

Characterisation of the mitochondrial RNase P complex and associated diseases



Ms Stephanie Oerum MSc BSc

St. Catherine's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2016

Characterisation of the mitochondrial RNase P complex and associated diseases

Stephanie Oerum

St. Catherine's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2016

ABSTRACT

Human mitochondrial RNase P protein 1 (MRPP1) and 2 (MRPP2) form a methyltransferase complex with N1-methylation activity towards purine-9 on mitochondrial tRNA. A third component, mitochondrial RNase P protein 3 (MRPP3), adds RNase P 5'-end ribonuclease processing activity to the complex. MRPP2 alone is a dehydrogenase reductase (SDR) involved in many reactions e.g. steroid metabolism and isoleucine β -oxidation. Inherited MRPP2 missense mutations cause the rare neurological HSD10 disease, where, due to the multifunctional nature of MRPP2, no clear genotype-phenotype relationship has been established. In this study, investigations are performed on the molecular properties of the individual proteins, as well as interactions within the methyltransferase/RNase P MRPP1-MRPP2-MRPP3 complex, to shed light on disease-causing mutations.

At the molecular level, the N-terminal region of MRPP1, with no known structural fold, is responsible for dimerisation, whilst the C-terminal region contains a methyltransferase (MT) domain. The structure of the MRPP1 MT-domain to 1.96 Å resolution reveals the binding of the methyl donor S-adenosyl-methionine (SAM). Assembly of the MRPP1-MRPP2 methyltransferase complex requires intact proteins, and formation of the MRPP1-MRPP2-MRPP3 RNase P complex further requires the pre-tRNA substrate. Low-resolution small angle X-ray scattering (SAXS) models of the complexes, with and without tRNA substrate, provide clues into the protein-tRNA interaction. Reported patient mutations in MRPP2 alters the biochemical properties of the protein with each effect correlating with the structural location of the mutation. From the biochemical data, a correlation exists between phenotype and MRPP2 protein abundance. A novel missense mutation, causing a mild phenotype, show devastating effects on all functions involving MRPP2, thus no clear correlation exists between disease severity and residual activity. A patient mutation in MRPP3, leading to the rare disorder Perrault syndrome, targets the active site and leads to a decrease in product formation. The mutation does not affect MRPP1-MRPP2-MRPP3 complex assembly, hence agreeing well with it being a catalytic mutant.

Together, these structural and biochemical data provide a better understanding of the molecular facets affected by HSD10 disease patient mutations, to further explore the diverse disease severity observed in patients. Long term, this will hopefully lead to novel treatments designed for patients suffering from HSD10 disease.

Declaration: I declare that there are no parts of this thesis or its research herein have been reproduced or accepted for another award or degree or diploma at any other university or learning institution. This thesis contains no other person's work except where stated in text.

Word count: approximately 47000 words (excluding bibliography, appendix, diagrams and tables).

ACKNOWLEDGEMENTS

I dedicate this thesis to my brother and my parents.

I would like to thank my main supervisor Wyatt Yue for valuable scientific discussions, for vital guidance on my project, for immense support the past four years, for always finding time to help, and for invaluable assistance in the thesis writing process.

I would also like to thank my co-supervisor Udo Oppermann for support throughout my DPhil work and in the final thesis writing process. He has been an invaluable personal support these years and made me feel at home in Oxford.

A huge thank goes to Jolanta Kopec, the main Postdoc in the Yue lab. I thank her for always meeting any questions with a good scientific debate, for being extremely helpful, for always putting others first, for being a very good friend and for making all my days in Oxford better and happier.

In the line of scientific acknowledgements, I would like to thank my farther, with whom I have had numerous fun and fulfilling scientific discussions. I thank him for his positive attitude towards science which I believe to have taken on from him, and for his patience when listening to my frustrations after a day of failed experiments.

I would like to thank my brother for being exceptionally caring and loving, and for being a great and important friend to me always. I thank my parents for always being there supporting me through everything, for overcoming obstacles with me, for listening to me, for sharing joys with me and for making these four years possible. A loving thank you.

A special thank goes to Ryan Manton, my love, with whom I have shared the ups and downs of DPhil life. Pernille Sogaard, the best friend one could have ever wished to meet, for making my days in Oxford feel more like home. The Yue group for all the fun in and outside the lab, the SGC for all the help, laughs, chats and late night lab hours together, my friends at home that have supported me from a distance, and all my new friends in Oxford. You are all so, so appreciated.

List of abbreviations

3',5'-ADP	3,5'-adenosine diphosphate
3'-AMP	3'-adenosine monophosphate
5'-GMP	5'-guanine monophosphate
A	adenine
Anion-EX	anion-exchange chromatography
ASU	asymmetric unit
BSA	bovine serum albumin
Cation-EX	cation-exchange chromatography
CV	column volume
DNA	deoxyribonucleic acid
ds	double stranded
DSF	differential scanning fluorimetry
DTT	dithiothreitol
EMSA	electrophoretic mobility shift assay
ESI-MS	electrospray ionisation mass spectrometry
G	guanine
HA	heavy atom
HP	high performance
HPLC	high-pressure liquid chromatography
HRP	horseradish peroxidase
IEX	ion exchange chromatography
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl β -D-1-thiogalactopyranoside
L	liter
LB	lysogeny broth
LIC	ligation independent cloning
LLG	log-likelihood-gain
M	molar
m	milli
MES	2-(N-morpholino)ethanesulphonic acid
MPD	2-methyl-2,4-pentandiol
MR	molecular replacement
mRNA	messenger RNA
MRPP	mitochondrial RNase P protein
MS/MS	tandem mass spectrometry
MTase	methyltransferase
MW	molecular weight

Na(ac)	sodium acetate
NaSuc	sodium succinate
Ni-NTA	Ni-charged nitrilotriacetic acid
NYN	Nedd4-YacP nuclease
NSD	normalised spatial discrepancy
OD	optical density
PCR	polymerase chain reaction
PEG	polyethylene glycol
PPR	RNA-binding pentatricopeptide repeat
PRORP	protein-only RNase P protein
Q	quaternary ammonium
RBS	ribosomal binding site
RFZ	rotation function Z-score
RMSD	root-mean-square-deviation
RNA	ribonucleic acid
RNAi	RNA interference
RNase P	Ribonuclease P
rpm	revolutions per minute
S	sulphonate
S200	superdex 200
S6	superose 6
S75	superdex 75
SAD	single-wavelength anomalous dispersion
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SAXS	small-angle X-ray scattering
SDM	site-directed Mutagenesis
SDR	short-chain dehydrogenase reductase
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SeMet	selenomethionine
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
SP	sulfopropyl
ss	single stranded
SV-AUC	sedimentation velocity analytical ultracentrifugation
TACS	tacsimate
TAL	transcription activator-like
TB	terrific broth
TCA (cycle)	tricarboxylic acid

TCA (compound)	trichloroacetic acid
TCEP	tris(2-chloroethyl) phosphate
TEV	tobacco Etch Virus
TFZ	translation function Z-score
TPR	tetratricopeptide repeat
Trm	tRNA methyltransferase
tRNA	transfer RNA
μ	micro

Contents

List of abbreviations	3
Contents	6
1 CHAPTER.....	15
INTRODUCTION.....	15
1.1 STRUCTURE AND MULTI-FUNCTIONALITY OF THE MRPP2 PROTEIN	15
1.1.1 <i>Dehydrogenase and reductase functions</i>	16
1.1.1.1 Steroids.....	16
1.1.1.2 Fatty acid and amino acid metabolites.....	19
1.1.2 <i>Summary of enzymatic activities</i>	22
1.2 STRUCTURE OF MRPP2.....	23
1.3 MRPP2 MOONLIGHTS IN MITOCHONDRIAL tRNA FUNCTIONS	25
1.3.1 <i>Mitochondrial tRNA</i>	26
1.3.2 <i>Mitochondrial tRNA methylation</i>	28
1.3.3 <i>Mitochondrial precursor tRNA 5'-end processing</i>	32
1.3.4 <i>Summary of non-enzymatic functions</i>	36
1.4 MITOCHONDRIAL DISEASES.....	36
1.4.1 <i>Diseases linked to MRPP2</i>	36
1.4.1.1 HSD10 disease	37
1.4.2 <i>Diseases linked to MRPP2 binding partners</i>	41
1.4.2.1 COXPD30 disease.....	41
1.4.2.2 Perrault syndrome	42
1.4.3 <i>Summary of diseases associated with MRPP2 and its binding partners</i>	42
1.5 THESIS OUTLINE	43

2	CHAPTER.....	46
	CHARACTERISATION OF MRPP1 ALONE.....	46
2.1	DOMAIN ORGANISATION	46
2.2	CLONING, SMALL SCALE EXPRESSION AND PURIFICATION	47
2.3	LARGE-SCALE EXPRESSION AND PURIFICATION	51
2.3.1	<i>Mature MRPP1</i>	52
2.3.2	<i>N-domain</i>	54
2.3.3	<i>MT-domain</i>	56
2.3.4	<i>MT+C-domain</i>	58
2.4	BIOCHEMICAL STUDIES	61
2.4.1	<i>Oligomeric state</i>	61
2.4.2	<i>Ligand binding</i>	64
2.5	STRUCTURAL STUDIES	65
2.5.1	<i>Crystallisation of the MT-domain of MRPP1</i>	65
2.5.1.1	MT-domain	66
2.5.1.2	MT+C-domain.....	67
2.5.2	<i>Structure determination</i>	71
2.5.3	<i>Overall structure</i>	74
2.5.4	<i>Analysis of the active site</i>	75
2.5.5	<i>Crystallisation attempts with the N-domain of MRPP1</i>	77
2.5.5.1	N-domain (residues 40-210).....	78
2.5.5.2	N-domain (residues 40-140).....	78
2.5.5.3	Structure predictions for the MRPP1 N-domain.....	78
2.5.6	<i>SAXS model of mature MRPP1</i>	80
2.6	SUBSTRATE RECOGNITION.....	82
2.6.1	<i>Purine binding</i>	82
2.6.2	<i>Soaking of MRPP1 crystals with nucleotide ligands</i>	84
2.6.3	<i>Co-crystallisation</i>	86
2.6.4	<i>Docking</i>	87

3	CHAPTER.....	88
	CHARACTERISATION OF THE MRPP1-MRPP2 COMPLEX	88
3.1	CLONING, LARGE-SCALE EXPRESSION AND PURIFICATION	88
3.1.1	<i>Reconstitution of complex from individual MRPP1 and MRPP2 expression</i>	<i>89</i>
3.1.2	<i>Co-expression of MRPP1-MRPP2</i>	<i>90</i>
3.2	INTERACTION STUDIES	92
3.2.1	<i>Mature MRPP1 and MRPP2.....</i>	<i>92</i>
3.2.2	<i>Domain mapping for complex formation using SEC.....</i>	<i>96</i>
3.2.3	<i>Sequence mapping of the interaction regions.....</i>	<i>99</i>
3.3	ACTIVITY STUDIES	103
3.3.1	<i>Mapping of MRPP1 domains in MTase activity.....</i>	<i>103</i>
3.3.2	<i>Mapping of catalytic residues in MRPP1.....</i>	<i>104</i>
3.3.3	<i>Investigation into the role of MRPP1 in MRPP2 SDR activity</i>	<i>107</i>
3.4	STRUCTURAL STUDIES	110
3.4.1	<i>X-ray crystallography.....</i>	<i>110</i>
3.4.1.1	<i>Crystallisation of MRPP1-MRPP2.....</i>	<i>110</i>
3.4.2	<i>SAXS.....</i>	<i>114</i>
3.4.2.1	<i>SAXS model of MRPP1-MRPP2.....</i>	<i>114</i>
3.4.2.2	<i>SAXS model of MRPP1-MRPP2-pre-(mt)tRNA^{Ile}</i>	<i>119</i>
3.5	TRNA BINDING	121

4	CHAPTER.....	125
	CHARACTERISATION OF MRPP3 ALONE.....	125
4.1	DOMAIN ORGANISATION	125
4.2	CLONING, SMALL-SCALE EXPRESSION AND PURIFICATION.....	126
4.3	LARGE-SCALE EXPRESSION AND PURIFICATION	127
4.4	SOLUTION STUDIES	128
4.4.1	<i>Oligomeric state</i>	128
4.4.2	<i>Ligand binding</i>	129
4.5	STRUCTURAL STUDIES	131
4.5.1	<i>Crystallisation attempts with MRPP3</i>	131
4.5.1.1	Mature MRPP3.....	132
4.5.1.2	Δ 85-MRPP3 and Δ 95-MRPP3	132
4.5.1.3	Optimising buffer conditions for crystallisation	134
4.5.1.4	Δ 273-MRPP3	136
4.5.2	<i>Outline of the published structure</i>	137
4.6	SAXS STUDIES OF MRPP3.....	141
	CHARACTERISATION OF THE MRPP1-MRPP2-MRPP3 COMPLEX.....	145
4.7	EXPRESSION AND PURIFICATION OF MRPP3 PROTEINS	145
4.7.1	<i>N-MRPP3</i>	146
4.7.2	<i>C-MRPP3</i>	146
4.8	SOLUTION STUDIES	147
4.8.1	<i>Reconstitution of the ternary complex with mature MRPP3</i>	147
4.8.2	<i>Domain mapping</i>	149
4.9	RNASE P ACTIVITY STUDIES	150
4.9.1	<i>MRPP3 regions important for activity</i>	150
4.9.2	<i>Regions of MRPP1 important for activity</i>	153
4.9.3	<i>Mapping of MRPP3 Catalytic residue</i>	154
4.10	SAXS STUDIES OF THE TERNARY COMPLEX	156

5	CHAPTER.....	160
	DISEASES ASSOCIATED WITH THE RNASE P COMPLEX.....	160
	HSD10 DISEASE	160
5.1	BIOCHEMICAL ANALYSIS OF 11 MISSENSE MUTATIONS.....	160
5.1.1	<i>Site-directed mutagenesis.....</i>	<i>161</i>
5.1.2	<i>Expression and solubility of MRPP2 mutant proteins.....</i>	<i>161</i>
5.1.3	<i>Oligomeric state by analytical size exclusion chromatography</i>	<i>163</i>
5.1.4	<i>Thermal stability by differential scanning fluorimetry</i>	<i>166</i>
5.1.5	<i>Ligand binding of MRPP2 mutant proteins.....</i>	<i>167</i>
5.2	SDR ACTIVITY OF MRPP2 WITH NOVEL MUTATIONS	171
5.2.1	<i>Large- scale expression and purification of V12L and V176M mutants</i>	<i>171</i>
5.2.2	<i>L-3-HAD activity of MRPP2</i>	<i>171</i>
5.2.3	<i>3α-HSD activity of MRPP2</i>	<i>173</i>
5.3	MOONLIGHTING ACTIVITIES OF MRPP2 WITH NOVEL MUTATIONS.....	174
5.3.1	<i>Complex assembly with MRPP1 and mutant MRPP2 proteins</i>	<i>174</i>
5.3.2	<i>m¹R9 MTase activity of MRPP1 with mutant MRPP2 proteins.....</i>	<i>177</i>
5.3.3	<i>RNase P activity of MRPP3 with mutant MRPP2 proteins</i>	<i>177</i>
	PERRAULT SYNDROME.....	179
5.4	MRPP3 MUTATIONS CAUSING PERRAULT SYNDROME	179
5.5	CLONING, EXPRESSION AND PURIFICATION.....	179
5.6	BIOCHEMICAL ANALYSIS	180
5.6.1	<i>Oligomeric state of the MRPP3 mutant protein</i>	<i>180</i>
5.6.2	<i>Thermal stability of the MRPP3 mutant protein</i>	<i>181</i>
5.7	CRYSTALLISATION ATTEMPTS	181
5.8	COMPLEX ASSEMBLY STUDIES WITH MUTANT MRPP3.....	182
5.9	ACTIVITY STUDIES WITH MUTANT MRPP3.....	183

6	CHAPTER.....	185
	CONCLUSIONS AND FUTURE AIMS.....	185
6.1.	<i>MRPP1</i>	186
6.1.1.	Oligomeric state	186
6.1.2.	Ligand binding.....	188
6.1.1.1	Active site	189
6.1.3.	Mapping of catalytic residues	189
6.2.	<i>MRPP1-MRPP2</i>	192
6.2.1.	MTase activity of mature MRPP1.....	192
6.2.2.	Interactions between MRPP1 and MRPP2.....	193
6.2.3.	MRPP1-MRPP2 summary	198
6.3.	<i>MRPP1-MRPP2-MRPP3</i>	198
6.3.1.	Biochemical analysis of MRPP3.....	198
6.3.2.	Complex assembly and pre-(mt)tRNA binding.....	199
6.3.3.	RNase P activity of the ternary complex.....	200
6.3.4.	MRPP1-MRPP2-MRPP3 outline	200
6.4.	<i>MRPP2 in the binary and ternary complex</i>	202
6.1.1.2	tRNA binding of MRPP2	202
6.1.1.3	Role of MRPP2 in the non-catalytic functions	202
6.5.	<i>Effect of mutations in MRPP2</i>	203
6.5.1.	Biochemical analysis of all missense mutations	203
6.5.2.	Functional analysis of two novel missense mutations.....	205
6.6.	<i>Effect of mutations in MRPP3</i>	212
6.6.1.	Investigations of a missense mutation causing Perrault syndrome	212
6.7.	<i>Future direction</i>	213

7	CHAPTER.....	217
	MATERIALS AND METHODS	217
7.1	CLONING	217
7.1.1	Ligation independent cloning (LIC)	217
7.1.2	Site-directed mutagenesis (SDM)	221
7.2	PROTEIN EXPRESSION	223
7.2.1	Bacterial small-scale expression	223
7.2.2	Bacterial large-scale expression	224
7.3	PROTEIN PURIFICATION	224
7.3.1	Purification buffers	224
7.3.2	SDS-polyacrylamide gel electrophoresis (SDS-PAGE).....	225
7.3.3	Protein extraction for small-scale purification	225
7.3.4	Protein extraction for large-scale purification	225
7.3.5	Small-scale protein purification	226
7.3.6	Immobilised metal affinity chromatography on batch column (IMAC).....	226
7.3.7	Size exclusion chromatography (SEC)	226
7.3.8	Treatment with Tobacco etch virus protease.....	227
7.3.9	Reverse IMAC on batch column	227
7.3.10	Ion exchange chromatography (IEX).....	227
7.3.11	Protein storage.....	228
7.3.12	Buffer exchange.....	228
7.3.13	Protein biochemical characterisation	228
7.4	ASSAY BUFFERS	228
7.4.1	Protein concentration determination	229
7.4.2	Electrospray ionisation mass spectrometry (ESI-MS).....	230
7.4.3	Tryptic digest - tandem mass spectrometry (MS/MS) analysis	230
7.4.4	Analytical SEC	230
7.4.5	Differential scanning fluorimetry (DSF)	231
7.4.6	Sedimentation velocity analytical ultracentrifugation (SV-AUC)	232
7.4.7	pre-(mt)tRNA in vitro transcription.....	235

7.4.8	<i>Electrophoretic mobility shift assay (EMSA)</i>	236
7.4.9	<i>Limited proteolysis</i>	237
7.4.10	<i>SPOT blot analysis</i>	237
7.5	X-RAY CRYSTALLOGRAPHY	238
7.5.1	<i>Protein crystallisation</i>	239
7.5.2	<i>Initial crystallisation trials</i>	241
7.5.3	<i>Follow-up crystallisation trials</i>	243
7.5.3.1	Crystallisation with additives	244
7.5.3.2	Crystallisation with micro-seeding	244
7.5.4	<i>Crystal mounting</i>	245
7.5.5	<i>Synchrotron data collection</i>	245
7.5.6	<i>Phasing</i>	246
7.5.7	<i>Model building and refinement</i>	248
7.6	HPLC-SMALL ANGLE X-RAY SCATTERING (HPLC-SAXS)	249
7.7	COMPUTATIONAL STUDIES	251
7.7.1	<i>Ligand docking</i>	251
7.7.2	<i>Intrinsic disorder profiling</i>	252
7.8	PROTEIN ACTIVITY CHARACTERISATION	252
7.8.1	<i>Buffers</i>	252
7.8.2	<i>MTase activity assay</i>	252
7.8.3	<i>SDR activity assay</i>	253
7.8.4	<i>RNase P activity assay</i>	254

BIBLIOGRAPHY	255
8 APPENDIX.....	282
8.1 NUMBERING KEY FOR PURINE BASES.....	282
8.2 OVERVIEW OF DISCUSSED MTASES SUBFAMILIES	282
8.3 SEQUENCE ALIGNMENT OF TRM10 MEMBERS.....	282
8.4 SPOT BLOT ANALYSIS.....	289
8.5 SAXS STATISTICS	290
8.6 <i>AB INITIO</i> MODELLING OF MRPP1-MRPP2.....	290

1 CHAPTER

INTRODUCTION

1.1 Structure and multi-functionality of the MRPP2 protein

17 β -hydroxysteroid-dehydrogenase type 10 (referred to as MRPP2 for mitochondrial RNase P Protein subunit 2) is an NAD⁺/NADH dependent protein belonging to the short-chain dehydrogenase/reductase (SDR) family [1]–[3]. The protein was first isolated from rat liver in 1995 [4], and three years later the gene *HSD17B10* encoding MRPP2 was cloned from human brain tissue and mapped to chromosome Xp2.11 [5]. MRPP2 is highly abundant in the liver and kidney, but less so in other tissues including brain, heart and gonads [6]. It has a mitochondrial targeting signal of 34 amino acids that facilitates its import and localisation in the mitochondria [6], [7]. MRPP2 performs a wide variety of enzymatic, as well as non-enzymatic functions, resulting in a very diverse nomenclature to reflect these various activities (**Table 1-1**). As such, MRPP2 is an example of a ‘moonlighting enzyme’: an enzyme with several functions, some of which may not relate to its enzyme activity [8].

Acronym	Name	Year	Ref
SC-HMAD	Short-chain 3-hydroxy-2-methylacyl-CoA dehydrogenase	1995	[4]
HADH II	3-Hydroxyacyl-CoA dehydrogenase type 2	1996	[9]
ERAB	Endoplasmic reticulum-associated amyloid β -binding protein	1997	[10]
SCHAD	Short chain 3-hydroxyacyl-CoA dehydrogenase	1998	[5]
ABAD	Amyloid β -peptide-binding alcohol dehydrogenase	1999	[11]
17 β -HSD10	17 β -hydroxysteroid dehydrogenase type 10	2001	[6]
MHBD	2-Methyl-3-hydroxybutyryl-CoA dehydrogenase	2003	[12]
MRPP2	Mitochondrial RNaseP protein subunit 2	2008	[13]
SDR5C1	Short-chain dehydrogenase/reductase 5C1	2012	[14]

Table 1-1. Various names used for MRPP2.

MRPP2 functions in distinct pathways such as fatty acid oxidation (**Section 1.1.1.2**), amino acid degradation (**Section 1.1.1.1**), metabolism of sex- and neuroactive steroids, and mitochondrial tRNA maturation (**Section 1.3**). A role in glucocorticoid and bile acid biosynthesis has also been proposed [7], and recently, MRPP2

was shown to have phospholipase C-like (PLCL) activity towards di- and monophosphatidylglycerol [15].

1.1.1 Dehydrogenase and reductase functions

The substrates that are enzymatically converted by MRPP2, display very different chemical compositions, e.g. steroids with aliphatic or aromatic rings, fatty acid and amino acid metabolites with aliphatic chain structures.

1.1.1.1 Steroids

MRPP2 can catalyse the conversion of progestogens, androgens, estrogens and corticoids. Formation of these substrates *in vivo* takes place from enzymatic conversion of progesterone, initially generated by a two-step conversion of cholesterol (**Figure 1-1**) [16]. The reactions catalysed by MRPP2 involve oxidation/reduction of hydroxyl/oxo-groups in various positions of the steroid backbone (**Figure 8-1**), allowing MRPP2 to be designated as 3 α -, 17 β -, 20 α -, 20 β -, 21-hydroxysteroid dehydrogenases (HSD) and as 3- and 17oxo-reductases (oxo-red) (**Table 1-2**).

Steroid substrate	Cofactor	Class	Activity type	Specific activity (nmol·mg ⁻¹ ·min ⁻¹)	k _{cat} /K _M (min ⁻¹ ·mM ⁻¹)
17 β -estradiol	NAD ⁺	Estrogen	17 β -HSD	15.6±0.8 †	15.3 †
5 α -DHT	NADH	Androgen	3oxo-red	6±1.2* 8.7±2.0 †	17.3 *,‡
5 α -DHT	NAD ⁺	Androgen	17 β -HSD	5±0.6* 1.0±0.1 †	N.D.
Adiol	NAD ⁺	Androgen	3 α /17 β -HSD	52.2±8.6*	N.D.
Adiol	NAD ⁺	Androgen	3 α -HSD	130±1.8 †	164 *
Testosterone	NAD ⁺	Androgen	17 β -HSD	1.1±0.2 †	N.D.
Adione	NADH	Androgen	3/17oxo-red	0.9±0.3 †	5.2‡
Androsterone	NAD ⁺	Androgen	3 α -HSD	12.1±0.9 †	1.0 * 14.7 †
5-androstenediol	NAD ⁺	Androgen	17 β -HSD	0.7±0.3 †	N.D.
20 α -dihydroprogesterone	NAD ⁺	Progestogen	20 α -HSD	3.0±0.6 †	N.D.
Allopregnanolone	NAD ⁺	Progestogen	3 α -HSD	N.D.	426.7 °
5 α -pregnan-20 β -ol-3-one	NAD ⁺	Progestogen	20 β -HSD	117 ± 4*	53*
Progesterone	NADH	Progestogen	3oxo-red	3.6±0.2 †	N.D.
3 α ,5 α -THDOC	NAD ⁺	Mineralocorticoid	3 α -HSD	N.D.	1381 °
Cortisol	NAD ⁺	Glucocorticoid	21-HSD	1.9 ± 0.2* 0.9 ± 0.1 †	N.D.
Cortisone	NAD ⁺	Glucocorticoid	21-HSD	2.3 ± 0.46 *	N.D.
Cortisone	NADH	Glucocorticoid	3oxo-red	6.5 ± 0.8 †	N.D.
Dehydrocorticosterone	NAD ⁺	Glucocorticoid	21-HSD	18.2 ± 0.7*	18*

Table 1-2. Some steroid substrates reported for MRPP2 in vitro. * Data from [7]. ‡ Data from [16]. † Data from [17]. ° Data from [18]. N.D. Not determined. 5 α -DHT: 5 α -dihydrotestosterone. Adiol: 3 α -androstenediol. 3 α ,5 α -THDOC: Allotetrahydrodeoxycortico-sterone. Adione: 5 α -androstenedione.

Figure 1-1 shows a selection of the more well-studied activities for MRPP2 acting as a 3α -HSD, 17β -HSD and 3α oxo-red.

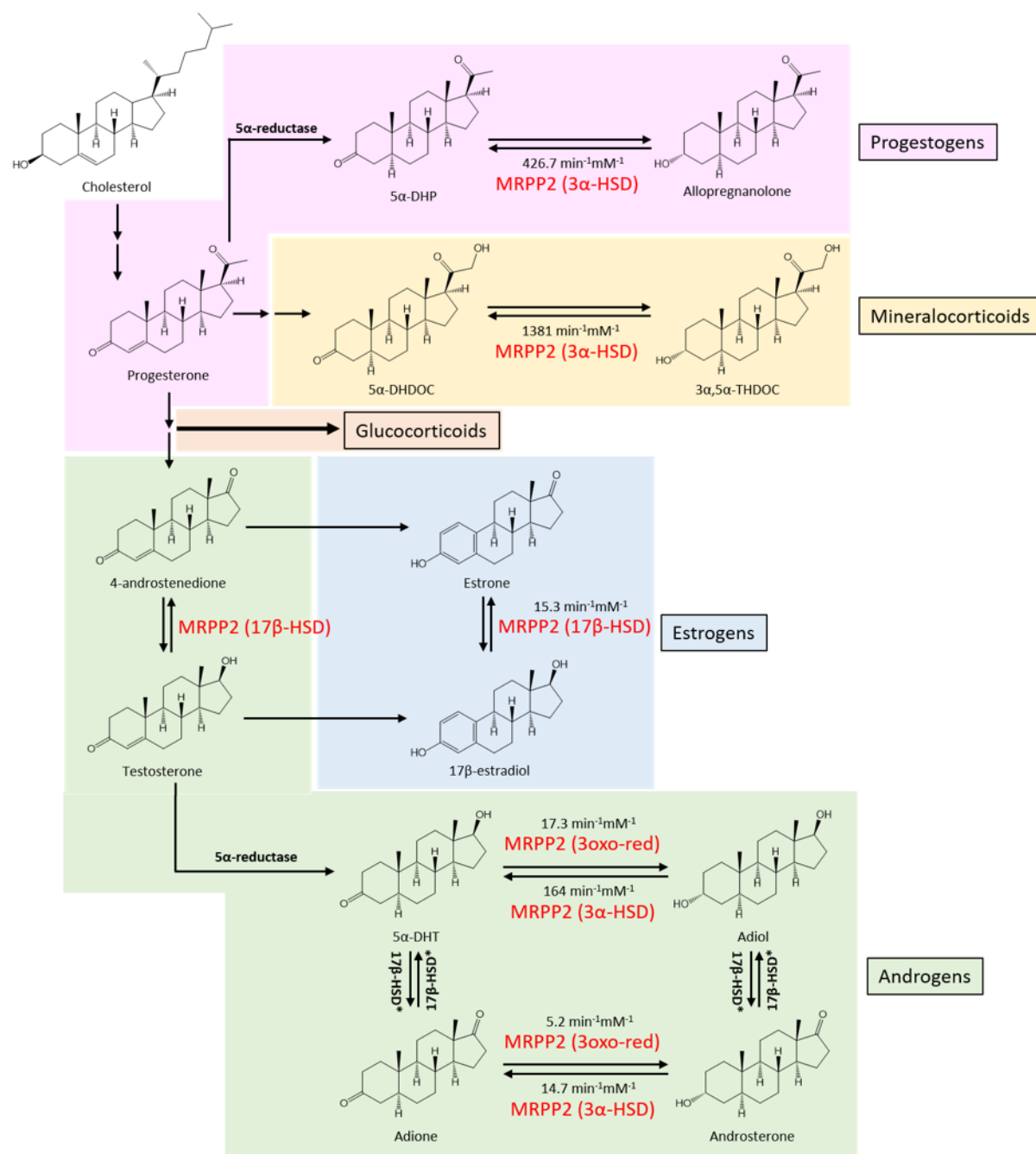


Figure 1-1. Some 3α -HSD, 17β -HSD and 3α oxo-red reactions performed by MRPP2 indicated in red. Catalytic efficiencies (k_{cat}/K_M) are shown where determined. *: 17β -HSD subtype is yet undetermined.

Sex hormones (estrogens and androgens)

17β -estradiol is the most potent female sex steroid and responsible for estrogen activity [19]. MRPP2 can inactivate this hormone by converting it to estrone via its 17β -HSD activity (**Figure 1-1, blue**) [17], [20]. Four additional HSDs (type 2, 4, 6 and 8)

also perform this reaction [21]–[23], but of all five enzymes, MRPP2 is the only one localised in mitochondria [6], [17].

The classical pathway for biosynthesis of the potent sex steroid 5 α -DHT is through reduction of the 4,5 double-bond of testosterone by the enzyme 5 α -reductase (**Figure 1-1, green**) [24]. An alternative, non-classical pathway has been discovered in which 5 α -DHT is synthesised through oxidation of the 3 α - or reduction of the 17 β -hydroxyl group of the weak androgens adiol or adione, and thus 5 α -DHT can also be formed in the absence of testosterone [25]. MRPP2 is involved in this alternative pathway by catalysing the oxidation and reduction of the 3 α -hydroxyl group on adiol and androsterone to 5 α -DHT and adione, respectively, via its 3 α -HSD and 3oxo-red activities. Based on the catalytic efficiencies (**Figure 1-1** and **Table 1-2**) for these reactions, the oxidation of adiol to 5 α -DHT is the dominant reaction for MRPP2 [7], [16], [17], [26].

Neurosteroids (progestogens and mineralocorticoids)

The naturally occurring neurosteroids allopregnanolone and allotetrahydrodeoxycorticosterone (3 α ,5 α -THDOC) are positive allosteric modulators for the γ -aminobutyric acid type A (GABA_A) receptor. When binding to the allosteric sites of GABA_A they increase the receptor opening frequency and duration, allowing a flow of Cl⁻ ions into the neuron. This influx of Cl⁻ ions changes the cell membrane potential, making the neuron less likely to generate an action potential [18], [27]. Both of these positive allosteric modulators are efficiently oxidised by MRPP2 to form 5 α -dihydroprogesterone (5 α -DHP) and dihydrodeoxycorticosterone (5 α -DHDOC), respectively (**Figure 1-1, pink and yellow**) [18], linking MRPP2 to key reactions in the central nervous system [18], [28]. The conversion of allopregnanolone to 5 α -DHP can also be catalysed by the enzyme 3 α -hydroxysteroid dehydrogenase type 3 (3 α -HSD3) [19]. The catalytic efficiency (k_{cat}/K_M) of this enzyme

($6.7 \text{ min}^{-1}\text{mM}^{-1}$) [29], is two orders of magnitude less than that of MRPP2 ($426.7 \text{ min}^{-1}\text{mM}^{-1}$) [18], suggesting that MRPP2 is the main enzyme catalysing this reaction [18], [25].

1.1.1.2 Fatty acid and amino acid metabolites

Glucose is the main carbon source for biosynthesis of acetyl-CoA in the tricarboxylic acid (TCA) cycle. When glucose levels are low (e.g. during fasting), straight- and branched-chain fatty acids, as well as branched amino acids (e.g. valine, leucine and isoleucine) [30], [31], can be metabolised to acetyl-CoA to feed the TCA cycle. MRPP2 has been linked to the metabolism of both straight-chain and branched-chain fatty acids, as well as the branched amino acid isoleucine.

Straight-chain fatty acids

The metabolism of *long straight-chain fatty acids* (**Figure 1-2**) [32] starts in the cytosol, where fatty acids are converted into acyl-CoA thioesters by the family of acyl-CoA synthetases (activation, **Figure 1-2, yellow**). These acyl-CoA thioesters are then shuttled into mitochondria via the carnitine cycle (transport, **Figure 1-2, orange**), to undergo cycles of β -oxidation with each cycle consisting of four enzymatic steps (β -oxidation, **Figure 1-2, blue**). Each cycle shortens the acyl-CoA thioester by two carbon atoms, generating reducing equivalents of NADH used in oxidative phosphorylation to produce ATP, and acetyl-CoA for the TCA cycle [32], [33].

The four enzymatic steps of the β -oxidation of straight-chain fatty acids include:

- Dehydrogenation (formation of a double bond) by an acyl-CoA dehydrogenase generating NADH reducing equivalents.
- Hydration (addition of H_2O across the double bond) by an enoyl-CoA hydratase.
- Dehydrogenation (reduction of the hydroxyl group to a keto group) by an L-3-hydroxyacyl-CoA dehydrogenase, generating NADH.

- Thiolytic cleavage of two carbon atoms by a thiolase to form one acetyl-CoA and a shorter acyl-CoA thioester for another cycle of β -oxidation.

MRPP2 can catalyse the third β -oxidation step for short straight-chain fatty acids with a k_{cat}/K_M of $3403 \pm 0.1 \text{ min}^{-1}\text{mM}^{-1}$ for L-3-hydroxyacyl-CoA *in vitro* [7], making it an L-3-hydroxyacyl-CoA dehydrogenase (L-3-HAD) [4], [34]. However, this reaction can also be catalysed by another L-3-hydroxyacyl-CoA dehydrogenase (HADH) enzyme [34] (**Figure 1-2**). *In vivo*, HADH is thought to be the main enzyme responsible for this reaction, as HADH-deficient patients accumulate straight-chain fatty acid metabolites [32], which is not observed for MRPP2-deficient patients.

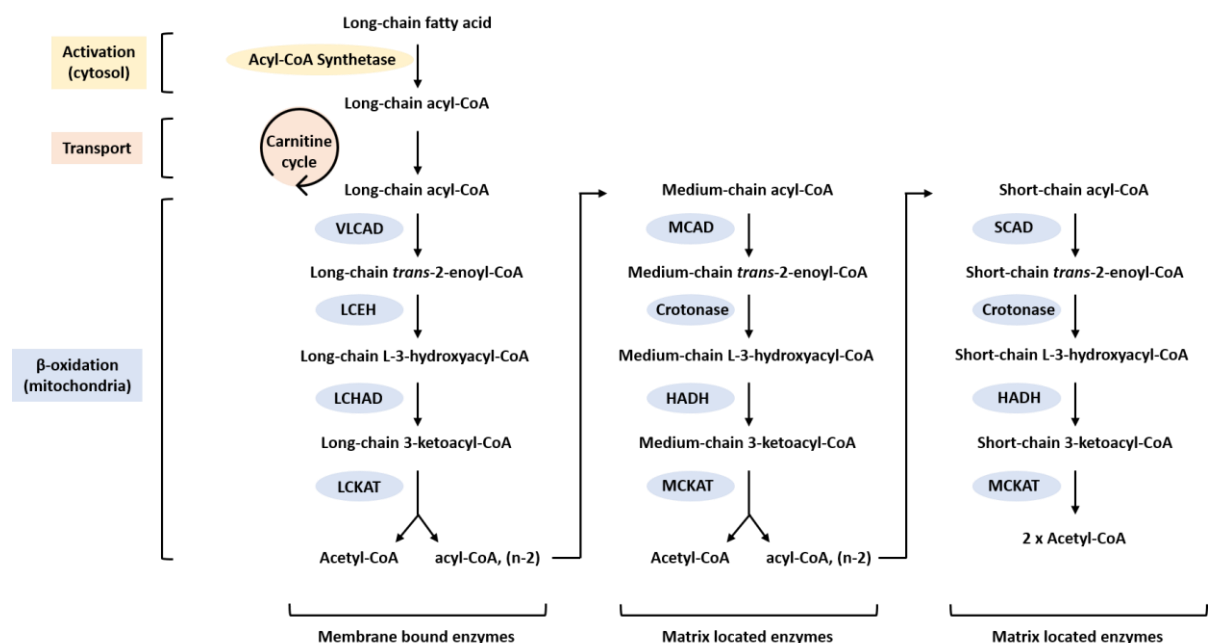


Figure 1-2. The three processes for metabolism of straight-chained fatty acids: activation, transport and β -oxidation. VLCAD: Very long chain acyl-CoA dehydrogenase. LCEH: long-chain enoyl-CoA hydratase. LCHAD: long-chain L)-3-hydroxyacyl-CoA dehydrogenase. LCKAT: long-chain 3-ketoacyl-CoA thiolase. MCAD: Medium-chain acyl-CoA dehydrogenase. Crotonase: Short-chain enoyl-CoA hydratase. HADH: Short-chain (S)-3-hydroxyacyl-CoA dehydrogenase SCHAD. MCKAT: medium-chain 3-ketoacyl-CoA thiolase. SCAD: Short-chain acyl-CoA dehydrogenase.

Branched-chain fatty acids and isoleucine

Metabolism of *long branched-chain fatty acids* follows a similar route as long straight-chain fatty acids, but with an initial round of β -oxidation taking place in peroxisomes, before transportation to mitochondria [30]. The final cycle of β -oxidation in

mitochondria produces acetyl-CoA and propionyl-CoA [30]. *Branched amino acids* only require one cycle of β -oxidation. They are transported to mitochondria, where the amino group is converted to a keto group with further activation for β -oxidation [31].

MRPP2 in fatty acid and isoleucine catabolism

MRPP2 is active on medium and short branched-chain fatty acids (e.g. 2-methylated acyl-CoA thioesters) [4], acting as an L-2-methyl-3-hydroxyacyl-CoA dehydrogenase (L-methyl-3-HAD). The MRPP2 substrate L-2-methyl-3-hydroxybutyryl-CoA originates from the degradation of the *branched amino acid* isoleucine (**Figure 1-3**), for which substrates for β -oxidation are prepared by mitochondrial branched chain amino-acid transaminase (BCATm) and branched-chain α -ketoacid dehydrogenase complex (BCKDC) [31]. MRPP2 converts L-2-methyl-3-hydroxybutyryl-CoA to 2-methyl-acetoacetyl-CoA with a k_{cat}/K_M of $56191.5 \text{ min}^{-1}\text{mM}^{-1}$ [35], suggesting that this function is biologically relevant. Indeed, MRPP2-deficient patients accumulate isoleucine metabolites upstream of the MRPP2-catalysed reaction (**Section 1.4.1.1**).

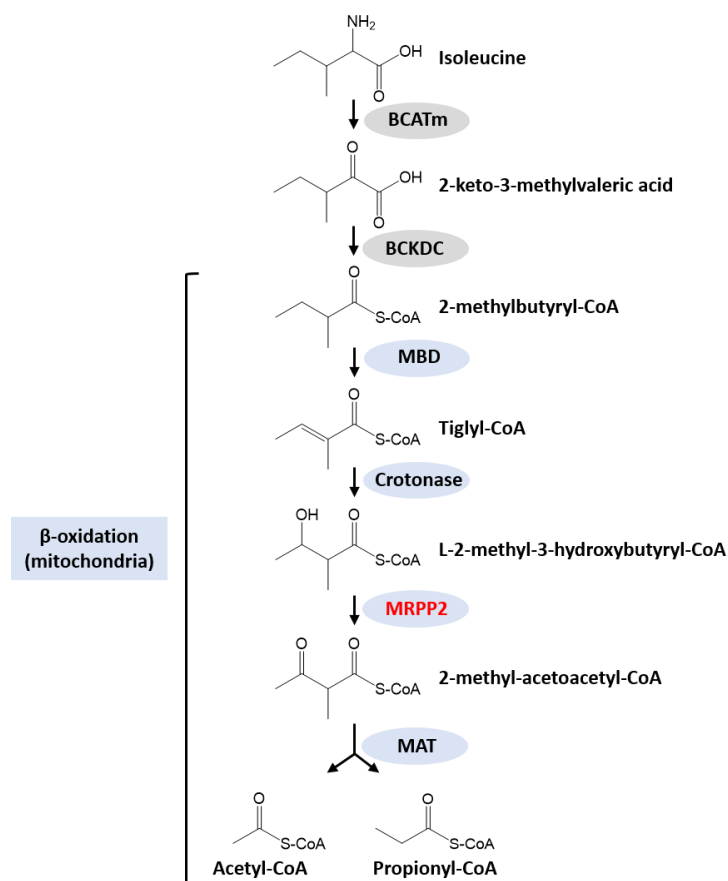


Figure 1-3. Metabolism of isoleucine. BCATm: mitochondrial branched chain amino-acid transaminase. BCKDC: branched-chain α -ketoacid dehydrogenase complex. MBD: 2-methylbutyryl-CoA dehydrogenase. Crotonase: enoyl-CoA hydratase. MRPP2: 17 β -hydroxysteroid dehydrogenase type 10. MAT: 3-ketothiolase [36].

The fatty acid degradation reactions described above are examples of MRPP2-catalysed oxidation reactions. However, the enzyme is also able to catalyse reduction reactions with appreciable rate. For example, the reduction of acetoacetyl-CoA to L-3-hydroxybutyryl-CoA proceeds with a catalytic efficiency (k_{cat}/K_M) of 24943.8 min⁻¹mM⁻¹ [5]. Similar to the oxidation reactions, MRPP2 also displays a preference for shorter substrates in its reduction reactions [5].

1.1.2 Summary of enzymatic activities

The sex steroid 17 β -estradiol, the neurosteroids allopregnanolone and 3 α ,5 α -THDOC, and the substrates from the non-classical 5 α -DHT pathway, represent the more well-studied steroid substrates for the SDR activity of MRPP2. For these reactions, the enzyme exhibits a higher catalytic efficiency, compared to reactions with other steroid

substrates (**Figure 1-2**). The catalytic efficiency in the inactivation of the sex steroid 17 β -estradiol matches that for HSDs with proven *in vivo* relevance on the same substrate [37], suggesting that MRPP2 could also be relevant for the reaction *in vivo*. MRPP2 is further involved in degradation of fatty acids and the amino acid isoleucine. The well-studied oxidation of the isoleucine metabolite L-2-methyl-3-hydroxybutyryl-CoA by MRPP2 proceeds with a high catalytic efficiency [35], and MRPP2-deficient patients accumulate the upstream isoleucine metabolites, altogether indicating its physiological relevance.

1.2 Structure of MRPP2

Proteins belonging to the SDR family are structurally grouped into the classical and extended forms based on differences in the NAD⁺-binding regions. MRPP2 belongs to the group of classical SDRs that share a conserved single Rossmann-fold of seven or eight β -strands flanked by three to four α -helices, with very few structural variations, largely limited to the C-terminal part of the proteins [1], [2]. The crystal structure of MRPP2, showing such a Rossmann-fold, was solved bound to amyloid- β (PDB: 1SO8) to 2.3 Å resolution in 2004 [38], and in complex with the cofactor NAD⁺ and inhibitor AG18051 to 2.0 Å (PDB: 1U7T, **Figure 1-4**) later that year [39]. In 2006 it was solved to a higher resolution of 1.2 Å (PDB: 2O23) in complex with NAD⁺ alone (Rauschenberger, K., 2010). Most SDRs are dimers or tetramers with subunits forming strong interactions through two long, conserved α -helices [1], [2]. MRPP2 is a tetramer, formed by a dimer of dimers, both in solution and in crystalline form, and each monomer is bound with one NAD⁺ cofactor molecule [5], [39] (**Figure 1-4**).

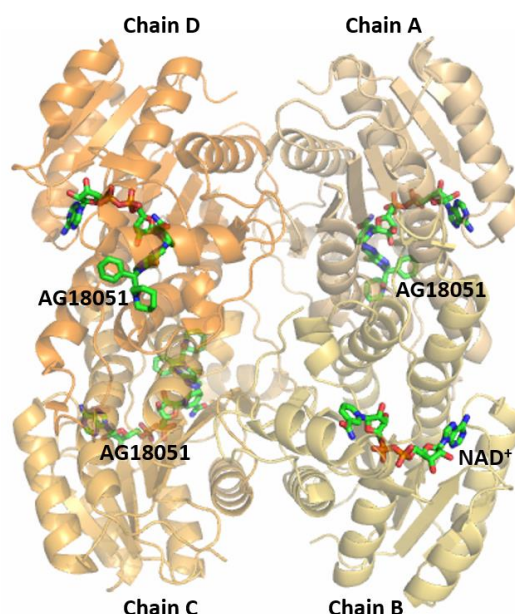


Figure 1-4. Crystal structure of the MRPP2 tetramer (PDB: 1U7T) with the inhibitor AG18051 bound in chains A, C and D, and with NAD⁺ bound in chain B.

Classical SDRs contain multiple conserved sequence motifs and residues, primarily involved in catalysis or interactions with the NAD⁺ cofactor (**Table 1-3**).

Motif/Residue	Residues in MRPP2	Location
TGxxxGxG	Thr16-Gly17-x-x-x-Gly21-x-Gly23	Loop between β A and α B
NNAG	Asn90-Cys91-Ala92-Gly93	End of β D
GxxxxxxS	Gly148-x-x-x-x-x-Ser155	β E
PG	Pro198-Gly199	End of β F
N	Asn121	α E
YxxxK	Tyr168-x-x-x-Lys172	α F
N	Asn64	Between β C and α D

Table 1-3. Conserved motifs and residues in classical SDRs. Table is adapted from Oppermann *et al.* [40], with residue numbering given for MRPP2.

The TGxxxGxG motif constitutes an essential part of the cofactor binding pocket [41], the GxxxxxxS motif makes up one side of the active site and the PG motif assists in formation of Van der Waals interactions with the nicotinamide moiety of NAD⁺ [40], [41]. Asn64 interacts with the adenine moiety of NAD⁺ to stabilise the loop in which Asn64 is located, ensuring proper formation of the binding pocket [41]. The NNAG motif (NCAG in MRPP2) is essential for maintaining the position of the β -sheet important for

cofactor positioning [41]. The YxxxK motif contains two of four catalytic residues (**Figure 1-5**): Tyr168, which is the catalytic base acting via a water molecule [1], and Lys172, which facilitates proton transfer via hydrogen bond formation to the NAD⁺ cofactor that lowers the pK_a of Tyr168 (OH group) [1], [42], [43]. These two residues are assisted by Asn121 and the Ser(155) residue from the GxxxxxS motif, altogether making up the catalytic tetrad of Tyr-Lys-Ser-Asn [1]. This tetrad is further assisted by a proton-relay formed within the active site, ensuring a constant supply of protons throughout catalysis [40]–[42]. Ser155 stabilises the substrate [41] and Asn121 forms a kink in helix α E that facilitates binding of a water molecule, necessary for the proton relay [41]. Binding of this water molecule also results in water-mediated stabilisation of the position of Lys172 [41].

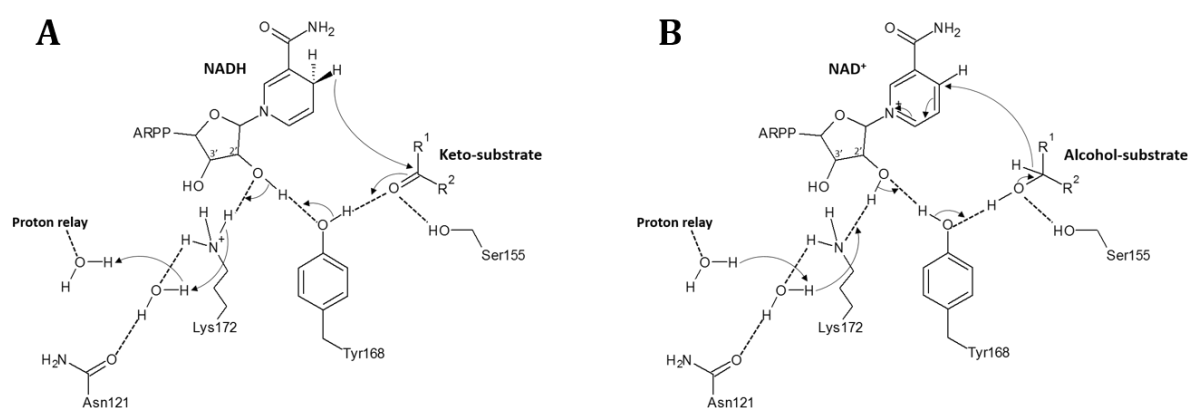


Figure 1-5. Reaction mechanism for MRPP2 catalytic A) reduction using NADH, and B) oxidation using NAD⁺.

1.3 MRPP2 moonlights in mitochondrial tRNA functions

In addition to the SDR enzymatic functions, MRPP2 has recently been found to play non-enzymatic roles in the maturation of mitochondrial tRNAs ((mt)tRNAs) via the formation of:

- 1) a binary complex between MRPP2 and the RNA methyltransferase protein MRPP1 (mitochondrial RNase P protein subunit 1), involved in (mt)tRNA methylation (**Section 1.3.2**).

2) a ternary complex comprising MRPP1, MRPP2 and the endoribonuclease protein MRPP3 (mitochondrial RNase P protein subunit 3), involved in 5'-end processing of precursor (mt)tRNAs (**Section 1.3.3**).

The ternary complex was discovered by the Rossmannith group (Holzmann *et al.*, 2008) [13] from *in vitro* reconstitution using chromatographic methods coupled with an activity assay performed on proteins, isolated from mitochondrial HeLa cells. The same group (Vilardo *et al.*, 2012) [14] later identified the separate (mt)tRNA methylation activity for the binary complex.

1.3.1 Mitochondrial tRNA

The human mitochondrial genome is a 16.6 kb circular, double stranded DNA that codes for 22 tRNAs, 13 proteins involved in the respiratory chain, and two ribosomal RNAs (rRNAs) [44], [45]. The 22 (mt)tRNAs make up the full set (one per amino acid, two iso-acceptors for leucine and serine) required for mitochondrial translation [46]. The two strands in the mitochondrial DNA are denoted light strand (C-rich) and heavy strand (G-rich) [44]. Of the 22 (mt)tRNAs, 14 are encoded by the light strand and transcribed as a single large transcript. The remaining 8 tRNAs are encoded by the heavy strand and transcribed as a small and a large polycistronic transcript comprising three tRNAs and five tRNAs, respectively. The precursor (mt)tRNAs in the polycistronic transcripts are denoted pre-(mt)tRNAs and are released by 5'- and 3'-end processing to become mature (mt)tRNAs (**Section 1.3.3**). Once released, (mt)tRNAs undergo post-transcriptional modifications (e.g. methylation, taurine coupling and addition of 3'-CCA) [47], [48] that contribute to the correct structure and function (**Section 1.3.2**), before they become involved in the protein translation process [49].

Conventional tRNAs, like those acting in the cytosol, contain an acceptor stem, a D-arm and -loop, a T-arm and -loop and an anti-codon stem and -loop (**Figure 1-6, left**), that folds up to the conventional L-shaped structure (**Figure 1-6, right**).

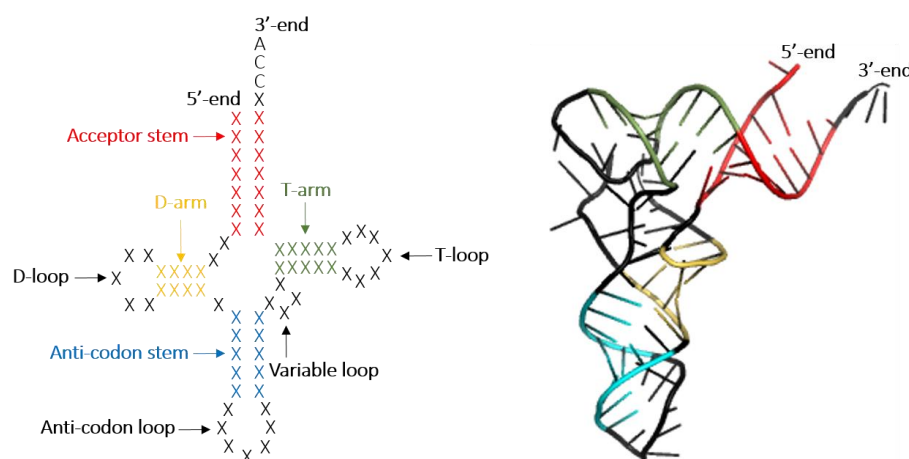


Figure 1-6. Schematic of a conventional, L-shaped tRNA. The 2D (left) 3D (right) structures are coloured in the same scheme. The 3D structure is taken from *E. coli* initiator tRNA (Met) (PDB: 3CW5).

(mt)tRNAs are often referred to as ‘bizarre’ due to their unorthodox sequence and structure compared to these conventional tRNAs [46] (**Figure 1-7**). The bizarre (mt)tRNAs have low G-C contents and their D-loop, T-loop and variable loop are either atypical in length, or completely absent, resulting in lowered structural stability compared to the conventional cytosolic tRNAs [46], [50].

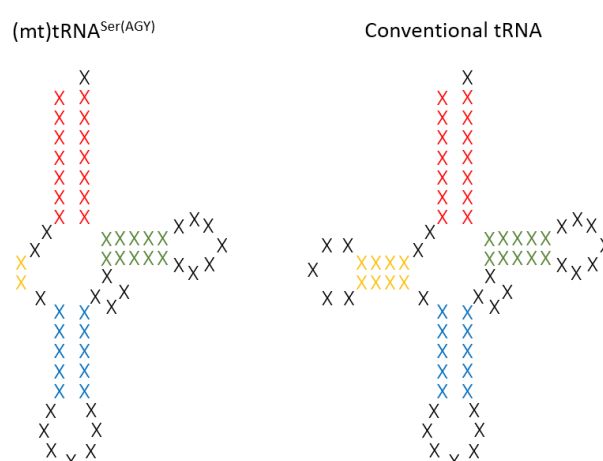


Figure 1-7. Schematic of a mitochondrial bizarre tRNA (left) and a conventional tRNA (right).

1.3.2 Mitochondrial tRNA methylation

One classical post-transcriptional modification of tRNA is methylation. Methylation occurs on almost all nitrogen sites of the nucleobases (**Figure 1-8**), and additionally on the oxygen of the ribose 2'-OH, C5-atom of pyrimidines, and C2- and C8-atoms on adenosine [51]. The methylations, along with all other known mitochondrial tRNA modifications, are shown on a (mt)tRNA, drawn with both the D-loop/arm, T-loop/arm and a 4-nucleotide variable loop (**Figure 1-8**) [45].

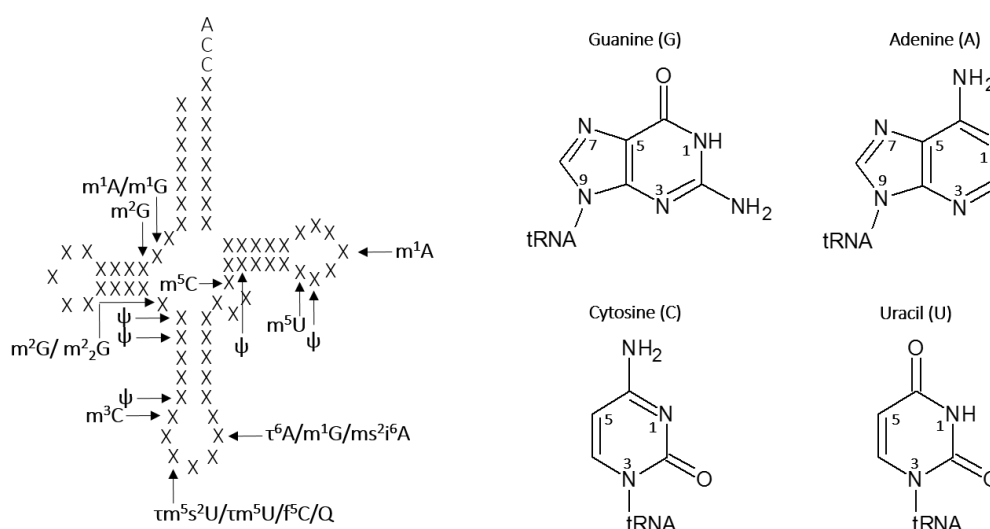


Figure 1-8. Known (mt)tRNA modifications annotated on a (mt)tRNA 2D structure. RNA purine nucleotides are (G) guanine and (A) adenine, and RNA pyrimidines are (C) cytosine and (U) uridine. m¹A, 1-methyladenosine; m²G, N²-methylguanosine; ψ, pseudouridine; Q, queuosine; m¹G, 1-methylguanosine; m²,N²-dimethylguanosine; t⁶A, N⁶-threonylcarbamoyladenosine; tm⁵s²U, 5-taurinomethyl-2-thiouridine; tm⁵U, 5-taurinomethyluridine; m⁵C, 5-methylcytidine; m⁵U, 5-methyluridine; f⁵C, 5-formylcytidine; ms²i⁶A, 2-methylthio-N⁶-isopentenyladenosine; m³C, 3-methylcytidine.

tRNA methylation is performed by the family of RNA methyltransferases (MTases), comprising more than 60 members in the human genome. Based on sequence and topology comparison, RNA MTases are classified into four subfamilies: Rossmann-fold MTases (RFM), SpoU-TrmD MTases (SPOUT), radical-SAM, and FAD/NAD(P)-dependent MTases [51], [52], of which the largest is the RFM family, followed by the SPOUT family [53], [54]. MTases are further divided into tRNA methyltransferase (Trm) subfamilies, where orthologue family members are described either with a *number* (often archaeal or eukaryotic

proteins, e.g. Trm5) or with a *letter* (often archaeal or bacterial proteins, e.g. TrmD) (overview: **Appendix 8.2**).

A non-SDR function of MRPP2 relies on its interaction with the MRPP1 protein, also known as tRNA methyltransferase 10 homologue C (TRMT10C) [14]. MRPP1 belongs to the Trm10 subfamily of SPOUT proteins and currently no bacterial orthologues have been identified for this subfamily. The complex of MRPP1-MRPP2 catalyses the methylation of the N1-atom in a purine (adenine or guanine) at position 9 (forming m¹A9 or m¹G9) on both precursor and mature mitochondrial tRNAs, utilising S-adenosyl-L-methionine (SAM) as the methyl donor (**Figure 1-9**) [14].

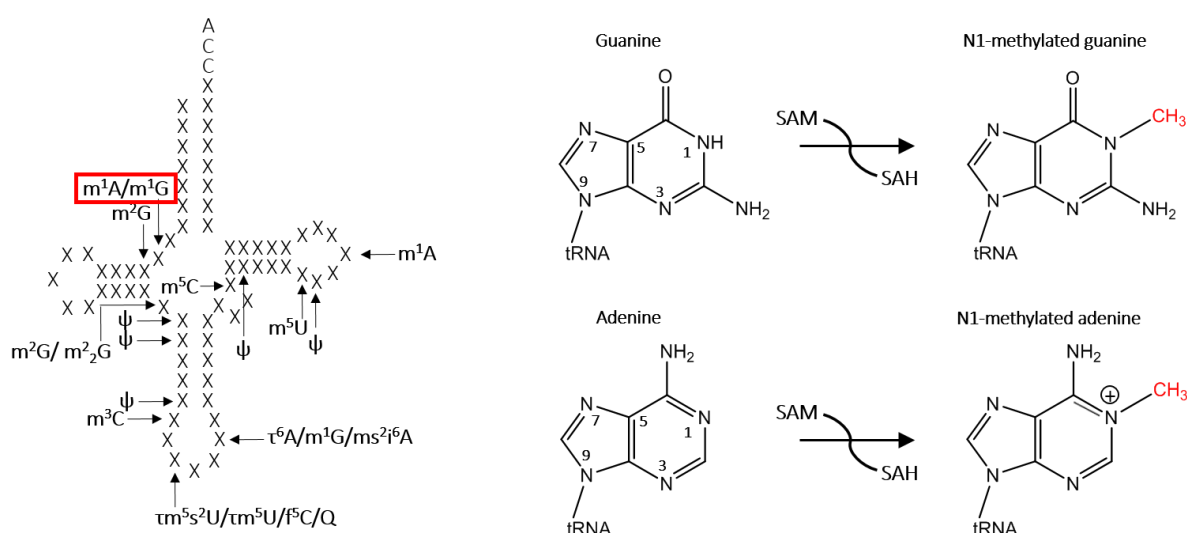


Figure 1-9. m¹A9 activity in (mt)tRNA^{Lys} and m¹G9 activity in (mt)tRNA^{Ile} for the MRPP1-MRPP2 complex.

The mechanism by which the MRPP1-MRPP2 complex methylates the purines in position 9 is not yet understood, but has been investigated for the m¹G37 specific SPOUT family member TrmD from the bacteria *H. influenza* [55], the m¹G37 specific RFM family member Trm5 from the archaea *M. jannaschii* [56], [57], and the m¹A58s specific RFM family member TrmI from the eubacteria *T. thermophilus* [58].

The m¹G37 mechanism involves a deprotonation of the N1-atom by an Asp or Glu residue (general base), a stabilisation of the resulting negative charge on the O6-atom by an

Arg residue (intermediate charge-stabiliser), and an interaction with the N2-atom by a Glu residue. Subsequently, the activated N1-atom makes a nucleophilic attack on the reactive methyl group of SAM (**Figure 1-10**) [59]. TrmD has recently been shown to further utilise a Mg^{2+} ion for stabilisation of the negatively charged intermediate [60].

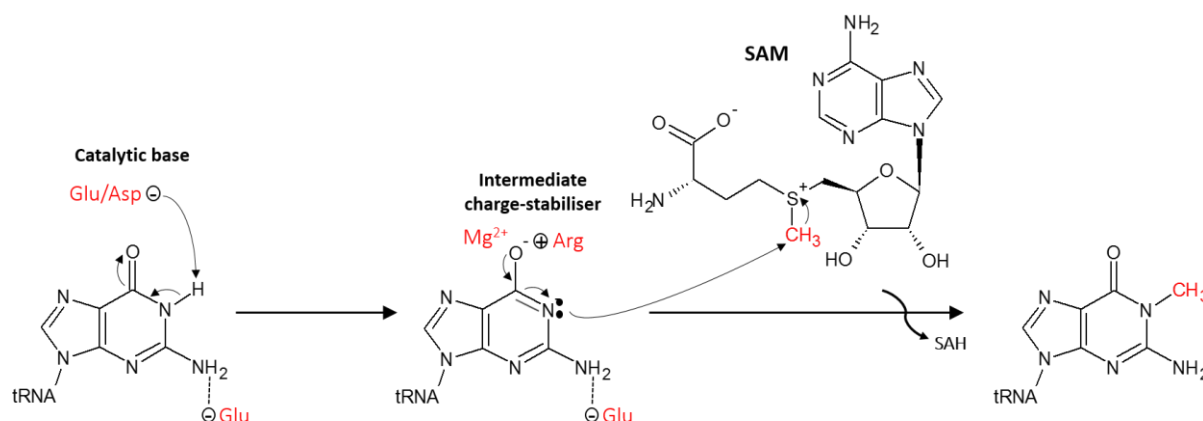


Figure 1-10. Mechanism for m¹G37 formation by TrmD (with Asp as catalytic base) or Trm5 (with Glu as catalytic base). Only the guanine base of tRNA is depicted for clarity.

The m¹A58 mechanism involves a deprotonation as the initial step, but as the N1-atom in the adenine is not protonated at physiological pH, the deprotonation by the Asp general base takes place on the exocyclic N6-atom, resulting in the formation of an imino tautomer. This formation activates the lone pair of the N1-atom for nucleophilic attack on the reactive methyl group of SAM, similar to the methylation of guanine (**Figure 1-11**) [58]. Another role was suggested for the Asp general base in the N1-methylation of adenine in which – rather than deprotonating the N6-atom – it serves merely to position the methyl donor SAM and the target adenine for methylation reaction [58].

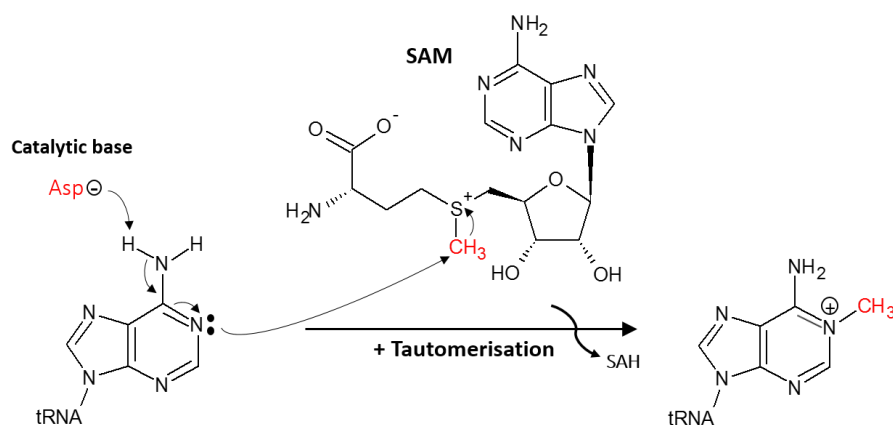


Figure 1-11. One suggested mechanism for m¹A58 formation by TrmI. Only the adenine base of tRNA is depicted. The tautomer of adenine is shown at physiological pH.

In human (mt)tRNA, the N1-methylation of purine-9 performed by the MRPP1-MRPP2 complex has been shown to be important for the stability of human (mt)tRNA^{Lys}, which exists *in vitro* as an extended hairpin structure [61], [62] in equilibrium with a conventional conformation [61], [62]. N1-methylation of purine-9 shifts this equilibrium towards the conventional structure by disrupting a destabilizing Watson-Crick interaction that would otherwise form between A9 and U65 [62]–[64]. The impairment of A9-U65 interaction allows U65 to bind A50 instead, facilitating a more conventional structure [65]. Structural plasticity, similar to what is observed for (mt)tRNA^{Lys}, was also reported for (mt)tRNA^{Leu(UUR)} and (mt)tRNA^{Asp}, indicating that post-transcriptional modifications may also affect the fold of these (mt)tRNAs [66]–[68].

In eukaryotes, m¹A9 is confined to (mt)tRNAs from animal [63], [69], supporting that this modification is important for stabilisation of the ‘bizarre’ (mt)tRNA conformations, but not of cytosolic tRNAs with conventional L-shape [45], [70]. 19 out of 22 human (mt)tRNAs comprise either guanine ($n = 5$) or adenine ($n = 14$) in position 9 [69], [71], therefore it is likely that N1-methylation will have a stabilising effect in all these (mt)tRNAs. A study revealed another important role for m¹G9 modification, related to translation termination efficiency [72]. In this study, the absence of the m¹G9 MTase from *S. cerevisiae* increased read-through at translational termination sites, leading to incorporation of

additional amino acids into the nascent polypeptide. The read-through shows that the non-methylated yeast tRNA outcompetes the translation termination factors that would otherwise recognise the stop codon, suggesting that the m¹G9 methylation decreases the affinity of tRNA for the translational machinery. Nucleotide-9 is located far from the anti-codon and could, if involved in interactions, result in binding of the tRNA to non-specific anti-codons by providing an additional interaction between the tRNA and the translational machinery. By methylation of nucleotide-9, such an interaction could be disrupted, thereby increasing the importance of the interaction with the anti-codon, ultimately decreasing non-specific binding.

The role of MRPP2 in the binary complex with MRPP1 is unclear at present, although the activity of the MRPP1-MRPP2 complex does not depend on either NAD⁺/NADH binding or SDR activity [14]. A hypothesis for a role of MRPP2 in direct interaction with the ‘bizarre’ (mt)tRNAs was supported by the observation that other Rossmann-fold proteins are involved in RNA binding [73], [74], and an *in vivo* crosslinking study that suggested a direct interaction between MRPP2 and some (mt)tRNAs [75]. However, an *in vitro* electrophoretic mobility shift assay (EMSA) with purified, recombinant proteins did not show any direct interaction between MRPP2 and (mt)tRNA [14]. Another hypothesis on the role of MRPP2 was to increase the affinity of MRPP1 for the (mt)tRNA substrate in the N1-methylation reaction [14], and an *in vitro* EMSA assay *did* reveal such an increase in the affinity of MRPP1 for the (mt)tRNAs in the presence of MRPP2 [14].

1.3.3 Mitochondrial precursor tRNA 5'-end processing

Human precursor (mt)tRNAs are transcribed as polycistronic transcripts that undergo processing at both the 5'-end (preceded by the *leader* sequence) and 3'-end (preceded by the *trailer* sequence) to release the mature (mt)tRNAs [76]. Cleavage at the 5'-end is carried out by the endoribonuclease P (RNase P) [77], leaving a phosphate group at

the tRNA 5'-terminus and a hydroxyl group at 3'-end of the removed leader sequence [78], [79] (**Figure 1-12A**). Cleavage at the 3'-end is carried out by the endoribonuclease Z (RNase Z) [80]–[82], leaving a 3'-hydroxyl group on the tRNA and a 5'-phosphate on the removed trailer sequence (**Figure 1-12B**). A few (mt)tRNAs are organised such that the 5'-end of one tRNA is adjacent to the 3'-end of another, and thus processing of the 3'-end of the upstream (mt)tRNA by RNase Z will simultaneously generate the mature 5'-end of the downstream (mt)tRNA without the need to involve RNase P, and *vice versa*. Such a mechanism has been demonstrated for the human (mt)tRNA^{Tyr}-(mt)tRNA^{Cys} and (mt)tRNA^{His}-(mt)tRNA^{Ser(AGY)}-(mt)tRNA^{Leu(CUN)} clusters [48], [83].

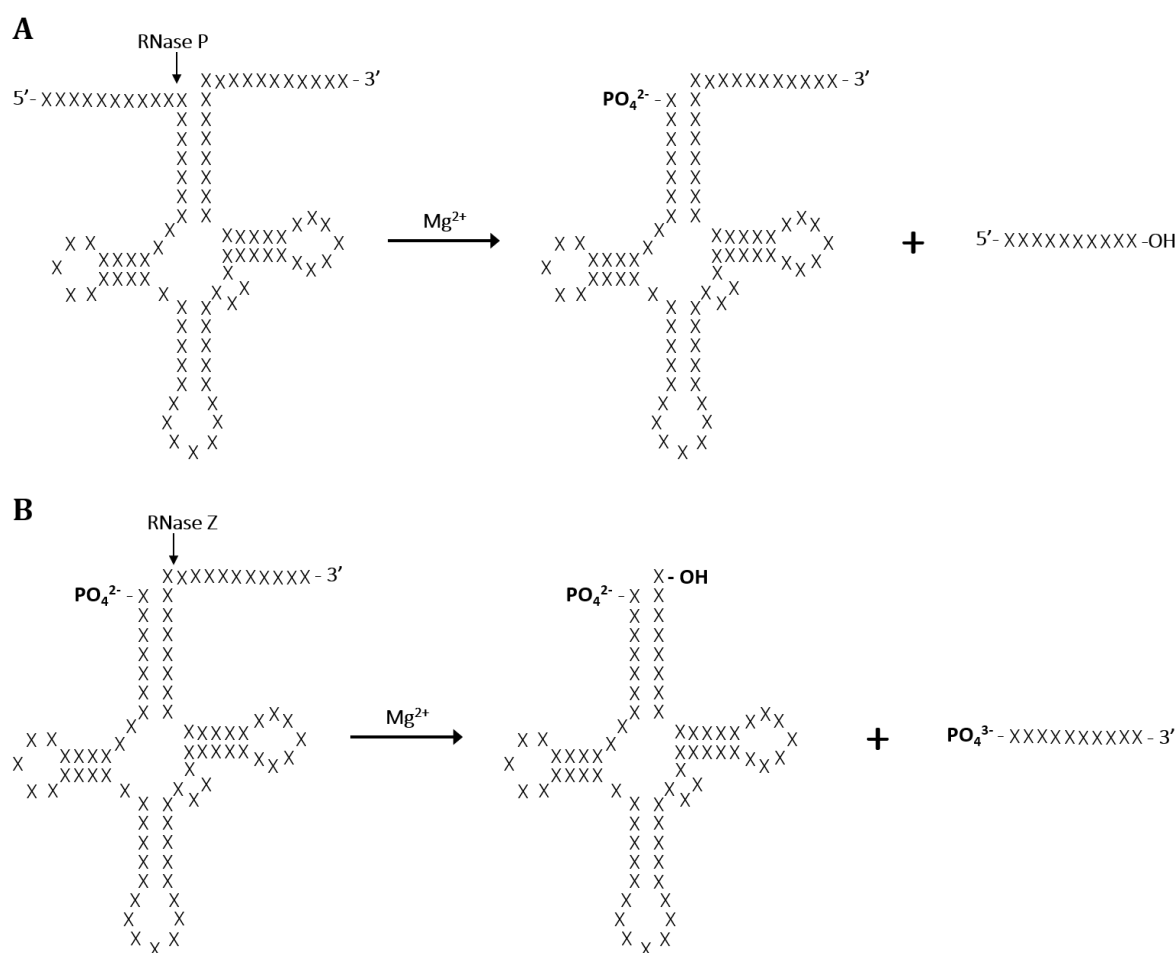


Figure 1-12. A) RNase P activity and B) RNase Z activity towards pre-(mt)tRNA.

For many years, the RNase P complex was thought to be a ribonucleoprotein in all species across the three kingdoms of life [84], [85], comprising a large catalytic RNA as well

as protein components that increase in the number from bacterial (1 protein) [86] to archaeal (4-5 proteins) [87] and to eukaryotic nuclear (>9 proteins) RNase P [88] (**Figure 1-13**).

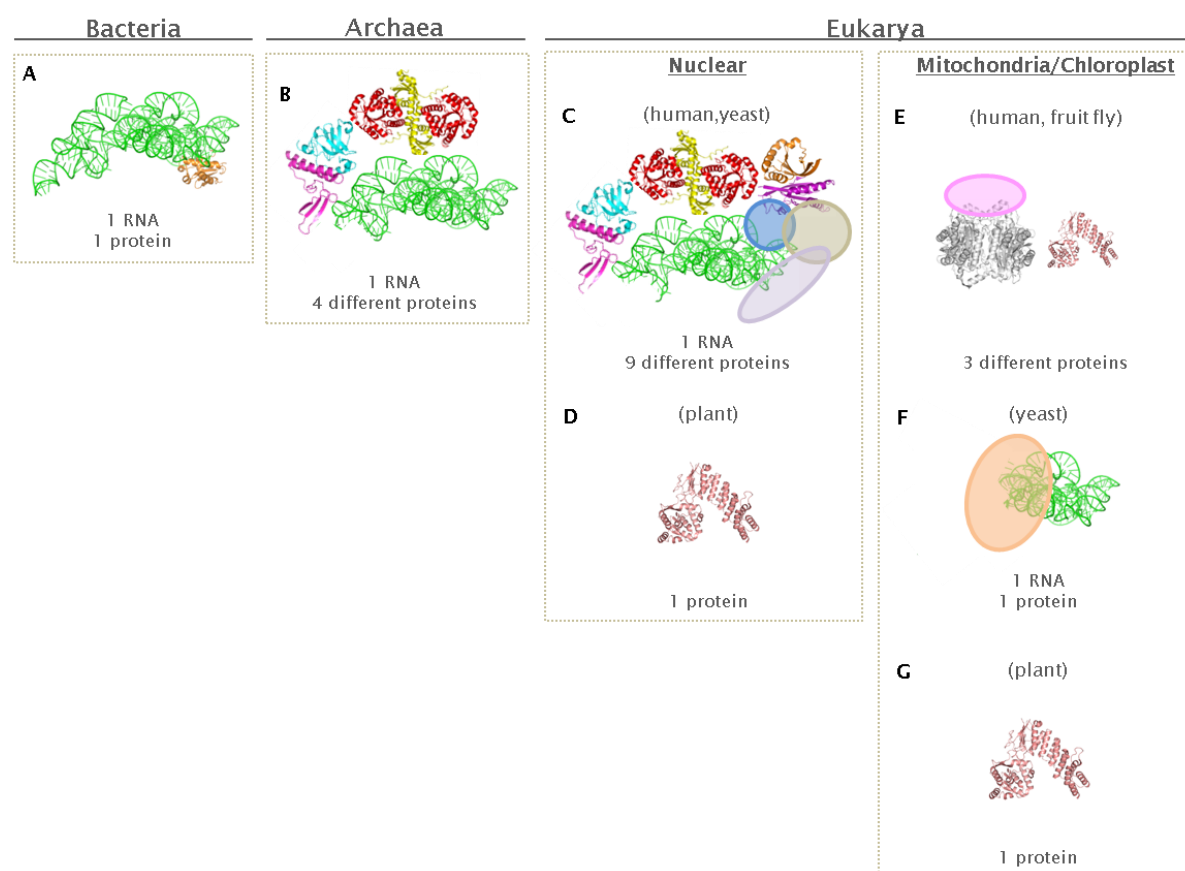


Figure 1-13. The RNase P complex from all domains of life. Crystal structures are shown if known, and proteins with no known structures are indicated as circles/ovals.

In 1998, Wang *et al.* suggested that RNase P in chloroplasts might not contain an RNA component [89], and in 2010, Giege *et al.* identified the first protein-only RNase P proteins in cytoplasm and organelles from *A. thaliana* [90]. These proteins are single-subunit, Mg^{2+} -dependent endoribonucleases named proteinaceous ribonuclease P (PRORP) [91]–[93]. Homologues of these plant PRORPs are found in the mitochondria of higher and lower eukaryotes [94], where they carry out RNase P activity either alone (e.g. in *T. brucei*) or in complex with other proteins and Human) (**Figure 1-13**) [13], [95]–[97]. The human PRORP homologue, MRPP3, was discovered in 2008 by Rossmanith *et al.* [13], and was shown to require MRPP1 and MRPP2 for RNase P activity. In addition to the RNase P activity, plant PRORPs play additional roles in maturation of

small nucleolar RNAs (snoRNAs) and mitochondrial messenger RNAs (mRNAs) by recognising t-elements (tRNA-like structures) [98], [99], suggesting that MRPP3 could likewise cover additional activities in human.

PRORPs, including MRPP3, are members of the Nedd4-YacP nuclease (NYN) family of proteins [100]. NYN proteins are found in all three kingdoms of life and are involved in core RNA metabolic processes such as ribosomal biogenesis and processing of tRNA and rRNA [100], [101]. The NYN domain is often fused to other RNA-binding domains [100], as is the case for PRORPs in which the NYN-domain is C-terminally linked to an RNA-binding pentatricopeptide repeat (PPR) domain [102]–[105]. PPR-domains contains small helix-turn-helix motifs [106] in tandem arrays of two to 30 repeats [101], [102], [107], to make up a super-helical PPR-domain with a central groove [102], capable of binding single stranded RNA in a modular fashion [101], [108]. PPR-domain containing proteins function in transcription and RNA processing, splicing, editing, stability and translation [104], [109], and so far, the PPR-domain has been shown to be a non-catalytic RNA-binding fold [104]. PPR proteins are grouped into the classical P-class and the less conserved PLP-class depending on the repeat size [104], [107]. PRORPs belong to the P-class with repeats of 35 amino acids in length [102].

The N1-methylation activity (m^1R9) of the MRPP1-MRPP2 complex has been shown to be dispensable for RNase P activity (5'-end ribonuclease) of the MRPP1-MRPP2-MRPP3 complex *in vitro* [14]. Therefore, processing can take place on unmethylated substrates, implying that the MRPP1-MRPP2 complex does not recognise the 3'-trailer or 5'-leader in the pre-(mt)tRNA. Indeed, a crosslinking study suggests a recognition site on the D-arm/loop and acceptor-stem/loop of the tRNA [75].

1.3.4 Summary of non-enzymatic functions

The multiple enzymatic and non-enzymatic functions of MRPP2, and indeed MRPP1 as well, constitute examples of ‘moonlighting enzymes’: proteins with several functions, some of which may not relate to their enzyme activities [8]. Their intricate interplay, and additional interactions with MRPP3, provide a novel, functional link between (mt)tRNA modification and processing with general amino acid and fatty acid metabolism, which merit further studies.

1.4 Mitochondrial diseases

Each human cell contains a varying number of mitochondria depending on cell type, e.g. 1000-2000 in liver cells. Each mitochondrion comprises from two to as many as 50 copies of the (mt)DNA [110]. The main role of the mitochondria is production of ATP via oxidative phosphorylation [110], and mitochondrial dysfunction therefore often relates to malfunction of tissues with high demand of energy such as brain, heart, liver, muscles and kidney [45]. Malfunction of these tissues gives rise to a wide range of clinical phenotypes such as loss of motor control, muscle weakness, retarded growth, cardiac diseases, seizures, lactic acidosis, developmental delays, visual/hearing problems, *etc.* [45]. Mitochondrial functions rely on nuclear encoded proteins (~1000) that are transported into the mitochondria after translation in the cytosol [110]. Mitochondrial dysfunction can therefore arise both from mutations in the (mt)DNA itself and from mutations in the part of the nuclear DNA that codes for proteins that function in the mitochondria [111].

1.4.1 Diseases linked to MRPP2

The production of ATP in mitochondria results from the initial breakdown of e.g. fatty acids involving MRPP2, providing one explanation why patients carrying catalytic missense mutations in MRPP2 often display the clinical hallmarks of mitochondrial dysfunction (**Table 1-4**). An additional explanation for the observed mitochondrial

dysfunction is as a direct result of faulty (mt)tRNA modifications. Patients often also present with ketoaciduria, a metabolic condition characterised by accumulating levels of keto acids from defects in amino acid and fatty acid breakdown [112].

1.4.1.1 HSD10 disease

The gene encoding MRPP2 (*HSD17B10*) maps to chromosome Xp11.2 [5] and, in females, the *HSD17B10* gene undergo random X-inactivation, like most other genes on X-chromosomes [113].

RNA interference (RNAi) knock-down of MRPP2 was shown to be embryonically lethal in *Drosophila* [97], [114] and mice [115], and loss of MRPP2 either by *HSD17B10* mutation in patient fibroblasts or small interfering RNA (siRNA) knock-down in HeLa, HEK-293 and HRT18 cells affected the processing of pre-(mt)tRNA [116]. Consistent with these findings, inherited mutations in the *HSD17B10* gene lead to the X-linked, rare genetic disorder HSD10 disease (MIM#300438). This disease was first reported in 2000 [36] and, to date, 31 patients, carrying 15 different missense mutations, have been identified (**Table 1-4**). Most patients are hemizygous male individuals, but some symptomatic heterozygous women are reported due to random X-inactivation [113]. The genotypes show different degrees of severity of the clinical phenotype which has led to the classification below [112], [117].

Neonatal form: this is the most severe form where patients show little neurological development, ketoaciduria, severe progressive cardiomyopathy and early death.

Infantile form: this is the classical form of HSD10 disease characterised by ketoaciduria and normal development in the first 6 to 18 months of life with subsequent retinopathy, cardiomyopathy and progressive neurodegeneration, leading to death at the age of 2-4 years or later.

Juvenile form: in this milder form patients show ketoaciduria, but are otherwise asymptomatic in the first five years of age with subsequent slow neurological deterioration.

Atypical form: patients suffering from this mild form exhibit little to no growth retardation and have normal development in childhood. They show typical urinary organic acid profiles and potential transient metabolic decompensation.

Heterozygous females show mild to moderate developmental delay and intellectual disabilities, but can also be asymptomatic carriers [118]–[120].

The identified patients with their corresponding nucleotide (nt)/amino acid (aa) changes, gender and clinical classification are shown in **Table 1-4**. The mutation p.R130C (c.388C>T) is the most frequently identified mutations, and is caused by spontaneous deamination to thymine [121].

Mutation		Clinical classification	Gender	Reference	Year of identification	
nt	aa					
c.34G>C	p.V12L	N.A.	♀	P.C.	2013	
c.194T>C	p.V65A	Atypical	♂	[120]	2011	
c.257A>G	p.D86G	Neonatal	♂	[115]	2010	
c.364C>G	p.L122V	Infantile	♂	[122]	2004	
			♂	[12]	2001	
c.388C>T	p.R130C	Infantile	♂	[112]	2012	
			♂	[123]	2007	
			♂	[119]	2009	
			♂	[36]	2000	
			♂	[124]	2002	
			♂	[125]	2004	
			♂	[126]	2003	
			♂	[12]	2001	
			Early infantile	♂	[118]	2005
			Late infantile	♂	[112]	2012
	I.D.	♀	[124]	2002		
c.460G>A	p.A154T	Atypical	♂	[117]	2014	
c.470C>T	p.A157V	Atypical	♂	[127], [128]	2015	
			♂	[127], [128]	2015	
			♀	[127], [128]	2015	
c.473C>T	p.A158V	Atypical	♂	[129]	2015	
c.495A>C	p.Q165H	Atypical	♂	[115]	2010	
			♂	[115]	2010	
c.626G>A	p.V176M	Atypical	♂	P.C.	2011	
c.628C>T	p.P210S	Early infantile	♂	[119]	2009	
		Infantile	♂	[119]	2009	
c.634A>G	p.K212E	Infantile	♂	[130]	2016	
c.677G>A	p.R226Q	Neonatal	♂	[119]	2009	
c.740A>G	p.N247S	Neonatal	♂	[118]	2005	
		Early onset psychomotor delay, ataxia	♀	[118]	2005	
c.745G>C	p.E249Q	Juvenile	♂	[131]	2002	

Table 1-4. The identified patients suffering from HSD10 disease and their corresponding mutation, gender and clinical classification. P.C.: Personal communication. I.D.: Intellectual disability. N.A. Not available.

HSD10 disease was originally characterised by urine testing from patients, showing elevated levels of metabolites (tiglylglycine and 2-methyl-3-hydroxybutyrate) found upstream of the MRPP2 enzymatic step in the isoleucine degradation pathway. HSD10 disease was therefore initially considered a classical aciduria, supported by *in vivo* findings of lowered L-methyl-3-HAD activity in cultured skin fibroblasts from patients [12]. Following this hypothesis, patients were subjected to isoleucine-restricted diet, which lowered the levels of isoleucine metabolites found in urine, but had little or no effect on neurological symptoms [36], [112], [124], [126], [131], suggesting that it was not a classical aciduria after all. In support of this conclusion, patients suffering from deficiency of the β -ketothiolase (β -KT) enzyme (MIM #203750), positioned immediately after the MRPP2 step in the isoleucine degradation pathway, does not present with any neurodegeneration as found in HSD10 disease, despite accumulation of the same upstream metabolites [132], [133]. Also, observations from a knock-out/rescue experiment in *Xenopus* embryos demonstrated no correlation between residual L-methyl-3-HAD activity of MRPP2 and severity of the clinical presentation. For example, the MRPP2 missense mutation p.D86G, giving rise to the severe neonatal form of HSD10 disease, showed 28% residual SDR enzyme activity, whereas the mutation p.Q165H, giving rise to the almost asymptomatic atypical form of the disease, showed no residual SDR activity [115].

The inconsistency between enzymatic defect of MRPP2/accumulation of metabolites and severity of the disease has led to the hypothesis that non-enzymatic functions of MRPP2 are the likely pathogenic drivers of HSD10 disease [115]. This hypothesis has gathered momentum with the discovery of MRPP2's role in (mt)tRNA modification and processing [13], [14]. A defect in (mt)tRNA maturation would also explain the respiratory-chain dysfunction and structural abnormalities of the mitochondria observed in some patient fibroblasts [112], [124], [131]. Vilardo *et al.* (2015) [134] and Falk *et al.* (2016) [130]

characterised five reported patient mutations *in vitro* by testing the recombinant mutant proteins against the oxo-red activity of MRPP2 (acetoacetyl-CoA reduction) and non-enzymatic m¹G9 MTase and RNase P activities on (mt)tRNA and pre-(mt)tRNA. In contrast with the *in vivo* knock-out/rescue studies, the *in vitro* characterisation suggested a potential link between residual SDR function and disease severity, further discussed in **Chapter 6**.

1.4.2 Diseases linked to MRPP2 binding partners

Missense mutations have recently been identified in the two binding partners of MRPP2, MRPP1 and MRPP3, giving rise to diseases with clinical manifestations consistent with mitochondrial dysfunction.

1.4.2.1 COXPD30 disease

The gene encoding MRPP1 (*RG9MTD1*) is located at chromosome 3q12.3. RNAi knock-down of MRPP1 was shown to be embryonically lethal in *Drosophila* [97], and result in reduced mitochondrial respiration in HeLa cells [135], respectively. Consistent with this, inherited mutations in MRPP1 give rise to the autosomal recessive disorder, combined oxidative phosphorylation deficiency 30 (COXPD30) (MIM#616974). A male patient was found to be compound heterozygous for c.542G>T/p.R181L and c.814A>G/p.T272A, whereas a female patient was found to be homozygous for the c.542G>T/p.R181L mutation (**Table 1-5**). The two patients both presented at birth with lactic acidosis, hypotonia, feeding difficulties, and deafness with death at 5 months of age after respiratory failure [136]. Analysis of patient fibroblasts showed decreased levels of MRPP1, indicative of lowered protein stability, and a slight accumulation of pre-(mt)tRNAs with no severe effect on steady-state levels of mature (mt)tRNA. No effect was observed on m¹R9-methylation of (mt)tRNAs catalysed by MRPP1, prompting the suggestion that the mutations in MRPP1 affect the pre-(mt)tRNA processing carried out by the ternary RNase P complex.

Mutation		Clinical presentation	Gender	Reference	Year of identification
nt	aa				
c.814A>G c.542G>T	p.T272A p.R181L	COXPD30 disease	♂	[136]	2015
c.542G>T c.542G>T	p.R181L p.R181L	COXPD30 disease	♀	[136]	2015

Table 1-5. The identified patients suffering from COXPD30 due to mutations in MRPP1 and their corresponding genotypes (heterozygous), gender and clinical presentation (disease form).

1.4.2.2 Perrault syndrome

Inherited mutations in a number of – mostly mitochondrial – proteins, give rise to the autosomal recessive disorder Perrault syndrome (MIM#233400), characterised by sensorineural deafness and ovarian dysgenesis. Patients can also present with neurologic manifestations, including mild mental retardation. Recently, the gene encoding MRPP3 (*KIAA0391*) was added to the growing list of genes that, when mutated, cause Perrault syndrome. RNAi knock-down of MRPP3 was shown to be embryonically lethal in *Drosophila*, and cause reduced levels of processed (mt)tRNA in HeLa cells [97]. The *KIAA0391* gene is located on chromosome 14q13.2, and the missense mutation c.1454C>T/p.A485V (Table 1-6) is reported in three homozygous female patients with Perrault syndrome [137], [138].

Mutation		Clinical presentation	Gender	Reference	Year of identification
nt	aa				
c.1454C>T	p.A485V	Perrault syndrome	♀ ♀ ♀	[137], [138] [137], [138] [137], [138]	2015 2015 2015

Table 1-6. The identified patients suffering from Perrault syndrome due to mutations in MRPP3 and their corresponding mutations, gender and clinical presentation (disease form).

1.4.3 Summary of diseases associated with MRPP2 and its binding partners

Missense mutations in MRPP2, leading to HSD10 disease, can result in reduction/impairment of both enzymatic and non-enzymatic functions of MRPP2. The underlying molecular basis for HSD10 disease is still unexplained and more studies at the

biochemical and cellular levels are clearly needed to understand the genotype-phenotype relationship. Missense mutations in the MRPP2 binding partners, MRPP1 and MRPP3, cause COXPD30 disease and Perrault syndrome, respectively, both of which are very novel, rare diseases that present with phenotypes typical for mitochondrial dysfunction.

1.5 Thesis outline

The aim of this DPhil thesis is to establish the effect of disease-causing missense mutations of MRPP2 at the protein level, using a structural and biochemical approach to first characterise the binary (MRPP1-MRPP2) and ternary (MRPP1-MRPP2-MRPP3) complexes, and second use the gained knowledge, in combination with activity assays, to understand how mutations in MRPP2 affects both the catalytic SDR function and the (mt)tRNA moonlighting function of MRPP2. The fundamental question set out for this thesis is: **Does a correlation exist between the structural location of the MRPP2 patient missense mutations and their effects on the MRPP2 functions?** I will address this question by breaking it down into three sub-questions, which form the basis of the thesis:

Aim 1) Interactions: How do the three proteins MRPP1, MRPP2 and MRPP3 interact with each other and with their respective cofactors and substrates?

Aim 2) Activity: What components and regions of the proteins in the binary (MRPP1-MRPP2) and ternary (MRPP1-MRPP2-MRPP3) complexes are important for (mt)tRNA methylation and processing activity?

Aim 3) HSD10 patient mutants: What are the effects of introducing the patient mutations in MRPP2 at the protein level, and can the effects be explained based on the obtained structural and biochemical results?

Chapter 2 describes the characterisation of the poorly understood MRPP1 protein to reveal its oligomeric state and ligand binding properties, and reports the crystal structure of the MRPP1 methyltransferase domain bound to the cofactor SAM. A low-resolution small-angle X-ray structure (SAXS) model for the mature protein is also presented.

Chapter 3 describes a mapping of the protein regions that are involved in formation of the binary complex of MRPP1-MRPP2 and the involvement of these regions in the m¹R9 MTase activity. A further characterisation of catalytic residues in MRPP1 is performed, and low-resolution SAXS models of the binary complex, with and without the bound (mt)tRNA substrate, are presented.

Chapter 4 is divided into two parts. The first part describes the MRPP3 alone in relations to the oligomeric state and ligand binding properties. The second part provides a low-resolution SAXS model of the ternary MRPP1-MRPP2-MRPP3 complex bound to the pre-(mt)tRNA substrate, and a mapping of the order of complex assembly along with investigations of the regions involved in RNase P activity and complex formation.

Chapter 5 is divided into two parts describing MRPP2 mutant proteins associated with 1) HSD10 disease and 2) Perrault syndrome. Both parts detail the biochemical effect of known patient mutations on the protein. The first part further presents an in depth characterisation of the activity of the novel patient mutations p.V12L, p.V176M, whereas the second part presents activity studies and complex assembly investigations for the novel mutation p.A485V in MRPP3.

Chapter 6 contains a discussion of the key findings from Chapters 2-5 and relates these findings to published work. The chapter further outlines the future directions for this project.

Chapter 7 contains the experimental procedures and background for each utilised method.

Appendix is provided at the end of the thesis with useful tables and figures for results and discussions.

2 CHAPTER

CHARACTERISATION OF MRPP1 ALONE

At the start of this project no structural information was available on the catalytically active subunit (MRPP1) of the binary methyltransferase complex MRPP1-MRPP2. MRPP1, also known as TRMT10C (tRNA methyltransferase type 10 C), belongs to the Trm10 family of the SPOUT superfamily of proteins that methylate bases and ribose moieties of rRNAs and tRNAs in all three kingdoms of life [51], [53], [54], [139]. The Trm10 subfamily of the SPOUT superfamily contains all MTases that methylate the N1-position in purine bases (A/G) located at position 9 of the tRNA (forming m¹A9 / m¹G9) [51]. In addition to the catalytically active MTase SPOUT domain, many SPOUT proteins contain nucleic acid binding domains, e.g. the pseudouridine synthase and archaeosine transglycosylase (PUA) domain or the thiouridine synthases, methylases and pseudouridine synthases (THUMP) domain, indicating that SPOUT proteins function primarily as MTases towards nucleic acids [54], [140], [141].

This chapter contains the X-ray crystallographic structure and SAXS model of MRPP1, along with a biochemical characterisation of the oligomeric state and cofactor binding properties of various domains.

2.1 Domain organisation

MRPP1 is a 403-amino-acid long protein (**Figure 2-1**) with an N-terminal cleavable mitochondrial targeting signal (MTS) peptide (residues 1-39). The mature protein, starting at residue 40, contains an N-terminal domain (N-domain; residues 40-213) with mainly helical secondary structures [142], followed by a SPOUT-type methyltransferase (MT) domain

(MT-domain; residues 213-380) [143] and a small, helical C-terminal extension (C-extension; residues 381-403) [144].

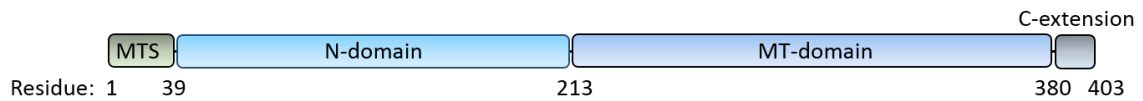


Figure 2-1. Outline of predicted domains in full-length MRPP1 (residue 1-403). MTS: mitochondrial targeting signal. MT: methyltransferase.

Using the DISOPRED3 [145] program in the PSIPRED server [146], protein disorder was predicted in the N-domain (residues 40-60, 74-86 and 153-175) and C-extension (residues 382-403) (**Figure 2-2**). These regions are predicted to be protein-interaction regions, suggesting that they could be involved in interaction with MRPP2 or MRPP3.

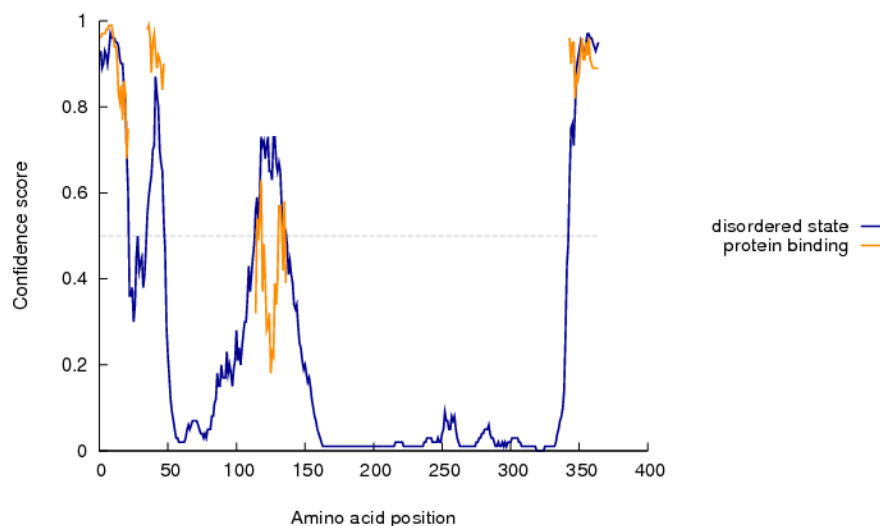


Figure 2-2. Intrinsic disorder profile for MRPP1 predicted by DISOPRED3 [145]. Amino acid position 0 corresponds to residue 40 in MRPP1.

2.2 Cloning, small scale expression and purification

Based on the structural predictions above, several recombinant protein constructs were designed spanning either mature MRPP1 (without the MTS peptide) or the separate N- and MT-domains with and without the C-extension (**Figure 2-3**). The constructs were initially cloned into a vector incorporating an N-terminal His₆-tag followed by a Tobacco Etch Virus (TEV) protease cleavage site.

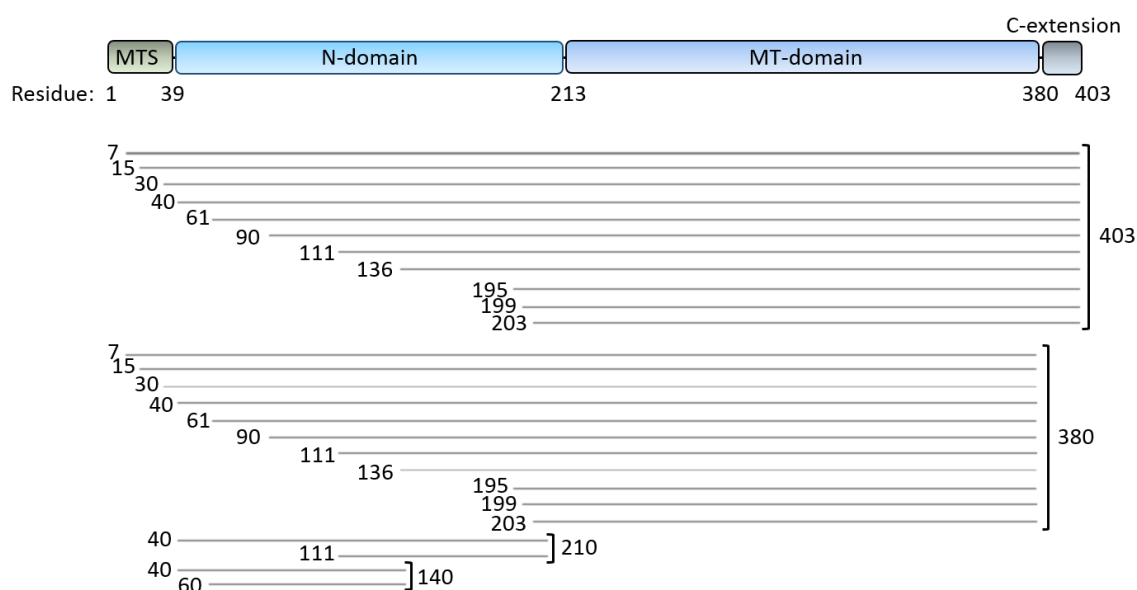


Figure 2-3. Constructs of MRPP1 cloned into *E. coli* expression vector. The boundaries for each construct are numbered.

To test the expression and solubility levels of the cloned constructs, *E. coli* cells containing these plasmids were cultured in 50 mL Terrific Broth (TB) media at 37°C and induced at 18°C with isopropyl β -D-1-thiogalactopyranoside (IPTG). The MRPP1 proteins were extracted by centrifugation followed by sonication of the cell pellet, and subsequently purified by immobilised metal affinity chromatography (IMAC) using a Ni-charged nitrilotriacetic acid (Ni-NTA) resin. The elution from the IMAC column was compared to a sample of the total lysate after sonication to determine 1) the expression as the amount of protein in the total lysate, and 2) final yield as the amount that is subsequently found in the eluted fraction, for each construct. The final yield is expected to correlate with protein solubility based on the assumption that most of the protein in the supernatant is captured during IMAC purification. The final yield of protein will therefore be referred to as the soluble protein in this work.

The result of the small-scale expression and purification is shown in **Figure 2-4**. All cloned constructs with a start boundary before residue 90 were insoluble (grey, **Figure 2-4**). After residue 90, several constructs were found to be soluble but

degrading from mid-sequence (**yellow**, **Figure 2-4**) whereas constructs with a start boundary at, or after, residue 195 had a tendency to degrade from the C-terminus (**orange**, **Figure 2-4**), even in the presence of high concentration of protease inhibitors. Constructs spanning only the predicted MT-domain with boundaries between residues 195 and 380 were soluble and stable (**green**, **Figure 2-4**).

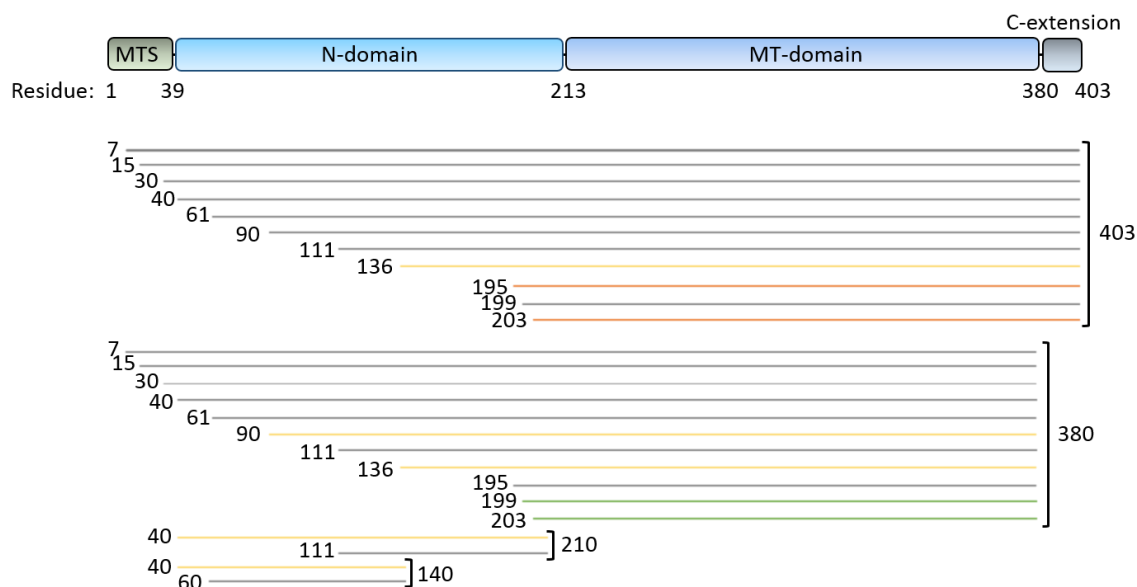


Figure 2-4. Cloned constructs of MRPP1 colour-coded for solubility levels. Insoluble constructs: grey. Constructs that are soluble but degrade from the N- and C-termini: **yellow**. Constructs that are soluble but degrade from the N-terminus: **orange**. Soluble, stable constructs: **green**.

In an attempt to obtain additional soluble protein fragments, several constructs (**Figure 2-5**) were cloned into a vector providing an N-terminal tandem His₆-tag and glutathione S-transferase (GST) fusion protein known to increase solubility [147]. Small-scale expression and purification for these constructs were performed and yielded soluble proteins of constructs with a start boundary at residue 111, which were previously found to be insoluble. However, these proteins were prone to degradation as observed for previous constructs.

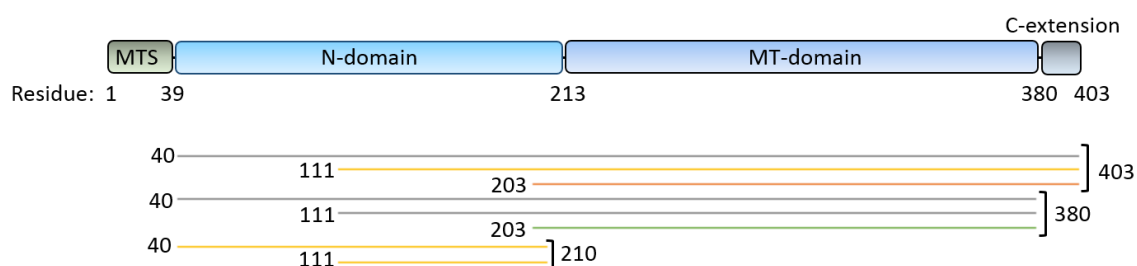


Figure 2-5. Cloned constructs of MRPP1. Insoluble constructs: grey. Soluble proteins that degrades from the N- and C-terminus: yellow. Soluble constructs that degrade from the N-terminus: orange. Soluble, stable constructs: green.

From the above, still no full-length, mature protein of MRPP1 was obtained. To produce this protein, constructs spanning the mature MRPP1 protein with and without the C-extension (40-403 and 40-380, respectively) were cloned into baculovirus vectors for expression trials in insect cells. The vectors all carried an N- or C-terminal His₆- or His₁₀-affinity tags with a TEV cleavage site. Some constructs additionally carried tags known to increase solubility (GST) or facilitate protein purification (Z_{basic}, Flag) [148], [149]. Expression of these constructs in insect cells was performed by Biotechnology group (SGC, Oxford) and small-scale protein extraction and purification was done similarly to the procedure outlined for the *E. coli* cultures.

Although the new expression system and vectors resulted in soluble protein for all clones, the majority of the purified proteins were prone to degradation as previously observed. The mature MRPP1-Flag-His₁₀ construct, expressed in insect cells, proved to be both soluble and less prone to degradation than the equivalent constructs bearing N-terminal tags, e.g. His₆-MRPP1. Based on this finding, the mature MRPP1 construct, with and without the C-extension, were re-cloned into an *E. coli* expression vector with a C-terminal His₆-tag and TEV protease cleavage site. These constructs (residues 40-403) proved to be soluble in *E. coli*, and less degraded during the course of purification than the corresponding constructs with the N-terminal His₆-tag. Degradation of the construct could be further diminished by addition of further purification steps and by freezing at -80°C for storage.

None of the N-terminally His₆-tagged constructs of MRPP1, with a start boundary between residues 195 and 203, could be cleaved by excess TEV protease and long incubation times even at room temperature. To facilitate tag-cleavage, the most soluble MT-domain constructs (203-403 and 203-380) were re-cloned into the *E. coli* expression vector offering a C-terminal His₆-tag and TEV protease cleavage site, with the rationale that a His₆-tag at this position for the mature protein construct (residues 40-403) was cleavable by TEV protease, which was indeed also the case for the re-cloned constructs.

In summary, the constructs used and described hereafter are denoted mature MRPP1 (residues 40-403), MT+C-domain(Δ 194) (residues 195-403), MT+C-domain(Δ 202) (residues 203-403), MT-domain (residues 203-380) and N-domain (residues 40-210) (**Figure 2-6**). The two soluble constructs obtained for the MT+C regions will both be used in assays throughout this thesis, but the N-boundary will be indicated as stated.

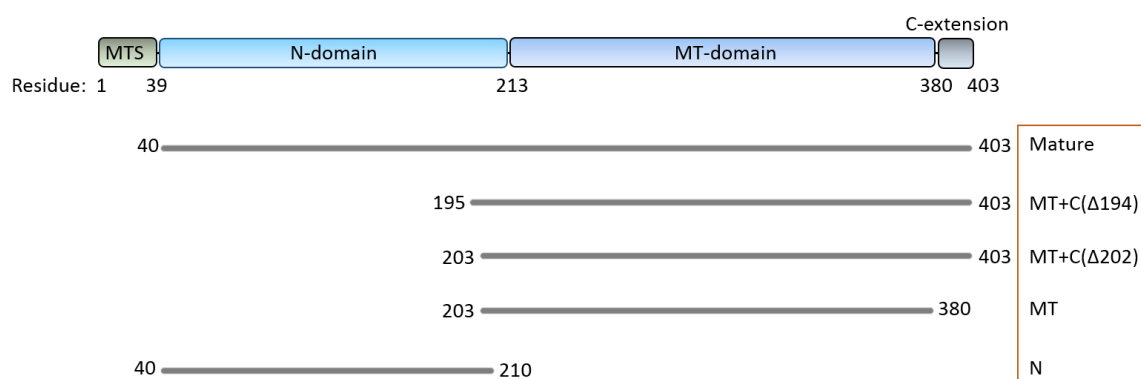


Figure 2-6. Cloned constructs of MRPP1 used throughout this chapter.

2.3 Large-scale Expression and Purification

The constructs for mature MRPP1 and the N-, MT-, MT+C(Δ 194)-, and MT+C(Δ 202)-domains were transformed into *E. coli* cells and the cells were cultured in batches of either 6L or 12L depending on solubility determined from the small-scale test. Cultures were grown in TB media at 37°C and induced at 18°C with IPTG. The cells were harvested by centrifugation and lysed by homogenisation with addition of protease inhibitors to stall degradation.

2.3.1 Mature MRPP1

The mature MRPP1 protein was purified by IMAC using Ni-NTA resin followed by size-exclusion chromatography (SEC) on a Superdex 75 (S75) column. The S75 column was chosen despite the higher molecular weight (MW) for the mature MRPP1 protein, in an attempt to separate the intact protein from the smaller degradation products. TEV protease treatment for tag cleavage of the protein resulted in *some* protein precipitation and thus this step was subsequently omitted. A final cation-exchange chromatography (cation-EX) step was performed on a HiTrap sulfopropyl (SP) high performance (HP) column. The purification procedure, along with the final electrospray ionisation mass spectrometry (ESI-MS) spectrum, for mature MRPP1 is outlined in **Figure 2-7**, and described in more detail in **Section 7.3**.

The mature MRPP1 protein had a tendency to degrade during purification. However, after the final cation-EX step the protein was stable even after 12 hours of incubation at room temperature. ESI-MS analysis of the purified sample showed a molecular mass of 45243.95 Da comparable to the theoretical molecular mass of 45374.2 Da. A molecular mass analysis using the PAWS software revealed that the difference in mass could correspond to loss of either the initial Met residue from the MRPP1 sequence or of the last, vector encoded, Lys residue.

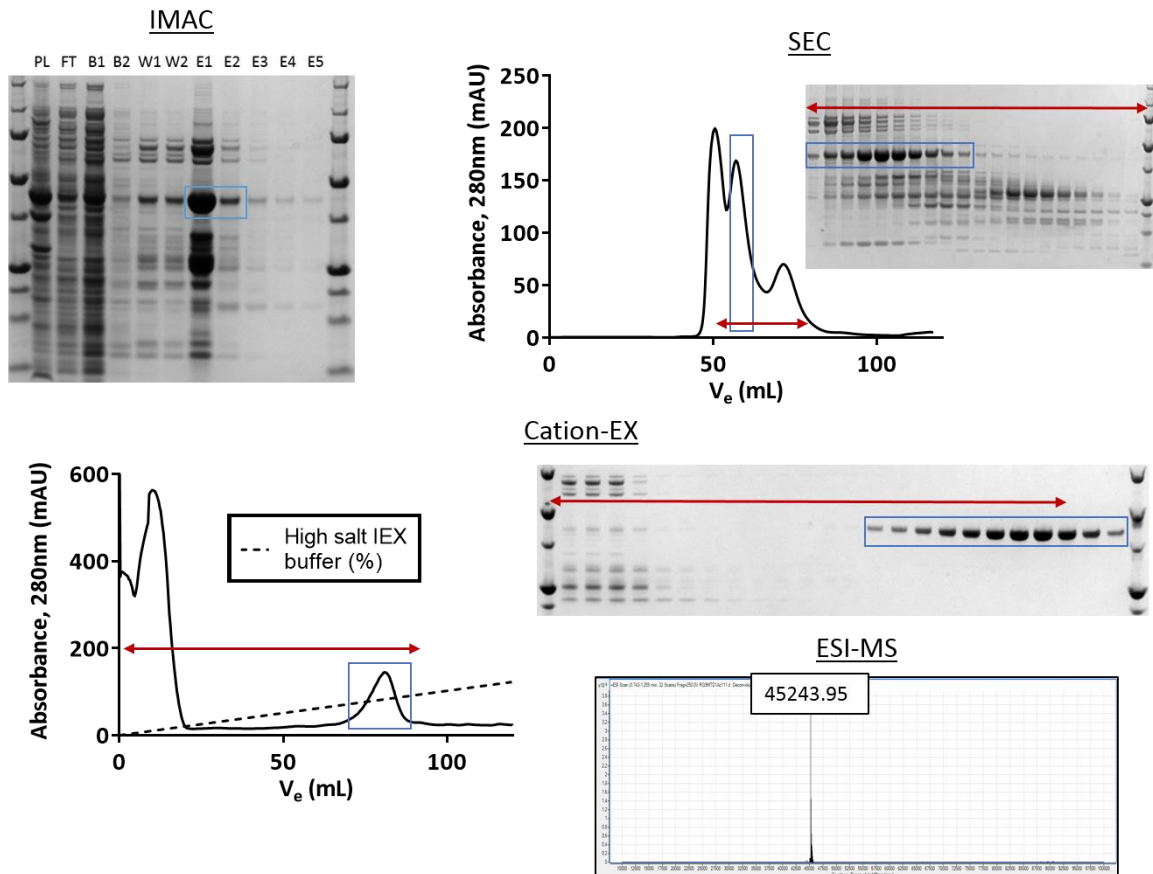


Figure 2-7. Large scale purification of mature MRPP1. The example shows SDS-PAGE analysis of the initial IMAC, chromatogram from SEC and SDS-PAGE analysis of peak fractions, and chromatogram from cation-EX and SDS-PAGE analysis of peak fractions. ESI-MS spectrum of the sample after purification is also shown. Red arrows correspond to fractions analysed on SDS-PAGE. Blue boxes indicate the pooled fractions. V_e : elution volume, PL: pellet, FT: flow through, B1 and B2: binding buffer fractions, W1 and W2: wash buffer fractions, E1-5: elution buffer fractions.

Tandem mass spectrometry (MS/MS) determination of tryptic digest products identified MRPP1 peptides covering almost the full sequence with residue 44-403 (**Figure 2-8**), indicating that the loss of mass in ESI-MS might indeed be the initial Met residue.

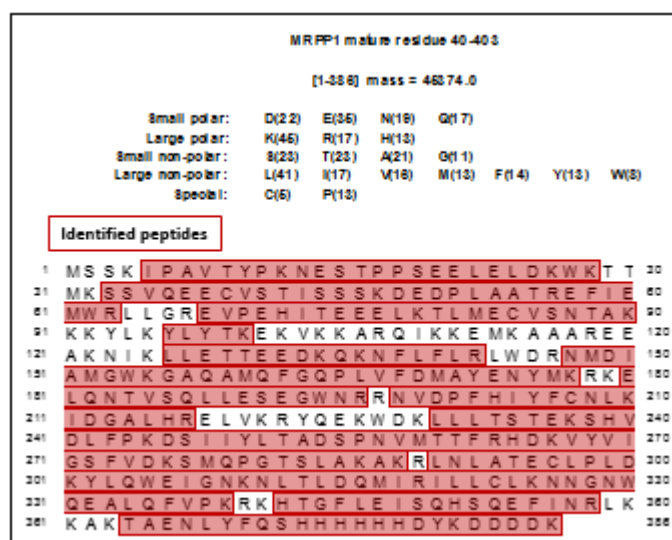


Figure 2-8. Peptides of the mature MRPP1 protein identified with MS/MS analysis of tryptic digest products (red boxes). Sequence numbering M1-T364 corresponds to MRPP1 M40-T403. The last 22 amino acids are encoded by the vector.

2.3.2 N-domain

The N-domain of MRPP1 (residues 40-210) was purified by IMAC with Ni-NTA resin followed by anion-exchange chromatography (anion-EX) on a Resource quaternary ammonium (Q) column. The protein was incubated overnight at 4°C in the presence of excess TEV protease to cleave off the His₆-tag and subsequently subjected to reverse IMAC. A final step of purification was SEC on a Superdex 200 (S200) column. The protein was prone to degradation and the degradation products can be observed from the sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the sample after SEC (Figure 2-9).

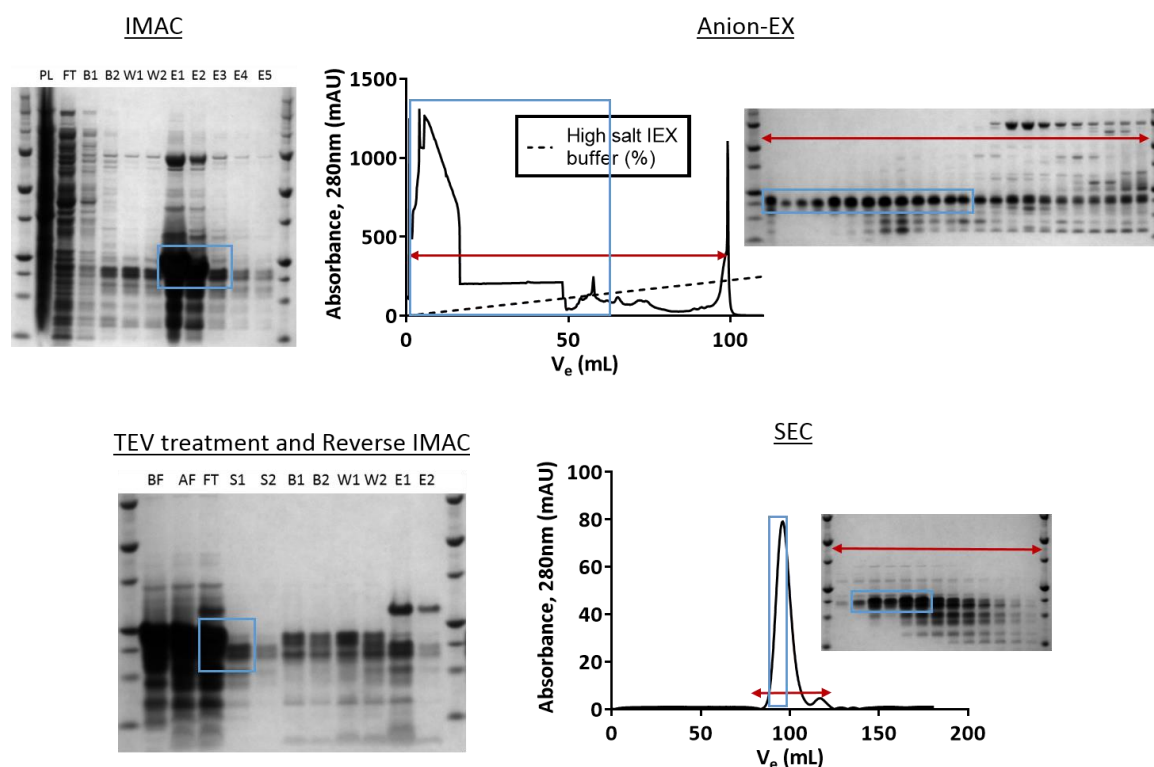


Figure 2-9. Large scale purification of the N-domain of MRPP1. The example shows SDS-PAGE analysis of the initial IMAC, chromatogram from anion-EX and SDS-PAGE analysis of peak fractions, SDS-PAGE analysis of Reverse IMAC with samples from before and after TEV protease treatment, the chromatogram from SEC and SDS-PAGE analysis of the peak. Red arrows correspond to fractions analysed on SDS-PAGE. Blue boxes indicate the pooled fractions. V_e : elution volume, PL: pellet, FT: flow through, B1 and B2: binding buffer fractions, W1 and W2: wash buffer fractions, E1-5: elution buffer fractions. BF: before TEV treatment, AF: after TEV treatment, S1 and S2: SEC buffer fractions.

The degradation profile for the N-domain was mapped by tryptic digest-MS/MS analysis of bands excised from SDS-PAGE, and by ESI-MS of the purified sample. The identified molecular masses from various peaks in ESI-MS (**Figure 2-10A**) were mapped onto the N-domain sequence (**Figure 2-10B**), and showed that the purified protein contains – at least – residues 40-165 and up to residues 40-191. The molecular mass peak of 15075.64 Da could not be mapped onto the sequence and is assumed to be a protein impurity from the purification procedure. Tryptic digest-MS/MS analysis of the N-domain identified peptides covering residues 40-181 (**Figure 2-10C**).

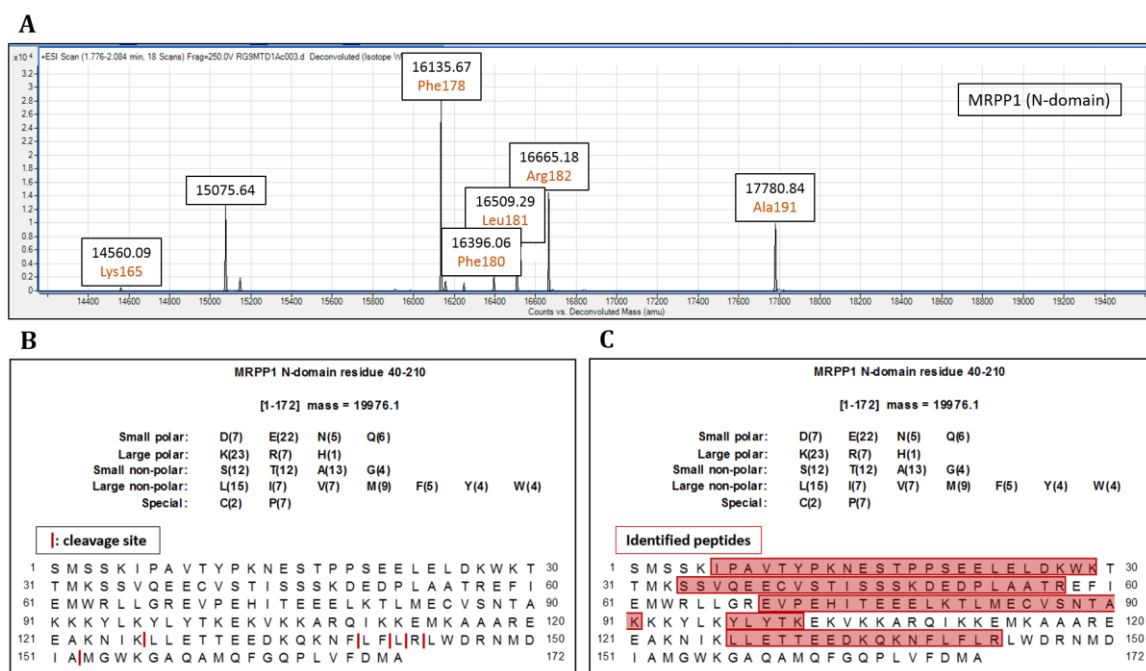


Figure 2-10. A) ESI-MS spectrum for the purified N-domain of MRPP1. B) PAWS analysis showing putative cleavage sites (red) in the N-domain sequence. Sequence numbering M2-A172 corresponds to MRPP1 M40-A210. First amino acid in the sequence is vector-encoded. C) Peptides of the N-domain of MRPP1 identified with tryptic digest-MS/MS analysis (red boxes). Sequence numbering M2-A172 corresponds to MRPP1 M40-A210. First amino acid in the sequence is vector-encoded.

The degradation of this construct occurred so quickly that no full-length protein was observed in the ESI-MS. The ESI-MS peak with the largest molecular mass maps to residues 40-191, which will be the proposed boundaries mentioned hereafter for the N-domain.

2.3.3 MT-domain

The MT-domain of MRPP1 (residues 203-380) was purified as for mature MRPP1, but with an S75 – rather than an S200 – column for the SEC step. The construct was not subjected to TEV protease treatment as it was found that the N-terminal His₆-tag could not be removed by the protease. This purification and subsequent assays were performed prior to cloning of the same construct into a vector incorporating a cleavable C-terminal His₆-tag. The purification procedure of the MT-domain is outlined in **Figure 2-11**.

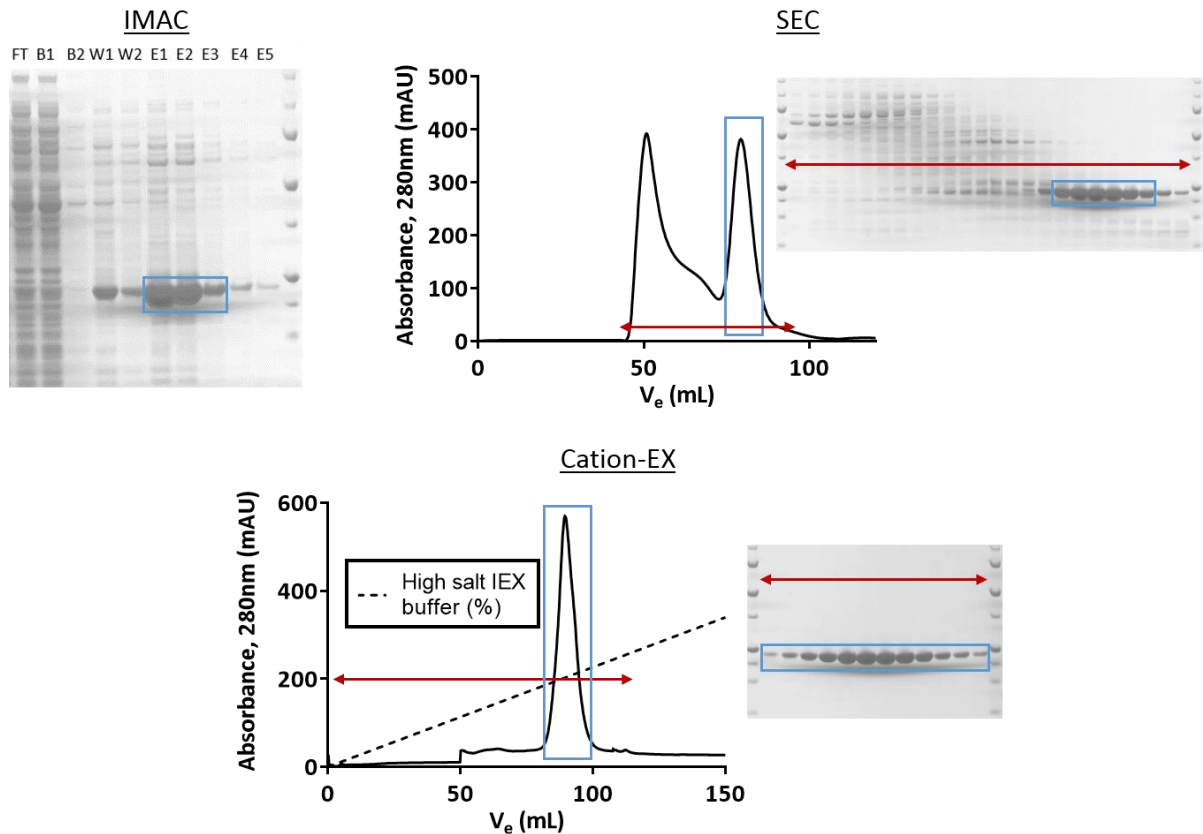


Figure 2-11. Large scale purification of the MT-domain of MRPP1. The example shows SDS-PAGE from the initial IMAC, chromatogram from SEC and SDS-PAGE analysis of peak fractions, and chromatogram from cation-IE and SDS-PAGE analysis of peak fractions. Red arrows, blue boxes and gel lanes are annotated as Figures 2-7 and 2-9.

ESI-MS analysis of the purified sample identified a molecular mass of 23607.32 Da, corresponding to the expected molecular mass for the intact protein region 210-380 (**Figure 2-12**).

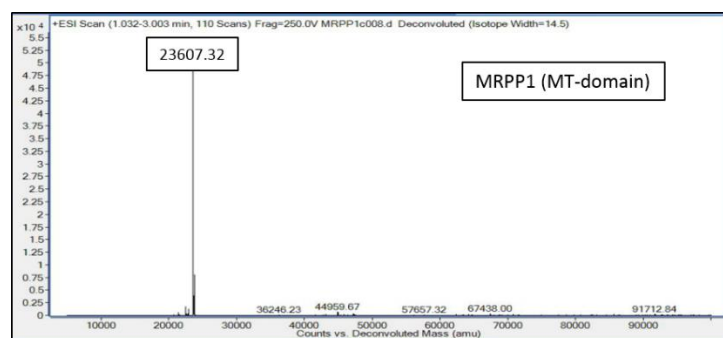


Figure 2-12. ESI-MS spectrum for the purified MT-domain of MRPP1.

2.3.4 MT+C-domain

The MT+C-domain($\Delta 194$) of MRPP1 was purified by IMAC with Ni-NTA resin, followed by SEC on an S75 column. The construct was prone to degradation from the C-terminus and thus long incubation times with TEV protease were avoided. A final cation-EX step was performed on a Resource sulphonate (S) column. The purification procedure is outlined in **Figure 2-13**.

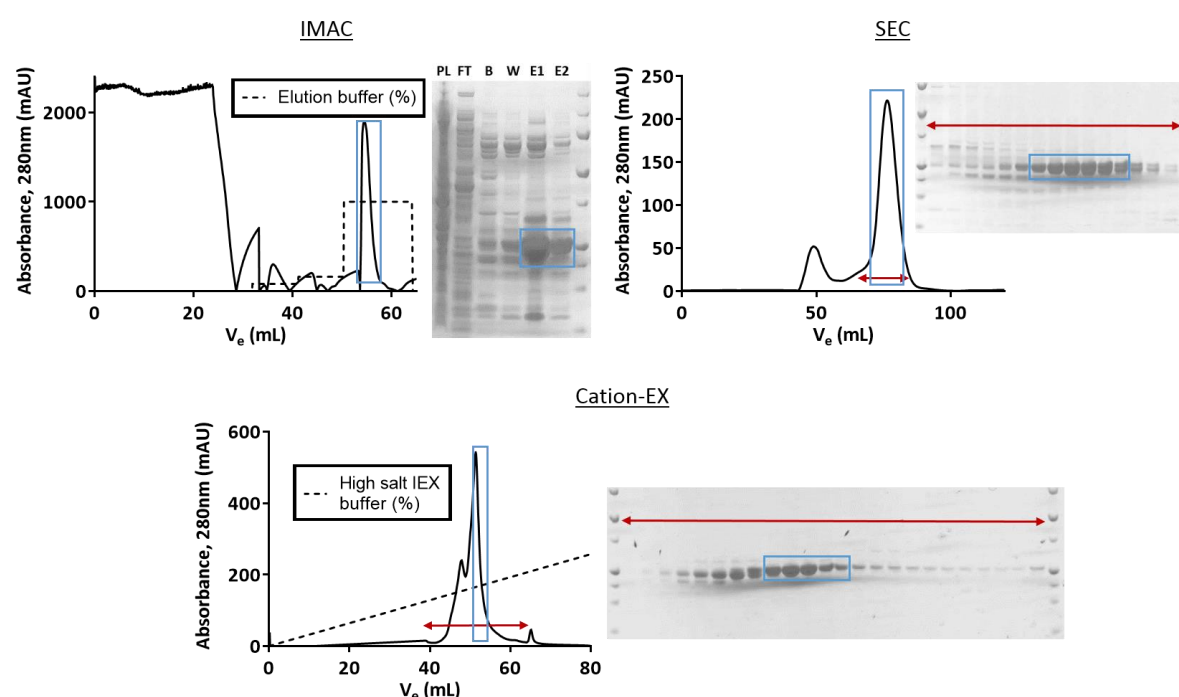


Figure 2-13. Large scale purification of the MT-domain($\Delta 194$) of MRPP1. The example shows chromatogram from the initial IMAC and SDS-PAGE analysis of peak fractions, chromatogram from SEC and SDS-PAGE analysis of peak fractions, and chromatogram from cation-EX and SDS-PAGE analysis of peak fractions. Red arrows, blue boxes and gel lanes are annotated as before.

ESI-MS analysis of the purified sample identified a molecular mass of 27109.1 Da corresponding to the expected molecular mass for residues 195-380 (**Figure 2-14A**). The lower molecular mass species observed are mapped onto the construct sequence and suggested cleavage from both the N-terminus within the vector-encoded sequence, and the C-terminus, resulting in protein fragments as short as residues 195-388 (**Figure 2-14B**). The tryptic digest-MS/MS analysis identified peptides covering residues 219-397 (**Figure 2-14C**).

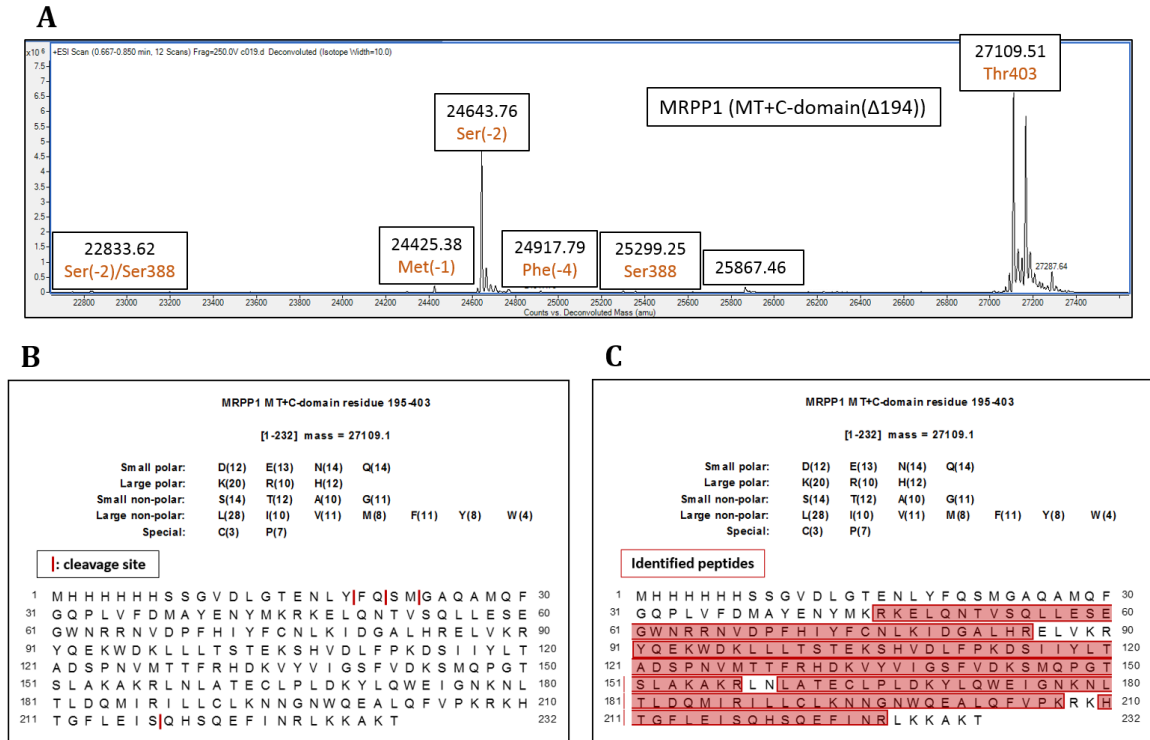


Figure 2-14. A) ESI-MS spectrum for the purified MT+C-domain($\Delta 194$) of MRPP1. B) PAWS analysis showing putative cleavage sites (red) in the sequence. Sequence numbering G24-T232 corresponds to MRPP1 G195-T403. The first 23 amino acids in the sequence are vector-encoded. C) Peptides of the MT+C-domain($\Delta 194$) of MRPP1 identified with tryptic digest analysis (red boxes). Sequence numbering G24-T232 corresponds to MRPP1 G195-T403. The first 23 amino acids in the sequence are vector-encoded.

The MT+C-domain($\Delta 202$) of MRPP1 was purified by IMAC followed by overnight TEV protease treatment at 4°C. The protein was subsequently applied to an S75 column followed by reverse IMAC. The purification procedure of the MT+C-domain($\Delta 202$) of MRPP1 is outlined in **Figure 2-15**.

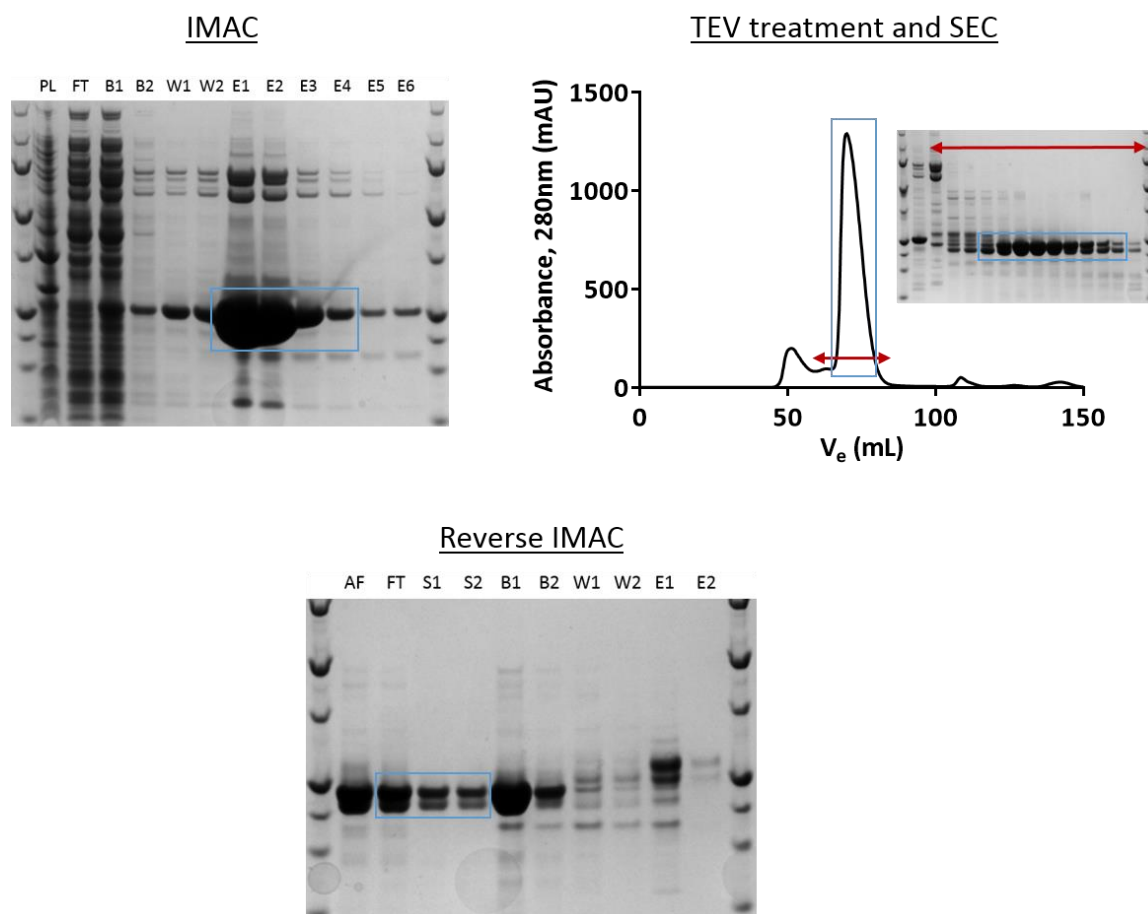


Figure 2-15. Large scale purification of the MT+C-domain($\Delta 202$) of MRPP1. The example shows SDS-PAGE analysis of the initial IMAC, chromatogram from SEC of the TEV protease treated sample and SDS-PAGE analysis of the peak fractions, and SDS-PAGE analysis of fractions from Reverse IMAC with a sample taken out after TEV protease treatment. Red arrows, blue boxes and gel lanes are annotated as before.

As described, this domain had a tendency to degrade. To map the degradation profile tryptic digest-MS/MS and ESI-MS analyses were performed. The ESI-MS spectrum showed molecular masses of various peaks (**Figure 2-16A**) which, when mapped onto the MRPP1 sequence of the MT+C-domain($\Delta 202$) with a calculated molecular mass of 24631.5 Da (**Figure 2-16B**), revealed that most of the protein is still intact, with little degradation a few residues at a time from the C-terminus down to a construct spanning residues 203-381. The part of this construct that is lost corresponds to the C-extension (residues 381-403), suggesting that this region at the C-terminus could be flexible. Tryptic digest/MS-MS analysis identified peptides covering residues 219-397 (**Figure 2-16C**), as found for the MT+C-domain($\Delta 194$) construct.

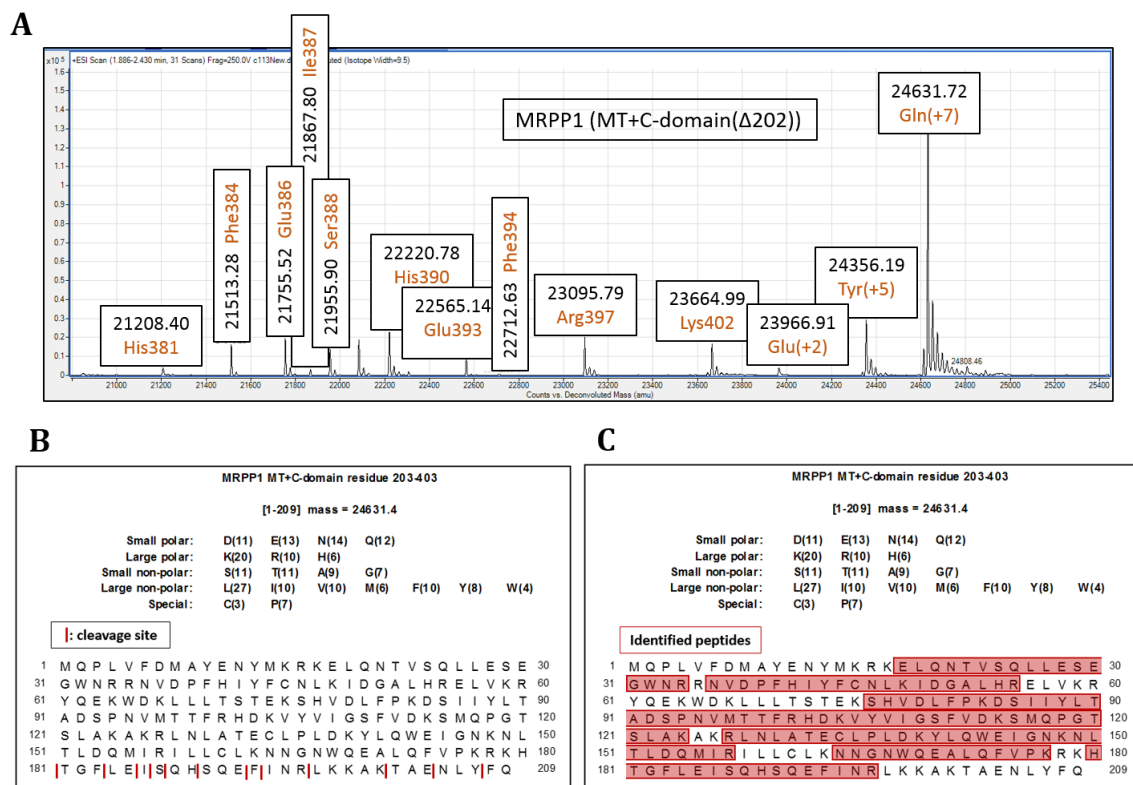


Figure 2-16. A) ESI-MS spectrum for the purified MT+C-domain($\Delta 202$) of MRPP1. B) PAWS analysis showing putative cleavage sites (red) in the sequence. Sequence numbering Q2-T202 corresponds to MRPP1 Q203-T403. The first and last 7 amino acids in the sequence are vector-encoded. C) Peptides of the MT+C-domain($\Delta 202$) of MRPP1 identified with tryptic digest analysis (red boxes). Sequence numbering Q2-T202 corresponds to MRPP1 Q203-T403. The first and last 7 amino acids in the sequence are vector-encoded.

2.4 Biochemical studies

2.4.1 Oligomeric state

The purified MRPP1 proteins were subjected to an analytical SEC assay to investigate their oligomeric states. Proteins of known MWs were first resolved on an analytical S200 column, and their $\text{LOG}^{10}(\text{MW})$ s were plotted against the obtained elution volume (V_e) for each protein, resulting in a linear function (**Figure 2-20**), from which experimental MWs could be deduced for any given protein eluting with a specific V_e . From the experimentally obtained MW, the oligomeric state could be deduced by comparison with the theoretical MW for a protein monomer.

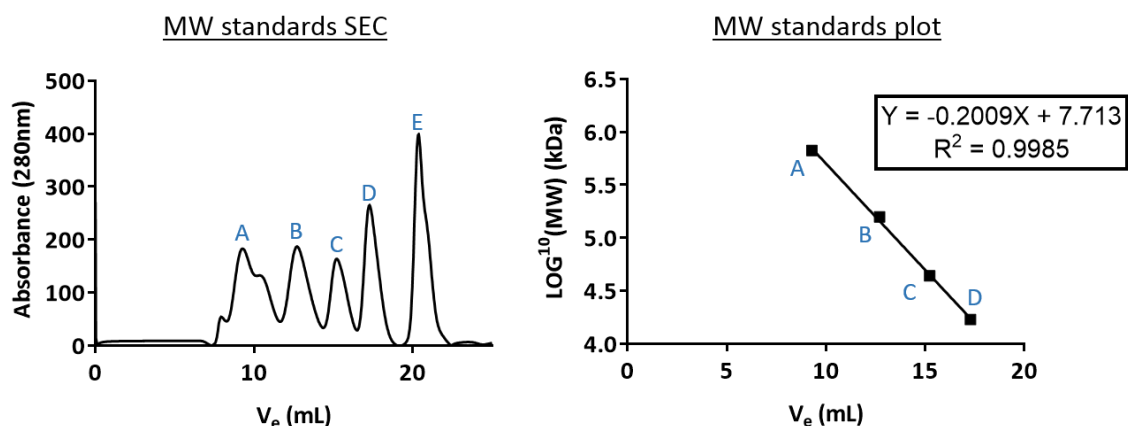


Figure 2-17. A) Chromatogram of proteins with known MWs. A: thyroglobulin (bovine) MW=670000 Da, B: γ -globulin (bovine) MW=158000 Da, C: ovalbumin (chicken) MW=44000 Da, D: myoglobin (horse) MW=17000 Da, E: vitamin B₁₂ MW=1350 Da. B) LOG^{10}MW plotted against V_e for each protein. Vitamin B₁₂ was excluded from the plot due to the poor resolution of the S200 column in the lower MW-region.

The mature MRPP1, N-domain, MT-domain and MT+C-domain(Δ 194) proteins were applied individually to the column and MWs were determined from the obtained V_e -values. The calculated MWs corresponds to that of dimeric mature MRPP1 of 109047 Da (theoretical MW of monomer = 45374.2 Da, residues 40-403), dimeric N-domain of 34258 Da (theoretical MW of monomer = 17824.6 Da, residues 40-191), monomeric MT-domain of 17172 Da (theoretical MW of monomer = 23623 Da, residues 203-380) and monomeric MT+C-domain(Δ 194) of 17712 Da (theoretical MW of monomer = 27109 Da, residues 195-403) (**Figure 2-18**).

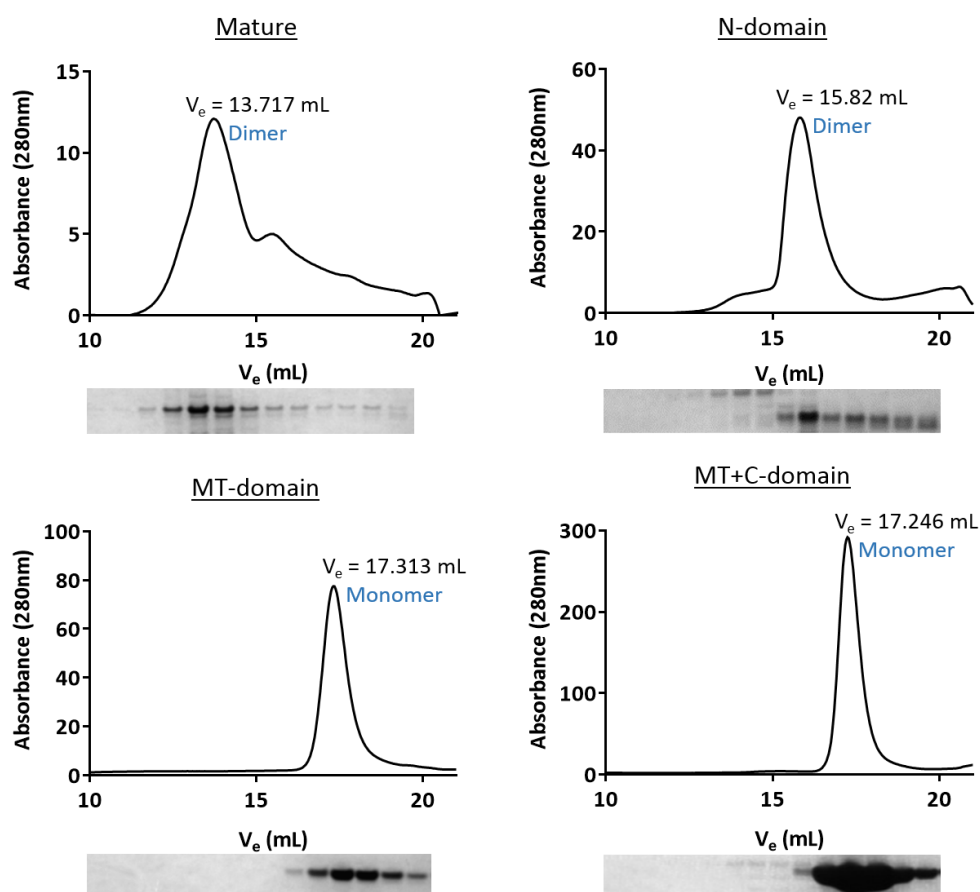


Figure 2-18. Analytical SEC chromatograms and SDS-PAGE analysis of mature, N-domain, MT-domain and MT+C-domain of MRPP1.

The mature MRPP1, N-domain and MT+C-domain($\Delta 194$) proteins were all prone to degradation during the analytical SEC assay. For the mature protein and MT+C-domain($\Delta 194$), the highest band observed on the SDS-PAGE gel corresponds to their full-length proteins determined by tryptic digest-MS/MS- and ESI-MS analysis. In the case of the N-domain, the protein in the chromatogram peak corresponds to a slightly truncated fragment, likely consisting of residues 40-191.

The SEC data support a model where the mature MRPP1 protein is a dimer in solution, and where the N-domain functions as the dimerisation domain with neither the MT-domain nor the C-extension being essential for this dimerisation.

The SEC method separates particles based on their hydrodynamic volumes and sizes rather than MWs. Accuracies in MWs of target proteins from V_e -values of standard proteins (as above) therefore depend strongly on similarities in size and shape between target and standard proteins. To circumvent these issues, the V_e -independent SEC multi-angle light scattering (MALS) technique could have been used. This method measures intensities of light scattered from a sample over a range of angles as it is eluted from a SEC column to determine the absolute MW for the protein of interest. SEC-MALS was attempted for mature MRPP1 and the MRPP1-MRPP2 complex in a phosphate/sucrose buffer, but in-system protein aggregation was encountered potentially due to the buffer composition or incompatibility of the protein with the SEC column material.

2.4.2 Ligand binding

The MT-domain of MRPP1 is predicted to contain the methyltransferase active site, and hence the binding site for the methylation cofactor S-adenosyl methionine (SAM). Ligand binding properties were investigated for the four constructs of MRPP1: mature, N-domain, MT-domain and MT+C-domain(Δ 194), in a differential scanning fluorimetry (DSF) assay. DSF measures the thermal unfolding of proteins in solution, by detecting the change in fluorescence from a probe upon its binding to hydrophobic residues exposed from the protein core, as the protein is unfolded using a stepwise gradient of increasing temperature. The probe fluorescence intensity can be plotted as a function of increasing temperature generating a sigmoidal curve where the midpoint corresponds to the melting temperature (T_m). If the protein is presented to a ligand before thermal unfolding, the protein-ligand interaction could stabilise the protein fold and cause a shift in the protein T_m (ΔT_m) [150].

The T_m was measured for each protein ‘as-purified’ and after addition of SAM. **Figure 2-19** shows that addition of SAM results in a T_m shift for the mature protein

($\Delta T_m = 6.04 \pm 0.09^\circ\text{C}$) and the MT-domain both with and without the C-extension ($\Delta T_{m_{MT+C}} = 9.75 \pm 0.48^\circ\text{C}$ and $\Delta T_{m_{MT}} = 1.34 \pm 0.26^\circ\text{C}$), indicative of SAM binding. No shift was observed for the N-domain suggesting SAM does not bind to this domain. These data support a model where the SAM binding-site resides within the MT-domain of MRPP1, and indicates that both the N-domain and the small C-extension are dispensable for SAM binding. The small C-extension does, however, contribute significantly to general protein stabilisation after SAM binding, as reflected in the $8.97 \pm 0.33^\circ\text{C}$ increase between the SAM bound proteins of MT-domain and MT+C-domain, but more investigations are needed to elucidate the role of the small C-extension in MRPP1.

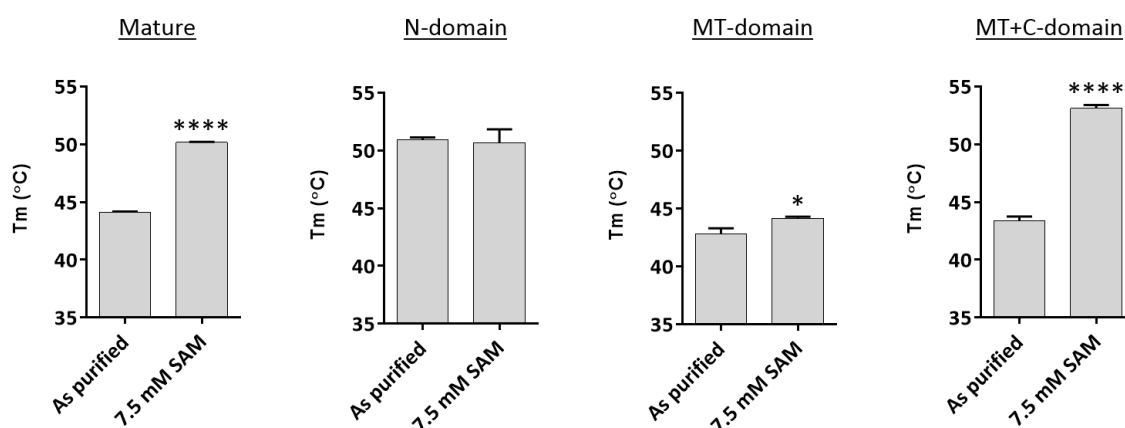


Figure 2-19. DSF assay showing T_m response upon addition of SAM for mature MRPP1, N-domain, MT-domain and MT+C-domain($\Delta 194$). For this experiment $n = 2-3$. * = $P < 0.05$, **** = $P < 0.0001$.

2.5 Structural studies

To further investigate the SAM binding to the MT-domain, and to obtain detailed structural information about the catalytic site of MRPP1, a crystal structure of the MT-domain was determined.

2.5.1 Crystallisation of the MT-domain of MRPP1

The soluble constructs of the MT- and MT+C-domain used in crystallisation experiments are outlined in **Figure 2-20**.

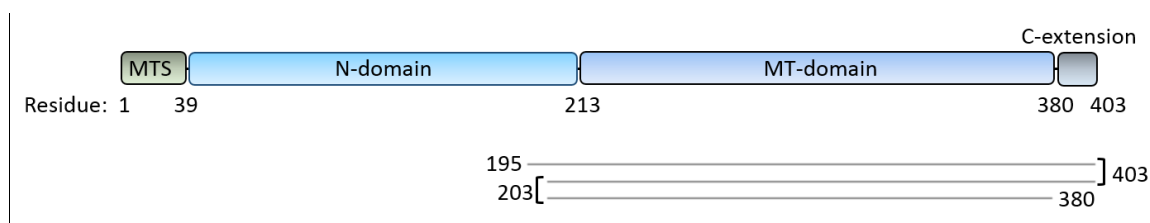


Figure 2-20. Cloned MRPP1 constructs covering the MT- and MT+C-domain, used in crystallisation trials.

2.5.1.1 *MT-domain*

The purified MT-domain protein was concentrated to 6 mg/mL and utilised to set up crystallisation plates of two random matrix screens: HIN and JCSG, at 4°C and 20°C, with different protein-to-precipitant ratios (1:2, 1:1, 2:1). No crystals appeared in these plates and most drops appeared clear even after two months of incubation. To produce fewer clear drops, protein concentration was increased to 15 mg/mL and to further facilitate crystallisation, protein was incubated with a 1.5-fold molar excess of the natural ligand SAM prior to setting up crystallisation plates of three random matrix screens: HCS, HIN and JCSG, at 4°C and 20°C, with different protein-to-precipitant ratios (1:2, 1:1, 2:1). No crystals appeared in these plates, but significant precipitation was observed after one day of incubation. Final crystallisation attempts were performed using an intermediate protein concentration of 10 mg/mL (again supplemented with SAM) and an expanded set of random matrix screens: MIDAS, LFS, HCS, HIN and JCSG (composition: **Section 7.5.2**). Crystalline precipitation formed in 0.1 M sodium succinate (NaSuc) and 15% polyethylene glycol (PEG) 3350 at 20°C and at a protein-to-precipitant ratio of 1:1. A follow-up plate was set up (**Figure 2-21**) surveying the immediate chemical environment around the crystalline precipitate condition. This screen failed to produce protein crystals and further crystallisation attempts were performed with other constructs.

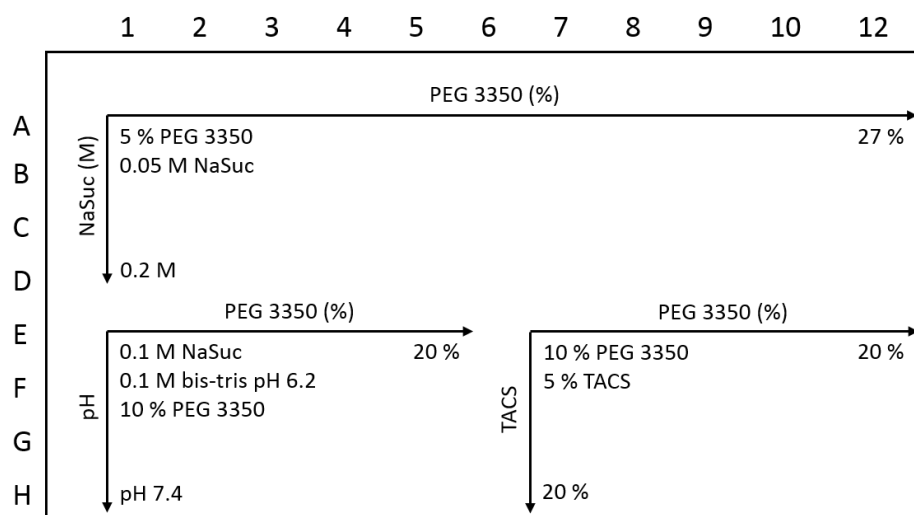


Figure 2-21. Follow-up screen surveying around the initial crystalline precipitate condition of 0.1 M NaSuc and 15% PEG3350. TACS: tacsimate (mixture of 1.8305 M malonic acid, 0.25 M ammonium citrate tribasic, 0.12 M succinic acid, 0.3 M *DL*-Malic acid, 0.4 M sodium acetate trihydrate, 0.5 M sodium formate, and 0.16 M ammonium tartrate dibasic).

2.5.1.2 MT+C-domain

The purified MT+C-domain(Δ 194) protein was concentrated to 19 mg/mL and incubated with 10-fold molar excess of SAM prior to setting up crystallisation plates in two random matrix screens: HCS and JCSG at 20°C. This yielded tiny crystals within 3 days in multiple conditions (**Figure 2-22**), but as larger crystals were simultaneously obtained with the MT+C-domain(Δ 202) these crystals were not further optimised.

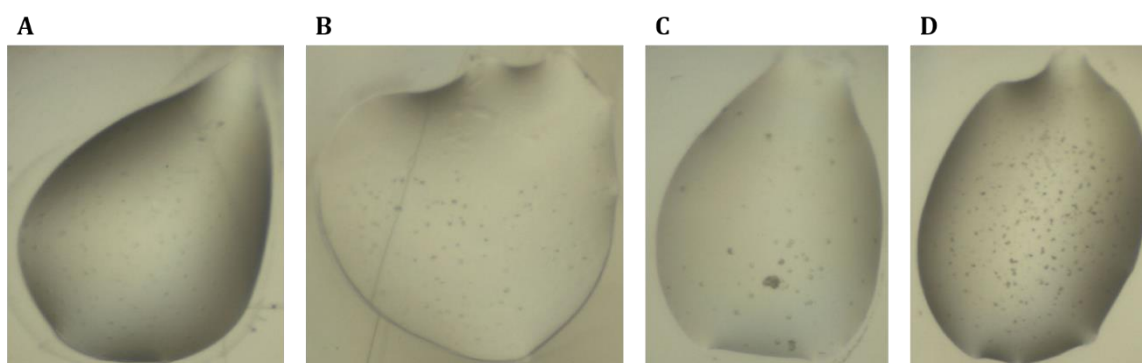


Figure 2-22. Crystals of the MT+C-domain(Δ 194) of MRPP1. A) 0.5 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M citrate pH 5.6, 1 M Li_2SO_4 . B) 0.2 M Na(ac), 0.1 M cacodylate pH 6.5, 30% PEG8000. C) 0.1 M bicine pH 9.0, 1 M MgCl_2 . D) 50% PEG200, 0.2 M MgCl_2 , 0.1 M cocadylate pH 6.5. Na(ac): sodium acetate.

The MT+C-domain(Δ 202) protein was concentrated to 11 mg/mL and used to set up crystallisation plates of two random matrix screens: HCS and JCSG, in two temperatures. These trials yielded small crystals within 3 days, in many conditions containing PEGs and

often salts and buffers with a broad pH-range of 4.2 to 8.5. Two different crystal forms (needles and diamonds) were obtained in various crystallisation conditions as exemplified in **Figure 2-23A** and **B**, respectively.

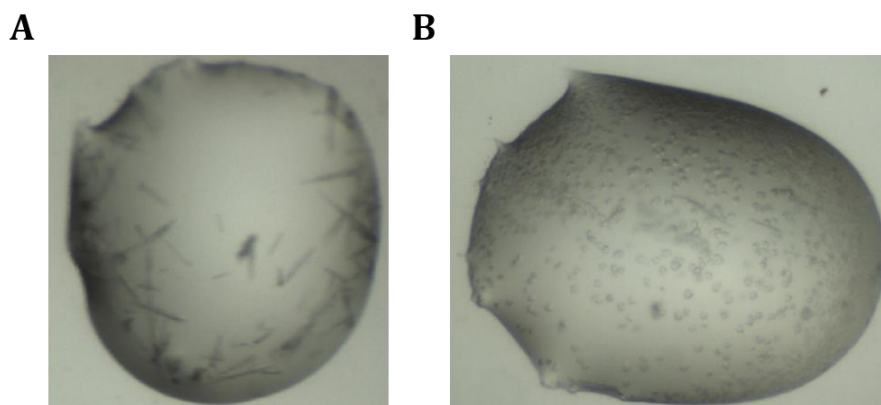


Figure 2-23. Crystals of the MT+C-domain(Δ 202) of MRPP1. A) Needle shaped crystals grown in 16% PEG8000, 20% glycerol, 0.16 M calcium acetate, 0.1 M cacodylate pH 6.5. B) Diamond shaped crystals grown in 25% PEG1000, 20% glycerol.

Follow-up screens were designed for seven hits and crystallisation plates were set up. This yielded larger diamond shaped crystals (**Figure 2-24**) grown in 4°C with precipitant conditions around the initial conditions of 25% PEG1000, 20% glycerol (**Figure 2-23B**).

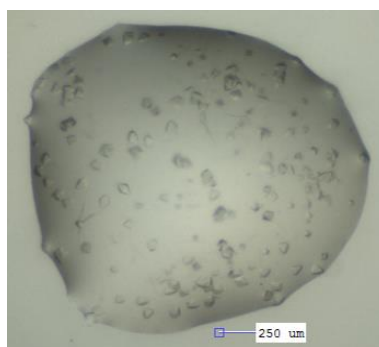


Figure 2-24. Diamond shaped crystals of the MT+C-domain(Δ 202) of MRPP1 from the follow-up screen grown in 15% PEG1000, 24% glycerol.

Both needle- and diamond-shaped crystals were observed in this screen with different protein-to-precipitant ratios. Higher protein concentrations (2:1 ratio) would lead to diamond shaped crystals and lower protein concentrations (1:2 ratio) would lead to needle formation. Furthermore, the two forms had a tendency to appear in the same drop at different

times (**Figure 2-24**), first diamonds and later needles, further suggesting a dependence on protein concentration.

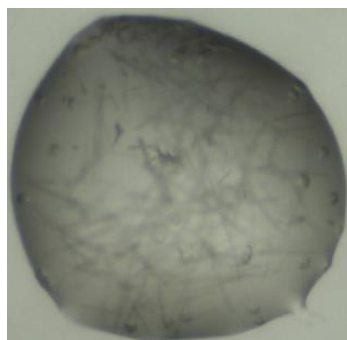


Figure 2-25. Needle- and diamond-shaped crystals of the MT+C-domain(Δ 202) of MRPP1 grown in the same condition of 21% PEG1000 and 20% glycerol.

These crystals were cryo-protected in mother liquor supplemented with 25% propylene glycol, and tested for diffraction at the Diamond Light Source beamline I24. The diamond shaped crystals showed diffraction to 3.5 Å resolution. The needle crystals showed no diffraction and all optimisation was therefore focussed on the diamond shaped crystals.

An analysis of crystal formation across the follow-up conditions (**Figure 2-26**) revealed that glycerol and PEG concentrations also affected the preference of one crystal form over another, as glycerol concentrations < 22% resulted in needle formation. In addition, very low glycerol concentrations in combination with high PEG concentrations resulted in precipitation, and PEG concentrations > 30% – or PEG concentrations > 23% when combined with high glycerol concentrations – resulted in clear drops.

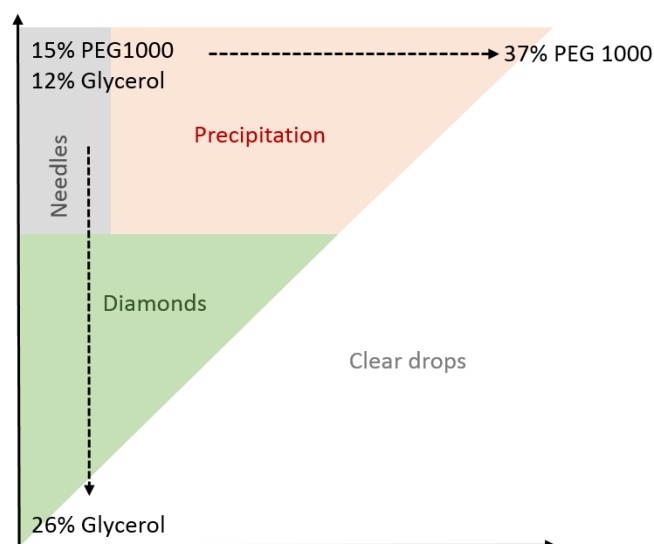


Figure 2-26. Analysis of the follow-up screen based on the initial crystallisation hit with precipitant condition 25% PEG1000 and 20% glycerol.

Based on this analysis, a second follow-up screen was produced focussing on the conditions yielding diamond shaped crystals (**green, Figure 2-26**). Using this new screen, larger crystals grew from fewer nucleation points with increasing size towards the area of lower PEG (down to 11%) and higher glycerol (up to 28%) concentrations (**Figure 2-27**).

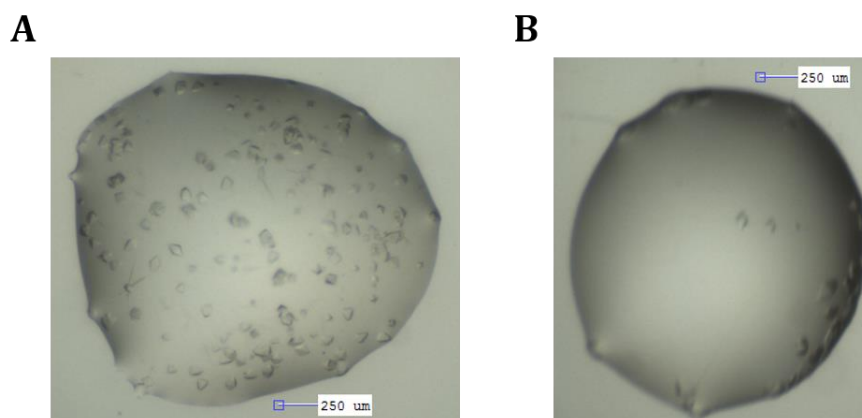


Figure 2-27. Crystals of the MT+C-domain(Δ 202) from (A) the first follow-up screen grown in 15% PEG1000, 24% glycerol and (B) the second follow-up screen grown in 12% PEG1000, 28% glycerol.

A dataset was collected at the Diamond Light Source beamline IO3 from a crystal diffracting to 2.79 Å resolution. To obtain better diffracting crystals, a third screen was made based on the condition from the second screen (12% PEG1000, 28% glycerol), further supplemented with a commercially available set of additives in the precipitation conditions. This screen yielded larger crystals in conditions with additives from many different chemical

groups, *i.e.* multi- and monovalent salts, osmolytes, volatile organic compounds, amino acids, dissociating agents, polyamines, polymers, carbohydrates and linkers. A dataset was collected at the Diamond Light Source beamline IO4 for a crystal grown with PEG3350 as additive diffracting to 2.2 Å resolution.

2.5.2 Structure determination

The collected data indicated that the crystal belonged to the tetragonal space group $P4_3$ or $P4_1$. $P4_3$ was chosen as the correct space group during molecular replacement (MR), where phases were attempted obtained from a cytosolic, human Trm10 protein structure (TRMT10A) with 26% sequence identity over residues 205-380 of the MRPP1 MT-domain [143]. The MR attempt failed, presumably due to the low sequence identity. Instead, phases were determined experimentally by the single-wavelength anomalous dispersion (SAD) technique, from a heavy atom (HA)-derivatised crystal. Crystals were grown in the previously identified best condition (12% PEG1000, 28% glycerol, 10% PEG3350) and soaked with 10 mM thiomersal ($\text{C}_9\text{H}_9\text{HgNaO}_2\text{S}$) for a duration between 1 sec to 1 h. A dataset was collected at the Diamond Light Source beamline IO4 for a crystal soaked for 5 minutes diffracting to 2.49 Å resolution. The collected data were isomorphous to the native crystal and a tetragonal space group, $P4_3$ or $P4_1$, was indicated. Analysis of the data processing statistics (anomalous correlation coefficient in the XDS scaling file) revealed a clear anomalous signal, therefore allowing six HA sites to be located using SHELXD. The hand discrimination ($P4_3$ or $P4_1$ space group), phasing procedure and automated model building was performed by SHELXE. $P4_3$ was chosen as the correct space group based on the map contrast (**Figure 2-28**) compared between the original ($P4_3$) hand, CC = 30.26% (cycle 3), and the inverted ($P4_1$) hand, CC = 13.36% (cycle 3).

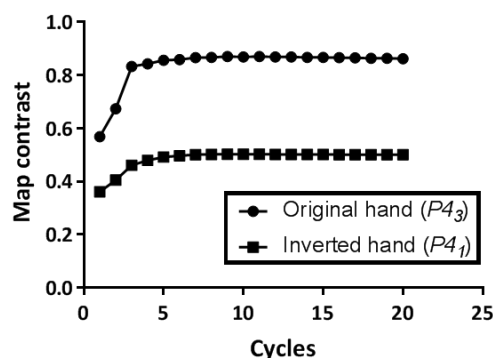


Figure 2-28. Contrast map vs. density modification cycle performed by SHELXE.

This procedure generated an initial poly-alanine model covering the entire protein chain. The phases from the HA-dataset were then combined with the 2.2 Å native dataset, and the model was completed with side-chains assigned to sequence in COOT [151]. Subsequently, better diffracting crystals were grown in the same conditions as before, and data collected to 1.96 Å resolution were used for the final refinement of the structure. Due to the nature of the tetragonal space group an indexing ambiguity exists (two possibilities) and by chance the final high resolution dataset was indexed differently [152]. To enforce consistent indexing, the dataset should have been re-indexed with POINTLESS using the reflections from the reference and the original set of indices. Subsequently, rigid body refinement should have been performed. However, in the interest of time, MR was used to place the model correctly. The data collection and final refinement statistics are given in **Table 2-1**.

For all datasets, the resolution cut-off was chosen where $CC\frac{1}{2}$ was above 0.5 with $I/\sigma(I)$ remaining above 1.0. Previously, R_{merge} was taken as the main measure of resolution cut-off, but this multiplicity-dependent factor was shown to inaccurately estimate data quality, and $CC\frac{1}{2}$ was suggested as a better value for determining resolution (**Section 7.5.5**). The R_{merge} value for the HA dataset with Friedel pairs treated as the same (I+ and I-) or as separate (I+/I-) reflections is 0.214 and 0.202, respectively, from AIMLESS [153]. For the

HA dataset in **Table 2-1**, the Friedel pairs are treated as separate reflections as Friedel's law is broken by the anomalous scattering from the HA [154].

	HA	Native	Native
Unit Cell			
Beamline	Diamond IO4	Diamond IO4	Diamond IO4-I
Space group	$P4_3$	$P4_3$	$P4_3$
a, b, c	81.72, 81.72, 146.60	81.44, 81.44, 146.59	82.64, 82.64, 148.62
α, β, γ	90, 90, 90	90, 90, 90	90, 90, 90
Molecule/ASU	3	3	3
Wavelength (Å)	0.97949	0.97949	0.92820
Resolution range (Å)	2.49-53.77 (2.55 – 2.49)	81.72-2.22 (2.299- 2.22)	58.44-1.96 (2.03- 1.96)
No. of reflections	911142 (67837)	354295 (24920)	494321 (33562)
No. of unique reflections	33167 (2450)	51121 (3773)	71314 (5279)
Multiplicity	27.5 (27.7)	6.9 (6.6)	6.9 (6.4)
Completeness (%)	98.7 (98.1)	99.97 (99.91)	99.98 (99.97)
I/σ(I)	12.40 (1.10)	11.75 (1.94)	16.95 (1.48)
CC_{1/2}	0.998 (0.550)	0.998 (0.659)	0.999 (0.638)
Wilson-B factor (Å²)	51.89	44.14	42.35
R_{merge}(%)	0.202 (3.671)	0.093 (1.312)	0.052 (1.267)
R_{meas}(%)	0.210 (3.942)	0.111 (1.587)	0.061 (1.516)
Anomalous completeness	97.9 (95.3)	99.5 (99.3)	99.4 (98.7)
Anomalous multiplicity	13.4 (13.7)	3.3 (3.2)	3.3 (3.0)
Anomalous correlation	0.520 (-0.024)	0.1 (0.001)	0.046 (0.001)
Refinement			
R_{work} (%)	0.1946 (0.3300)		
R_{free} (%)	0.2173 (0.3675)		
RMSD_{bond} (Å)	0.008		
RMSD_{angle} (°)	1.23		
No. of non-hydrogen atoms	4772		
macromolecules	4497		
Protein residues	534		
Ligands	58		
Water	217		
Ramachandran favoured	98		
Ramachandran outliers	0		
Clash score	3.63		
Average B-factor (Å²) of	54.20		
Macromolecules	54.20		
Ligands	71.30		
Solvent	50.60		

Table 2-1. Data collection and refinement statistics for the structure of the MT+C-domain(Δ 202) of MRPP1. Data for highest resolution shell are shown in parenthesis.

2.5.3 Overall structure

The MT+C-domain crystal contains three MRPP1 monomers in the asymmetric unit (ASU) each bound to a molecule of SAM (**Figure 2-29**). The three MRPP1 monomers are not related by simple rotation symmetry. They align with an RMSD between 0.204-0.238 Å for C $^{\alpha}$ atoms, with the main differences attributed to parts of the flexible loop at residues Cys335-Thr352 and segments involved in crystal contacts at residues Tyr214-Leu271 (A:B, A:C, B:C), Lys216-Asp239 (B:C), Pro319-Tyr244-Thr321 (A:B, A:C, B:C), Arg328-Trp266 (B:C) or -Ser273 (B:C) or -Leu267 (A:B, A:C), or -Leu269 (A:C), Gln318-Met202 (A:B, B:C) or -His242 (A:C) showing that A:B $\neq$ A:C $\neq$ B:C. The packing displayed a complex formation significance score (CSS) of 0.0 for all three interfaces computed by the PISA server [155], altogether indicating that the ASU does not contain a physiologically significant multimer. This is in line with the finding that the MT+C-domain protein exists as a monomer in solution.

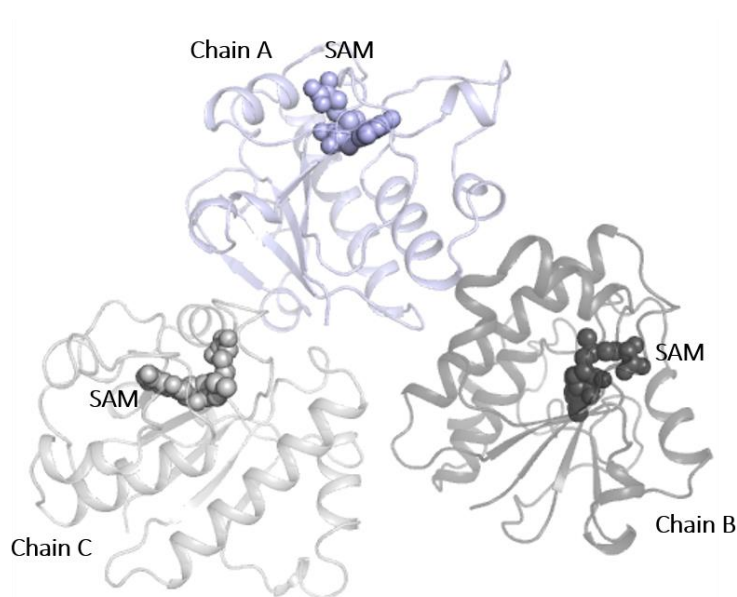


Figure 2-29. Asymmetric unit for MRPP1 MT Domain crystal. Each monomer holds one SAM methyl-donor (spheres).

In the final refined model of the SAM bound MT+C-domain, no electron density was visible for the C-extension (C-terminal 23 amino acids), which was therefore not modelled. An MS/MS analysis of the crystallised protein sample indicated partial truncation

of the C-extension in line with previous findings (**Section 2.3.4**). The missing electron density could therefore reflect proteolytic removal of the residues before and/or during crystallisation. Alternatively, if the crystallised protein *does* contain the C-extension, the missing electron density likely indicates flexibility of this region in the crystal packing. A loop region (residues 335-352) also had less visible electron density compared to the rest of the structure and was therefore built as poly-alanine.

2.5.4 Analysis of the active site

The MT-domain of MRPP1 contains the characteristic SPOUT (α/β) knot fold [53], [54], [156], consisting of a central 6-stranded β -sheet (β 1- β 6) flanked on both sides by α -helices (α 1- α 6) (**Figure 2-30A**). The SAM molecule bound in each MRPP1 monomer is clearly visible in the electron density map (Fo-Fc) (**Figure 2-30B**, contoured at 2.0σ). The SAM molecule is accommodated in a bent conformation as generally observed for SPOUT proteins [139], [156]. It is bound via a deep trefoil knot (**Figure 2-30**) [156], [157] that has been proposed to be associated with cofactor shuffling during catalysis by allowing the protein to undergo constant isomerisation [59].

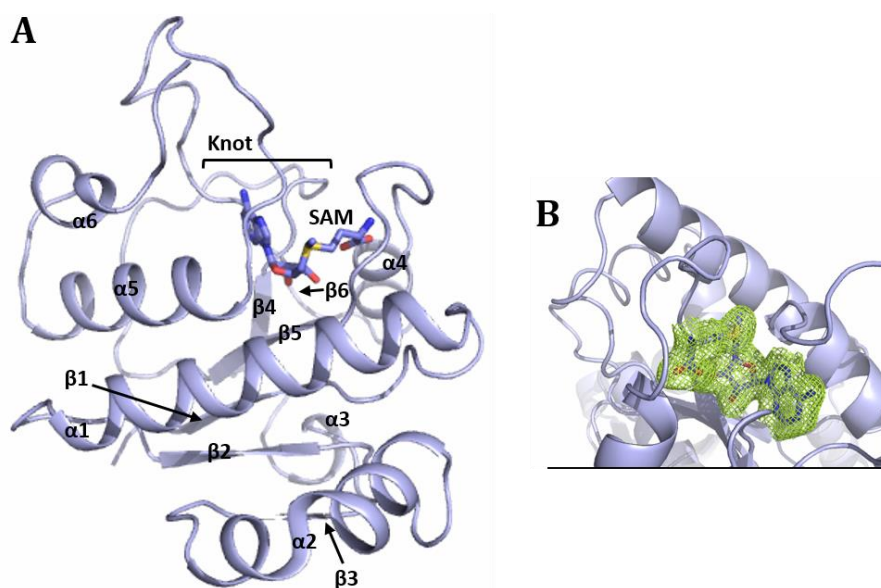


Figure 2-30. Structure of a monomer of the MRPP2 MT-domain with the 6-stranded β -sheet (β 1- β 6) and 6 flanking α -helices (α 1- α 6) indicated along with the ligand SAM (sticks) bound in the active site. The trefoil knot is indicated. The SAM molecule is clearly visible in the electron density map (Fo-Fc contoured at 2.0σ).

The closest structural homologues of MRPP1, based on DALI search, are Trm10 members including human TRMT10A (PDB 4FMW, Z-score 26.4, C^α-RMSD 1.5 Å, 26% sequence identity), *S. pombe* Trm10 (4JWF, 25.4, 1.7 Å, 28%), *S. Cerevisiae* Trm10 (4JWJ, 23.6, 1.9 Å, 23%) and *S. acidocaldarius* Trm10 (5A7T, 14.0, 2.5 Å, 16%). Sequence alignment of MRPP1 with these three Trm10 proteins and other known Trm10 family members (**Section 8.3**), revealed four conserved motifs within β4-β5 (motif I, residues 288-293), β5-α4 (motif II, residues 307-323) and β6-α5 (motif III, residues 335-338 and motif IV, residues 350-360) of the MT-domain (**Section 8.3, red boxes**) [49]. These motifs are highlighted in the sequence of full-length MRPP1 (residue 1-403) in **Figure 2-31**.

1	MAAFLKMSVSVNFFRPFTRFLVPFTLHRKRNNLTILQRYMSSKIPAVTYPKNESTPPSEELELDKWKTTM	70
71	KSSVQEECVSTISSKDEDEPLAATREFIEMWRLLGREVPEHITEEELKTLMECVSNTAKKKYLKYLTKKEK	141
142	VKKARQIKKEMKAAAREEAKNIKLLETTEEDKQKNFLRLWDRNMDIAMGWKGAQAMQFGQPLVFDM	209
210	AYENYMKRKELQNTVSQLESEGWNRRNVDPFHIYFCNLKIDGALHRELVKRYQEKWDKLLTSTEKSH	278
279	VDLFPKDSIIYLTADSPNVMTTFRHDKVYVIGSFVDKSMQPGTSLAKAKRLNLATECLPLDKYLQWEIGNK	349
350	NLTLDQMIRILLCLKNNGNWQEALQFVPKRKHTGFLEISQHSQEFINRLKKAKT	403

Figure 2-31. The sequence for MRPP1 outlining the four identified motifs. The beginning of the crystal structure is indicated.

Structurally, the four conserved motifs (**Figure 2-32A, orange**) overlap with parts of MRPP1 that make up the SAM binding pocket. The adenosine ring of SAM binds in a hydrophobic pocket formed by Leu290, Ala292, Ile309, Gly310, Cys335, Leu336, Leu338, Leu351 and Met356 (**Figure 2-32B, pink**). Additionally, hydrogen bonds are formed (**Figure 2-32B, inset**) between the 2'- and 3'-OH on the ribose moiety and backbone amide of Gly310, between the N-7 and N-1 of the adenine ring and backbone amide of Leu351 and Leu336, respectively, and between the amine on C-6 of the adenine and backbone carbonyl of Lys349. The methionyl moiety of the SAM ligand binds in a pocket formed by Thr291,

Ser322, Phe312, Gln318, Asn350, Asp314 and Thr321 (**Figure 2-32B**, inset) with hydrogen bonds between the COO⁻-group of the methionyl moiety of SAM and backbone amide and carbonyl of Ser322 and Thr321, and between the methionyl N-atom and backbone carbonyl of Gln318.

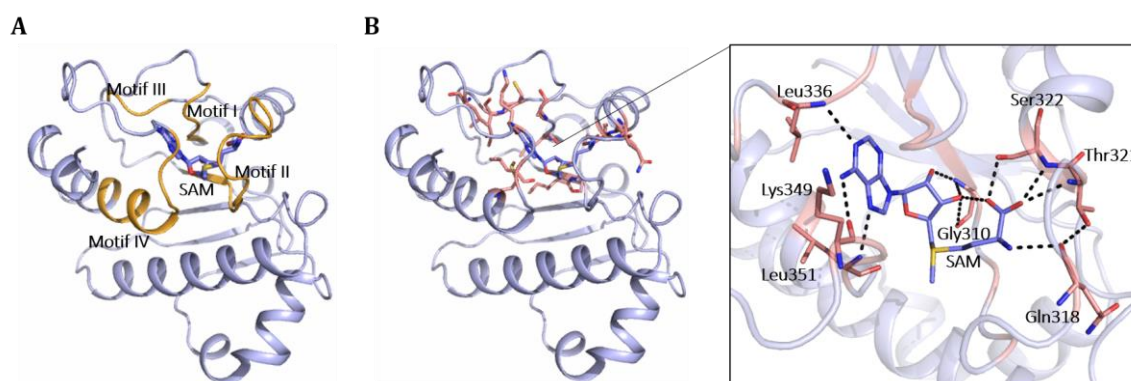


Figure 2-32. A) Conserved motifs I-IV (orange) of Trm10 family members with SAM shown in blue sticks. B) Residues forming the SAM binding pocket (pink sticks) shown with SAM (blue sticks). Inset, Residues involved in hydrogen-bond formation (pink sticks) to the SAM ligand (blue sticks).

In summary, the crystal structure of the MT-domain provides an overview of the catalytic site of MRPP1, and enables an analysis of potential important residues for m¹G9 and m¹A9 activity investigated in **Chapter 3** on the MTase complex MRPP1-MRPP2.

2.5.5 Crystallisation attempts with the N-domain of MRPP1

The N-domain of MRPP1 was found to be responsible for protein dimerisation. To determine the fold of this domain, and the novel dimer interaction surface, crystallisation experiments were performed. At the time of carrying out these experiments, mature MRPP1 had not yet been obtained in amounts and quality satisfactory for crystallography, and thus trials were performed with constructs of the N-domain only (**Figure 2-20**).

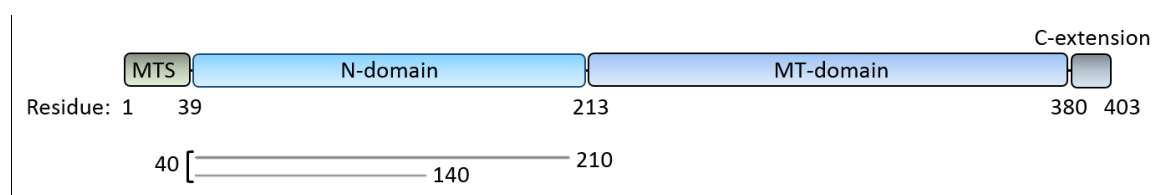


Figure 2-33. Cloned MRPP1 constructs used in crystallisation trails.

2.5.5.1 N-domain (residues 40-210)

The purified, untagged N-domain protein was concentrated to 12.5 mg/mL and used to set up crystallisation plates of two random matrix screens: JCSG and LFS, at 4°C, with different protein-to-precipitant ratios (1:2, 1:1, 2:1). After 2 months of incubation, no crystals were observed in any drops. Precipitation was observed in most wells as regular precipitation or lumpy precipitation in conditions containing 2-methyl-2,4-pentandiol (MPD) [158].

2.5.5.2 N-domain (residues 40-140)

The construct covering a smaller part of the N-domain was expressed and purified by Elzbieta Rembeza (Metabolic and Rare Diseases group, SGC). ESI-MS identified an intact protein consistent with the lack of degradation products in SDS-PAGE analysis during purification. The purified protein was concentrated to 17 mg/mL and used to set up crystallisation plates as before of two random matrix plates: HCS and JCSG in two temperatures and different protein-to-precipitant ratios. After 2 months of incubation, no crystals were observed, but most drops showed precipitation, indicating that a lower protein concentration could be used in future crystallisation trials.

2.5.5.3 Structure predictions for the MRPP1 N-domain

Like MRPP1, other Trm10 proteins contain unique domains or sequences in the N-terminal. Of the structurally characterised Trm10 members, only the archaeal Trm10 is solved as the full-length protein including this additional region (**Figure 2-34B**). The N-domain of Trm10 from archaea consists of both α -helices and β -strands [159], and circular dichroism (CD) on the N-domain of Trm10 proteins from yeast predicted these to consist *mainly* of β -strands [49]. PSIPRED structure predictions (**Figure 2-34A**) suggested that the N-domain in MRPP1 contains *solely* β -strands, suggesting a different structural arrangement compared to other Trm10 members, as expected from the dimerisation of this

domain only in MRPP1. Homology modelling was attempted for the N-domain using MODELLER [160], but due to the low structural similarity of this domain to known folds, no satisfactory model could be obtained.

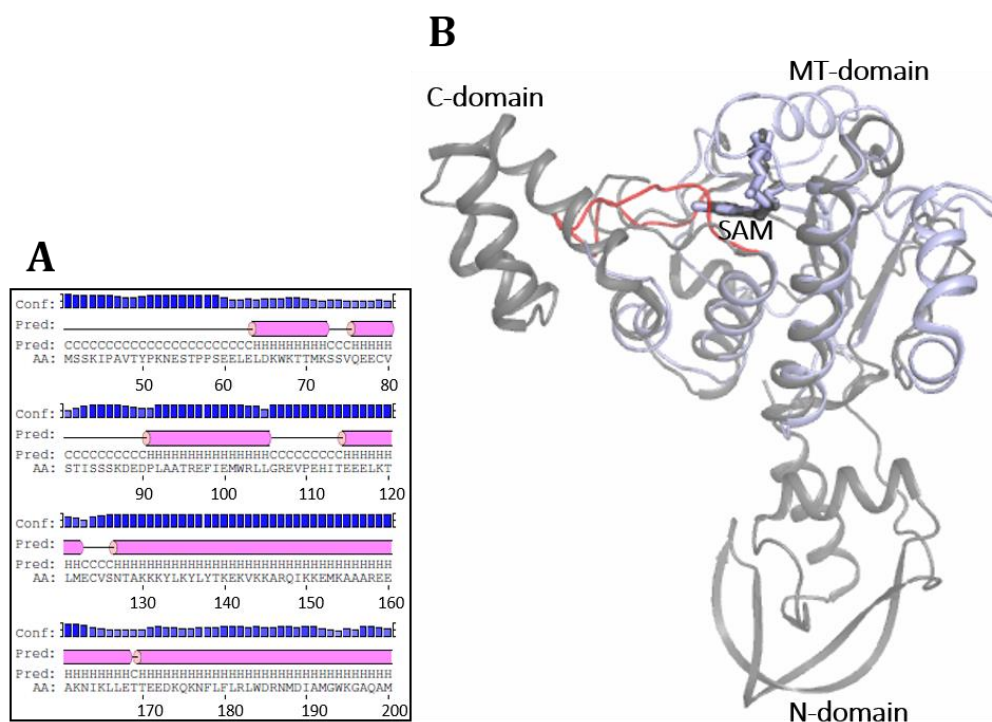


Figure 2-34. A) PSIPRED secondary structure predictions of the N-domain of MRPP1. B) Alignment of the MT-domain of MRPP1 (blue) bound to SAM (sticks) and full-length Trm10 from archaea (*S. acidocaldarius*) (black) bound to SAH (sticks). The loop shown in red is the disordered loop (residue 335-352) in the MRPP1 structure.

In addition to the N-domain, MRPP1 and other Trm10 members contain unique domains or sequences in the C-terminal, and the C-domain of Trm10 from archaea (65 residues) is shown in the structure (**Figure 2-34**). This domain shows no sequence similarity with the small C-extension of MRPP1, not visible in the structure presented here, suggesting that these regions are unlikely to be structurally homologous.

After all crystallisation attempts, mature MRPP1 became available at quantities allowing generation of a low-resolution model of this construct using alternative X-ray scattering methods.

2.5.6 SAXS model of mature MRPP1

Small angle X-ray scattering (SAXS) measures the scattering from macromolecules in solution (**Section 7.6**). When initially analysing the SAXS data in a model-independent fashion, basic parameters can be determined, e.g. the radius of gyration, R_g , which is a measure of the protein radius relative to its centre-of-mass. Another evaluation of protein shape is the Kratky plot, which differentiates qualitatively between folded and disordered particles and number of domains in a protein. The maximal dimension of the particle, $D_{\max}$, is determined from the shape of a function of the distribution of inter-atomic distances, or $P(r)$ distribution, in which $D_{\max}$ is the upper limit of the integral. After evaluation of the SAXS data, a model can be reconstructed by *ab initio* modelling.

Purified mature MRPP1 protein was applied for SAXS analysis in line with an HPLC column, prior to scattering measurements. The protein was applied in a buffer containing sucrose and phosphate, mimicking the tRNA backbone, to increase stability for the protein at high concentrations required for SAXS analysis. Scattering data were collected at the Diamond Light Source beamline B21 (**Figure 2-35**).

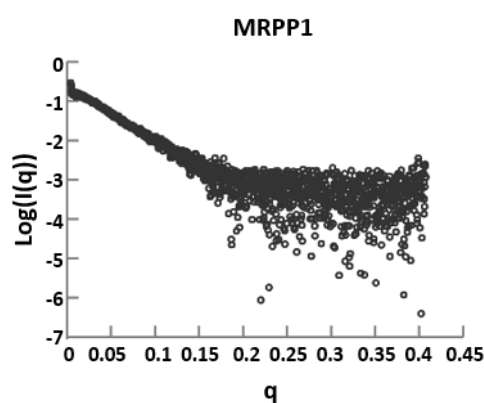


Figure 2-35. SAXS scattering plot: $\log(I(q))$ vs. q , for mature MRPP1.

The processed scattering data resulted in a linear Guinier fit (**Figure 2-36A**) giving rise to an R_g of 34.40 ± 1.885 Å. The Kratky plot (**Figure 2-36B**) indicated a folded, partly

flexible protein in solution. A D_{max} of 100 Å gave rise to the $P(r)$ distribution function shown in **Figure 2-36C**. Additional statistics can be found in **Appendix, Section 8.5**.

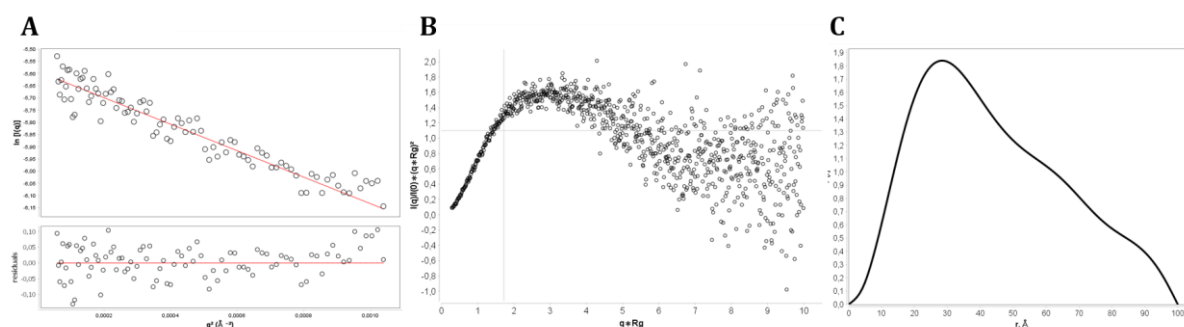


Figure 2-36. Scattering data plots for mature MRPP1. A) Guinier plot, B) $P(r)$ distribution function, C) Kratky plot.

The MW of the protein was calculated directly from the scattering curve to 97 kDa [161], supporting the previous findings of an MRPP1 dimer using SEC (theoretical MW of MRPP1 dimer: 91 kDa). The calculated SAXS model for dimeric MRPP1 is shown in **Figure 2-37**. The *monomeric* nature of the isolated *MT-domains* of MRPP1 in solution suggests that these domains are presumably not in close proximity to each other in the full-length MRPP1 dimer. The crystal structure of this domain (177 residues each) fits size-wise into the two ends of the model, leaving room for the small C-extensions (23 residues each) at the ends. The two N-domains (163 residues each) involved in the dimerisation interface can be accommodated in the centre space of the model, between the fitted MT-domains. It is possible that the N- and MT-domains are connected by a flexible linker as found in many methyltransferases, allowing inter-domain movement for substrate binding [139], [162]–[164]. In line with this, the Kratky plot indicates a flexible protein and the model seems rather extended considering the expected size of the dimer. It is possible that the protein will become more compact upon binding to MRPP2, as the N-domain and MT-domain in each monomer might otherwise not be tightly connected.

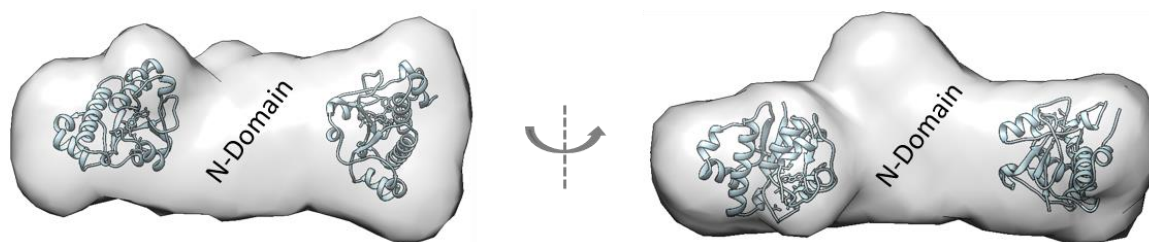


Figure 2-37. The SAXS model for dimeric, mature MRPP1. The crystal structure of the MRPP1 MT-domains were fitted into the SAXS model using the fit-in-map function in the Chimera software [165].

A comparison of the experimentally determined SAXS scattering plot of mature MRPP1 with the scattering plot calculated from the crystal structure of full-length Trm10 from archaea, using the Debye formula in FoxS [166], [167], results in a relatively high χ -value of 4.76 for q -values between 0.01° - 0.40° . A χ -value of this order indicates structural differences between the proteins and as their respective MT-domains align with an C^α -RMSD of 2.5 Å (**Figure 2-34**), these differences likely falls within the N-domains and C-terminal extensions.

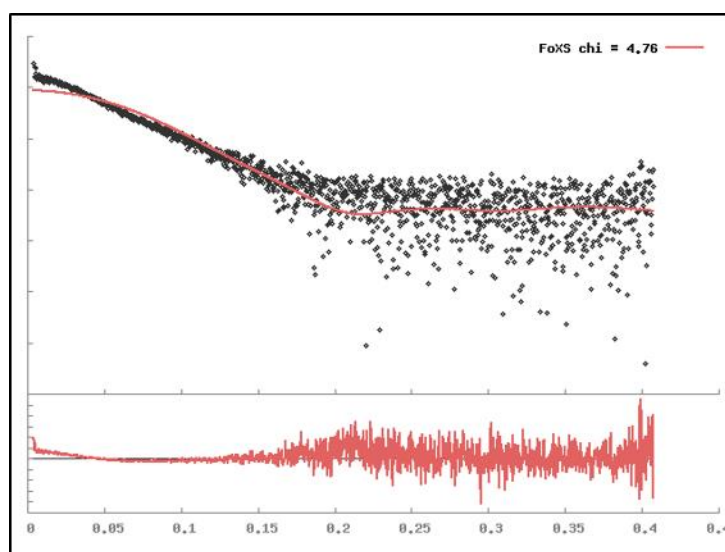


Figure 2-38. Comparison of the experimentally determined SAXS scattering plot ($\log I(q)$ versus q) of full-length MRPP1 (black), with the calculated SAXS scattering plot ($\log I(q)$ versus q) from the archeon Trm10 full-length crystal structure (red) computed with the Debye formula.

2.6 Substrate recognition

2.6.1 Purine binding

Upon methylation of purine-9 in (mt)tRNA by the MRPP1-MRPP2 complex, the purine moiety must be bound close to the SAM molecule within the MRPP1 protein. An

obvious binding pocket for this purine is visible in the structure of the MT-domain of MRPP1 (**Figure 2-39**). This pocket is sufficiently large ($\sim 8 \times 10 \times 8$ Å) to accommodate both the planar guanine (width: 6.5 Å between H-N1 and H-C8, height: 5.5 Å between H-(NH-C2) and O=C6) and planar adenine (width: 6.5 Å between H-C2 to H-C8, height: 5.0 Å between H-C2 and H-(NH-C6)).

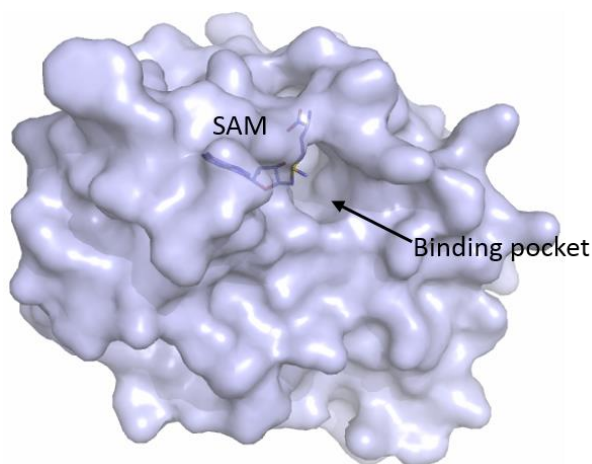


Figure 2-39. Potential purine binding pocket in MRPP1 in close proximity to the SAM cofactor (sticks).

To test whether the MT+C-domain of MRPP1 binds such purines, DSF was performed with the purine-containing compound 3'-adenosine monophosphate (3'-AMP) mimicking the adenosine in tRNA to be methylated (**Figure 2-40**).

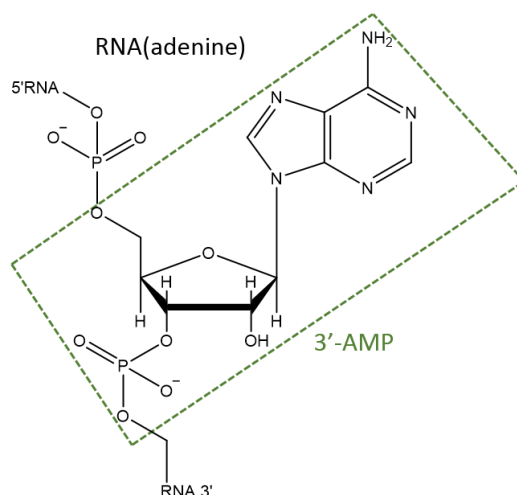


Figure 2-40. Chemical structure of RNA backbone with an adenosine base. The 3'-adenosine monophosphate (AMP) part of the RNA is outlined in green.

DSF was performed with the MT+C-domain($\Delta 202$) alone as this was the domain that had previously been crystallised, and in the presence of 1 mM 3'-AMP and/or 1 mM

S-adenosine-L-homocysteine (SAH) – the cofactor product of SAM. SAH was added to confirm that 3'-AMP does not occupy the pocket for the SAM cofactor, and was used instead of SAM to conclusively ensure no 3'-AMP protein turnover, though the MT-domain was shown to be inactive on its own (**Section 3.3.1**). **Figure 2-41** shows that 3'-AMP and SAH each binds to the MT+C-domain resulting in a ΔT_m of $3.65 \pm 0.16^\circ\text{C}$ and $5.0 \pm 0.18^\circ\text{C}$, respectively, and that the addition of both compounds resulted in a higher shift (ΔT_m of $5.65 \pm 0.20^\circ\text{C}$) indicating that they bind to different pockets of the protein.

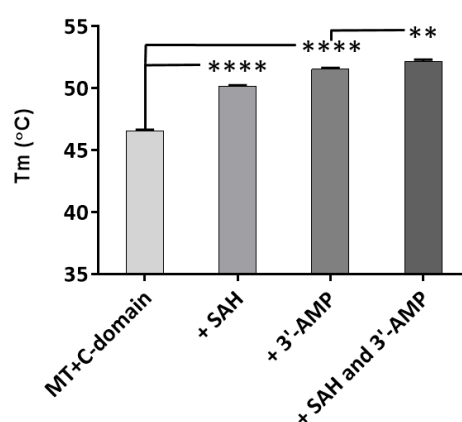


Figure 2-41. DSF assay of the MRPP1 MT+C-domain($\Delta 202$) added with SAH, 3'-AMP or both. For this experiment $n = 3$. **= $P < 0.01$, ****= $P < 0.0001$.

2.6.2 Soaking of MRPP1 crystals with nucleotide ligands

In an attempt to obtain a co-crystal structure with a bound purine-containing compound, such as the one identified in DSF, the well-diffracting diamond shaped crystals of the MT+C-domain($\Delta 202$) were grown. The crystals were soaked with either 10 mM 3'-AMP, 4 mM 5'-guanine monophosphate (5'-GMP) or 10 mM 3',5'-adenosine diphosphate (3',5'-ADP) (**Figure 2-42**), but they readily dissolved and diffraction was reduced to 3.03 Å resolution at best (1 min soak with 3',5'-ADP), which produced a dataset with no apparent density for the ligand.

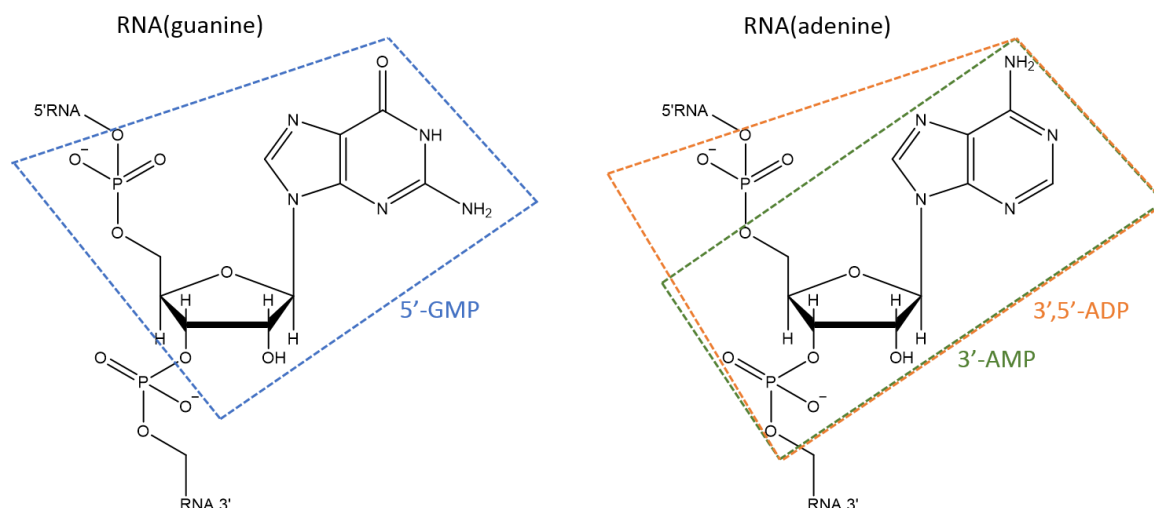


Figure 2-42. RNA backbone with either a guanine or an adenine base. The 5'-guanine monophosphate (5'-GMP, blue), 3'-adenosine monophosphate (3'-AMP, green) and 3',5'-adenosine diphosphate (3',5'-ADP, orange) part of the RNA is outlined.

To optimise soaking, new crystal forms were searched for. Due to the previously observed concentration dependence of crystal forms (**Section 2.5.1.2**), crystallisation plates were set up with a much higher protein concentration than before (30 mg/mL). These trials yielded crystals of new shapes (e.g. rectangular, almond shaped) in many drops (**Figure 2-43**). No clear correlation could be found between crystallisation condition and crystal shapes.

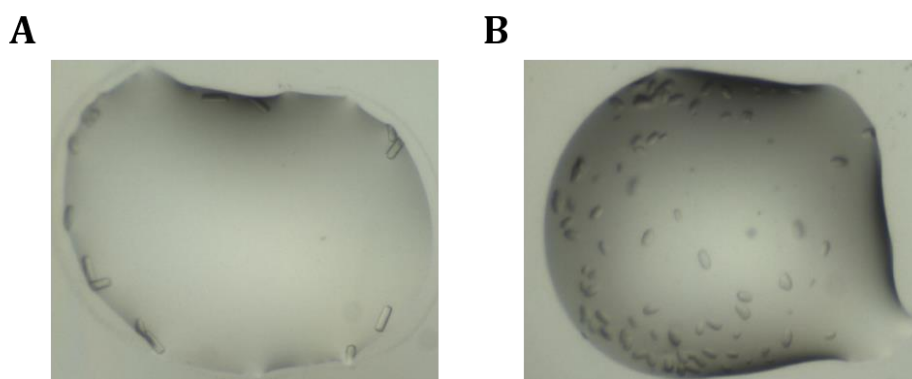


Figure 2-43. Crystals of the MT+C-domain(Δ 202) of MRPP1 grown at 30 mg/mL. A) rectangular crystals grown in 35% tacsimate. B) almond shaped crystals grown in 0.05 M ammonium sulphate, 0.05 M bis-tris pH 6.5, 30% pentaerythritol ethoxylate.

These crystals proved to be more stable for soaking in a solution of 1.2 mM SAH with either 10 mM 3'-AMP, 4 mM 5'-GMP or 10 mM 3',5'-ADP (**Figure 2-42**). The crystals were soaked for 1 sec to 1 h, washed in mother liquor, cryo-protected as before and

datasets were collected at the Diamond Light Source beamline IO2 for rectangular crystals diffracting to 1.9 Å (3'-AMP), 2.1 Å (5'-GMP) and 2.1 Å (3',5'-ADP) resolution. No diffraction was observed for the almond shaped crystals. The rectangular crystal belonged to the $P3_1$ or $P3_2$ trigonal space groups and phasing was performed by the MR technique with the MRPP1 MT-domain structure as search model, which identified $P3_1$ as the correct space group. The calculated difference maps ($F_o - F_c$) showed no density that could account for the soaked purine-containing compounds. Density was observed for the SAH ligand in the same conformation as previously found for SAM.

2.6.3 Co-crystallisation

The failed soaking attempts prompted co-crystallisation with 3-fold molar excess of 3'-AMP or 1.5-fold excess of 3',5'-ADP, using the same crystallisation conditions as before, with prior addition of equimolar SAH to the protein.

The MT+C-domain($\Delta 202$) protein co-crystallized with 3'-AMP in multiple drops, adopting a different, triangular shape (**Figure 2-44**). Unlike previous conditions, these crystals were grown using PEG4000 as the crystallisation precipitant. Unfortunately, no diffraction was observed for these crystals and optimisation is ongoing.

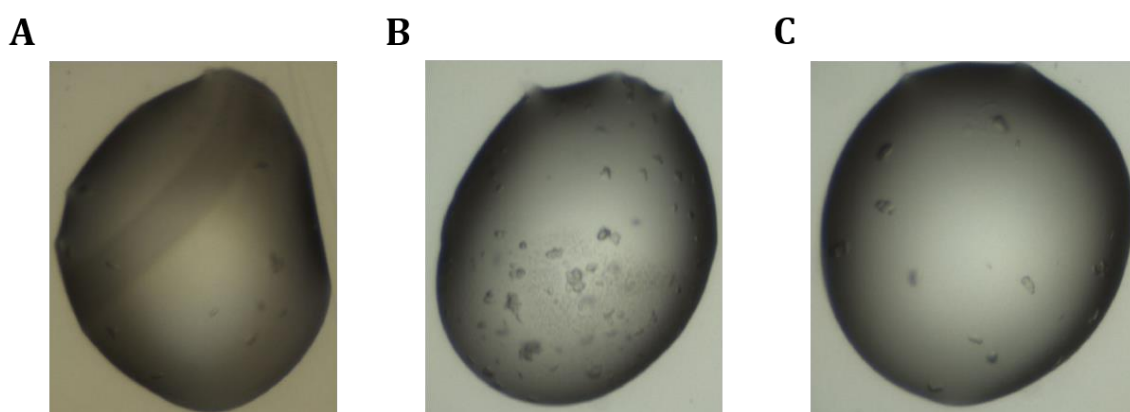


Figure 2-44. Crystals of the MT+C-domain($\Delta 202$) of MRPP1 from crystallisation attempt of co-crystallisation with SAH and either 3'-AMP or 3',5'-ADP. A (20°C) and B (4°C) of 0.2 M sodium acetate, 0.1 M Tris pH 8.5, 30% PEG4000, and C (4°C) of 0.1 M ammonium acetate, 0.1 M citrate pH 5.6, 30% PEG4000.

The sample added 3',5'-ADP yielded needle-shaped crystals similar to those obtained originally with addition of SAM only (**Figure 2-23A**). This crystal form was previously found not to diffract, and in the interest of time these crystals were therefore not further pursued.

2.6.4 Docking

After failed soaking and co-crystallisation attempts, 5'-GMP was docked into the putative binding pocket of MRPP1 MT-domain structure, in order to obtain some structural insight into optimally located residues for m¹R9 catalysis. The docking results placed the phosphate group of 5'-GMP into the putative purine binding site for seven of the nine models, which were excluded. The best docking model based on predicted binding affinity (-6.1 kcal/mol) displayed the purine part of the ligand bound into the binding site (**Figure 2-45**). Docking of 5'-AMP resulted in a similar fit as the 5'-GMP, and analysis of the active site residues in **Chapter 3** will be based on the docking model with guanine.

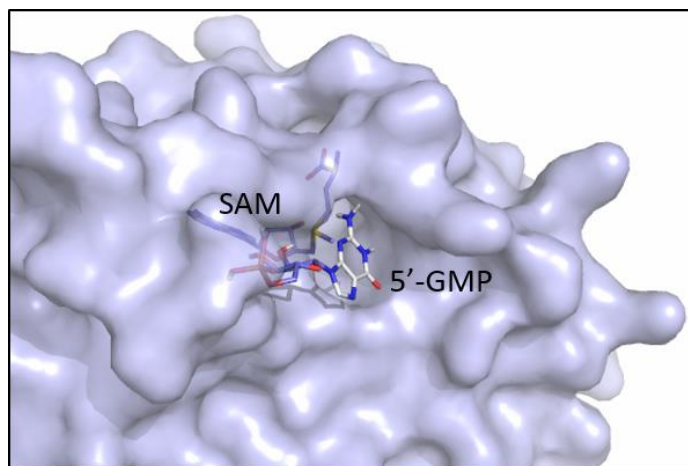


Figure 2-45. Docking of 5'-GMP into the presumed purine binding site of the MT-domain of MRPP1.

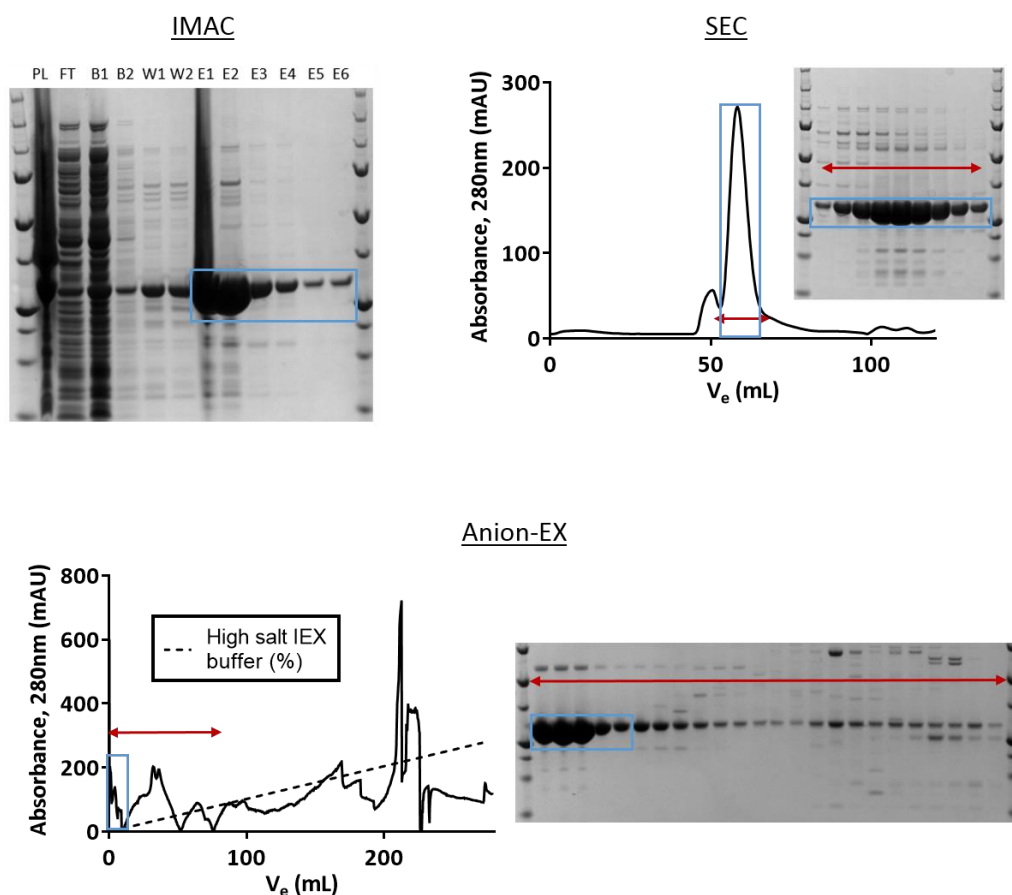


Figure 3-2. Large scale purification of full-length MRPP2. The example shows SDS-PAGE analysis of initial IMAC, chromatogram from SEC and SDS-PAGE analysis of peak fractions, and chromatogram from anion-EX and SDS-PAGE analysis of peak fractions. Red arrows correspond to the region analysed on SDS-PAGE. Blue boxes indicate the pooled fractions. V_e : elution volume, PL: pellet, FT: flow through, B1 and B2: binding buffer fractions, W1 and W2: wash buffer fractions, E1-6: elution buffer fractions.

3.1.1 Reconstitution of complex from individual MRPP1 and MRPP2 expression

Plasmids containing His₆-tagged mature MRPP1 or His₆-tagged full-length MRPP2 were separately transformed into *E. coli* and expressed. Cell pellets for the two proteins were co-lysed by homogenisation to enable complex formation ahead of purification. The protein complex was purified (**Figure 3-3**) by IMAC using Ni-NTA resin, followed by SEC on an S200 column and cation-EX on a HiTrap SP HP.

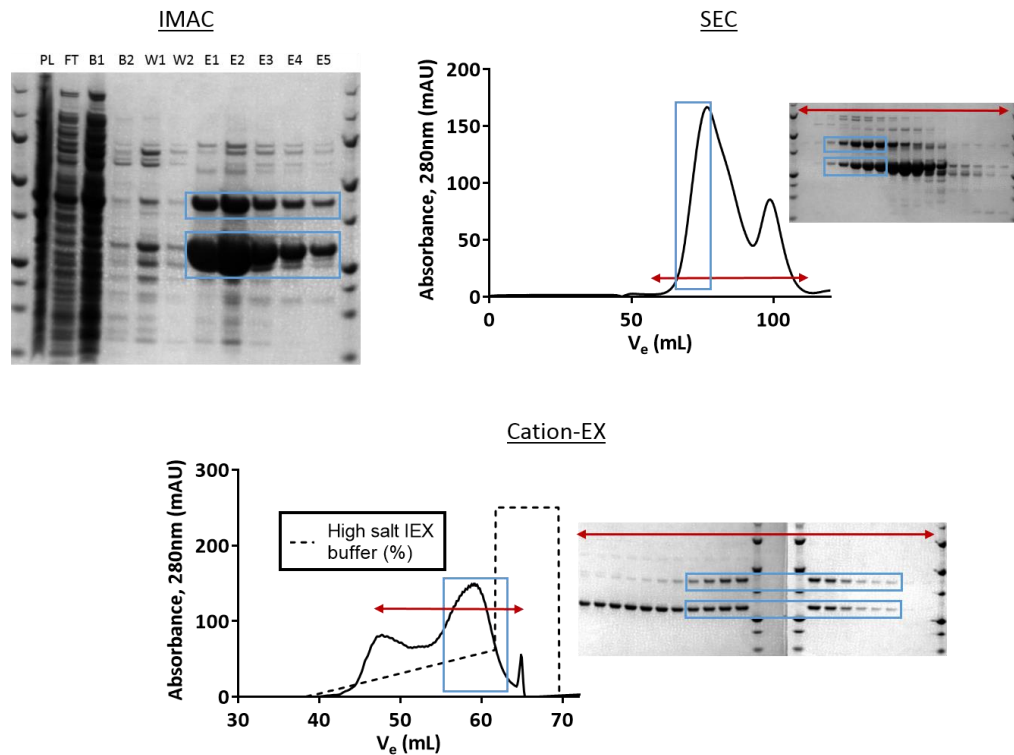


Figure 3-3. Large scale purification of mature MRPP1 co-lysed with full-length MRPP2. The example shows SDS-PAGE analysis of the initial IMAC, chromatogram from SEC and SDS-PAGE analysis of peak fractions, and chromatogram from cation-EX and SDS-PAGE analysis of peak fractions. Red arrows correspond to the region analysed on SDS-PAGE. Blue boxes indicate the pooled fractions. V_e : elution volume, PL: pellet, FT: flow through, B1 and B2: binding buffer fractions, W1 and W2: wash buffer fractions, E1-5: elution buffer fractions.

In this purification procedure, the SEC step revealed co-elution of the complex and excess uncomplexed MRPP2 tetramer. The excess MRPP2 protein was to a large extent removed in the last cation-EX step, but the reduction of NaCl concentration to below 400 mM prior to cation-EX caused loss of protein due to precipitation. Further issues included further precipitation upon TEV protease treatment and removal of only 50% of the His₆-tags on MRPP2 proteins of the remaining complex.

3.1.2 Co-expression of MRPP1-MRPP2

To bypass the above issues of excess uncomplexed proteins and precipitation during complex reconstitution, an attempt to co-express the two proteins was made. A DNA fragment that contains both mature MRPP1 and full-length MRPP2 coding regions, with a ribosomal binding site between them, was generated. This fragment was cloned into the

pNIC28-Bsa4 vector, resulting in a bicistronic transcript where the two genes are transcribed from a single promoter, but with individual translation initiation sites, to generate N-terminal His₆-tagged MRPP1 and untagged MRPP2.

This bicistronic co-expression plasmid was expressed, and the complex was purified (**Figure 3-4**) by IMAC using Ni-NTA resin, followed by SEC on a Superose 6 (S6) column and cation-EX on a HiTrap SP HP column to which the protein was applied at 300 mM NaCl, which proved to be the lowest NaCl concentration attained without protein precipitation. TEV protease treatment was also omitted to avoid the previously observed protein precipitation.

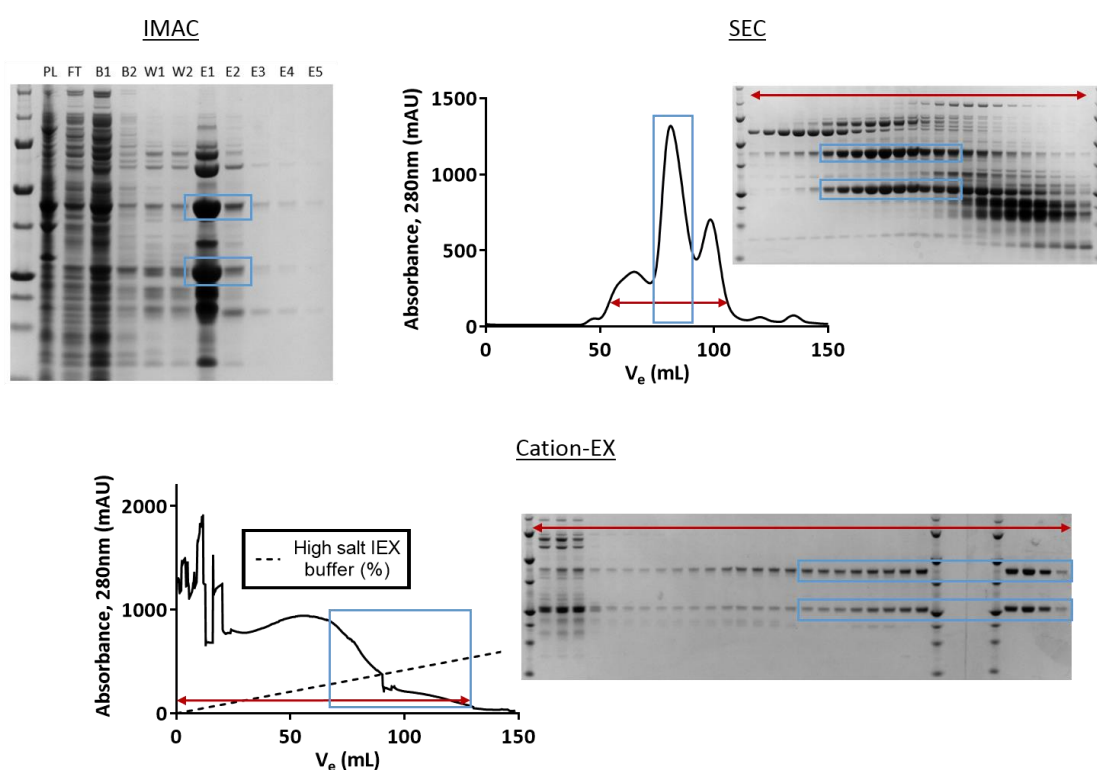


Figure 3-4. Large scale purification of co-expressed mature MRPP1 and full-length MRPP2. The example shows SDS-PAGE analysis of the initial IMAC, chromatogram from SEC and SDS-PAGE analysis of peak fractions, and chromatogram from cation-EX and SDS-PAGE analysis of peak fractions. Red arrows, blue boxes and gel lanes are annotated as before.

The strategy of co-expressing His₆-tagged mature MRPP1 and untagged full-length MRPP2 circumvented the issue of contaminating the complex sample with excess uncomplexed MRPP2 protein, as only MRPP1-complexed MRPP2 was retained in the

Ni-NTA resin. The co-expressed proteins also avoided precipitation in buffers with NaCl concentration down to 300 mM, which reduced the loss of protein in the cation-EX step. This purification protocol yielded large amount of pure protein complex for assays. The need for protein buffers with high concentration of NaCl could render later complex crystallisation by vapour diffusion equilibration difficult, by excluding random matrix screen conditions with $[Na^+]$ and $[Cl^-] < 300$ mM (**Section 7.5.1**).

Unless otherwise stated, the binary complex used throughout this chapter is prepared from co-expression of MRPP1 and MRPP2 proteins where the MRPP1 carries an N-terminal His₆-tag. Further reconstitution studies involving MRPP1 proteins utilise His₆-tagged mature MRPP1 and MT+C-domain(Δ 194), and untagged MT+C-domain(Δ 202) and N-domain produced as described in **Chapter 2**.

3.2 Interaction studies

3.2.1 Mature MRPP1 and MRPP2

The purified proteins of mature MRPP1 and MRPP2 were used in an analytical SEC with the purpose of investigating the interaction between them. SDS-PAGE analysis was performed on the eluted fractions collected throughout the analytical SEC experiment. If a complex is formed between the mixed proteins, these should be detected in earlier fractions compared to when they are applied as individual, uncomplexed proteins.

To establish complex formation, SEC was performed with mature MRPP1 alone, MRPP2 alone and a mixture of the two after extended incubation (1-2 h) on ice. The chromatogram, when the two proteins were mixed (**Figure 3-5, green box**), showed the appearance of a peak to the left of the corresponding peaks for the individual proteins, indicative of complex formation. No cofactors or substrates were added to the mixture, suggesting that the complex formation is not dependent on neither SAM nor the (mt)tRNA substrate. Complex formation is further evidenced in the SDS-PAGE gel by the appearance

of MRPP1 and MRPP2 bands in earlier SEC fractions (**Figure 3-5, green box**) and depletion of MRPP1 and MRPP2 bands in the elution fractions of the individual proteins (**Figure 3-5**). Altogether, a larger species (a complex) is formed between the two proteins without the need of added cofactors or substrate.

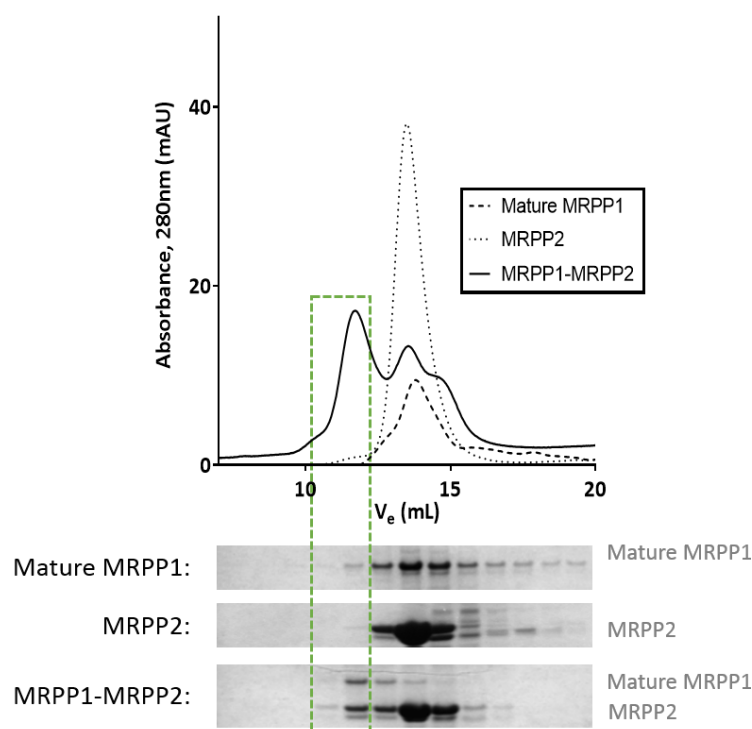


Figure 3-5. Analytical SEC with the mature MRPP1 and full-length MRPP2 proteins applied together and individually. The green box indicates the elution fractions in which proteins should be co-eluting if they form a complex.

The effect of cofactors and substrates on complex *dissociation*, if any, had not previously been studied. Analytical SEC was next used to test stability of the co-expressed MRPP1-MRPP2 complex, applied to the column after extended incubation (3 h) with excess amounts of cofactors and substrates for each protein, *i.e.* SAM for MTase activity of MRPP1, NAD⁺/NADH and allopregnanolone for SDR activity of MRPP2, and (mt)tRNA for the MTase activity of the MRPP1-MRPP2 complex. To ensure that the cofactors were in fact bound to the proteins, a DSF assay was performed alongside the analytical SEC. The DSF assay (**Figure 3-6B**) showed an increase in T_m when incubating the MRPP1-MRPP2 complex with both SAM ($\Delta T_m = 3.32 \pm 0.61^\circ\text{C}$) and NADH ($\Delta T_m = 3.73 \pm 0.58^\circ\text{C}$). No T_m

shift was observed for NAD^+ under these assay conditions ($\Delta T_m = 0.47 \pm 0.47^\circ\text{C}$), which is consistent with DSF data obtained on active MRPP2 protein alone (**Section 5.1.5**) under same assay conditions, and thus likely does not reflect a lack of binding. When comparing the obtained analytical SEC data with data from individual MRPP1 and MRPP2 proteins, no complex *dissociation* was observed in the chromatogram or SDS-PAGE (**Figure 3-5A**), showing that the availability of excess amounts of cofactor and substrates does not cause notable *dissociation*. It remains to be established if the presence of these cofactors in large excess would affect complex *association*.

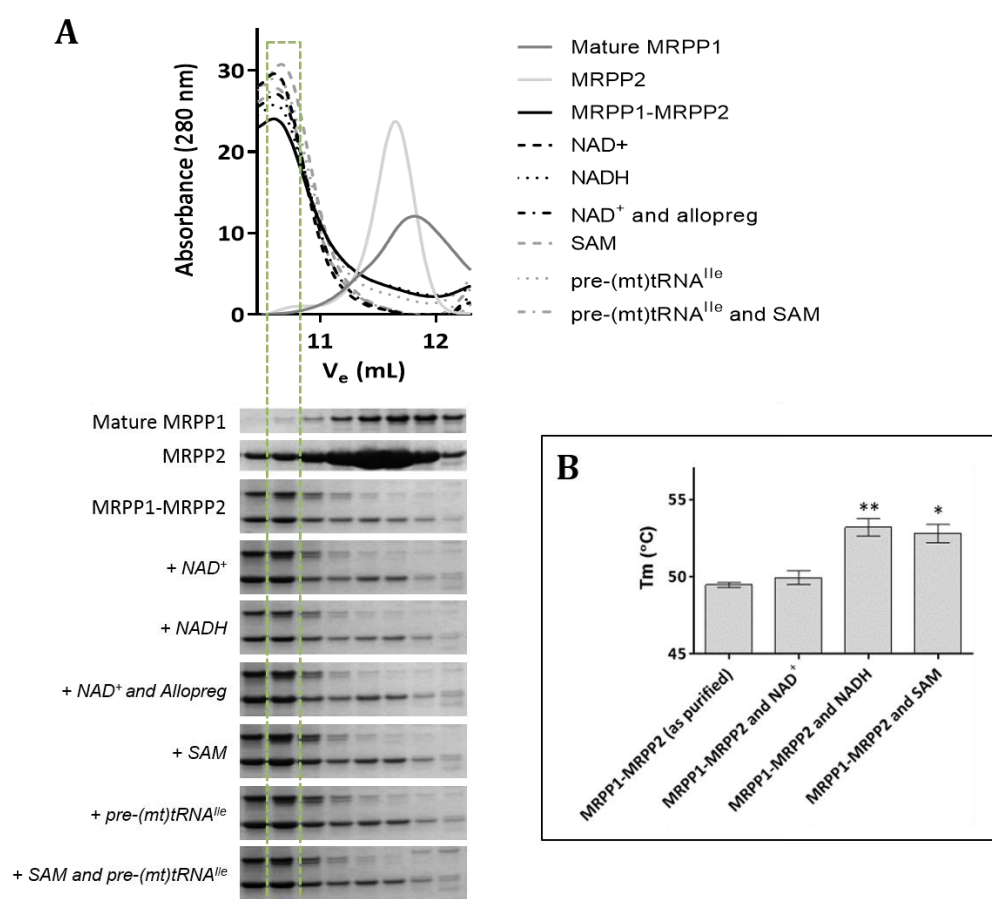


Figure 3-7. A) Analytical SEC with mature MRPP1, full-length MRPP2, or co-expressed MRPP1-MRPP2. The co-expressed complex was further applied with various combinations of cofactors and MRPP1/MRPP2 substrates. The green box indicates the elution fractions in which proteins should be co-eluting if they form a complex. B) T_m of co-expressed MRPP1-MRPP2 proteins in the presence and absence of cofactors used in the analytical SEC. * = $P < 0.05$, ** = $P < 0.01$.

Having established that mature MRPP1 and full-length MRPP2 form a stable complex, the stoichiometry of this complex, generated from co-expression, was investigated

by sedimentation velocity analytical ultracentrifugation (SV-AUC). SV-AUC studies the sedimentation of macromolecules in solution in high centrifugal fields using hydrodynamic theory to define the size, shape and interactions of the macromolecules in solution [168]. For AUC, buffer subtraction is crucial for the data quality and so, to ensure that the reference buffer and protein buffer were of equal composition, the MRPP1-MRPP2 complex was dialysed overnight and the dialysis buffer was subsequently used as reference. The phosphate buffer used for SAXS analysis of MRPP1 was chosen in the SV-AUC experiment to mimic the potential protein stabilising effects upon binding of the tRNA sugars and phosphate moieties. The obtained sedimentation coefficient distribution function, $C(s)$ (**Figure 3-8**), shows a peak corresponding to a MW for the binary complex of 201 kDa. This indicates a 2xMRPP1:4xMRPP2 stoichiometry comparable to the expected MW of 208 kDa for such a complex (MW dimeric His₆-MRPP1 = 90312.2 Da, MW tetrameric untagged MRPP2 = 117903.6 Da). However, as the SV-AUC technique is protein-shape dependent, the determined stoichiometry was further evaluated using other methods (**Section 3.4.2.1**). A small peak was also observed corresponding to a MW of 120 kDa, which could be explained by a tetramer of MRPP2 of 118 kDa.

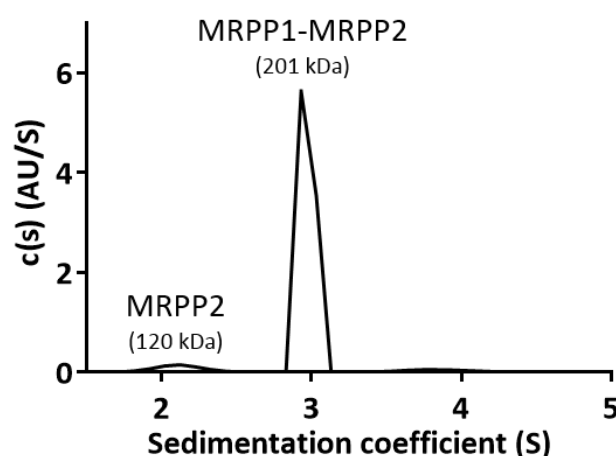


Figure 3-8. $C(s)$ distribution function from an SV-AUC experiment with the co-expressed, binary complex of MRPP1-MRPP2. RMSD = 0.007802.

The SV-AUC method is dependent on the shape of the molecule as e.g. a plate-like structure would displace a larger mass as it sediments (F_b), compared to a rod-like structure, ultimately affecting the sedimentation coefficient (s) (further described in **Section 7.4.6**). Instead, the shape-independent sedimentation equilibrium AUC (SE-AUC) method could have been used to more accurately determine the MW of the complex. In SE-AUC, the sample is allowed to slowly distribute in the cell at low centrifugal speed to create a concentration gradient across the cell (equilibrium), making this method independent of the frictional forces and thereby of the hydrodynamic shape of the protein [169].

3.2.2 Domain mapping for complex formation using SEC

To investigate the interaction between MRPP2 and different domains of MRPP1, purified proteins of MRPP1 N-domain, MT+C-domain and MRPP2 were used in the analytical SEC assay. MRPP1 proteins were applied to the SEC column with or without MRPP2, and in the presence or absence of SAM and *in vitro* transcribed pre-(mt)tRNA^{Ile} substrate to understand if these would affect protein interaction.

The first experiment set out to determine if the N-domain of MRPP1 alone was sufficient for complex formation with MRPP2. As described in **Chapter 2**, some degradation was observed for the MRPP1 N-domain, with the protein being predominantly present as a fragment of residues 40-191. This protein was applied to the SEC column alone or with MRPP2, in the presence or absence of SAM and pre-(mt)tRNA^{Ile}. The chromatogram (**Figure 3-9, green box**) showed no peak shift upon mixing of the proteins and cofactor/substrates, and SDS-PAGE analysis did not reveal both MRPP1 N-domain and MRPP2 bands in the earlier elution fractions, which would be expected for a formed complex, compared to where they would otherwise be found as individual proteins. This indicates that the N-domain alone is not sufficient for complex formation with MRPP2 either in the presence or absence of cofactor and/or substrate.

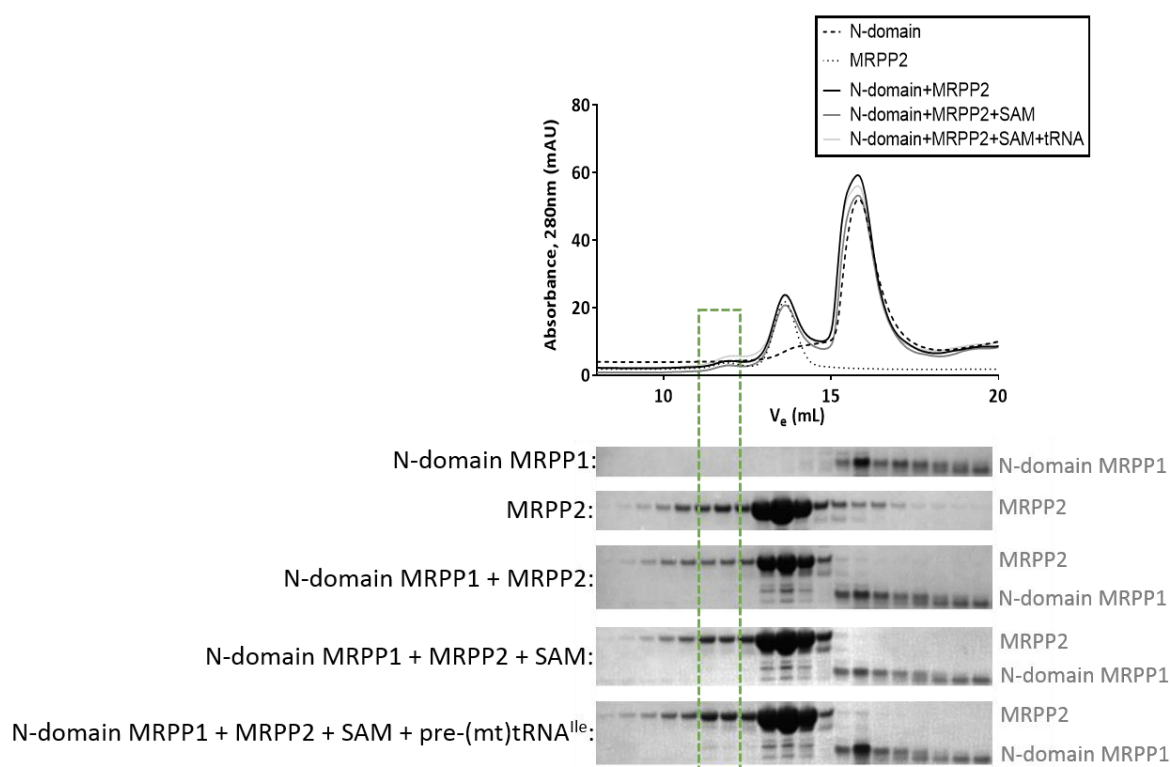


Figure 3-9. Analytical SEC with the N-domain of MRPP1 applied with and without MRPP2 and in the presence and absence of the cofactor and substrate. The green box indicates the elution fractions in which proteins should be co-localising if they form a complex.

The second experiment investigated if the primary interaction region with MRPP2 lies within the MT+C-domain of MRPP1. The MT+C-domain($\Delta 194$) of MRPP1 was applied to the SEC column alone or with MRPP2, in the presence or absence of SAM and pre-(mt)tRNA^{Ile}. The chromatogram (**Figure 3-10, green box**) showed no peak shift upon mixing of the proteins and cofactor/substrates, and the SDS-PAGE analysis (**Figure 3-10, green box**) again did not identify MT+C-domain of MRPP1 or MRPP2 proteins in earlier elution fractions where a complex would be expected. This indicates that the MRPP1 MT+C-domain alone – like the N-domain – is not sufficient for complex formation with MRPP2, either in the presence or in the absence of cofactor and/or substrate.

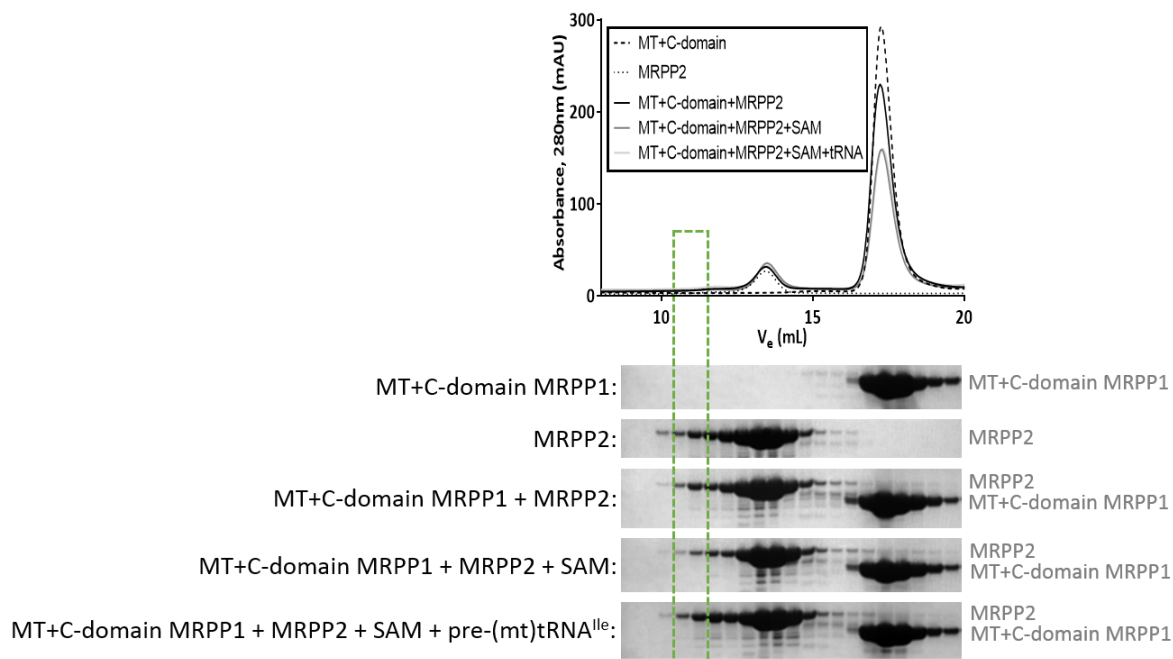


Figure 3-10. Analytical SEC with the MT+C-domain(Δ 194) of MRPP1 applied with and without MRPP2, in the presence and absence of cofactor and substrate. The green box indicates the elution fractions in which proteins should be co-localising if they form a complex.

Overall, the two SEC experiments demonstrate that none of the individual MRPP1 domains alone are sufficient for complex assembly with MRPP2, thus regions from both the N-domain and the MT+C-domain must be required. Interestingly, the interaction between mature MRPP1 and full-length MRPP2 eliminated all degradation otherwise observed for MRPP1 (**Figure 3-11**) suggesting that a role of MRPP2 in this complex could be to stabilise regions of MRPP1 that would otherwise be prone to cleavage by *E. coli* proteases. From the degradation pattern observed for full-length MRPP1 alone (**Figure 3-11**, left), tryptic digest analysis suggests that the protease cleavage sites are found between the N-domain and MT-domain and between the MT-domain and C-extension.

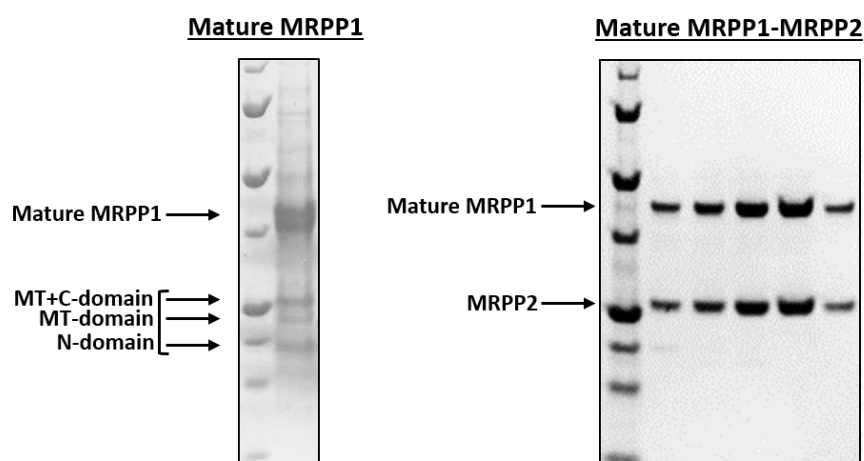


Figure 3-11. Left) SDS-PAGE analysis of mature MRPP1 showing the degradation observed for this construct in the absence of MRPP2. Tryptic digest-MS/MS analysis identified Mature MRPP1 as residues Ile44-Arg397, MT+C-domain as residues Asn187-Arg397, MT-domain as residues Asn187-Lys377 and N-domain as residues Ile44-Lys177. Right) SDS-PAGE analysis of co-expressed mature MRPP1 and MRPP2 showing the halted protein degradation for mature MRPP1 in the presence of MRPP2.

3.2.3 Sequence mapping of the interaction regions

The analytical SEC results indicated that regions from both the MRPP1 N-domain and MT+C-domain were important for the interaction with the MRPP2 protein. To further map the sequence regions of MRPP1 involved in this interaction, a SPOT blot analysis was performed. An array of short peptides, together covering the full protein sequence to be analysed for protein binding, are synthesised and chemically coupled to a membrane with each peptide appearing as a 'spot'. For the investigations of MRPP1 and MRPP2 interaction, the entire protein sequence for mature MRPP1 was translated into overlapping 15-mer peptides with maximum overlap of 13 amino acids (*i.e.* a 2 amino acid shift in register between peptides). The SPOT membrane with the coupled peptides was incubated with His₆-tagged MRPP2 protein overnight. The next day, the membrane was washed with PBS/PBST to remove unbound protein, probed with an HRP-conjugated anti-His antibody, and visualised by chemiluminescence. The spots that appeared dark on the membrane (**Figure 3-12A**) reacted to the antibody, and therefore represent the sequence regions at which His₆-MRPP2 is bound to an MRPP1 peptide. The membrane was subjected to densitometry analysis to deduce the intensity of each spot, subtracted from control

background spots. The densitometry data are represented as a bar chart in **Figure 3-12B**. The areas identified as MRPP2 interacting regions are numbered 1-7 in **Figure 3-12**, and a mapping of the corresponding domains in MRPP1 (N, MT and C) is shown in **green**. The crystallised part of MRPP1 is outlined as a **blue** dashed line in **Figure 3-12A**.

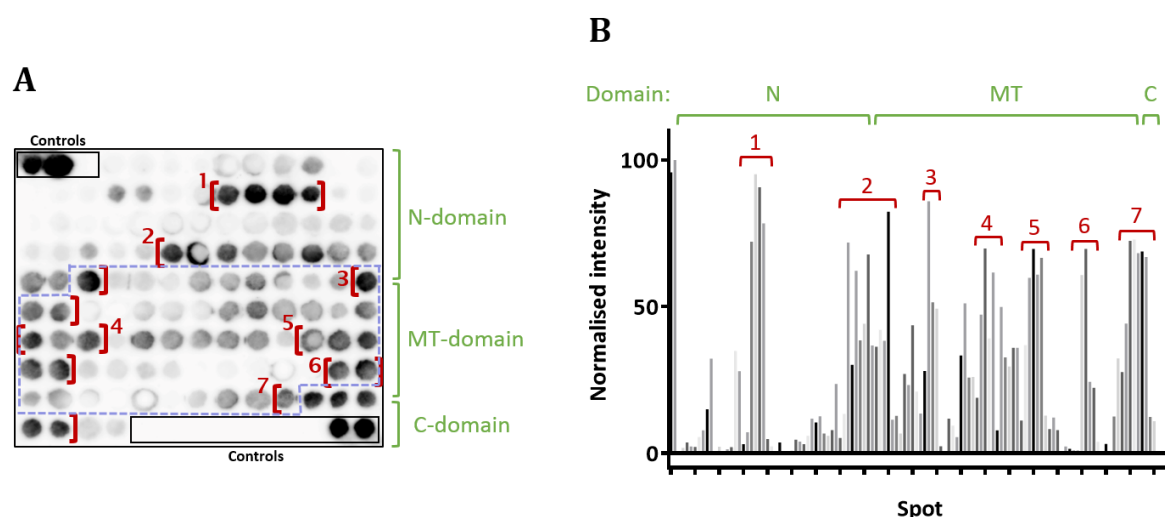


Figure 3-12. A) MRPP1 peptide membrane with bound His₆-tagged MRPP2 labelled with HPR-conjugated anti-His antibody and visualised by chemiluminescence. Control peptides are shown in black boxes: as positive control the first two and last two spots contains a His₆-peptide, and as negative control the third and last seven spots before the His₆-controls are blank (no peptides coupled). B) Densitometry data from the raw membrane represented as a bar chart. **Red** brackets indicate regions identified to interact with MRPP1. The **blue** dashed line indicates the crystallised MT-domain region. The domains of MRPP1 are annotated in **green**.

The SPOT blot analysis revealed at least one region from the N-domain as a strong candidate for interaction with MRPP2 (**Region 1**), which interestingly lies within the only region of the N-domain that is predicted not to be intrinsically disordered (residue 87-152 compared to **Region 1** of residues 100-106) (**Section 2.1**). Furthermore, the region towards the end of the N-domain and before start of the MT-domain (**Region 2**) contained many dark spots and could represent a further binding site for MRPP2. Unsuccessful homology modelling for the N-domain restricted structural mapping of **Region 1** and **Region 2**. Such mapping was only possible within the obtained crystal structure of the MT-domain, where more dark spots were observed, corresponding to **Regions 3-6** (**Figure 3-13A**), with **Region 5** not being surface exposed. The surface exposed **Region 6** covered the flexible loop, built as poly-alanine in the structure due to weak electron density (residues 335-352). The

equivalent loop in Trm10 proteins from archaea and yeast have been studied by mutagenesis and docking studies, that revealed an important role of this loop in tRNA binding and activity, and thus this seems an unlikely candidate for MRPP2 interaction [49], [159]. **Region 3** and **Region 4** clustered into a patch near the beginning of the MT-domain (**Figure 3-13A**, ‘structure start’) only seven residues N-terminal of **Region 2**. Altogether, **Regions 2-4** represent a likely collaborative patch for MRPP2 interaction (**Figure 3-13A** and **B**). The strong candidate from the N-domain (**Region 1**) could either form part of this patch by structurally localising close to **Region 2-4**, or independently interact with MRPP2. Involvement of both the N-domain and MT-domain for interaction with MRPP2 in the SPOT assay is entirely consistent with the previous analytical SEC data showing no complex formation with either fragment alone.

In addition to the above regions, the SPOT blot analysis also identified the C-extension as an interaction site for MRPP2 (**Region 7**). This extension was either disordered (flexible) or not present in the crystal structure (prone to protease attack), but protected from degradation in solution when complexed with MRPP2. Together, these data suggest that the C-extension is involved in interaction with MRPP2. The beginning of the C-extension is structurally distant from the predicted interaction patch, but its flexibility may allow long-range interactions with other parts of MRPP1, e.g. the proposed interaction patch for MRPP2 (**Figure 3-13A**). Alternatively, this C-extension could interact independently with MRPP2.

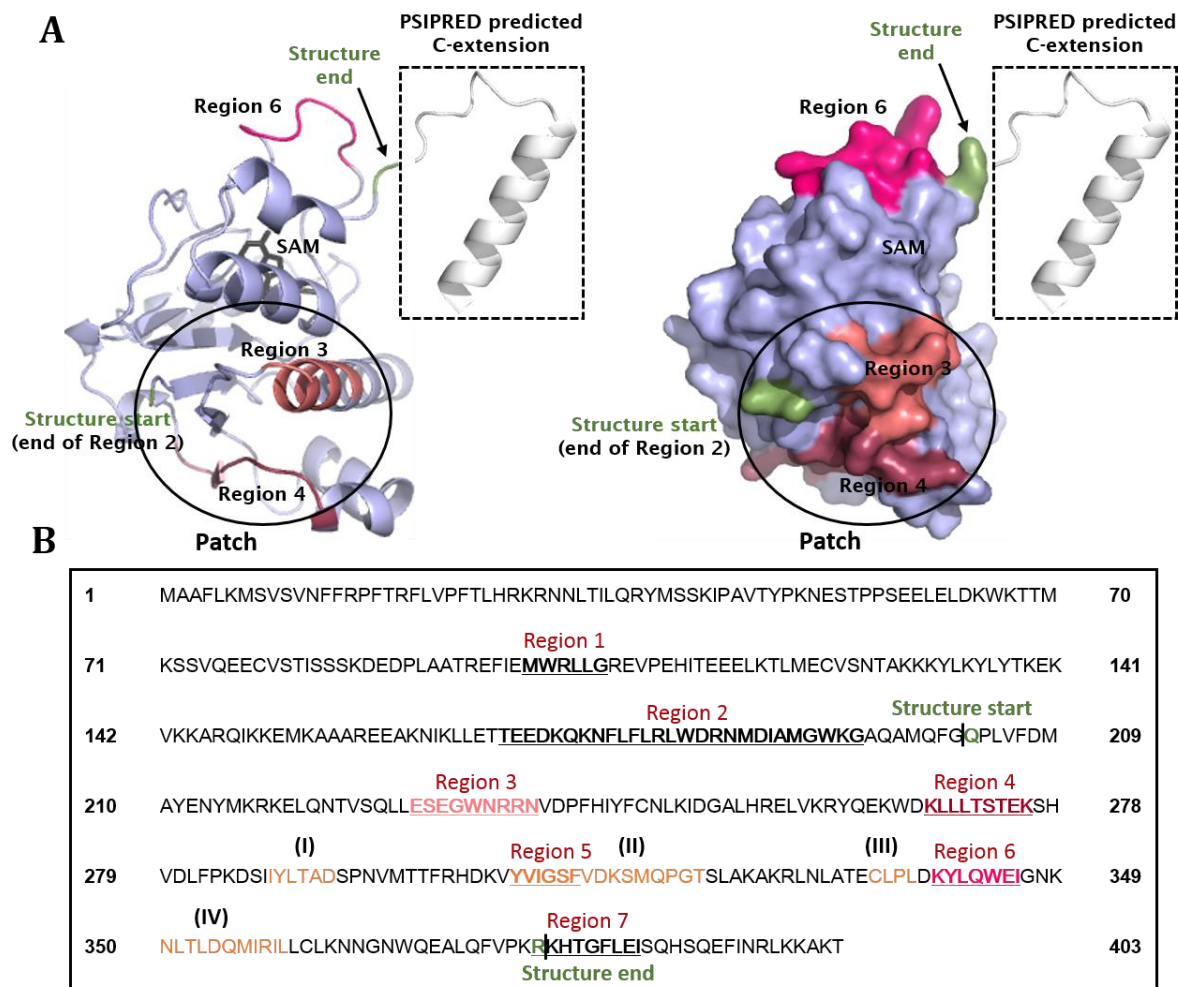


Figure 3-13. A) Mapping of the sequences identified in the spot blot analysis as potential MRPP2 interaction sites, onto the MRPP1 MT structure in cartoon (left) and surface (right) representation. The main interaction patch is shown. The start and end of the structure are indicated in green. B) Mapping of the identified interaction regions onto the mature MRPP1 protein sequence. The regions are shown in bold and coloured in the same scheme as A).

The detection limit in a SPOT blot analysis for protein-peptide interaction is typically $K_D = 500 \mu\text{M}$ [170], thus some weak interactions might not be detected. Also, steric constraints of the immobilised peptides, lack of sufficient N- and C-terminal flanking regions, and synergistic binding involving different regions in binding, could be other reasons for potential unidentified interactions. It is therefore possible that additional, unidentified, regions are important for the MRPP1-MRPP2 interaction. The SPOT blot analysis cannot conclusively exclude random protein-peptide interactions, and such false positives could be excluded by repeating the assay with a number of standard proteins to

identify any often-occurring hits or by K_D determination using biochemical interaction methods such as isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR).

After having mapped the binding sites of MRPP2 to MRPP1, the SPOT blot assay was attempted performed in reverse with peptides of MRPP2 bound to the membrane. Unfortunately, the mature MRPP1 protein was not stable in the buffer required for blotting.

Having concluded that none of the MRPP1 domains on their own can form a complex with MRPP2, enzymatic assays were performed to investigate if this property was also reflected in the MTase activity.

3.3 Activity studies

3.3.1 Mapping of MRPP1 domains in MTase activity

A MTase activity assay was set up via collaboration with the Droogmans' group at the Université Libre de Bruxelles, Belgium, using, as reagents, the *in vitro* transcribed pre-(mt)tRNA^{Ile} substrate, different MRPP1 fragments, full-length MRPP2 and a SAM cofactor tritium-labelled at the methyl-group ([methyl-³H]SAM). The pre-(mt)tRNA^{Ile} substrate contains a guanine base in position 9, thus this assay is testing the m¹G9 activity of the MRPP1-MRPP2 complex. Activity was measured in counts per minute (cpm), with counts arising from the radioactively labelled methyl-groups from [methyl-³H]SAM incorporated into the pre-(mt)tRNA substrate.

To test if mature MRPP1 alone is active on pre-(mt)tRNA^{Ile}, the m¹G9 MTase activity assay was performed with mature MRPP1 in the presence or absence of MRPP2. The obtained cpm-values (**Figure 3-14**) show that the co-expressed complex of mature MRPP1 and full-length MRPP2, but not the individual proteins, has MTase activity, confirming previous findings that MRPP1 requires MRPP2 for catalysis [14]. The observed activity for the MRPP1-MRPP2 complex further confirms that the *E. coli* expression system produces recombinant proteins with an enzymatically relevant fold.

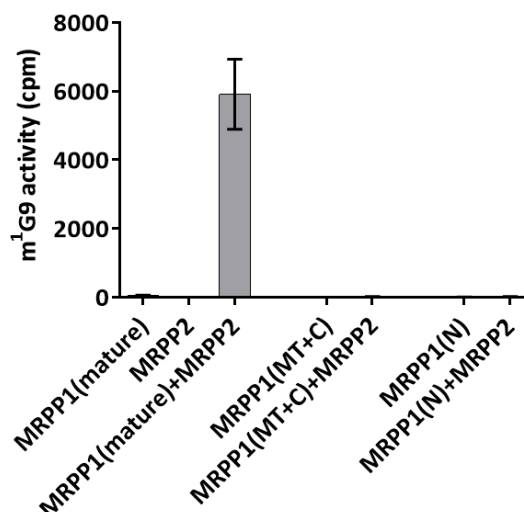


Figure 3-14. MTase (m¹G9) activity of MRPP1 mature protein, MT+C-domain(Δ 203) and N-domain towards pre-(mt)tRNA^{lle} in the absence and presence of MRPP2. For this experiment n = 3.

The MTase activity assay was then utilised to test the m¹G9 activity on the MT+C-domain of MRPP1, harbouring the MTase active site. The assay was performed with the MT+C-domain(Δ 203) protein in the presence or absence of MRPP2 and showed no activity in both instances (**Figure 3-14**), consistent with the inability of this domain to form a complex with MRPP2, necessary for activity of mature MRPP1. The lack of activity found for the MT+C-domain meant that the assay was not performed with the MT-domain alone, as it is doubtful that activity would be gained upon further removal of the C-extension.

The m¹G9 MTase activity was further tested for the N-domain of MRPP1 in the presence or absence of MRPP2 (**Figure 3-14**). Again, no activity was observed, which is not surprising as 1) the active site and SAM binding site reside in the MT-domain of MRPP1 that is not present in this protein fragment, and 2) the N-domain is not able to form a complex with the MRPP2 protein necessary for activity.

3.3.2 Mapping of catalytic residues in MRPP1

MRPP1 is known to catalyse both m¹G9 and m¹A9 activities, and its crystal structure represents the first of a protein from the Trm10 family with dual specificity, since all previous Trm10 structures represent enzymes that catalyse only m¹G9 or only m¹A9

[14], [159], [171]–[173]. Therefore, inspection of the MRPP1 MT-domain structure, and comparison with other Trm10 structures, could potentially provide clues as to how MRPP1 catalyses the methylation of two differently charged substrates (adenine and guanine), while other members have single specificities.

The mechanism for N1-methylation by MRPP1 has not yet been determined, but the N1-methylation reactions of guanine and adenine in tRNA are studied in other organisms. These studies show that methylation of N1 of adenine and guanine requires two different mechanisms. N1-methylation of guanine by TrmD/Trm5 requires an Asp or Glu residue as the catalytic base and an Arg as an intermediate charge-stabiliser. TrmD has further been shown to utilise a Mg^{2+} ion for stabilisation of the charged intermediate [60]. N1-methylation of adenine by TrmI likewise requires an Asp as the catalytic base, but as there is no intermediate charge created, no intermediate charge-stabiliser would be needed (**Section 1.3.2**).

A sequence alignment of the Trm10 proteins (**Section 8.3**) revealed a conserved Asp residue in position 314 and a conserved Gln residue in position 226 of MRPP1 (conserved in both m¹A9, m¹G9 and dual-specific Trm10 MTases). Asp314 is located within the active site (**Figure 3-16**), forming hydrogen bonds to the backbone amide of Ser316, Met317 and Gln218 as well as the NH_3^+ -group of the methionyl moiety of SAM, and thus appears a good candidate for the catalytic base. Gln226 is located in strand β 1 outside the active site, but the side-chain points towards the active site, 3.1 Å from the O6-atom of the docked guanine (**Section 2.6.4**), and could thus potentially function as the intermediate charge-stabiliser in the methylation of guanine (**Figure 3-16**).

To test this hypothesis, an Asn-substituted mutant of the potential catalytic base Asp314 (p.D314N) and an Ala-substituted mutant of the potential intermediate charge-stabiliser Gln226 (p.Q226A) were generated by site-directed mutagenesis using the

co-expression vector, encoding both MRPP1 and MRPP2, as template. The mutant proteins were expressed and purified as described for wildtype and proved to be soluble and stable against degradation. SDS-PAGE analysis of the final purified sample of the MRPP1 mutants (with p.D314N or p.Q226A) in complex with MRPP2 is shown in **Figure 3-15** next to a sample of wildtype MRPP1 in complex with MRPP2.

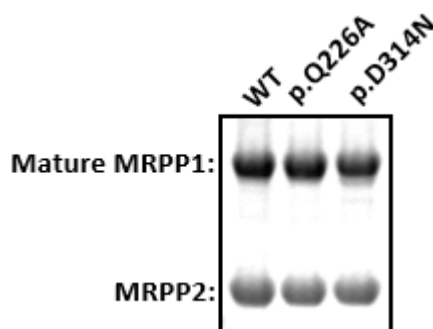


Figure 3-15. SDS-PAGE analysis of the final purified sample of MRPP1 wildtype, D314N, and Q226A co-expressed with MRPP2.

The MTase activity assay was carried out on the wildtype and mutant proteins using *in vitro* transcribed pre-(mt)tRNA^{Ile} with a guanine base in position 9, and pre-(mt)tRNA^{Val} with an adenine base in this position. The MTase assay tested the hypothesis that mutation of the presumed catalytic base Asp314 would abolish both m¹G9 and m¹G9 activity, while mutation of Gln226, if it were to function as an intermediate charge-stabiliser, would affect only m¹G9 activity.

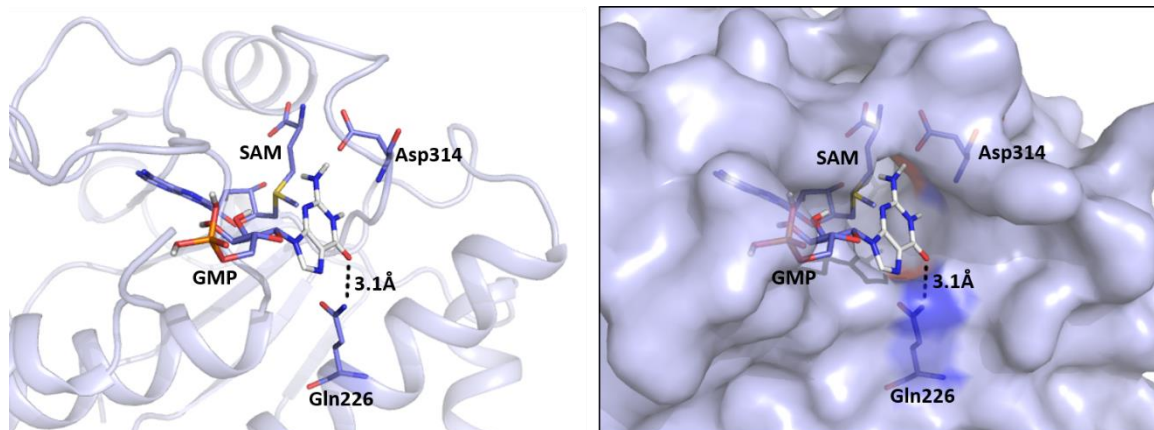


Figure 3-16. The potential catalytic base, Asp314, and intermediate charge-stabiliser, Gln226, are shown along with the SAM ligand and a docked 5'-GMP molecule mimicking the guanosine to be methylated and the backbone from the tRNA.

As expected, the MRPP1(D314N)-MRPP2 complex, containing the presumed catalytic base mutation, abolished m¹A9 and m¹G9 activities, compared to the wildtype MRPP1-MRPP2 control (**Figure 3-17**). The complex of MRPP1(Q226A)-MRPP2, containing the presumed intermediate charge-stabiliser mutation, also showed no m¹G9 activity. Surprisingly, the p.Q226A mutation also abolished m¹A9 activity of the complex (**Figure 3-17**), suggesting that this residue is implicated in both mechanisms rather than having the proposed role of intermediate charge-stabiliser solely for guanine. The essential requirement for both Asp314 and Gln226 in MRPP1 catalysis suggests that the adenine methylation proceeds via a mechanism distinctively different from the mechanism utilised by the structurally unrelated MTase TrmI. The mechanism of MRPP1 is further discussed in **Chapter 6**.

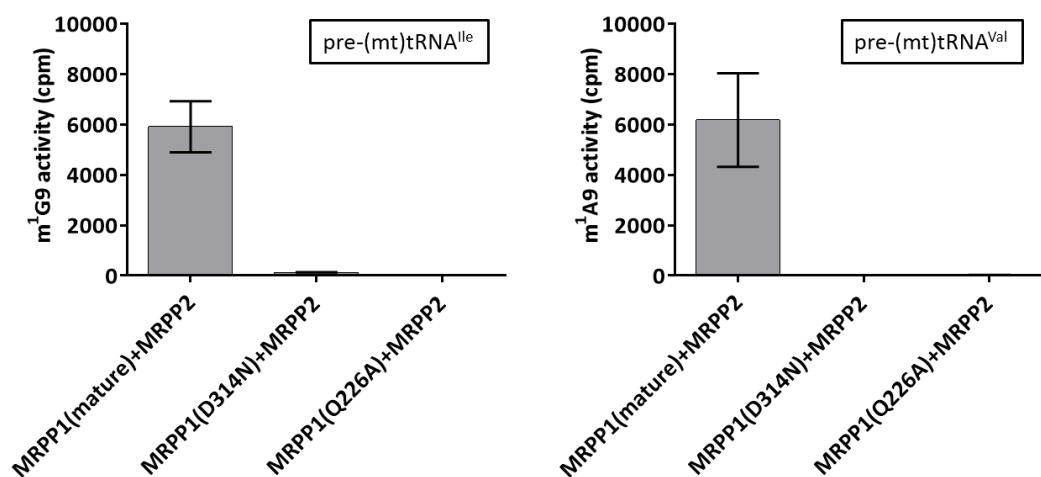


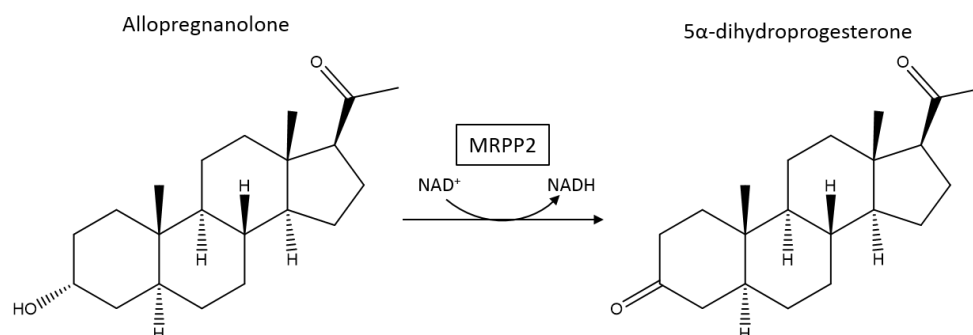
Figure 3-17. N1-methylation by the binary complex of MRPP1-MRPP2 with wild-type, D314N and Q226A MRPP1 on pre-(mt)tRNA^{Ile} and pre-(mt)tRNA^{Val}. For this experiment $n = 3$.

3.3.3 Investigation into the role of MRPP1 in MRPP2 SDR activity

Having established that the MTase activity of MRPP1 depends on MRPP2, an activity study was performed to determine if the presence of MRPP1 would influence the SDR function of MRPP2. Earlier data established that MRPP2 in the MRPP1-MRPP2

complex was still able to accommodate the cofactor for SDR activity and that the complex did not dissociate in the presence of excess amounts of either the cofactor or a substrate for SDR activity (allopregnanolone). This suggested that that MRPP2 may be able to perform its SDR functions even in the presence of MRPP1.

Allopregnanolone is a neurosteroid and a natural substrate for MRPP2 which, via its 3 α -HSD activity (**Section 1.1.1.1**), oxidises this to form 5 α -dihydroprogesterone (5 α -DHP) using NAD⁺ as cofactor (**Reaction Scheme 1**).



Reaction Scheme 1. Conversion of allopregnanolone to 5 α -dihydroprogesterone (5 α -DHP) by MRPP2 using NAD⁺.

The MRPP2 alone and MRPP1-MRPP2 co-expressed proteins were used in a spectrophotometric activity assay measuring the fluorescence (350 nm excitation and 450 nm emission) from NADH derived during catalysis. The assay was executed with equimolar amounts of MRPP2 protein alone and in complex, in the presence of excess NAD⁺. The allopregnanolone substrate and NAD⁺ cofactor were mixed and diluted with assay buffer before addition of protein. The substrate was used at final concentrations ranging from 100 to 0.05 μ M, with the concentration of NAD⁺ and protein held constant. Kinetic parameters, K_M and V_{max} , were obtained by fitting the data to the Michaelis-Menten equation (**Figure 3-18A**). To obtain Michaelis-Menten kinetics, the system must adopt steady-state between substrate, enzyme and the substrate-enzyme complex ($E + S \rightleftharpoons ES$) [174]. This is not the case for the above system at the lower concentrations (below

3.7 μM = $[E]$) where the system exhibit $[E] > [S]$. The kinetic constants are therefore estimates that should be utilised only to compare enzymes tested in this work rather than compared between publications. As it is unknown how many active sites of MRPP2 are accessible for catalysis due to complex formation, no k_{cat} values were determined. A fraction of the complex protein sample used in the spectrophotometric activity assay was applied for analytical SEC with SDS-PAGE analysis to ensure that no uncomplexed MRPP2 protein was present to perform the catalysis on its own (**Figure 3-18B**).

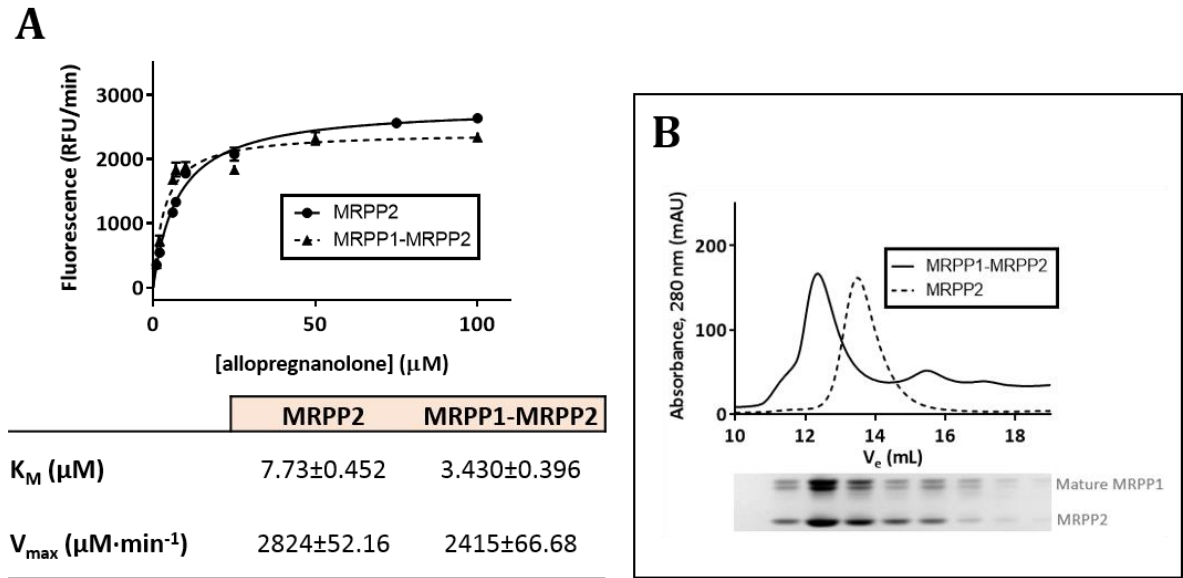


Figure 3-18.A) Michaelis-Menten plot for the catalysis of allopregnanolone by MRPP2 using NAD^+ . Both complexed (with MRPP1) and uncomplexed (alone) MRPP2 was tested. RFU: relative fluorescence units. B) Analytical SEC chromatogram and SDS-PAGE analysis of the purified complex used for activity studies.

The SDR activity assay revealed a similar V_{max} between free and complexed MRPP2 indicative of that no active sites are blocked by the interaction with MRPP1. A 2-fold reduced K_M -value is observed for the complex, compared to the MRPP2 protein alone, suggesting that MRPP1 has a stabilising effect on the enzymatic activity of MRPP2. The analytical SEC chromatogram and SDS-PAGE analysis showed MRPP1 and MRPP2 proteins eluting with a constant complex-stoichiometry, indicating that no uncomplexed MRPP2 was present. These data showed that the presence of MRPP1 does not hinder MRPP2 in performing its 3α -HSD function and potentially other SDR functions.

3.4 Structural studies

Using either individually expressed or co-expressed MRPP1 and MRPP2 proteins, efforts were made to obtain structural information on the complex, with and without (mt)tRNA, using X-ray crystallography and the SAXS technique.

3.4.1 X-ray crystallography

3.4.1.1 Crystallisation of MRPP1-MRPP2

The first target for crystallisation was the MRPP1-MRPP2 complex reconstituted from individually expressed components at the lysis stage. The protein complex was concentrated to 9.5 mg/mL (calculated for a 2xMRPP1:4xMRPP2 stoichiometry) and used in crystallisation trials in five random matrix screens: HCS, JCSG, LFS, BCS and HIN, at 4°C and 20°C, with different protein-to-precipitant ratios (1:2, 1:1, 2:1). This procedure yielded crystals in one drop, with the condition: 50% MPD, 0.2 M ammonium phosphate dibasic, 0.1 M Tris pH 8.5. However, these crystals gave rise to diffraction patterns characteristic of small molecules or salts. Subsequent strategies to facilitate crystallisation by supplementing the protein sample with SAM ligand, and/or performing *in situ* proteolysis (trypsin-to-protein ratio of 1:100) [175], were unsuccessful.

After cloning of the MRPP1-MRPP2 co-expression plasmid, new crystallisation trials were performed with the co-expressed protein complex at 6.5 mg/mL. This yielded crystals after 1-3 days of incubation in one condition (**Figure 3-19**): 50% PEG200, 0.2 M MgCl₂, 0.1 M cacodylate pH 6. These crystals showed no diffraction at the Diamond synchrotron. Optimisation of crystals grown in this condition is planned as future work.



Figure 3-19. Crystals formed in a drop from crystal plates set up with co-expressed MRPP1-MRPP2 protein complex. Condition: 50% PEG200, 0.2 M MgCl_2 , 0.1 M cacodylate pH 6.5 grew after 3 days of incubation.

Until this point, all crystallisation attempts for the MRPP1-MRPP2 complex had been performed at protein concentrations below 10 mg/mL. To survey the crystallisation space above this concentration, the co-expressed complex was next concentrated to 39 mg/mL, which was the highest concentration that could be reached with still leaving enough sample to set up crystallisation plates. These trials yielded small crystals in three conditions, grown at 20°C with a 1:2 protein-to-precipitant ratio (**Figure 3-20**).

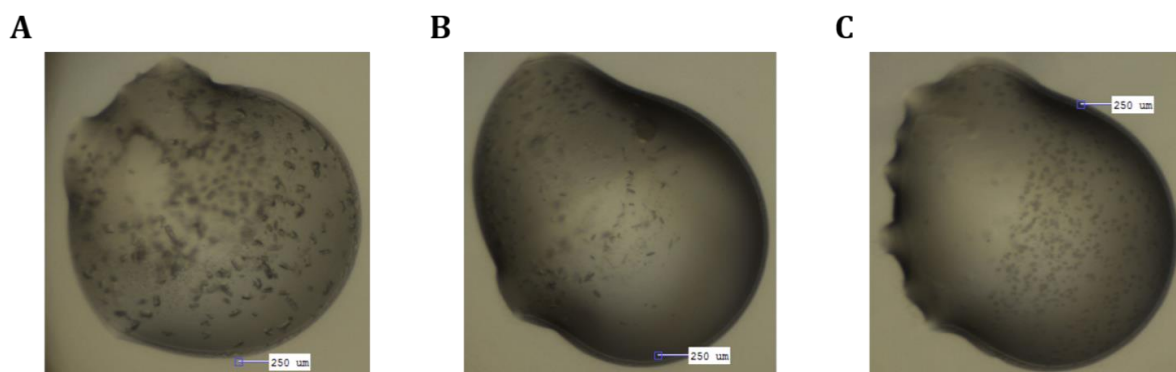


Figure 3-20. Crystal drops from plates set up with co-expressed MRPP1-MRPP2 complex at 39 mg/mL. A) 0.2 M ammonium acetate, 0.1 M citrate pH 5.6, 30% PEG4000, B) 0.05 M CaCl_2 , 0.1 M MES pH 6.0, 20% PEG6000, 10% ethylene glycol, C) 0.2 M NH_4Cl , 0.1 M MES pH 6.0, 20% PEG6000, 10% ethylene glycol.

A follow-up screen was designed for the larger crystals (**Figure 3-20A**) surveying the immediate chemical environment around the initial hit. Crystals appeared as blocks in conditions with low concentrations of ammonium acetate (0.1 M) and pH 5.4-5.6 (**Figure 3-21B**), which were larger than crystals formed in the initial screen (**Figure 3-21A**).

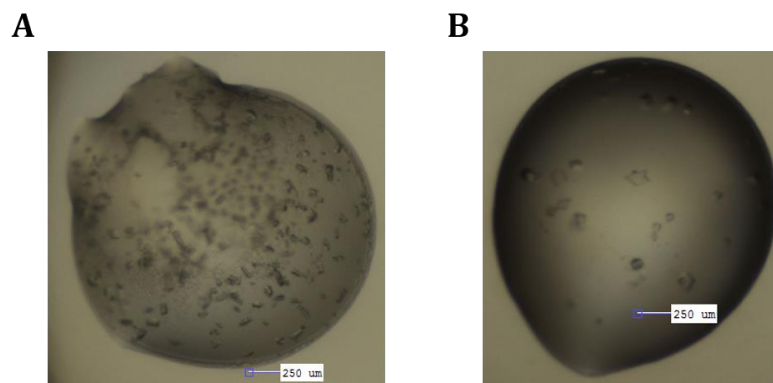


Figure 3-21. Crystals grown in a drop with the co-expressed MRPP1-MRPP2 protein complex. A) initial hit grown in 0.2 M ammonium acetate, 0.1 M citrate pH 5.6, 30% PEG4000, B) the follow-up hit grown in 0.1 M ammonium acetate, 0.1 M citrate pH 5.6, 34% PEG4000.

Crystals were mounted with no cryo-protectant or cryo-protected in mother liquor supplemented with 5% propylene glycol. The crystals were screened at the microfocus beamline I24 of Diamond Light Source, due to their relatively small sizes, and demonstrated better diffraction in the absence of cryo-protectant. A dataset was collected to 5.03 Å resolution on a crystal without cryo-protectant at the same beamline. The diffraction data were indexed to the $P2_12_12_1$ spacegroup based on absent reflections with even indexing along each axis ($h00$, $h=2n$; $0k0$, $k=2n$, $00l$, $l=2n$) with *no* strong reflections at odd indexing (Figure 3-22).

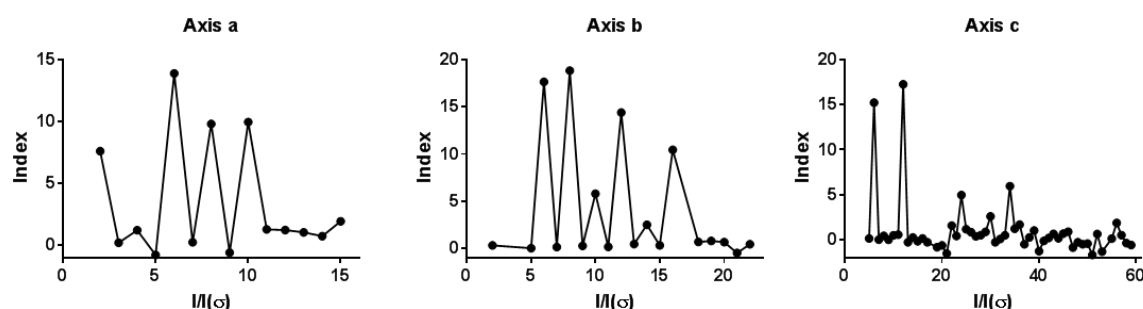


Figure 3-22. The screw axis along each axis (a, b and c) shown as Index vs. $I/I(\sigma)$ outputted from POINTLESS.

Cell parameters of $a = 79.79$ Å, $b = 125.54$ Å and $c = 299.18$ Å (α , β , $\gamma = 90^\circ$, 90° , 90°) were very different (e.g. larger c dimension) to those previously characterised for MRPP2 alone, thus these crystals could represent MRPP1-MRPP2 complexes. In the absence of experimental phases, molecular replacement was attempted with the MRPP2

structure (PDB: 1U7T) as search model. The output solution placed three MRPP2 molecules in the *ASU*. The log-likelihood-gain (LLG) and rotation and translation function Z-scores (RFZ and TFZ, respectively) from Phaser [176] for each placed molecule, before and after refinement, are shown in **Table 3-1**. The LLG-value increases with each placed molecule and the final TFZ-value > 7, indicating a correct solution.

	Before refinement			After refinement	
	LLG	RFZ	TFZ	LLG	TFZ
Tetramer 1	209	10.1	11.8	272	13.1
Tetramer 2	585	*	22.5	608	21.7
Tetramer 3	500	5.9	6.8	647	9.2

Table 3-1. The LLG- and Z-scores from Phaser for each placed molecule of MRPP2 before and after refinement. *The second tetramer was found using rotation peaks from the first search as there were more than one solution with good Z-scores (>8.0).

The solution did not seem to leave room for any molecules of MRPP1 (**Figure 3-23**).

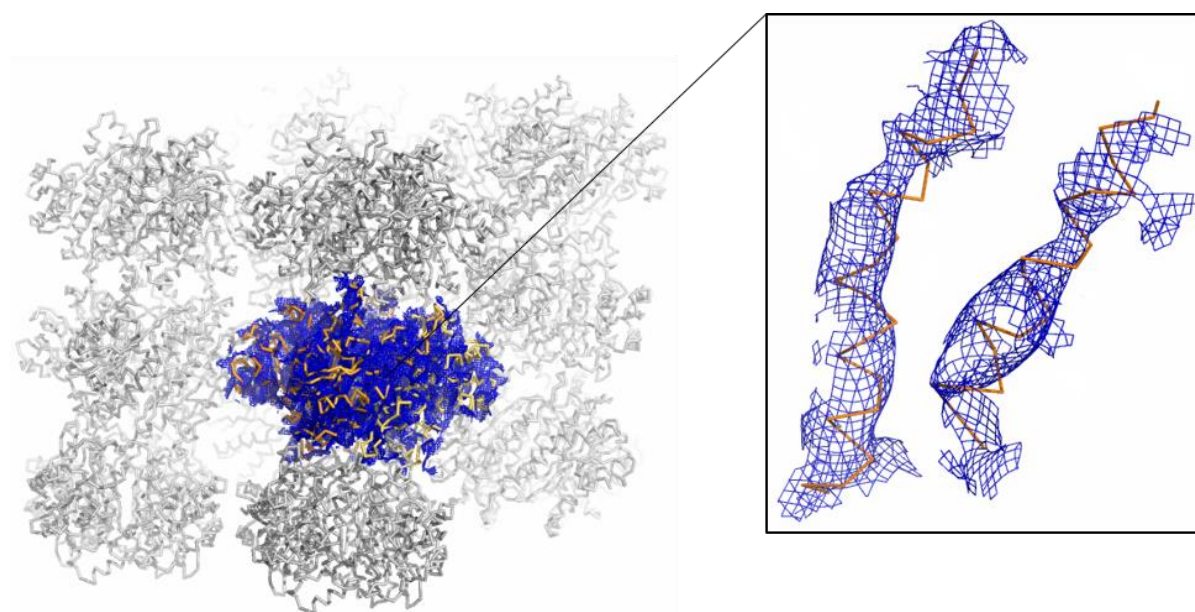


Figure 3-23. Unrefined crystal structure of MRPP2 to 5 Å resolution. The electron density map 2Fo-Fc for two helices of one MRPP2 monomer is shown contoured at 0.8 σ .

The formation of MRPP2-alone crystals could suggest either excess MRPP2 from purification, or complex dissociation under the crystallisation conditions. In case of dissociation, this was attempted avoided by incubating the MRPP1-MRPP2 sample

(at 26 mg/mL), with 1.5-fold excess of SAM and an equimolar amount of *in vitro* transcribed pre-(mt)tRNA^{lle}, prior to setting up crystallisation plates. These crystallisation screens yielded initial nucleation from crystallisation plates incubated at 4°C, in two precipitant conditions (**Figure 3-24**). Optimisation from nucleation is planned as future work.

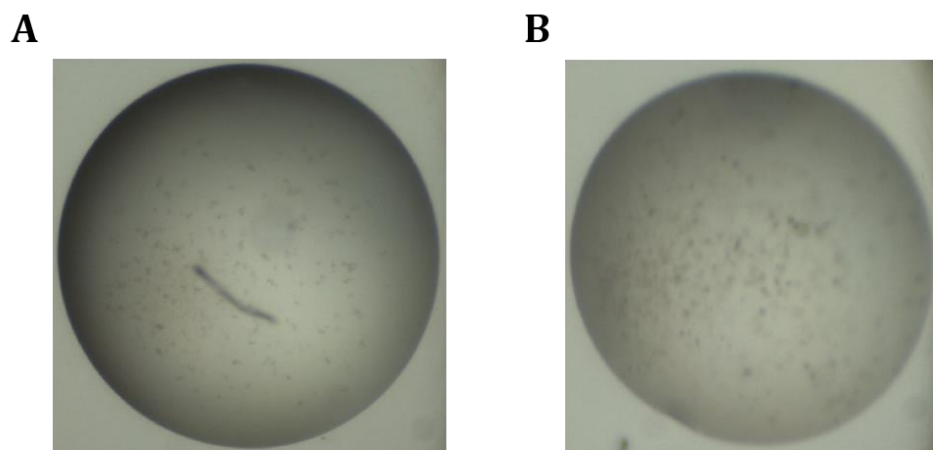


Figure 3-24. Crystals grown from co-expressed MRPP1-MRPP2 protein added with SAM and pre-(mt)tRNA^{lle}. A) 22% polyacrylic acid 5100, 0.02 M MgCl₂, 0.1 M HEPES pH 7.5 and, B) 20% PEG3350, 0.2 M potassium citrate tribasic.

3.4.2 SAXS

3.4.2.1 SAXS model of MRPP1-MRPP2

In the absence of X-ray structures for the complex, the low resolution method SAXS was applied. In these studies, co-expressed MRPP1-MRPP2 proteins were used in which MRPP2 is untagged. As a control sample for SAXS, untagged MRPP2 protein was generated using a different construct (sub-cloned into p11 vector by Naeem Shafqat at SGC) from that described in **Section 3.1**, to facilitate His₆-tag removal. The co-expressed and purified MRPP1-MRPP2 complex, and untagged MRPP2 protein alone, were applied for in-line SAXS analysis at the Diamond Light Source beamline B21 in a phosphate buffer with 1% sucrose and scattering data were collected (**Figure 3-25**).

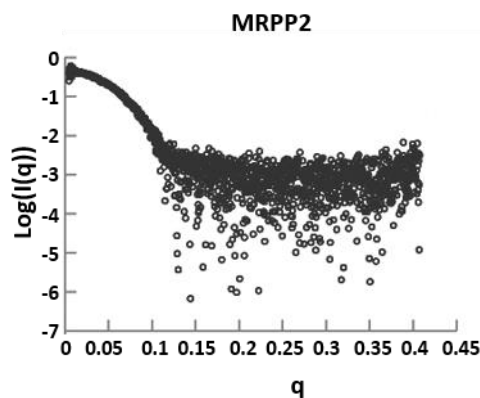


Figure 3-25. SAXS scattering plot: $\log(I(q))$ vs. q , for MRPP2.

The obtained and processed scattering data for the MRPP2 protein alone, resulted in a linear Guinier fit (**Figure 3-26A**) giving rise to an R_g for the protein of 28.44 ± 0.4415 Å. The Kratky plot (**Figure 3-26B**) indicated a folded, globular species in solution as expected from the crystal structure of the MRPP2 tetramer. A $D_{\max}$ of 80 Å gave rise to the $P(r)$ distribution function shown in **Figure 3-26C**.

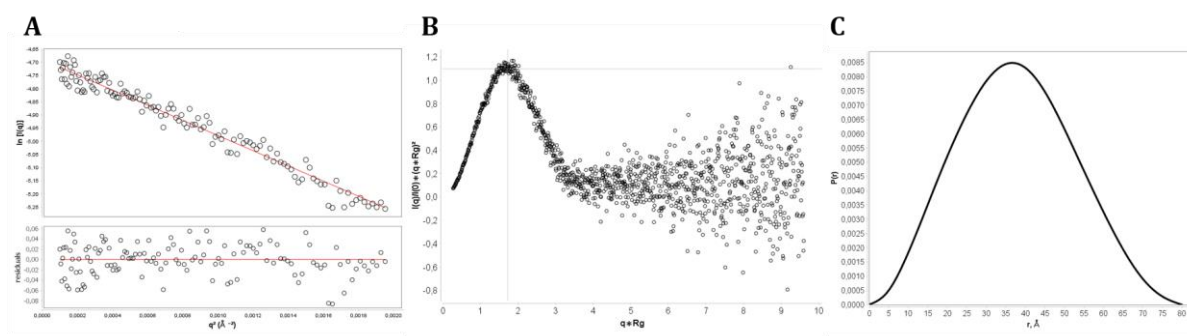


Figure 3-26. SAXS data untagged MRPP2 alone. A) Guinier plot, B) Kratky plot, C) $P(r)$ distribution function.

A back calculation of the theoretical scattering plot for MRPP2 based on its crystal structure showed a good fit (χ -value of 0.952 for q -values between 0.01° - 0.45°) with the experimentally determined scattering plot (**Figure 3-27**). A slight misfit was observed for higher q -values indicating some flexibility of the protein in solution.

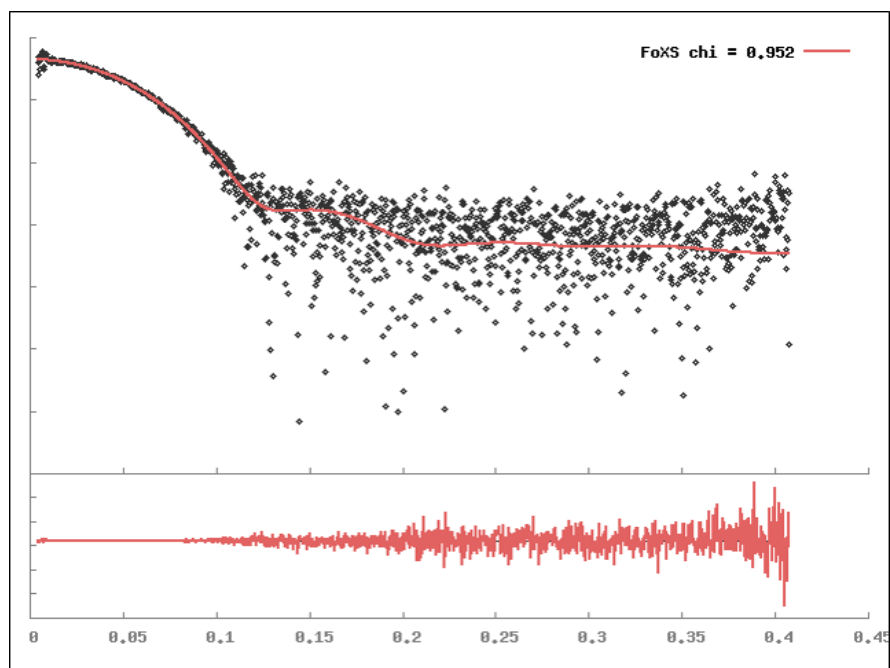


Figure 3-27. Comparison of the experimentally determined SAXS scattering plot ($\log I(q)$ versus q) of MRPP2 (black), with the calculated SAXS scattering plot ($\log I(q)$ versus q) from the tetrameric MRPP2 (PDB: 1U7T) crystal structure (red) computed with the Debye formula.

The crystal structure of tetrameric MRPP2 (PDB: 1U7T) was fitted to the SAXS model of MRPP2 (**Figure 3-28**) using the fit-in-map function in the Chimera software [165].

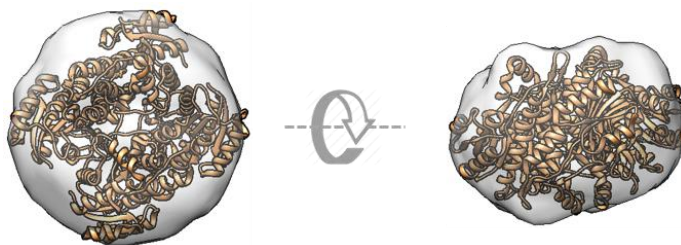


Figure 3-28. The SAXS model for tetrameric, full-length MRPP2. Fitted using Chimera.

Scattering data were collected for the complex of MRPP1-MRPP2 (**Figure 3-29**) using in-line SAXS analysis as for MRPP2.

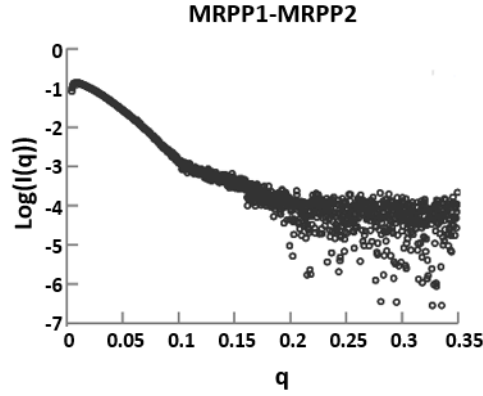


Figure 3-29. SAXS scattering plot: $\log(I(q))$ vs. q , for the MRPP1-MRPP2 complex.

The scattering data gave rise to a linear Guinier fit (**Figure 3-30A**) resulting in an R_g of the protein complex of 45.59 ± 0.9720 Å. The Kratky plot (**Figure 3-30B**) indicated a folded, partially flexible protein in solution. The $D_{\max}$ of 153 Å gave the $P(r)$ distribution function shown in **Figure 3-30C**. The R_g and $D_{\max}$ values of the complex are larger than those obtained for the individual proteins (MRPP1: $R_g = 34.40 \pm 1.885$ Å, $D_{\max} = 100$ Å and MRPP2: $R_g = 28.44 \pm 0.4415$ Å, $D_{\max} = 80$ Å). The MW of the MRPP1-MRPP2 complex was calculated from the SAXS scattering curve to 214 kDa, agreeing with the previously identified 4xMRPP2:2xMRPP1 stoichiometry for a theoretical MW of 208 kDa.

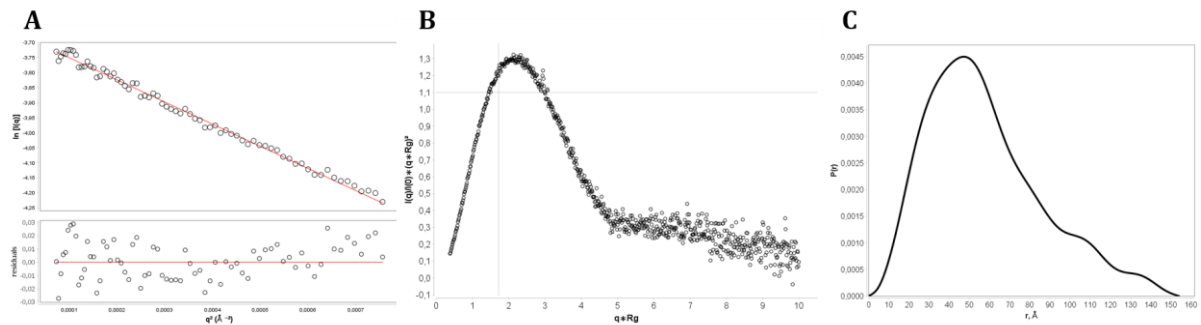


Figure 3-30. SAXS data for the MRPP1-MRPP2 complex. A) Guinier plot, B) Kratky plot, C) $P(r)$ distribution function.

SAXS models for MRPP1-MRPP2 were generated with $P1$, $P2$ and $P4$ point symmetry to account for all possible tetramer-dimer symmetry related binding modes (**Appendix, Section 8.6**). An increase in symmetry from $P1$ to $P2$ or $P4$ increased the normalised spatial discrepancy (NSD) between initial *ab initio* models, demonstrating that

enforcing higher symmetry generates models that are more frequently systematically different from one another. The final χ^2 for the fit between the model and raw data was 1.712 (*P1*), 1.683 (*P2*) and 1.743 (*P4*). The SAXS models for mature MRPP1 and MRPP2 were fitted to the three SAXS models of the MRPP1-MRPP2 complex (**Appendix, Section 8.6**). In this fit, only the MRPP1-MRPP2 *P1*-model could accommodate mature MRPP1 as a dimer while, at the same time, accounting for the full SAXS pocket. The *P1*-model was therefore chosen for further studies. The obtained SAXS *P1*-model for the MRPP1-MRPP2 complex is shown in **Figure 3-31** with the crystal structure of tetrameric MRPP2 fitted in using the Chimera software [165]. The model is shown alongside that of mature MRPP1 and MRPP2.

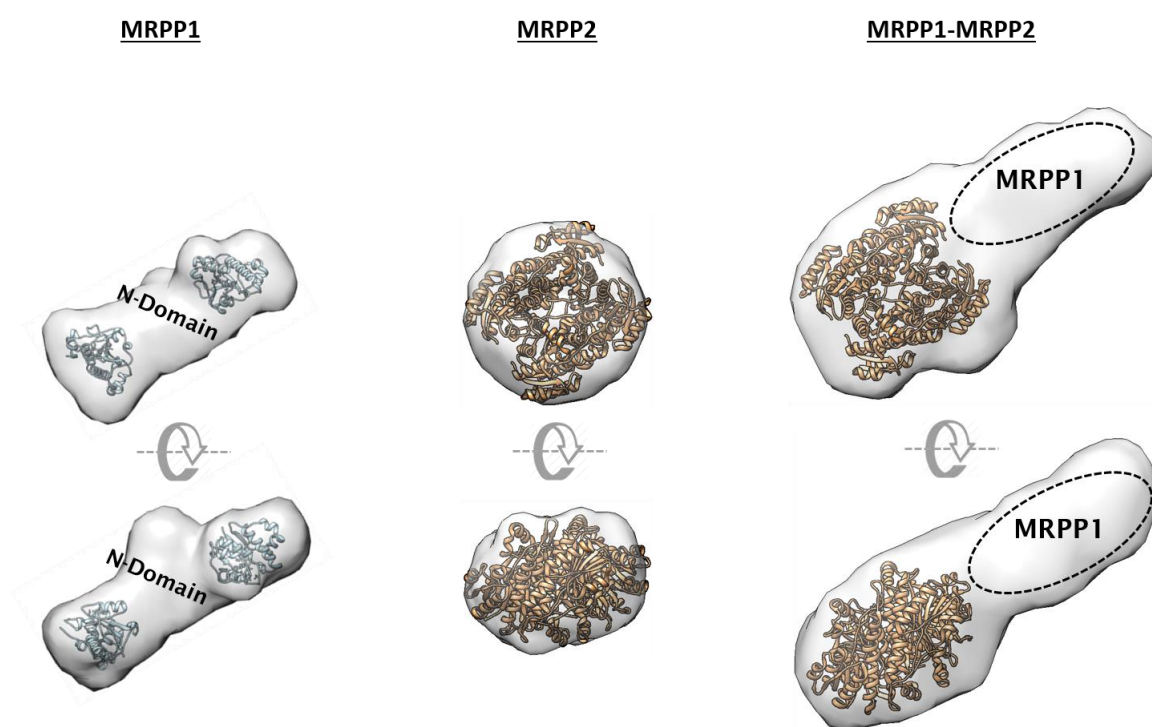


Figure 3-31. SAXS models for mature MRPP1, full-length MRPP2 and the binary complex of MRPP1-MRPP2 built in *P1* symmetry. Crystal structures are fitted to the SAXS models using Chimera.

Both the crystal structure and the SAXS model of MRPP2 fit nicely into one region of the complex model, and the remainder of the model shows an extended shape that could accommodate the mature MRPP1 protein. Fitting of the MRPP1 MT-domain crystal

structure into the model is discussed in **Chapter 6**. When comparing the SAXS model for mature MRPP1 protein alone, with that for MRPP1-MRPP2, it seems that the MRPP1 protein within the complex becomes more compact along the long axis and perhaps this compaction could be the cause of the protease-protective effect of MRPP2 on MRPP1 when complexed (**Section 3.2.2**).

3.4.2.2 SAXS model of MRPP1-MRPP2-pre-(mt)tRNA^{Ile}

To investigate the binding of (mt)tRNA to the complex using SAXS, the MRPP1-MRPP2 complex was mixed with excess *in vitro* transcribed pre-(mt)tRNA^{Ile} in buffer conditions that facilitate tRNA binding (low NaCl concentrations) and stability (presence of Mg²⁺). The final mixture was applied for in-line SAXS analysis at the Diamond Light Source beamline B21 and scattering data were obtained (**Figure 3-32**).

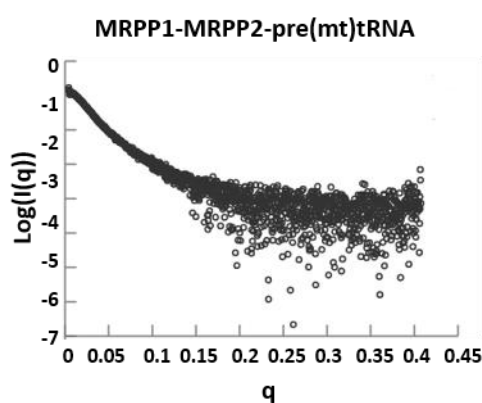


Figure 3-32. SAXS scattering plot: $\log(I(q))$ vs. q , for the MRPP1-MRPP2-pre-(mt)tRNA^{Ile} complex.

The scattering data for the MRPP1-MRPP2-pre-(mt)tRNA^{Ile} complex resulted in a linear Guinier fit (**Figure 3-33A**) giving rise to an R_g for the complex of 69.67 ± 2.766 Å. The Kratky plot (**Figure 3-33B**) indicated a folded, but non-globular, extended species in solution. The $D_{\max}$ of 216 Å gave rise to the $P(r)$ distribution function shown in **Figure 3-33C**.

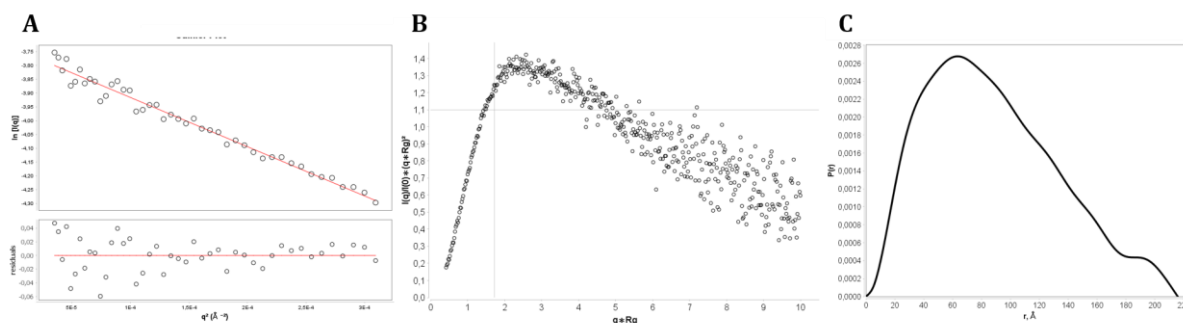


Figure 3-33. SAXS plots for in-line SAXS of MRPP1-MRPP2-pre-(mt)tRNA^{Ile}. A) Guinier plot, B) Kratky plot, C) P(r) distribution function.

The obtained SAXS model for the MRPP1-MRPP2-pre-(mt)tRNA^{Ile} complex is shown in **Figure 3-34** alongside that for the MRPP1-MRPP2 complex. As expected for a larger species, the R_g and $D_{\max}$ values for the MRPP1-MRPP2-pre-(mt)tRNA^{Ile} complex ($R_g = 69.67 \pm 2.766$ Å, $D_{\max} = 216$ Å) are larger than those for the MRPP1-MRPP2 complex ($R_g = 45.59 \pm 0.9720$ Å, $D_{\max} = 154$ Å). When overlaying the SAXS model for MRPP1-MRPP2 within that for MRPP1-MRPP2-pre-(mt)tRNA^{Ile}, the remaining space accounts for one pre-(mt)tRNA molecule. From the overlay it seems that MRPP1 undergoes conformational changes upon binding of the pre-(mt)tRNA substrate which is in agreement with studies on the Trm10 homologues of MRPP1 from yeast [49]. These studies showed that the free Trm10 proteins exist in a more open and extended conformation in solution, compared to when they are bound with the tRNA substrate. Conformational changes have also been observed for other SPOUT MTases upon tRNA binding [139]. The interesting suggestion that the SAXS model accommodates *one* pre-(mt)tRNA molecule per MRPP1(dimer)-MRPP2(tetramer)-SAM (2 molecules) is further discussed in **Chapter 6**.

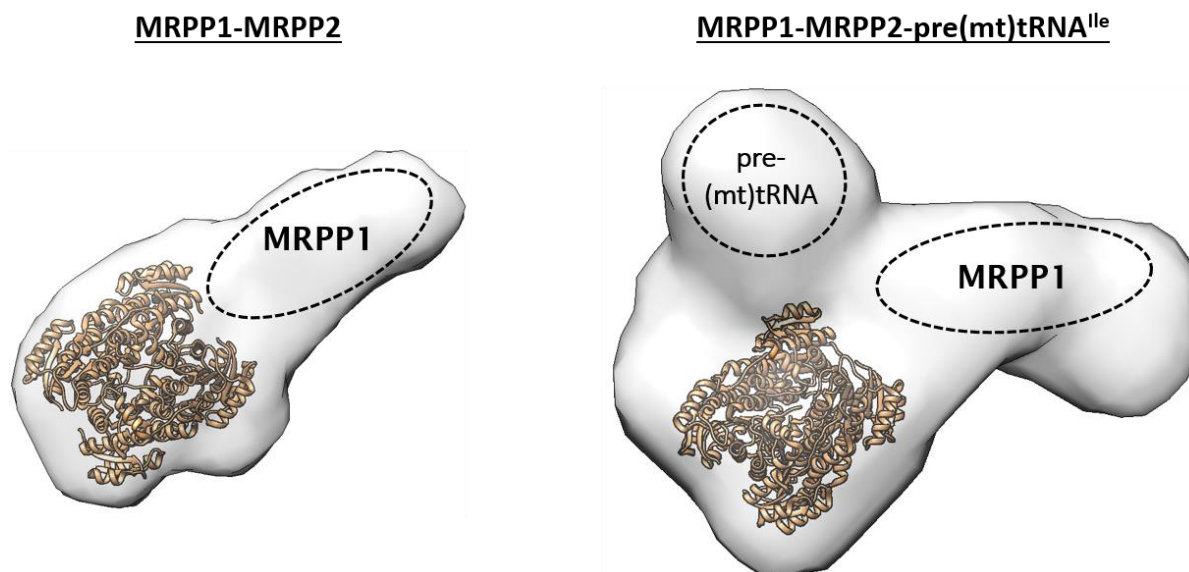


Figure 3-34. SAXS model for the complex of MRPP1-MRPP2-pre-(mt)tRNA^{Ile} shown alongside the SAXS model of the binary complex of MRPP1-MRPP2.

3.5 tRNA binding

To further examine the binding of (mt)tRNA to the MRPP1-MRPP2 complex an electrophoretic mobility shift assay (EMSA) was performed looking for shifts in its electrophoretic mobility caused by an interaction of protein with the *in vitro* transcribed pre-(mt)tRNA^{Ile} substrate.

The EMSA assay was first performed with the binary complex of MRPP1-MRPP2 as a single point experiment, and as a titration of increasing protein-to-(mt)tRNA ratio. From the NATIVE-PAGE gel in **Figure 3-35A**, stained for both RNA (**Lane 1-3**) and protein (**Lane 4-6**), it was observed that the binary complex of MRPP1-MRPP2 interacts with the pre-(mt)tRNA substrate, as indicated by a shift of the RNA stained band (**Lane 1** compared to **Lane 3**, **Figure 3-35A**). A titration of an increasing amount of complex protein to an excess amount of pre-(mt)tRNA (**Figure 3-35B**) further revealed a dose-dependent increase in complex formation.

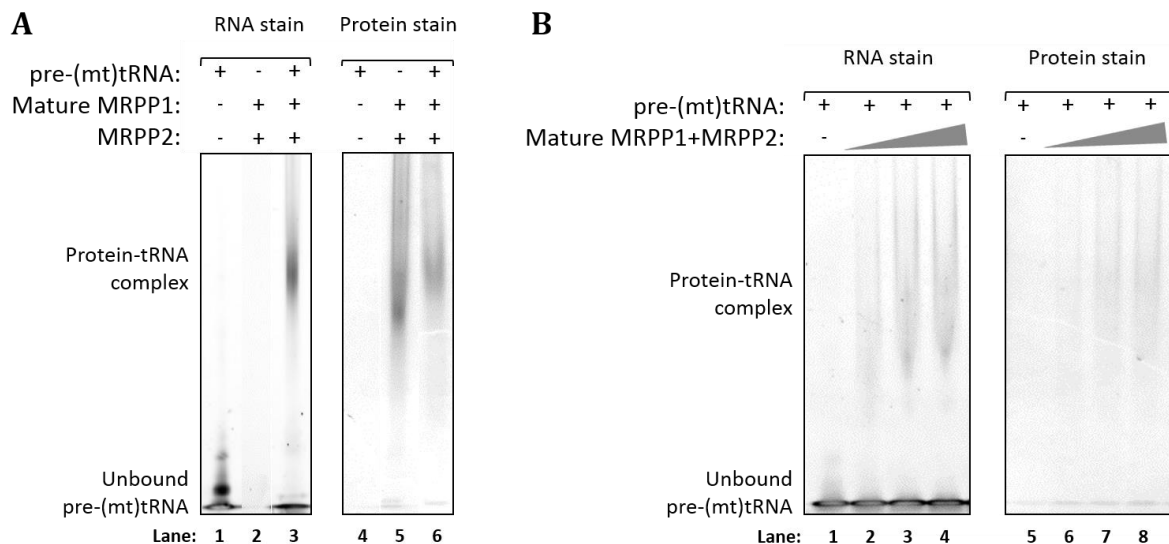


Figure 3-35. A) EMSA of pre-(mt)tRNA^{lle} and the MRPP1-MRPP2 complex. B) EMSA of pre-(mt)tRNA^{lle} titrated with increasing amount of MRPP1-MRPP2 complex. SYBRO GOLD staining for RNA (Lane 1-3) and Coomassie Brilliant Blue staining for protein (Lane 4-6).

The assay was subsequently performed to test if the N-domain of MRPP1, in the presence or absence of MRPP2, could interact with pre-(mt)tRNA^{lle}. Interestingly, **Figure 3-36** shows that the N-domain of MRPP1 alters the electrophoretic mobility of pre-(mt)tRNA^{lle} in the presence *and* absence of MRPP2, showing that this domain is able to interact with the pre-(mt)tRNA substrate independently of both MRPP2 and the MT+C-domain of MRPP1.

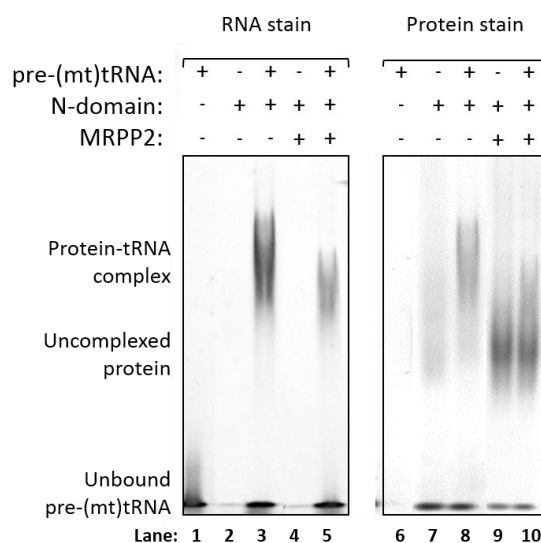


Figure 3-36. EMSA of pre-(mt)tRNA^{lle} with the N-domain of MRPP1, in the presence and absence of MRPP2. SYBRO GOLD staining for RNA (Lane 1-5) and Coomassie Brilliant Blue staining for protein (Lane 6-10).

The assay was next attempted with the MT+C-domain of MRPP1 in the presence or absence of MRPP2. However, despite numerous attempts, this protein construct failed to enter into the gel matrix, presumably due to the pI of this protein (predicted 9.55) being much higher than the pH of the gel/buffer system (pH 8.1). However, as the MT+C domain harbours the active site for MRPP1, it is likely engaging in some direct interactions with the pre-(mt)tRNA substrate during the catalytic cycle. This is confirmed by point mutation studies of the yeast and archaea Trm10 proteins showing that Lys- or Arg-to-Glu mutations of surface exposed residues in helix $\alpha 1$ of the MT-domain, conserved in MRPP1, abolish tRNA binding [49], [177].

The SAXS model for the MRPP1-MRPP2-pre(mt)tRNA^{Ile} complex placed MRPP2 in vicinity of the tRNA substrate. Previous data obtained *in vitro* [14] and *in vivo* [75] have been contradictory on MRPP2s ability to bind (mt)tRNA in the absence of MRPP1. To identify any (mt)tRNA binding properties of recombinant MRPP2 *in vitro*, the EMSA was performed on MRPP2 alone with pre-(mt)tRNA^{Ile}. From the NATIVE-PAGE gel (**Figure 3-37**), stained for RNA (**Lane 1-3**) and protein (**Lane 4-6**), it was observed that no pre-(mt)tRNA substrate was bound to MRPP2 alone in this experimental set up, as indicated by the lack of shift of the RNA stained band (**Lane 1** compared to **Lane 3, Figure 3-37**). That does not, however, exclude, that MRPP2 forms direct interactions with the tRNA substrate upon complex formation with MRPP1. The interaction of MRPP2 with tRNA is further discussed in **Chapter 6**.

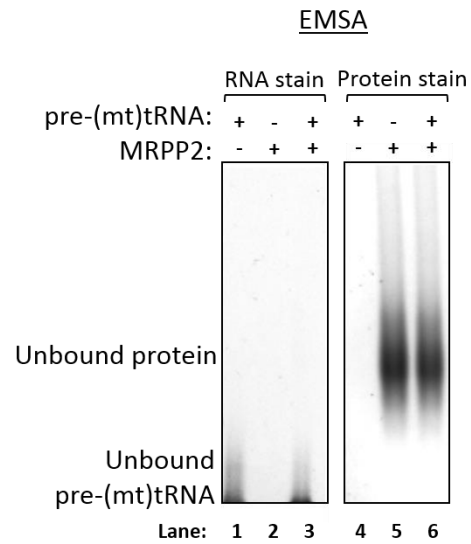


Figure 3-37. EMSA with full-length MRPP2 and pre-(mt)tRNA^{le}. SYBRO GOLD staining for RNA (Lane 1-3) and Coomassie Brilliant Blue staining for protein (Lane 4-6).

4 CHAPTER

CHARACTERISATION OF MRPP3 ALONE

At the start of this project no structural information was available on the Mg^{2+} dependent endoribonuclease MRPP3, which is the catalytic subunit of the ternary RNase P complex of MRPP1-MRPP2-MRPP3, and thus contains the 5'-end pre-(mt)tRNA processing activity (RNase P activity). Furthermore, no information was available on the interaction between this subunit and either the binary complex of MRPP1-MRPP2 or the pre-(mt)tRNA substrate. In the first part of the chapter, the MRPP3 protein is being characterised alone using structural and biochemical methods.

4.1 Domain organisation

MRPP3 belongs to the Protein-only RNase P (PRORP) protein subfamily of Nedd4-YacP nucleases (NYN's) and contains, in addition to the NYN-domain, an RNA-binding pentatricopeptide repeat (PPR) domain. PRORP proteins have so far been identified in higher and lower eukaryotes [94], with those from *A. thaliana*, *T. brucei*, *Drosophila* and human being the most well-studied [13], [90], [96], [97].

The PRORP protein in human mitochondria, MRPP3, is 583 residues long and contains, like MRPP1, an N-terminal cleavable MTS peptide (residues 1-50)[104]. The mature protein (residues 51-583) is predicted to comprise two PPR repeats covering residues 221-253, followed by a core-domain of unknown fold (residues 254-403) and an NYN-domain (residues 403-451). Flanking these domains are extensions at the N-terminus (N-extension, residues 51-219) and the C-terminus (C-extension, residues 452-583) [178] with unknown folds. MRPP3 is also predicted to bind a structural Zn^{2+} ion in a

Cys-Cys-His-Cys (CCHC) motif shared between the core-domain (Cys348-Cys351) and the C-extension (His557-Cys578) (**Figure 4-1**) [100], [178].

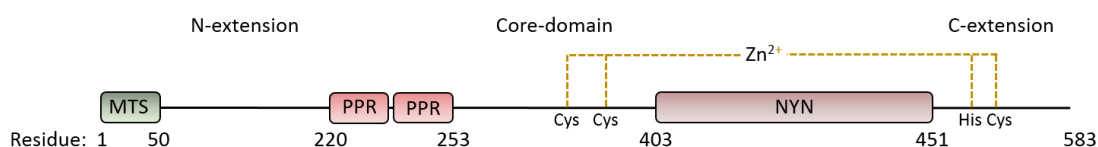


Figure 4-1 Predicted domain organisation of human MRPP3. MTS: mitochondrial targeting signal. PPR: pentatricopeptide repeat. NYN: Nedd4-YacP nuclease.

Using DISORPRED3 [145] via the PSIPRED server [146], protein disorder was predicted for in the N-extension (residues 51-109) of the mature protein (**Figure 4-2**) [145]. This region was further predicted to be involved in protein interactions, possibly with the MRPP1-MRPP2 complex.

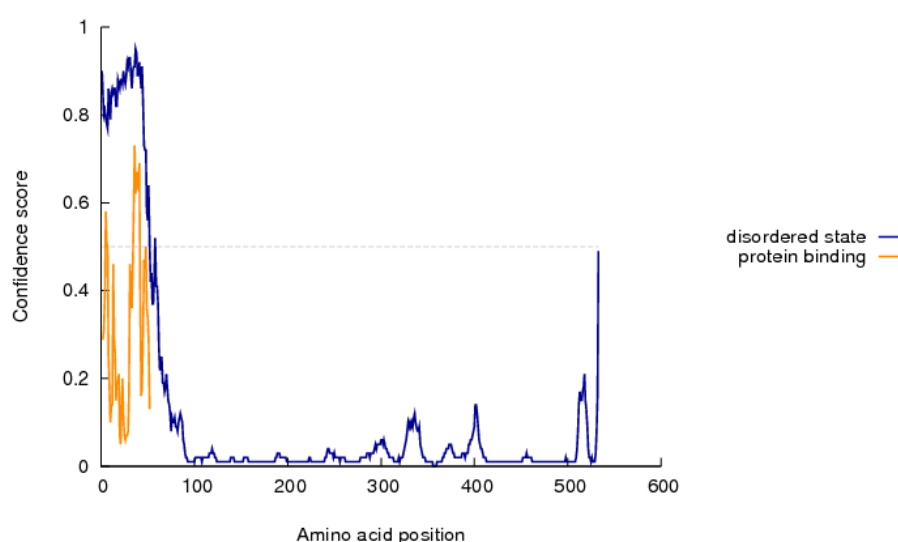


Figure 4-2. Intrinsic disorder profile for MRPP3 predicted by DISOPRED3 via the PSIPRED server. Amino acid position 0 corresponds to residue 51 in MRPP3.

4.2 Cloning, small-scale expression and purification

Based on the structural predictions above, several constructs were designed covering the mature MRPP3 protein (without the MTS peptide) and fragments of the PPR- and NYN-domains with and without the core-domain and N- and C-extensions (**Figure 4-3**). The constructs were cloned into the *E. coli* expression vector pNIC28-Bsa4 incorporating an N-terminal His₆-tag with a TEV protease cleavage site, and tested for expression and solubility in small-scale (50 mL cell cultures). The small-scale tests revealed that many

constructs with an intact C-terminus (residue 583) and varying N-terminal boundaries were soluble (**Figure 4-3**, green). Any truncation from the C-terminus rendered the protein insoluble, unless the entire core-domain, NYN-domain and C-extension were removed, resulting in a few short soluble constructs for the PPR-domain with part of the N-extension. No degradation was observed for any of the constructs, and the N-terminal His₆-tag was cleavable by TEV protease except for constructs with a start boundary at residue 274 for which only partial (~50%) protease cleavage was obtained even after long incubation with the protease.

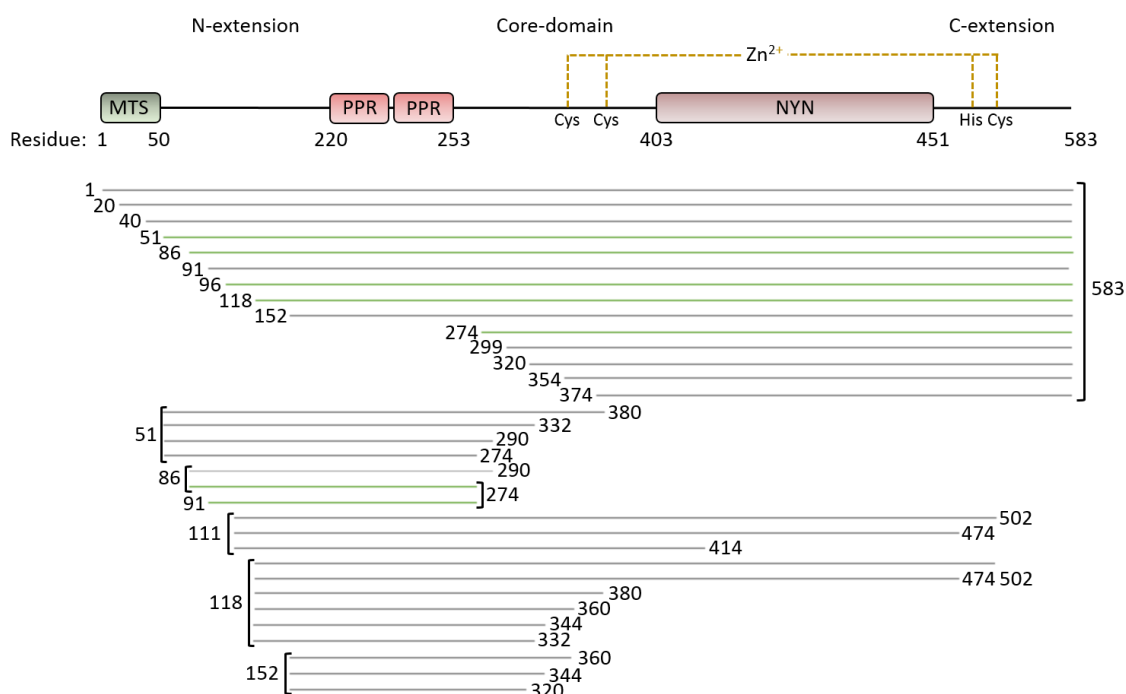


Figure 4-3. Cloned constructs of MRPP3. Soluble constructs: green. Insoluble: grey.

4.3 Large-scale expression and purification

The soluble construct of mature MRPP3 (residues 51-583) was expressed and purified in large-scale for usage in biochemical assays, and expression and cell lysis was performed as described for MRPP1 and MRPP2 proteins. The protein was purified (**Figure 4-4**) by IMAC (Ni-NTA resin) followed by SEC (S200 column) and cation-EX (HiTrap SP HP column). The protein was treated with TEV protease to remove the His₆-tag

and subsequently subjected to reverse IMAC. Tryptic digest-MS/MS and ESI-MS analysis identified the protein as intact (molecular mass: 61608.78 Da), untagged mature MRPP3 of residues 51-583.

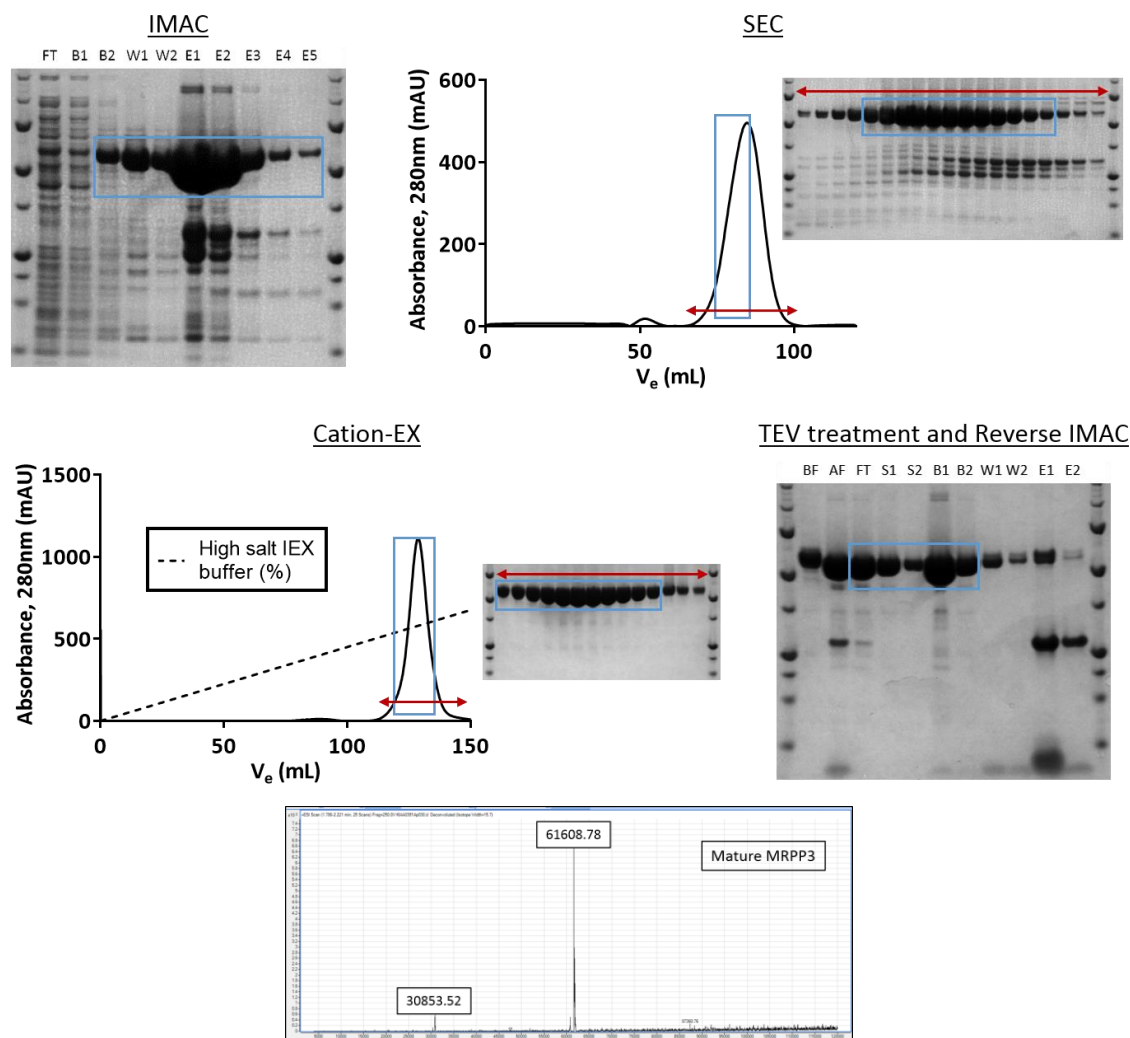


Figure 4-4. Large-scale purification of mature MRPP3. The example shows SDS-PAGE of initial IMAC, chromatogram from SEC and SDS-PAGE of peak fractions, chromatogram from cation-EX and SDS-PAGE of peak fractions, and SDS-PAGE of fractions from reverse IMAC. ESI-MS spectrum of a sample after purification is shown. Red arrows correspond to the region analysed on SDS-PAGE. Blue boxes indicate pooled fractions with the protein of interest. V_e : elution volume, FT: flow through, B1 and B2: binding buffer fractions, W1 and W2: wash buffer fractions, E1-5: elution buffer fractions. BF, AF: sample before and TEV protease treatment, respectively. S1 and S2: SEC buffer.

4.4 Solution studies

4.4.1 Oligomeric state

The purified, mature MRPP3 protein was applied to analytical SEC (S200) with the purpose of investigating the oligomeric state of the protein. The V_e was determined to

15.184 mL from the elution profile (**Figure 4-5**), which gave rise to an experimental MW of 45976.3 Da, corresponding to that of a monomeric protein (theoretical MW of monomer = 61608.7 Da).

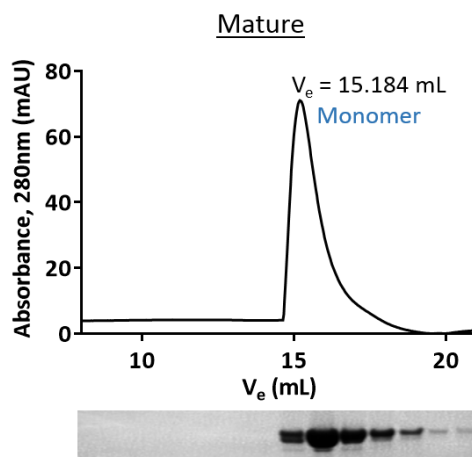


Figure 4-5. A) Analytical SEC chromatograms of mature MRPP3.

4.4.2 Ligand binding

The NYN-domain is a ribonuclease fold, proposed to be the catalytically active subunit of MRPP3. NYN-domains comprise four conserved acidic residues [100], involved in binding of a metal-ion utilised for cleavage of the phosphodiester bond of the RNA backbone. In MRPP3, these four residues corresponds to D409, D478, D479 and D499 [179], [180] and it is likely that the metal-ion utilised by MRPP3 for RNase P activity, Mg²⁺, binds to this site.

The metal-binding properties of mature MRPP3 was investigated using DSF by measuring the T_m of the protein ‘as-purified’ and after addition of Mg²⁺. The data (**Figure 4-6**) showed no change in protein T_m between the as-purified and Mg²⁺ added sample, suggesting that mature MRPP3 protein alone does not bind Mg²⁺-ions under the tested experimental conditions. If the metal-ion is, in fact, bound to the protein, the lack of T_m shift could be a result of no binding-induced thermostabilising effects, though metal-ions most often result in a ΔT_m. Alternatively, the protein in the as-purified sample could be *pre-saturated* with co-purified Mg²⁺ ions, though this is unlikely as a similar construct

(residues 46-583), expressed and purified as for mature MRPP3 in this work, was reported to be only 38.6% saturated with metal-ions after the initial IMAC step [92].

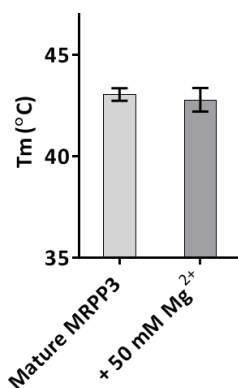


Figure 4-6. DSF assay of mature MRPP3 protein as-purified and upon addition of 50 mM Mg²⁺.

Even if the protein is not fully pre-saturated with Mg²⁺ ions from co-purification, it might still be metal-bound, as addition of the metal-chelater ethylenediaminetetraacetic acid (EDTA) to the DSF assay conditions, resulted in a change in Tm ($\Delta T_m = 4.87 \pm 0.01^\circ\text{C}$) (**Figure 4-7**). This change could be created by chelation of bound Mg²⁺ or, alternatively, the structural Zn-atom predicted to bind in MRPP3. An EDTA-chelation of the Zn-atom could force the protein into a more kinetically favourable fold that Zn-binding would otherwise prohibit and thereby cause the observed increase in Tm. With previous experiments showing only 38.6% saturation with metal-ions, and the possibility of EDTA removing the structural Zn-atoms, no further dialysis experiments were performed to ensure full removal of co-purified Mg²⁺ before DSF.

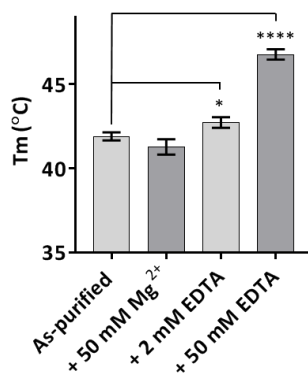


Figure 4-7. DSF assay of mature MRPP3 protein as-purified and upon addition of either 50 mM MgCl₂, 2 mM EDTA or 50 mM EDTA.

The thermal unfolding curve of the as-purified protein (**Figure 4-8**) reveals a prominent unfolding state around 55°C (**unfolding state 1**) and a subtle, second unfolding state around 75°C (**unfolding state 2**). This biphasic protein unfolding proposes two independent folding events for the MRPP3 structure, each constituting one or more of the predicted domains (**Figure 4-1**). A similar biphasic unfolding is observed for shorter MRPP3 constructs (see **Section 4.5.1.3**).

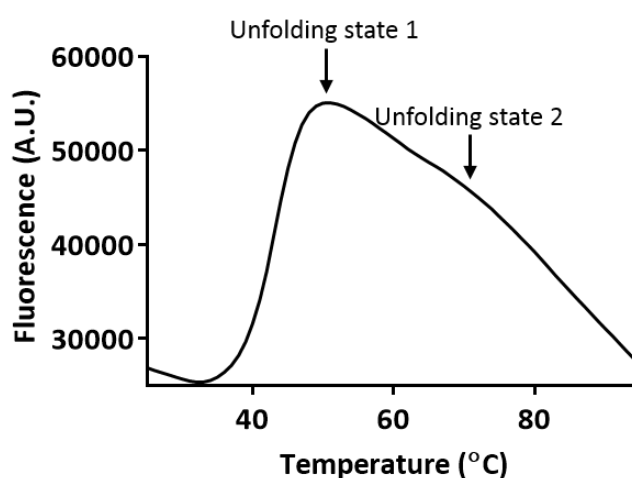


Figure 4-8. Thermal unfolding curve obtained with as-purified mature MRPP3.

4.5 Structural studies

4.5.1 Crystallisation attempts with MRPP3

In addition to biochemical assays, crystallisation trials for MRPP3 proteins were undertaken in an attempt to understand the protein structure. The cloned, soluble MRPP3 constructs outlined in **Figure 4-9** were expressed and purified in large-scale for crystallisation experiments. These include the mature MRPP3 protein and shorter constructs that will be denoted by their N-terminal truncation, *i.e.* $\Delta 85$ -MRPP3, $\Delta 95$ -MRPP3, $\Delta 117$ -MRPP3 and $\Delta 273$ -MRPP3.

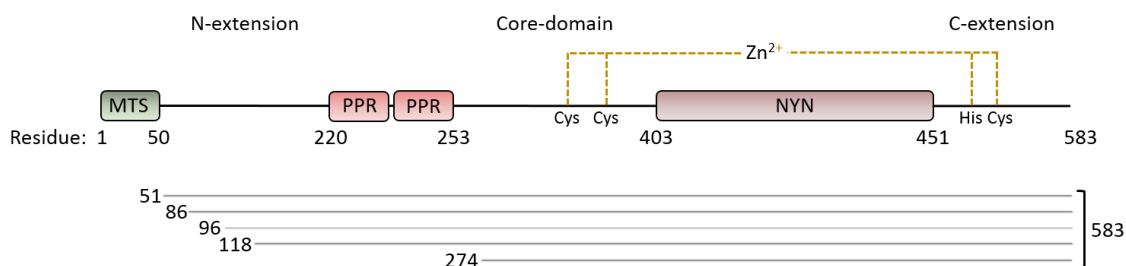


Figure 4-9. Cloned and soluble MRPP3 constructs used in crystallisation trials.

4.5.1.1 Mature MRPP3

The purified, untagged mature MRPP3 protein was concentrated to 5, 7, 10, 12.5, 20 and 32 mg/mL and used to set up 58 crystallisation plates of seven random matrix screens: JCSG, HCS, HIN, LFS, BCS, MIDAS and SaltRX at different temperatures and protein to precipitant ratios. After two months of incubation, no crystals were observed in any of the drops, and no particular trend could be deduced for the precipitation-and-clear drops patterns.

The DSF data suggested that mature MRPP3 did not bind Mg^{2+} ions, but as the DSF method measures changes to the protein thermal stability, rather than the ligand binding constant, the obtained DSF data cannot conclusively exclude metal-binding. Therefore, in the second round of crystallisation attempts, the protein sample was supplemented with Mg^{2+} and used to set up five random matrix screens: LFS, BCS, SaltRX, JCSG and MIDAS at different temperatures and protein-to-precipitant ratios. Unfortunately, the supplement of Mg^{2+} ions did not improve the crystallisation.

The observation that the mature construct is refractory to crystallisation was later confirmed for a similar construct covering residues 45-583 [179]. Crystallisation efforts were therefore next focused on shorter constructs of MRPP3.

4.5.1.2 $\Delta 85$ -MRPP3 and $\Delta 95$ -MRPP3

The soluble $\Delta 85$ -MRPP3 (**Figure 4-10**), $\Delta 95$ -MRPP3 (**Figure 4-11**) and $\Delta 117$ -MRPP3 (**Figure 4-12**) constructs were expressed and purified as for mature MRPP3,

and the identities and intact molecular masses were identified using tryptic digest-MS/MS and ESI-MS. Protein expression and solubility was lower for $\Delta 117$ -MRPP3 compared to $\Delta 85$ - and $\Delta 95$ -MRPP3, and this protein had a tendency to precipitate at concentrations above 6 mg/mL and during storage at -80°C .

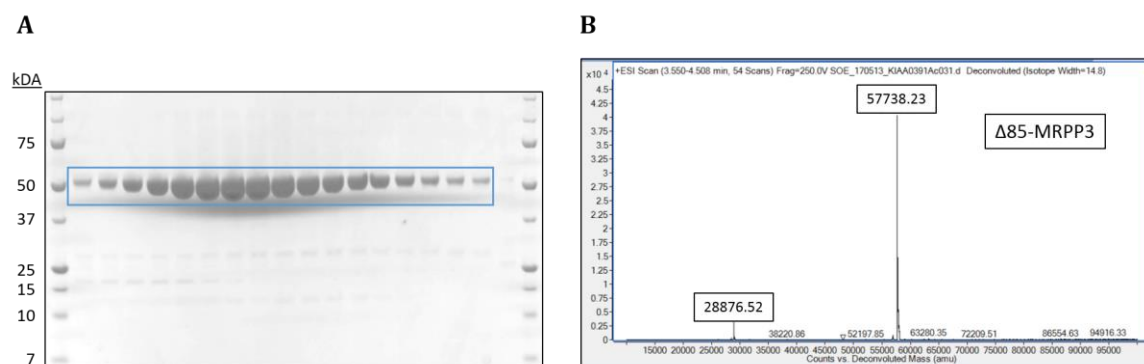


Figure 4-10. A) SDS-PAGE analysis from the final cation-EX purification step of $\Delta 85$ -MRPP3. The pooled fractions are indicated with a blue box. B) ESI-MS spectrum of purified $\Delta 85$ -MRPP3.

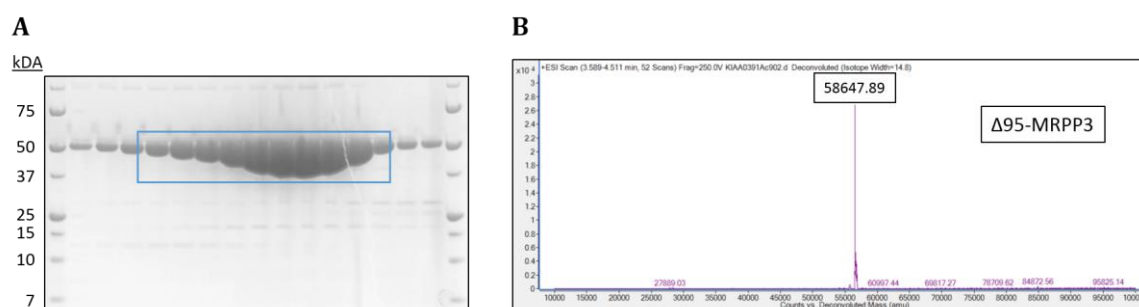


Figure 4-11. SDS-PAGE analysis from the final cation-EX purification step of $\Delta 95$ -MRPP3. The pooled fractions are indicated with a blue box. B) ESI-MS spectrum of purified $\Delta 95$ -MRPP3.

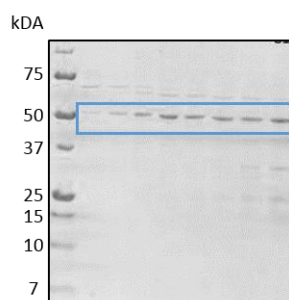


Figure 4-12. SDS-PAGE analysis from the final SEC purification step of $\Delta 117$ -MRPP3. The pooled fractions are indicated with a blue box.

The purified, untagged proteins were concentrated to 6 mg/mL ($\Delta 117$ -MRPP3) or 10 and 14 mg/mL ($\Delta 85$ and $\Delta 95$ -MRPP3), incubated with 2-fold molar excess of Mg^{2+} and utilised for crystallisation trials with four random matrix screens: HIN, JCSG, HCS and

MIDAS, at different temperatures and protein-to-precipitant ratios. After two months of incubation, no crystals were observed in any of the drops.

4.5.1.3 Optimising buffer conditions for crystallisation

The generic protein buffer ingredients such as buffer-ions, salts and detergents can interact non-covalently with the protein and may, via these interactions, prevent partial protein denaturation, formation of heterogeneous aggregation or conformational flexibility, all of which would lead to a more structurally heterogeneous sample [181]. A homogenous sample would experience a lower entropic barrier for crystallisation compared to a heterogeneous sample, and thus a search was conducted for a protein stabilising buffer. Thermal stability was taken as a measure of sample homogeneity through protein stability [150], and DSF was therefore used to screen for buffers that would thermally stabilise the protein [150]. The result of the buffer screen is outlined in **Figure 4-13** with colour coded T_m-values (in °C) in which dark green corresponds to the highest identified T_m and dark red corresponds to the lowest identified T_m. The T_m of the protein as-purified (in the original storage buffer of 25 mM HEPES pH 7.5, 350 mM NaCl, 5% glycerol) was measured separately to be 42.84±1.65°C.

The buffer screen showed higher T_m-values for the Δ95-MRPP3 protein in buffers of lower pH and with addition of glycerol (**well B4**), glycine (**well A5**), sucrose (**well B6**) and (NH₄)₂SO₄ (**well B9**). Differences in the NaCl concentration did not seem to affect the protein (**columns 1-3**). The addition of Zn²⁺ and Fe³⁺ caused the protein to precipitate so no thermal unfolding curve could be obtained (**Figure 4-14**). The addition of Mg²⁺ and Ca²⁺ to the storage buffer (H-SEC) did not stabilise the protein further.

		1	2	3	4	5	6	7	8	9	10	11	12
A	pH 5.5 50mM citrate										Binding buffer	L-SEC buffer	H-SEC buffer (10mM CaCl ₂)
B	pH 6.5 50mM MES												H-SEC buffer (10mM (NH ₄) ₂ SO ₄)
C	pH 7.5 50mM HEPES	25mM NaCl	150mM NaCl	500mM NaCl	100mM NaCl, 10% glycerol	100mM NaCl, 50mM glycine	100mM NaCl, 50mM sucrose	100mM NaCl, 60μM DOM (.4xCMC)	100mM NaCl, 100mM urea	100mM NaCl, 100mM (NH ₄) ₂ SO ₄	Elution buffer	H-SEC buffer (10mM ZnSO ₄)	H-SEC buffer (10mM FeCl ₃)
D	pH 8.5 50mM bicine											H-SEC buffer (10mM MgCl ₂)	H-SEC buffer (10mM FeCl ₃)
		43.17	43.17	39.44	45.24	47.46	44.42	40.77	42.88	44.24	43.36	45.16	46.23
		43.47	44.74	45.12	47.18	44.10	46.08	43.56	44.68	47.39	43.15	45.06	42.97
		40.61	42.17	42.62	40.69	38.23	39.49	39.15	42.49	42.45	43.02	46.18	33.63
		33.36	35.97	37.55	39.37	34.82	36.76	34.07	38.9	41.14	41.45	46.26	N.A.
Tm:		> 36.99	37.00-38.99	39.00-40.99	41.00-42.83	42.84-43.99	44.00-45.99	46.00 <					

Figure 4-13. Layout of the DSF buffer screen for Δ95-MRPP3. Binding buffer: 300 mM NaCl, 50 mM HEPES pH 7.5, 10% glycerol, 20 mM imidazole, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP); Elution buffer: 300 mM NaCl, 50 mM HEPES pH 7.5, 10% glycerol, 250 mM imidazole, 0.5 mM TCEP; L-SEC buffer: 300 mM NaCl, 50 mM HEPES pH 7.5, 10% glycerol, 0.5 mM TCEP; H-SEC buffer: 500 mM NaCl, 50 mM HEPES pH 7.5, 10% glycerol, 0.5 mM TCEP. DOM: 2,5-Dimethoxy-4-methylamphetamine, CMC: critical micelle concentration. Tm values are mean from $n = 2$ experiments.

The thermal unfolding curve for Δ95-MRPP3 measured in H-SEC buffer (**Figure 4-14**), reveals a biphasic unfolding as previously observed for mature MRPP3 (**Figure 4-8**), supporting the structure prediction of multiple domains.

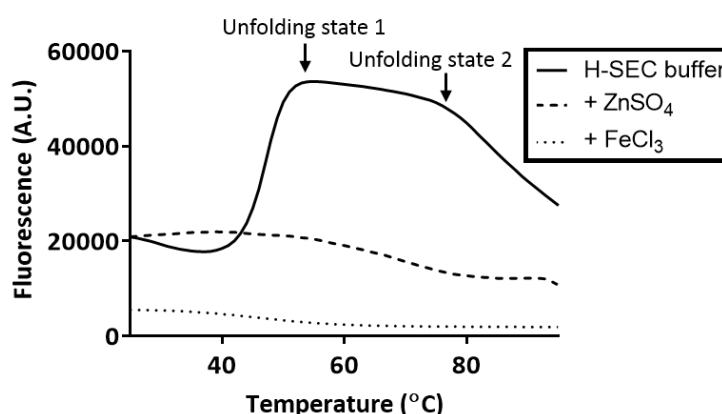


Figure 4-14. Thermal unfolding curve for Δ94-MRPP3 in H-SEC buffer: 500 mM NaCl, 50 mM HEPES pH 7.5, 10% glycerol, 0.5 mM TCEP, with the further addition of ZnSO₄ or FeCl₃. For this experiment $n = 2$.

An additional buffer screen was designed, surveying the observed trends from the initial buffer screen (**Figure 4-15**). As the protein is a presumed metallonuclease, further

metal-ions were implemented in the screen despite the initially tested ions not showing a stabilising effect and DSF showing no binding of Mg^{2+} .

		1	2	3	4	5	6	7	8	9	10	11	12
A	50mM citrate pH 5.5 100 mM NaCl												
B	50mM citrate pH 5.5 100 mM NaCl, 50mM glycine		400mM NaCl	5% glycerol	10% glycerol	10mM MgCl ₂	10% glycerol, 10mM MgCl ₂	10mM CaCl ₂	10% glycerol, 10mM CaCl ₂	10mM MnCl ₂	10&glyce rol, 10mM MnCl ₂	10mM CoCl ₂	10% glycerol, 10mM MgCl ₂
C	50mM MES pH 6.5, 100 mM NaCl												
D	50mM MES pH 6.5, 100 mM NaCl, 100mM (NH ₄) ₂ SO ₄												

48.79	47.53	49.52	50.04	48.77	50.42	49.19	48.48	48.59	49.45	48.30	48.72
48.63	47.87	49.6	49.95	45.36	50.24	49.17	48.28	48.74	49.8	48.46	48.82
46.59	48.36	47.98	48.84	47.27	48.97	47.21	48.79	46.39	48.58	N.A.	N.A.
46.92	49.34	50.52	50.88	49.57	51.39	49.57	51.16	49.20	50.55	N.A.	N.A.

> 47	47.00-48.00	48.01-49.00	49.01-50.00	50.01-51.00	< 51.00
------	-------------	-------------	-------------	-------------	---------

Figure 4-15. Buffer screen of $\Delta 94$ -MRPP3 optimised from an initial buffer screen. For this experiment $n = 2$.

This screen resulted in overall higher T_m values (range 45.36-51.39°C) compared to the previous screen (range 33.36-47.46°C). The largest ΔT_m , between MRPP3 in storage buffer and in a new buffer, was $8.55 \pm 1.65^\circ\text{C}$ in 50 mM MES pH 6.5, 100 mM NaCl, 100 mM (NH₄)₂SO₄, 10% glycerol, 10 mM MgCl₂. This corresponded to a 1 point decrease in pH, changes to the buffer-ion, a 2-fold increase in glycerol concentration, a 3.5-fold decrease in NaCl concentration and supplementation of other salts.

The $\Delta 94$ -MRPP3 protein was exchanged into this buffer and concentrated to 17 mg/mL for crystallisation trials. The buffer exchange and higher protein concentration did not, however, improve crystallisation and after two months no crystals were observed in any drops, potentially due to the – now – presence of both (NH₄)₂SO₄ and NaCl at 100 mM concentrations, excluding a number of random matrix screen conditions with concentrations of these salts < 100 mM (**Section 7.5.1**).

4.5.1.4 $\Delta 273$ -MRPP3

The MRPP3 construct of residues 274-583, with the N-extension and PPR domain removed, was expressed and purified as for mature MRPP3 (**Figure 4-16**), and its identity

and intact molecular mass was identified using tryptic digest-MS/MS and ESI-MS. The purified, untagged protein was concentrated to 12 and 14 mg/mL and used in crystallisation trials. Again, no crystal appeared even after long incubation times.

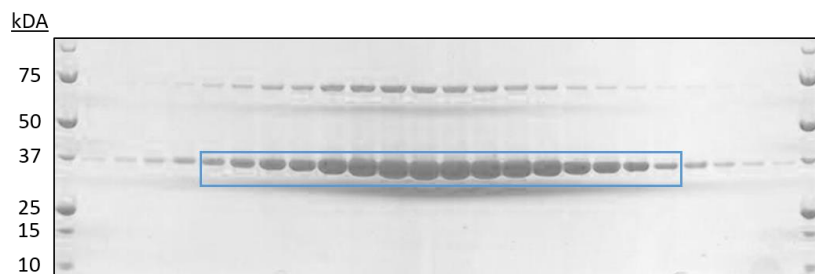


Figure 4-16. SDS-PAGE analysis from the final purification step of $\Delta 273$ -MRPP3. The pooled fractions are indicated with a blue box.

Whilst the crystallisation attempts for $\Delta 273$ -MRPP3 were ongoing, a crystal structure was published for the same construct solved to 2.7 Å resolution by Li *et al.* [180] and a structure was published for a $\Delta 206$ -MRPP3 construct solved to 1.8 Å resolution by Reinhard *et al.* [179]. The successful crystallisation of the $\Delta 273$ -MRPP3 required a higher protein concentration than used here (20 mg/mL vs. 14 mg/mL), but in otherwise similar crystallisation conditions.

4.5.2 Outline of the published structure

The two published MRPP3 structures display the same fold and metal-binding pattern, so the structural analysis presented herein is based on the longer crystallised construct ($\Delta 206$ -MRPP3) by Reinhard *et al.* [179] (PDB: 4XGL). The structure (**Figure 4-17**) confirms previous structure prediction and DSF biphasic unfolding data showing the presence of multiple protein domains.

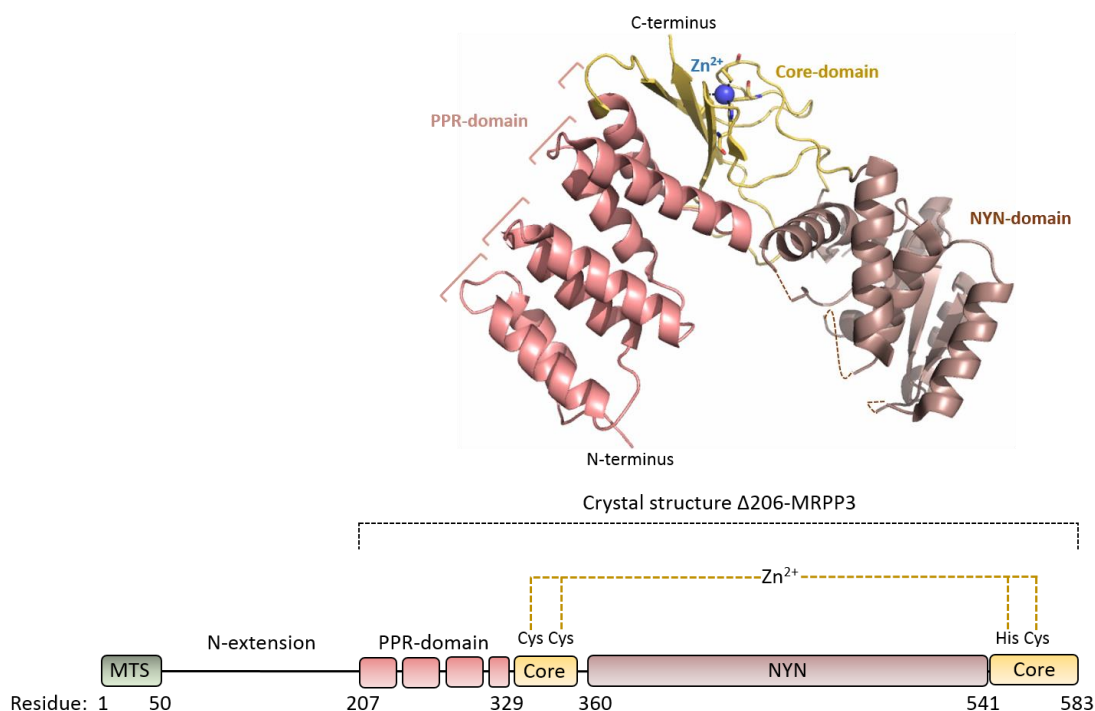


Figure 4-17. Domain organisation of mature MRPP3 based on the published structure of $\Delta 206$ -MRPP3.

The structure shows three PPR-repeats and a single PPR-terminating-helix together making up the PPR-domain (**Figure 4-17**, pale red). It further reveals a large NYN-domain (**Figure 4-17**, brown) and a C-extension that interacts with the middle part of the protein to form *one* collaborative core-domain (**Figure 4-17**, yellow). The folding of this core domain is assisted by the structural Zn^{2+} ion (**Figure 4-17**, blue) bound in the Cys-Cys-His-Cys motif (**Figure 4-17**, shown as sticks) with residues from *both* parts of the core-domain [180]. The formation of *one* core-domain from *two* sequentially disjoint parts of the protein offers an explanation to the finding that any C-terminal truncation of MRPP3 rendered the protein insoluble unless both the NYN-domain and the core-domain were fully removed. The crystal structure revealed no metal-ions bound in the active site despite efforts to crystallise the protein in buffers/precipitant solutions devoid of metal-chelators, and with excess amounts of Mg^{2+} ions [179]. This agrees with work presented here that mature MRPP3 alone does not cause a T_m shift in the presence of Mg^{2+} under the DSF conditions.

Recently, a crystal structure of the MRPP3 plant homologue organellar-PRORP1 from *A. thaliana* has been reported to 1.95 Å resolution (PDB: 4G24) [93]. The structure covers residue 95-572 (Δ 94-PRORP1) and show a similar domain organisation to MRPP3 of a PPR-domain, a central core-domain, an NYN-domain and a structural Zn-atom bound in a CCHC motif (**Figure 4-18**) [93]. PRORP1 is crystallised as a longer construct, compared to MRPP3, revealing two additional PPR-repeats in the N-terminal region. It is possible that such additional repeats are present in the MRPP3 N-extension, not covered by the Δ 206-MRPP3 crystal structure, resulting in an expanded PPR-domain and shortened N-extension. PRORP1 does not need assisting proteins to perform RNA cleavage (RNase P activity), and thus must be able to accommodate the metal-ion cofactors within the protein for catalysis. Indeed, the crystal structure of PRORP1 shows two Mn^{2+} ions in the active site in positions presumably occupied by Mg^{2+} during catalysis. These ions bind to PRORP1 using the four conserved Asp residues that are also present in MRPP3.

Comparing the crystal structures of MRPP3 and PRORP1, it becomes apparent that the metal-binding discrepancy seen here and by others can be explained by dual conformations of the protein. Specifically, the spatial organisation of the four Asp residues varies between the metal-free (MRPP3) and metal-bound (PRORP1) structures, with one Asp residue in the metal-unbound MRPP3 even pointing away from the active site (D499) (**Figure 4-18**). The salt bridge formed between D478 and R498 in MRPP3 brings the Arg residue into the space that would, from the PRORP1 structure, otherwise be occupied by a metal-ion, prohibiting metal-binding in MRPP3 for this conformation [180].

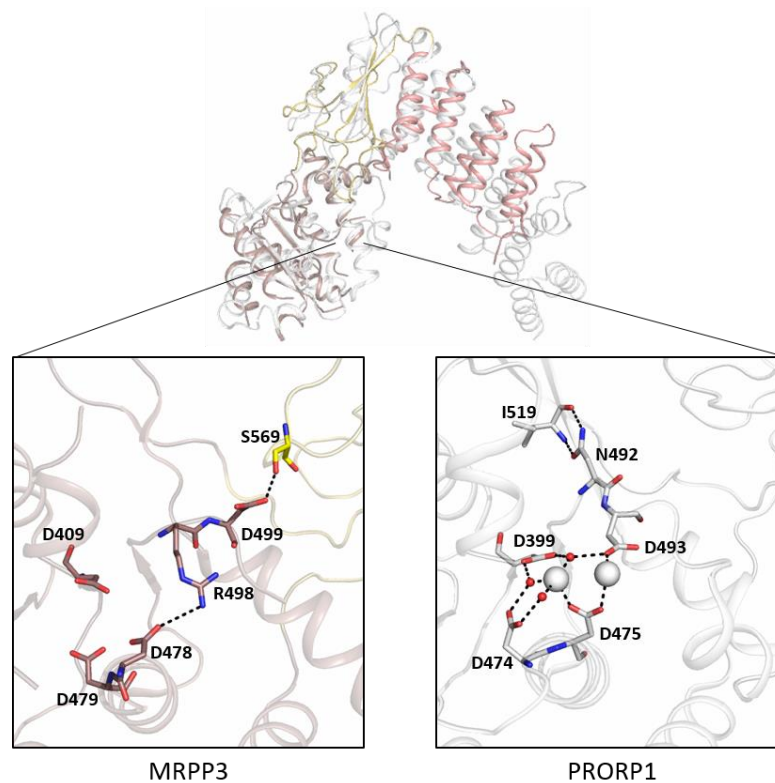


Figure 4-18. Active site view of MRPP3 (left, coloured) and PRORP1 (right, grey), enlarged from a superposition of the two structures. The two Mn^{2+} ions bound to PRORP1 are shown as spheres (grey) along with three water molecules (red). The catalytic Asp residues are shown in sticks along with residues interacting with structural elements in other parts of the proteins that assist in active site formation (S569/R498 for MRPP3 and I519/N492 for PRORP1). Figure adapted from Reinhard *et al.* [179].

Apart from the difference in metal-binding properties, the two structures also reveal different distances between the NYN- and PPR-domains. MRPP3 adopts a more relaxed, L-shaped conformation whereas PRORP1 is in a more compact V-shaped form (**Figure 4-19**).

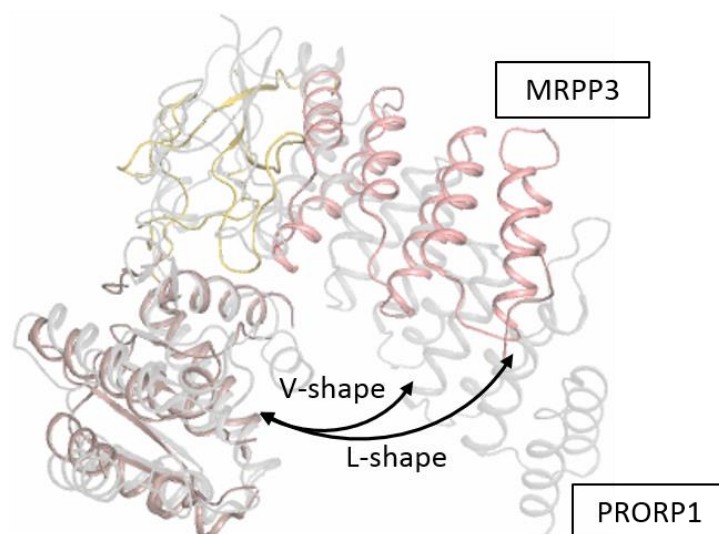


Figure 4-19. L-shape and V-shape conformations of MRPP3 (in colour) and PRORP1 (grey), respectively. In this figure, the NYN-domains of the two structures are aligned.

Overall, it seems that uncomplexed MRPP3 is in an inactive conformation that requires conformational changes in the active site – potentially by changing the orientation of individual domains towards each other – to accommodate the Mg^{2+} cofactor necessary for RNase P activity [179], [180]. Such conformational change may be mediated by the binding of MRPP1-MRPP2 and/or pre-(mt)tRNA substrate. Therefore, structural information on the ternary RNase P complex with pre-(mt)tRNA is needed to elucidate the activation mechanism of MRPP3.

4.6 SAXS studies of MRPP3

With the $\Delta 206$ -MRPP3 structure solved, and the mature MRPP3 protein not amenable to crystallisation, structural characterisation of MRPP3 was subsequently undertaken with the low resolution SAXS technique. For this, two MRPP3 constructs were used: the mature protein as well as the $\Delta 206$ -MRPP3 construct used in the published crystal structure. $\Delta 206$ -MRPP3 was cloned, expressed, purified and verified as for mature MRPP3.

Purified mature MRPP3 and $\Delta 206$ -MRPP3 proteins were subjected to in-line SAXS analysis at the Diamond Light source beamline B21 and scattering data were collected (**Figure 4-20A** and **Figure 4-20B**, respectively).

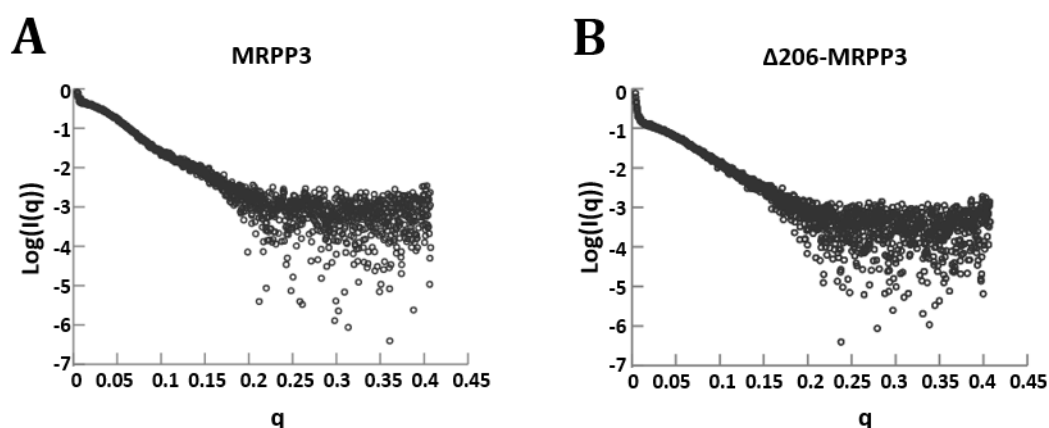


Figure 4-20. SAXS scattering plot: $\log(I(q))$ vs. q , for A) MRPP3 and B) $\Delta 206$ -MRPP3.

The scattering data for $\Delta 206$ -MRPP3 gave a linear Guinier fit (**Figure 4-21A**) with an R_g for the protein of 28.83 ± 1.470 Å. The Kratky plot (**Figure 4-21B**) indicated a folded, partly flexible protein in solution. The $D_{\max}$ of 98 Å gave rise to the $P(r)$ distribution function shown in **Figure 4-21C**.

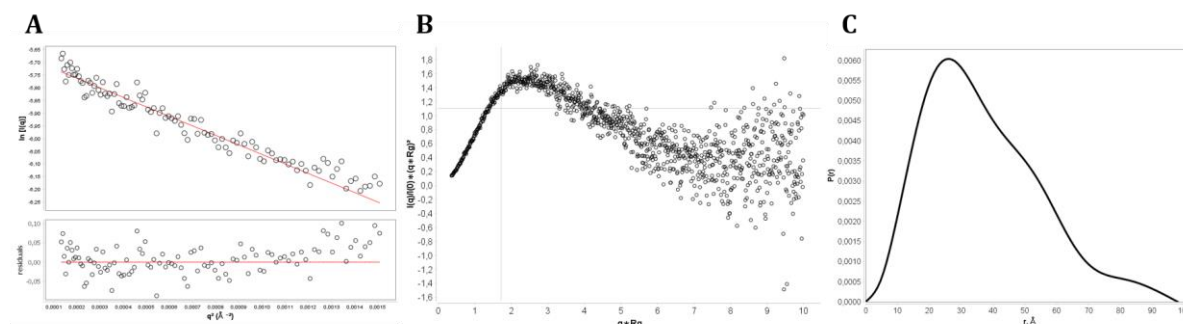


Figure 4-21. SAXS data plots for $\Delta 206$ -MRPP1. A) Guinier plot, B) Kratky plot, C) $P(r)$ distribution function.

A calculated theoretical scattering plot for $\Delta 206$ -MRPP3, derived from its crystal structure, showed a fit to the experimental scattering plot with an χ -value of 1.74 for q -values between 0.01° - 0.45° (**Figure 4-22**), suggesting that the structure represents a relevant form of the protein in solution. A poor fit is observed for the lower q -values due to radiation damage induced protein aggregation, and at slight misfit is found for higher q -values indicating protein flexibility confirmed by missing electron density in the crystal structure of $\Delta 206$ -MRPP3 for several large loops [179].

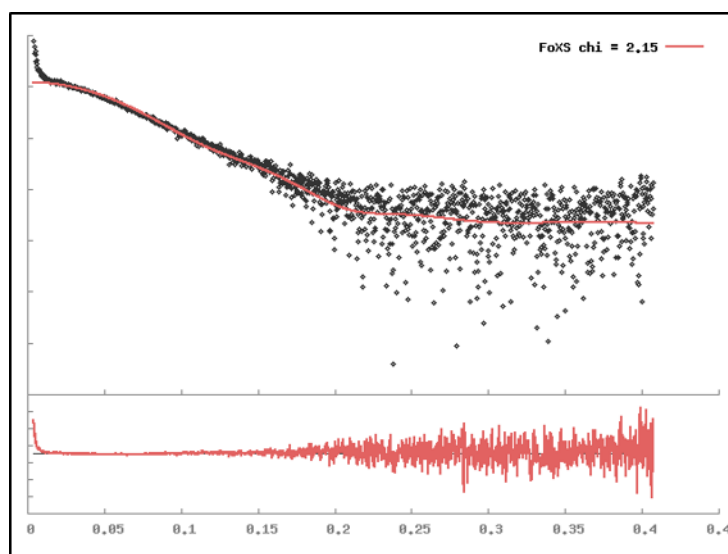


Figure 4-22. Comparison of the experimentally determined SAXS scattering plot ($\log I(q)$ versus q) of $\Delta 206$ -MRPP3 (black), with the calculated SAXS scattering plot ($\log I(q)$ versus q) from the $\Delta 206$ -MRPP3 (PDB: 4XGL) crystal structure (red) computed with the Debye formula.

For mature MRPP3, the scattering data gave rise to a linear Guinier fit (**Figure 4-23A**) resulting in an R_g of 33.18 ± 1.118 Å. The Kratky plot (**Figure 4-23B**) indicated a folded, partly flexible protein in solution and the $D_{\max}$ of 111 Å gave rise to the $P(r)$ distribution function shown in **Figure 4-23C**.

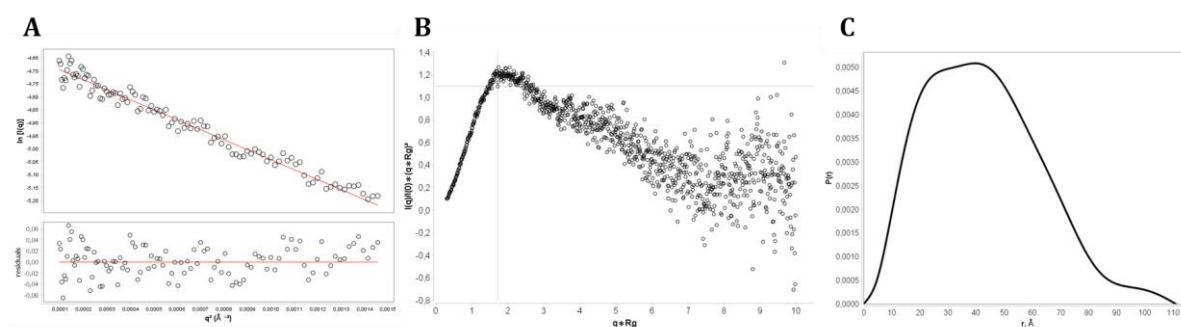


Figure 4-23. Mature MRPP3 protein SAXS data. A) Guinier plot, B) Kratky plot, C) $P(r)$ distribution function.

The larger R_g and $D_{\max}$ values obtained for mature MRPP3 ($R_g = 33.18 \pm 1.118$ Å, $D_{\max} = 111$ Å), compared to $\Delta 206$ -MRPP3 ($R_g = 28.83 \pm 1.470$ Å, $D_{\max} = 98$ Å), are consistent with their expected differences in size. The SAXS model for mature MRPP3 (**Figure 4-24**) accommodates the $\Delta 206$ -MRPP3 crystal structure leaving room for the N-extension (residues 51-206) not present in the crystal structure (**Figure 4-24**, green box).

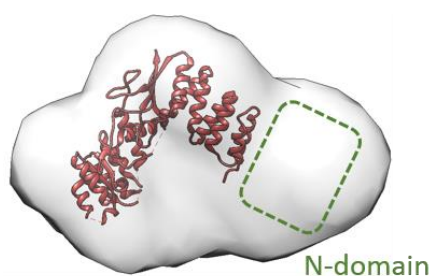


Figure 4-24. SAXS model of mature MRPP3 with $\Delta 206$ -MRPP3 fitted into the envelope. PDB: 4XGL.

At the time of thesis preparation, a 3.05 Å resolution crystal structure was published of an MRPP3 plant homologue, cytosolic-PRORP2 from *A. thaliana*, missing only the N-terminal residues 1-27 (PDB: 5FT9) [182]. This protein has a similar domain architecture

as MRPP3, but PRORP2 does not contain an MTS sequence due to its cytosolic location [90], or a long N-extension (**Figure 4-25**). Moreover – like PRORP1 – this protein also encompasses two additional PPR repeats, increasing the probability that MRPP3 could also contain two such additional repeats.

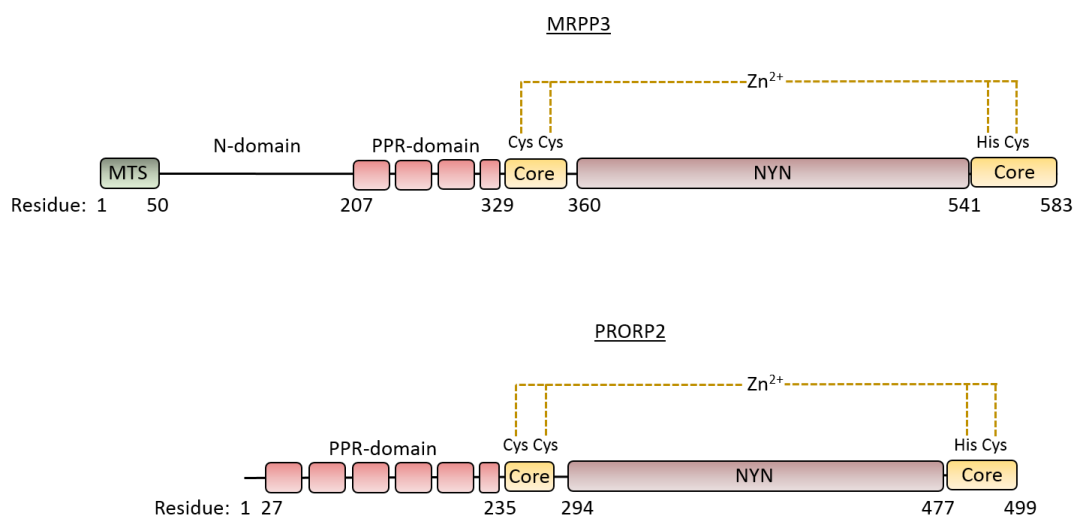


Figure 4-25. Domain organisation of PRORP2 and mature MRPP3.

If MRPP3 *does* contain two additional PPR-repeats of 35 amino acids each, the N-extension would be reduced to residues 51-136. Modelling two further PPR-repeats from PRORP2 onto the $\Delta 206$ -MRPP3 structure, this longer MRPP3 model occupies more of the SAXS model, while still leaving room for the, now shorter, N-extension (**Figure 4-26**).

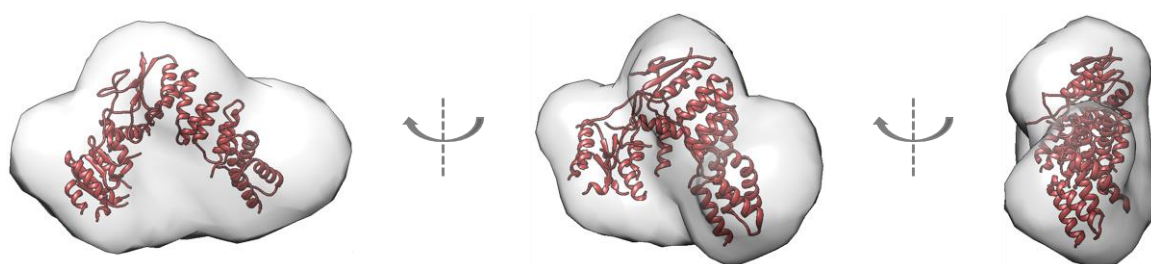


Figure 4-26. SAXS envelope of mature MRPP3 aligned with a composite structure containing human MRPP3 (residues 207-583 from PDB: 4XGL) and plant PRORP2 (residues 28-108 from PDB: 5FT9).

CHARACTERISATION OF THE MRPP1-MRPP2-MRPP3 COMPLEX

This second part of **Chapter 4** presents a SAXS model of the MRPP1-MRPP2-MRPP3 complex with pre-(mt)tRNA, along with biochemical characterisation for the order of complex assembly and a mapping of domains for protein-protein interaction and RNase P activity.

4.7 Expression and Purification of MRPP3 proteins

A pre-requisite for biochemical and structural characterisation of the complex is the availability of large quantities of highly purified proteins. In this work, the ternary complex will be reconstituted by mixing MRPP3 with the two proteins MRPP1 and MRPP2 either expressed individually or co-expressed (**Section 3.1**). The various MRPP3 proteins utilised for complex reconstitution and biochemical characterisation are shown in **Figure 4-27**. To map the functional regions of MRPP3, two constructs, additional to the mature protein, are included. The MRPP3 construct covering the inseparable core-domain/NYN-domain (residues 274-583) was used to represent the C-terminal part of the protein and is denoted in this chapter as C-MRPP3. The MRPP3 construct covering the PPR-domain and part of the N-extension (residues 86-274), was used to represent the N-terminal part of the protein and is denoted N-MRPP3.

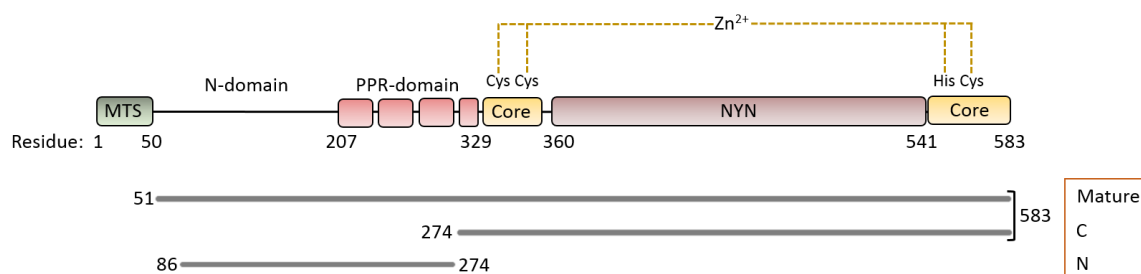


Figure 4-27. Cloned, soluble constructs of MRPP3 utilised in assays throughout this chapter. Soluble constructs: green.

4.7.1 N-MRPP3

N-MRPP3 was purified by IMAC (Ni-NTA resin) followed by SEC (S200 column) (**Figure 4-28**). Due to low protein yield no further purification steps were performed. The identity and intact molecular mass were confirmed as for other constructs.

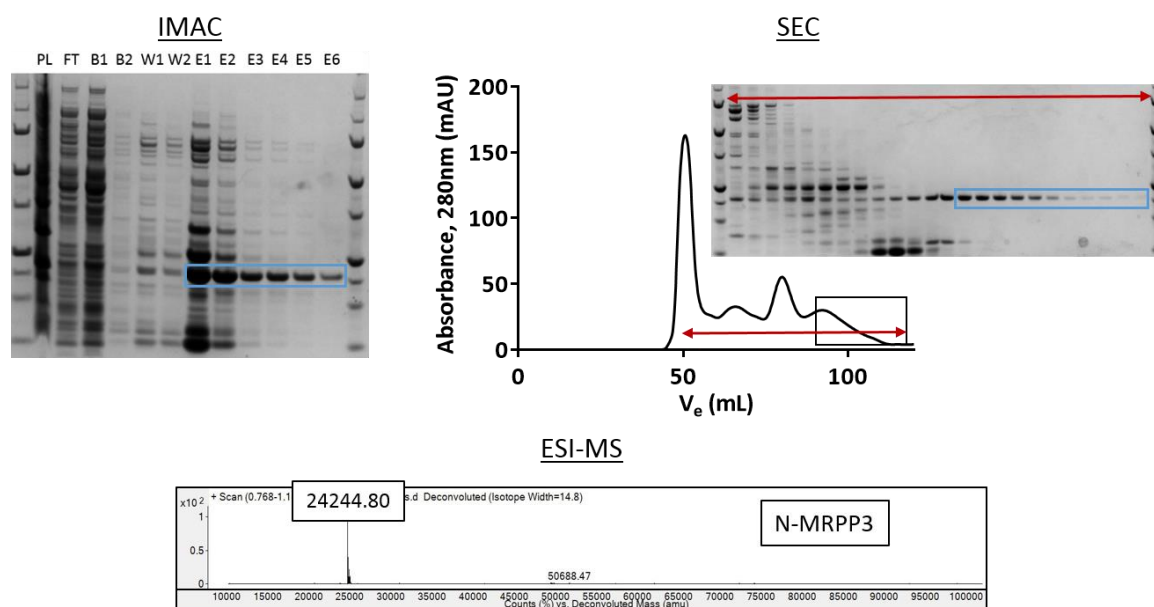


Figure 4-28. Large-scale purification of N-MRPP3. The example shows SDS-PAGE analysis of initial IMAC, and chromatogram from SEC and SDS-PAGE analysis of peak fractions. The ESI-MS spectrum of a sample after purification is shown. Red arrows, blue boxes and gel lanes are annotated as before.

4.7.2 C-MRPP3

C-MRPP3 was purified by IMAC (Ni-NTA resin), SEC (S200 column) and cation-EX (HiTrap SP HP column) (**Figure 4-29**). The His₆-tag proved difficult to remove by protease cleavage and was left intact. The identity and intact molecular mass were confirmed as before.

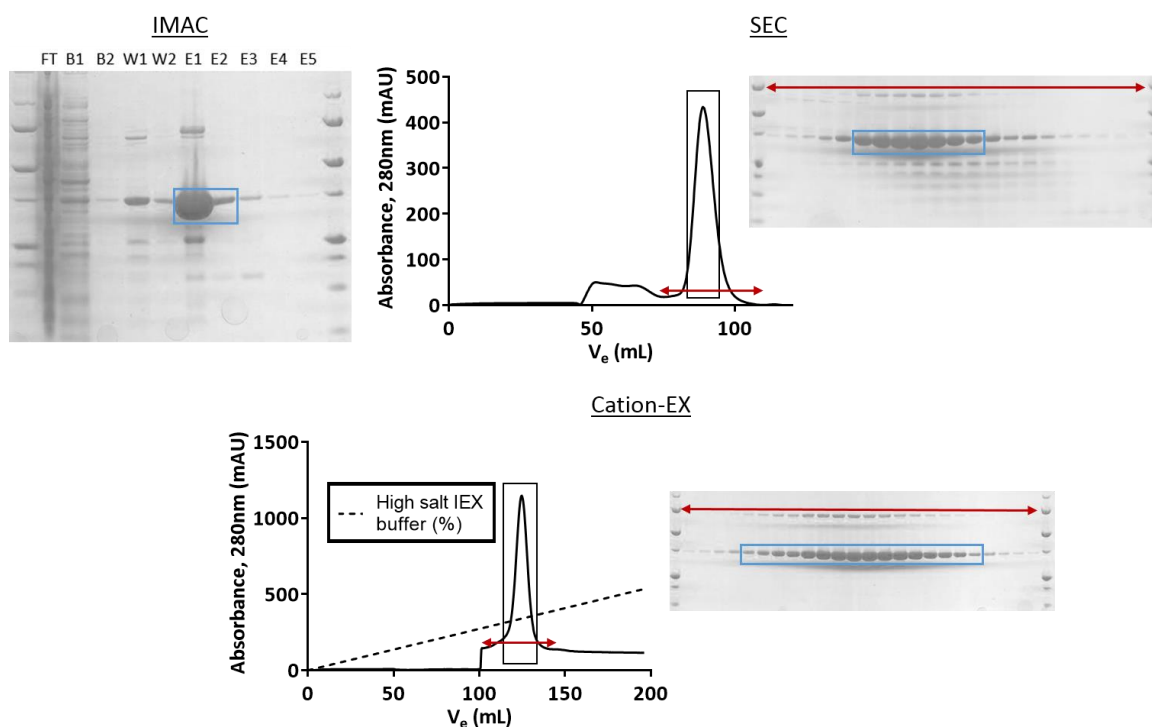


Figure 4-29. Large scale purification of C-MRPP3 ($\Delta 206$ -MRPP3). The example shows SDS-PAGE analysis of initial IMAC, chromatogram from SEC and SDS-PAGE analysis of peak fractions, and chromatogram from cation-EX and SDS-PAGE analysis of peak fractions. Red arrows, blue boxes and gel lanes are annotated as before.

4.8 Solution studies

4.8.1 Reconstitution of the ternary complex with mature MRPP3

To investigate the interaction between the proteins in the ternary complex, mature MRPP3, mature MRPP1 and full-length MRPP2 proteins, individually prepared in large scale, were subjected to analytical SEC in the presence of Mg^{2+} for RNase P activity, and in the presence and absence of *in vitro* transcribed pre-(mt)tRNA^{Ile} substrate.

The three proteins were applied to the column alone and in a mixture with and without pre-(mt)tRNA substrate. Upon loading of a pre-incubated mixture of all three proteins, no peak shift was observed, relative to individual protein peaks, beyond that of the binary MRPP1-MRPP2 complex, indicating that MRPP3 does not form a ternary complex with the two partner proteins in the absence of ligands. However, a further peak shift – beyond the shift of MRPP1-MRPP2 alone – was observed upon incubation of all three proteins *with* excess pre-(mt)tRNA substrate (**Figure 4-30**, green box). SDS-PAGE analysis

showed the appearance of all three proteins (MRPP1, MRPP2 and MRPP3) in earlier fractions than the binary complex (**Figure 4-30, green box**), and disappearance of the proteins in the elution fractions where the individual proteins would otherwise be found. This, and the peak shift on the chromatogram, show that a larger MW species (*i.e.* a complex) is formed amongst the three proteins upon addition of the pre-(mt)tRNA substrate.

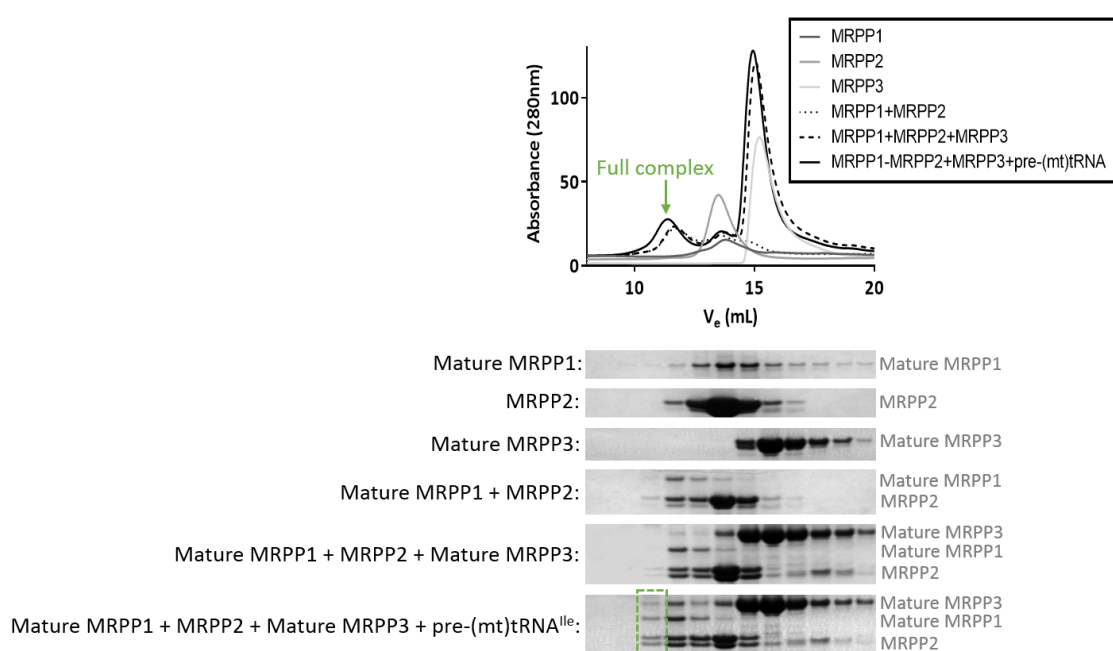


Figure 4-30. Analytical SEC with the mature MRPP3 protein applied in various combinations with mature MRPP1 and full-length MRPP2 with and without pre-(mt)tRNA^{lle}. The green box indicates the elution fractions in which proteins would co-localise upon complex formation.

The above result suggests an ordered complex assembly where a pre-formed MRPP1-MRPP2 complex initially binds pre-(mt)tRNA (**Section 3.5**) and subsequently accommodates the MRPP3 protein. Having elucidated the assembly order and requirement of pre-(mt)tRNA substrate for complex formation, domain mapping of MRPP3 was performed to investigate if the N- or C-MRPP3 protein alone was sufficient for complex formation.

4.8.2 Domain mapping

Analytical SEC was intended for the N-MRPP3 and C-MRPP3 proteins to identify the interaction regions between the MRPP1-MRPP2-pre-(mt)tRNA^{Ile} complex and MRPP3. Unfortunately, due to protein precipitation, N-MRPP3 could not be concentrated to a level required for SEC and therefore only C-MRPP3 was used in this assay.

C-MRPP3 and co-expressed MRPP1-MRPP2 (**Section 3.1.2**) proteins were prepared in large scale. The MRPP1-MRPP2-pre-(mt)tRNA^{Ile} complex was formed prior to addition of the C-MRPP3 protein and the assay was conducted in the presence of Mg²⁺. The chromatogram (**Figure 4-31, green box**) shows no peak shift upon mixing of C-MRPP3 with the MRPP1-MRPP2 proteins and pre-(mt)tRNA^{Ile} substrate, and SDS-PAGE analysis of the elution fractions did not reveal any of the three proteins in earlier elution fractions compared to where they would otherwise be eluted as individual proteins (**Figure 4-31, green box**). This indicates that the C-MRPP3 alone is not sufficient for complex formation with the pre-formed MRPP1-MRPP2-pre-(mt)tRNA^{Ile} protein-substrate complex.

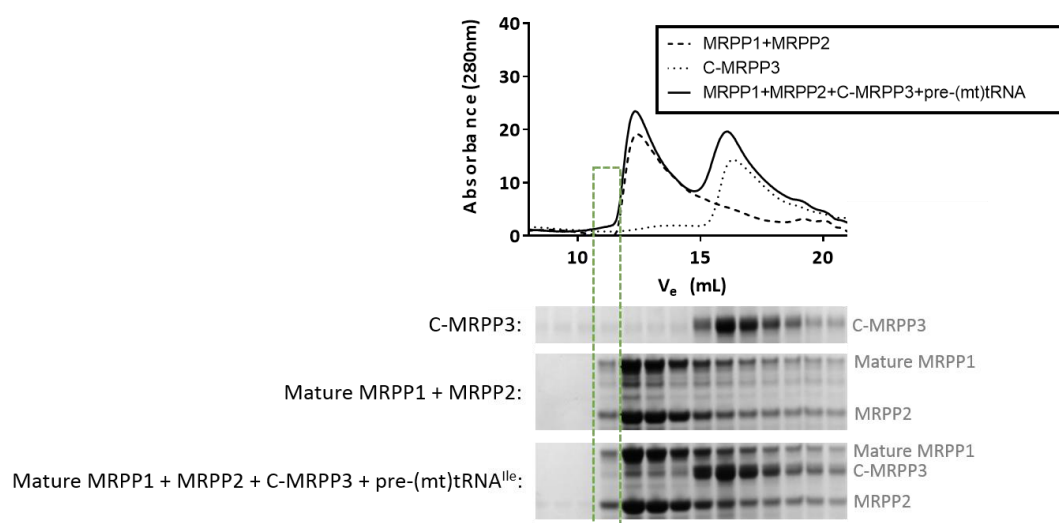


Figure 4-31. Analytical SEC with C-MRPP3 applied alone and with the co-expressed MRPP1-MRPP2 complex in the presence and absence of pre-(mt)tRNA^{Ile}. The green box indicates the elution fractions where the proteins should be co-localising if they form a complex.

As this assay could not be performed with the low concentration of N-MRPP3 protein, an RNase P activity assay – requiring lower protein concentrations – was set up to determine the role of the MRPP3 regions in the complex.

4.9 RNase P activity studies

4.9.1 MRPP3 regions important for activity

The RNase P assay identifies the 5'-end cleavage products of pre-(mt)tRNA substrate (*i.e.* the 5'-leader and processed tRNA) resulting from the reaction catalysed by the active RNase P complex (MRPP1-MRPP2-MRPP3) in the presence of Mg^{2+} (**Figure 4-32**). If no reaction occurs, the intact pre-(mt)tRNA will be present as the only species.

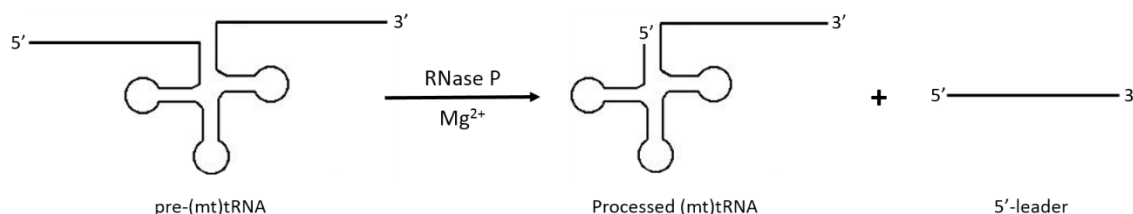


Figure 4-32. Outline of RNase P activity on a mitochondrial pre-tRNA.

The assay was performed with *in vitro* transcribed pre-(mt)tRNA^{lle} as substrate and the ternary complex was formed by incubating co-expressed MRPP1-MRPP2 with different MRPP3 constructs in the presence of Mg^{2+} . The various MRPP3 proteins include, mature MRPP3, N-MRPP3 and C-MRPP3, along with a shorter construct lacking part of the N-extension ($\Delta 95$ -MRPP3) (**Figure 4-33**). Detection of substrate and cleavage products was carried out by denaturing UREA-PAGE gel stained for RNA. A negative control was included by pre-incubating the MRPP1-MRPP2 and MRPP3 proteins with EDTA before addition of pre-(mt)tRNA substrate. Additionally, the proteins were tested for activity individually (except for mature MRPP1 that was unavailable at the time).

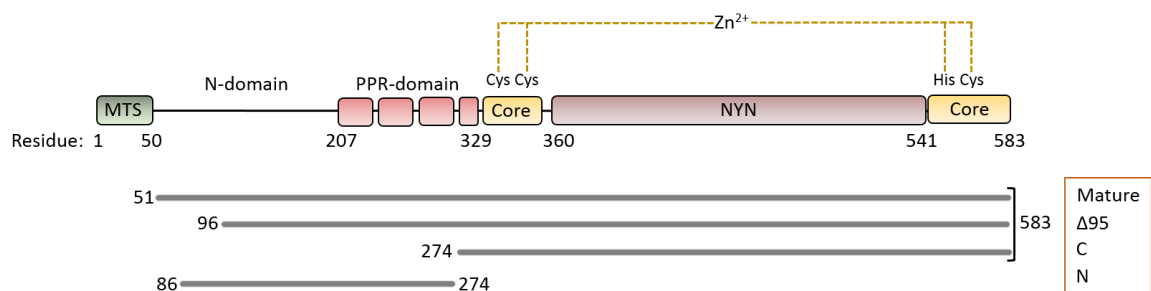


Figure 4-33. Cloned, soluble constructs of MRPP3 utilised in the presence of MRPP1-MRPP2 complex for the RNase P assay.

The denaturing UREA-PAGE gel (**Figure 4-34**) shows clear appearance of a 5'-leader band and a shift in the size of the pre-(mt)tRNA substrate from before and after processing, in reaction duplicates containing all three proteins (**lanes 6 and 7**). The observed RNase P activity confirms that the *E. coli* expressed, recombinant MRPP3 protein adopts an active fold as also observed for the MRPP1 and MRPP2 proteins in the MTase assay (**Section 3.3.1**). MRPP3 protein did not show any RNase P activity alone (**lane 3**), or only with MRPP2 (**lane 5**), towards this pre-(mt)tRNA, as observed by the lack of band shift. As expected, MRPP2 alone (**lane 2**), MRPP1-MRPP2 complex without MRPP3 (**lane 4**) and the negative control with all three proteins in the presence of EDTA (**lane 1**) also did not show any activity.

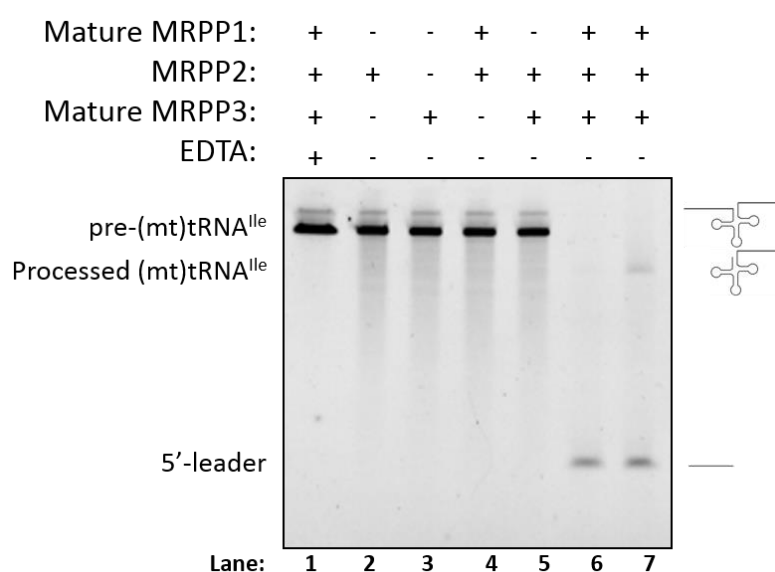


Figure 4-34. RNase P activity tested for various complex components alone and together, represented by the separation of reaction products on a denaturing UREA-PAGE gel stained for RNA. The pre-(mt)tRNA substrate and cleavage products are schematically shown next to the gel.

After establishing that the MRPP1-MRPP2-MRPP3 complex involving mature MRPP3 protein has RNase P activity towards pre-(mt)tRNA^{Ile}, and that neither component are active alone, the smaller fragments of MRPP3 were tested by incubating these with co-expressed MRPP1-MRPP2 and pre-(mt)tRNA^{Ile}. Again, the denaturing UREA-PAGE gel (**Figure 4-35**) shows no RNase P activity for either the negative control (**lane 1**) or the sample without MRPP3 (**lane 2**). RNase P activity was observed for the ternary complex involving mature MRPP3 (**lane 3**) and $\Delta 94$ -MRPP3 (**lane 4**). The role of the N-extension remains unclear since its partial removal ($\Delta 94$ -MRPP3) still leaves a functional protein. The role of the N-extension is further discussed in **Chapter 6**.

Despite containing an intact catalytic NYN-domain, a complex involving C-MRPP3 showed no RNase P activity (**lane 6**), suggesting that the PPR-domain of MRPP3 plays an important, direct or indirect role in catalysis. The PPR-domain in the MRPP3 homologue PRORP1 is involved in direct RNA-binding via the tRNA D-loop and T-loop [183], [184], and it is likely that the PPR-domain of MRPP3 plays a similar role, thus allowing efficient binding of the pre-(mt)tRNA. No RNase P activity was observed for a complex involving the MRPP3 PPR-domain alone (**lane 5**).

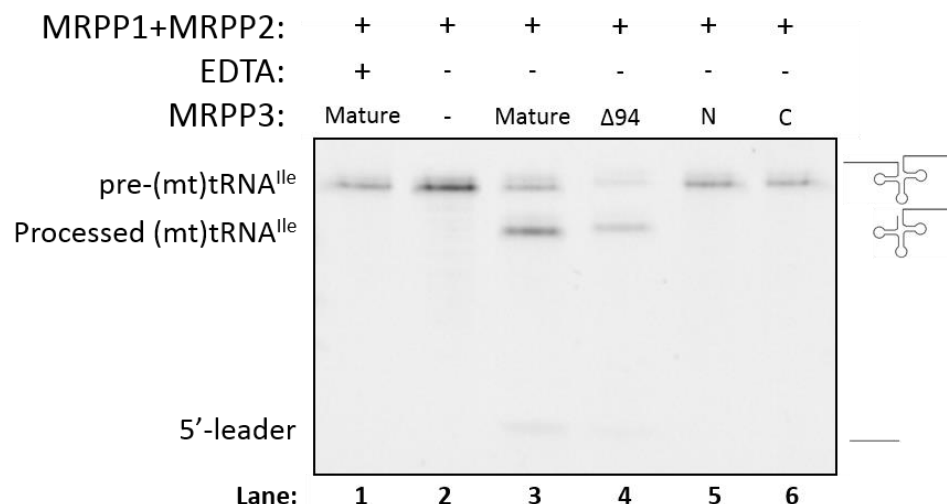


Figure 4-35. RNase P activity of MRPP3 fragments, in complex with MRPP1-MRPP2, represented by the separation of reaction products on a denaturing UREA-PAGE gel stained for RNA. The pre-(mt)tRNA substrate and cleavage products are schematically shown next to the gel.

4.9.2 Regions of MRPP1 important for activity

The activity assay was next performed to test the influence of MRPP1 on RNase P activity, by removing various domains of this protein before complex formation. The assay was conducted by incubating different MRPP1 constructs (N-domain and MT+C-domain) with MRPP2 and mature MRPP3 in the presence of Mg^{2+} before addition of pre-(mt)tRNA^{Ile}. The reaction was also performed for mature MRPP1 with MRPP2 and MRPP3 in the presence or absence of EDTA, serving as positive and negative controls, respectively.

The UREA-PAGE gel (**Figure 4-36**) showed that mature MRPP1, combined with MRPP2 and MRPP3, exhibited RNase P activity towards the pre-(mt)tRNA in the absence of EDTA (**lane 4**), but no activity in the presence of EDTA (**lane 1**), as expected for controls. No pre-(mt)tRNA processing takes place when only MRPP1 N-domain (**lane 2**) or MT+C-domain (**lane 3**) is combined with MRPP2 and MRPP3. Data presented here showed that no binary complex could be formed between MRPP2 and either of these two domains of MRPP1 alone (**Section 3.2.2**), and the lack of RNase P activity for these fragments thus shows that MRPP3 requires an intact MRPP1-MRPP2 complex for activity.

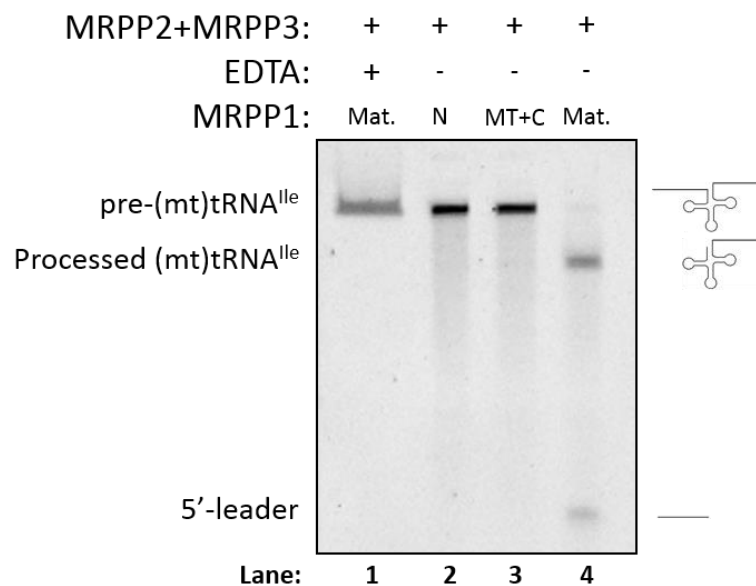


Figure 4-36. RNase P activity of MRPP1 fragments in a mixture with MRPP2 and MRPP3, represented by the separation of reaction products on a denaturing UREA-PAGE gel stained for RNA. Mat.: mature.

4.9.3 Mapping of MRPP3 Catalytic residue

The ternary protein complex could only be formed in the presence of substrate. To trap a fully stable complex that avoid any possible substrate turnover and/or complex dissociation, a catalytically inactive mutant of MRPP3 was cloned for structural studies. A mutation candidate was identified as one of the catalytic residues: Asp479, previously shown to abolish RNase P activity [179].

Ala substitution of this potential catalytic residue, p.D479A, was performed by site-directed mutagenesis, and the mutant construct was expressed and purified (**Figure 4-37A**) as described for mature wildtype MRPP3 and proved to be soluble and stable against degradation. The presence of the mutation was confirmed by Sanger sequencing and ESI-MS (**Figure 4-37B**) identified a molecular mass of 61565.79 Da which accounts for the Asp479-to-Ala substitution.

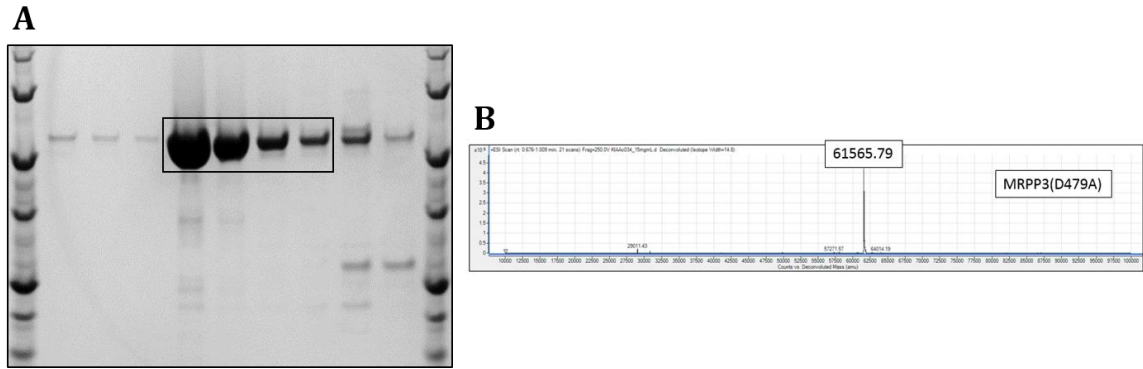


Figure 4-37. A) SDS-PAGE analysis from the final purification step of mature MRPP3(D479A) mutant. The pooled fractions are indicated with a black box. B) ESI-MS of purified MRPP3(D479A).

The RNase P activity of MRPP3(D479A) was tested by incubating co-expressed MRPP1-MRPP2, MRPP3(D479A) and pre-(mt)tRNA^{Ile} in the presence of Mg²⁺, alongside a positive control (complex with mature wildtype MRPP3) and a negative control (complex with mature wildtype MRPP3 + EDTA).

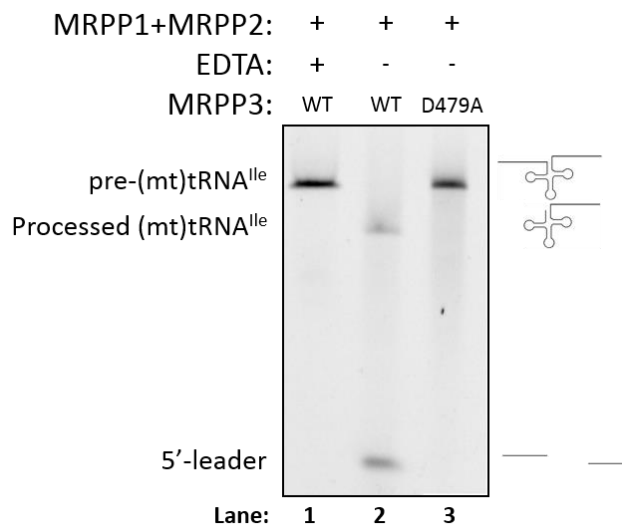


Figure 4-38. RNase P activity of mature MRPP3 wildtype (WT) and D479A mutant, in a mixture with MRPP1-MRPP2 complex, represented by the separation of reaction products on a denaturing UREA-PAGE gel.

RNase P activity was observed (**Figure 4-38**) for the positive control (**lane 2**) but not for the negative control (**lane 1**). Importantly, no 5'-leader or processed pre-(mt)tRNA product was observed for the complex involving the MRPP3(D479A) mutant (**lane 3**), confirming that this mutant was catalytically inactive. This inactive MRPP3(D479A) mutant protein was subsequently utilised for structural studies.

4.10 SAXS studies of the ternary complex

The SAXS technique had previously proven useful to obtain low resolution structural information on the MRPP1-MRPP2-pre-(mt)tRNA^{Ile} complex, and thus this method was also applied to the ternary complex including MRPP3. The complex was formed by incubating the MRPP1-MRPP2 binary complex with pre-(mt)tRNA^{Ile} and subsequently adding MRPP3(D479A) in the presence of Mg²⁺. The complex was applied for in-line SAXS analysis at the Diamond Light source beamline B21 in a buffer facilitating tRNA binding and complex formation (low NaCl concentration, presence of Mg²⁺), and scattering data were collected (**Figure 4-39**).

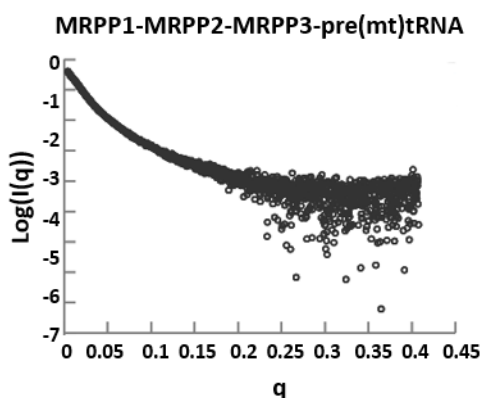


Figure 4-39. SAXS scattering plot: $\log(I(q))$ vs. q , for the MRPP1-MRPP2-MRPP3-pre-(mt)tRNA^{Ile}-MRPP3(D479A) complex.

From the scattering data, a Guinier plot could be obtained showing potential inter-particle interactions at the lower scattering angles indicative of high sample concentration (**Figure 4-39A**). The Guinier plot gave rise to an R_g of 82.60 ± 0.9538 Å, the Kratky plot (**Figure 4-39B**) indicated a flexible particle in solution and the $D_{\max}$ of 248 Å gave rise to the $P(r)$ distribution function shown in **Figure 4-39C**.

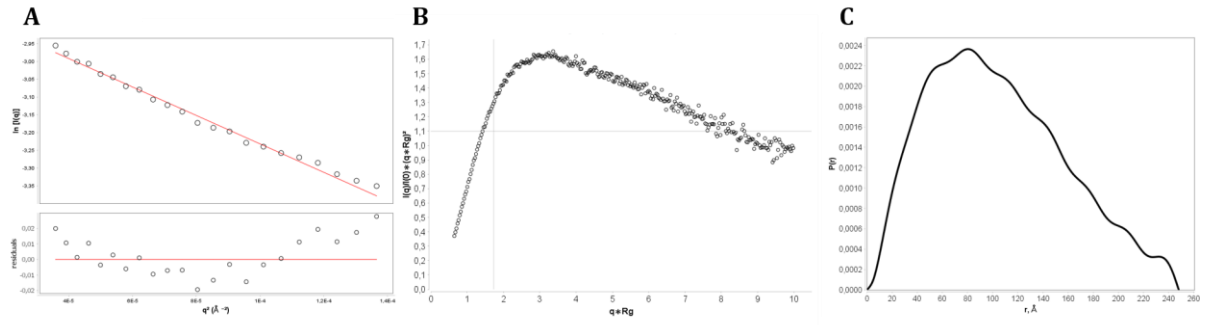


Figure 4-40. SAXS plots for MRPP1-MRPP2-pre-(mt)tRNA^{Ile}-MRPP3(D479A) complex. A) Guinier plot, B) Kratky plot, C) P(r) distribution function.

As expected for a larger species, the R_g and $D_{\max}$ values increased for the full MRPP1-MRPP2-pre-(mt)tRNA^{Ile}-MRPP3(D479A) complex ($R_g = 82.60 \pm 0.9538 \text{ \AA}$, $D_{\max} = 248 \text{ \AA}$), compared to the complex of MRPP1-MRPP2-pre-(mt)tRNA^{Ile} ($R_g = 69.67 \pm 2.766 \text{ \AA}$, $D_{\max} = 216 \text{ \AA}$). The obtained SAXS model for the MRPP1-MRPP2-pre-(mt)tRNA^{Ile}-MRPP3(D479A) complex is shown in **Figure 4-41 (bottom)** alongside the SAXS models for the mature wildtype MRPP3 protein alone and the complex of MRPP1-MRPP2-pre-(mt)tRNA^{Ile} (**top**), in order to guide the placement of MRPP3 within the ternary complex.

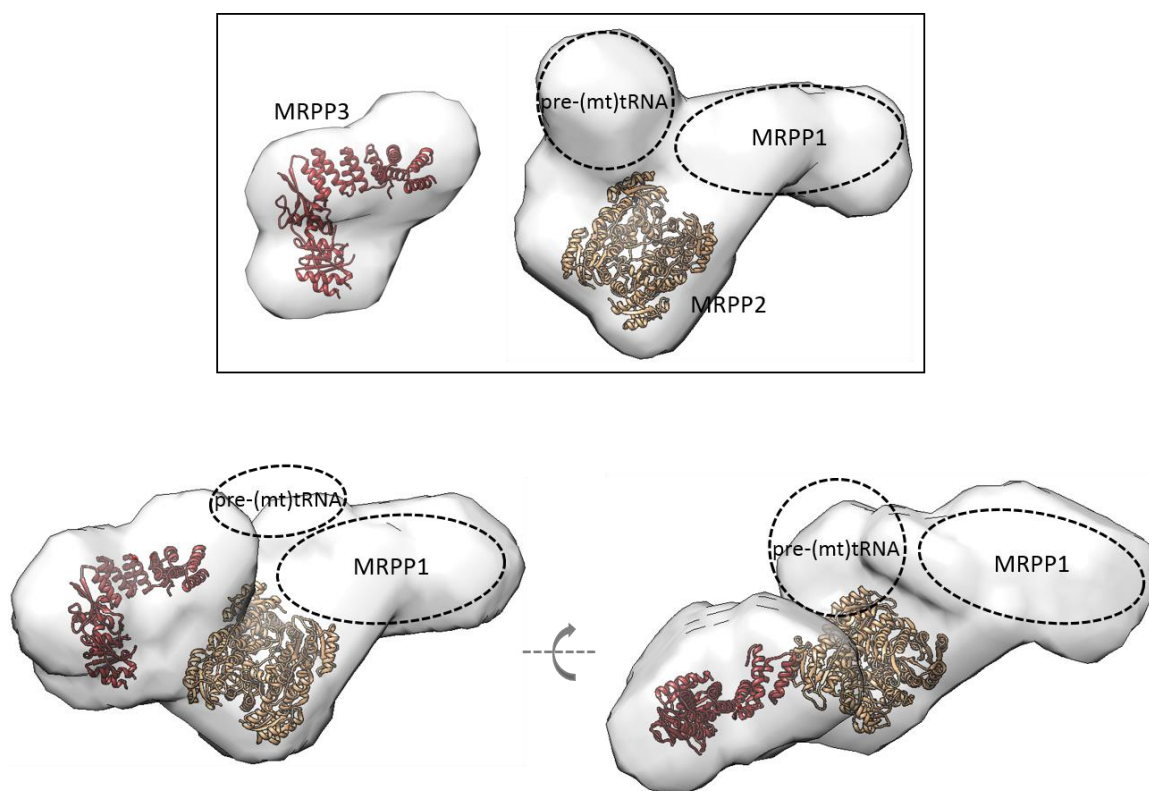


Figure 4-41. SAXS models for the complex of MRPP1-MRPP2-pre-(mt)tRNA^{lle}-MRPP3(D479A), shown alongside those of the mature MRPP3(WT) protein alone and the complex of MRPP1-MRPP2-pre-(mt)tRNA^{lle}.

The binding stoichiometry, as suggested by fitting known structures into the SAXS models, appears to be one MRPP3 molecule per complex of MRPP1(dimer):MRPP2(tetramer):pre-(mt)tRNA^{lle}(one molecule). The binding of one monomer of MRPP3 corresponds with one bound pre-(mt)tRNA substrate. The complex stoichiometry and MRPP3 binding is further discussed in **Chapter 6**.

From the SAXS model of the complex, it seems that MRPP2 and MRPP3 are located in close proximity, and could possibly interact with each other even in the absence of MRPP1. To test this possibility, analytical SEC was performed with full-length MRPP2 and mature wildtype MRPP3 in the absence of pre-(mt)tRNA substrate. The chromatogram and SDS-PAGE analysis of the elution fractions, shown in **Figure 4-42**, revealed no interaction between these two proteins in the absence of MRPP1 and pre-(mt)tRNA substrate.

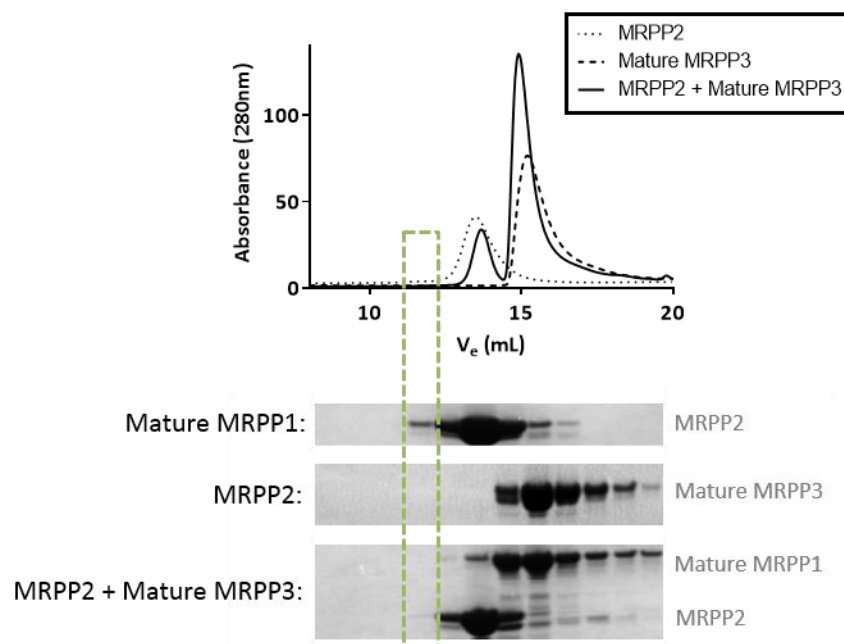


Figure 4-42. Analytical SEC with full-length MRPP2 and mature MRPP3 proteins applied individually and together. The green box indicates the elution fractions in which proteins would co-localise upon complex formation.

5 CHAPTER

DISEASES ASSOCIATED WITH THE RNASE P COMPLEX

Patients have been identified with mutations in the genes coding for MRPP2 (*HSD17B10*) and MRPP3 (*KIAA0391*) giving rise to HSD10 disease (**Section 1.4.1.1**) and Perrault syndrome (**Section 1.4.2.2**), respectively. This chapter presents structural, biochemical and activity studies carried out on proteins containing the patient mutations, to understand their underlying effects towards protein structure and function. The chapter is divided into a first part on HSD10 disease and a second part on Perrault syndrome. The first part is additionally divided into a first part on biochemical studies of all known missense mutations and a second part on activity studies with two novel missense mutations.

HSD10 DISEASE

5.1 Biochemical analysis of 11 missense mutations

At the start of this project, nine missense patient mutations in MRPP2 were known (p.V65A, p.D86G, p.L122V, p.R130C, p.Q165H, p.P210S, p.R226Q, p.N247A, p.E249Q) [36], [112], [115], [118]–[120], [122]–[126], [131]. Whilst this thesis work was ongoing, two novel missense mutations (p.V12L, p.V176M) were identified by our collaborator Prof. Oliver Sass (University of Applied Sciences, Germany). These 11 missense mutations distribute widely across the protein (**Figure 5-1**) with locations in core structural elements, in close proximity to catalytic residues or harbouring monomer-monomer interfaces. The experimental objectives in this thesis chapter are to first characterise the structural properties of the 11 patient mutations at the recombinant protein level, followed by more detailed activity studies *in vitro* of the two novel patient mutations: p.V12L and p.V176M.

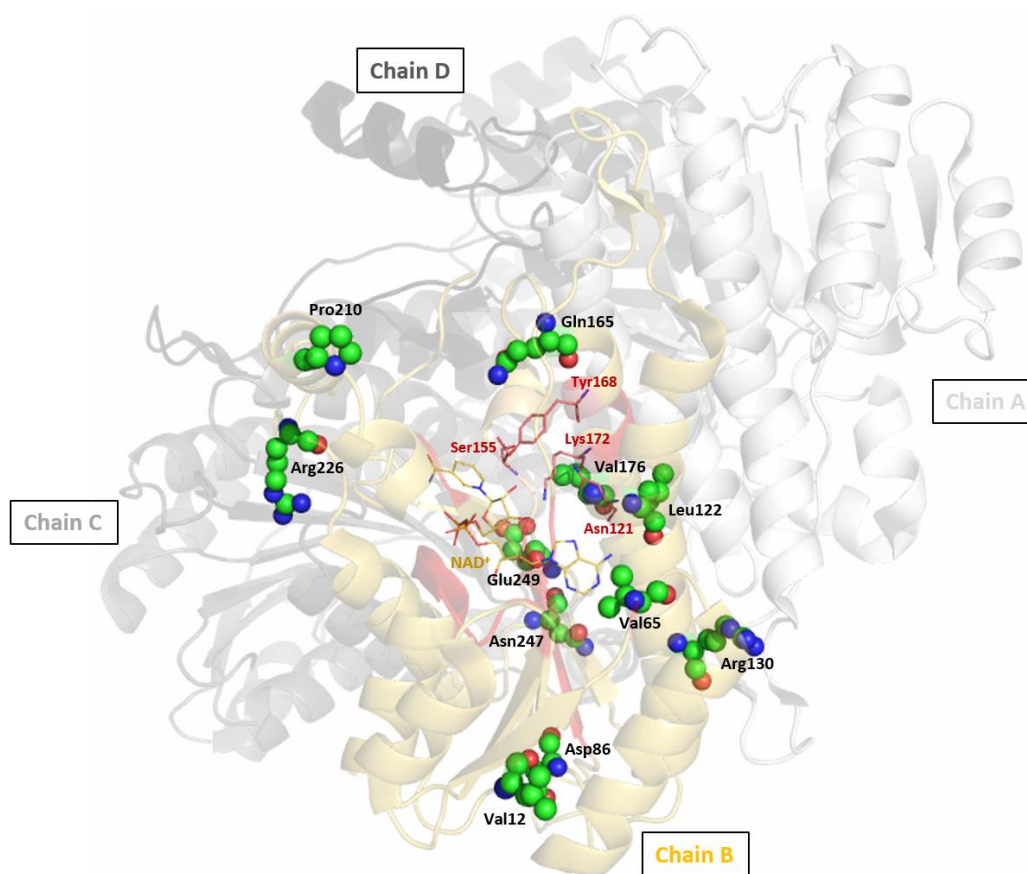


Figure 5-1. Mapping of conserved motifs/residues (red cartoon), catalytic tetrad (red sticks), NAD⁺ cofactor (yellow sticks) and patient mutations (green spheres) on one monomer (chain B, yellow) structure of the MRPP2 tetramer. The other three monomers (A, C and D) are shaded grey, (PDB: 1U7T).

5.1.1 Site-directed mutagenesis

Constructs for MRPP2 carrying the patient missense mutations were generated by site-directed mutagenesis, based on the pNIC28-Bsa4 vector containing full-length His₆-MRPP2 or both mature His₆-MRPP1 and full-length MRPP2 in a bicistronic transcript. The presence of the DNA changes was verified by Sanger sequencing.

5.1.2 Expression and solubility of MRPP2 mutant proteins

E. coli cells containing MRPP2 wildtype or mutant constructs were expressed and purified at small-scale to determine the expression (from total lysate) and solubility (from the elution fraction) levels of each construct, based on SDS-PAGE analysis (**Figure 5-2C**). Intensities of bands in the lysate (**Figure 5-2A**) and elution (**Figure 5-2B**) fractions,

corresponding to MRPP2 proteins, were integrated, normalised against wildtype, and plotted as a bar chart.

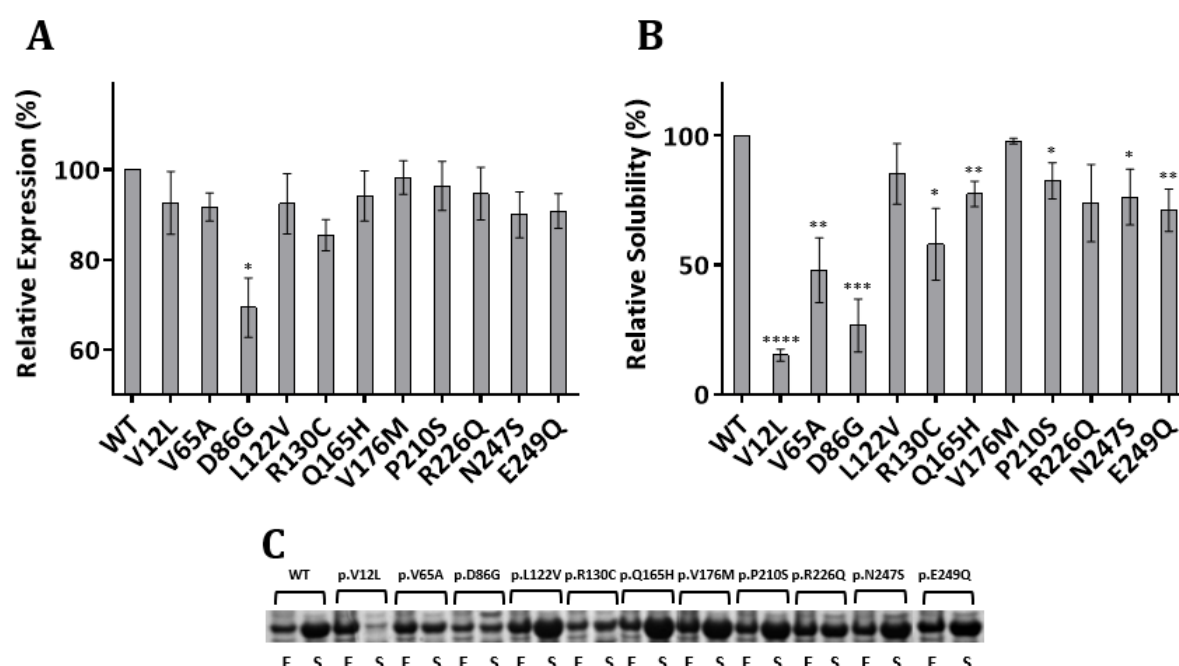


Figure 5-2. Expression and solubility studies for the disease-causing MRPP2 mutant proteins. For this experiment $n = 3$. A) Intensities of protein bands on the SDS-PAGE gel of A) the total lysate and B) the elution fraction from the initial IMAC. Wildtype (WT) is taken as 100%. C) SDS-PAGE analysis of total lysate (E) and elution fractions (S) from initial IMAC for each mutant protein. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.005$, **** = $P < 0.0001$.

SDS-PAGE analysis of the crude extract showed a reduced expression for MRPP2(D86G) compared to wildtype protein (**Figure 5-2A**), suggesting that the p.D86G mutation negatively affect transcription and/or translation of MRPP2 in *E. coli*. The observed reduction in expression agrees with previous findings that this mutant protein could not be detected in fibroblasts from an HSD10 disease patient carrying this mutation [116].

Many mutations in MRPP2 reduces the solubility of the protein (**Figure 5-2B**), with the most severe reduction observed for p.V12L and p.D86G. Val12 and Asp86 are both located in the central β -sheet that forms the core structure of the Rossmann-fold (**Figure 5-3**). This β -sheet is likely involved in stabilising the overall fold of MRPP2 and mutational changes to this element could result in protein mis-folding leading to the observed reduced solubility.

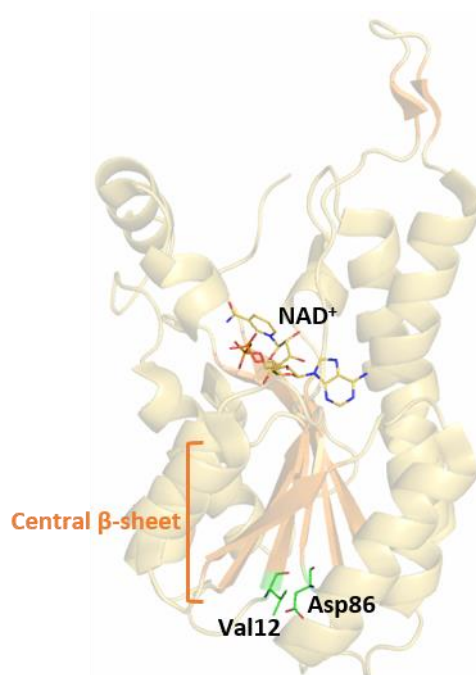


Figure 5-3. Structural location of Val12 and Asp86 (green sticks), sites of the disease-causing mutations p.V12L and p.D86G. The central β -sheet of the MRPP2 Rossmann fold is coloured orange. The cofactor NAD^+ is shown in yellow sticks, (PDB: 1U7T).

5.1.3 Oligomeric state by analytical size exclusion chromatography

The analytical SEC assays performed so far were carried out by absorbance measurements at 280 nm for aromatic amino acids. The absorbance signal for wild-type His₆-MRPP2 protein is low at this wavelength, due to the extension coefficient of $\epsilon = 5960 \text{ M}^{-1}\text{cm}^{-1}$ (**Section 7.4.1, Equation 1**). Therefore, to perform analytical SEC with wild-type and mutant MRPP2, purified from small-scale cultures, an HPLC system able to measure absorbance at 215 nm, corresponding to the protein peptide bonds, was used in connection with a high resolution analytical SEC column (Sepax SRT SEC-300). The obtained chromatograms for mutant proteins are shown in **Figure 5-4**.

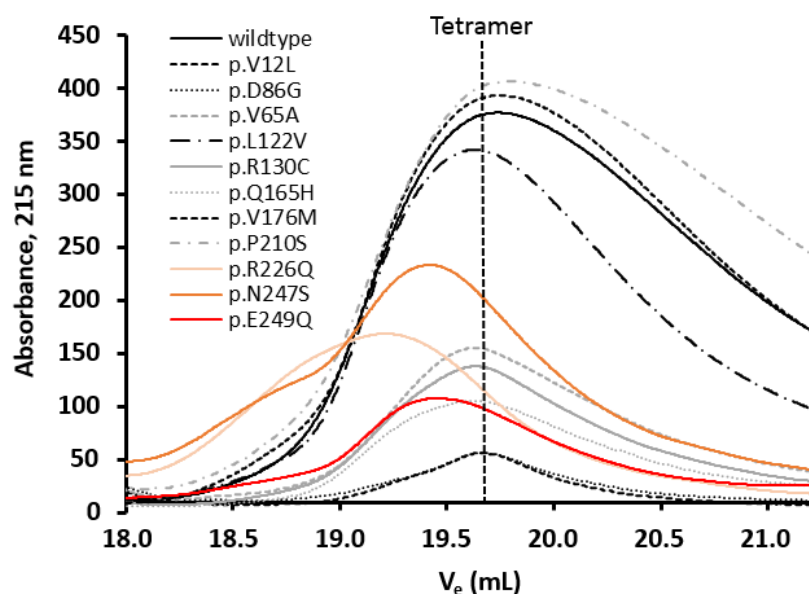


Figure 5-4. Analytical SEC assay analysing the oligomeric state for the mutant MRPP2 proteins giving rise to HSD10 disease in different patients. Proteins of known MWs were also applied to the column to create a standard curve of $\text{LOG}^{10}(\text{MW})$ against V_e for MW-determination ($y = -0.2673x + 10.441$ with $R^2 = 0.9973$).

Based on the standard curve, the V_e for wildtype MRPP2 of 19.79 mL corresponds to a tetrameric protein of 109227 Da (theoretical MW of monomer: 29475.9 Da) as expected from the crystal structure. The obtained V_e -values for the mutant proteins revealed that MRPP2 protein carrying the p.R226Q, p.N247S or p.E249Q mutations formed higher oligomeric species (**Figure 5-4**, coloured graphs). These residues are all located in the BxC dimer interface of the tetramer, and their side-chain atoms are involved in hydrogen bond formation (**Figure 5-5A** and **B**). For Asn247 and Glu249, these hydrogen bonds involve residues from the opposite monomer (chain C), while Arg226 forms intra-monomer hydrogen bonds. Therefore, the Asn247-to-Ser and Glu249-to-Gln changes could have a more destabilising effect on the tetramer than mutations in Arg226, leading to faster aggregation/precipitation. This could offer an explanation why only MRPP2(N247S) and MRPP2(E249Q), but not MRPP2(R226Q), exhibited lowered solubility already after IMAC purification.

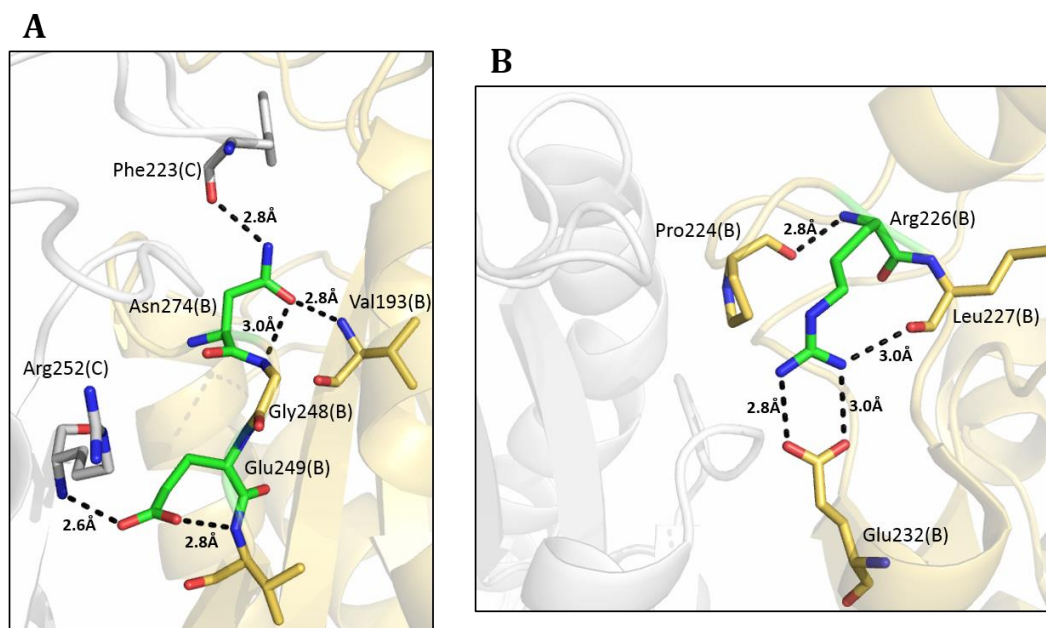


Figure 5-5. Structural locations of Arg226, Asn247 and Glu249 (green sticks), sites of the disease-causing p.R226Q, p.N247S and p.E249Q mutations respectively. Chain C of the MRPP2 tetramer is coloured white and chain B yellow, (PDB: 1U7T).

To further test if the overall quaternary fold of these three MRPP2 mutant proteins differ from wildtype MRPP2, a limited proteolysis assay was performed in which protein was incubated with an increasing amount of the protease trypsin at a set time, and the resulting proteolytic products were analysed by SDS-PAGE (**Figure 5-6**).

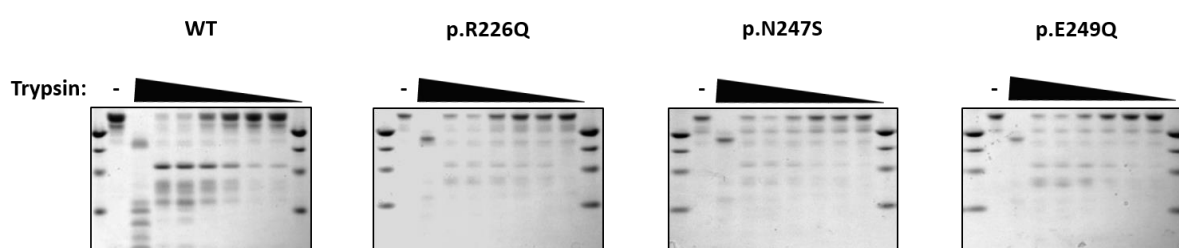


Figure 5-6. Limited proteolysis using trypsin for MRPP2 wildtype (WT), p.R226Q, p.N247S and p.E249Q. Trypsin was used in a 1:1, 1:50, 1:75, 1:100, 1:500, 1:750 and 1:1000 ratio of protein-to-trypsin.

The three MRPP2 mutant proteins exhibited a similar proteolytic pattern, with fewer cleavage products compared to wildtype protein, especially at the highest concentration of trypsin. Trypsin cleaves the peptide bond after the hydrophilic residues Lys and Arg, often found on the protein surface. The fewer cleavage products observed for MRPP2(R226Q), MRPP2(N247S) and MRPP2(E249Q), compared to wildtype protein, suggests that surface

exposed Lys and Arg residues in wildtype protein, are buried in the mutant structures, possibly due to the observed formation of higher oligomeric species/aggregation.

5.1.4 Thermal stability by differential scanning fluorimetry

The SEC-purified proteins of MRPP2 wildtype and mutants were utilised in a DSF assay to determine their thermal stability in the as-purified form. The T_m -values obtained from the DSF assay were plotted in a bar chart (**Figure 5-7**), annotated with positive T_m shifts (green stars) and negative T_m shifts (red stars), compared to wildtype.

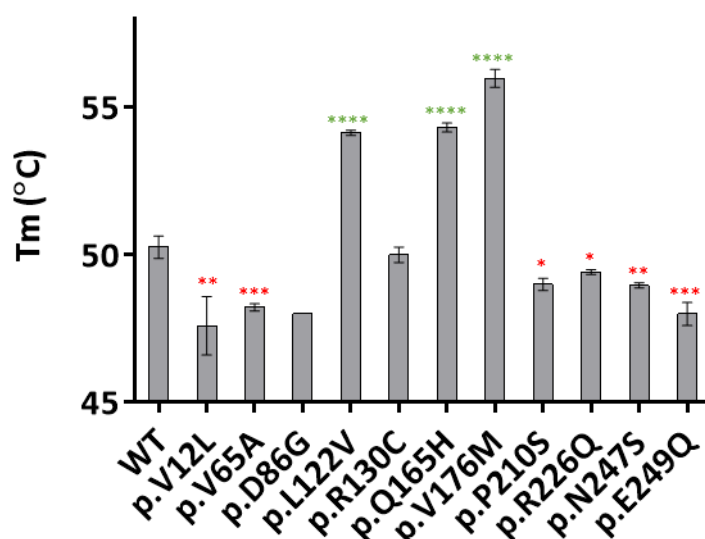


Figure 5-7. T_m values for MRPP2 mutant proteins in the as-purified state. For this experiment $n=3-5$ (p.D86G is $n=1$). Positive ΔT_m values relative to wildtype are indicated with green stars, and negative values are indicated with red stars. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.005$, **** = $P < 0.0001$.

Many as-purified MRPP2 mutant proteins were found to be less thermostable than wildtype, however, MRPP2 protein carrying either of the p.L122V, p.Q165H or p.V176M mutations showed an increased thermal stability. These mutated residues are located in the two long α -helices (αE and αF) (**Figure 5-8**) harbouring three of the catalytic residues, Asn121, Tyr168 and Lys172 (**Figure 5-8**), and residues involved in the proton-relay system as well as substrate binding further described in **Section 6.5.2**.

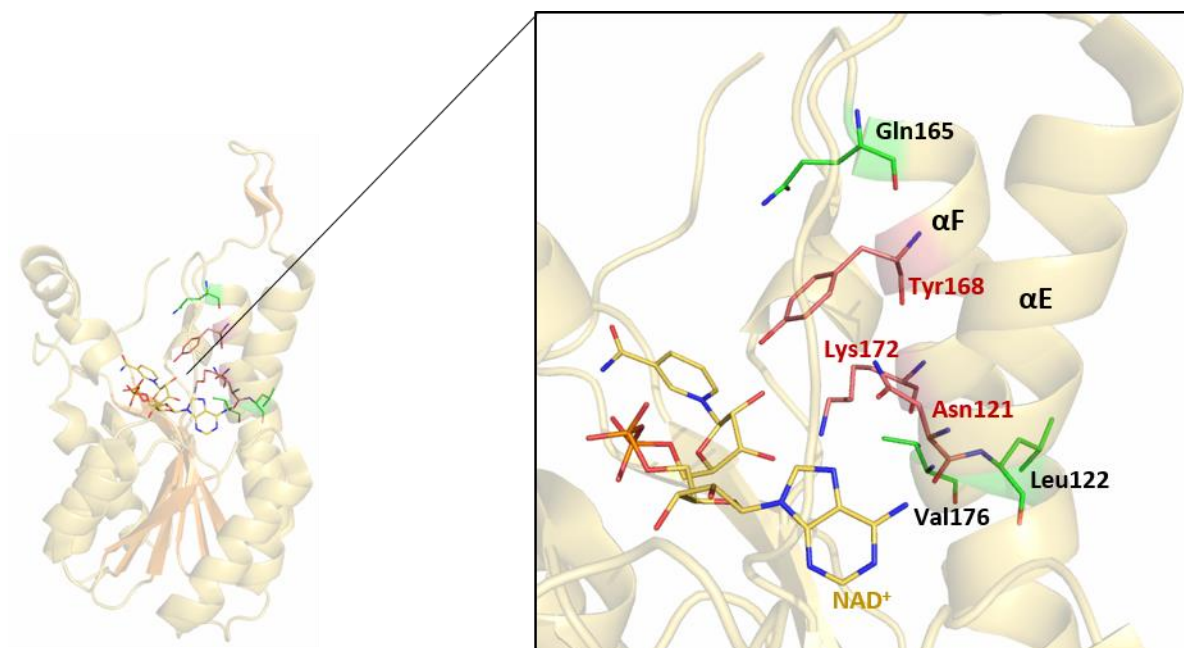


Figure 5-8. Structural locations of Leu122, Gln165 and Val176 (green sticks), sites of the disease-causing p.L122V, p.Q165H and p.V176M mutations. Three residues from the catalytic tetrad are shown in red (Asn121, Tyr168 and Lys172) along with the NAD⁺ cofactor in yellow sticks, (PDB: 1U7T).

The increased thermal stability observed for these mutant proteins could be a result of global conformational changes originating from the mutations within the two long helical regions. To test this possibility, the limited proteolysis assay was performed using trypsin (**Figure 5-9**) on MRPP2 protein with either p.L122V, p.Q165H or p.V176M mutations. All three mutant proteins exhibited a similar proteolytic pattern to wildtype protein, indicating no global structural changes to the protein, and suggesting that the increased thermal stability is caused by smaller local changes.

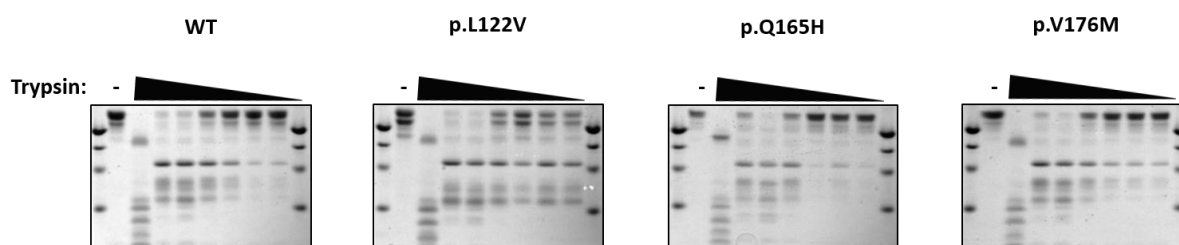


Figure 5-9. Limited proteolysis using trypsin for MRPP2 wildtype (WT), p.L122V, p.Q165H and p.V176M. Trypsin was used in similar ratios as shown for Figure 5-6.

5.1.5 Ligand binding of MRPP2 mutant proteins

The DSF assay was further used to determine the ligand binding properties of each mutant compared to the wildtype protein. Two possible ligands could be tested – either the

natural cofactor substrate NAD^+ or its product NADH – for the SDR enzymatic reactions of MRPP2. The DSF assay was initially performed for wildtype MRPP2 in a dose-responderent manner, to determine which ligand resulted in a larger T_m shift and the optimal concentration for further investigation of mutant proteins. The thermal unfolding-curve for wildtype MRPP2 as-purified and after addition of either NAD^+ or NADH (**Figure 5-10**) revealed T_m -values of $50.09 \pm 0.08^\circ\text{C}$ for the as-purified protein, $50.28 \pm 0.03^\circ\text{C}$ in the presence of NAD^+ , and $60.33 \pm 0.14^\circ\text{C}$ in the presence of NADH. Under the assay condition, NAD^+ did not result in ΔT_m , even at the highest $[\text{NAD}^+]$ of 3 mM [115], whereas binding to NADH gave a ΔT_m of $10.24 \pm 0.16^\circ\text{C}$ even at $[\text{NADH}] = 1 \text{ mM}$. Based on these findings, NADH was chosen for subsequent investigations with MRPP2 mutant proteins.

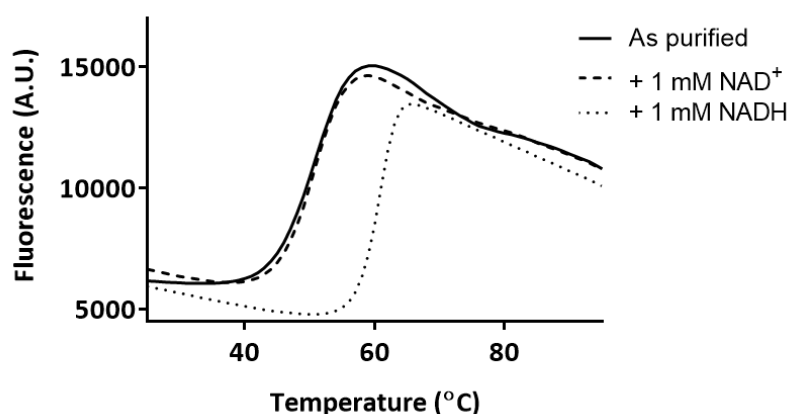


Figure 5-10. Thermal unfolding curve for MRPP2 in the absence of ligand (as purified) and in the presence of NAD^+ or NADH.

The DSF assay was next performed with MRPP2 mutant proteins alongside wildtype in the presence of NADH. ΔT_m -values were determined from the difference in NADH-bound and as-purified forms for each protein, and plotted in a bar chart (**Figure 5-11**).

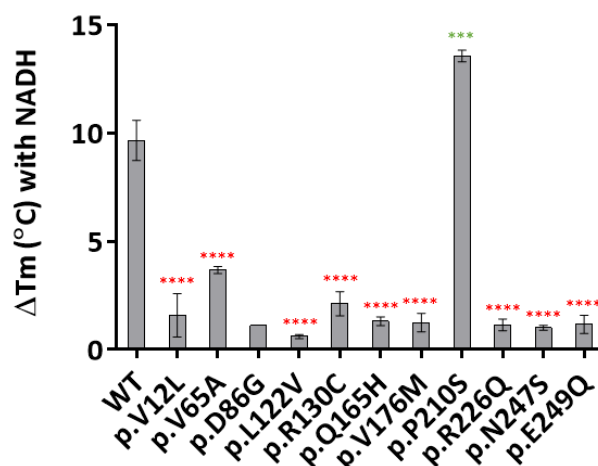


Figure 5-11. DSF assay analysis in the presence of NADH for MRPP2 wildtype and mutant proteins. For this experiment $n = 3 - 6$ (p.D86G is $n = 1$). Positive ΔT_m values are indicated with green stars and negative values are indicated with red stars. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.005$, **** = $P < 0.0001$.

Most mutant proteins showed significantly lower T_m shifts with NADH than wildtype, likely indicating a decreased binding affinity for NADH in these mutant proteins. Only MRPP2(P210S) retained – at least – wildtype level thermal stability in the presence of NADH. In fact, this mutant protein exhibited a larger ΔT_m than wildtype, indicative of more structural rigidity in the mutant protein, compared to wildtype, after binding of the ligand. Pro210 is located in a segment (residue 208-217) that does not display any electron density in wildtype MRPP2 crystal structures (PDB: 2O23) unless a Cys214-to-Arg mutation is introduced (PDB: 1U7T) (**Figure 5-12**), indicative of structural mobility in this segment. The p.P210S mutation could have a similar, stabilising effect for this segment as p.C214R, after binding of NADH, resulting in the observed increased ΔT_m compared to wildtype. In the MRPP2 paralogue *Scully* from *Drosophila* (1E3W, 44.6, 0.5 Å over 251 C $^\alpha$, 88%) a corresponding segment became ordered in the crystal upon binding of fatty acid metabolites [185], suggesting that the p.P210S mutation could also affect substrate binding.

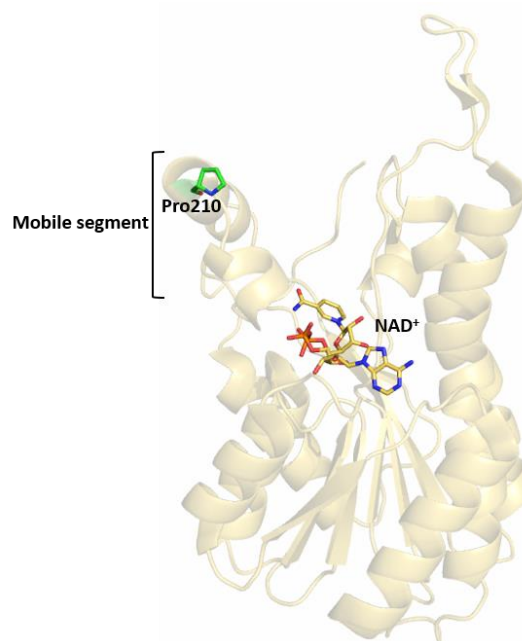


Figure 5-12. The structural location of Pro210 found to be mutated to Ser (p.P210S) in some patients suffering from HSD10 disease, (PDB: 1U7T).

The DSF assay was next performed with the MRPP2(P210S) mutant protein in the presence of the cofactor substrate NAD^+ to compare with wildtype protein. As observed for NADH, NAD^+ also resulted in a T_m -shift for the mutant protein (**Figure 5-13**), contrasting with no T_m -shift observed for the wildtype protein, in the presence of NAD^+ . This again supports a different structural arrangement of the cofactor-bound MRPP2(P210S), compared to wildtype, which makes the mutant protein more rigid. It would be interesting for future studies to determine the overall structure of this mutant protein in the presence of both cofactor forms.

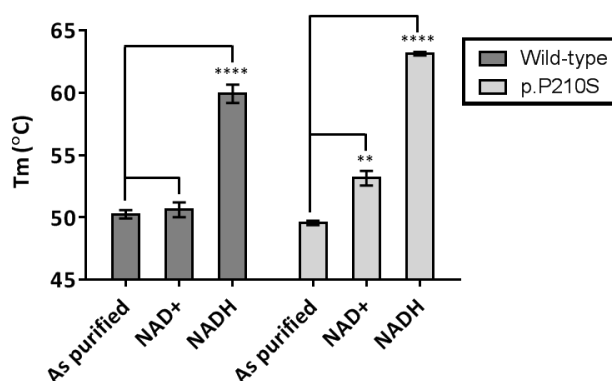


Figure 5-13. T_m values for MRPP2(P210S) and wildtype MRPP2 proteins in the presence of the cofactor substrate (NAD^+) and product (NADH). For this experiment $n = 3 - 5$. **= $P < 0.01$, ****= $P < 0.0001$.

5.2 SDR activity of MRPP2 with novel mutations

The novel mutations p.V12L and p.V176M, identified by our collaborator, have not been characterised at the protein level. This section describes their effect, *in vitro*, on the MRPP2 protein with respect to: 1) SDR activity, 2) m¹R9 MTase activity and 3) RNase P activity.

5.2.1 Large- scale expression and purification of V12L and V176M mutants

The constructs of MRPP2(V12L) or MRPP2(V176M) were expressed at large scale, purified and verified as for other constructs (**Figure 5-14** and **Figure 5-15**).

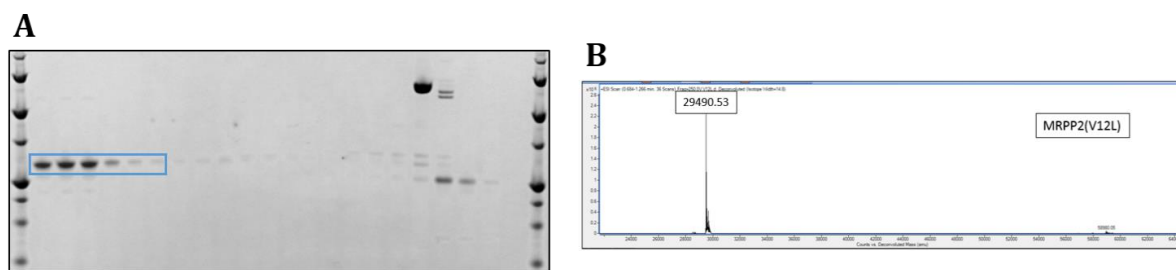


Figure 5-14. Example of a large scale purification of the MRPP2(V12L) mutant protein.

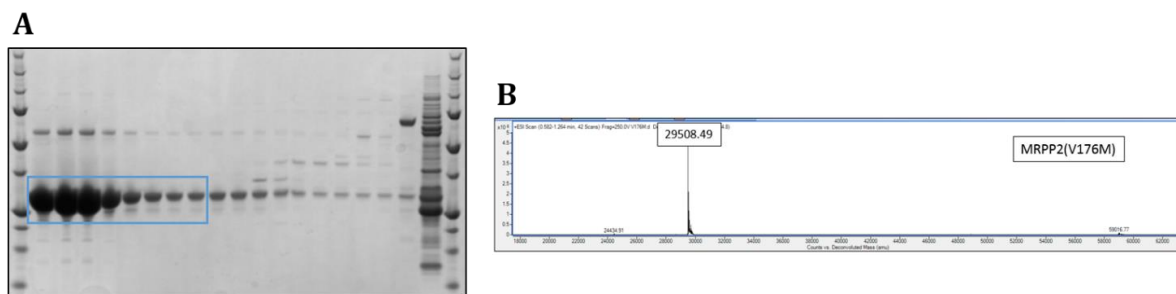


Figure 5-15. Example of a large scale purification of the MRPP2(V176M) mutant protein.

The SDR activity of the purified MRPP2 wildtype and mutant proteins was established with two different substrates, reflecting the involvement of the MRPP2 enzyme in fatty acid degradation and steroid conversion.

5.2.2 L-3-HAD activity of MRPP2

An important *in vivo* function of MRPP2 is its enzymatic L-methyl-3-HAD activity towards short *branched-chain* acyl-CoA thioesters, such as L-2-methyl-3-hydroxybutyryl-

CoA, which accumulates from β -oxidation of long branched-chain fatty acids (**Section 1.1.1.2**). Unfortunately, this substrate was not commercially available and thus the unmethylated (short straight-chain fatty acid) acyl-CoA thioester was used (L-3-OH-but-CoA) (**Figure 5-16**), as this has been shown to be a substrate for the L-3-HAD activity of MRPP2 *in vitro*.

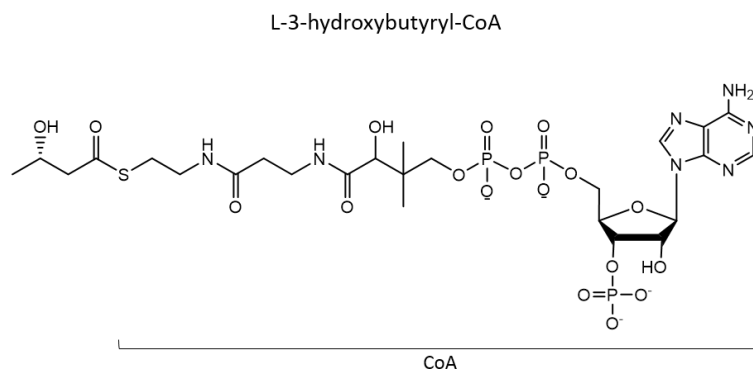
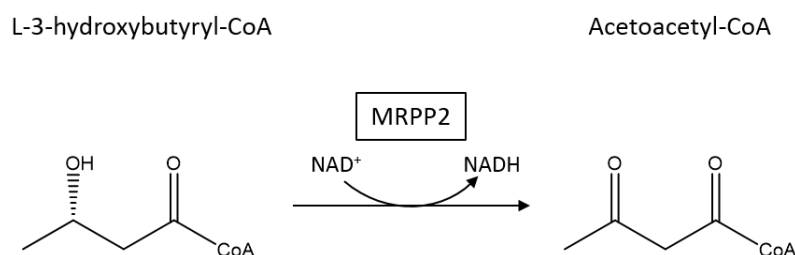


Figure 5-16. L-3-hydroxybutyryl-CoA (L-3-OH-but-CoA).

This substrate is synthesised as a racemic mixture of the D- and L-enantiomers (also known as *R* and *S*, respectively), but will be depicted as the biologically relevant L-enantiomer substrate throughout this chapter [6], [186]. The oxidation of this short *straight-chain* (unmethylated) substrate to acetoacetyl-CoA using NAD^+ as the cofactor is shown in **Reaction Scheme 2**.



Reaction Scheme 2. The oxidation of L-3-OH-but-CoA to acetoacetyl-CoA by MRPP2 utilising NAD^+ .

Purified wildtype MRPP2 and mutant MRPP2(V12L) and MRPP2(V176M) were used to run the spectrophotometric activity assay on the L-3-OH-but-CoA substrate with NAD^+ cofactor. Kinetic parameters, K_M and $V_{\max}$, were obtained by fitting the data to the Michaelis-Menten equation and k_{cat} was calculated with four active sites indicated for

MRPP2 (**Figure 5-17**). As for previous SDR activity assays presented here, the system exhibit $[E] > [S]$ at the lower concentrations (below $3.7 \mu\text{M} = [E]$) and kinetic constants should therefore be utilised only for comparison of enzymes tested in this work.

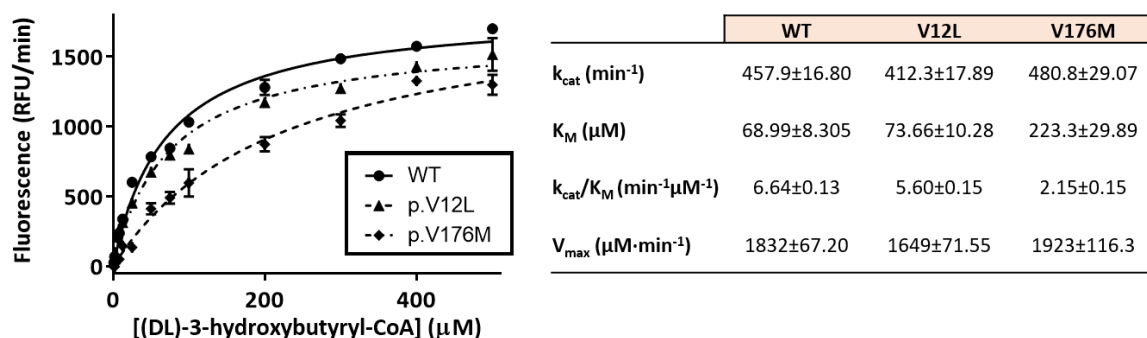


Figure 5-17. Michaelis-Menten plot for the catalysis of (DL)-3-hydroxybutyryl-CoA by MRPP2 wildtype and mutant proteins using NAD^+ . RFU: relative fluorescence units.

All three proteins were active towards the acyl-CoA thioester L-3-OH-but-CoA and displayed similar k_{cat} values, but the K_M of MRPP2(V176M) was increased 3-fold leading to an overall reduced catalytic efficiency (k_{cat}/K_M) for this mutant protein, compared to wildtype.

5.2.3 3α -HSD activity of MRPP2

The 3α -HSD activity of MRPP2 in conversion of allopregnanolone to 5α -dihydroprogesterone (5α -DHP) using NAD^+ , was previously investigated for wildtype MRPP2 and the complex of MRPP1-MRPP2 (**Section 3.3.3**) using the spectrophotometric activity assay. The 3α -HSD activity was further tested for MRPP2(V12L) and MRPP2(V176M), also using allopregnanolone as substrate.

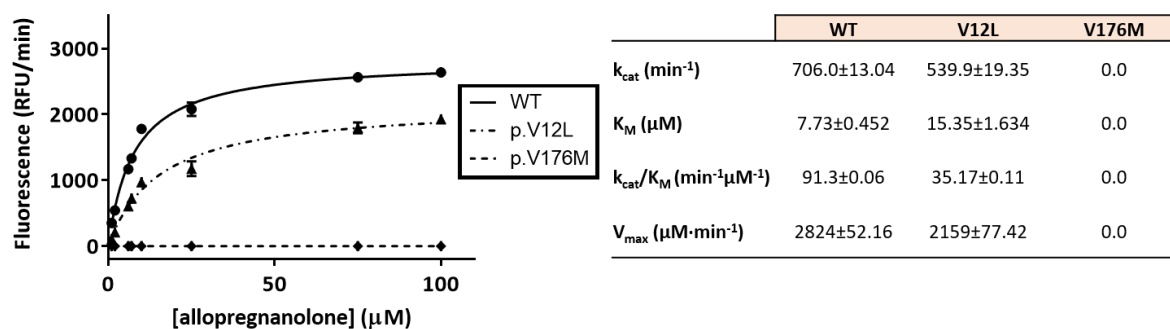


Figure 5-18. Michaelis-Menten plot for the catalysis of allopregnanolone by MRPP2 wildtype and mutant proteins using NAD^+ . RFU: relative fluorescence units.

The kinetic data (**Figure 5-18**) gave rise to catalytic efficiencies (k_{cat}/K_M) of $91.3 \pm 0.06 \text{ min}^{-1}\mu\text{M}^{-1}$ for wildtype protein compared to $35.17 \pm 0.11 \text{ min}^{-1}\mu\text{M}^{-1}$ for p.V12L. No activity was observed for MRPP2(V176M) with this substrate.

5.3 Moonlighting activities of MRPP2 with novel mutations

Having established the effect of the p.V12L and p.V176M mutations on the L-3-HAD and 3α -HSD activities of MRPP2, subsequent studies were focussed on the moonlighting properties of MRPP2, in the context of the binary MRPP1-MRPP2 complex for its $m^1\text{R9}$ MTase activity, and the ternary MRPP1-MRPP2-MRPP3 complex for its RNase P activity.

5.3.1 Complex assembly with MRPP1 and mutant MRPP2 proteins

Wildtype MRPP2, MRPP2(V12L) and MRPP2(V176M) were each co-expressed with N-terminally His₆-tagged mature MRPP1 and purified as previously described for the wildtype construct. The co-expressed wildtype and mutant protein complexes were purified by IMAC using Ni-NTA and the first elution fraction from each purification was analysed by SDS-PAGE (**Figure 5-19**). In the IMAC procedure, His₆-tagged MRPP1 (**Figure 5-19, blue**) pulled down similar amounts of untagged wildtype MRPP2, MRPP2(V12L) and MRPP2(V176M) (**Figure 5-19, orange**), indicative of complex formation for both mutants at levels comparable to wildtype protein.

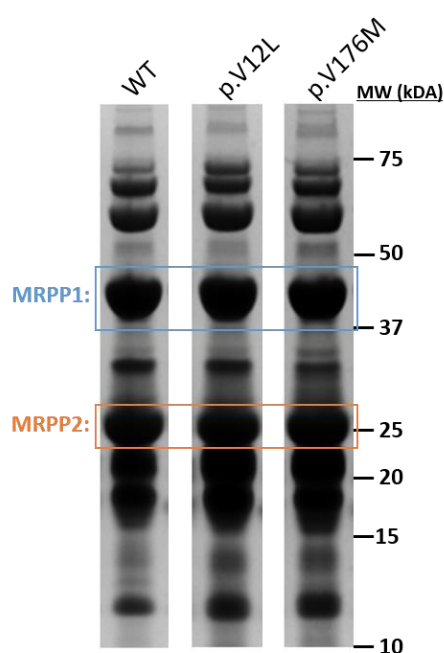


Figure 5-19. SDS-PAGE analysis of the first elution fraction from the initial IMAC for MRPP1-MRPP2(WT), MRPP1-MRPP2(V12L) and MRPP1-MRPP2(V176M). Mature, His₆-tagged MRPP1 is indicated in **blue** along with untagged full-length MRPP2 mutant or wildtype protein in **orange**. Size of molecular weight (MW) markers is shown to the right.

The elution fraction containing the pulled down complexes of His₆-MRPP1 with untagged MRPP2 wildtype or mutant proteins, were incubated overnight at 4°C to evaluate the stability of the formed complexes. Subsequently, equal amounts of the complexes were loaded onto an S6 column for SEC analysis. The chromatograms and corresponding SDS-PAGE analyses of elution fractions (**Figure 3-5**) showed complex formation for MRPP1 protein with both wildtype MRPP2 and MRPP2(V12L) (**Figure 3-5, green box**), with the presence of some uncomplexed proteins (**Figure 3-5, red box**) for both. For the complex of MRPP1 with MRPP2(V176M), most of the proteins were found in fractions corresponding to individual proteins, indicating that the p.V176M mutation, with time, significantly alters the equilibrium of complex formation towards the dissociated, uncomplexed proteins.

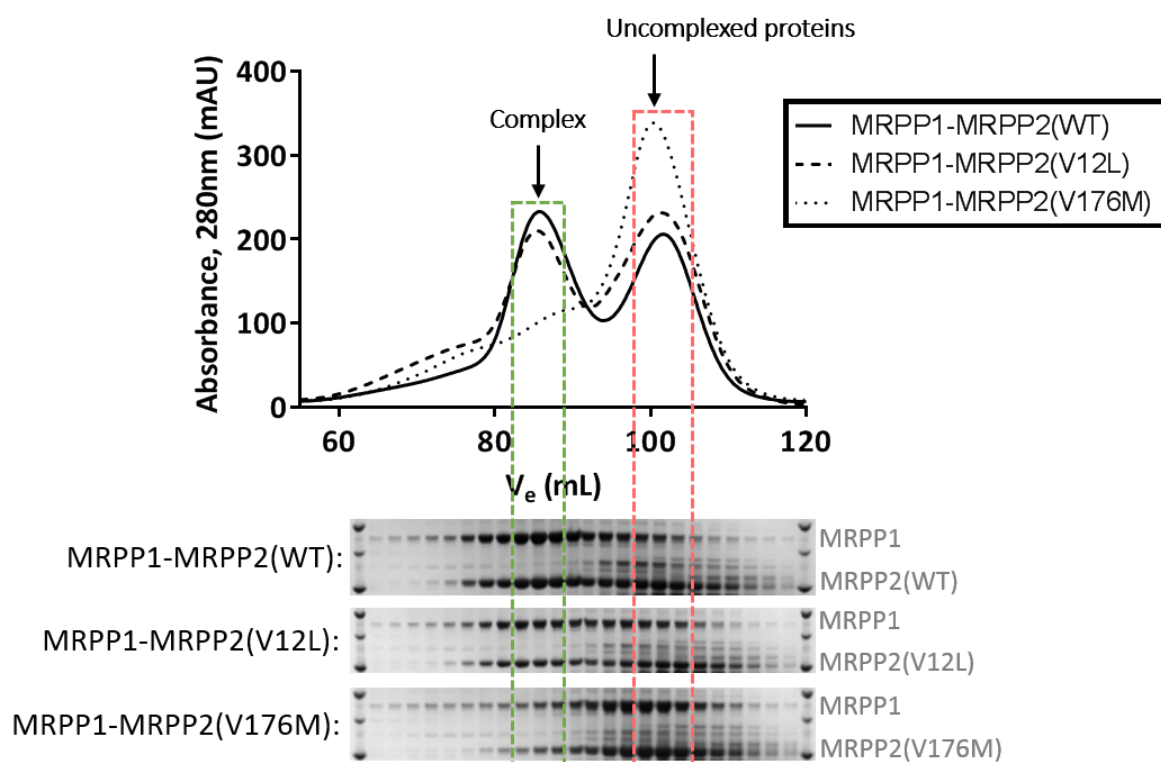


Figure 5-20. SEC with the mature MRPP1 protein co-expressed with either wildtype MRPP2, MRPP2(V12L) or MRPP2(V176M). The green box indicates the elution fractions in which proteins should be co-localising if they form a complex. The red box indicates the elution fractions in which proteins should be found if they are present as separate species in solution.

The SEC fractions corresponding to complexes of MRPP1 with wildtype MRPP2 or MRPP2(V12L) were pooled and subjected to cation-EX, resulting in pure protein of both complexes. For the partially dissociated proteins of MRPP1 with MRPP2(V176M), fractions of both the individual proteins and remaining complex were pooled and subjected to cation-EX. This resulted in binding and elution of only MRPP1 (mutant MRPP2 protein in flow through sample), suggesting that any residual complexed proteins had now dissociated. The pure, mature MRPP1 protein from the cation-EX step was pooled and added with separately purified MRPP2(V176M) from the flow-through, in a 2x(MRPP1):4x(MRPP2(V176M)) stoichiometry to reconstitute a complex for further activity studies.

5.3.2 m¹R9 MTase activity of MRPP1 with mutant MRPP2 proteins

The m¹G9 and m¹A9 MTase activities were determined for mature MRPP1 in the presence of MRPP2(V12L) or MRPP2(V176M), with complex of wildtype MRPP2 included as positive control. m¹G9 activity was tested with *in vitro* transcribed pre-(mt)tRNA^{Ile} as substrate, and m¹A9 activity with *in vitro* transcribed pre-(mt)tRNA^{Val}. The obtained cpm-values for each complex, from which cpm-values for tRNA-alone samples were subtracted as background, are presented as a bar chart (**Figure 5-21**).

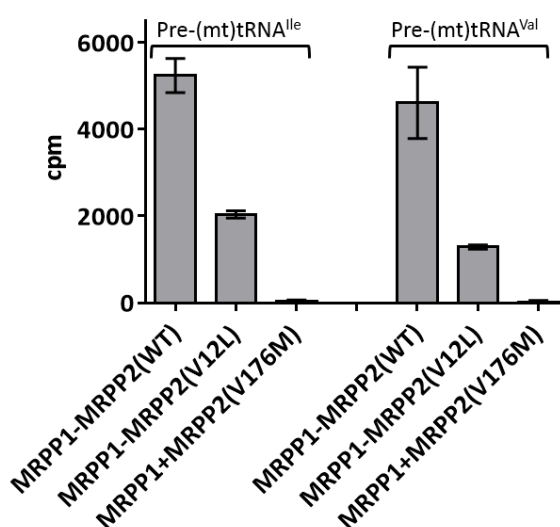


Figure 5-21. m¹G9 and m¹A9 MTase activity of the co-expressed MRPP1-MRPP2(WT), MRPP1-MRPP2(V12L) and reconstituted MRPP1+MRPP2(V176M) complexes.

The assay showed a 2.5-fold reduction in m¹G9 MTase activity and a 3.5-fold reduction in m¹A9 MTase activity for MRPP1-MRPP2(V12L), compared to the protein complex with wildtype MRPP2. For the MRPP1-MRPP2(V176M) complex, almost no MTase activity was observed for formation of either m¹G9 or m¹A9.

5.3.3 RNase P activity of MRPP3 with mutant MRPP2 proteins

To understand if the p.V12L and p.V176M mutations influence RNase P activity of the MRPP1-MRPP2-MRPP3 complex, the assay was carried out by incubating the reconstituted complexes of MRPP1-MRPP2(wildtype/mutants) with mature MRPP3 in the presence of Mg²⁺ and *in vitro* transcribed pre-(mt)tRNA^{Ile} substrate. A negative control was

included by pre-incubating the co-expressed MRPP1-MRPP2(WT) and MRPP3 proteins with EDTA before addition of pre-(mt)tRNA^{Ile} substrate. UREA-PAGE analysis of unprocessed and processed (mt)tRNA fragments (**Figure 4-32**) revealed that both mutations reduced RNase P activity of the ternary complex. The ternary complex with MRPP2(V12L) yielded $29.14 \pm 5.72\%$ processed pre-(mt)tRNA^{Ile} which is a 2-fold decrease from the ternary complex with wildtype MRPP2 that yielded $59.93 \pm 1.09\%$ processed pre-(mt)tRNA^{Ile}. This reduction in activity is comparable to the reduction for the m¹R9 MTase activity of the MRPP1-MRPP2(V12L) complex.

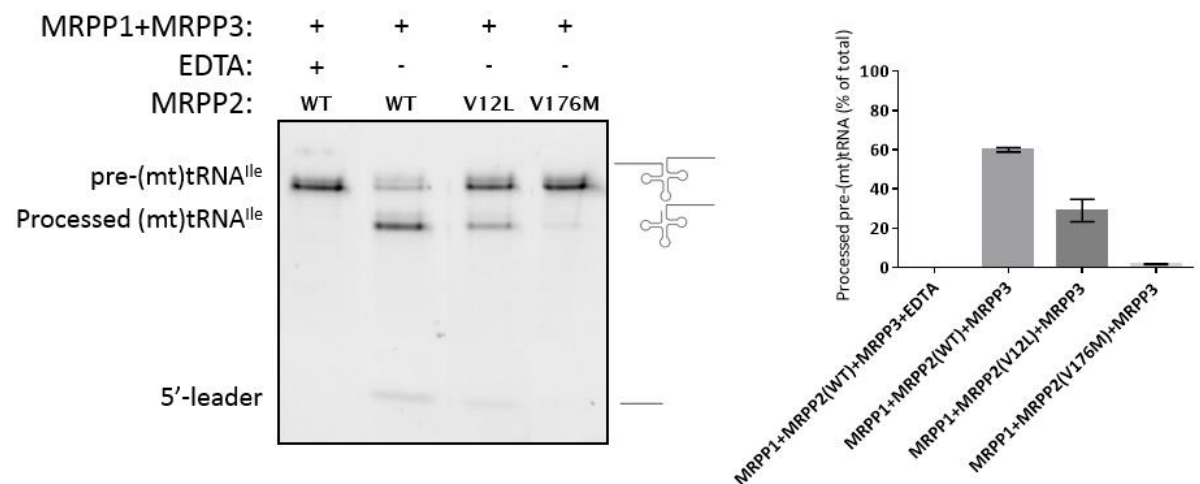


Figure 5-22. RNase P activity of MRPP3 towards mitochondrial pre-(mt)tRNA^{Ile}, in the presence of MRPP1-MRPP2 complexes with either wildtype MRPP2, p.V12L or p.V176M.

The p.V176M mutation in MRPP2 resulted in reduced processing of (mt)tRNA^{Ile} to only $1.86 \pm 0.22\%$ of wildtype, following the trend for the MTase activities. This agrees with the findings in **Chapter 4 (Section 4.9.1)** that mature MRPP3 protein requires an intact MRPP1-MRPP2 complex for catalysis.

PERRAULT SYNDROME

5.4 MRPP3 mutations causing Perrault Syndrome

During the course of this work, a novel patient mutation causing the Perrault Syndrome, p.A485V in MRPP3, was communicated by our collaborator Prof. William Newman (University of Manchester, UK). The Ala485 residue (**Figure 5-23**, purple) is highly conserved across NYN proteins [100], and is located in close proximity to the four active site Asp residues (**Section 4.5.2**) of MRPP3 (**Figure 5-23**). Therefore, a mutation of this residue could affect the RNase P activity. Studies into the biochemical properties of mutant protein, ternary complex formation, and RNase P activity of this mutant will be the subject in the following sections.

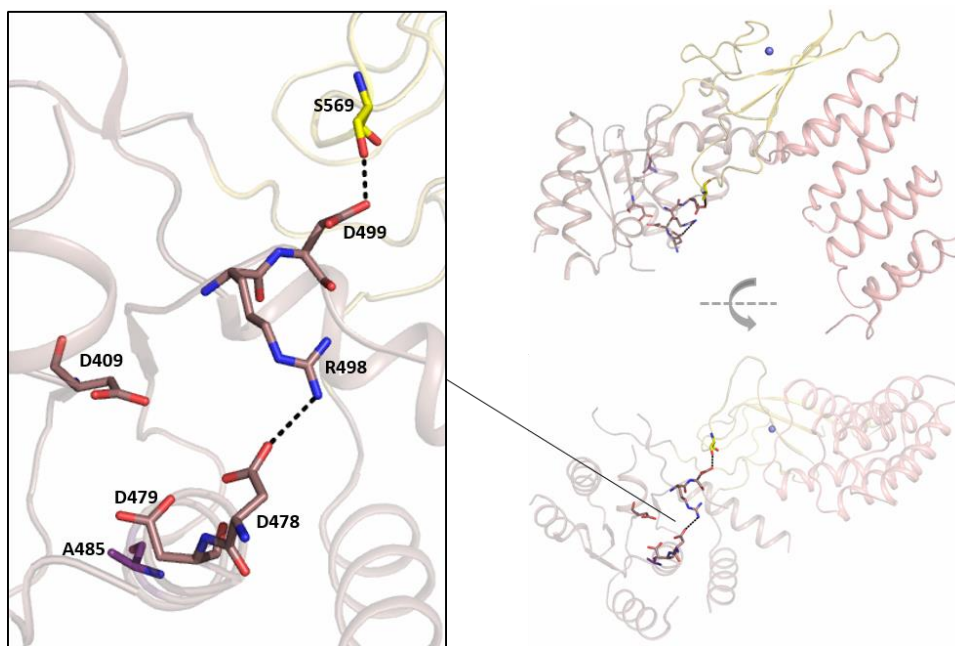


Figure 5-23. Location of the MRPP3 patient mutation p.A485V in the MRPP3 protein, (PDB: 4XGL).

5.5 Cloning, expression and purification

The Ala485-to-Val change in MRPP3 was generated by site-directed mutagenesis on the plasmid encoding mature MRPP3 (residues 51-583) with an N-terminal His₆-tag and a TEV cleavage site. The presence of the mutation was identified by Sanger sequencing.

The mature MRPP3(A485V) was expressed and purified at large scale as for wildtype MRPP3 (**Section 4.3**), and verified by tryptic digest-MS/MS and ESI-MS (**Figure 5-24**). The MRPP3(A485V) protein exhibited *in vitro* expression and solubility levels similar to wildtype.

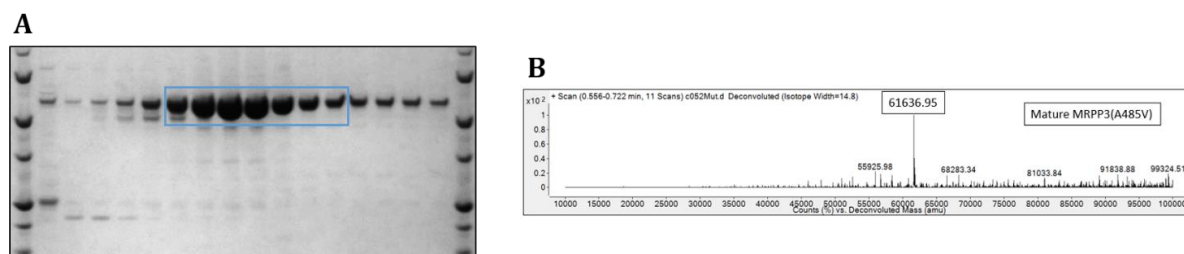


Figure 5-24. A) SDS-PAGE analysis from the last purification step of mature MRPP3(A485V) protein. B) ESI-MS of the purified protein sample.

5.6 Biochemical analysis

5.6.1 Oligomeric state of the MRPP3 mutant protein

Purified, untagged, MRPP3(A485V) was subjected to analytical SEC (S200) (**Figure 5-25**) resulting in a V_e of 14.99 mL, corresponding to a MW of 50399.7 Da. This MW is equivalent to a monomeric protein (theoretical mass of monomer: 61636.95 Da), as expected from the monomeric state of wildtype MRPP3. The wildtype and mutant proteins further showed a comparable elution profile indicative of similar structural arrangements.

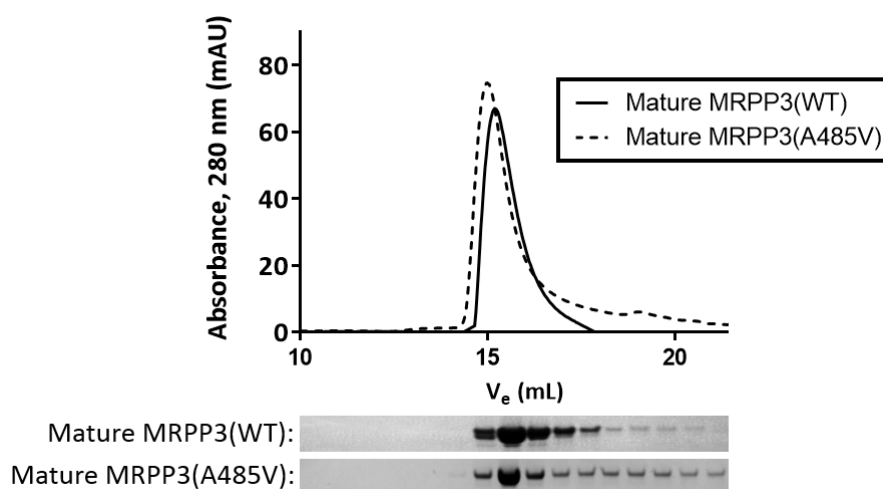


Figure 5-25. Analytical SEC profile for MRPP3(WT) and MRPP3(V485A).

5.6.2 Thermal stability of the MRPP3 mutant protein

The thermal stability of MRPP3(A485V) was compared to wildtype MRPP3 using DSF. Both proteins were applied in the standard buffer used for DSF (150 mM NaCl, 10 mM HEPES pH 7.5) and in a buffer identified to yield the most thermostable $\Delta 94$ -MRPP3 protein (50 mM MES pH 6.5, 5% glycerol, 100mM NaCl, 100mM (NH₄)₂SO₄) (Section 4.5.1.3). The obtained T_m-values (Figure 5-26) showed, that the mutant protein was consistently more thermostable than the wildtype protein, suggesting reduced protein flexibility as a result of the Ala485-to-Val mutation.

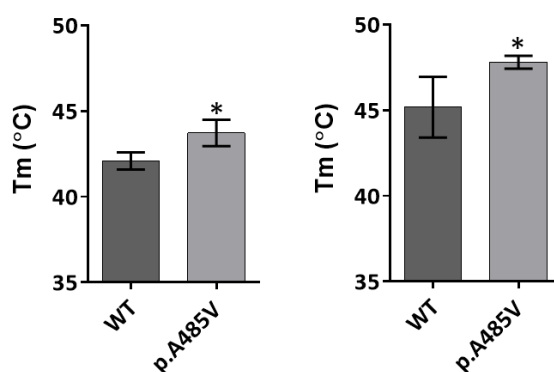


Figure 5-26. DSF assay showing the protein T_m for MRPP3(WT) and MRPP3(A485V) in a buffer of 150 mM NaCl, 10 mM HEPES pH 7.5 (left), and a buffer of 50 mM MES pH 6.5, 5% glycerol, 100mM NaCl, 100mM (NH₄)₂SO₄ (right). For this experiment $n = 3 - 4$. * = $P < 0.05$.

5.7 Crystallisation attempts

In an attempt to identify any conformational changes in the mutant protein, X-ray crystallography was employed. Mature wildtype MRPP3 had proven difficult to crystallise and therefore, to facilitate crystallisation, the Ala485-to-Val mutation was introduced into the previously crystallised $\Delta 206$ -MRPP3 construct. This construct was expressed and purified as described for mature wildtype MRPP3 (Figure 5-27).

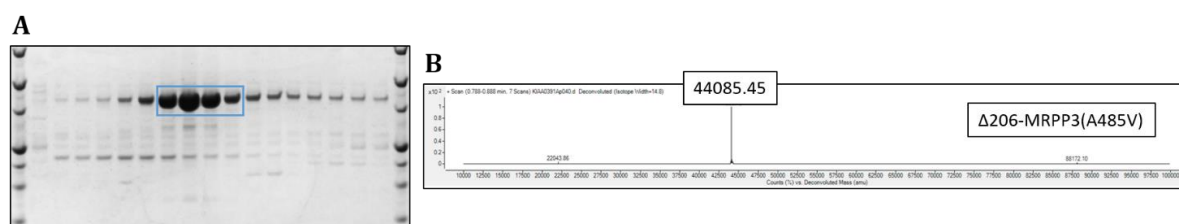


Figure 5-27. A) SDS-PAGE analysis from the last purification step of mutant $\Delta 206$ -MRPP3. B) The ESI-MS spectrum of the purified sample.

Crystallisation trials were carried out using random matrix screens and manually designed screens covering the chemical space reported to produce crystals for wildtype $\Delta 206$ -MRPP3 [179]. Unfortunately, no crystals were observed in any drops after two months of incubation.

5.8 Complex assembly studies with mutant MRPP3

To investigate if the mutant protein was able to form the ternary complex, the mature MRPP3(A485V) protein was mixed with co-expressed MRPP1-MRPP2 proteins in the presence of the Mg^{2+} cofactor and the pre-(mt)tRNA^{Ile} substrate and applied for SEC. Mature MRPP1, MRPP2 and mature MRPP3(A485V) proteins were applied to the column individually, and as a mixture of either MRPP1-MRPP2, MRPP1-MRPP2 + MRPP3(A485V) or MRPP1-MRPP2 pre-incubated with pre-(mt)tRNA^{Ile} and subsequently added with MRPP3(A485V). As for wildtype mature MRPP3, the chromatograms and SDS-PAGE analyses of the eluted fractions (**Figure 5-28**) showed that the mutant protein could form a ternary complex in the *presence* of pre-(mt)tRNA^{Ile}. Interestingly, in contrast to wildtype MRPP3, the mutant protein was also found in elution fractions corresponding to the complex, even in the *absence* of pre-(mt)tRNA^{Ile}. This could reflect that the Ala485-to-Val mutation induces conformational changes to the protein that allows MRPP3 to interact strongly with the binary complex of MRPP1-MRPP2 via regions that are either not interaction points, or weak interaction points, for wildtype MRPP3. More studies are needed to understand the interactions with MRPP3 in the RNase P complex.

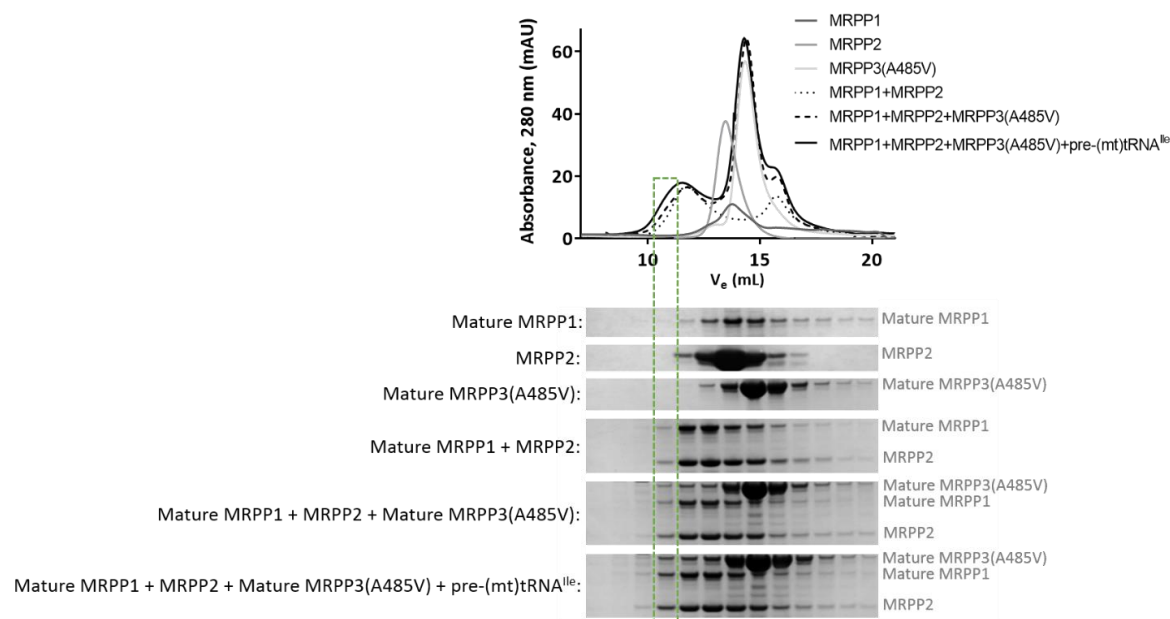


Figure 5-28. Analytical SEC with the mature MRPP3(A485V) protein applied in various combinations with mature MRPP1 and full-length MRPP2 with and without pre-(mt)tRNA^{Ile}. The green box indicates the elution fractions in which proteins co-localise upon complex formation.

5.9 Activity studies with mutant MRPP3

The RNase P assay was performed for MRPP3(A458V) by incubation of co-expressed MRPP1-MRPP2 complex with either mature wildtype or mutant MRPP3 in the presence of Mg²⁺. After incubation, the *in vitro* transcribed pre-(mt)tRNA^{Ile} substrate was added. The reaction products were monitored at various time points by denaturing UREA-PAGE gel stained for RNA (**Figure 5-29C**). The fractions of processed pre-(mt)tRNA^{Ile} at different time-points were fitted by non-linear regression as a one phase exponential association (**Figure 5-29A**), resulting in the value of maximum amount of processed pre-(mt)tRNA, $Y_{\max}$, (**Figure 5-29B**). Compared to the complex with wildtype MRPP3, the complex containing MRPP3(A485V) exhibited a 19.75% overall decrease in the amount of processed pre-(mt)tRNA^{Ile}, as reflected in the reduced $Y_{\max}$. This is in line with the suggestion that the Ala485-to-Val mutation affects activity, due to the location in close proximity to the catalytic Asp residues.

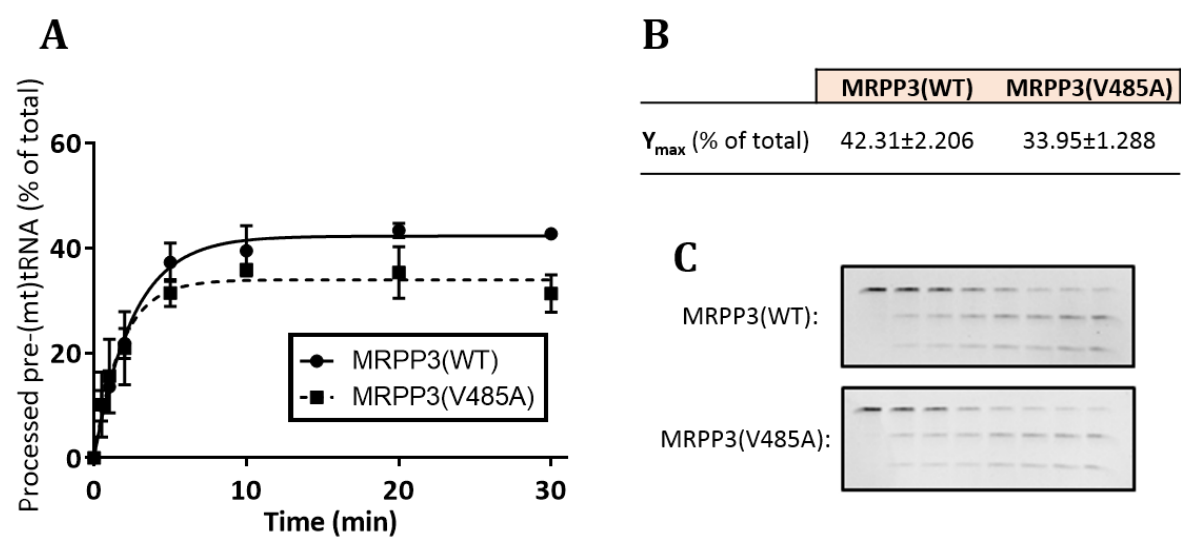


Figure 5-29. RNase P activity of MRPP3(A485V) compared to wildtype MRPP3. The kinetic parameter $Y_{\max}$ is shown along with the denaturing UREA-PAGE gel. For this experiment $n = 2$.

6 CHAPTER

CONCLUSIONS AND FUTURE AIMS

MRPP2 is a versatile protein with numerous catalytic as well as non-catalytic functions. When acting as an enzyme, MRPP2 functions as an SDR, using a variety of substrates with little similarity in chemical structures. When not acting as an enzyme, MRPP2 is involved in N1-methylation and 5'-end processing of mitochondrial tRNA. These activities require MRPP2 to form either a binary complex with MRPP1 (methylation) or a ternary complex with both MRPP1 and MRPP3 (processing).

Missense mutations in MRPP2 give rise to the rare, X-linked HSD10 disease often characterised by neurodegeneration, growth retardation and ketoaciduria. To date, 15 missense mutations have been identified in patients who present with different degrees of disease severity, varying from the most severe **neonatal form**, through the **infantile form** and **juvenile form**, to the almost asymptomatic **atypical form**. The severity of these phenotypes does not correlate with the residual enzymatic activity in mutant MRPP2 *in vivo*, suggesting that effects on its non-enzymatic functions might be the principal driver of the clinical phenotype.

The aim of the work presented in this thesis is to try to understand how these patient mutations affect MRPP2 functions at the molecular level, which led to the overall thesis question stated in Chapter 1: “**Does a correlation exist between the structural location of the MRPP2 patient missense mutations and their effects on the MRPP2 functions?**”.

The enzymatic reactions of MRPP2 were well-studied before the start of this project, but little was known about its non-enzymatic functions as part of the MRPP1-MRPP2 MTase complex and MRPP1-MRPP2-MRPP3 RNase P complex. To obtain a better

understanding of the non-catalytic functions of MRPP2, a structural and biochemical characterisation was undertaken of the individual proteins and two complexes using either wildtype MRPP2, or MRPP2 carrying patient mutations.

6.1. MRPP1

MRPP1 is a methyltransferase belonging to the Trm10 subfamily of the SPOUT family of MTases, and comprises an N-domain, an MT-domain and a small C-terminal extension. The N-domain and MT-domain might be connected by a flexible interdomain linker that makes the two domains structurally mobile towards one another.

6.1.1. Oligomeric state

Using analytical SEC and SAXS, mature MRPP1 was shown to exist as a dimer in solution, with analytical SEC suggesting that dimerisation occurs through the N-domain. This is in contrast to ‘regular’ SPOUT MTases from other subfamilies, e.g. TrmD, TrmH and TrmL (**Figure 6-1**), that form homodimers via the MT-domain [55], [59], [157], [187], [188]; a domain that was monomeric in MRPP1 when isolated and investigated using SEC.

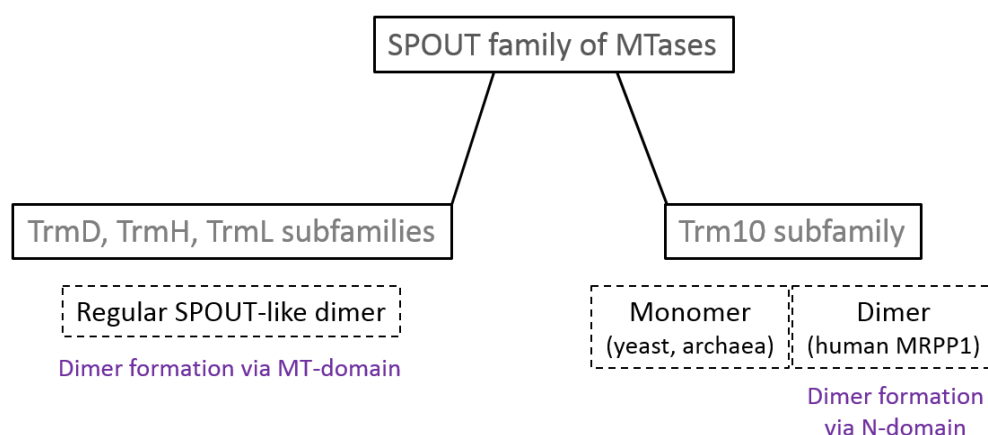


Figure 6-1. Overview of oligomeric states of different SPOUT subfamily members and the domains involved in dimer formation.

The homodimerisation in ‘regular’ SPOUT MTases takes place via a helix (helix $\alpha 5$) from each MT-domain, and the dimerisation is referred to as symmetrical or

asymmetrical based on the relative orientation of the helices ($\alpha 5$) towards each other (**Figure 6-2**) [59].

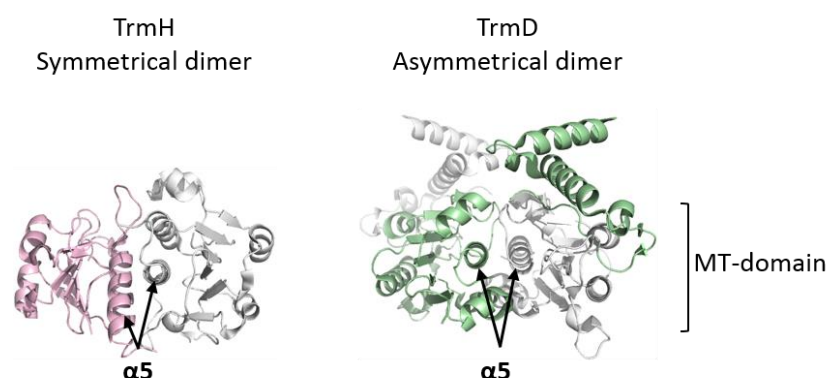


Figure 6-2. Dimeric SPOUT proteins TrmL (PDB: 4JAL) and TrmH (PDB: 4KGN) from *E. coli* and TrmD from *H. Influenza*. The helix involved in dimer formation is indicated ($\alpha 5$).

To investigate why MRPP1 presumably does not dimerise via helix $\alpha 5$ in the MT-domain, efforts were made to crystallise the MT-domain of MRPP1, which provided a structure of the MT-domain to 1.96 Å. Structure analysis revealed that the MT-domain of MRPP1 contains an additional helix ($\alpha 6$) after helix $\alpha 5$, that, when aligned with the MT-domain of e.g. dimeric TrmH from *B. pseudomallei* and TrmL from *E. coli*, sterically clashes with the opposite monomer, blocking dimerisation via helix $\alpha 5$ (**Figure 6-3**). This provides an explanation why MRPP1 dimerises in a manner different from other SPOUT MTases, *i.e.* via the N-domain rather than the MT-domain.

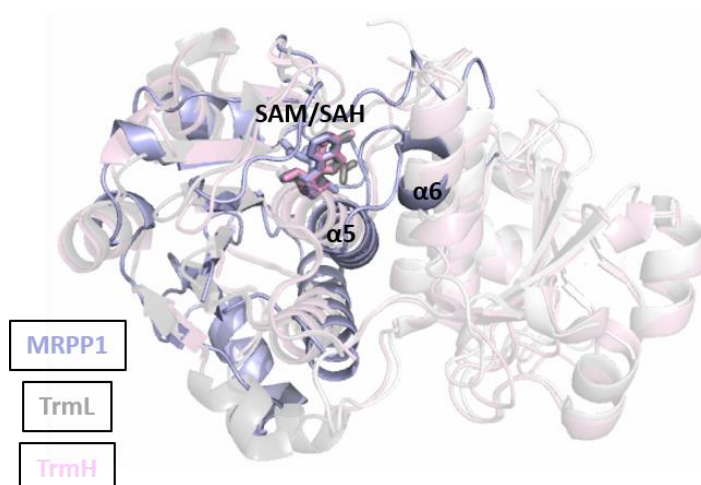


Figure 6-3. Structural alignment of the monomeric MT-domain of MRPP1 with the dimeric SPOUT proteins TrmL (PDB: 4JAL) and TrmH (PDB: 4KGN) from *E. coli*. The SAM ligand from MRPP1 and SAH ligand from TrmL and TrmH are shown as sticks. The two helices $\alpha 5$ and $\alpha 6$ are indicated on MRPP1.

The dimeric nature of MRPP1 also contrasts with the otherwise monomeric nature of the Trm10 subfamily to which MRPP1 belongs (**Figure 6-1**). To date, the characterised Trm10 family members from fission yeast (*S. pombe*), baker's yeast (*S. cerevisiae*) and the archaea *S. acidocaldarius* are monomeric in solution [49], [159]. These structures exhibit the same structural arrangement as MRPP1, prohibiting the 'regular' SPOUT-like dimerisation. A sequence alignment of known Trm10 family members (**Section 8.3**) revealed that these proteins all contain N- and C-extensions of different lengths, in addition to their MT-domain. However, low sequence identity is found across these extensions, which might explain why MRPP1 can dimerise via the N-domain unlike other Trm10 members. It would be interesting to investigate the oligomeric state of further Trm10 proteins from higher eukaryotes (metazoans), e.g. the human cytosolic TRMT10A and TRMT10B, as it is possible that such MTases could form an evolutionarily distinct subclass of dimeric Trm10 proteins.

It is also worth noting that MRPP1 is the only Trm10 member, to date, that has been found to require a binding partner (MRPP2) for MTase activity, supporting a slightly different organisation for this protein.

6.1.2. *Ligand binding*

The MT-domain of MRPP1 is predicted to contain the methyltransferase catalytic site, and consistent with this, the methylation cofactor SAM was shown by DSF to bind the isolated MT-domain as well as the MT-domain with the C-extension (MT+C). The thermal stabilisation upon SAM binding was significantly greater for the MT+C-domain than for the MT-domain alone, suggesting a potential role of the C-extension in ligand binding by either direct interaction with the ligand or by formation of protein structure stabilising interactions with the SAM bound MT-domain. The obtained crystal structure of the MT+C-domain confirmed SAM binding to the MT-domain, but could not identify any direct role of the C-extension in this binding, as no electron density was observed for this region.

6.1.1.1 Active site

‘Regular’ dimeric SPOUT proteins accommodate SAM in a similar fashion to MRPP1 (**Figure 6-3**). However, the ‘regular’ SPOUT proteins involve residues from *both* MT-domains in *catalysis* and in *binding of the purine to be modified* [139], and an intact MT-domain dimer surface has been shown by mutation studies to be required for tRNA binding [189]. As the MT-domains of MRPP1, and other Trm10 members, are monomeric in solution when isolated, the above active site organisation and tRNA binding mode is likely not possible. The differences in active site organisation is also reflected in the SAM cofactor accessibility. **Figure 6-4** shows how the cofactor in dimeric TrmL and TrmH proteins is bound in a pocket that is almost entirely closed in MRPP1, due to the presence of a loop [189]. An alternative entrance to the same cofactor binding pocket exists in MRPP1, and other Trm10 members, which is buried in the monomer-monomer interface of the dimeric TrmH/TrmL proteins.

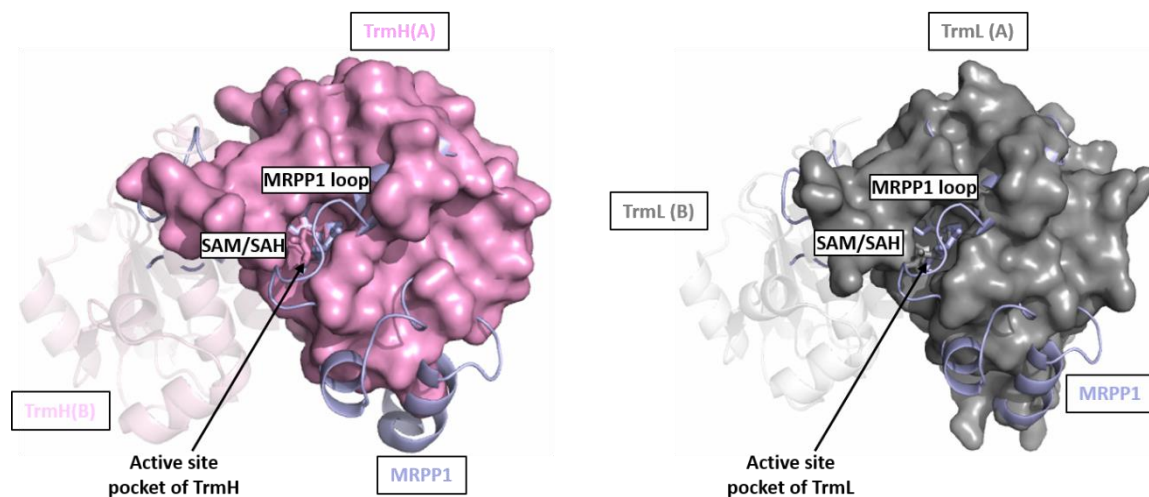


Figure 6-4. Structural alignment of the MT-domain of MRPP1 with A) dimeric SPOUT protein TrmH from *B. Pseudomallei* and B) dimeric SPOUT protein TrmL from *E. coli*. The surface is shown for one monomer of TrmH and TrmL. The SAM ligand from MRPP1 and SAH ligand from TrmH and TrmL are shown as sticks. The loop from MRPP1 that blocks the active site pocket of TrmH and TrmL is indicated.

6.1.3. Mapping of catalytic residues

MRPP1 has dual-specificity in its methyltransferase activity, *i.e.* methylates both guanine and adenine. Both reactions could require a catalytic base and, in the case of

guanine, a further positively charged intermediate charge-stabiliser. *In silico* docking of 5'-GMP, combined with sequence alignment of Trm10 proteins, revealed that the conserved Asp314 amino acid could act as the catalytic base for m¹A9 and m¹G9 MTase activity, and that the conserved Gln226 could be a candidate for the intermediate charge-stabiliser.

Catalytic base

An Asp314-to-Asn mutant MRPP1 protein (together with MRPP2) showed no *in vitro* activity in either reaction, supporting the notion that this is likely the catalytic base used in both reactions. These findings agree with previous mutation studies of human and yeast Trm10 proteins on the corresponding residue [14], [49].

Intermediate charge-stabiliser

An Gln226-to-Ala mutant MRPP1 protein (together with MRPP2) also did not show *in vitro* activity for m¹G9 formation, confirming previous studies of the corresponding residue in yeast Trm10 [49]. Surprisingly, the p.Q226A mutation also abolished m¹A9 activity as later confirmed by others [49], suggesting that Gln226 is not merely the stabiliser for the charged guanine intermediate, but rather plays an important role for both substrates.

The Gln226 residue in MRPP1 could be involved in binding of the phosphate group in purine 9 of the (mt)tRNA, as suggested for a corresponding Asn residue in TrmL and TrmH from bacteria [189], [190]. If so, the point of interaction is likely different between the Gln and Asn residues, as an Asn-to-Gln substitution in TrmH decreased MTase activity, due to sterically hindered binding of the phosphate group by the larger Gln residue [190]. While it cannot be excluded that the p.Q226A mutation affects overall pre-(mt)tRNA and/or SAM binding in MRPP1, a similar Gln-to-Ala mutation in Trm10 proteins from yeast confirmed that these functions were unaffected in these homologue proteins [49].

The equivalent Gln residue is also present in the crystallised m¹A9-specific Trm10 from the archaea *S. acidocaldarius* and in the dual-specific m¹A9/m¹G9 Trm10 protein from

the archaea *T. kodakaraensis*, where no crystal structure is available. In Trm10 from *S. acidocaldarius*, this Gln residue also protrudes towards the active site, and it would be intriguing to see if mutation of this Gln residue in both archaeal Trm10 proteins had the same activity-abolishing effect observed in MRPP1.

Purine specificity

Two Trm10 members have confirmed m¹A9 activity (*S. acidocaldarius* and *T. kodakaraensis*). An additional Asp residue (Asp184 [159] and Asp206 [171], respectively), is found in both proteins, but not in the Trm10 yeast proteins harbouring only m¹G9 activity (**Section 8.3, green star**), and could thus be a determinant for m¹A9 activity. In MRPP1, this additional Asp residue is replaced by Leu that, due to its uncharged nature, is likely to act differently in catalysis than a charged Asp residue. Therefore, this amino acid position cannot be the sole determinant for m¹A9 activity. Another Asp residue is located deep in the active site pocket of MRPP1 (Asp293) (**Section 8.3, purple star**), in a position corresponding to a Lys(81) residue in TrmL from *E. coli*, which was previously shown to be involved in catalysis but not tRNA binding [189]. Although the Asp293 residue could be involved in the MTase function in MRPP1, it seems unlikely to be the determinant for m¹A9 activity as this position is Asp in the guanine specific Trm10 proteins from yeast, and is not conserved in m¹A9 active Trm10 proteins from *S. acidocaldarius* and *T. kodakaraensis*. Aside from the residues discussed above, alignment of the available Trm10 protein structures and their primary sequences show no obvious amino acid candidates in or around the active site that could account for the differences between m¹G9 specific (*S. cerevisiae* and *S. pombe*), m¹A9 specific (*S. acidocaldarius*) and m¹A9/m¹G9 dual-specific (MRPP1 and *T. kodakaraensis*) Trm10 MTases. It is possible that the purine specificity could simply be due to differences in surface charge around the active site and size of the purine-binding pocket (**Figure 6-5**), which allows different Trm10 family

members to accommodate different purine substrates, rather than to specific residues for catalysis. The active site pocket is more open for the m¹G9-specific TRMT10A and m¹A9-specific Trm10, compared to other Trm10 proteins. However, no obvious similarities are observed within the m¹G9 specific group of proteins that are also clearly different from the m¹A9 specific Trm10 and further altered compared to the m¹G9/m¹A9 dual-specific protein. A detailed understanding on how the substrate specificities are generated would benefit from future high resolution, substrate-bound crystal structures of Trm10 family members.

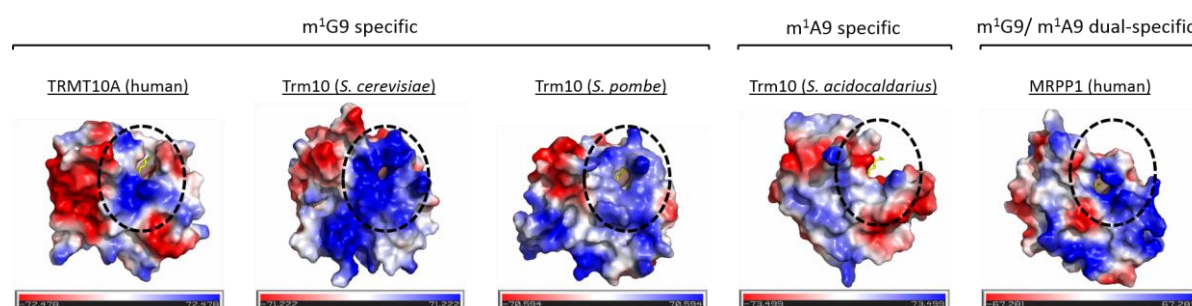


Figure 6-5. The electrostatic potential of the MT-domain for each Trm10 member for which a crystal structure is available. The SAM cofactor is shown in yellow sticks. The N1-methylation specificity is shown for each Trm10 member.

6.2. MRPP1-MRPP2

Studies were next performed on the complex of MRPP1 and MRPP2 and revealed a 2xMRPP1:4xMRPP2 stoichiometry from both SV-AUC and SAXS. Co-expression of MRPP1 and MRPP2 further revealed that the presence of MRPP2 completely eliminated the previously observed degradation of MRPP1 during purification, suggesting a structure stabilising effect of MRPP2 on MRPP1.

6.2.1. MTase activity of mature MRPP1

An MTase activity assay showed that both MRPP2 and the entire, mature protein of MRPP1 were required for m¹G9 MTase activity. The fact that MRPP1 requires an additional protein for catalysis contrasts that of all other functionally characterised Trm10 subfamily members. An example of the latter is the human cytosolic member of the Trm10 family,

TRMT10A which *alone* can catalyse m¹G9 formation *in vitro*, on some (mt)tRNAs predicted to have similar sequences but potentially different secondary structures to conventional tRNAs [14], [50]. TRMT10A and MRPP1 have similar domain organisation (**Figure 6-6**), and their MT-domains have high structural homology (C^α RMSD = 1.5 Å).

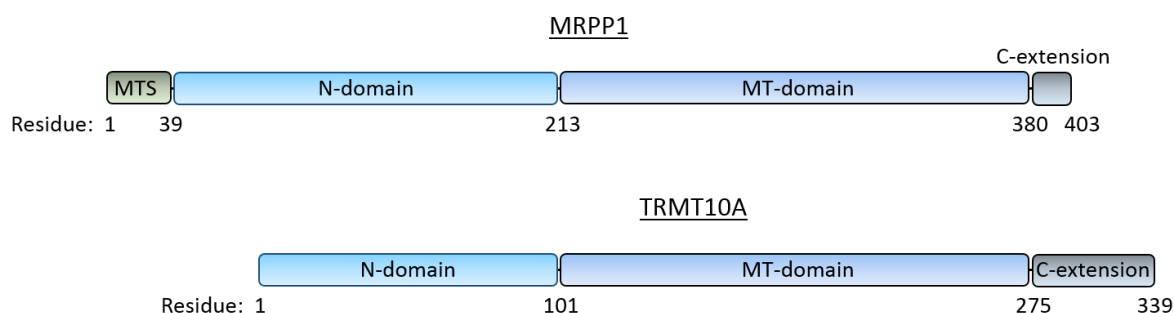


Figure 6-6. Domain organisation in full-length MRPP1 and TRMT10A.

The fact that TRMT10A – but not MRPP1 – can perform the MTase reaction on (mt)tRNAs without the help of a partner protein such as MRPP2 is interesting, and must be due to differentiating features in the N-domains and C-extensions of the two proteins, as their MT-domains are very similar. The N-domain of MRPP1 was found to be involved in interactions with both tRNA and MRPP2, and a lack of this domain abolished MTase activity. It is interesting to speculate that the smaller N-domain in TRMT10A associates tightly with the MT-domain and is able to ‘grab’ the tRNA without additional proteins for assistance, whereas the larger N-domain in MRPP1 relies on MRPP2 to direct it towards the tRNA for binding. Alternatively, the larger C-extension of TRMT10A could play a role in tRNA binding as previously suggested for the Trm10 homologue from archaea [159] – an interaction point which could be lost for MRPP1 that contains a much smaller C-extension (**Figure 6-6**).

6.2.2. Interactions between MRPP1 and MRPP2

Mapping of regions in MRPP1 for MRPP2 interaction

SEC data showed that the entire protein of MRPP1 was required for interaction with MRPP2. Following this finding, the potential sequence motifs in MRPP1, required for

interaction with MRPP2, were identified using a SPOT blot assay. Entirely consistent with the finding that complex formation with MRPP2 requires all domains of MRPP1, the SPOT blot analysis identified potential MRPP2-binding sites in the N-domain, MT-domain and C-extension of MRPP1. The identified regions within the MT-domain were mapped onto the crystal structure, and revealed two regions clustered together to form a surface patch near the beginning of the structure. These two regions, together with a third region seven residues N-terminal of the structure start, could form a collaborative patch for MRPP2 binding.

Interestingly, the two identified patient mutations in MRPP1 (R118L and T272A), leading to COXPD30 disease, both fall directly in this predicted interaction patch, potentially indicating that these mutations could affect complex formation. However, no *in vivo* effect of these mutations was shown on m¹R9 MTase activity by MRPP1-MRPP2 [136].

Two other likely candidate-binding regions were identified with the SPOT blot analysis in the N-domain and C-extension, respectively, but neither of these are encompassed in the present crystal structure. Based on susceptibility to degradation, secondary structure predictions, and lack of electron density in both the present MRPP1 crystal structure and structures of homologues proteins, these segments were hypothesised to be flexible and could thus either 1) spatially orient to form part of the identified MRPP2 interaction patch or 2) interact independently with MRPP2. Long-range interactions between C- and N-extensions have previously been found for other SPOUT family members, e.g. TrmH [191], [192], suggesting that this could also be the case for MRPP1.

It has been suggested that a 65 amino acid large C-domain in an archaea homologue of MRPP1 is responsible for tRNA moulding for catalysis [159]. Considering the comparably small size of the C-extension in MRPP1, it is less likely that this part plays a similar role. This could be investigated by performing the MTase activity assay using the

soluble proteins of MRPP1 with and without the C-extension (residues 40-403 and 40-380, respectively), made during the work described in this thesis.

SAXS model of MRPP1-MRPP2

Attempts to obtain a crystal structure of the full MRPP1-MRPP2 complex were unsuccessful. In the absence of such a structure, SAXS was employed to obtain further structural details of the complex. The best SAXS model was calculated with *P1* point symmetry, and fitted an MRPP1(dimer):MRPP2(tetramer) stoichiometry confirmed by calculations from the SAXS scattering curve and obtained from SV-AUC. The generated SAXS model indicated that the MRPP1 dimer binds on *one* side of the MRPP2 tetramer.

The SPOT blot identified protein-interaction regions in both the N- and MT-domains of MRPP1 suggest a cross-over arrangement of the N-domains in MRPP1, allowing interactions of MRPP2 with both the N- and MT-domain, but from different MRPP1 monomers. The crystal structure of the MT-domain of MRPP1 is manually fitted into the SAXS model of MRPP1-MRPP2 with the identified interaction-region facing MRPP2 (**Figure 6-7**). The N-domains are oriented such that one is in close proximity to MRPP2 for interaction via the identified region.

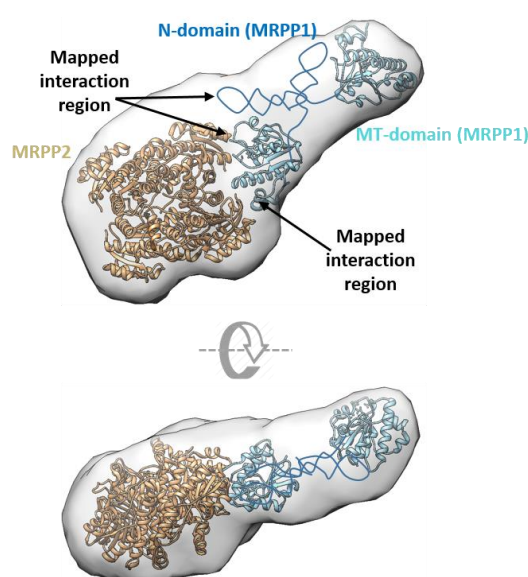


Figure 6-7. SAXS models for the binary complex of MRPP1-MRPP2 with a suggested placement of the protein components based on interactions studies.

Since the MRPP2 tetramer has a symmetrical shape, it is rather surprising that it only interacts with an MRPP1 dimer at *one end* of the MRPP2 tetramer, rather than with one MRPP1 dimer or monomer at *both ends* of the MRPP2 tetramer. The fact that an MRPP1 dimer is not observed on both sides of MRPP2 could mean that 1) binding of MRPP1 induces conformational changes in MRPP2, which would prevent further MRPP1 binding, or 2) binding of MRPP1 sterically hinders the binding of a second MRPP1. The fact that MRPP2 still exhibit SDR activity when bound to MRPP1 suggests that no major conformational changes are induced in this protein. However, before a higher resolution structure is available for this complex, it is difficult to conclusively favour one hypothesis over another.

SAXS model of MRPP1-MRPP2-pre-(mt)tRNA^{lle}

A SAXS model of the MRPP1-MRPP2 complex bound to the pre-(mt)tRNA substrate showed *one* pre-(mt)tRNA molecule bound per MRPP1(dimer):MRPP2(tetramer), likely with one SAM molecule bound in each MRPP1 MT-domain. The binding of only one tRNA per dimer of SAM-bound MRPP1 is consistent with the stoichiometry 1) predicted for the dimeric TrmL MTase based on co-crystallisation of purine mimicking ligands [189], 2) identified with analytical SEC studies for dimeric TrmH [190], and 3) confirmed with structural and functional studies for dimeric TrmD [139], [193]. The binding of only one tRNA in SAM-bound TrmD was explained by conformational changes in the protein homodimer-tRNA complex that rendered the protein asymmetrical after tRNA binding [139]. The structural differences in dimer formation between MRPP1 and the ‘regular’ SPOUT-like dimer of TrmD means that the tRNA-bound TrmD crystal structure does not provide sufficient clues for the specific tRNA binding mode of MRPP1.

The N-domain in tRNA binding

Both TrmD and TrmH proteins involve domains/extensions, additional to the MT-domain, in tRNA binding [139], [194]. An EMSA showed that MRPP1, similarly to both TrmD and TrmH, uses the N-domain for binding of pre-(mt)tRNA, and the suggested cross-over arrangement for the N-domain allows for such tRNA interaction (**Figure 6-8**), with *one* tRNA molecule binding on *one* side of the MRPP1-MRPP2 complex. The suggested arrangement further allows a MT-domain from *one* MRPP1 monomer to collaborate with the N-domain from *another* monomer for substrate binding, offering an explanation why MRPP1 functions as a dimer. In this model, the active site located in the MT-domain is in close proximity to the purine-9 to be methylated (**Figure 6-8**). The tRNA is shown as gray ovals, as it could not accurately be fitted into the SAXS pocket due to long (~30 nt), likely flexible, 3'-trailer and 5'-leader.

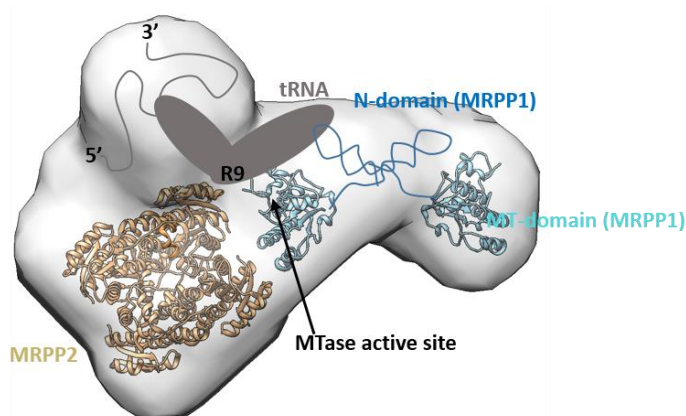


Figure 6-8. Potential binding mode of MRPP1 and pre-(mt)tRNA that allows the nucleotide to be methylated (R9) to be in close proximity to the MTase active site and the tRNA-binding N-domain to be near the substrate.

The assistance in tRNA binding by domains, additional to the enzymatic domain, connected by flexible interdomain linkers, is shown for many SPOUT family members and other tRNA modifying enzymes [139], [159], [162]–[164]. For Trm10 proteins from yeast and archaea, the N-domains have been shown to directly interact with the tRNA substrate [49], [159]. It is tempting to speculate that the role of these domains is to recognise a more universally conserved parts of the tRNAs, if the immediate environment around the modified

nucleotide shows low structural conservation. The immediate environment around the Trm10 methylation site purine-9 consists of the variable D- and T-arm/loops and the variable loop with low structural conservation, and it is possible that the MRPP1 N-domain assists in tRNA binding by recognising other conserved parts of the tRNA. Consistent with this possibility, a docking model for the Trm10 protein from archaea suggested that the N-domain-tRNA interaction of this protein is through the conserved anti-codon stem/loop [159].

6.2.3. *MRPP1-MRPP2 summary*

Based on work presented here, and studies on MRPP1 homologues, the below roles for each MRPP1 domain is proposed. The role of MRPP2 in the complex is discussed later (Section 6.1.1.3).

- **The MT-domain is the catalytic domain of the protein** binding the cofactor.
- **The N-domain assists in tRNA binding**, potentially by recognising more universally conserved parts of the tRNA.
- The role of **the C-extension** is not entirely clear, but data suggest **a role in MT-domain stabilisation upon SAM binding**. The C-extension was also found to interact with MRPP2. Further investigations into a potential role in catalysis could be performed.

6.3. *MRPP1-MRPP2-MRPP3*

The MRPP3 endoribonuclease contains an N-extension, a PPR-domain likely involved in pre-(mt)tRNA binding, an NYN-domain containing the active site and a core domain orienting the PPR- and NYN-domain relative to each other for RNase P activity.

6.3.1. *Biochemical analysis of MRPP3*

Analytical SEC identified MRPP3 to be monomeric, which is in line with findings for other MRPP3 homologous PRORP proteins from *A. thaliana* (PRORP1 and PRORP2)

[184]. DSF data suggested that the mature MRPP3 protein did not bind Mg^{2+} necessary for catalysis, which agrees with the MRPP3 structures by Li *et al.* [180] and Reinhard *et al.* [179] in 2015, where no Mg^{2+} was found in the active site.

6.3.2. Complex assembly and pre-(mt)tRNA binding

Analytical SEC experiments demonstrated that mature MRPP3, but not the core/NYN-domains alone, was able to form the ternary complex with MRPP1 and MRPP2. The ternary complex of MRPP1-MRPP2-MRPP3, however, could only be formed in the *presence* of the pre-(mt)tRNA substrate. The transient, substrate-bound, ternary complex was captured using a catalytically inactive MRPP3 mutant (p.D479A) for SAXS analysis, to investigate the spatial organisation of components in the final substrate-bound ternary complex. The SAXS model supported a stoichiometry of MRPP2(dimer):MRPP2(tetramer):MRPP3(monomer):pre-(mt)tRNA(one molecule) with binding of MRPP3 on the opposite side of MRPP2, relative to MRPP1. MRPP3 is a monomer in solution and in the crystal, and it is therefore not surprising that it would function as such in the complex. The obtained substrate-bound SAXS models of the binary and ternary complexes both appear to accommodate only one pre-(mt)tRNA molecule, and it therefore seems reasonable that the ternary complex would accommodate just one molecule of MRPP3 for processing of this substrate. The leader and trailer sequences of the pre-(mt)tRNA substrate are ~30 nucleotides long and likely flexible, and therefore the tRNA could not accurately be fitted into the SAXS model for more detailed investigations of the nucleotide-recognition of MRPP3. Unpublished data suggest that the MRPP3 plant homologue PRORP1 recognises only the first two nucleotides in the 5'-leader sequence and not the 3'-trailer [195]. Therefore, future SAXS experiments could be performed using a tRNA substrate with a similarly short 5'-leader and no 3'-trailer, allowing easier fitting to the SAXS model.

6.3.3. RNase P activity of the ternary complex

Activity of mature MRPP3

An RNase P activity assay showed that the reconstituted complex of MRPP1-MRPP2 and MRPP3, but not MRPP3 alone, had RNase P activity towards the tested pre-(mt)tRNA^{Ile}. These findings were confirmed by unpublished data for pre-(mt)tRNA^{Ile} (and pre-(mt)tRNA^{Lys}), but the study additionally found that the protein alone can inefficiently cleave *some* pre-(mt)tRNAs (Tyr, His, Glu, Ser(UCN) and Met) [91], [92], [196].

Activity of MRPP3 fragments

No RNase P activity was observed for the ternary complex in the absence of either the PPR-domain or the core/NYN-domains of MRPP3, but partial removal of the N-extension (43 residues) still left a functional protein. The retained activity upon removal of the N-extension was confirmed for the MRPP3 homologue, PRORP1 from plant, that has an equivalent 89 residue N-extension [183]. N-extensions are found in many single stranded (ss) RNA binding PPR-domain proteins [101], and in double-stranded (ds) DNA-binding repeat proteins such as tetratricopeptide repeat (TPR) [197] proteins or transcription activator-like (TAL) effectors [198], [199]. In TAL effectors, the N-extension adopts an α -helical repeat structure and contributes to dsDNA binding specificity [198], [199]. The function of the N-extension in MRPP3 remains to be determined, including the effect of further truncations, using e.g. the soluble Δ 118-MRPP3 made in the course of this thesis.

6.3.4. MRPP1-MRPP2-MRPP3 outline

It is interesting that the RNase P activity of MRPP3 depends so strongly on two additional proteins (MRPP1 and MRPP2), when the MRPP3 homologues from other eukaryotes efficiently perform the 5'-end processing of pre-tRNA on their own (e.g. *A. thaliana* PRORP1 and PRORP2). As described in the introduction (**Chapter 1**), (mt)tRNAs often adopt 'bizarre' structures – especially before they undergo posttranslational

modifications. It is possible that the MRPP1-MRPP2 complex activates the pre-(mt)tRNA for recognition by MRPP3, by ‘re-moulding’ the (mt)tRNA in a way that offers a similar recognition surface for MRPP3 across all ‘bizarre’ (mt)tRNAs. In support of the hypothesis that MRPP1-MRPP2 plays an activating role for RNase P activity, unpublished studies showed that addition of MRPP1-MRPP2 increased the catalytic efficiency of MRPP3 (>30,000 fold) by increasing its binding affinity for pre-(mt)tRNA by >300-fold [196]. Structural studies of MRPP3 have further suggested that the pre-(mt)tRNA binding of MRPP3 activates the protein by inducing conformational changes that allow the protein to accommodate the Mg^{2+} cofactor. The role of MRPP1-MRPP2 as the pre-(mt)tRNA recognition component of the RNase P complex is in line with findings herein showing that the MRPP1-MRPP2 complex can bind the pre-(mt)tRNA prior to binding of MRPP3. Based on studies presented here, and published work on homologues, the assembly of the ternary complex is proposed to proceed as follows:

Step 1) MRPP1 and MRPP2 form a complex.

Step 2) The binary complex binds the pre-(mt)tRNA substrate and potentially re-mould this for optimised recognition by MRPP3.

Step 3) The binary complex-bound – potentially re-moulded – pre-(mt)tRNA is recognised by MRPP3 to form the ternary complex.

Step 4) The complex-bound MRPP3 protein adopts a conformation that allows binding of the Mg^{2+} cofactor necessary for pre-(mt)tRNA 5'-end processing.

Step 5) Processing takes place before MRPP3 dissociates and processed (mt)tRNA is released.

The above steps highlight the possibility that some pre-(mt)tRNAs could be processed without being methylated, as confirmed by activity studies showing that the two

processes act independently of each other [14]. To avoid improperly methylated mature tRNA from being formed *in vivo* it is possible that another step, further in the processing, e.g. 3'-end processing or CCA-addition, requires N1-methylation at R9.

6.4. MRPP2 in the binary and ternary complex

6.1.1.2 tRNA binding of MRPP2

The exact role of MRPP2 in mature (mt)tRNA/pre-(mt)tRNA methylation and processing is currently incompletely understood and the subject of much discussion. A direct interaction between MRPP2 and the (mt)tRNA substrate has been suggested, based on the discovery that protein binding pockets for RNA-like di-nucleotides, such as NAD⁺/NADH, could accommodate RNA itself [200]. Work presented here did not detect any direct interaction *in vitro* between recombinant MRPP2 alone and pre-(mt)tRNA using an EMSA. Similar findings were reported by Vilardo *et al.* [14], using the same assay set-up with both pre-(mt)tRNA as well as mature (mt)tRNAs. Conversely, *in vivo* data obtained via the iCLIP method [75] indicated a direct interaction between MRPP2 and various (mt)tRNAs. In further characterising the role of MRPP2, it should be considered that – even if MRPP2 does not bind the (mt)tRNA alone – it does not exclude an interaction with the substrate in the presence of MRPP1.

6.1.1.3 Role of MRPP2 in the non-catalytic functions

An increase in K_D of MRPP1 towards the mature (mt)tRNA with addition of MRPP2, was previously observed using an EMSA [14], which could be due to either 1) direct interaction of MRPP2 with the substrate itself (as discussed above) or 2) binding-induced changes of the structure of MRPP1 to increase its substrate affinity. The data presented in this thesis suggest that the role of MRPP2 in the complex is to stabilise, amongst other parts, the nuclease sensitive N-domain of MRPP1, shown to be involved in pre-(mt)tRNA binding. The stabilisation of the tRNA-interacting N-domain favours the

possibility that the increased K_D of the complex arise from an MRPP2 binding-induced change in MRPP1 that increase the proteins substrate affinity.

6.5. Effect of mutations in MRPP2

6.5.1. Biochemical analysis of all missense mutations

Of the 15 missense mutations of MRPP2, characterised in HSD10 disease patients, introduction of 11 of these (the remaining 4 only became known recently) separately into the recombinant MRPP2, markedly affected the behaviour of the mutant proteins (**Table 6-1**).

	Expression	Solubility	Oligo. state	T _m as-purified	T _m with NADH	Location in structure	Phenotype
p.V12L	As WT	Severely reduced	As WT	Reduced	Reduced	Core β-sheet	Unknown
p.V65A	As WT	Reduced	As WT	Reduced	Reduced	NAD ⁺ cofactor binding loop	Atypical
p.D86G	Reduced	Severely reduced	As WT	Reduced	Reduced	Core β-sheet	Neonatal
p.L122V	As WT	As WT	As WT	Increased	Reduced	Helices α5 and α6	Infantile
p.R130C	Reduced	Reduced	As WT	Lower	Reduced	BxA interface	Infantile
p.Q165H	As WT	Reduced	As WT	Increased	Reduced	Helices α5 and α6	Atypical
p.V176M	As WT	As WT	As WT	Increased	Reduced	Helices α5 and α6	Atypical
p.P210S	As WT	Reduced	As WT	Lower	As WT	Mobile segment	Infantile
p.R226Q	As WT	As WT	Aggregation	Lower	Reduced	BxC interface	Neonatal
p.N247S	As WT	Reduced	Aggregation	Lower	Reduced	BxC interface	Neonatal
p.E249Q	As WT	Reduced	Aggregation	Lower	Reduced	BxC interface	Juvenile

Table 6-1. Overview of data obtained from biochemical studies on 11 known missense mutations identified in HSD10 disease patients.

With the exception of p.V12L for which no phenotype is available, all mutations that were observed to cause a *significant reduction in protein abundance*, by either low expression/severely reduced solubility (p.D86G) or protein aggregation (p.R226Q, p.N247S and p.E249Q), are associated with the *juvenile form* or severe *neonatal form* of HSD10 disease. In contrast, mutations with near wildtype expression, solubility and oligomeric state (p.V65A, p.L122V, p.R130C, p.Q165H, p.V176M and p.P210S) are associated with the *infantile form* or almost asymptomatic *atypical form* of disease. At the structural level (**Figure 6-9**), the mutations that cause a similar biochemical effect were typically structurally clustered together into:

- Mutations causing significant reductions in protein solubility (p.V12L and p.D86G) being located in the core β -sheet of the Rossmann fold.
- Mutations causing increased T_m for the as-purified protein (p.L122V, p.Q165H, p.V176M) being located in either of the two long helices, $\alpha 5$ or $\alpha 6$, which contain the catalytic residues.
- Mutations causing increased T_m for the NADH-bound protein (p.P210S) being located in the mobile segment.
- Mutations causing protein aggregation (p.R226Q, p.N247S, p.E249Q) being located in the BxC dimer interface.

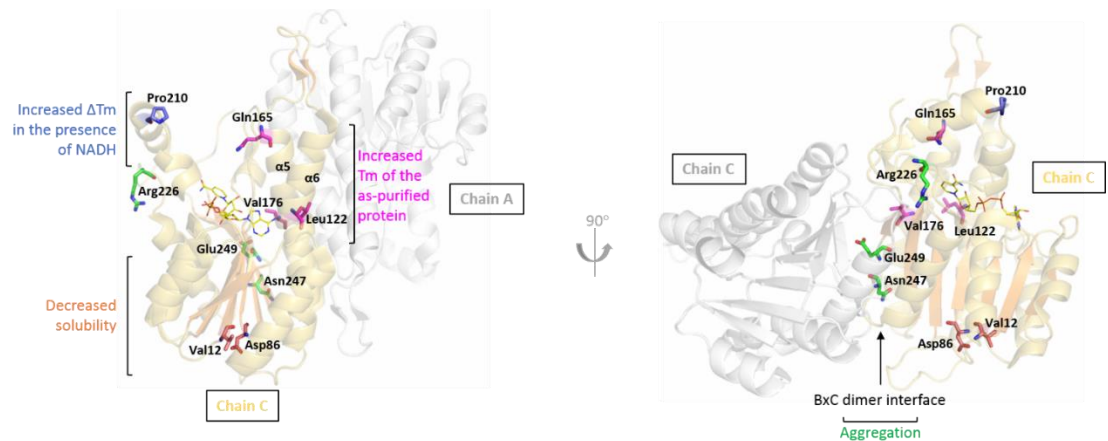


Figure 6-9. Location and biochemical categorisation of missense mutations identified in HSD10 disease patients.

The p.R130C and p.V65A mutations did not fall into any of the above groups. Val65 is conserved amongst MRPP2 proteins from bacteria to mammals, and the Val12-to-Ala mutation has been shown *in vivo* to result in 50% decreased enzymatic activity [120]. A LigPlot [201] analysis of the MRPP2 structure (PDB: 1U7T) in complex with NAD^+ revealed that the backbone amide in Val65 forms a hydrogen bond to the adenine ring on the cofactor. The Val65-to-Ala mutation could change the orientation of the backbone and thus alter the hydrogen-bonding pattern, and subsequently affect the cofactor binding. Altered

NADH binding properties were identified in the DSF assay for this mutant, but many other mutant proteins exhibited even lower thermal stability than p.V65A.

Arg130 is located in the BxA dimer interface and the p.R130C mutation has been suggested to result in protein instability, due to disruption of hydrogen bonds between Arg130 in monomer B and residues from monomer A [121]. Such protein instability was reported in an *E. coli* expressed recombinant MRPP2(R130C) protein that showed a steady loss of enzymatic activity over time at room temperature [115], and agrees with findings of very low levels of MRPP2(R130C) in fibroblasts from an HSD10 disease patient [116]. MRPP2(R130C) protein did not show sign of either altered oligomeric state or thermal stability (as-purified), compared to wildtype, that could confirm this protein instability.

After completion of the thesis experimental work, four new patient mutations were identified, namely p.A154T, p.A157V, p.A158V and p.K212E [127]–[130]. These mutations are located in the mobile segment (p.K212E) or helices $\beta 5$ and $\beta 6$ (p.A154T, p.A157V, p.A158V) and are thus, based on the above discussion, expected *not to significantly affect protein abundance*. Accordingly, patients with these mutations would be expected to display either of the phenotypes: infantile or atypical form. Indeed this is the case, with p.K212E giving rise to the infantile form and p.A154T, p.A157V, p.A158V giving rise to the atypical form.

6.5.2. Functional analysis of two novel missense mutations

Two of the mutations, p.V12L and p.V176M, were further characterised enzymatically and found to have an effect on the SDR, MTase and RNase P activities (Figure 6-10).

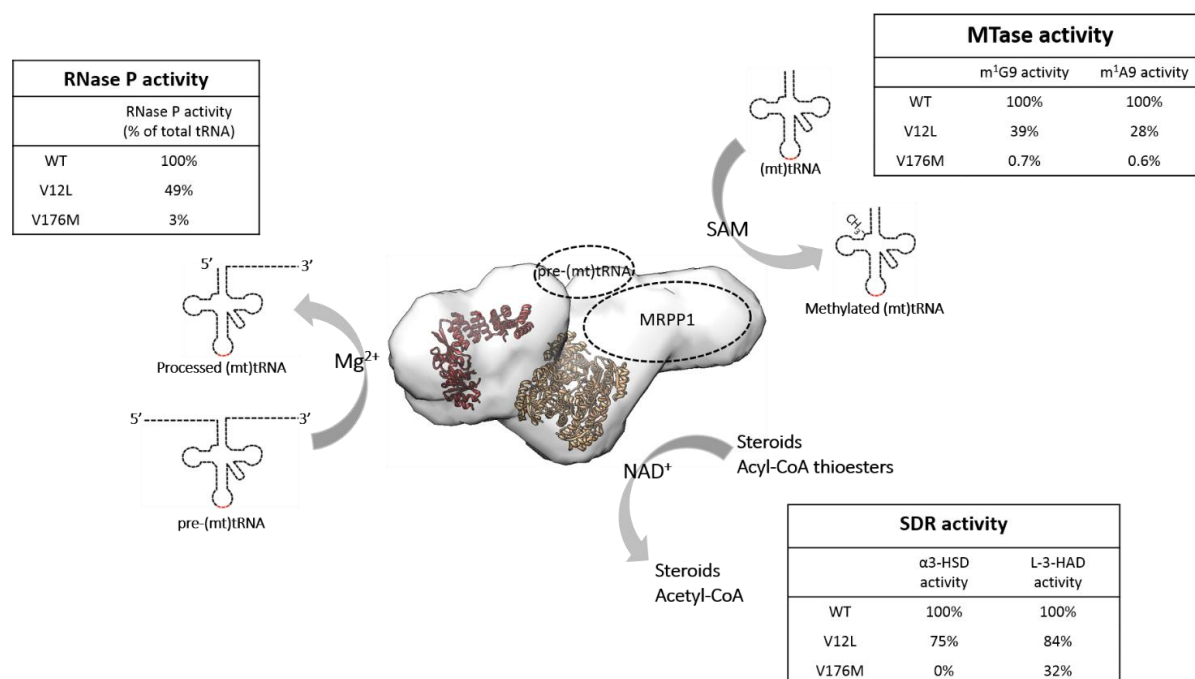


Figure 6-10. Outline of the residual activity given as percentage (%) of wildtype (WT), exhibited for mutated (V12L or V176M) MRPP2 proteins.

p.V176M: Structure-based interpretation for enzymatic functions

MRPP2(176M) exhibited a 3-fold decreased L-3-HAD activity, compared to wildtype, and no α 3-HSD activity. The Val176 residue is located in an α -helix (α E) that contributes to one side of the substrate binding pocket (**Figure 6-11**). α E is followed by a loop-helix-loop segment (between β E and α E) that contains many catalytically important residues, including Gln165 implicated in substrate binding of both acyl-CoA thioesters and steroids [185], [201], Ser155 from the catalytic tetrad [1], [41], Thr153 forming part of the proton-relay system [40], [41], Ala154 forming part of the cofactor binding pocket [201] and Gln162 interacts with the co-crystallised inhibitor AG18051 bound in the substrate binding pocket of MRPP2 [201], [202] (which could also play a role in binding of natural substrates). Mutation of Val176-to-Met could cause severe clashes with residues from this loop-helix-loop segment, *i.e.* the side-chain C ^{α} -atom of Ala156 and the side-chain ring of Phe159 (**Figure 6-11**). Such steric clashes could rearrange the loop-helix-loop segment and affect the position of the catalytically important residues resulting in altered kinetic parameters. The observed 3-fold increase in K_M towards L-3-OH-but-CoA of

MRPP2(V176M), and lack of activity towards the steroid substrate, compared to wildtype, shows poor substrate accommodation for this mutant protein, with a more severe effect for the steroid substrate (allopregnanolone) than for the CoA thioester (L-3-OH-but-CoA). This might be due to fewer contact points in the protein for the small steroid molecule compared to the larger CoA thioester substrate. A steroid-bound crystal structure of the MRPP2 paralogue from *Drosophila* (Scully), identified only one hydrogen bond and two hydrophobic interactions between the steroid and the protein, supporting the above hypothesis of few contact points [185], [201]. Interestingly, the residue that forms hydrogen bonds to the steroid is Gln165 (human MRPP2 numbering) located in the loop-helix-loop, which was proposed to be affected by the Val176-to-Met mutation. In support of a structural link between p.Q165H and p.V176M, proteins with these mutations showed similar behaviour in the biochemical studies. It would be interesting to consolidate the above finding in the future by testing the activity of MRPP2(V176M) towards other steroid hormone substrates such as β -estradiol.

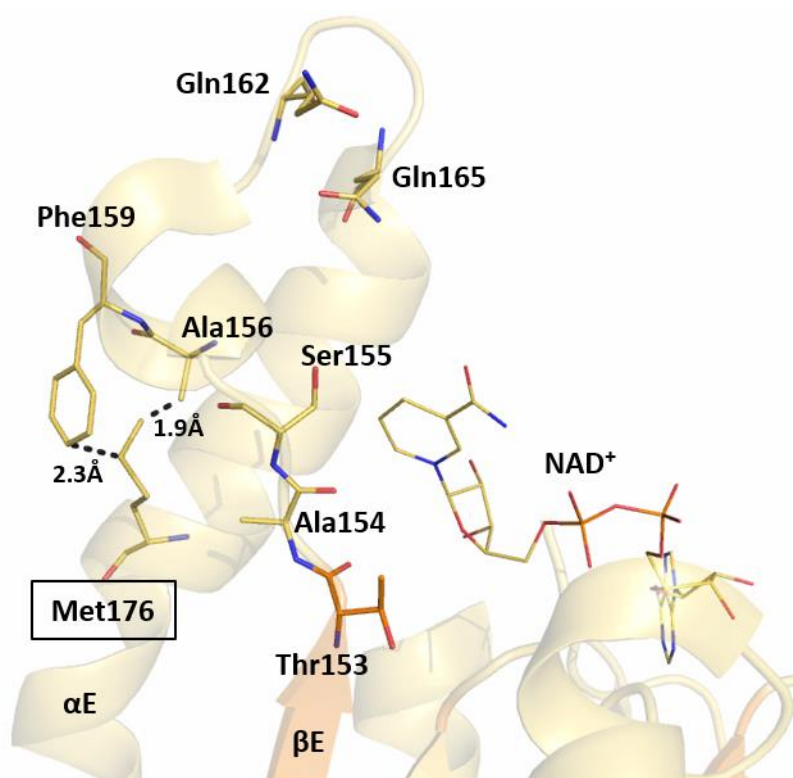


Figure 6-11. Effect of mutation of Val176 to a Met residue. Clashing or catalytically important residues are outlined and shown as sticks. The cofactor NAD⁺ is shown as lines, (PDB: 1U7T).

p.V176M: Structure-based interpretation for non-enzymatic functions

By comparison with wildtype MRPP1-MRPP2, the p.V176M mutation in MRPP2 caused the complex to dissociate which seems a likely explanation for the almost abolished m¹R9 MTase and RNase P activities observed for the MRPP1-MRPP2(V176M) complex. Biochemical studies performed in this work did not identify any overall structural differences (proteolytic pattern, oligomeric state) for MRPP2(V176M), compared to wildtype protein. The Val176 residue is also not surface-exposed where it could directly disturb MRPP1-MRPP2 complex formation. It is therefore puzzling how this mutation, initially thought to solely affect enzymatic catalysis, is able to disrupt the complex formation.

The abolished activities would suggest a severe phenotype for p.V176M, like the neonatal form. However, the patient with this mutation showed the almost asymptomatic atypical form. This inconsistency in the expected genotype-phenotype relationship could indicate that other factors in the patient compensates for the defect in MRPP2. A recent study identified 13 healthy individuals carrying mutations known to cause one of eight severe, genetic childhood disorders with complete penetrance, suggesting that genotypic rescue mechanisms exist in these patients that substantially diminish the consequences of their expected, severe disorder [203].

p.V12L: Structure-based interpretation for enzymatic functions

MRPP2(V12L) showed wildtype-like L-3-HAD activity and a 2.5-fold reduction in the α 3-HSD activity, compared to wildtype. Val12 is located in the beginning of the first β -strand (β A), one of the seven strands making up the core β -sheet of the Rossmann fold (**Figure 6-12**). When mutated to Leu, the residue would be in closer Van der Waals contact with the Arg84 main-chain and Phe82 side-chain atoms, at the end of the α C-helix. The

shorter distances could cause a slight rearrangement of all implicated residues resulting in destabilisation of the core β -sheet and possibly the overall structure. Such a destabilisation can explain the significantly lowered solubility of MRPP2(V12L) (**Section 5.1.2**) compared to wildtype. The various SDR activity assays were performed over the course of two weeks, during which tested activities saw continual reduction in activity compared to wildtype, *i.e.* the initial L-3-HAD activity was wildtype-like and the later measured 3 α -HSD activity was reduced 2-fold relative to wildtype. Though not tested systematically over time on the same substrate, the reduction in activity could reflect protein destabilisation over time, with corresponding loss of activity, as observed for the MRPP2(R130C) protein *in vitro* [115]. Altogether, the effect on enzymatic activities compared with the significantly lowered protein solubility suggests, that the Val12-to-Leu mutation mainly affects protein abundance rather than enzymatic function.

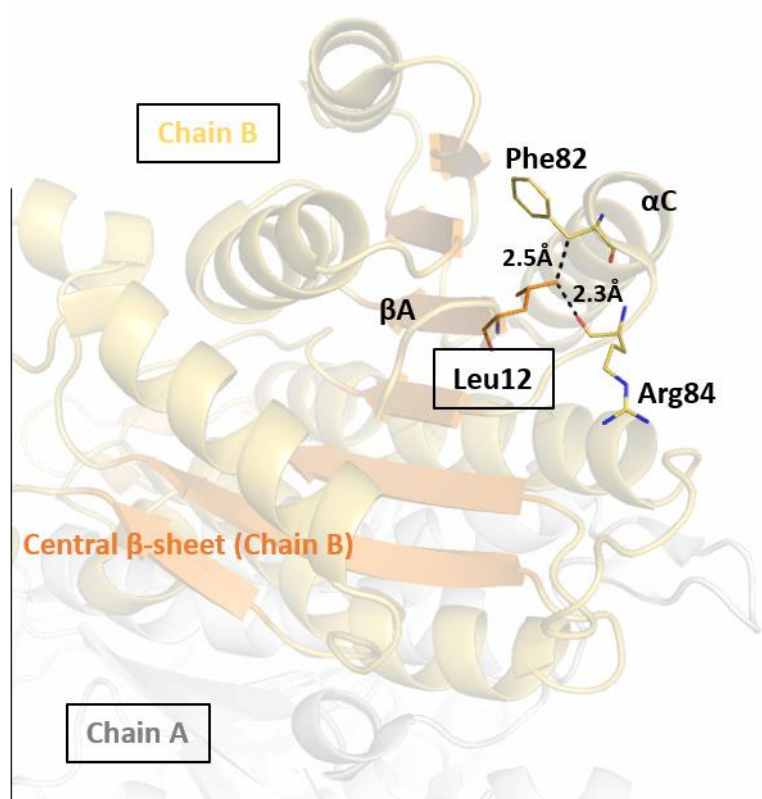


Figure 6-12. Effect of mutation of Val12 to a Leu residue (orange sticks). Clashing residues are outlined and shown as yellow sticks. Chain A is shown in grey and chain B is shown in yellow. The central β -sheet is outlined in orange, (PDB: 1U7T).

p.V12L: Structure-based interpretation for non-enzymatic functions

The p.V12L mutation also led to a reduction in non-enzymatic activity with a time-dependency as observed for the SDR functions, *i.e.* the co-expressed MRPP1-MRPP2(V12L) complex showed 2-fold reduction in RNase P activity and 2.5- and 3.5-fold reduction in the later measured m¹G9 and m¹A9 MTase activity, respectively. The decrease in activity for complexes involving this mutant protein could reflect a time-dependent decrease in the amount of the complex caused by mis-folding of MRPP2(V12L).

In addition to being located in the core β -sheet of the Rossmann-fold, Val12 is exposed on the protein surface (**Figure 6-13**), and from the SAXS model it could be in close proximity to both the (mt)tRNA substrate, MRPP1 and MRPP3. The decreased activity for the non-enzymatic functions could therefore also reflect a disruption of protein-protein or protein-substrate interactions, caused by the Val-to-Leu mutation. A high-resolution structure of MRPP1-MRPP2 in complex with the (mt)tRNA substrate is needed to confirm this possibility.

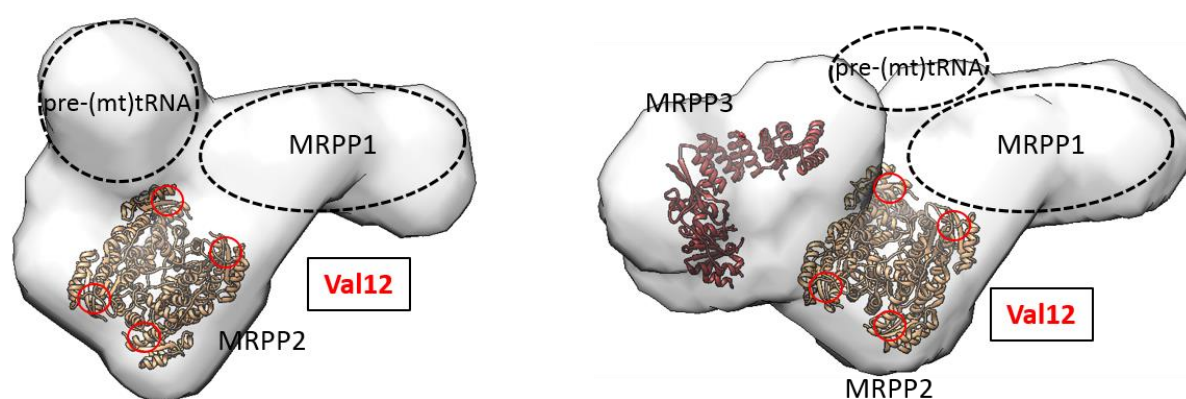


Figure 6-13. Location of Val12 in all four monomers of tetrameric MRPP2 shown in the SAXS model of the MRPP1-MRPP2(WT)-pre-(mt)tRNA^{Ile} and MRPP1-MRPP2(WT)-pre-(mt)tRNA^{Ile}-MRPP3 complexes.

The p.V12L missense mutation was characterised in a female patient and the clinical picture is not yet clear, preventing a correlation between residual activity and disease severity.

Comparison with published *in vitro* data

In this thesis, two of 15 known HSD10 disease mutations were functionally characterised *in vitro*. A similar *in vitro* characterisation was previously performed for five other known patient mutations [130], [134]. These data, together with the work presented here, are summarised in **Table 6-2**.

Mutation	Phenotype	Activities						Ref.
		SDR			MTase		RNase P	
		Oxo-red (NADH)	L-3-HAD (NAD ⁺)	3 α -HSD (NAD ⁺)	m ¹ G9	m ¹ A9		
WT	Normal	100%	100%	100%	100%	100%	100%	*
p.V12L	Unknown	N.D.	84.3%	38.5%	38.9%	27.9%	48.6%	T.S.
p.R130C	Infantile	20.6%	N.D.	N.D.	0%	N.D.	0%	[134]
p.V176M	Atypical	N.D.	32.4%	0%	0.66%	0.55%	3.10%	T.S.
p.P210S	Infantile	24.7%	N.D.	N.D.	11.4%	N.D.	12.7%	[134]
p.K212E	Infantile	26.4%	N.D.	N.D.	51.0%	N.D.	22.5%	[130]
p.R226Q	Neonatal	1.51%	N.D.	N.D.	0%	N.D.	0%	[134]
p.N247S	Neonatal	1.56%	N.D.	N.D.	0%	N.D.	0%	[134]

Table 6-2. Activity data from various studies of MRPP2 patient missense mutations. The activities are given as the percentage of respective wildtype protein studies. Oxo-red reaction: acetoacetyl-CoA to L-3-hydroxyacetyl-CoA. Substrates for L-3-HAD and 3 α -HSD reactions are described in Chapter 5. *This and other referenced studies. T.S.: This study.

The *in vitro* studies showed no clear picture of the effect of RNA methylation and processing, e.g. mutations causing *both severe and milder phenotypes* showed almost no residual MTase and RNase P activity (p.R130C vs. p.R226Q), agreeing with the lack of correlation between the almost abolished activities for MRPP2(V176M) and the associated *mild phenotype*. It is worth noting that these studies are performed solely on one or two (mt)tRNA substrates, and differences in activity likely exist across the 22 (mt)tRNAs.

Despite the lack of correlation between phenotype and RNA related activities, the *in vitro* studies suggested a potential correlation between phenotype severity and residual SDR (oxo-red) activity, *i.e.* mutations giving rise to the milder infantile form of disease (p.R130C, p.P210S, p.K212E) showed higher residual oxo-red activity compared to mutations causing the severe neonatal form (p.R226Q, p.N247S). Assuming that the reduction and oxidation of the CoA thioester (oxo-red and L-3-HAD activity, respectively) are catalysed similarly, this correlation agrees with data presented here that MRPP2(V176M), associated with a mild phenotype, retained some L-3-HAD activity. It does, however, clearly contrast with observations from the knock-out/rescue experiment in *Xenopus* embryos that demonstrated *no* correlation between residual L-methyl-3-HAD activity of MRPP2 and severity of the clinical presentation [115]. This, and the fact that p.V176M further completely abolish the (SDR) 3 α -HSD activity, clearly shows that more studies are needed to fully understand the molecular mechanism behind this disease.

6.6. Effect of mutations in MRPP3

6.6.1. Investigations of a missense mutation causing Perrault syndrome

During this thesis work, a mutant was identified in MRPP3 leading to Perrault syndrome (p.A485V). The Ala485 residue is highly conserved amongst NYN family proteins and is, in MRPP3, located in close proximity to the conserved Asp residues involved in catalysis. *In vitro* investigations of MRPP3(A485V), showed that this mutant protein was able to form the ternary complex with MRPP1-MRPP2 and the pre-(mt)tRNA substrate. Interestingly, the mutant protein seemed to form this complex even in the *absence* of substrate, which contrasts findings for the wildtype protein (**Section 4.8.1**). The reason for this effect is not understood but suggests that MRPP3 might form protein-protein interactions with the MRPP1-MRPP2 complex, even in the wildtype complex. The consistently higher T_m identified for the mutant protein found in DSF could reflect a

rearrangement in and/or around the catalytic site from the Ala485-to-Val substitution, reflecting the position of Ala485 in close proximity to the four catalytic Asp residues in MRPP3. A change to the catalytic site could further be the reason for the observed 19.75% overall reduction in the amount of processed (mt)tRNA, compared to wildtype MRPP3.

6.7. Future direction

This thesis has contributed structural and biochemical information on the uncharacterised MRPP1 protein and new insight into the structure, activity and assembly of the binary and ternary complexes of MRPP1, MRPP2, MRPP3 and pre-(mt)tRNA. In addition, it has provided new information on how clinically relevant mutants in MRPP2 affect these parameters. In doing so, the thesis has contributed to a deeper understanding of the molecular mechanism behind the devastating, rare genetic disorder, HSD10 disease, and enabled the suggestion that the disease phenotype might correlate with the level of MRPP2 protein in the patient that is able to engage in functional complex formation with MRPP1 and MRPP3.

More studies are needed on the functional side of characterisation, which could be expanded to cover the additional, less well-studied, SDR functions of MRPP2, not tested here, that provides MRPP2 with a potential role in e.g. glucocorticoid and bile acid biosynthesis [7], and its newly identified phospholipase C-like activity in removal of di- and monophosphatidylglycerol [15]. Such studies could also include testing the MTase and RNase P activity of the binary and ternary complex using mutant MRPP2 proteins on the full set of (mt)tRNAs.

A significant step towards elucidating the genotype-phenotype relationship would be the successful crystallisation of the binary and ternary complexes involving MRPP2. Such work would benefit from further studies confirming the 4xMRPP2:2xMRPP2, 4xMRPP2:2xMRPP1:1x(mt)tRNA and 4xMRPP2:2xMRPP1:1x(mt)tRNA:1xMRPP3

complex stoichiometries using e.g SEC-MALS. Crystallography attempts, in this work, for the full MRPP1-MRPP2 complex yielded only MRPP2 crystals, due to either excess MRPP2 not fully removed in purification or complex dissociation during crystallisation. In the case of dissociation, this problem could be overcome in future crystallisation attempts by crosslinking the protein components. Complex crystallisation could be further facilitated by using the catalytically inactive mutants of MRPP1 (p.D314N or p.Q226A) in crystallisation trials, or by supplementation of the SAM-analogues sinefungin, previously found to promote crystal formation of MTase-tRNA complexes [139]. The crystallisation process could also be further optimised by excluding reservoir solutions with similar chemical content, but with higher concentrations, compared to the mother liquor, as this directs the vapor diffusion towards the drop rather than the reservoir. Alternatively, the conditions could be included but spiked with the common content to increase concentration in the reservoir solution, compared to the drop.

In addition to understanding the structure and function of MRPP2, alone and as a member of the binary and ternary complex, it is important to understand how its enzymatic and non-enzymatic activities are regulated in the cell. As a starting point it is puzzling why a metabolic enzyme, acting on such diverse substrates, is moonlighting in complexes involved in mature (mt)tRNA/pre-(mt)tRNA methylation and processing. The modified (mt)tRNAs, produced by the binary and ternary MRPP1-MRPP2-(MRPP3) complexes, are used in synthesis of the 13 mitochondrial proteins that are all part of the oxidative phosphorylation pathway (OXPHOS), the principal output of which is ATP, but also NAD^+ . MRPP2 catalyses a diversity of reactions that all uses NAD^+ as a co-factor and thus, it is tempting to speculate that MRPP2 may act as a sensor for broadly assessing cellular NAD^+ levels. MRPP2 could use these levels to determine whether to participate in the binary and ternary complex (and facilitate NAD^+ synthesis) or act as an enzyme (and consume NAD^+). A very

simple outline of such a feedback mechanism is shown in **Figure 6-14**, where cellular levels of NAD^+ determine whether or not MRPP2 engage in complex formation.

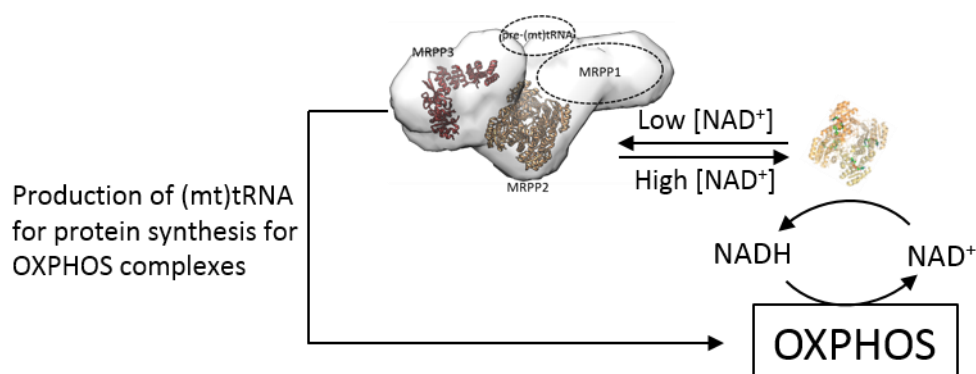


Figure 6-14. Schematic outline of a proposed regulatory link between NAD^+ levels and protein synthesis, with an inhibitory effect on complex

As a minimum requirement for such a regulatory model, one would expect the enzymatic function of MRPP2 to be suppressed when in complex. To assess this possibility, the 3α -HSD activity of MRPP2 towards allopregnanolone was tested alone and in the presence of MRPP1 under conditions that favoured complex formation. Interestingly, MRPP2 retained its catalytic activity under these conditions. Furthermore, analytical SEC showed that the MRPP1-MRPP2 complex remained intact in excess NAD^+ and NADH, suggesting that the above model is, at best, an oversimplification. Since MRPP2 retained SDR activity when bound to MRPP1, it is possible that MRPP2 remains in complex during enzymatic activity, simultaneously blocking MTase and RNase P activity. To assess such a regulation, the three different activity assays set up during this thesis could be used to measure activities in the presence of increasing amount of substrates and co-factors utilised by complex partners for other activities, *i.e.* addition of NAD^+ and Mg^{2+} during MTase activity measurements and addition of SAM and NAD^+ when measuring RNase P activity.

In any case, understanding the cellular regulation of the binary MTase and ternary RNase P complex constitutes another important line of research, which, in combination with a detailed understanding of the structure and function of these complexes, will eventually

provide enough insight to hopefully develop novel and effective treatments for the devastating HSD10 disease as well as other diseases associated with proteins in these complexes.

7 CHAPTER

MATERIALS AND METHODS

7.1 Cloning

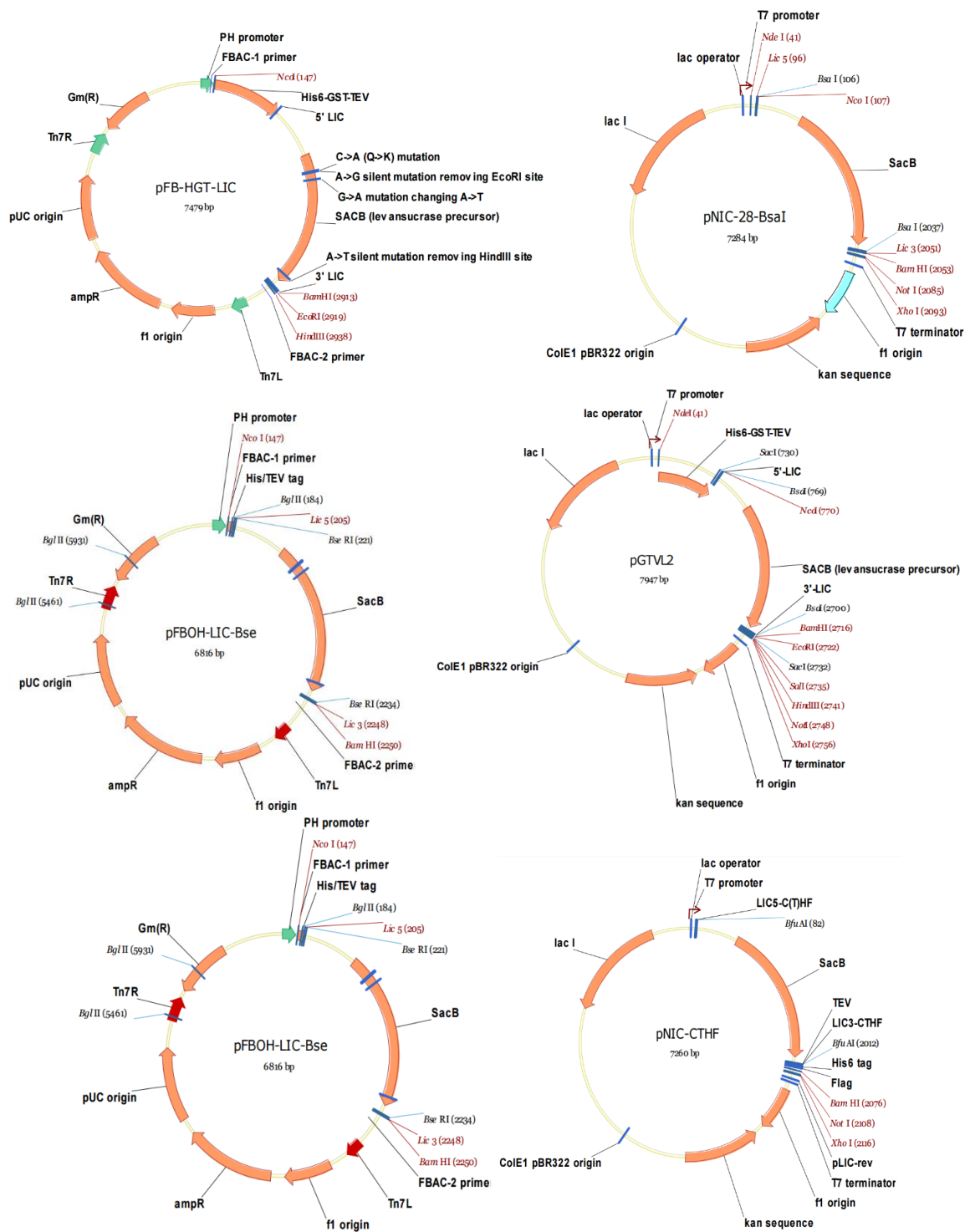
cDNA clones for MRPP1 (IMAGE: 4719195), MRPP2 (IMAGE: 2819721) and MRPP3 (IMAGE: 5206545) were available in-house before the start of this project.

7.1.1 Ligation independent cloning (LIC)

LIC cloning was performed by the SGC Biotechnology Group for the MRPP1, MRPP2 and MRPP3 constructs shown in **Table 7-1**. The vectors utilised in the LIC cloning are specified with the corresponding tag and protease cleavage site position. pNIC28-Bsa4 (Genbank EF198106), pGTVL2 (Genbank JF522100), p11 and pNIC-CTHF (Genbank EF199844) are *E. coli* expression vectors and pFB-LIC-Bse (Genbank EF199842), pFB-L21-LIC-Bse, pFB-HGT-LIC, pFB-6HZB and pFB-CT10HF-LIC are baculovirus transfer vectors. The vector maps are shown in **Figure 7-1**.

Construct boundaries	Vector	Tag/protease site
MRPP1		
7-403, 7-380, 15-403, 40-403. 40-403, 40-380, 40-210, 61-403, 90-403, 111-403, 111-380, 111-210, 136-403, 195-403, 199-403, 203-403, 203-380	pNIC28-Bsa4	His ₆ -TEV-N
40-403, 40-380, 111-403, 111-380, 203-403, 203-380	pFB-LIC-Bse	His ₆ -TEV-N
40-403. 40-380, 40-210, 111-403, 111-380, 111-210, 203-403 203-380	pGTVL2	His ₆ -GST-TEV-N
40-403, 40-380, 203-403, 203-380	pNIC-CTHF	Flag-His ₆ -TEV-C
40-403, 40-380	pFB-L21-LIC-Bse	His ₆ -TEV-N
40-403, 40-380	pFB-HGT-LIC	His ₆ -GST-TEV-N
40-403, 40-380	pFB-6HZB	His ₆ -ZBc-TEV-N
40-403, 40-380	pFB-CT10HF-LIC	His ₁₀ -Flag-TEV-C
MRPP2		
1-261	pNIC28-Bsa4	His ₆ -TEV-N
	p11	His ₆ -TEV-N
MRPP3		
1-583, 20-583, 40-583, 51-583, 51-380, 51-332, 51-290, 51-274, 86-583, 86-290, 86-274, 91-583, 91-274, 96-583, 111-502, 111-474, 111-414, 118-583, 118-502, 118-474, 118-380, 118-360, 118-344, 118-332, 152-583, 152-360, 152-344, 152-320, 274-583, 299-583, 320-583, 354-583, 373-583.	pNIC28-Bsa4	His ₆ -TEV-N

Table 7-1. Construct boundaries, vector, tag and protease cleavage site for vector for all constructs cloned for MRPP1, MRPP2 and MRPP3.



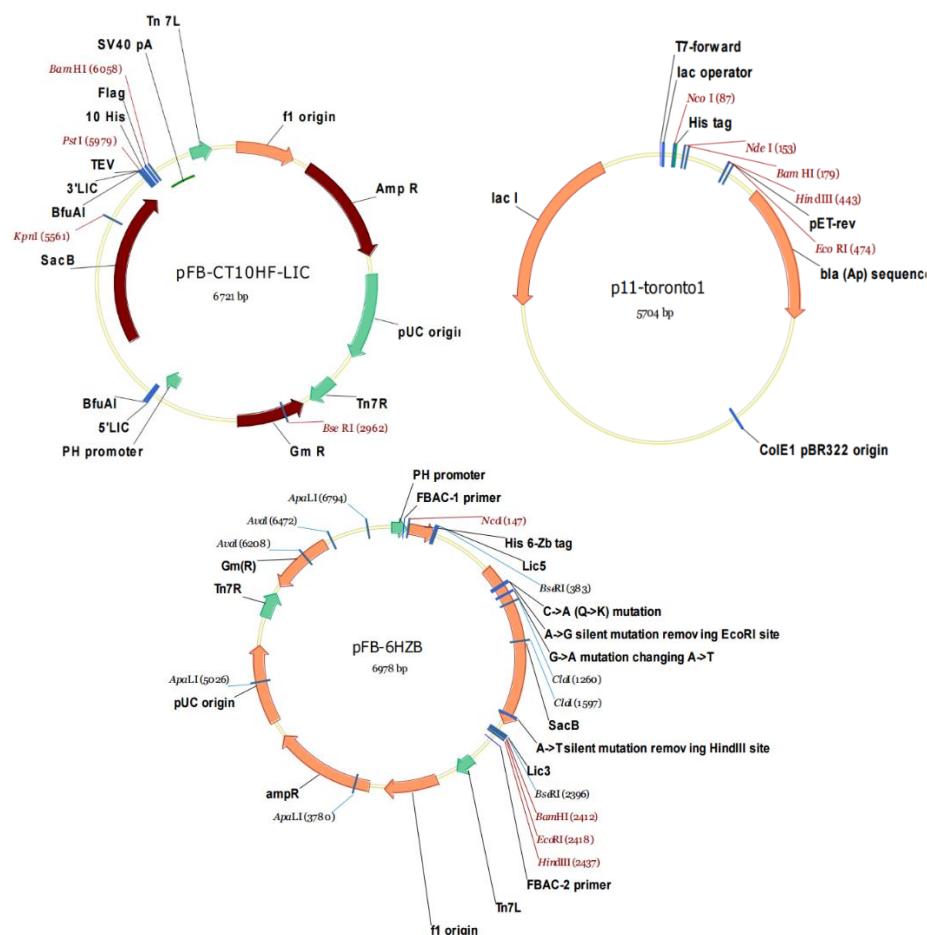


Figure 7-1. Vector maps for *E. coli* expression and baculovirus transfer vectors used in cloning.

Cloning of an *E. coli* co-expression plasmid was performed by Dr. Claire Strain-Damerell at the SGC (**Figure 7-2**). An insert is produced from a polymerase chain reaction (PCR) amplified DNA fragment encoding mature MRPP1 and full-length MRPP2, separated by a ribosomal binding site (RBS). The insert is cloned into the pNIC28-Bsa4 vector resulting in a plasmid with two coding regions: an N-terminally His₆-tagged MRPP1 and an untagged MRPP2. The resulting plasmid contains one T7-promotor region and two RBSs (one for each coding region).

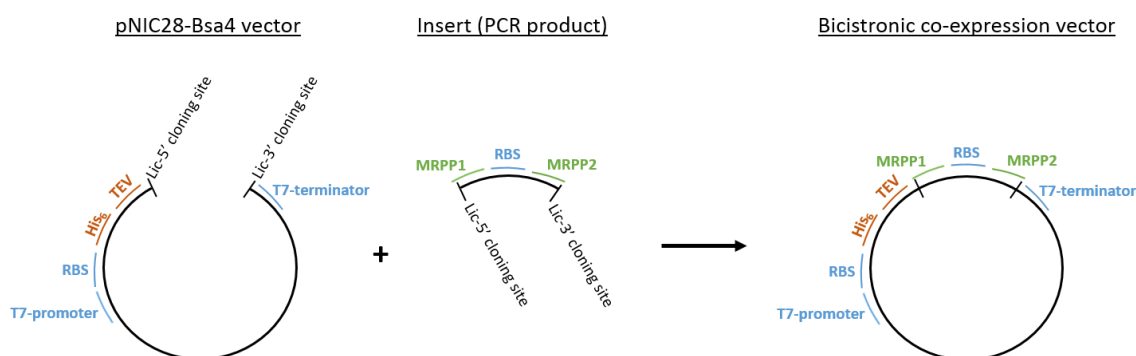


Figure 7-2. Cloning procedure of the *E. coli* bicistronic co-expression vector system.

7.1.2 Site-directed mutagenesis (SDM)

Clones of MRPP2 patient mutants p.V65A, p.D86G, p.L122V, p.R130C, p.Q165H, p.P210S, p.R226Q, p.N247S and p.E249Q in the pNIC28-Bsa4 vector were generously provided by Dr. Albert Amberger, Dr. Andrea Deutschmann and Prof. Johannes Zschocke from Division of Human Genetics, Innsbruck Medical University.

Mutagenesis of the inactive MRPP3 mutant p.D479A onto the pNIC28-Bsa4 vector was performed by the SGC Biotechnology Group.

SDM was performed on already cloned constructs of full-length MRPP2 (residues 1-261) and mature MRPP3 (residues 51-583) in the pNIC28-Bsa4 vector, and on mature MRPP1 (residues 41-403) and full-length MRPP2 in the bicistronic pNIC28-Bsa4 vector, using primers shown in **Table 7-2**.

Forward and reverse primer, at stock concentrations of 100 μ M, were diluted with ultrapure water to a final concentration of 10 μ M of each primer. 1 μ L of this mixture was added to a mastermix of 0.35 μ L Platinum Pfx polymerase (2.5 U/ μ L) in 5 μ L 10xPCR buffer, 5 μ L PCR enhancer, 1 μ L 40 mM MgSO₄, 1.25 μ L 10 mM dNTPs, 35.40 μ L ultrapure water and 1 μ L DNA template plasmid (50 ng/ μ L) to a final volume of 50 μ L. PCR was performed in a thermocycler (Multigene Labnet) with 5 min at 94°C for initial denaturation followed by 30 cycles of 30 seconds denaturation at 94°C, 30 sec annealing at

53°C, and 4-5 min (depending on template length) extension at 68°C with a final extension for 5 min at 68°C.

The mixture with the amplified PCR product was subsequently treated with 1 µL *DpnI* (10 U/µL) at 37°C for 2 h in a thermocycler (Bio-Rad C1000 Touch). DNA concentration was determined by measuring the absorbance at 260 nm on a NanoDrop spectrophotometer (Thermo Scientific, 2000) (**Section 7.4.1**).

5 µL of the amplified PCR product was added to 50 µL of *E. coli* Mach1 competent cells [204] and incubated on ice for 30 min. The transformation sample was heat shocked at 42°C for 45 sec and chilled on ice for 2 min, before addition of 250 µL Miller's Lysogeny broth (LB-Miller) and further incubation at 37°C for 90 min at 200 revolutions per minute (rpm). The transformation mixture was plated onto LB-Miller/agar plates containing 50 µg/mL kanamycin, and incubated at 37°C overnight. The next day, single colonies were used to inoculate overnight cultures, from which cell pellets were harvested the next morning by centrifugation (Eppendorf centrifuge 5810R) at 4°C for 30 min at 4000 rpm. The plasmid was extracted from the cell pellet and purified using a Mini-prep Kit (Qiagen). Introduction of the mutation was confirmed by Sanger-sequencing of the purified plasmid using a T7-promoter primer (TAATACGACTCACTATAGGG) for MRPP2 and T7-terminator primer (GCTAGTTATTGCTCAGCGG) for MRPP3 in the pNIC28-Bsa4 vector, and with internal primers (TTGCTTTTAACATCAACAGAAAAGTCTCATGTAG and CAAGCTGCTAGAAAC CACTGAGGAAGATAAACAG) for the p.D314N and p.Q226A mutation site of MRPP1, respectively, in the bicistronic pNIC28-Bsa4 vector.

Mutation	Forward primer	Reverse primer	Template
MRPP1			
p.Q226A	CAGAATACTGTTTCCGC GCTTTTAGAAAGTGA	TCACTTTCTAAAAGCG CGGAAACAGTATTCTG	MRPP1(40-403) Bicistronic pNIC28-Bsa4
p.D314N	ATTGGGTCTTTTGTTAA TAAGAGTATGCAGC	GCTGCATACTCTTATT AACAAAAGACCCAAT	MRPP1(40-403) Bicistronic pNIC28-Bsa4
MRPP2			
p.V12L	GTCAAGGGCCTGCTGGC GGTAATAACC	GGTTATTACGCGCAGC AGGCCCTTCAC	Full-length MRPP2(1-261) MRPP1(40-403) Bicistronic pNIC28-Bsa4
p.A154T	GTCATCATCAACACTAC CAGTGTGGCTGCCTTC	GAAGGCAGCCACACT GGTAGTGTGATGATG AC	Full-length MRPP2(1-261)
p.V176M	TCCAAGGGGGGAATAA TGGGCATGACACTGGCC	GGGCAGTGTGATTATT CCCCCTTGGA	Full-length MRPP2(1-261) MRPP1(40-403) Bicistronic pNIC28-Bsa4
MRPP3			
p.V485V	CCATTCCTTCTGTATGT CACACTGCACTCCGGG	CCCGGAGTGCAGTGT GACATACAGAAGGAA TGG	Mature MRPP3(51-583)

Table 7-2. Mutation, vector and DNA template plasmid for all mutants generated of MRPP1, MRPP2 and MRPP3.

7.2 Protein expression

7.2.1 Bacterial small-scale expression

1 μ L of plasmid at $\sim$ 50 ng/ μ L was transformed into 30 μ L of the *E. coli* BL21(DE3) competent cells [205] containing the pRARE2 plasmid for *E. coli* expression. A single colony was used to inoculate a starter culture of 2 mL LB-Miller containing 50 μ g/mL kanamycin and 34 μ g/mL chloramphenicol grown in a 50 mL falcon tube. The cells were cultured overnight at 37°C in a Multitron Standard incubator (InforsHT) shaking at 180 rpm. The next morning, 500 μ L of the culture was added to 50 mL TB containing 50 μ g/mL kanamycin in a 250 mL baffled shaker flask, and cultured at 37°C, until the optical density (OD) reached 0.8, at which point temperature was reduced to 18°C. At OD 1.8, 0.1 mM

IPTG was added to each culture to induce protein expression overnight. The following morning, cell pellets were harvested by centrifugation (Beckman Coulter Avanti JXN-26) at 4°C for 30 min at 4000 rpm and the pellets were stored in a 50 mL falcon tubes at -80°C.

7.2.2 Bacterial large-scale expression

12 mL starter culture was prepared for each 1 L culture to be produced grown in 250mL baffled shaker flasks. 10 mL of the overnight culture was added to each of 1 L TB (up to 24 L) containing 50 µg/mL kanamycin in 1L baffled shaker flask. The cells were cultured and induced for protein expression as before. The following morning, the cell pellets were harvested by centrifugation (Beckman Coulter Avanti JXN-26) at 4°C for 12 min at 5000 rpm, and the pellets were stored in 50 mL falcon tubes at -80°C.

7.3 Protein purification

7.3.1 Purification buffers

Lysis buffer: 500 mM NaCl, 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 5% glycerol, 0.5 mM tris(2-chloroethyl) phosphate (TCEP), 20 mM imidazole

Wash buffer: 500 mM NaCl, 50 mM HEPES pH 7.5, 5% glycerol, 0.5 mM TCEP, 40 mM imidazole

Elution buffer: 500 mM NaCl, 50 mM HEPES pH 7.5, 5% glycerol, 0.5 mM TCEP, 250 mM imidazole

SEC buffer: 500 mM NaCl, 50 mM HEPES pH 7.5, 5% glycerol, 0.5 mM TCEP

Zero salt IEX buffer: 50 mM HEPES pH 6.8-7.5, 5% glycerol, 0.5 mM TCEP

Low salt IEX buffer: 50 or 300 mM NaCl, 50 mM HEPES pH 6.8-7.5, 5% glycerol, 0.5 mM TCEP

High salt IEX buffer: 2 M NaCl, 50 mM HEPES pH 6.8-7.5, 5% glycerol, 0.5 mM TCEP

Loading buffer: 1% w/v Sodium dodecyl sulphate (SDS), 31.5 mM Tris-HCl pH 6.8, 10% w/v glycerol, 180 mM dithiothreitol (DTT), 0.005% bromophenol blue

MES buffer: 0.1 M 2-(N-morpholino)ethanesulphonic acid (MES) pH 6.1

7.3.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

20 μ L sample was added with 5 μ L loading buffer and boiled for 3 min at 95°C. 10 μ L of the boiled sample was loaded onto a 4-12% bis-tris polyacrylamide gel (Invitrogen) electrophoresed for 60 min at 165 V in MES buffer. 10 μ L of standard protein markers (Bio-Rad) were run alongside the protein samples. The SDS-PAGE gel was stained with Coomassie Brilliant Blue (Generon) and imaged using the Bio-Rad ChemiDoc Imaging System.

7.3.3 Protein extraction for small-scale purification

Frozen cell pellets were thawed in a 50 mL falcon tube placed in lukewarm water and resuspended on ice in 2 mL Lysis buffer. Cells were lysed by sonication (Sonics, Vibra Cell) over 12 min at 35% amplitude with the sonicator pulsing 'ON' for 5 seconds and 'OFF' for 10 seconds. Cell lysate was clarified by centrifugation (Eppendorf centrifuge 5810R) at 4°C for 30 min at 4000 rpm. A small amount of the centrifuged cell pellet was resuspended in 20 μ L binding buffer and subjected to SDS-PAGE, to ensure that the target protein was not present in the insoluble fraction. The supernatant (containing clarified lysate) was utilised for further purification.

7.3.4 Protein extraction for large-scale purification

Frozen cell pellets were thawed and resuspended on ice in 50 mL lysis buffer containing 1 mL protease inhibitor cocktail set III (Calbiochem) per L of harvested cell culture. The cells were lysed by passing five times on a homogeniser (EmulsiFlex-C5-Avestin) to a pressure of 1000 bar. The cell lysate was clarified by centrifugation

(Beckman Coulter Avanti JXN-26) at 4°C for 1 h at 16500 rpm. A sample of the centrifuged cell pellet was taken for SDS-PAGE and the supernatant was utilised for further purification.

7.3.5 Small-scale protein purification

The clarified lysate from small scale extraction was applied under gravity flow to a 0.5x10 cm (DxH) column (Bio-Rad) containing 75 µL Ni-NTA affinity resin (GE Healthcare) pre-equilibrated with Lysis buffer. The Ni-NTA resin was washed with 2 x 8 column volumes (CVs) Lysis buffer and 2 x 8 CVs Wash buffer. The bound protein was eluted with 1 x 16 CVs of Elution buffer. 20 µL of each wash and elution sample were subjected to SDS-PAGE.

7.3.6 Immobilised metal affinity chromatography on batch column (IMAC)

The clarified lysate from large scale extraction was added with 1.25-3 mL of Ni-NTA affinity resin (GE Healthcare) and incubated at 4°C for 1 h rotating at 8 rpm. The lysate, together with Ni-NTA resin, was centrifuged (Beckman Coulter Avanti JXN-26) for 30 min at 5000 rpm at 4°C. The Ni-NTA pellet was applied onto a 1.5x10 cm (DxH) column (Bio-Rad) under gravity flow and the flow-through was collected for SDS-PAGE analysis. The Ni-NTA resin was washed with 2 x 10 CVs Lysis buffer and 2 x 10 CVs Wash buffer. The bound protein was eluted with 5-6 x 2 CVs of Elution buffer. 20 µL Lysis buffer was added with 0.75 µL of the captured flow-through and this sample, along with 20 µL of each wash and elution fraction, were subjected to SDS-PAGE. The fractions containing the protein of interest were pooled for further purification.

7.3.7 Size exclusion chromatography (SEC)

The pooled fractions from IMAC were concentrated to 5 mL using an Amicon Ultra centrifugal filter device (Merck Millipore) with a 10-30 kDa cut-off depending on the target protein size. The concentrated sample was loaded onto a 120 mL size exclusion

chromatography (SEC) column (Sephadex 200 HiLoad 16/60, Sephadex 75 HiLoad 16/60 or Superose 6 prep grade XK 16/70) connected to an AKTA-Express (GE Healthcare) in 4°C. Before loading the protein, the column was pre-equilibrated with SEC buffer. Protein was loaded and eluted at a rate of 1 mL/min and 1.8 mL fractions were collected throughout the run. Protein was detected by UV absorbance and was observed in the chromatogram as an increase in absorbance (peak) when eluted. 20 µL of each peak fraction were subjected to SDS-PAGE. The fractions containing the protein of interest were pooled for further purification.

7.3.8 Treatment with Tobacco etch virus protease

The concentration of protein in the pooled fractions from SEC was determined (**Section 7.4.1**). To remove the affinity tag from target proteins, Tobacco etch virus (TEV) protease was added to the eluates in a 1:20 (w/v) TEV-to-protein ratio and the solution was incubated overnight at 4°C. 20 µL samples of before and after TEV protease treatment were evaluated for cleavage efficiency by SDS-PAGE.

7.3.9 Reverse IMAC on batch column

The TEV protease treated eluent was applied twice under gravity flow to a 2 x 10 cm column containing 125 µL Ni-NTA affinity resin (GE Healthcare) equilibrated with Lysis buffer and the flow-through was collected for SDS-PAGE. The Ni-NTA resin was washed with 2 x 8 CVs Lysis buffer and 2 x 8 CVs Wash buffer. TEV protease, and any uncut resin-bound protein, was eluted with 2 x 8 CVs Elution buffer. From SDS-PAGE analysis, fractions containing the protein of interest were pooled.

7.3.10 Ion exchange chromatography (IEX)

The pooled fractions from Reverse IMAC were concentrated to 5 mL using an Amicon Ultra centrifugal filter device with a 10-30 kDa cut-off. The concentrated sample was diluted with Zero salt IEX buffer to 50 mL to reduce NaCl concentration and loaded

onto an anion- or cation chromatography column (HiTrap SP HP 5 mL/1 mL, HiTrap Q HP 5 mL/1 mL, Resource S 1 mL or Resource Q 1 mL) (GE Healthcare) connected to an AKTA-Purifier (GE Healthcare) at 4°C. Before loading the protein, the column was pre-equilibrated with Low salt IEX buffer. The protein was loaded onto the column at 0.5 mL/min via a 50 mL superloop and the flow-through was collected for SDS-PAGE. The column-bound protein was eluted at a rate of 1 mL/min with a linear gradient of 0-35% High salt IEX buffer over 30-40 CVs. 1 mL fractions were collected throughout elution. 20 µL of the flow-through and peak fractions were subjected to SDS-PAGE.

7.3.11 Protein storage

The fractions pooled from the final purification step were concentrated, and the final protein concentration was measured (**Section 7.4.1**). The concentrated protein was flash frozen in liquid nitrogen and stored at -80°C in 50 µL aliquots.

7.3.12 Buffer exchange

Protein was exchanged into buffers designated for different assays by loading the sample onto a HiPrep 26/10 Desalting column connected to an AKTA-Purifier (GE Healthcare) at 4°C and eluted at 4 mL/min with 1.8 mL fraction collection throughout. The fractions containing the protein were pooled, concentrated and stored until use.

7.3.13 Protein biochemical characterisation

7.4 Assay buffers

Analytical SEC buffer: 100 mM NaCl, 50 mM Tris-HCl pH 8.0, 3 mM DTT, 3 mM MgCl₂

TCA resuspension buffer: 4% w/v SDS, 0.2 M Tris-base, 0.15 M NaOH

DSF buffer: 150 mM NaCl, 10 mM HEPES pH 7.5

AUC buffer: 50 mM sodium phosphate pH 7.4, 300 mM NaCl, 5% glycerol, 1% sucrose, 0.5 mM TCEP

EMSA buffer: 100 mM NaCl, 50 mM Tris-HCl pH 8.0, 3 mM DTT, 3 mM MgCl₂

Limited proteolysis buffer: 10 mM HEPES pH 7.8, 150 mM NaCl

PBS buffer: 8mM Na₂HPO₄, 150mM NaCl, 2mM KH₂PO₄, 3mM KCl, pH 7.4

PBST buffer: 8mM Na₂HPO₄, 150mM NaCl, 2mM KH₂PO₄, 3mM KCl, 0.05% Tween-20, pH 7.4

Loading buffer: 1% w/v SDS, 31.5 mM Tris-HCl pH 6.8, 10% w/v glycerol, 180 mM DTT, 0.005% bromophenol blue

EMSA loading buffer: 100 mM Tris-HCl, 10% glycerol, 0.0025% Bromophenol Blue pH 8.6

MES buffer: 0.1 M MES pH 6.1

TAE buffer: 40mM Tris-base pH 8.3, 20 mM acetic acid, and 1 mM EDTA

Tris-glycine buffer: 25 mM Tris-base, 192 mM glycine pH 8.3

7.4.1 Protein concentration determination

Protein concentrations was determined by measuring the absorbance at 280 nm on a NanoDrop spectrophotometer (Thermo Scientific, 2000) against a buffer blank, using the molar extinction coefficient (ϵ) and MW of the protein derived from the primary sequence in the ExPaSy ProtParam [206]. The ϵ -values were estimated from standard values for Tyr ($\epsilon = 1490 \text{ M}^{-1}\text{cm}^{-1}$) and Trp ($\epsilon = 5500 \text{ M}^{-1}\text{cm}^{-1}$) residues in the protein. Protein concentrations were calculated within the NanoDrop software using Lambert-Beer's Law (**Equation 1**), where A is the absorbance of the protein, ϵ is the molar extinction coefficient ($\text{M}^{-1} \text{ cm}^{-1}$), c is the protein concentration in molar (M) and l is the path length of which the light has to travel in centimetres (cm).

$$\text{Equation 1: } A = \epsilon \cdot l \cdot c$$

7.4.2 Electrospray ionisation mass spectrometry (ESI-MS)

Intact molecular masses were determined by ESI-MS on an Agilent LC/MSD TOF system with reversed-phase high-performance liquid chromatography (HPLC) coupled to electrospray ionization and an orthogonal time-of-flight mass analyser. Proteins were desalted prior to mass spectrometry by rapid elution off a C3 column with a gradient of 5-95% isopropanol in water with 0.1% formic acid. Spectra were analysed using the Agilent Masshunter Qualitative Analysis software. Molecular masses were mapped onto the protein sequence using the PAWS Software, with a search tolerance of 100 parts per million (ppm).

7.4.3 Tryptic digest - tandem mass spectrometry (MS/MS) analysis

Protein gel bands from SDS-PAGE were excised and dissolved in 200 μ L 10% methanol. The mixture was subjected to tryptic digest-MS/MS analysis performed by Dr. Rod Chalk and Ms. Oktawia Borkowska at the SGC. The obtained data were analysed using the MASCOT software[207], and identified peptides were mapped on to the sequence manually to determine the covered protein sequence for each protein gel band.

7.4.4 Analytical SEC

100 μ L protein sample of varying concentrations (1 -20 mg/mL) was injected via a 100 μ L loop onto a 24 mL SEC column (Superdex S200 10/300 GL, Superdex S75 10/300 GL) (GE Healthcare) pre-equilibrated in Analytical SEC buffer. The protein was loaded and eluted at a rate of 0.4 mL/min and 1 or 0.5 mL fractions were collected throughout the run. Protein elution was detected as a UV absorbance peak with the specific elution volume (V_e) given as the maximum absorbance value. 500 μ L of each collected fraction was trichloroacetic acid (TCA) precipitated by addition of 36 μ L 100% TCA and incubated at room temperature for 15 min. The fractions were centrifuged (Labnet PrismR microcentrifuge) at 14000 rpm for 15 min, and the precipitated protein pellet was resuspended in 15 μ L TCA resuspension buffer and subjected to SDS-PAGE analysis.

Five proteins of known MW (Bio-Rad) (**Table 7-3**) were applied to analytical SEC under the same conditions as for the protein of interest. A plot of $\text{LOG}^{10}(\text{MW})$ against the obtained V_e -values for each protein produces a linear function from which MWs for other proteins can be deduced based on their V_e -values.

Protein	MW (Da)
Thyroglobulin (bovine)	670000
γ -globulin (bovine)	158000
Ovalbumin (chicken)	44000
Myoglobin (horse)	17000
Vitamin B12	1350

Table 7-3. Proteins with known MWs used to create the standard curve for an analytical SEC assay.

7.4.5 Differential scanning fluorimetry (DSF)

DSF measures the thermal unfolding of proteins in solution by detecting the change in fluorescence from a probe upon binding of this probe to hydrophobic residues exposed from the protein core when it is being unfolded. The change in probe fluorescence is monitored over a stepwise thermal gradient from 25-95°C. Fluorescence intensity can be plotted as a function of temperature, resulting in a sigmoidal curve that can be described as a two-state transition (unfolding and precipitation/aggregation) [150].

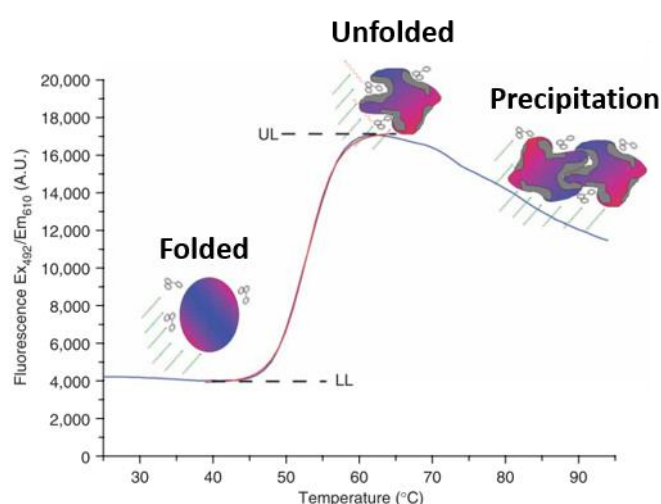


Figure 7-3. Sigmoidal curve resulting from plotting probe fluorescence intensity as a function of temperature. The protein unfolding is indicated along with the probe binding upon unfolding. LL and UL are the values of minimum and maximum intensities, respectively. Adapted from Niesen *et al.* [150].

The inflection point of the sigmoidal curve corresponds to the protein melting temperature (T_m) and can be calculated by fitting the curve to the Boltzmann equation (**Equation 2**) where LL and UL are the values of minimum and maximum intensities, respectively, and a is the slope of the curve at the T_m [150].

$$\text{Equation 2: } y = LL + \frac{(UL-LL)}{e^{\frac{T_m-x}{a}}}$$

The T_m represents the point at which there exists an equal amount of folded and unfolded proteins in solution. If the protein is presented to a ligand before heating, a protein-ligand interaction could stabilise the protein and cause a shift in the protein T_m (ΔT_m) proportional to the binding affinity of the ligand [150].

In this study, the fluorescence probe of choice was SYPRO ORANGE due to its high signal-to-noise ratio, and relatively high excitation wavelength (near 500 nm). The high excitation wavelength decreases the likelihood of any small molecules interfering with the optical properties of the dye causing a quenching of the fluorescence signal [150].

DSF was performed in a 96-well plate using an Mx3005p RT-PCR machine (Stratagene) with excitation and emission filters of 492 and 610 nm, respectively. Each well consisted of 2 μ L protein of 2 μ M in DSF buffer, 2 μ L SYPRO ORANGE diluted 1000-fold in DSF buffer from the manufacturers stock (Invitrogen) and 2 μ L ligand at various concentrations between 1 and 40 mM. Fluorescence intensities were measured from 25 to 96°C with a ramp rate of 3°C/min. The T_m was determined by curve-fitting to **Equation 2** using Excel 2016 and GraphPad Prism (v.5.01) software.

7.4.6 Sedimentation velocity analytical ultracentrifugation (SV-AUC)

SV-AUC studies the real-time movement (sedimentation) of macromolecules in solution in high centrifugal fields using hydrodynamic theory to define the size, shape and interactions of the macromolecules in solution. During AUC, macromolecules are

characterised in their native state in any buffer of choice, making this non-invasive method ideal to investigate biologically relevant conditions [168].

The redistribution of mass in a gravitational field in an SV-AUC experiment depends on the rate at which molecules move during the redistribution. This rate depends on three forces (**Figure 7-4**): F_s that acts on the particle from the gravitational field which is influenced by the mass of the particle, the rotor speed and the distance from the centre of the rotor. A counterforce, F_b , that corresponds to the mass displayed by the particle as it sediments. A frictional force, F_f , is developed by the particle as it travels through the buffer [168], [208].

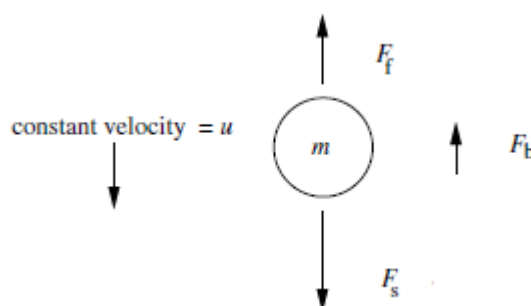


Figure 7-4. Forces working on a particle as it sediments in an SV-AUC experiment. Adapted from Introduction to Analytical Ultracentrifugation, G. Ralston, 1993 [208].

Balancing these forces gives a ratio of the velocity to the centrifugal field is described as the sedimentation coefficient (s). The radial concentration distribution during sedimentation show spreading of the sedimentation boundaries due to diffusion over time, and decrease of the concentration in the plateau region decreases from radial dilution (**Figure 7-5**) [168], [208].

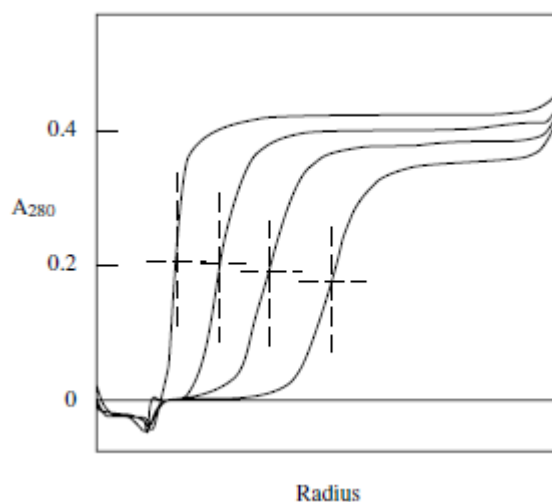


Figure 7-5. Movement of the sedimentation boundaries over time in an SV-AUC experiment. Adapted from Introduction to Analytical Ultracentrifugation, G. Ralston, 1993 [208].

During the AUC experiment, the radial concentration distributions are measured at intervals ranging from min. to hours. As the rotor spins, a measurement of the concentration of molecules at closely spaced radial intervals in the cell takes place every time the cell passes through the optical paths of absorbance detector. These scans serve as the input for data analysis in which the sedimentation coefficient distribution ($c(s)$) function can be obtained from a direct fit to the data by assuming a constant shape and an equal frictional ratio for all species. The $c(s)$ distribution can then be converted to a distribution of molar masses ($c(M)$ distribution) [168], [209].

To prepare the samples for SV-AUC, 400 μL of MRPP1-MRPP2 protein complex was dialysed in AUC buffer. 300 μL of the dialysed sample at an OD of 1.050, and a 300 μL reference sample from the AUC buffer used in dialysis, were loaded into a conventional double sector filled Epon centerpiece (path length 1.2 cm) with quartz windows, and mounted in a Beckman An60 titanium rotor. The SV-AUC experiment was carried out in a Beckman XL-I analytical ultracentrifuge at 4°C with a rotor speed of 27,000 rpm, and the data were collected in continuous mode for 6 h, at a single wavelength of 280 nm. Multiple scans at different time points were fitted to a continuous size distribution model using the program SEDFIT [210]. The root mean square deviation was 0.007802, and the residuals

were randomly distributed. Solvent densities and viscosities were computed using SEDNTERP (<http://sednterp.unh.edu/>). The partial specific volume of the MRPP1-MRPP2 protein complex (0.743173 mL/g at 4°C) was calculated from its amino acid composition.

7.4.7 pre-(mt)tRNA in vitro transcription

The pre-(mt)tRNA DNA templates (T7-DNA) containing a 5' T7 RNA polymerase binding site (TAATACGACTCACTATAGGG) were generated by PCR with the primers outlined in **Table 7-4**.

pre-(mt)tRNA	Forward primer	Reverse primer
Ile	<u>TAATACGACTCACTATAGGG</u> CCAT TACAATCTCCAGCATTCC	GATTCTCAGGGATGGGTTCG
Tyr	<u>TAATACGACTCACTATAGGG</u> TGGT TTGTAGAGAATAGTCAACGG	AAATCACCTCGGAGCTGGTAA
Gln	<u>TAATACGACTCACTATAGGG</u> ACCA ACATTTTCGGGGTATG	TTCTAGGACTATGAGAATCGAACC
Val	<u>TAATACGACTCACTATAGGG</u> TCGT AACATGGTAAGTGTAAGTGG	GTAAGTAAGGTGGAGTGGGTTTGG
Leu	<u>TAATACGACTCACTATAGGG</u> CCTTA GTATTATACCCACACCCACC	AATGAGGAGTAGGAGGTTGG
Asp	<u>TAATACGACTCACTATAGGG</u> CCAT GGCCTCCATGACTTT	GTCTTGTAAGACCTACTTGCG

Table 7-4. Primers used to amplify T7-DNA templates for in vitro transcription of pre-(mt)tRNAs.

A 100 μ M stock solution of forward and reverse primer were diluted in ultrapure water to a final concentration of 10 μ M. 1.5 μ L of each primer was then added to a mastermix of 0.25 μ L Phusion polymerase (2 U/mL) in 10 μ L 5xPhusion HF Red Top polymerase buffer, 1 μ L 10 mM dNTPs, 34.25 μ L ultrapure RNase-free water and 1 μ L DNA template (40 ng/ μ L) to a final volume of 50 μ L. PCR was performed with 2 min at 95°C for initial denaturation, followed by 40 cycles of 45 sec denaturation at 95°C, 30 sec annealing at 60°C and 30 sec extension at 72°C with a final extension for 5 min at 72°C. The amplified T7-DNA products were checked for correct amplification on a 1% agarose/TAE buffer gel with addition of 1 μ L peqGREEN DNA/RNA in-gel dye (PEQLAB), alongside a

1 kb DNA ladder (New England Biolabs). The agarose gel was electrophoresed for 30 min at 120 V. The T7-DNA products were used subsequently un-purified.

In vitro pre-(mt)tRNA transcription was performed using the Transcript Aid T7 High yield transcription kit (Thermo Scientific). A mastermix of 2 μ L Transcript Aid Enzyme Mix, 4 μ L 5 x Reaction buffer, 8 μ L 100 mM NTPs, 3.5 μ L ultrapure RNase-free water and 2.5 μ L T7-DNA template (1.5 μ g) was mixed to a final volume of 20 μ L and incubated at 37°C for 4 h in a thermocycler (Bio-Rad C1000 Touch). After incubation, 2 μ L DNase1 (1 U/ μ L) was added to the mixture, which was then further incubated in the thermocycler for 15 min at 37°C. A final incubation was performed at 65°C for 10 min after addition of 2 μ L 0.5 M EDTA pH 8.0.

The RNA transcript products were purified using the RNeasy Mini Kit (Qiagen) and eluted in 30 μ L RNase-free water. RNA concentration was determined by measuring the absorbance at 260 nm on the NanoDrop spectrophotometer. Prior to assay use, the obtained pre-(mt)tRNAs were heat-denatured at 85°C for 5 min and annealed on ice for an additional 5 min in the presence of 2mM MgCl₂ and 100 mM NaCl.

7.4.8 Electrophoretic mobility shift assay (EMSA)

EMSA was performed in a 12 μ L reaction mixture with 2 μ M pre-(mt)tRNA^{lle} and 16 μ M of the investigated protein(s) diluted in EMSA buffer. For titration experiment, 1 μ M *in vitro* transcribed pre-(mt)tRNA^{lle} (**Section 7.4.7**) was mixed with protein in a 10:1, 8:1 and 6:1 tRNA-to-protein ratio. The components were incubated on ice for 10 min, mixed with 6 μ L EMSA loading buffer, and immediately subjected to NATIVE-PAGE by applying 10 μ L of the mixture to a non-denaturing 3-8% tris-acetate polyacrylamide gel (Invitrogen) electrophoresed for 2.5 h at 150 V at 4°C in Tris-glycine buffer. SYBRO GOLD (Invitrogen) was diluted 1000-fold in Tris-glycine buffer and used to stain for RNA for 30 min and RNA bands were visualised with the Bio-Rad ChemiDoc Imaging System. Coomassie Brilliant

Blue (Generon) was used to subsequently stain for protein overnight and protein was visualised using the Bio-Rad ChemiDoc Imaging System.

7.4.9 Limited proteolysis

Limited proteolysis was performed in 20 μ L reaction of 10 μ M protein, 16 μ L Limited proteolysis buffer and 2 μ L trypsin protease (Thermo Scientific) in a protein-to-trypsin ratio of 1:1 to 1:1000. The mixture was incubated at room temperature for 30 min before being added with 5 μ L loading buffer, heat denatured for 5 min at 95°C and subjected to SDS-PAGE analysis.

7.4.10 SPOT blot analysis

In a SPOT blot analysis, interactions can be mapped between an intact soluble protein and peptides from a binding-partner immobilised on a solid support. The entire amino acid sequence of a protein is translated into overlapping peptides of defined length, where each spot corresponds to a 15-mer peptide with maximum overlap of 13 amino acids from the peptide before or after (*i.e.* a 2 amino acid shift in register). In this experiment, the sequence for mature MRPP1 was used:

MSSKIPAVTYPKNESTPPSEELDKWKTMTMKSSVQEECVSTISSKDEDPLAATREFIEMWRLLGREVPEHITE
EELKTLMECVSNTAKKKYLKYLTKYKVKARQIKKEMKAAAREEAKNIKLLETTEEDKQKNFLFLRLWDRNMDIAMG
WKGAQAMQFGQPLVFD MAYENYMKRKELQNTVSQ LLESEGWNRRNVDPFHIYFCNLKIDGALHRELVKRYQEKWDKL
LLTSTEKSHVDLFPKDSIIYLTADSPNVM TTFRHDKVYVIGSFVDKSMQPGTSLAKAKRLNLATECLPLDKYLQWEIGNK
NLTL DQMIRILLCLKNNGNWQEALQFVPRKRKHTGFLEISQHSQEFINRLKKA KT

To synthesise the peptides, Fmoc-protected amino acids were iteratively added to a functionalised cellulose membrane resulting in 50–100 nmol of peptide in spots with a diameter of about 8 mm, dispensed in drops of up to 1 μ L [211]. The membrane was synthesised by Dr. Sarah Picaud at the SGC on a derivatised cellulose membrane (Intavis) using a MultiPep SPOT peptide arrayer (Intavis). In the synthesis, some spots were left blank (no peptide coupled) and some were coupled only to His₆ to serve as controls.

The membrane was rehydrated in 100% ethanol and equilibrated into PBST buffer 3 times for 5 min each. The membrane was blocked with 10 mL 5% bovine serum albumin (BSA) in PBST buffer for 8 h at room temperature on a rocking table. The membrane was then washed with PBST buffer twice for 5 min each and once in PBS buffer for 5 min, before being mixed with 10 mL His₆-tagged MRPP2 protein to a concentration of 1 μ M in PBS buffer and left overnight on a rocking table at 4°C. The next day, the membrane was washed with PBST buffer three times for 5 min each to remove unbound protein, and blocked with 10 mL 5% BSA in PBST buffer for 1 h at room temperature. Membrane washing was repeated with PBST buffer three times for 5 min each. The MRPP2 protein was detected with horseradish peroxidase (HRP)-conjugated anti-His antibody in a 1:3000 dilution in 2.5% BSA PBST buffer and incubated for 1 h. After labelling, the membrane was washed with PBST buffer three times for 20 min each, dabbed dry with a towel and developed using an ECL kit (Thermo Scientific). With the kit, equal amounts of peroxide solution and luminol enhancer solution was used, and the solution-covered membrane was left to develop for 1 min. Membrane spots were visualised with an ImageQuant Las-4000 camera (FUJIFILM) on the chemiluminescence setting. Densitometry was performed using E-editor 2.0 software.

7.5 X-ray Crystallography

X-ray crystallography is the study of the 3D molecular structures of proteins. Today, obtaining a protein (crystal) structure requires a number of steps from initial bioinformatics, through protein purification and crystallisation to validation of the final structure (**Figure 7-6**) [154].

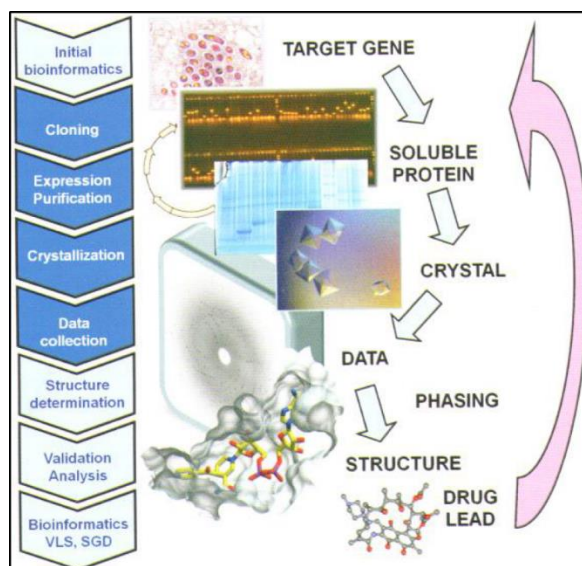


Figure 7-6. Pipeline for solving protein crystal structures from initial bioinformatics to the final structure.
Figure taken from Biomolecular Crystallography, B. Rupp [154].

Some of the biggest obstacles on this path are to produce large amounts of soluble, pure protein, to obtain well-diffracting crystals of this purified protein and to overcome the phase problem to solve the crystal structure [154].

7.5.1 Protein crystallisation

To produce a protein that can be crystallised, many constructs are designed based on bioinformatics analysis. A crystallisable construct is one that folds into a protein structure that can self-assemble into a well-formed crystal lattice. Once a pure, soluble, stable protein is obtained at large quantities, crystallisation experiments can be performed. The crystallisation set-up is designed to sample as much of the crystallisation space as possible using as little protein as possible. Crystals are grown by the vapour diffusion technique with either hanging- or sitting-drop. In this technique a plate is used consisting of a well for a precipitant solution (called the reservoir solution) and a subwell for the protein mixed with precipitant solution (called the drop) (**Figure 7-7**) [154].

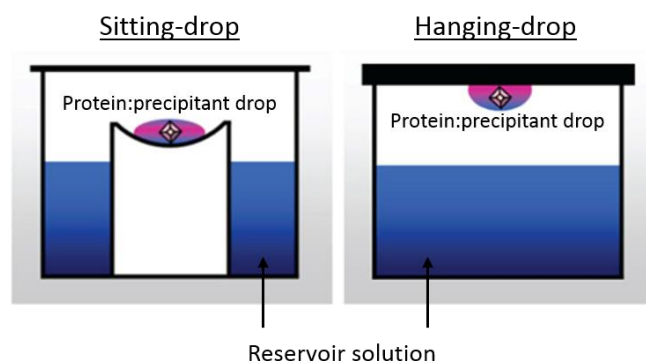


Figure 7-7. Schematic outline of the sitting- and hanging-drop vapour diffusion techniques. Figure adapted from (<http://crystal.csiro.au/User-Guide/Soluble-Proteins/Vapour-Diffusion.aspx>).

The precipitant solution contains various chemicals such as salts, polymers and buffers. Crystallisation screens with various precipitant solutions are commercially available as random matrix screens. After dispensing the protein-and-precipitant drop into the subwell and the reservoir solution into the main well, the plate is sealed and left to equilibrate through vapour diffusion. During this process, water evaporates from the less concentrated drop into the reservoir solution causing an up-concentration of both precipitant and protein in the drop. At some point, the proteins solubility limit is exceeded and the solution becomes supersaturated, forcing the protein out of solution as phase separation, nucleation or regular protein precipitation. The precipitant solution in the drop, in which a protein is grown, is called the mother liquor. It is important to optimise the concentration of ions in the mother liquor according to the concentration of identical ions in the reservoir solution, to ensure that the diffusion direction is towards the reservoir rather than the drop. If protein buffers with e.g. high NaCl concentrations are used, the direction of diffusion can be cotroled by spiking the reservoir solution with the same salt until $[\text{NaCl}]_{\text{reservoir}} > [\text{NaCl}]_{\text{drop}}$, prior to setting up crystallisation plates.

The process of crystallisation can be described with a phase diagram (**Figure 7-8**), where protein concentration is plotted on the y-axis and precipitant concentration on the x-axis. Origin (0,0) corresponds to pure water. The single phase region corresponds to

soluble protein, the metastable phase region corresponds to supersaturated protein solution and the two-phase region corresponds to soluble and saturated protein in thermodynamic equilibrium. An initial mixture of protein and precipitant, in certain concentrations of the sitting- or hanging-drop, represents a point in the phase diagram and as water evaporates this point moves up the diagram towards higher concentrations of both components. If the point was initially placed in the single phase region and slowly moves into the metastable phase region, chances are that the protein will nucleate and form crystals (or precipitate) to reach equilibrium [154].

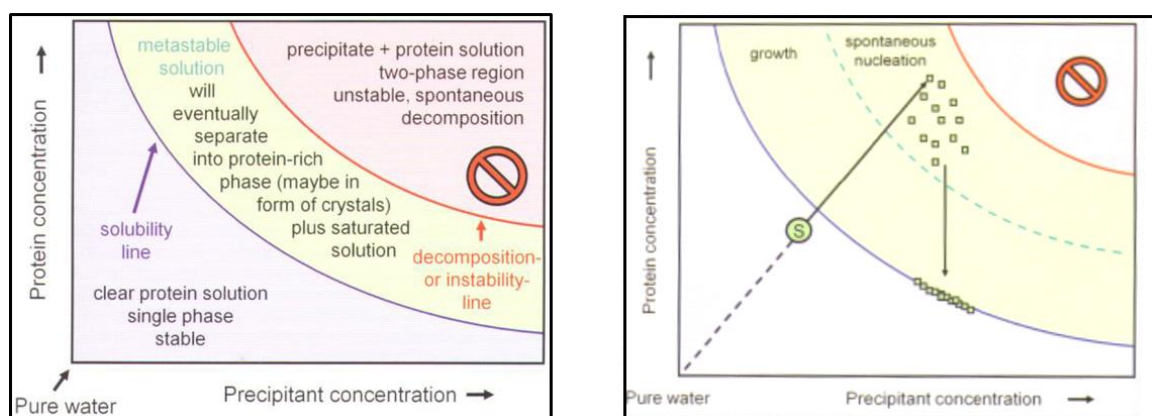


Figure 7-8. The phase diagram with protein concentration plotted against precipitant concentration. Origin, (0,0), corresponds to pure water. The single phase region (purple) corresponds to soluble protein, the metastable phase region (green) corresponds to super-saturated protein solution and the two-phase region (red) corresponds to soluble and saturated protein in thermodynamic equilibrium. Figures taken from Biomolecular Crystallography, B. Rupp [154].

7.5.2 Initial crystallisation trials

Sitting drop vapour-diffusion crystallisation was performed in 96-well, three-subwell SWISSCI plates (SWISSCI AG), with random matrix screens covering 96 precipitant solutions. 20 μ L of the precipitant solution was pipetted into each plate well to be equilibrated against 3 x 150 nL drops in the three subwells. The subwell drops were set up by dispensing protein-and-precipitant in ratios of 2:1, 1:1 and 1:2 using the TTP Labtech's mosquito Crystal liquid handling robot. The crystallisation plates were incubated at 4°C or 20°C, and imaged at set intervals over a period of two months using an automatic inspection system (Rigaku).

Each crystallisation experiment was set up varying one or more of the following parameters: precipitant, protein concentration, temperature, buffer contents, protein-to-precipitant ratio, purification protocol or addition of ligands/substrates.

The random matrix screens used in the crystallisation experiments were commercially available or produced in-house, and based on conditions known to facilitate protein crystallisation:

- Hampton Research Crystal screen (HCS)
- Hampton Research Index (HIN) screen
- Hampton Research SaltRX screen
- JCSG screen [212]
- Ligand friendly screen (LFS) [212] (equivalent to the Qiagen PACT screen)
- Basic ChemSpace (BCS) screen [213]
- Molecular dimensions modern intelligent dynamic alternative screen (MIDAS) [214]

The HCS and HIN screens were designed by Hampton Research as coarse screens for protein crystallisation surveying pH, salts, organics and polymers.

The SaltRx screen was likewise developed by Hampton Research and is based on salts and pH variations with the salts being the only primary precipitant reagents.

The JCSG screen was developed jointly by the Joint Center for Structural Genomics (JCSG) and the SGC and designed to extend the conditions from commercially available screens to maximise the coverage and minimise redundancy of the chemical environment surveyed in each crystallisation experiment [212].

The LFS screen is equivalent to the commercially available PACT screen that aims to systematically screen each component in the crystallisation condition in an uncoupled fashion that allows information to be gained about the protein even in the absence of crystals [212].

The BCS screen designed in-house samples the widely used polyethylene glycol (PEG) precipitants through the use of PEG smears. PEG smears are mixtures of different PEGs with a requirement of either neutral or cooperatively positive effects of each component on crystal growth. The PEG smears allow a wide coverage of properties of these polymers while reducing the PEG variability resulting in greater sampling of other parameters such as buffers and additives [213].

The MIDAS screen was developed by Molecular Dimensions to complement already existing crystallisation screens by replacing widely used inorganic salts and PEGs as precipitants with alternative organic polymers, thus reducing redundancy in the screened crystallisation parameters [214].

7.5.3 Follow-up crystallisation trials

Follow-up crystallisation screens were designed around conditions in which crystals or crystalline precipitation was observed, by varying different components of the original condition as shown in **Figure 7-1**.

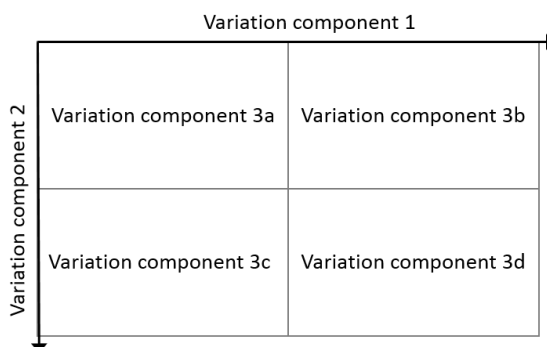


Figure 7-9. The follow-up crystallisation screen design when surveying the chemical environment around an initial crystallisation hit.

Follow-up screens were prepared using ready-made stock solutions (Hampton Research/Molecular Dimensions) dispensed using the Multiprobe II plus liquid handler (PerkinElmer) to make up 0.5 mL solution per solution.

7.5.3.1 *Crystallisation with additives*

Small molecules, in addition to precipitant, buffer or salt, can be introduced into the crystallisation conditions to stabilise the protein, avoid partial denaturation, heterogeneous aggregation or conformational flexibility resulting in a more homogenous sample [181]. Such additives can lower the entropic barrier for crystallisation and thus serve to increase the solubility of the protein, which in turn affects the phase diagram. They can also reduce the number of nucleation points leading to larger and better diffracting crystals [215].

A follow-up screen was designed covering 96 equal conditions similar to previous conditions that generated crystals. This screen was dispensed into the 96 wells (17 μ L) each supplemented with a different additive solution (3 μ L) (Hampton Research).

7.5.3.2 *Crystallisation with micro-seeding*

Crystal nucleation and crystal growth can be described as two separate processes that require different levels of protein supersaturation for optimal progression (**Figure 7-8**). Low levels of supersaturation promotes slow growth, thus avoiding incorporation of crystal defects [154]. To separate the process of nucleation from growth, the micro-seeding technique can be deployed. Already grown crystals are crushed to form micro-seeds

(microscopic crystal nuclei) that are transferred to conditions under which spontaneous nucleation is otherwise absent or rare (e.g. conditions that produces clear drops). The slower crystal growth, around fewer nucleation points, can lead to larger and better diffracting crystals [215], [216].

To produce the micro-seed solution, a drop containing crystals was mixed with 200 μ L of its own precipitant solution and vortexed for 5 min. Using conditions found to yield clear drops, crystallisation trials were performed with drops containing 140 nL protein-and-precipitant and 10 nL of micro-seed solution.

7.5.4 Crystal mounting

Once a suitable crystal is obtained, the crystal is harvested, cryo-protected and flash-cooled. Harvesting consists of fishing out the crystal from its mother liquor using a nylon fiber loop. The cryo-protection and flash-cooling is performed to minimise X-ray radiation damage to the crystal and to avoid formation of crystalline ice on the protein crystal. The cryo-protectant (e.g. large MW PEGs, cryosalts, glycerol or sucrose) [154] is supplemented to the mother liquor and applied to the drop before crystals are flash cooled in liquid nitrogen.

In this study mounting was done by supplementing 1.5 μ L of the reservoir solution with 0.5 μ L 100% cryo-protectant. 1 μ L of this solution was added to the crystal drop and from this mixture single crystals were fished out and flash-cooled in liquid nitrogen in which they were also stored.

7.5.5 Synchrotron data collection

Cryo-protected and flash-cooled crystals are mounted on the goniometer of a diffractometer in a cryo-stream. The goniometer allows rotation of the crystal to attempt sampling of the entire asymmetric unit (ASU) in reciprocal space, and thus ensure collection of a complete dataset. A dataset is collected by exposing the crystal to X-rays during crystal

rotation and recording the reflections on an area detector. The degree of crystal rotation necessary to obtain a complete dataset depends primarily on the Laue group (crystal point group plus inversion). The recorded diffraction images (or frames) are indexed, integrated and scaled and the unit cell and space group are determined [154]. In this process, the resolution cut-off can be estimated from the signal-to-noise ratio ($I/\sigma(I)$) of merged intensity measurements, the correlation coefficient ($CC_{1/2}$) which describes the agreement of intensity measurements between two randomly determined halves of a dataset [217], and – more old-fashioned and impersice – by the R_{merge} , which describes the mean error of an intensity measurement, compared to the mean intensity, averaged over the chosen group of reflections [218]. R_{merge} is determined during merging of data from all intensity measurements, and when a reflection is measured many times (multiplicity), each individual observation has a larger error than the final mean intensity, causing the R_{merge} to increase. Instead of R_{merge} , a multiplicity-independent R-value, R_{meas} , that looks at individual measurements, can be used to evaluate the quality of the data, though $CC_{1/2}$ should be the preferred measure of resolution cut-off [217].

Diffraction data for this thesis were collected at the Diamond Light Source on beamlines IO4, IO4-1 and I24. The collected frames were indexed and integrated using XDS [219] ($P4_3$ space group) or iMOSFLM [220] ($P3_1$ space group) and scaled using AIMLESS [153]. The distribution of the presumed number of molecules in the *ASU* was determined using Matthews's coefficient in the CCP4 suite of programs [221].

7.5.6 Phasing

At this point only the measured intensities of the diffraction spots (or reflections) are obtained, providing only the amplitude for the diffracted X-ray. To reconstruct the electron density, the phases for each reflection are also needed. The lack of phases in the diffraction pattern renders it impossible to directly reconstruct the protein electron density via Fourier

transforms from the amplitude data. This is known as the phase problem. The two main methods deployed to obtain the missing phases in macromolecular crystallography are 1) molecular replacement in which a previously obtained structure of high sequence identity to the target protein is used to calculate phases, that are then applied in the initial reconstruction of the electron density, and 2) experimental phasing that exploits isomorphous differences between native and derivative datasets or anomalous differences from anomalous scatters within a data sets. These intensity differences can be introduced by e.g. heavy atom (HA) soaking of an already obtained crystal or by incorporating selenomethionine (SeMet) into the protein by expression in SetMet-rich media. The HA/SeMet substructure provides the initial phases used to reconstruct the electron density.

In this study, HA-soaking was used. Soaking was performed by mixing 1.5 μL reservoir solution with 1 μL 10mM $\text{C}_9\text{H}_9\text{HgNaO}_2\text{S}$. 1 μL of this solution was added to the crystal drop, and single crystals were picked at various time points from 1 sec to 1 h and transferred to a drop of reservoir solution supplemented with 0.5 μL 100% propylene glycol for washing and cryo-protection. From this drop, the crystals were mounted and flash cooled in liquid nitrogen. Soaking was also used in attempt to obtain a ligand bound structure. The soaking of these ligands was done as described and applied to the drop in 2-20 mM concentrations.

A variety of methods to handle the experimental phasing are available and together with density modifications techniques, a protein structure might be solved from a single anomalous dataset [154].

In this study, phases were determined by either single-wavelength anomalous dispersion (SAD) with the SHELX package [222] using the HA dataset ($P4_3$ space group) or by molecular replacement ($P3_1$ space group) done in PHASER [176] with the search model generated from SAD phasing.

7.5.7 Model building and refinement

With the structure factor amplitudes, F_{hkl} , and phase angle, α_{hkl} , determined, the electron density can be obtained by Fourier reconstruction from reciprocal diffraction space back into direct space [154]. From the reconstructed electron density, a protein structure can be built computationally or manually, and the resulting structure is refined with the use of stereochemical and geometrical restraints. The quality of refinement is evaluated by the R-factor, R , given as the sum of differences between observed and calculated structure amplitudes divided by the sum of observed structure amplitudes (**Equation 3**) [218].

$$\text{Equation 3: } R = \frac{\sum_h ||F_{obs}| - |F_{calc}||}{\sum_h |F_{obs}|}$$

Other evaluation statistics are available, such as the unbiased R_{free} factor (**Equation 4**). In this factor, a set of reflections, chosen at random, are excluded from model refinement, and a later comparison of these data with the final model gives an estimate of refinement quality [218].

$$\text{Equation 4: } R_{free} = \frac{\sum_{test\ set} ||F_{obs}| - |F_{calc}||}{\sum_{test\ set} |F_{obs}|}$$

In this study, the initial model from SAD phasing was built using the auto tracing module of SHELXE [222]. Refinement and modification to the models was performed with alternate rounds of REFMAC5 [223] and COOT [151]. The final round of refinement was done in PHENIX [224].

The refined structures were validated with MolProbity [225] and the atomic coordinate files deposited in the Protein Data Bank. Molecular graphics figures were generated using PyMOL [226].

The protein structure analysis was undertaken using CryCo (v.5.0) to determine the crystal packing interfaces and LigPlot [201] for identification of hydrogen bonds and hydrophobic interactions.

7.6 HPLC-Small Angle X-ray Scattering (HPLC-SAXS)

Macromolecules in solution scatter X-rays in a pattern that is circularly symmetric around the direction of the incident beam read by a charge-coupled detector (**Figure 7-10**).

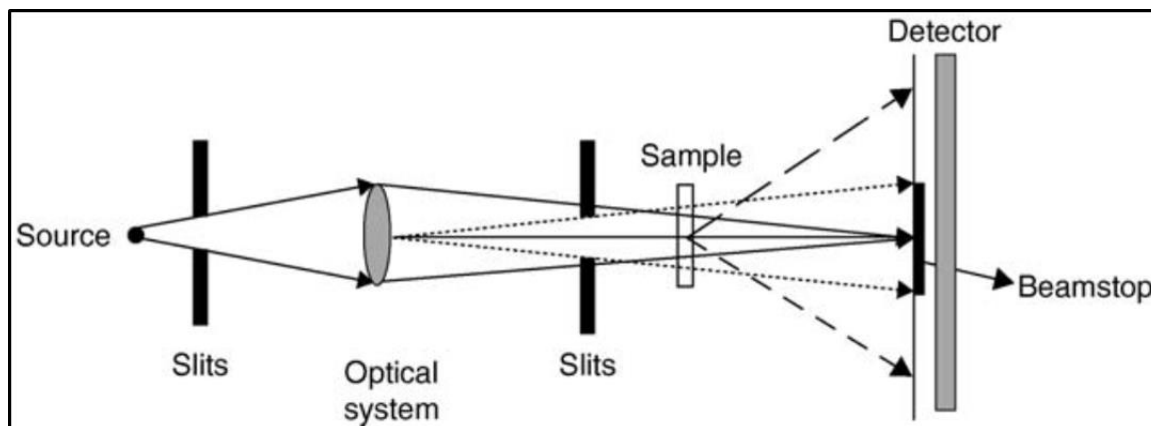


Figure 7-10. Outline of a SAXS setup. Figure adapted from Soft Matter Characterization, R. Das and R. Doniach, Chapter 21[227].

The resulting scattering profiles are plotted as the intensity of scattered photons, $I(q)$, vs the scattering wave vector, q . Sometimes the intensity is also plotted as a function of s which is defined as the wave vector divided by 2π [227].

In an SAXS experiment, the scattering from the protein buffer needs to be subtracted from the obtained protein SAXS data. Once this subtraction is done, the Fourier transform relates the intensity of scattered photons to the electron density of the scattering molecules, upon assuming that all molecules in the sample are oriented and positioned randomly. This Fourier transform is known as the Debye formula [227].

When initially analysing the SAXS data in a model-independent fashion, basic parameters can be determined including the radius of gyration, R_g , which is a measure of the protein radius relative to its centre-of-mass. Measurement of R_g depends only on data at the lowest scattering angles and an accurate measure require protein concentrations where no interparticle interaction or protein aggregation is observed [227].

An evaluation of the protein shape is the Kratky plot derived as the squared scattering wave vector, q^2 , times $I(q)$ plotted against q (**Figure 7-11**). This plot allows qualitative differentiation between globular and disordered particles [228], [229].

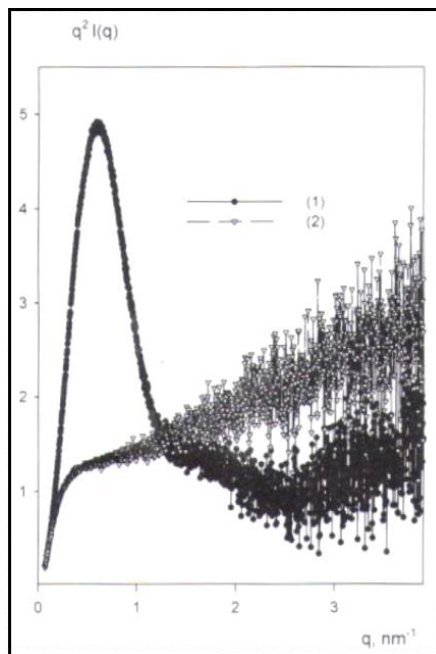


Figure 7-11. Kratky plot for a folded (1) and disordered (2) protein. Figure adapted from Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules, D. Svergun, M. Koch, P. Timmins and R. May [229].

The scattering intensity of non-interacting particles in solution can further be described as an integral that is limited by the maximal dimension (maximal cord length) in the particle, D_{max} , as is dependent on the electron-density-weighted intramolecular distances [229] and distribution of inter-atomic distances, or $P(r)$ distribution, which is a function in real space and thus macromolecule features can more often be recognised in this function rather than the scattering curve [229].

After evaluation of the SAXS data, an *ab initio* model can be reconstructed using available SAXS modelling software.

Individual proteins, protein complexes and protein-pre(mt)tRNA complexes were subjected to HPLC-SAXS on a KW403-4F column at 4.5 mg/mL in 50 mM sodium phosphate pH 7.4, 300 mM NaCl, 5% glycerol, 1% sucrose, 0.5 mM TCEP

(for MRPP1, MRPP2 and MRPP1-MRPP2) or 100 mM NaCl, 50 mM Tris-HCl pH 8, 3 mM MgCl₂, 3 mM DTT (for MRPP1-MRPP2-pre-(mt)tRNA^{lle} with and without MRPP3), by flowing sample through an in-vacuum quartz capillary of 1.6mm diameter at 0.16 mL/min. Data were collected using a Pilatus2M detector (Dectris, CH) with a sample-detector distance of 3914 mm and a wavelength of $\lambda=1$ Å. The range of momentum transfer of $0.1 < s < 5 \text{ nm}^{-1}$ was covered. The data covering the peak were radially averaged and the scattering of buffer was subtracted using Scatter(v.3.0) (Robert Rambo, <http://www.bioisis.net>). The forward scattering I(0), real and reciprocal R_g and Perod Volume were obtained using Scatter(v. 3.0), in which the P(r) distribution function and the D_{max} were also analysed, before *ab initio* modelling was performed with the ATSAS suite of programs. In the *ab initio* modelling, 13 initial models were built using DAMMIF in fast computation mode with P1, P2 or P4 point symmetry. The NSD values were calculated for each generated model using SUPCOMB and subsequently compared to find the most probable ones using DAMSEL, with exclusion of models with NSD > mean + 2·SD. The most probable models were aligned using DAMSUP, averaged using DAMAVER and filtered according to volume by DAMFILT [230]. From the filtered model, a core structure was generated in DAMSTART from which a refined model was built in slow computational mode using DAMMIN, with P1, P2 or P4 point symmetry [231]. Protein MWs were determined directly from the scattering curve using SAXS MoW2, with the assumption that the proteins adopt a globular- rather than a plate-like structure [161].

7.7 Computational studies

7.7.1 Ligand docking

Ligand coordinates were downloaded from the PDBChem Ligand library online. A ligand file in .pdbqt format was created in AutoDockTools-1.5.6 with no changes to the default allowed bond rotations. Docking of ligand to an apo protein structure was performed

in AutoDockVina [232] with the area for docking chosen around the presumed ligand binding pocket. The docking algorithm outputted a list of nine models with various ligand conformations scored based on predicted binding affinity in kcal/mol.

7.7.2 Intrinsic disorder profiling

Intrinsic disorder profiles for MRPP1 and MRPP3 were predicted by DISOPRED3. Putative protein binding regions were calculated from an initial PSI-BLAST search over a filtered sequence database. Each residue was then encoded by the profile for a window of 15 positions in the sequence and classified using a neural network. The classifier was trained using a support vector machine learning algorithm and outputted a probability estimate of the residues being disordered [145].

7.8 Protein activity characterisation

7.8.1 Buffers

MTase activity buffer: 50 mM Tris-HCl pH 8, 5 mM MgCl₂, 20 mM NaCl, 1 mM DTT

SDR activity buffer: 100 mM NaCl, 20 mM Tris-HCl pH 8.0

RNase P activity buffer: 30 mM Tris-HCl pH 8, 40 mM NaCl, 4.5 mM MgCl₂, 2 mM DTT

Tris-Borate-EDTA (TBE)-urea sample buffer: 89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.0, 12% ficoll, 0.01% bromophenol blue, 0.02% xylene cyanole FF, 7 M urea.

TBE buffer: 89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.0.

7.8.2 MTase activity assay

MTase activity assay was performed by Dr. Martine Roovers at the Institute Microbiologique Wiame. The assay monitors the amount of ³H that is transferred from the [methyl-³H]SAM methyl donor to the *in vitro* transcribed pre-(mt)tRNA^{Val/Ile}. 1 μCi [methyl-³H]SAM (15 Ci/mmol) (PerkinElmer), 1 μM individual or complexed protein(s) and 1 μg *in vitro* transcribed pre-(mt)tRNAs (**Section 7.4.7**) were diluted in MTase activity

buffer to 200 μL . The solution was incubated for 60 min at 37°C in a heat block after where the reaction was stopped by phenol extraction and the pre-(mt)tRNA substrates were TCA-precipitated. Radioactive methylated pre-(mt)tRNA was captured on a filter and washed three times with ethanol prior to the measurement of the radioactivity in a scintillation counter.

7.8.3 SDR activity assay

100 μM NAD^+ and substrate (allopregnanolone or DL-3-hydroxybutyryl-CoA) of various concentrations were mixed in SDR activity buffer to a final volume of 50 μL , and dispensed to black 384-well, polystyrene, non-binding surface plates (Corning). 3.7 μM protein was added and the fluorescence from NAD^+ conversion to NADH was continuously measured at 350 nm excitation/450 nm emission in a PHERAstar FSX microplate reader (BMG LABTECH) for 30 min. The substrates 3-hydroxybutyryl-CoA and allopregnanolone (Sigma) were assayed at final concentrations ranging from 100 to 0.05 μM . The obtained fluorescence data were analysed with the MARS software (BMG LABTECH) and Excel (Microsoft) program. Kinetic parameters (K_M , $V_{\max}$) were obtained by fitting the data to the Michaelis-Menten equation (**Equation 5**) where v is the reaction rate, K_M is the Michaelis constant, $[S]$ is the substrate concentration and $V_{\max}$ is the maximum rate [174]. The fit was done using the GraphPad Prism (v.5.01) software.

$$\text{Equation 5: } v = \frac{V_{\max} \cdot [S]}{K_M + [S]}$$

The maximum number of substrate molecules that an enzyme molecule can convert into products per time measurement, k_{cat} , was calculated for four active sites indicated per MRPP2 tetrameric protein from **Equation 6**, where $[E]_0$ is the total enzyme concentration [174].

$$\text{Equation 6: } k_{\text{cat}} = \frac{V_{\max}}{[E]_0}$$

7.8.4 RNase P activity assay

RNase P cleavage was performed by mixing 300 nM MRPP1-MRPP2, 150 nM MRPP3, 10 units of ribonuclease inhibitors (RNasin from Promega) and 400 nM *in vitro* transcribed pre-(mt)tRNA^{Ile} (**Section 7.4.7**) in RNase P activity buffer to a total reaction volume of 8.25 μ L. The reaction was performed at room temperature and stopped at set times by transferring 1 μ L of the reaction mixture into 5 μ L 500 mM EDTA and heating to 95°C. This sample was added with 10 μ L TBE-urea sample buffer (Invitrogen), boiled for 6 min at 70°C and spun for 1 min at maximum speed in a tube-strip picofuge (Stratagene). 10 μ L of this sample was analysed by denaturing-PAGE on a 6% TBE-urea gel (Invitrogen) and run at room temperature for 1 h at 180 V in TBE buffer (Invitrogen). The gel was stained for 30 min with SYBRO GOLD (Invitrogen) diluted 1000-fold in TBE buffer from the manufactures stock, and imaged using the Bio-Rad ChemiDoc Imaging System. The bands corresponding to processed (mt)tRNA were quantified using ImageJ [233] and plotted as a fraction of the amount of unprocessed pre-(mt)tRNA^{Ile} by non-linear regression fitting with one phase exponential association (**Equation 7**) in Graph Pad prism, where Y_{max} is the maximum amount of processed pre-(mt)tRNA and K is the initial rate of reaction.

$$\text{Equation 7: } y = Y_{max} \cdot (1 - e^{-K \cdot x})$$

BIBLIOGRAPHY

- [1] H. Jörnvall *et al.*, “Short-chain dehydrogenases/reductases (SDR),” *Biochemistry*, vol. 34, no. 18, pp. 6003–6013, 1995.
- [2] Y. Kallberg, U. Oppermann, H. Jörnvall, and B. Persson, “Short-chain dehydrogenase/reductase (SDR) relationships: A large family with eight clusters common to human, animal, and plant genomes,” *Protein Sci.*, vol. 11, pp. 636–641, 2002.
- [3] K. L. Kavanagh, H. Jörnvall, B. Persson, and U. Oppermann, “Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes,” *Cell. Mol. Life Sci.*, vol. 65, no. 24, pp. 3895–906, Dec. 2008.
- [4] M. J. Luo, L.-F. Mao, and H. Schulz, “Short-chain 3-Hydroxy-2-methylacyl-CoA Dehydrogenase from Rat Liver: Purification and Characterization of a Novel Enzyme of Isoleucine Metabolism,” *Biochem. Biophys.*, vol. 321, no. 1, pp. 214–220, 1995.
- [5] X. He, H. Schulz, and S.-Y. Yang, “A human brain l-3-hydroxyacyl-coenzyme A dehydrogenase is identical to an amyloid β -peptide-binding protein involved in Alzheimer’s disease,” *J. Biol. Chem.*, vol. 273, pp. 10741–10746, 1998.
- [6] X. Y. He, G. Merz, Y. Z. Yang, P. Mehta, H. Schulz, and S. Y. Yang, “Characterization and localization of human type10 17 β -hydroxysteroid dehydrogenase,” *Eur. J. Biochem.*, vol. 268, no. 18, pp. 4899–907, Sep. 2001.
- [7] N. Shafqat *et al.*, “Expanded substrate screenings of human and *Drosophila* type 10 17 β -hydroxysteroid dehydrogenases (HSDs) reveal multiple specificities in bile acid and steroid hormone metabolism: characterization of multifunctional 3 α /7 α /7 β /17 β /20 β /21-HSD,” *Biochem. J.*, vol. 376, no. Pt 1, pp. 49–60, Nov. 2003.
- [8] C. J. Jeffery, “Moonlighting proteins,” *Trends Biochem. Sci.*, vol. 24, no. 1, pp. 8–11,

- Jan. 1999.
- [9] a Kobayashi, L. L. Jiang, and T. Hashimoto, "Two mitochondrial 3-hydroxyacyl-CoA dehydrogenases in bovine liver.," *J. Biochem.*, vol. 119, no. 4, pp. 775–82, 1996.
 - [10] S. Du Yan *et al.*, "An intracellular protein that binds amyloid-beta peptide and mediates neurotoxicity in Alzheimer's disease.," *Nature*, vol. 389, pp. 689–695, 1997.
 - [11] S. D. Yan *et al.*, "Role of ERAB/L-3-hydroxyacyl-coenzyme A dehydrogenase type II activity in Abeta-induced cytotoxicity.," *J. Biol. Chem.*, vol. 274, no. 4, pp. 2145–56, Jan. 1999.
 - [12] R. Ofman *et al.*, "2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency is caused by mutations in the HADH2 gene.," *Am. J. Hum. Genet.*, vol. 72, no. 5, pp. 1300–7, May 2003.
 - [13] J. Holzmann, P. Frank, E. Löffler, K. L. Bennett, C. Gerner, and W. Rossmanith, "RNase P without RNA: Identification and Functional Reconstitution of the Human Mitochondrial tRNA Processing Enzyme," *Cell*, vol. 135, pp. 462–474, Oct. 2008.
 - [14] E. Vilardo, C. Nachbagauer, A. Buzet, A. Taschner, J. Holzmann, and W. Rossmanith, "A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase - extensive moonlighting in mitochondrial tRNA biogenesis," *Nucleic Acids Res.*, vol. 40, no. 22, pp. 11583–93, Dec. 2012.
 - [15] T. O. H. Boynton and L. J. Shimkets, "Myxococcus CsgA , Drosophila Sniffer , and human HSD17B10 are cardiolipin phospholipases," *Genes Dev.*, vol. 29, no. 18, pp. 1903–1914, 2015.
 - [16] X. Y. He *et al.*, "Function of human brain short chain L-3-hydroxyacyl coenzyme A dehydrogenase in androgen metabolism.," *Biochim. Biophys. Acta*, vol. 1484, no. 2–3, pp. 267–77, Apr. 2000.
 - [17] X. He, G. Merz, P. Mehta, H. Schulz, and S. Yang, "Human brain short chain l-3-

- hydroxyacyl coenzyme A dehydrogenase is a single-domain multifunctional enzyme: characterization of a novel 17 β -hydroxysteroid dehydrogenase,” *J. Biol. Chem.*, vol. 274, pp. 15014–15019, 1999.
- [18] X.-Y. He, J. Wegiel, Y.-Z. Yang, R. Pullarkat, H. Schulz, and S.-Y. Yang, “Type 10 17 β -hydroxysteroid dehydrogenase catalyzing the oxidation of steroid modulators of γ -aminobutyric acid type A receptors,” *Mol. Cell. Endocrinol.*, vol. 229, no. 1–2, pp. 111–7, Jan. 2005.
- [19] P. Lukacik, K. L. Kavanagh, and U. Oppermann, “Structure and function of human 17 β -hydroxysteroid dehydrogenases,” *Mol. Cell. Endocrinol.*, vol. 248, no. 1–2, pp. 61–71, Mar. 2006.
- [20] X. He, Y. Yang, H. Schulz, and S. Yang, “Intrinsic alcohol dehydrogenase and hydroxysteroid dehydrogenase activities of human mitochondrial short-chain,” vol. 345, pp. 139–143, 2000.
- [21] L. Wu, M. Einstein, W. M. Geissler, and H. K. Chan, “Expression Cloning and Characterization of Human 17 β -Hydroxysteroid Dehydrogenase Type 2 , a Microsomal Enzyme Possessing 20 α -Hydroxysteroid Dehydrogenase Activity,” *J. Biol. Chem.*, vol. 268, no. 17, pp. 12964–12969, 1993.
- [22] J. Adamski *et al.*, “Molecular cloning of a novel widely expressed human 80 kDa 17 β -hydroxysteroid dehydrogenase IV,” *Biochem. J.*, vol. 311 (Pt 2, pp. 437–443, 1995.
- [23] J. Fomitcheva, M. E. Baker, E. Anderson, G. Y. Lee, and N. Aziz, “Characterization of Ke 6, a new 17 β -hydroxysteroid dehydrogenase, and its expression in gonadal tissues,” *J. Biol. Chem.*, vol. 273, no. 35, pp. 22664–22671, 1998.
- [24] F. Azzouni, A. Godoy, Y. Li, and J. Mohler, “The 5 α -Reductase Isozyme Family: A Review of Basic Biology and Their Role in Human Diseases,” *Adv. Urol.*,

- vol. 2012, pp. 1–18, 2012.
- [25] S.-Y. Yang, X.-Y. He, and H. Schulz, “Multiple functions of type 10 17beta-hydroxysteroid dehydrogenase.,” *Trends Endocrinol. Metab.*, vol. 16, no. 4, pp. 167–75, 2005.
 - [26] H. K. Lin, J. M. Jez, B. P. Schlegel, D. M. Peehl, J. a Pachter, and T. M. Penning, “Expression and characterization of recombinant type 2 3 alpha-hydroxysteroid dehydrogenase (HSD) from human prostate: demonstration of bifunctional 3 alpha/17 beta-HSD activity and cellular distribution.,” *Mol. Endocrinol.*, vol. 11, no. 13, pp. 1971–1984, 1997.
 - [27] J. J. Lambert, D. Belelli, C. Hill-Venning, and J. a. Peters, “Neurosteroids and GABAA receptor function,” *Trends Pharmacol. Sci.*, vol. 16, no. 9, pp. 295–303, 1995.
 - [28] X.-Y. He, J. Wegiel, and S.-Y. Yang, “Intracellular oxidation of allopregnanolone by human brain type 10 17beta-hydroxysteroid dehydrogenase.,” *Brain Res.*, vol. 1040, no. 1–2, pp. 29–35, Apr. 2005.
 - [29] J. W. Trauger, a Jiang, B. a Stearns, and P. V LoGrasso, “Kinetics of allopregnanolone formation catalyzed by human 3 alpha-hydroxysteroid dehydrogenase type III (AKR1C2),” *Biochemistry*, vol. 41, no. 45, pp. 13451–13459, 2002.
 - [30] R. J. a Wanders, J. Komen, and S. Ferdinandusse, “Phytanic acid metabolism in health and disease.,” *Biochim. Biophys. Acta*, vol. 1811, no. 9, pp. 498–507, 2011.
 - [31] S. H. Korman, “2-Ethylhydracrylic Aciduria in Short/Branched-Chain Acyl-CoA Dehydrogenase Deficiency: Application to Diagnosis and Implications for the R-Pathway of Isoleucine Oxidation,” *Clin. Chem.*, vol. 51, no. 3, pp. 610–617, 2005.
 - [32] S. M. Houten, S. Violante, F. V. Ventura, and R. J. A. Wanders, “The Biochemistry

- and Physiology of Mitochondrial Fatty Acid β -Oxidation and Its Genetic Disorders,” *Annu. Rev. Physiol.*, vol. 78, pp. 23–44, 2016.
- [33] H. Schulz, “Beta oxidation of fatty acids,” *Biochim. Biophys. Acta - Lipids Lipid Metab.*, vol. 1081, pp. 109–120, 1991.
- [34] X. Y. He, S. Y. Yang, and H. Schulz, “Assay of L-3-hydroxyacyl-coenzyme A dehydrogenase with substrates of different chain lengths,” *Anal. Biochem.*, vol. 180, no. 1, pp. 105–9, 1989.
- [35] S.-Y. Yang *et al.*, “Mental retardation linked to mutations in the HSD17B10 gene interfering with neurosteroid and isoleucine metabolism,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 35, pp. 14820–4, Sep. 2009.
- [36] J. Zschocke *et al.*, “Progressive infantile neurodegeneration caused by 2-methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency: a novel inborn error of branched-chain fatty acid and isoleucine metabolism,” *Pediatr. Res.*, vol. 48, no. 6, pp. 852–5, Dec. 2000.
- [37] S. Nobel, L. Abrahmsen, and U. Oppermann, “Metabolic conversion as a pre-receptor control mechanism for lipophilic hormones,” *Eur. J. Biochem.*, vol. 268, no. 15, pp. 4113–4125, 2001.
- [38] J. W. Lustbader *et al.*, “ABAD Directly Links A β to Mitochondrial Toxicity in Alzheimer’s Disease,” *Science*, vol. 304, no. 5669, pp. 448–52, Apr. 2004.
- [39] C. R. Kissinger *et al.*, “Crystal Structure of Human ABAD/HSD10 with a Bound Inhibitor: Implications for Design of Alzheimer’s Disease Therapeutics,” *J. Mol. Biol.*, vol. 342, no. 3, pp. 943–52, Sep. 2004.
- [40] U. C. Oppermann, C. Filling, and H. Jörnvall, “Forms and functions of human SDR enzymes,” *Chem. Biol. Interact.*, vol. 130–132, no. 1–3, pp. 699–705, Jan. 2001.
- [41] C. Filling *et al.*, “Critical residues for structure and catalysis in short-chain

- dehydrogenases/reductases.,” *J. Biol. Chem.*, vol. 277, no. 28, pp. 25677–84, Jul. 2002.
- [42] R. Ladenstein, J. O. Winberg, and J. Benach, “Medium- and short-chain dehydrogenase/reductase gene and protein families: Structure-function relationships in short-chain alcohol dehydrogenases,” *Cell. Mol. Life Sci.*, vol. 65, no. 24, pp. 3918–3935, 2008.
- [43] J. Benach, S. Atrian, R. González-Duarte, and R. Ladenstein, “The catalytic reaction and inhibition mechanism of *Drosophila* alcohol dehydrogenase: observation of an enzyme-bound NAD-ketone adduct at 1.4 Å resolution by X-ray crystallography.,” *J. Mol. Biol.*, vol. 289, no. 2, pp. 335–355, 1999.
- [44] C. Florentz, B. Sohm, P. Tryoen-Tóth, J. Pütz, and M. Sissler, “Human mitochondrial tRNAs in health and disease,” *Cell. Mol. Life Sci.*, vol. 60, no. 7, pp. 1356–75, Jul. 2003.
- [45] T. Suzuki, A. Nagao, and T. Suzuki, “Human Mitochondrial tRNAs: Biogenesis, Function, Structural Aspects, and Diseases,” *Annu. Rev. Genet.*, vol. 45, pp. 299–329, Jan. 2011.
- [46] R. Giegé, F. Jühling, J. Pütz, P. Stadler, C. Sauter, and C. Florentz, “Structure of transfer RNAs: similarity and variability,” *WIREs RNA*, vol. 3, no. 1, pp. 37–61, 2012.
- [47] T. Suzuki, A. Nagao, and T. Suzuki, “Human mitochondrial diseases caused by lack of taurine modification in mitochondrial tRNAs,” *WIREs RNA*, vol. 2, no. 3, pp. 376–386, 2011.
- [48] A. Reichert, U. Rothbauer, and M. Mörl, “Processing and Editing of Overlapping tRNAs in Human Mitochondria,” *J. Biol. Chem.*, vol. 273, no. 48, pp. 31977–31984, 1998.
- [49] Z. Shao *et al.*, “Crystal structure of tRNA m1G9 methyltransferase Trm10: insight

- into the catalytic mechanism and recognition of tRNA substrate,” *Nucleic Acids Res.*, vol. 42, no. 1, pp. 509–25, Jan. 2014.
- [50] M. Helm, H. Brulé, D. Friede, R. Giegé, D. Pütz, and C. Florentz, “Search for characteristic structural features of mammalian mitochondrial tRNAs,” *RNA*, vol. 6, no. 10, pp. 1356–1379, 2000.
- [51] Y. Motorin and M. Helm, “RNA nucleotide methylation,” *WIREs RNA*, vol. 2, no. 5, pp. 611–631, 2011.
- [52] A. Czerwoniec, J. M. Kasprzak, K. H. Kaminska, K. Rother, E. Purta, and J. M. Bujnicki, “Folds and Functions of Domains in RNA Modification Enzymes,” in *DNA and RNA Modification Enzymes: Comparative Structure, Mechanism, Functions, Cellular Interactions and Evolution.*, H. Grosjean, Ed. Austin, Texas, USA: Landes Bioscience, 2009, pp. 289–302.
- [53] V. Anantharaman, E. V. Koonin, and L. Aravind, “SPOUT: a Class of Methyltransferases that Includes spoU and trmD RNA Methylase Superfamilies, and Novel Superfamilies of Predicted Prokaryotic RNA methylases,” *J. Mol. Microbiol. Biotechnol.*, vol. 4, no. 1, pp. 71–75, 2002.
- [54] K. L. Tkaczuk, S. Dunin-Horkawicz, E. Purta, and J. M. Bujnicki, “Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases,” *BMC Bioinformatics*, vol. 8, p. 73, 2007.
- [55] H. J. Ahn, H.-W. Kim, H.-J. Yoon, B. Lee II, S. W. Suh, and J. K. Yang, “Crystal structure of tRNA (m1G37) methyltransferase: insights into tRNA recognition,” *Embo J*, vol. 22, no. 11, pp. 2593–2603, 2003.
- [56] T. Christian, G. Lahoud, C. Liu, K. Hoffmann, J. J. Perona, and Y.-M. Hou, “Mechanism of N-methylation by the tRNA m1G37 methyltransferase Trm5,” *RNA*, vol. 16, no. 12, pp. 2484–92, 2010.

- [57] S. Goto-Ito, T. Ito, M. Kuratani, Y. Bessho, and S. Yokoyama, "Tertiary structure checkpoint at anticodon loop modification in tRNA functional maturation.," *Nat. Struct. Mol. Biol.*, vol. 16, no. 10, pp. 1109–1115, 2009.
- [58] P. Barraud *et al.*, "Crystal Structure of *Thermus thermophilus* tRNA m1A58 Methyltransferase and Biophysical Characterization of Its Interaction with tRNA," *J. Mol. Biol.*, vol. 377, no. 2, pp. 535–550, 2008.
- [59] P. A. Elkins *et al.*, "Insights into Catalysis by a Knotted TrmD tRNA mMethyltransferase," *J. Mol. Biol.*, vol. 333, no. 5, pp. 931–949, 2003.
- [60] R. Sakaguchi, G. Lahoud, T. Christian, H. Gamper, and Y. M. Hou, "A Divalent Metal Ion-Dependent N1-Methyl Transfer to G37-tRNA," *Chem. Biol.*, vol. 21, no. 10, pp. 1351–1360, 2014.
- [61] M. Helm *et al.*, "The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA," *Nucleic Acids Res.*, vol. 26, no. 7, pp. 1636–1643, 1998.
- [62] F. Voigts-Hoffmann *et al.*, "A Methyl Group Controls Conformational Equilibrium in Human Mitochondrial tRNA^{Lys}," *J. Am. Chem. Soc.*, vol. 129, no. 44, pp. 13382–13383, 2007.
- [63] M. Helm and G. Attardi, "Nuclear Control of Cloverleaf Structure of Human Mitochondrial tRNA^{Lys}," *J. Mol. Biol.*, vol. 337, no. 3, pp. 545–560, 2004.
- [64] A. Y. Kobitski, M. Hengesbach, M. Helm, and G. U. Nienhaus, "Sculpting an RNA conformational energy landscape by a methyl group modification--a single-molecule FRET study.," *Angew. Chem. Int. Ed. Engl.*, vol. 47, no. 23, pp. 4326–30, 2008.
- [65] M. Helm, R. Giegé, and C. Florentz, "A Watson–Crick Base-Pair-Disrupting Methyl Group (m¹A9) Is Sufficient for Cloverleaf Folding of Human Mitochondrial tRNA^{Lys} †," *Biochemistry*, vol. 38, no. 40, pp. 13338–13346, 1999.

- [66] L. M. Wittenhagen and S. O. Kelley, “Impact of disease-related mitochondrial mutations on tRNA structure and function,” *Trends Biochem. Sci.*, vol. 28, no. 11, pp. 605–11, Nov. 2003.
- [67] B. Sohm, M. Frugier, H. Brulé, K. Olszak, A. Przykorska, and C. Florentz, “Towards understanding human mitochondrial leucine aminoacylation identity,” *J. Mol. Biol.*, vol. 328, no. 5, pp. 995–1010, 2003.
- [68] B. Sohm, M. Sissler, H. Park, M. P. King, and C. Florentz, “Recognition of human mitochondrial tRNA^{Leu}(UUR) by its cognate leucyl-tRNA synthetase,” *J. Mol. Biol.*, vol. 339, no. 1, pp. 17–29, 2004.
- [69] F. Jühling, M. Mörl, R. K. Hartmann, M. Sprinzl, P. F. Stadler, and J. Pütz, “tRNAdb 2009: compilation of tRNA sequences and tRNA genes,” *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D159–D162, Jan. 2009.
- [70] Y. Motorin and M. Helm, “tRNA stabilization by modified nucleotides,” *Biochemistry*, vol. 49, no. 24, pp. 4934–4944, 2010.
- [71] J. Pütz, B. Dupuis, M. Sissler, and C. Florentz, “Mamit-tRNA, a database of mammalian mitochondrial tRNA primary and secondary structures,” *RNA*, vol. 13, no. 8, pp. 1184–90, 2007.
- [72] N. Torabi and L. Kruglyak, “Variants in SUP45 and TRM10 underlie natural variation in translation termination efficiency in *Saccharomyces cerevisiae*,” *PLoS Genet.*, vol. 7, no. 7, p. e1002211, 2011.
- [73] A. Barbas, A. Popescu, C. Frazão, C. M. Arraiano, and A. M. Fialho, “Rossmann-fold motifs can confer multiple functions to metabolic enzymes: RNA binding and ribonuclease activity of a UDP-glucose dehydrogenase,” *Biochem. Biophys. Res. Commun.*, vol. 430, no. 1, pp. 218–224, 2013.
- [74] M. R. Fernández *et al.*, “Human and yeast ζ -crystallins bind AU-rich elements in

- RNA,” *Cell. Mol. Life Sci.*, vol. 64, no. 11, pp. 1419–27, Jun. 2007.
- [75] B. M. Beckmann *et al.*, “The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs,” *Nat. Commun.*, vol. 6, p. 10127, 2015.
- [76] J. W. Taanman, “The mitochondrial genome: structure, transcription, translation and replication,” *Biochim. Biophys. Acta*, vol. 1410, no. 2, pp. 103–123, 1999.
- [77] H. D. Robertson, S. Altman, and D. Smith, “Purification and properties of a specific escherichia coli ribonuclease which cleaves a tyrosine transfer ribonucleic acid precursor,” *J. Biol. Chem.*, vol. 247, no. August, pp. 5243–5251, 1972.
- [78] W. Rossmannith, A. Tullo, T. Potuschak, R. Karwan, and E. Sbisà, “Human mitochondrial tRNA processing,” *J. Biol. Chem.*, vol. 270, no. 21, pp. 12885–12891, 1995.
- [79] D. Walczyk, M. Gößringer, W. Rossmannith, T. S. Zatsepin, T. S. Oretskaya, and R. K. Hartmann, “Analysis of the Cleavage Mechanism by Protein-Only RNase P Using Precursor tRNA Substrates with Modifications at the Cleavage Site,” *J. Mol. Biol.*, vol. 428, no. 24 Pt B, pp. 4917–4928, 2016.
- [80] E. B. Dubrovsky, V. A. Dubrovskaya, L. Levinger, S. Schiffer, and A. Marchfelder, “Drosophila Rnase Z processes mitochondrial and nuclear pre-tRNA ’ ends in vivo,” *Nucleic Acids Res.*, vol. 32, no. 1, pp. 255–262, 2004.
- [81] L. K. Brzezniak, M. Bijata, R. J. Szczesny, and P. P. Stepień, “Involvement of human ELAC2 gene product in 3’ end processing of mitochondrial tRNAs,” *RNA Biol.*, vol. 8, no. January, pp. 616–626, 2011.
- [82] M. I. Lopez Sanchez *et al.*, “RNA processing in human mitochondria,” *Cell Cycle*, vol. 10, no. 17, pp. 2904–2916, 2011.
- [83] W. Rossmannith, “Processing of human mitochondrial tRNA(Ser(AGY))GCU: a novel pathway in tRNA biosynthesis,” *J. Mol. Biol.*, vol. 265, no. 4, pp. 365–71, Jan. 1997.

- [84] J. W. Brown and N. R. Pace, "Ribonuclease P RNA and protein subunits from bacteria.," *Nucleic Acids Res.*, vol. 20, no. 7, pp. 1451–6, 1992.
- [85] S. C. Walker and D. R. Engelke, "A protein-only RNase P in human mitochondria.," *Cell*, vol. 135, no. 3, pp. 412–414, Oct. 2008.
- [86] S. C. Walker and D. R. Engelke, "Ribonuclease P: the evolution of an ancient RNA enzyme.," *Crit. Rev. Biochem. Mol. Biol.*, vol. 41, no. 2, pp. 77–102, 2006.
- [87] T. a Hall and J. W. Brown, "Archaeal RNase P has multiple protein subunits homologous to eukaryotic nuclear RNase P proteins.," *RNA*, vol. 8, no. April 2002, pp. 296–306, 2002.
- [88] J. R. Chamberlain, Y. Lee, W. S. Lane, and D. R. Engelke, "Purification and characterization of the nuclear RNaseP holoenzyme complex reveals extensive subunit overlap with RNase MRP.," *Genes Dev.*, vol. 12, pp. 1678–1690, 1998.
- [89] M. J. Wang, N. W. Davis, and P. Gegenheimer, "Novel mechanisms for maturation of chloroplast transfer RNA precursors.," *EMBO J.*, vol. 7, no. 6, pp. 1567–1574, 1988.
- [90] A. Gobert *et al.*, "A single Arabidopsis organellar protein has RNase P activity.," *Nat. Struct. Mol. Biol.*, vol. 17, no. 6, pp. 740–4, Jun. 2010.
- [91] E. Vilardo and W. Rossmanith, "Roles and interplay of the subunits of human mitochondrial RNase P.," in *RNA2016 Conference*, 2016.
- [92] W. H. Lim, "Importance of substrate recognition and metal ions in the Ribonuclease P catalysis.," The University of Michigan, 2011.
- [93] M. J. Howard, W. H. Lim, C. A. Fierke, and M. Koutmos, "Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 40, pp. 16149–54, Oct. 2012.
- [94] W. Rossmanith, "Of P and Z: mitochondrial tRNA processing enzymes.," *Biochim.*

- Biophys. Acta*, vol. 1819, no. 9–10, pp. 1017–26, 2012.
- [95] R. Salavati, A. K. Panigrahi, and K. D. Stuart, “Mitochondrial ribonuclease P activity of *Trypanosoma brucei*,” *Mol. Biochem. Parasitol.*, vol. 115, no. 1, pp. 109–17, Jun. 2001.
 - [96] A. Taschner, C. Weber, A. Buzet, R. K. Hartmann, A. Hartig, and W. Rossmannith, “Nuclear RNase P of *Trypanosoma brucei*: A Single Protein in Place of the Multicomponent RNA-Protein Complex,” *Cell Rep.*, vol. 2, no. 1, pp. 19–25, Jul. 2012.
 - [97] A. Sen, A. Karasik, A. Shanmuganathan, E. Mirkovic, M. Koutmos, and R. T. Cox, “Loss of the mitochondrial protein-only ribonuclease P complex causes aberrant tRNA processing and lethality in *Drosophila*,” *Nucleic Acids Res.*, vol. 44, no. 13, pp. 6409–22, 2016.
 - [98] B. Gutmann, A. Gobert, and P. Giegé, “PRORP proteins support RNase P activity in both organelles and the nucleus in *Arabidopsis*,” *Genes Dev.*, vol. 26, no. 10, pp. 1022–7, May 2012.
 - [99] J. Forner, B. Weber, S. Thuss, S. Wildum, and S. Binder, “Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: T-elements contribute to 5′ and 3′ end formation,” *Nucleic Acids Res.*, vol. 35, no. 11, pp. 3676–3692, 2007.
 - [100] V. Anantharaman and L. Aravind, “The NYN Domains Novel Predicted RNAses with a PIN Domain-Like Fold,” *RNA Biol.*, vol. 3, no. 1, pp. 18–27, 2006.
 - [101] P. Yin *et al.*, “Structural basis for the modular recognition of single-stranded RNA by PPR proteins,” *Nature*, vol. 504, no. 7478, pp. 168–71, Dec. 2013.
 - [102] L. F. Landweber, “The PPR motif – a TPR-related motif prevalent in plant organellar proteins,” *Protein Seq. Motifs*, vol. February, no. 25, pp. 45–47, 2000.
 - [103] F. Pinker *et al.*, “PPR proteins shed a new light on RNase P biology,” *RNA Biol.*, vol.

- 10, no. 9, pp. 1457–1468, Jun. 2013.
- [104] S. Manna, “An overview of pentatricopeptide repeat proteins and their applications,” *Biochimie*, vol. 113, pp. 93–99, 2015.
- [105] S. Aubourg, N. Boudet, M. Kreis, and A. Lecharny, “In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants,” *Plant Mol. Biol.*, vol. 42, no. 4, pp. 603–613, 2000.
- [106] K. Schwinghammer, A. C. M. Cheung, Y. I. Morozov, K. Agaronyan, D. Temiakov, and P. Cramer, “Structure of human mitochondrial RNA polymerase elongation complex,” *Nat. Struct. Mol. Biol.*, vol. 20, no. 11, pp. 1298–303, 2013.
- [107] C. Lurin *et al.*, “Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis,” *Plant Cell*, vol. 16, no. 8, pp. 2089–2103, 2004.
- [108] E. Delannoy, W. a Stanley, C. S. Bond, and I. D. Small, “Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles,” *Biochem. Soc. Trans.*, vol. 35, no. Pt 6, pp. 1643–1647, 2007.
- [109] C. Schmitz-Linneweber and I. Small, “Pentatricopeptide repeat proteins: a socket set for organelle gene expression,” *Trends Plant Sci.*, vol. 13, no. 12, pp. 663–670, 2008.
- [110] B. Albert, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of THE CELL*, 5th ed. New York: Science, Garland, 2008.
- [111] M. R. Duchon and G. Szabadkai, “Roles of mitochondria in human disease,” *Essays Biochem.*, vol. 47, pp. 115–37, 2010.
- [112] J. Zschocke, “HSD10 disease: clinical consequences of mutations in the HSD17B10 gene,” *J. Inherit. Metab. Dis.*, vol. 35, no. 1, pp. 81–9, Jan. 2012.
- [113] J. García-Villoria *et al.*, “X-inactivation of HSD17B10 revealed by cDNA analysis in two female patients with 17 β -hydroxysteroid dehydrogenase 10 deficiency,” *Eur. J.*

- Hum. Genet.*, vol. 18, no. 12, pp. 1353–5, Dec. 2010.
- [114] L. Torroja, D. Ortuño-Sahagún, A. Ferrús, B. Hämmerle, and J. A. Barbas, “scully, an essential gene of *Drosophila*, is homologous to mammalian mitochondrial type II L-3-hydroxyacyl-CoA dehydrogenase/amyloid-beta peptide-binding protein.,” *J. Cell Biol.*, vol. 141, no. 4, pp. 1009–17, 1998.
- [115] K. Rauschenberger *et al.*, “A non-enzymatic function of 17 β -hydroxysteroid dehydrogenase type 10 is required for mitochondrial integrity and cell survival,” *EMBO Mol. Med.*, vol. 2, no. 2, pp. 51–62, Feb. 2010.
- [116] A. J. Deutschmann *et al.*, “Mutation or knock-down of 17 β -hydroxysteroid dehydrogenase type 10 cause loss of MRPP1 and impaired processing of mitochondrial heavy strand transcripts,” *Hum. Mol. Genet.*, vol. 23, no. 13, pp. 3618–28, Feb. 2014.
- [117] T. Fukao *et al.*, “The first case in Asia of 2-methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency (HSD10 disease) with atypical presentation,” *J. Hum. Genet.*, vol. 59, no. 11, pp. 609–614, 2014.
- [118] C. Perez-Cerda *et al.*, “2-Methyl-3-hydroxybutyryl-CoA dehydrogenase (MHBD) deficiency: an X-linked inborn error of isoleucine metabolism that may mimic a mitochondrial disease.,” *Pediatr. Res.*, vol. 58, no. 3, pp. 488–91, Sep. 2005.
- [119] J. García-Villoria *et al.*, “Study of patients and carriers with 2-methyl-3-hydroxybutyryl-CoA dehydrogenase (MHBD) deficiency: difficulties in the diagnosis.,” *Clin. Biochem.*, vol. 42, no. 1–2, pp. 27–33, 2009.
- [120] L. H. Seaver *et al.*, “A novel mutation in the HSD17B10 gene of a 10-year-old boy with refractory epilepsy, choreoathetosis and learning disability.,” *PLoS One*, vol. 6, no. 11, p. e27348, Jan. 2011.
- [121] S.-Y. Yang, C. Dobkin, X.-Y. He, M. Philipp, and W. T. Brown, “A 5-methylcytosine

- hotspot responsible for the prevalent HSD17B10 mutation.,” *Gene*, vol. 515, no. 2, pp. 380–4, Feb. 2013.
- [122] B. T. Poll-The *et al.*, “Spastic diplegia and periventricular white matter abnormalities in 2-methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency, a defect of isoleucine metabolism: differential diagnosis with hypoxic-ischemic brain diseases.,” *Mol. Genet. Metab.*, vol. 81, no. 4, pp. 295–9, 2004.
- [123] M. R. Cazorla, A. Verdú, C. Pérez-Cerdá, and A. Ribes, “Neuroimage findings in 2-methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency.,” *Pediatr. Neurol.*, vol. 36, no. 4, pp. 264–7, 2007.
- [124] R. Ensenuer *et al.*, “Clinical variability in 3-hydroxy-2-methylbutyryl-CoA dehydrogenase deficiency,” *Ann. Neurol.*, vol. 51, no. 5, pp. 659–9, 2002.
- [125] J. O. Sass, R. Forstner, and W. Sperl, “2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency: impaired catabolism of isoleucine presenting as neurodegenerative disease,” *Brain Dev.*, vol. 26, no. 1, pp. 12–14, Jan. 2004.
- [126] V. R. Sutton, W. E. O’Brien, G. D. Clark, J. Kim, and R. J. A. Wanders, “3-Hydroxy-2-methylbutyryl-CoA dehydrogenase deficiency,” *J. Inherit. Metab. Dis.*, vol. 26, pp. 69–71, 2003.
- [127] T. Fukao *et al.*, “Two patients with atypical form and one with infantile form of HSD10 disease were identified in Japan,” *Abstr. J Inherit Metab Dis*, vol. 38, no. Suppl. 1, p. 151, 2015.
- [128] S. Akagawa *et al.*, “Japanese Male Siblings with 2-Methyl-3-Hydroxybutyryl- CoA Dehydrogenase Deficiency (HSD10 Disease) Without Neurological Regression,” *JIMD Rep.*, pp. 113–116, 2016.
- [129] C. F. Lorea *et al.*, “Case report: atypical juvenile parkinsonism and basal ganglia calcifications due to HSD10 disease,” *Abstr. J Inherit Metab Dis*, vol. 38, no. Suppl.

- 1, p. 210, 2015.
- [130] M. J. Falk *et al.*, “A novel HSD17B10 mutation impairing the activities of the mitochondrial RNase P complex causes X-linked intractable epilepsy and neurodevelopmental regression,” *RNA Biol.*, vol. 13, no. 5, pp. 477–485, 2016.
- [131] S. E. OLPIN *et al.*, “2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency in a 23-year-old man,” *J. Inherit. Metab. Dis.*, vol. 25, pp. 477–482, 2002.
- [132] S. H. Korman, “Inborn errors of isoleucine degradation: A review,” *Mol. Genet. Metab.*, vol. 89, no. 4, pp. 289–299, 2006.
- [133] T. Fukao, C. R. Scriver, and N. Kondo, “The clinical phenotype and outcome of mitochondrial acetoacetyl-CoA thiolase deficiency (beta-ketothiolase or T2 deficiency) in 26 enzymatically proved and mutation-defined patients,” *Mol. Genet. Metab.*, vol. 72, pp. 109–114, 2001.
- [134] E. Vilardo and W. Rossmanith, “Molecular insights into HSD10 disease: impact of SDR5C1 mutations on the human mitochondrial RNase P complex,” *Nucleic Acids Res.*, pp. 1–9, 2015.
- [135] M. I. G. Sanchez *et al.*, “RNA processing in human mitochondria,” *Cell Cycle*, vol. 10, no. 17, pp. 2904–2916, Sep. 2011.
- [136] M. D. Metodiev *et al.*, “Recessive Mutations in TRMT10C Cause Defects in Mitochondrial RNA Processing and Multiple Respiratory Chain Deficiencies,” *Am. J. Hum. Genet.*, vol. 98, no. 5, pp. 993–1000, 2016.
- [137] L. A. M. Demain *et al.*, “A Homozygous Variant in the Mitochondrial RNase P Subunit PRORP Causes the Hearing Loss and Premature Ovarian Insufficiency of Perrault syndrome,” *Conf. Mol. Biol. Hear. Deaf. 2016*, 2016.
- [138] I. Hochberg *et al.*, “A Variant in a Mitochondrial-tRNA Processing Gene Causes Ovarian Dysgenesis and sensorineural Deafness (Perrault syndrome),” in *ENDO*

Conference 2016, 2016.

- [139] T. Ito *et al.*, “Structural basis for methyl-donor-dependent and sequence-specific binding to tRNA substrates by knotted methyltransferase TrmD,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 31, pp. E4197–E4205, 2015.
- [140] L. Aravind and E. V. Koonin, “THUMP--a predicted RNA-binding domain shared by 4-thiouridine, pseudouridine synthases and RNA methylases,” *Trends Biochem. Sci.*, vol. 26, no. 4, pp. 215–7, 2001.
- [141] L. Aravind and E. V. Koonin, “Novel predicted RNA-binding domains associated with the translation machinery,” *J. Mol. Evol.*, vol. 48, no. 3, pp. 291–302, 1999.
- [142] D. T. Jones, “Protein secondary structure prediction based on position-specific scoring matrices,” *J. Mol. Biol.*, vol. 292, pp. 195–202, 1999.
- [143] S. F. Altschul *et al.*, “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,” *Nucleic Acids Res.*, vol. 25, no. 17, pp. 3389–402, Sep. 1997.
- [144] D. T. Jones, “Protein secondary structure prediction based on position-specific scoring matrices,” *J. Mol. Biol.*, vol. 292, pp. 195–202, 1999.
- [145] J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, and D. T. Jones, “The DISOPRED server for the prediction of protein disorder,” *Bioinformatics*, vol. 20, no. 13, pp. 2138–2139, 2004.
- [146] D. W. A. Buchan, F. Minneci, T. C. O. Nugent, K. Bryson, and D. T. Jones, “Scalable web services for the PSIPRED Protein Analysis Workbench,” *Nucleic Acids Res.*, vol. 41, no. Web Server issue, pp. 349–357, 2013.
- [147] K. S. Johnsonb, “Single-step purification S-transferase,” vol. 67, pp. 31–40, 1988.
- [148] M. Hedhammar and S. Hober, “Z(basic)--a novel purification tag for efficient protein recovery,” *J. Chromatogr. A*, vol. 1161, no. 1–2, pp. 22–8, Aug. 2007.

- [149] T. P. Hopp *et al.*, “A short polypeptide marker sequence useful for recombinant protein identification and purification,” *Nat. Biotechnol.*, vol. 6, pp. 709–712, 1988.
- [150] F. H. Niesen, H. Berglund, and M. Vedadi, “The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability,” *Nat. Protoc.*, vol. 2, no. 9, pp. 2212–2221, 2007.
- [151] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, “Features and development of Coot,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. Pt 4, pp. 486–501, 2010.
- [152] Z. Dauter, “Data-collection strategies,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 55, no. 10, pp. 1703–1717, 1999.
- [153] P. R. Evans and G. N. Murshudov, “How good are my data and what is the resolution?,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 69, no. 7, pp. 1204–1214, 2013.
- [154] B. Rupp, *BIOMOLECULAR CRYSTALLOGRAPHY Principles, Practice, and Application to Structural Biology*. New York: Garland Science, Taylor & Francis Group, LLC, 2010.
- [155] E. Krissinel and K. Henrick, “Interference of macromolecular assemblies from crystalline state,” *J. Mol. Biol.*, vol. 372, no. 3, pp. 774–97, 2007.
- [156] H. L. Schubert, R. M. Blumenthal, and X. Cheng, “Many paths to methyltransfer: a chronicle of convergence,” *Trends Biochem. Sci.*, vol. 28, no. 6, pp. 329–335, 2003.
- [157] O. Nureki *et al.*, “An enzyme with a deep trefoil knot for the active-site architecture,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 58, no. 7, pp. 1129–1137, 2002.
- [158] K. Anand, D. Pal, and R. Hilgenfeld, “An overview on 2-methyl-2, 4-pentanediol in crystallization and in crystals of biological macromolecules An overview on 2-methyl-2, 4-pentanediol in macromolecules,” *Acta Crystallogr. D*, vol. D58, pp.

1722–1728, 2002.

- [159] B. Van Laer *et al.*, “Structural and functional insights into tRNA binding and adenosine N1-methylation by an archaeal Trm10 homologue,” *Nucleic Acids Res.*, vol. 44, no. 2, pp. 940–53, 2016.
- [160] A. Sali and T. L. Blundell, “Comparative Protein Modelling by Satisfaction of SPatial Restraints,” *J Mol Biol*, vol. 234, pp. 779–815, 1993.
- [161] H. Fischer, M. de Oliveira Neto, H. B. Napolitano, I. Polikarpov, and A. F. Craievich, “Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale,” *J. Appl. Crystallogr.*, vol. 43, pp. 101–109, 2010.
- [162] O. Schilling, B. Späth, B. Kostecky, A. Marchfelder, W. Meyer-Klaucke, and A. Vogel, “Exosite modules guide substrate recognition in the ZiPD/ElaC protein family,” *J. Biol. Chem.*, vol. 280, no. 18, pp. 17857–17862, 2005.
- [163] B. Kostecky, E. Pohl, A. Vogel, O. Schilling, and W. Meyer-Klaucke, “The crystal structure of the zinc phosphodiesterase from *Escherichia coli* provides insight into function and cooperativity of tRNase Z-family proteins,” *J. Bacteriol.*, vol. 188, no. 4, pp. 1607–1614, 2006.
- [164] R. J. Liu, T. Long, M. Zhou, X. L. Zhou, and E. D. Wang, “tRNA recognition by a bacterial tRNA Xm32 modification enzyme from the SPOUT methyltransferase superfamily,” *Nucleic Acids Res.*, vol. 43, no. 15, pp. 7489–7503, 2015.
- [165] E. F. Pettersen *et al.*, “UCSF Chimera - A visualization system for exploratory research and analysis,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, 2004.
- [166] D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, and A. Sali, “Accurate SAXS profile computation and its assessment by contrast variation experiments,” *Biophys. J.*, vol. 105, no. 4, pp. 962–974, 2013.

- [167] D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, and A. Sali, “FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles,” *Nucleic Acids Res.*, vol. 44, no. May, pp. 424–429, 2016.
- [168] J. L. Cole, J. W. Lary, T. Moody, and T. M. Laue, “Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium,” *Methods Cell Biol.*, vol. 84, no. 7, pp. 143–179, 2009.
- [169] D. J. Houde and S. A. Berkowitz, *BIOPHYSICAL CHARACTERIZATION OF PROTEINS IN DEVELOPING BIOPHARMACEUTICALS*. Elsevier, 2015.
- [170] P. Filippakopoulos *et al.*, “Histone recognition and large-scale structural analysis of the human bromodomain family,” *Cell*, vol. 149, no. 1, pp. 214–231, 2012.
- [171] M. Kempenaers, M. Roovers, Y. Oudjama, K. L. Tkaczuk, J. M. Bujnicki, and L. Droogmans, “New archaeal methyltransferases forming 1-methyladenosine or 1-methyladenosine and 1-methylguanosine at position 9 of tRNA,” *Nucleic Acids Res.*, vol. 38, no. 19, pp. 6533–43, Oct. 2010.
- [172] J. E. Jackman, R. K. Montange, H. S. Malik, and E. M. Phizicky, “Identification of the yeast gene encoding the tRNA m1G methyltransferase responsible for modification at position 9,” *RNA*, vol. 9, no. 5, pp. 574–585, May 2003.
- [173] W. E. Swinehart, J. C. Henderson, and J. E. Jackman, “Unexpected expansion of tRNA substrate recognition by the yeast m1G9 methyltransferase Trm10,” *RNA*, vol. 19, no. 8, pp. 1137–46, Aug. 2013.
- [174] P. Atkins and J. D. Paula, *Atkins - Physical Chemistry*, 8th ed. Oxford: OUP, 2006.
- [175] A. Dong, X. Xu, and A. M. Edwards, “In situ proteolysis for protein crystallization and structure determination,” vol. 4, no. 12, pp. 1019–1021, 2007.
- [176] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and

- R. J. Read, "Phaser crystallographic software," *J. Appl. Crystallogr.*, vol. 40, no. Pt 4, pp. 658–674, 2007.
- [177] B. Van Laer *et al.*, "Structural and functional insights into tRNA binding and adenosine N1-methylation by an archaeal Trm10 homologue," *Nucleic Acids Res.*, p. gkv1369, 2015.
- [178] M. Punta *et al.*, "The Pfam protein families database.," *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D290-301, Jan. 2012.
- [179] L. Reinhard, S. Sridhara, and B. M. Hallberg, "Structure of the nuclease subunit of human mitochondrial RNase P," *Nucleic Acids Res.*, vol. 43, no. 11, pp. 5664–72, 2015.
- [180] F. Li, X. Liu, W. Zhou, X. Yang, and Y. Shen, "Auto-inhibitory Mechanism of the Human Mitochondrial RNase P Protein Complex," *Sci. Rep.*, vol. 5, p. 9878, 2015.
- [181] R. Sousa, "Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization," *Acta Crystallogr. - Sect. D Biol. Crystallogr.*, vol. 51, no. 3, pp. 271–277, 1995.
- [182] F. Pinker, P. Giegé, and C. Sauter, "Crystallization and crystallographic analysis of an *Arabidopsis* nuclear proteinaceous RNase P," *Acta Crystallogr. Sect. F Struct. Biol. Commun.*, vol. 71, no. 11, pp. 1372–1377, 2015.
- [183] T. Imai *et al.*, "Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in *Arabidopsis thaliana* play a crucial role in recognition of nucleotide bases at T Ψ C loop in precursor tRNAs," *Biochem. Biophys. Res. Commun.*, vol. 450, no. 4, pp. 1541–1546, Jul. 2014.
- [184] A. Gobert *et al.*, "Structural insights into protein-only RNase P complexed with tRNA," *Nat. Commun.*, vol. 4, p. 1353, Jan. 2013.
- [185] A. J. Powell *et al.*, "Recognition of structurally diverse substrates by type II 3-

- hydroxyacyl-CoA dehydrogenase (HADH II)/Amyloid- β binding alcohol dehydrogenase (ABAD),” *J. Mol. Biol.*, vol. 303, no. 2, pp. 311–327, 2000.
- [186] D. Chelius, C. Loeb-Hennard, S. Fleischer, and J. O. McIntyre, “Phosphatidylcholine Activation of Human Heart (R)-3-Hydroxybutyrate Dehydrogenase Mutants Lacking Active Center Sulfhydryls: Site-Directed Mutagenesis of a New Recombinant Fusion Protein,” *Biochemistry*, vol. 39, pp. 9687–9697, 2000.
- [187] G. Michel, V. Sauvage, R. Larocque, Y. Li, A. Matte, and M. Cygler, “The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot,” *Structure*, vol. 10, no. 10, pp. 1303–1315, 2002.
- [188] K. Lim *et al.*, “Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot,” *Proteins*, vol. 51, no. 1, pp. 56–67, Apr. 2003.
- [189] R. J. Liu, M. Zhou, Z. P. Fang, M. Wang, X. L. Zhou, and E. D. Wang, “The tRNA recognition mechanism of the minimalist SPOUT methyltransferase, TrmL,” *Nucleic Acids Res.*, vol. 41, no. 16, pp. 7828–7842, 2013.
- [190] K. Watanabe *et al.*, “Roles of Conserved Amino Acid Sequence Motifs in the SpoU (TrmH) RNA Methyltransferase Family,” *J. Biol. Chem.*, vol. 280, no. 11, pp. 10368–10377, 2005.
- [191] E. Pleshe, J. Truesdell, and R. T. Batey, “Structure of a class II TrmH tRNA-modifying enzyme from *Aquifex aeolicus*,” *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, vol. 61, no. 8, pp. 722–728, 2005.
- [192] O. Nureki *et al.*, “Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme,” *Structure*, vol. 12, no. 4, pp. 593–602, 2004.
- [193] T. Christian, G. Lahoud, C. Liu, and Y. M. Hou, “Control of Catalytic Cycle by a Pair

- of Analogous tRNA Modification Enzymes,” *J. Mol. Biol.*, vol. 400, no. 2, pp. 204–217, 2010.
- [194] H. Hori *et al.*, “Identification and characterization of tRNA (Gm18) methyltransferase from *Thermus thermophilus* HB8: domain structure and conserved amino acid sequence motifs,” *Genes Cells*, vol. 7, no. 3, pp. 259–72, 2002.
- [195] B. Klemm *et al.*, “Precursor tRNA recognition and alternative cleavage site selection by protein-only RNase Ps from *Arabidopsis thaliana*,” in *RNA2016 Conference*, 2016.
- [196] N. Wu, X. Liu, and C. Fierke, “Ribonuclease P-catalyzed tRNA maturation in human mitochondria,” in *RNA2016 Conference*, 2016.
- [197] T. Z. Grove, A. L. Cortajarena, and L. Regan, “Ligand binding by repeat proteins: natural and designed,” *Curr. Opin. Struct. Biol.*, vol. 18, no. 4, pp. 507–515, 2008.
- [198] D. Deng *et al.*, “Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors,” *Science (80-.)*, vol. 335, no. February, pp. 720–723, 2012.
- [199] H. Gao, X. Wu, J. Chai, and Z. Han, “Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region,” *Cell Res.*, vol. 22, no. 12, pp. 1716–20, 2012.
- [200] M. W. Hentze, “Enzymes as RNA-binding proteins: a role for (di)nucleotide-binding domains?,” *Trends Biochem. Sci.*, vol. 19, no. 3, pp. 101–3, Mar. 1994.
- [201] A. C. Wallace, R. A. Laskowski, and J. M. Thornton, “LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions,” *Protein Eng.*, vol. 8, no. 2, pp. 127–134, 1995.
- [202] S.-Y. Yang, X.-Y. He, C. Isaacs, C. Dobkin, D. Miller, and M. Philipp, “Roles of 17 β -hydroxysteroid dehydrogenase type 10 in neurodegenerative disorders,” *J. Steroid Biochem. Mol. Biol.*, vol. 143, pp. 460–472, 2014.
- [203] R. Chen *et al.*, “Analysis of 589,306 genomes identifies individuals resilient to severe

- Mendelian childhood diseases,” *Nat. Biotechnol.*, vol. 34, no. 5, pp. 531–538, 2016.
- [204] M. J. Casadaban and S. N. Cohen, “Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*,” *J. Mol. Biol.*, vol. 138, no. 2, pp. 179–207, 1980.
- [205] F. W. Studier and B. A. Moffatt, “Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes,” *J. Mol. Biol.*, vol. 189, no. 1, pp. 113–130, 1986.
- [206] B. A. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., “Protein Identification and Analysis Tools on the ExPASy Server,” *Proteomics Protoc. Handb.*, pp. 571–607, 2005.
- [207] D. N. Perkins, D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell, “Probability-based protein identification by searching sequence databases using mass spectrometry data,” *Electrophoresis*, vol. 20, no. 18, pp. 3551–3567, 1999.
- [208] G. Ralston, *Introduction to Analytical Ultracentrifugation*. Sidney, Australia: Department of Biochemistry, The university of Sidney, 1993.
- [209] W. Mächtle and L. Börger, *Analytical ultracentrifugation of polymers and nanoparticles*. 2006.
- [210] P. Schuck, “Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling,” *Biophys. J.*, vol. 78, no. 3, pp. 1606–1619, 2000.
- [211] S. Picaud and P. Filippakopoulos, “SPOTting Acetyl-Lysine Dependent Interactions,” *Microarrays*, vol. 4, no. 3, pp. 370–388, 2015.
- [212] J. Newman *et al.*, “Towards rationalization of crystallization screening for small- To medium-sized academic laboratories: The PACT/JCSG+ strategy,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 61, no. 10, pp. 1426–1431, 2005.
- [213] A. Chaikuad, S. Knapp, and F. von Delft, “Defined PEG smears as an alternative

- approach to enhance the search for crystallization conditions and crystal-quality improvement in reduced screens,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 71, no. 8, pp. 1627–1639, 2015.
- [214] C. Grimm, A. Chari, K. Reuter, and U. Fischer, “A crystallization screen based on alternative polymeric precipitants,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 6, pp. 685–697, 2010.
- [215] T. Bergfors, “Seeds to crystals,” *J. Struct. Biol.*, vol. 142, no. 1, pp. 66–76, 2003.
- [216] E. A. Stura and I. A. Wilson, “Applications of the streak seeding technique in protein crystallization,” *J. Cryst. Growth*, vol. 110, no. 1–2, pp. 270–282, 1991.
- [217] P. A. Karplus and K. Diederichs, “Assessing and maximizing data quality in macromolecular crystallography,” *Curr Opin Struct Biol*, vol. 34, pp. 60–68, 2015.
- [218] D. M. Blow, *Outline of crystallography for biologists*. Oxford University Press, 2002.
- [219] W. Kabsch, “Integration, scaling, space-group assignment and post-refinement,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 133–144, 2010.
- [220] T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. W. Leslie, “iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 67, no. 4, pp. 271–281, 2011.
- [221] E. Potterton, P. Briggs, M. Turkenburg, and E. Dodson, “A graphical user interface to the CCP4 program suite,” *Acta Crystallogr. - Sect. D Biol. Crystallogr.*, vol. 59, no. 7, pp. 1131–1137, 2003.
- [222] G. M. Sheldrick, “A short history of SHELX,” *Acta Crystallogr. Sect. A Found. Crystallogr.*, vol. 64, no. Pt 1, pp. 112–122, 2008.
- [223] G. N. Murshudov *et al.*, “REFMAC5 for the refinement of macromolecular crystal structures,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 67, no. 4, pp. 355–367,

- 2011.
- [224] P. V. Afonine *et al.*, “Towards automated crystallographic structure refinement with phenix.refine,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 68, no. 4, pp. 352–367, 2012.
- [225] V. B. Chen *et al.*, “MolProbity: All-atom structure validation for macromolecular crystallography,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 1, pp. 12–21, 2010.
- [226] W. L. Delano, “The PyMOL molecular graphics system,” *Version 1.8 Schrödinger, LLC*, 2002.
- [227] R. Das and S. Doniach, “Structural Studies of Proteins and Nucleic Acids in Solution Using Small Angle X-Ray Scattering (SAXS),” in *Soft Matter Characterization*, R. Borsali and R. Pecora, Eds. Stanford University, CA, USA: Springer-Verlag Berlin Heidelberg, 2008.
- [228] M. Levantino, B. A. Yorke, D. C. F. Monteiro, M. Cammarata, and A. R. Pearson, “Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules,” *Curr. Opin. Struct. Biol.*, vol. 35, no. August 2016, pp. 41–48, 2015.
- [229] D. I. Svergun, M. H. Koch, P. A. Timmins, and R. P. May, *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*, First. Oxford Science Publications, 2013.
- [230] V. V. Volkov and D. I. Svergun, “Uniqueness of ab initio shape determination in small-angle scattering,” *J. Appl. Crystallogr.*, vol. 36, pp. 860–864, 2003.
- [231] M. V. Petoukhov *et al.*, “New developments in the ATSAS program package for small-angle scattering data analysis,” *J. Appl. Crystallogr.*, vol. 45, no. Pt 2, pp. 342–350, 2012.

- [232] O. Trott and A. Olson, “AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading,” *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–461, 2010.
- [233] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, “Image processing with imageJ,” *Biophotonics Int.*, vol. 11, no. 7, pp. 36–41, 2004.
- [234] S. L. Hiley *et al.*, “Detection and discovery of RNA modifications using microarrays,” *Nucleic Acids Res.*, vol. 33, no. 1, p. e2, 2005.
- [235] J. Anderson, L. Phan, and A. G. Hinnebusch, “The Gcd10p/Gcd14p complex is the essential two-subunit tRNA(1-methyladenosine) methyltransferase of *Saccharomyces cerevisiae*,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 10, pp. 5173–5178, 2000.
- [236] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, “ExPASy: The proteomics server for in-depth protein knowledge and analysis,” *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3784–3788, 2003.
- [237] J. Notredame, C., Higgins, D. G., & Heringa, C. Notredame, D. G. Higgins, and J. Heringa, “T-coffee: a novel method for fast and accurate multiple sequence alignment,” *J. Mol. Biol.*, vol. 302, no. 1, pp. 205–217, 2000.
- [238] X. Robert and P. Gouet, “Deciphering key features in protein structures with the new ENDscript server,” *Nucleic Acids Res.*, vol. 42, no. W1, pp. 320–324, 2014.

8 APPENDIX

8.1 Numbering key for purine bases

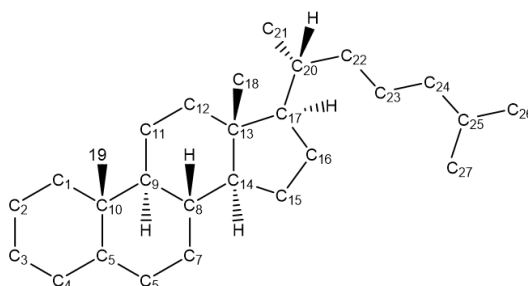
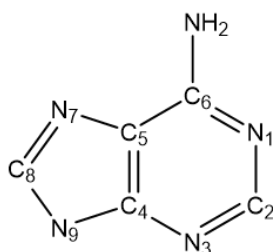


Figure 8-1. Numbering key for steroid chemical structures.

Adenine



Guanine

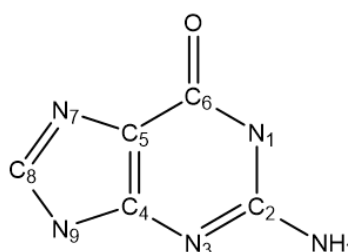


Figure 8-2. Numbering key for purine base chemical structures.

8.2 Overview of discussed MTases subfamilies

In this thesis, many MTase subfamilies are included in discussions. **Table 8-1** provides an overview of these subfamilies with information relevant for discussions.

Superfamily	SPOUT	SPOUT	SPOUT	RFM	SPOUT
Family (Eukaryote/archaea)	Trm3	Trm7	Trm5	Trm6-Trm61	Trm10
Family (Bacteria)	TrmH	TrmL	TrmD	TrmI	N.I.
Catalytic function	m ²⁺ -OH ₁₈	m ²⁺ -OHG34	m ¹ G37	m ¹ A58	m ¹ A9/m ¹ G9
Oligomeric state	Sym. dimer	Sym. dimer	Asym. dimer	Tetramer	Monomer /This work

Table 8-1. Overview of MTase families discussed throughout this thesis work [58], [164], [190], [234], [235]. N.I.: Not identified. Sym.: symmetrical. Asym.: asymmetrical.

8.3 Sequence alignment of Trm10 members

32 Trm10 protein DNA sequences were identified using the ProSite ExPaSY server [236] with ENTRY: PS51675 belonging to the SAM-dependent MTase Trm10-type. An

alignment of the identified sequences is presented in **Figure 8-3**. Sequence alignment was performed using the T-coffee server [237] and the sequence alignment figure was generated in ESPrnt [238].

```

MRRP1_HUMAN
Trm10_YEAST
Trm10_SCHPO
Trm10_SULAC
      oooooo..o..oo.....
1      10      20      30      40      50
MRRP1_HUMAN  MAFLKMS.V..SVNFFRPFTRFLVFPFTLHRKRNNL..TIL.QRYMSSK.IPA..VITYPKNE..STPPSE
TRMT10A_HUMAN MSSEMLPA.F..IET.....S.....N.....
TRMT10B_HUMAN MDWKLEGS.T..QKV.....E.....S.....PV.....
MRRP1_BOVIN  MPVLLKMS.V..SITFLRPFARVLVFPFTLHRKRRVLYSTIM.QRYMSSK.IPA..ASYPNKE..STPPSE
TRMT10A_BOVIN MSSEMLPA.F..SET.....C.....N.....
TRMT10B_BOVIN MDWKLEGS.A..QKT.....E.....S.....RV.....
MRRP1_RAT    MNVTVR.....FLRPFARYLVPLYTFHRTSRNSYSRVL.QRYVSSK.VPS..LPCHNKD..STSPPE
TRMT10A_RAT  MSSEVLPA.S..IEP.....P.....N.....
TRMT10B_RAT  MDCEWEGR.P..QRA.....G.....S.....RA.....
MRRP1_MOUSE  MNVTVR.....FLRPFARCLVPYTFHRRKSHLYSGVL.QRYMSSK.APS..LSCHNKD..SASPPE
TRMT10A_MOUSE MSSEMLPA.S..IES.....P.....N.....
TRMT10B_MOUSE MDCKSEES.A..QRT.....D.....S.....QA.....
MRRP1_XENTR  MAFVNT.....LLRTIRCSAVHTLVQEGRSLSLLKASHQLTQSRK..IML..SNHVRKEEAKSEPNE
TRMT10A_XENTR MSLEAVTE.P..GKC.....S.....I.....
MRRP1_DROME  MLKHFA.....RWRLGSQLLKGCAPVRQAS.K.TSSAENLIA.GTEEPQKK..FVNPF.
TRMT10A_DROME MDTTEKAL.....KEV.....T.....PA.....
TRMT10B_DANRE MEEQLRVG.G..CSP.....E.....
Trm10_YEAST  MSNDEINQNE..E.....T.N.....
Trm10_SCHPO  MENKDALD.I..GKD.....D.....
Trm10_CANAL  MTEQTSEA.T..VVN.....N.....S.....
Trm10_DEBHA  MENQDTEQ.S..QKR.....S.....G.....
Trm10_CRYNJ  MKFHLGLT.EALQER..E.....PN.....
Trm10_CRYNB  MRFHLGLT.EALQER..E.....PN.....
Trm10_CANGA  MSDTSDLV.D..G.....ET.....
Trm10_GIBZE  MDLDPAHK.P..SQA.....E.....GSSE.....
Trm10_EMENI  MEEEEERPR.K..IQK.....L.....NHDE.....
Trm10_ASPFU  MEDDDRPR.K..YPK.....L.....A.....
Trm10_YARLI  MVDVTEPV.A..PV.....E.....
Trm10_ASHGO  MTPETNND.E..T.....LSE.....
Trm10_KLULA  MSDTSENS.N..A.....
Trm10_SULAC  MTLAKVF..S..QK.....
Trm10_THEKO  MKTLADVF.R..EA.....
consensus>70  M.....

MRRP1_HUMAN
Trm10_YEAST
Trm10_SCHPO
Trm10_SULAC
      .....oo..o
60      70      80      90      100
MRRP1_HUMAN  ELELDKWKTTMKSSVQ.E..E.C...VST..ISSKDEDP.....LAATREFIE.....M
TRMT10A_HUMAN .....VD.K.KQ.G...INE..D.....QE
TRMT10B_HUMAN .....LQ.G.QE.G...ILE.E.T..GEDG.....LPEGFQLLQIDAE..G
MRRP1_BOVIN  ELELDKWKITMKSSVQ.E..E.D...VST..ATSEDEDP.....LAATRELVE.....M
TRMT10A_BOVIN .....VE.R.QK.N...LSE..D.....GE
TRMT10B_BOVIN .....LQ.E.HE.I...TLD..D.P..GEDG.....VSESFQLLQIDVG..C
MRRP1_RAT    QLELDGWKTTMKSSIQ.E..N.G...VSV..VS.DKDEDS.....LAATRELIE.....M
TRMT10A_RAT  .....VE.G.KP.G...ASD..S.....EE
TRMT10B_RAT  .....SQ.D.PE.G...LPE..A...RDDG.....LSESFLLQMDVE..Y
MRRP1_MOUSE  QLELDGWKATMKSSIQ.E..D.G...VSE..VS.DKDEDS.....LASTRELIE.....M
TRMT10A_MOUSE .....VE.E.KL.G...TSO..G.....EE
TRMT10B_MOUSE .....FQ.E.PD.G...LPE..A.G..GEDG.....LSESFQLLQVDVE..Y
MRRP1_XENTR  TLDLEEWKSIKSDIG.N..A.E...MVK..TE.TQEDSS.....LNEMQELVE.....M
TRMT10A_XENTR .....DP.N.TKDL...LAS...Q..H...AGNNTPLQE
MRRP1_DROME  .SQ.....PAPALS.N..D.T...ISE..NK.EERDKR.....LKVQLQLEAD...I
TRMT10A_DROME .....ML.N.....S.L.....
TRMT10B_DANRE .....D.G...GFT..I.V..NEAP.....VADLLEFLHIDVE..L
Trm10_YEAST  .....KV.K.RT.P.....ADV..K
Trm10_SCHPO  .....TS.E.....TP.....EEIE..E
Trm10_CANAL  .....PA.P..T.IPK...IK.REKP.....SEVE..K
Trm10_DEBHA  .....INAS.SKDL...SPLHFE.LYRRSEP.....I.....MDID..E
Trm10_CRYNJ  .....INAS.SKDL...SLLHFE.LYRRSEP.....I.....MDID..E
Trm10_CRYNB  .....KW.Q.....
Trm10_CANGA  .....KE.QGNE.Q...GQVE..Q.NQAQDDPTTG..PQAETEKAIIPS...QGSTIP..Q
Trm10_GIBZE  .....AH.E.SE.P...LITG..A.IKGSQDDTSPAEQ.TNSQKAI.P.GAT..TDNS..K
Trm10_EMENI  .....VQ.EGSE.P...VMTG..A.VGAVHDDDAE..S.ASDSNRASPSNDRIDNDVK..Q
Trm10_ASPFU  .....VK.E.TA.GASSATPE...VDRTSEP.....HPKN...PQNVP..A
Trm10_YARLI  .....LS.R.PK.P...RA.....
Trm10_ASHGO  .....EI.P..ADTSD...VKDKPKP.....IVRA...P.....
Trm10_KLULA  .....LR.E..L.G...ISS..IYI.....G..H
Trm10_SULAC  .....LK.E..K.G...ISS..IG.....
Trm10_THEKO  .....
consensus>70  .....

```

MRRP1_HUMAN
Trm10_YEAST
Trm10_SCHPO
Trm10_SULAC

	110	120	130	140
MRRP1_HUMAN	WRL.....GREVP..EH	ITEELKTLMECV	SN TAKKKYLY	LYLTKE
TRMT10A_HUMAN	E.....S.....QKPR.....L.GEGC..EP	I.....SKRQM	KKL	KQKWEE
TRMT10B_HUMAN	EC.....QEGE.....ILATGST.AWC	I.....SKNVO	..RKQRHW	EK
MRRP1_BOVIN	WRL.....GKEVP..EH	FSEELKTLMECV	SKSSK	RYLYIKE
TRMT10A_BOVIN	Q.....N.....QKPG.....S.SERF..QP	I.....SKRQM	KKLM	KQKWE
TRMT10B_BOVIN	EH.....WEQE.....TLPTGSA.AWC	SKNVO	..RKQRHW	EK
MRRP1_RAT	WRL.....GKEVP..EH	ITEELKTLMECA	SKSAK	KKYLYLYGKE
TRMT10A_RAT	E.....R.....QKQR.....V.DAEA..QP	I.....SKRQL	KKLM	KQRLWEE
TRMT10B_RAT	DP.....QE.K.....RRPANS.AWS	SKNVO	..RKQRHW	ER
MRRP1_MOUSE	WRL.....GKEVP..EH	ITEEDLKTLMECA	SKSAK	KKYLYLYGKE
TRMT10A_MOUSE	E.....R.....QEPR.....V.DAGA..EP	I.....SKRQL	KKLM	KQKWE
TRMT10B_MOUSE	ER.....PE.E.....TSPANS.VWS	SKNMQ	..RKQRHW	ER
MRRP1_XENTR	WRLA.....GRAVP..QS	ITTEQLQVLMELP	TKTAR	KKYLYLSVRE
TRMT10A_XENTR	N.....S.....SAPR.....SECKAQ..DAM	SKRQM	KKFL	KQKWE
MRRP1_DROME	AHQE.....GRRVPSLEFF	KDHHWEHVLTLP	TKSARI	KYFYGLWQIE
TRMT10A_DROME	NNCP.....GTTPG..TPM	SKNQL	KKQR	LAEF.AE
TRMT10B_DANRE	EH.....ARDD.....R.....EE.SAR	SKNVM	..RKQRNW	ER
Trm10_YEAST	P.....LPPVP..EGM	SKKQW	KKMC	KQRWE
Trm10_SCHPO	N.....E.....TQEQ..PVL	SKSAL	KKLR	QQEW.DA
Trm10_CANAL	R.....R.....KLKQ.....LEVVP..EGF	SKKQW	KKEL	KQRWQD
Trm10_DEBHA	D.....D.....FKRQ.....KVVPV..EGM	TKREY	KKQL	KQRWEE
Trm10_CRYNJ	E.....S.....YLNAGPSAPSKIPQGGEGDRL	QGM	SKKAM	KRAAKQARLEE
Trm10_CRYNB	E.....S.....YLNAGPSAPSKIPQGGEGDGL	QGM	SKKAM	KRAAKQARLEE
Trm10_CANGA	R.....LPPVP..EGM	SKSQW	KKQWR	QRMQYEA
Trm10_GIBZE	K.....RS.....AEDEPA..QPM	SKNAL	KKLR	KQQWYEA
Trm10_EMENI	ESSEE.PASSTATD.....GKETTE..QKI	SKRQL	RRQA	KLQWEA
Trm10_ASPFU	ACDEEG.....QDAH.....GKDPG..PTL	SKNQL	KKLR	KEHWEA
Trm10_YARLI	P.....Y.....NTK.....TAVIP..EGM	SKNEW	KKAK	KAIWES
Trm10_ASHGO	A.....LPPVP..EGM	SKSQW	KKQW	KKEQFEL
Trm10_KLULA	Q.....LPPVP..EGI	SKSQW	KKIC	RRKRFEE
Trm10_SULAC	F.....ER..PSL	QSLAI	KMLL	KNYGLVEERREGML
Trm10_THEKO	TL	SKRFR	KS	KNKL
consensus>70			sk	k.k.e

MRRP1_HUMAN
Trm10_YEAST
Trm10_SCHPO
Trm10_SULAC

	150	160	170	180
MRRP1_HUMAN	KVKKARQIKKEMKAAAREE	AK...NI.KL	LETT.EED.KOKNFLFLR	..LW
TRMT10A_HUMAN	QRELRLKQKKEKRRKKLE	RQC.Q.M	EPN.S.DGH.D.R	
TRMT10B_HUMAN	IVAANKSKRKQEKERRKAN	RAE.N.P	G.I.CPQ.H.SKRFLR	..AL
MRRP1_BOVIN	KMKKARQIKKEMKAAKEE	PK...KD.QL	PETI..KED.QKQNFLFLR	..LW
TRMT10A_BOVIN	QRELRLKQKKEKRRKKLE	RQC.Q.P	ESN.S.DGS.D.R	
TRMT10B_BOVIN	IVAANKSKRKQEKERRKAN	RVE.N.S	G.I.YPQ.H.SKRFLR	..SL
MRRP1_RAT	MMKKAKQMKKEMKAAAREE	AK...RA.RS	LEPS..TGE.EQDFMFLR	..LW
TRMT10A_RAT	QREQRKEKKEKRRKKLE	RRC.Q.L	ESN.S.DGN.D.R	
TRMT10B_RAT	IVSSKSKRKQERERRKIK	RAE.D.L	GNG.T.CPQ.H.SKRFLK	..AL
MRRP1_MOUSE	KAKKAKQVKKEMKAAAREE	AK...RA.RL	LETT..AEE.QQDFMFLR	..LW
TRMT10A_MOUSE	QREQRKEKKEKRRKKLE	RR..Q.L	ESN.S.DGN.D.R	
TRMT10B_MOUSE	IVSSKSKRKQERERRKAK	RAE.D.P	GNG.T.CPQ.H.SKRFLK	..AL
MRRP1_XENTR	VMKTN...RKEKKKELKES	KSKIESL.DQ	LETK..EDTPEKKNTFLLH	..VW
TRMT10A_XENTR	QRELRLKQKKEKRRKKLE	RQA.Q.A	EHN.I.DAN.S.R	
MRRP1_DROME	MKKEADQKKAERAKEAER	RV.AEMRKER	EENTHIIYG.LGHTSLFLR	..IY
TRMT10A_DROME	LRKLRREREREKKQKRE	AKE	LG.L..PV.R.TGP	
TRMT10B_DANRE	RLEVKKSKKEKLRKKLN	RQD.K.D	V.S.DAQ.L.SKRVIK	..AI
Trm10_YEAST	NKAKYNAERVRKKRLRHE	RSA.K.IQEY	IDRG.E.EVPQE.LIRE	
Trm10_SCHPO	GREKRAEMREKRLRKEE	RKR.K.I	EAG.E.VVKSQ.KK	
Trm10_CANAL	TKQEYLEVQREKRLARQR	KRE.R.LKDL	DEN.....DE.L	
Trm10_DEBHA	TKDEYKQKKEKKAARER	RKE.R.IKEA	EANG.....ETNEE.LYNYHQ	..M
Trm10_CRYNJ	IKPLKRAAERERRRQRTAQ	LAE.....G.YAAG.T	LSE.A.DKELVERRRRV	
Trm10_CRYNB	IKPLKRAAERERRRQRTAQ	LAE.....G.YAAG.T	LSE.A.DKELVERRRRV	
Trm10_CANGA	KKEEYAEVRKEKRRAKEN	RRK.K.IQEY	IDRG.E.EVPAE.LKRE	
Trm10_GIBZE	GKEDRKLRKDSRIARKVVR	KRE.E.RDAL	IAQG.I.NPY	
Trm10_EMENI	KREERKIIKQKTAARKER	RRE.L.WEEA	KREG.K.DPNEE.LNKLF	..MT
Trm10_ASPFU	MREQRKVKKEKLVAKRER	RRA.A.LEQA	KQEG.A.EATEE.TRKA	FE..S
Trm10_YARLI	KKDEIAAVKKEKKAARKR	KQL.A.ISRG	EIPA.....PIPD	
Trm10_ASHGO	NKPLYAKIRKEKQKAREQ	RRE.R.LQKALE	ENG.G.EIPEE.LRRT	
Trm10_KLULA	TRAEYAQIRKEKRNRAKLA	RRE.K.LKEY	TDRG.E.EIPEE.LKRP	
Trm10_SULAC	ITQDHGIKLISGKGTESTR.....YTFRKGG			
Trm10_THEKO	QDIAIEIVHGKGA.VFRVPEKTAVAWDLNGN	RVD.G.S	YYA.Y.APLCM	REKFEP..VL
consensus>70	ke			


```

MRRP1_HUMAN      000000.0
Trm10_YEAST      0.0      00000000
Trm10_SCHPO      0.0      00000000
Trm10_SULAC      0.0      00000000

370
MRRP1_HUMAN      N.NG.NWQEAALQ.FV.....PKRK
TRMT10A_HUMAN    E.TR.DWQEAFFFTIL.....PQRK
TRMT10B_HUMAN    E.TH.NWPEALKKGV.....SSGK
MRRP1_BOVIN      N.TG.SWEEALK.FV.....PQRK
TRMT10A_BOVIN    E.TR.DWQEAFFFTIL.....PQRK
TRMT10B_BOVIN    E.TQ.NWPEALKKGV.....SSRK
MRRP1_RAT        N.TG.NWEEALK.FV.....PQRK
TRMT10A_RAT      E.TG.DWQEAFFFTIL.....PQRK
TRMT10B_RAT      E.TR.DWPEALKKGV.....SPGK
MRRP1_MOUSE      N.TG.NWEEALK.FV.....PQRK
TRMT10A_MOUSE    E.TR.DWQEAFFFTIL.....PQRK
TRMT10B_MOUSE    E.TR.NWPEALKKGV.....SPGK
MRRP1_XENTR      D.TG.DWKKALS.FV.....PNRK
TRMT10A_XENTR    E.KK.EWKEAFFSVL.....PQRK
MRRP1_DROME      K.TG.DWDTALK.HV.....PQRK
TRMT10A_DROME    A.GQ.DWTTAILETI.....PMRK
TRMT10B_DANRE    D.TR.DWTKALMAGI.....PPGK
Trm10_YEAST      D.DH.NWKNAFESVL.....PQRK
Trm10_SCHPO      E.YR.DWKEAFMEVI.....PQRK
Trm10_CANAL      ENDG.DWETAFFNKVL.....PQRK
Trm10_DEBHA      ENDK.DWGKAFFNEVL.....PQRK
Trm10_CRYNJ      H.LG.DWAAAFEAIV.....PIRK
Trm10_CRYNB      H.LG.DWAAAFEAIV.....PIRK
Trm10_CANGA      E.SR.DWKEAFESVI.....PSRK
Trm10_GIBZE      E.YE.NWGEAFMSVI.....PQRK
Trm10_EMENI      Q.LR.DWAEAFMQTL.....PQRK
Trm10_ASFFU      E.LG.DWGAFFIQVI.....PQRK
Trm10_YARLI      E.LK.DWKAAFEAVL.....PMRK
Trm10_ASHGO      D.NH.DWKEAFESVL.....PARK
Trm10_KLULA      D.NK.DWKEAFESVL.....PQRK
Trm10_SULAC      E.LDQSLEEAIIISLQSKSDKI.....SR.....LLHDVQLYGM.EVLEEEARWLRADDKVI
Trm10_THEKO      V.EGKSMDAIVYEMQEPHARWRLRKELPKRATRYMVEGKVYRVVEKELFDEYSKWLKI.RWEDFVK.VL
consensus>70    e...dw..a.....p.rk

MRRP1_HUMAN
Trm10_YEAST
Trm10_SCHPO
Trm10_SULAC

390
MRRP1_HUMAN      ....HTGFL.....EISQHSQEFIN.....RL.....KKAKT.....
TRMT10A_HUMAN    ...GAVPTD.....KACESASH.D.....NQS.....V.....RME...
TRMT10B_HUMAN    ...GYILR.....NSV.....E.....
MRRP1_BOVIN      ...HAGYL.....EISQHSQEFIN.....RM.....KKSSTFN.....
TRMT10A_BOVIN    ...GAVPTD.....QACESCSDH.D.....KKF.....A.....RVE...
TRMT10B_BOVIN    ...GYVLR.....NSV.....E.....
MRRP1_RAT        ...HTGYL.....EVPEHSQAAFR.....KL.....KKTKTLN.....
TRMT10A_RAT      ...GAVPAH.....KACESSPO.D.....HQA.....LP...
TRMT10B_RAT      ...GYILQ.....NSV.....E.....
MRRP1_MOUSE      ...HTGYL.....EVSEQSQELVR.....KL.....KKTKTLN.....
TRMT10A_MOUSE    ...GAVPAH.....KACESSPO.D.....HQS.....LP...
TRMT10B_MOUSE    ...GYVLQ.....NSAE.....G.....
MRRP1_XENTR      ...HDGFA.....EPAASKKRNIK.....NDGGMEYAASTKRNLOKNSKMY.....
TRMT10A_XENTR    ...GAIPLT.....ETGEQPEC.R.....ASE.....QE...
MRRP1_DROME      ...VVQNE.....FQPRREKDHWS.....TG.....TRARKLNMRIDHL
TRMT10A_DROME    ...GAKAK.....ITDKK.EPN.H.....CL.....E.....QQ...
TRMT10B_DANRE    ...GFVLA.....SAASKPLEET.....
Trm10_YEAST      ...LDAEA.....KSASSSPAP.K.....D.....
Trm10_SCHPO      ...GILLK.....SDSEFDVSE.D.....TR.....S.....QSN...
Trm10_CANAL      ...IKSKS.....PS.....
Trm10_DEBHA      ...VKGKL.....THGSDPEKSIE.....PS.....E.....VSE...
Trm10_CRYNJ      ...YAPGR.....KTKRAKTET.K.....RN.....E.....KEE...
Trm10_CRYNB      ...YAPGR.....KAKRAKTET.K.....RN.....E.....KVE...
Trm10_CANGA      ...LDPVK.....EKEQQQQQQ.Q.....QQ.....
Trm10_GIBZE      ...GGKLK.....EQQGASGET.Q.....ET.....E.....EAE...
Trm10_EMENI      ...GGALK.....DSVKGQRDS.T.....PR.....D.....DVE...
Trm10_ASFFU      ...GGKLK.....SADHESEDQ.T.....PR.....E.....SVE...
Trm10_YARLI      ...LDPANYGRAHRRARKKARLAEGAEGNDEE.DDEDDDEEE...
Trm10_ASHGO      ...LAELA.....DHAQESNSS.S.....PA.....
Trm10_KLULA      ...LEVDS.....TKEDSET..A.....
Trm10_SULAC      EIV.....
Trm10_THEKO      RELDLVAL.....ERKRIHHLN.....KI.....SNARIIN.....
consensus>70

```

```

MRRP1_HUMAN
Trm10_YEAST
Trm10_SCHPO
Trm10_SULAC

MRRP1_HUMAN
TRMT10A_HUMAN
TRMT10B_HUMAN
MRRP1_BOVIN
TRMT10A_BOVIN
TRMT10B_BOVIN
MRRP1_RAT
TRMT10A_RAT
TRMT10B_RAT
MRRP1_MOUSE
TRMT10A_MOUSE
TRMT10B_MOUSE
MRRP1_XENTR
TRMT10A_XENTR
MRRP1_DROME
TRMT10A_DROME
TRMT10B_DANRE
Trm10_YEAST
Trm10_SCHPO
Trm10_CANAL
Trm10_DEBHA
Trm10_CRYNJ
Trm10_CRYNB
Trm10_CANGA
Trm10_GIBZE
Trm10_EMENI
Trm10_ASPFU
Trm10_YARLI
Trm10_ASHGO
Trm10_KLULA
Trm10_SULAC
Trm10_THEKO
consensus>70
MRRP1_HUMAN
TRMT10A_HUMAN
TRMT10B_HUMAN
MRRP1_BOVIN
TRMT10A_BOVIN
TRMT10B_BOVIN
MRRP1_RAT
TRMT10A_RAT
TRMT10B_RAT
MRRP1_MOUSE
TRMT10A_MOUSE
TRMT10B_MOUSE
MRRP1_XENTR
TRMT10A_XENTR
MRRP1_DROME
TRMT10A_DROME
TRMT10B_DANRE
Trm10_YEAST
Trm10_SCHPO
Trm10_CANAL
Trm10_DEBHA
Trm10_CRYNJ
Trm10_CRYNB
Trm10_CANGA
Trm10_GIBZE
Trm10_EMENI
Trm10_ASPFU
Trm10_YARLI
Trm10_ASHGO
Trm10_KLULA
Trm10_SULAC
Trm10_THEKO
consensus>70

```

Figure 8-3. Sequence alignment of 33 Trm10 protein homologues of MRPP1 using ESPrpt[238]. Known secondary structure elements are displayed for human MRPP1 (MRPP1_HUMAN), *Saccharomyces cerevisiae* (Trm10_YEAST), *Schizosaccharomyces pombe* (Trm10_SCHPO) and *Sulfolobus acidocaldarius* (Trm10_SULAC) structures. Identical and similar residues are boxed in black and grey, respectively. Red boxes: conserved motif I-IV. Blue boxes: potential interaction regions for MRPP2. Green star: residue suggested to determine N1-methylation towards adenine. Purple star: residue that could be involved in MTase activity (Asp293). Red stars: residues mutated in this study. Residue numbering represents human MRPP1. HUMAN: *Homo sapiens*, BOVIN: *Bos Taurus*, RAT: *Rattus norvegicus*, MOUSE: *Mus musculus*, XENTR: *Xenopus tropicalis*, DROME: *Drosophila melanogaster*, DANRE: *Danio rerio*, YEAST: *Saccharomyces cerevisiae*, SCHPO: *Schizosaccharomyces pombe*, CANAL: *Candida albicans*, DEBHA: *Debaryomyces hansenii*, CRYNJ: *Cryptococcus neoformans* var. *neoformans* serotype D, strain JEC21 / ATCC MYA-565, CRYNB: *Cryptococcus neoformans* var. *neoformans* serotype D, strain B-3501A, CANGA: *Candida glabrata*, GIBZE: *Gibberella zeae*, EMENI: *Emericella nidulans*, ASPFU: *Neosartorya fumigata*, YARLI: *Yarrowia lipolytica*, ASHGO: *Ashbya gossypii*, KLULA: *Kluyveromyces lactis*, SULAC: *Sulfolobus acidocaldarius*, THEKO: *Thermococcus kodakarensis*.

8.4 SPOT blot analysis

Each spot in the SPOT blot membrane (**Figure 8-4A**) corresponds to a 15-mer peptide of mature MRPP1 (residues 40-403). The spots that appeared dark on the membrane reacted to the antibody, and therefore represent the sequence regions at which His₆-MRPP2 is bound to the MRPP1 peptide. These sequences are overlaid to detect overlapping regions (**Figure 8-4C, red**) and thus identify the exact MRPP2 targeting peptides. The peptides are then mapped onto the sequence (**Figure 8-4B**) and structure (**Section 3.2.3**) of mature MRPP1 (residues 40-403).

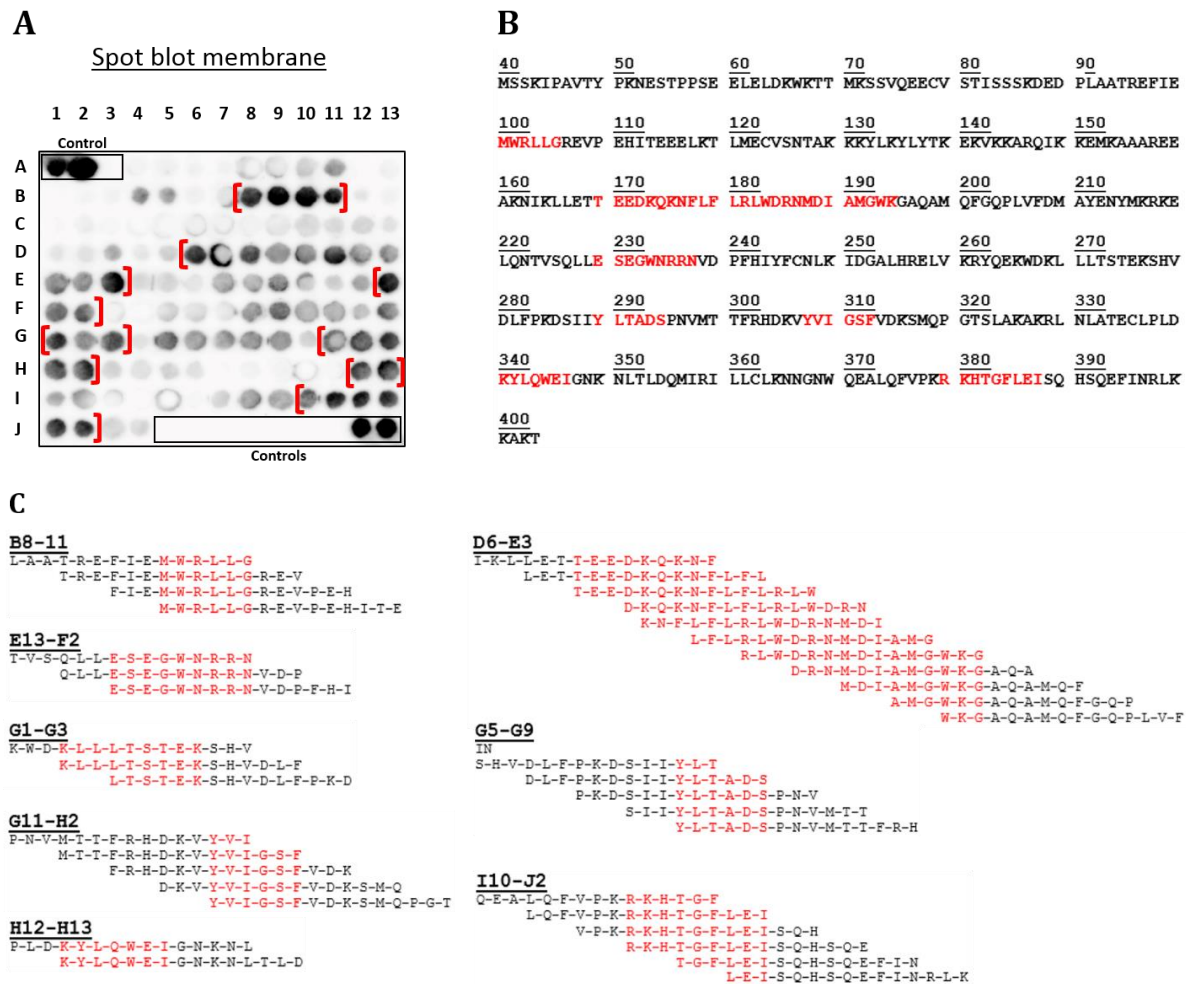


Figure 8-4. A) Peptide membrane with His₆-tagged MRPP2 labelled with HPR-conjugated anti-His antibody and visualised by chemiluminescence. The red boxes indicate regions identified to interact with MRPP1. B) Sequence of mature MRPP1 (residues 40-403) with the identified MRPP2 targeting sequences highlighted. C) SPOT blot peptide analysis of 15-mers corresponding to spots that appeared dark on the membrane.

8.5 SAXS statistics

Table 8-2 shows statistics for SAXS data collection, data processing and *ab initio* model building for all models generated in this work. MWs were calculated using SAXS MoW2 [161] assuming globularity rather than plate-like structures.

	MRPP1	MRPP2	MRPP3	$\Delta 206$ -MRPP3	MRPP1-MRPP2	MRPP1-MRPP2- pre(mt)tRNA ^{lle}	MRPP1-MRPP2- MRPP3-pre(mt)tRNA ^{lle}
Guinier R_g	40.10±2.799	29.24±0.9457	33.79±1.152	33.30±1.519	46.77±1.160	72.71±5.450	108.8±7.249
Real space $I(0)$ (a.u.)	3.016·10 ⁻³ ± 2.894·10 ⁻⁴	9.068·10 ⁻³ ± 2.527·10 ⁻⁴	9.223·10 ⁻³ ± 4.676·10 ⁻⁴	2.868·10 ⁻³ ± 4.573·10 ⁻⁴	2.342·10 ⁻² ± 7.144·10 ⁻⁴	2.171·10 ⁻² ± 1.530·10 ⁻³	4.287·10 ⁻² ± 4.009·10 ⁻³
Real space R_g (Å)	34.40±1.885	28.44±0.4415	33.18±1.118	28.83±1.470	45.59±0.9720	69.67±2.766	82.60±0.9538
Porod V (Å ³)	174930	167454	306000	167576	873102	1279622	4077408
Porod Exponent	2.478	3.340	2.231	2.280	1.914	2.398	2.192
Porod V mass (Å ³)**	85462-105379	99980-100875	179777-184337	85836-100949	488409-525965	705819-770856	1795663-2456269
D_{max} (Å)	100	80	111	98	154	216	248
MW (Da)	97167	86105	62967	66274	214389	436597	630158
Model building							
Point symmetry	P1	P1	P1	-	P1	P1	P1
Models generated	13	13	13	-	13	13	13
Models included*	12	13	13	-	12	13	12
NSD _{mean}	0.673±0.043	0.488±0.006	0.604±0.022	-	0.562±0.032	0.792±0.036	0.765±0.067
χ^2 vs. raw data	2.722	1.044	2.080	-	1.712	1.374	1.500

Table 8-2. Values are obtained as described in **Section 7.6**. $q \cdot R_g < 1.3$ for all datasets. Guinier regions correspond to the Guinier plots represented in **Chapter 2-4**.

8.6 *Ab initio* modelling of MRPP1-MRPP2

The SAXS scattering data obtained for the MRPP1-MRPP2 complex was used to generate *ab initio* models of the protein complex with *P1*, *P2* and *P4* point symmetry to account for all possible tetramer-dimer binding modes (**Figure 8-5**). The statistics for the refined models are shown in **Table 8-3**.

	MRPP1- MRPP2	MRPP1- MRPP2	MRPP1- MRPP2
Model building			
Point symmetry	<i>P1</i>	<i>P2</i>	<i>P4</i>
Models generated	13	13	13
Models included*	13	12	12
NSD _{mean}	0.562±0.032	0.933±0.285	1.190±0.297
χ^2 vs. raw data	1.712	1.683	1.743

Table 8-3. Values are obtained as described in **Section 7.6**. Statistics are shown for SAXS models of MRPP1-MRPP2 build with point symmetry *P1*, *P2* or *P4*. *Models discarded when $NSD > NSD_{mean} + 2 \cdot SD$.

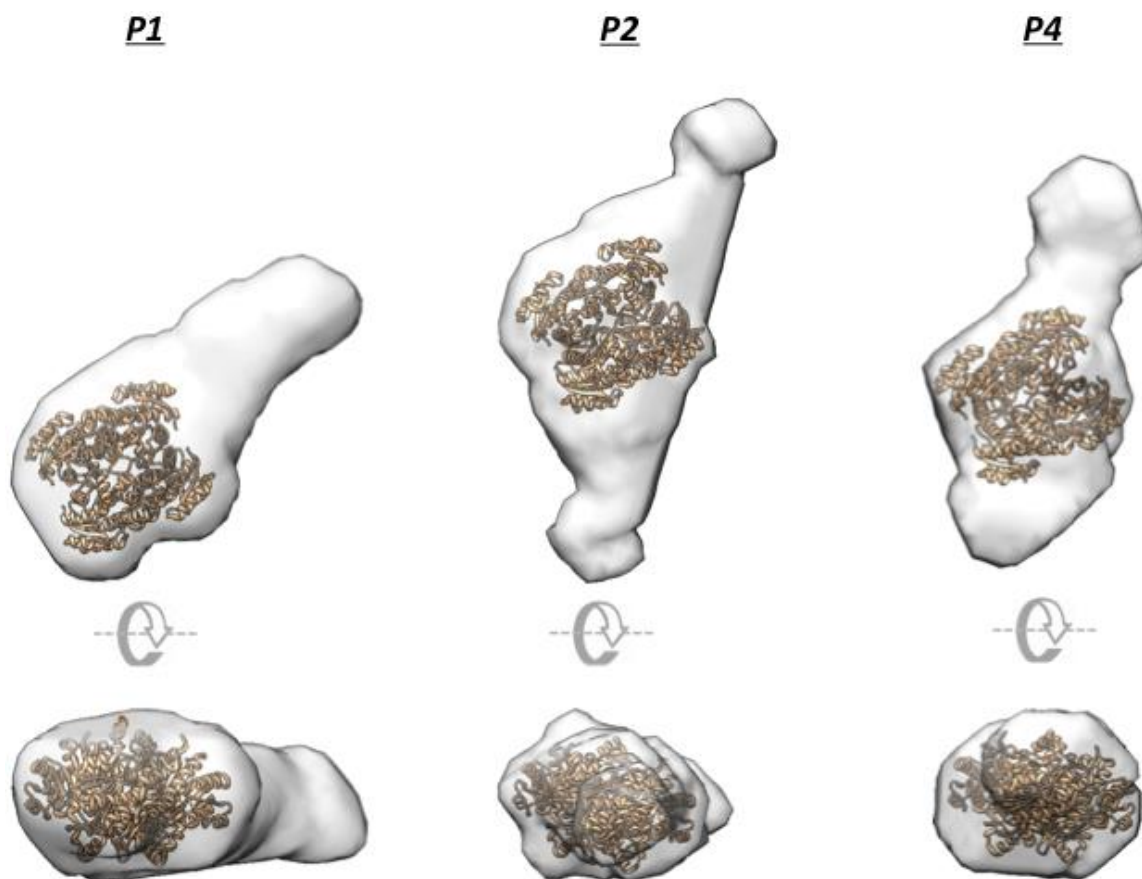


Figure 8-5. SAXS models of the MRPP1-MRPP2 complex generated with either no symmetry (*P1*) or with enforced *P2* or *P4* point symmetry. The crystal structure of tetrameric MRPP1 (PDB: 1U7T) was fitted into the SAXS model of MRPP1-MRPP2 using the fit-in-map function in the Chimera software [165].

The SAXS model of MRPP2 (**Figure 8-6, green**) was fitted with Chimera [165] to the SAXS models of MRPP1-MRPP2 (**Figure 8-6, red**) generated with *P1*, *P2* and *P4* symmetry. The SAXS model of mature MRPP1 (**Figure 8-6, grey**) was subsequently manually fitted to the vacant parts of all three SAXS pockets. The *P2*-model does not accommodate the dimer of mature MRPP1 on either side of the SAXS model of tetrameric MRPP2. The MRPP1 dimer is better accommodated in the *P4*-model, but leaves unaccountedfor parts of the SAXS pocket around the MRPP2 tetramer that does not fit with additional MRPP1 dimers. The *P1*-model mostly accommodates the MRPP1 dimer in this conformation, and does not leave unaccountedfor parts of the SAXS pocket.

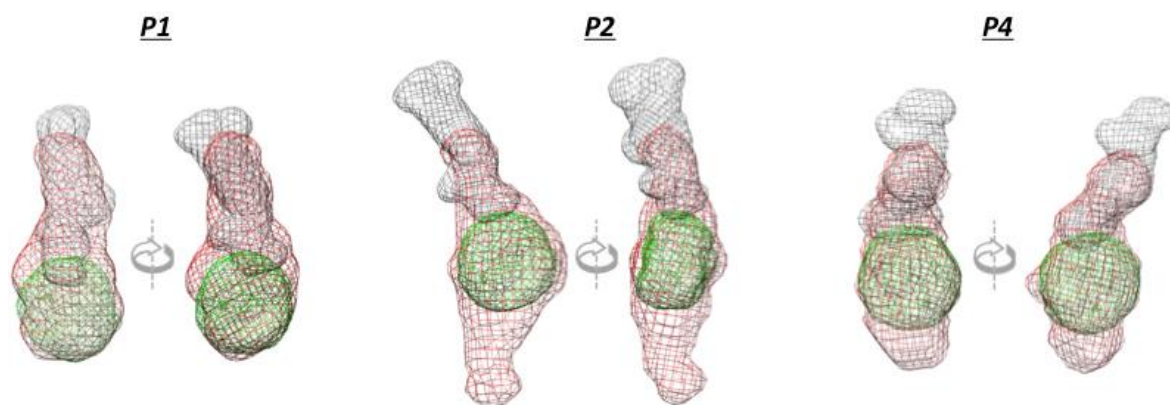


Figure 8-6. SAXS models of mature MRPP1 and full-length MRPP2 fitted into the SAXS model of MRPP1-MRPP2 built with *P1*, *P2* or *P4* point symmetry.