

The Structure and Evolution of Non-canonical Coiled Coils

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Jyoti Adlakha
aus Faridabad, Haryana
Indien

Tübingen
2018

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 14.11.2018

Dekan: Prof. Dr. Wolfgang Rosenstiel

1. Berichterstatter: Prof. Dr. Andrei Lupas

2. Berichterstatter: Prof. Dr. Karl Forchhammer

SUMMARY

Coiled coils are ubiquitous structural units of proteins which fulfill a wide range of biological functions. They can serve as molecular spacers, oligomerization motifs, mechanical levers in membrane fusion, components of cytoskeleton, as well as facilitate ion transport and signal transduction. Canonical coiled coils are highly regular, left-handed supercoiled bundles of two or more α -helices, which follow a characteristic heptad repeat pattern. However, other periodicities engendering different supercoils are possible. Insertion of two (nonads) or six (hexads) residues in a regular heptad repeat increases the supercoil strain so far, that the helical structure locally breaks. In trimeric coiled coils, the single helices continue as short β -strands, and assemble as a triangular structural element called β -layer. A previous study showed that nonad repeats yield a new structure, the α/β coiled coil, with regularly alternating α - and β -segments and only one backbone hydrogen bond per repeat in common with a heptad-repeat coiled coil. In this first project, I present the crystal structures of two hexad-repeat families. These minimalistic α/β coiled coils do not share any backbone hydrogen bond with a heptad-repeat coiled coil. Furthermore, conversion of hexads to heptads by simple insertion of one residue per repeat leads to the formation of a canonical coiled coil. These results further support previous data showing that novel backbone structures are possible within the allowed regions of Ramachandran space with minor mutations to a canonical coiled coil.

In the second project, I investigate the structural and functional characteristics of the mempromCC family, a group of conserved integral membrane proteins in prokaryotes and mitochondria. They exhibit a characteristic head-neck-stalk-anchor architecture, where a membrane-anchored trimeric coiled-coil stalk projects the N-terminal head domains via a β -layer neck. Humans express two mempromCC paralogs, MCUR1 and CCDC90B. Here, I present the crystal structure of the head domain of human CCDC90B and a full-length model of MCUR1. Cellular localization studies show that the prokaryotic and eukaryotic proteins localize to the cytoplasmic and inner mitochondrial membranes respectively, with an N-in C-out orientation in both cases. Using human MCUR1, an essential regulator of Ca^{2+} uptake through the mitochondrial calcium uniporter (MCU), I study the role of individual domains and find that the head interacts directly with MCU. Upon Ca^{2+} binding, MCUR1 head domain is destabilized which then accelerates its conversion to β -amyloid fibrils. Furthermore, I find

that MCUR1 is processed *in vivo* and show that a major part of its N-terminal disordered region is cleaved. Both the full-length and processed forms of MCUR1 interact with MCU; however, it is still unclear if they are functionally equivalent. Finally, comparison of mempromCC homologs with unrelated prokaryotic proteins at the structural and sequence level identifies the head domain as the family defining element.

In the third project, I study the effect of frameshift resistant (FSR) repeat amplification on the structure and function of existing and novel proteins. This particular type of repetition comprising units of $n \times 3$ base-pairs and lacking stop codons, encodes the same protein repeat of n residues in all three frames of equal sense. I focus on heptad FSR repeats which are significantly enriched and conform to coiled-coil periodicity. Through biophysical and biochemical methods, I show that these repeat insertions in proteins are mostly unstructured and have mainly deleterious effects. Using *Microcystis aeruginosa*, I investigate the *in vivo* expression of FSR repeat ORFs with proteome and transcriptome analyses and find that a number of them are transcribed, but undetectable at the protein level. From these results, it appears that FSR repeat amplification in bacterial genomes is a recent evolutionary event, whose products are initially unstructured and non-functional. Eventually they can obtain beneficial mutations to become more structured, giving rise to novel cellular functions.

ZUSAMMENFASSUNG

Coiled-Coils sind ubiquitäre Proteinstruktur-Motive, welche verschiedenste biologische Funktionen erfüllen. Sie dienen als molekulare Abstandshalter und Oligomerisierungsmodule, erfüllen Hebefunktion in Membranfusionsprozessen, stellen Komponenten des Zellskeletts dar und sind am Ionentransport und der Signalübertragung an Membranen beteiligt. Kanonische Coiled-Coils sind regelmäßige, linkshändig gedrehte spiralförmige Bündel von mindestens zwei α -Helices, deren Aminosäuresequenz ein typisches Heptaden-Muster aufweist. Neben diesem sich wiederholenden Muster von sieben Resten sind weitere Periodizitäten möglich, welche entsprechende Änderungen im Grad der Verdrehung der Helices aufweisen. Das Einfügen von zusätzlich zwei (Nonade) oder sechs (Hexade) Aminosäureresten in eine Heptade führt zur lokalen Überbeanspruchung der einzelnen Helices und letztlich zu ihrem Bruch. In trimeren Coiled-Coils werden die einzelnen Helices an der Insertionsstelle als kurze β -Stränge fortgesetzt und bilden ein dreieckförmiges Strukturelement, genannt β -layer. Vorherige Arbeiten zeigten, dass die repetitive Anordnung von Nonaden zur Ausbildung einer α/β -Coiled-Coil führt. Diese neuartige Struktur besteht aus alternierenden α - und β -Abschnitten und besitzt, im Gegensatz zu Heptaden, bei denen jede Aminosäure (i) eine solche Wasserstoffbrückenbindung mit dem vierten nachfolgenden Rest ($i+4$) ausbildet, je Wiederholungsmotiv nur noch eine Wasserstoffbrückenbindung im Backbone der Aminosäurekette. In diesem Projekt wurden die Kristallstrukturen zweier Motive, die repetitive β -layer im Hexaden-Abstand haben, gelöst. Diese zeigen minimalistische α/β -Coiled-Coils, die keine der genannten Wasserstoffbrückenbindung im Backbone der Aminosäurekette mehr aufweisen. Durch Einfügen von nur einer zusätzlichen Aminosäure wurden die Hexaden in Heptaden überführt und die α/β -Coiled-Coil in eine kanonische Coiled-Coil umgewandelt. Diese Ergebnisse zeigen, dass es möglich ist, durch die Einführung nur weniger Mutationen in das Heptaden-Muster einer Coiled-Coil neue Backbone-Strukturen, beschreibbar mit dem Ramachandran-Plot, zu generieren.

Das zweite Projekt beschäftigt sich mit strukturellen und funktionellen Untersuchungen der mempromCC-Proteinfamilie mit Homologen in Prokaryoten und Mitochondrien. Die Mitglieder dieser Familie sind trimere integrale Membranproteine mit einer charakteristischen Domänenarchitektur, bestehend aus Kopf, Hals, Stiel und Membrananker. Die N-terminale Kopfdomäne ist über einen β -layer mit einer Coiled-Coil (Stiel) verbunden, welche wiederum

in den C-terminalen Membran-Anker mündet. Im Menschen gibt es mit MCUR1 und CCDC90B zwei Paraloge. In dieser Arbeit werden die Kristallstruktur der Kopfdomäne von CCDC90B und ein Homologiemodell des gesamten MCUR1-Proteins präsentiert. Durchgeführte Studien bestätigen die Lokalisation prokaryotischer und eukaryotischer Homologer an der inneren Cytoplasma-Membran bzw. der inneren mitochondrialen Membran, mit dem C-Terminus verankert in der Membran und dem Stiel-Kopf-Abschnitt im Cytoplasma bzw. der mitochondrialen Matrix. Eine funktionelle Analyse der einzelnen Domänen von MCUR1, einem essentiellen Regulator des *mitochondrial calcium uniporter* (MCU), zeigt, dass die Kopfdomäne sowohl die direkte Bindung an MCU vermittelt als auch Ca^{2+} bindet. Ca^{2+} -Ionen destabilisieren die MCUR1-Kopfdomäne und fördern die Umwandlung α -helikaler Sekundärstruktur in β -amyloide Fibrillen. Weiterhin wird gezeigt, dass ein Großteil des laut Vorhersagen ungefalteten N-Terminus von MCUR1 *in vivo* durch zelluläre Proteasen abgespalten wird. Sowohl unprozessiertes als auch prozessiertes MCUR1 bindet an MCU. Es bleibt allerdings unklar, ob beide Formen auch funktionell äquivalent sind. Wie der Vergleich mit Strukturmodellen, die für bakterielle mempromCC- Vertreter erstellt wurden, zeigt, sind prokaryotische Homologe strukturell sehr ähnlich mit einer gewissen Variabilität in ihren Stieldomänen. Vergleichende Sequenz- und Strukturanalysen mit nicht verwandten Proteinen derselben Domänenarchitektur identifizieren die Kopfdomäne als das die mempromCC-Familie definierende Element.

Das dritte Projekt untersucht die Auswirkungen der Amplifikation sogenannter Frameshift-resistenter (FSR) Motive auf die Struktur und Funktion evolutionär konservierter als auch jüngerer Proteine. Dieser besondere Motiv-Typ umfasst repetitive Einheiten bestehend aus n Nukleotiden, die für keines der drei Stopp-Codons kodieren und wobei n nicht durch drei teilbar ($n \nmid 3$) ist. Diese Repeats kodieren Aminosäuresequenz-Motive, die aus n Resten bestehen und sind in allen drei Leserahmen identisch. Mit Fokus auf repetitive FSR-Motive, die für Heptaden kodieren, sollte ein möglicher struktureller und damit evolutionärer Bezug zu Coiled-Coils untersucht werden. Mittels biophysikalischer und biochemischer Methoden wurde anhand natürlich vorkommender Beispiele gezeigt, dass die repetitive Insertion solcher Motive fast immer zur Zerstörung der nativen Struktur und Beeinträchtigung der Funktion der Proteine führt. Durchgeführte Transkriptom- und Proteom-Analysen zu Proteinen mit FSR-Motiven in *Microcystis aeruginosa* zeigen, dass viele der Gene *in vivo* transkribiert werden, während die entsprechenden Proteine meist nicht nachweisbar sind. Diese Ergebnisse lassen

vermuten, dass es sich bei derartigen FSR-Repeat-Amplifikationen in bakteriellen Genomen sehr wahrscheinlich um jüngere evolutionäre Ereignisse handelt, die zu Strukturverlust und Funktionsbeeinträchtigung führen. Dennoch stellen sie möglicherweise einen der Ausgangspunkte für die Evolution von strukturell und funktionell neuartigen Proteinen dar.

CONTENTS

SUMMARY	7
ZUSAMENFASSUNG	9
CONTENTS	13
LIST OF FIGURES	17
LIST OF TABLES	19
ABBREVIATIONS	21
Chapter 1: Introduction to Coiled Coils	
1.1 Brief History.....	25
1.2 Structure of canonical coiled coils.....	25
1.3 The GCN4 leucine zipper.....	26
1.4 Non-canonical coiled coils.....	27
1.4.1 Discontinuities.....	27
1.4.2 Bifaceted coiled coils.....	29
1.4.3 Polar residues at core positions.....	29
1.5 Folding and stability.....	30
1.6 Functional roles of coiled coils.....	31
Chapter 2: Structural characterization of α/β coiled coils derived from hexad repeats	
2.1 Introduction	
2.1.1 What is a β -layer?	34
2.1.2 Single β -layer insertions.....	36
2.1.3 The α/β coiled coil from repetitive nonads.....	38
2.1.4 β -layers as capping structures.....	38
2.2 Results	
2.2.1 Structural characterization of the “KAD-VYT-LYT” hexad module.....	40
The trimeric autotransporter adhesin NadA5 from <i>Neisseria meningitidis</i>	48
2.2.2 Structural characterization of the “TAT” repetitive hexads module.....	50
Crystal structure of ParcuUY81 with GCN4-N16V fusion.....	52
NMR-spectroscopy for ParcuUY81.....	54
2.3 Contributions to this work	55
2.4 Discussion	56
2.5 Methods	
2.5.1 Molecular cloning.....	58
2.5.2 Protein expression and purification.....	60

2.5.3 Homology modelling.....	61
2.5.4 Crystallization, data collection and structure determination.....	61

Chapter 3: Structural and functional characterization of MempromCC proteins – a new family of membrane proteins in prokaryotes and mitochondria

3.1 Introduction

3.1.1 The mempromCC protein family.....	64
3.1.2 Human mempromCC proteins - MCUR1 and CCDC90B.....	68
3.1.2.1 MCUR1 as a regulator of calcium uptake by MCU.....	68
The MCU complex in mitochondria.....	69
Calcium uptake occurs at MAMs.....	69
Components of the MCU complex.....	71
A. MCU – the central pore-forming protein.....	71
B. MCUb – the dominant negative paralog of MCU.....	73
C. MICU1, MICU2 and MICU3 – the gatekeepers.....	73
D. EMRE – the scaffold factor.....	74
3.1.2.2 MCUR1 as a cytochrome c oxidase (COX) assembly factor.....	75
3.1.2.3 MCUR1 as a regulator of Ca ²⁺ threshold for mitochondrial permeability transition.....	76
3.1.3 Pathophysiological implications of the MCU complex.....	76
3.1.4 Aims of the study.....	77

3.2 Results

3.2.1 The human mempromCC proteins – MCUR1 and CCDC90B.....	78
3.2.1.1 Bioinformatic sequence analyses.....	78
3.2.1.2 Crystal structure of human mempromCC protein CCDC90B.....	80
3.2.1.3 Structural characterization of MCUR1.....	83
3.2.1.4 Homology model of MCUR1.....	84
3.2.1.5 <i>In vitro</i> Ca ²⁺ binding to MCUR1 and CCDC90B.....	84
3.2.1.6 The head domain of MCUR1, but not CCDC90B, can form β-amyloid fibrils	
3.2.1.7 Effect of Ca ²⁺ on MCUR1 amyloid fibril formation.....	88
3.2.1.8 Does MCUR1 directly interact with MCU-NTD?.....	93
Co-immunoprecipitation of MCU with MCUR1.....	94
<i>In vitro</i> binding assay using MST.....	95
3.2.1.9 Subcellular localization of human MCUR1 and CCDC90B.....	97
3.2.1.7 MCUR1 is proteolytically processed in human mitochondria.....	99
Is this proteolysis mediated by Cyclophilin D?	99
Does MCUR1 head domain bind to Cyclophilin D?	101
3.2.2 Prokaryotic mempromCC proteins – structural and functional studies.....	103
3.2.2.1 Structural characterization of <i>Caulobacter</i> mempromCC homologs.....	103
3.2.2.2 Structural characterization of archaeal Kcr-0859 showing mempromCC-	

related domain architecture.....	104
3.2.2.3 Subcellular localization and orientation of mempromCC in bacteria.....	106
3.2.2.4 Functional characterization of mempromCC proteins in <i>Caulobacter</i> <i>crenscentus</i> NA1000.....	108
Pull-down to identify interaction partners.....	108
3.3 Contributions to this work.....	110
3.4 Discussion	
3.4.1 MempromCC are obligate trimers with a head-neck-stalk-anchor architecture.....	112
3.4.2 Eukaryotic mempromCC proteins localize to IMM and prokaryotic homologs to the cytoplasmic membrane, with an N-in C-out topology.....	112
3.4.3 Eukaryotic mempromCC proteins have diverse N-terminal disordered extensions	
3.4.4 Structure of the conserved head domain of human CCDC90B and MCUR1.....	113
3.4.5 MCUR1 head domain is destabilized by Ca ²⁺ binding which accelerates its conversion to β -amyloid fibrils.....	114
3.4.6 MCUR1 head domain interacts with the N-terminal domain of MCU.....	115
3.4.7 MCUR1 is proteolytically processed in human mitochondria.....	116
3.4.8 Cellular function of prokaryotic mempromCC homologs.....	117
3.5 Methods	
3.5.1 Molecular cloning.....	118
3.5.2 Protein expression and purification.....	118
3.5.3 CD spectroscopy.....	119
3.5.4 Microscale Thermophoresis (MST).....	120
3.5.5 SEC-MALS.....	120
3.5.6 Mammalian cell culture and transfection.....	121
3.5.7 Immunoprecipitation.....	121
3.5.8 Mitochondria isolation and subfractionation.....	121
3.5.9 Subcellular fractionation of <i>C. crescentus</i> NA1000.....	122
3.5.10 Proteinase K assay.....	123
3.5.11 Electron microscopy.....	123
3.5.12 NMR spectroscopy.....	124
3.5.13 Crystallization and structure determination.....	124
3.5.14 Homology modelling.....	125
 Chapter 4: Investigating frameshift-resistant repeat amplification in functional proteins	
4.1 Introduction	
4.1.1 Defining tandem repeats.....	128
4.1.2 Frameshift resistant tandem repeats.....	129
4.1.3 Mechanism of tandem repeat instability.....	130
4.1.4 Functional impact of tandem repeats.....	131

4.1.5 Aims of this study.....	133
4.2 Results	
4.2.1 Expression and characterization of frameshift resistant (FSR) repeat containing proteins and their non-FSR homologs.....	134
4.2.1.1 FSR repeats inserted between functional protein domains	
1. NADH dehydrogenase (<i>Moorea producens</i> 3L).....	137
2. Serine/threonine protein kinase (<i>Moorea producens</i> 3L).....	140
4.2.1.2 FSR repeats replacing part of a functional protein domain	
1. Restriction endonuclease-like hypothetical protein MAE_23560 (<i>M. aeruginosa</i> NIES-843).....	142
2. HIRAN domain protein (<i>Moorea producens</i> 3L).....	143
4.2.1.3 FSR repeats spanning whole protein domain.....	144
4.2.2 Crystallizing FSR repeat peptides.....	144
4.2.3 Proteome analysis to detect FSR repeat proteins.....	145
4.2.4 Transcriptome analysis for <i>M. aeruginosa</i> strains NIES-843 and NIES-44.....	146
4.3 Contributions to this work.....	150
4.4 Discussion.....	151
4.5 Methods	
4.5.1 Molecular cloning.....	155
4.5.2 Protein expression and purification.....	155
4.5.3 CD spectroscopy.....	156
4.5.4 Crystallization, data collection and structure determination.....	156
4.5.5 Culturing <i>Microcystis aeruginosa</i> cells.....	157
4.5.6 Proteome analysis.....	157
4.5.7 Immunoprecipitation by coupling antibody to NHS-activated HP column.....	157
4.5.8 Total RNA isolation.....	158
4.5.9 cDNA preparation.....	159
4.5.10 qRT-PCR.....	159
4.5.11 Transcriptome analysis.....	159
ACKNOWLEDGEMENT.....	163
BIBLIOGRAPHY.....	167
APPENDIX I.....	183
APPENDIX II.....	184
APPENDIX III.....	185
APPENDIX IV.....	187
CURRICULUM VITAE.....	191

LIST OF FIGURES

Fig. No.	Figure legend	Page No.
2.1	Periodicity plot for the insertion of 1-6 residues in a canonical heptad repeat	34
2.2	Single nonad β -layer in <i>Actinobacillus</i> OMP100 stalk	37
2.3	The α/β coiled-coil from nonad repeats	39
2.4	Identified sequences with KAD-VYT-LYT co-optimized module	41
2.5	Crystal structures for KAD-VYT-LYT family hexad motifs	42
2.6	Crystal structure of Mmor β -layers	43
2.7	Interaction network of the KAN-VYT-VYT β -layer module of Mmor	44
2.8	Mutation of conserved hexad β -layer residues	46
2.9	Crystal structure of Mglob hexad β -layers	47
2.10	Model of NadA5	49
2.11	Biophysical characterization of ParcuUY81	51
2.12	Sequence and crystal structure of ParcuUY81-heptad	52
2.13	Crystal structure and interaction network of ParcuUY81-GCN4	53
3.1	Bioinformatic classification of mempromCC protein family	65
3.2	The head-neck segment of mempromCC family is evolutionarily conserved	67
3.3	Mitochondria-Associated Membranes (MAMs)	70
3.4	Structure of pore-forming MCU protein	72
3.5	Disorder prediction in human MCUR1	79
3.6	Biophysical characterization of CCDC90B ₄₃₋₁₂₅ -GCN4 N16V	80
3.7	Crystal structure of human CCDC90B	82
3.8	NMR 1D ¹ H spectrum of MCUR1 ₁₆₀₋₂₃₀	83
3.9	Structural model of human MCUR1	85
3.10	Low-affinity Ca ²⁺ binding to MCUR1	86
3.11	Characteristics of cross- β amyloid fibrils	87
3.12	MCUR1 is destabilized upon divalent cation binding	89
3.13	Secondary structure of CCDC90B ₄₃₋₁₂₅ -GCN4-N16V and CCDC90B ₄₃₋₁₂₅	90
3.14	TEM micrographs showing β -amyloid fibril formation of MCUR1 ₁₆₀₋₂₃₀	91
3.15	Prediction of aggregation and amyloid propensity of MCUR1	92
3.16	Structural model of β -amyloid formation in MCUR1	93
3.17	MCU complex representation	93
3.18	MCUR1 head domain directly interacts with N-terminal domain of MCU	95
3.19	Oligomer formation of MCU ₇₅₋₂₃₃	96
3.20	<i>In vitro</i> MCU binding to MCUR1 and CCDC90B	97
3.21	Subcellular localization of human MCUR1 and CCDC90B	98
3.22	Mass spectrometric analysis of MCUR1 fragments	100
3.23	Structure and mechanism of action of cyclophilin D	101
3.24	MCUR1 proteolysis and binding to cyclophilin D	102

3.25	Structure of prokaryotic mempromCC homologs	105
3.26	Subcellular localization of prokaryotic mempromCC homologs	107
3.27	Pull-down to identify MpcC interacting partners	109
4.1	Schematic representation of Frameshift resistant (FSR) repeats	129
4.2	FSR repeat insertion in NADH dehydrogenase	139
4.3	Characterization of S/T kinase FSR repeat insert	141
4.4	Biophysical characterization of SUMO-MAE_23560	143
4.5	Covalent immobilization chemistry of purified antibody to the NHS-activated sepharose HP-column	146
4.6	RT-PCR analysis for MAE_23560	147

LIST OF TABLES

No.	Title	Page No.
2.1	List of primers for PCR amplification	58
2.2	Sequence of hexad repeat constructs	59
2.3	Crystallization and cryo conditions	62
3.1	Data collection and refinement statistics	126
4.1	List of FSR repeat containing sequences selected for structural and functional characterization	135
4.2	List of FSR repeat peptides set-up for crystallization	136
4.3	Identified FSR transcripts in <i>Microcystis aeruginosa</i> NIES-843	148
4.4	Identified FSR transcripts in <i>Microcystis aeruginosa</i> NIES-44	149

ABBREVIATIONS

AFG3L2	ATPase family gene 3-like 2
AHTC	anhydrotetracycline
ANT	adenine nucleotide translocase
ADP	adenosine diphosphate
ATP	adenosine triphosphate
β-ME	β-mercaptoethanol
BN-PAGE	blue native polyacrylamide gel electrophoresis
CD	circular dichroism
cDNA	complementary DNA
CypD	cyclophilin D
CCDC90A	coiled-coil domain containing protein 90A, or MCUR1
CCDC90B	coiled-coil domain containing protein 90B
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMRE	essential MCU regulator
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
FSR	frame-shift resistant
GCN4	general control nonderepressible 4 protein
HEK293	human embryonic kidney cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC	hydrophobic interaction chromatography
HIRAN	HIP1 16 Rad5p N-terminal
HPLC	high performance liquid chromatography
HRP	horse-radish peroxidase
IF	intermediate filaments
IMAC	immobilized metal affinity chromatography
IMM	inner mitochondrial membrane
IMS	inter-membrane space (mitochondria)
IPTG	isopropyl β-D-1-thiogalactopyranoside
KIH	knobs-into-holes
LC-MS/MS	liquid chromatography tandem mass spectrometry
MCU	mitochondrial calcium uniporter
MCUb	dominant-negative MCU regulator
MCUR1	mitochondrial calcium uniporter regulator 1
MICU1	mitochondrial calcium uptake protein 1
mRNA	messenger RNA
miRNA	micro RNA

MRE	mean residue ellipticity
MS	mass spectrometry
MST	microscale thermophoresis
NADH	nicotinamide adenine dinucleotide (reduced)
NIES	National Institute for Environmental Studies, Japan
NiNTA	nickel nitrilotriacetic acid
NMR	nuclear magnetic resonance
OMM	outer mitochondrial membrane
ORF	open reading frame
PBS	phosphate buffered saline
PDB	protein databank
PMSF	phenylmethylsulfonyl fluoride
PTP	permeability transition pore
qRT-PCR	quantitative real-time polymerase chain reaction
RMSD	root-mean-square deviation
RNA	ribonucleic acid
rRNA	ribosomal RNA
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SPG7	spastic paraplegia 7
SUMO	small ubiquitin-like modifier proteins
TAA	trimeric autotransporter adhesin
TBS	tris-buffered saline
TCEP	tris(2-carboxyethyl)phosphine
TEV	tobacco etch virus protease
tRNA	transfer RNA
TSPO	translocator protein

CHAPTER 1

Introduction to Coiled coils

1.1 Brief History

Coiled coils are highly versatile protein structural elements involved in numerous biological activities. The first structural investigations in coiled coils came from William Astbury, who obtained X-ray diffraction patterns for natural fibers such as wool (in native and denatured forms), as well as horns, tendons, hair, and porcupine quills. His work revealed three main diffraction patterns: an α -form shown by unstretched wool, a β -form from stretched wool and a γ -form corresponding to tendons (collagen). α -form, the most common diffraction pattern with strong meridional arcs at 5.15 Å and equatorial reflections at 10 Å and 27 Å, was generated by a group of proteins referred to as ‘k-m-e-f’ for keratin, myosin, epidermin and fibrinogen. By 1953, Linus Pauling and Francis Crick independently proposed the model of supercoiled helices for this group of proteins which we now refer to as coiled coils (Pauling et al., 1951; Pauling and Corey, 1953; Crick, 1952, 1953a, 1953b). While Pauling only considered backbone periodic configuration and envisaged a set of periodicities (4/1, 7/2, 11/3, 15/4) and stoichiometries for constituent helices, Crick provided the first fully parameterized model for the sequence periodicity of 7/2 (seven residues over two helical turns). He placed side-chain packing interactions at the core of his model recognizing that when α -helices are twisted around each other by 20°, their side-chains would interlock systematically along the fiber with the same interactions repeated every 7 residues (or 2 helical turns). He referred to this regular side-chain packing as “knobs-into-holes”, in which one knob from a helix packs into a hole formed by four side chains of the neighboring helix.

1.2 Structure of canonical coiled coils

Coiled coils are superhelical bundles composed of two or more α -helices, running in a parallel or anti-parallel direction. As proposed by Crick (Crick, 1953a, b), helices in a coiled coil interact via knobs-into-holes (KIH) arrangement of residue side-chains, which occupy equivalent positions along the bundle interface. This geometry contrasts with the more irregular ridges-into-grooves packing of regular α -helices in globular proteins, where a residue packs

above or beneath the equivalent residue from the opposite helix (Chothia et al., 1977). For this reason, KIH packing is sometimes referred to as ‘in-register’ and ridges-into-grooves as ‘out-of-register’. Regular KIH packing requires periodically recurring residues along the interface. Coiled coils achieve this by giving a left-handed twist to right-handed α -helices, effectively reducing the periodicity from 3.63 residues per turn (r/t) of undistorted α -helices to 3.5 r/t which corresponds to seven residues repeating every two helical turns. The seven positions of a heptad repeat are labelled *a-g*, where *a* and *d* core positions are occupied by hydrophobic residues and the remaining solvent exposed positions (*b*, *c*, *e*, *f* and *g*) mostly by hydrophilic residues. This Crick model of a heptad repeat based coiled-coil is termed as “canonical”. Owing to the regular nature of coiled coils, their structures can be fully described by parametric equations. As a consequence, coiled coils have been exploited in protein design efforts, to understand the general relationship between sequence and structure.

1.3 The GCN4 leucine zipper

GCN4 (General Control Nonderepressible), a eukaryotic transcriptional activator protein, belongs to the bZIP (basic leucine zipper domain containing) family of proteins. It is responsible for the activation of more than 30 genes involved in amino acid biosynthesis under starvation conditions in yeast (Hope and Struhl, 1985; Arndt and Fink, 1986). GCN4 leucine zipper is one of the best studied coiled coil domains. It is a 31-residue parallel dimer, built of four complete heptad repeats (O’Shea et al., 1991; Landschulz et al., 1988). The *d* position of heptad repeats is primarily occupied by the leucine residue, hence the given name. GCN4 also features a conserved asparagine (N16) at one of the *a* positions which confers dimer specificity and forces zipper in a parallel orientation which promotes stability by self-complementary hydrogen bonding. Polar residues can be easily accommodated in a dimer as the core residue in position *a* is solvent accessible. In higher oligomeric forms such as trimers or tetramers, this position is buried deep in the core and therefore prefers hydrophobic residues.

Harbury et al. in 1993 established rules governing the oligomerization specificity of GCN4 leucine zipper. In a dimeric parallel coiled-coil, as the C_{α} - C_{β} bond vector of the residue in position *a* is parallel to the equivalent bond from the facing helix, it favors β -branched residues such as Ile, Val, and Thr; whereas in position *d*, it directly points into the core of the superhelix (perpendicular geometry) and prefers γ -branched residue leucine. For a parallel tetramer, core packing geometry is exactly reversed: *a* residue favors perpendicular and *d*

favors parallel packing. In a trimer, core packing angle is intermediate of that in the parallel and perpendicular arrangement; therefore it is called ‘acute’. Acute positions do not display any residue preferences. Since then, a number of studies have reported different structural forms brought about by minor changes in the sequence of GCN4-p1 leucine zipper. For example, mutation of N16 to a valine resulted in a mixture of stable dimers and trimers (Harbury et al., 1993; Knappenberger et al., 2002). In another study, replacing hydrophilic residues at both *e* and *g* positions (involved in salt-bridge interactions in the oligomeric state) with nonpolar alanines gave a soluble seven-helical coiled-coil bundle (Liu et al., 2006).

I have employed this knowledge of oligomer preference of core residues of the GCN4 leucine zipper and their high stability to express and purify difficult proteins. In the next chapters, I will discuss a number of cases where GCN4 variants have been fused to the proteins of interest for structural and biophysical characterization. A modified version of pASK-IBA expression vector was used to fuse N- or C-terminal (or both) of the protein of interest to stabilizing GCN4 adaptors in the correct heptad register (Hernandez-Alvarez et al., 2008). Prior information of the native coiled-coil oligomer state is helpful in selecting the appropriate GCN4 variant. Such a system is especially useful for proper folding of protein domains flanked by coiled coils in their native context or coiled-coil domains that lack trigger sequences (see section 1.5) (Steinmetz et al., 1998; Kammerer et al., 1998) which would otherwise give an unfolded, insoluble product.

1.4 Non-canonical coiled coils

A significant proportion of naturally identified coiled coils deviate from the canonical heptad pattern, periodicity and core packing geometry as described by Crick. In this section, I describe the most commonly encountered non-canonical coiled coils – (i) non-heptad periodicities arising out of the insertion or deletion of one or more residues from a regular heptad repeat; (ii) bifaceted coiled coils harboring more than one hydrophobic interface; and (iii) polar residues occupying the canonical hydrophobic positions of the heptad repeat.

1.4.1 Discontinuities

Although coiled coils are fairly regular structures, discontinuities can arise from the insertion or deletion of residues in the heptad pattern (Lupas and Gruber, 2005). Such discontinuities can be structurally accommodated by perturbations in the packing of coiled coils. Insertion of

three and four residues, called stammers and stutters respectively (Brown et al., 1996), are close to the periodicity of an undistorted α -helix (3.63 r/t) which allows their accommodation within the helical structure. However, they tend to distort KIH core packing interactions. Stutters raise the local periodicity to 3.67 r/t resulting in local unwinding of the superhelix. Stammers have an opposite effect; they reduce the periodicity to 3.33 r/t, leading to an overwinding of constituent helices. Both insertions, however, have the same effect on core packing: they lead to a local adoption of x - da geometry. Stutters shift residues in position a towards the center of the core (x layer), while moving d outwards and e inwards (da layer) yielding an a - d - e core. Similarly, in stammers, counterclockwise drift (as seen from the N-terminus) results in d residues occupying the x layer position, and a and g residues forming the da layer, yielding an overall a - d - g core. In both cases, x layer residues point towards the coiled-coil axis while the da layer residues form a ring of interacting residues enclosing a central cavity. In this arrangement, KIH packing locally transforms to knobs-into-knobs interaction (Lupas and Gruber, 2005).

Single insertions can also be delocalized over multiple heptads leading to higher coiled-coil periodicities (e.g. $(7+7+4)/5 = 18/5$ or $(7+7+7+4)/7 = 25/7$), and multiple stutter and stammer insertions can also combine with a single heptad repeat (e.g. pentadecads $(7+4+4)/4 = 15/4$ or 3.75). Coiled coils can only assume a certain range of periodicities, which is limited by the supercoil strain tolerated by the constituent helices. The lower limit of this range is around 3.33, i.e. insertion of a stammer into a heptad repeat ($10/3$; $3.63-3.33 = 0.3$ r/t less than an undistorted helix). From this, the estimated upper limit is around 3.9. Insertions of 1 residue can be either looped out of the α -helix (skip) resulting in the formation of a π -turn at the site of insertion or delocalized over multiple heptad repeats (Lupas and Gruber, 2005). Similarly, accommodation of 5 residue insertion by delocalization over multiple heptads results in periodicities that fall within the accessible range. The most demanding for a coiled-coil bundle are the insertions of 2 and 6 residues. In parallel trimeric coiled coils, they were shown to locally form short β -strands which rotate the path of each chain by 120° around the supercoil axis and associate to form a new structural element known as the β -layer (Hartmann et al., 2016). Repetitive arrangement of β -layers generate a new fibrous structure called α/β coiled-coil. In Chapter 2, I will undertake a detailed examination of the structural features and interaction networks formed by α/β coiled coils derived from tandem nonad (7+2) and hexad (7+6 or 7-1) discontinuities.

1.4.2 Bifaceted coiled coils

For coiled coils, positions flanking the core residues equally influence the oligomerization state and helix orientation. In two and three-stranded coiled coils, positions *e* and *g* are mostly occupied by charged residues which shield the hydrophobic core and provide stabilizing interactions by forming salt-bridges. Inclusion of non-polar residues at either position results in a broader hydrophobic surface which favors the formation of tetramers over dimers or trimers. This broader interface can be pictured as having two seams of core residues, (*g-d* and *d-a*) or (*d-a* and *a-e*), with one common position (Walshaw et al., 2001; Walshaw and Woolfson, 2003; Woolfson et al., 2012). Known as Type I bifaceted coiled coils, their core geometry differs from canonical Type N interfaces. The axially symmetric packing causes the shared position to point directly towards the central axis (*x* geometry), while the other two positions point sideways enclosing a central cavity (*da* geometry). We have already discussed this packing geometry in section 1.4.1, for stutters and stammers. The advantage of combining *x* positions in one pair of diagonally opposite helices together with the *da* positions of others (or complementary *x-da* packing; Hulko et al., 2006) provides the preference for an anti-parallel orientation.

As the two hydrophobic seams move further apart, they prefer higher oligomeric states. Adjacent hydrophobic seams (*g-d* and *a-e*) or Type II interfaces, lead to the formation of pentamers, hexamers and heptamers (Thomson et al., 2014; Huang et al., 2014). Including a fifth residue results in two hydrophobic seams separated by an intervening residue (*g-d* and *e-b*, separated by *a*; or *c-g* and *a-e*, separated by *d*). This arrangement, known as Type III bifaceted coiled coils, allows for the formation of even higher oligomers. The largest known protein under this category is a 12-helical anti-parallel barrel in the multidrug efflux pore protein TolC (Koronakis et al., 2000). Bifaceted helices can therefore, produce structures that range from α -helical fibers (or bundles), to tubes with large diameters and solvent-filled pores (Thomson et al., 2014), and to even fully open α -sheets (Egelman et al., 2015).

1.4.3 Polar residues at core positions

Core residues, *a* and *d*, of a regular heptad coiled coil are primarily occupied by hydrophobic residues, whose close packing interactions provide the energy required to distort α -helices. However, nearly 25% of coiled coil sequences were identified to contain polar residues at either position (Conway et al., 1991). Although thermodynamically disfavored, they seem to impart

structural specificity at the expense of reduced stability. An example is the GCN4 leucine zipper. Trimeric autotransporter adhesins (TAAs) also contain a high content of polar core residues, especially occupying position d of the heptad repeat (Hartmann et al., 2009). Asparagine is the most commonly identified residue. It results in the formation of a special motif namely $N@d$ layers which can coordinate monovalent anions (iodide, chloride, bromide, nitrate) at the center. Consecutive insertion of $N@d$ layers is found in many TAAs. In Chapter 2 (section 2.2.1), we discuss the full-length structure of NadA5, a TAA found in *Neisseria meningitidis* (Malito et al., 2014). NadA5 shows two non-canonical coiled coil features: tandem insertions of 2 and 6 residues which result in repetitive β -layer formation (called α/β coiled-coil), and the presence of multiple $N@d$ layers which coordinate iodide ions. Insertion of $N@d$ layers has been biochemically confirmed to reduce structural stability. GCN4 mutants with engineered $N@d$ layers unfolded at much lower temperatures, and showed concentration-dependent folding (Hartmann et al., 2009). Based on this property, it has been proposed that multiple $N@d$ insertions maintain the coiled coils of TAAs in a natively unfolded, export-competent state until complete autotransport has occurred through the bacterial outer membrane to avoid problems in proper folding and transport of long fibers. Apart from $N@d$, other widely represented polar motifs in TAAs and various homotrimeric coiled coils include SxxNTxx, NxxQDxx, QxxHxxx, QxxDxxx etc., where polar residues are seen to occupy both a and d positions.

1.5 Folding and stability

As a result of their regularly repeating interaction network, coiled coils are generally very stable proteins. The most stable coiled coil known to date is tetrabrachion, homotetrameric stalk domain of a surface-layer protein of *Staphylothermus marinus*. It can withstand heating up to 130°C in 6M guanidinium chloride, denaturing only in 70% sulfuric acid as a result of the hydrolysis of peptide bonds (Peters et al., 1995). Although this is exceptional, other known coiled coils also display high stability and resistance to chemical and thermal denaturation, for instance, the GCN4 leucine zipper and its variants. Primary factors that contribute to coiled-coil stability include helical propensity, core hydrophobicity, tightness of the core packing and favorable ionic interactions that shield the core from solvent molecules (Lupas and Gruber, 2005).

However, it is quite surprising that coiled coils are frequently predicted to be natively unstructured by various disorder prediction programs (Lupas et al., 2017). The reason lies in the highly repetitive, solvent-exposed structures of coiled coils; they have a low proportion of hydrophobic residues (required only at 2 positions out of 7) compared to globular proteins, and a reduced sequence complexity – both factors in common with the natively disordered proteins. A question arises – why have coiled coils evolved to resemble natively unstructured sequences? One possible explanation could be the need to ensure proper in-register folding of long fibers. For example, in a myosin rod that extends for nearly 1000 residues, packing interactions are the same along the entire length of the rod. If local interactions formed rapidly and randomly, they would trap the chains in a natively out-of-register conformation. To prevent this, specialized “trigger sequences” have evolved in coiled coils (Steinmetz et al., 1998; Kammerer et al., 1998; Steinmetz et al., 2007). They are short, autonomous folding units that serve as nucleation sites and guarantee an in-register assembly of constituent helices. While trigger sequences do not display any consensus sequence motif, they are characterized by short segments of high α -helical propensity which are capable of forming many interactions stabilizing the correct oligomeric form (Steinmetz et al., 1998).

1.6 Functional roles of coiled coils

Coiled coils are widespread structural elements found in nearly 10% of eukaryotic and up to 5% of prokaryotic proteins (Walshaw and Woolfson, 2001; Liu and Rost, 2001). They fulfill a wide array of biological functions (Lupas, 1996; Truebestein and Leonard, 2016; Hartmann, 2017). As coiled-coil fibers display high mechanical strength, they are found in hair, feathers, horns, and nails (Lupas and Gruber, 2005). In eukaryotes, they constitute the basic building blocks of intermediate filaments, which are the essential components of cytoskeleton and nucleoskeleton. Coiled-coil fibers are also frequently found on bacterial cell surfaces (e.g. flagellins, pilins and adhesins). As elongated structures, they can serve as molecular spacers projecting domains across large distances. An important role of coiled coils lies in their ability to mediate oligomerization, for example in transcription factors (leucine zippers), signaling molecules (G protein $\beta\gamma$), and molecular motors. In motor proteins, myosin and kinesin, extended coiled-coil domains contribute to cellular motility (Squire et al., 2017). They can function as levers in processes like vesicle tethering, membrane fusion and chromosome segregation (Matityahu and Onn, 2018; Witkos and Lowe, 2017). pH-dependent conformational change of an unstructured loop into a trimeric coiled coil was shown for

membrane fusion mediating influenza hemagglutinin (Carr and Kim, 1993; Bullough et al., 1994), which projects the fusion peptide away from viral surface into the host membrane at a distance of 100Å.

Coiled coils are also found as components of channels where they facilitate ion transport across membranes. A number of protein design efforts have focused in this direction; for example, design of Rucker, a zinc/H⁺ antiporter which is a homotetrameric, membrane-embedded coiled-coil (Joh et al., 2014). Recent studies have also employed bifaceted coiled coils in designing soluble enzyme catalytic barrels. For example, a designed seven-helical barrel with a central Cys-His-Glu catalytic triad was capable of hydrolyzing p-nitrophenyl acetate with a significantly high catalytic rate (Burton et al., 2016). Finally, as components of cellular receptors, coiled coils can mediate signal transduction (Hulko et al., 2006; Ferris et al., 2011; Gushchin and Gordeliy, 2018). Transmission of signals across membrane, for e.g. in the HAMP domain of two-component signal transduction receptors, utilizes the mechanism of axial helix rotation. Transition between canonical knobs-into-holes and complementary *x-da* packing (two isoenergetic states) upon ligand binding, drives the long-range signal transmission. In short, coiled coils are omnipresent structural elements.

CHAPTER 2

Structural characterization of α/β coiled coils derived from hexad repeats

The present chapter deals with a special type of coiled-coil discontinuity, namely the β -layer. As already discussed in Chapter 1, canonical coiled coils follow a heptad repeat pattern. Removal of one or more residues from a regular heptad repeat results in perturbation in the packing of coiled coils. Many naturally occurring coiled coils harbor such discontinuities. One such type of a discontinuity is the insertion of two or deletion of one residue, giving rise to a nonad or hexad repeat respectively. In a previous report from our group (Hartmann et al., 2016), it was shown that such perturbations in regular coiled coils overstrain the single helices. The increased conformational strain locally breaks the α -helices into short β -strands forming a triangular network of interaction known as the β -layer. These β -layers are also found arranged repetitively in long protein fibers, and form α/β coiled coils with regularly alternating α - and β -segments. An example of a fiber built from repetitive nonads with only one backbone hydrogen bond per repeat in common with a heptad-repeat coiled coil has been studied crystallographically. But we are still missing the structural information of repetitive hexad insertions in coiled-coil fibers. Here, we undertake the structural characterization of two prominently identified hexad repeat families: (i) KAD-VYT-LYT co-optimized module, and (ii) TAT repetitive module. We present the first examples of a novel coiled-coil backbone with β -layers in hexad spacing, which do not share any backbone hydrogen bond with a heptad coiled coil. Conversion of hexads to heptads by simple insertion of one residue per repeat leads to the formation of a canonical coiled coil. This work expands the repertoire of available structures and will be useful for structural prediction and protein design efforts in the future.

2.1 INTRODUCTION

2.1.1 What is a β -layer?

Coiled coils are among the best understood and widely present structural features in proteins whose backbone structures can be easily computed from parametric equations. They contain two or more α -helices wound around in a superhelical bundle. Primarily they follow a heptad repeat pattern *hpphppp* (hydrophobic, h and polar, p). Positions on the heptad repeat are generally denoted as *abcdefg* where hydrophobic residues *a* and *d* are involved in the knobs-into-holes packing, forming the interaction core of this left-handed supercoiled bundle. Naturally occurring coiled coils frequently deviate from this standard model in periodicity, core packing and handedness (Lupas and Gruber, 2005). They can obtain discontinuities through insertions or deletions during the protein evolution process, resulting in varying degrees of perturbation in the packing of coiled-coil helices. As α -helices can tolerate a limited degree of supercoiling before they reach a breakpoint, only a range of periodicities is accessible while accommodating such insertions (Fig. 2.1).

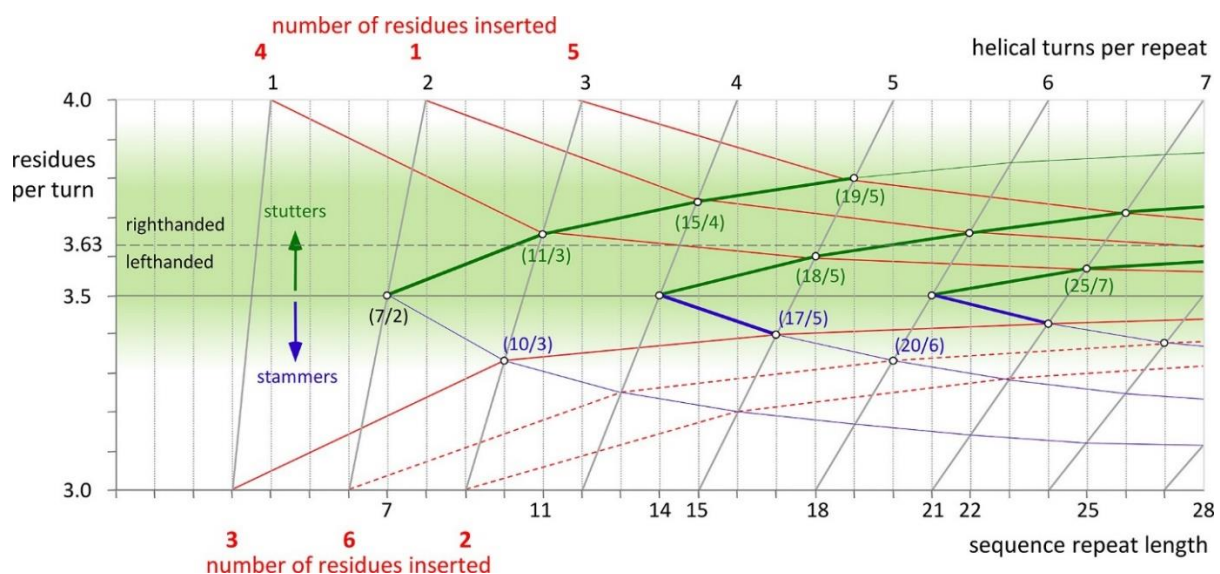


Figure 2.1: **Periodicity plot for the insertion of 1-6 residues in a canonical heptad repeat.** The figure has been reproduced from (Hartmann et al., 2016). Green area in the plot marks the estimated periodicities which can be accessed by α -helical coiled coils. The dotted line marks the periodicity of 3.63 residues per turn of unperturbed α -helix. Values higher than 3.63 result in right-handed supercoiling and lower values lead to left-handed supercoils. Consecutive insertions of 3 residues (stammers) and 4 residues (stutters) to a heptad are followed in blue and green lines respectively. The red lines follow the heptad periodicity and delocalization of 1 to 6 residues insertion over multiple heptad repeats. Intersection of the red lines from 1 and 5 residue insertion (delocalized over two heptads) with the green lines at 15/4 or 19/5 implies that these periodicities can also be brought about by consecutive insertion of stutters. Red dotted lines corresponding to the insertion of 2 and 6 residues requires a strong delocalization to reach the accessible periodicity green region.

As already described in detail in chapter 1, irregularities in a regular heptad coiled coil can be seen as insertion of one to six residues. The most commonly observed discontinuities are the insertion of 3 or 4 residues. Insertion of 4 residues, called a stutter, can be easily accommodated by slight underwinding of the helices. 11 residues are adjusted over 3 turns giving a periodicity of 3.67, which is very close to the periodicity of an undistorted α -helix (3.63). Therefore, hendecads are rather straight. Consecutive insertion of stutters increases the periodicity and thus the right-handedness of coiled coils. A pentadecad (periodicity $15/4 = 3.75$ residues/turn, $3.75 - 3.63 = 0.12$) brought about by insertion of two stutters to a heptad, shows a similar degree of right-handedness as the left-handedness of a heptad repeat coiled coil ($3.5 - 3.63 = -0.13$). Insertion of 3 residues, called stammer, results in overwinding of the left-handed supercoil and decrease in periodicity ($(7+3)/3 = 3.33$). From the solved crystal structures of stammer insertions in coiled coils, it could be seen that local overwinding was sufficient to strain the helices to form short 3_{10} -helical segments. Therefore, we assume that the value 3.33 defines the lower limit of the range of periodicities accessible to α -helices. From this, we can estimate that the upper limit would be around 3.9, i.e. 0.3 residues per turn more than the periodicity of a perfectly straight helix (Hartmann et al., 2016). To date, the most extreme example of a right-hand supercoil occurs in YadA, where 19 residues are accommodated over 5 turns leading to a local periodicity of 3.8 (Alvarez et al., 2010). Unlike stutters and stammers, insertion of 1 or 5 residues is more demanding on the coiled-coil bundle. They are found to be delocalized over two or more heptads, giving periodicities of 3.75 ($(7+7+1)/4$) and 3.8 ($(7+7+5)/5$). Insertion of 1 residue (skip residue) can also be alternatively accommodated by formation of a π -turn at the site of insertion without affecting the rest of coiled coil.

While the above mentioned discontinuities can be easily accommodated within the coiled-coil structure, insertion of 2 (nonad) or 6 (hexad) residues locally overstrain the helical geometry. The periodicities resulting from the insertion of 2 ($(7+2)/2 = 4.5$; $(7+2)/3 = 3.0$) and 6 ($6/2 = 3.0$; $(7+6)/3 = 4.33$; $(7+6)/4 = 3.25$) residues do not fall into the accessible range for α -helical coiled coils. The increased strain locally breaks the α -helices into short β -strands which associate to form a triangular structural element, named β -layer (Hartmann et al., 2016). The three β -strands rotate counter-clockwise by $\sim 120^\circ$ around the coiled-coil axis (as seen from the N-terminus) and continue as α -helices again, downstream of the neighboring helices. According to the coiled-coil nomenclature, positions on a nonad (7+2) can be described as *a-b-c- β 1- β 2- β 3-e-f-g* and for a hexad (7-1) as *β 1- β 2- β 3-e-f-g*. This implies that the three residues forming β -strands always arrange themselves in the core *d* position. In the next sections, we

will look at the previously solved crystal structures of single nonad and hexad discontinuities and the first α/β coiled-coil fiber comprising of repetitive nonad β -layers in more detail (Hartmann et al., 2016).

2.1.2 Single β -layer insertions

During previous work in our group on trimeric autotransporter adhesins (TAA), the fibrous surface proteins of gram-negative bacteria, OMP100, a putative TAA from *Actinobacillus actinomycetemcomitans*, was found to contain an unusual insertion of 2 residues within the heptad repeats of its coiled-coil stalk (This work was done by Prof. Andrei Lupas, Dr. Marcus Hartmann and Dr. Birte Hernandez) (Hartmann et al., 2016). Extending the heptad to a nonad, insert IENKADKAD was natively inserted between non-canonical heptad repeats containing $N@d$ motif (described in Sect. 1.4.3). The local overstrain generated by the insertion of two residues resulted in a sharp break at the three central residues KAD and a local departure from α -helical to β -strand character (Fig. 2.2). The three β -strands cross each other rotating in the counter-clockwise direction. We call this triangular plane formed by the three β -strands perpendicular to the central axis, the β -layer. The succeeding heptad coiled-coil continues downstream of the neighboring chain. The first three residues IEN occupy the a , b and c positions of the heptad repeat respectively. The central KAD residues, referred to as $\beta 1$, $\beta 2$ and $\beta 3$, take the d position leading into e , f and g positions for the last three residues. Within the coiled-coil, β -layer is stabilized by backbone interactions between the c residue of N-terminal segment and the e residue of the C-terminal neighboring chain (Fig. 2.2C). $\beta 2$ position is mostly occupied by a small hydrophobic residue, such as alanine or valine. In this case, the central alanine forms a backbone hydrogen bonding network with the corresponding alanines of the neighboring chains, stabilizing the β -layer (Fig. 2.2E).

Using GCN4 leucine zipper as a model, the authors also showed that β -layers can be simply generated by insertion of 2 or 6 residues in a heptad coiled-coil (Hartmann et al., 2016). Two different sequence motifs: IENKADKAD from *Actinobacillus* OMP100 and MATKDDIAN from *Thermus carboxydivorans* Tcar0761, along with their derived hexad and heptad variants (IENKAD and IENKKAD; MATKDD) were fused between GCN4-N16V stabilizing adaptors. The structures of the resulting constructs uncovered rather unexpected details. Unlike the central KAD of IENKADKAD, both MAT and KDD in MATKDDIAN could form β -layers. MATKDD formed a true hexad with MAT occupying the β -layer position

and KDD completing the *e*, *f* and *g* positions of the heptad. The preceding residue for this hexad, leucine (*g* residue of the previous heptad) was rotated by $\sim 15^\circ$ outwards from the core of the bundle towards the *c* position, facilitating β -layer formation at the *d* position. In contrast, IENKAD hexad borrowed three residues from the C-terminal GCN4 to complete a nonad with the central residues forming the β -layer and a stutter accommodation in the next heptad of GCN4. We find that β -layers strictly dictate the next residue to occupy *e* position of the heptad repeat irrespective of a nonad or hexad insertion.

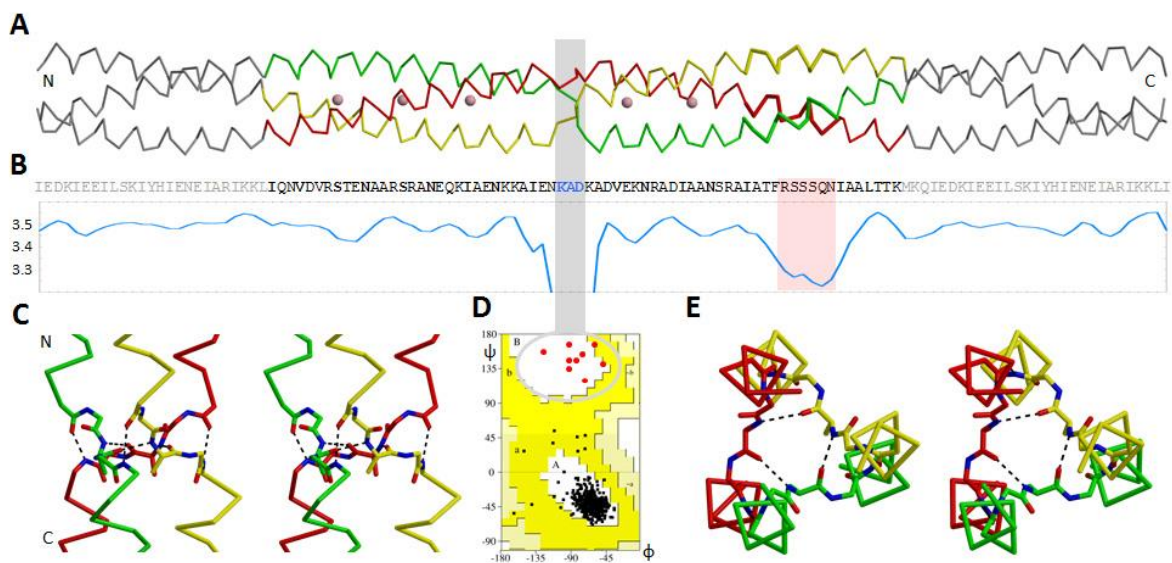


Figure 2.2: **Single nonad β -layer in *Actinobacillus* OMP100 stalk.** (A) Structure of the *Actinobacillus* OMP100 stalk aligned to its sequence and periodicity plot in (B). OMP100 trimer is colored by chain; GCN4 adaptors are in grey. Pink bar shows the area of a C-terminal stammer. The three β -layer residues highlighted by a grey bar fall in the β -region (red dots) of the Ramachandran plot (D). (C) Side and (E) top views showing the β -layer interaction network. (Reproduced from Hartmann et al., 2016)

Single β -layers are also found as connectors between the head and stalk domains of TAAs, which mediate smooth transition between larger diameter heads and smaller diameter stalks (Hartmann et al., 2012; Bassler et al., 2015). While β -layers are frequently identified capping either the N- or C-terminal end of coiled coils, they can also occur embedded within them. Systematic bioinformatic searches identified β -layers in various bacterial and phage fibrous proteins, some of them belonging to a family of prokaryotic endonucleases with PD-(D/E)XK motif (DUF3782), and a family of membrane proteins conserved in prokaryotes and mitochondria annotated as DUF1640 (see Chapter 3 for detailed characterization) etc. All identified and solved β -layer containing proteins form homotrimeric bundles, with the exception of SLH domain, a monomer with pseudo-threefold symmetry.

2.1.3 The α/β coiled coil from repetitive nonads

Repetitive arrangement of nonad insertions in coiled coils results in a fiber with alternating α - and β -elements. We refer to such fibers as α/β coiled coils. A protein from *Thermosinus carboxydivorans*, Tcar0761, contains 14 consecutive IANMATKDD repeats inserted between heptad segments. 4 nonad repeats including flanking heptads were cloned between GCN4-N16V adaptors and crystal structure was determined for the first α/β coiled coil containing β -layers in nonad spacing (Fig. 2.3). In the homotrimeric bundle, the central MAT residues of each nonad formed the alternating β -layers. We proposed that it should be possible to derive a minimalistic α/β coiled coil formed of hexads with three residues in the α -helical region and three in the β region of the Ramachandran plot.

2.1.4 β -layers as capping structures

β -Layers are frequently found capping the ends of coiled coils, stabilized by an extensive molecular interaction network. In the most common N-capping interaction network, the $\beta 3$ residue (mostly D, S, N or T) forms backbone hydrogen bond with the g residue (commonly D, E or Q) of the following α -helix, which coordinates back with its sidechain closing a ring of interaction. However, with conserved K as the $\beta 1$ residue as seen in the stalks of TAAs or phage proteins, a C-capping interaction network is formed with the preceding helix. β -Layer insertions offer the advantage of increasing the resilience of fibers by tightly interweaving the monomeric chains. They further increase the structural and functional complexity of fibrous proteins supported by the high frequency of β -layer occurrence in bacterial and phage surface fibrous proteins (Hartmann et al., 2016).

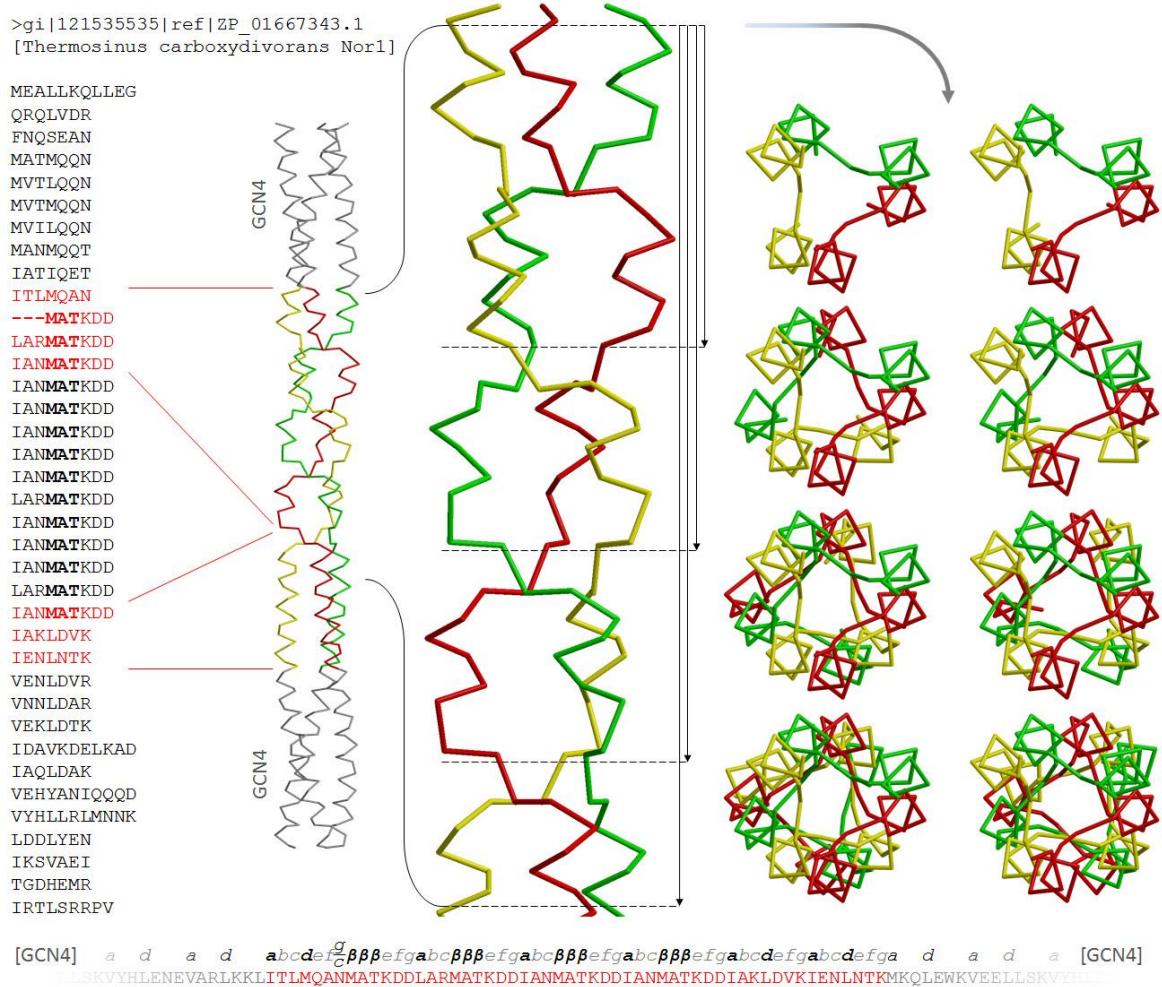


Figure 2.3: The α/β coiled coil from nonad repeats. **(Left)** The insert fused between GCN4 adaptors is shown in red on the full sequence of *Thermosinus* Tcar0761. **(Center)** Structure of the four consecutive β -layers is shown enlarged. **(Right)** Top view of β -layers as seen from the N-terminus. They show 1, 2, 3 or all 4 β -layers as indicated by the arrows on the enlarged side-view. (from Hartmann et al., 2016)

2.2 RESULTS

The primary objective of this study was to elucidate the structural characteristics of conserved tandem hexad insertions in coiled-coil proteins. Two of the most commonly found modules in bacterial and phage fibrous proteins are (i) $(x)_3KAD(x)_3VYT(x)_3LYT(x)_3$, a co-optimized module containing a small hydrophobic residue (mostly alanine) as the central residue in the first β -layer nonad and a large polar residue such as tyrosine in the successive hexads, (ii) $(x)_3TAT(x)_3TAT(x)_3$ repetitive β -layers. In the next sections, I discuss important features of these two modules based on protein structures solved using X-ray crystallography. The modules have been named by the sequence motif of central three residues forming the β -layer.

2.2.1 Structural characterization of the “KAD-VYT-LYT” hexad module

In previous searches for proteins containing repetitive hexad insertions, we identified a highly conserved motif with the consensus sequence – $LxxKADKxxVYTKxE$ – to occur in many bacterial ORFs and viral fibrous proteins, suggesting a co-optimized module (Hartmann et al., 2016). Based on these data, we extended our searches and identified more of such motifs inserted one or multiple times in the coiled-coil stalk of uncharacterized bacterial proteins. Some examples of these sequences containing $(x)_3KAD(x)_3VYT(x)_3LYT(x)_3$ insertion are presented in Fig 2.4A. The first β -layer, annotated as $a-b-c-\beta1-\beta2-\beta3-e-f-g$, is always a nonad and starts at position a . The $\beta1$ residue, primarily occupied by a conserved lysine residue (KAD/KAN), forms a C-capping structure. The central β -layer residue, $\beta2$, features a conserved small hydrophobic residue, such as alanine or valine. Successive β -layers occur in hexad spacing with the following register $\beta1-\beta2-\beta3-e-f-g$. The first hexad is preceded by the last residues ($e-f-g$) of the previous nonad repeat. We find that β -layers always occur at the d position and follow either a c or g position. In hexads, the residue preceding β -layer occupies the g position of the previous repeat. It moves slightly outwards from the core towards the c position, thereby facilitating the formation of β -layer in the d position and yielding an overall register $e-f-(g/c)-\beta1-\beta2-\beta3-e-f-g$. The $\beta1$ position in hexads is always occupied by a hydrophobic residue (V or L) and the central core residue by a bulky residue such as tyrosine. Only one previously solved protein, a tail-fiber protein with hyaluronidase activity from *Streptococcus pyogenes* prophage SF370.1 (PDB code 2C3F) (Smith et al., 2005), contains the $(x)_3KAD(x)_3VYT(x)_3$ conserved module. It contains four β -layers embedded in the coiled coil:

an N-terminal nonad, two central β -layers with the sequence LQQKADKETVYTKAE, and another one at the C-terminus.

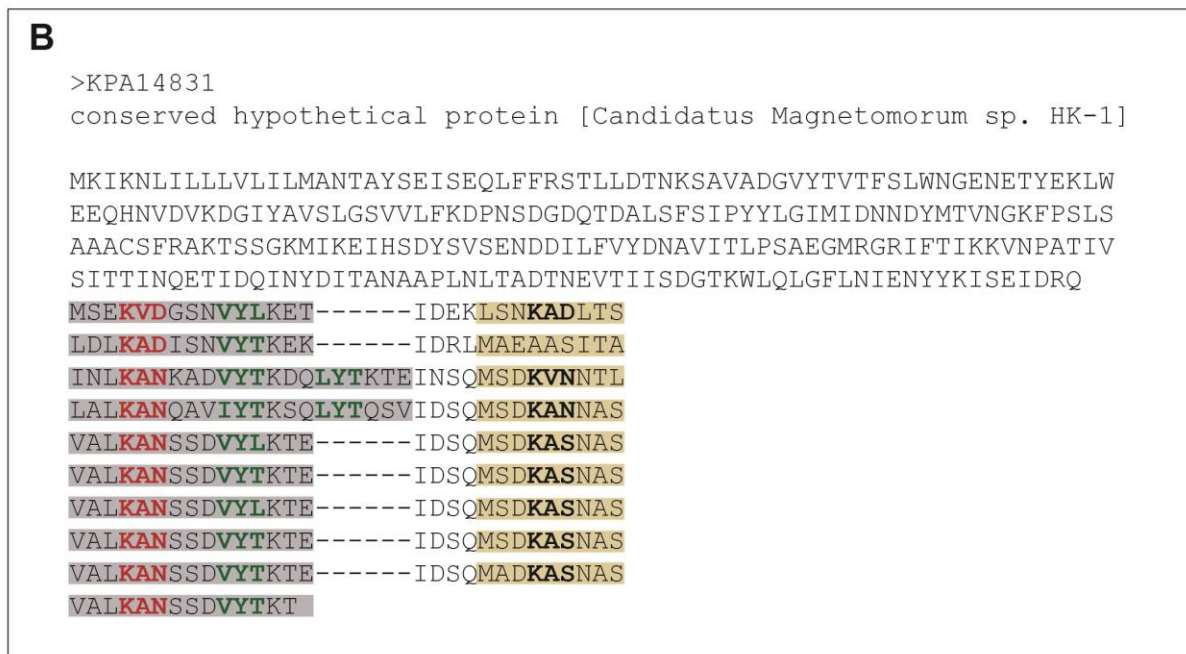
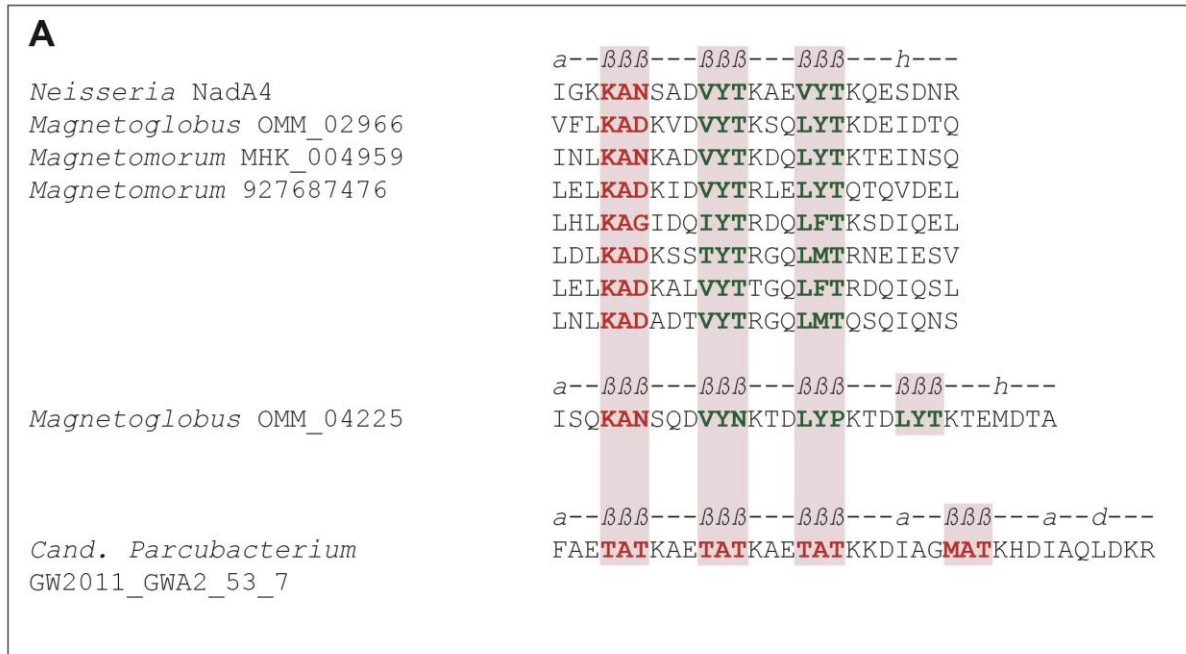


Figure 2.4: **(A)** Representative examples of identified sequences with KAD-VYT-LYT co-optimized module or a TAT repetitive module with assigned coiled-coil registers. **(B)** Annotated sequence of *Magnetomorum* sp. HK-1 MHK_004959.

For structural characterization of repetitive β -layers in hexad spacing, we selected two sequences (highlighted in pink in Fig. 2.4A) from the hypothetical proteins MHK_004959 from *Candidatus Magnetomorum* sp. HK-1 (GenBank KPA14831.1) and OMM_04225 from *Candidatus Magnetoglobus multicellularis* str. *Araruama* (GenBank ETR68995.1). In the C-terminal coiled-coil stalk of MHK_004959 protein, multiple KAD-VYT modules are inserted in a recurring fashion (Fig. 2.4B). Similar regular insertions can be observed in the N-terminal coiled-coil sequence of OMM_04225. Single modules from both proteins comprising three and four β -layers from *Magnetomorum* (referred further as Mmor) and *Magnetoglobus* (referred as Mglob) respectively were fused in between trimeric GCN4-pII adaptors for structural characterization. We included a short linker sequence GGSG in the designed construct in between the TEV protease site and the GCN4 adaptor, to ease the cleavage of N-terminal histidine tag. Small amount of proteins purified from the supernatant from overexpressing *E. coli* ArcticExpress cells, were sufficient to obtain well-diffracting crystals for Mmor and Mglob.

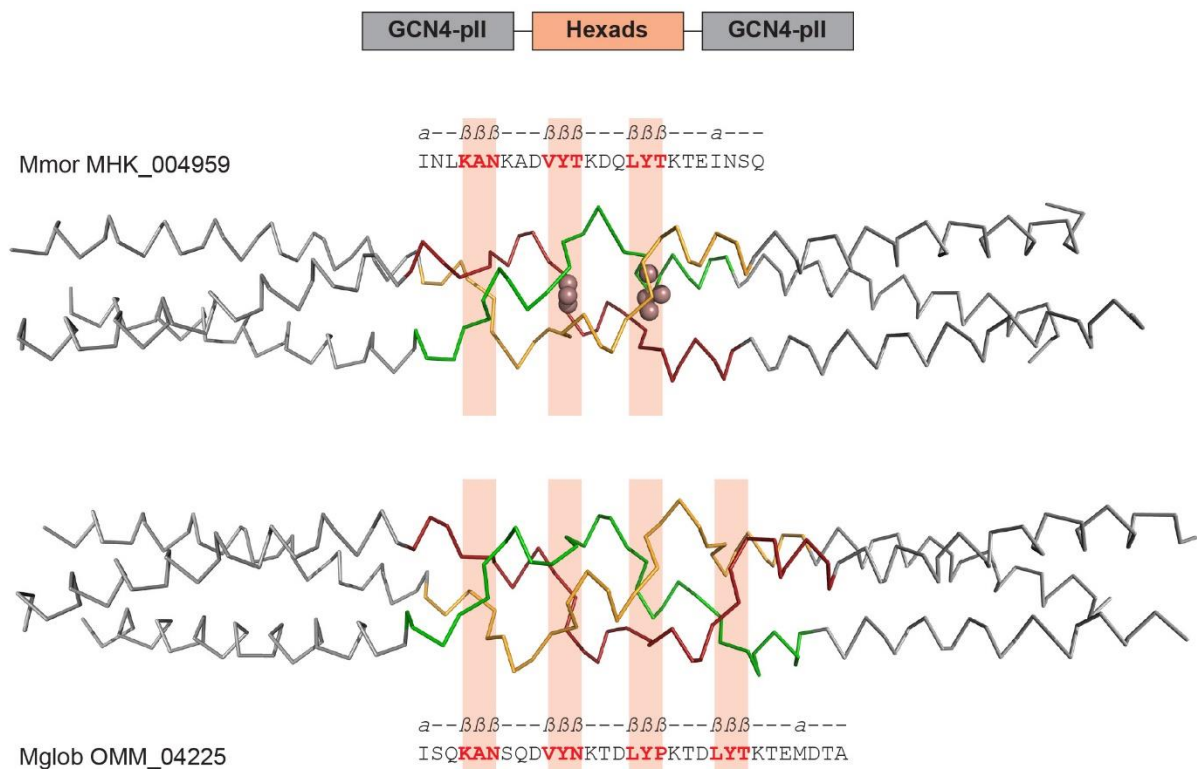


Figure 2.5: **Crystal structures for KAD-VYT-LYT family hexad motifs.** Schematic diagram for the designed constructs is shown at the top. Hexad β -layer module was fused between N- and C-terminal GCN4-pII trimeric coiled coils. Solved crystal structures for Mmor (MHK_004959) and Mglob (OMM_04225) are shown as ribbon representation. β -Layers are highlighted in pink boxes with the corresponding sequences. Water molecules are shown as lightpurple spheres.

Crystal structure of Mmor, solved at a resolution of 1.5Å using GCN4-pII as a molecular replacement model, shows a parallel trimer with three β -layers formed between the flanking coiled coils (Fig. 2.5). As expected, residues KAN of the nonad sequence INLKANKAD form the first, canonical β -layer. A33, the central β -layer residue, forms characteristic backbone hydrogen bonding network with the corresponding alanines of neighboring chains (Fig. 2.6A).

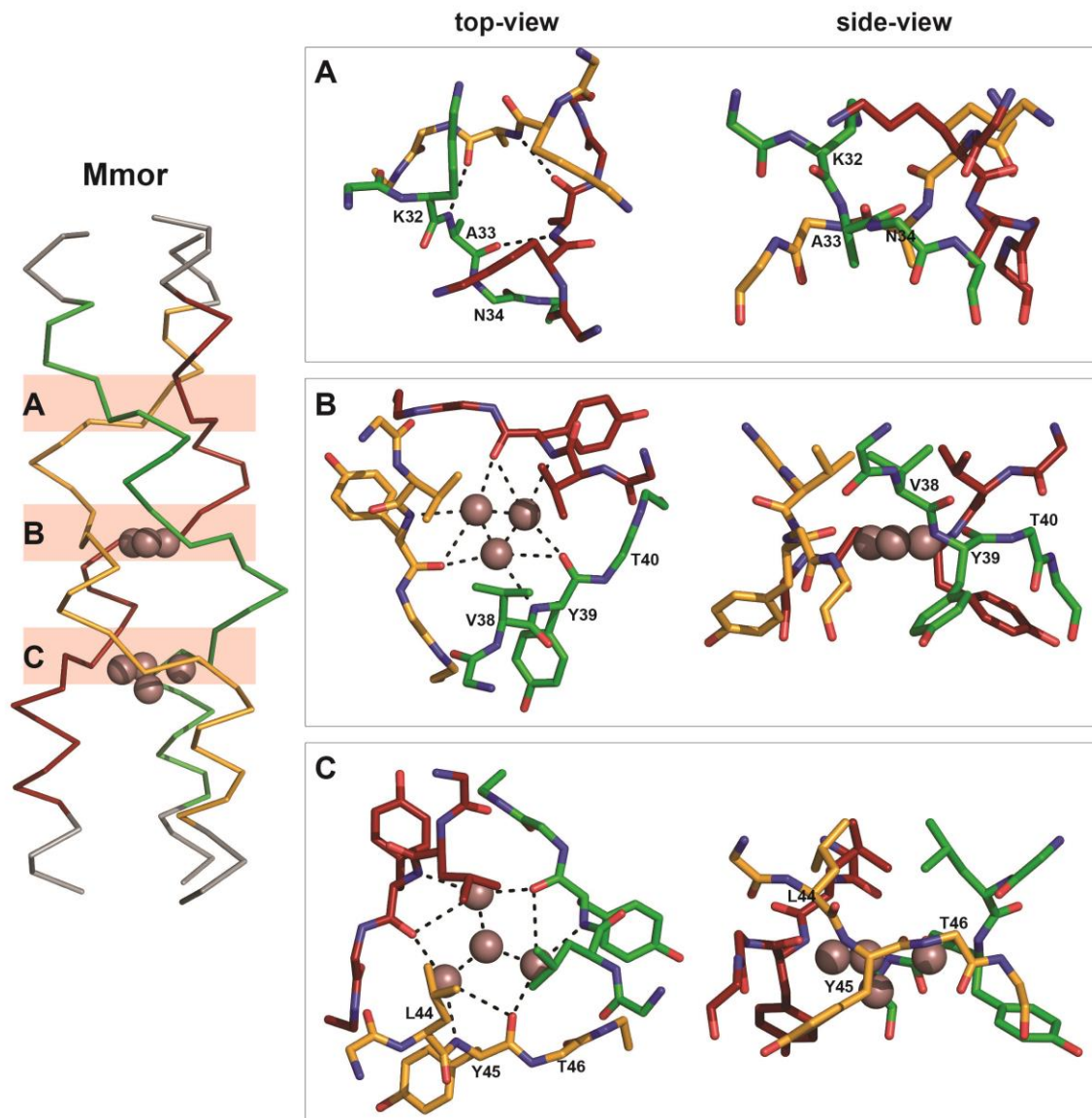


Figure 2.6: **Crystal structure of Mmor β -layers.** (Left) Structure of KAN-VYT-LYT β -layer insert of Mmor. Red, yellow and green are the three monomeric chains. Flanking grey residues are a part of GCN4 coiled-coil adaptor. The three β -layers are highlighted in pink and marked by letters A-C (from N-terminus). (Right) Top and side views of the three β -layers (A-C). Water molecules are shown as lightpurple spheres. Hydrogen bonds are shown as dotted lines.

The second and third β -layers of Mmor, arranged in a hexad spacing, are formed by the VYT and LYT residues respectively. These β -layers deviate from the canonical β -layer structure, with a largely increased diameter spanned by a water network in the core of trimer. In the second β -layer, backbone amide and carbonyl moieties of the central tyrosine residues (Y39) from the three chains coordinate three ordered water molecules in the center (Fig. 2.6B). The third β -layer, slightly wider than the second, is able to accommodate a tetrahedral water network (Fig. 2.6C). In contrast to the canonical β -layer, the central residue tyrosine of both β -layers formed from hexads, is bent out of the core, pushing the $\beta 1$ residue (V or L) into the core, thereby promoting the increased diameter of the second and third β -layers.

The conserved $\beta 1$ residue K32, in the first β -layer, is involved in forming a C-capping network with the preceding helix (Fig. 2.7A). Such an interaction network is only favored with lysines occupying the $\beta 1$ position. In the Mmor structure, K32' reaches across the trimeric core and coordinates the carbonyls of $\alpha 1$ (or a ; I29) and $\beta 1$ (K32) residues, and the sidechain of $\beta 3$ residue (N34), all from the neighboring chain (clockwise as seen from the N-terminus). K32 amide also forms backbone hydrogen bond with I29 from the same chain. C-capping network is finally completed by the $\alpha 3$ residue (or c) L31 forming backbone hydrogen bond with e residue K35'' of the neighboring chain (clockwise). The $\beta 3$ residue N34, then forms an N-cap for the next β -layer, coordinating the backbone and sidechain of g residue D37.

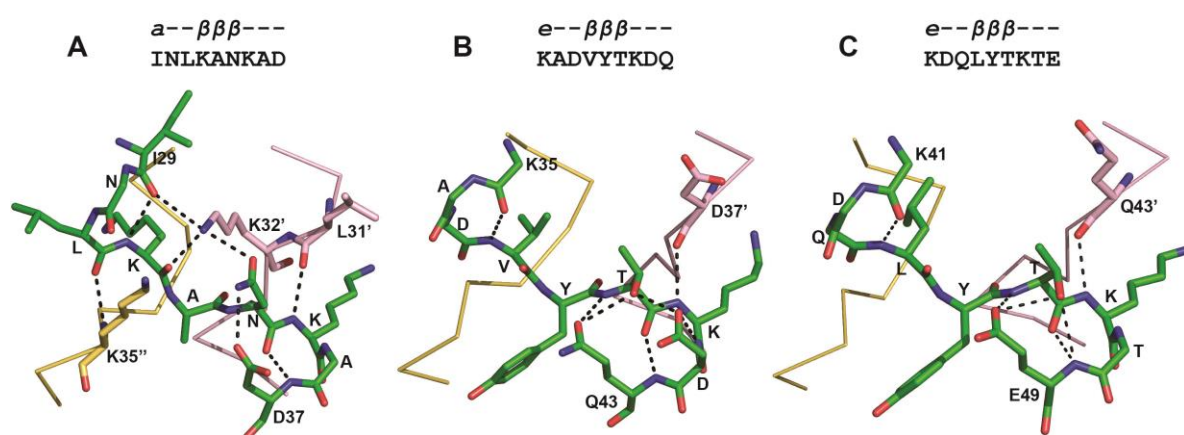


Figure 2.7: **Interaction network of the KAN-VYT-VYT β -layer module of Mmor.** (A) C-capping network formed by the first β -layer KAN lysine residue. (B) and (C) N-capping network formed by threonines of VYT (second β -layer) and LYT (third β -layer) respectively.

Looking at the interaction network established by tyrosine β -layers (Fig. 2.7B and 2.7C), two definite interactions are found: (i) hydrophobic residue in $\beta 1$ position (V or L) forms backbone

hydrogen bond with *e* residue K of the previous layer (lysines found to be conserved), and (ii) $\beta 3$ residue forms N-capping interaction network with the *g* residue. The main chain carbonyl of $\beta 3$ coordinates the amide of *g* residue, which in turn forms hydrogen bonding with $\beta 3$ amide through its sidechain, thereby completing a closed ring of interaction network. Additional bonds at the N-capping site of tyrosine β -layers are formed by the sidechain of $\beta 3$ residue, threonine in both instances.

To visualize the significance of conserved residues at the $\beta 1$ and $\beta 2$ positions of tyrosine β -layers, we designed two Mmor mutants where we replaced either the $\beta 1$ (V and L) or the $\beta 2$ (Y) residues for both hexad β -layers. Crystals for Mmor-mut1 diffracted well to a resolution of 1.9Å. In the electron density map, while GCN4 chains could be fitted with confidence, no peaks were observed for the three β -layers highlighting the flexibility or disorder of these residues and suggesting that substitution of both tyrosines in the second and the third layers with alanines completely abolished β -layer formation. This confirms the significance of conserved tyrosines as the central residue in β -layers with hexad spacing. In contrast, for Mmor-mut2, with mutations of $\beta 1$ residues V38 and L44 to alanines, β -layers identical to the native structure were formed (Fig. 2.8). This shows that both tyrosine β -layers can tolerate a smaller hydrophobic residue such as alanine at the $\beta 1$ position without significant effects to the backbone geometry. Tyrosine residues in these layers are bent out of the core identical to those of native structure.

The crystal structure of Mglob, featuring a canonical nonad β -layer formed by residues KAN and followed by three tyrosine β -layers, displays an identical parallel trimer with properly established β -layers between GCN4 coiled-coil adaptors. The interaction network formed by these layers is similar to that described for Mmor (Fig. 2.9). Due to poor diffraction quality of the crystals, the water network could not be solved for Mglob structure. In summary, our crystal structures of Mmor and Mglob present the first examples of a minimalistic α/β coiled coil built of tandem hexad repeats, with three residues present in the α region and three in the β region of the Ramachandran plot. The first nonad-based β -layer with conserved KAx motif favors a C-capping interaction network and the successive layers in hexad spacing favor an N-capping network.

e--ββ---ββ---

Mmor wt KADVYTKDQLYTKTE

Mmor mut2 KADAYTKDQAYTKTE

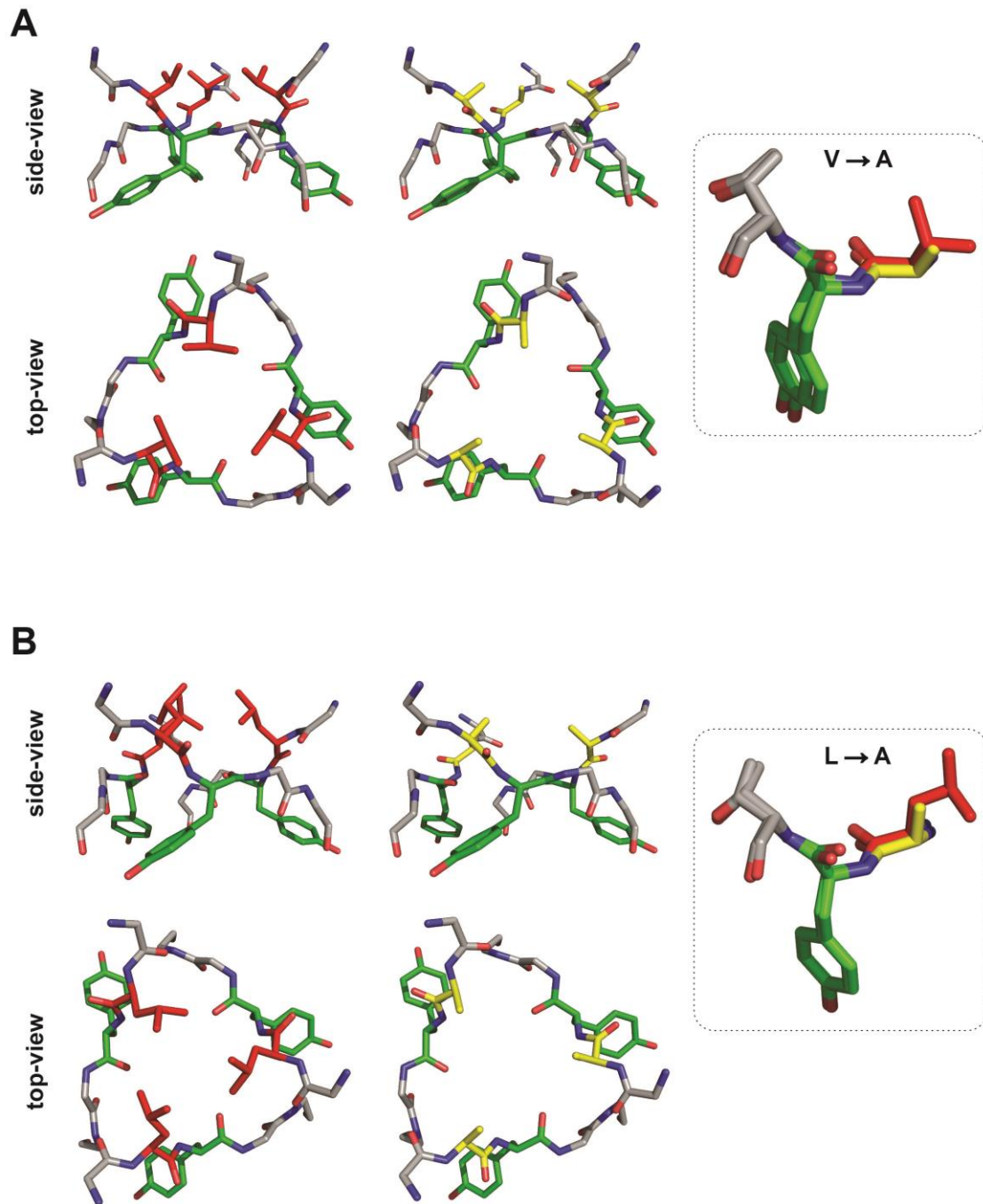


Figure 2.8: **Mutation of conserved hexad β-layer residues.** Structural comparison of the second and third hexad β-layers of Mmor-wt and Mmor-mut2 constructs. In Mmor-mut2, both β1 residues Val (in A) and Leu (in B) have been replaced with Ala.

Mglob

a---*βββ*---*βββ*---*βββ*---*βββ*---*a*---
 ISQ**KAN**SQD**VYN**KTD**LYP**KTD**LYT**KTEMDTA

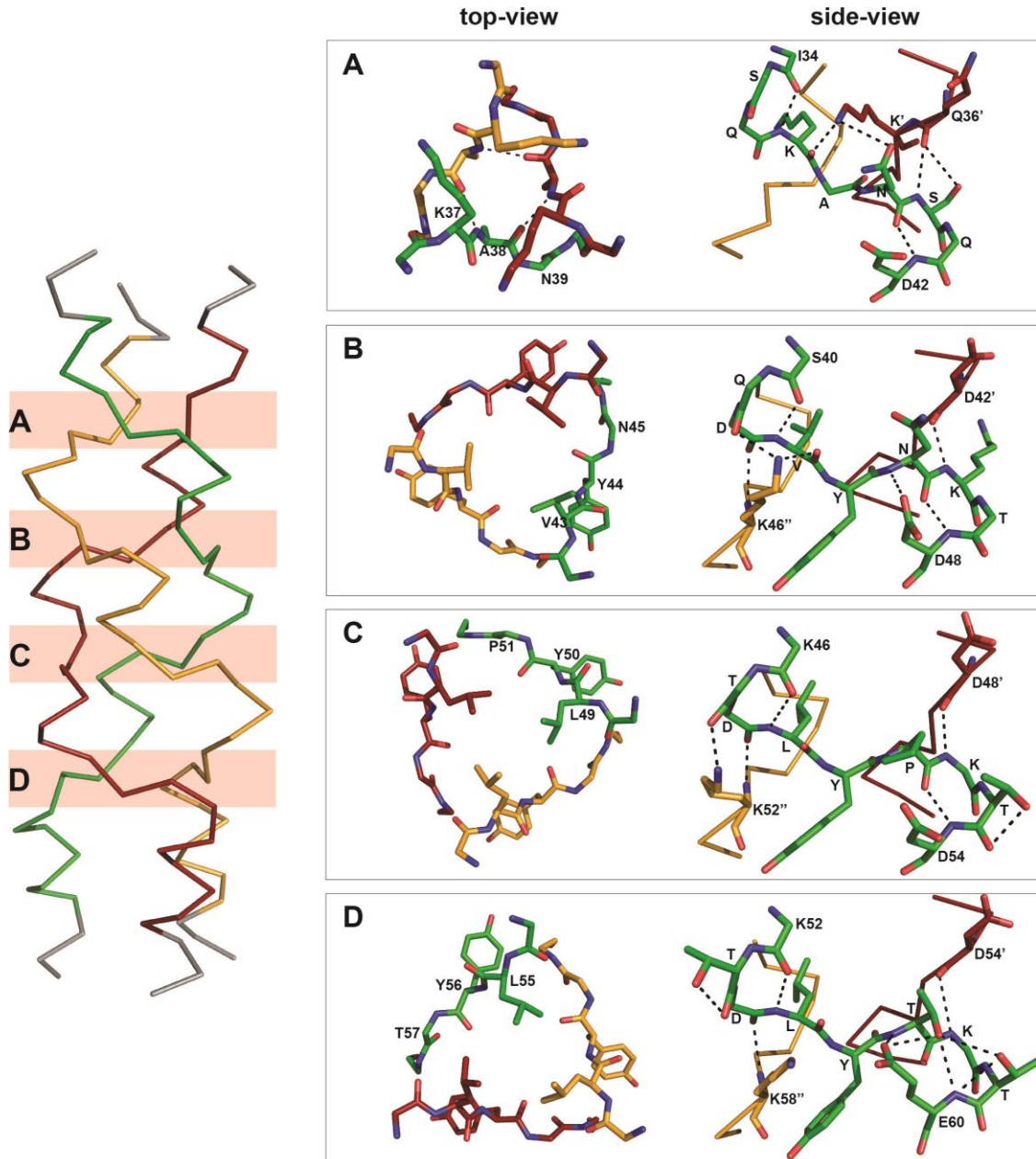


Figure 2.9: **Crystal structure of Mglob hexad β-layers** module fused between N- and C-terminal trimeric GCN4-pII coiled-coil adaptors is shown on the left. Sequence of the insert is shown at the top with its coiled-coil register. The four β-layers (starting from the N-terminus) are marked by letters A-D, A is the canonical KAN β-layer and B to D are tyrosine β-layers. The right box diagrams display the top view of individual β-layers and the side-view of the stabilizing interaction network formed by each layer.

The trimeric autotransporter adhesin NadA5 from *Neisseria meningitidis*

Another interesting example comprising the above described co-optimized hexad module (x)₃KAD(x)₃VYT(x)₃VYT(x)₃ is a trimeric autotransporter adhesin NadA. TAAs are a family of extracellular proteins in gram-negative bacteria that mediate attachment to and invasion of host cells (Cotter et al., 2005; Linke et al., 2006; Łyskowski et al., 2011). TAAs are obligate homotrimers. Their general domain architecture includes a C-terminal β -barrel integral membrane domain which anchors the proteins to outer membrane, an N-terminal signal peptide for export and a “passenger” domain. The passenger domain includes at least one head which is primarily responsible for host cell adhesion and a long coiled-coil stalk, variable in length, which acts as molecular spacer projecting the adhesive heads away from the bacterial surface. Multiple head and stalk domains can be repetitively arranged in different combinations generating the molecular diversity necessary to evade host cellular defenses (Reiss et al., 2004; Szczesny and Lupas, 2008). The head-stalk transition is mediated by a short, highly conserved neck, comprising a β -layer which functions as the adaptor connecting the larger diameter transversal heads to the smaller diameter coiled-coil stalks. At the C-terminus, four TM β -strands from the three monomeric chains assemble into a 12-stranded β -barrel (Meng et al., 2006). The membrane anchor is conserved in all TAAs, therefore it is considered as the family-defining element (Hoiczuk et al., 2000). During the autotransport activity, after secretion of TAAs into the periplasmic space, initially the β -barrel assembles into the bacterial outer membrane forming the pore through which the three unstructured chains exit into the extracellular region (Szabady et al., 2005). After the export is complete, monomers properly fold into a long fiber with a short coiled-coil segment at the end of passenger domain occluding the pore domain.

NadA5 is expressed in *Neisseria meningitidis*, a gram-negative encapsulated bacterium which causes severe sepsis and meningococcal meningitis. NadA5 (PDB code 4CJD) forms an elongated 320 Å long helical trimer (Malito et al., 2014). It features a unique coiled-coil head, interrupted by an insertion sequence (residues N49–G84) which assembles into protruding wing-like structures. The head coiled coil directly continues into the stalk. Whereas high-resolution electron density was observed for the head and the N-terminal region of the stalk, the latter half of coiled-coil stalk which includes four β -layers - one independent nonad (residues I138-E146) flanked by heptad repeats and the other being the KAD-VYT-VYT optimized module (residues I161-E181), only discontinuous blobs with a three-fold symmetry

feature were observed. It was not possible to solve β -layers with the available data. Although one could clearly observe regions of helices, the peaks were broad and overlapping among the trimeric chains. This observation suggests a partially-unfolded, highly flexible region in the C-terminal half of the coiled-coil stalk. As TAAs are modular in structure, we can build new models by sequence homology to existing solved domains. To visualize the complete protein structure, I have reconstructed an *in silico* model of full-length NadA5 using the program MODELLER (Sali et al., 1995) (Fig. 2.10).

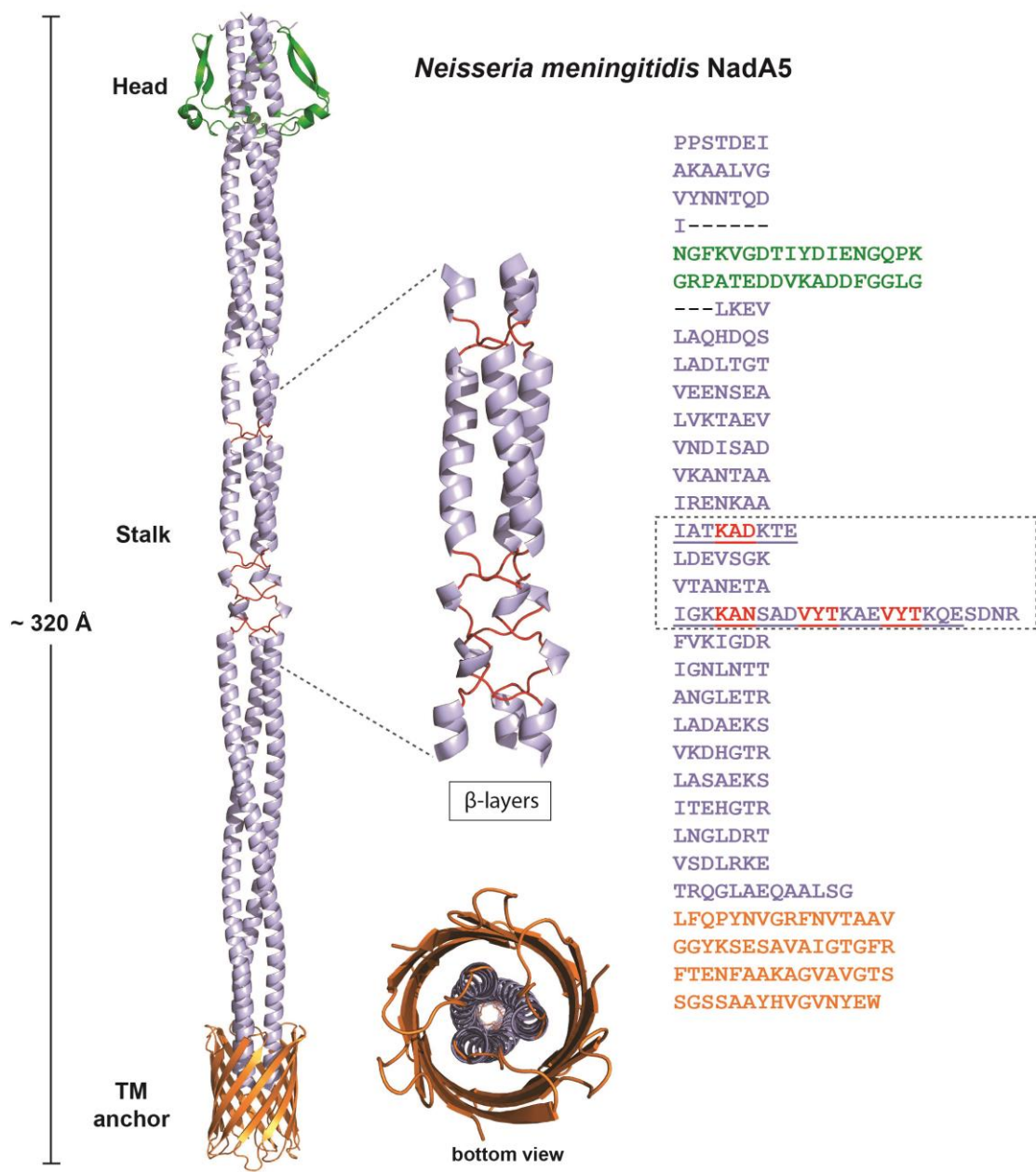


Figure 2.10: **Model of NadA5.** Full length model of trimeric autotransporter adhesin NadA5 from *Neisseria meningitidis* is shown, with the side and bottom views, and its sequence. The segment containing β -layers is enlarged.

2.2.2 Structural characterization of the “TAT” repetitive hexads module

The second, not so frequently observed, repetitive hexad module is built of “TATKxx” residues. Previous attempts to obtain constructs by fusing similar MATKDD repeats between GCN4-N16V adaptors failed, as the recombinantly expressed protein turned out to be insoluble and not amenable to refolding (Hartmann et al., 2016). Screening for new proteins containing tandem hexads, we prominently found the TATK motif instead of MATK in coiled coils which led us to believe that the former could be more easily accommodated within β -layers in hexad spacing. As there are no structures of coiled-coil proteins with repetitive TATK insertions in the PDB, we selected a hypothetical protein UY81 from *Candidatus Parcubacteria* (*Giovannonibacteria*) (GenBank ID: KKW34673.1) for structural characterization. This protein contains four β -layers - two hexads (**TATKAE** and **TATKGD**) sandwiched between two nonads (**FAETATKAE** and **IAGMATKGD**) (Fig. 2.11A). These β -layers are flanked by canonical coiled-coil sequences.

Full-length ParcuUY81 (ParcuUY81-fl) fused to an N-terminal 6xHis tag was recombinantly expressed in *E. coli* C41 (DE3) and purified as a soluble protein. Far-UV CD spectra showed that the protein adopts a primarily α -helical structure with characteristic minima at 208 nm and 222 nm (Fig. 2.11C). Upon thermal melting, native ParcuUY81-fl unfolds in two-steps with inflection points at 57.5°C and 81.5°C. From analytical SEC, the calculated molecular mass corresponded to a trimer in solution. ParcuUY81-fl crystallized under multiple buffer screening conditions. However, the crystals, even after grid optimization, diffracted at best to a resolution of 3.6 Å.

Subsequently, ParcuUY81-fl was subjected to limited proteolysis using trypsin (1:20). After one hour incubation at room temperature, cleaved fragments were separated by loading on a Superdex 75 gel filtration column. A stable fragment, ~ 1-2 kDa smaller in size than ParcuUY81-fl, was obtained (Fig. 2.11B). CD spectroscopy and thermal melting experiments on trypsin-cleaved fragment showed an identical spectra and two-state unfolding as the full-length protein (Fig. 2.11D). Mass-spectrometric analysis confirmed that only 4-5 flexible residues at both N- and C-terminal ends were cleaved. Crystals obtained from trypsin-cleaved ParcuUY81-fl, containing residues 3-88 of native construct, diffracted to a higher resolution (~ 2.5 Å).

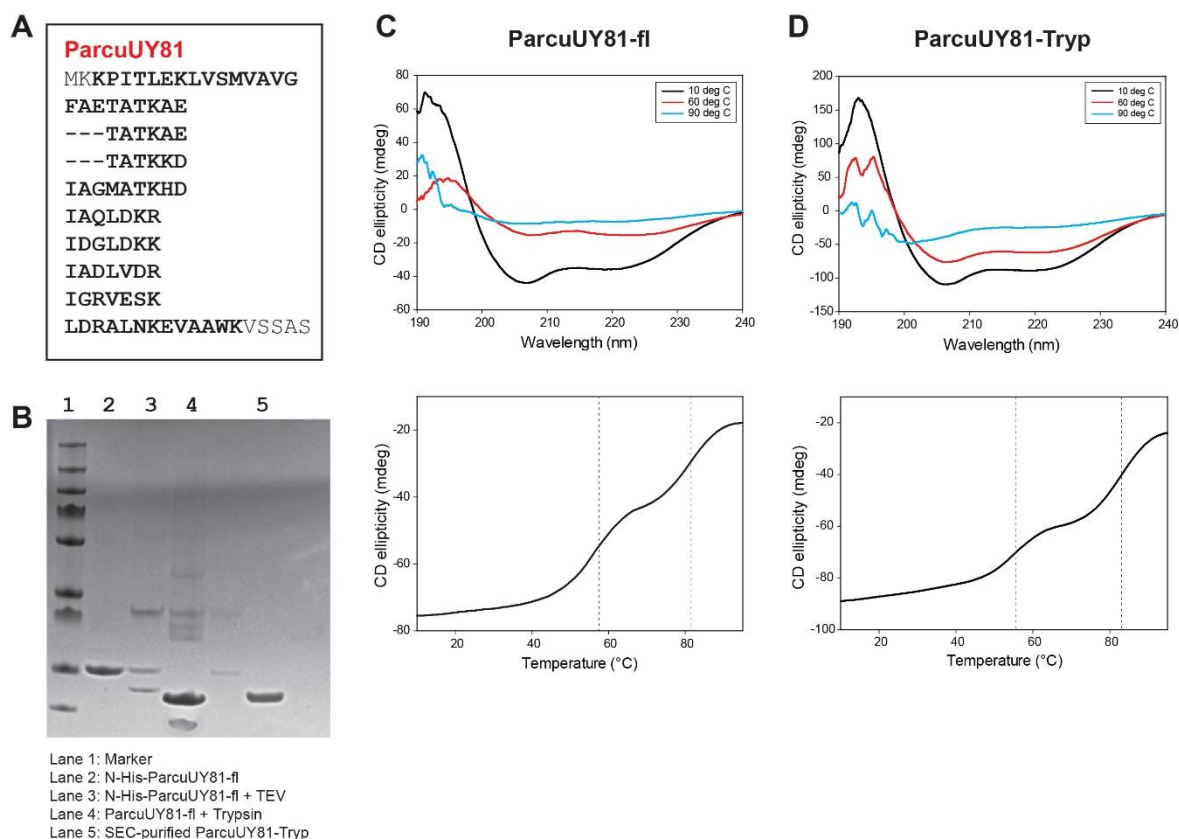


Figure 2.11 **Biophysical characterization of ParcuUY81.** (A) Sequence of ParcuUY81-fl. Residues in bold are the residues remaining after tryptic digestion. (B) ParcuUY81-fl and tryptic product run on an SDS-PAGE gel. Samples corresponding to marked lanes have been shown below. (C) and (D) CD and melting curves for ParcuUY81-fl and ParcuUY81-tryptic respectively. The inflection temperatures in the melting curve are marked with dotted lines.

Initial attempts to solve the structure of ParcuUY81 using trimeric GCN4 as a molecular replacement model for native coiled coil failed. Heavy-atom soaking trials primarily resulted in the disruption of the crystal lattice, with the exception of cadmium and uranium salts. Initial observations from the self-rotation Patterson function hinted towards a tetrameric structure, which was contradictory to the expected trimer. All previously identified β -layer containing coiled coils assembled into trimeric bundles with the single β -strand of each monomer in a β -layer rotating 120° around the coiled-coil axis before continuing into α -helices again. Such an arrangement of residues does not fit the tetrameric prediction. Next, we designed a control protein ParcuUY81-heptad, where all β -layer forming nonads and hexads were replaced by canonical heptad pattern either by deletion or insertion of residues (see Table 2.2). As expected, ParcuUY81-heptad formed a trimeric bundle in both solution and crystalline state (Fig. 2.12).

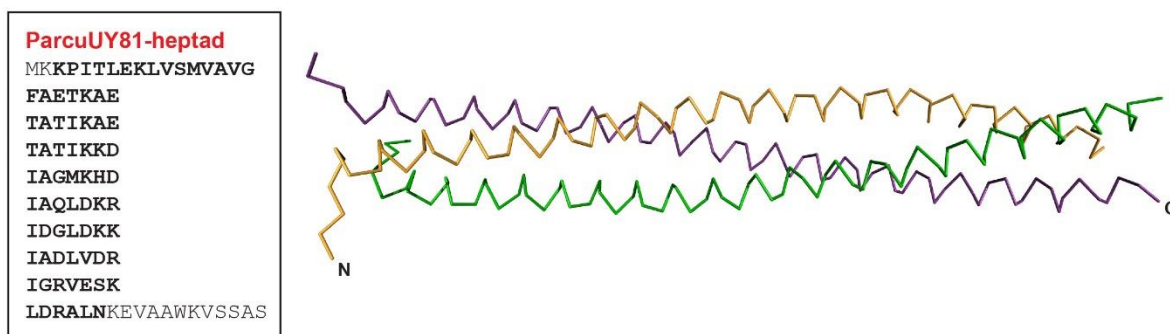


Figure 2.12: **Sequence and structure of ParcuUY81-heptad**, a designed control protein. Conversion of hexad repeats to heptads results in a canonical coiled coil. Bold letters indicate the sequence of the solved structure.

Assuming the tetrameric configuration to be a crystallization artefact, we shortened the C-terminal region by inserting a premature stop codon in the coiled-coil stalk to obtain ParcuUY81₁₋₇₀ (residues 1-70). The construct could be expressed as a soluble protein but similar to the previous constructs, the crystal structure of ParcuUY81₁₋₇₀ (data collected at a resolution of $\sim 2.4\text{\AA}$) could not be solved by molecular replacement. Platinum-salts dissolved or fragmented the fragile ParcuUY81₁₋₇₀ crystals, and the stable crystal fragments diffracted poorly as well. Facing the difficulties to solve the native ParcuUY81 crystal structure, I designed an *in silico* model using MODELLER (Sali et al., 1995) using sequence homology to coiled-coil fragments in the sequence of ParcuUY81-heptad and to previously solved crystal structures containing nonad and hexad insertions in coiled-coil sequences. Molecular replacement trials with this model failed to replace a proper length of the protein. From these results, three reasons appear plausible: the crystal was formed by a proteolytically cleaved fragment, or the crystal lattice space group identified was incorrect, or ParcuUY81 is actually a tetramer and replacement by a trimeric model could not give the right structure solution.

Crystal structure of ParcuUY81 with GCN4-N16V fusion

Given the inability to solve the crystal structure of TATK β -layers in their native coiled-coil environment, we fused a segment (residues 18-54) of ParcuUY81-fl comprising the four predicted β -layers including their neighboring heptads in register between two GCN4-N16V adaptors, a variant of leucine zipper which forms a mixture of dimers and trimers in solution (Harbury et al., 1993). ParcuUY81-GCN4N16V was found to be well-expressed in the insoluble fraction. After purification under denaturing conditions, the protein was refolded and set-up for crystallization at 5 and 10 mg/ml. Crystals obtained under the condition 0.1 M sodium acetate (pH 4.6), 2 M sodium formate, diffracted to a maximum resolution of 2.75\AA .

The crystal structure of ParcuUY81-GCN4, solved using GCN4-N16V as a molecular replacement model, shows a parallel trimeric bundle with β -layers formed at the junction of two coiled-coil adaptors (Fig. 2.13). All four β -layers appear canonical, with the central alanine residues forming characteristic inter-chain backbone hydrogen-bonding network. The N-terminal GCN4 coiled coil directly proceeds into the first β -layer formed by a nonad sequence FAETATKAE, where TAT occupies the *d* position.

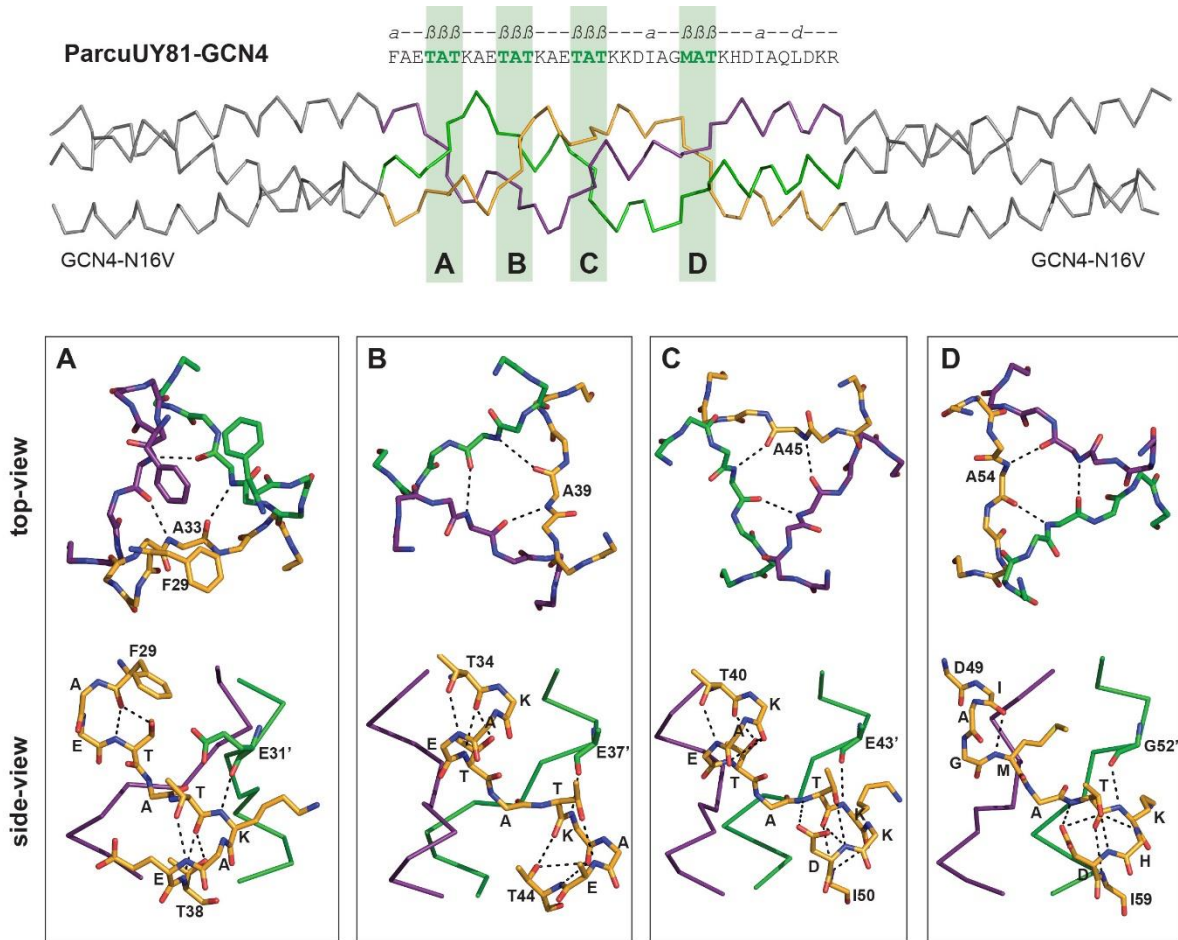


Figure 2.13: **Crystal structure and interaction network of ParcuUY81-GCN4.** (Top) Structure of ParcuUY81 tandem β -layers (colored) fused within GCN4-N16V adaptors (grey). Sequence of insert is shown above. The four β -layers (A-D) are highlighted with green bars. (Below) Top and side-views of individual β -layers (A-D). While top view displays the exact picture of each layer and relative rotation of three chains from A-D as viewed from the N-terminus, bottom side-view shows the interaction network formed by β -layer residues within the same (yellow) chain. Hydrogen bonds and salt-bridges are shown as dotted lines.

As visualized from the crystal structure, phenylalanine (F29) residues from the three monomers form an asymmetric hydrophobic core (Fig. 2.13A). While two F29 residues point away from the core, the third residue directly points towards the center. Threonine (T32) in the $\beta 1$ position forms polar interactions with the *a* residue F29 through its backbone NH and sidechain

hydroxyl groups. The typical C-capping interaction, as is seen for K in $\beta 1$ of the KAD-VYT-LYT module, is not formed with T in this position. The $\beta 3$ residue T34 coordinates the backbone NH group of residue g (E37), thereby forming a stable N-cap for the following helix. Both E31 and E37 extend to form salt-bridge interactions with the conserved lysine residues (K35 and K41 respectively) of the neighboring chain (clockwise as seen from the N-terminus).

For the successive hexad β -layer TATKAE, E37 occupying the g position of the previous layer slightly moves out of the core towards c position (g/c position), facilitating TAT (residues 38-40) to occupy the regular d position (Fig. 2.13B). T38, in an identical fashion to the first β -layer T32, coordinates the backbone carbonyl of $\beta 3$ residue T34 with its NH and sidechain hydroxyl. Such an interaction network is repeated within the successive β -layers, wherein the first threonine ($\beta 1$ residue) coordinates either the a position residue (in nonads) or the $\beta 3$ residue of the previous β -layer (in hexads) (Fig. 2.13C-D). $\beta 3$ residue in all β -layers, in this particular case always T, forms N-capping interactions with the g residue (D or E), which further stabilizes the β -layer by forming salt-bridge interactions with the e residues (usually a conserved K or R) of the neighboring chain.

NMR-spectroscopy for ParcuUY81

It is well-known that coiled coils can attain non-physiological structures, which are separated by low-energy barriers on a flat energy landscape, when crystallized out of their native context (Lupas et al., 2017). Therefore, I was keen to solve the structure of ParcuUY81 TATK module flanked by its native coiled-coil sequence to confirm if the protein exists in a trimeric or tetrameric state in its physiological context. Ultimately, I resorted to solution NMR spectroscopy. Preliminary experiments collecting the 1D ^1H spectrum showed good peak dispersion, consistent with a properly folded protein. Subsequently, ^{13}C and ^{15}N isotope-labelled ParcuUY81-fl was used to acquire spectra for sequential assignment. ParcuUY81-fl purified from minimal media was largely present as soluble aggregates, with a small fraction eluting in the right molecular weight range. Although ^{15}N -HSQC looked an ideal spectra with sharp signals, well-dispersed suggestive of well-folded oligomer, residues could not be definitely assigned. In future, other representative examples of the “TAT” conserved module should be tested to characterize such repetitive hexad insertions in coiled-coil proteins in their native environment.

2.3 CONTRIBUTIONS TO THIS WORK

All bioinformatics work was done by Prof. Andrei Lupas. The sequences used in this work for experimental study were identified or designed by Prof. Andrei Lupas. Dr. Birte Hernandez conceptualized the construct design within GCN4 coiled-coil adaptors. I carried out the molecular cloning, protein expression and purification for constructs expressed in LB or minimal media. Biophysical characterization of proteins using CD spectroscopy, thermal melting and limited proteolysis experiments were done by me. Crystallization screens were set-up by Dr. Reinhard Albrecht and Kerstin Bär. Crystal structures were XDS processed and phased using molecular replacement by Dr. Marcus Hartmann. I then manually refined the crystal structures using COOT. For ParcuUY81-GCN4, I manually built the four central β -layers and refined the crystal structure. NMR isotope-labelled sample was prepared by me and spectra measurements were done by Dr. Murray Coles. Finally, I designed the homology model of full-length NadA5 using previously solved crystal structure of its head domain (PDB 4CJD) using the program MODELLER.

2.4 DISCUSSION

Canonical coiled-coil fibers built of heptad repeats are frequently interrupted by the insertion or deletion of one or more residues. These discontinuities can be accommodated by local distortion of α -helices, assuming a range of periodicities which is however limited by the strain imposed on the individual helices. Insertion of 2 and 6 residues are the most demanding for regular coiled-coil bundles. They overstrain the single helices to a break point, locally disrupting the α -helical structure and leading to the formation of short β -strands in trimeric coiled coils. These strands rotate the path of each chain by 120° counterclockwise around the coiled-coil axis, forming a triangular supersecondary structural element called the β -layer. This structure is stabilized by hydrogen bonds formed between the central residues of the three strands, usually a small hydrophobic residue. β -Layers derived from the insertion of 2 or 6 residues are identical.

Tandem insertion of β -layers into a regular fiber gives an α/β coiled coil with alternating α and β structural elements. While a former study solved the crystal structure of an α/β coiled coil built completely from repetitive nonad insertions (Hartmann et al., 2016), similar information was missing for tandem hexad discontinuities to complete the picture. In this work, I have structurally characterized two bioinformatically identified hexad elements – the KAD-VYT-LYT co-optimized module and the repetitive TAT insertion module. The solved crystal structures show that repetitive hexad insertions form a minimalistic α/β coiled coil with three residues occupying the α and the next three β region of the Ramachandran plot and no backbone hydrogen bond in common with a heptad repeat coiled coil. Widely observed in bacterial and phage surface fibrous proteins, the first module KAD-VYT-LYT comprises a canonical nonad β -layer followed by one or more tyrosine β -layers in hexad spacing, which differ substantially from the former. Characterized by a large tyrosine residue occupying the central β -layer position, compared to a small hydrophobic residue (such as alanine or valine) in canonical β -layers, hexad layers are much wider in diameter likely promoted by the bulky tyrosine residue which is bent out of the core. Mutation of tyrosine to alanine completely disrupted the formation of β -layer network, signifying the importance of a bulky sidechain residue occupying the central position in hexad β -layers. Apart from tyrosine, phenylalanine (F) and methionine (M) can also be found occupying this position. An extensive bonding network, characterized by a C-capping interaction network formed by the conserved $\beta 1$ lysines of the first layer and the N-capping interactions established by the $\beta 3$ threonines of successive hexad layers, allows

for a stable assembly of repetitive β -layers within regular coiled-coil fibers. The second type of hexads module built from repetitive TAT β -layers is seldom observed. In the solved crystal structure of a fragment of a hypothetical protein from *Candidatus Parcubacterium*, where two TAT hexads are sandwiched between nonad xxxTATxxx and xxxMATxxx β -layer elements, we see four canonical β -layers stabilized by backbone hydrogen bonds formed among the central alanine residues of the individual layers. While no C-capping interaction network was formed with threonine as the $\beta 1$ residue in the first β -layer, the $\beta 3$ threonines in all layers formed N-caps to the following α -helical segments of their respective chains. Each $\beta 1$ threonine in hexad layers coordinates the $\beta 3$ threonine of the previous β -layer with conserved backbone and sidechain interactions. The additional sidechain contacts made by threonines, which cannot be brought about by MATxxx hexad repeats, could explain the stable formation of TAT tandem hexad β -layers.

The KAD-VYT-LYT co-optimized module, which is frequently observed in a recurring fashion in the stalks of trimeric autotransporter adhesins and other bacterial and viral surface proteins, can now be simply modelled for new proteins using our solved crystal structures as templates. Such β -layer recurrences in surface protein fibers led us to propose their involvement in providing resilience to environmental stress conditions by tight interleaving of the monomeric chains. By integrating β -stranded elements into long α -helical chains, they further provide structural and functional complexity. Many pathogenic bacteria and viruses use this simple mechanism of repetitively arranged complex modules in surface proteins to generate the molecular diversity in order to evade host immune responses.

Summarizing the above results, it is clear that repetitive insertion of both nonad and hexad elements in a canonical coiled-coil fiber results in the formation of an α/β coiled coil. The data confirms that minor mutations such as simple insertion of two residues or deletion of one residue per repeat in a canonical coiled coil result in a novel backbone structures within the allowed regions of Ramachandran space.

2.5 METHODS

2.5.1 Molecular cloning

Unless otherwise mentioned, genes were commercially synthesized (Eurofins) with *E. coli* codon optimization.

The construct Mglob encoding residues 244-274 of a hypothetical protein OMM_04225 from *Candidatus Magnetoglobus multicellularis* str. *Araruama* (GenBank ETR68995.1) was fused between two GCN4-pII coiled-coil fragments by cloning within the XbaI-HindIII sites of the pASK-IBA GCN4-pII plasmid.

Construct Mmor containing residues 311-335 of a conserved hypothetical protein MHK_004959 from *Candidatus Magnetomorum* sp. HK-1 (GenBank KPA14831.1) was designed identical to Mglob with flanking trimeric GCN4-pII adaptors and cloned within the BamHI and HindIII sites of pASK-IBA GCN4-pII-Mglob plasmid. Mmor-mut1 and Mmor-mut2 constructs were derived from site-directed mutagenesis using Mmor construct as the template. Mmor-mut1 features mutation of the two conserved tyrosine residues to alanines Y321A and Y327A. Mmor-mut2 has V320A and L326A mutations.

Table 2.1: List of primers for PCR amplification

Construct	Primer
Mmor-mut1	Y321A fw: 5'- CGAACAAAGCAGATGTGGCTACGAAAGACCAGTTATAC
	Y321A rv: 5'- GTATAACTGGTCTTTCGTAGCCACATCTGCTTTGTTTCG
	Y327A fw: 5'- GGCTACGAAAGACCAGTTAGCCACCAAACTGAGATTAACAGTCAA
	Y327A rv: 5'- TTGACTGTTAATCTCAGTTTTGGTGGCTAACTGGTCTTTCGTAGCC
Mmor-mut2	V320A fw: 5'- CAAAGCGAACAAAGCAGATGCGTATACGAAAGACCAGTTA
	V320A rv: 5'- TAACTGGTCTTTCGTATACGCATCTGCTTTGTTTCGCTTTG
	L326A fw: 5'- GCATACACCAAACTGAGATTAACAGT
	L326A rv: 5'- CTGGTCTTTCGTATACGCATCTGC
ParcuUY81 ₁₋₇₀	fw: 5'- TAAGTGAATCCAACTGGAC
	rv: 5'- ACCGATACGGTCCACTAAATC
ParcuUY81-GCN4	fw: 5'- GACCATGGTCTCCACTGACAGCTACCAAAGCCGAACTGC
	rv: 5'- CCCAAGCTTCAGCTCGCACTAGAGACTTTCC

Table 2.2: Sequence of hexad repeat constructs

Construct	GenBankID	Residues	Sequence
Mmor	KPA14831.1	311-335	(6xHis-TEV) -GGSG- (GCN4-pII) _N -INLKANKAD ---VYTKDQ ---LYTKTEINSQ- (GCN4-pII) _C
Mmor-mut1	KPA14831.1	311-335	(6xHis-TEV) -GGSG- (GCN4-pII) _N -INLKANKAD ---VATKDKQ ---LATKTEINSQ- (GCN4-pII) _C
Mmor-mut2	KPA14831.1	311-335	(6xHis-TEV) -GGSG- (GCN4-pII) _N -INLKANKAD ---AYTKDKQ ---AYTKTEINSQ- (GCN4-pII) _C
Mglob	ETR68995.1	244-274	(6xHis-TEV) -GGSG- (GCN4-pII) _N -ISQKANSQD ---VYNKTD ---LYPKTD ---LYTKTEMDTA- (GCN4-pII) _C
ParcuUY81-GCN4	KKW34673.1	18-54	(6xHis-TEV) -GGSG- (GCN4-N16V) _N -FAETATKAE ---TATKAE ---TATKKD IAGMATKHD IAQLDKR- (GCN4-N16V) _C
ParcuUY81-fl	KKW34673.1	1-93	(6xHis-TEV) - MKKPITLEKLVSMAVAVG FAETATKAE ---TATKAE ---TATKKD IAGMATKHD IAQLDKR IDGLDKK IADLVDR IGRVESK LDRALNKEVAAWKVSSAS
ParcuUY81₁₋₇₀	KKW34673.1	1-70	(6xHis-TEV) - MKKPITLEKLVSMAVAVG FAETATKAE ---TATKAE ---TATKKD IAGMATKHD IAQLDKR IDGLDKK IADLVDR IG
ParcuUY81-heptad			(6xHis-TEV) - MKKPITLEKLVSMAVAVG FAETKAE TATIKAE TATIKKD IAGMKHD IAQLDKR IDGLDKK IADLVDR IGRVESK LDRALNKEVAAWKVSSAS

ParcuUY81-fl, a hypothetical protein UY81_C0065G0003 from *Candidatus Parcubacteria* (*Giovannonibacteria*) GW2011_GWA2_53_7, and its control design ParcuUY81-heptad were gene synthesized and cloned into the NcoI and BamHI sites of pETHis1a expression vector (Bogomolovas et al., 2009) allowing for the expression of the construct with an N-terminal His₆ tag and a TEV protease cleavage site. ParcuUY81₁₋₇₀ was amplified from ParcuUY81-fl by round-the-horn mutagenesis using phosphorylated primers. The resulting PCR product was ligated and selected after transformation to obtain the final clone. ParcuUY81-GCN4 was also derived from ParcuUY81 by PCR amplification of the residues 18-54 and cloned within the BsaI/HindIII of the pASK IBA2-HisTEV GCN4-N16V vector.

2.5.2 Protein expression and Purification

Constructs Mmor and Mglob were expressed in *E. coli ArcticExpress* (DE3) at 37°C until O.D.₆₀₀ of 0.8 and induced with 0.1 mM AHTC (anhydrotetracycline) at 12°C for 24 hours. Mmor-mut1 and Mmor-mut2 were expressed in *E. coli* Top10. ParcuUY81-fl, ParcuUY81-heptad, ParcuUY81₁₋₇₀ and Parcu-GCN4 were all expressed in *E. coli* C41 (DE3) strain. Cells were grown at 37°C until O.D.₆₀₀ = 0.5, and then induced with 0.5 mM IPTG (for ParcuUY81-fl, ParcuUY81-heptad and ParcuUY81₁₋₇₀) or 0.1 mM AHTC (for ParcuUY81-GCN4). Finally, the cells were harvested by centrifugation after 4 hours.

Proteins Mmor, Mglob, ParcuUY81-fl, ParcuUY81-heptad, ParcuUY81₁₋₇₀ and ParcuUY81-GCN4 were purified from supernatant under native conditions. Cells were resuspended in lysis buffer (20 mM Tris (pH 7.5), 150 mM NaCl, 4 mM MgCl₂, DnaseI and protease inhibitor cocktail) and lysed using sonication (50% duty cycle and output control 5). After centrifugation, the supernatant was loaded on Ni-NTA Agarose column pre-equilibrated with Buffer A (20 mM Tris (pH 7.5), 150 mM NaCl) and the bound proteins were eluted with increasing concentrations of imidazole. Eluted protein was dialyzed against Buffer A and subsequently digested with TEV protease (TEV protease: protein 1:20). The cleaved protein was reloaded on Ni-NTA and the flow-through was collected. Finally, the protein was subjected to gel filtration chromatography on Superdex 75.

For Mmor-mut1 and Mmor-mut2, purification was performed under denaturing conditions. Cell lysate was dissolved in 6 M guanidine hydrochloride for 1 h at room temperature. The cleared supernatant was loaded on Ni-NTA Agarose column pre-equilibrated with Buffer A*

(20 mM Tris (pH 7.5), 300 mM NaCl and 6 M Gua.Cl) and eluted with imidazole. The collected protein was refolded by dialysis against Buffer A (20 mM Tris (pH 7.5), 150 mM NaCl) and processed to remove the N-terminal histidine tag by TEV proteolysis.

For NMR analysis of ParcuUY81-fl, cells were grown at 37°C in M9 minimal medium supplemented with ¹³C-glucose and ¹⁵N-ammonium chloride, until O.D.₆₀₀ = 0.7-0.8. The cells were then induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) at 20°C and harvested by centrifugation after 18 hours. ¹³C and ¹⁵N-labelled ParcuUY81-fl was then purified under native conditions on a NiNTA affinity column and finally cleaned on a Superdex 75 size-exclusion chromatography column to remove larger aggregates. 250 μl of the sample was mixed with 50 μl of D₂O and filled in a Shigemi symmetrical NMR microtube (Product No. Z543349, Sigma Aldrich, Germany).

2.5.3 Homology modelling

The crystallographic structure of (residues A34-A137) was used together with Hia (PDB code 3EMO) translocation pore domain, the nonad β-layer (PDB code 5APT: IENKADKAD fused between the GCN4 leucine zippers) and the hexad repetitive module KAD-VYT-VYT from Mmor structure (described in section 2.2.1) to build the model using MODELLER (Sali et al., 1995). The missing stalk residues (138-220) were modelled using the solved N-terminal coiled-coil (residues 27-137) of NadA5 as the template.

2.5.4 Crystallization, data collection and structure determination

Crystallization trials were set up in 96-well sitting-drop plates with drops consisting of 300 nl protein solution and 300 nl reservoir solution (RS), and reservoirs containing 50 μl RS. All crystals were cryo-protected, then loop mounted and flash-cooled in liquid nitrogen. Data were collected at 100 K and a wavelength of either 1.07 Å at beamline X10SA of the Swiss Light Source (Villigen, Switzerland), using a PILATUS 6M-F hybrid pixel detector (Dectris Ltd.). All data were indexed, integrated and scaled using XDS (Kabsch, 2010). Structures were solved by molecular replacement with MOLREP (Vagin and Teplyakov, 2000) They were finalized in cycles of manual modeling with Coot (Emsley and Cowtan, 2004), and refinement with REFMAC5 (Murshudov et al., 1999). This work was done together with Dr. Marcus Hartmann. The structures have not yet been submitted to the PDB.

Table 2.3: Crystallization and cryo conditions.

Protein*	Reservoir solution**	Cryo solution**
Mmor	0.1 M HEPES pH 7.5, 30% (w/v) PEG 300	Mother liquor
Mmor-mut1	0.1 M tri-sodium citrate pH 5.5, 20% (w/v) PEG 3000	20% PEG 2000
Mmor-mut2	0.1 M sodium citrate , 8% (w/v) PEG 8000	20% PEG 200
Mglob	Morpheus HT-96 screen, well D2***	Mother liquor
ParcuUY81-fl	0.1 M SPG buffer pH 6, 25% (w/v) PEG 1500	10% glycerol
ParcuUY81-Tryp	25% (w/v) PEG 2000 MME	20% PEG 400
ParcuUY81-short	0.1 M Bis-Tris pH 6.5, 2 M (NH ₄) ₂ SO ₄	25% glycerol
ParcuUY81-heptad	0.2 M K ₂ SO ₄ , 20% (w/v) PEG 3350	20% glycerol
ParcuUy81-GCN4	0.1 M sodium acetate (pH 4.6), 2 M sodium formate	30% glycerol

*All protein samples were prepared in buffer containing 20 mM Tris, pH 7.6, and 150 mM NaCl.

**The reservoir and cryo conditions are shown for the best diffracting crystals.

*** Morpheus screen HT-96/ FX-96, Molecular Dimensions

CHAPTER 3

Structural and functional characterization of MempromCC proteins – a new family of membrane proteins in prokaryotes and mitochondria

In the previous chapter, we examined the structural features of β -layers in regular coiled coils. β -layers are found widely distributed in proteins ranging from phages and bacteria up to eukaryotes and display conserved characteristic sequence motifs. Probing new proteins containing β -layers, we identified a new family of integral membrane proteins displaying a head-neck-stalk-anchor architecture (Pfam domain annotation DUF1640), broadly distributed in prokaryotes and localized to the mitochondria of eukaryotes. We refer to this protein family as mempromCC, for membrane-attached proteins of prokaryotes and mitochondria containing coiled coils. Besides mostly uncharacterized proteins, DUF1640 includes prominent members, such as human MCUR1 (mitochondrial calcium uniporter regulator 1 or CCDC90A) and its paralog CCDC90B (coiled-coil domain containing protein 90B). MCUR1 is an essential regulator of Ca^{2+} uptake through the mitochondrial calcium uniporter (MCU) whereas CCDC90B does not demonstrate a similar effect. Furthermore, MCUR1 was found to regulate the calcium threshold in permeability transition. These mitochondrial Ca^{2+} modulations play a significant role in maintaining cellular bioenergetics, mediating cellular differentiation and initiation of programmed cell death pathways. In this chapter, we examine the structural and biophysical characteristics of human MCUR1 and CCDC90B, and the effects of divalent cations Ca^{2+} and Mg^{2+} . We further investigate the interaction of human mempromCC paralogs with MCU and the permeability transition pore complex component Cyclophilin D. As MCU lacks homologs in bacteria, yeast and certain fungi lineages, we explore the possibly evolutionary conserved function(s) of mempromCC proteins in the second part of this chapter using the bacterium *Caulobacter crescentus* as a model organism.

3.1 INTRODUCTION

3.1.1 The mempromCC protein family

Automated and manual motif-based sequence searches in combination with secondary structure prediction were employed previously in our group to identify new β -layer containing proteins in non-redundant databases (Hartmann et al., 2016). Starting with a known β -layer sequence motif, a heterogeneous group of integral membrane proteins was found in prokaryotes and eukaryotes (Bioinformatics was done by Prof. Andrei Lupas and Ioanna Karamichali). The identified set of membrane proteins comprises in total 1085 proteins from 270 species from all the three domains and five viruses. Majority of the prokaryotic sequences belong to proteobacteria and a minor part are archaeal proteins. The viral homologs are found in bacteriophages and share a high sequence similarity. And all eukaryotic proteins are nuclear-encoded with a mitochondrial target sequence. In contrast to the more diverse prokaryotic proteins, all identified eukaryotic proteins are annotated as DUF1640 homologs.

Although highly diverse in sequence and length, all family members are characterized by a common head-neck-stalk-anchor (HNSA) architecture. A membrane anchored trimeric coiled-coil stalk projects an N-terminal helical head domain. The head-stalk junction is intermediated by the presence of one or multiple β -layer connectors. The sequences differ remarkably in size. The longest representatives, found in the fungal divisions *Ascomycota* and *Basidiomycota*, comprise up to 320 residues and the shortest example, from *Xylella*, consists of only 60 residues. The size of helical head domains varies between 40 and 220 residues. Similarly, the trimeric coiled-coil stalks of the proteins diverge highly in length and in their repeat patterns even between closely related species. Although we mostly identify left-handed supercoiled stalks built from canonical heptad repeats, segments indicative for a right-handed supercoil built from repeats of eleven or fifteen residues could also be seen.

In order to analyze the phylogeny of the obtained sequences, collected sequences were clustered by pairwise sequence comparison of the head-neck region using CLANS (Frickey and Lupas, 2004) (Fig. 3.1). Nearly 70% of the identified proteins clustered within the central main group. This subcluster comprises many proteobacterial proteins belonging to α -, β -, γ - and δ -classes together with all eukaryotic homologs that specifically localize to the mitochondria. The remaining sequences form small peripheral groups. They exclusively

contain the prokaryotic proteins. Whereas the sequences in the central group display strong connections to each other, they show little or no connections to these peripheral prokaryotic groups. These few evolutionary connections exist between the proteobacterial sequences of the main cluster and the groups of *Pseudomonas*, *Delftia* and *Chromatiales*. Among the peripheral groups, only the proteobacterial clusters of *Enterobacteriales* and *Pseudomonas*, and the *Aquificales* group of *Sulfurihydrogenibium* and the archaeal *Methanocaldococcus* clusters are related to each other.

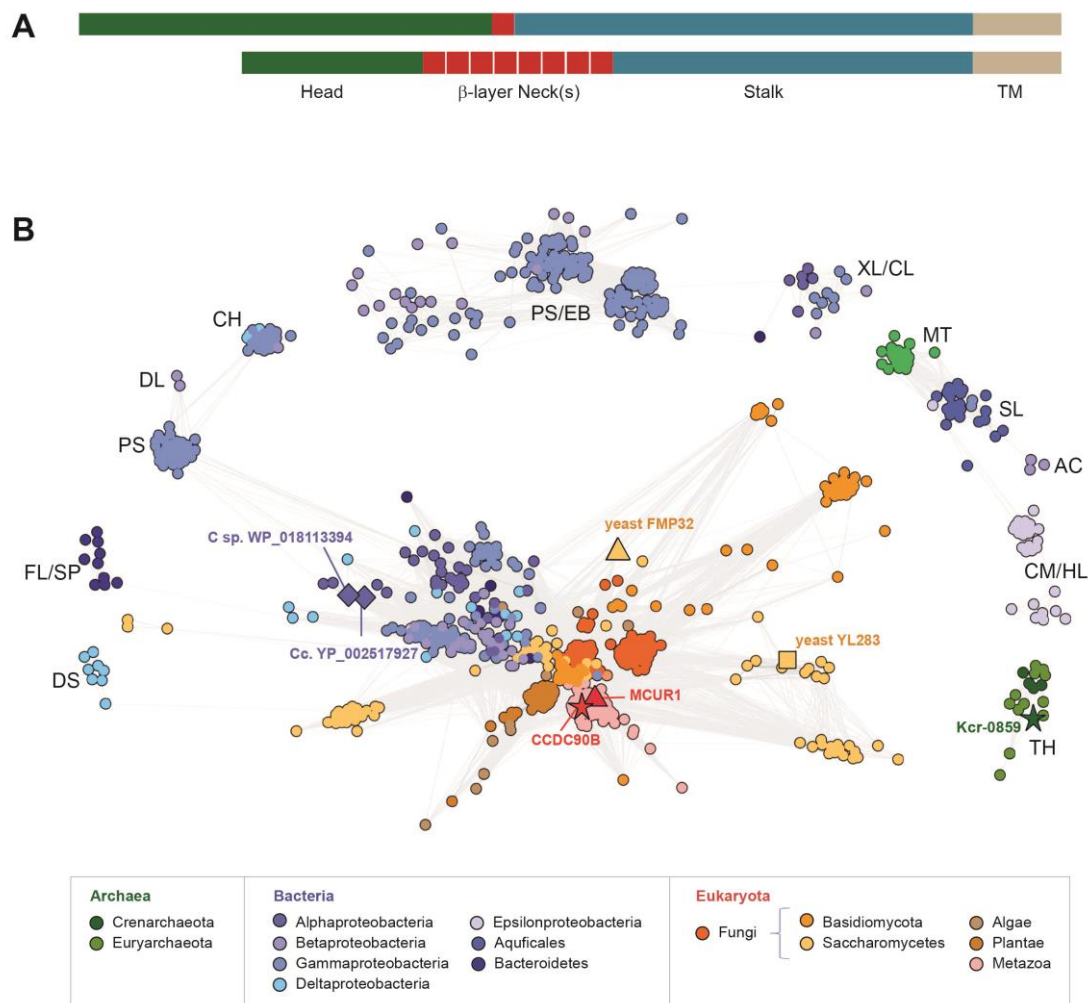


Figure 3.1: **Bioinformatic classification of mempromCC protein family.** (A) Domain architecture of mempromCC homologs characterized by a head domain (green), one or multiple β -layer necks (red), a coiled-coil stalk (blue) and a transmembrane anchor (grey). (B) Cluster map of the head-neck segments of identified β -layer containing proteins generated using CLANS. Sequences are shown as circles and colored according to taxonomic classification as described below. Evolutionary related sequences are connected by grey lines based on BLAST score values. Triangles and squares highlight proteins of particular interest in the scope of this work. Stars indicate proteins with solved crystal structures. The central cluster represents the mempromCC family. Peripheral clusters are abbreviated as *Achromobacter* (AC), *Chromatiales* (CH), *Candidatus liberibacter* (CL), *Campylobacter* (CM), *Delftia* (DL), *Desulfovibrio* (DS), *Enterobacteriales* (EB), *Flavobacteria* (FL), *Helicobacter*

(HL), *Methanocaldococcus* (MT), *Pseudomonas* (PS), *Sulfurihydrogenibium* (SL), *Sphingobacteria* (SP), *Thermoprotei* (TH), and *Xylella* (XL).

Multiple sequence alignment with Clustal Omega (Sievers et al., 2011) and secondary structure predictions using Quick2D (Zimmermann et al., 2017) of the head-neck region of sequences selected from the central main group (Fig. 3.2) reveal important information. The conserved head domain, chiefly 40-50 residues in length, is predicted to be helical. Some residues are conserved in nearly two-thirds of the sequences and include a conserved “FDT” (Phe-Asp-Thr) motif located close to the N-terminal region of the head. More distantly related proteins, including sequences from *Pseudomonas* and *Chromatiales*, lack this motif but have the same secondary structure prediction. The β -layer necks are characterized by the conserved sequence motif [aliphatic]-A-T-K-[polar]-[DE]. On the basis of evolutionary relationship established among the head-neck sequences, we classify the members of central cluster and peripheral groups of *Pseudomonas*, *Delftia* and *Chromatiales* as mempromCC (membrane-attached proteins of prokaryotes and mitochondria containing coiled coils) family. Members of this family are obligate trimers displaying a head-neck-stalk-anchor architecture and comprise of a conserved α -helical head domain that is projected by a membrane-anchored long coiled-coil stalk. The head-stalk junction is invariably mediated by one or multiple repetitively arranged β -layers, therefore clearly differentiating them from other protein families with similar domain architecture.

While prokaryotes seem to possess only one homolog per organism, eukaryotes have at least two paralogous mempromCC proteins. These proteins, classified under DUF1640 (Pfam annotation as domain of unknown function), are predominantly uncharacterized. DUF1640 represents an entirely helical domain of unknown function that is about 160 residues in length and encompasses the conserved head, neck, stalk and membrane anchor. Prokaryotic mempromCC homologs lack signal sequences as well as any N-terminal extensions which indicates their localization to the cytoplasmic membrane. On the other hand, eukaryotic mempromCC homologs are characterized by an N-terminal mitochondrial target signal (MTS). MTS is often succeeded by a non-conserved, variable length region which is predicted to be mostly disordered.

mempromCC family

```

HS1 [55] QRKLTFDTHALVQDLETHGFDKTQAETIVSALTALSNVSLDTIYKE--MVTQAAQ [147]
HS2 [160] SRKLYFDTHALVCLLEDNGFATQQAEIIVSALVKILEANMDIVYKD--MVTKMQ [147]
AT [49] RRAFLVDTLALVRSLEAQGVPSKQAEAITSAITEVLNDSLENVSES--FVSKAE [72]
CR [59] SNHLLVDTLELSKSFEKAGLPRDTAEKLAKDITTLIVLNKEKMEGA--FVKTVV [134]
RS [66] SRINAFDTHRFVAALERN-FPTSIAQTLMRATRALLVGRFGRVKQE--IFGVRD [89]
SC [96] TSQNQLDTIKFYQMLRERGNFSDEQCKIIIALLLQLLNDQFYSCYNDLFLRDME [164]
SP [35] RKYHGFNSLRFVRVLQEAGIDDKKSETLMRLISNVYSDMHEKISDF--SVTKEQ [129]
CC MTAIGFDTLSASKRLREAGMDQPVAEAIVELVQTTMLPDISD-----LATKSE [71]
CS -MNVAFDTLNASRRLRDAGLEERAAEAIVELVQSAAHFPDISG-----LATKTD [123]
CG MAHAAFDTLEFVDELEKSGIPEKQARALAAAVRKAHESSD-----LATKAD [35]
HP -MSIRFDKLRFVKKLQEANQPPEVAEAFATALDEALEQSQSG-----LVNKSD [49]
PC ----MKTELALYQALISINVPEQKANAVIDALETDMHSL-----LATKAD [46]
EH ----MTSVIELYEQLSSAPDDKTRARLIAEAFEQMEQRYSEVTD----LATGAA [85]
CE -MAILFDTLKIRNRLTQKGMNEAVAEVELLNEAQIAD-----LATRRD [40]

```

---h-h--h-hh--h---h---h-hhh-hh--hh---h--h-----hh----

Pseudomonas / Enterobacteriales group

```

PS ----MKSRVEALEKVLPEIRRLVRVETRLDGIEQA----MATKMD [45]
JA -MESMEHRVEKLELAVDAKQRLVRMETRLGGIDHRLDTMASKAD [75]
PA -MENFETRFERLEDNHQIKIDLALFSGQAMA-----FATKAD [198]
EC MSDKLERRIERLEGDLSLTRNDLALAERATN-----LSTKAD [111]
SE ---MLEKRVQKLEDDLAAIRTDLAVIKSN-----YATKED [38]

```

---h---h--h---h--h---hhhh-----hh----

Crenarchaeota group

```

MY -MINVSLARQLLEEFSKDNEAYQEFIRRISVGIAQDDKVRLLILDSVIKEVATKSD [109]
SA --MSVSVTRQLLDEISKDEKLYDELVKKIAVGIAKDDEVRVLVLNSIIKDVATKND [156]
TA --MQSITAEDVVRLFEADARARKRLAELLVGEPDVRLAMINAVLRD----VATKSD [130]
PF ---MESLIDRLLEDAARHPEKRRRLARMLALDIATEETLRSLLEGLLRDVATKQD [78]
KC ----MSNGRQLLEELRKDEELRRALAELIPEVLRNRELRRAILLALSREMATKED [117]

```

-----h---h--h-----h---hh-h-----hh----

Figure 3.2: **The head-neck segment of mempromCC family is evolutionarily conserved.** Sequence alignment of head domain and the first β -layer neck of mempromCC homologs, along with two unrelated groups of *Pseudomonas/Enterobacteriales* and *Crenarchaeota* are shown. Secondary structure α -helix prediction is colored in pink with color intensity proportional to the prediction score. Conserved residues are marked in bold. Blue and black letters indicate residues conserved in at least two-thirds or half of the sequences respectively. Hydrophobic residues conserved in at least 50% of sequences are marked below with “h”. Species names are colored as in (Fig. 1B) and abbreviated as follows: HS1 (NP_068597.2, CCDC90B, *Homo sapiens*), HS2 (NP_001026883.1, MCUR1, *Homo sapiens*), AT (NP_973473.1, *Arabidopsis thaliana*), CR (XP_001694431.1, Rat1, *Chlamydomonas reinhardtii*), RS (CCO26633.1, *Rhizoctonia solani* AG-1 IB), SC (Q05867.1, YL283, *Saccharomyces cerevisiae* S288c), SP (O14042.1 *Schizosaccharomyces pombe* 972h-), CC (YP_002517927.1, *Caulobacter crescentus* NA1000), CS (WP_018113394.1, *Caulobacter* sp. JGI 0001013-D04), CG (WP_006683118.1, *Candidatus glomeribacter gigasporarum*), HP (WP_021111668.1, *Haemophilus parasuis*), PC (AIC21840.1, *Pseudomonas chlororaphis*), EH (WP_029762635.1, *Ectothiorhodospira haloalkaliphila*), CE (ETX04478.1, *Candidatus Entotheonella* Sp. Tsy2), PS (WP_016781459.1, *Pseudomonas fragi*), JA (CDG82950.1, *Janthinobacterium agaricidamnosum* NBRC 102515 = DSM 9628), PA (WP_014605518.1, *Pantoea ananatis*), EC (WP_001737198.1, *Escherichia coli*), SE (WP_000890813.1, *Salmonella enterica*), MY (ZP_09704076.1, *Metallosphaera yellowstonensis* MK1), SA (WP_024084599.1, *Sulfolobus acidocaldarius*), TA (WP_020962471.1, *Thermofilum adornatus*), PF (WP_014025983.1, *Pyrolobus fumarii*), KC (WP_012309502.1, Kcr-0859, *Candidatus Korarchaeum cryptofilum*).

3.1.2 Human mempromCC proteins - MCUR1 and CCDC90B

The only functionally studied members of the mempromCC family to this date, are the human paralogs MCUR1 and CCDC90B. A recent study on MCUR1 (mitochondrial calcium uniporter regulator 1) demonstrated its function as an essential regulator of mitochondrial calcium uniporter-mediated calcium uptake (Mallilankaraman et al., 2012a). It is a 40 kDa protein localized to the inner mitochondrial membrane (IMM). Identified in a directed human RNAi screen, knockdown of MCUR1 reduced the calcium uptake by energized mitochondria in intact and permeabilized HEK293 cells (Mallilankaraman et al., 2012a). A similar effect on Ca^{2+} influx, however, was not observed with the knockdown of CCDC90B (Tomar et al., 2016). Various other studies have proposed different functions for MCUR1: as a cytochrome C oxidase assembly factor (Paupé et al., 2015) as well as a regulator of Ca^{2+} threshold for mitochondrial permeability transition (Chaudhuri et al., 2016). All functional studies are discussed in detail below.

3.1.2.1 MCUR1 as a regulator of calcium uptake by MCU

MCUR1 is ubiquitously expressed in all mammalian tissues. It was first identified in an RNAi screen of 45 mitochondrial membrane proteins to identify candidates that led to a significant reduction in mitochondrial calcium uptake (Mallilankaraman et al., 2012a). The mitochondrial calcium uptake occurs through a highly Ca^{2+} -selective uniporter complex residing in the inner mitochondrial membrane that operates by the huge transmembrane potential generated across its IMM. Knockdown of MCUR1 led to a dramatic reduction (~85%) in the mitochondrial $[\text{Ca}^{2+}]$ without modifying the cytosolic Ca^{2+} content (Mallilankaraman et al., 2012a), which could be rescued by expressing RNAi insensitive MCUR1 cDNA. In further studies, MCUR1 was shown to directly interact with the central pore-forming MCU protein and proposed to function as a scaffold factor for the complex (Lee et al., 2015; Tomar et al., 2016). BN-PAGE analysis showed that MCU oligomerizes in IMM in a complex with an apparent molecular weight of ~480 kDa (Baughman et al., 2011). Deletion of MCUR1 perturbed this higher-order complex assembly. Although there has been a considerable interest in the molecular structure and assembly of MCU complex in the last few years, the regulatory mechanism of calcium uptake by MCUR1 and CCDC90B are still not understood.

The MCU complex in mitochondria

In the early 1960s, it was discovered that mitochondria isolated from cells accumulated large amounts of Ca^{2+} (Deluca et al., 1961; Vasington et al., 1962). A similar finding was reported later for the mitochondria of eukaryotic cells *in vivo*. Finally in 2011, after 50 years of extensive research in the field of mitochondrial calcium signaling, the central pore-forming protein of the MCU complex was identified (De Stefani et al., 2011; Baughman, Perocchi et al., 2011). Subsequently, other regulatory proteins were recognized. Human MCU complex, known to date, is made up of the following components – MCU (pore-forming subunit), MCUb (a dominant negative paralog of MCU), MICU1 and MICU2 (paralogous EF-hand Ca^{2+} -binding gatekeeper proteins), EMRE (a metazoan specific single-pass transmembrane protein), and paralogs MCUR1 and CCDC90B.

Mitochondrial calcium signaling is an essential process that regulates crucial cellular processes like bioenergetics, metabolism, cell differentiation, and cell death pathways. In the cell, mitochondria work as large capacity calcium buffers on the cytoplasmic face (Williams et al., 2013). The rapid compartmentalization of local calcium increases in cytosol through the uniporter and its efflux shapes the cellular calcium signals (Babcock et al., 1997; Putney Jr. and Thomas, 2006; Rizzuto et al., 2012). Mitochondria utilize these spatio-temporal patterns of Ca^{2+} signals to regulate diverse physiological processes and enzymatic activities (Iino, 2010). For instance, entry of mitochondrial calcium signals the production of ATP to match the cytosolic energy requirements (Jouaville et al., 1999; Tarasov et al., 2012).

Calcium uptake occurs at MAMs (Mitochondria-Associated Membranes)

Calcium uptake through MCU is driven by the electrochemical gradient maintained across IMM. This gradient is generated by pumping of protons by the respiratory complexes towards the intermembrane space which results in the creation of a large transmembrane potential, Ψ_m , usually between -150 and -180 mV (Patergnani et al., 2011). Substituting this value in the Nernst equation,

$$V_{\text{eq}} = \frac{RT}{zF} \ln \frac{[\text{Ca}^{2+}]_o}{[\text{Ca}^{2+}]_i}$$

where V_{eq} is the equilibrium potential, R is the universal gas constant ($8.314 \text{ J K}^{-1} \text{ mol}^{-1}$), z is the valence of ionic species, F is the Faraday's constant ($96,485 \text{ C mol}^{-1}$), translates into the equilibrium concentration gradient of $\sim 10^6$ for Ca^{2+} . Under resting cellular conditions, cells maintain a low $[\text{Ca}^{2+}]_{\text{cyt}} \sim 100 \text{ nM}$. If allowed to equilibrate freely, $[\text{Ca}^{2+}]$ in the matrix can

reach up to 100 μM which can depolarize the membranes completely. As these values are not observed experimentally, it shows that the uniporter is tightly regulated to allow a response only above a certain Ca^{2+} threshold concentration.

Ca^{2+} release from the major intracellular Ca^{2+} -stores such as endo/sarcoplasmic reticulum (ER/SR [Ca^{2+}] $\sim 250\text{-}600\ \mu\text{M}$) increases the global cytosolic calcium levels from $0.1\ \mu\text{M}$ to only about $1\text{-}2\ \mu\text{M}$ (Somlyo, 1984; de la Fuente et al., 2013). As MCU complex has a very low affinity for Ca^{2+} ($K_d \sim 20\text{-}30\ \mu\text{M}$) (Kirichok et al., 2004) on the IMS side, this cytosolic increase is insufficient to illicit a stimulatory response. This apparent discrepancy between the low affinity of MCU, low global cytosolic levels of Ca^{2+} and the highly efficient calcium uptake in mitochondria led to the proposal of the “hotspot hypothesis” (Rizzuto et al., 1993, 1998, 2004; Bononi et al., 2012). This hypothesis states that mitochondria preferentially accumulate calcium at regions of close apposition to the ER Ca^{2+} -release sites where domains of high local [Ca^{2+}] exist (Rizzuto et al., 1998; Patergnani 2011) (Fig. 3.3). Known as “Mitochondria-Associated Membranes” (MAMs), transient [Ca^{2+}]_{cyt} at these microdomains can reach $> 10\ \mu\text{M}$.

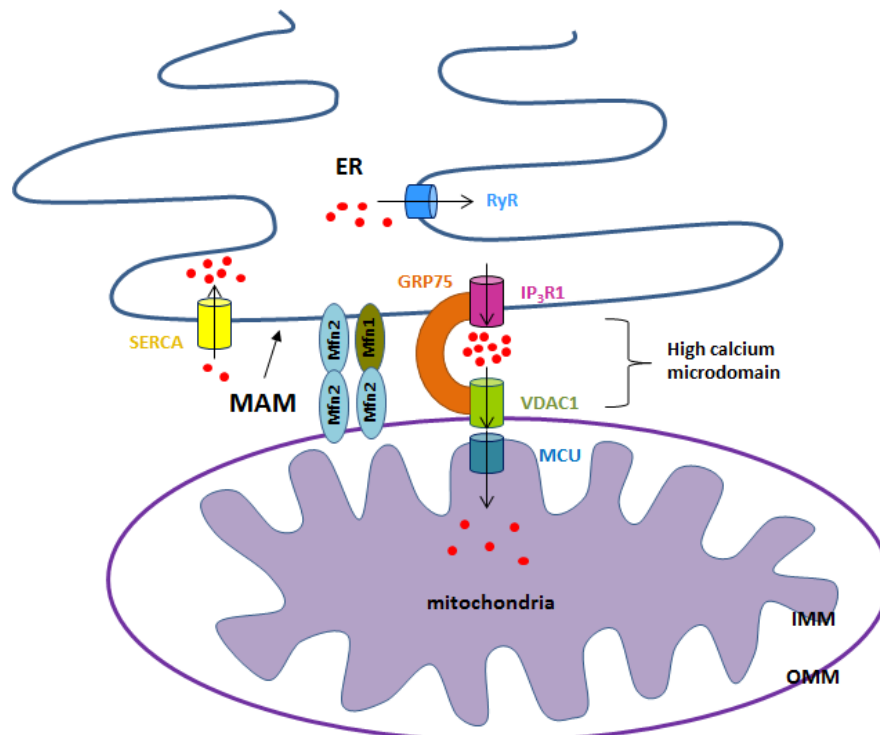


Figure 3.3: **Mitochondria-Associated Membranes (MAMs)** are regions of close apposition between ER and mitochondria. Local microdomains of high calcium concentration at these sites facilitate fast and efficient Ca^{2+} signal transmission. IP₃R (inositol 1,4,5-triphosphate receptor); VDAC1 (voltage-dependent anion-selective channel protein 1); MCU (mitochondrial calcium uniporter); SERCA (sarco/endoplasmic reticulum Ca^{2+} -ATPase); RyR (ryanodine receptor); Mfn1 and Mfn2 (mitofusins); Grp75 (glucose-regulated protein). Adapted from (Chernorudskiy et al., 2017).

Components of the MCU complex

A. MCU – the central pore-forming protein

Mitochondrial calcium uniporter (MCU) is the channel-forming and Ca^{2+} -conducting subunit of MCU complex. It was identified in a screen searching for proteins related to MICU1, which was the first identified component of the MCU complex. Using a combination of whole-genome phylogenetic profiling, genome-wide RNA co-expression analysis and organelle-wide protein expression analysis, two independent groups identified MCU *in silico* in the MitoCarta database (an inventory of nuclear and mitochondrial DNA genes encoding proteins for mitochondrial localization) (Baughman, Perocchi et al., 2011; De Stefani et al., 2011). It was later experimentally verified by whole-mitoplast patch clamp electrophysiology measurement of calcium currents (Chaudhuri et al., 2013). In the cell, MCU works together with accessory proteins that tightly regulate the pore and influence Ca^{2+} influx. The MCU channel is ruthenium-red sensitive and residue S259 was identified to be essential for this sensitivity (Baughman et al., 2011).

Homologs of MCU are present in all major branches of eukaryotes, found in all plants and metazoan, but absent in yeast and some fungi and protozoan lineages (Bick et al., 2012). MCU is ubiquitously expressed in mammals. It comprises of a globular soluble N-terminal domain, two conserved transmembrane helices TM1 and TM2 which are flanked by two coiled-coil helices CC1 and CC2 (Fig. 3.4A). The N and C-terminal domains reside in the mitochondrial matrix and only a short linker (EYSWDIMEP), which features a conserved DIME motif and connects the two TM helices, faces the intermembrane space (Baughman et al., 2011; Oxenoid et al., 2016). Mutation of these negatively charged residues in the DIME motif abolished MCU activity (Baughman et al., 2011; De Stefani et al., 2011).

Recently, four independent groups published the cryo-EM structures of MCU from different fungal species (Baradaran et al., 2018; Fan et al., 2018; Nguyen et al., 2018; Yoo et al., 2018). MCU assembles into a homotetrameric channel (Fig. 3.4B). The transmembrane helix TM2 forms the central hydrophilic core and continues into CC2 in the matrix. The N-terminal domain of MCU is connected to TM1 via an elongated CC1 coiled-coil domain, which also acts as a stabilizing peripheral helix to the core bundle (Fig. 3.4C). The critical DxxE motif

lines the entrance of the pore on the IMS side with protruding carboxylate rings from D225 and E228 forming the Ca^{2+} selectivity filter (Fig. 3.4D).

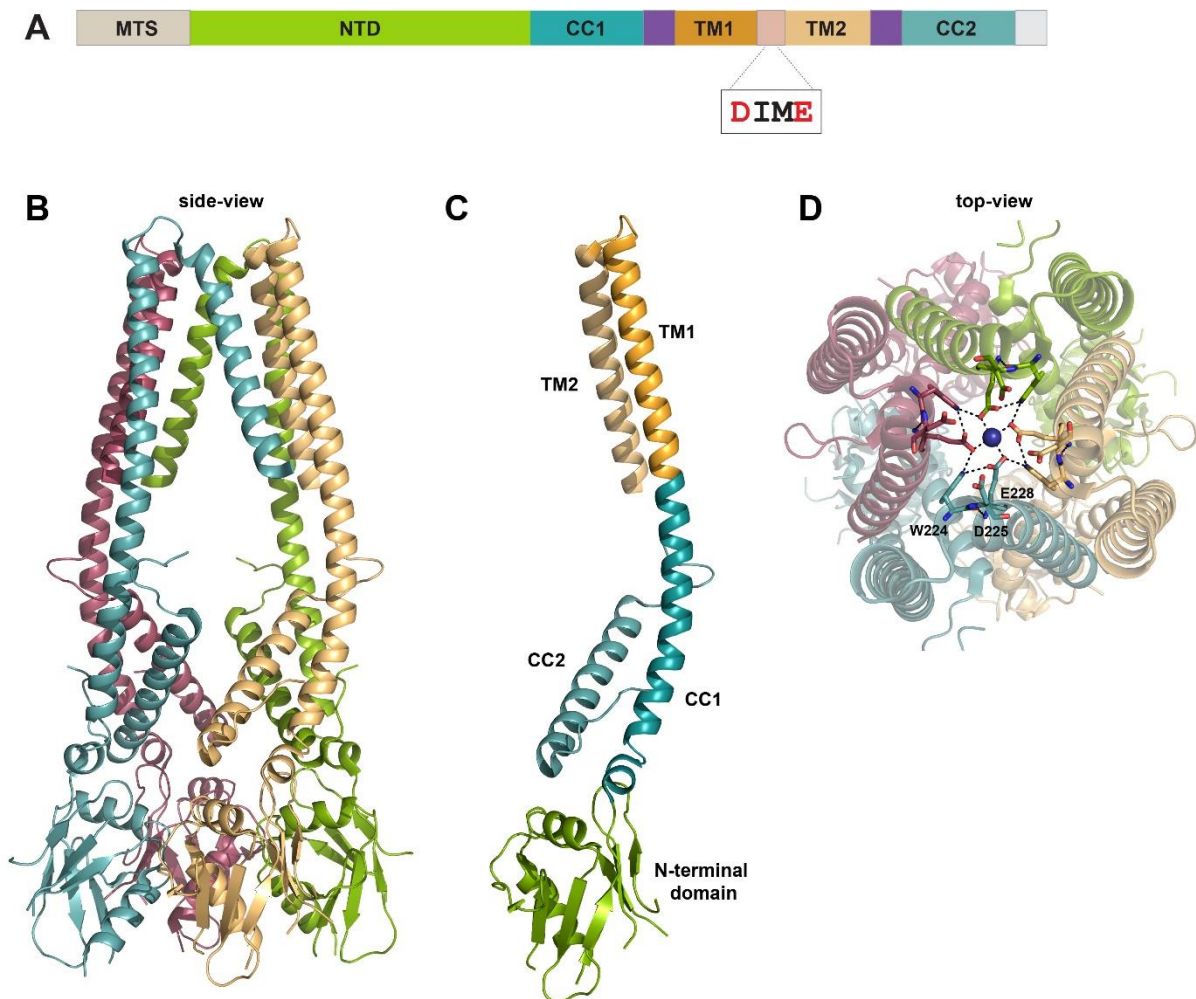


Figure 3.4: **Structure of pore-forming MCU protein.** (A) Domain architecture of MCU. It consists of a mitochondrial target signal (MTS), N-terminal globular domain (NTD), two transmembrane helices (TM1 and TM2), and two coiled-coil helices (CC1 and CC2). The loop connecting TM1 and TM2 features a conserved DIME motif (B) Tetrameric structure of fungal MCU solved by cryo-EM (PDB 6dnf) (Baradaran et al., 2018). Individual chains are colored differently. CC2 and TM2 form the central pore, while CC1 and TM1 stabilize the core as peripheral helices. (C) MCU monomer colored by its domain organization. CC1 helix directly continues into TM1. (D) The two rings of negatively charged residues D225 and E228, part of the conserved DIME motif, act as the selectivity filter for Ca^{2+} . Top view shows the interaction network of W224 and E228 with a central calcium ion (blue sphere).

The structure of the N-terminal soluble domain of human MCU (residues 72-189) was solved by X-ray crystallography. It displays a β -grasp-like fold wherein two central helices, nearly perpendicular to each other, are sandwiched between 3-stranded β -sheets (Lee et al., 2015; Lee et al., 2016). Deletion of MCU N-terminal domain severely abolished calcium uptake through the uniporter without affecting its oligomer assembly. Electrostatic surface

charge analysis identified two prominent acidic and two basic patches on the surface of MCU-NTD which may facilitate higher-order assembly of the protein (Lee et al., 2016). Divalent cations Ca^{2+} and Mg^{2+} bind to the acidic surface of MCU-NTD with a near millimolar affinity (Lee et al., 2016). This binding translates into an auto-regulatory mechanism of MCU, conveying a negative feedback response to close the channel when exposed to prolonged high $[\text{Ca}^{2+}]$ at the immediate base of microdomains. It was also shown that MCU-NTD directly interacts with MCUR1, but how this interaction is mediated remains to be identified.

B. MCUB – the dominant negative paralog of MCU

MCUB (or CCDC109B), a paralog of MCU, shows 50% sequence similarity to MCU and shares the same protein structure and topology, comprising of 2 TM helices connected by a short loop facing IMS. This loop, however, bears a critical amino acid substitution E257V (VYSWDIMEP) that leads to a drastic change in the channel properties owing to one reduced negative charge. MCUB acts as dominant-negative form, and the insertion of one or more MCUB subunits in the uniporter channel was shown to reduce the Ca^{2+} influx.

C. MICU1, MICU2 and MICU3 – the gatekeepers

MICU1 (mitochondrial calcium uniporter 1) was the first MCU complex component to be identified using a comparative proteomics approach (Perocchi et al., 2010). It is an inter-membrane space localized soluble protein that is approximated to the IMM by its association with the C-terminal polyaspartate tail of EMRE. It contains two-highly conserved predicted EF-hand Ca^{2+} -binding domains (Perocchi et al., 2010; Csordás et al., 2013). The crystal structure of MICU1 in both apo and Ca^{2+} -bound forms has been solved (Wang et al., 2014). Apo-MICU1 forms a hexamer. Upon Ca^{2+} binding, it undergoes a conformational change to form multiple oligomeric species which are then suspected to activate MCU.

MCU and MICU1 proteins show the same evolutionary pattern of expression (Bick et al., 2012). MICU1 has two paralogous proteins: MICU2 and MICU3, sharing around 25% sequence identity to MICU1. While MICU2 is expressed ubiquitously in human tissues like MICU1, MICU3 is predominantly expressed in the central nervous system (CNS) (Plovanich et al., 2013). MICU2 forms an obligate dimer with MICU1 and is the dominant gatekeeper protein. While at low $[\text{Ca}^{2+}]$, both MICU1 and MICU2 exercise their tight regulation on the MCU activity; upon physiological stimulation, Ca^{2+} inhibits MICU2 and activates MICU1

allowing a prompt response to calcium accumulation in the matrix (Mallilankaraman et al., 2012b; Patron et al., 2014). The rates of calcium entry show a sigmoidal behavior – slow at low $[Ca^{2+}]$ and exponential at higher $[Ca^{2+}]$ ($>10 \mu M$) (Rizzuto et al., 1998; Csordás et al., 1999, 2010; Giacomello et al., 2010).

Interaction studies reported that MCUR1 and MICU1 do not pull-down together (Mallilankaraman et al., 2012a). This could suggest that the two proteins do not exist in the MCU complex at one time. It leaves us with open questions – Do different pools of MCU complexes containing either of the two proteins - MCUR1 or MICU1 - exist? Does the cell require a certain fraction of the complexes without a gatekeeper protein at the resting cellular conditions? The answers still need to be investigated.

D. EMRE – the scaffold factor

EMRE (essential MCU regulator) was identified using a combination of SILAC and mass spectrometry-based proteomics approach (Sancak et al., 2013). It is a 10 kDa single-pass transmembrane protein in IMM with a highly conserved aspartate-rich C-terminus. In the IMS, this acidic C-terminal tail strongly associates with the basic region of MICU1. EMRE also interacts with the TM1 helix of MCU protein with its single transmembrane helix. Therefore, it acts as a bridging subunit keeping the Ca^{2+} -sensing gatekeeper MICU1/MICU2 complex attached to the functional MCU pore, just like a lid on the top of the channel (Tsai et al., 2016). Homologs of EMRE are not found in any plants, protozoa or fungi, highlighting that it is a metazoan innovation (Sancak et al., 2013). Human MCU cannot form functional channels in the absence of EMRE (Kovács-Bogdán et al., 2014). MCUR1 could also pull-down EMRE (Tomar et al., 2016), but which domain of MCUR1 is involved in the interaction to the N-terminus of EMRE is not known.

The assembly of MCU channels is regulated by the proteolytic activity of m-AAA proteases. Expression of EMRE, the bridging subunit in MCU-MICU1 assembly, in excess can result in the assembly of incomplete uniporter sub-complexes MCU-EMRE and EMRE-MICU1. This leads to the accumulation of constitutively active MCU-EMRE channels, a condition that has been identified to lead to a muscular disorder in humans (Tsai et al., 2017). m-AAA proteases AFG3L2 and SPG7 were found to be involved in EMRE turnover, rapidly degrading any non-assembled EMRE using the energy from ATP hydrolysis. This guarantees

the proper assembly of fully functional MCU channels, preventing Ca^{2+} leakage into the mitochondria thereby regulating Ca^{2+} homeostasis.

3.1.2.2 MCUR1 as a cytochrome c oxidase (COX) assembly factor

A 2015 study reported that MCUR1 plays a role in COX assembly and does not directly regulate MCU activity (Paupe et al., 2015). Knockdown of MCUR1 (by siRNA and lentiviral mediated shRNA) in fibroblasts resulted in a COX IV assembly defect, which was found to be responsible for the decreased mitochondrial membrane potential and indirectly to the reduced calcium current. Steady state levels of COXI and COXII, components of COX IV complex, were significantly reduced in MCUR1 KD cells whereas other subunits remained unaffected. As COXI and COXII are the earliest subunits to enter into the complex IV assembly pathway, the instability of these newly synthesized components suggested that MCUR1 might function in chaperoning COXI and COXII into an early COX complex (Paupe et al., 2015). These data imply that MCUR1 is likely involved in incorporation or maturation of these subunits, or plays a role in addition of prosthetic factors necessary to stabilize the COX assembly intermediate.

The authors further studied the COX assembly defect in budding yeast. As the yeast species lack MCU orthologs, presence of MCUR1 and CCDC90B homologs in yeast suggests a function unrelated to mitochondrial calcium uptake. Deletion of *fmp32* (an ortholog of CCDC90B in *Saccharomyces cerevisiae*) resulted in a growth defect on non-fermentable carbon source (glycerol) which substantially increased when grown at higher temperature (36°C compared to 28°C), indicating a COX defect and reduced OXPHOS activity. The authors concluded that the abrogation of Ca^{2+} uptake in MCUR1 deficient cells could likely be a secondary effect due to reduced $\Delta\Psi_m$ as a result of COX assembly defect.

Direct MCU-mediated Ca^{2+} currents recorded by patch clamp electrophysiology of isolated mitoplasts, however, did not support the above claim that reduction of mitochondrial membrane potential is exclusively responsible for reduced Ca^{2+} uptake (Vais et al., 2015). Stable knockdown of MCUR1 (by ~75%) reduced the Ca^{2+} currents through the uniporter by ~65% which could be rescued by expression of shRNAi-insensitive MCUR1. This confirmed the finding that MCUR1 is a direct regulator of calcium uptake by MCU.

3.1.2.3 MCUR1 as a regulator of Ca²⁺ threshold for mitochondrial permeability transition

Calcium entry is crucial for ATP production and cell survival. But excessive calcium uptake can trigger the cell's apoptotic process. Overload of calcium in mitochondria leads to the opening of permeability transition pore (PTP) with the disruption of membrane potential resulting in the loss of solutes (<1.5 kDa). This osmotic imbalance swells up and disrupts the mitochondria releasing cytochrome *c*, which in turn activates pro-apoptotic factors.

In 2016, a study identified that *Drosophila* MPT response is resistant to calcium overload which translates into improved cell survival (Chaudhuri et al., 2016). By comparing the differential behavior of PTP opening in *Drosophila* and humans, they found that MCUR1 conferred the permeability transition sensitive to electrophoretic Ca²⁺ uptake in human cells. As MCUR1 has no homolog in *Drosophila* genome, similar sensitivity could be induced in *Drosophila* cells with the expression of human MCUR1. Clearly, MCUR1 is not essential for regulating mitochondrial calcium transport in all species, since *Drosophila* cells display an intact Ca²⁺ uptake despite lacking MCUR1 homolog.

Originally identified by Haworth and Hunter (1979), PTP is a large conductance channel in the inner mitochondrial membrane. Although the molecular identity of MPT is controversial, key components include Cyclophilin D (CypD), TSPO, ADP/ATP translocase, F1-F0-ATP synthase and spastic paraplegia 7 (SPG7). Cyclophilin D is an important and the only confirmed regulator of MPTP. Full length human MCUR1 was found to interact with CypD in addition to MCU (Chaudhuri et al., 2016). It was proposed that MCUR1 could act as a bridging subunit between MCU and CypD, where approximation of the two complex subunits exposes the Ca²⁺ sensor in PTP to a region of higher local [Ca²⁺] near the matrix face of the uniporter pore (Chaudhuri et al., 2016).

3.1.3 Pathophysiological implications of the MCU complex

As we know, mitochondrial calcium buffering plays a critical role in the regulation of cellular metabolism, ATP production and cell survival (Bonora et al., 2012; Giorgi et al., 2012). During mitochondrial Ca²⁺ overload, cells are subjected to apoptotic/necrotic cell death pathways (De Stefani et al., 2011). Several oncogenes and tumor suppressors modulate the mitochondrial calcium signals to exert their anti/pro-apoptotic activities. One example is the regulation of MCU expression by miRNA miR-25 (Marchi et al., 2013). miR-25 can reduce the calcium

uptake through specific down-regulation of MCU, providing resistance against Ca^{2+} -induced apoptosis. miR-25 is found to be upregulated in a variety of human cancers, including prostate and colon carcinomas. Other pathophysiological effects result from the mutation of one or more components of the human MCU complex. Homozygous loss-of-function mutation of MICU1 has been associated with proximal myopathy, learning difficulties and a progressive extrapyramidal movement disorder, stressing the requirement of a proper MCU gating (Logan et al., 2014).

3.1.4 Aims of the study

The primary objective of this study is the structural, biophysical and functional characterization of eukaryotic and bacterial proteins of the mempromCC family. Using a combination of X-ray crystallography and NMR solution spectroscopy, I aim to solve the structures of human paralogs MCUR1 and CCDC90B, and bacterial homologs from *Caulobacter* species. Subsequently, I probe the cellular localization of mempromCC proteins. For functional characterization, I use human MCUR1, which has been identified to be an essential regulator of calcium uptake through the mitochondrial calcium uniporter, as a model protein. I investigate the molecular interaction between MCUR1 and MCU and identify the domain(s) involved in this interaction. Furthermore to elucidate the regulatory function of MCUR1 on Ca^{2+} influx, I analyze its structure and biophysical properties, as well as its molecular interaction with MCU, in dependence of the divalent cations Ca^{2+} and Mg^{2+} . Later, I investigate MCUR1 interaction with cyclophilin D, an essential component of the mitochondrial permeability transition complex involved in initiating programmed cell death pathway. As MCU lacks homologs in bacteria, yeast and certain fungi lineages, in the second part of the study, I explore the function of prokaryotic mempromCC homologs using the bacterium *Caulobacter crescentus* as a model organism.

3.2 RESULTS

3.2.1 The human mempromCC proteins – MCUR1 and CCDC90B

3.2.1.1 Bioinformatics sequence analyses

Human cells express two mempromCC homologs: CCDC90A (or MCUR1 for Mitochondrial Calcium Uniporter Regulator 1) and CCDC90B. The two proteins share a high sequence similarity of 72% (sequence identity of 53%), featuring an N-terminal mitochondrial target signal (MTS), a conserved head, one β -layer (neck) connecting the head and the long coiled-coil stalk (124 residues) which is finally followed by a transmembrane helix at the C-terminus (Pfam domain annotation DUF1640). The head domain is predicted to be made of two α -helices connected by a short loop. One prominent conserved “FDT” (Phe-Asp-Thr) motif is identified at the N-terminus of the head domain. The coiled-coil stalks of MCUR1 and CCDC90B are identical in length and share the same repeat periodicities. However, MCUR1, in contrast, harbors an additional N-terminal domain comprised of nearly 100 residues. Secondary structure estimation using Quick2D (Zimmermann et al., 2017) predicts minor secondary structure propensity in this region. Furthermore, it is predicted to be disordered in consensus by multiple predictors: PONDR VL-XT, PONDR VSL2b, PrDOS, PV2, and Espritz-N (data collected from D²P² database) (Oates et al., 2013). The predicted disordered region encompasses this additional N-terminal region (Fig. 3.5A). Interestingly, PONDR VL-XT identifies some probability of order within this disordered segment (Fig. 3.5B). This usually indicates the presence of one or more small regions with a tendency for disorder-to-order transition upon some environmental change, such as binding to an interacting partner. A similar analysis of this region using the program ANCHOR (Dosztányi et al., 2009) also identified a *Molecular Recognition Feature* (MoRF) within residues 67-76 (corresponding sequence “SPLLLLLLVP”) of MCUR1. Such regions potentially serve as the initial binding sites for molecular recognition and undergo disorder-to-order transitions upon binding to proteins or ligands. Previous immunoprecipitation studies have identified that both MCU (mitochondrial calcium uniporter) (Lee et al., 2015; Tomar et al., 2016) and PPIF (peptidyl-prolyl-isomerase F or cyclophilin D) (Chaudhuri et al., 2016) can bind to MCUR1. I suspect that these interaction partners or some other unidentified factors can mediate this disorder-to-order transition of the long disordered chain of MCUR1 upon binding under certain conditions in the mitochondria.

Quite interestingly, I found that this disordered region is highly conserved among mammals, but differs significantly from the yeast, fungi and plant proteins.

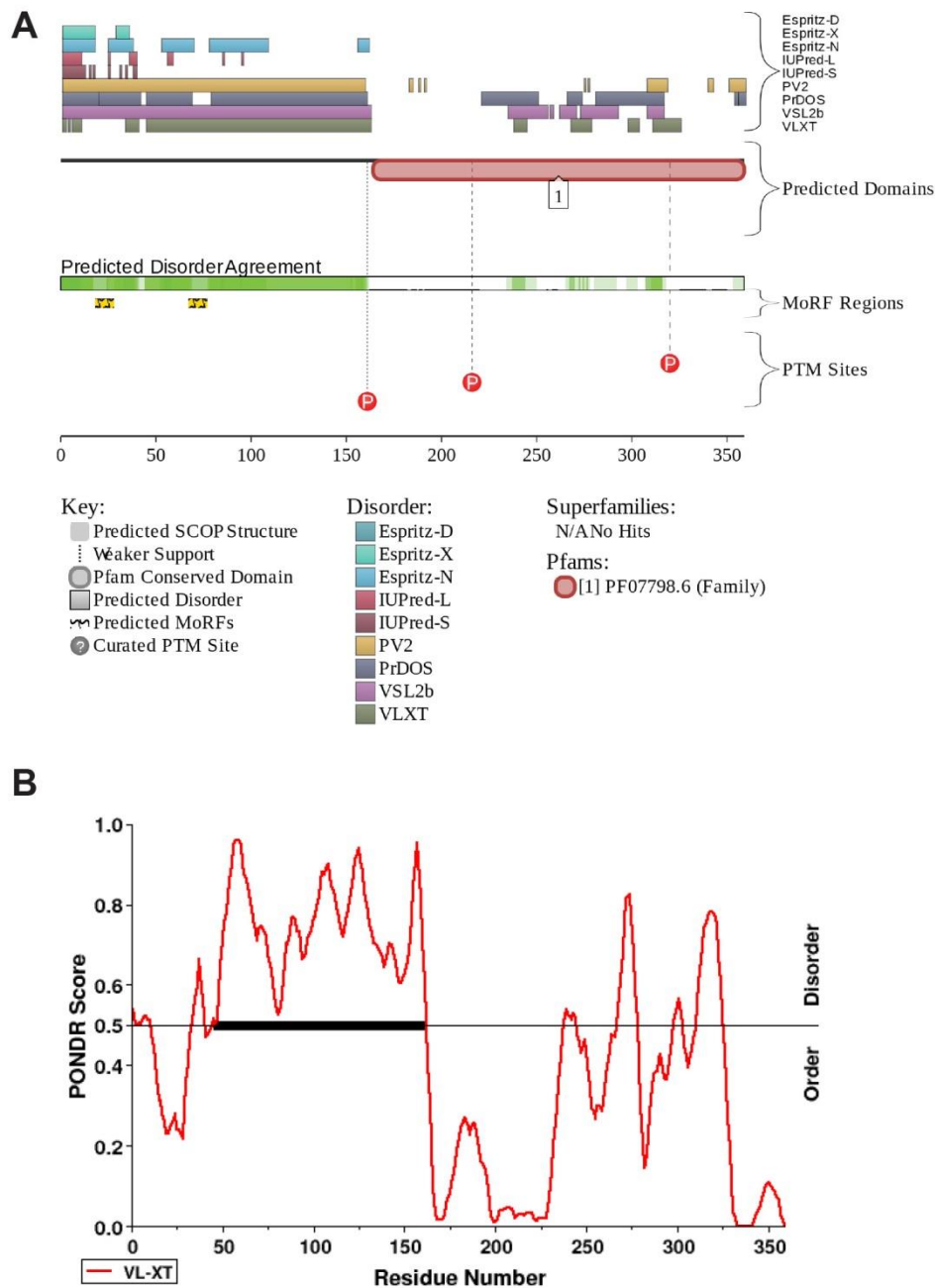


Figure 3.5: **Disorder prediction in human MCUR1. (A)** Comparative MCUR1 native disorder prediction using multiple predictors: Espritz-D, Espritz-X (green), Espritz-N (blue), IUPred-L (magenta), IUPred-S (brown), PV2 (yellow), PrDOS (dark blue), VSL2b (purple) and VLXT (olive green). The pre-computed data was collected from Database of Disordered Protein Predictions (D^2P^2) (Oates et al., 2013). DUF1640 Pfam domain predicted in the C-terminal region of MCUR1 which comprises the conserved head, neck and coiled-coil stalk. The predicted disordered region with 50% agreement among the predictors is marked in green. Predicted molecular recognition features (MoRF regions) and putative phosphorylation sites are marked below. **(B)** Disorder prediction in MCUR1 by PONDR VL-XT. Scores above the threshold score of 0.5 indicate disordered region. Black horizontal bar shows the region with a high-propensity to undergo disorder-to-order transition.

3.2.1.2 Crystal structure of human mempromCC protein CCDC90B

To realize our primary objective of structural characterization of the conserved head domain of mempromCC family including the β -layer neck, we selected homologs ranging from bacteria, yeasts to human. I initially focused on human paralogous proteins MCUR1 and CCDC90B. As it was notoriously difficult to purify sufficient amounts of soluble protein, I designed multiple constructs of the head-neck segments, variable in length and type of fusion, and tested their expression in various *E. coli* strains. Finally, I succeeded to solubly express CCDC90B₄₃₋₁₂₅, a fragment comprising the head-neck segment and the first 21 residues of the coiled-coil stalk, in *E. coli* ArcticExpress (DE3). The coiled-coil fragment of the construct was fused in-register to a GCN4-N16V adaptor, which forms a mixture of dimers and trimers in solution (Harbury et al., 1993). In the past, our lab have frequently used such in-register fusions of GCN4 variants to stabilize coiled-coil fragments for structural characterization (Deiss et al., 2014; Hernandez Alvarez et al., 2008).

Far-UV CD spectroscopy showed that CCDC90B₄₃₋₁₂₅-GCN4-N16V adopts a primarily α -helical structure with characteristic spectral minima at 208 nm and 220 nm (Fig. 3.6A). Upon thermal melting, the two domains of the fusion protein unfold independently in two steps. CCDC90B₄₃₋₁₂₅ fragment melts apparently with a melting temperature T_m of 71°C, whereas unfolding of GCN4-N16V is known to take place at higher temperatures with a melting temperature $T_m > 95^\circ\text{C}$ (Knappenberger et al., 2002).

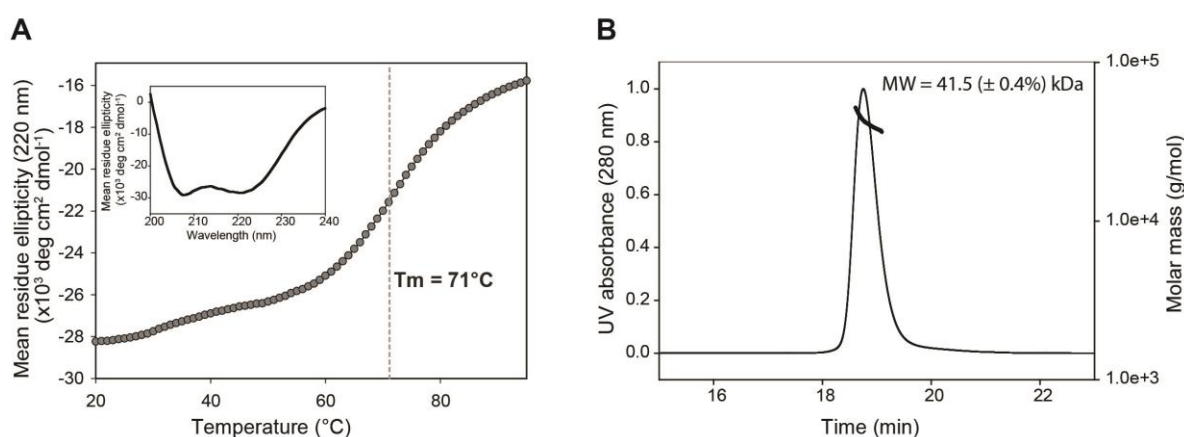


Figure 3.6: **Biophysical characterization of CCDC90B₄₃₋₁₂₅-GCN4 N16V.** (A) Melting curve of CCDC90B₄₃₋₁₂₅-GCN4 N16V monitored at 220 nm using far-UV CD spectroscopy. The CCDC90B₄₃₋₁₂₅ fragment unfolds with a melting temperature T_m of 71°C, Thermal unfolding of the GCN4 N16V segment is $T_m > 95^\circ\text{C}$ and not visible in the applied temperature range. The inset shows a single far-UV CD spectrum. (B) SEC-MALS data of CCDC90B₄₃₋₁₂₅-GCN4 N16V in 20 mM Tris pH 7.5, 150 mM NaCl. The calculated molecular mass corresponding to a trimer is indicated.

The trimeric oligomerization state of the protein was verified by SEC-MALS. Loaded on an HPLC column (AdvanceBio SEC 130Å, Agilent Technologies), the protein eluted as a single major peak with an apparent molecular weight of 41.5 kDa, which is equivalent to three times the theoretical molecular mass of 13.5 kDa (Fig. 3.6B).

For the obtained crystals, we collected a dataset to a resolution of 2.1 Å which we could solve via molecular replacement using trimeric coiled-coil segments as search models (work done by Dr. Marcus Hartmann). Crystal structure of CCDC90B₄₃₋₁₂₅ fused to GCN4-N16V shows an elongated parallel trimer and displays a head-neck-stalk architecture with a primarily α -helical head, one β -layer neck connecting the N-terminal helix of the head to the coiled-coil stalk (Fig. 3.7B). The solved structure starts at residue D62, part of the conserved FDT motif, at the beginning of the head domain. The first 19 residues of CCDC90B, corresponding to residues 43–61, are not resolved in the electron density. This can be attributed to the flexibility of the N-terminal region in crystallographic state, as the SDS-PAGE analysis of crystals show integrity of the protein. Helices α 1 (residues 62-72) and α 2 (residues 77-101) of the head domain are spaced by a short loop four residues in length. Overall, the head domain is structurally divided in two parts: an N-terminal six-helix bundle that extends into a trimeric coiled-coil segment, which then directly connects to the β -layer neck. In the antiparallel six-helix bundle, α 1-helices fold back and pack against the two α 2-helices of the same and one neighboring chain.

Loop L1 features a conserved hydrophobic residue phenylalanine (F75) which is positioned like a cap on top of the head shielding the hydrophobic core of the helix bundle (Fig. 3.7C). In other eukaryotic and proteobacterial mempromCC homologs, this position is occupied by a large hydrophobic residue, most commonly phenylalanine, leucine or methionine. F75 is preceded by glycine and both residues are conserved in more than 50% of mempromCC homologs. At the transition from loop L1 to α 2, the structure is stabilized by a closed network of intra- and inter-chain hydrogen bonds formed by residue Q79 with the backbone amide and carboxyl moieties of neighboring residues D76 and G74 respectively (Fig. 3.7C).

The C-terminal part of the α 2-helices of each chain (residues 87-101), which is not in contact with the α 1-helices, is composed of 15 residues and accommodated as a pentadecad over four helical turns in a right-handed coiled coil. The β -layer neck comprises the first three residues of the motif MVTQAQ (residues 102-107) with the central β -layer residue V103 forming the

inter-chain backbone hydrogen bonds and coordinating a water molecule in the center (Fig. 3.7D). The β -layer is followed by a segment of the natural stalk comprising of a hendecad and a heptad, which shows a transition from a slightly right-handed to a left-handed coiled coil and merges continuously into the GCN4-N16V adaptor.

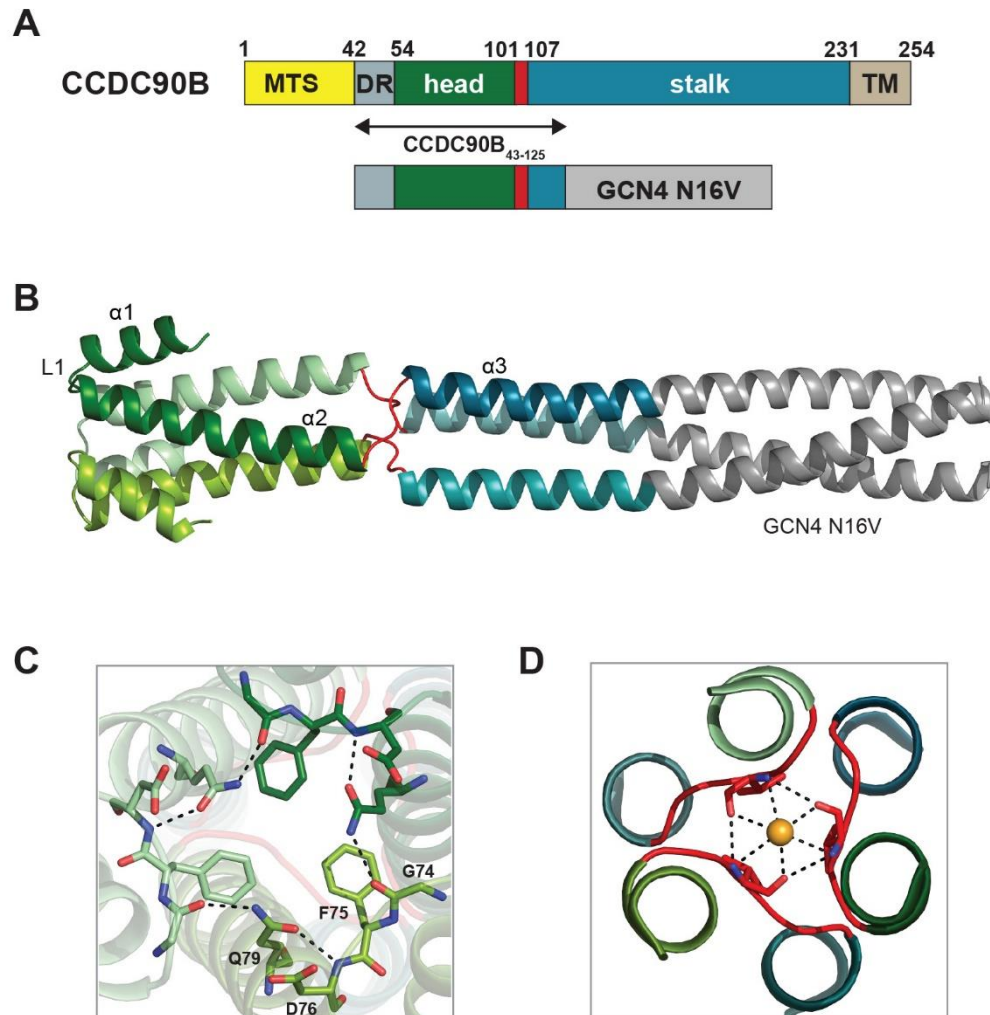


Figure 3.7: **Crystal structure of human CCDC90B.** (A) Domain organization of CCDC90B (NP_068597.2) and CCDC90B₄₃₋₁₂₅-GCN4 N16V fusion construct. Single domains include mitochondrial target signal (MTS), disordered region (DR), head domain (head), coiled-coil stalk (stalk) and a transmembrane helix (TM). β -layer neck is shown in red. Residue range is indicated at the top. (B) Cartoon representation of the crystal structure of CCDC90B₄₃₋₁₂₅-GCN4 N16V. Structure starts at residue D62. Domains are colored according to the schematic representation in (A), with head domain (helices $\alpha 1$ and $\alpha 2$ connected by a loop L1) in green, β -layer neck in red, native stalk (helix $\alpha 3$) in blue and GCN4-N16V adaptor in grey. Monomers are colored in different shades. (C) Detailed top view of the head domain showing conserved F75 along with stabilizing closed interaction network formed by residues G74, D76 and Q79. (D) Close-up of β -layer neck, highlighting the hydrogen bonding network involving central β -layer residues (V103) and a water molecule.

3.2.1.3 Structural characterization of MCUR1

Human paralogs CCDC90B and MCUR1 show a high sequence similarity of 72% in the sequences covering the head, stalk and membrane anchor domains. Individual domains of the paralogs are identical in length, with the stalks even sharing the same repeat pattern (Fig. 3.9A). As already mentioned in Sect. 3.2.1.1, the N-terminal extensions preceding the head domains differ remarkably in length and sequence, representing the most distinctive feature of the paralogs (Fig. 3.9A). Based on secondary structure prediction, they comprise of about 60 residues in CCDC90B and 160 residues in MCUR1, which includes a mitochondrial signal peptide followed by an intrinsically disordered segment.

For structural characterization of MCUR1, I designed multiple expression constructs comprised of the conserved head domain, but excluding the N-terminal disordered region. Similar to CCDC90B, the obtained MCUR1 constructs were either expressed at low levels or impaired in solubility. A fusion construct of MCUR1₁₆₀₋₂₃₀ with GCN4-N16V, analogous to the successfully utilized CCDC90B₄₃₋₁₂₅ construct, was found to be well-expressed in the insoluble fraction, but not amenable to refolding. The same fragment lacking a C-terminal fusion, MCUR1₁₆₀₋₂₃₀ (Fig. 3.9A), could be refolded successfully in physiological buffer, but did not yield any diffracting crystals.

Given the lack of diffracting crystals, we conducted preliminary experiments to assess the suitability of MCUR1₁₆₀₋₂₃₀ for solution NMR spectroscopy. At first glance, the 1D ¹H spectrum showed adequate dispersion and diffusion coefficients consistent with an oligomer, most likely the expected trimer. We then

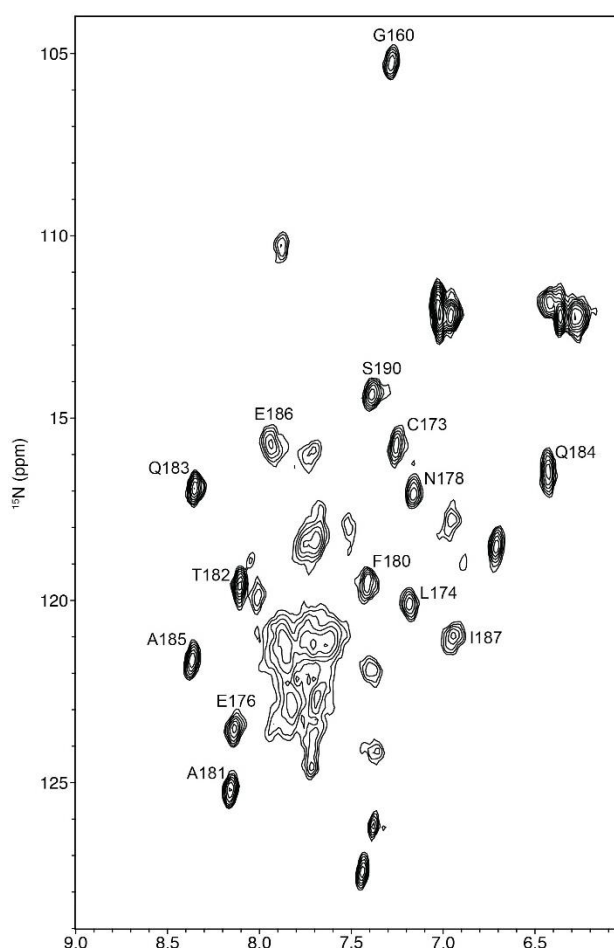


Figure 3.8: NMR 1D ¹H spectrum of MCUR1₁₆₀₋₂₃₀. Assigned residues are labelled.

prepared isotope-labelled samples and acquired spectra for sequential assignment. In the ^{15}N -HSQC, a stretch of residues corresponding to the core of the head domain (C173-S190) showed sharp signals with good dispersion, indicative of folded environment and protection of amide protons from solvent (Fig. 3.8). Other residues were both much broader with low dispersion, could not be definitively assigned or were not observed. Under these circumstances, it was not possible to proceed to a high-resolution structure.

3.2.1.4 Homology model of MCUR1

Facing the difficulties of structural characterization, I designed a homology model of MCUR1 with MODELLER (Sali et al., 1995) using CCDC90B₄₃₋₁₂₅ crystal structure as the template. Lacking the flexible N-terminal extension preceding the head, the MCUR1 model mirrors CCDC90B in length as well as in domain architecture (Fig. 3.9B). The electrostatic surface charge distribution of both paralogous proteins show significant differences in their head domains. CCDC90B₄₃₋₁₂₅ surface displays a prominent acidic patch localized at the loop of the head α -hairpin with residues D69, E71, D76 and E81 contributing the bulk of negative charge. (Fig. 3.9E). While the corresponding glutamate residues E176 and E186 in wild-type MCUR1 are conserved, the equivalent aspartates have been replaced by hydrophobic residues L174 and A181 which therefore projects the top of head as apparently neutral (Fig. 3.9D). The small negative patches in MCUR1 are contributed primarily by α 1-helix residues E176 and D177 of each monomer.

3.2.1.5 *In vitro* Ca²⁺ binding to MCUR1 and CCDC90B

Given the regulatory function of MCUR1 in mitochondrial Ca²⁺ transport, I suspected that this surface-charge polarity may contribute to potential Ca²⁺ binding. Therefore, I tested the effect of Ca²⁺ on oligomer formation and stability by SEC-MALS. MCUR1₁₆₀₋₂₃₀ eluted as a single peak of a calculated molecular mass of 22.6 kDa, corresponding to a trimer in solution (theoretical mass of monomer is 8.32 kDa) (Fig. 3.10A). Only a minor fraction was present as aggregate in the void volume. So both, MCUR1 and CCDC90B, exist as trimers in solution. Addition of 5 mM Ca²⁺ did not have any visible effect on oligomerization of MCUR1 or CCDC90B as the proteins eluted at the same molecular weight.

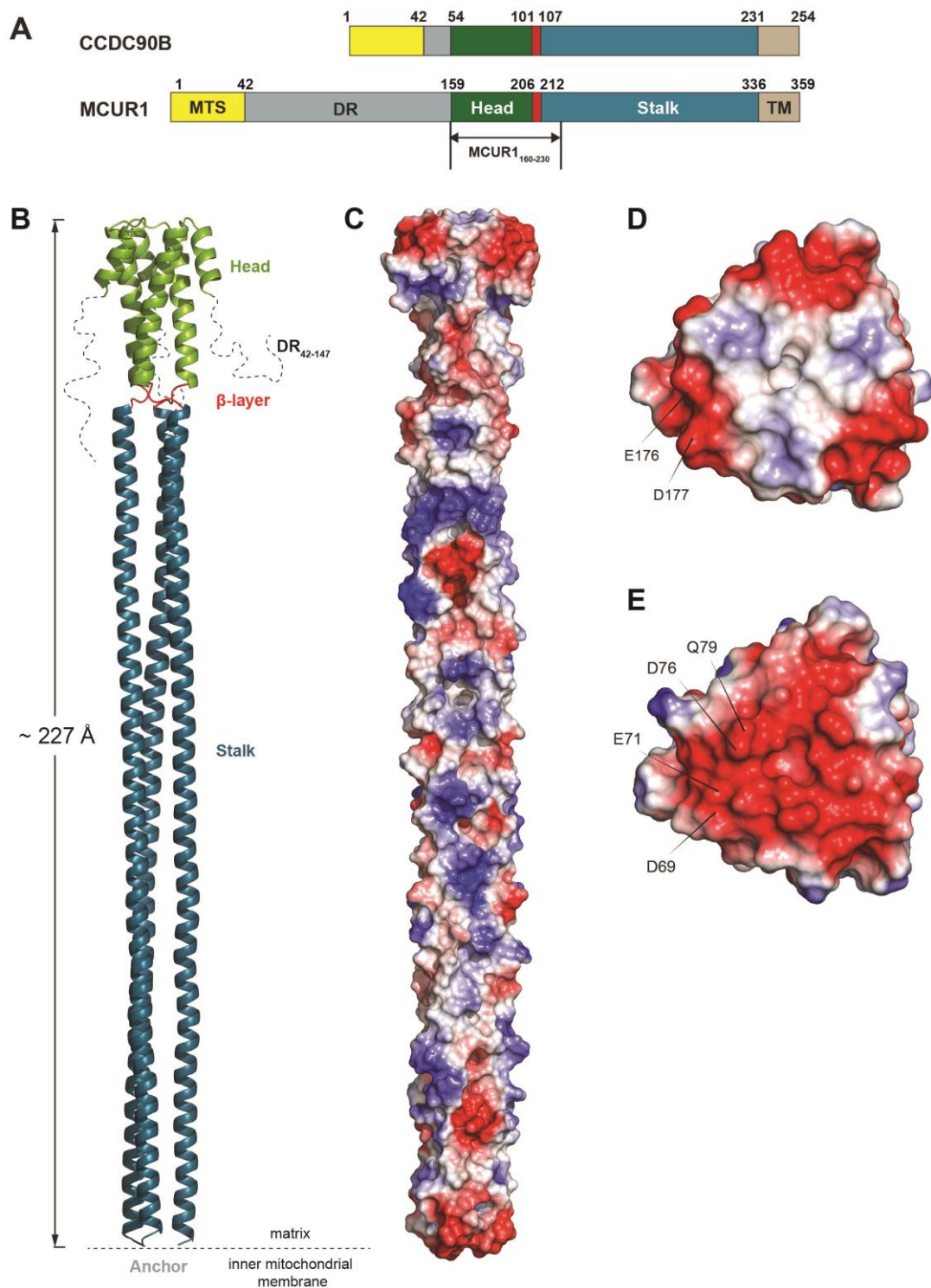


Figure 3.9: **Structural model of human MCUR1.** (A) Comparative scheme showing the domain organization of human paralogs CCDC90B and MCUR1. MCUR1₁₆₀₋₂₃₀ fragment used for biophysical characterization is indicated. (B) Homology model of MCUR1 spanning the head-neck-stalk region. The N-terminal disordered region is marked as dotted lines. MCUR1 is C-terminally anchored to the inner mitochondrial membrane (IMM) by a TM helix (not shown) and projects into the matrix. (C) Electrostatic surface charge distribution of MCUR1. Gradient is marked from acidic (red), through neutral (white) to basic (blue). (D) and (E) Top view of the head domains of MCUR1 and CCDC90B respectively, showing charge distribution. Negatively charged residues are labelled.

I then utilized Microscale Thermophoresis (MST) to examine any structural changes in MCUR1₁₆₀₋₂₃₀ and CCDC90B₄₃₋₁₂₅ in response to CaCl₂ titration. MCUR1₁₆₀₋₂₃₀ and CCDC90B₄₃₋₁₂₅ (with and w/o C-terminal GCN4 fusion) were fluorescently labelled with red-NHS Alexa Fluor 647 dye by covalent crosslinking to primary amines. Labelled-proteins were titrated with CaCl₂ in the 1 μ M – 40 mM concentration range. Whereas both CCDC90B constructs do not show any appreciable binding to Ca²⁺, the curve monitored for MCUR1₁₆₀₋₂₃₀ clearly showed Ca²⁺ binding (Fig. 3.10B). Next, I determined the affinity constants of the MCUR1 head for Ca²⁺ and Mg²⁺ with K_d ~ 0.6 mM for CaCl₂ and ~ 5.0 mM for MgCl₂ (Fig. 3.10C). This indicates that Ca²⁺ specifically binds to MCUR1 head domain with a nearly 10-fold higher affinity compared to Mg²⁺. At concentrations higher than 5 mM CaCl₂, MCUR1 tends to precipitate slightly. This effect was, however, observed to be less pronounced in the presence of Mg²⁺.

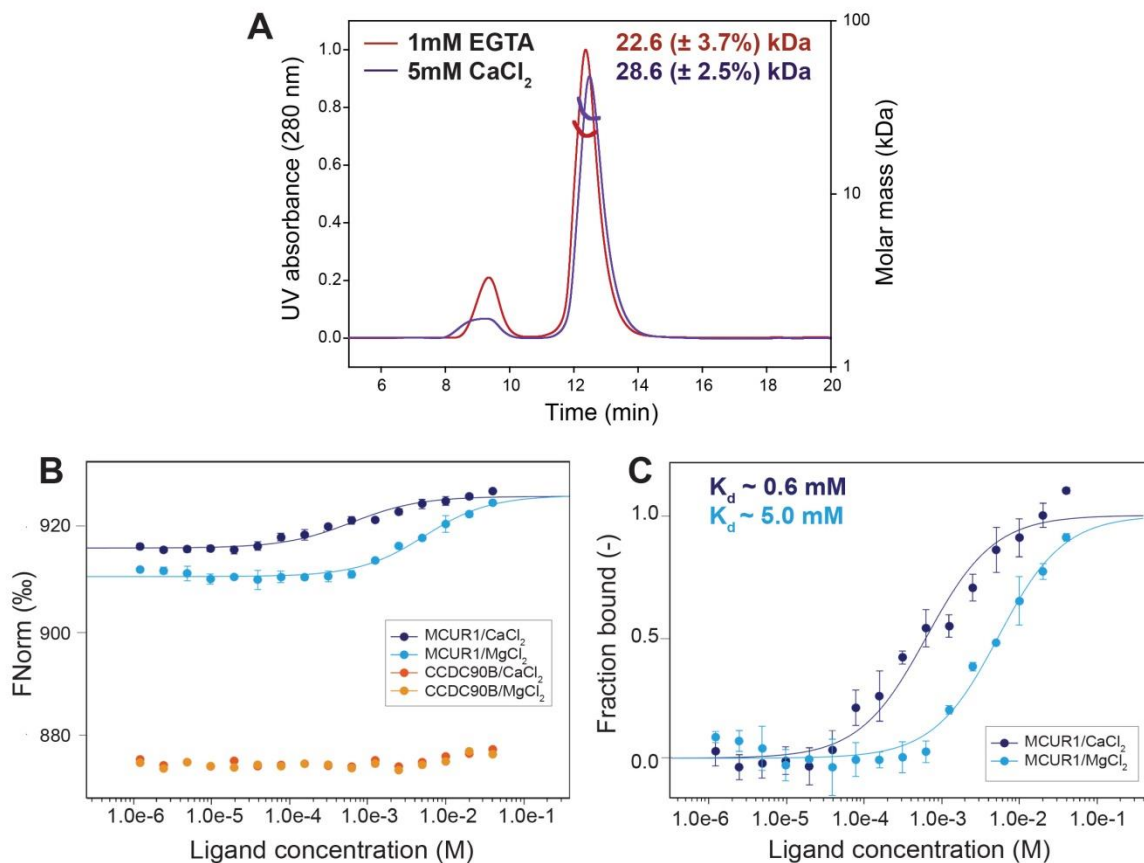


Figure 3.10: **Low-affinity Ca²⁺ binding to MCUR1.** (A) SEC-MALS plot and calculated molecular mass of MCUR1₁₆₀₋₂₃₀ in the presence of 1 mM EGTA (red) and 5 mM CaCl₂ (purple). (B) and (C) MST titration curves for Ca²⁺ and Mg²⁺ binding to MCUR1₁₆₀₋₂₃₀ (dark and light blue) and CCDC90B₄₃₋₁₂₅ (dark and light orange) used to determine the dissociation constants (K_d). Data points are mean of three replicates, shown with standard error.

3.2.1.6 The head domain of MCUR1, but not CCDC90B, can form β -amyloid fibrils

Next, I assessed the effect of Ca^{2+} on the secondary structure and stability of MCUR1 using far-UV CD spectroscopy. MCUR1₁₆₀₋₂₃₀ displayed a typical α -helical spectra with characteristic minima at 208 nm and 220 nm (Fig. 3.12A). Upon heating, the protein unfolds with a melting temperature T_m of 71 °C (Fig. 3.12B). In contrast to CCDC90B₄₃₋₁₂₅ (Fig. 3.13), single CD spectra of MCUR1₁₆₀₋₂₃₀ monitored at increasing temperatures show that thermal unfolding of α -helical structure was accompanied by irreversible formation of soluble β -like structure, which becomes visible as an emerging sharp signal at 216 nm, typical for β -amyloid formation (Fig. 3.14). At this point, I suspected that the translucent gel-like precipitates formed by incubating protein at room temperature over the course of time or during protein concentration could be amyloid fibrils.

What are β -amyloid fibers?

The term “amyloid” refers to aggregates of proteins arranged in a fibrillar form. They are characterized by the formation of long unbranched fibers, in which multiple parallel β -stranded chains of that protein arrange in an orientation perpendicular to the fiber axis. Amyloids have been associated with more than 50 human diseases and neurodegenerative disorders such as Alzheimer’s, Parkinson’s, Diabetes mellitus type 2, Huntington’s disease, Creutzfeldt-Jakob disease (CJD), and Rheumatoid arthritis (Knowles et al., 2014).

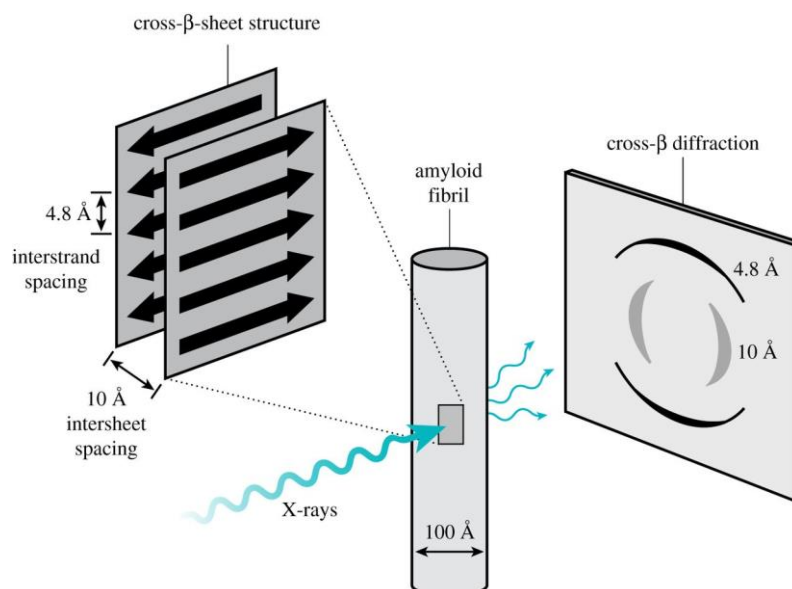


Figure 3.11: **Characteristics of cross- β amyloid fibrils.** Schematic representation of the cross- β structure of amyloid fibrils. Characteristic X-ray diffraction arcs seen at 4.8 Å and 10 Å corresponding to the interstrand and intersheet spacing respectively. Reproduced from Dharmadana et al., 2017.

β -Amyloids have a distinctive cross- β structure. The “gold standard” test to verify the cross- β quaternary structure is X-ray diffraction pattern of the amyloid fibrils. Two characteristic diffraction reflection arcs at 4.8 Å and ~10 Å, corresponding to the inter-strand and sheet stacking distances respectively, are observed for well-oriented fibrils (else they blur into circular rings) (Fig. 3.11). The neighboring chains pack tightly, thereby excluding water from the core. Opposing strands are slightly offset from each other which allows their side-chains to interdigitate and create a zipper interface.

3.2.1.7 Effect of Ca^{2+} on MCUR1 amyloid fibril formation

Examining the effect of Ca^{2+} addition on the formation of β -amyloid-like fibrils, we recorded single CD spectra and melting curves at 208/216/220/224 nm with 1 mM and 5 mM CaCl_2 concentrations (Fig. 3.12A). While Ca^{2+} does not affect the α -helical secondary structure of MCUR1₁₆₀₋₂₃₀, it strongly impairs its thermal stability (Fig. 3.12B). The melting temperature T_m of ~71°C measured in Ca^{2+} -free buffer is remarkably reduced to ~50°C and ~43°C in the presence of 1 mM and 5 mM CaCl_2 , respectively. White precipitates, composed of high molecular weight β -fibrillar aggregates, were seen after complete heat denaturation in direct proportion to the amount of calcium added. This explains the end low β -signal at 216 nm (Fig. 3.12C). This effect could be reversed upon addition of the Ca^{2+} chelator EGTA. At concentrations higher than 5 mM CaCl_2 , MCUR1 starts to aggregate and fall out of the solution. To check if it was a specific effect of Ca^{2+} , we performed CD measurements in the presence of Mg^{2+} . We obtained a near identical end result of thermal denaturation. With 5 mM MgCl_2 , MCUR1₁₆₀₋₂₃₀ melts with a T_m of ~44°C identical to Ca^{2+} addition. Both CCDC90B₄₃₋₁₂₅ (with and w/o GCN4 fusion) constructs, however, denature to form random coil structure with no change observed even upon adding Ca^{2+} (Fig. 3.13).

The nature of β -amyloid fibrils formed by MCUR1₁₆₀₋₂₃₀ was further analyzed by transmission electron microscopy (TEM). 120 μM protein in physiological pH buffer (20 mM Tris pH 7.6, 150 mM NaCl, 1 mM TCEP) was incubated at 25°C for 24 hours with or without the addition of Ca^{2+} . As observed by visual inspection, cloudy aggregates formed in the presence of Ca^{2+} within 3 hours of incubation, whereas no precipitates were visible in the Ca^{2+} -free samples even after 24 hours. TEM micrographs of MCUR1₁₆₀₋₂₃₀ fibrils show the presence of primarily short individual protofibrils in protein samples with no addition or addition of 1 mM EGTA (Fig. 3.14).

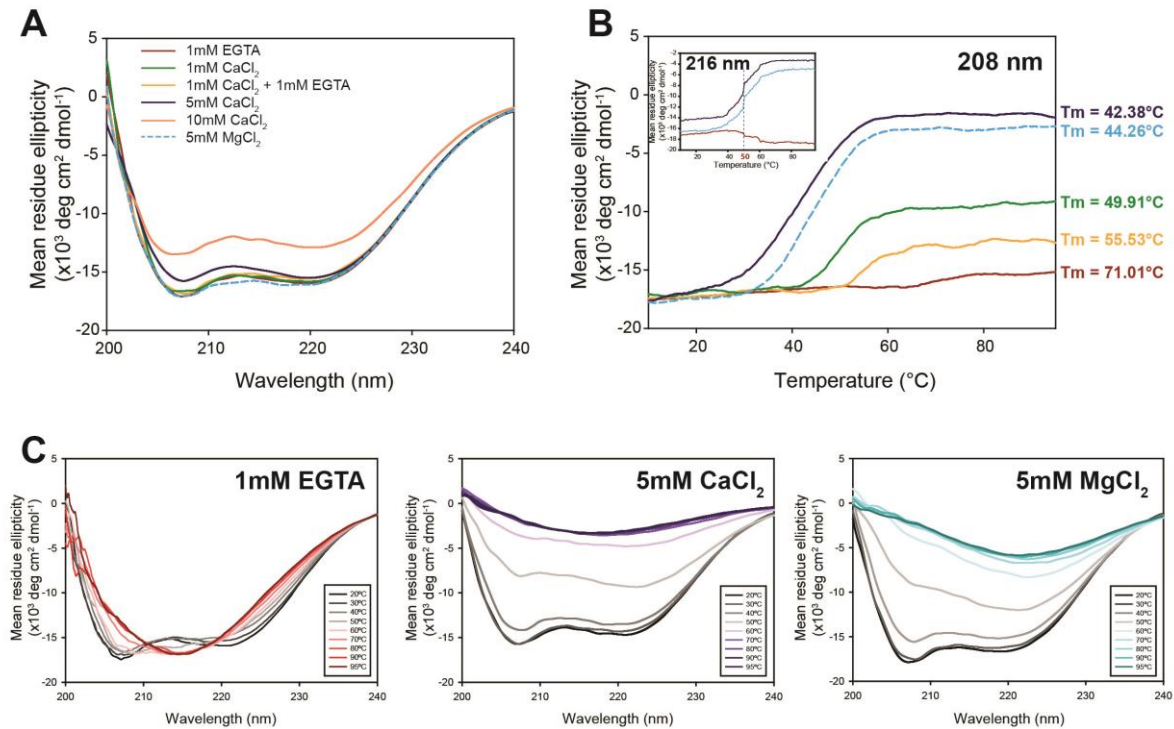


Figure 3.12: **MCUR1 is destabilized upon divalent cation binding.** (A) Far-UV CD spectra of MCUR1₁₆₀₋₂₃₀ at 20°C in the presence of EGTA, MgCl₂ and CaCl₂ at indicated concentrations. For the yellow curve (1 mM CaCl₂ + 1 mM EGTA), CaCl₂ was added to the sample and chelated by addition of equal molar amounts of EGTA immediately before measurement. (B) Thermal melting curves for MCUR1₁₆₀₋₂₃₀ in the presence of EGTA, MgCl₂ or CaCl₂ at 216 nm (inset) and 208 nm. Respective concentrations and the color code are as in panel (A). Calculated melting temperatures (T_m) are indicated. (C) Single far-UV CD spectra of MCUR1₁₆₀₋₂₃₀ showing loss of α -helical structure upon heating measured at different temperatures in the presence of 1 mM EGTA, 5 mM CaCl₂ and 5 mM MgCl₂.

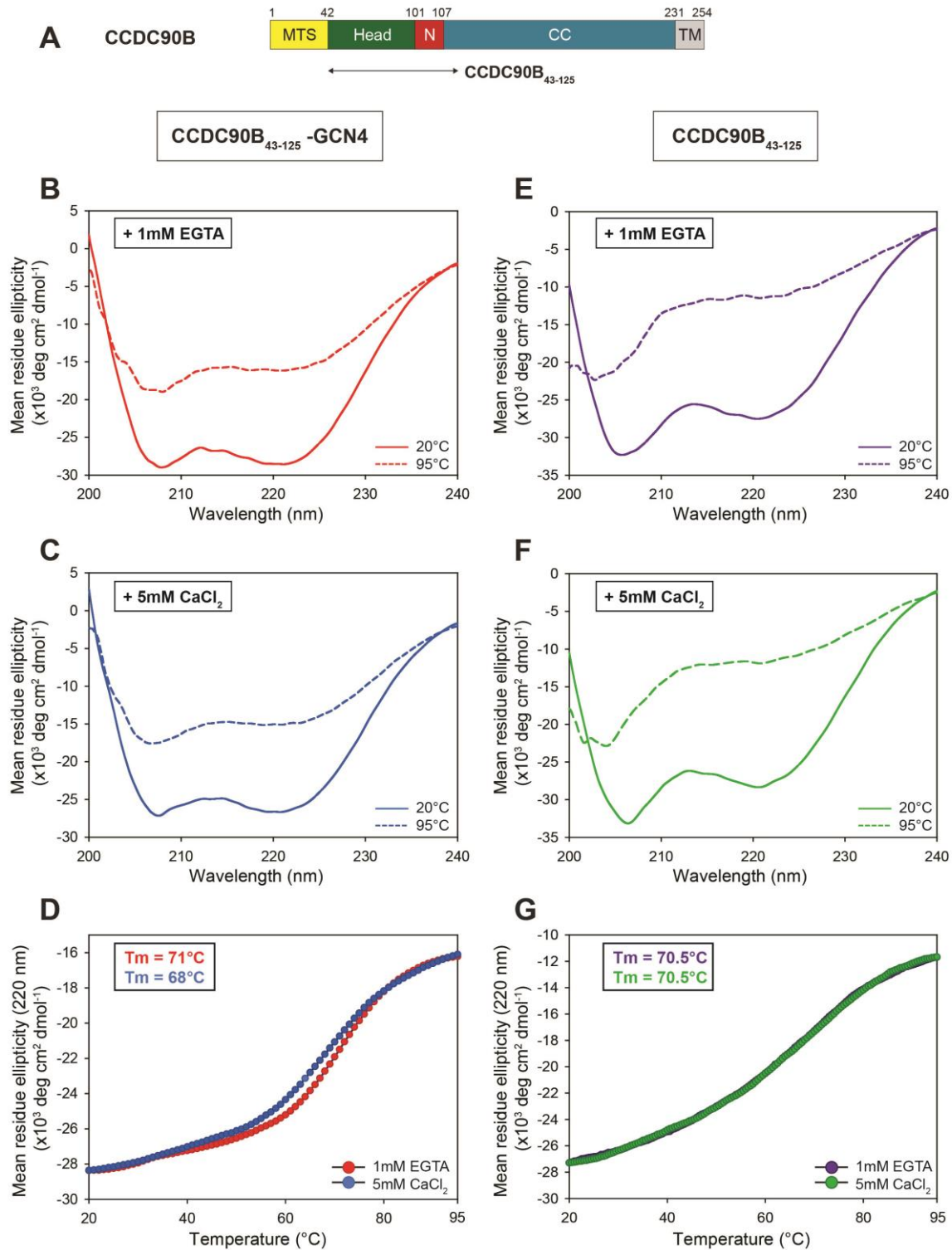


Figure 3.13: **Secondary structure analysis of CCDC90B₄₃₋₁₂₅-GCN4 N16V and CCDC90B₄₃₋₁₂₅.** (A) Domain organization of CCDC90B with residue range marked at the top. (B) and (C) Far-UV spectra of CCDC90B₄₃₋₁₂₅-GCN4 N16V, and (E) and (F) of CCDC90B₄₃₋₁₂₅ in the absence (1 mM EGTA) or presence of Ca²⁺ (5 mM CaCl₂) recorded at 20°C and 95°C. Panels (D) and (G) show melting curves of CCDC90B₄₃₋₁₂₅-GCN4 N16V and CCDC90B₄₃₋₁₂₅, respectively, recorded for each in the presence of 1 mM EGTA and 5 mM CaCl₂. Calculated melting temperatures (T_m) for each condition are indicated.

In sample containing 1 mM Ca^{2+} , already after 24 hours, we observed long extended curvilinear protofibrillar structures which started to transform into mature amyloid fibrils in 5 mM Ca^{2+} sample. Such a pronounced effect was not observed with Mg^{2+} . Thus, presence of calcium specifically seems to accelerate the process of fibrillar formation for MCUR1₁₆₀₋₂₃₀. Similar effect of Ca^{2+} has been previously reported for amyloid β -peptide ($\text{A}\beta$) involved in the toxic plaque formation in Alzheimer's disease (Isaacs et al., 2006). Our results show that increase in temperature or addition of Ca^{2+} accelerates the destabilization of MCUR1₁₆₀₋₂₃₀ which then assists the fast kinetics of early fibril formation.

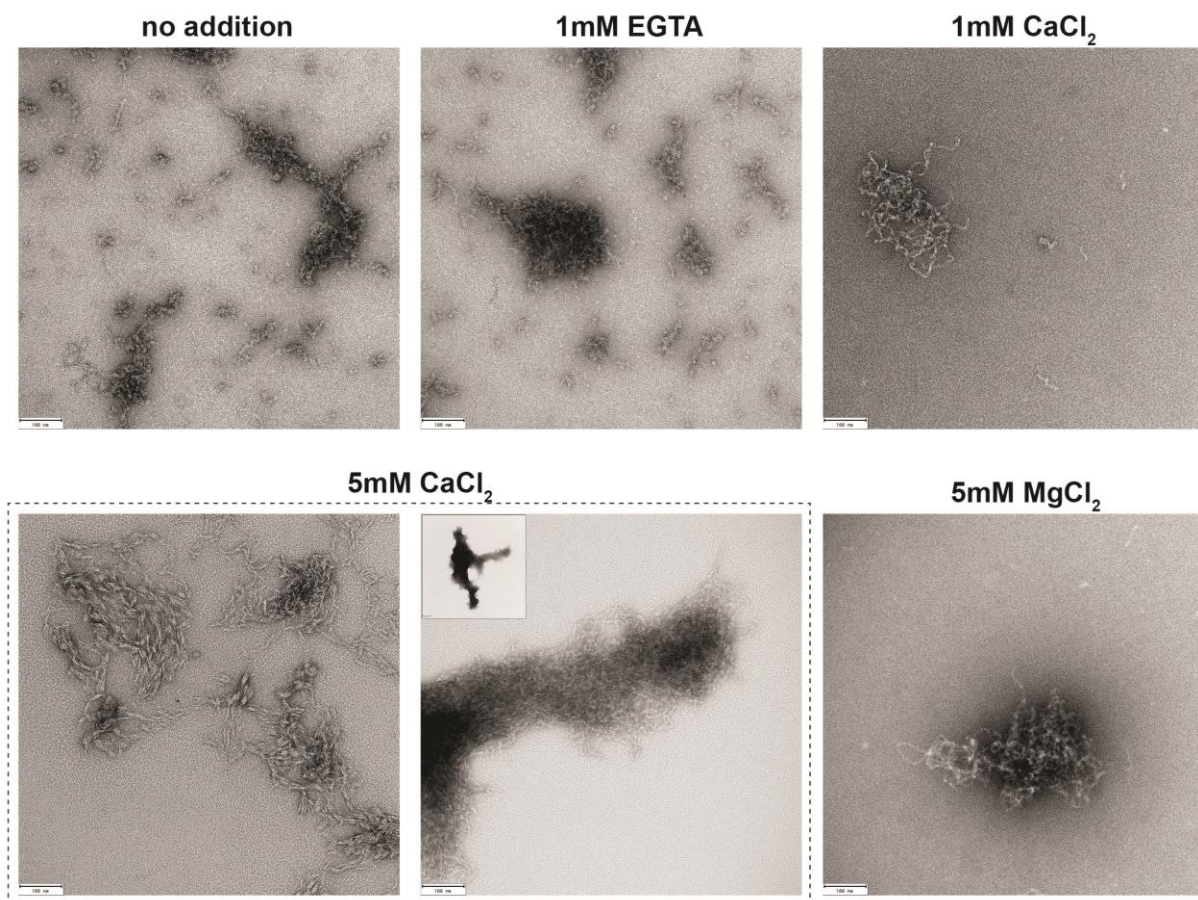


Figure 3.14: TEM micrographs showing β -amyloid fibril formation of MCUR1₁₆₀₋₂₃₀ in the presence of 1 mM EGTA, 1 mM and 5 mM CaCl_2 , and 5 mM MgCl_2 after incubation for 24 hours at 25°C. Scale bar is 100 nm (for inset, scale bar is 500 nm).

Many studies have reported that the protein sequence influences the formation of amyloid fibrils. In the partially unfolded state, a certain stretch of MCUR1 may possess β -formation propensity. Analyses of aggregation and amyloid propensity using multiple computational prediction tools, such as Amylpred2, TANGO, and ZipperDB (Fernandez-Escamilla et al., 2004; Tsois et al., 2013), identified a common amyloidogenic hotspot in the α 2-helix of the

MCUR1 head domain (residues 184-196). Figure 3.15 plots beta-aggregation score generated by TANGO for MCUR1₁₆₀₋₂₃₀ and CCDC90B₅₅₋₁₂₅ residues. The equivalent residues in CCDC90B head region, with slightly dissimilar sequence, show significantly lower β -aggregation propensities. These primary structural differences can be useful in understanding sequence-amyloid relationships in the future.

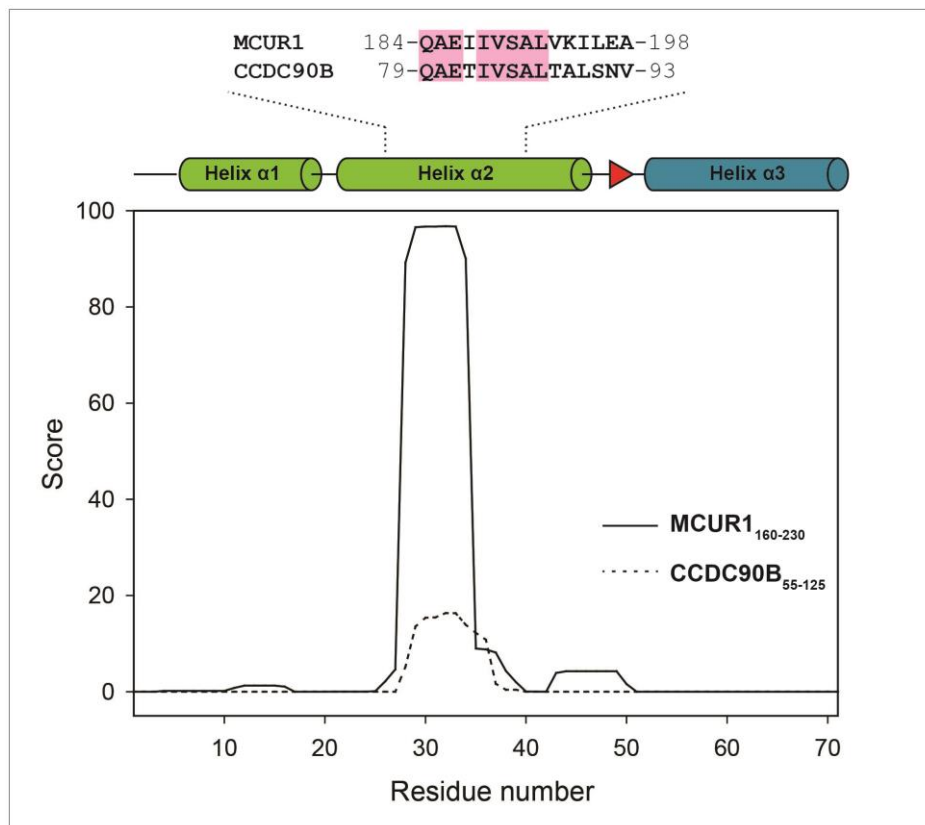


Figure 3.15: **Prediction of aggregation and amyloid propensity of MCUR1.** The prediction tool TANGO was used to analyze the β -aggregation propensity of MCUR1₁₆₀₋₂₃₀ (solid line) and CCDC90B₅₅₋₁₂₅ (dotted line). The diagram shows the calculated TANGO scores for each protein fragment plotted against the residue numbers. The upper panel aligns the identified amylogenic hotspot of MCUR1 with the corresponding region of CCDC90B in relation to secondary structure prediction data.

In a representative model of MCUR1 amyloid fibrils generated from ZipperDB (Fig. 3.16), we can visualize the tight packing of the hydrophobic side-chains of anti-parallel sheets creating a zipper-like interface, excluding the water molecules from the core.

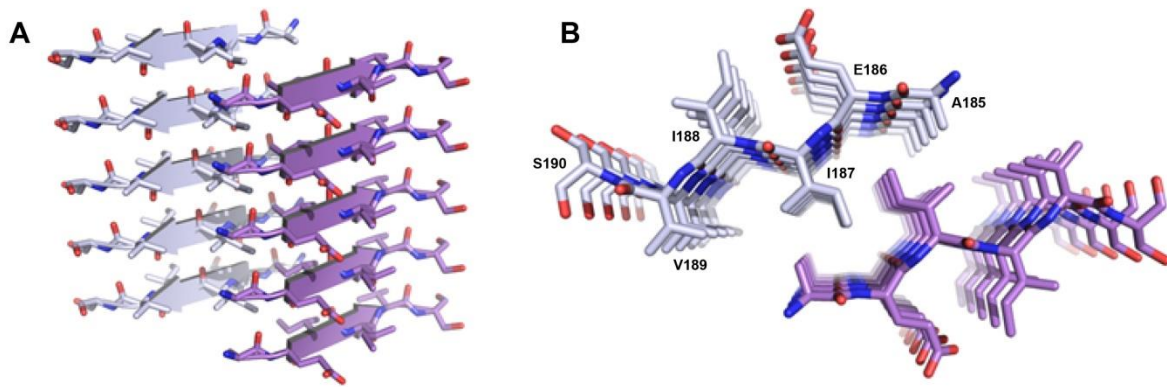


Figure 3.16: **Structural model of β -amyloid formation in MCUR1.** (A) side-view and (B) top-view cartoon representation of highest scoring β -fibrillation propensity hexad sequence A185-S190 in MCUR1 head domain. Calculations and model were generated using ZipperDB.

3.2.1.8 Does MCUR1 directly interact with MCU-NTD?

MCUR1 essentially regulates mitochondrial Ca^{2+} uptake upon interaction with MCU (Mallilankaraman et al., 2012a). Several studies mapping the binding region between both proteins identified MCUR1 to bind to the N-terminal domain of MCU (Lee et al., 2015; Tomar et al., 2016). However, there was discrepancy among the previous studies with regard to the MCUR1 domain(s) involved in this interaction. As we had much better knowledge of domain boundaries now, we addressed the question to identify the domain(s) of MCUR1 which mediates its binding to MCU (Fig. 3.17) (work done together with Dr. Birte Hernandez). We first identified the potential region involved with an immunoprecipitation experiment and later confirmed the same with an *in vitro* binding test using MST.

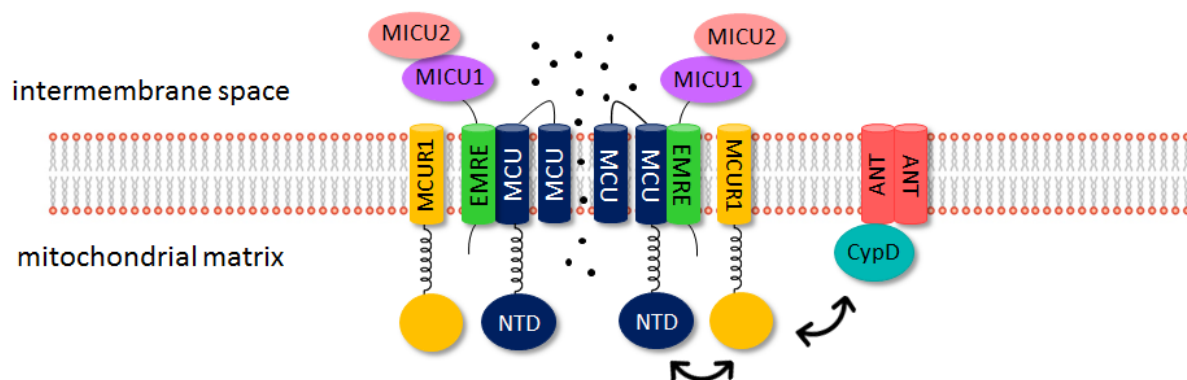


Figure 3.17: **MCU complex representation.** Components of the mitochondrial calcium uniporter (MCU) and the permeability transition pore (MPTP) complexes are shown with their submitochondrial localization and orientation. We aim to investigate the proposed interaction between the N-terminal domain of MCU with MCUR1, and MCUR1 with Cyclophilin D.

Co-immunoprecipitation of MCU with MCUR1

Considering the domain organization based on the structure of MCUR1, we designed variants with substitution and/or deletion of single/multiple domains including the N-terminal disordered region, the head, the β -layer neck and the stalk (Fig. 3.18A).

FLAG-tagged MCUR1 variants were co-expressed with full-length HA-tagged MCU in HEK293 cells, co-immunoprecipitated using anti-FLAG magnetic beads and subsequently analyzed on a western blot (Fig. 3.18B). As expected, full-length MCUR1 was able to pull-down MCU. Deletion of six residues “MVTKMQ” (residues 207-212) involved in the β -layer formation (MCUR1- $\Delta\beta$) displayed no significant effect on the binding of MCUR1 to MCU. This indicates dispensability of the β -layer neck for interaction with the uniporter. Similarly, for MCUR1-GCN4pII, where 91 residues of the stalk (residues 231-321) were replaced by thirteen heptads derived from the sequence of trimeric GCN4pII variant (Harbury et al., 1993), no significant reduction in MCU binding was observed. However, deletion of almost the entire coiled-coil stalk (MCUR1- Δ stalk, corresponding to residues 224-321) not only decreased the protein stability as seen from lower expression levels, but also clearly reduced MCU binding. We can safely conclude here that MCUR1 stalk length and stability, and not the primary coiled-coil sequence, are critical for interaction.

MCUR1- Δ DR lacking the disordered region (residues 43-159) was also expressed at low levels, but co-precipitated equivalent amounts of MCU. In contrast, MCUR1- Δ DR-Head, lacking the disordered region together with the conserved head domain, shows complete abrogation of binding to MCU, inferring that the head domain together with the preceding disordered region is necessary for binding of MCUR1 to MCU. A former study reported that CCDC90B could also be immunoprecipitated with MCU, but unlike MCUR1, it does not modulate the activity of the uniporter (Tomar et al., 2016). To identify if any difference exists in the binding of the head region of MCUR1 and CCDC90B to MCU, we substituted the conserved head in full-length MCUR1 construct with the corresponding region from paralogous CCDC90B. MCUR1-90Bhead also binds to MCU but with a lower affinity than full-length MCUR1.

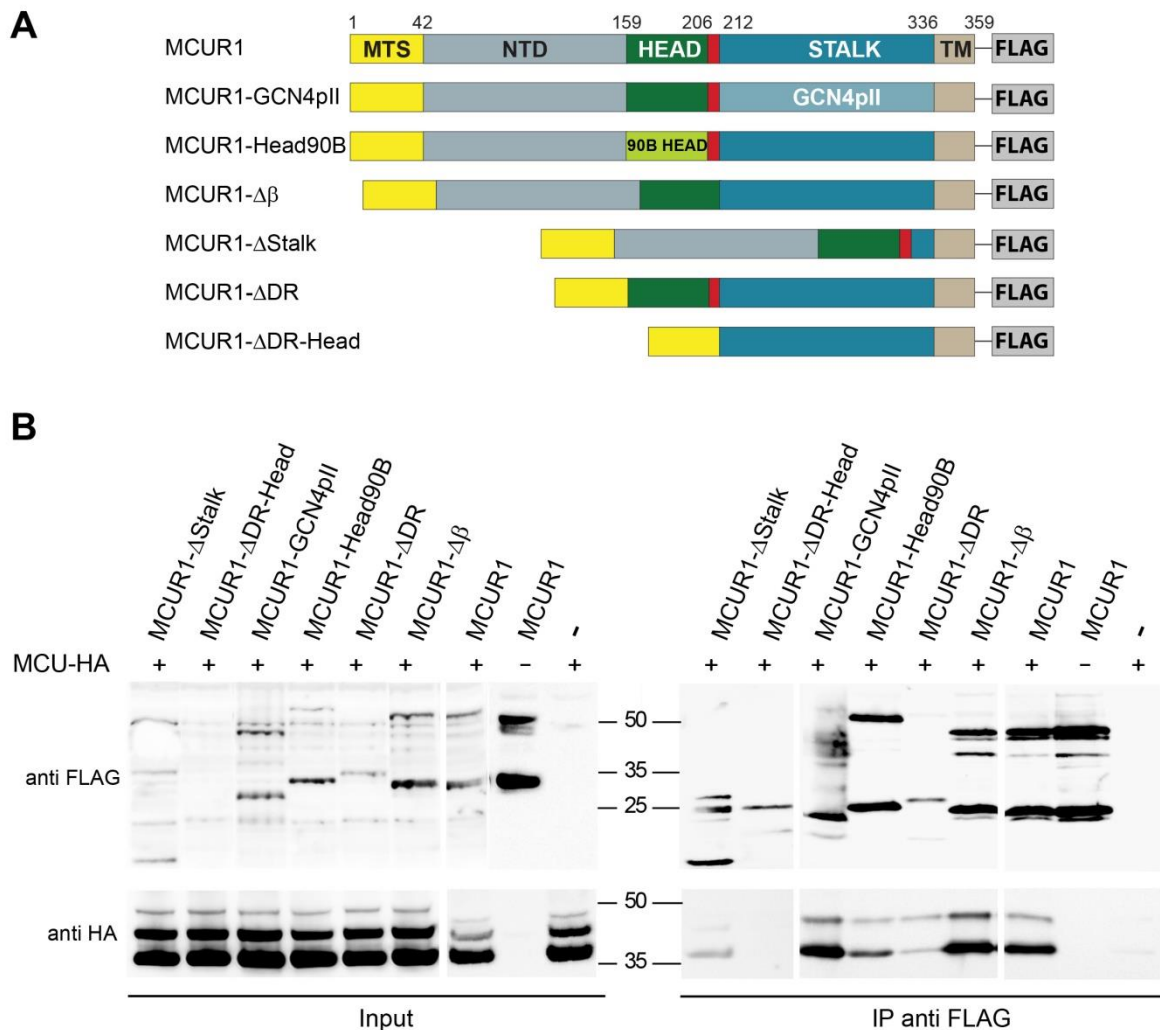


Figure 3.18: **MCUR1 head domain directly interacts with N-terminal domain of MCU.** (A) Schematic representation of MCUR1 constructs used for pull-down assay. (B) Co-immunoprecipitation (Co-IP) of MCU-HA with MCUR1-FLAG variants using anti-FLAG antibody. Input and IP samples were analyzed by western blotting using anti-FLAG and anti-HA antibodies.

***In vitro* binding assay using MST**

Based on our finding that the head domains of both paralogs mediate binding to MCU, the binding affinities of MCUR1 and CCDC90B heads were estimated *in vitro* using MST. MCUR1 has been identified to interact with the N-terminal soluble domain of MCU. Therefore, for the *in vitro* binding study, I expressed MCU₇₅₋₂₃₃ (residues 75-233) comprising the N-terminal domain of MCU including the succeeding CC1 coiled-coil domain (Fig. 3.19A). MCU₇₅₋₂₃₃ was expressed in inclusion bodies, so the protein was purified under denaturing conditions and refolded as described previously, in buffer containing 20 mM Tris (pH 8.8), 150 mM NaCl, 1 mM DTT (Lee et al., 2016). The purified protein was well-folded and displayed an α -helix/ β -strand mix spectrum in far-UV CD spectroscopic analysis (Fig. 3.19C). From

SDS-PAGE, we observed that MCU₇₅₋₂₃₃ exists as a ladder of monomer and stable higher-order oligomers, present in an equilibrium state (Fig. 3.19B). It shifts towards forming higher order oligomers upon incubation with $[Ca^{2+}] \geq 0.5\text{mM}$ as visualized on BN-PAGE (Fig. 3.19D). This was rather surprising as Ca^{2+} was shown to destabilize the N-terminal domain of MCU and to promote its disassembly previously (Lee et al., 2016). However, it must be noted that the construct used in this study additionally includes CC1, implicating this domain to mediate Ca^{2+} -dependent oligomerization.

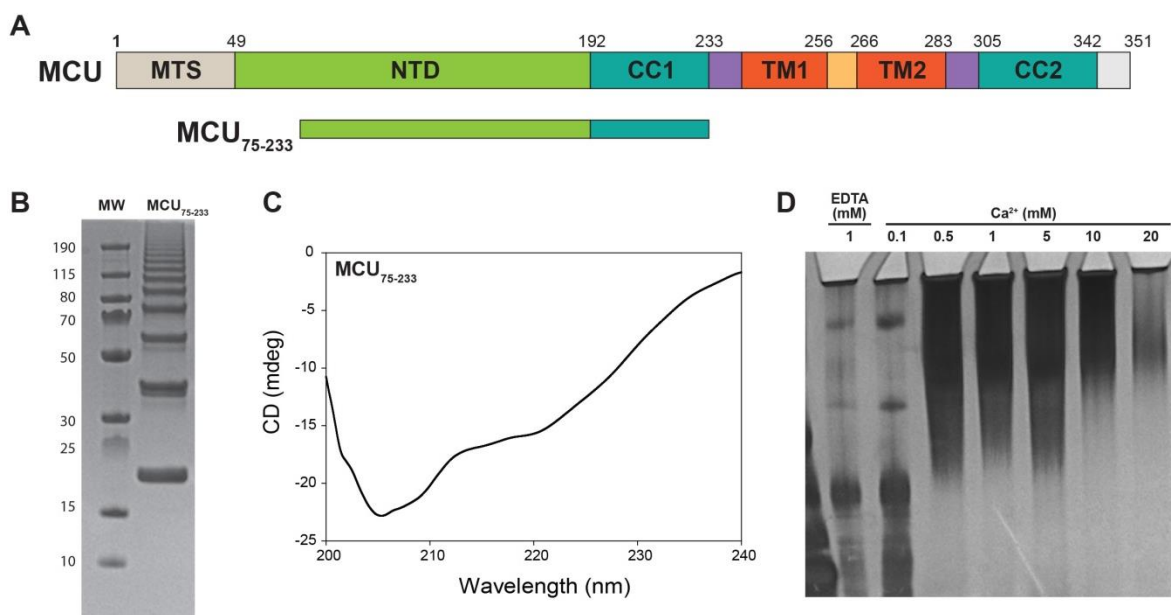


Figure 3.19: **Oligomer formation of MCU₇₅₋₂₃₃.** (A) Domain organization of human MCU (NP_612366.1) comprising of a mitochondrial target signal (MTS), an N-terminal soluble matrix domain (NTD), two transmembrane helices (TM1 and TM2), and two coiled-coil domains (CC1 and CC2). As indicated, MCU₇₅₋₂₃₃ includes residues 75-192 of the NTD succeeded by the CC1 domain. (B) SDS-PAGE analysis of MCU₇₅₋₂₃₃, which was purified under denaturing conditions and refolded. The protein forms a protein ladder of stable oligomers representing monomers, dimers, trimers, tetramers etc. according to their molecular weight in the gel. (C) Far-UV CD spectrum of MCU₇₅₋₂₃₃ measured at 20°C. (D) Blue-native PAGE analysis of oligomer and aggregate formation of MCU₇₅₋₂₃₃ in dependence of increasing concentrations of $CaCl_2$. Aggregation was observed to start at concentrations of $> 3\text{ mM } CaCl_2$.

In MST experiments, fluorescence-labelled MCU₇₅₋₂₃₃ was titrated against MCUR1₁₆₀₋₂₃₀ and CCDC90B₄₃₋₁₂₅. The dissociation constants of MCU binding to the head domains of MCUR1 and CCDC90B were estimated to be $12.71 \pm 3.51\ \mu\text{M}$ and $58.70 \pm 1.58\ \mu\text{M}$ respectively (Fig. 3.20A), which corroborates our pull-down result showing that MCUR1 binds with a comparatively higher affinity to MCU. Addition of 0.1 mM $CaCl_2$ decreased the K_d of MCUR1 binding to $7.49 \pm 1.59\ \mu\text{M}$ which did not change appreciably at 1 mM $CaCl_2$ ($7.76 \pm 1.55\ \mu\text{M}$)

(Fig. 3.20B). At higher $[Ca^{2+}]$, both MCUR1 and MCU purified proteins are prone to higher order oligomer formation and subsequent aggregation as seen from MST and SDS-PAGE, therefore measurements were unfeasible.

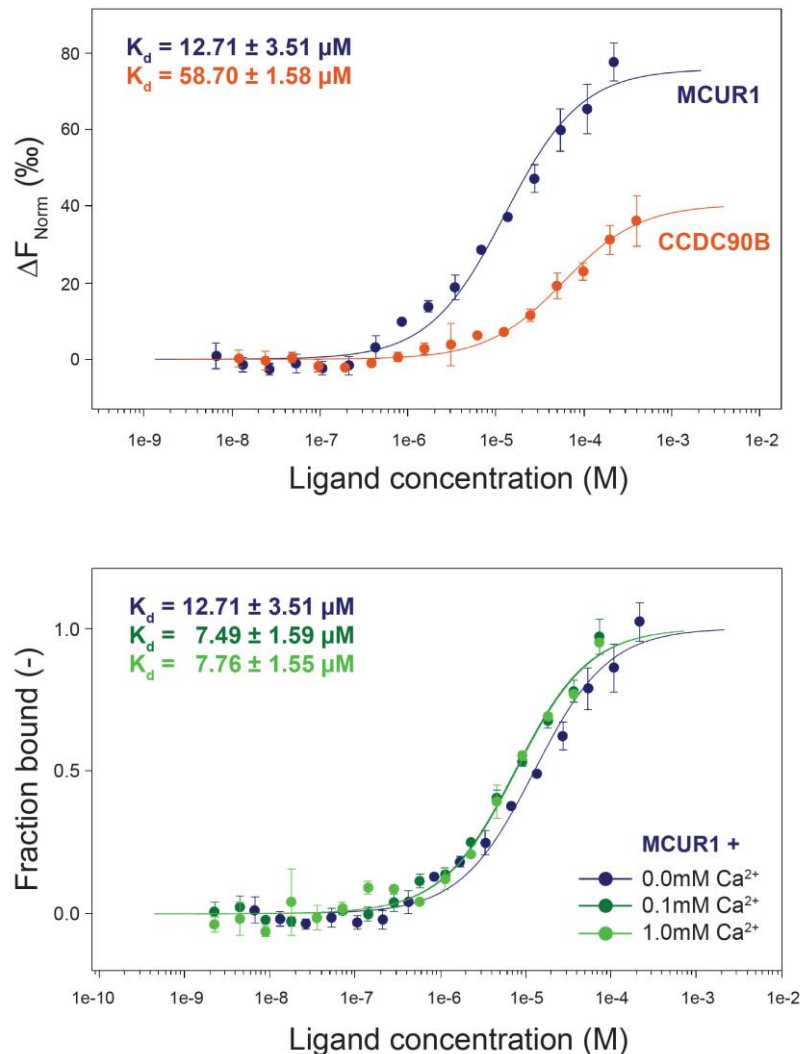


Figure 3.20: *In vitro* MCU binding to MCUR1 and CCDC90B. (A) MST measurement for MCUR1₁₆₀₋₂₃₀ (blue) and CCDC90B₄₃₋₁₂₅ (red) binding to fluorescence-labelled MCU₇₅₋₂₃₃. (B) MST experiment analyzing the effect of Ca^{2+} on MCUR1₁₆₀₋₂₃₀ binding to MCU₇₅₋₂₃₃. Titration curves for absence (blue) and presence of 0.1 mM (dark green) and 1 mM (light green) $CaCl_2$ are shown with calculated dissociation constants. In (A) and (B) Data points are mean of three measurements.

3.2.1.9 Subcellular localization of human MCUR1 and CCDC90B

Human mempromCC paralogs are nuclear-encoded and contain a predicted N-terminal mitochondrial target signal (MTS), therefore, the proteins are transported from the cytosol to mitochondria. Most transmembrane prediction programs (TMHMM, Phobius, HMMTOP) detect only one C-terminal TM region. However, a study in 2012 by Madesh and colleagues (Mallilankaraman et al., 2012a) claimed the presence of a second TM in the low complexity

disordered region close to the N-terminus of the MTS processed MCUR1. Analyzing the membrane insertion of full-length MCUR1 by performing a proteinase-K treatment of the mitoplasts, they identified a 6 kDa fragment which likely corresponds to residues at the N-terminus protruding on the IMS side. To confirm the presence of a second N-terminal TM helix, I decided to localize full-length MCUR1 and a C-terminal TM helix truncation variant MCUR1 Δ TM (residues 1 - 331) (Fig. 3.21).

Mitochondria were isolated from HEK293 cells transfected with C-terminal FLAG-tagged MCUR1 full-length or MCUR1 Δ TM and sub-fractionated into OMM, IMM and matrix components. In agreement with previous findings (Mallilankaraman et al., 2012a), we confirmed that full-length MCUR1 together with its processed fractions localized exclusively to IMM. Surprisingly, the C-terminal truncation variant MCUR1 Δ TM still partly localized to the IMM. Although from this result, it cannot be yet concluded if MCUR1 indeed features a second TM helix closer to the N-terminus, the obtained results are in contrast to our assumption that MCUR1 is anchored to the IMM exclusively by its C-terminal TM helix.

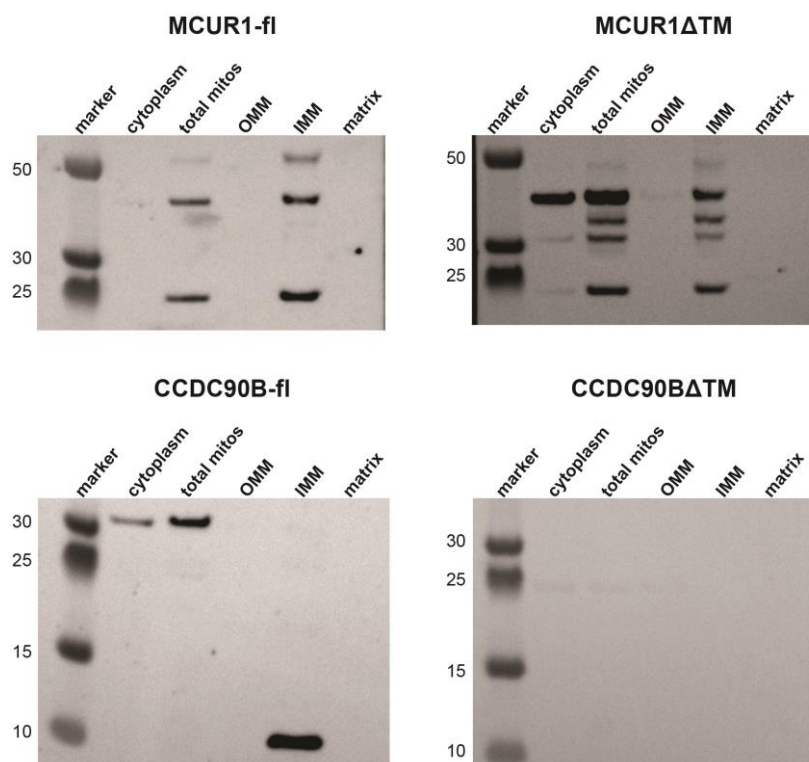


Figure 3.21: **Subcellular localization of human MCUR1 and CCDC90B.** HEK293 cells overexpressing MCUR1-FLAG, CCDC90B-V5 ('fl' indicates full-length proteins) and their C-terminal truncation variants MCUR1 Δ TM-FLAG and CCDC90B Δ TM-FLAG were subfractionated into cytoplasm, total mitochondria, OMM (outer mitochondrial membrane), IMM (inner mitochondrial membrane) and matrix components, and analyzed by Western blot using anti-FLAG (1:2000) or anti-V5 (1:1000) antibodies.

As a control, we decided to test CCDC90B in the same experimental setup. From sequence analysis, we know that CCDC90B harbors only one C-terminal TM helix and lacks the putative disordered region. Similar to MCUR1, we transfected C-terminally V5-tagged full-length CCDC90B and FLAG-tagged CCDC90B Δ TM (residues 1-227) in HEK293 cells, isolated mitochondria followed by sub-fractionation. A completely different localization pattern for full-length CCDC90B was seen compared to MCUR1 (Fig. 3.21). MTS-processed CCDC90B was present in two different pools – one in cytosol and the other localized to IMM (V5 tag was cleaved off during fractionation but is seen in IMM fraction); CCDC90B Δ TM was not expressed. Future experiments are necessary to analyze the observed differences in the expression pattern of both paralogs in more detail.

3.2.1.10 MCUR1 is proteolytically processed in human mitochondria

Analyzing our co-immunoprecipitation and cellular localization results, we observed proteolytically cleaved fractions of MCUR1 identical to previous reports (Chaudhuri et al., 2016; Tomar et al., 2016). To narrow down the region underlying proteolytic cleavage, we expressed MCUR1-FLAG in HEK293 cells and precipitated three prominent protein fragments using anti-FLAG antibody which were separated by SDS-PAGE and digested with trypsin (Fig. 3.22). Analysis of the tryptic peptides by mass spectrometry identified the two larger bands with an apparent molecular weight of 41 kDa and 37 kDa respectively as full-length MCUR1 with uncleaved and cleaved mitochondrial targeting signal peptide. For the smallest, most prominent band at around 25 kDa, no peptides corresponding to the N-terminal 140 residues of MCUR1 were found, suggesting proteolytic processing of a major part of the disordered region.

Is this proteolysis mediated by Cyclophilin D?

Cyclophilin D (CypD), or peptidyl-prolyl cis-trans isomerase F (PPIF), is an essential component of the mitochondrial permeability transition pore complex (see section 3.1.2.3) (Schinzel et al., 2005). In a previous report, full-length human MCUR1 was shown to immunoprecipitate matrix-localized CypD in addition to MCU, implicating that MCUR1 could act as bridging unit between MCU and MPTP complexes by sensing local Ca²⁺ elevations (Chaudhuri et al., 2016).

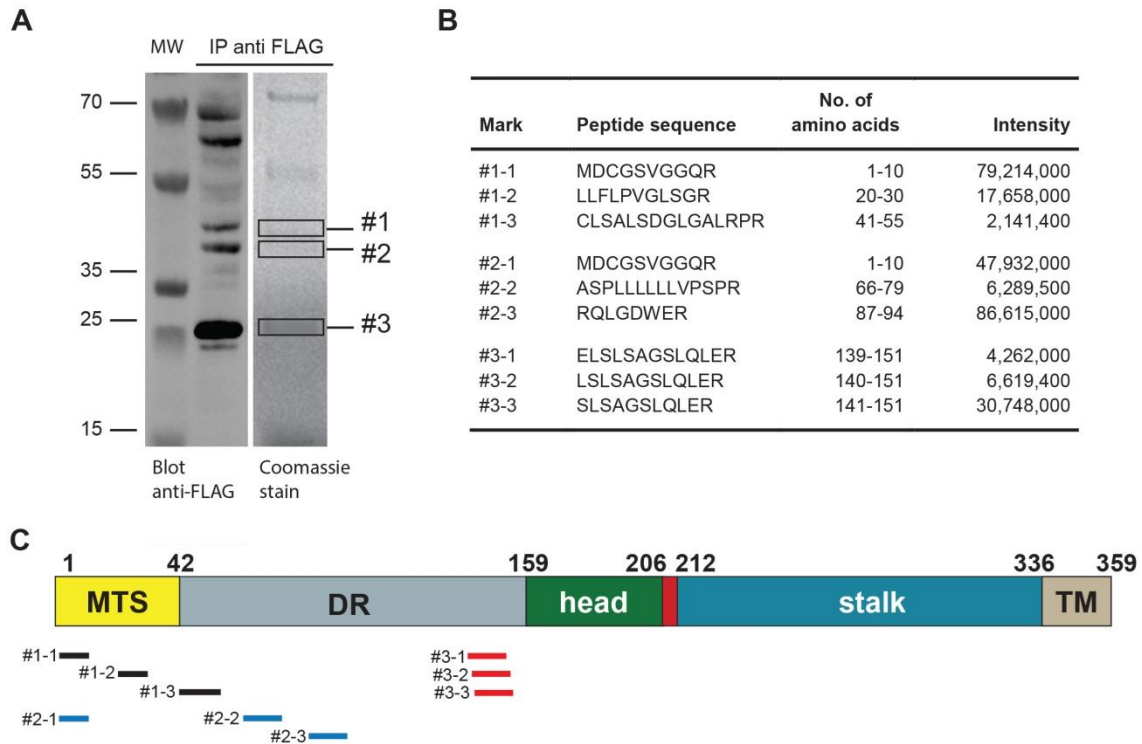


Figure 3.22: **Mass spectrometric analysis of MCUR1 fragments.** (A) Lysate of HEK293 cells, transiently transfected with wild-type MCUR1-FLAG, was subjected to immunoprecipitation using anti-FLAG antibody. A small sample was analyzed by western blotting using anti-FLAG antibody. The remaining sample was separated by SDS PAGE and stained with Coomassie Brilliant Blue. Comparing western blot signals and protein bands of the stained SDS gel, the corresponding regions containing differentially processed fragments of MCUR1-FLAG were cut from the gel (#1, #2 and #3) and analyzed by mass spectrometry. (B) The table summarizes the sequence, residue numbers and intensity values determined by mass spectrometry for three most N-terminal fragments identified from each gel slice. (C) Schematic presentation of identified peptides in relation to the domain architecture of MCUR1.

By catalyzing the cis-trans isomerization of proline imidic peptide bonds, PPIases accelerate protein folding, assist in proteolysis and post-translational modifications. Therefore, we suspected that one possible function of CypD could be to assist MCUR1 N-terminal cleavage, by allowing accessibility for a proteolytic attack. Structural insights into CypD reveal the presence of two conserved pockets which contribute to substrate selectivity and enzymatic catalysis (Davis et al., 2010). S1, the proline interaction pocket, acts as the docking surface for target proline, and S2, a deep scaffold binding pocket where disparate sidechain residues can bind, accommodates the successive residue (Fig. 3.23). Human MCUR1 disordered region (residues 42-159) contains 12 prolines. So, we designed 12 MCUR1 variants with single prolines mutated to alanine residues, and checked their expression profiles (Fig 3.24A; all data not shown) (Dr. Birte Hernandez and Dr. Claire Bedez, personal communication). Initial

expression tests identified mutant MCUR1-P85A to undergo incomplete proteolysis, showing that a large fraction still retained the complete disordered N-terminal region in comparison to other mutants. Sequence alignment of mammalian MCUR1 orthologs shows Pro85 to be conserved among species, further pointing to a possible significance of CypD in proteolytic regulation.

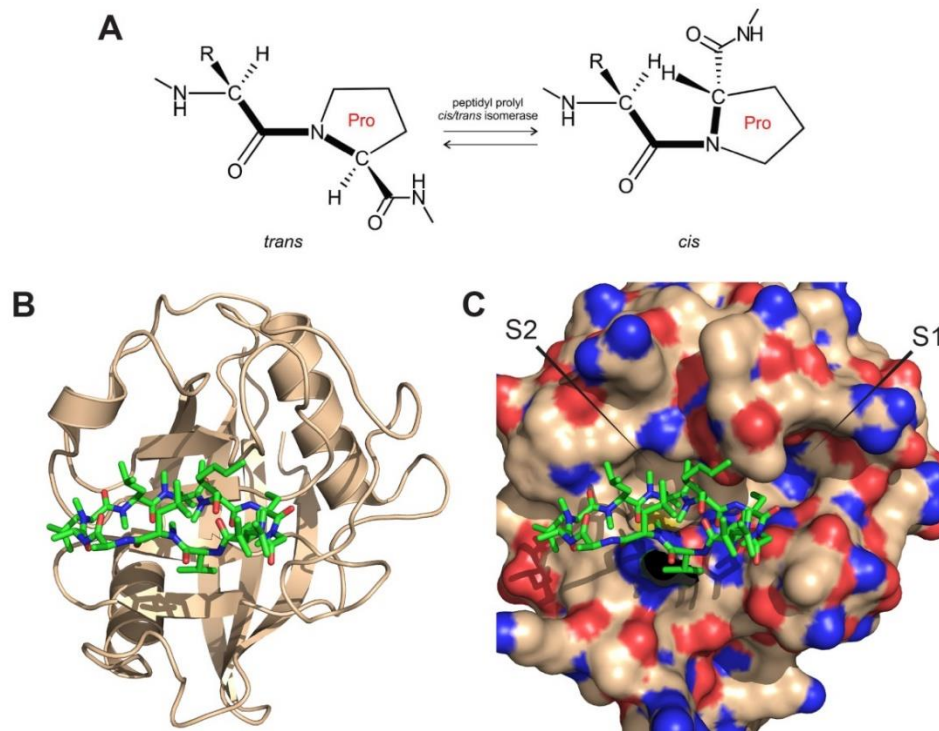


Figure 3.23: **Structure and mechanism of action of cyclophilin D.** (A) Cyclophilin D, or peptidyl-prolyl cis-trans isomerase F (PPIF), catalyzes the isomerization of proline peptide bonds. (B) Structure of human cyclophilin D bound to its inhibitor molecule Cyclosporin A. (C) The macrocyclic ring of Cyclosporin A mimics the placement of substrate molecule in cyclophilin D active site cavity. S1 pocket binds target proline and S2, a considerably broad pocket binds to the next residue. A small hydrophobic residue (as Ala) is preferred preceding proline, but successive residue shows no specificity.

Does MCUR1 head domain bind to Cyclophilin D?

As a next step, we tested *in vitro* binding of MCUR1 head domain to CypD using MST. CypD ($\Delta 1-43$ K133I) was expressed in *E. coli* BL21 (DE3) cells and purified as described in (Valasani et al., 2014). Estimated by SEC-MALS, CypD ($\Delta 1-43$ K133I) exists as a monomer in solution. Alexa Fluor-647 fluorescence-labelled CypD was titrated against MCUR1₁₆₀₋₂₃₀ and CCDC90B₄₃₋₁₂₅. Compared to MCU, CypD bound weakly to the head domains of MCUR1₁₆₀₋₂₃₀ and CCDC90B₄₃₋₁₂₅ with dissociation constants of $72.94 \pm 1.5 \mu\text{M}$ and $204.84 \pm 1.28 \mu\text{M}$ respectively (Fig. 3.24B). This suggests that residues from the intrinsically disordered region of MCUR1 might also be involved in binding. This suggestion is supported

by the finding that only full-length MCUR1 and not the 25 kDa proteolytically cleaved fragment lacking the disordered residues, could be immunoprecipitated with CypD (Chaudhuri et al., 2016).

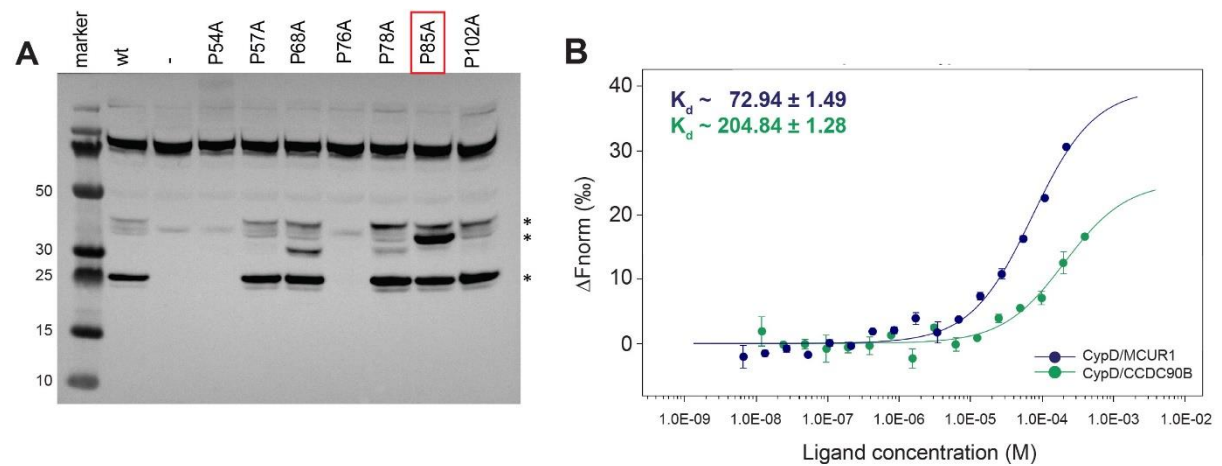


Figure 3.24: **MCUR1 proteolysis and binding to Cyclophilin D.** (A) MCUR1-FLAG with proline mutations to alanine (residue number marked at top) expressed in HEK293 cells analyzed on a western blot with anti-FLAG antibody. (B) MST measurement for MCUR1₁₆₀₋₂₃₀ (blue) and CCDC90B₄₃₋₁₂₅ (green) binding to Cyclophilin D. Data points are average of three measurements. Calculated dissociation constants K_d are shown.

3.2.2 Prokaryotic mempromCC proteins – structural and functional studies

As mempromCC homologs were identified only in pathogenic strains of *E. coli* and *S. typhimurium* and other species which we cannot culture in our laboratory, we selected *Caulobacter crescentus* for structural and functional studies. *Caulobacter* species cluster together with the metazoan homologs within the central group of the CLANS map generated from the head-neck segments from a heterogeneous group of membrane-anchored proteins sharing common domain architecture (Fig. 1). Proteobacterial mempromCC proteins share high sequence similarity in their head domains with their eukaryotic counterparts. They only differ in terms of an absent signal peptide, and the variability in sequence, length and periodicity of their coiled-coil stalks. Their head domains start nearly at the conserved FDT motif. All prokaryotic homologs are presently annotated as hypothetical proteins. As bacterial species lack MCU homologs, the presence of mempromCC proteins hints at other cellular function(s).

3.2.2.1 Structural characterization of *Caulobacter* mempromCC homologs

MempromCC proteins from two *Caulobacter* species: WP_047412812 from *C. sp.* OV484 (referred as MpcC-OV484) and WP_029916579 from *C. sp.* UNC358MFTsu5 (referred as MpcC-UNC) were selected for structural characterization. Complete length of proteins including the head, neck(s) and stalk, and excluding the transmembrane region were fused to an N-terminal histidine tag and purified under native conditions from *E. coli* C41 (DE3) overexpressing cells. A major fraction of the purified proteins tended to form aggregates. While MpcC-OV484 did not crystallize, MpcC-UNC crystallized as thin plates which diffracted at best to 3.1 Å. Attempts to refine the crystals using grid-screen optimization failed to produce good quality crystals. Subsequently, I tried to solve the structure with NMR spectroscopy. MpcC-UNC was an interesting choice, as the protein featured “TATK” β -layer in hexad spacing, similar to those described in Chapter 2. In MpcC-UNC, TATKAD is sandwiched within two nonad β -layers ISGLATKAD and LANMATKAD. MpcC-UNC purified from isotope-labelled minimal-medium eluted as three major peaks from Superdex 75. While peak 1 and peak 2 were high molecular weight aggregates, the minor fraction peak 3 constituted the properly assembled oligomer, which was confirmed from 1D ^1H spectra. In preliminary experiments, peak 3 displayed a well-dispersed ^{15}N -HSQC spectrum, however, the inability to purify sufficient amounts of well-folded protein hampered further progress.

Due to the difficulty in obtaining well-diffracting crystals, homology models of bacterial mempromCC proteins from *Caulobacter* species: *C. crescentus* NA1000 (YP_002517927) and *C. sp.* JGI 0001013-D04 (WP_018113394) (Fig. 3.25) were designed using CCDC90B₄₃₋₁₂₅ crystal structure as a template (work done by Ioanna Karamachali). The models clearly display the natural variability of the stalk domains in prokaryotes in terms of length and the number of β -layers. Showing identical head-neck segments, the stalk of *Caulobacter sp.* JGI 0001013-D04, which contains eight successive β -layer necks, is with a length approximately 10 nm longer than the *Caulobacter crescentus* NA1000 protein.

3.2.2.2 Structural characterization of archaeal Kcr-0859 showing mempromCC-related domain architecture

We further investigated the molecular structure of proteins from peripheral prokaryotic clusters that share the helical domain architecture, but not sequence homology with members of the mempromCC family. The crystal structure of Kcr-0859 (WP_012309502), a crenarchaeal protein of unknown function from *Candidatus Korarchaeum cryptofilum* was solved (Dr. Birte Hernandez, Dr. Marcus Hartmann, personal communication). Kcr-0859- Δ TM (residues 1-136) lacking the C-terminal membrane anchor was recombinantly expressed in *E. coli* C41 strain in M9 minimal media supplemented with selenomethionine (SeMet). Crystals were obtained under three conditions and the structure was solved at a resolution of 2.5Å via SAD phasing.

In agreement with our expectation, the Kcr-0859 Δ TM structure shows an elongated parallel trimer comprising of a helical head domain, which is connected via a β -layer neck to a coiled-coil stalk (Fig. 3.25A). Each monomer of the head bundle consists of three short helices (α 1, α 2, α 3), which are connected by short turns and arranged perpendicular to each other. Within the trimer, helices are packed in parallel with helix α 2 of each monomer accommodated between α 1' and α 3'' of the other chains. The central axis of the head domain is kinked by 40 degrees relative to the coiled-coil axis of the adjacent stalk. This kinked arrangement is likely a result of crystal packing constraints, but is also indicative of a certain degree of flexibility of the neck region, as reported for β -layer-mediated transitions in DALL domains of TAAs (Hartmann et al., 2012; Koiwai et al., 2016). Formed by the first three residues of the motif MATKED, the β -layer directly succeeds the α 3 helices and forms the characteristic inter-chain hydrogen bonds between the central alanine residues.

Following the neck, the structure shows 76 residues of the 96-residue cytosolic part of the coiled-coil stalk. This whole region does not possess noticeable supercoiling over large extents, as anticipated from its annotation, which indicated an overall sequence periodicity of 3.64 residues per turn, only marginally different from 3.63 r/t expected for an undistorted α -helix. Towards its C-terminal end, the stalk contains two YxD motifs, both of which are resolved in the structure. These polar motifs are commonly found in right-handed coiled coils, where they convey structural specificity and stability by forming inter-chain hydrogen bonds between the hydroxyl groups of the tyrosines and the carboxyl groups of the aspartates (Alvarez et al., 2010).

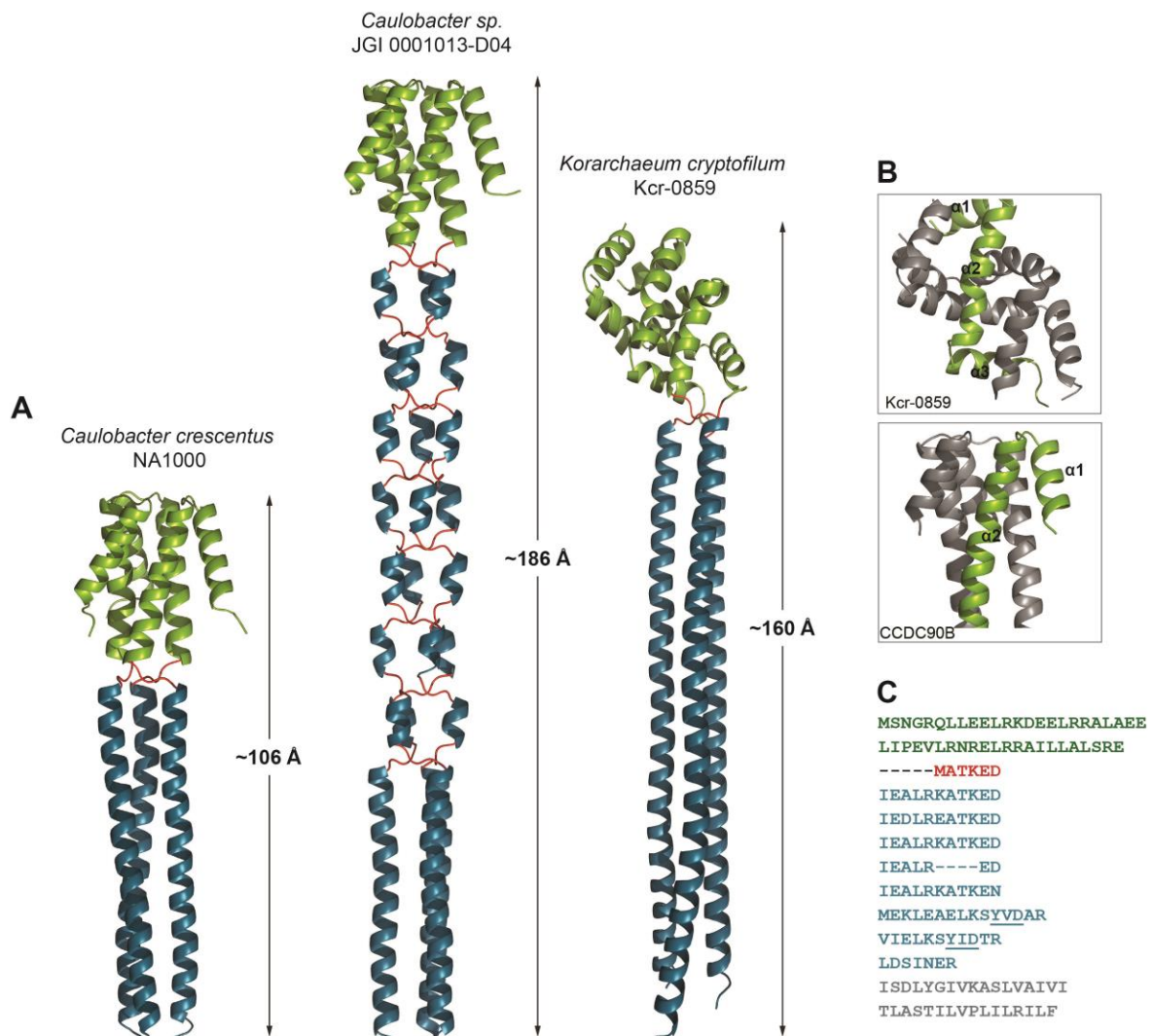


Figure 3.25: **Structure of prokaryotic mempromCC homologs.** (A) Homology models of *Caulobacter crescentus* NA1000, *Caulobacter sp.* JGI 00010113-D04, and crystal structure of Kcr-0859: head domain (green), β -layer neck (red) and stalk (blue). Approximate length is calculated. (B) Comparison of head domains of human CCDC90B and *K. cryptofilum* Kcr-0859. One chain is colored in green, other two chains are in grey. Helices ($\alpha 1$, $\alpha 2$ and $\alpha 3$) are marked for one chain. (C) Amino acid sequence of Kcr-0859. Residues are colored according to their domain affiliation. Polar YxD motifs are underlined.

3.2.2.3 Subcellular localization and orientation of mempromCC in bacteria

MempromCC localizes to the cytoplasmic membrane in bacteria with an N-in C-out topology

Secondary structure prediction of mempromCC proteins in bacteria predicts one transmembrane (TM) domain at the C-terminus but no signal sequence at the N-terminus for export across lipid bilayer indicating that mempromCC proteins very likely anchor to the inner cell membrane. We experimentally analyzed the cellular localization of mempromCC homolog CCNA_02554 (YP_002517927) from *C. crescentus* NA1000, thereafter referred to as MpcC, using electron microscopy and subcellular fractionation.

To determine the intracellular localization by EM, we designed an N-terminal HA-MpcC fusion construct in a xylose-inducible high-copy plasmid pBXMCS4 and utilized the leaky expression to ensure that low amounts of the fusion protein are expressed in the cells. Cryo-sections of HA-MpcC expressing *Caulobacter* cells were stained with polyclonal anti-HA antibody, followed by immunogold labelled secondary antibody and analyzed by transmission electron microscopy (work done by Dr. Juthaporn Sangwallek). We observed an accumulation of electron-dense particles in close proximity to the inner side of cytoplasmic membrane (Fig. 3.26A), supporting our assumption that mempromCC protein is C-terminally anchored to the bacterial cytoplasmic membrane projecting the HA-tagged head into cytoplasm.

This result was confirmed by subcellular fractionation of *Caulobacter* cells expressing wild-type MpcC into membrane and soluble parts. The inner and outer membranes were separated using selective detergent solubilization (using 2% Triton X-100) of the IM. The resulting fractions were probed by Western blotting with antisera for MpcC, inner membrane protein TimA, and outer membrane protein CpaC. The control proteins TimA and CpaC are most abundant in the expected fractions. We clearly observe that mempromCC is enriched in the inner membrane fraction (Fig. 3.26B).

I then investigated the membrane topology of mempromCC using Proteinase K treatment. MpcC was fused to an N-terminal HA and C-terminal 3x-FLAG tag in a high-copy plasmid pBXMCS4. Cells were treated with lysozyme shortly to disrupt outer cell membrane and incubated with Proteinase K for 30 min at room temperature followed by subcellular fractionation. Fractionated spheroblasts were analyzed by western blotting using anti-HA and

anti-FLAG antibodies (Fig. 3.26C). In control sample, full-length HA-MpcC-FLAG₃ and a proteolytic fragment (~1 kDa smaller, equivalent to one FLAG epitope) are detected by both antibodies, implying an intact N-terminus but apparent degradation at the C-terminus. For Proteinase K treated samples, I observed an additional band detected by anti-HA, but not with anti-FLAG antibody.

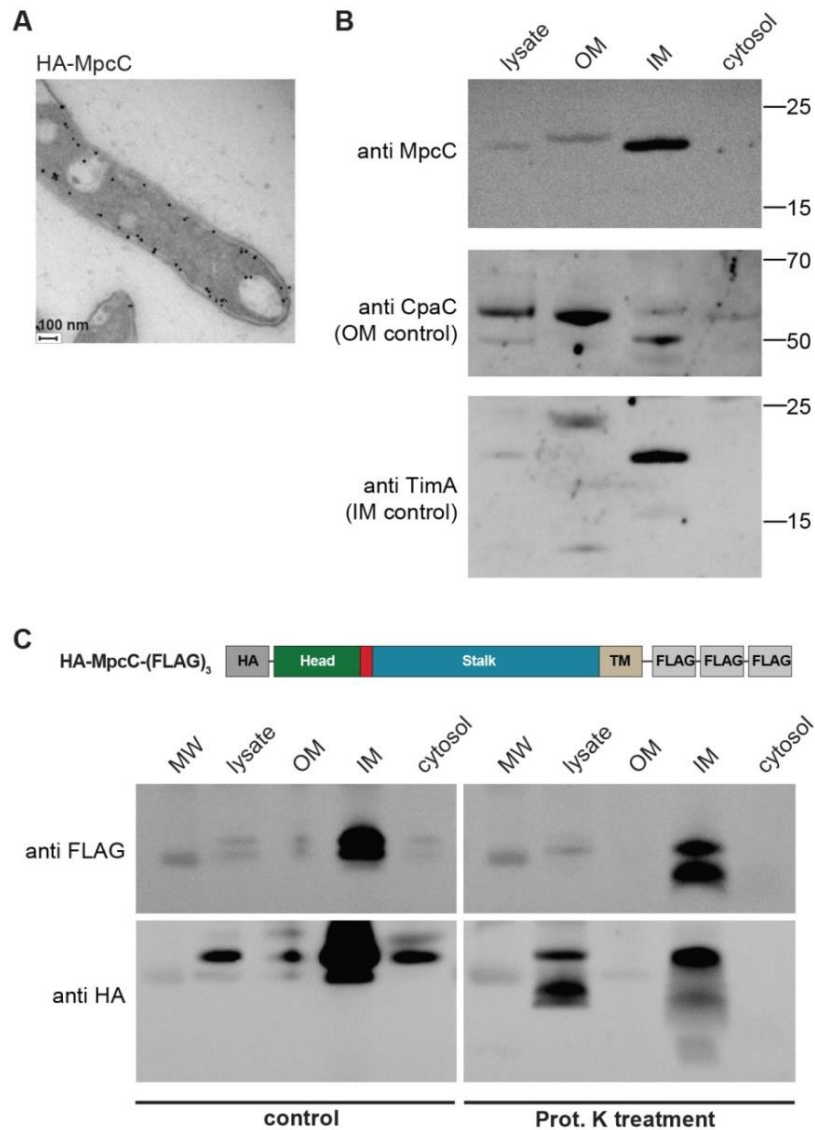


Figure 3.26: **Subcellular localization of prokaryotic mempromCC homologs.** (A) TEM micrograph showing MpcC localization in *Caulobacter* cells. (B) Subcellular fractionation of *C. crescentus* NA1000 cells expressing HA-MpcC. Whole cell lysate (lysate), outer membrane (OM), inner membrane (IM) and cytosol fractions were analyzed on western blot against anti-MpcC, anti-CpaC (OM control) and anti-TimA (IM control) antisera. (C) Proteinase-K treatment of *C. crescentus* NA1000 cells expressing HA-MpcC-FLAG₃. Fractionation samples were analyzed on western blot using anti-HA and anti-FLAG antibodies.

This fragment of the smallest size has all three C-terminal FLAG-epitopes cleaved by Proteinase K. The incomplete digestion, as observed for other proteolytic fragments, can be explained to be a result of close proximity of (FLAG)₃-tag to the membrane surface preventing accessibility to the protease. Thus, both EM and subcellular fractionation methods confirm that mempromCC proteins in bacteria are localized to the cytoplasmic membrane with the coiled-coil stalk projecting the head in the cytoplasm.

3.2.2.4 Functional characterization of mempromCC proteins in *C. crescentus* NA1000

Pull-down to identify interaction partners

To identify MpcC (CCNA_02554) binding partners in *C. crescentus* NA1000, I performed a pull-down experiment using soluble MpcC variants as bait. Two different constructs – MpcC₁₋₈₅ (residues 1-85) and MpcC₁₋₆₀ (residues 1-60) – with different coiled-coil stalk lengths and lacking the transmembrane anchor were fused to 3x-FLAG at the C-terminus (Fig. 3.27A). GCN4-pII-(FLAG)₃ was used as the negative control. All constructs were recombinantly expressed in *E. coli* BL21 (DE3), purified from supernatant and used at a concentration of 1.5 mg/ml in the pull-down experiment. Bait proteins attached to the anti-FLAG M2 magnetic beads (Sigma) fished out the interacting partners from the *Caulobacter crescentus* NA1000 lysate. The collected eluates were dimethyl labelled for comparative proteomic analysis, digested with trypsin and subjected to mass-spectrometric analysis.

From the statistically significant set of bound peptides identified in pairwise sample comparison of MpcC₁₋₈₅, MpcC₁₋₆₀ and control group, only FtsZ (a protein involved in cell division) was identified to be significantly enriched (Fig. 3.27B). A previous study identified FtsZ to be a substrate of ClpA protease (Williams et al., 2014), which is located just upstream of the MpcC gene. To confirm the interaction between full-length MpcC₁₋₈₅ and FtsZ, I conducted a co-immunoprecipitation experiment using – (i) *C. crescentus* co-expressing MpcC₁₋₈₅-FLAG on pBXMCS4 and genomic integrated mCherry-FtsZ or (ii) using soluble full-length MpcC₁₋₈₅-FLAG bound to magnetic beads and incubated with lysate of cells expressing genomic integrated mCherry-FtsZ. The eluted samples were analyzed by western blot using anti-FLAG and anti-mCherry antibodies (Fig. 3.27C). In both cases, mempromCC MpcC₁₋₈₅ failed to immunoprecipitate FtsZ.

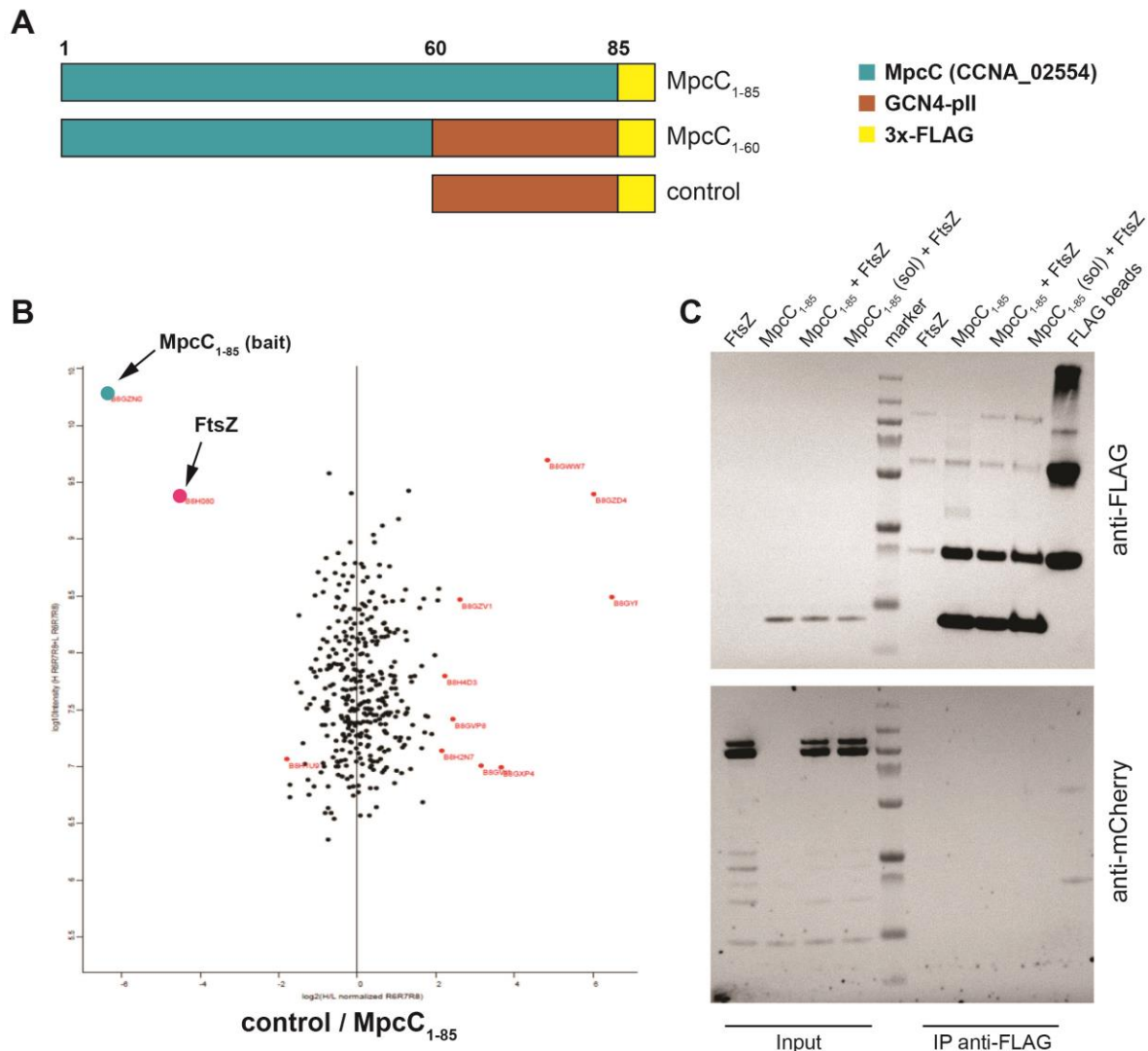


Figure 3.27: **Pull-down to identify MpcC interacting partners.** (A) Designed constructs C-terminal FLAG-tagged MpcC₁₋₈₅ and MpcC₁₋₆₀, without C-terminal membrane domain, were purified as soluble proteins. GCN4-pII was used as the control protein. (B) Abundance of MS identified proteins compared in MpcC₁₋₈₅ and control sample. FtsZ (pink dot) was the only significant hit. Proteins with statistical significance ($p < 0.05$) have been marked as red dots. (C) Co-immunoprecipitation study of MpcC-FLAG and mCherry-FtsZ.

Parallel studies to identify proteins interacting with mempromCC in *C. crescentus* NA1000 using bacterial-2-hybrid system also failed to establish definite binding partners (Dr. Juthaporn Sangwallek, personal communication). First data from ongoing work studying a mempromCC deletion mutant at the transcriptomic and proteomic level are more promising to reveal the function of these proteins in prokaryotic systems.

3.3 CONTRIBUTIONS TO THIS WORK

All bioinformatics work including identification of mempromCC proteins, phylogenetic analysis using CLANS and multiple sequence alignment was carried out by Prof. Andrei Lupas and Ioanna Karamichali. Proteins used for experimental study were selected by Prof. Andrei Lupas.

Human MempromCC project

- Molecular cloning of human CCDC90B₄₃₋₁₂₅-GCN4, MCUR1₁₆₀₋₂₃₀ and MCU₇₅₋₂₃₃, followed by protein expression, protein purification (under native or denaturing conditions) and biophysical characterization (SEC-MALS, CD spectroscopy and thermal melting experiments) was carried out by me.
- Crystallization screens were set-up by Dr. Reinhard Albrecht and Kerstin Bär. Crystal structure of CCDC90B₄₃₋₁₂₅-GCN4 was solved by Dr. Marcus Hartmann. I manually refined the structure using COOT.
- I designed the MCUR1 full-length homology model with the program MODELLER using the crystal structure of CCDC90B. NMR isotope-labelled sample was prepared by me; spectra measurements and residue assignments were done by Dr. Murray Coles.
- I studied the effect of Ca²⁺/Mg²⁺ on CCDC90B and MCUR1 using SEC-MALS, CD spectroscopy and microscale thermophoresis (including preparing fluorescent-labelled samples, *in vitro* binding studies and analysis), and prepared the amyloid fibril samples for electron microscopy imaging. EM was done by Dr. Katharina Hipp.
- For pull-down studies of MCU with MCUR1, I contributed to the construct designs. Cloning and pull-down experiments were carried out by Kerstin Bär, Silvia Deiss and Dr. Birte Hernandez. *In vitro* binding study for MCU and MCUR1/CCDC90B using MST was done by me.
- I carried out the pull-down for MCUR1-FLAG expressed in HEK293 cells for MS analysis of proteolysis products.
- For cellular localization of human MCUR1 and CCDC90B, Dr. Birte Hernandez transfected and cultured HEK293 cells. I performed the mitochondrial isolation and subfractionation, followed by Western blot analysis for all constructs.
- CyclophilinD was expressed and purified by Kerstin Bär. Using this, I performed MST experiments to study binding of CypD with MCUR1/CCDC90B. 7 out of 12 MCUR1 proline mutants (P→A) were cloned by Kerstin. I analyzed their cellular expression by western blot.

Bacterial MempromCC project

- Molecular cloning, protein expression and purification for two *Caulobacter* proteins MpcC-UNC and MpcC-OV484 were carried out by me. NMR spectroscopy measurements were done by Dr. Murray Coles and Manish Chaubey.
- The homology models of *Caulobacter* proteins were designed by Ioanna Karamichali. The structural characterization of archaeal mempromCC protein was carried out by Dr. Birte Hernandez and Dr. Marcus Hartmann.
- I carried out the cellular localization and Proteinase K topology experiments for *Caulobacter* MpcC. EM sample was prepared by Dr. Juthaporn Sangwallek.
- I performed the pull-down experiments using soluble and plasmid-expressed mempromCC proteins to identify their interaction partners in *Caulobacter crescentus* NA1000 by MS and western blot.

3.4 DISCUSSION

3.4.1 MempromCC are obligate trimers with a head-neck-stalk-anchor architecture

The mempromCC family contains mostly uncharacterized homologous coiled-coil containing membrane proteins conserved in prokaryotes and eukaryotic mitochondria. All mempromCC homologs are predominantly helical and comprise of an N-terminal head domain projected by a C-terminal membrane-anchored coiled-coil stalk. The head domains are evolutionary conserved in sequence and structure, posing the unique features of mempromCC family and clearly demarcating it from other families. The head-stalk transition is invariably mediated by one or more β -layer necks. The stalk domains of prokaryotic mempromCC homologs are more diverse in sequence and length than their eukaryotic counterparts. Structural models of *Caulobacter* proteins give a clear impression of the diversity of prokaryotic stalk domains in terms of length and number of β -layer necks.

3.4.2 Eukaryotic mempromCC proteins localize to IMM and prokaryotic homologs to the cytoplasmic membrane, with an N-in C-out topology

Majority of the eukaryotic homologs contain a predicted mitochondrial target sequence. From the submitochondrial fractionation technique, I confirmed the inner mitochondrial membrane localization of human MCUR1, in agreement with previous reports (Mallilankaraman et al., 2012a; Paupe et al., 2015; Chaudhuri et al., 2016). While MCUR1 is exclusively localized to IMM, results show that CCDC90B localizes partly to both cytoplasm and IMM. Previous electrophysiology experiments have shown that only MCUR1 is involved in the regulation of Ca^{2+} uptake through MCU. Overexpression of MCUR1 increases the Ca^{2+} current through the uniporter channel, whereas CCDC90B overexpression or knockdown does not modulate this influx. Although its function remains presently unknown, mitochondrial CCDC90B has been shown to exist in a complex with MCU and MCUR1. The second fraction of CCDC90B which localizes to cytoplasm, may have a completely different function from cellular calcium uptake. It could be the evolutionary conserved function of the mempromCC family, shared with bacterial and yeast species which do not possess MCU homologs.

In addition to the conserved C-terminal transmembrane helix, a 2012 study (Mallilankaraman et al., 2012a) claimed the presence of a second TM in the low complexity disordered region of MCUR1 close to the N-terminus. To verify this, I tested the cellular localization of its C-

terminal truncation variant MCUR1 Δ TM and found that it localizes partly to IMM. Possible reasons may include that it strongly tethers to an interaction partner in the IMM and therefore evades the high pH carbonate extraction into the soluble matrix component. Therefore, future experiments are warranted to confirm the presence of a second N-terminal TM. A major fraction of MCUR1 Δ TM, however, was observed to be present in cytosol. This could arise from the translocation problem of truncation variant across mitochondrial membrane. Closer observation also showed that MCUR1 Δ TM was more susceptible to proteolysis compared to the full-length MCUR1.

Prokaryotic mempromCC proteins lack any signal peptide preceding their head domains. From electron microscopy and subcellular fractionation experiments in *Caulobacter crescentus* NA1000, I confirmed that prokaryotic homologs are anchored to the cytoplasmic membrane via their C-terminal transmembrane helices with an N-in C-out topology, i.e. they project their heads into the cytoplasm.

3.4.3 Eukaryotic mempromCC proteins have diverse N-terminal disordered extensions

Whereas prokaryotes encode only one mempromCC homolog per organism, most eukaryotes express at least two paralogous proteins which primarily differ in their N-terminal sequence extensions preceding the conserved head domain. Human MCUR1, in contrast to CCDC90B, contains an additional 100 residues. Highly variable in sequence and length, these extensions are predicted to be natively disordered, which are not necessarily “unstructured”, i.e. devoid of any secondary structural features. Computational algorithms such as PONDR VL-XT and ANCHOR identified short segments, known as *Molecular Recognition Features* (MoRFs), within these disordered regions (Dosztanyi et al., 2009; van der Lee et al., 2014). These segments act as nucleation sites for folding as they display a high propensity to undergo disorder-to-order transition in response to environmental changes such as binding to an interaction partner or post-translational modifications. This diversity in N-terminal extensions among eukaryotic paralogous proteins may point to differences in their cellular functions.

3.4.4 Structure of the conserved head domain of human CCDC90B and MCUR1

The head domain of CCDC90B belongs to the biggest subgroup comprising more than two-thirds of the head sequences identified in this study, including all from eukaryotes and many from proteobacteria. Crystal structure shows that the protein forms a parallel homotrimer. Its

head domain is formed of a six-helical antiparallel bundle that connects to the coiled-coil stalk via a β -layer neck. This six helix bundle is assembled by the α 1-helices packing against the trimeric bundle formed by the α 2-helices of the each chain. As the two paralogs share high sequence similarity, I reconstructed a full-length homology model for the conserved head-neck-stalk domains of MCUR1 using CCDC90B crystal structure as the template. The intrinsically disordered region of MCUR1 can be envisaged as a highly flexible unstructured N-terminal extension. The coiled-coil stalk of MCUR1 acts as a molecular spacer in IMM and projects the head domain at a distance of 23 nm into the matrix.

3.4.5 MCUR1 head domain is destabilized by Ca^{2+} binding which accelerates its conversion to β -amyloid fibrils

Despite their structural similarity, head domains of MCUR1 and CCDC90B clearly differ in their surface charge distribution and biophysical properties. Binding of Ca^{2+} or Mg^{2+} to MCUR1 head domain significantly impaired its thermal stability. Interestingly, the thermal unfolding process of MCUR1, and not CCDC90B, was accompanied by the formation of β -amyloid fibrils. The destabilization caused by the addition of Ca^{2+} further accelerated the fibrillar formation. A similar effect of Ca^{2+} on β -fibril formation has been previously reported for amyloid β -peptide ($\text{A}\beta$) and α -synuclein involved in the toxic plaque formation in Alzheimer's disease and Parkinson's disease respectively (Isaacs et al., 2006; Han et al., 2018). Clearly, the slight difference in the sequences of human MCUR1 and CCDC90B dictates the inability of the latter to form amyloid fibers. Comparative sequence analysis with multiple computational tools such as Amylpred2, TANGO, and ZipperDB identified a common amyloidogenic hotspot in the conserved head region of MCUR1 (residues 184-196 corresponding to α 2-helix in the structure) which is slightly mutated in CCDC90B.

While the pathophysiological relevance of human MCUR1 amyloid formation is not known at present, multiple studies have shown that mitochondria exposed to amyloid fibrils exhibit calcium dysregulation. Small protein aggregates formed at the earlier stages constitute the toxic form of amyloids. These aggregates have the ability to integrate into the lipid bilayer and form ion channels. Indiscriminate leakage of ions across these channels in the inner mitochondrial membrane disrupts the calcium homeostasis, which ultimately signals the opening of the permeability transition pore and cellular death.

In vitro studies show that MCUR1 head can bind to Ca^{2+} in the upper micromolar range and with a lower affinity to Mg^{2+} in the single-digit millimolar range, with no effect on its oligomeric state. CCDC90B, on the other hand, does not show any appreciable binding to both divalent cations. While global free $[\text{Ca}^{2+}]$ in the matrix would never reach millimolar levels, diffusion theory predicts that $[\text{Ca}^{2+}]$ can reach as high as 0.1 mM at the microdomains of Ca^{2+} exit (at a distance of ~ 10 nm from the pore) (Tadross et al., 2013). Considering the spatial proximity of MCUR1 to the calcium-releasing side of the uniporter channel, a functional relevance of Ca^{2+} binding in the regulation of MCUR1 stability and activity *in vivo* seems plausible, similar to MCU. As present experiments were performed with a single domain separated from its natural context, it remains unclear whether Ca^{2+} exerts the same effects on MCUR1 secondary structure in the matrix. Future studies will be required to map potential Ca^{2+} binding sites on the MCUR1 head domain and to assess their significance for MCUR1 stability and activity *in vivo*.

3.4.6 MCUR1 head domain interacts with the N-terminal domain of MCU

Although the prokaryotic mempromCC proteins have not been functionally characterized so far, studies on human paralogs MCUR1 and CCDC90B have identified their essential role in various cellular processes. Human MCUR1 functions as a scaffold factor required for the proper assembly of multi-protein complexes in the inner mitochondrial membrane, including the mitochondrial calcium uniporter (MCU) and the cytochrome c oxidase (COX) complexes (Mallilankaraman et al., 2012a; Paupe et al., 2015; Tomar et al., 2016). As a component of MCU complex, the function of MCUR1 has been extensively studied. Paralogs MCUR1 and CCDC90B interact with each other, and also with essential components of the MCU channel complex, including the Ca^{2+} -selective channel subunit MCU and the regulatory single-pass transmembrane protein EMRE. Although interacting with the same proteins, paralogs differ in their biological activities: only MCUR1, but not CCDC90B, was shown to be essential for active MCU complex formation and cellular Ca^{2+} uptake (Chaudhuri et al., 2016; Mallilankaraman et al., 2012; Tomar et al., 2016).

Based on these studies, we used MCUR1 as a model to examine the functional significance of the individual domains of mempromCC homologs. Previous attempts to map the region of MCUR1 interacting with MCU were performed with constructs designed on the basis of secondary structure prediction. Furthermore, there was discrepancy among the identified

regions ranging from the conserved head domain to the coiled-coil stalk (Tomar et al., 2016; Chaudhuri et al., 2016). With a better understanding of the domain boundaries now, we identify here that the conserved head domain is indispensable for mediating MCU-MCUR1 interaction. Previous studies reported that MCU could immunoprecipitate both MCUR1 and CCDC90B (Tomar et al., 2016; Chaudhuri et al., 2016). We showed here that the head domains of both paralogs are interchangeable for MCU binding in pull-down experiments. *In vitro* binding experiments confirmed that MCU directly interacts via its N-terminal domain with the head domains of MCUR1 and CCDC90B, with a five-fold higher affinity for MCUR1. This observation together with the finding that CCDC90B, in contrast to MCUR1, does not modulate mitochondrial Ca²⁺ uptake (Tomar et al., 2016) supports our assumption that both paralogs fulfill different cellular functions. It is likely that the N-terminal disordered region present only in MCUR1 plays a critical role. Importance of this disordered region at present remains unclear from our experiments as a variant lacking this segment is strongly impaired in stability. The stalk domain which can be simply replaced by a non-related coiled-coil of similar length serves essentially as a projector of the head. Complete deletion of the stalk strongly reduces binding to MCU, suggesting the stalk length to be a critical factor in facilitating the interaction by bridging defined distances. Addition of Ca²⁺ did not appreciably affect the MCU-MCUR1 interaction.

3.4.7 MCUR1 is proteolytically processed in human mitochondria

In agreement with previous studies (Tomar et al., 2016; Chaudhuri et al., 2016), we observed two processed forms of MCUR1 – a full-length variant with cleaved mitochondrial target signal and a second smaller fragment where a major part of the disordered residues are missing. As the two processed elements retain the conserved head domain, it explains the observation that both of these forms can interact with MCU (Tomar et al., 2016). However, so far it is unclear if they are functionally equivalent or possess different biological activities. It is conceivable that MCUR1 activity is proteolytically regulated, similar to other known proteins such as dynamin-like GTPase OPA1, a key regulator of mitochondrial dynamics, and the Parkinson disease related Ser/Thr kinase PINK1 (Ali and McStay, 2018; Greene et al., 2012; MacVicar and Langer, 2016). Moreover, a recent study showed that mAAA proteases AFG3L2 and SPG7 strictly regulate the level of non-assembled EMRE for the proper assembly of MCU complex (Tsai et al., 2017). This supports the idea that MCU complex assembly and activation are likely proteolytically regulated.

MCUR1 was also shown to interact with Cyclophilin D (CypD), an essential component of the permeability transition pore complex. As CypD is known to assist proteolytic activities by the cis-trans isomerization of proline peptide bonds, we probed its involvement in MCUR1 N-terminal cleavage. Initial expression tests of MCUR1 single-proline mutants identified a conserved proline (Pro85) mutant to be resistant to complete proteolysis, which clearly highlights the possible function of CypD in MCUR1 proteolytic regulation. Further *in vitro* studies found a weak binding affinity of CypD to MCUR1 head domain in the absence of its N-terminal disordered region. In future, studies should aim to confirm the MCUR1 residues involved in its interaction with CypD, identify the mitochondrial proteases which regulate MCUR1 proteolysis and its physiological importance on Ca²⁺ uptake regulatory activity.

3.4.8 Cellular function of prokaryotic mempromCC homologs

Prokaryotic species lack MCU orthologs; therefore, the presence of mempromCC homologs suggests a cellular function different from mitochondrial calcium uptake regulation. To identify the interacting partners of mempromCC *in vivo*, I carried out a pull-down assay using soluble MpcC from *Caulobacter crescentus* NA1000 as the bait protein and identified a cell-division protein FtsZ as the only significant hit in mass spectrometric analysis. In subsequent pull-down studies using *C. crescentus* cell lysate expressing genomic integrated FtsZ-mCherry and analysis by western blot, MpcC failed to pull-down FtsZ. Possible reasons could be that the large C-terminal mCherry tag in FtsZ interfered in its interaction with MpcC *in vitro*, or it could also be a false-positive hit. Future experiments should test for *in vitro* interaction with purified FtsZ attached to a smaller (His)₆ or HA tag. A parallel study to identify mempromCC interaction partners using bacterial-2-hybrid system identified mostly false-negatives (Dr. Juthaporn Sangwallek, personal communication), as the overexpressed protein could not properly integrate into the membrane and accumulated in the cell as inclusion bodies. As a result, MpcC non-specifically interacted with other cellular proteins. Our present work to identify mempromCC binding partners by *in vivo* cross-linking, together with studies on the mempromCC deletion mutant at the transcriptome and proteome level will hopefully reveal the function of these proteins in prokaryotic systems in the future.

In conclusion, the mempromCC family proteins share the same domain architecture, but, based on their sequence diversity, they are likely to participate in different cellular processes. Data

suggests that they fulfil their functions generally via their head domains, which are projected by the membrane-anchored coiled-coil stalk and constitute the family-defining element.

3.5 METHODS

3.5.1 Molecular Cloning

For recombinant expression of proteins in *E. coli*, the corresponding DNA fragments were codon optimized and custom synthesized by gene synthesis. Fragments encoding MCU₇₅₋₂₃₃ (NP_612366), MCUR1₁₆₀₋₂₃₀ (NP_001026883), CCDC90B₄₃₋₁₂₅ (NP_068597), CCDC90B₄₃₋₁₂₅-GCN4 N16V were cloned in pETHis_1a (G. Stier, EMBL Heidelberg) for overexpression with an N-terminal 6xHis-tag cleavable by TEV protease. DNA fragments encoding tagged variants of MpcC (CCNA_02554, YP_002517927), were synthesized by gene synthesis and cloned in pBXMCS4 (Thanbichler et al., 2007). For expression of HA-MpcC-(FLAG)₃ and HA-MpcC, the plasmids were transformed in *C. crescentus* NA1000 by electroporation (Ely, 1991). For transient expression in HEK293 cells, DNA fragments encoding MCU-HA (full length MCU with C-terminal HA-tag), FLAG-MCUR1 (full-length MCUR1 with N-terminal FLAG-tag) and MCUR1 variants fused to an N-terminal FLAG-tag were synthesized by gene synthesis and cloned in vector pCDNA3.1. The MCUR1- $\Delta\beta$ construct lacks residues 207-212, encoding the β -layer neck motif MVTKMQ. In MCUR1-GCN4pII, the fragment 231-321 was replaced by same number of residues of sequence (MKQIEDKIEEILSKIYHIENE-IARIKKL)₃-MKQIEDK derived from the trimeric GCN4pII variant (Harbury et al., 1993). Residues 224-321, 43-159 and 43-233 were deleted in constructs MCUR1- Δ stalk, MCUR1- Δ DR and MCUR1- Δ DR-Head, respectively. In MCUR1-Head90B, fragment spanning residues 162-234 of MCUR1 was replaced by residues 45-129 of the head region of CCDC90B.

3.5.2 Protein expression and purification

Recombinant MCU₇₅₋₂₃₃ was expressed in *E. coli* C41 strain. CCDC90B₄₃₋₁₂₅-GCN4 N16V, CCDC90B₄₃₋₁₂₅ and MCUR1₁₆₀₋₂₃₀ were expressed in *E. coli* ArcticExpress (DE3) cells. *E. coli* strains were grown in Luria broth (LB) supplemented with kanamycin at 37°C for C41 and ArcticExpress (DE3). Protein expression was induced with 1 mM isopropyl β -D-thiogalactoside (IPTG) at an optical density of OD₆₀₀ = 0.5. Following incubation for 4 h at 37°C for *E. coli* C41 and 24 h at 12°C for ArcticExpress (DE3), cells were harvested by centrifugation. The cell pellet was resuspended in lysis buffer containing 20 mM Tris, pH 7.5,

150 mM NaCl, 4 mM MgCl₂, DNaseI, 1 mM PMSF and cOmplete EDTA-free Protease Inhibitor Cocktail (Roche), and subsequently lysed by sonication.

CCDC90B₄₃₋₁₂₅-GCN4 N16V and CCDC90B₄₃₋₁₂₅ were purified under native conditions. Following centrifugation of cell lysate to remove cell debris, supernatant was loaded on a Ni-NTA Agarose column pre-equilibrated with Buffer A (20 mM Tris, pH 7.6, 150 mM NaCl). Bound proteins were eluted with a two-step gradient including a step of 5% buffer B followed by linear gradient of 5-100% buffer B (20 mM Tris, pH 7.6, 300 mM NaCl, 0.5 M imidazole). Protein containing fractions were dialyzed against Buffer A and incubated with TEV protease for His-tag cleavage. Cleaved protein was separated from 6xHis-tagged TEV protease and proteolytic fragments by re-loading the sample on Ni-NTA. Fractions containing the cleaved protein were pooled and purified to homogeneity by gel filtration on Superdex 75.

MCU₇₅₋₂₃₃ and MCUR1₁₆₀₋₂₃₀ were purified under denaturing conditions. Cell lysate was stirred in 6M guanidine hydrochloride (Gua-HCl) at room temperature for 1 h. Following centrifugation, the supernatant was loaded on Ni-NTA Agarose column equilibrated with 20 mM Tris, pH 8.0, 300 mM NaCl, 6 M Gua-HCl, and bound proteins were eluted with a linear gradient of 0-0.5 M imidazole in the same buffer. MCUR1₁₆₀₋₂₃₀ and MCU₇₅₋₂₃₃ were refolded by dialysis against buffer containing 20 mM Tris, pH 7.6, 150 mM NaCl and 1 mM DTT. Following cleavage by TEV protease, the protein was purified to homogeneity performing a second Ni-NTA column followed by gel filtration on Superdex 75, as described above.

For structural characterization of MCUR1₁₆₀₋₂₃₀ by NMR, *E. coli* C41 cells were grown in M9 minimal medium supplemented with ¹³C-glucose and ¹⁵N-ammonium chloride. Protein expression was induced at OD₆₀₀ = 0.6 with 1 mM IPTG. Following incubation at 20°C for 18 h, the cells were harvested by centrifugation. Protein purification was performed under denaturing conditions as described above.

3.5.3 CD spectroscopy

Circular dichroism (CD) spectra were recorded using a Jasco J-810 spectropolarimeter equipped with a JASCO-423S Peltier Controller. CD measurements were performed at a protein concentration of 0.5 mg/ml for MCUR1₁₆₀₋₂₃₀, and 0.2 mg/ml for CCDC90B₄₃₋₁₂₅-GCN4 N16V and CCDC90B₄₃₋₁₂₅ in 10 mM Tris, pH 7.5, 20 mM NaCl buffer using a cuvette

with a path length of 1 mm. Single CD spectra were recorded at a speed of 100 nm/min with a data pitch of 0.5 and a response time of 1s. Each spectrum represents the average of five scans corrected by the signal of buffer scan. Thermal melting curves were recorded by monitoring ellipticity at indicated wavelengths (208 nm, 216 nm, 220 nm or 224 nm) in a temperature range from 10-95°C applying a ramp of 0.5°C/min. Blank correction, smoothing of data, and calculation of molecular ellipticities and melting temperatures were performed using Spectra Manager Software (JASCO).

3.5.4 Microscale Thermophoresis (MST)

MST binding experiment was carried out by titrating dilution series of ligand against a fluorescently labelled biomolecule. 100 µl of 20 µM protein sample in 20 mM HEPES pH 7.6, 150 mM NaCl buffer was mixed with Alexa Fluor 647 red-NHS amine-reactive dye dissolved in 100 µl DMSO (430 µM). Reaction was carried out in the dark for 1 h. In the meantime, buffer exchange column (NanoTemper Gravity Flow Column B) was equilibrated with MST buffer (50 mM Tris, pH 7.6, 150 mM NaCl and 0.05% Tween-20). 200 µl of reaction mixture followed by 300 µl of MST buffer was applied to the equilibrated column, finally eluted with 600 µl of MST buffer and stored at -80 °C as 10 µl aliquots. Concentration of the labelled protein used for MST experiment was optimized to get initial fluorescence counts in the range 400-1000 units. 16 serial dilutions of non-fluorescent ligand were prepared in MST buffer and mixed with the labelled partner in 1:1 ratio. Tubes were incubated for 15 min, centrifuged at 13000 rpm for 10 min to remove aggregates and filled in Monolith NT 'Premium coated' capillaries. The experiment was carried out on MicroScale Thermophoresis instrument Monolith NT.115 (NanoTemper Technologies) with in-built MO.Control software. Analysis of data was done using MO.Affinity Analysis software.

3.5.5 SEC-MALS

Size Exclusion Chromatography and Multi Angle Light Scattering (SEC-MALS) experiments were performed to calculate the absolute molecular mass of proteins and their oligomeric states in solution using a 1260 Infinity II HPLC (Agilent) coupled to a miniDawn TREOS and Optilab T-rEX refractive index detector (Wyatt Technologies). Proteins were applied at a concentration of 5 mg/ml on an AdvanceBio SEC 130 Å (for MCUR1) or AdvanceBio SEC 300 Å (for CCDC90B) column equilibrated with 20 mM Tris, pH 7.5, 150 mM NaCl and 0.2% NaN₃ and separated at a flow-rate of 0.5 ml/min at 18°C. For MCUR1₁₆₀₋₂₃₀ analyses, 1 mM TCEP was

added to the buffer. Data analysis and molecular mass calculation was performed using ASTRA software package (Wyatt Technologies).

3.5.6 Mammalian cell culture and transfection

HEK293 cells were cultivated in DMEM containing 10% FBS and 2 mM glutamine. Cells were transiently transfected with pCDNA3.1 plasmid encoding MCU-HA and MCUR1-FLAG variants using Lipofectamine 2000 according to the manufacturer's protocol (Thermo Fisher Scientific).

3.5.7 Immunoprecipitation

HEK293 cells were harvested 20-24 h post-transfection and washed twice with ice-cold PBS. The cell pellet was resuspended in lysis buffer containing 20 mM HEPES, pH 7.3, 150 mM NaCl, 10 mM KCl, 1% Nonidet P-40, 8% glycerol, 1 mM PMSF and cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) and incubated for 10 min on ice. Following sonication, cell debris were harvested in a microcentrifuge at 13,000 rpm. The soluble fraction was incubated with Anti-FLAG M2 Magnetic Beads for 2 h at 4 °C. The beads were collected and washed three times in lysis buffer using a magnetic separator. Proteins bound to beads were eluted in 0.1 M glycine, pH 3.0 for 5 min at room temperature. Following addition of 1 M Tris, pH 8.5 for neutralization of samples, beads were separated. The protein-containing supernatants were concentrated using Amicon Ultra 0.5 ml 3K centrifugal filters, separated on a NuPAGE 4-12% Bis-Tris Protein Gel and blotted. Membranes were probed with anti-FLAG antibody (F7425, SIGMA) and anti-HA antibody (H6908, SIGMA), both produced in rabbit.

3.5.8 Mitochondria isolation and subfractionation

Cells were harvested 20 hours after transfection at $500 \times g$ for 4 min at 4°C and washed in ice-cold PBS. Total mitochondria was isolated using the Mitochondrial Isolation Kit for Cultured Cells (Thermo Fisher). 800 µl of Reagent A was added to the harvested cells and vortexed at medium speed. To this, 10 µl of Reagent B was added. The sample was incubated on ice for 5 min with vortexing at maximum speed every minute. Afterwards, 800 µl of Reagent C was added and components mixed by inverting the tube several times. After centrifugation at $700 \times g$ for 10 min to remove cellular debris, supernatant was transferred to a new tube and spun down at $12,000 \times g$ for 15 min. The pellet contained isolated mitochondria and the supernatant was the cytosolic fraction. 100 µl of supernatant was precipitated using ice-cold acetone

overnight at -80°C , washed with ice-cold 70% ethanol, dried and dissolved in 50 μl of 1x NuPAGE Sample Buffer (Thermo Fisher Scientific). Mitochondrial pellet was washed once in 500 μl Reagent C and collected by centrifugation at $12,000 \times g$ for 20 min. One tube containing the mitochondria was dissolved in 50 μl of 1x NuPAGE Sample Buffer for SDS-PAGE. The second was fractionated into OMM (outer mitochondrial membrane), IMM (inner mitochondrial membrane), and the matrix.

Mitochondrial pellet was resuspended in 800 μl of hypotonic shock buffer (5 mM sucrose, 5 mM HEPES, pH 7.2 with KOH and 1 mM EGTA) and subjected to osmotic shock for 10 min. 200 μl of high salt storage buffer (750 mM KCl, 100 mM HEPES, pH 7.2 with KOH and 2.5 mM EGTA) was added. Mitoplasts were sedimented by centrifugation at $20,000 \times g$ for 20 min. Supernatant containing OMM and IMS proteins was collected and precipitated using ice-cold acetone.

Proteins of IMM and matrix were separated using the alkaline carbonate extraction method. Mitoplasts were resuspended in 500 μl carbonate extraction buffer (120 mM Na_2CO_3 , pH 11.5 with NaOH) and incubated at 4°C for 2 hrs. Sample was ultracentrifuged in Beckmann table top ultracentrifuge TLA 100.3 rotor at $110,000 \times g$ for 2 hrs. The supernatant contains proteins extracted by carbonate, while the membrane pellet retains integral inner membrane proteins. All the samples were analyzed by western blot using anti-FLAG (1:2000) or anti-V5 (1:1000) antibodies, both produced in rabbit.

3.5.9 Subcellular fractionation of *C. crescentus* NA1000

C. crescentus NA1000 cells expressing tagged MpcC were grown in PYE medium to an OD_{660} of 0.8 and fractionated as described previously with slight modifications (Anwari, 2012). Pellet from 5 ml culture was resuspended in 280 μl of ice-cold spheroplasting buffer (10 mM Tris, pH 7.5, 0.75 M sucrose). Lysozyme was added to a final concentration of 100 $\mu\text{g}/\text{ml}$ and cells were incubated on ice for 2 min. Following addition of cOmplete EDTA-free Protease Inhibitor Cocktail (Roche), 25 $\mu\text{g}/\text{ml}$ DNaseI and 10 mM MgCl_2 , two volumes of ice-cold lysis buffer (1.5 mM EDTA, pH 7.5) was added slowly to the cells with constant mixing. Cell lysis was completed using glass beads and unbroken cells were removed by two centrifugation steps at 4000 rpm for 10 min. Performing ultracentrifugation at 50,000 rpm for 2 hrs at 4°C using a

TL100.3 rotor, the supernatant containing the soluble fraction (periplasm and cytoplasm) was separated from the crude membrane fraction present in the pellet.

Cell membranes were fractionated according to (Thein et al., 2010). The inner membrane fraction was solubilized by resuspension of the membrane pellet in solubilization buffer containing 50 mM Tris, pH 8.0, 2% (w/v) Triton X-100, 10 mM MgCl₂, and finally separated from the outer membrane fraction by centrifugation at 50,000 rpm for 2 h at 4 °C. The outer membrane pellet was washed in solubilization buffer followed by water and finally dissolved in Laemmli sample buffer. Proteins of the soluble and inner membrane fractions were precipitated with ice-cold acetone and solubilized in Laemmli sample buffer. Samples were analyzed by SDS-PAGE and western blotting using polyclonal anti-MpcC (1:1000), anti-TimA (1:2000), and anti-CpaC (1:1000, L. Shapiro, Stanford) antisera.

3.5.10 Proteinase K assay

C. crescentus NA1000 cells expressing HA-MpcC-(FLAG)₃ were grown in PYE medium to an OD₆₆₀ = 0.8. Spheroblasts were prepared as described above. Following lysozyme incubation, Proteinase K was added to the cell suspension at a concentration of 100 µg/ml. Following incubation on ice for 30 min, Proteinase K was inactivated with addition of 10 mM PMSF. Spheroblasts were fractionated as described and analyzed on a Western blot using rabbit anti-FLAG antibody (F7425, SIGMA) and anti-HA antibody (H6908, SIGMA).

3.5.11 Electron microscopy

For visualization of protein localization by Electron Microscopy (EM), *C. crescentus* NA1000 strain expressing HA-MpcC was cultivated in M2G medium to exponential phase. Protein expression was induced upon addition of 0.3% (w/v) xylose and cells were harvested 2 hrs after induction. Cells were washed twice in PBS and fixed in 0.1 M phosphate buffer, pH 7.4, containing 2% paraformaldehyde and 0.05% glutaraldehyde for 2 hrs at room temperature. After washing twice in 0.1% glycine, samples were fixed and prepared for cryosectioning as described (Tokuyasu, 1973). Pellets were mixed with 10% warm gelatin and solidified on ice. Cut into blocks of about 1 mm³, the solid mixtures were cryoprotected in 2.3 M sucrose at 4°C overnight. The infiltrated blocks were frozen on cryosectioning stubs in liquid nitrogen and sections of 55–70 nm were cut using a Leica Ultracut UCT microtome equipped with a Reichert FCS cryo attachment. After retrieval with a 1:1 mixture of 2% methyl-cellulose and 2.3 M

sucrose, the cryosections were placed on carbon/piloform-coated EM support grids and floated upside down in PBS at 40°C. For immunogold labeling, sections were incubated with rabbit anti-HA antibody (Clontech), followed by incubation with secondary goat anti-rabbit IgG-ultra small gold antibody (Aurion). The sections, contrasted in methyl cellulose/uranyl acetate, were analyzed on a FEI Tecnai G2 Spirit TEM (FEI, Hillsboro, Oregon, USA) operating at 120 kV. Images were taken with a Gatan Ultrascan 4000 (Pleasanton, CA, USA) camera at maximum resolution using manufacturer's software.

5 µl of MCUR1 fibril samples were directly applied to carbon/piloform-coated EM support grids at a concentration of 50 µg/ml (in buffer containing 20 mM Tris, pH 7.5, 150 mM NaCl and 1 mM TCEP), washed with water and coated with 1% uranyl acetate. After drying, grids were imaged as described.

3.5.12 NMR spectroscopy

All spectra were recorded on Bruker AVIII-600 and AVIII-800 spectrometers. ¹⁵N HSQC spectra were acquired over a temperature range from 298 K to 313 K. Diffusion ordered spectroscopy (DOSY) experiments were acquired to assess the translational diffusion times and obtain estimates of the effective molecular weight. Choosing the lower temperature, standard triple resonance experiments were acquired to perform backbone sequential assignment and 3D TOCSY spectra for sidechain assignment. This work was done by Dr. Murray Coles.

3.5.13 Crystallization and structure determination

Crystallization trials were set up in 96-well sitting-drop plates with drops consisting of 300 nl protein solution and 300 nl reservoir solution (RS), and reservoirs containing 50 µl RS. Crystals of selenomethionine-labeled Kcr-0859ΔTM were obtained with a RS containing 100 mM sodium cacodylate, pH 6.5, 30% (v/v) MPD and 5% (w/v) PEG 2000. Crystals of CCDC90B₄₃₋₁₂₅-GCN4_{N16V} were obtained using the Morpheus HT-96 screen (Molecular Dimensions), well F2. Prior to mounting, crystals of Kcr-0859ΔTM were transferred into a droplet of RS supplemented with 20% (v/v) glycerol for cryo-protection. All crystals were loop mounted and flash-cooled in liquid nitrogen. Data were collected at 100 K and a wavelength of either 1.07 Å (CCDC90B₄₃₋₁₂₅-GCN4_{N16V}) or 0.979 Å at the Selenium K-edge (Kcr-0859ΔTM) at beamline X10SA of the Swiss Light Source (Villigen, Switzerland), using a PILATUS 6M-F

hybrid pixel detector (Dectris Ltd.). All data were indexed, integrated and scaled using XDS (Kabsch, 2010), with the statistics given in Table 3.1.

For the phasing of the Selenomethionine-labeled Kcr-0859 Δ TM, we employed SHELXD (Sheldrick, 2008) for heavy atom location, locating six selenium sites in the asymmetric unit. After phasing and density modification with SHELXE, one Kcr-0859 Δ TM trimer could be traced with Buccaneer (Cowtan, 2006). The structure of CCDC90B₄₃₋₁₂₅-GCN4_{N16V} was solved by molecular replacement with MOLREP (Vagin and Teplyakov, 2000), using trimeric coiled-coil fragments of PDB entry 5APQ search models. The structure was completed using Buccaneer. Both structures were finalized in cycles of manual modeling with Coot (Emsley and Cowtan, 2004), and refinement with REFMAC5 (Murshudov et al., 1999). Refinement statistics are given in Table 3.1, together with PDB accession codes. This work was done together with Dr. Marcus Hartmann.

3.5.14 Homology modelling

The comparative homology model of MCUR1 (residues 167-336) was generated with Modeller (Sali et al., 1995) using CCDC90B₄₃₋₁₂₅ crystal structure as the template for head-neck segment. The backbone of coiled-coil stalk was built using fragments of identical periodicities from previously solved crystal structures.

Table 3.1: Data collection and refinement statistics

	Kcr-0859ΔTM	CCDC90B₄₃₋₁₂₅-GCN4_{N16V}
Wavelength (Å)	0.979	1.07
Space group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Cell dimensions (Å)	a=197.5, b=48.9, c=51.1	a=36.0, b=298.2, c=29.3
Monomers / ASU	3	3
Resolution range data collection (Å)	39.3 - 2.19 (2.32 – 2.19)	37.3 - 2.10 (2.23 – 2.10)
Completeness (%)	99.0 (95.7)	99.8 (98.8)
Redundancy	3.40 (3.27)	6.27 (6.46)
I/σ(I)	9.23 (1.74)	13.4 (1.89)
Rmerge (%)	8.6 (58.5)	8.9 (86.7)
CC(1/2)	99.8 (79.8)	99.9 (83.5)
Resolution range refinement (Å)	39.3 - 2.19 (2.25 – 2.19)	37.3 - 2.10 (2.15 - 2.10)
Rcryst (%)	24.3 (31.7)	23.1 (38.6)
Rfree (%)	27.3 (33.3)	25.7 (40.7)
RMSD Bond angles / lengths	1.04 / 0.0065	1.34 / 0.012
Ramachandran statistics (%)	100 / 0 / 0	97.8 / 2.2 / 0
PDB accession Code	6H9L	6H9M

CHAPTER 4

Investigating frameshift-resistant repeat amplification in functional proteins

Repetition is an important mechanism of gene evolution, producing proteins whose repeating units span all levels of complexity, from single residues to whole domains. In this chapter, we focus on a special type of repetition in protein evolution, the frameshift resistant repeats (FSR), which are simply generated by the repetition of n base-pairs, where n is not divisible by 3 and does not contain any stop codons. Repetitions of this type encode the same protein repeat of n residues in all three frames of equal sense. FSR repetition is common in many genomes across all branches of life. Among bacteria, they are found to be significantly enriched in cyanobacteria as well as opportunistic and pathogenic organisms. They are found both within existing genes, where they appear to be mostly unstructured and deleterious, and as new ORFs. While most of the latter seem to be purged quickly, some have clearly survived purifying selection and have become real genes encoding new proteins. Here, we examine the effects of frameshift resistant repeat amplification on the structure and function of existing and new proteins in bacteria. Through bioinformatic sequence analysis, we have selected a set of candidate FSR repeat containing proteins from two cyanobacterial species *Microcystis aeruginosa* and *Moorea producens*, both abundant in FSR repeat containing ORFs, for a structural and functional study. The FSR repeats in these proteins are found to be either inserted within a functional protein domain or replacing a part of it. Alongside, we characterize their non-FSR parent homologs for a comparative study. Finally, we check for the expression of FSR repeat containing ORFs in different bacterial species by transcriptome and proteome analysis.

4.1 INTRODUCTION

4.1.1 Defining tandem repeats

Tandem repeats (TRs) are nucleotide sequence units that are consecutively repeated two or more times in a DNA. They are often referred to as “satellite DNA”, and are abundant in both coding and intergenic regions. TRs are usually classified on the basis of unit length into three subcategories – microsatellites, minisatellites and satellite DNA. Microsatellites, also called simple sequence repeats (SSRs), contain a unit length between 1-10 nucleotides. Minisatellites have a unit length between 10-100 bp and have been successfully employed as markers for genetic profiling. TRs with even longer unit lengths of ~100-1000 bp are termed as satellite DNA. On the basis of sequence conservation, TRs can also be classified as perfect or imperfect/degenerated repeats (Zhou et al., 2014).

Whereas TRs are mostly found in non-coding regions, recent evidence suggests their sizeable presence in protein coding sequences. For instance, repetitive elements are present in nearly 17% of genes in the human genome (Gemayel et al., 2010; Jansen et al., 2012). Prokaryotic genomes also possess around 10% repetitive regions (van Belkum et al., 1998), which is a significant fraction considering their small sizes. Translation of TRs in coding regions results in amino acid tandem repeats, which organize as repetitive structural elements in proteins (e.g. leucine rich repeats, ankyrin repeats, coiled coils etc.). Tri- and hexa-nucleotide (or a multiple of 3) TRs have been specifically enriched as a result of the selection pressure to maintain the downstream region in frame. Although previously discarded as non-functional junk DNA, recent studies have highlighted the important role played by repeat duplication in genomes. Studies on human genome have shown that some TRs are hypermutable, i.e. prone to increase or decrease of TR copy number as a result of strand-slippage replication and recombination events, and therefore the cause of diseases such as fragile X syndrome (CGG trinucleotide repeats in 5' UTR of *FMRI*), Huntington's disease (polyQ expansion in exon 1 of *IT-15* gene) and spinobulbar muscular atrophy (polyQ expansion in *AR* gene) (Hannan, 2010). In bacteria, TR variations drive the rapid adaptation strategies which can range from evading host cellular immune response, tissue tropism and environmental stress tolerance (Gemayel et al., 2010, 2012; Zhou et al., 2014). While most of the repeat insertion events are deleterious, some with positive selection may eventually give rise to new proteins with unique functional properties.

4.1.2 Frameshift resistant tandem repeats

Frameshift resistant (FSR) repeats are a special subset of TRs, wherein the repetition of a certain unit of nucleotides results in the repeat of the same unit of amino acids in all three frames of equal sense. First reported by S. Ohno in the early 1980s, FSR repeat phenomena is a byproduct of the repeat unit having any length that is not a multiple of three and whose tandem repetition does not contain any stop codons (Ohno, 1983, 1984a, b). The inherent advantage of FSR repeats lies in their imperviousness to random base substitutions, insertions and deletions; hence the given name. As tandem FSR units are translated to identical polypeptides in all three frames, any mutation that is not a multiple of 3 can only cause a local perturbation with the original periodicity resuming thereafter. Even acquisition of internal stop codons by such random mutations can only silence one reading frame. Fig. 4.1A shows the schematic representation of FSR repeats. An example illustrates how the repetition of a four nucleotide sequence TTTA gives a repeat of FIYL residues in all three reading frames (Fig. 4.1B). For certain FSR repeats, the repeating unit encodes a palindromic sequence (Fig. 4.1C), which eventually translates into the same amino acid sequence in all six reading frames. Sequence conservation at both the DNA and protein levels points out that this repetition is a very recent evolutionary event.

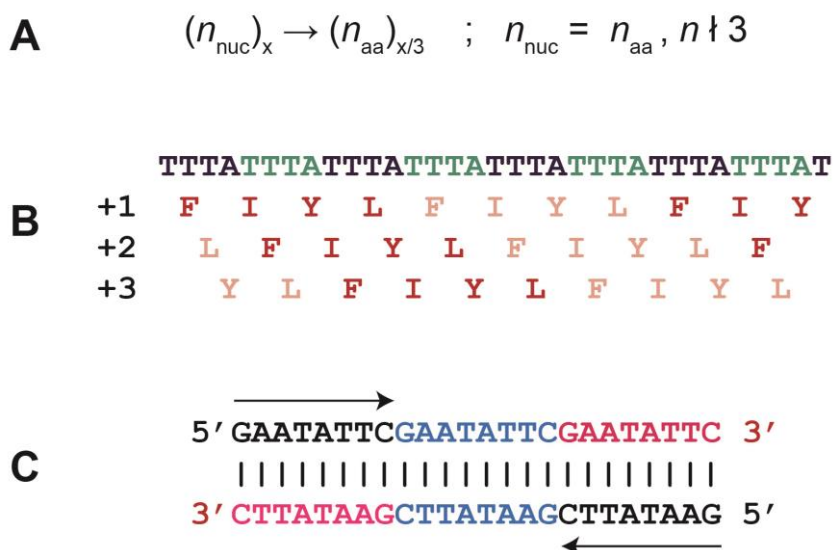


Figure 4.1: Schematic representation of Frameshift resistant (FSR) repeats. (A) Frameshift resistant repeats follow the basic rule: a repeat of n nucleotides translates to a repeat of n amino acids given that n is not divisible by 3. (B) Tandem repetition of 4 nt FSR sequence TTTA gives a repeat of 4 aa FIYL in all three frames of equal sense. (C) A palindromic FSR repeat sequence GAATATTC gives the same polypeptide sequence in all six frames.

Even though this idea was proposed more than three decades ago, no systematic study has been conducted to understand the FSR repeat occurrence in genomes. Briefly encountered while analyzing *Nematostella vectensis* (a small sea anemone) protein coding sequences, a previous study reported that 18% (or 806 TRs) of all the identified tandem repetitive elements are FSR repeats (Naamati et al., 2009). In collaboration with Mateusz Korycinski, we have undertaken a detailed study to investigate the abundance, characteristic features and structural and functional implications of FSR tandem repeats in proteins across organisms.

4.1.3 Mechanisms of tandem repeat instability

Tandem repeats are hypermutable; in humans, microsatellites mutate at a rate of 10^{-3} to 10^{-4} per locus per cellular generation (Weber and Wong, 1993) as compared to the mutability rate of 10^{-8} per generation for single nucleotide substitutions (Drake et al., 1998). TR polymorphisms are primarily the result of addition or deletion of repeating units, instead of nucleotide substitutions. Two major mechanisms have been proposed to explain TR instability: strand-slippage replication and unequal recombination (Levinson and Gutman, 1987; Pâques et al., 1998; Bichara et al., 2006).

In the strand-slippage replication (or DNA slippage or slipped-strand mispairing) model, mispairing occurs between the template and nascent strands. The newly synthesized strand detaches from the template and reattaches at another position, resulting in the formation of an unpaired repeats hairpin structure either on the nascent or the template strand. As DNA replication proceeds on this structure, it leads to TR expansion or contraction. When the bulge is present on the template strand, it will result in TR unit loss in the newly synthesized DNA. In the vice-versa scenario, TR expansion occurs. As a third mechanism, TRs may alter the location of Okazaki initiation which then influences the formation of hairpins and leads to an asymmetric sequence replication (Pearson et al., 2005). The rate of TR gain/loss may also depend on the relative position of TR i.e. either on the leading or the lagging strand, due to different fork stalling effects during replication (Kang et al., 1995; Aguilera et al., 2013). More complex models even include DNA double-strand break repair mechanism. For stalled replication machinery at the repeats, breaks can be repaired by single-strand annealing to homologous TRs on either the same strand (TR loss) or on the other strand (TR gain).

Apart from strand-slippage replication, recombination events (including unequal crossovers and gene conversion) are another mechanism to explain TR instability. During homologous recombination, ssDNA anneals to the neighboring intact sister chromatid as a template for repair; however, it may be shifted by one or multiple TR units. While several studies suggest that strand-slippage replication is generally associated with microsatellite (1-10 bp) instability and recombination events dominate minisatellite (10-100 bp) instability, the underlying precise molecular mechanisms remain unclear.

A number of factors influence TR instability and variability. Prominent among them are: (i) number of repeat units; the greater the number, the higher the TR instability, (ii) length of the repeat unit and (iii) repeat purity; the longer and purer repeating unit has a higher mutation frequency. Legendre et al. (2007) reported that the presence of a single repeat impurity in a pure polyGT tract increased its stability by five-fold. Base composition, or GC content, is also an important factor in determining TR stability (Gragg et al., 2002). Various cellular processes such as high transcription rates and external factors such as environmental stresses can also enhance repeat mutability (Wierdl et al., 1996; Rosenberg, 2001; Schmidt and Mitter, 2004; Mittelman et al., 2010).

4.1.4 Functional impact of tandem repeats

Variable TRs can influence gene expression and function depending on their genomic localization either in non-coding or coding regions. A number of recent studies have suggested that repeat variation in promoter and cis-regulatory elements could be an evolutionary conserved mechanism of regulating gene expression. One frequently observed method is the alteration in number of transcription factor binding sites. Two prominent examples of this mechanism are: (i) TAAA repeat number variation in *Neisseria meningitidis* nadA promoter element influences gene expression of NadA adhesin protein (Martin et al., 2005), and (ii) variable TCC repeats in human EGFR promoter region alter the number of Sp1 (transcriptional regulator) binding sites (Johnson et al., 1988). Variable TRs can also affect the spacing between critical promoter elements. Furthermore, tandem repeated tracts in human promoter regions were found to inhibit nucleosome formation promoting open chromatin structures which clearly affect gene transcription (Vinces et al., 2009). Another interesting study reported that variable TRs in intronic regions can regulate gene expression through alternative mRNA splicing (Shang et al., 2011).

Intragenic tandem repeats are highly enriched in proteins associated with cellular signaling, communication and cell-surface localization, and underrepresented in proteins associated with cellular metabolism and bioenergetics. Minisatellite abundance in cell surface genes (LPS, adhesins, pili, and fimbriae) is found to be evolutionarily conserved from bacteria to humans, whose polymorphisms in fact generate functional diversity (Legendre et al., 2007). For example, *S. cerevisiae FLO1* gene, encoding a cell-surface adhesin, contains tandem repeated units of approximate length 100 bp. Phenotypic studies identified a positive correlation between *FLO1* TR repeat length and the adhesion intensity (Verstrepen et al., 2005). Similarly, TR enrichment in *S. cerevisiae FLO11* gene increases the buoyancy of yeast strains. *FLO11*, also encoding an adhesin, is involved in biofilm formation. Glycosylation of its Ser/Thr-rich TRs renders the cell wall hydrophobic giving a floating biofilm phenotype (Fidalgo et al., 2006).

Variable intragenic TRs can even mediate phase variation, a strategy particularly employed by pathogenic bacteria to evade host defense mechanism. For example, modulations in the CTCTT pentanucleotide repeat (also Frameshift resistant) at the 5' end of the coding region of surface membrane proteins in *N. gonorrhoeae* can lead to proper (ON state) or improper (OFF state) translation of proteins (Stern et al., 1986). Switching between these two states during infection generates the population phenotypic variability that allows certain individuals to survive host immune response. TRs can form extended structures in proteins such as in keratins, collagens, anti-freeze proteins, spider silk and the FG-rich proteins of the nuclear pore. Trinucleotide polyQ repeats are involved in neurodegenerative disorders in humans; longer TRs correlate with an early disease onset and severe symptoms (Gatchel and Zoghbi, 2005). Intragenic TRs have also been implicated in generating natural variation such as circadian clock tuning in fruit flies (Johnsen et al., 2007) and skeletal morphology in animals (Fondon and Garner, 2004). These studies clearly point out that abundance of TRs in functional protein-coding regions can be involved in beneficial roles. In this study, I intend to focus only on ORFs encoding intragenic frameshift resistant micro and minisatellite repeats.

4.1.5 Aims of this study

The primary objective of this study was to understand the structure and role of tandem FSR repeats in protein evolution. For this, I selected a set of FSR repeat containing proteins for structural, biophysical and functional characterization. For a comparative functional analysis, I simultaneously expressed and characterized their non-FSR parent homologs, which share high sequence identity at both the protein and DNA level and can be found in the same or a closely related strain. Next, I aimed to experimentally verify the expression of FSR repeat containing proteins *in vivo* using whole-cell transcriptome and proteome analysis. With the above experimental strategies in mind, I intended to address the following questions: Are FSR repeat containing proteins expressed in organisms? How do the proteins structurally accommodate FSR repeat insertions and domain replacements? Do these events affect protein stability and function compared to non-FSR parent homologs? What is the general role of FSR repetitions in protein evolution?

4.2 RESULTS

4.2.1 Expression and characterization of frameshift resistant (FSR) repeat containing proteins and their non-FSR homologs

An extensive bioinformatic search was carried out to identify and collect sequences containing FSR repetitive elements in model organisms covering all phylogenetic groups including 1430 bacteria and archaea, 67 animals, 29 plants and 260 fungal species (work done by Mateusz Korycinski). FSR repeats were found in all kingdoms of life and no correlation was observed between proteome size and the number of ORFs containing FSR repeats. However, a few organisms such as *Nematostella vectensis*, *Microcystis aeruginosa*, and *Burkholderia pseudomallei* displayed a significant enrichment of FSR elements compared to other species, with over 100 FSR repeat containing ORFs per genome. Primarily, these repeats are short microsatellites ranging from 4 to 8 nucleotides, with an overrepresentation of 7 nt periodicity. Regarding their specific localization within ORFs, no clear preference was observed. They could occupy any position starting from the 5' to the 3' end of an ORF. Whereas the majority of FSR repeats span only 10% or less of the total gene length, a few examples were found to be completely built of such repetitive elements.

In order to understand the effect of FSR repeat amplification on the structure and function of existing and new proteins, we characterized a few representative examples. Proteins listed in Table 4.1 were selected for structural and functional characterization. Three different categories of FSR repeats were classified in proteins: (i) FSR repeats found to be inserted within a functional protein domain (e.g. NADH dehydrogenase) or between two domains in a multi-domain protein (e.g. S/T kinase), (ii) FSR tandem repeat sequences replacing a part of functional protein domains, usually the C-terminal region (e.g. restriction endonuclease-like and HIRAN domain proteins) and (iii) FSR repeat duplications constituting an entire new ORF. For most FSR repeat proteins, closest non-FSR orthologs/paralogs were also characterized for a comparative study. The strategy for selection of most proteins was simply based on the presence of existing well-defined assays which can be utilized to verify their residual activity.

	Reference ID	Name	Organism	
Frameshift repeats as domain insertions				
1.	EGJ33944.1	serine/threonine protein kinase/putative ATPase	<i>Moorea producens</i> 3L	FSR
2.	WP_008190547.1	NADH dehydrogenase	<i>Moorea producens</i> 3L	FSR
3.	WP_017715270.1	nitroreductase family protein	<i>Oscillatoria</i> sp. PCC 10802	non-FSR
Frameshift repeats as domain insertions				
4.	WP_012265514/ BAG02178.1.1	Uma2 family endonuclease MAE_23560	<i>Microcystis aeruginosa</i> NIES-843	FSR
5.	EPF16067.1	hypothetical protein MAESPC_05199	<i>Microcystis aeruginosa</i> SPC777	non-FSR
6.	EGJ29873.1	HIRAN domain protein LYNGBM3L_59020	<i>Moorea producens</i> 3L	FSR
7.	EGJ29882.1	HIRAN domain protein LYNGBM3L_59090	<i>Moorea producens</i> 3L	non-FSR
Frameshift Repeats spanning whole domains				
8.	EGJ30341.1	hypothetical protein LYNGBM3L_51520	<i>Moorea producens</i> 3L	FSR
9.	ADI66200.1	hypothetical protein Aazo_5130	<i>Nostoc azollae</i> 0708	FSR

Table 4.1: List of FSR repeat containing sequences selected for structural and functional characterization

Peptide name	Protein	Organism	Peptide sequence	Crystallization trial	Secondary structure in CD
LSVISYQ	MAE_23560 (BAG02178.1.1)	<i>Microcystis aeruginosa</i> NIES-843	KSVISYQ LSVISYQ LSDVSFK LKLF	8mg/ml in 25% Acetonitrile; no crystals	appeared β -character in 10% acetonitrile, but did not unfold
RTYV			(RTYV) ₈	No crystals	unstructured in CD; slight induction of alpha by addition of 20% TFE
ASRIAHR	YP_991638.1	<i>Burkholderia mallei</i> SAVP1	ASVHAHR -(ASRIAHR) ₄	crystals diffract at best to 8Å	unstructured in CD
VTSHKSQ	YP_034083.1	<i>Bartonella henselae</i> str. Houston-1	(VTSHKSQ) ₅	No crystals	unstructured in CD
TSNIKHQ	WP_053093658.1	<i>Proteus mirabilis</i> 646- PMIR	KQ-(TSNIKHQ) ₅	No crystals	unstructured in CD

Table 4.2: List of FSR repeat peptides set-up for crystallization

Appropriate synthetic genes for selected proteins were cloned in vectors allowing for overexpression of recombinant protein with an N-terminal His-tag or with a solubility enhancing SUMO fusion domain. Proteins were expressed using *E. coli* C41 (DE3) or *E. coli* ArcticExpress (DE3) cells, where the latter expresses cold-adapted chaperones which assist in protein folding at low temperatures. The proteins were purified optionally under native or denaturing conditions with nickel affinity chromatography. After refolding, proteins were cleaved with TEV- or SUMO-proteases and set up for crystallization. Refolding conditions included variation of buffer components, pH, salt concentration, temperature, surfactants or non-detergent sulfobetaines, etc. Detailed studies of candidate FSR and their non-FSR parent homologs (see Table 4.1) are described below.

4.2.1.1 FSR repeats inserted between functional protein domains

1. NADH dehydrogenase (*Moorea producens* 3L)

The flavoprotein NADH dehydrogenase (WP_008190547) from cyanobacterium *Moorea producens* 3L belongs to the Nitro-FMN-reductase superfamily. Proteins of this family are usually found to be homodimers and catalyse the reduction of nitrogen containing compounds using NAD(P)H as an electron donor in an obligatory two-electron transfer utilizing FMN or FAD as a cofactor. *M. producens* NADH dehydrogenase features a 32 residue insertion (Fig. 4.2A), derived from the combination of two tandem repeats each of a 25 bp FSR repeat (“tgggggaaaccacggcagtcgctca”) and a 19 bp FSR repeat (“ccaagaccgcgctgctc”), in the extended surface loop covering the cofactor binding site. The total length of insert (96 bp), being a multiple of 3, precludes a frameshift mutation for the C-terminal part of the functional domain. For this protein, the closest sequence homolog found using a BLAST search (at the time of conceptualization) is a nitroreductase family protein present in *Oscillatoria* sp. PCC 10802. Both proteins share 67% amino acid sequence identity. Recent genome annotation efforts have also identified a nitroreductase in *Moorea producens*, which shares 80% sequence identity to its FSR homolog. Both non-FSR nitroreductase homologs completely lack the FSR repeat insertion, while the flanking amino acids are nearly identical. It is interesting to note here that we did not find any other close relatives of *M. producens* NADH dehydrogenase which contain FSR repeat insertion. Bioinformatics search for 25 bp and 19 bp FSR minisatellites in *M. producens* identified a hypothetical protein (WP_008184368) which contains a tandem of 25 bp, 19 bp and 17 bp repeats constituting the major chunk of its ORF.

As identical domain boundary (25 bp to 19 bp transition) was observed in NADH dehydrogenase, it can be postulated that this FSR insertion could have occurred by recombination or retrotransposition activity.

For structural characterization, both the FSR (*M. producens* 3L) protein and its non-FSR counterpart (*Oscillatoria* sp. PCC 10802) were recombinantly expressed in *E. coli*. The non-FSR nitroreductase purified as a yellow-coloured soluble protein with bound flavin cofactor. Far-UV CD spectroscopy shows that it primarily adopts an α -helical secondary structure with characteristic minima at 208 nm and 222 nm, and melts cooperatively in a single step at approximately 69°C. (Fig 4.2D) The non-FSR protein crystallized as yellow cuboids which diffracted to a resolution of 1.2 Å. Its structure was solved by molecular replacement using the crystal structure of a putative nitroreductase from *Exiguobacterium sibiricum* (PDB 3GE6). The protein forms a homodimer, with two such dimers occupying the asymmetric unit cell of the crystal (Fig. 4.2B). Each monomer is bound to an FMN cofactor molecule, with the loop (the one where insertion occurred in FSR homolog) covering the FMN binding site. Only K199 and R204 residues of this loop, located C-terminal to the site of FSR insertion (marked with arrow in Fig. 4.2C), are involved in coordinating FMN.

In contrast to the well-expressed, soluble non-FSR nitroreductase, FSR repeat containing protein was primarily expressed in inclusion bodies from which it was impossible to obtain folded soluble material for further biophysical or structural characterization. Repeated attempts to refold the protein with varying buffer conditions proved to be futile. It has been previously shown that FMN binding in the pocket of NADH dehydrogenases can initiate the proper folding process. Therefore, we tested if the protein could be refolded upon the addition of flavin cofactors FMN or FAD. On-column refolding in the presence of excess amount of flavin cofactors eluted only unfolded polypeptide as confirmed by secondary structure estimation using circular dichroism spectroscopy.

As a next step to gauge the effect of tandem frameshift resistant repeat insertion in the cofactor binding site, we deleted the FSR repeat from *M. producens* NADH dehydrogenase. The result was a stable, yellow-coloured flavin-bound protein. The purified deletion variant was α -helical as seen from far-UV CD spectra and did not completely melt even upto 95°C (Fig. 4E).

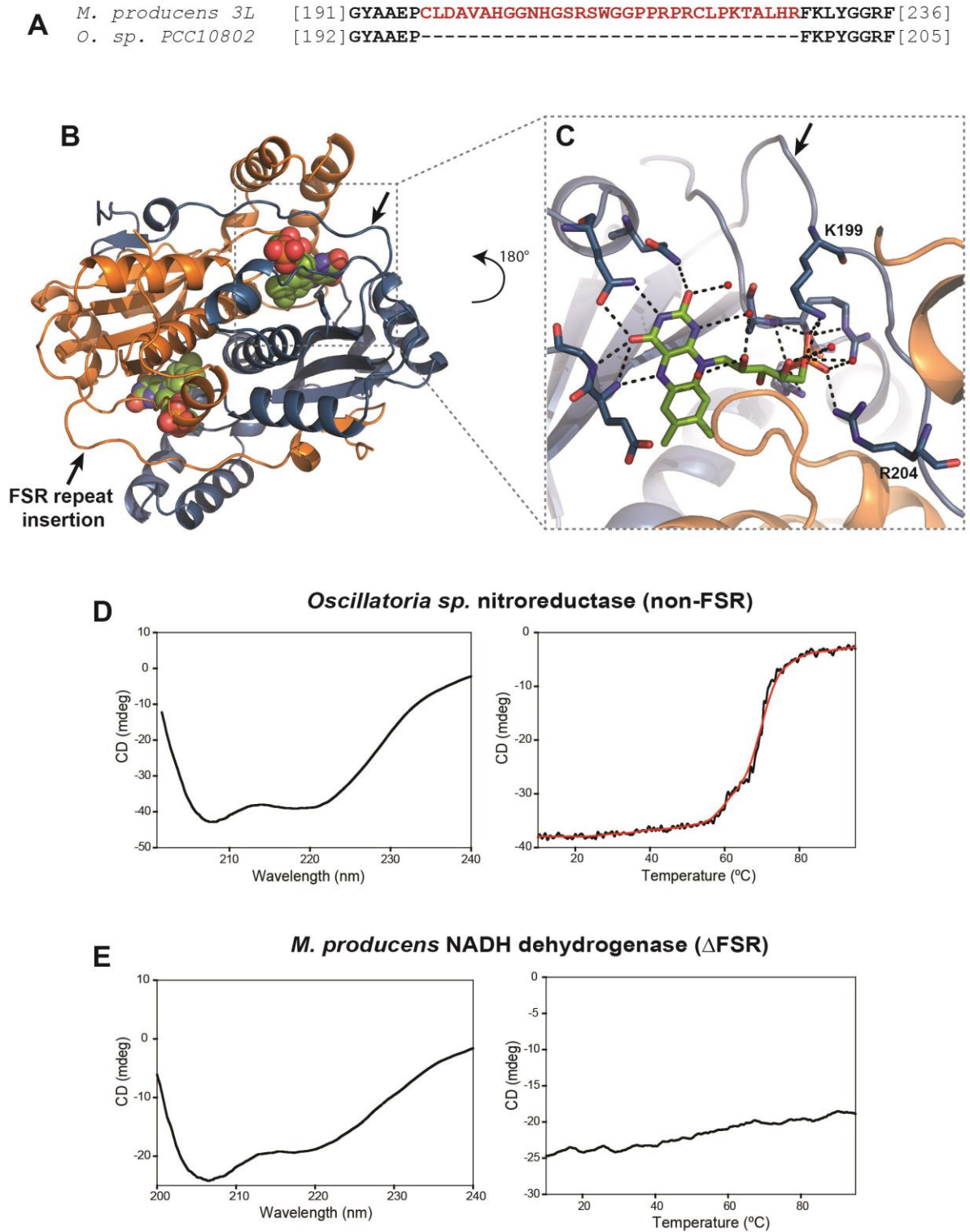


Figure 4.2: **FSR repeat insertion in NADH dehydrogenase.** **(A)** Sequence comparison of FSR and non-FSR nitroreductases from *Moorea producens* 3L (WP_008190547) and *Oscillatoria sp.* PCC10802 (WP_017715270) respectively, at the site of repeat insertion. **(B)** Crystal structure of *O. sp.* PCC10802 nitroreductase homodimer bound to FMN cofactor molecules. Individual chains are coloured in blue and orange. Black arrows point at the site of FSR repeat insertion. **(C)** Enlarged view of the surface covering the cofactor binding site. Hydrogen bond interactions are shown for FMN molecule. Small red spheres are water molecules. **(D)** and **(E)** CD and melting curves for *O. sp.* PCC10802 nitroreductase and *M. producens* NADH dehydrogenase Δ FSR respectively. Red line shows the best fit smooth curve.

This result suggests that the 32 residue FSR insert in the loop inhibited the binding of flavin cofactor molecule and proper folding of NADH dehydrogenase, resulting in the localization of the expressed unfolded polypeptide in inclusion bodies. Our assumption was further confirmed by insertion of the exact FSR repeat sequence in the non-FSR *Oscillatoria sp.* homolog followed by tests for stability of the protein in solution. As expected, most of the protein ended up in inclusion bodies with no possible recovery of well-folded functional protein. The above results clearly show that FSR repeat insertion in *M. producens* NADH dehydrogenase resulted in a non-functional variant.

2. Serine/threonine protein kinase /putative ATPase (*Moorea producens* 3L)

A second interesting candidate is a 239 kDa single-chain multi-domain serine/threonine protein kinase signal transduction protein LYNGBM3L_20360 (EGJ33944.1) from *M. producens* 3L. Seven repeats of FSR heptad sequence “QSAISNQ” have been inserted between the AAA+ ATPase and TPR domains of this multi-domain protein. Similarly to dehydrogenase, 147 bp (multiple of 3) insertion in this kinase prevents frameshift mutation in the translation of essential C-terminal functional domains. Whereas homologs are found broadly distributed in other cyanobacteria, this FSR insertion is unique to *Moorea*. As the heptamer FSR sequence is strongly predicted to form a coiled coil, the underlying question here is: Do the tandem repeats of this insert possess the ability to assemble into a folded coiled-coil domain and if so, does it lead to a change in the oligomeric state of this multi-domain protein?

To answer the above questions, we aimed at the structural and biophysical characterization of the FSR insert. I fused the seven heptad repeats of FSR insert in between stabilizing GCN4-N16V coiled-coil adaptors. To address the two possible heptad registers of the FSR insert, I designed two different constructs: QSAISNQ with Q occupying the *a* and I at the *d* position of the heptad repeat, and ISNQQSA with I and Q at *a* and *d* positions respectively. Both constructs were recombinantly expressed in *E. coli* C41 cells, purified under denaturing conditions and refolded. Whereas the QSAISNQ construct could be refolded in the presence of high salt buffer (20 mM Tris, pH 7.6 and 500 mM NaCl), the other construct ISNQQSA could not be refolded to yield a stable protein indicating the former to present the right coiled-coil register. QSAISNQ purified protein displayed a typical α -helical spectra (Fig. 4.3). Upon thermal melting, the protein showed a gradual decline in the CD ellipticity with residual α -helical character even at 95°C, likely originating from the GCN4 adaptors which have a melting point higher than 95°C.

The observed decrease in α -helical signal clearly corresponds to the unfolding of central FSR insert.

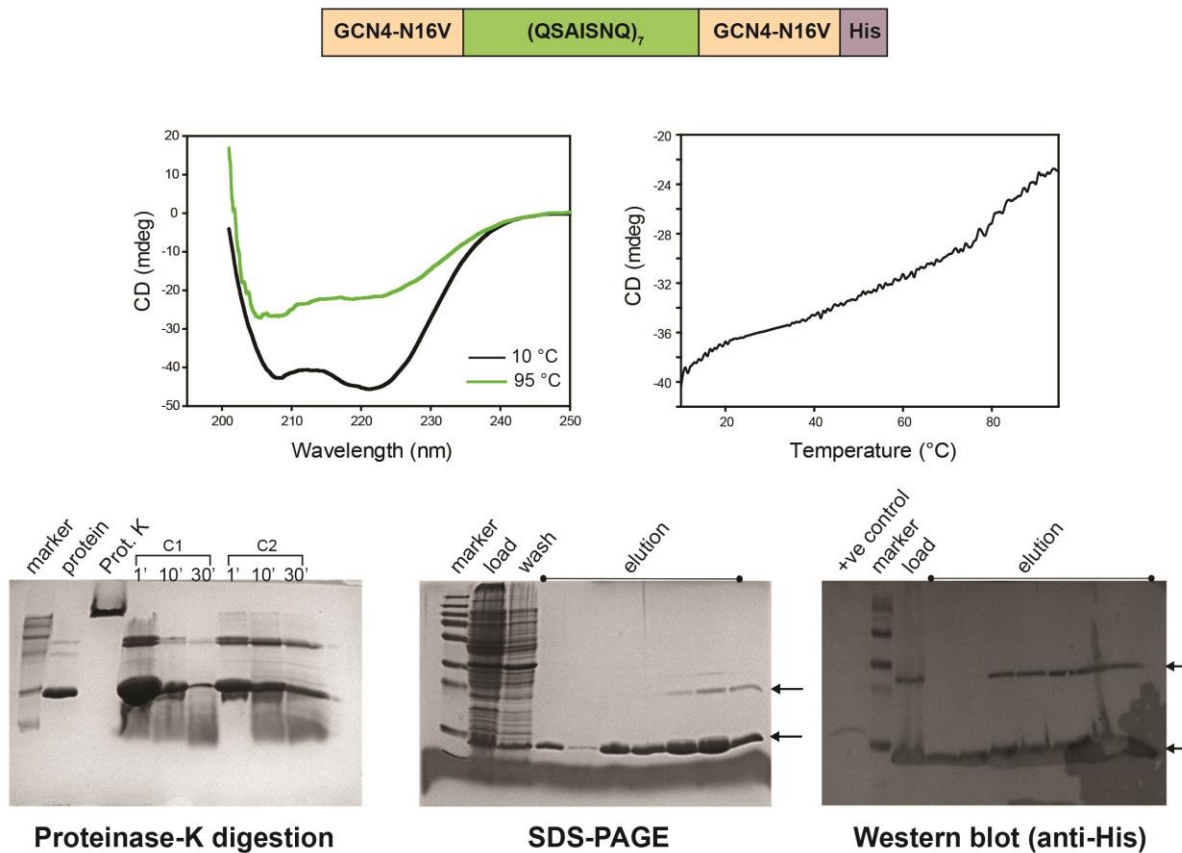


Figure 4.3: **Characterization of S/T kinase FSR repeat insert.** **(Top)** Schematic representation of the designed construct. Seven tandem repeats of the sequence QSAISNQ were fused between two GCN4-N16V coiled-coil adaptors. **(Middle)** Far-UV CD and melting curves. **(Bottom)** Proteinase K limited proteolysis of the refolded protein at different Prot. K concentrations C1 and C2 for increasing time intervals (C1 = 100 μ g/ml, C2 = 300 μ g/ml). SDS-PAGE and Western blot analyses of the purified protein denatured in 6M guanidinium chloride. (+ve control is a random purified protein containing an N-terminal His tag).

Proteinase K limited proteolysis experiment also confirmed that the insert was folded, as the protein was resistant to complete digestion even after 30 min of incubation. Furthermore, we observed on SDS-PAGE that the protein formed a stable oligomer which did not dissociate even in the presence of 6M guanidinium chloride. Finally, we set-up the protein for crystallization at different concentrations in the range of 1-10 mg/ml; however, we could not obtain any diffracting crystals. Our results suggest that the heptanucleotide FSR tandem repeat insertion in the multi-domain signal transduction protein is likely to assemble as a folded coiled coil.

4.2.1.2 FSR repeats replacing part of a functional protein domain

1. Restriction endonuclease-like hypothetical MAE_23560 (*M. aeruginosa* NIES-843)

Among the identified FSR repeat-containing sequences, we frequently found examples where FSR tandem repeats replaced a part of functional protein domain. One such example is a putative Uma2 family restriction endonuclease MAE_23560 present in a cyanobacterium *Microcystis aeruginosa* NIES-843. Members of this endonuclease family, generally annotated as hypothetical proteins, have greatly expanded in a number of cyanobacterial species. A non-FSR homolog of MAE_23560 with near sequence identity at the amino acid as well as DNA level, MAESPC_05199, exists in a closely related strain *M. aeruginosa* SPC777. Two full repeats of LSVISYQ replace the C-terminal residues of the endonuclease in MAE_23560. This FSR insert, also follows the coiled-coil heptad pattern where leucine and isoleucine occupy the *a* and *d* positions respectively. Other strains of *M. aeruginosa*, PCC 9809 and PCC 9432 and *M. panniformis* FACHB-1757 share this truncated ortholog. The FSR insertion LSVISYQ in all these strains occurred at the same hotspot, but the total repeat length slightly varies.

The closest structural homolog for MAE_23560, sharing 54% sequence identity, is a putative nuclease (PDB 3OT2) from *Anabaena variabilis* ATCC 29413 belonging to the DUF820 protein family. Visualizing its crystal structure, we observed that FSR repeat in MAE_23560 replaces the long helix in the fold (yellow α -helix in Fig. 4.4A) with truncation of the long C-terminal chain (towards red). We proposed that LSVISYQ repeat could form an α -helical coiled coil and stabilize the truncated construct. For structural characterization, the two proteins MAE_23560 (FSR) and MAESPC_05199 (non-FSR) were recombinantly expressed as SUMO-fusions in various *E. coli* strains. In contrast to its soluble, well-folded non-FSR ortholog, protein MAE_23560 was very difficult to handle. Although I could once obtain a small amount of soluble protein (Fig. 4.4B), it was not reproducible. Far-UV CD analysis of the purified SUMO-cleaved FSR protein showed a mixture of α -helical and β -strand character, which unfolded at temperatures $> 80^\circ\text{C}$ (Fig. 4.4C).

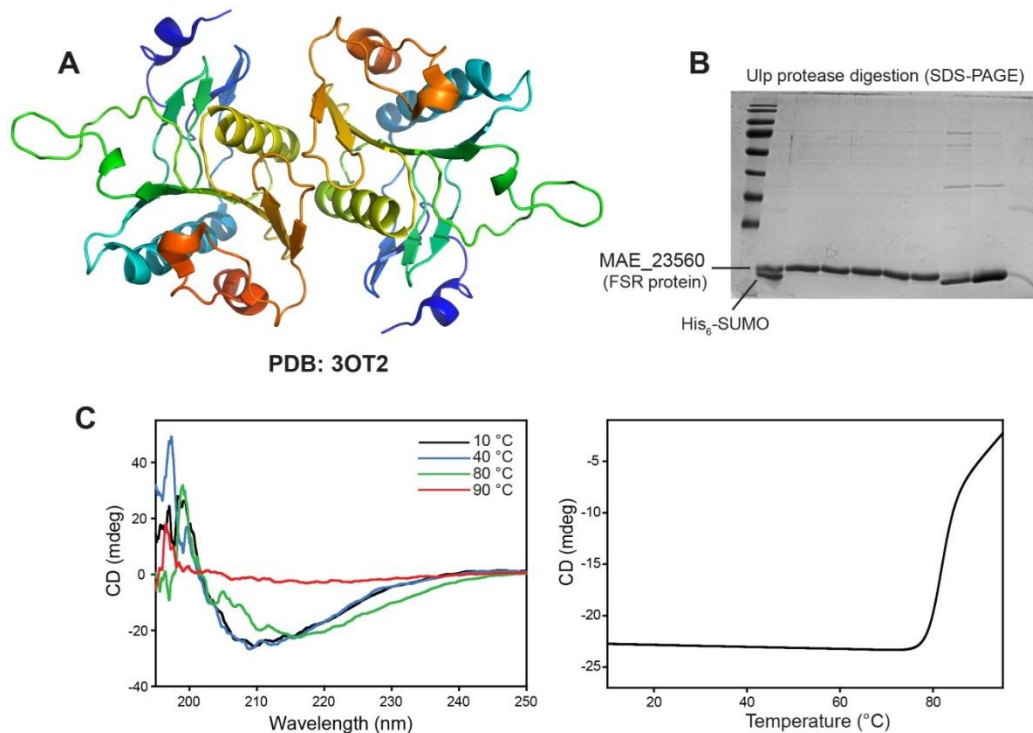


Figure 4.4: **Biophysical characterization of SUMO-MAE_23560.** (A) Closest structural homolog of MAE_23560 – a putative nuclease belonging to DUF820 (Ava_3926) from *Anabaena variabilis* ATCC 29413 (PDB ID 3OT2). FSR tandem repeat LSVISYQ in MAE_23560 replaces the corresponding C-terminal region (coloured yellow-to-red) that forms the dimer interface. (B) SDS-PAGE analysis of Ulp proteolysed SUMO-MAE_23560. (C) CD and melting curves for MAE_23560.

2. HIRAN domain protein LYNGNBM3L_59020 (*Moorea producens* 3L)

A similar example of a FSR repeat replacing part of a functional domain belongs to *Moorea producens* 3L. The parent non-FSR homolog consists of an N-terminal segment of a HipA domain fused C-terminally to a HIRAN domain. While HIRAN domain is involved in recognizing features associated with DNA damage or stalled replication forks (Iyer et al., 2006; Chavez et al., 2018), HipA which possesses a serine/threonine kinase-like fold is proposed to affect translation by phosphorylating EF-Tu thereby promoting cell persistence (Schumacher et al., 2009). A copy of this protein, containing a FSR replacement, is present in the same organism ~ 10kb upstream (both on the minus strand) of the non-FSR paralog. Tandem repetition of heptad FSR repeat LSAISYQ with a helical secondary structure propensity replaces the last β -strand of DNA-binding HIRAN domain (PDB 3K2Y); this would correlate into a fold change. We find that the FSR repeat LS(A/V)ISYQ which represents one of the most prominently identified frameshift resistant insertion sequence in cyanobacterial species conforms well to a heptad coiled-coil repeat pattern. An identical result was seen here during

protein expression as for the previous case: while parent non-FSR homolog was soluble, well-folded and crystallizable, FSR paralog was primarily expressed in inclusion bodies from which stable folded product could not be recovered. It seems likely that FSR insertion in the *M. producens* 3L HIRAN domain protein strongly affects the folding of the protein thereby disrupting its function.

4.2.1.3 FSR repeats spanning whole protein domain

Among the collected set of FSR sequences, we identified ORFs encoding hypothetical proteins which were totally composed of frameshift resistant tandem repeats. Two examples (see Table 4.1) from *Moorea producens* 3L and *Nostoc azollae* 0708 were selected for structural and biophysical studies. While SUMO-fusions of these proteins were slightly soluble, they co-purified with a large number of non-specifically bound proteins. Even a combination of IMAC, ion-exchange chromatography, HIC, SEC, affi-chromatography (Blue, Red) could not yield pure protein. These results suggest that the purified products were primarily unfolded with a high tendency for aggregation (work done together with Kerstin Bär).

4.2.2 Crystallizing FSR repeat peptides

As a next step, we obtained chemically-synthesized peptides (GeneCust, Luxembourg) for a few FSR repeat sequences, which were mostly heptad repeat based in conformity with the periodic coiled-coil pattern, and set them up for crystallization. The detailed list of these peptides can be found in Table 4.2. For most peptides, either we did not obtain any crystals or only poorly diffracting crystals.

It is interesting to note here that some of the FSR peptides, e.g. QSAISNQ, LSVISYQ, ASRIAHR and TSNIKHQ, encode a special heptad repeat pattern $a(b')-b(a')-c(g')-d(f')-e(e')-f(d')-g(c')$ with two hydrophobic seams offset by 1 residue. Such coiled-coil helices with broader interfaces are called “bifaceted” and have been described in detail in Chapter 1 (Walshaw et al., 2001; Lupas and Bassler, 2017; Woolfson et al., 2017). Three possible bifaceted coiled coils are denoted as Type I (seams share a common residue), Type II (adjacent seams) and Type III (seams separated by one residue). Our selected FSR repeats fall in the Type III category. As a consequence of wide angular separation of the seams, bifaceted helices can produce structures ranging from large diameter closed tubes (or α -cylinders) to fully open α -sheets.

4.2.3 Proteome analysis to detect FSR repeat proteins

Are FSR proteins expressed *in vivo*? To answer this question, we conducted a total protein analysis for three different organisms showing a high incidence of FSR repeat proteins: *Microcystis aeruginosa* NIES-843, *Burkholderia thailandensis* E264 and *Yersinia pseudotuberculosis* YPIII. *M. aeruginosa* NIES-843 genome features more than 100 ORFs containing FSR repeat sequences, while other two have over 50. As we described in the previous sections, FSR repeat containing proteins are highly enriched in *Moorea producens* 3L, some of which we selected for structural and functional characterization. This organism would have been an ideal candidate for proteome analysis, but unfortunately we could not access this cyanobacterial strain from the isolators.

Mass spectrometric (MS) analyses of the isolated proteome of all three organisms identified two categories of FSR containing proteins: (i) FSR repeats located near the N-terminus of the protein, which contained an alternate start codon downstream of the repeat sequence, and (ii) C-terminal FSR repeats in proteins, whose non-FSR homologs exist in the same organism. As MS analyses did not identify any peptides covering the FSR repeat sequences, it was impossible to draw any conclusions regarding their expression in these organisms.

Assuming that the reason for non-identification of FSR repeat peptides is the low expression level of these proteins, we decided to enrich FSR proteins prior to MS analysis. We synthesized an antibody against a synthetic FSR peptide (LSVISYQ)₃ (Davids Biotechnologies, Germany) and used it to immunoprecipitate and enrich LSVISYQ repeat containing sequences and its slight variants from *M. aeruginosa* strains NIES-843 (15) and NIES-44 (35). (Based on database searches, strains NIES-843 and NIES-44 are predicted to contain 15 and 35 ORFs encoding LSVISYQ repeat sequences respectively) The affinity-purified antibody was coupled to HiTrap NHS-activated HP-column (GE Healthcare) and incubated with total cell lysate (Fig. 4.5). After washing residual substrate, bound proteins were eluted from the column at low pH using glycine buffer (pH 2.5-3.0) and quickly neutralized by Tris buffer (pH 8.5). Low pH assists in protein unfolding, thereby lowering the protein interaction between antibody chains and bound specific proteins. MS analysis of eluate from both samples containing enriched FSR repeat proteins again did not identify any peptides covering LSVISYQ.

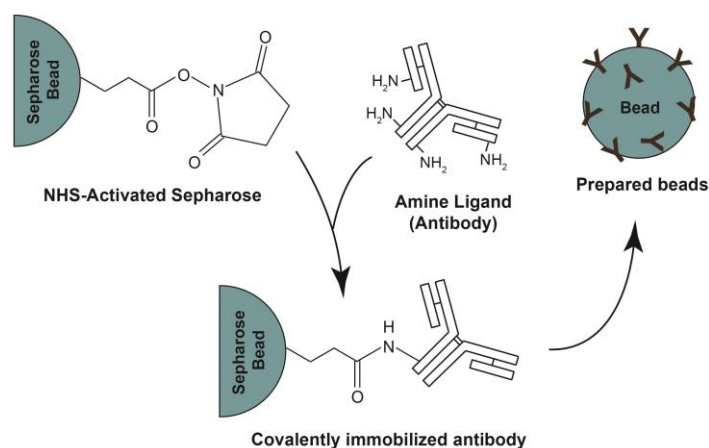


Figure 4.5: **Covalent immobilization chemistry of purified antibody to the NHS-activated sepharose HP-column** (GE Healthcare). Prepared beads were incubated with cell lysate to enrich LSVISYQ epitope proteins.

4.2.3 Transcriptome analysis for *M. aeruginosa* strains NIES-843 and NIES-44

In order to characterize the FSR repeat containing sequences at the transcriptional level, I initially selected only one ORF for mRNA detection – the restriction-endonuclease like protein domain MAE_23560 from *M. aeruginosa* NIES-843. As a first step, total RNA was isolated from the harvested cyanobacterial cells using TRIZol-chloroform/isopropanol method. The quality of isolated RNA was determined by agarose gel electrophoresis. Subsequently, the sample was depleted of any genomic DNA contamination, and reverse transcribed to obtain cDNA which was finally utilized for Real-Time PCR analysis using SYBR Green fluorescent dye. During DNA replication in PCR cycles, the dye integrates into the double-stranded DNA and an increase in fluorescence signal over time can be monitored. Using two sets of primer pairs, one covering the N-terminal region (P3/P4) and the second one spanning the C-terminal FSR repeat region (P1/P2) of the transcribed gene, we confirmed specific amplification of MAE_23560 cDNA (Fig. 4.6). The negative controls were clean, implying that the sample had no genomic DNA contamination. This result clearly confirms the expression of MAE_23560 at the mRNA level.

Later together with Dr. Birte Hernandez, we carried out a whole cell transcriptome analysis to identify all transcribed FSR repeat containing ORFs in *M. aeruginosa* strains NIES-843 and NIES-44. Total mRNA (which is usually less than 8% of total RNA) was enriched and reverse transcribed. After mRNA quality control, DNA library was prepared and sequenced on MiSeq Illumina sequencing platform. Data processing and analysis was done by Max Collenberg (Department 6, Max Planck Institute for Developmental Biology).

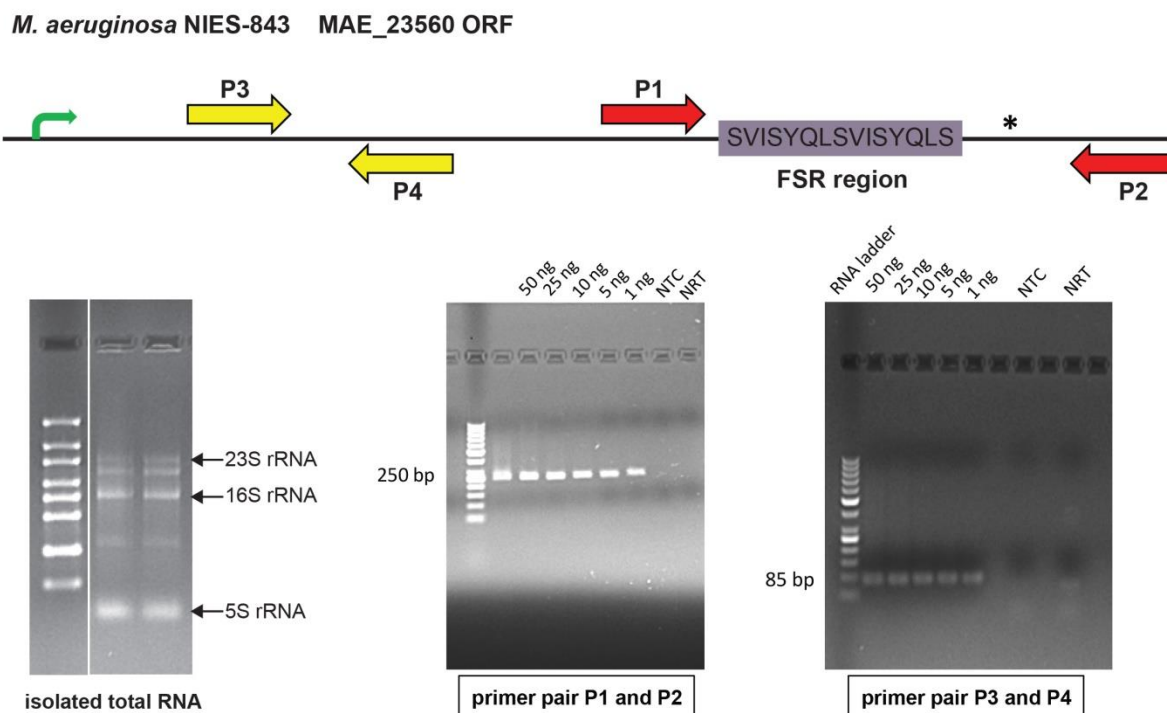


Figure 4.6: **RT-PCR analysis for MAE_23560.** (Top) Schematic representation of the *M. aeruginosa* MAE_23560 ORF. Start (green arrow) and stop (asterisk) sites are marked. Designed primer pairs (P3/P4 for the N-terminal region; P1/P2 for the C-terminal region which spans the FSR repeats) for cDNA PCR amplification are shown. (Bottom left) Total isolated RNA analyzed on 2% agarose gel confirms a high-quality, non-degraded RNA. (Bottom middle and right) Agarose gel analysis of PCR-amplified MAE_23560 cDNA using the two primer pairs.

The list of all identified transcripts for both *M. aeruginosa* strains along with their FSR repeat DNA/amino acid sequence and a relative transcriptional value is given in Table 4.3 and Table 4.4. While most of the identified transcripts encode hypothetical proteins, we identified among others a few functionally characterized proteins containing FSR repeats such as hemolysin secretion protein, a penicillin-binding protein, GP63-like surface protein, DNA gyrase subunit A, FdxN element excision controlling factor protein-like, phosphoglucomutase and dihydrofolate synthase. It is interesting to note that DNA gyrase subunit A containing the same FSR repeat insertions “GDRRQET” and “PHTPHPT” at identical locations in protein, was identified to be expressed in both *M. aeruginosa* NIES-843 and NIES-44 strains, albeit at a low level. We even detected transcripts of MAE_23560, the restriction endonuclease-like hypothetical protein containing the LSVISYQ repetition, which we had previously confirmed by qRT-PCR analysis.

Table 4.3: Identified FSR transcripts in *Microcystis aeruginosa* NIES-843

Gene ID	Protein name	Repeat sequence (unit length)	Rel. value
BAG00445	MAE_06230	TGGGGAGG/ LPHFPTSP (8) GTGGGGT/ PHTPHPT (7)	23.19
BAG00766	Unknown	ACCCAC/ VGCGVWG (7)	3413.02
BAG01160	MAE_13380	CAGGAGA/ QETGDRR (7) CCCCACA/ PHTPHPT (7)	55.88
BAG01427	MAE_16050	CCCTTATCAAGGGGGATCCCCCGCCTATCGGCACCC/ PYQGGSPRLSAPPLSRGIPPIGTPLIKGDPPAYRH (37)	5840.04
BAG01492	MAE_16700	(A/C)TTCCCCA/ WGIGELGN or WSGEVGK (8) GGGTGATAGGGTTTTG/ HHPITPKPYHPKTLSP (16)	74.58
BAG01820	MAE_19980	TCCCCACT/ SPLPHFPT (8)	47.68
BAG01932	Unknown	CTCCCCA/ LPHFPTSP (8)	731.75
BAG02178	MAE_23560	ATCAGTT/ LSVISYQ (7)	114.38
BAG02224	MAE_24020	AGGAGTC/ RSQESGV (7) CCCCACA/ PHTPHPT (7) CCCACTT/ PSHFPL (7) CCCCACTT/ PHFPTSPL (8)	44.51
BAG02505	MAE_26830	GGGTGT(A/G)/ PTPHTPH (7) GGGGAAGT/ TSPLPHFP (8) CTGTCTC/ QETGDRR (7)	14.51
BAG03192	MAE_33700	CTGATAA/ LSVISYQ (7)	339.59
BAG03436	MAE_36140	ACACCCT/ TPYTLHP (7) CACACCC/ HTPHPTP (7)	361.26
BAG03731	penicillin-binding protein	AGGAGAC/ RRQETGD (7) CCCCA(C/A)T(A/C)/ PHYPTSPT (8) CCCCACTTCCCTTT/ PHFPPTSLSPLPF (14)	74.79
BAG04476	MAE_46540	(C/G)GGGTGTT/ RVLGCSGV (8) CCCCACTT/ PHFPTSPL (8)	865.29
BAG04618	GP63-like protein	TGGGGTG/ HTPHPTP (7) TCCTGA(A/C)/ FRIQDSG or VRSQESG (7)	27.98
BAG05059	putative transposase ImeAB' protein	CCCA(T)CAC/ PTPHTPH or VWGVGCG (7)	154.36
BAG05134	MAE_53120	GGGGTGT/ TPHTPH (7) CTGACTCCTGAATA/ YSGVSIQESVFRSQ (14)	56.94
BAG05278	hemolysin secretion protein	ACCCAC/ TPHTPH (7)	398.51

Table 4.3 continued

Gene ID	Protein name	Repeat sequence (unit length)	Rel. value
BAG05441	FdxN element excision controlling factor protein like	AGATAAC/ VICYLLS (7)	6.85
BAG05611	MAE_57890	TTATCAG/ LSVISYQ (7)	6462.11
BAG05787	MAE_59650	GTGGGG(T/A)(A)/ TPHPTPH (7) CTCCTGA/ SGVRSQE (7) CTCCCGAATACTGA/ SVFGSQYSGVSIRE (14)	39.02
BAG05811	DNA gyrase subunit A	GGGTGTG/ HTPHPTP or GCGVWGV (7) CTGTCTC/ ETGDRRQ (7)	41.36
BAG05861	Unknown	ACCCAC/ VGCGVWG or TPHPTPH (7)	1498.50

Table 4.4: Identified FSR transcripts in *Microcystis aeruginosa* NIES-44

Gene ID	Protein name	Repeat sequence (unit length)	Rel. value
GAL91980	haloacid dehalogenase/epoxide hydrolase family	AGGCAAA/ FAFCLLP (7) AGGCAAT/ IAYCLLP (7)	25.62
GAL93050	Phosphoglucomutase	AGGAGAC/ RRQETGD (7) AGGAGTC/ RSQESGV (7) AGTATTC/ SIQYSVF (7)	83.55
GAL93477	putative holliday junction resolvase YggF	TGTCTCC/ GDRRQET (7) CGTCTCC/ GDGRRET (7)	123.27
GAL93975	N44_02555	TGGGGTG/ HTPHPTP (7) AACTGAT/ LSVISYQ (7)	1442.31
GAL94435	N44_03015	CGTAGGGATGATTC/ ESSLRNHPYGIPT (14)	206.19
GAL94705	DNA gyrase subunit A	CCCCACA/ PHTPHPT (7) GACAGAG/ GDRRQET (7)	12.45
GAL95418	dihydrofolate synthase	GTGGGGT/ TPHPTPH or VGCGVWG (7)	0.82
GAL95552	probable membrane protein NMA1128	CTGATAA/ LSVISYQ (7)	396.70

4.3 CONTRIBUTIONS TO THIS WORK

- All bioinformatics work was carried out by Prof. Andrei Lupas and Mateusz Korycinski. Sequences used for structural and functional characterization were selected by Prof. Andrei Lupas.
- I carried out the molecular cloning, protein expression, purification (under native or denaturing conditions), refolding, biophysical characterization (CD spectroscopy, thermal melting curves) and limited proteolysis experiments for all constructs.
- Crystal screens were set-up by Dr. Reinhard Albrecht and Kerstin Bär. The crystal structure for nitroreductase from *Oscillatoria sp.* PCC10802 was initially phased by Dr. Marcus Hartmann. I carried out the manual refinement using COOT and REFMAC5, and a second round of molecular replacement to identify and solve all four chains in the asymmetric unit cell.
- I prepared the total protein samples for *Microcystis aeruginosa* NIES-843, *Burkholderia thailandensis* E264 and *Yersinia pseudotuberculosis* YPIII for MS analysis. Analysis of the identified peptides was done by Mateusz Korycinski against his FSR protein database.
- I carried out the immunoprecipitation of FSR proteins from *M. aeruginosa* NIES-843 and *M. aeruginosa* NIES-44 strains using antibody ordered against LSVISYQ peptide. Both cyanobacterial strains were cultured by me.
- Total RNA isolation, cDNA synthesis and qRT-PCR for *M. aeruginosa* NIES-843 ORF MAE_23560 was carried out by me. Whole cell transcriptome preparation was carried out by Dr. Birte Hernandez, and data analyzed by Max Collenberg (Department 6, Max Planck Institute for Developmental Biology).

4.4 DISCUSSION

In this chapter, I have presented my work on the frameshift resistant repeats in proteins. This special phenomena occurs when repetition of a DNA motif, whose repeat unit length is not a multiple of 3, gives the same amino acid repeat of identical unit length in all three frames of equal sense. FSR repeats are impervious to random base mutations. Insertions or deletions, which are not a multiple of 3, usually translate into frameshift mutation for the C-terminal region of expressed proteins. On the contrary, pure FSR tandem repeats display only local sequence perturbations, with the original repeat sequence and periodicity resuming shortly afterwards. Comprehensive bioinformatic sequence analysis identified that FSR repetition is a common and widespread phenomena across all phylogenetic groups, ranging in repeat unit length from 4 bp microsatellites to, in rare cases, extending over a hundred base-pairs. This project focused only on FSR repeats of short unit length occurring in bacteria. Such FSR repeats were found to be specifically enriched in a few bacterial species, especially some cyanobacteria including *Microcystis aeruginosa* and *Moorea producens*, and various opportunistic and pathogenic bacteria such as *Burkholderia pseudomallei*, *Yersinia pseudotuberculosis*, *Salmonella typhimurium*, and other enterobacteria.

While a number of previous reports have highlighted the phenomena and high frequency of short DNA tandem repeat insertions in cyanobacteria, none of them have structurally and biochemically characterized the resulting protein products. An example of FSR repeat protein is pyruvate:flavodoxin oxidoreductase NifJ required for nitrogen fixation in cyanobacteria. NifJ catalyzes the transfer of electrons from pyruvate to flavodoxin, which then reduces nitrogenase. *Nostoc sp.* 7120 *nifJ* gene contains a heptanucleotide FSR sequence CCCAGT tandemly repeated within its ORF (Bauer et al., 1993). It results in the insertion of 12 aa sequence (peptide repeat PQSPVPS) within its functional protein domain (mol. wt. 132 kDa) without any frameshift mutation occurring in its C-terminal region. The crystal structure of its closest structural homolog, the pyruvate:ferrodoxin oxidoreductase (PDB 6CIN) (Chen et al., 2018), shows that this insertion occurred in a surface loop of domain II which forms part of the dimer interface. Other closely related cyanobacterial strains of *Nostoc sp.* 7120 contain similar insertions at the same site within their *nifJ* genes, but they slightly differ in their size and sequence. Interestingly, *nifJ* was identified in a heterocyst-specific cDNA library, confirming the read-out of this gene under nitrogen-limited growth conditions. This poses interesting

questions: Is this FSR repeat containing protein expressed and functional? How does the protein accommodate this insertion structurally? Is the protein function and stability affected compared to its parent non-FSR homolog? If answer to the first question is no, then another question arises: What role do such repetitions play in protein evolution and why have they been evolutionarily conserved for so long?

To answer these questions, I carried out the structural and biochemical characterization of a few selected FSR repeat containing proteins from the two different cyanobacterial species: *Moorea producens* 3L and *Microcystis aeruginosa* NIES-843. The FSR repeats were classified in three main categories – as insertions in functional protein domains, domain replacements or repeats constituting entire new ORFs. Special emphasis was placed on heptad FSR repeat sequences which displayed high coiled-coil forming probabilities; most frequent among them being LSAISYQ, LSVISYQ, QSAISNQ, ASRIAHR, VTSHKSQ, and TSNIKHQ. Alongside, I expressed and characterized the closest sequence non-FSR parent homologs of selected FSR repeat proteins. The results of *in vitro* protein characterization studies suggest that FSR repeats in proteins are mostly unstructured. While parent non-FSR homologous proteins were properly folded and stable in solution, FSR repeat proteins were insoluble after expression. Attempts to refold these proteins under various buffer conditions failed to yield stable folded products. Therefore, it appears that FSR repeats are rather deleterious at first, affecting overall protein folding and function.

M. producens NADH dehydrogenase features a combination of 25 bp and 19 bp FSR repeat insertion in a surface loop which covers the FMN binding site. Deletion of this repeat insert, thereby making it identical to its non-FSR parent homolog, resulted in a yellow-colored, flavin-bound, well-folded domain. On the opposite, insertion of this FSR repeat sequence in the parent protein gave an unfolded, insoluble protein product which could not be rescued by refolding in the presence of flavins. The results inarguably suggest that the 32 aa FSR insertion could not be accommodated within the present surface loop as it would be conceivable by formation of a novel local secondary structure or a simple extrusion. It rather affected the structure of the cofactor binding pocket and the overall protein fold resulting in the loss of protein function. Unsurprisingly, as NADH dehydrogenase represents an essential component of the oxidative phosphorylation pathway, a non-FSR homolog exists in the same organism. For a second protein, the serine/threonine protein kinase/putative ATPase, FSR insertion occurs between two domains of this multi-domain signal transduction protein. Seven tandem repeats of the

central heptamer FSR insertion, QSAISNQ, were fused between stabilizing GCN4-N16V coiled-coil adaptors. Both protein unfolding and limited proteolysis experiments revealed that the central insert formed a well-folded core. As the refolded protein was somewhat unstable and tended to aggregate into higher molecular mass complexes, it was difficult to establish the native oligomeric state of this heptad repeat. However, I could clearly observe stable oligomers formed even under denaturing conditions. We observed that most of these frequently identified heptad FSR repeats comprise broad hydrophobic interfaces as in bifaceted coiled coils; suggesting that they can assemble into higher oligomeric α -helical barrels.

Other FSR candidates, including a putative restriction endonuclease and a HIRAN domain protein contain heptad FSR repeats which replace a part of their C-terminal DNA binding domains. In both cases, FSR repeat proteins were expressed in the insoluble fraction, in contrast to their parent homologs. Similarly, *de novo* proteins generated by FSR repeat duplication were found to be mostly unstructured. From whole cell transcriptome analysis for *M. aeruginosa*, it was clear that FSR repeat proteins, including novel ORFs, are certainly transcribed. However, at the proteome level, we could not detect any *in vivo* FSR repeat peptides. This suggests that they are either expressed at low levels or only under certain environmental conditions, which is difficult to ascertain at present.

In the cyanobacterial transcriptome analysis, FSR repeat length of 7 and 8 bp was seen predominantly with a few ORFs even containing their multiples i.e. 14 and 16 bp. As heptamer TRs are overrepresented in many prokaryotes (Mrázek et al., 2007; Zhou et al., 2014); it has been hypothesized that 7bp length of the repeat unit corresponds to the size of DNA segment that interacts with DNA polymerase active site, thereby facilitating replication slippage. For one hypothetical protein, MAE_16050, encoding mRNA had the longest FSR repeat run of 37 bp length. For this highly transcribed ORF, the FSR sequence repeated almost 7 times constitutes nearly the entire length of the gene. The most frequently observed FSR repeat sequences included (repeat unit length in brackets): PHTPHPT (7), QETGDRR (7), VGCGVWG (7), LSVISYQ (7), VICYLLS (7), WSGGEVVGK (8), and their slight variants, with a number of ORFs possessing more than one repeat type. As tandem repetition of only particular heptanucleotide sequences (GC rich) is significantly enriched, it suggests that DNA replication slippage is not random but a highly sequence specific mechanism. The results further show that *Microcystis* strains frequently reutilize the same repeat sequences within their

genome, unlike the sea anemone *N. vectensis*, where majority of the tandem repeat sequences are utilized only once (Naamati et al., 2009).

In summary, our data suggest that FSR repeat amplification in bacterial genomes is a rather recent evolutionary event. Occurrence of similar repetitions across different bacterial genomes and the presence of orthologous protein-coding genes with similar FSR repeat elements suggest that at least a small fraction has survived purifying selection. In agreement with a previously proposed ‘grow slow and moult’ model of *de novo* protein emergence (Bornberg-Bauer et al., 2015), it appears that FSR repetition starts by extending an open reading frame or generating novel protein-coding genes whose products are initially unstructured and non-globular, with the potential to non-specifically interact with other cellular components. Over evolutionary time-scales, some of these genes may acquire beneficial mutations and the encoded proteins become more structured, thereby assuming novel functions, which might be important in adaptation to host-specific ecological niches.

4.5 METHODS

4.5.1 Molecular cloning

For recombinant expression of proteins in *E. coli*, DNA fragments were codon optimized and custom synthesized by gene synthesis (Eurofins). Full-length sequences for MAE_23560 (WP_012265514), MAESPC_05199 (EPF16067), LYNGBM3L_59020 (EGJ29873), LYNGBM3L_59090 (EGJ29882), LYNGBM3L_51520 (EGJ30341), Aazo_5130 (ADI66200), *Oscillatoria sp.* PCC10802 nitroreductase (WP_017715270) and *Moorea producens* 3L NADH dehydrogenase (WP_008190547) were cloned in pETHis_1a (G. Stier, EMBL Heidelberg) using NcoI/BamHI restriction sites for overexpression with an N-terminal 6xHis-tag (cleavable by TEV protease) and in pET28b-SUMO between AgeI/HindIII sites for overexpression with an N-terminal 6x-His-SUMO tag (cleavable by Ulp protease). The fragment corresponding to residues 711-759 of *M. producens* S/T protein kinase (EGJ33944) fused to N and C-terminal GCN4-N16V adaptors was cloned in pASK-IBA vector between Eco31I restriction sites. This construct contains a non-cleavable C-terminal 6x-His tag.

4.5.2 Protein expression and purification

The plasmids were transformed into *E. coli* strains BL21 (DE3), C41 (DE3) or ArcticExpress (DE3). Cells were grown in LB medium containing kanamycin ($50 \mu\text{g ml}^{-1}$) at 37°C until $\text{OD}_{600} = 0.6-0.8$ and induced with 0.5-1 mM IPTG. Following incubation for 4 h at 37°C for *E. coli* BL21 and C41 strains and for 24 h at 12°C for ArcticExpress, cells were harvested by centrifugation. The cell pellets were resuspended in lysis buffer containing 20 mM Tris, pH 7.6, 150 mM NaCl, 4 mM MgCl_2 , DNaseI (RNase-free), 1 mM PMSF and cOmplete EDTA-free Protease Inhibitor Cocktail (Roche), and subsequently lysed by sonication on a Branson Sonifier 250/Microtip 5 at output control 5 and 50% duty cycle.

SUMO-fused *Oscillatoria sp.* PCC10802 nitroreductase and LYNGBM3L_59090 constructs were purified under native conditions. Following centrifugation of the cell lysate to remove cell debris, the supernatant was loaded on a Ni-NTA Agarose column pre-equilibrated with buffer A (20 mM Tris, pH 7.6, 150 mM NaCl). Bound proteins were eluted with a two-step gradient including a step of 5% buffer B (20 mM Tris, pH 7.6, 500 mM NaCl, 0.5 M imidazole) followed by linear gradient of 5-100% buffer B. Protein containing fractions were dialyzed against buffer A and incubated with Ulp protease for His-SUMO-tag cleavage. Cleaved protein

was collected by re-loading the sample on Ni-NTA. Fractions containing the cleaved protein were pooled and purified to homogeneity by gel filtration on Superdex 200. Other constructs were purified under denaturing conditions. Cell lysate was stirred in 6 M guanidinium hydrochloride (Gua-HCl) at room temperature for 1 h. Following centrifugation, supernatant was loaded onto Ni-NTA Agarose column equilibrated with 20 mM Tris, pH 8.0, 300 mM NaCl, 6 M Gua-HCl. Bound proteins were eluted with a linear gradient of 0-0.5 M imidazole in the same buffer. Purified proteins were refolded by dialysis against different buffers varying in their composition, pH, salt concentration etc.

4.5.3 CD spectroscopy

Circular dichroism (CD) spectra were recorded on a Jasco J-810 spectropolarimeter equipped with a JASCO-423S Peltier Controller. CD measurements were performed using a cuvette with a path length of 1 mm. Single CD spectra were recorded at a speed of 100 nm/min with a data pitch of 0.5 and a response time of 1s. Each spectrum represents the average of five scans corrected by the signal of buffer scan. Thermal melting curves were recorded by monitoring ellipticity at indicated wavelengths with a temperature range of 10-95°C, applying a ramp of 0.5°C/min. Blank correction, data smoothing, and calculation of molar ellipticities and melting temperatures was performed using Spectra Manager Software (JASCO).

4.5.4 Crystallization, data collection and structure determination

Crystallization trials were set up in 96-well sitting-drop plates with drops consisting of 300 nl protein solution and 300 nl reservoir solution (RS), and reservoirs containing 50 µl RS. All crystals were cryoprotected, loop mounted and flash-cooled in liquid nitrogen. Best crystals for *Oscillatoria sp.* PCC10802 nitroreductase were obtained under the condition 0.1 M MES pH 6.5, 30% (w/v) PEG4000 and soaked in 5% PEG200 before harvesting. Data were collected at 100 K and a wavelength of either 1.07 Å at beamline X10SA of the Swiss Light Source (Villigen, Switzerland), using a PILATUS 6M-F hybrid pixel detector (Dectris Ltd.). Data was indexed, integrated and scaled using XDS (Kabsch, 2010). The structure was solved by molecular replacement with MOLREP (Vagin and Teplyakov, 2000), using PDB entry 3GE6 as the search model. After manual modeling of Chain A with Coot (Emsley and Cowtan, 2004) and refinement with REFMAC5 (Murshudov et al., 1999), other chains in the asymmetric unit cell were replaced by molecular replacement with Chain A.

4.5.5 Culturing *Microcystis aeruginosa* cells

The cyanobacterial cultures *Microcystis aeruginosa* NIES-843 and NIES-44 were obtained from the National Institute for Environmental Sciences, Japan. Conditions for culturing are followed according to the guidelines of authors. The composition of media can be found in Appendix III. Cell density was measured at 750 nm.

Microcystis aeruginosa NIES-843 (isolated from Lake Kasumigaura freshwater by Shigeto Otsuka, 1997): MA liquid medium, 25°C, light intensity 10-20 $\mu\text{mol photons/m}^2/\text{s}$, Light/Dark cycle: 10L: 14D (Otsuka et al., 2001)

Microcystis aeruginosa NIES-44 (isolated from Lake Kasumigaura freshwater by Makoto M. Watanabe, 1974): CB liquid medium, 25°C, light intensity 20-30 $\mu\text{mol photons/m}^2/\text{s}$, Light/Dark cycle: 10L: 14D, unicellular, 6 μm - 8 μm cell size. (Ichimura et al., 1971)

4.5.6 Proteome analysis

Microcystis aeruginosa NIES-843, *Burkholderia thailandensis* E264 and *Yersinia pseudotuberculosis* YPIII were cultivated to late logarithmic phase and harvested. Cell pellets were resuspended in SDS protein extraction buffer (4% (w/v) SDS, 5 mM glycerol-2-phosphate, 5 mM sodium fluoride, 5 mM sodium orthovanadate, 10 mM EDTA in 100 mM Tris/HCl pH 8.0). The samples were incubated at 95°C for 10 min, vortexed every 2-3 min and later chilled on ice. Cells were lysed using sonication, two rounds for 30 seconds on a Branson Sonifier 250/Microtip 5 at output control 4 and 40% duty cycle. Samples were incubated with 10 mM DTT at room temperature for 45 min to reduce disulfide bonds and later in 5.5 mM IAA at room temperature in dark for 45 min to alkylate the reduced cysteine disulfide bonds. Cell extracts were centrifuged at 13,000 rpm for 5 min and the supernatants collected. Total protein was precipitated by mixing with 8 sample volume ice-cold acetone and 1 sample volume ice-cold methanol and kept at -80°C overnight. Protein precipitates were washed several times with 5 ml ice-cold 80% acetone and centrifuged at 1,000 rpm for 5min at 4°C. The pellets were air dried and rehydrated in urea buffer (6 M Urea, 2 M thiourea in 100 mM Tris/HCl pH 7.5) before analyzing in MS.

4.5.7 Immunoprecipitation by coupling antibody to NHS-activated HP column

HiTrap NHS-activated HP 1 ml column (GE healthcare) coupled to the antibody was prepared according to the instructions provided by the supplier. Affinity-purified antibody was obtained against the synthetic peptide “(LSVISYQ)₃” (Davids Biotechnologie, Germany). The antibody

was exchanged into the coupling buffer (0.2 M NaHCO₃, 0.5 M NaCl, pH 8.3) by buffer exchange using a PD-10 column and concentrated to 2 mg/ml. Column was washed with 3×2 ml 1 mM HCl and incubated with 1 ml antibody solution for 1 h at room temperature. Afterwards, non-specifically bound ligand and uncoupled excess active groups in the column were deactivated by washing the column multiple times with Buffer A (0.5 M ethanolamine, 0.5 M NaCl, pH 8.3) and Buffer B (0.1 M sodium acetate, 0.5 M NaCl, pH 4). Finally the column was charged with washing buffer (50 mM Tris pH 7.5, 150 mM NaCl, protease inhibitor, 1 mM PMSF). Simultaneously, cell lysates from *M. aeruginosa* strains NIES-843 and NIES-44 were prepared. Harvested cells were resuspended in 1 ml of the extraction buffer (25 mM HEPES pH 7.3, 130 mM NaCl, 10 mM KCl, 4 mM MgCl₂, 0.5% NP-40 supplemented with protease inhibitor cocktail plus 1 mM PMSF) and lysed using a combination of freeze-thaw (3 rounds liquid N₂/70°C) and sonication (3 times, 30 seconds on a Branson Sonifier 250/Microtip 5 at output control 4 and 40% duty cycle). The antibody-coupled columns were injected with cyanobacterial cell lysates and incubated for 4 hrs in cold. Subsequently the columns were washed thrice with washing buffer and eluted with 2×1 ml of 0.1 M glycine-HCl (pH 2.4), which was neutralized later with 1 M Tris (pH 8.5). The eluted proteins were then subjected to mass spectrometric analysis.

4.5.8 Total RNA isolation

Total RNA was isolated using TRIzol Max Bacterial RNA Isolation Kit (Ambion, Life Technologies). 100 µl *Microcystis aeruginosa* NIES-843 wet cell pellet was resuspended in 100 µl of preheated Max Bacterial Enhancement Reagent and incubated at 95°C for 4 min. To this, 500 µl of TRIzol Reagent was mixed. Additional freeze/thaw cycles (freeze in liquid nitrogen, then thaw at 70°C) were carried out to complete the cell lysis. This was followed by phase separation to isolate total RNA. 200 µl of ice-cold chloroform was added to the total cell lysate, mixed vigorously and centrifuged at 12,000×g for 15 min at 4°C. The mixture separated into a lower red phenol-chloroform phase and an upper RNA containing aqueous phase. The aqueous phase was gently transferred to a new tube and mixed with 0.5 ml of cold isopropanol to precipitate RNA. After 10 min incubation at room temperature, the tubes were centrifuged at 15,000×g for 10 min at 4°C. RNA pellet was washed once in 75% ethanol, air-dried and finally resuspended in 50 µl of RNase-free water supplemented with 0.5 µl Ribolock RNase Inhibitor before freezing at -80°C. RNA quality and quantity was calculated from absorbance

at 260 nm and 280 nm on a Nanodrop spectrophotometer. For pure, non-degraded RNA samples, the value A_{260}/A_{280} should be around 2.0-2.2.

4.5.9 cDNA preparation

2 μ l DNase-I (RNase free) was added to the isolated RNA and incubated at 37°C for 2 hrs to deplete the sample of genomic DNA contamination. After heat inactivation of DNase-I at 75°C for 10 min, 4 μ l SuperScript VILO Master Mix (contains SuperScript™ III RT, RNaseOUT™ Recombinant Ribonuclease Inhibitor, a proprietary helper protein, random primers, MgCl₂ and dNTPs) was added to upto 2.5 μ g RNA in a 20 μ l total reaction volume, gently mixed and incubated at 25°C for 10 min, followed by 42°C for 1 h and then 85°C for 5 min to terminate the reaction. A negative RT control for qPCR experiment was processed simultaneously, containing equal amounts of RNA but heat-inactivated VILO reverse transcriptase. Finally 1 μ l of RNase A and 1 μ l of RNase H were added to both reaction tubes and incubated at 37°C for 1-2 hrs to degrade RNA.

4.5.10 qRT-PCR

The obtained cDNA was used in a 20 μ l PCR reaction mixture containing 2 μ l of RT reaction (cDNA), 1 μ l (0.3 μ mol) each of forward and reverse primer, 6 μ l sterile water and 10 μ l SYBR Green PCR Master Mix (Thermo Fisher). The qRT-PCR experiment was carried out in a Topical thermocycler (Analytik Jena), and the program set-up and data analysis performed with the qPCRsoft3.1 software. The PCR cycle was: 1 cycle at 95°C for 10 min, and 40 cycles of [denaturation at 95°C for 20s, annealing at 58°C for 30s, and extension at 72°C for 1 min]. Blue channel fluorescence signal was recorded during the extension step, with excitation at 470 nm and detection at 520 nm. At the end, DNA melting curves were recorded in the temperature range 60-95°C with a heating rate of 5°C/s. No-cDNA template control (NTC) and no-RT sample (NRT) were used as negative controls for qRT-PCR.

4.5.11 Transcriptome analysis

Total RNA was isolated using a combination of TRIzol reagent and Direct-Zol RNA Mini-Prep Plus kit. The upper colorless aqueous phase containing RNA (see sect. 4.4.8) was transferred to Zymo-Spin IIIICG Column2 and centrifuged. The column was washed twice each with Direct-zol RNA PreWash and RNA Wash buffers. Then, 80 μ l of DNase/RNase-free water was added directly to the column matrix and eluted in an RNase-free Eppendorf. 1.5 μ l of RiboLock

RNase Inhibitor was added to the isolated RNA. To deplete any genomic DNA contamination, the total isolated RNA was treated with 10 μ l DNaseI (Zymo Research) for 1 h at room temperature. The sample was purified and concentrated using a ZYMO Research RNA Clean & Concentrator RCC-5 kit. Before storing at -80°C , eluted total RNA was again supplemented with 0.75 μ l of RiboLock RNase Inhibitor. Final amount of isolated RNA was estimated to be $\sim 1 \mu\text{g}$.

Subsequently, rRNA was depleted from the total RNA with Ribo-Zero rRNA removal kit. RNA quality and quantity were checked on a Bioanalyzer using RNA Pico Chip. This was followed by cDNA library preparation using TruSeq RNA Sample Preparation v2 kit (Illumina). 5 μ l of mRNA at a concentration of 10-15 ng/ μ l was used for library preparation and final quality control was done using a DNA 1000 Chip on Bioanalyzer. After sequencing on the Illumina MiSeq V3 platform, data processing and analysis was carried out by Max Collenberg.

ACKNOWLEDGEMENT

First and foremost, I would like to express my profound sense of gratitude to my doctoral advisor Prof. Dr. Andrei Lupas, Director, Department of Protein Evolution, Max Planck Institute for Developmental Biology, for giving me the opportunity to work under his supervision. I thank him for his invaluable guidance, discussions and continuous support throughout this endeavor.

I also deeply acknowledge the support of my co-advisor Dr. Birte Hernandez Alvarez. I sincerely thank her for her guidance, scientific insights, feedback, time and innumerable discussions over the past years. I also thank her for critical reading of this thesis and valuable suggestions.

I am grateful to the additional members of my thesis advisory committee, Prof. Dr. Karl Forchhammer and Dr. Oliver Weichenrieder, for their valuable suggestions regarding the projects. I also thank Prof. Dr. Karl Forchhammer for being my thesis supervisor at the Eberhard Karls Universität Tübingen. I further acknowledge Prof. Dr. Volkmar Braun for his encouraging words and helpful suggestions for my projects.

I would like to express my special gratitude to the departmental crystallography unit. I thank Dr. Marcus Hartmann for his help with structural data processing, and further encouraging and supporting me to learn the skill; and Dr. Reinhard Albrecht and Kerstin Bär for setting up the crystallization screens. I thank Dr. Murray Coles and Manish Chaubey for performing NMR studies. Next, I wish to acknowledge the support of the following people at the Eberhard Karls Universität Tübingen: Prof. Dr. Boris Maček, Dr. Mirita Franz-Wachtel and Silke Wahl for their help in planning and performing mass spectrometric analyses at the Proteome Center Tübingen; Prof. Dr. Karl Forchhammer and Jan Bornikeol for their help with culturing *Microcystis* cyanobacterial strains; Dr. Sandra Schwarz, Teresa Meffert, and Dr. Monika Schütz, Johanna Weirich, for providing *Burkholderia thailandensis* E264 and *Yersinia pseudotuberculosis* YPIII cells, respectively, for proteome analysis. I sincerely thank Prof. Lucy Shapiro at the Stanford University for providing antisera for cellular localization studies.

Next, I would like to thank my project collaborators at the Max Planck Institute for Developmental Biology, Tübingen; Mateusz Korycinski for bioinformatics in the frameshift resistant repeat project; Dr. Juthaporn Sangwallek and Ioanna Karamichali in mempromCC project; Silvia Deiss and Kerstin Bär for their technical assistance; Katharina Hipp for electron microscopy studies; Rebecca Schwab for providing the facility for cyanobacterial stress experiments; Christa Lanz for whole transcriptome cDNA library preparation and sequencing; and Max Collenberg for analyzing the transcriptome data.

I also thank Dr. Vikram Alva, Dr. Edgardo Sepulveda and Dr. Jörg Martin for their valuable suggestions. I am grateful to all the past and present members of the Department of Protein Evolution for a wonderful work environment. I would like to specially thank my friends and colleagues at the institute – Amit Kumar, Edgardo Sepulveda, Manish Chaubey, Vikram Alva, Vrinda Venu, Reinhard Albrecht, Christopher Heim, Silvia Deiss, Kerstin Bär, Mohammad ElGamacy, Juthaporn Sangwallek, Harshul Arora, Maria Logotheti, Mateusz Korycinski, Claire Bedez, Adrian Fuchs, Lorena Maldoner, and Astrid Ursinus.

I deeply acknowledge the support of my parents, Mr. Brijinder Kumar Adlakha and Mrs. Anita Adlakha, for their ever extending love, cooperation and emotional support. They are the ones who with their constant support made me capable enough to reach this level. And I specially thank my elder brother Dheeraj Adlakha, scientist at the Indian Space Research Organization, who has always been the greatest source of inspiration for me and who has helped me in decision making in all walks of life. Words fail to express the profound sense of gratitude I feel for my family. Without their affection and encouragement, this thesis would not have been possible. Finally, I thank my God, for all the strength, for all this has been possible because of *him*.

Jyoti

BIBLIOGRAPHY

Ahmed A.B., Kajava A.V. (2013), Breaking the amyloidogenicity code: methods to predict amyloids from amino acid sequence, *FEBS Lett*, 587, 1089-1095

Ali S., McStay G.P. (2018), Regulation of mitochondrial dynamics by proteolytic processing and protein turnover, *Antioxidants*, 7

Alvarez B.H., Gruber M., Ursinus A., Dunin-Horkawicz S., Lupas A.N., Zeth K. (2010), A transition from strong right-handed to canonical left-handed supercoiling in a conserved coiled-coil segment of trimeric autotransporter adhesins, *J. Struct Biol.*, 170(2), 236-45

Anwari K. (2012), Isolate and sub-fractionate cell membrane from *Caulobacter crescentus*, *Bio Protoc.*, 2

Arndt K.T., Styles C.A., Fink G.R. (1986), A suppressor of a *HIS4* transcriptional defect encodes a protein with homology to the catalytic subunit of protein phosphatases, *Cell*, 56(4), 527-537

Babcock D.F., Herrington J., Goodwin P.C., Park Y.B., Hille B. (1997), Mitochondrial participation in the intracellular Ca²⁺ network, *J. Cell Biol.*, 136(4): 833

Baines C.P., Kaiser R.A., Purcell N.H., Blair N.S., Osinska H., et al. (2005), Loss of cyclophilin D reveals a critical role for mitochondrial permeability transition in cell death, *Nature*, 434, 658-662

Baradaran R., Wang C., Siliciano A.F., Long S.B. (2018), Cryo-EM structures of fungal and metazoan mitochondrial calcium uniporters, *Nature*, 559(7715), 580-584

Bassler J., Hernandez Alvarez B., Hartmann M.D., Lupas A.N. (2015), A domain dictionary of trimeric autotransporter adhesins, *Int. J. Med. Microbiol.*, 305(2), 265-275

Bauer C.C., Scappino L., Haselkorn R. (1993), Growth of the cyanobacterium *Anabaena* on molecular nitrogen: NifJ is required when iron is limited, *Proc. Natl. Acad. Sci. USA*, 90, pp. 8812-8816

Baughman J.M., Perocchi F., Girgis H.S., Plovanich M., Belcher-Timme C.A. et al. (2011), Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter, *Nature*, 476,341-345

Bichara M., Wagner J., Lambert I.B. (2006), Mechanisms of tandem repeat instability in bacteria, *Mutat. Res.*, 598(1-2), 144-63

- Bick A.G., Calvo S.E., Mootha V.K. (2012), Evolutionary diversity of the mitochondrial calcium uniporter, *Science*, 336(6083), 886
- Bononi A., Missiroli S., Poletti F., Suski J.M., Agnoletto C., Bonora M. et al. (2012), Mitochondria-associated membranes (MAMs) as hotspot Ca²⁺ signaling units, *Adv Exp Med Biol.*, 740, pp 411-37
- Bornberg-Bauer E., Schmitz J., Heberlein M. (2015), Emergence of *de novo* proteins from ‘dark genomic matter’ by ‘grow slow and moult’, *Biochem. Soc. Trans.*, 43, 867–873
- Brown J.H. et al. (1996), Heptad breaks in alpha-helical coiled coils: stutters and stammers, *Proteins*, 26, 134–145
- Bullough P.A. et al. (1994), Structure of influenza haemagglutinin at the pH of membrane fusion, *Nature*, 371, 37–43
- Burkhard P., Stetefeld J., Strelkov S.V. (2001), Coiled coils: a highly versatile protein folding motif, *Trends Cell Biol.*, 11(2), p82-88
- Burton A.J. et al. (2016), Installing hydrolytic activity into a completely *de novo* protein framework, *Nat. Chem.*, 8, 837–844
- Carr C.M. and Kim P.S. (1993), A spring-loaded mechanism for the conformational change of influenza hemagglutinin, *Cell*, 73, 823–832
- Chaudhuri D., Sancak Y., Mootha V.K., Clapham D.E. (2013), MCU encodes the pore conducting mitochondrial calcium currents, *eLife*, 2:e00704
- Chaudhuri D., Artiga D.J., Abiria S.A., Clapham D.E. (2016), Mitochondrial calcium uniporter regulator 1 (MCUR1) regulates the calcium threshold for the mitochondrial permeability transition, *Proc. Natl. Acad. Sci.*, 113(13), E1872-E1880
- Chavez D.A., Greer B.H., Eichman B.F. (2018), The HIRAN domain of helicase-like transcription factor positions the DNA translocase motor to drive efficient DNA fork regression, *J. Biol. Chem.*, 293, 8484-94
- Chen P.Y.T., Aman H., Can M., Ragsdale S.W., Drennan C.L. (2018), Binding site for coenzyme A revealed in the structure of pyruvate:ferredoxin oxidoreductase from *Moorella thermoacetica*, *Proc. Natl. Acad. Sci.*, 115(15), 3846-3851
- Chothia, C., Levitt, M., Richardson D. (1977), Structure of proteins: packing of α -helices and pleated sheets, *Proc. Natl. Acad. Sci.*, 74, 4130-4134
- Clapham D.E. (2007), Calcium Signaling, *Cell*, 131(6), p1047-1058

- Conway J.F., Parry D.A. (1991), Three-stranded alpha-fibrous proteins: The heptad repeat and its implications for structure, *Int. J. Biol. Macromol.*, 13(1), 14–16
- Cotter S.E., Surana N.K., St. Geme III J.W. (2005), Trimeric autotransporters: a distinct subfamily of autotransporter proteins, *Trends Microbiol.*, 13(5), 199-205
- Crick, F.H.C. (1952), Is α -keratin a coiled coil? *Nature*, 170, 882-883
- Crick F.H.C. (1953a), The Fourier transform of a coiled-coil, *Acta Crystallogr.*, 6, 685-689
- Crick F.H.C. (1953b), The packing of α -helices: simple coiled-coils, *Acta Crystallogr.*, 6, 689-697
- Csordás G., Golenár T., Seifert E.L., Kamer K.J., Sancak Y. et al. (2013), MICU1 controls both the threshold and cooperative activation of the mitochondrial Ca^{2+} uniporter, *Cell Metab.*, 17(6), 976-987
- Davis T.L., Walker J.R., Campagna-Slater V., Finerty Jr P.J., Paramanathan R. et al. (2010), Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases, *PLoS Biol.*, 8(7), e1000439
- De Stefani D., Patron M., Rizzuto R. (2015), Structure and function of the mitochondrial calcium uniporter complex, *Biochim. Biophys. Acta*, 1853(9), 2006-2011
- De Stefani D., Raffaello A., Teardo E., Szabò I., Rizzuto R. (2011), A forty-kilodalton protein of the inner membrane is the mitochondrial calcium uniporter, *Nature*, 476, 336-340
- Deiss S., Hernandez Alvarez B., Bär K., Ewers C.P., Coles M., Albrecht R., and Hartmann M.D. (2014), Your personalized protein structure: Andrei N. Lupas fused to GCN4 adaptors, *J. Struct. Biol.*, 186, 380-385
- Deluca H.F., Engstrom G.W. (1961), Calcium uptake by rat kidney mitochondria, *Proc. Natl. Acad. Sci.*, 47(11): 1744–1750
- Dosztányi Z., Mészáros B., Simon I. (2009), ANCHOR: web server for predicting protein binding regions in disordered proteins, *Bioinformatics*, 25(20), 2745-2746
- Drake J.W., Charlesworth B., Charlesworth D., Crow J.F. (1998), Rates of spontaneous mutation, *Genetics*, 148(4), 1667-1686
- Egelman E.H., Xu C., DiMaio F., Magnotti E., Modlin C., Yu X., Wright E., Baker D., Conticello V.P. (2015), Structural plasticity of helical nanotubes based on coiled-coil assemblies, *Structure*, 23, 280–289

- Ely B. (1991). Genetics of *Caulobacter crescentus*, *Methods Enzymol.*, 204, 372-84
- Emsley P. and Cowtan K. (2004), Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr.*, 60(Pt 12 Pt 1), 2126-2132
- Engene N., Rottacker E.C., Kaštovský J., Byrum T., Choi H., Ellisman M.H., Komárek J., Gerwick W.H. (2012), *Moorea producens* gen. nov., sp., nov. and *Moorea bouillonii* comb. Nov., tropical marine cyanobacteria rich in bioactive secondary metabolites, *Int J Syst Evol Microbiol.*, 62(5), 1171-1178
- Fan C., Orlando B.J., Fastman N.M., Zhang J., Xu Y., Chambers M.G., Xu X., Perry K., Liao M., Feng L. (2018), X-ray and cryo-EM structures of the mitochondrial calcium uniporter, *Nature*, 559(7715), 575-579
- Fernandez-Escamilla A.M., Rousseau F., Schymkowitz J., Serrano L. (2004), Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins, *Nat. Biotechnol.*, 22(10), 1302-1306
- Ferris H.U. et al. (2011), The mechanisms of HAMP mediated signaling in transmembrane receptors, *Structure*, 19, 378–385
- Fidalgo M., Barrales R.R., Ibeas J.I., Jimenez J. (2006), Adaptive evolution by mutations in the FLO11 gene, *Proc. Natl. Acad. Sci. USA*, 103, 11228-11233
- Fieni F., Lee S.B., Jan Y.N., Kirichok Y. (2012), Activity of the mitochondrial calcium uniporter varies greatly between tissues, *Nat. Commun.*, 3, 1317
- Fondon J.W. and Garner H.R. (2004), Molecular origins of rapid and continuous morphological evolution, *Proc. Natl. Acad. Sci.*, 101, 18058-18063
- Frickey T., Lupas A.N. (2004), CLANS: a Java application for visualizing protein families based on pairwise similarity, *Bioinformatics*, 20, 3702-3704
- Garg V., Kirichok Y. (2016), Keeping a lid on calcium uptake, *eLife*, 5:e17293
- Gemayel R., Cho J., Boeynaems S., Verstrepen K.J. (2012), Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences, *Genes*, 3, 461-480
- Gemayel R., Vences M.D., Legendre M., Verstrepen K.J. (2010), Variable tandem repeats accelerate evolution of coding and regulatory sequences, *Annu. Rev. Genet.*, 44, 445-477
- Gibbons J.G. and Rokas A. (2009), Comparative and functional characterization of intragenic tandem repeats in 10 *Aspergillus* genomes, *Mol. Biol. Evol.*, 26(3), 591-602

Giorgi C., De Stefani D., Bononi A., Rizzuto R., Pinton P. (2009), Structural and functional link between the mitochondrial network and the endoplasmic reticulum, *Int. J. Biochem. Cell Biol.*, 41, 1817-1827

Gragg H., Harfe B.D., Jinks-Robertson S. (2002), Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in *Saccharomyces cerevisiae*, *Mol. Cell Biol.*, 22(24), 8756-8762

Gruber M. and Lupas A.N. (2003), Historical review: another 50th anniversary – new periodicities in coiled coils, *Trends Biochem Sci.*, 28(12), 679-85

Gunter T.E. and Pfeiffer D.R. (1990), Mechanisms by which mitochondria transport calcium, *Am. J. Physiol.*, 258(5), C755-C786

Gushchin I., and Gordeliy V. (2018), Transmembrane signal transduction in two-component systems: piston, scissoring, or helical rotation? *Bioessays*, 40

Gutiérrez-Aguilar M., Baines C.P. (2014), Structural mechanisms of cyclophilin D-dependent control of the mitochondrial permeability transition pore, *Biochim. Biophys. Acta*, 1850(10), 2041-2047

Han J.Y., Choi T.S., Kim H.I. (2018), Molecular role of Ca²⁺ and hard divalent metal cations on accelerated fibrillation and interfibrillar aggregation of α -synuclein, *Sci Rep.*, 8, 1895

Hannan A.J. (2010), Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for ‘missing heritability’, *Trends Genet.*, 26(2), 59-65

Harbury P.B., Zhang T., Kim P.S., Alber T. (1993), A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants, *Science*, 262(5138), pp. 1401-1407

Harbury P.B., Kim P.S., Alber T. (1994), Crystal structure of an isoleucine-zipper trimer, *Nature*, 371, 80-83

Hartmann M.D. (2017), Functional and structural roles of coiled coils. In: Parry D., Squire J. (eds) *Fibrous proteins: Structures and mechanisms*. Subcellular Biochemistry, vol 82. Springer, Cham

Hartmann M.D., Grin I., Dunin-Horkawicz S., Deiss S., Linke D., Lupas A.N., Alvarez B.H. (2012), Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria, *Proc. Natl. Acad. Sci.*, 109(51), 20907-20912

Hartmann M.D., Mendler C.T., Bassler J., Karamichali I., Ridderbusch O., Lupas A.N., Alvarez B.H. (2016), α/β coiled coils, *eLife*, 5:e11861

- Hartmann M.D., Ridderbusch O., Zeth K., Albrecht R., Testa O. et al. (2009), A coiled-coil motif that sequesters ions to the hydrophobic core, *Proc. Natl. Acad. Sci.*, 106(40), 16950-16955
- Haworth R.A. and Hunter D.R. (1979), The Ca²⁺-induced membrane transition of rat liver mitochondria. II. Nature of the Ca²⁺ trigger site, *Arch. Biochem. Biophys.*, 195, 460–467
- Hernandez-Alvarez B., Hartmann M.D., Albrecht R., Lupas A.N., Zeth K., Linke D. (2008), A new expression system for protein crystallization using trimeric coiled-coil adaptors, *Protein Eng. Des. Sel.*, 21, 11-18
- Hicks M.R., Walshaw J., Woolfson D.N. (2002), Investigating the tolerance of coiled-coil peptides to nonheptad sequence inserts, *J. Struct. Biol.*, 137, 73-81
- Hoiczuk E., Roggenkamp A., Reichenbecher M., Lupas A., Heesemann J. (2000), Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins, *EMBO J.*, 19(22), 5989–5999
- Hope I.A. and Struhl K. (1985), GCN4 protein, synthesized in vitro, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast, *Cell*, 43(1), 177-188
- Huang P.S. et al. (2014), High thermodynamic stability of parametrically designed helical bundles, *Science*, 346, 481–485
- Hulko M. et al. (2006), The HAMP domain structure implies helix rotation in transmembrane signaling, *Cell*, 126, 929–940
- Ichimura T., Watanabe M.M. (1977), An axenic clone of *Microcystis aeruginosa* Kutz. emend. Elenkin from Lake Kasumigaura, *Bull. Jpn. Soc. Phycol.*, 25, 177-181
- Iino M. (2010), Spatiotemporal dynamics of Ca²⁺ signaling and its physiological roles, *Proc. Jpn Acad. Ser B Phys. Biol. Sci.*, 86(3), 244–256
- Isaacs A.M., Senn D.B., Yuan M., Shine J.P., Yankne B.A. (2006), Acceleration of amyloid β -peptide aggregation by physiological concentrations of calcium, *J. Biol. Chem.*, 281(38), 27916-27923
- Iyer L.M., Babu M.M., Aravind L. (2006), The HIRAN domain and recruitment of chromatin remodeling and repair activities to damaged DNA, *Cell Cycle*, 5(7), 775-782
- Jansen A., Gemayel R., Verstrepen K.J. (2012), Unstable microsatellite repeats facilitate rapid evolution of coding and regulatory sequences, *Genome Dyn.*, 7, 108-125

- Jerabek-Willemsen M., Wienken C.J., Braun D., Baaske P., Duhr S. (2011), Molecular interaction studies using microscale thermophoresis, *Assay Drug Dev. Technol.*, 9, 342-353
- Jhun B.S., Mishra J., Monaco S., Fu D., Jiang W., Sheu S.S., O-Uchi J. (2016), The mitochondrial Ca²⁺ uniporter: regulation by auxiliary subunits and signal transduction pathways, *Am. J. Physiol. Cell Physiol.*, 311: C67-C80
- Joh N.H. et al. (2014), De novo design of a transmembrane Zn²⁺-transporting four-helix bundle, *Science*, 346, 1520–1524
- Johnsen A., Fidler A.E., Kuhn S., Carter K.L., Hoffmann A. et al. (2007), Avian CLOCK gene polymorphism: evidence for a latitudinal cline in allele frequencies, *Mol. Ecol.*, 16, 4867-4880
- Jones A.C., Monroe E.A., Podell S., Hess W.R., Klages S. et al. (2011), Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*, *Proc. Natl. Acad. Sci.*, 108(21), 8815-8820
- Jorda J., Xue B., Uversky V.N., Kajava A.V. (2010), Protein tandem repeats: the more perfect the less structured, *FEBS J.*, 277(12), 2673-2682
- Jouaville L.S., Pinto P., Bastianutto C., Rutter G.A., Rizzuto R. (1999), Regulation of mitochondrial ATP synthesis by calcium: Evidence for a long-term metabolic priming, *Proc. Natl. Acad. Sci.*, 96(24), 13807-13812
- Kabsch W. (2010), XDS, *Acta Crystallogr. D Biol. Crystallogr.*, 66(Pt 2), 125-132
- Kamer K.J., Grabarek Z., Mootha V.K. (2017), High-affinity cooperative Ca²⁺ binding by MICU1-MICU2 serves as an on-off switch for the uniporter, *EMBO Rep.*, 18(8), 1397-1411
- Kammerer R.A., Schulthess T., Landwehr R., Lustig A., Engel J., Aebi U., Steinmetz M.O. (1998), An autonomous folding unit mediates the assembly of two-stranded coiled coils, *Proc. Natl. Acad. Sci.*, 95, 13419–13424
- Kaneko T., Nakajima N., Okamoto S., Suzuki I., Tanabe Y. et al. (2007), Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843, *DNA Res.*, 14(6), 247-256
- Kang S., Ohshima K., Shimizu M., Amirhaeri S., Wells R.D. (1995), Pausing of DNA synthesis *in vitro* at specific loci in CTG and CGG triplet repeats from human hereditary disease genes, *J. Biol. Chem.*, 270, 27014–27021
- Kirichok Y., Krapivinsky G., Clapham D.E. (2004), The mitochondrial calcium uniporter is a highly selective ion channel, *Nature*, 427, 360-364

Knappenberger J.A., Smith J.E., Thorpe S.H., Zitzewitz J.A., Matthews C.R. (2002), A buried polar residue in the hydrophobic interface of the coiled-coil peptide, GCN4-p1, plays a thermodynamic, not a kinetic role in folding, *J. Mol. Biol.*, 321(1), 1-6

Koronakis V. et al. (2000), Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export, *Nature*, 405, 914–919

Landschulz W.H., Johnson P.F., McKnight S.L. (1988), The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins, *Science*, 240(4860), 1759-64

Lee S.K., Shanmughapriya S., Mok M.C.Y., Dong Z., Tomar D., Carvalho E., Rajan S., Junop M.S., Madesh M., Stathopoulos P.B. (2016), Structural insights into mitochondrial calcium uniporter regulation by divalent cations, *Cell Chem. Biol.*, 23(9), p1157-1169

Lee Y., Min C.K., Kim T.G., Song H.K., Lim Y. et al. (2015), Structure and function of the N-terminal domain of the human mitochondrial calcium uniporter, *EMBO Rep.*, 16(10), 1318-1333

Legendre M., Pochet N., Pak T., Verstrepen K.J. (2007), Sequence-based estimation of minisatellite and microsatellite repeat variability, *Genome Res.*, 17(12), 1787-1796

Levinson G. and Gutman G.A. (1987), Slipped-strand mispairing: a major mechanism for DNA sequence evolution, *Mol. Biol. Evol.*, 4(3), 203-221

Linke D., Riess T., Autenrieth I.B., Lupas A., Kempf V.A.J. (2006), Trimeric autotransporter adhesins: variable structure, common function, *Trends Microbiol.*, 14(6), 264-270

Liu J.C., Liu J., Holmström K.M., Menazza S., Parks R.J. et al. (2016), MICU1 serves as a molecular gatekeeper to prevent in vivo mitochondrial calcium overload, *Cell Rep.*, 16(6), 1561-1573

Liu J. and Rost B. (2001), Comparing function and structure between entire proteomes, *Protein Sci.*, 10(10), 1970-1979

Liu J., Zheng Q., Deng Y., Cheng C.S., Kallenbach N.R., Lu M. (2006), A seven-helix coiled coil, *Proc. Natl. Acad. Sci.*, 103(42), 15457-15462

Logan C.V., Szabadkai G., Sharpe J.A., Parry D.A., Torelli S. et al. (2014), Loss-of-function mutations in MICU1 cause a brain and muscle disorder linked to primary alterations in mitochondrial calcium signaling, *Nat Genet.*, 46(2), 188-93

Lupas, A. (1996), Coiled coils: new structures and new functions, *Trends Biochem Sci.*, 21, 375-382

- Lupas A.N., Bassler J., Dunin-Horkawicz S. (2017), The structure and topology of α -helical coiled coils. In: Parry D., Squire J. (eds) *Fibrous Proteins: Structures and Mechanisms. Subcellular Biochemistry*, vol 82. Springer, Cham
- Lupas A.N. and Gruber M. (2005), The structure of alpha-helical coiled coils, *Adv. Protein Chem.*, 70, 37-78
- Łyskowski A., Leo J.C., Goldman A. (2011), Structure and biology of trimeric autotransporter adhesins, *Adv. Exp. Med. Biol.*, 715, 143-58
- Makin O.S., Atkins E., Sikorski P., Johansson J., Serpell L.C. (2005), Molecular basis for amyloid fibril formation and stability, *Proc. Natl. Acad. Sci.*, 102(2), 315-320
- Mallilankaraman K., Cárdenas, C., Doonan P.J., Chandramoorthy H.C., Irrinki K.M. et al. (2012a), MCUR1 is an essential component of mitochondrial Ca^{2+} uptake that regulates cellular metabolism, *Nat. Cell Biol.*, 14(12), 1336-43
- Mallilankaraman K., Doonan P., Cárdenas, C., Chandramoorthy H.C., Müller M. et al. (2012b), MICU1 is an essential gatekeeper for MCU-mediated mitochondrial Ca^{2+} uptake that regulates cell survival, *Cell*, 151(3), p630-644
- Malito E., Biancucci M., Faleri A., Ferlenghi I., Scarselli M. et al. (2014), Structure of the meningococcal vaccine antigen NadA and epitope mapping of a bactericidal antibody, *Proc. Natl. Acad. Sci.*, 111(48), 17128-17133
- Mammucari C., Gherardi G., Rizzuto R. (2017), Structure, activity regulation, and role of the mitochondrial calcium uniporter in health and disease, *Front. Oncol.*, 7:139
- Marchi S., Patergnani S., Pinton P. (2014), The endoplasmic reticulum-mitochondria connection: One touch, multiple functions, *Biochim. Biophys. Acta*, 1837, 461-469
- Marchi S., Pinton P. (2014), The mitochondrial calcium uniporter complex: molecular components, structure and pathophysiological implications, *J. Physiol.*, 592(5), 829-839
- Marchi S., Pinton P. (2013), Mitochondrial calcium uniporter, miRNA and cancer: live and let die, *Commun Integr Biol.*, 6(3):e23818
- Martin P., Makepeace K., Hill S.A., Hood D.W., Moxon E.R. (2005), Microsatellite instability regulates transcription factor binding and gene expression, *Proc. Natl. Acad. Sci. USA*, 102, 3800-3804
- Matityahu A. and Onn I. (2018), A new twist in the coil: functions of the coiled-coil domain of structural maintenance of chromosome (SMC) proteins, *Curr. Genet.*, 64(1), 109-116

- Mazel D., Houmard J., Castets A.M., Tandeau de Marsac N. (1990), Highly repetitive DNA sequences in cynaobacterial genomes, *J. Bacteriol.*, 172(5), 2755-2761
- Meng G., Surana N.K., ST Geme J.W. 3rd, Waksman G. (2006), Structure of the outer membrane translocator domain of the Haemophilus influenzae Hia trimeric autotransporter, *EMBO J.*, 25(11), 2297-2304
- Mittelman D., Sykoudis K., Hersh M., Lin Y. Wilson J.H. (2010), Hsp90 modulates CAG repeat instability in human cells, *Cell Stress Chaperones*, 15, 753-759
- Mrázek J., Guo X., Shah A. (2007), Simple sequence repeats in prokaryotic genomes, *Proc. Natl. Acad. Sci.*, 104(20), 8472-8477
- Murshudov G.N., Skubák P., Lebedev A.A., Pannu N.S., Steiner R.A., Nicholls R.A., Winn M.D., Long F., Vagin A. A. (2011), *REFMAC5* for the refinement of macromolecular crystal structures, *Acta Cryst.*, D67, 355-367
- Naamati G., Fromer M., Linial M. (2009), Expansion of tandem repeats in sea anemone *Nematostella vectensis* proteome: A source for gene novelty?, *BMC Genomics*, 10:593
- Nguyen N.X., Armache J.P., Lee C., Yang Y., Zeng W., Mootha V.K., Cheng Y., Bai X., Jiang Y. (2018), Cryo-EM structure of a fungal mitochondrial calcium uniporter, *Nature*, 559(7715), 570-574
- Oates M.E., Romero P., Ishida T., Ghalwash M., Mizianty M.J. et al. (2013), D²P²: database of disordered protein predictions, *Nucleic Acids Res.*, 41: D508-D516
- Offer G. and Sessions R. (1995), Computer modelling of the α -helical coiled coil: packing of side-chains in the inner core, *J. Mol. Biol.*, 249, 967-987
- Ohno S. and Epplen J.T. (1983), The primitive code and repeats of base oligomers as the primordial protein-encoding sequence, *Proc. Natl. Acad. Sci. USA*, 80, 3391-3395
- Ohno S. (1984a), Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence, *Proc. Natl. Acad. Sci. USA*, 81, 2421-2425
- Ohno S. (1984b), Repeats of base oligomers as the primordial coding sequences of the primeval earth and their vestiges in modern genes, *J. Mol. Evol.*, 20, 313-321
- O'Shea E., Klemm J.D., Kim P.S., Alber T. (1991), X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil, *Science*, 254(5031), pp. 539-544

Oxenoid K., Dong Y., Cao C., Cui T., Sancak Y. et al. (2016), Architecture of the mitochondrial calcium uniporter, *Nature*, 533(7602), 269-273

Pan X., Liu J., Nguyen T., Liu C., Sun J., Teng Y. et al. (2013), The physiological role of mitochondrial calcium revealed by mice lacking the mitochondrial calcium uniporter, *Nat Cell Biol.*, 15(12), 1464-72

Pâques F., Leung W.Y., Haber J.E. (1998), Expansions and contractions in a tandem repeat induced by double-strand break repair, *Mol Cell Biol.*, 18(4), 2045-2054

Patergnani S., Suski J.M., Agnoletto C., Bononi A., Bonora M. (2011), Calcium signaling around Mitochondria Associated Membranes (MAMs), *Cell Commun Signal.*, 9:19

Patron M., Checchetto V., Raffaello A., Teardo E., Reane D.V., Mantoan M., Granatiero V., Szabò I., De Stefani D., Rizzuto R. (2014), MICU1 and MICU2 finely tune the mitochondrial Ca²⁺ uniporter by exerting opposite effects on MCU activity, *Mol. Cell*, 53(5), 726–737

Pauling L., Corey R.B., Branson H.R. (1951), The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain, *Chemistry*, 37, 205-211

Pauling L., Corey R.B. (1953), Compound helical configurations of polypeptide chains: structure of proteins of the α -keratin type, *Nature*, 171, 59-61

Paupé V., Prudent J., Dassa E.P., Rendon O.Z., Shoubridge E.A. (2015), CCDC90A (MCUR1) is a cytochrome c oxidase assembly factor and not a regulator of the mitochondrial calcium uniporter, *Cell Metab.*, 21(1), p109-116

Pearson C.E., Edamura K.N., Cleary J.D. (2005), Repeat instability: mechanisms of dynamic mutations, *Nat. Rev. Genet.*, 6, 729-742

Penna E., Espino J., De Stefani D., Rizzuto R. (2018), The MCU complex in cell death, *Cell Calcium*, 69, 73-80

Perocchi F., Gohil V.M., Girgis H.S., Bao X.R., McCombs J.E., Palmer A.E., Mootha V.K. (2010), MICU1 encodes a mitochondrial EF hand protein required for Ca²⁺ uptake, *Nature*, 467(7313), 291-296

Piot C., Croisille P., Staat P., Thibault H., Rioufol G., Mewton N. et al. (2008), Effect of cyclosporine on reperfusion injury in acute myocardial infarction, *N Engl J. Med.*, 359, 473-481

Putney Jr. J.W. and Thomas A.P. (2006), Calcium signaling: double duty for calcium at the mitochondrial uniporter, *Curr Biol.*, 16(18), R812–R815

Raffaello A., De Stefani D., Sabbadin D., Teardo E., Merli G., Picard A., Checchetto V., Moro S., Szabò I., Rizzuto R. (2013), The mitochondrial calcium uniporter is a multimer that can include a dominant-negative pore-forming subunit, *EMBO J.*, 32(17), 2362-2376

Raffaello A., Mammucari C., Gherardi G., Rizzuto R. (2016), Calcium at the center of cell signaling: interplay between endoplasmic reticulum, mitochondria and lysosomes, *Trends Biochem. Sci.*, 41(12), 1035-1049

Rizzuto R., Brini M., Murgia M., Pozzan T. (1993), Microdomains with high Ca^{2+} close to IP_3 -sensitive channels that are sensed by neighboring mitochondria, *Science*, 262, 744-747

Rizzuto R., De Stefani D., Raffaello A., Mammucari C. (2012), Mitochondria as sensors and regulators of calcium signaling, *Nat. Rev. Mol. Cell Biol.*, 13, 566-578

Rizzuto R., Duchen M.R., Pozzan T. (2004), Flirting in little space: the ER/mitochondria Ca^{2+} liaison, *Sci. STKE*, 215 (re1)

Rizzuto R., Pinton P., Carrington W., Fay F.S., Fogarty K.E., Lifshitz L.M., Tuft R.A., Pozzan T. (1998), Close contacts with the endoplasmic reticulum as determinants of mitochondrial Ca^{2+} responses, *Science*, 280(5370), 1763-1766

Rosenberg S.M. (2001), Evolving responsively: adaptive mutation, *Nat. Rev. Genet.*, 2, 504-515

Sakmann B. and Neher E. (1984), Patch clamp techniques for studying ionic channels in excitable membranes, *Ann. Rev. Physiol.*, 46:455-472

Sali A., Potterton L., Yuan F., van Vlijmen H., Karplus M. (1995), Evaluation of comparative protein modeling by MODELLER, *Proteins*, 23(3), 318-326

Sancak Y., Markhard A.L., Kitami T., Kovács-Bogdán E., Kamer K.J. et al. (2013), EMRE is an essential component of the mitochondrial calcium uniporter complex, *Science*, 342(6164), 1379-82

Santo-Domingo J. and Demarex N. (2010), Calcium uptake mechanisms of mitochondria, *Biochim. Biophys. Acta*, 1797, 907-912

Schinzel A.C., Takeuchi O., Huang Z., Fisher J.K., Zhou Z., Rubens J., Hetz C., Danial N.N., Moskowitz M.A., Korsmeyer S.J. (2005), Cyclophilin D is a component of mitochondrial permeability transition and mediates neuronal cell death after focal cerebral ischemia, *Proc. Natl. Acad. Sci. USA*, 102(34), 12005-12010

Schmidpeter P.A.M., Jahreis G., Geitner A.J., Schmid F.X. (2011), Prolyl isomerases show low sequence specificity toward the residue following the proline, *Biochemistry*, 50, 4796-4803

Schmidt A.L. and Mitter V. (2004), Microsatellite mutation directed by an external stimulus, *Mutat. Res. Fundam. Mol. Mech. Mutagen*, 568, 233-243

Schumacher M.A., Piro K.M., Xu W., Hansen S., Lewis K., Brennan R.G. (2009), Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB, *Science*, 323(5912), 396-401

Sievers F., Wilm A., Dineen D.G., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J.D., Higgins D.G. (2011), Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.*, 7:539

Smith N.L., Taylor E.J., Lindsay A.M., Charnock S.J., Turkenburg J.P., Dodson E.J., Davies G.J., Black G.W. (2005), Structure of a group A streptococcal phage-encoded virulence factor reveals a catalytically active triple-stranded β -helix, *Proc. Natl. Acad. Sci.*, 102(49), 17652-57

Söding J. and Lupas A.N. (2003), More than the sum of their parts: on the evolution of proteins from peptides, *BioEssays*, 25(9), 837-846

Spät A., Szanda G., Csordás G., Hajnóczky, G. (2009), High and low calcium-dependent mechanisms of mitochondrial calcium signaling, *Cell Calcium*, 44(1), 51–63

Squire J.M., Paul D.M., Morris E.P. (2017), Myosin and Actin Filaments in Muscle: Structures and Interactions, *Subcell Biochem*, 82, 319-371

Steinmetz M.O., Stock A., Schulthess T., Landwehr R., Lustig A., Faix J., Gerisch G., Aebi U., Kammerer R.A. (1998), A distinct 14 residue site triggers coiled-coil formation in cortexillin I, *EMBO J.*, 17, 1883–1891

Steinmetz M.O., Jelesarov I., Matousek W.M., Honnappa S., Jahnke W., Missimer J.H., Frank S., Alexandrescu A.T., Kammerer R.A. (2007), Molecular basis of coiled-coil formation, *Proc. Natl. Acad. Sci.*, 104(17), 7062-7067

Stern A., Brown M., Nickel P., Meyer T.F. (1986), Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation, *Cell*, 47, 61-71

Strelkov S.V. and Burkhard P. (2002), Analysis of α -helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation, *J. Struct. Biol.*, 137, 54-64

Szabady R.L., Peterson J.H., Skillman K.M., Bernstein H.D. (2005), An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter, *Proc. Natl. Acad. Sci. USA*, 102, 221–226

- Szczesny P. and Lupas A. (2008), Domain annotation of trimeric autotransporter adhesins—daTAA, *Bioinformatics*, 24(10), 1251-1256
- Tadross M.R., Tsien R.W., Yue D.T. (2013), Ca²⁺ channel nanodomains boost local Ca²⁺ amplitude, *Proc. Natl. Acad. Sci.*, 110(39), 15794-15799
- Tarasov A.I., Griffiths E.J., Rutter G.A. (2012), Regulation of ATP production by mitochondrial Ca²⁺, *Cell Calcium*, 52(1), 28–35
- Thanbichler M., Iniesta A.A., Shapiro L. (2007), A comprehensible set of plasmids for vanillate- and xylose-inducible gene expression in *Caulobacter crescentus*, *Nucleic Acids Res.*, 35(20), e137
- Thomson A.R., Wood C.W., Burton A.J., Bartlett G.J., Sessions R.B., Brady R.L., Woolfson D.N. (2014), Computational design of water-soluble α -helical barrels, *Science*, 346, 485–488
- Tomar D., Dong Z., Shanmughapriya S., Koch D.A., Thomas T. et al. (2016), MCUR1 is a scaffold factor for the MCU complex function and promotes mitochondrial bioenergetics, *Cell Rep.*, 15(8), 1673-1685
- Truebestein L. and Leonard T.A. (2016), Coiled-coils: The long and short of it, *Bioessays*, 38(9), 903-916
- Tsai C.W., Wu Y., Pao P.C., Phillips C.B., Williams C., Miller C., Ranaghan M., Tsai M.F. (2017), Proteolytic control of the mitochondrial calcium uniporter complex, *Proc. Natl. Acad. Sci.*, 114(17), 4388-4393
- Tsai M.F., Phillips C.B., Ranaghan M., Tsai C.W., Wu Y., Williams C., Miller C. (2016), Dual functions of a small regulatory subunit in the mitochondrial calcium uniporter complex, *eLife*, 5:e15545
- Tsolis A.C., Papandreou N.C., Ionomidou V.A., Hamodrakas S.J. (2013), A consensus method for the prediction of ‘aggregation-prone’ peptides in globular proteins, *PLoS One*, 8(1):e54175
- Vagin A. and Teplyakov A. (2000), An approach to multi-copy search in molecular replacement, *Acta Crystallogr. D Biol. Crystallogr.*, 56(Pt 12), 1622-1624
- Vais H., Mallilankaraman K., Mak D.D., Hoff H., Payne R., Tanis J.E., Foskett J.K. (2016), EMRE is a matrix Ca²⁺ sensor that governs gatekeeping of the mitochondrial Ca²⁺ uniporter, *Cell Rep.*, 14(3), p403-410

Vais H., Tanis J.E., Müller M., Payne R., Mallilankaraman K., Foskett J.K. (2015), MCUR1, CCDC90A, is a regulator of the mitochondrial calcium uniporter, *Cell Metab.*, 22(4), p533-535

van Belkum A., Scherer S., van Alphen L., Verbrugh H. (1998), Short-sequence DNA repeats in prokaryotic genomes, 62(2), 275-293

Vasington F.D., Murphy J.V. (1962), Ca ion uptake by rat kidney mitochondria and its dependence on respiration and phosphorylation, *J. Biol. Chem.*, 237(8), 2670-7

Verstrepen K.J., Jansen A., Lewitter F., Fink G.R. (2005), Intragenic tandem repeats generate functional variability, *Nature Genet.*, 37(9), 986-990

Vinces M.D., Legendre M., Caldara M., Hagihara M., Verstrepen K.J. (2009), Unstable tandem repeats in promoters confer transcriptional evolvability, *Science*, 324, 1213-1216

Walshaw J., Shipway J.M., Woolfson D.N. (2001), Guidelines for the assembly of novel coiled-coil structures: alpha-sheets and alpha-cylinders, *Biochem. Soc. Symp.*, 68, 111-123

Walshaw J. and Woolfson D.N. (2003), Extended knobs-into-holes packing in classical and complex coiled-coil assemblies, *J. Struct. Biol.*, 144, 349-361

Wang L., Yang X., Li S., Wang Z., Liu Y., Feng J., Zhu Y., Shen Y. (2014), Structural and mechanistic insights into MICU1 regulation of mitochondrial calcium uptake, *EMBO J.*, 33(6), 594-604

Weber J.L. and Wong C. (1993), Mutation of human short tandem repeats, *Hum. Mol. Genet.*, 2(8), 1123-1128

Wienken C.J., Baaske P., Rothbauer U., Braun D., Duhr S. (2010), Protein-binding assays in biological liquids using microscale thermophoresis, *Nat. Commun.*, 1:100

Wierdl M., Greene C.N., Datta A., Jinks-Robertson S., Petes T.D. (1996), Destabilization of simple repetitive DNA sequences by transcription in yeast, *Genetics*, 143, 713-721

Williams B., Bhat N., Chien P., Shapiro L. (2014), ClpXP and ClpAP proteolytic activity on divisome substrates is differentially regulated following the *Caulobacter* asymmetric cell division, *Mol Microbiol.*, 93(5), 853-66

Williams G.S.B., Boyman L., Chikando A.C., Khairallah R.J., Lederer W.J. (2013), Mitochondrial calcium uptake, *Proc. Natl. Acad. Sci.*, 110(26), 10479-10486

Witkos T.M. and Lowe M. (2017), Recognition and tethering of transport vesicles at the Golgi apparatus, *Curr. Opin. Cell Biol.*, 47, 16-23

Wong R., Steenbergen C., Murphy E. (2012), Mitochondrial permeability transition pore and calcium handling, *Methods Mol. Biol.*, 810:235-242

Woolfson D.N. (2005), The design of coiled-coil structures and assemblies, *Adv. Protein Chem.*, ISSN: 0065-3233, vol. 70, 79-112

Woolfson D.N. et al. (2012), New currency for old rope: from coiled-coil assemblies to α helical barrels, *Curr. Opin. Struct. Biol.*, 22, 432–441

Woolfson D.N. (2017) Coiled-coil design: updated and upgraded. In: Parry D., Squire J. (eds) *Fibrous Proteins: Structures and Mechanisms. Subcellular Biochemistry*, vol 82. Springer, Cham

Yoo J., Wu M., Yin Y., Herzik M.A. Jr., Lander G.C., Lee S.Y. (2018), Cryo-EM structure of a mitochondrial calcium uniporter, *Science*, 361(6401), 506-511

Zhou K., Aertsen A., Michiels C.W. (2014), The role of variable DNA tandem repeats in bacterial adaptation, *FEMS Microbiol. Rev.*, 38, 119-141

Zimmermann L., Stephens A., Nam S.Z., Rau D., Kübler J., Lozajic M., Gabler F., Söding J., Lupas A.N., Alva V. (2017), A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core, *J. Mol. Biol.*, S0022-2836(17), 30587-9

Appendix I

Organisms and strains

Strain	Description/Genotype	Source
Bacteria		
<i>E. coli</i> TOP10	F- <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) Φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 araD139</i> Δ(<i>araI</i>)7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i>	Invitrogen
<i>E. coli</i> XL-1 Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac</i> [F' <i>proAB lacIq ZΔM15 Tn10</i> (Tetr)]	Agilent
<i>E. coli</i> BL21 (DE3)	F ⁻ <i>ompT gal dcm lon hsdS_B(r_B⁻m_B⁻) λ(DE3</i> [<i>lacI lacUV5-T7p07 ind1 sam7 nin5</i>]) [<i>malB</i> ⁺] _{K-12} (λ ^S)	Agilent
<i>E. coli</i> BL21-Gold (DE3)	B F ⁻ <i>ompT hsdS(rB – mB –) dcm+ Tetr gal λ(DE3)</i> <i>endA The</i>	Agilent
<i>E. coli</i> C41 (DE3)	F ⁻ <i>ompT gal dcm hsdSB(rB- mB-)(DE3)</i>	Lucigen
<i>E. coli</i> Arctic Express (DE3)	B F ⁻ <i>ompT hsdS(rB – mB –) dcm+ Tetr gal λ(DE3)</i> <i>endA Hte [cpn10 cpn60 Gentr]</i>	Agilent

Appendix II

Plasmids used in this study

Vector	Application/ Description	Cloning sites	Antibiotic Resistance	Source/ Reference
pET28b	C-ter 6x-His tag	NcoI/XhoI	Kanamycin	Novagen
pETHis1a	N-ter 6x-His tag	NcoI/BamHI	Kanamycin	Bogomolovas et al., 2009
pETHis1a-NdeI	N-ter 6x-His tag	NdeI/BamHI	Kanamycin	(derived from pETHis1a)
pET28b-SUMO	N-ter 6x-His-SUMO tag	AgeI/HindIII	Kanamycin	(inserted SUMO3 in NdeI/BamHI of pET28b)
pASK-IBA2	C-ter strep tag	XbaI/HindIII	Ampicillin	IBA Lifesciences
pASK-IBA2-HisTEV GCN4-pII	N-ter 6x-His-TEV-GCN4-pII C-ter GCN4-pII	BsaI		(derived from pASK-IBA2)
pASK-IBA2-HisTEV GCN4-N16V	N-ter 6x-His-TEV-GCN4N16V C-ter GCN4N16V	BsaI/HindIII	Ampicillin	(derived from pASK-IBA2)
pBXMCS4	High-copy; xylose-inducible; <i>C. crescentus</i>	NdeI/XbaI	Gentamycin	Thanbichler et al., 2007
pCDNA3.1	Mammalian gene expression	HindIII/AgeI	Neomycin/ Ampicillin	Invitrogen

Appendix III

Media composition

Name	Composition (per L)	Organism
LB	10 g BD Bactotryptone 5 g BD Bactoyeast extract 5 g NaCl, pH 7.0 with NaOH 15 g agar for plates	<i>E. coli</i>
M9	8.5 g Na ₂ HPO ₄ ·2H ₂ O 3.0 g KH ₂ PO ₄ 1.0 g NH ₄ Cl (or ¹⁵ NH ₄ Cl) 0.5 g NaCl 4.0 g glucose (or ¹³ C-glucose) 0.1 mM CaCl ₂ 2.0 mM MgSO ₄	<i>E. coli</i>
PYE	2 g Bactopeptone 1 g Yeast extract 0.2 g MgSO ₄ ·7H ₂ O	<i>Caulobacter crescentus</i> NA1000
MA	5 mg Ca(NO ₃) ₂ ·4H ₂ O 10 mg KNO ₃ 5 mg NaNO ₃ 4 mg Na ₂ SO ₄ 5 mg MgCl ₂ · 6H ₂ O 10 mg β-Na ₂ glycerophosphate.5H ₂ O 0.5 mg Na ₂ EDTA.2H ₂ O 0.05 mg FeCl ₃ ·6H ₂ O 0.5 mg MnCl ₂ ·4H ₂ O 0.05 mg ZnCl ₂ 0.5 mg CoCl ₂ ·6H ₂ O 0.08 mg Na ₂ MoO ₄ ·2H ₂ O 2 mg H ₃ BO ₃ 50 mg Bicine 100 mL Distilled water pH 8.6	<i>Microcystis aeruginosa</i> NIES-843

Media composition (continued)

Name	Composition (per L)	Organism
CB	15 mg Ca(NO ₃) ₂ .4H ₂ O 10 mg KNO ₃ 4 mg MgSO ₄ .7H ₂ O 5 mg β-Na ₂ glycerophosphate.5H ₂ O 0.01 µg Vitamin B ₁₂ 0.01 µg Biotin 1 µg Thiamine HCl 0.3 mg Na ₂ EDTA.2H ₂ O 0.05 mg FeCl ₃ .6H ₂ O 0.01 mg MnCl ₂ .4H ₂ O 0.03 mg ZnCl ₂ 0.001 mg CoCl ₂ .6H ₂ O 0.0008 mg Na ₂ MoO ₄ .2H ₂ O 50 mg Bicine 100 mL Distilled water pH 9.0	<i>Microcystis aeruginosa</i> NIES-44

Appendix IV

SUPPLEMENTARY METHODS

Preparation of chemically competent *Escherichia coli*

100 ml LB media containing the appropriate resistance gene antibiotic was inoculated with 5 ml overnight bacterial pre-culture and incubated at 37°C, 180 rpm until OD_{600 nm} of 0.4-0.6 was reached. Bacterial growth and cell density was monitored with Genesys 10S UV-Vis spectrophotometer (Thermo Scientific). Cells were harvested by centrifugation at 4000 rpm for 5 min at 4°C, washed once in ice-cold 0.1 M CaCl₂, and incubated overnight on ice. Afterwards cells were centrifuged and gently resuspended in fresh 0.1 M CaCl₂ supplemented with 10% glycerol. 100 µl aliquots were flash frozen in liquid nitrogen and stored at -80°C.

Transformation of competent *Escherichia coli* cells

100-300 ng of plasmid DNA was added to 100 µl of competent cells and incubated on ice for 10 min. Cells were heat shocked for 90 s at 42°C and again kept on ice for 2 min. 1 ml of warm LB was added and the cells were incubated at 37°C for 1 h to express the antibiotic resistance gene. Afterwards, cells were harvested at 4000 rpm for 5 min, plated onto selection media plates and incubated overnight at 37°C. Positive colonies were tested by plasmid DNA sequencing using appropriate forward and reverse primers.

Preparation of electrocompetent *Caulobacter crescentus* cells

100 ml PYE media containing 0.5 µg/ml gentamycin was inoculated with 5 ml overnight pre-culture and incubated at 30°C, 180 rpm until OD_{660 nm} of 0.4-0.6 was reached. After cooling on ice, cells were harvested by centrifugation at 4000 rpm for 10 min at 4°C. The pellet was washed twice in ice-cold water, and once in ice-cold 10% glycerol. Finally, the cells were resuspended in 10% glycerol to a final concentration of 10¹¹ cells/ml. The cell suspension was kept on ice for 30-60 mins, and later 50 µl aliquots were flash frozen in liquid nitrogen and stored at -80°C.

Electroporation of electrocompetent *Caulobacter crescentus* cells

50 ng of plasmid was added to 50 μ l of *C. crescentus* electrocompetent cells and mixed by gentle tapping. Cells were subjected to an electric pulse on a Bio-Rad GenePulser with the following settings: 25 μ F, 2.5 kV, 400 Ω . The typical time constant is \sim 9.1 ms. Immediately, 950 μ l of PYE media was added to the cuvette, mixed gently and transferred to a new Eppendorf tube. The cells were incubated at 30°C, 200 rpm, for 2 hrs, later harvested by centrifugation at 4000 rpm for 5 min and spread on selective media plates.

CURRICULUM VITAE

Name: Jyoti Adlakha
Date of Birth: 09.07.1990
Nationality: Indian
Address: Max-Hartmann-Str. 17, 72076 Tübingen

Education and research experience

- 02.2014 - present - Max Planck Institute for Developmental Biology, Tübingen
Department of Protein Evolution
Thesis: “The Structure and Evolution of Non-canonical Coiled coils”
- 07.2013 – 01.2014 - Internship, CSIR-Institute of Genomics and Integrative Biology, Delhi
- 07.2008 – 06.2013 - Indian Institute of Technology Delhi
B. Tech and M. Tech in Biochemical Engineering and Biotechnology
Cumulative CGPA 9.221 (B. Tech), 9.469 (M. Tech) on scale of 10
- 05.2006 – 05.2008 - Senior secondary school, final score 92.8%
Modern Vidya Niketan Sr. Sec. School, Faridabad
- 1993 – 05.2006 - Secondary school, final score 96.4%
St. Anthony’s Sec. School, Faridabad

