16	2.222.43			57.24	30.28	36.24			
49	Met	8.16	121.32			4.23	2.26			57.39	30.18	36.23			
50	Glu	8.18	119.59	119.67		4.24	2.202.28			56.28	30.54	36.23			
51	Glu	8.41	122.48			1.861.99				55.82	29.63	36.23			
52	Glu	8.41	122.48			1.861.99				55.82	29.63	36.23			
53	Glu	8.41	121.41			4.3	2.202.28			56.27	30.33	36.23			
54	Phe	8.19	121.13			4.6	2.202.28			56.37	30.2	36.24			
55	Ala	8.21	127.59			4.43	0.480.68	1.33		57.31	39.96	36.24			
56	Ala	8.16	120.6			4.2	2.041.97			52.16	19.24	35.97			
57	Glu	8.07	119.59			4.23	0.820.96	1.00	0.95	55.82	31.06	36.17			
58	Gly	8.07	110.75			4.23	2.191.94			55.82	31.06	36.17			
59	Phe					4.44	1.672.21		3.58	62.85	32.23	27.25	49.67		
60	Leu	8.41	120.61			4.32	1.121.55			53.67	45.04	26.48			
61	Leu	8.39	124.15	124.05		4.83	1.83		10.1	56.14	55.89	26.48	26.76	23.75	
62	Ile	8.39	123.26			4.97	3.143.03	0.800.76		60.42	40.45	30.63	13.05		
63	Glu	8.94	127.34			4.54	0.480.68	1.33	7.27	54.34	32.75	27.97	16.95		
64	Ala	8.98	125.6			4.77	0.650.65	0.55		54.34	32.75	27.97	16.95		
65	Glu	8.98	125.22			4.76	2.041.98		6.71	59.59	19.11	33.75			
66	Ala	8.97	127.59			4.65	1.24			59.59	19.11	33.75			
67	Leu	8.59	124.67			4.8	1.881.83	1.83	1.69	59.59	19.11	33.75			
68	Asp	8.58	120.81			4.44	1.672.21			59.59	19.11	33.75			
69	Gly	8.03	110.39			4.61	1.451.41		3.02	58.04	40.78	27.66	25.73	23.39	23.27
70	Gly	8.03	116.74			5.03	1.161.31	1.181.31		56.58	33.32	25.14	28.18		42.19
71	Leu	8.03	116.74			5.03	0.710.83	0.710.88		52.85	45.62	26.29	25.87	25.82	24.97

Table 3.2 Assignments of all resonances of Meca NTDL, residues 72-126

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	NE (ppm)	NE (ppm)	HD (ppm)	Hβ (ppm)	Hγ (ppm)	Hδ (ppm)	He (ppm)	Cα (ppm)	Cβ (ppm)	Cγ (ppm)	Cδ (ppm)	Cε (ppm)
72	Glu	8.79	120.52	-	-	4.98	1.91,1.88	2.06,1.90	-	-	54.79	33.26	37.92	-	-
73	Ile	8.8	127.24	-	-	4.56	1.55	1.07,0.39,0.62,1.08,0.61	0.43	-	60.24	38.96	28.19,18.09	13.58	-
74	Ile	8.75	126.15	-	-	4.67	1.55	0.70,1.44,1.44,0.74,0.76	0.71	-	60.53	40.13	28.32,18.20	13.86	-
75	Val	9.25	129.28	-	-	4.63	1.5	0.14,0.36,0.13,0.34	-	-	61.29	33.65	21.68,21.72,20.84,20.88	-	-
76	Thr	8.91	121.29	-	-	5.75	4.07	1.32	-	-	59.88	71.72	20.68	-	-
77	Lys	8.76	124.3	-	-	4.6	1.82,1.87	1.52,1.45	1.52	2.80,2.74	55.74	34.08	24.69	29.71	42
78	Ala	8.14	124.87	-	-	3.99	1.04	-	-	-	52.02	19.7	-	-	-
79	Gln	8.02	119.35	-	112.18	4.09	1.9	2.09	-	6.75,7.36	55.22	29.74	33.58	-	-
80	Leu	8.22	123.72	-	-	4.42	1.50,1.56,1.54	1.54	0.63,0.77	-	54.79	42.83	27.03	23.51,24.99	-
81	Ser	8.26	116.56	-	-	4.3	3.75,3.78	-	-	-	58.43	63.69	-	-	-
82	Lys	8.3	122.97	-	-	4.29	1.75,1.82	1.36	1.63	2.96	56.3	32.7	24.63	29.07	42.21
83	Asp	8.21	119.41	-	-	4.56	2.71,2.53	-	-	-	54.41	41.01	-	-	-
84	Gly	8.22	108.84	-	-	-	-	-	-	-	45.63	-	-	-	-
85	Gln	8.1	119.38	-	-	4.51	2.64,2.67	-	-	-	55.74	29.75	-	-	-
86	Lys	8.37	122.66	-	-	4.26	1.72,1.82	1.36,1.41	1.63	2.96	56.22	33.11	24.67	29.04	42.11
87	Leu	8.32	124.58	-	-	4.41	1.55,1.65,1.50,1.55	1.54	0.76,0.88,0.81,0.83	-	54.8	42.72	27.05	23.72,23.34,24.97,24.93	-
88	Glu	8.46	122.75	-	-	4.38	4.38	-	-	-	55.63	30.61	-	-	-
89	Leu	8.25	125.02	-	-	4.54	2.25,1.48,1.51,1.92	1.55	0.85,0.82	-	52.78	41.87	27.07	23.45,25.17	-
90	Pro	-	-	-	-	4.15	1.82,1.68	1.6	3.15	-	57.25	31.63	27.19	43.44	-
91	Ile	8.19	122.84	-	-	4.38	1.81	0.93,1.16,1.51	0.65	-	58.62	36.7	17.16,27.06	12.81	-
93	Glu	8.51	121.49	-	-	-	-	-	-	-	58.86	30.51	-	-	-
94	Asp	8.4	121.35	-	-	-	-	-	-	-	54.29	41.02	-	-	-
95	Lys	8.17	121.88	-	-	-	-	-	-	-	56.02	32.92	-	-	-
96	Lys	8.29	122.98	-	-	-	-	-	-	-	56.36	32.91	-	-	-
97	Gln	8.4	121.93	-	-	-	-	-	-	-	55.49	29.64	-	-	-
98	Glu	8.46	124.28	-	-	-	-	-	-	-	54.37	29.73	-	-	-
100	Ala	8.48	124.74	-	-	-	-	-	-	-	52.28	19.33	-	-	-
101	Ser	8.34	115.69	-	-	-	-	-	-	-	58.52	63.85	-	-	-
102	Glu	8.54	122.61	-	-	-	-	-	-	-	56.77	30.38	-	-	-
103	Asp	8.31	121.32	-	-	-	-	-	-	-	54.28	40.94	-	-	-
104	Leu	8.17	123.03	-	-	4.25	1.58,1.67	1.64	0.92,0.85	-	55.8	42.17	26.99	23.33,25.03	-
105	Asp	8.28	120.5	-	-	-	-	-	-	-	54.97	40.94	-	-	-
106	Ala	7.91	123.49	-	-	-	-	-	-	-	52.82	19.14	-	-	-
107	Leu	8.04	120.28	-	-	4.3	1.58,1.68	1.64	0.66,0.92	-	55.26	42.06	26.99	25.05,23.46	-
108	Leu	8.01	122.28	-	-	4.34	1.57,1.67	1.63	0.92,0.85	-	55.18	42.41	26.97	25.06,23.37	-
109	Asp	8.18	121.05	-	-	-	-	-	-	-	54.48	41.42	-	-	-
110	Asp	8.2	120.35	-	-	-	-	-	-	-	54.68	40.99	-	-	-
111	Phe	8.08	120.24	-	-	-	-	-	-	-	58.34	39.22	-	-	-
112	Gln	8.21	121.46	-	-	-	-	-	-	-	56.05	29.22	-	-	-
113	Lys	8.15	122.29	-	-	-	-	-	-	-	56.77	33.1	-	-	-
114	Glu	8.38	122.02	-	-	-	-	-	-	-	56.8	30.24	-	-	-
115	Glu	8.42	122.25	-	-	-	-	-	-	-	56.91	30.32	-	-	-
116	Gln	8.35	121.15	-	-	-	-	-	-	-	55.84	29.47	-	-	-
117	Ala	8.29	125.55	-	-	-	-	-	-	-	52.64	19.16	-	-	-
118	Val	8.1	119.13	-	-	4.08	2.06	0.92,0.91,0.93	-	-	62.4	32.82	21.14,20.57	-	-
119	Asn	8.46	122.2	-	-	-	-	-	-	-	53.31	38.88	-	-	-
120	Gln	8.42	121.32	-	-	-	-	-	-	-	56.21	29.38	-	-	-
121	Glu	8.44	121.24	-	-	-	-	-	-	-	57.09	30.13	-	-	-
122	Asn	8.34	118.79	-	-	-	-	-	-	-	53.26	38.75	-	-	-
123	Leu	7.97	121.96	-	-	4.2	1.29,1.43	1.43	0.77,0.84	-	55.33	42.58	26.82	24.92,23.37	-
124	Tyr	7.95	119.61	-	-	-	-	-	-	-	57.98	38.78	-	-	-
125	Phe	7.98	121.5	-	-	-	-	-	-	-	57.54	39.63	-	-	-
126	Gln	7.74	126.19	-	-	-	-	-	-	-	57.35	30.62	-	-	-

Table 3.3: Assignment of all ILV-methyl groups of MecA NTDL

<sup>1</sup> H (ppm)	<sup>13</sup> C (ppm)	Assign F1
0, 70331	21, 30396	(A/B) 11Val (Hgb*/Hga*)
0, 66478	20, 61949	(A/B) 64ValHgb*
0, 69247	11, 54177	(B/A) 61IleHd1:
0, 86775	21, 19314	(B/A) 48ValHga*
0, 70054	13, 57271	(B/A) 74IleHd1:
0, 60544	13, 6206	A3IleHd1*
0, 59789	20, 21015	A11ValHga*
0, 69374	13, 92667	A20IleHd1*
0, 42358	11, 36796	A30IleHd1*
0, 39249	22, 47062	A40LeuHda*
0, 55934	25, 95337	A40LeuHdb*
0, 23874	20, 55808	A44ValHga*
0, 47149	21, 78686	A44ValHgb*
0, 89671	20, 11313	A48ValHgb*
0, 95641	21, 16299	A56ValHga*
1, 00229	21, 72847	A56ValHgb*
0, 76146	25, 50818	A60LeuHda*
0, 79963	23, 48068	A60LeuHdb*
0, 53044	12, 79069	A62IleHd1*
0, 557	22, 39305	A64ValHga*
0, 80984	23, 15121	A67LeuHda*
0, 84261	25, 42871	A67LeuHdb*
0, 71517	25, 63915	A71LeuHda*
0, 88156	24, 68239	A71LeuHdb*
0, 42557	13, 32123	A73IleHd1*
0, 34252	20, 62798	A75ValHgb*
0, 77611	23, 15418	A80LeuHda*
0, 8361	24, 72587	A80LeuHdb*
0, 78854	23, 4442	A87LeuHda*
0, 83498	24, 64941	A87LeuHdb*
0, 82655	23, 1399	A89LeuHda*
0, 86091	24, 78547	A89LeuHdb*
0, 842	12, 49169	A91IleHd1*
0, 85076	23, 05906	A104LeuHda*
0, 91844	24, 74171	A104LeuHdb*
0, 85267	23, 12521	A107LeuHda*
0, 92211	24, 78357	A107LeuHdb*
0, 84458	23, 06935	A108LeuHda*
0, 91497	24, 77846	A108LeuHdb*
0, 91804	20, 87224	A118ValHga*
0, 93608	20, 24985	A118ValHgb*
0, 5653	13, 52051	B3IleHd1*
0, 70759	20, 0837	B11ValHgb*
0, 27565	20, 62011	B44ValHga*
0, 5212	21, 80151	B44ValHgb*
0, 90254	20, 2763	B48ValHgb*
0, 84411	20, 38368	B56ValHga*
0, 85737	21, 24063	B56ValHgb*
0, 50074	12, 78398	B62IleHd1*
0, 60139	22, 38936	B64ValHga*
0, 39614	20, 65508	B75ValHgb*

Table 3.4: Assignment of backbone resonances of Nsp1, residues 0-45

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	C α (ppm)	C β (ppm)
0	Ala	-	-	52, 85	19, 38
1	Met	8, 53	119, 39	55, 81	32, 65
2	Glu	8, 43	121, 93	57	30, 25
3	Ser	8, 25	116, 3	58, 3	63, 79
4	Leu	8, 21	124, 2	55, 24	42, 41
5	Val	8, 12	122, 28	59, 82	32, 59
6	Pro	-	-	63, 54	32, 14
7	Gly	8, 49	109, 15	45, 24	-
8	Phe	7, 82	118, 6	57, 01	39, 94
9	Asn	8, 63	121, 39	52, 79	38, 9
10	Glu	8, 61	123, 47	57, 6	29, 55
11	Lys	8, 43	118	57, 64	32, 63
12	Thr	7, 7	107, 47	62, 35	70
13	His	8, 25	122, 24	56, 5	34, 9
14	Val	8, 96	117, 15	59, 85	35, 17
15	Gln	8, 61	125, 95	55, 21	30, 11
16	Leu	8, 95	125, 58	53, 83	45, 85
17	Ser	8, 63	117, 04	56, 62	63, 37
18	Leu	9, 18	126, 86	51, 47	43, 75
19	Pro	-	-	62, 16	32, 03
20	Val	8, 43	124, 89	61, 07	31, 62
21	Leu	8, 87	126, 08	51, 83	43, 02
22	Gln	9, 12	121, 07	55, 37	29, 08
23	Val	9, 08	125, 79	65, 77	31, 05
24	Arg	8, 08	117, 76	57, 98	29, 69
25	Asp	7, 68	117, 08	54, 77	42, 08
26	Val	6, 88	118, 28	64, 04	31, 86
27	Leu	9, 37	131, 8	57, 29	42, 9
28	Val	10, 02	126, 68	62, 81	31, 6
29	Arg	8, 04	119, 78	57, 13	29, 96
30	Gly	7, 23	106, 01	45, 31	-
31	Phe	9, 69	115, 62	59, 49	41, 04
32	Gly	7, 49	106, 84	45, 66	-
33	Asp	8, 44	115, 6	55, 11	43, 02
34	Ser	7, 51	112, 16	56, 25	66, 15
35	Val	9, 13	123, 99	67, 35	31, 5
36	Glu	8, 71	117	59, 94	29, 06
37	Glu	7, 54	119, 1	59, 05	30, 95
38	Val	8, 49	112, 31	65, 69	30, 15
39	Leu	7, 58	120, 78	58, 17	41, 86
40	Ser	7, 47	113, 75	61, 56	62, 72
41	Glu	8, 69	121, 26	58, 63	30, 71
42	Ala	8, 44	124, 1	55, 7	18, 4
43	Arg	7, 72	114, 94	60, 47	29, 94
44	Gln	7, 69	119, 59	58, 77	28, 37
45	His	8, 35	116, 69	59, 4	28, 85

Table 3.4: Assignment of backbone resonances of Nsp1, residues 45-90

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	C α (ppm)	C β (ppm)
45	His	8,35	116,69	59,4	28,85
46	Leu	8,74	121,86	58,28	41,77
47	Lys	7,63	120,16	59,34	32,15
48	Asp	7,96	116,78	54,82	41,53
49	Gly	7,85	108,78	46,9	-
50	Thr	8,9	110,63	60,17	67,47
51	Cys	8,48	121,8	57,68	30,82
52	Gly	9,79	113,41	43,69	-
53	Leu	8,78	120,31	53,41	47,05
54	Val	9,45	125,31	60,24	33,59
55	Glu	8,97	126,69	57,71	29,69
56	Val	8,64	122,04	64,37	-
57	Glu	6,89	124,23	54,18	-
58	Lys	8,96	123,73	59,12	31,92
59	Gly	-	-	45,46	-
60	Val	7,95	122,12	66,49	32,77
61	Leu	8,76	119,34	-	-
63	Gln	-	-	56,13	30,14
64	Leu	7,97	121,67	53,48	43,67
65	Glu	7,3	118,09	55,54	31,14
66	Gln	8,24	121,59	53,62	28,56
67	Pro	-	-	61,49	34,8
68	Tyr	8,05	119,72	55,92	41,42
69	Val	9,04	120,49	61,95	32,49
70	Phe	9,29	123,32	56,3	41,67
71	I1	8,04	112,9	58,72	39,73
72	Lys	8,96	125,42	53,81	36,19
73	Arg	8,97	123,08	57,43	31,21
74	Ser	8,74	120,88	58,52	64,26
75	Asp	8,31	122,31	54,3	41,99
76	Ala	8,36	124,79	52,85	19,55
77	Arg	8,43	119,01	56,73	30,77
78	Thr	8,04	114,11	61,82	69,79
79	Ala	8,28	127,78	50,42	18,58
80	Pro	-	-	63,68	31,74
81	His	8,36	117,85	56,58	29,59
82	Gly	8,28	108,82	45,46	-
83	His	8,08	119,39	55,98	31,59
84	Val	8,08	121,66	61,77	33,15
85	Met	8,77	127,89	55,83	33,51
86	Val	8,81	117,32	58,62	36,78
87	Glu	8,94	116,87	54,05	33,91
88	Leu	8,54	121,69	54,66	43,78
89	Val	8,73	120,01	60,71	36,32
90	Ala	8,69	126,04	50,46	22,8

Table 3.4: Assignment of backbone resonances of Nsp1, residues 91-135

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	C α (ppm)	C β (ppm)
91	Glu	8, 86	118, 47	55, 44	34, 72
92	Leu	8, 84	123, 87	57, 08	43, 28
93	Glu	-	-	57, 09	30, 67
94	Gly	8, 22	106, 37	46, 21	-
95	Il	7, 57	119, 6	62	38, 47
96	Gln	-	-	56, 74	29, 15
97	Tyr	8, 88	120, 44	58, 51	-
98	Gly	7, 94	108, 99	46, 09	-
99	Arg	7, 83	118, 65	55, 63	30, 76
100	Ser	7, 87	115, 69	58, 35	63, 73
101	Gly	8, 69	112, 79	-	-
103	Thr	-	-	63, 2	69, 25
104	Leu	9, 03	126, 76	56, 24	42, 9
105	Gly	7, 44	103, 93	45, 47	-
106	Val	9, 16	121, 03	59, 93	34, 71
107	Leu	9	126, 54	53, 53	43, 57
108	Val	8, 34	113, 37	58, 09	33, 39
109	Pro	-	-	63, 56	31, 52
110	His	8, 03	120, 57	53, 04	32, 88
111	Val	10, 39	130, 71	60, 6	31, 94
112	Gly	6, 74	107, 44	45, 15	-
113	Glu	10, 56	124, 33	58, 92	29, 5
114	Il	9, 06	125, 99	58, 63	37, 27
115	Pro	-	-	62, 3	33, 08
116	Val	8, 84	114, 3	61, 38	32, 49
117	Ala	7, 72	120, 66	51, 18	20, 88
118	Tyr	8, 46	116, 05	57, 78	42, 05
119	Arg	9, 36	122, 27	53, 23	32, 04
120	Lys	8, 61	126, 69	56, 2	32, 49
121	Val	8, 96	116, 94	58, 77	36, 19
122	Leu	8, 74	123, 21	53, 2	44, 88
123	Leu	8, 89	125, 07	53, 66	45, 56
124	Arg	8, 79	126, 21	56, 92	30, 68
125	Lys	8, 44	125, 92	57, 64	33, 53
126	Asn	8, 4	118, 85	-	39, 48
127	Gly	-	-	45, 39	-
128	Asn	8, 41	118, 81	53, 38	39, 11
129	Lys	8, 45	121, 75	56, 69	32, 85
130	Gly	8, 39	109, 81	45, 25	-
131	Ala	8, 24	123, 82	52, 68	19, 38
132	Gly	8, 44	108, 1	45, 42	-
133	Gly	8, 23	108, 43	45, 28	-
134	His	8, 2	118, 96	55, 97	30, 66
135	Ser	8, 3	117, 04	58, 19	63, 82

Table 3.4: Assignment of backbone resonances of Nsp1, residues 136-180

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	C α (ppm)	C β (ppm)
136	Tyr	8, 33	122, 26	58, 23	38, 79
137	Gly	8, 36	110, 53	45, 41	-
138	Ala	8, 11	123, 66	52, 77	19, 39
139	Asp	8, 35	118, 66	54, 24	40, 97
140	Leu	8, 05	122, 22	55, 44	42, 13
141	Lys	8, 25	121, 31	56, 39	32, 86
142	Ser	8, 11	116, 1	58, 37	63, 81
143	Phe	8, 14	121, 65	57, 73	39, 71
144	Asp	8, 28	121, 83	54, 2	41, 5
145	Leu	8, 22	122, 82	55, 52	42, 31
146	Gly	8, 43	109, 12	45, 55	-
147	Asp	8, 2	120, 45	54, 46	41, 4
148	Glu	8, 46	120, 85	56, 86	30, 04
149	Leu	8, 27	122, 43	55, 42	42, 39
150	Gly	8, 37	109, 32	45, 42	-
151	Thr	7, 98	113, 15	61, 59	70, 08
152	Asp	8, 48	124, 47	52, 42	41, 39
153	Pro	-	-	63, 55	31, 98
154	Tyr	8, 23	119, 35	58, 05	38, 35
155	Glu	7, 94	121, 96	56, 81	30, 59
156	Asp	8, 23	121, 07	54, 36	41, 23
157	Phe	8, 11	120, 77	58, 29	39, 33
158	Gln	8, 11	121, 3	56, 1	29, 41
159	Glu	8, 29	121, 16	57, 08	29, 97
160	Asn	8, 33	118, 77	53, 62	38, 79
161	Trp	8, 12	121, 32	58, 17	29, 31
162	Asn	8, 24	119, 05	53, 8	38, 54
163	Thr	7, 95	113, 73	62, 57	69, 56
164	Lys	8, 11	122, 57	56, 75	32, 8
165	His	8, 2	119, 57	56, 12	30, 31
166	Ser	8, 24	116, 99	58, 48	63, 84
167	Ser	8, 47	117, 9	58, 77	63, 88
168	Gly	8, 42	110, 73	45, 5	-
169	Val	7, 96	119, 43	62, 61	32, 75
170	Thr	8, 11	117, 9	62, 22	69, 82
171	Arg	8, 4	123, 3	56, 98	30, 66
172	Glu	8, 44	121, 32	57, 35	29, 95
173	Leu	8, 14	122, 58	55, 78	42, 24
174	Met	8, 25	120, 25	56, 05	32, 48
175	Arg	8, 15	121, 57	56, 87	30, 77
176	Glu	8, 36	121, 27	57, 04	30, 06
177	Leu	8, 25	122, 49	55, 57	42, 27
178	Asn	8, 38	118, 68	53, 47	39, 11
179	Gly	8, 3	109, 57	45, 57	-
180	Gly	7, 97	115, 14	46, 07	-



Table 3.5 Assignments of all resonances of Nsp1, residues 1-60

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	H $\alpha$ (ppm)	H $\beta$ (ppm)	H $\gamma$ (ppm)	H $\delta$ (ppm)	He (ppm)	Ca (ppm)	C $\beta$ (ppm)	C $\gamma$ (ppm)	C $\delta$ (ppm)	C $\epsilon$ (ppm)
0													
1	Ala	8.53	119.39	4.34	1.4	2.54,2.61	-	-	52.86	19.39	32.1	-	-
2	Met	8.43	121.93	4.44	2.02,2.08	2.23,2.26	-	-	55.81	32.65	36.37	-	-
3	Glu	8.25	116.3	4.28	2.04,1.96	3.83,3.83	-	-	56.99	30.21	-	-	-
4	Ser	8.21	124.2	4.37	1.56,1.64	1.6	0.90,0.83	-	58.3	63.8	-	-	-
5	Leu	8.12	122.28	4.37	2.05	0.89,0.91	-	-	59.82	42.41	27.03	24.98,23.49	-
6	Val	-	-	4.36	1.84,2.23	1.94,2.01	3.84,3.66	-	59.82	32.54	21.16,20.46	-	-
7	Pro	-	-	3.92,3.92	-	-	-	-	45.24	32.13	27.47	50.98	-
8	Gly	7.82	109.15	4.5	2.87,2.81	-	-	-	57.02	39.95	-	-	-
9	Phe	8.63	121.39	4.59	2.66,3.17	-	-	-	52.8	38.91	-	-	-
10	Asn	8.61	123.48	4.4	2.23,2.23	2.06,2.08	-	-	57.61	29.51	36.34	-	-
11	Glu	8.43	118.01	4.3	1.89,1.89	1.54,1.43	1.70,1.70	3.00,3.00	57.61	32.65	25.44	29.11	42.18
12	Lys	8.25	107.46	4.31	4.25	1.22	-	-	62.34	69.99	21.72	-	-
13	Thr	8.25	122.24	5.37	2.73,2.83	-	-	-	56.49	34.89	-	-	-
14	His	8.96	117.15	4.54	1.9	0.89,0.89	-	-	59.85	35.18	20.45,20.76	-	-
15	Val	8.61	125.95	4.79	1.77,1.61	1.78,1.90	-	-	53.83	30.13	34.33	-	-
16	Gln	8.95	125.58	4.85	1.54,1.49	1.54	0.84,0.94	-	53.83	45.83	26.72	27.36,23.94	-
17	Leu	8.63	117.04	5.04	3.65,3.70	1.61	-	-	56.81	63.38	-	-	-
18	Ser	8.18	126.68	5.02	1.16,1.73	1.61	0.71,0.82	-	51.47	43.75	26.86	24.36,26.66	-
19	Leu	-	-	4.03	1.19,1.76	1.96,2.01	4.08,4.33	-	62.17	31.99	26.83	-	-
20	Pro	8.43	124.89	4.68	1.52	0.77,0.67	-	-	61.07	31.61	21.53,21.21,21.21	-	-
21	Val	8.87	126.08	4.66	1.20,1.20	1.4	0.29,0.61	-	51.84	43.02	27.33	25.47,21.96	-
22	Leu	9.12	121.07	4.43	1.96,2.13	2.22,2.49	-	-	55.36	29.08	35.18	-	-
23	Gln	9.08	125.8	3.72	2.11	1.00,0.94	-	-	65.76	31.07	23.06,19.57	-	-
24	Val	8.08	117.76	4.17	1.78,1.84	1.59,1.65	3.18,3.18	-	57.97	29.69	27.1	43.46	-
25	Arg	7.68	117.07	4.53	2.76,2.86	-	-	-	54.76	42.09	-	-	-
26	Asp	6.87	118.28	3.9	1.85	0.71,0.85	-	-	64.04	31.88	23.39,23.61	-	-
27	Val	9.37	131.8	4.23	1.73,1.46	1.76	0.87,0.91	-	57.29	42.89	27.09	23.66,24.66	-
28	Leu	10.02	126.68	4.04	1.96	0.96,0.83	-	-	62.81	31.63	20.55,22.62	-	-
29	Val	8.04	119.77	3.92	1.59,2.19	1.57,1.29	3.28,3.39	-	57.12	29.86	29.53	43.13	-
30	Arg	7.23	106.01	3.80,4.36	-	-	-	-	45.3	-	-	-	-
31	Gly	9.69	115.62	4.56	2.80,3.40	-	-	-	59.49	41.04	-	-	-
32	Phe	7.49	106.84	3.81,4.18	-	-	-	-	55.11	43.02	-	-	-
33	Gly	8.44	115.6	4.91	2.75,2.66	-	-	-	56.24	66.14	-	-	-
34	Asp	7.51	112.16	4.83	4.17,3.92	-	-	-	67.34	31.52	22.70,21.16	-	-
35	Ser	9.13	123.99	3.45	1.97	0.91,0.82	-	-	59.94	29.11	36.74	-	-
36	Val	8.71	116.99	3.96	2.01,1.96	2.35,2.29	-	-	59.03	30.97	36.7	-	-
37	Glu	7.54	119.11	4.18	2.04,2.11	2.24,2.36	-	-	65.69	30.14	20.20,22.11	-	-
38	Leu	8.49	112.31	3.82	2.32	0.95,0.85	-	-	58.16	41.87	27.07	23.26,25.55	-
39	Val	7.58	120.78	3.9	1.79,1.37	1.66	0.85,0.75	-	61.56	62.76	-	-	-
40	Ser	7.47	113.75	4.11	3.99,4.01	-	-	-	55.68	30.77	36.41	-	-
41	Glu	8.68	121.26	4.16	2.11,2.05	2.55,2.61	-	-	58.63	18.4	-	-	-
42	Leu	8.44	124.1	2.91	0.89	-	-	-	60.47	29.93	29.38	43.36	-
43	Ala	7.72	114.94	3.77	1.78,1.89	1.46,1.83	3.16,3.12	-	58.77	28.36	33.24	-	-
44	Arg	7.69	119.58	3.9	2.16,2.16	2.24,2.36	-	-	59.4	28.86	-	-	-
45	Gln	8.35	116.68	4.41	2.98,3.06	-	-	-	58.29	29.11	-	-	-
46	His	8.74	121.86	4.04	1.25,2.00	1.78	0.88,0.71	-	59.34	32.14	26.85	26.55,24.84	-
47	Leu	7.63	120.16	3.96	1.83,1.92	1.35,1.46	1.62,1.62	2.91,2.91	58.29	32.14	24.74	29.56	42.08
48	Lys	7.96	116.77	4.65	2.74,2.74	-	-	-	54.82	41.52	-	-	-
49	Asp	7.85	108.78	4.25,4.25	-	-	-	-	60.16	67.46	21.42	-	-
50	Gly	8.9	110.63	4.89	5.12	1.16	-	-	57.67	30.82	-	-	-
51	Thr	8.48	121.8	5.36	2.98,2.48	-	-	-	43.69	47.05	-	-	-
52	Cys	9.79	113.41	4.72,3.17	-	-	-	-	53.41	47.05	27.33	25.56,25.97	-
53	Gly	8.78	120.31	5.12	1.08,1.52	1.3	0.87,0.63	-	60.23	33.58	21.92,22.17	-	-
54	Val	9.45	125.31	4.51	1.83	0.73,0.90	-	-	57.7	29.61	37.89	-	-
55	Glu	8.97	126.7	4.26	1.89,1.97	2.10,2.31	-	-	64.37	31.4	22.10,22.25	-	-
56	Val	8.64	122.04	3.66	2.01	0.84,0.90	-	-	54.2	32.97	36.12	-	-
57	Glu	6.89	124.22	4.57	1.86,2.18	2.13,2.19	-	-	59.09	31.94	24.64	29.36	42.08
58	Glu	8.96	123.73	4.06	1.79,1.83	1.39,1.50	1.70,1.70	3.00,3.00	59.09	31.94	24.64	29.36	42.08
59	Gly	-	-	4.27,3.70	-	-	-	-	45.43	-	-	-	-
60	Val	7.95	122.12	3.82	2.12	1.08,0.90	-	-	66.49	32.82	23.61,20.93	-	-

Table 3.5 Assignments of all resonances of Nsp1, residues 61-120

Residue index	Amino acid	H <sup>1</sup> (ppm)	N (ppm)	Ha (ppm)	Hb (ppm)	Hv (ppm)	Hd (ppm)	Ca (ppm)	Cb (ppm)	Cv (ppm)	Co (ppm)	Ce (ppm)
61	Leu	8.76	119.33	3.98	1.32,1.84	1.64	0.29,0.75	59.76	37.85	26.64	22.43,25.52	-
62	Pro	4.59	-	4.59	1.74,2.45	1.96,2.02	3.73,3.30	65.08	31.56	28.11	50.67	-
63	Gln	-	-	4.38	2.14,2.23	2.27,2.42	-	56.1	30.12	34.97	-	-
64	Leu	7.96	121.67	4.77	1.86,2.08	1.65	0.71,0.83	53.49	43.68	26.8	26.80,22.78	-
65	Glu	7.3	118.09	4.4	1.66,1.88	2.17,2.11	-	55.55	30.85	36.16	-	-
66	Gln	8.24	121.59	3.36	1.72,1.72	1.84,2.11	-	53.63	28.57	33.99	-	-
67	Pro	-	-	4.37	1.77,2.17	1.46,1.51	3.37,3.66	61.49	34.81	24.43	50.14	-
68	Trp	8.05	119.72	4.53	2.31,2.29	0.78,0.93	-	55.92	41.43	22.74,22.61	-	-
69	Val	9.04	120.48	4.48	2.06	0.78,0.93	-	61.93	32.48	-	-	-
70	Phe	9.29	123.32	5.44	2.96,2.75	-	-	56.3	41.66	-	-	-
71	Ile	8.04	112.9	5.12	1.53	1.62,0.96,0.82	0.72	58.72	39.75	26.12,18.54	14.04	-
72	Lys	8.96	125.42	5.12	1.54,1.53	1.30,1.23	1.49,1.49	53.8	36.21	24.34	29.43	41.84
73	Arg	8.97	123.09	4.49	1.75,2.03	1.78,1.66	3.28,3.25	57.42	31.21	28.2	43.71	-
74	Ser	8.74	120.88	4.47	3.85,3.91	-	-	58.52	64.14	-	-	-
75	Asp	8.31	122.31	4.61	2.71,2.78	-	-	54.26	42.02	-	-	-
76	Ala	8.36	124.79	4.32	1.41	-	-	52.82	19.61	-	-	-
77	Arg	8.43	119.01	4.32	1.76,1.85	1.60,1.66	3.15,3.21	66.73	30.86	27.18	43.41	-
78	Thr	8.04	114.1	4.31	4.15	1.17	-	61.8	69.8	21.84	-	-
79	Ala	8.28	127.78	4.62	1.24	-	-	50.41	18.57	-	-	-
80	Pro	-	-	4.3	1.71,2.18	1.84,1.84	3.50,3.68	63.69	31.71	27.56	50.47	-
81	His	8.36	117.85	4.48	3.22,3.17	-	-	56.59	29.6	-	-	-
82	Gly	8.28	108.82	4.00,3.65	-	-	-	45.44	-	-	-	-
83	His	8.08	119.4	4.39	2.87,2.61	-	-	55.98	31.59	-	-	-
84	Val	8.07	121.66	4.14	2.06	0.92,1.01	-	61.75	33.11	21.31,20.92	-	-
85	Met	8.77	127.89	5.13	1.94,2.21	2.45,2.48	-	55.81	33.56	32.68	-	-
86	Val	8.61	117.33	5.57	2.11	0.71,0.78	-	56.82	36.78	18.12,22.22	-	-
87	Glu	8.84	116.88	4.75	1.86,2.03	2.07,2.22	-	54.05	33.86	35.81	-	-
88	Leu	8.54	121.69	5.08	1.59,1.28	1.55	0.71,0.64	54.66	43.78	27.86	25.13,25.39	-
89	Val	8.73	120.02	4.52	1.9	0.63,0.77	-	60.7	36.31	20.46,21.45	-	-
90	Ala	8.69	126.04	5.83	1.32	-	-	50.45	22.78	-	-	-
91	Glu	8.86	118.47	4.93	2.40,2.18	2.39,2.39	-	55.43	34.66	36.73	-	-
92	Leu	8.84	123.87	4.29	1.82,1.59	1.6	0.92,0.92	57.06	43.36	27.33	24.20,24.70	-
93	Glu	-	-	4.46	2.07,2.21	2.29,2.29	-	57.1	30.61	36.77	-	-
94	Gly	8.22	106.36	4.06,3.97	-	-	-	46.22	-	-	-	-
95	Ile	7.57	119.61	3.94	1.85	1.12,1.41,0.73	0.81	61.99	38.45	17.38,27.73	12.56	-
96	Gln	-	-	4.24	1.88,1.79	2.14,1.92	-	56.76	28.98	33.76	-	-
97	Trp	8.88	120.45	4.34	2.37,2.56	-	-	58.52	37.77	-	-	-
98	Gly	7.94	108.99	3.94,3.69	-	-	-	46.1	-	-	-	-
99	Arg	7.84	118.66	4.4	1.67,1.89	1.56,1.60	3.15,3.15	55.63	30.66	27.11	43.45	-
100	Ser	7.86	115.7	4.39	3.90,3.87	-	-	58.35	63.74	-	-	-
101	Gly	8.69	112.79	-	-	-	-	-	-	-	-	-
103	Thr	-	-	4.34	4.14	1.02	-	63.16	69.25	23.02	-	-
104	Leu	9.03	126.76	4.5	1.58,1.40	1.7	0.78,0.75	56.24	42.89	28.85	24.36,25.65	-
105	Gly	7.43	103.93	4.09,4.24	-	-	-	45.46	-	-	-	-
106	Val	9.16	121.03	4.75	1.61	0.57,0.86	-	53.53	34.73	22.01,22.39	-	-
107	Leu	9	126.55	5.29	1.52,1.66	1.64	0.81,0.83	58.09	43.51	27.12	23.73,26.21	-
108	Val	8.34	113.37	3.97	2.22	0.48,0.63	-	63.55	33.37	23.44,17.83	-	-
109	Pro	-	-	4.2	1.65,2.02	1.06,1.66	3.21,1.39	53.05	32.92	27.68	48.52	-
110	His	8.03	120.57	5.1	2.87,2.47	0.94,0.89	-	60.6	31.85	22.02,21.81	-	-
111	Val	10.39	130.71	4.31	2.28	-	-	45.15	-	-	-	-
112	Gly	6.74	107.44	3.47,4.09	-	2.49,2.27	-	58.92	29.46	37.18	-	-
113	Glu	10.56	124.34	4.12	2.04,2.22	1.40,1.63,0.95	0.9	58.64	37.26	16.85,27.48	11.32	-
114	Ile	9.07	125.98	4.36	1.98	2.09,1.50	3.64,4.53	62.31	33.11	28.15	51.59	-
115	Pro	-	-	4.58	2.2	0.94,0.73	-	61.37	32.5	21.92,18.64	-	-
116	Val	8.84	114.3	4.31	1.38	-	-	51.18	42.07	-	-	-
117	Ala	7.72	120.66	4.41	2.62,2.62	-	-	57.77	42.07	-	-	-
118	Trp	8.46	116.05	5.24	1.46,1.65	1.39,1.36	3.06,3.24	53.23	32.03	26.8	42.26	-
119	Arg	9.36	122.27	4.74	1.54,0.94	1.17,1.38	1.43,1.43	56.18	32.37	24.96	29.42	-
120	Lys	8.61	126.7	5.08	-	-	2.83,2.83	-	-	-	-	41.97

Table 3.5 Assignments of all resonances of Nsp1, residues 121-180

Residue index	Amino acid	H <sup>N</sup> (ppm)	N (ppm)	H $\alpha$ (ppm)	H $\beta$ (ppm)	H $\gamma$ (ppm)	H $\delta$ (ppm)	He (ppm)	C $\alpha$ (ppm)	C $\beta$ (ppm)	C $\gamma$ (ppm)	C $\delta$ (ppm)	C $\epsilon$ (ppm)
121	Val	8.96	116.94	5.25	2.23	0.66,0.68	-	-	56.76	36.17	18.61,21.51	-	-
122	Leu	8.74	123.2	5.16	1.29,1.41	1.28	0.38,0.41	-	53.21	44.89	27.14	24.69,25.28	-
123	Leu	8.89	125.07	4.8	1.14,1.78	1.6	0.83,0.77	-	53.66	45.6	26.97	28.80,25.05	-
124	Arg	8.79	126.22	3.93	1.81,1.69	1.48,1.56	3.15,3.15	-	56.9	30.65	27.71	43.47	-
125	Lys	8.44	125.92	4.08	1.46,1.72	1.24,1.14	1.51,1.51	2.79,2.79	57.65	33.51	25.6	29.46	42.03
126	Asn	8.4	118.85	4.68	2.72,2.80	-	-	-	53.38	39.5	-	-	-
127	Gly	-	3.90,3.99	-	-	-	-	-	45.39	-	-	-	-
128	Asn	8.41	118.81	4.63	2.73,2.77	-	-	-	53.38	39.09	-	-	-
129	Lys	8.45	121.75	4.29	1.75,1.85	1.38,1.44	1.65,1.65	2.97,2.97	56.7	32.86	24.72	29.02	42.17
130	Gly	8.4	109.81	3.97,3.91	-	-	-	-	45.25	-	-	-	-
131	Ala	8.24	123.82	4.32	1.39	-	-	-	52.67	19.39	-	-	-
132	Gly	8.44	108.1	3.94,3.94	-	-	-	-	45.42	-	-	-	-
133	Gly	8.23	108.43	3.90,3.90	-	-	-	-	45.28	-	-	-	-
134	His	8.2	118.96	4.63	3.06,3.00	-	-	-	55.97	30.61	-	-	-
135	Ser	8.3	117.04	4.43	3.76,3.78	-	-	-	58.2	63.85	-	-	-
136	Tyr	8.33	122.26	4.56	3.07,2.94	-	-	-	56.22	36.78	-	-	-
137	Gly	8.36	110.53	3.86,3.92	-	-	-	-	45.41	-	-	-	-
138	Ala	8.11	123.66	4.27	1.37	-	-	-	52.78	19.39	-	-	-
139	Asp	8.35	118.66	4.56	2.71,2.61	-	-	-	54.25	40.99	-	-	-
140	Leu	8.05	122.23	4.27	1.63,1.57	1.58	0.88,0.82	-	55.46	42.12	27	25.04,23.51	-
141	Lys	8.25	121.31	4.28	1.76,1.86	1.36,1.40	1.65,1.65	2.97,2.97	56.39	32.79	24.73	29.06	42.18
142	Ser	8.11	116.11	4.38	3.76,3.78	-	-	-	58.37	63.8	-	-	-
143	Phe	8.14	121.65	4.61	2.98,3.14	-	-	-	57.72	39.7	-	-	-
144	Asp	8.28	121.84	4.61	2.55,2.69	-	-	-	54.2	41.53	-	-	-
145	Leu	8.22	122.82	4.32	1.60,1.63	1.6	0.85,0.91	-	55.56	42.36	27.07	23.75,25.04	-
146	Gly	8.43	109.11	3.93,3.93	-	-	-	-	45.54	-	-	-	-
147	Asp	8.2	120.45	4.61	2.62,2.69	-	-	-	54.46	41.4	-	-	-
148	Glu	8.46	120.85	4.25	1.96,2.06	2.24,2.29	-	-	56.84	30.05	36.34	-	-
149	Leu	8.27	122.43	4.34	1.68,1.58	1.63	0.91,0.85	-	55.43	42.27	26.97	23.41,25.10	-
150	Gly	8.37	109.31	3.96,3.99	-	-	-	-	45.42	-	-	-	-
151	Thr	7.98	113.15	4.33	4.18	1.16	-	-	61.58	70.09	21.7	-	-
152	Asp	8.48	124.47	4.86	2.75,2.55	-	-	-	52.4	41.41	-	-	-
153	Pro	-	-	4.34	2.14,1.65	1.75,1.90	3.70,3.82	-	63.54	31.97	27.02	50.76	-
154	Tyr	8.23	119.36	4.5	2.95,3.07	-	-	-	58.05	38.38	-	-	-
155	Glu	7.94	121.97	4.15	1.87,1.92	2.17,2.15	-	-	56.81	30.57	36.18	-	-
156	Asp	8.23	121.07	4.54	2.52,2.63	-	-	-	54.37	41.23	-	-	-
157	Phe	8.11	120.77	4.5	3.03,3.10	-	-	-	58.29	39.31	-	-	-
158	Gln	8.11	121.3	4.13	1.87,2.01	2.21,2.22	-	-	56.1	29.4	33.77	-	-
159	Glu	8.29	121.16	4.11	1.86,1.86	2.19,2.21	-	-	57.08	29.98	36.19	-	-
160	Asn	8.33	118.77	4.63	2.73,2.76	-	-	-	53.61	38.79	-	-	-
161	Trp	8.11	121.32	4.53	3.24,3.30	-	-	-	58.18	29.29	-	-	-
162	Asn	8.24	119.05	4.61	2.66,2.57	-	-	-	53.79	38.55	-	-	-
163	Thr	7.95	113.73	4.23	4.23	1.16	-	-	62.56	69.56	21.66	-	-
164	Lys	8.11	122.57	4.23	1.70,1.67	1.27,1.33	1.58,1.59	2.91,2.91	56.73	32.8	24.71	29.04	42.14
165	His	8.2	119.57	4.63	2.94,3.13	-	-	-	56.13	30.3	-	-	-
166	Ser	8.24	116.99	4.47	3.89,3.85	-	-	-	56.5	63.84	-	-	-
167	Ser	8.47	117.9	4.47	3.90,3.93	-	-	-	56.77	63.87	-	-	-
168	Gly	8.42	110.73	3.96,3.98	-	-	-	-	45.5	-	-	-	-
169	Val	7.96	119.43	4.16	0.92,0.92	-	-	-	62.61	32.76	21.23,20.61	-	-
170	Thr	8.11	117.9	4.32	4.26	1.21	-	-	62.22	69.83	21.85	-	-
171	Arg	8.4	123.3	4.23	1.78,1.85	1.59,1.63	3.19,3.17	-	57	30.72	27.17	43.39	-
172	Glu	8.44	121.32	4.2	1.96,2.02	2.26,2.26	-	-	57.36	29.96	36.33	-	-
173	Leu	8.14	122.58	4.29	1.65,1.60	1.64	0.86,0.92	-	55.78	42.29	27.01	24.20,23.57	-
174	Met	8.25	120.25	4.39	2.04,2.06	2.52,2.61	-	-	56.07	32.07	32.07	-	-
175	Arg	8.15	121.57	4.25	1.78,1.85	1.64,1.59	3.18,3.19	-	57.07	30.72	27.18	43.38	-
176	Glu	8.36	121.27	4.26	1.96,2.06	2.28,2.24	-	-	57.07	30.06	36.36	-	-
177	Leu	8.24	122.48	4.32	1.58,1.67	1.63	0.91,0.86	-	55.59	42.27	27	23.44,25.04	-
178	Asn	8.38	118.67	4.74	2.87,2.78	-	-	-	53.46	39.12	-	-	-
179	Gly	8.3	109.57	3.97,3.97	-	-	-	-	45.57	-	-	-	-
180	Gly	7.97	115.14	3.76,3.76	-	-	-	-	46.06	-	-	-	-

## Acknowledgements

Here, I would like to take this opportunity to thank all the great people I have met here in Germany, without your generous help and support, I would have never come this far.

First, I would like to thank my supervisor, Prof. Dr. Teresa Carlomagno, for offering me this opportunity to work in structural biology by using NMR spectroscopy although I had zero experience with NMR before coming to Germany. Her support, invaluable patience and encouragement mean a lot to me and enable me to accomplish my projects through these four years. Her rigorous attitude toward science greatly influenced me and my future career. It is of great honor to be trained scientifically under her supervision. Furthermore, I also want to thank Prof. Dr. Mark Brönstrup and Prof. Dr. Kürsad Turgay for being part of my thesis committee and taking interest into my research by reading my PhD thesis.

Major thanks to Dr. John Kirkpatrick, for the introduction of both theoretical and practical aspects of biomolecular NMR and the patience in helping me with different NMR experiments and discussing key ideas of my research. I would also like to thank Dr. Megha Karanth, for taking care of the entire laboratory management and offering technical helps no matter what happened in the lab.

It was a great pleasure to work with Dr. Petra Hinse, who initiated the project for me and helped me a lot during the first one year and a half of my PhD. I also want to thank all the present and past members that I met in the Carlomagno group during the past four years for creating a great working atmosphere and sharing their expertise. It is a great honor to work with all of you (John, Mumdooh, Michelangelo, Georg, Megha, Luca, Deepshikha, Veena, Neha, Simone, Natasha, Maria, Marco, Alex, Olga, Arun, Vittoria, Phillip, and Kim).

Last, I would like to thank my parents for their unconditional love and support. Due to the COVID-19 pandemic, I could not go back to China to meet them, but they are always there for me no matter what. I also want to thank my husband XiaoWei, Zhang who always believes in me, supports me, and helps me go through all the difficulties in my life.

谢谢你们 (Thank you) !!!

## Curriculum Vitae

**Ying Wang**

Citizenship: China

Gender: Female

**Center of Biomolecular Drug Research**

**(BMWZ), Leibniz University Hannover,**

**Schneiderberg 38, 30167 Hannover, Germany**

**Email: [ying.wang@oci.uni-hannover.de](mailto:ying.wang@oci.uni-hannover.de)**

**Mobile Phone: (+49) 15203097600**

### Education:

2018-

**Leibniz University Hannover (BMWZ, Germany)**

Ph.D. Structural biology

2015-2017

**Sookmyung Women's University (South Korea)**

M.Sc. Organic and Biochemistry

2010-2014

**Changchun University of Chinese Medicine (China)**

B.Sc. Pharmaceutical engineering

### Research

**Center of Biomolecular Drug Research (BMWZ),**

### Experience:

**Leibniz University Hannover, Germany**

Doctoral Researcher

Thesis: Structural insights into the activity regulation of full-length non-structural protein 1 from SARS-CoV-2 and Substrate recruitment by the hexameric MecA–ClpC complex.

Supervisor: Prof. Dr. Teresa Carlomagno

**Department of Chemistry, College of Natural Science**

**Sookmyung Women's University, South Korea**

Graduate Researcher

Thesis: Structural and Functional analyses of novel esterases from *Lactobacillus acidophilus*, *Lactococcus garvieae* and *Leuconostoc citreum*.

Supervisor: Prof. Dr. T. Doohun Kim

**College of Pharmacy,**

**Changchun University of Chinese Medicine, China**

Undergraduate Student Researcher

Thesis: Study on the mechanism of toxicity of cinnabar and its medication safety.

Supervisor: Prof. Dr. Bo Sun

### **Conferences and Posters:**

2021.07.05-08      **EUROMAR 2021 (European Conference on Magnetic Resonance)**

Virtual event, Germany

Poster presenter

2017.04.19-21      **Korean Chemical Society <sup>119</sup>th General Meeting**

Seoul, South Korea

Poster presenter

### **Research Skills:**

Molecular biology: Bacterial/Cell line culture; Cloning; PCR; Gel electrophoresis.

Recombinant protein expression and purification; Site directed mutagenesis; Affinity chromatography; biochemical and biophysical assays.

Structural Biology: Protein crystallization; NMR spectroscopy of proteins

### **References:**

#### **1. Prof. Dr. Teresa Carlomagno (PhD supervisor)**

Professor and Academic Lead of HWB-NMR. Integrative Structural Biology School of Biosciences, University of Birmingham, B15 2TT (UK).

Email: t.carlomagno@bham.ac.uk

#### **2. Prof. Dr. Doohun Kim (Master supervisor)**

Professor of Biochemistry, Department of Chemistry, College of Natural Science, Sookmyung Women's University (South Korea).

Email: doohunkim@sookmyung.ac.kr

## List of publications:

1. **Wang, Y.**, Kirkpatrick, J., Zur Lage, S., Korn, S.M., Neissner, K., Schwalbe, H., Schlundt, A., and Carlomagno, T. (2021).  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  backbone chemical-shift assignments of SARS-CoV-2 non-structural protein 1 (leader protein). *Biomol NMR Assign* 15, 287-295.
2. Altincekic N, Sophie M. Korn, Zur Lage S, Kirkpatrick J, Schwalbe H, **Wang Y**, Carlomagno T, Schlundt A, et al., Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. *Front.Mol. Biosci.*, 8: 653148 (2021)
3. Le L.T.H.L, Yoo W, **Wang Y**, Jeon S, Kim K.K, Kim H-W, Kim T.D, Dual functional roles of a novel bifunctional  $\beta$ -lactamase/esterase from *Lactococcus garvieae*. *Int. J. Biol. Macromol.* 206, 203-212 (2022)
4. Do H, **Wang Y**, Lee C W, Yoo W, Jeon S, Hwang J, Lee M J, Kim K K, Kim H-W, Lee J H, Kim T.D, Sequence Analysis and Preliminary X-ray Crystallographic Analysis of an Acetylerase (LgEstI) from *Lactococcus garvieae*. *Crystal.* 12(1),46 (2022)
5. **Wang Y**, Ryu B H, Yoo W, Lee C W, Lee J H, Kim T.D, Identification, characterization, immobilization, and mutational analysis of a novel acetylerase with industrial potential (LaAcE) from *Lactobacillus acidophilus*. *Biochim Biophys Acta.* 1862(1):197-210 (2018)