

Investigating Metabolite-RNase Communication

Carlanne Margaret Stone

School of Biological Sciences



UNIVERSITY *of* PORTSMOUTH

The thesis is submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of
Portsmouth

February 2017

Abstract

Maintaining cellular homeostasis involves a repertoire of intricate systems being able to respond to internal changes and environmental stimuli. Co-ordinating the process of post-transcriptional gene regulation is a number of ribonucleases, including polynucleotide phosphorylase (PNPase). PNPase controls steady-state transcript levels and thus regulates the production of various proteins, including enzymes involved in central metabolism. A feedback mechanism between central metabolism and RNA turnover has been previously suggested for the bacterium *Escherichia coli*. The Krebs cycle metabolite citrate was observed to modulate the activity of *E. coli* PNPase *in vitro* and *in vivo*. To discover whether such interactions are conserved across evolution, PNPase homologs from bacteria, eukarya and archaea were studied. Notably, citrate co-crystallises within the active site of *Homo sapiens* PNPase, suggesting that the citrate-PNPase communicative link may be conserved in eukaryotes.

In the current study, a combination of bioinformatics and *in silico* molecular docking approaches, show that citrate is predicted to bind PNPase and related exoribonucleolytic proteins, from diverse bacterial species, eukaryotic organelles and archaea. Furthermore, *in vitro* results suggested that PNPase, from another bacterial species *Synechocystis sp.*, may also be susceptible to inhibition/attenuation by citrate, and that this attenuation may therefore be commonplace amongst prokaryotes. Moreover, both eukaryotic PNPase from human mitochondria and the archaeal exosome complex from *Sulfolobus solfataricus*, is similarly inhibited/attenuated by citrate. The recurring interaction between citrate and PNPase homologs across all three domains, may represent an ancient and evolutionarily conserved mechanism of regulating RNA turnover.

Using the same *in silico* and *in vitro* approaches, the tricarboxylic acid (TCA) metabolites acetyl-CoA and succinyl-CoA were also shown to affect hPNPase and EcPNPase 3'-5' phosphorolytic activity. Results indicated that the nucleotide component of CoA in these metabolites, may bind and occlude the active site in a similar way to citrate. Accordingly, other nucleotide-based metabolites were investigated; phosphate-rich nucleotides and signalling molecules (GTP, ppppG, ppGpp) were predicted to bind to the active site of hPNPase. The results from gel-based assays then demonstrated that GTP, ppppG and ppGpp could affect the activity of both hPNPase and EcPNPase. It was also observed that the activity of hPNPase was more affected by these metabolites than EcPNPase and this was supported by previous research that suggested that PNPase homologs, across evolutionarily diverse organisms, have different phosphate preferences. Whether other PNPases can similarly interact with phosphate-rich nucleotides needs to be investigated. Likewise, the *in vivo* effects and physiological relevance of these metabolite-PNPase interactions remain to be discovered.

In summary, this study demonstrates that a metabolite-PNPase regulatory mechanism has the capacity to be conserved amongst all three domains of life and proposes that metabolite-mediated, post-transcriptional mechanisms are widespread. A system where central metabolism can influence RNA stability in a feedback loop, provides another tier of added complexity to the current hierarchal process governing the cellular flow of information. This mechanism potentially facilitates the fine-tuned response that is required to modify cellular functioning for adaptation and or survival. A greater understanding of the intricate network of interactions, occurring in cells, is invaluable for developing novel medical and biotechnological applications.

Contents

Abstract	i
Contents	iii
Declaration	x
List of Figures	xi
List of Tables	xiv
Abbreviations	xv
Acknowledgements and Dedications	xvii
Dissemination	xix
1 Introduction	1
1.1 Cellular Information Flow and Gene Regulation	1
1.2 Ribonucleases.....	2
1.3 RNA Turnover	3
1.3.1 RNA Turnover in Prokaryotes	3
1.3.2 RNA Turnover in Eukaryotes	4
1.3.3 RNA Turnover in Archaea	6
1.3.4 RNA Turnover Summary.....	6
1.4 RNase Complexes	6
1.4.1 RNA Degradosome	6
1.4.2 RNA Exosome.....	11
1.5 Polynucleotide Phosphorylase	11
1.5.1 PNPase Activity.....	11
1.5.2 PNPase Structure.....	12
1.5.3 Origin of the PNPase Domains	14
1.5.4 PNPase Structural Conservation	15
1.5.5 PNPase Roles	17
1.6 Regulating RNA Turnover in Response to Cellular Metabolism	18
1.6.1 PNPase Activity May Be Modulated by Citrate in E. coli	19
1.6.2 Is Citrate Binding Conserved in PNPases Homologs?	22
1.6.3 Conservation of Citrate Binding Residues in PNPase Homologs	24
1.6.4 Use of Techniques, Including Molecular Docking, in Ligand-Protein Interaction Research	27
1.6.5 Exploring PNPase Activity.....	29
1.6.6 TCA Metabolites Affecting PNPase	30

1.6.6.1	The TCA Cycle	30
1.6.6.2	Evolution of The TCA Cycle	31
1.6.7	Regulating RNA Turnover in Response to Cellular Metabolism Summary.....	35
1.7	Thesis Focus: Exploring the Impact of a Metabolite-PNPase Conserved Link	35
1.7.1	Thesis Aims.....	37
2	Materials and Methods.....	39
2.1	Chapter Aims.....	39
2.2	Chemical and Biochemical Reagents	39
2.3	Buffer and Media Recipes	39
2.4	Marker Proteins and Nucleic Acids	39
2.5	Antibiotics	39
2.6	Antibodies.....	40
2.7	Nucleic Acids: Deoxyribonucleic acid (DNA)	40
2.7.1	Provided Plasmid DNA.....	40
2.7.2	Chemical Synthesis: DNA	40
2.7.3	DNA Integrity and Concentration	41
2.7.4	DNA Sequencing.....	41
2.7.5	Site Directed Mutagenesis (SDM).....	41
2.8	Nucleic Acids: Ribonucleic acid (RNA)	44
2.8.1	Chemical Synthesis: RNA	44
2.8.2	RNA Integrity Concentration	45
2.9	Prokaryote Cell Strains	45
2.9.1	General Microbiological Techniques.....	45
2.9.1.1	Lysogeny Broth (LB) Agar & Plate Preparation	46
2.9.1.2	Lysogeny Broth (LB) Liquid Media Preparation	46
2.9.2	Cell Strains	46
2.9.3	Cell Strain Maintenance	47
2.9.4	Preparation of Chemically Competent E. coli Cell Strains	47
2.9.5	Transformation of Chemically Competent E. coli Cell Strains with Plasmid DNA. ...	47
2.9.6	Plasmid DNA Maintenance and Purification	48
2.10	Gel Electrophoresis	50
2.10.1	Agarose Gel Electrophoresis	51
2.10.2	Polyacrylamide Gel Electrophoresis (PAGE).....	52
2.10.2.1	Native-PAGE	52
2.10.2.2	Urea-PAGE	52
2.10.2.3	SDS-PAGE.....	53
2.11	Western Blot Analysis	54
2.12	Conclusion.....	55
3	Recombinant Protein Preparation	56
3.1	Introduction.....	56
3.1.1	Applications, Methods and Limitations of Protein Research.....	56

3.1.2	Recombinant Protein Technologies	57
3.1.2.1	Existing Strategies: Bacterial Expression Systems.....	57
3.1.2.2	Existing Strategies: Other Expression Systems.....	58
3.1.2.3	Strategies for Preparing PNPase and Archaeal Exosome.....	59
3.1.3	Summary and Aims.....	64
3.2	Methods.....	64
3.2.1	DNA Preparation	64
3.2.1.1	<i>H. sapiens</i> PNPase	64
3.2.1.2	<i>S. solfataricus</i> Exosome.....	64
3.2.1.3	<i>E. coli</i> PNPase	65
3.2.2	DNA Plasmid Transformation.....	65
3.2.3	Protein Expression: Small-Scale.....	65
3.2.4	Protein Expression: Large-Scale.....	66
3.2.5	Protein Purification	66
3.2.5.1	<i>E. coli</i> Cell Lysis	67
3.2.5.2	Isolation of Soluble Proteins.....	67
3.2.5.3	RNase & DNase Treatment	67
3.2.5.4	Heat Step & Centrifugation	67
3.2.5.5	Ammonium Sulphate Precipitation	67
3.2.5.6	Dialysis	67
3.2.5.7	Immobilised Metal Affinity Chromatography	67
3.2.5.8	Anion Exchange Chromatography	68
3.2.5.9	Size Exclusion Calibration.....	68
3.2.5.10	Protein Concentration.....	68
3.2.5.11	Size Exclusion Chromatography	69
3.2.6	Protein Purification: Protein-Specific.....	69
3.2.6.1	hPNPase	69
3.2.6.2	EcPNPase	69
3.2.6.3	SsoExosome	70
3.2.7	Protein Identification.....	70
3.2.7.1	SDS-PAGE and Western Blot.....	71
3.2.7.2	Size Exclusion Chromatography	71
3.2.7.3	Denaturing Mass Spectrometry (LC-MS/MS)	71
3.2.8	3'-5' RNA Degradation Activity of Recombinant Proteins	71
3.3	Results	72
3.3.1	Recombinant DNA Preparation.....	72
3.3.2	Protein Expression	73
3.3.3	Size Exclusion Chromatography Calibration Curve	76
3.3.4	Protein Purification: Protein-Specific.....	79
3.3.4.1	<i>Escherichia coli</i> PNPase	79
3.3.4.2	<i>Homo sapiens</i> PNPase	82
3.3.4.3	<i>Sulfolobus solfataricus</i> Exosome	86
3.3.5	Protein Identification.....	90

3.3.5.1	SDS-PAGE and Western Blot	90
3.3.5.2	Denaturing Mass Spectrometry (LC-MS/MS)	91
3.3.5.3	Non-Denaturing Mass Spectrometry	91
3.3.6	3'-5' RNA Degradation Activity of Recombinant Proteins	92
3.4	Conclusions	92
4	Inhibition of Homologous Phosphorolytic Ribonucleases by Citrate May Represent an Evolutionarily Conserved Communicative Link between RNA Degradation and Central Metabolism.....	94
4.1	Chapter Aims.....	94
4.2	Methods.....	94
4.2.1	Bioinformatics: Protein Sequence Alignments	95
4.2.2	In-Silico Molecular Docking.....	96
4.2.2.1	Structure Preparation	96
4.2.2.2	In-Silico Citrate Docking	97
4.2.2.3	Protein Ligand Interaction Fingerprints (PLIF) and Heat Map Generation	98
4.2.3	X-Ray Crystallography	99
4.2.4	Gel-Based In Vitro RNA Degradation Assays.....	100
4.3	Results	100
4.3.1	Investigating Citrate-PNPase Interactions	101
4.3.1.1	E. coli and H. sapiens Structures	101
4.3.1.2	Structural Conservation of PNPase Homologs	109
4.3.2	Citrate-PNPase Interactions Are Conserved in all Domains of Life	115
4.3.2.1	Active Site Sequence Alignments	115
4.3.2.2	Vestigial Site Sequence Alignment	118
4.3.3	Citrate is Predicted to Dock In Silico into Prokaryote, Eukaryote and Archaeal PNPase Homologs	119
4.3.3.1	Candidate Selection for In Silico Docking	119
4.3.3.2	Protein Structure Preparation: Magnesium Placement.....	119
4.3.3.3	Protein Structure Preparation: Active Site Finder	126
4.3.3.4	Active Site In-Silico Citrate Docking.....	132
4.3.4	Crystals Diffract by X-Ray Crystallography and Require Further Optimisation	139
4.3.5	Modulation of PNPase Homologs' Exoribonuclease Activity by Citrate.....	143
4.3.6	Sequence Analysis of Citrate Binding Residues	146
4.4	Conclusion.....	150
5	Metabolites Other Than Citrate May Modulate the Phosphorolytic Activity of Polynucleotide Phosphorylase from Homo sapiens	152
5.1	Chapter Aims.....	153
5.2	Methods.....	153
5.2.1	In-Silico Molecular Docking.....	153
5.2.1.1	In-Silico Metabolite Database Docking	153
5.2.1.2	Protein Ligand Interaction Fingerprints (PLIF) and Heat Map Generation	155

5.2.1.3	Pharmacophore Generation.....	156
5.2.2	Gel-Based Degradation Assays.....	156
5.3	Results.....	156
5.3.1	Some TCA Metabolites Are Predicted In Silico to Dock into PNPase from <i>H. sapiens</i> 157	
5.3.2	Some TCA Metabolites Inhibit <i>H. sapiens</i> PNPase Degradation Activity In Vitro ..	162
5.3.3	Correlation of In Silico Docking Calculations and In Vitro Degradation Activity	164
5.3.4	Some TCA Metabolites Inhibit <i>E. coli</i> PNPase Degradation Activity In Vitro.....	166
5.3.5	Predicting Metabolite Features That May Be Important for Interactions with PNPase from <i>H. Sapiens</i>	167
5.3.6	Phosphate-Rich Metabolites Are Predicted to Dock In Silico into PNPase from <i>H.</i> <i>Sapiens</i> 170	
5.3.7	Phosphate-Rich Metabolites Inhibit <i>H. sapiens</i> and <i>E. coli</i> PNPase Degradation Activity In Vitro.....	173
5.4	Conclusion.....	175
6	Developing a High-Throughput 3'-5' RNase Kinetic Assay	180
6.1	Introduction.....	180
6.1.1	3'-5' RNase Catalysis.....	181
6.1.2	Enzyme Kinetics Equations.....	181
6.1.3	Determining 3'-5' RNase Kinetics	184
6.1.3.1	Plate Reader Set-up and Optimisation.....	186
6.1.3.2	Optimisation of Assay Conditions	186
6.2	Aims	187
6.3	Methods.....	187
6.3.1	Instrument Set-up.....	188
6.3.2	Determining Detection Conditions (Q-Ab & ADP*)	188
6.3.3	ADP Standard Curve.....	188
6.3.4	Enzyme Titration	189
6.3.5	RNA Substrate Titration	190
6.4	Results	191
6.4.1	Instrument Set-up Optimisation.....	191
6.4.2	Optimising Q-Ab & ADP* Detection Conditions	192
6.4.2.1	Bellbrook Recommended Conditions for End-Point ADP Detection.....	192
6.4.2.2	Optimal Conditions for Real-Time ADP Detection	192
6.4.3	Determining Optimal Plate Parameters	195
6.4.4	ADP Standard Curve.....	199
6.4.5	<i>Ec</i> PNPase Michaelis-Menten Kinetics.....	202
6.4.5.1	Factors Affecting the <i>Ec</i> PNPase RFU Window	202
6.4.5.2	<i>Ec</i> PNPase Titration.....	204
6.4.5.3	<i>Ec</i> PNPase Degradation Activity (0.5 mM Phosphate).....	209
6.4.5.4	<i>Ec</i> PNPase Degradation Activity (10 mM Phosphate).....	212
6.4.5.5	<i>Ec</i> PNPase Degradation Activity (5 mM Phosphate).....	217

6.4.5.6	EcPNPase Michaelis-Menten Kinetics Summary	220
6.4.6	SsoExosome Michaelis-Menten Kinetics	222
6.4.6.1	Factors Affecting the SsoExosome RFU Window.....	222
6.4.6.2	SsoExosome Titration	224
6.4.6.3	SsoExosome Degradation Activity (0.5 mM Phosphate)	229
6.5	Conclusions.....	233
7	Discussion and Future Perspectives	236
7.1	Regulating RNA Turnover in Response to Cellular Metabolism	237
7.1.1	Interactions Between Citrate and Homologous 3'-5' Phosphorolytic Ribonucleases 237	
7.1.2	Interactions Between TCA Metabolites and 3'-5' Phosphorolytic Ribonucleases ..	239
7.1.3	Interactions Between Phosphate-Rich Metabolites and 3'-5' Phosphorolytic Ribonucleases	241
7.2	Future Perspectives	244
7.2.1	Discovering Novel Metabolite-RNase Interactions.....	244
7.2.2	Examining Metabolite-RNase Interactions In Vivo.....	245
7.3	Summary	247
8	References	248
9	Appendix	260
9.1	Certificate of Ethics Review	260
9.2	Gel Electrophoresis Ladders.....	261
9.3	Primer Sequences.....	264
9.3.1	Mutagenesis Primers	264
9.3.2	Sequencing Primers.....	265
9.4	Cell Strain Genotypes	265
9.5	Vector Maps	267
9.6	Gene Cloning Strategy.....	269
9.6.1	Vector Sequence: pET-28b (+) (Reverse Complement).....	269
9.6.2	Vector Sequence: pET-MCN-EAVNH.....	271
9.6.3	Homo sapiens Polynucleotide Phosphorylase (hPNPase)	274
9.6.3.1	Amino Acid Sequence: hPNPase.....	274
9.6.3.2	pET-28b [hPNPase] Transcript	275
9.6.3.3	pET-28b [hPNPase] Plasmid Map	276
9.6.3.4	pET-28b [hPNPase] Translation	277
9.6.3.5	ProtParam Output: H ₆ -hPNPase.....	277
9.6.4	Sulfolobus solfataricus Exosome (SsoExo4_41_42)	277
9.6.4.1	Amino Acid Sequence: SsoExo4_41_42	277
9.6.4.2	pET-MCN-EAVNH [SsoExo4_41_42] Transcript	278
9.6.4.3	pET-MCN-EAVNH [SsoExo4_41_42] Plasmid Map	279
9.6.4.4	pET-MCN-EAVNH [SsoExo4_41_42] Translation	280

9.6.4.5	ProtParam Output: H ₆ -SsoExo4.....	281
9.6.4.6	ProtParam Output: SsoExo41.....	281
9.6.4.7	ProtParam Output: SsoExo42.....	281
9.6.4.8	ProtParam Output: H ₆ - SsoExo4_41_42.....	281
9.6.5	E. coli Polynucleotide Phosphorylase (EcPNPase)	282
9.6.5.1	Amino Acid Sequence: EcPNPase	282
9.6.5.2	ProtParam Output: EcPNPase.....	282
9.7	PACT Premier™ Crystallography Screen.....	283
9.8	Protein Accession Codes	285
9.9	Metabolite Structures	285

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Signed

A handwritten signature in black ink, appearing to read 'Carlanne', with a long horizontal flourish extending to the right.

Carlanne Margaret Stone

Word count: <77,000

List of Figures

Figure 1.1 Cellular Information Flow: A Complex Regulatory System.....	1
Figure 1.2 RNase-Mediated RNA Degradation.....	3
Figure 1.3 Eukaryotic mRNA Decay	5
Figure 1.4 RNA Degradosome Complex	7
Figure 1.5 RNA Degradosome Evolution.....	9
Figure 1.6 PNPase Activity	12
Figure 1.7 PNPase Domain Organisation and Crystal Structure.....	13
Figure 1.8 PNPase Domain Evolution	15
Figure 1.9 PNPase and Exosome Structural Homology	16
Figure 1.10 EcPNPase Crystal Structure.....	20
Figure 1.11 Structural Conservation of 3'-5' PNPase Homologs.....	26
Figure 1.12 TCA Cycle.....	31
Figure 1.13 TCA Cycle Evolution.....	34
Figure 1.14 Cellular Information Flow: Metabolite-RNase Communication.....	36
Figure 2.1 Overview of Site-Directed Mutagenesis (SDM)	42
Figure 2.2 Fluorescein Structure.....	45
Figure 2.3 Plasmid DNA Purification.....	50
Figure 3.1 EcPNPase, hPNPase and SsoExosome Protein Expression Summary	74
Figure 3.2 SEC Calibration	77
Figure 3.3 EcPNPase AEC	80
Figure 3.4 EcPNPase SEC	81
Figure 3.5 EcPNPase Purification Stages	82
Figure 3.6 hPNPase IMAC.....	84
Figure 3.7 hPNPase SEC	85
Figure 3.8 hPNPase Purification Stages.....	86
Figure 3.9 SsoExosome IMAC.....	88
Figure 3.10 SsoExosome SEC	89
Figure 3.11 SsoExosome Purification Stages.....	90
Figure 3.12 Protein Purification Summary	91
Figure 3.13 Recombinant Protein Degradation Activity Summary	92
Figure 4.1 MOE Flow Diagram	96
Figure 4.2 PCT Recommended Actions	99
Figure 4.3 EcPNPase Active Site Citrate Interactions	102
Figure 4.4 EcPNPase Active Site Citrate PLIF	104
Figure 4.5 EcPNPase Vestigial Site Citrate Interactions	106
Figure 4.6 hPNPase Active Site Citrate Interactions	108
Figure 4.7 Structural Conservation of PNPase Homologs.....	110
Figure 4.8 PNPase Homologs PH-1/Rrp41 Domain Conservation.....	111
Figure 4.9 PNPase Homologs PH-1/Rrp41 Domain Variations (part 1 of 2).....	112
Figure 4.10 PNPase Homologs PH-1/Rrp41 Binding Motifs Conservation	114
Figure 4.11 PNPase Homologs' Binding Motifs Sequence Conservation	117
Figure 4.12 PNPase Vestigial Site Sequence Alignment	118

Figure 4.13 EcPNPase Structure Preparation	121
Figure 4.14 EcPNPase Mn ²⁺ Coordination	122
Figure 4.15 EcPNPase Structure Preparation; Mg ²⁺ Placement	123
Figure 4.16 PNPase Homologs Structure Preparation; Mg ²⁺ Placement	125
Figure 4.17 SsoExosome Structure Preparation; Mg ²⁺ Placement	126
Figure 4.18 EcPNPase Structure Preparation; Site Finder.....	128
Figure 4.19 hPNPase Structure Preparation; Site Finder.....	129
Figure 4.20 SsoExosome Structure Preparation; Site Finder.....	130
Figure 4.21 SanPNPase Structure Preparation; Site Finder and Site Selection Conservation.....	130
Figure 4.22 Citrate 2D Structure	132
Figure 4.23 Active Site Citrate Docking	134
Figure 4.24 PNPase Homologs Active Site Citrate Docking; All Hits S Values.....	135
Figure 4.25 Active Site Citrate Docking Overlay.....	136
Figure 4.26 PNPase Homologs Active Site Citrate Docking; Heat Map.....	138
Figure 4.27 SsoExosome Crystallisation Trials; PCT Test.....	140
Figure 4.28 Crystals of SsoExosome + Citrate	141
Figure 4.29 SsoExosome and Lysozyme X-Ray Crystallography Diffraction.....	142
Figure 4.30 3'-5' Degradation Assay; 5'FAM(A) ₂₀ RNA Substrate Concentration Range.....	143
Figure 4.31 3'-5' Degradation Assay; hPNPase Concentration Range	144
Figure 4.32 3'-5' Degradation Assay; <i>In Vitro</i> Inhibition of PNPase Homologs by Citrate	145
Figure 4.33 Conservation of PNPase Homologs Binding Motifs Sequence Logos	148
Figure 4.34 PNPase Homologs Vestigial Site Sequence Logos	149
Figure 4.35 Citrate-PNPase Communication Summary	151
Figure 5.1 <i>In Silico</i> Docking and <i>In Vitro</i> Degradation Activity of hPNPase with TCA Metabolites	158
Figure 5.2 Lowest Predicted Binding Conformation of Citrate, Acetyl-CoA and Succinyl-CoA in the hPNPase Active Site.....	160
Figure 5.3 Metabolite-hPNPase Heat Map	161
Figure 5.4 <i>In Vitro</i> Degradation Activity of hPNPase with TCA Metabolites.....	163
Figure 5.5 <i>In Silico</i> Docking Vs <i>In Vitro</i> Inhibition of hPNPase by TCA Metabolites	165
Figure 5.6 3'-5' Degradation Assay; <i>In Vitro</i> Inhibition of EcPNPase Homologs by Citrate, Acetyl-CoA and Succinyl-CoA.....	166
Figure 5.7 hPNPase Pharmacophore Generation	168
Figure 5.8 hPNPase Pharmacophore: Phosphate Positioning	170
Figure 5.9 <i>In Silico</i> Docking of hPNPase with Phosphate-Rich Metabolites	171
Figure 5.10 <i>In Vitro</i> Degradation Activity of hPNPase and EcPNPase with Phosphate-Rich Metabolites GTP, ppppG and ppGpp.....	174
Figure 5.11 <i>In Vitro</i> Degradation Activity of PNPase Homologs with ppGpp	178
Figure 6.1 PNPase Activity	183
Figure 6.2 ADP Detection Plate Assay	185
Figure 6.3 Minitab Logistic Growth Function.....	189
Figure 6.4 Plate Optimisation: ADP* & Q-Ab Concentration Range.....	194
Figure 6.5 Optimising ADP* Quenching by Q-Ab.	196
Figure 6.6 Assay Window Data.....	198

Figure 6.7 Assay Window Fit	199
Figure 6.8 ADP _{20x} Standard Curve.....	200
Figure 6.9 RFU Window with EcPNPase Assay Components	203
Figure 6.10 EcPNPase Titration: RFU vs Time	205
Figure 6.11 EcPNPase Titration: [ADP _{20x}] vs Time.....	206
Figure 6.12 EcPNPase Titration: Linear Fit [ADP _{20x}] vs Time	207
Figure 6.13 EcPNPase Titration: Rate of ADP _{20x} Production vs [EcPNPase]	208
Figure 6.14 RNA _{20mer} Degradation by EcPNPase (0.5 mM Phosphate): RFU vs Time.....	210
Figure 6.15 RNA _{20mer} Degradation by EcPNPase (0.5 mM Phosphate): [ADP _{20x} (nM)] vs Time.....	211
Figure 6.16 RNA _{20mer} Degradation by EcPNPase (0.5 mM Phosphate): Michaelis-Menten	212
Figure 6.17 RNA _{20mer} Degradation by EcPNPase (10 mM Phosphate): RFU vs Time.....	214
Figure 6.18 RNA _{20mer} Degradation by EcPNPase (10 mM Phosphate): [ADP _{20x} (nM)] vs Time.....	215
Figure 6.19 RNA _{20mer} Degradation by EcPNPase (10 mM Phosphate): Michaelis-Menten	216
Figure 6.20 RNA _{20mer} Degradation by EcPNPase (5 mM Phosphate): RFU vs Time.....	218
Figure 6.21 RNA _{20mer} Degradation by EcPNPase (5 mM Phosphate): [ADP _{20x} (nM)] vs Time.....	219
Figure 6.22 RNA _{20mer} Degradation by EcPNPase (5 mM Phosphate): Michaelis-Menten	220
Figure 6.23 RFU Window with SsoExosome Assay Components.....	223
Figure 6.24 SsoExosome Titration: RFU vs Time	225
Figure 6.25 SsoExosome Titration: Raw Data [ADP _{20x}] vs Time.....	226
Figure 6.26 SsoExosome Titration: Linear Fit [ADP _{20x}] vs Time.....	227
Figure 6.27 SsoExosome Titration: Rate of ADP _{20x} Production vs [SsoExosome]	228
Figure 6.28 RNA _{20mer} Degradation by SsoExosome (0.5 mM Phosphate): [ADP _{20x} (nM)] vs Time.....	230
Figure 6.29 RNA _{20mer} Degradation by SsoExosome (0.5 mM Phosphate): [ADP _{20x} (nM)] vs Time.....	231
Figure 6.30 RNA _{20mer} Degradation by SsoExosome (0.5 mM Phosphate): Michaelis-Menten.....	232
Figure 7.1 Control Points and Metabolites Regulating The TCA Cycle	240
Figure 7.2 (p)ppGpp Metabolism in <i>E. coli</i>	243
Figure 9.1 100 bp DNA Ladder	261
Figure 9.2 Low Molecular Weight (LMW) DNA Ladder.....	262
Figure 9.3 Low Range ssRNA Ladder	262
Figure 9.4 SeeBlue® Pre-Stained Standard Protein Molecular Weight Ladder	263
Figure 9.5 SeeBlue® Plus2 Pre-Stained Standard Protein Molecular Weight Ladder.....	263
Figure 9.6 BenchMark™ Standard Protein Ladder.....	264
Figure 9.7 <i>E. coli</i> Cell Strain Genotype.....	266
Figure 9.8 pET-28a-c (+) Vector Map	267
Figure 9.9 pETMCN-EAVNH Vector Map	268
Figure 9.10 pETDuet-1 Vector Map	268
Figure 9.11 hPNPase Plasmid Map.....	276
Figure 9.12 SsoExosome Plasmid Map.....	280
Figure 9.13 Metabolite 2D Structures	286
Figure 9.14 Metabolite 2D Structures Continued.....	287

List of Tables

Table 2.1 Antibiotic Preparation	40
Table 2.2 <i>In vitro</i> SDM Reaction Preparation	43
Table 2.3 QuikChange ® II XL SDM Cycling Parameters	44
Table 2.4 Antibiotic Selection	46
Table 2.5 Transformation Plates	48
Table 2.6 Agarose Gel Preparation	51
Table 2.7 Native-PAGE Preparation	52
Table 2.8 Denaturing Urea-PAGE Preparation	53
Table 2.9 SDS-PAGE Preparation	54
Table 3.1 Protein-Specific Expression Conditions	75
Table 3.2 SEC Protein Standards	78
Table 4.1 Protein Accession Codes	95
Table 4.2 RMSD Comparison of PDB vs Docking Structure	124
Table 4.3 MOE Site Selection	131
Table 4.4 PNPase Homologs Active Site Citrate Docking; Top Hits S Value	136
Table 4.5 3'-5' Degradation Assay; Quantifying the <i>In Vitro</i> Inhibition of PNPase Homologs by Citrate	146
Table 5.1 Metabolite Database Identification Codes	155
Table 5.2 Degradation Assay; Quantifying the <i>In Vitro</i> Inhibition of hPNPase by TCA Metabolites	164
Table 5.3 Degradation Assay; Quantifying the <i>In Vitro</i> Inhibition of EcPNPase by Citrate, Acetyl-CoA and Succinyl-CoA	167
Table 5.4 Degradation Assay; Quantifying the <i>In Vitro</i> Inhibition of hPNPase and EcPNPase by Phosphate-Rich Metabolites	174
Table 6.1 ADP _{20x} Standard Accuracy	201
Table 6.2 RFU Window with EcPNPase Assay Components	204
Table 6.3 EcPNPase Kinetics Summary	221
Table 6.4 RFU Window with SsoExosome Assay Components	224
Table 6.5 SsoExosome Kinetic Parameters	232
Table 6.6 PNPase Homologs Kinetics Summary	233
Table 9.1 PACT Premier™ Crystallography Screen	284
Table 9.2 Protein Accession Codes	285

Abbreviations

A	Adenosine
ADP* or (A*)	ADP Alexa594 Tracer (ADP*)
Q-Ab	ADP2 Antibody-IRDye® QC-1 (Q-Ab) (mouse monoclonal)
α	alpha
aa	amino acid
NTR/NTD	Amino Terminal Region/Domain
APS	Ammonium persulfate
Amp	Ampicillin
Å	Ångström (s)
TM	Annealing Temperature
bp	Base Pair (s)
β	beta
CaCl ₂ .2H ₂ O	Calcium Chloride Dihydrate
CTR/CTD	Carboxyl Terminal Region/Domain
Δ	Change in
CAM	Chloramphenicol
C	Cytidine
Da	Dalton
°C	Degree Celsius
dH ₂ O	Deionised water
dNTP	Deoxynucleic triphosphate
DNA	Deoxyribonucleic acid
Kd	Dissociation constant
DTT	Dithiothreitol
ds	double stranded
<i>E. coli</i> and EcPNPase	<i>Escherichia coli</i> PNPase
EtOH	Ethanol
EtBr	Ethidium Bromide
EDTA	Ethylenediaminetetraacetic acid
ϵ	extinction coefficient
FI	Fluorescent Intensity
γ	gamma
g	gram (s)
G	Guanosine
H ₆ or His ₆ -Tag	Histidine tag (s)
<i>H. Sapiens</i> and hPNPase	<i>Homo Sapiens</i> PNPase
HRP	Horse Radish Peroxidase
h	Hour (s)
HCl	Hydrochloric Acid
H	Hydrogen
OH	Hydroxyl
V _o	Initial Rate
pI	Isoelectric point
IPTG	Isopropyl- β -D-thiogalactopyranoside
Kan	Kanamycin
k	kilo
l or L	litre
LB	Lysogeny Broth
MgCl ₂	Magnesium chloride
MgCl ₂ .6H ₂ O	Magnesium Chloride Hexahydrate

MnCl ₂ ·4H ₂ O	Manganese(II) Chloride Tetrahydrate
V _{max}	Maximal Velocity
mRNA	messenger ribonucleic acid
<i>M. thermotrophicus</i> and MthExosome	<i>Methanothermobacter thermotrophicus</i> Exosome
m	Metre (s)
K _m	Michaelis Constant
μ	Micro
μM	Micro molar
μg	Microgram (s)
μl	Microliter (s)
ml	millilitre
m	milli
mAU	milli absorbance unit (s)
min/sec	Minute/Second
M	Molar
MW	Molecular Weight
TEMED	N, N, N', N' - Tetramethylethylenediamine
n	nano
NDP	Nucleoside di-phosphate
NTP	Nucleoside tri-phosphate
nt	Nucleotide (s)
RNA _{20mer}	Poly(A) 20 mer RNA substrate
OD ₆₀₀	Optical Density at 600 nm
%	Percentage
P	Phosphate (s)
PCR	Polymerase Chain Reaction
KOAc	Potassium acetate
KCl	Potassium chloride
psi	<i>pound-force per square inch</i>
PDB	Protein Data Bank
RFU	Relative Fluorescence Units
RNA	Ribonucleic acid
RT	Room Temperature
s	Second (s)
ss	single stranded
sRNA	small ribonucleic acid
NaCl ₂	Sodium chloride
SDS	Sodium Dodecyl Sulphate
<i>S. antibioticus</i> and SanPNPase	<i>Streptomyces antibioticus</i> PNPase
<i>S. solfataricus</i> and SsoExosome	<i>Sulfolobus solfataricus</i> exosome
<i>S. sp</i> and SspPNPase	<i>Synechocystis sp</i> PNPase
[E _T]	Total Enzyme Concentration
K _{cat}	Turnover Rate
ADP _{20x}	Twenty ADP molecules
U	Uridine
(v/v)	volume per volume
λ	Wavelength
w/v	weight/volume
σ	sigma
TBE	Tris, Boric acid and EDTA
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
X-gal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

I would like to dedicate this thesis to my family

Acknowledgements

Writing these acknowledgments was incredibly hard as the immense amount of support and guidance that I have received throughout my PhD can't be easily summarised. I am forever grateful to all my friends, family and colleagues who have helped, mentored and motivated me during this challenging, but rewarding, time of my life. I could not have completed this work, and thesis, without your generous contributions.

Firstly, I would like to thank my supervisor Professor Anastasia Callaghan. During my undergraduate degree, you kindly and patiently helped me to develop the skills I needed to be a researcher. The encouragement and the enjoyment that I experienced whilst working in the Callaghan group inspired me to pursue a PhD. Since then, you have provided guidance on this research topic, ensured continuous funding and promoted development opportunities, and for this I will always be grateful. Thanks to the Institute of Biomedical and Biomolecular Science for funding my PhD studies. I would also like to make a special mention to my second supervisor Professor Matt Guille. I am extremely thankful for your help proof reading this thesis and for all your advice.

To all Callaghan group members, past and present, you have been amazing people to work with. Jack, Charlotte and Helen thank you for answering all my questions and teaching me during those early days. To the PNPase team, Louise, Josh, Daniel and Helen, I appreciate all the hard work everyone put in to conduct and publish our work. Charlotte Mardle/Caddick, you have been an incredible friend and have always been there for scientific and life advice! Louise Butt-Stone Senior, I am really thankful for all your support over my PhD, especially for reading all my thesis drafts. I have enjoyed the coffee and cake meetings and the healthier swimming trips with you both. Pete, Martin and Rich, thanks for helping to make my research run as smoothly as possible. Andy and Darren, I have really appreciated all your scientific advice and unwavering support. I also extend my gratitude to all the members of the Biophysics Department, thanks for making my studies so memorable!

Once more, thank you to all my amazing friends and family. Shauna and Megan thanks for listening and for being incredible friends. Nicky, Chris and Georgia I have enjoyed our time completing charity challenges each year. To my grandparents Jeanette, Iris and Jim, you have generously provided your wisdom and support throughout the years. Thanks for fixing broken laptops, the great holidays and always being at the end of the phone. Matthew, you have been extremely thoughtful and caring, and you have kept me very happy in this time of my life (looking forward to celebrating finishing my PhD with you in Budapest!). To my little sister Molly, you have always made me smile and laugh when I have needed it the most. Finally, to my wonderful Mum and Dad, you have always encouraged me to work hard and achieve my goals. Whenever I have needed help and advice you have always been there for me. Through the tough times you have taught me to be resilient and courageous and have given me plenty to look forward to with all our exciting family holidays. I could never thank you enough for all the incredible things you do!

Dissemination

Publications:

Stone, C.M., Butt, L.E., Bufton, J.C., Lourenco, D.C., Gowers, D.M., Pickford, A.R., Cox, P.A., Vincent, H.A., and Callaghan, A.J. (2017). Inhibition of homologous phosphorolytic ribonucleases by citrate may represent an evolutionarily conserved communicative link between RNA degradation and central metabolism. *Nucleic Acids Res*, 45(8), 4655-4666. Doi: 10.1093/nar/gkx114

Other publications are currently in preparation.

Oral presentations:

Investigating citrate-PNPase communication: a focus on mechanism conservation

1. Presented by C. Stone at the South West Structural Biology Consortium Meeting (SWSBC), University of Portsmouth (2016). This presentation received a commendation.
2. Presented by C. Stone at the Institute of Biomolecular and Biomedical Sciences (IBBS) conference, University of Portsmouth (2016). This presentation was awarded first prize.

Poster presentations:

Investigating citrate-PNPase communication: a focus on mechanism conservation

1. Presented by C. Stone at the Institute of Biomolecular and Biomedical Sciences (IBBS) conference, University of Portsmouth (2015).
2. Presented by C. Stone at the Graduate School Research Conference, University of Portsmouth (2016). Awarded first prize.
3. Presented by C. Stone at the Post-Transcriptional Control of Gene Expression: mRNA Decay Conference, Federation of American Societies for Experimental Biology (FASEB) Conference, Lisbon Portugal (2016). Awarded second prize

Grants awarded for research dissemination:

Grants which funded attendance at the 2016 FASEB mRNA decay conference in Lisbon, Portugal

1. Biochemical Society General Travel Grant, awarded £450.00
2. University of Portsmouth, Science Faculty, Research Students Conference Bursary, awarded £450.00

1 Introduction

1.1 Cellular Information Flow and Gene Regulation

It is well established that the cellular information flow in all domains of life, involves deoxyribonucleic acid (DNA) being transcribed into ribonucleic acid (RNA) before being translated into proteins (Figure 1.1 (a-c)). As synthesised proteins can include the enzymes responsible for creating metabolites, the central dogma can be further expanded to include central metabolism (Figure 1.1 (c-d)). These metabolites are also known to regulate gene expression and therefore the system can be seen to be comprised of a complex and dynamic feedback loop.

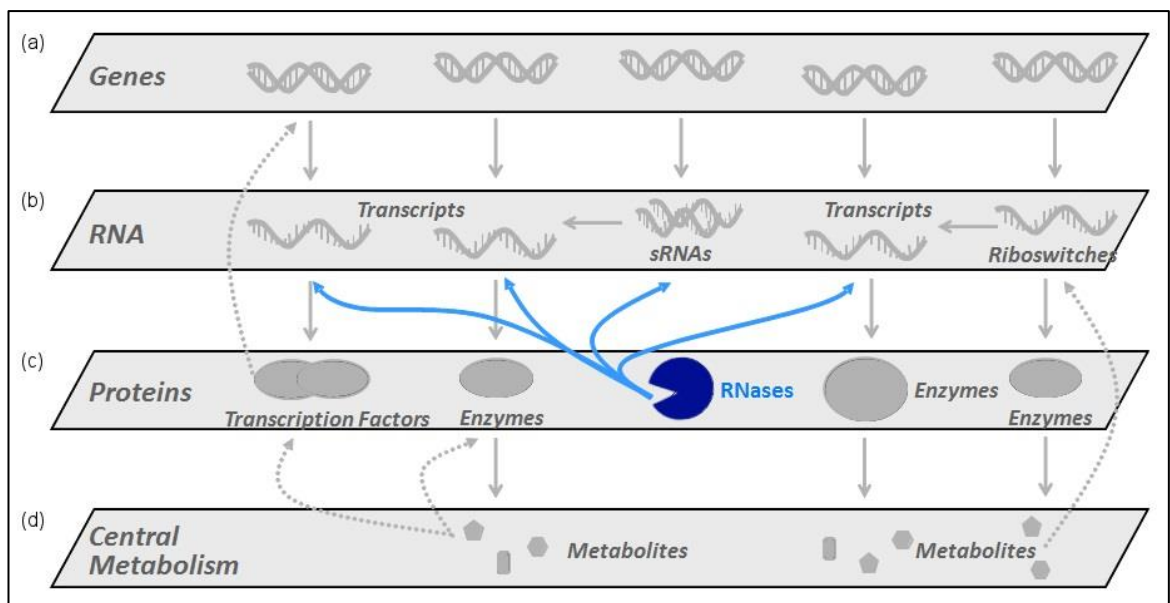


Figure 1.1 Cellular Information Flow: A Complex Regulatory System

The flow of cellular information shows that (a) DNA genes can be transcribed into (b) transcripts including messenger RNA (mRNA), small non-coding RNAs (sRNAs), long non-coding RNAs and riboswitches. The latter two can post-transcriptionally control gene expression (side arrow) thus regulating the translation of the transcript into protein. (c) Furthermore, synthesised proteins can include transcription factors and ribonucleases (RNases-blue), which can regulate gene expression of the DNA and RNA level, respectively. The collection of proteins in the cell (proteome) controls (d) central metabolism, generating metabolites which can then relay back and control transcription factors and riboswitches (dotted lines). Image adapted from Prof. Anastasia Callaghan.

Since the discovery of DNA, there has been a plethora of research published which demonstrates that at each stage of cellular information flow, shown in Figure 1.1, tight regulation occurs. There have also been reports of detrimental effects on cellular function when information is not transferred in an efficient and correct manner. This has revealed the importance of regulating gene expression, not only for normal cellular growth and differentiation, but also cellular senescence.

Precise control of transcript levels relies on both transcription and degradation rates. Specifically, controlling the level of RNA turnover is vital for normal cellular homeostasis, since the pool of RNA available effectively dictates both the cellular pool of proteins and metabolites, termed the proteome and metabolome respectively (Figure 1.1 (c)-(d) respectively). Therefore, by altering

RNA turnover, a cell is able to respond to temporal cellular requirements and or environmental changes. Acting together with DNA transcription control, RNA turnover regulation facilitates the essential adaptations that the cell needs to survive (Clements *et al.*, 2002; Ibrahim, Wilusz, & Wilusz, 2008; Nurmohamed *et al.*, 2011; Witharana, Roppelt, Lochnit, Klug, & Evguenieva-Hackenberg, 2012). Extensive research has revealed that ribonuclease enzymes (RNases) have an important and broad role in regulating gene expression at the RNA level and this is highlighted Figure 1.1 (blue RNase) and reviewed subsequently in Section 1.2.

1.2 Ribonucleases

RNases are ubiquitous enzymes that play central roles in RNA metabolism and control turnover across all domains of life. They are required for the control of gene expression, primarily through the degradation of mRNAs, the processing and degradation of regulatory RNAs, and for the maturation and quality control of stable RNAs (reviewed in Arraiano *et al.*, 2010). However, they also have the potential to be incredibly destructive and consequently their activity is carefully regulated (reviewed in Deutscher, 2015).

Across evolutionarily divergent organisms, in all domains of life, there are a wide variety of RNA substrates. As a result, there are also a number of different degradation mechanisms available and this is reflected by the large size and functional diversity of the RNase enzyme group (Ibrahim *et al.*, 2008; Nurmohamed *et al.*, 2011; Zuo & Deutscher, 2001). Nevertheless, the general ways in which RNases work are highly conserved. Collectively, RNases either degrade RNA by endoribonuclease or exoribonuclease activity; essentially breaking phosphodiester bonds within the RNA molecule internally, or by removing nucleotides from the 5' or 3' ends respectively (Figure 1.2) (Deutscher, 1993). These RNA degradation processes result in the overall destruction of transcripts and release of nucleotide diphosphates for phosphorolytic mechanism (NDPs, red segments (Figure 1.2)) or monophosphate for hydrolytic mechanisms, from RNA that is no longer needed or damaged, thus providing more nucleotides that can be reutilised for RNA synthesis (Deutscher, 1993).

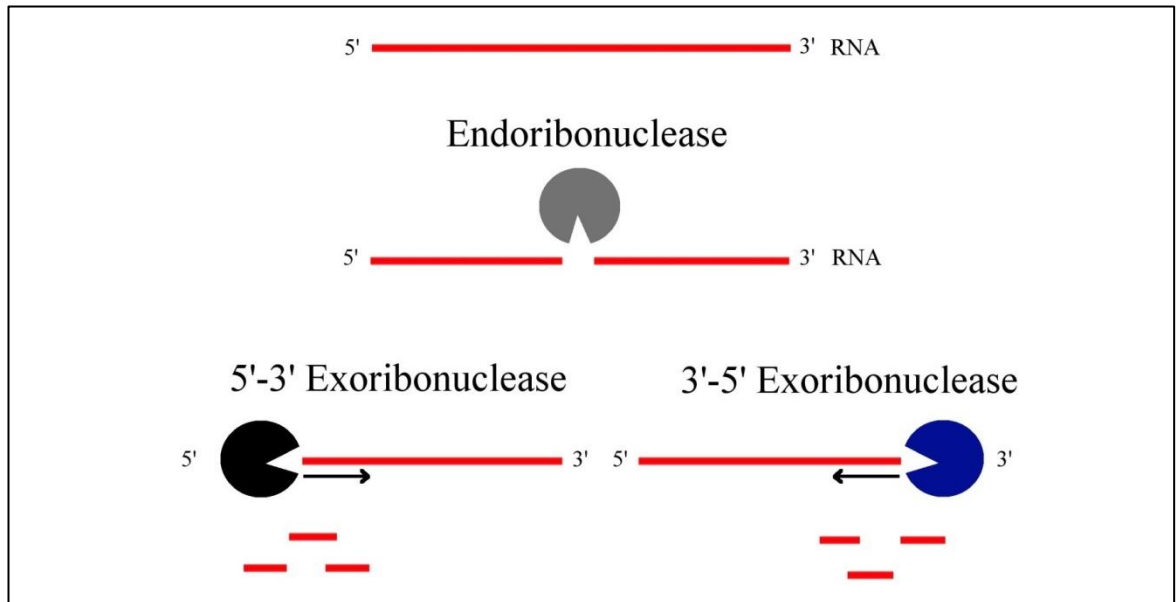


Figure 1.2 RNase-Mediated RNA Degradation

The process of RNase-mediated RNA degradation is shown in this schematic. Endoribonucleases (grey) break internal phosphodiester bonds within the RNA molecule (red). Whereas exoribonucleases remove nucleotides from the 5' end (black) or 3' end (blue). This process releases nucleotide diphosphates (NDPs, red segments) from RNA degraded by phosphorolytic enzymes and nucleotide monophosphates from hydrolytic enzymes. Image generated in GIMP (Version 2)(GIMP, n.d.).

RNases can degrade RNA independently or as part of multi-protein complexes. However, the net result, the regulation of cellular RNA levels and thus protein production within the cell, is the same (Zuo & Deutscher, 2001). For example, in bacteria, the main RNA degrading machine is the degradosome, whereas in archaea and eukaryotes it is the exosome (Carpousis, 2007; Rajmakers, Schilders, & Pruijn, 2004; Sarkar & Fisher, 2006). Prior to discussing these RNase complexes in more detail, the processes by which each domain of life, including prokaryotes, eukaryotes and archaea, can regulate RNA turnover are briefly reviewed in Sections 1.3.1-1.3.3 respectively.

1.3 RNA Turnover

1.3.1 RNA Turnover in Prokaryotes

The rate of turnover dictates RNA stability, this can vary considerably between organisms and even individual transcripts within a cell. *Escherichia coli* is the best studied prokaryote model, so much of the research on RNA turnover has been conducted using this organism. In *E. coli*, RNAs which are known to be stable include transfer and ribosomal RNAs (tRNAs and rRNAs), conversely, messenger RNAs (mRNAs) are typically unstable with an average half-life ranging from 2-25 minutes (Janga & Babu, 2009). The 5' end of mRNAs are also reported to be less stable than the 3' end, producing a model where a net 5'-3' directionality of degradation is suggested (Selinger, 2003). Additionally, the existence of structural features or biochemical factors, that target specific classes of mRNA for decay, have been suggested to affect stability. This was following the observation that some *E. coli* mRNAs are degraded by individual RNases, while others are targeted by multi-protein complexes *in vivo* (Bernstein, Lin, Cohen, & Lin-Chao, 2004).

The length of the poly adenosine (poly(A)) tail at the 3' end of a mRNA has also been linked to prokaryotic transcript stability. Previous data found that in a wild type *E. coli* strain, poly(A) tails ranged between 10-50 nts and in strains where the poly(A) polymerase (PAP) enzyme which synthesises these tails was absent, mRNA transcripts were found to be more stable (O'Hara *et al.*, 1995). This research indicated that the post-transcriptional modification of polyadenylation in *E. coli*, decreases stability by signalling transcripts for degradation (Sarkar, 1997). It was suggested that these poly (A) tails, produced from polyadenylation enzymes, help prokaryotic RNases to turnover RNA by clearing the barrier of secondary structures and providing a platform for RNase binding (Mohanty & Kushner, 2000a). Interestingly, polyadenylation is highly indiscriminate; essentially targeting virtually all unprotected mRNA 3' ends for efficient control of gene expression (Sarkar, 1997).

Rapid mRNA turnover is thought to be essential for an organism's survival in environments subject to rapid changes (Wilusz, Wormington, & Peltz, 2001). As such, the RNA of the cyanobacteria *Prochlorococcus* has been discovered to have a short median half-life of 2.4 minutes, which is suggested to allow rapid recycling of nucleotides that would be advantageous in nutrient poor oceans (Steglich *et al.*, 2010).

1.3.2 RNA Turnover in Eukaryotes

The average half-life of eukaryotic mRNA is longer when compared with that of transcripts from prokaryotes (Rauhut & Klug, 1999). This is likely due to several factors, for example unlike unicellular prokaryotic organisms, which are highly exposed to variations in the surrounding environment; multi-cellular eukaryotic organisms are typically more protected from their environment. Hence the rapid RNA turnover required for prokaryotes to survive in harsh conditions is less crucial for eukaryote survival. Additionally, eukaryotes have evolved a more complex gene expression system, where mRNA is transcribed within the nucleus and transported to the cytoplasm for translation into proteins. Therefore, post-transcriptional modifications, which are not required for prokaryote mRNA stability, are required for eukaryote transcript stability. This increased level of eukaryote transcript processing, through the addition of protective entities, likely explains why differences in mRNA stability control exist between the RNA turnover systems of complex eukaryotic and simpler prokaryotic organisms.

Eukaryotic mRNA processing involves the addition of a long poly(A) tail and covalently linked 5'-methylguanosine-cap structure. In contrast to prokaryotes, the long 3' Poly(A) tail does not target the mRNA for degradation, instead these modifications promote transcript stabilisation and are crucial for efficient processing, nuclear export and translation (Schmid & Jensen, 2008; Wilusz *et al.*, 2001). As a result, the length of mRNA Poly(A) tails and efficiency of polyadenylation is significantly higher in eukaryotes than prokaryotes, with tails on eukaryotic transcripts ranging from 80-200 nt in length (Steege, 2000). However, the precise mechanism in eukaryotes for

mRNA decay, can vary significantly across eukaryotes. For example, in the yeast *Saccharomyces cerevisiae*, polyadenylation of aberrant transcripts can lead to degradation rather than stabilisation (Ibrahim *et al.*, 2008). Interestingly, the addition of a poly(A) tail can even have opposite functions within the same cell (Borowski, Szczesny, Brzezniak, & Stepien, 2010).

Evidently, with added mRNA processing there is added complexity in eukaryotic mRNA degradation pathways. RNases which cleave RNA internally can leave it susceptible to further external degradation by other RNases. However, due to their additional modifications, degradation of eukaryotic transcripts usually occurs following the deadenylation and decapping events shown in Figure 1.3 (Wilusz *et al.*, 2001).

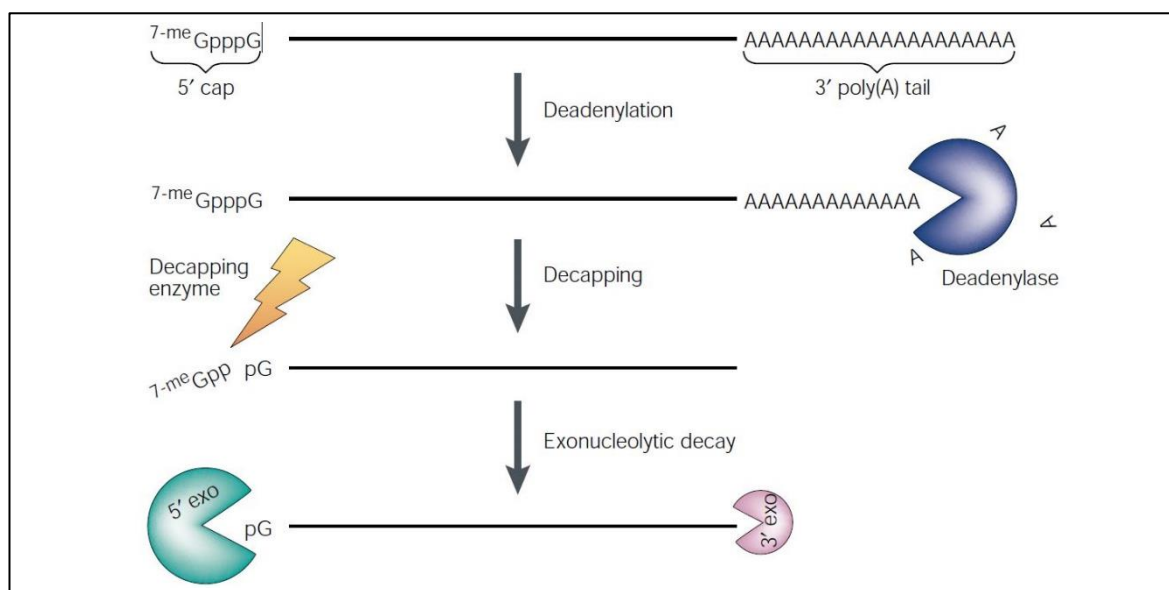


Figure 1.3 Eukaryotic mRNA Decay

Processed eukaryotic mRNA containing a 5' cap structure (7-meGpppG) and a 3' poly(A) tail is deadenylated by deadenylase (blue), decapped by decapping enzymes (yellow bolt) and then subject to decay at the 5' and/or 3' ends by exoribonucleases (exo, green and pink respectively). Image taken directly from (Wilusz *et al.*, 2001).

As it is in prokaryotes, effective RNA turnover is an essential process in eukaryotes; abnormal mRNA stability has been implicated in disease states, including cancer, chronic inflammatory responses and coronary disease (Wilusz *et al.*, 2001). As a result, processes including nonsense-mediated decay, which remove mutated transcripts with premature termination codons, exist to prevent erroneous gene expression. Also, processes commonly termed no-go decay and non-stop decay are also in place to remove transcripts with defective transcription and termination elements respectively (Belasco, 2010; Shoemaker & Green, 2012). This effective turnover of harmful defective transcripts is crucial, as it prevents gene expression which can result in genetic diseases such as cystic fibrosis and Duchenne muscular dystrophy (Wilusz *et al.*, 2001).

To add more complications, in addition to nuclear-encoded transcripts, eukaryotes also transcribe genes within their mitochondria and disturbance of mitochondrial gene expression has been linked

to aging, cancer and neurodegenerative diseases in humans (Borowski *et al.*, 2010). Furthermore, the RNA degradation process within mitochondria has been reported to be different to the processes occurring within the cytoplasm and this is most likely due to the mitochondria's evolutionary roots. For example, plant and algae chloroplasts were found to have a poly(A)-mediated RNA degradation mechanism very similar to that of prokaryotes (Hayes, Kudla, & Gruissem, 1999). Since mRNA degradation occurring in organelles can also involve nuclear-encoded proteins, there must be fundamental differences between this system and that existing in bacteria (Hayes *et al.*, 1999).

1.3.3 RNA Turnover in Archaea

The process of RNA turnover in archaea is still not fully understood. For example, the presence of polyadenylation enzymes and the role of polyadenylation within archaea appears to be conflicting; neither PAP nor polynucleotide phosphorylase (PNPase) homologs have been identified in any archaeal class. However, a multi-protein complex similar to PNPases, termed the archaeal exosome, has been found in some classes. Additionally, different members of the archaea have been discovered to have very different RNA turnover machinery. For example, the halophilic archaeal organism *Haloferax volcanii* is not known to contain homologs of the eukaryotic degrading exosome complex, although the hyperthermophile *Sulfolobus solfataricus* possesses an archaeal exosome. In addition, no polyadenylation was detected in *H. volcanii*, whereas polyadenylation was discovered to occur in *S. solfataricus* by the archaeal exosome complex (Portnoy *et al.*, 2005).

1.3.4 RNA Turnover Summary

In summary, although the mechanisms of RNA turnover have diversified over evolution, leading to fundamental differences between the kingdoms, there are many basic features in common, including the use of ribonucleases to mediate mRNA degradation. The subsequent Sections 1.4.1-1.4.2 review the plethora of RNase complexes that exist and how they degrade transcripts, to control gene expression, across evolutionarily distinct organisms. It focuses on how the RNA degradosome and exosome complexes control RNA turnover across prokaryotes, eukaryotes and archaea.

1.4 RNase Complexes

1.4.1 RNA Degradosome

In prokaryotes, a complex of RNases termed the degradosome is known to effectively couple exoribonucleases and endoribonucleases together with other proteins, to produce a degrading machine that can efficiently coordinate RNA turnover. The most well studied degradosome complex is that of *E. coli* and this is reported to control global gene expression post-transcriptionally (Callaghan *et al.*, 2005; Marcaida, DePristo, Chandran, Carpousis, & Luisi, 2006).

The *E. coli* degradosome is assembled upon the endoribonuclease RNase E (EC 3.1.26.12), which has two distinct domains; a structured catalytic N-terminal domain (NTD) and a mostly natively unstructured RNA and protein binding C-terminal domain (CTD) (Callaghan *et al.*, 2004; Marcaida *et al.*, 2006) (Figure 1.4)(a). It is within this extensive scaffolding CTD that the other principal components of the degradosome including, the exoribonuclease PNPase (EC 2.7.7.8), a DEAD-box RNA-dependent ATPase/helicase called RhlB (EC 3.6.4.13), and the glycolytic enzyme Enolase (EC 4.2.1.11) are assembled (Figure 1.4) (Carpousis, Vanhouwe, Ehretsmann, & Krisch, 1994; Py, Causton, Mudd, & Higgins, 1994; Py, Higgins, Krisch, & Carpousis, 1996; Vanzo *et al.*, 1998). Remarkably the RNase E protein does not need to exhibit RNase activity to act as the essential scaffold in the reconstitution process; indicating that the enzyme is bifunctional (Coburn, Miao, Briant, & Mackie, 1999).

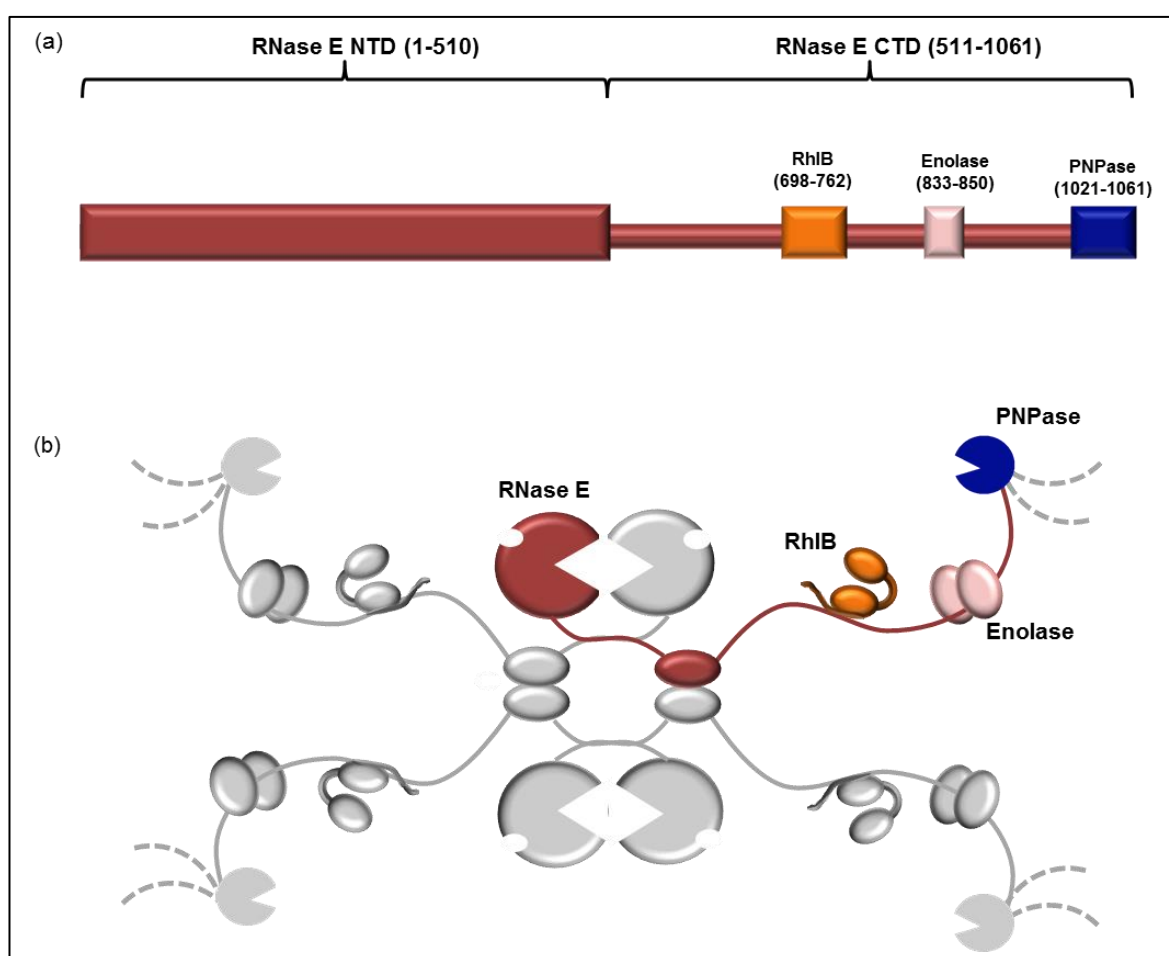


Figure 1.4 RNA Degradosome Complex

(a) A schematic of the domain organisation of the RNA degradosome is provided, showing the RNase E N-terminal domain (NTD, 1-510) and its C-terminal domain (CTD, 511-1061). The RhlB (698-762), Enolase (833-850) and PNPase (1021-1061) RNase E microdomains, within which these proteins assemble, are indicated (coloured orange, pink and blue respectively). (b) The organisation of these proteins upon the CTD of RNase E is shown as a schematic, with one RNase E monomer highlighted in red. The other monomers of the tetrameric protein are shown in grey for clarity. Figure kindly provided by Dr Heather Bruce and adapted.

The main enzyme of the degradosome, RNase E, has an evolutionarily conserved NTD; however its CTD exhibits sequence divergence and researchers have suggested that the two halves are thus evolving at different rates (Erce, Low, March, Wilkins, & Takayama, 2009). The CTD has been

observed to exhibit weak conservation even in closely related phyla (Hardwick, Chan, Broadhurst, & Luisi, 2011; Marcaida *et al.*, 2006). This sequence diversity, and the natively unstructured characteristics of the CTD, are thought to help modulate the interactions of proteins assembled within the degradosome. Although degradosomes and degradosome-like assemblies are conserved throughout proteobacteria, the composition of assembled proteins is variable and it is thought that the CTD provides the capacity for rapid adaptations. A summary of RNA degradosome evolution is provided in Figure 1.5 and the diversity of proteins which associate with the CTD are shown (Hardwick *et al.*, 2011).

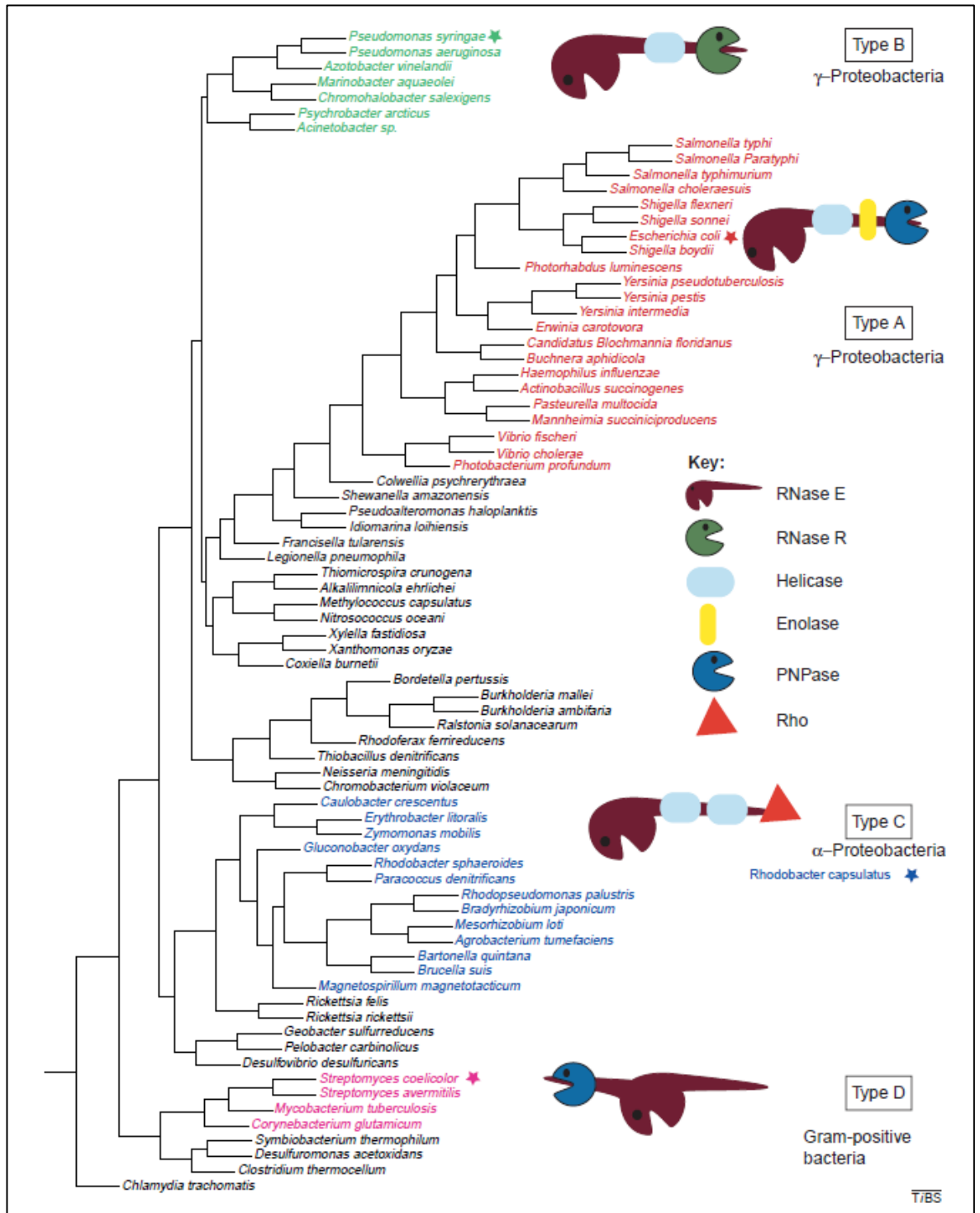


Figure 1.5 RNA Degradosome Evolution

A phylogenetic tree for bacterial RNase E which shows the degradosome as an assembly that varies greatly in the course of evolution. RNase E homologs with a helicase, enolase and PNPase associated at the C-terminal domain, similar to *E. coli* (red star/cartoon), are colour-coded as the red branches in the phylogenetic tree (Type A). The characteristic microdomains for the branches that are not well conserved are labelled in black. The branches coloured green are also γ -proteobacteria (Type B), like *E. coli*, however the 3'-5' exoribonuclease RNase R is instead present within this degradosome. The *Rhodobacter capsulatus* RNase E (blue star/cartoon) sequence is not available but its homologs are indicated within the blue branch. The α -proteobacteria (Type C) and gram-positive bacteria (Type D) are most divergent from *E. coli* with either no PNPase, or PNPase at the N-terminal domain respectively. Image taken directly from (Marcaida *et al.*, 2006).

The compositional variability of the degradosome may explain how individual prokaryotic organisms differentially regulate RNA metabolism. It may also suggest how these bacteria are able

to adapt and survive when exposed to different environmental and cellular metabolic fluxes (Carpousis, 2007; Marcaida *et al.*, 2006). This is especially important, since bacteria are known to inhabit a wide variety of environmental niches, which can be extreme in nature and/or rapidly changing. For example, alternative assemblies of the *E. coli* degradosome have been suggested to form in different growth conditions; under cold shock, RNA helicase CsdA co-purifies with the degradosome in place of RhlB (Prud'homme-G enreux *et al.*, 2004). Other minor components of the degradosome have been discovered including the protein chaperones Dnak and GroEl. Interestingly these proteins have been linked not only to the heat shock survival of *E. coli* (Miczak, Kaberdin, Wei, & Lin-Chao, 1996), but also to the heat and antibiotic shock survival of the opportunistic human pathogen *Acinetobacter baumannii* (Cardoso *et al.*, 2010). This information further indicates that RNA degradosomes are adjustable and capable of adapting to different environmental conditions.

Although many variations are available, in *E. coli*, a minimal degradosome of RNase E, PNPase, and RhlB is suggested to form spontaneously *in vitro* in the absence of all other cellular components (Coburn *et al.*, 1999). This minimal RNA degradosome is therefore capable of both endoribonuclease and exoribonuclease activity through the presence of RNase E and PNPase respectively. Additionally, the presence of RhlB in the minimal degradosome has been shown to facilitate PNPase-mediated degradation *in vitro*, presumably by helping unwind RNA stem loops (Py *et al.*, 1996). As mentioned previously, enolase was originally discovered as a RNA degradosome component, however the observation that only 5-10% of the total cellular enolase co-purified with the degradosome, led to questions about its role (Py *et al.* 1996). Indeed, later a minimal degradosome was shown to assemble and exhibit activity without enolase; suggesting enolase was not crucial *in vitro* (Coburn *et al.*, 1999). Although a potential role in regulating response to growth conditions has been suggested for enolase (Vanzo *et al.*, 1998), the activity this enzyme has on RNA metabolism is poorly understood and the relevance/importance of this metabolic enzyme on RNA turnover still requires understanding (Hardwick *et al.*, 2011).

In addition to the presence of the ATP-sensing RNA helicase and the metabolic enzyme enolase within the *E. coli* degradosome, the discovery of the Krebs enzyme aconitase within the *Caulobacter crescentus* degradosome indicated another link between RNA degradation and metabolism and environmental sensing (Carpousis *et al.*, 2007; Hardwick *et al.*, 2011; Liou *et al.*, 2001; Nurmohamed *et al.*, 2009; Nurmohamed *et al.*, 2011). Furthermore, although *Bacillus subtilis* lacks an RNase E homolog, it possesses a degradosome-like assembly with a functionally analogous endoribonuclease RNase Y, which can also recruit two glycolytic enzymes, namely enolase and phosphofructokinase (Commichau *et al.*, 2009). From these observations, researchers have suggested that the conserved association of metabolic enzymes with RNases in bacteria further indicates that the interaction has functional consequence (Hardwick *et al.*, 2011). This is reviewed in more detail later in Section 1.6. However, it is important to note the reoccurrence of

the interactions between enzymes involved in central metabolism and the main machinery controlling post-transcriptional gene regulation in bacteria.

In summary, the overall composition of the degradosome complex is diverse across prokaryotes, however the enzymes involved in this complex are highly conserved. The equivalent RNA degrading complex in eukaryotes is called the exosome and its role in regulating post-transcriptional gene expression is briefly introduced below in Section 1.4.2 before more detailed structural information is provided in Section 1.5.

1.4.2 RNA Exosome

All eukaryotic and archaeal exosomes identified to date contain proven or predicted RNase and RNA binding subunits. These exosome subunits associate with a range of cofactors, not only in a species-dependent manner, but also dependent on their cellular localisation (Ng, Waterman, Antson, & Ortiz-Lombardia, 2010). The exosome core assembly itself has a remarkable resemblance to the bacterial RNase and polyadenylation enzyme PNPase (Hardwick *et al.*, 2011). The archaeal exosome is most similar to PNPase both structurally and functionally. In contrast, although the eukaryotic exosome has structural homology to PNPase and the archaeal exosome, it has lost its phosphorolytic catalytic activity and requires additional domains to mediate its hydrolytic activity (Lykke-Andersen *et al.*, 2009). The structural and functional conservation of the RNA exosome and PNPase is discussed in Section 1.5.4 in more detail.

The presence of PNPase in the prokaryote RNA degradosome, and its homology to the eukaryotic and archaeal exosome structures, highlights the importance of this ribonuclease in RNA turnover. Hence the subsequent Section 1.5 describes the activity, structure, evolution and role of PNPase in regulating gene expression.

1.5 Polynucleotide Phosphorylase

Polynucleotide Phosphorylase (PNPase, also known as polyribonucleotide nucleotidyltransferase) and RNase PH are two enzymes within the phosphate dependent exoribonuclease (PDX) super-family. The PDX family use inorganic phosphate (PO_4^{3-}) to degrade RNA, in contrast to the other 6 super-families: RBN, DEDDM, RNR, RRP4 and 5PX, which are known to use hydrolytic cleavage for this process (as reviewed in Zuo & Deutscher, 2001).

1.5.1 PNPase Activity

PNPase is an important processive 3'-5' bacterial exoribonuclease, which can act both independently within the cell cytoplasm, and as part of the bacterial degradosome complex (Bernstein *et al.*, 2004). The activity of PNPase in *E. coli* (EcPNPase) is well-studied and its mode of activity is shown in Figure 1.6. This enzyme has an interesting role in that it can reversibly act as either a metal-dependent phosphorolytic nuclease (degrading RNA) or a template-independent

polymerase (synthesising RNA) (Littauer & Soreq, 1982; Mohanty & Kushner, 2000b). In the presence of excess inorganic phosphate (PO_4^{3-}) and either of the metal cofactors; magnesium (Mg^{2+}) or manganese (Mn^{2+}), PNPase can progressively degrade single stranded RNA in a 3' to 5' direction, releasing NDPs in the process (Figure 1.6, forward arrow) (Nurmohamed *et al.*, 2009). More specifically, the metal ion in EcPNPase is thought to be coordinated by two conserved aspartate residues, and these residues may act in conjunction with the bound metal to support general acid/base catalysis (Nurmohamed *et al.*, 2009). However, in the presence of excess NDPs the enzyme can reverse its function (Figure 1.6, reverse arrow) and systematically add NDPs onto the 3' end of RNA, releasing inorganic phosphate in the process (Mohanty & Kushner, 2000a; Nurmohamed *et al.*, 2011).

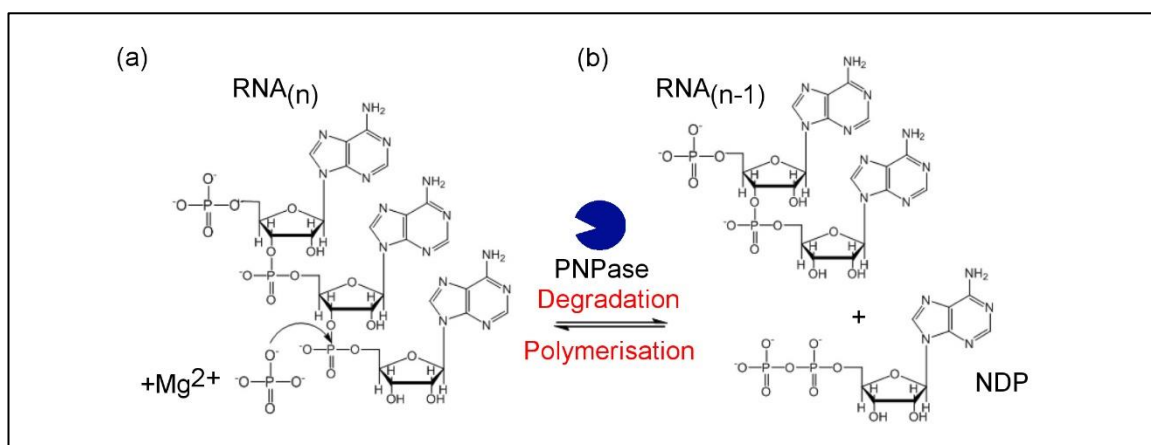


Figure 1.6 PNPase Activity

(a) PNPase can act to either degrade RNA substrate at the 3' end using inorganic phosphate (PO_4^{3-}) and metal ions such as Mg^{2+} or (b) in the presence of NDPs it can synthesise RNA from the 3' end in the reverse reaction. Figure adapted from (Nurmohamed *et al.*, 2011) using GIMP (v2)(GIMP, n.d.).

1.5.2 PNPase Structure

Structural studies of PNPase from *E. coli* (EcPNPase; (Nurmohamed *et al.*, 2009; Shi, Yang, Lin-Chao, Chak, & Yuan, 2008)), *Streptomyces antibioticus* (SanPNPase; (Symmons, Jones, & Luisi, 2000)), *C. crescentus* (Hardwick, Gubbey, Hug, Jenal, & Luisi, 2012), *Coxiella burnetii* (Franklin *et al.*, 2015) and *Homo sapiens* (hPNPase; (Lin *et al.*, 2012)) have revealed a ring-shaped homotrimeric complex, with a core hexamer of PH domains that form a central channel that is able to accommodate single-stranded RNA. The RNA-binding domains are positioned on one side of the PH ring where they can guide RNA into the channel towards the active site (Franklin *et al.*, 2015; Hardwick *et al.*, 2012; Lin *et al.*, 2012; Shi *et al.*, 2008; Symmons *et al.*, 2000).

Focusing on EcPNPase, a schematic of the full-length domain organisation and solved core enzyme structure is shown in Figure 1.7 (a) and (b) respectively. The domain schematic in Figure 1.7 (a) shows the full length protein, with the catalytically inactive phosphorolysis or RNase PH-like domain (PH-1), a linking α helical domain (H), a catalytically active RNase PH-like domain 2 (PH-2), followed by two RNA binding domains; a K-homology domain (KH) and an S1 domain (S1) (Leszczyniecka, DeSalle, Kang, & Fisher, 2004; Symmons *et al.*, 2000). The 3D structure of

EcPNPase is highlighted Figure 1.7 (b); only the core enzyme (Δ KH/S1; KH and S1 domains truncated) could be determined by X-ray crystallography. This may be a result of flexibility within the KH and S1 domains (3GCM) (Nurmohamed *et al.*, 2009). EcPNPase is a homotrimeric protein and interactions between each monomer (coloured dark blue, blue and cyan respectively) collectively form a central pore. The domains which contribute to a single ‘core’ monomer are highlighted with dotted lines; this includes the PH-1 and 2 domains and a linking α helical domain (Figure 1.7 (b)).

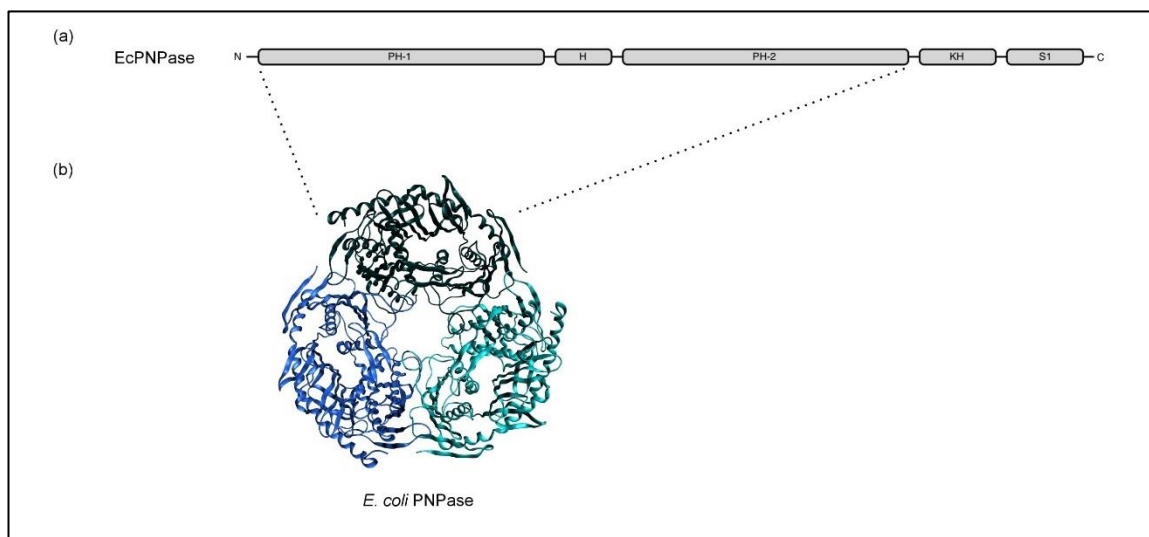


Figure 1.7 PNPase Domain Organisation and Crystal Structure

(a) PNPase domain architecture is shown with the RNase PH-like domain (PH-1) domain, an α helical domain (H), PH-2 domain and the KH and S1 RNA binding regions. (b) The PNPase crystal structure from *E. coli* (PDB: 3GCM) with individual monomers coloured in dark blue, blue and cyan. The domains in (a) which constitute one monomer in (b) are indicated with dotted lines (Nurmohamed *et al.*, 2009). Image created using GIMP (v2) (GIMP, n.d.).

The catalytic activity of EcPNPase resides in the PH-2 domain only; however it has been shown that both PH domains are required for full catalytic activity (Briani *et al.*, 2007; Jarrige *et al.*, 2002). Interestingly, both domains are also required for maintaining the enzyme’s activity in response to cold temperatures, affecting *E. coli* stress survival (Awano *et al.*, 2008). Due to difficulty determining the crystal structure of full-length EcPNPase, the KH and S1 RNA binding domains are proposed to assemble on the top of the core protein in a similar manner to *C. crescentus* PNPase (Hardwick *et al.*, 2012). Moreover, the proposed mechanism of how the KH and S1 domains help feed RNA substrates into the active site for cleavage, is provided by the crystal structures of RNA-bound and apo *C. crescentus* PNPase. The RNA-free form suggests that RNA capture is facilitated by a ‘splayed’ conformation of S1 domains. The RNA bound form of *C. crescentus* PNPase suggests that the three KH domains collectively close upon the substrate RNA and direct the 3’ end towards the constricted aperture at the entrance of the central channel. It is proposed that upon RNA binding, the substrate is channeled to the active site through mechanical ratcheting (Hardwick *et al.*, 2012). The KH pore of hPNPase is also proposed to trap long RNA tails and feed them into the central channel for the degradation process in a similar manner to *C. crescentus* PNPase (Lin *et al.*, 2012).

The enzyme EcPNPase shown in Figure 1.7 is highly conserved. It shares a similar structure across all PNPase homologs and its evolution is described below in Section 1.5.3.

1.5.3 Origin of the PNPase Domains

Through extensive phylogenetic comparison, which highlighted the similarities between PNPase RNase PH-like domains and the domains in the RNase PH protein itself, a model of PNPase domain evolution has been proposed (Figure 1.8) (Leszczyniecka *et al.*, 2004). Leszczyniecka and colleagues conducted phylogenetic analysis to provide the model which suggested that a common ancestral RNase PH-like domain (Figure 1.8, red) underwent an initial duplication event (Figure 1.8 (a) black arrow), which gave rise to the physically linked RNase PH-like domains 1 and 2. A later divergence event of the RNase PH-like domain 2 was suggested (Figure 1.8 (b), blue), followed by a second duplication event (Figure 1.8 (c) arrow), thus creating two copies of the gene. It has been hypothesised that for one of these duplicates, a stop codon was introduced within the 5' untranslated region (UTR) of the RNase PH-like domain 1; causing this domain to be silenced (Figure 1.8 (d) black circle). Following separate evolutionary changes; a gene transfer event relocated one gene to a different location of the genome, now called RNase PH (green), and the other gene, now called PNPase (red and blue), acquired KH and S1 domains (orange and yellow respectively). Although two different proteins resulted, similarities still exist; the RNase PH and PNPase RNase PH-like 2 domains are most similar (Figure 1.8 (e)). Functionally, both these enzymes belong to the same PDX family of 3'-5' phosphorolytic exoribonucleases.

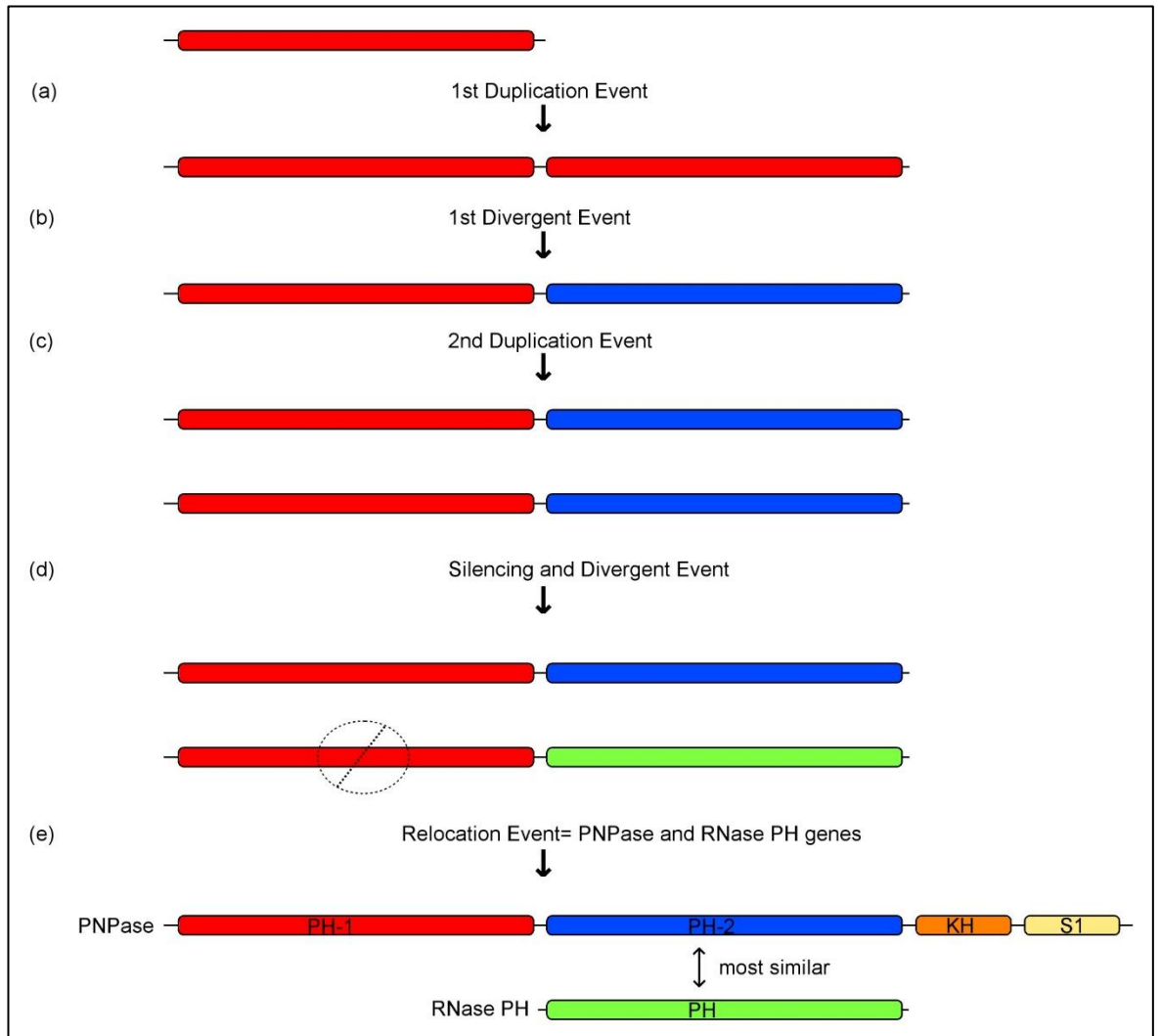


Figure 1.8 PNPase Domain Evolution

The domain evolution of PNPase and RNase PH proteins is proposed within this schematic from the extensive phylogenetic analysis conducted by Leszczyniecka *et al.*, 2004. (a) The model suggests an initial gene duplication event, (b) followed by a divergent event gave rise to the physically linked RNase PH-like domains 1 and 2 (red and blue). (c) A second duplication event, (d) a subsequent gene silencing event (black circle) and divergent event (green), then gave rise to the (e) two separate PNPase and RNase PH genes. The PNPase gene also acquired the additional KH and S1 domains (orange and yellow respectively). Image adapted from (Leszczyniecka *et al.*, 2004) using GIMP (v2)(GIMP, n.d.).

In summary, over evolution the RNase PH domain has been retained and has even evolved to form the ribonuclease PNPase. This is important to understand, as RNase PH-like proteins form not only the basis of PNPase in prokaryotes, such as *E. coli*, and eukaryotes, such as *H. sapiens*, but they are also the main component of exosome complexes. The structural conservation of PNPase and exosome complexes are discussed in more detail in the following Section 1.5.4.

1.5.4 PNPase Structural Conservation

The PNPase enzyme itself is evolutionarily conserved, it is found in both prokaryotes and eukaryotes and can be located within the cytoplasm, chloroplasts and mitochondria; an N-terminal signal peptide targets PNPases to these organelles (Schuster & Stern, 2009; Yehudai-resheff, Hirsh, & Schuster, 2001). Whilst archaea do not possess PNPase (Leszczyniecka., 2004), their exosome consists of RNase PH-like proteins called Rrp, which are thought to originate from a common

ancestor of PNPase (Evguenieva-Hackenberg, Hou, Glaeser, & Klug, 2014; Liu, Greimann, & Lima, 2006). Furthermore, the domains of the eukaryotic exosome also consist of Rrp proteins and so it is no surprise that the overall organisation of RNase PH, PNPase, archaeal and eukaryotic exosomes are all very similar, as shown in Figure 1.9 (Slomovic *et al.*, 2008). Thus the structure, and possibly the function, of RNA degrading machines are seen to be evolutionarily linked across the three domains of life.

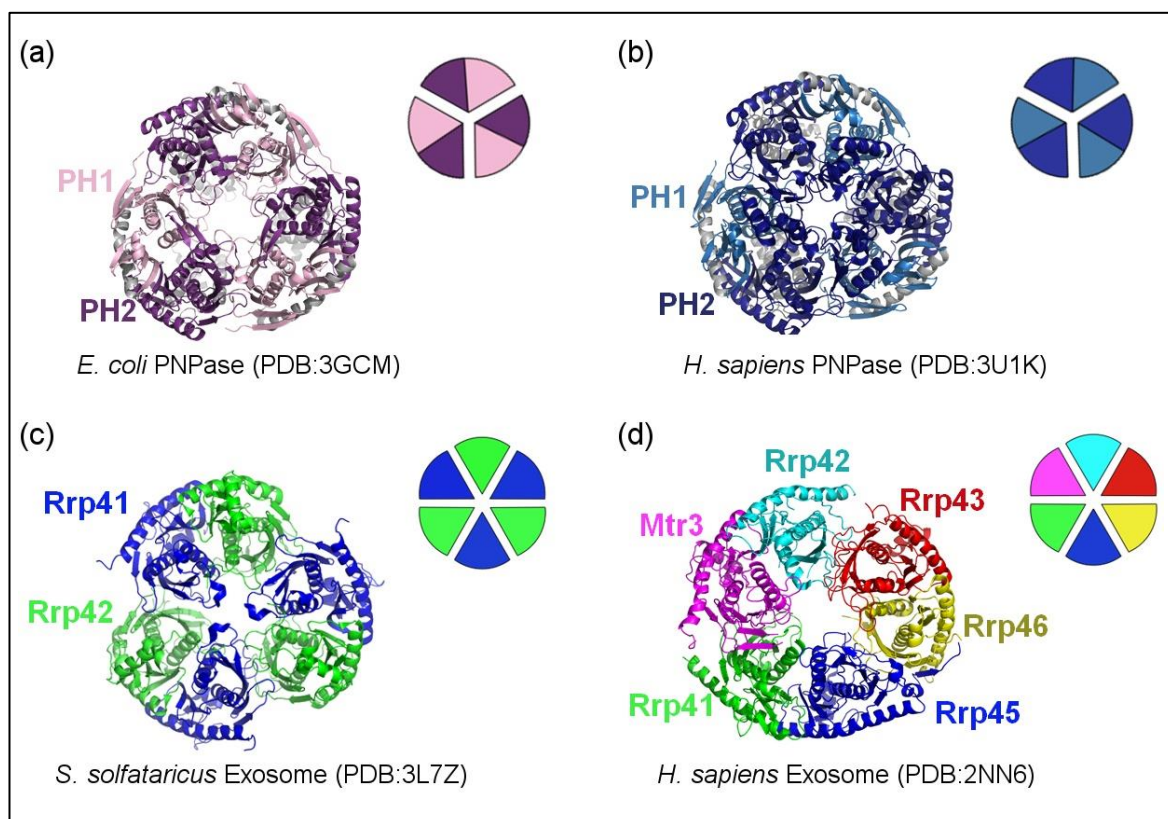


Figure 1.9 PNPase and Exosome Structural Homology

The crystal structures of (a) *E. coli* (3GCM), (b) *H. sapiens* (3U1K) PNPase, (c) *S. solfataricus* (3L7Z) and (d) *H. sapiens* (2NN6) exosomes. RNase PH-like domains (PH) of PNPases and homologous domains (Rrp) of exosome are indicated. Two PH domains (PH-1 and PH-2) collectively form one PNPase monomer in (a) *E. coli* (light pink and purple respectively) and (b) *H. sapiens* PNPase (light and dark blue respectively). Two Rrp domains (Rrp41 and Rrp42) form an equivalent exosome monomer in (c) *S. solfataricus* (blue and green respectively). The individual domains which make up (d) *H. sapiens* exosomes are Rrp41 (green), Rrp42 (cyan), Rrp43 (red), Rrp45 (blue), Rrp46 (yellow) and Mtr3 (pink). (Lin *et al.*, 2012; Liu *et al.*, 2006; Lu, Ding, & Ke, 2010; Nurmohamed *et al.*, 2009). Adapted from an image provided by Dr Louise Butt.

The overall crystal structures of PNPase and exosome homologs is shown in Figure 1.9 and since no crystallography data are available for EcPNPase S1 and KH domains, the RNA binding domains of the other enzymes are not shown. Comparison of the structures for EcPNPase and hPNPase show that both enzymes consists of two different PH-1 and 2 domains (coloured light pink and purple and light blue and dark blue respectively), which collectively form one protomer (Figure 1.9 (a-b)). For both EcPNPase and hPNPase, three protomers collectively form a homotrimeric protein with a pseudo-hexameric ring-like structure, which is very similar to the archaeal *S. solfataricus* exosome and eukaryotic *H. sapiens* exosomes (Figure 1.9 (c-d)). The RNase PH protein is not shown within this figure, but is also known to form a hexameric ring of six identical RNase PH domains.

The archaeal exosome complex contains five different proteins: Rrp42 and Rrp41 (equivalent to PH-1 and PH-2 respectively) and three RNA-binding proteins Rrp4 (contains S1 and KH domains), Csl4 (contains S1 and Zn-finger domains) and DnaG (binds polyA) (Evguenieva-Hackenberg, Walter, Hochleitner, Lottspeich, & Klug, 2003). Structural studies of exosomes from *S. solfataricus* (SsoExosome; Lorentzen *et al.*, 2005a; Lorentzen, Dziembowski, Lindner, Seraphin, & Conti, 2007a; Lorentzen & Conti, 2005, 2012; Lu, Ding, & Ke, 2010), *Archaeoglobus fulgidus* (Büttner, Wenig, & Hopfner, 2005; Hartung, Niederberger, Hartung, Tresch, & Hopfner, 2010), *Pyrococcus abyssi* (Navarro, Oliveira, Zanchin, & Guimarães, 2008), and *Methanothermobacter thermautotrophicus* (Ng *et al.*, 2010) have revealed a similar architecture to PNPase. A trimer of Rrp41/Rrp42 dimers assemble to form the ring-shaped PH-domain hexamer of the archaeal exosome core (Figure 1.9 (c)). Three RNA-binding proteins, Rrp4 and/or Csl4-DnaG can then assemble on one face of the ring where they channel RNA towards the active site (not shown) (Büttner *et al.*, 2005; Lee, Hartung, Hopfner, & Ha, 2010; Lorentzen & Conti, 2012; Lorentzen *et al.*, 2007; Lu *et al.*, 2010). The eukaryotic core exosome has a hexameric ring structure that is similar, except it consists of six different RNase PH-like proteins (Rrp41 (green), Rrp42 (light blue), Rrp43 (red), Rrp45 (blue), Rrp46 (yellow) and Mtr3 (pink), Figure 1.9 (d)) (Liu *et al.*, 2006).

Whilst the PNPase and exosome proteins are seen to be structurally similar, functionally only the archaeal exosome exhibits similar enzymatic activity to PNPase. The eukaryotic exosome core has lost its RNA phosphorolytic degradation activity and requires additional accessory protein domains, Rrp6 and Rrp44, to hydrolytically cleave the RNA (Dziembowski, Lorentzen, Conti, & Séraphin, 2007; Januszyk & Lima, 2014; Liu *et al.*, 2006; Lykke-Andersen *et al.*, 2009; Wasmuth, Januszyk, & Lima, 2014). Despite this, the possibility of a 3'-5' RNA feeding mechanism through the central pore of the eukaryotic exosome is suggested to be similar to PNPase and the archaeal exosome (Wasmuth *et al.*, 2014). Nevertheless, the eukaryotic exosome was not the focus of this PhD research and PNPase homologs and the homologous archaeal exosome were only considered as examples of 3'-5' phosphorolytically active enzymes. In order to appreciate the importance of studying PNPase, its cellular roles are outlined below in Section 1.5.5.

1.5.5 PNPase Roles

EcPNPase can function independently but it is also a component of the multiprotein degradosome complex. About 10-20% of PNPase is found in the canonical degradosome along with RNase E, the RNA helicase RhlB and the glycolytic enzyme enolase (Aït-Bara & Carpousis, 2015; Bandyra, Bouvier, Carpousis, & Luisi, 2013; Sarkar & Fisher, 2006). This association with another ribonuclease and a helicase is likely to assist EcPNPase in its degradative capacity. However, localisation to the cytoplasmic membrane as part of the degradosome, is thought to restrict its activity (Aït-Bara & Carpousis, 2015; Bandyra *et al.*, 2013).

In terms of physiological function, the most studied PNPase is that from *E. coli*. In this bacterium PNPase is required for growth at low temperatures (Awano *et al.*, 2008) and its activity has been implicated in all aspects of RNA metabolism including; mRNA turnover (Donovan & Kushner, 1986; Mohanty & Kushner, 2003), degradation of small regulatory RNAs (sRNAs; (Andrade & Arraiano, 2008; Andrade, Pobre, Matos, & Arraiano, 2012) and the quality control of stable RNAs (Cheng & Deutscher, 2003; Li, Reimers, Pandit, & Deutscher, 2002). In bacterial species, PNPase also affects complex processes, such as the tissue-invasive virulence of *Salmonella enterica* (Clements *et al.*, 2002) and virulence factor secretion in *Yersinia* (reviewed in Matos, Bárria, Pobre, Andrade, & Arraiano, 2012).

As mentioned previously, PNPase in eukaryotes can be localised to the chloroplast and mitochondria. The exact role of hPNPase in the mitochondria remains to be fully understood and there is some controversy over its activity within this organelle. This is mainly due to its presence in the intermembrane space (IMS), where it is not thought to be involved in regulating mitochondrial RNA levels (Chen *et al.*, 2006), but it may instead play a role in RNA import (Wang *et al.*, 2010). Interestingly, a mutation in *PNPT1*, which encodes hPNPase, is reported to impair RNA import into mitochondria and cause respiratory-chain deficiency and hereditary hearing loss (Alodaib *et al.*, 2016). PNPase is clearly involved in many essential processes and although its activity within the mitochondria still requires much more research, it's important to understand the roles it plays in both RNA turnover and metabolism.

Not surprisingly, given these diverse roles, PNPase activity is tightly controlled at both post-transcriptional and post-translational levels. In *E. coli*, the stability of the *pnp* mRNA transcript and therefore expression of the *pnp* gene is auto-regulated by PNPase in a mechanism involving the endoribonucleases RNase III and RNase E (Carzaniga *et al.*, 2009; Jarrige *et al.*, 2001). Not only does PNPase regulate itself, but other cellular components have been shown to interact with the enzyme and control its activity. The next Section 1.6 reviews in more detail how cellular metabolism may regulate PNPase activity, and thus RNA turnover.

1.6 Regulating RNA Turnover in Response to Cellular Metabolism

The initial discovery that the glycolytic enzyme enolase and the Krebs cycle aconitase, can also form part of the degradosome in *E. coli* and *C. crescentus* respectively, hinted at a connection between RNA degradosome activity and central metabolism (Hardwick *et al.*, 2011; Miczak *et al.*, 1996; Nurmohamed *et al.*, 2011; Vanzo *et al.*, 1998). Also suggestive of a link between RNA degradation and metabolism is the finding that PNPase activity can be regulated by nucleotides and secondary signalling molecules. For example, EcPNPase is inhibited by ATP nucleotides (Del Favero *et al.*, 2008) and cyclic di-GMP secondary messenger molecules *in vitro* (c-di-GMP; Tuckerman *et al.*, 2011), and its homologues from *Nonomuraea sp.* and *Streptomyces* are inhibited

by stringent response alarmone signalling molecules (Gatewood & Jones, 2010; Siculella *et al.*, 2010). More specifically, PNPase in *S. antibioticus* was found to respond and coordinate its function in accordance to the stringent response alarmone guanosine tetraphosphate (ppGpp) and its precursor guanosine pentaphosphate (pppGpp) (Bralley & Jones, 2003), collectively termed (p)ppGpp (Hauryliuk, Atkinson, Murakami, Tenson, & Gerdes, 2015). This is interesting since (p)ppGpp has been linked to antibiotic production in response to amino acid and energy starvation, and suggests a link between environmental sensing and PNPase activity (Bralley & Jones, 2003; Mechold, Cashel, Steiner, Gentry, & Malke, 1996). Although PNPase from *S. antibioticus* and *E. coli* have significant structural conservation (Nurmohamed *et al.*, 2009; Symmons *et al.*, 2000), remarkably (p)ppGpp was not found to inhibit EcPNPase (Gatewood & Jones, 2010) and the significance of this alarmone's effect on other PNPase homologs is unknown.

Although the observations described previously may indicate a communicative link between RNA degradative machines and central metabolism, direct evidence of such a link was missing. This was true until research studying the structural and functional roles of PNPase in *E. coli*, independent of the degradosome, discovered that the Krebs cycle metabolite citrate could modulate its RNase activity both *in vitro* and *in vivo* (Nurmohamed *et al.*, 2011). Cellular metabolism was also discovered to be dependent upon EcPNPase activity (Nurmohamed *et al.*, 2011). The findings presented within this publication by our group and collaborators in Cambridge, laid the foundations for this PhD project. Hence the details of the citrate-PNPase interaction, discovered in bacteria, are described initially below and then the question of whether the citrate-PNPase interaction is a conserved communication mechanism in other organisms is addressed.

1.6.1 PNPase Activity May Be Modulated by Citrate in *E. coli*

The initial suggestion of a citrate-PNPase interaction in *E. coli* occurred upon solving the crystal structure of EcPNPase (3GCM). It was in this publication that Nurmohamed and colleagues observed that, like the *S. antibioticus* homologue (SanPNPase), EcPNPase forms a homotrimer with a ring-like architecture and a central channel, as shown in Figure 1.10 (a) (Nurmohamed *et al.*, 2009; Symmons *et al.*, 2000). Interestingly, they also noticed two citrate molecules co-crystallised within each of the EcPNPase three active sites and within one of three 'vestigial' active sites, which were not previously seen for SanPNPase (Figure 1.10). They also noted that, although these active and vestigial sites are related through approximate molecular symmetry, only the former possessed catalytic activity (Nurmohamed *et al.*, 2009).

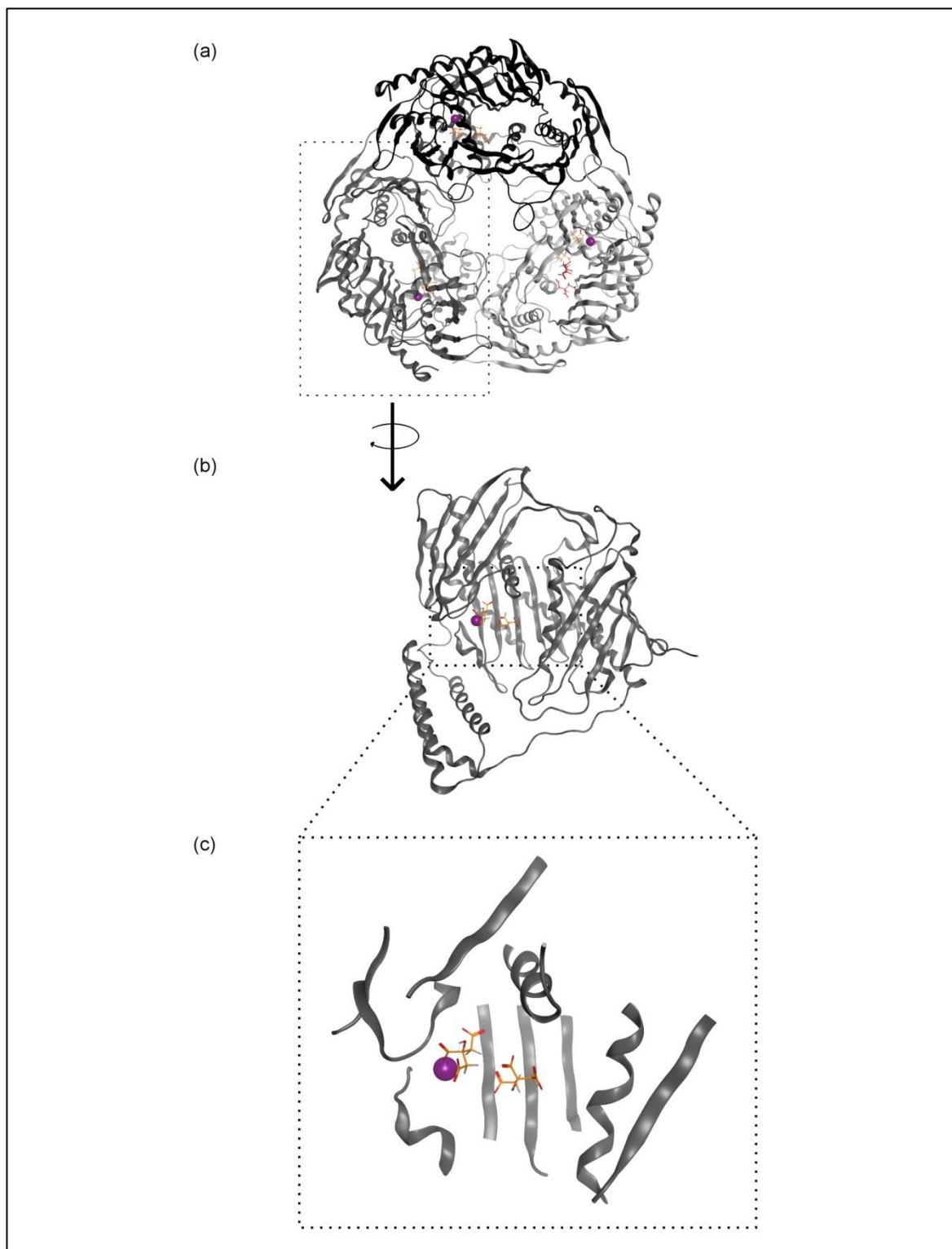


Figure 1.10 EcPNPase Crystal Structure

(a) EcPNPase homotrimer with each monomer coloured black, dark grey and light grey respectively. Highlighted within the structure are Mg²⁺ ions (purple sphere) and citrate molecules (at active and vestigial sites; orange and red sticks respectively). (b) The individual monomer and (c) a zoomed in image of the active site (dark grey ribbons) are shown. This image was created in MOE (Molecular Operating Environment, 2013) using the 3GCM structure deposited within the PDB (Nurmohamed *et al.*, 2009) and prepared in GIMP (v2)(GIMP, n.d.).

Upon discovery, the question of whether citrate was just an artefact of the crystallisation buffer, which contained 20 mM sodium citrate, or if it represented a physiologically relevant interaction was examined further by our research group. Accordingly, the subsequent study suggested that EcPNPase activity was indeed modulated *in vitro* and *in vivo* by physiologically relevant

concentrations of citrate (Nurmohamed *et al.*, 2011). More specifically, experiments *in vivo* found that *E. coli* cell strains (RNase II/R⁻), which are known to be dependent on PNPase for viability (Cheng & Deutscher, 2003; Donovan & Kushner, 1986), exhibited suppressed growth in the presence of increased intracellular citrate concentrations. Doubling times of the untreated PNPase dependant strain (RNase II/R⁻) increased from 49.0 (+/- 2.2) to 59.4 (+/- 3.2) minutes in the presence of 50 mM Mg²⁺ citrate. Contrastingly, there was no effect measured for WT and PNPase null strains when the same concentration of Mg²⁺ citrate was applied. It is understood that both RNases II and R, which are present in the WT and PNPase null strains, are able to compensate for loss in PNPase activity. Hence the negligible effect of citrate on these strains was unsurprising, and suggested that the negative effect observed specifically in the RNase II/R⁻ strain, was likely a result of citrate impact on PNPase function (Cheng & Deutscher, 2003; Nurmohamed *et al.*, 2011).

Utilisation of these strains for subsequent microarray analysis of mRNA levels revealed a PNPase-mediated response to citrate, when comparing a PNPase null mutant +/- citrate, against a WT strain +/- citrate. Genes generally affected by PNPase and those displaying a PNPase-mediated response to citrate were compared. Individual genes were assigned to gene ontology terms and from this it was clear that a PNPase-mediated citrate response broadly affected genes involved in cellular metabolic processes. More specifically, genes which were affected by citrate in a PNPase-dependent manner included those involved in cellular amino acid and derivative metabolic processes, and cellular biosynthetic and organic acid metabolic processes (Nurmohamed *et al.*, 2011). Transcripts *cirA* and *fkpA*, encode an outer membrane receptor and heat shock chaperone protein respectively and have been previously reported to be regulated by PNPase, were upregulated in a PNPase-mediated response to citrate in microarray analysis (Nurmohamed *et al.*, 2011). To complement the microarray analysis, RT-PCR was conducted to validate the relative abundance of mRNAs known to be regulated by PNPase. The levels of these transcripts, including *cirA* and *fkpA*, were also found to be upregulated in the presence of Mg²⁺ citrate, in a PNPase-dependant manner. Collectively, results from RT-PCR and microarray analysis showed that not only were specific transcripts affected in a PNPase-dependent response to citrate, but RNA metabolism was also broadly altered *in vivo* (Nurmohamed *et al.*, 2011).

Additionally, cellular metabolism was also observed to be affected widely by PNPase activity; the impact of PNPase on the metabolome was examined using ¹H NMR spectroscopy and gas chromatography mass spectrometry. The results of which suggested many metabolites were affected by PNPase loss, when comparing the null strain with the WT. The Krebs cycle metabolite succinate was found to decrease whereas citrate increased. It is important to recapitulate that a metabolite, in this case citrate, can affect PNPase activity and that PNPase itself can in turn affect metabolite levels. This potentially suggests that these interactions represent a novel feedback loop, where RNases may sense the metabolic status of the cell and modify it accordingly (Nurmohamed *et al.*, 2011).

Following *in vivo* work, the effect of citrate on PNPase activity was also investigated *in vitro*. It is known that PNPase requires metal ions for activity and that it can use either Mg^{2+} or Mn^{2+} for catalysis. The effect of both Mg^{2+} citrate and Mn^{2+} citrate on EcPNPase activity were therefore investigated and compared *in vitro*. Interestingly, only Mg^{2+} citrate was found to inhibit PNPase degradation activity; even in conditions of a 2-fold excess of Mg^{2+} to citrate. These findings indicate that citrate does not simply chelate active site metal ions and the inhibition seen actually requires specific ligand geometry.

It is important to acknowledge that citrate is also found ‘free’ within the cell and this ‘free’ citrate can also bind EcPNPase, yet unlike Mg^{2+} citrate, the free form was discovered to stimulate EcPNPase activity. The authors investigated the binding affinity of these interactions further and established that ‘free’ citrate bound EcPNPase with a weaker affinity than Mg^{2+} citrate; ~25 and ~3 mM respectively. Additionally, rather than binding in the active site, like Mg^{2+} citrate, free citrate was found to bind the ‘vestigial’ active site, potentially acting allosterically. It is well established that a metal ion is required for catalytic activity at the active site, hence the effect of the Mg^{2+} citrate was the focus of this PhD work and the effects of free citrate were not investigated.

In summary, the results described by Nurmohamed *et al* clearly revealed a communicative link between PNPase and metabolism in *E. coli*; whether this represents a mechanism that is conserved in other prokaryotes and/or higher organisms, remains to be identified and was one of the main research questions of this thesis.

1.6.2 Is Citrate Binding Conserved in PNPases Homologs?

As discussed previously in Section 1.5.4, PNPase is present in prokaryotic and eukaryotic domains and, although PNPase is absent in archaea, a homologous protein, termed the archaeal exosome, is known to exist. The structure and function of these PNPase homologs are highly conserved, all forming ring like structures, with similar domain architecture and phosphorolytic activity. It is important to note that these archaeal exosomes described are distinct from the eukaryotic exosomes, which have lost their phosphorolytic activity and instead utilises hydrolytic activity to degrade their RNA substrates.

The protein sequences and crystal structures of a number of PNPase homologs across the three domains of life are available, allowing easy comparison of their overall structure and proposals of their evolution. Upon examination of the solved eukaryotic hPNPase structure (PDB: 3U1K), citrate was also found to co-crystallise within the active site, similar to EcPNPase (Lin *et al.*, 2012). When the two enzymes are compared, they both possess a similar homotrimeric shape and citrate binds similarly within the active sites. However, unlike EcPNPase, the role of citrate within the hPNPase active site was not examined by Lin *et al*; they simply stated that co-crystallisation of

citrate was a result of its presence in the crystallisation buffer. Following the understanding that citrate has a physiological role within EcPNPase, it would be interesting to see if the same physical association of citrate with hPNPase has analogous effects on its activity. Furthermore, the localisation of PNPase to the mitochondria, the site of general metabolic activities in humans including the tricarboxylic acid (TCA) cycle and thus citrate synthesis, suggests that citrate could be available for hPNPase to mediate a regulatory role (Piwowarski *et al.*, 2003). The intracellular citrate concentration in *E. coli* has been described to range between 2-20 mM when grown on glucose/glycerol or acetate respectively (Bennett *et al.*, 2009). Although the level of inhibition of EcPNPase by citrate observed *in vitro* correlates with citrate concentration in this range (Nurmohamed *et al.*, 2011), the intracellular concentration of citrate in human cells is not clear. The human metabolome online database resource suggests that citrate is found in all cell types. It is present in both the mitochondria and cytoplasm and also in extracellular biofluids including blood, urine and saliva (Wishart *et al.*, 2013) and this may explain why it is difficult to determine cellular concentration of citrate. Although there has been some controversy over the intracellular citrate concentration, alongside the localisation of hPNPase to the mitochondria, and its role within this organelle (Chen, Koehler, & Teitell, 2007), there is not enough evidence to exclude the existence of a citrate-hPNPase regulatory interaction and as a result hPNPase was selected as a eukaryotic model to study within this work.

In contrast to EcPNPase and hPNPase, which were co-crystallised with citrate, the structure of PNPase from gram-positive *S. antibioticus* (SanPNPase) was solved in the free form as citrate wasn't present in the crystallisation buffer (PDB: 1E3P, Symmons *et al.*, 2000). Like hPNPase, SanPNPase also has a similar structure to PNPase from the gram-negative bacterium *E. coli*. It therefore represented a bacterial PNPase homolog that could be used for *in silico* investigations comparing citrate-PNPase structural conservation across structurally diverse prokaryotes. Additionally, recombinant PNPase protein, from *Synechocystis sp* (cyanobacteria), was commercially available and therefore could be used for *in vitro* experimental studies, comparing citrate-PNPase communication across evolutionarily diverse prokaryotes.

The crystal structure of the archaeal homolog SsoExosome was also solved in the free form (PDB: 4BA1, Lorentzen & Conti, 2012). Whether citrate can bind to the SsoExosome and regulate activity is currently unknown. There is evidence that *S. solfataricus* possess a TCA cycle, and citrate synthase is encoded within the genome, suggesting the presence of citrate within this organism (She *et al.*, 2001; Ulas, Riemer, Zaparty, Siebers, & Schomburg, 2012). The SsoExosome also has a similar structure and active site to EcPNPase and hPNPase, so it may potentially interact with citrate. For these reasons, all three enzymes were utilised for *in silico* and *in vitro* experimental studies within this thesis, as they represent well characterised models, spanning across evolution.

In summary, the structures of several PNPases are available, including EcPNPase, hPNPase and SsoExosome, which are representatives that span across the three domains of life including prokaryotes, eukaryotes and archaea respectively. Furthermore, there is evidence for each of these organisms, that citrate synthesis occurs in the same location that these enzymes are found and thus it is potentially available for binding/regulation. Whether the citrate-PNPase communication in *E. coli* represents an important regulatory effect, that is conserved across evolution, is an unresolved question. Accordingly, the following Section 1.6.3 reviews bioinformatic techniques that have previously been used to study PNPase sequence conservation and that can be applied to investigate the conservation of citrate-binding residues.

1.6.3 Conservation of Citrate Binding Residues in PNPase Homologs

Bioinformatics can provide insights into the structure and function of macromolecules such as proteins and nucleic acids (Xiong, 2006). The ability to compare individual residues in proteins, which may be involved in ligand binding or enzymatic catalysis, across a variety of evolutionarily distinct species, is invaluable. For instance, comparing individual amino acid sequences of a protein of interest and a previously characterised protein through basic alignments and/or more complex bioinformatics can shed light on possible domains, motifs and even functions which are conserved. Most multiple sequence alignment (MSA) tools use a hierarchical or tree-based method, including the publicly accessible and popular Clustal web server systems (Higgins & Sharp, 1988). These types of systems build an MSA using pairwise comparison; starting with the most similar sequence pair and progressing to more distantly related sequences. This progressive algorithm can be problematic, as errors, in any stage, are propagated during the growing MSA. An alternative method, which is less efficient but improves the accuracy of MSA, involves an iterative process of pairwise alignments. Iterative MSA reduce the errors inherent in progressive methods, by repeatedly realigning the initial sequence, as the MSA grows (Hirosawa, Totoki, Hoshida, & Ishikawa, 1995).

Protein sequence alignments, using the techniques described above and structural comparisons previously published in the literature, have revealed that hPNPase, EcPNPase and SsoExosome all have similar domain architecture and overall structure (Figure 1.11). As mentioned previously in Section 1.5.4, PNPases contain five different protein domains; the PH-1 domain (equivalent to archaeal Rrp42), the PH-2 domain (equivalent to archaeal Rrp41), a PNPase unique helical domain (H), and a KH and S1 RNA binding domain (equivalent to archaeal Rrp4), and this is summarised in Figure 1.11 (a). Archaeal exosomes can also have a Csl4-DnaG RNA binding domain; however, this contains a S1 and Zn-finger domain that is not equivalent to the PNPase homologs. Moreover, structural studies, summarised in Figure 1.11 (b), highlight the similar architecture of PNPases to the archaeal exosome; PNPase forms a hexamer ring-like structure from a homotrimer of PH-1/PH-2 domains, whereas the archaeal exosome forms a similar hexameric structure with a trimer of Rrp41/Rrp42 dimers. Additionally, like the KH and S1 RNA binding domains, which form their

own ring on the top face of the PH-domain hexamer in PNPase, a heterotrimer of Rrp4/Csl4-DnaG RNA-binding proteins form a flexible ring on one face of the archaeal exosome (not shown) (SsoExosome; Lorentzen & Conti, 2005, 2012; Lorentzen *et al.*, 2005; Lorentzen, Dziembowski, Lindner, Seraphin, & Conti, 2007; Lu, Ding, & Ke, 2010).

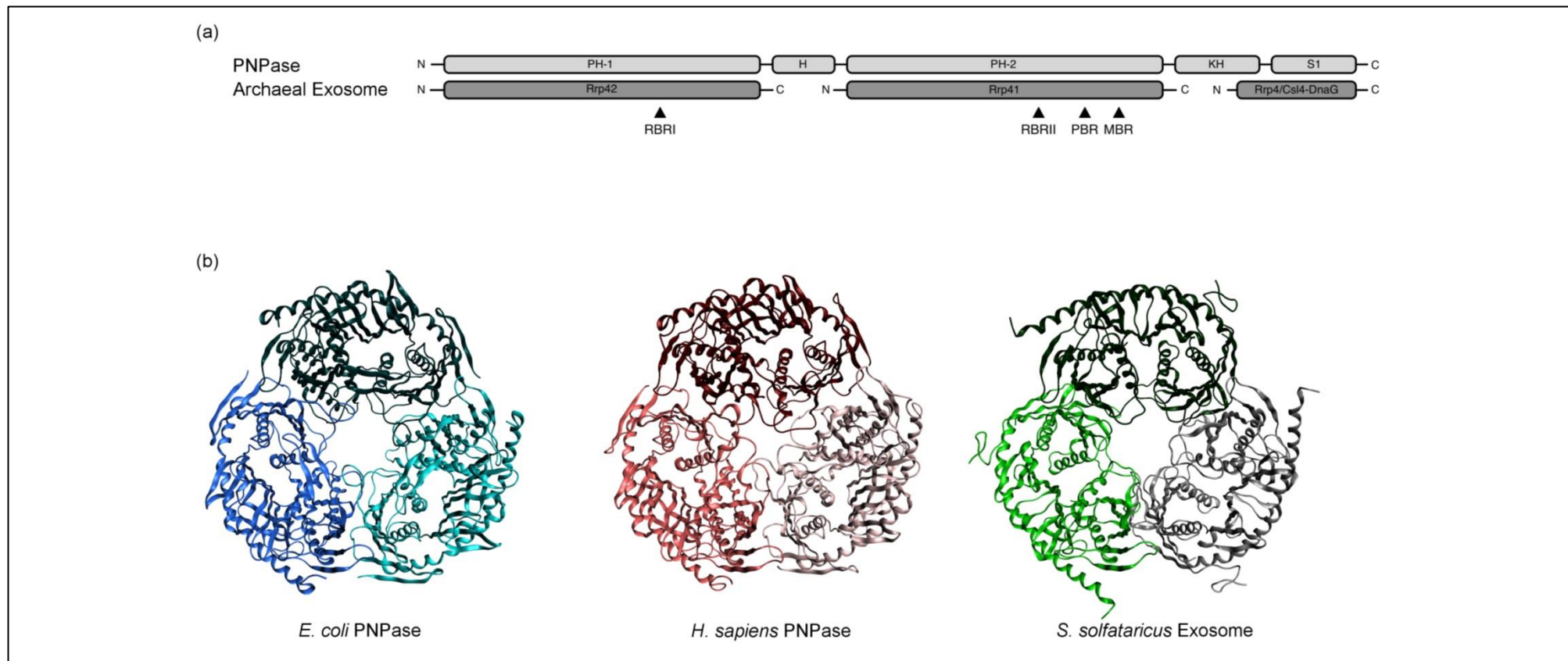


Figure 1.11 Structural Conservation of 3'-5' PNPase Homologs

(a) PNPase and Archaeal Exosome protein domains are illustrated in the schematic. The RNA Binding Region (RBRI) in PNPase PH-1 domain and its homologous archaeal exosome protein Rrp42 are indicated. A helical domain is shown to link the PH-1 and PH-2 domains in PNPase only. A second RNA Binding Region (RBRII), a Phosphate and Metal Ion Binding Region (PBR and MBR, respectively) are indicated within the PNPase PH-2 domain and its homologous archaeal exosome protein Rrp41. The KH and S1 domains in PNPase and corresponding archaeal exosome protein homolog Rrp4 are also shown. These proteins form a trimeric ring; (b) the solved crystal structures of EcPNPase (3GCM), hPNPase (3U1K) and SsoExosome (4BA1) are shown, with individual monomers, in each protein, coloured different shades of blue, red and green respectively. Structural images were generated in MOE (Molecular Operating Environment, 2013) and prepared using GIMP (v2)(GIMP, n.d.).

Although the domain architecture and structural similarity of enzyme homologs are a good early indication of a functional similarity, a more detailed sequence alignment is required to evaluate the level of sequence conservation. Previous publications, utilising both structural and mutagenesis approaches, have helped define the architecture of the PNPase active site. This research has shown that the active sites of EcPNPase, hPNPase and SsoExosome are all composed of four sub-sites: two RNA Binding Regions (RBRI and RBRII), an inorganic Phosphate Binding Region (PBR) and a divalent Metal ion Binding Region (MBR), which are highlighted within Figure 1.11 (a) (Portnoy, Palnizky, Yehudai-Resheff, Glaser, & Schuster, 2007). The catalytic centre is located entirely within the PH-2 domain and comprises the MBR and PBR. This alignment and the information gained from the study conducted by Portnoy and co-workers was valuable, however, in order to compare the other PNPase species utilised within this work, another alignment was required.

By comparing the conservation of important active site residues, which are known through previous mutagenesis and crystallography experiments to interact with key ligands, deductions can be made about PNPase homologs for which functional information is limited or not available. In theory this may help predict if the characterised citrate interaction in the prokaryote *E. coli* PNPase, and thus its regulatory effects, are conserved in eukaryotes and/or archaeal homologs.

1.6.4 Use of Techniques, Including Molecular Docking, in Ligand-Protein Interaction Research

The previous Section 1.6.3 described how initial protein sequence alignments using bioinformatics tools can help provide information about a protein of interest. If structural data are available, even more information may be acquired; potentially domains, or even individual residues of a protein can be discovered which may be of interest to study further.

As mentioned previously, in terms of PNPase a variety of structures are available, however unlike hPNPase and EcPNPase, the SsoExosome was not co-crystallised with citrate bound. In order to see if citrate does interact with this enzyme, sophisticated molecular characterisation techniques such as X-ray crystallography or NMR spectroscopy could be used. In terms of NMR, the SsoExosome is above the typical molecular weight limits of solving its overall 3D structure, although ligand binding may be examined through observing chemical shift perturbations in the presence of citrate. To examine citrate binding to SsoExosome, X-ray crystallography can also be conducted. If citrate is present in the crystallisation buffer or soaked in after SsoExosome crystals have formed, the resulting crystals may produce more information about where citrate binds and what residues are involved.

Alternatively, if a structure is already solved, an alternative approach can be used to predict ligand binding *in silico*. This is much easier to conduct compared to crystallography, as getting crystals

can sometimes be difficult. For example, if a structure is already solved, *in silico* ligand binding can be predicted using molecular docking calculations relatively quickly. Alternatively, if the structure is not available and only the protein sequence is known, a model can be predicted based on a homologous protein using molecular modelling tools.

Once a structure is solved experimentally by one of the various sophisticated techniques mentioned above i.e. NMR or X-ray crystallography, the information is deposited within the RCSB protein data bank (PDB), which is a repository of atomic coordinates and other information describing proteins and other important biological macromolecules (Berman *et al.*, 2000). Important, detailed information about the location of every atom in a molecule relative to one another is recorded in a text file and is accessible for further use. More specifically, pdb.txt files contain a list, including the type of atom (ATOM standard residues: amino acid/nucleic acid and HETATM nonstandard residues: water, inhibitors, cofactors, ions, and solvent respectively), the atom's serial number and its name (i.e. carbon (C), nitrogen (N)). Details of the residue type, chain identifier, residue number, and X, Y, Z orthogonal Å coordinates are also provided, along with structural features such as the presence of α -helices, β -sheets and disulphide bond linkages. Collectively, molecular visualisation programs (e.g. PyMOL) use all this information to generate 3D structures from input data and create images like those shown in Figure 1.11 (b). This process is incredibly valuable; picturing 3D structural information may not only provide insight into how a molecule such as a protein may function, just by its overall shape and size, but comparison to other similar structures previously determined, may also shed light on its role. For example, comparison of the full-length *H. sapiens* and *C. crescentus* PNPase structures with the core *E. coli* PNPase structure (KH truncation), suggested that the RNA-binding KH domain in PNPase regulates the size of the central pore channel. This work also proposed that a more constructed central pore is more efficient for RNA binding and cleavage; highlighting the use of comparative studies for revealing potential mechanisms of enzyme activity. In addition, the appearance of well-defined, ligand accessible pockets within a protein can suggest a possible means of functional regulation, especially if ligands/inhibitors are known.

Molecular modelling and docking programs, for example the Molecular Operating Environment (MOE) program, can utilise PDB.txt file information to generate the same 3D structures seen in visualisation programs such as PyMOL (Molecular Operating Environment, 2013). However, they possess the added prospect of creating *in silico* 3D models from protein sequences (with an unknown structure) using the solved structure as a template. These sophisticated programs (e.g. MOE) can also dock individual molecules or databases containing large libraries of molecules into the structure, using a number docking algorithms. These small molecule databases are easily accessible for virtual screening from public resources such as the Zinc database (Sterling & Irwin, 2015) and commercial websites such as Sigma-Aldrich. Utilisation of these kinds of databases allows a variety of commercially available ligands to be docked into the pockets of a receptor

protein. It is the docking algorithms which then allow docking of these ligands to be conducted, and they use various ligand orientations, to predict if there is a ligand conformation which favourably binds to the target site. Strengths of association (binding affinities) between a ligand and a receptor can be predicted using scoring functions, which suggest if the interaction will likely form a stable complex. There are a variety of different programs and algorithms available to conduct *in-silico* structural molecular modelling and docking and MOE is an example of a well-accepted program with a broad range of features and free online resources (Molecular Operating Environment, 2013). Other researchers have used MOE for molecular docking to study enzyme-metabolite interactions in evolutionarily distinct organisms; specifically to investigate the interactions of endocrine active substances (EASs) and their metabolites towards the ligand binding domain (LBD) of the androgen receptor (AR) in three distantly related vertebrate species (Galli *et al.*, 2014). In summary, MOE is an accepted approach to study the conservation of protein-ligand interactions, which is applicable to the studies investigating citrate-binding studies in evolutionarily distinct organisms.

1.6.5 Exploring PNPase Activity

The 3'-5' degradation activity of PNPase has been well established and the methods of these assays can be applied to study metabolite-RNase interactions. For example, the inhibition of the TCA metabolite citrate on canonical PNPase degradation activity in *E. coli* has been documented (Nurmohamed *et al.*, 2011). Although the effect of other TCA metabolites hasn't been examined, the energy containing molecules ATP and GTP, have been shown to inhibit PNPase (Del Favero *et al.*, 2008). Also this inhibition appears to be specific to purine nucleoside triphosphates; the pyrimidine nucleotides CTP and UTP did not exert any appreciable effect (Del Favero *et al.*, 2008).

Despite the high structural conservation of PNPase (Section 1.6.3), other studies have published results which suggest differences exist between homologs from different organisms. For example, nucleoside diphosphates (a mixture of 20 μ M ADP, CDP, GDP and UDP) were observed to effect the activity of *Streptomyces coelicolor* PNPase but not EcPNPase (Chang, Cozad, Mackie, & Jones, 2008), and alarmones have been suggested to regulate PNPases in some prokaryotic species but not others (Gatewood & Jones, 2010). The study by Chang and colleagues also reported differences in RNA substrate preference; PNPase from *E. coli* was 2-fold more effective at degrading the single stranded 3' tail of a substrate derived from the rpsO-pnp operon of *S. coelicolor* than PNPase from either *S. coelicolor* or *S. antibioticus*. Although the *S. coelicolor* PNPase was more effective than EcPNPase in degrading a substrate with a 3' stem-loop structure (Chang *et al.*, 2008). Other factors which regulate PNPase activity have also been described to vary over evolution. For example, EcPNPase displays optimal activity at phosphate concentrations exceeding 10 mM, whereas the optimal activity of hPNPase was found at 0.1 mM, and is inhibited at higher concentrations (Portnoy *et al.*, 2007). It is thought that the localisation of these two enzymes may play a role in their different preferences for inorganic phosphate; EcPNPase may

have evolved to function in the presence of high concentrations of phosphate (10-20 mM) in bacteria and organelle matrices, and although the concentration of phosphate in the IMS in human cells is unknown, hPNPase may have acquired a low phosphate preference for its function within this location (Portnoy *et al.*, 2007).

1.6.6 TCA Metabolites Affecting PNPase

The information provided so far within this Section 1.6 has mainly focused on the various factors effecting PNPase activity and whether the citrate-PNPase interaction observed in *E. coli* is conserved. The following section now focuses on the TCA cycle and whether other TCA metabolites, similar to citrate, may also regulate RNA turnover.

1.6.6.1 The TCA Cycle

Citrate was established by Hans Krebs to be an intermediate of the aerobic process which yields the energy required for living organisms (reviewed in Kornberg, 2000). Named after its discoverer, the Krebs Cycle, also commonly referred to as the citric acid or TCA cycle contains a number of metabolite intermediates, as shown in Figure 1.12.

The TCA cycle is the second of three stages which utilise organic fuel molecules in a series of chemical reactions leading to the overall generation of energy. More specifically, in aerobic organisms, carbohydrates, proteins and fats can be oxidised to yield CO₂ and H₂O via glycolysis, followed by the TCA cycle and finally the respiratory chain (also termed the electron transport chain). In summary, pyruvate derived from glucose and other sugars by glycolysis, is initially converted to acetyl-CoA in an irreversible oxidation process, where the carboxyl group is removed as CO₂ and the acetyl group moiety is linked to CoA (Figure 1.12). This two carbon acetyl group can now be fed into the TCA cycle and is donated from acetyl-CoA to the four carbon compound oxaloacetate, to form the six carbon citrate molecule (Figure 1.12). The multiple chemical reactions, within each turn of the cycle, utilise one acetyl group and generate two molecules of CO₂. Furthermore, the energy released is conserved in nucleoside triphosphate (ATP or GTP) and in the reduced electron carriers NADH (3x) and FADH₂ (1x), which are subsequently oxidised, giving up protons (H⁺) and electrons that are used in the respiratory chain to generate ATP. Not only does this process result in the overall generation of energy required for the cell, some intermediates of the process are utilised as precursors for a wide variety of products (as reviewed in Lehninger, Nelson, & Cox, 2000).

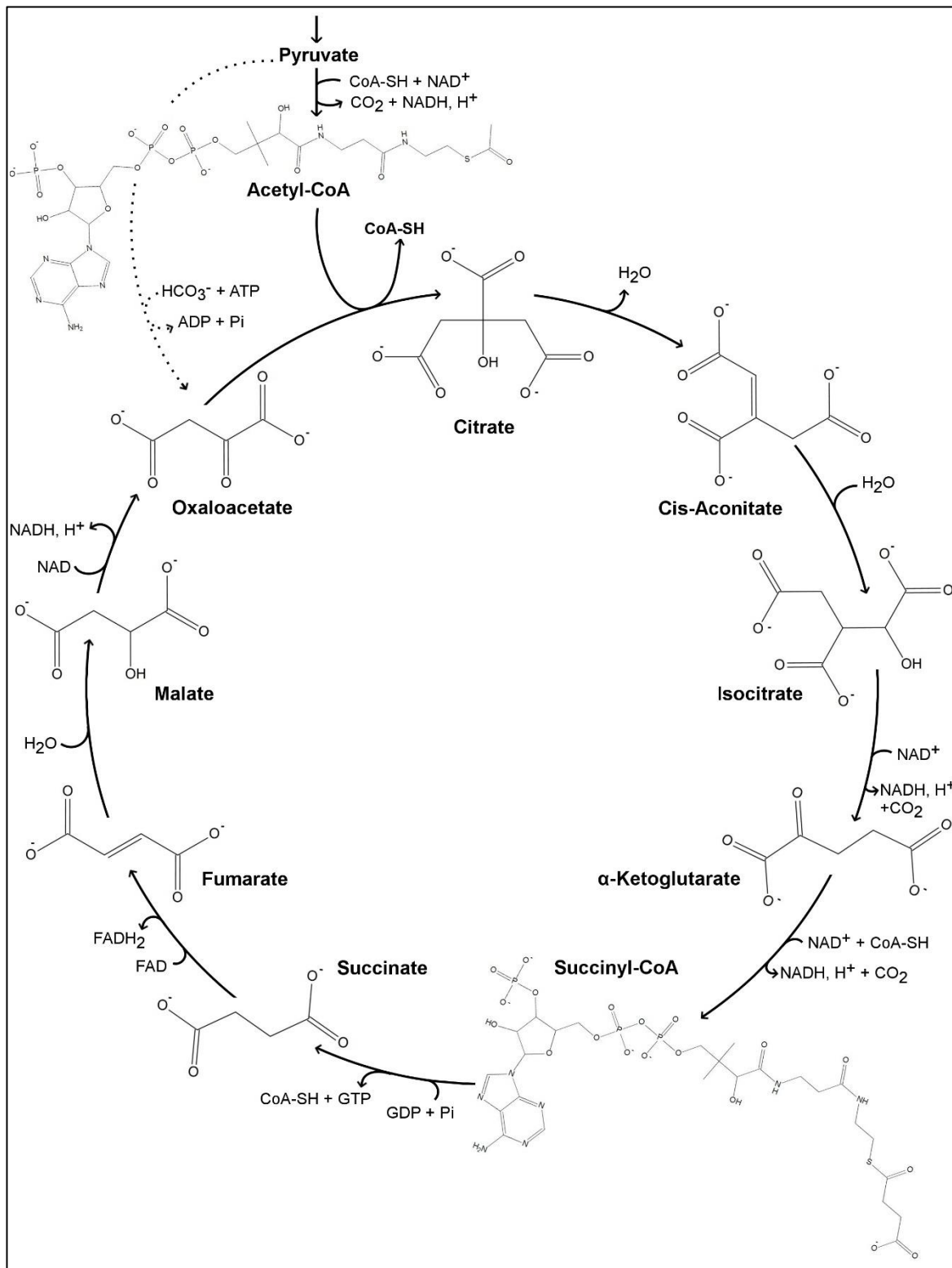


Figure 1.12 TCA Cycle

The main stages of the TCA cycle. The multiple chemical reactions, within each turn of the cycle, utilise one acetyl group and generate two molecules of CO_2 . The energy released is conserved in nucleoside triphosphates (ATP or GTP), NADH (3x) and FADH_2 (1x), which are subsequently used in the respiratory chain to generate more ATP and thus energy. 2D representations of metabolites were generated in MOE (Molecular Operating Environment, 2013) and the figure was prepared in GIMP (v2)(GIMP, n.d.).

1.6.6.2 Evolution of The TCA Cycle

The discovery of the TCA cycle explained how microorganisms could oxidise acetate to CO_2 and H_2O , but failed to explain how these organisms also survived on acetate as a sole carbon source (Kornberg, 2000). This was until a bypass step was discovered in which acetyl-CoA reacted with

another TCA metabolite, isocitrate, to generate two further TCA metabolites, malate and succinate. The sequence was thus: 1) isocitrate \rightarrow succinate + glyoxylate and then 2) glyoxylate + acetyl-CoA \rightarrow malate + CoA. Effectively this 'glyoxylate bypass/shunt' reaction meant carbon was not lost as CO₂ and the two cycles could be combined and termed the 'glyoxylate cycle' (Kornberg, 2000).

The evolution of the 'complete' TCA cycle was studied in detail by Huynen, Dandekar, & Bork in 1999. In this study, Huynen and colleagues utilised the availability of completely sequenced genomes and compared the presence of genes, encoding the enzymes that catalyse the series of TCA reactions, over evolutionarily distinct organisms. The genomes of four archaea, fourteen bacteria and one eukaryote were examined and variations such as the reductive and branched TCA cycle, the glyoxylate shunt and the pathways connecting the TCA cycle to pyruvate and phosphoenolpyruvate were examined. A graphical display of the reaction steps provided within this publication, is shown in Figure 1.13. It is obvious that the reactions of the TCA cycle in each species are diverse and have evolved with varying degrees of 'completeness' (Figure 1.13). For example, only the four largest genomes, including *E. coli*, *Bacillus subtilis*, *Mycobacterium tuberculosis* and *S. cerevisiae* and the smaller genome from *Rickettsia prowazekii* contain a complete TCA cycle, while others contain gaps or the TCA cycle is completely absent. Within the genomes lacking a complete cycle, there seems to be more conservation between steps 6-8 (Figure 1.13); from succinate to oxaloacetate (oxidative cycle) and less conservation in steps 1-3, from acetyl-CoA to 2-ketoglutarate (Figure 1.13). The research noted that although incomplete, the metabolites usually generated in the reactions which are missing, are instead produced in other ways. This means that metabolites such as 2-ketoglutarate, oxaloacetate and succinyl-CoA, which are starting points for other pathways including the synthesis of glutamate, aspartate and porphyrin respectively, are still produced from pyruvate. They observed that autotrophic bacteria generated these metabolites via the right branch of the TCA (oxidative direction, clockwise Figure 1.13(a)), whereas the methanogenic archaea and *Archaeoglobus fulgidus* generate them via the left branch (reductive direction, anti-clockwise Figure 1.13(a)).

The variations of the TCA cycle presented within Figure 1.13 provide an indication of its evolutionary origins (Huynen, Dandekar, & Bork in 1999). It is thought that the TCA cycle originally evolved as two separate pathways, stemming from pyruvate to either 2-ketoglutarate as an oxidative branch, or succinyl-CoA in a reductive branch. Huynen and colleagues also comment that alternative potential intermediate stages, are present in other pathways. Particularly two archaeal species *Methanococcus jannaschii* and *Methanobacterium thermoautotrophicum* have homologs of isocitrate dehydrogenase and aconitase in the leucine biosynthesis pathway. They noted that the methanogenic lifestyle of these organisms is presumably very old; as reflected by its presence throughout the euryarchaeal phylogeny. The only archaeal isocitrate dehydrogenase found in *A. fulgidus* probably arose through horizontal gene transfer from the bacteria. In the bacteria the

TCA cycle is reported to have been complete by the time of the proteobacteria emergence, and the incomplete TCA in pathogens is likely due to a secondary loss of genes.

Since the publication by Huynen and colleagues, more evidence has suggested our understanding of TCA cycles, in individual species, still remains to be improved. Previously it was generally accepted that cyanobacteria, including *Synechocystis sp.* PCC 6803 had an incomplete TCA cycle, however the discovery of novel genes has changed this view (Zhang & Bryant, 2011).

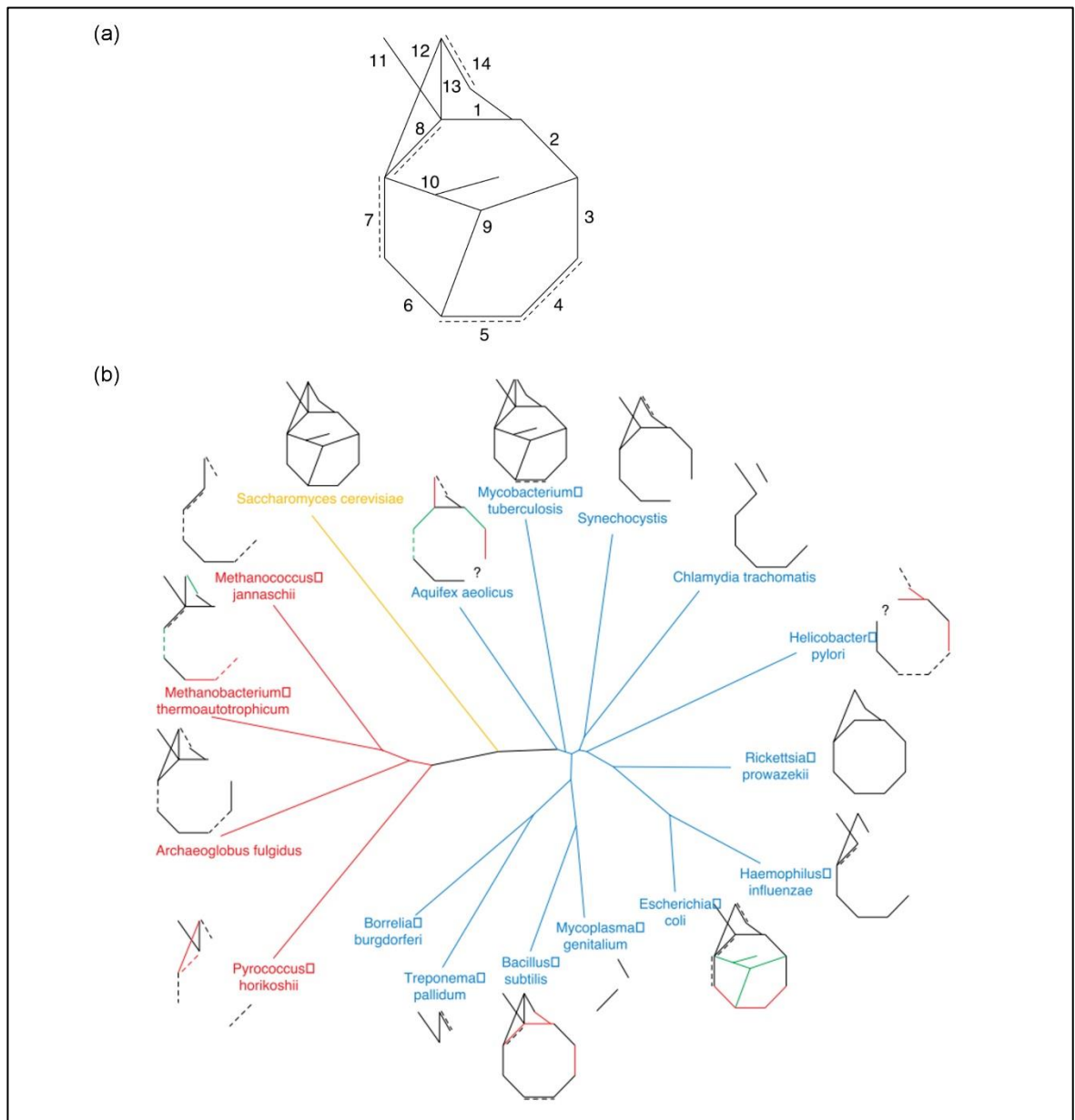


Figure 1.13 TCA Cycle Evolution

(a) TCA cycle reactions. When there are two enzymes that are not homologous but catalyse the same reaction (non-homologous gene displacement), one is marked with a solid line and the other with a dashed line. The oxidative direction is clockwise. The enzymes with their EC numbers are as follows: 1, citrate synthase (4.1.3.7); 2, aconitase (4.2.1.3); 3, isocitrate dehydrogenase (1.1.1.42); 4, 2-ketoglutarate dehydrogenase (solid line; 1.2.4.2 and 2.3.1.61) and 2-ketoglutarate ferredoxin oxidoreductase (dashed line; 1.2.7.3); 5, succinyl-CoA synthetase (solid line; 6.2.1.5) or succinyl-CoA-acetoacetate-CoA transferase (dashed line; 2.8.3.5); 6, succinate dehydrogenase or fumarate reductase (1.3.99.1); 7, fumarase (4.2.1.2) class I (dashed line) and class II (solid line); 8, bacterial-type malate dehydrogenase (solid line) or archaeal-type malate dehydrogenase (dashed line) 3 (1.1.1.37); 9, isocitrate lyase (4.1.3.1); 10, malate synthase (4.1.3.2); 11, phosphoenolpyruvate carboxykinase (4.1.1.49) or phosphoenolpyruvate carboxylase (4.1.1.32); 12, malic enzyme (1.1.1.40 or 1.1.1.38); 13, pyruvate carboxylase or oxaloacetate decarboxylase (6.4.1.1); 14, pyruvate dehydrogenase (solid line; 1.2.4.1 and 2.3.1.12) and pyruvate ferredoxin oxidoreductase (dashed line; 1.2.7.1). (b) TCA cycles of individual species, together with the phylogeny of the species. The major kingdoms of life are indicated in red (Archaea), blue (Bacteria) and yellow (Eukarya). Question marks represent reactions for which there is biochemical evidence in the species itself or in a related species but for which no genes could be found. Genes that lie in a single operon are shown in the same colour (black, red or green). Figure and legend taken directly from (Huynen *et al.*, 1999).

As reviewed in Lehninger *et al.*, 2000, the location of the TCA cycle varies over evolution; in prokaryotes the TCA cycle occurs in the cytosol, with the proton gradient for ATP production (electron transport pathway) occurring across the plasma membrane. Contrastingly, in eukaryotes,

the TCA cycle is present in the matrix of the mitochondria and ATP production occurs in the inner membrane.

In summary, although over evolution the exact location, the diversity of reaction steps and the completeness of the TCA cycle is varied, there seems to be a tendency, for a wide range of organisms, to converge to this method of energy production (Huynen *et al.*, 1999). Thus, the mechanisms regulating the TCA cycle and in turn, the downstream processes affected by this metabolic pathway, are clearly important to understand. The exact enzymes involved in the TCA cycle, and thus the pool of metabolites available to an organism, appear to vary between different species (Figure 1.13 (b)). Whether RNases can sense the pool of available metabolites and coordinate their activity accordingly, remains to be discovered. The TCA cycle in *E. coli* and *H. sapiens* is well characterised and they are examples of excellent prokaryote and eukaryote model systems with a full TCA cycle. However, the completeness of the TCA cycle in archaea is less understood and so the effects of TCA metabolites on PNPase was the main focus of this work.

1.6.7 Regulating RNA Turnover in Response to Cellular Metabolism Summary

In summary, the information provided within this Section 1.6 highlights the recurrent evolution of physical associations between the enzymes of RNA degradation and central metabolism and suggests a possible conserved metabolite-RNase communicative link. The impact of a communicative link between RNases and central metabolism is now discussed in Section 1.7, so that the importance of studying these interactions in more detail can be identified.

1.7 Thesis Focus: Exploring the Impact of a Metabolite-PNPase Conserved Link

The proposed work will not only provide new information and scientific advancement in the field of post-transcriptional gene regulation research, its potential impact could also be to improve the development of biotechnology applications.

In terms of advancing our understanding of RNase regulation, if a conserved metabolite-RNase communication link truly exists it would add key information to the post-transcriptional gene regulation network shown in Figure 1.14. If a cell were able to sense the pool of metabolites available and control gene expression accordingly, it would add another tier, and feedback loop, to the hierarchal process which regulates the flow of cellular information. It would also provide the basis for further research; although this study aims to determine if a conserved communication link exists *in vitro*, undeniably the field would also benefit from a more detailed *in vivo* characterisation.

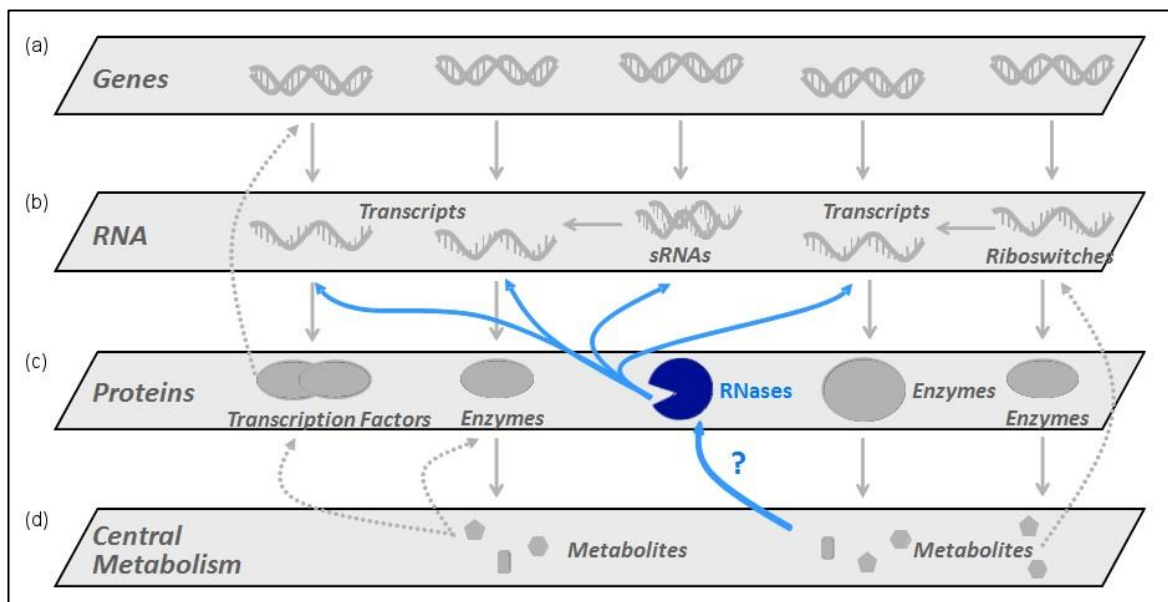


Figure 1.14 Cellular Information Flow: Metabolite-RNase Communication

The flow of cellular information. Whether central metabolism can control RNases (blue arrow/question mark) remains to be discovered. Image adapted from Prof. Anastasia Callaghan.

The general field of systems biology may also be impacted; with important *in silico* validation and data generated, computational modelling specialists in universities and research institutes, may be able to use the methods of metabolite docking into PNPase homologs for similar studies. Potentially revealing inhibitory ligands, that were not testing within this work, and could subsequently be tested experimentally. If metabolite-PNPase and metabolite-archaeal exosome communicative links are conserved, it may infer an ancient and important communication mechanism. In which case, other RNases could be investigated; possibly including elements of the bacterial degradosome.

A greater understanding of the factors effecting metabolic changes within a cell, potentially can be utilised with both therapeutic and industrial benefits in mind. A wide variety of industries thrive on exploiting cellular systems to produce products, including drugs, biofuels and food. Understanding these cellular systems not only allows improvements of current processes but also the development of novel technologies. For example, by targeting and artificially manipulating the metabolome of host systems in the process of manufacturing drugs, the amount of starting culture materials may be reduced and as a result of increased product yields, profits may be increased. Furthermore, a deeper understanding of global regulatory mechanisms controlling gene expression may also provide invaluable insight into how disease states arise. For example, the role of PNPase in cancer, chronic inflammatory responses and coronary disease (Wilusz *et al.*, 2001) may be revealed. This could help towards the development of medical countermeasures against the plethora of disease states associated with faulty gene expression and metabolism.

With wide-spread impact on scientific understanding, biotechnology engineering and advancement of medical therapeutics, the role of metabolites in ribonuclease activity regulation was the subject of this thesis; the aims of which are provided in the subsequent Section 1.7.1.

1.7.1 Thesis Aims

Regulation of post-transcriptional gene expression is essential for maintaining cellular homeostasis and co-ordinating the process of RNA turnover is a number of ribonucleases, including polynucleotide phosphorylase (PNPase). Following the observation that the TCA cycle metabolite citrate was observed to modulate the activity of *E. coli* PNPase *in vitro* and *in vivo*, Nurmohamed and colleagues proposed that a feedback mechanism may exist between central metabolism and RNA turnover. This is further substantiated by the studies reviewed in this chapter, that identify nucleotides (ATP and GTP), a secondary signalling molecule (c-di-GMP) and a stringent response alarmone signalling molecules (ppGpp) interacting with proteins involved in post-transcriptional gene regulation (Del Favero *et al.*, 2008; Gatewood & Jones, 2010; Tuckerman, Gonzalez, & Gilles-Gonzalez, 2011). Therefore, the primary purpose of this study is to both explore the conservation of citrate-PNPase interactions across evolution and to examine whether other metabolites may also play a role in affecting RNA turnover. This study will provide evidence that citrate regulation may be conserved in all three domains of life, across prokaryotes, eukaryotes and archaea. Furthermore, by primarily focusing on metabolites of central metabolism and looking at the key functional groups/features that may be important in PNPase regulation; this study will also propose that a metabolite-PNPase interaction may be particularly maintained in complex multicellular organisms such as *H. sapiens*.

Details of all the materials and methods required for this study, including the reagents and general biochemical/microbiology techniques utilised for preparing and analysing the synthesis of various biological samples, will be provided in Chapter 2. Initially, this study involves cloning, expression and purification of recombinant PNPase homolog representatives across the three domains of life, including the prokaryotic *E. coli* PNPase, eukaryotic *H. sapiens* PNPase and an archaeal *S. solfataricus* exosome. This work is described in detail in Chapter 3, along with assessment of protein purity and identification of canonical 3'-5' exoribonuclease activity *in vitro*, using gel-based electrophoresis techniques. In Chapter 4, *in silico* protein sequence alignments and molecular docking calculations will be utilised to investigate the conservation of citrate-binding residues across evolutionarily distinct PNPase homologs. The effect of citrate on 3'-5' RNase activity, is then examined using *in vitro* gel-based assays. By combining evidence from *in silico* and *in vitro* techniques, this work will propose that the attenuation of PNPase activity by citrate is widespread across evolution. To investigate the effects of metabolites on PNPase activity, a similar approach of combining molecular docking predictions of metabolite-binding with gel-based assays, is applied in the subsequent Chapter 5. The given evidence suggests that other metabolites may also play a role in affecting RNA turnover. A potential method for improving the kinetic analysis of RNA

degradation activity, mediated by PNPase homologs, is explored in Chapter 6. The application of a commercially available high-throughput plate reader assay, for determining 3'-5' RNase activity in real-time, is examined so that important kinetic activity parameters can be rapidly and accurately calculated. Finally, the data presented within this study will be summarised in Chapter 7 and the impact of the proposed communicative link, between metabolites and PNPase homologs, is discussed.

2 Materials and Methods

2.1 Chapter Aims

This chapter provides details of the preparation of materials and experimental methods used as standard, unless otherwise specified, throughout this thesis. Within the following chapter subsections, in addition to the detailed methods, a basic description of the rationale/theory behind each technique is provided if appropriate; for example, if a method/technique is extensively used in subsequent chapters.

2.2 Chemical and Biochemical Reagents

All chemicals and biochemical reagents were provided by Fisher Scientific, except where stated otherwise.

2.3 Buffer and Media Recipes

Solutions, buffers and growth media were prepared with 18.2 Ω deionised water (dH₂O) (Purite) and filter sterilised using a 0.22 μ m filter (Millipore). Non heat-sensitive components were autoclaved at 120°C, 15 pound-force per square inch (psi) for 20 minutes, if they were required for bacterial growth or for techniques requiring sterile conditions.

2.4 Marker Proteins and Nucleic Acids

BenchMark™ and Pre-Stained SeeBlue® (Standard and Plus2) protein molecular weight markers were supplied by Life Technologies. 100 base pair (bp) DNA ladder, low molecular weight (LMW) DNA ladder and Low Range ssRNA Ladder were supplied by New England BioLabs (NEB N3231S, N3233S and N0364S respectively). Details of all ladders for gel electrophoresis are provided in Appendix 9.2.

2.5 Antibiotics

Antibiotics were prepared as concentrated stocks (1000 x) and were dissolved in either dH₂O or 100% EtOH, filter sterilised using a 0.22 μ m filter and stored at -20°C in 1 ml aliquots, see Table 2.1 for details.

Antibiotic	Stock Concentration (mg/ml)	Solvent	Dilution Factor Required	Final Working Concentration (µg/µl)
Chloramphenicol	34	100 % EtOH	1/1000	34
Kanamycin	25	Water	1/1000	25
Ampicillin	100	Water	1/1000	100
Tetracyclin	10	Water	1/1000	10

Table 2.1 Antibiotic Preparation

Antibiotics are listed with the stock concentrations (mg/ml) and dilution factors required to obtain the final working concentrations (µg/µl). The solvent in which they were diluted is also provided.

2.6 Antibodies

Poly-histidine tagged recombinant proteins were probed with either a primary antibody (1° antibody): His₆-probe polyclonal IgG (G-18) Sc-804 produced in rabbit (Santa Cruz Biotechnologies), followed by a secondary antibody (2° antibody): anti-rabbit IgG (whole molecule) peroxidase conjugate, produced in goat (Sigma A0545), or directly with a single His₆-probe polyclonal IgG horseradish peroxidase (HRP) conjugate, produced in rabbit (Abcam AB1187). Human PNPase was probed with a 1° antibody PNPase polyclonal IgG, produced in rabbit (GeneTex, Source Bioscience GTX116542S), and the same anti-rabbit-peroxidase 2° antibody mentioned previously (Sigma A0545).

2.7 Nucleic Acids: Deoxyribonucleic acid (DNA)

DNA stocks were prepared by the variety of methods listed below.

2.7.1 Provided Plasmid DNA

E. coli PNPase, cloned into an ampicillin resistant pET-Duet-1 vector, was kindly supplied by Prof. Ben Luisi (University of Cambridge). The provided plasmid DNA was transformed into bacterial cell strains for downstream purification, and subsequent sequencing, as described in Sections 2.9.5, 2.9.6 and 2.7.4 respectively.

2.7.2 Chemical Synthesis: DNA

DNA Primers for DNA sequencing were independently designed if Source Bioscience stock primers were not available. Using the guidelines provided by Source Bioscience, primers were designed to be 18-23 nucleotides (nt) long and bind specifically to the sequence ~ 50 nts upstream of region of interest (“Source BioScience,” n.d.). Primers were designed to have no self-complementarity, a Guanosine/Cytosine (GC) content of ~40-60% (with 2/3 G/Cs at the start of the designed primer sequence) and an annealing temperature (T_M) of 55-60°C. Mutagenic oligonucleotide primers were individually designed for site-directed mutagenesis (SDM) as recommended by the QuikChange® XL SDM kit manual (Agilent Technologies, n.d.). All oligonucleotides were synthesised by Invitrogen and were diluted with double deionised water (ddH₂O) to 1 mM and stored at -20°C. Primer sequences are listed in Appendix 9.3.

2.7.3 DNA Integrity and Concentration

DNA integrity and size were analysed by agarose gel electrophoresis, as described in Section 2.10.1 and DNA concentration was determined from the absorbance $A_{260\text{ nm}}$ measured using the NanoDrop 2000c Spectrophotometer (Thermo Scientific) and the Beer-Lambert equation below.

$$A = \epsilon_{260} Cl$$

Equation 2.1 Beer-Lambert Law

The equation describes Absorbance (A) as a function of the concentration (C, ng/μl), path length (l, cm) and molar extinction coefficient at 260_{nm} (ϵ_{260} , ng-cm/μl). The generally accepted extinction coefficients, at a wavelength of 260_{nm}, for double-stranded DNA and single-stranded DNA is 50 and 33 ng-cm/μl respectively. When using a NanoDrop 2000c Spectrophotometer (Thermo Scientific), with a path length of 1 cm and when $A=1$, the concentration for double-stranded DNA is 50 ng/μl and single stranded DNA is 33 ng/μl (Thermo Scientific, 2009).

2.7.4 DNA Sequencing

Sanger sequencing (Sanger, Nicklen, & Coulson, 1977) of plasmid DNA constructs was outsourced to either Life Technologies (provided with GeneArt synthesis service) or Source Bioscience. DNA sequences were then visualised and analysed using FinchTV™ sequencing chromatogram trace viewer V1.4, provided by Geospiza® (Geospiza, n.d.).

2.7.5 Site Directed Mutagenesis (SDM)

In vitro SDM was used to modify plasmid DNA to allow site specific mutation of the nucleotide sequence (Hutchison *et al.*, 1978). These mutations consequently change the translated amino acid sequence, thus either correcting sequence errors identified through sequencing (Section 2.7.4) or introducing point mutations for protein structural and or/ functional analysis. A summary of the SDM steps recommended within the QuikChange XL SDM Kit Instruction Manual (Agilent Technologies, n.d.) is provided in Figure 2.1

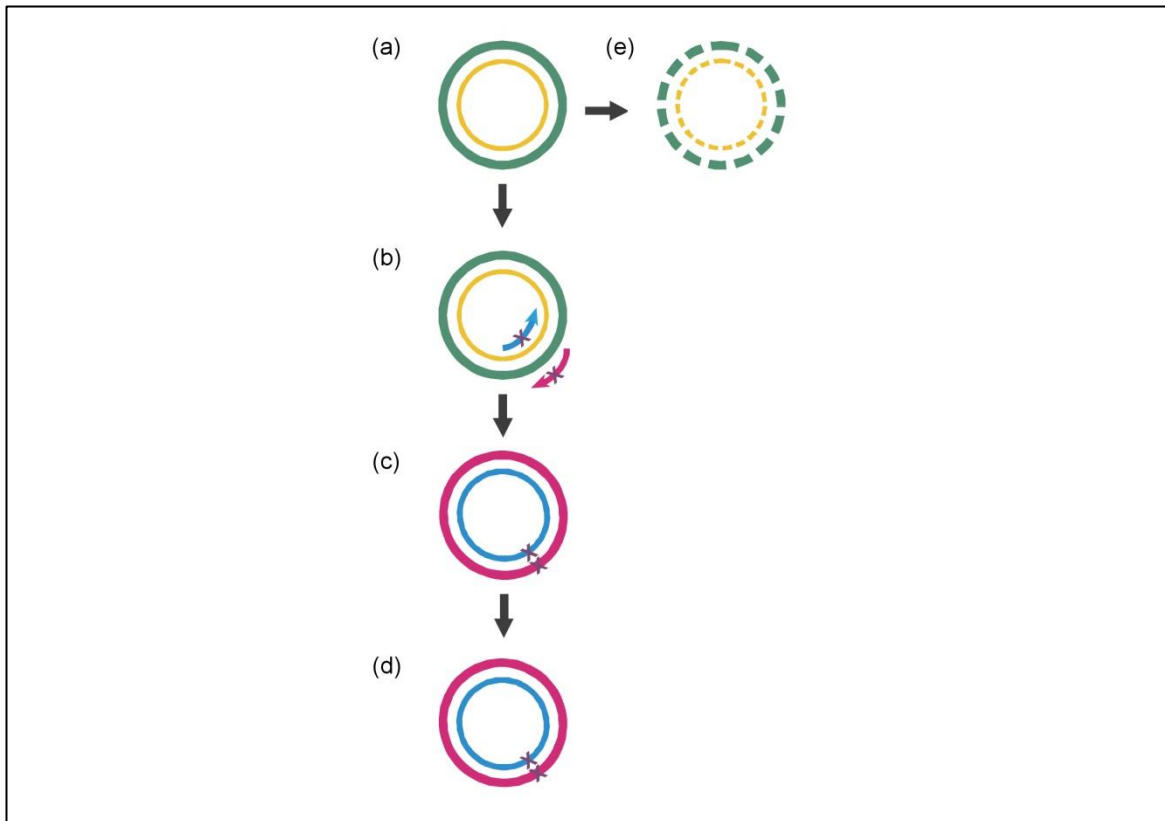


Figure 2.1 Overview of Site-Directed Mutagenesis (SDM)

(a) Double stranded DNA plasmid (green and yellow circles), selected for SDM, was denatured and (b) the mutagenic primers (pink and blue arrows) with the desired mutation (x) annealed. (c) Primers were extended with *PfuUltra* high fidelity (HF) DNA polymerase, generating a new DNA plasmid (pink and blue circles) with the desired mutation (x) within the sequence. (d) The DNA was further amplified by multiple rounds of PCR. (e) The parental methylated and hemimethylated DNA was digested with DpnI restriction endonuclease enzyme (green and yellow dotted circle) leaving only the intact mutated plasmid, which was transformed into competent *E. coli* cells for nick repair. Image modified from the QuikChange XL SDM Kit Instruction Manual (Agilent Technologies, n.d.).

Oligonucleotide primers were individually designed for SDM as described within the guidelines of the QuikChange XL SDM kit (Agilent) instruction manual (Agilent Technologies, n.d.). In brief, primers were designed to be between 25-45 bases in length, with the mutated nucleotide (s) located in the middle; with at least 10-15 bases on each side. Both primers were designed to anneal to the same sequence, but on opposite strands of the parental double-stranded DNA vector, with a $T_m \geq 78^\circ\text{C}$. Primers also had one or more terminal G/Cs with an overall GC content of 40%. Designed oligonucleotide primers were synthesized by Invitrogen™ and were diluted with nuclease-free H_2O , to a concentration of 50 ng/ μl , prior to storage at -20°C . The primer sequences used for SDM are listed in the Appendix 9.3. Mutagenesis reactions were set up, on ice, in 0.5 ml thin-walled PCR tubes as described below in Table 2.2. Reagent concentrations and volumes were used as recommended in the QuikChange SDM kit protocol and a pWhitescript mutagenesis control was used at all times.

(a)

Reagents	Stock Concentration	Volume (μ l)	Final Concentration
Reaction Buffer	10 x	5	1x
pWhitescript 4.5-kb Control Plasmid	5 ng/ μ l	2	10 ng
Control Primer #1 (34-mer)	100 ng/ μ l	1.25	125 ng
Control Primer #2 (34-mer)	100 ng/ μ l	1.25	125 ng
dNTP Mix		1	
QuikSolution Reagent		3	
ddH ₂ O		36.5	
Total		50	

(b)

Reagents	Stock Concentration	Volume (μ l)	Final Concentration
Reaction Buffer	10 x	5	1x
ds DNA Template	x ng/ μ l	x	10 ng
Primer #1	x ng/ μ l	x	125 ng
Primer #2	x ng/ μ l	x	125 ng
dNTP mix		1	
QuikSolution Reagent		3	
ddH ₂ O		x	
Total		50	

Table 2.2 *In vitro* SDM Reaction Preparation

SDM reagents are listed with the volumes, stock concentrations and the final concentrations for both the (a) control and (b) SDM PCR reactions. The deoxynucleotide mix (dNTP), specific primers and proprietary SDM kit reagents were added to double stranded (ds) DNA/control plasmid, as recommended in the QuikChange XL SDM Kit Instruction Manual (Agilent Technologies, n.d.).

To 50 μ l SDM reactions, 1 μ l *PfuUltra* high fidelity (HF) DNA polymerase (2.5 U/ μ l) was added for strand synthesis and the cycling parameters, outlined in Table 2.3, were used for DNA amplification in a GeneAmp® PCR System 9700 thermal cycler (Applied Biosystems). Following temperature cycling, the reaction tubes were placed on ice for 2 minutes to cool. SDM DNA amplification was confirmed by agarose gel electrophoresis; 10 μ l of the SDM product was run on a 1% agarose gel as described in Section 2.10.1. As recommended in the manual, even if DNA amplification could not be confirmed by gel, the DpnI endonuclease restriction digestion of the parental template was still conducted.

Segment	Cycles	Temperature (°C)	Time
1	1	95	1 minute
		95	50 seconds
2	18	60	50 seconds
		68	1 minute/kb of plasmid length
3	1	68	7 minutes

Table 2.3 QuikChange® II XL SDM Cycling Parameters

The target size of each plasmid DNA was considered (1 minute/ kb of plasmid length) when deciding SDM cycling parameters. The other cycling conditions were used, as recommended in the QuikChange XL SDM kit instruction manual (Agilent Technologies, n.d.).

To each 51 µl SDM reaction, 1 µl of Dpn I restriction enzyme (10 U/µl) was added and thoroughly mixed by gentle pipetting. Reactions were centrifuged briefly for 1 minute in a Thermo Scientific Heraeus Fresco 17 benchtop centrifuge, prior to incubating at 37°C for 1 hour in order to digest the template DNA (target sequence 5'-Gm⁶ATC-3'). Post-Dpn I treatment, 2 µl of undigested vector, containing the desired mutation, for both the sample and control reactions, were then transformed into chemically competent *E. coli* DH5α cells, alongside a pUC18 transformation control (Section 2.9.5). Following transformation, appropriate volumes of transformed cells were plated on selective LB agar plates as described later in Section 2.9.5; (Table 2.5) and for blue-white colony screening, the SDM control reaction was plated on X-Gal/IPTG plates (Section 2.9.1.1). All plates were then incubated at 37°C for >16 hrs and mutagenic efficiency was calculated from the number of colonies with β-galactosidase activity, as indicated by the presence of a blue colour upon screening. After plasmid replication and purification, as described in Section 2.9.6 respectively, all recombinant plasmid DNA constructs were sequenced following SDM to confirm the correct mutation(s) (Section 2.7.4).

2.8 Nucleic Acids: Ribonucleic acid (RNA)

RNA stocks were prepared by the variety of methods listed below.

2.8.1 Chemical Synthesis: RNA

All chemically synthesised, single stranded, RNA molecules were provided by GE healthcare (Dharmacon); including the 5' Fluorescein (Figure 2.2) labelled Poly(A)₂₀ RNA (5'F Poly(A)_{20mer}) and unlabelled RNA (Poly(A)_{20mer}), which were utilised in determining RNase 3'-5' degradation activity.

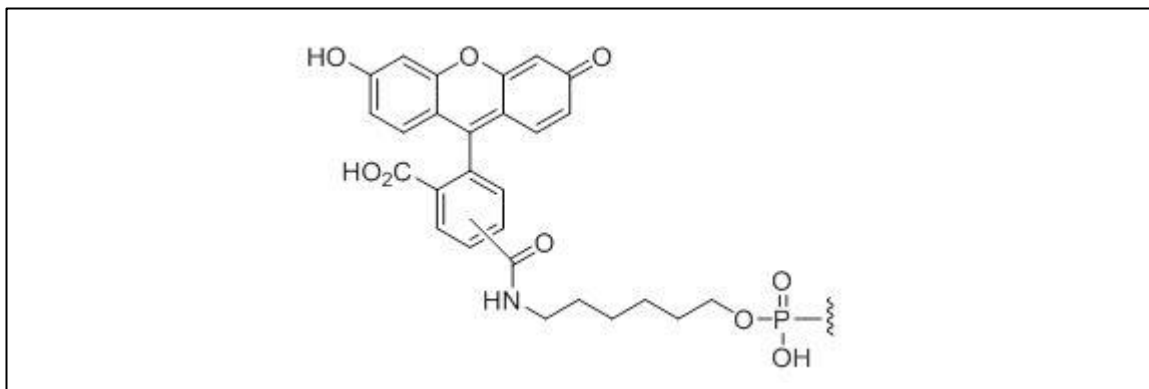


Figure 2.2 Fluorescein Structure

RNA substrates were chemically modified with the addition of a 5' fluorescein (F) label for detection after gel electrophoresis. Fluorescein (MW: 538.47) has an excitation/emission maximum of 494 nm and 525 nm respectively and so the Fujifilm FLA-5000 phosphorimager (blue laser/filter) was utilised for visualisation (Dharmacon.gelifesciences, n.d.).

2.8.2 RNA Integrity Concentration

RNA integrity, purity and size was analysed by polyacrylamide gel electrophoresis, as described later in Section 2.10.2.2 and RNA concentration was determined from the absorbance $A_{260\text{nm}}$ measured using the NanoDrop 2000c Spectrophotometer (Thermo Scientific). Absorbance measurements were conducted in triplicate and averaged to ensure reproducibility.

2.9 Prokaryote Cell Strains

The general microbiological techniques used to maintain prokaryotic cell strain stocks are provided below. The process of preparing these strains for transformation with recombinant plasmid DNA and subsequent DNA replication/purification is also described.

2.9.1 General Microbiological Techniques

Any media required for bacterial growth was autoclaved at 120°C, 15 psi for 20 minutes as standard. Where selective Lysogeny Broth (LB) agar and or liquid media were used for cell strain and/or plasmid selection, appropriate antibiotics were added. For antibiotic preparation and selection see Section 2.5 and Table 2.4 respectively.

		Antibiotic Resistance	Final Concentration (µg/µl)
<i>E. coli</i> Cell Strain	DH5α	None	0
	BL21(DE3)pLysS	Chloroamphenicol	34
	LOBSTR-BL21(DE3)	None	0
Plasmid	pET-Duet-1	Ampicillin	100
	pET28b	Kanamycin	25
	pETMCN-EAVNH	Ampicillin	100
	EcPNPase (pET-Duet-1)	Ampicillin	100
	hPNPase (pET28b)	Kanamycin	25
	SsExosome (pETMCN-EAVNH)	Ampicillin	100

Table 2.4 Antibiotic Selection

Various *E. coli* cell strains and plasmid DNA types are listed with their respective antibiotic resistances, the final working antibiotic concentrations (µg/µl) are also provided.

2.9.1.1 Lysogeny Broth (LB) Agar & Plate Preparation

In order to make one LB agar plate, typically between 30-40 ml of LB agar (LB Agar, Miller components per 1 L: 10 g Tryptone, 5 g Yeast extract, 10 g NaCl₂ and 15 g Agar, pH 7 at 25°C) was used. The LB agar was prepared as per the manufacturer's instructions (40 g to be dissolved in 1 L of dH₂O), the volume was adjusted as required and autoclaved as standard. Autoclaved selective LB agar was poured into petri dishes using aseptic technique. Agar plates were left to set at room temperature (RT) and were subsequently dried upside-down at 65°C for approximately 1 hour. If plates were not used immediately they were stored at 4°C and then re-dried, as described above, prior to use. Additionally, if blue-white colony screening was required, 100 µl of 10 mM IPTG (prepared in sterile dH₂O and filtered) followed by 100 µl of 2% X-Gal (prepared in dimethylformamide (DMF)) was spread on the LB agar. These were added 30 minutes prior to use, ensuring they did not mix to avoid precipitation. Depending on the cell culture plated onto the LB agar, the appropriate volumes of bacterial cells require to generate individual colonies varied; examples of which are listed in Table 2.5.

2.9.1.2 Lysogeny Broth (LB) Liquid Media Preparation

LB liquid media (LB Broth, Miller components per 1 L: 10 g Tryptone, 5 g Yeast extract, 10 g NaCl₂, pH 7 at 25°C) was prepared as per the manufacturer's instructions (25 g to be dissolved in 1 L of H₂O) and autoclaved.

2.9.2 Cell Strains

The *E. coli* DH5α cell strain (F⁻ Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK⁻, mK⁺) phoA supE44 λ⁻ thi-1 gyrA96 relA1) was used for plasmid preparation and long term plasmid storage. The *E. coli* BL21(DE3)pLysS ((F⁻ ompT gal dcm lon hsdSB(rB⁻ mB⁻) λ(DE3) pLysS(cmR)) cell strain was used for protein expression, for more details of cell strain genotypes see Appendix 9.4 ("*E. coli* genotypes," 2015). Strains were stored at -80°C except when in use, where they were kept on ice for the minimal amount of time to avoid thawing. The original DH5α

and BL21(DE3)pLysS cell strains were kindly provided by Dr James Youell and Dr Luke Evans (University of Portsmouth). As discussed in Swords (2003), although a variety of procedures for the transformation of *E. coli* are available; the modified calcium chloride chemical method utilised was selected due to ease, relative efficiency, and lack of need for a specialized apparatus such as an electroporator. Thus, fresh glycerol stocks of the donated bacterial strains were prepared and then made chemically competent for transformation as described in Sections 2.9.3-2.9.4 respectively.

2.9.3 Cell Strain Maintenance

Frozen *E. coli* bacterial strains were streaked out on selective LB agar plates. Following streaking, agar plates were incubated at 37°C for 12-16 hours. Large, single, well-isolated colonies displaying typical morphology were selected and these were used to inoculate 10 ml of selective LB liquid media. Bacterial cell cultures were grown at 37°C overnight (O/N) using an Innova 4400 shaking incubator at ~220- 225 revolutions per minute (rpm). 1 ml of the O/N culture was aspirated into a 1.5 ml Eppendorf tube, cells were centrifuged at 3,000 rpm in an Eppendorf Centrifuge 5810 R for 5 minutes and the supernatant discarded. Cell pellets were re-suspended in 1 ml of 30% (v/v) sterile glycerol and these glycerol stocks were stored at -80°C for future use.

2.9.4 Preparation of Chemically Competent *E. coli* Cell Strains

A baffled flask containing 250 ml of selective LB liquid media was inoculated with 250 µl of relevant O/N bacterial culture (1:1000; standard dilution). Bacterial cells were grown at 37°C in a shaking incubator at 220 rpm. The rate of bacterial growth was monitored using a Biowave CO8000 cell density meter until an optical density at 600 nm (OD_{600}) of 0.6 was acquired, at which point the culture was decanted into a pre-chilled 500 ml flat-bottomed centrifuge bottle. This was stored, at all times, on an ice slurry. Cells were harvested by centrifugation at 5,000 rpm at 4°C for 10 minutes in a Beckman Coulter Allegra 25R refrigerated centrifuge (rotor AT-14-10). The resulting supernatant was removed and discarded, and the bacterial cell pellet was gently re-suspended in 80 ml of ice cold CCMB80 buffer (10 mM KOAc, 80 mM $CaCl_2 \cdot 2H_2O$, 20 mM $MnCl_2 \cdot 4H_2O$, 10 mM $MgCl_2 \cdot 6H_2O$, 10% (v/v) glycerol, adjusted to pH 6.4 with 1 mM HCl) and incubated on ice for 20 minutes. Bacterial cells were re-centrifuged and the supernatant discarded as before. The pellet was once again re-suspended in 10 ml of ice-cold CCMB80 buffer and the cell density OD_{600} was calculated using a 10-fold dilute sample, as the Biowave CO8000 meters error increases at $OD_{600} > 1$. Additional CCMB80 buffer was added to reach a final OD_{600} of 1-1.5. For long term storage, 250 µl aliquots of each cell strain were stored in 1.5 ml Eppendorf tubes and frozen at -80°C.

2.9.5 Transformation of Chemically Competent *E. coli* Cell Strains with Plasmid DNA.

Chemically competent *E. coli* cell strains were thawed on ice and 50 µl was decanted into a pre-chilled 15 ml falcon tube. Typically, recombinant plasmid DNA was prepared as 10 ng/µl stocks

and 10 ng (1 μ l) of this was added to the competent cells alongside the following two controls; 1) a positive plasmid DNA control, which should transform efficiently (10 ng pET-28b) and 2) a dH₂O negative control. The cell suspension with added plasmid DNA/water was incubated on ice for 30 minutes. The reactions were then heat-shocked at 42°C for 45 seconds in a water bath and immediately cooled on ice for 3 minutes. 900 μ l of warm LB liquid media (un-supplemented with antibiotics) was added to the transformed cells, incubated at 37°C (225 rpm) for 1 hour to promote cell recovery. During this stage cells recover their antibiotic resistance allowing subsequent analysis of transformation efficiency on selective LB agar plates. Typically, 200 μ l of transformed cells were spread and plates were incubated at 37°C O/N. However, some cell strains transformed more efficiently with certain selected DNA plasmids and thus a lesser volume was required for adequate growth, see Table 2.5 for volumes plated.

Reaction Type	Volume Plated (μ l)
Mutagenesis Control (pWhitescript)	250
Transformation Control (pUC18 or pET28b)	5:200 dilution
Standard Transformation	200
SDM Transformation	250 (2x)

Table 2.5 Transformation Plates

Examples of cell culture volumes typically plated on selective LB agar plates are listed for different types of transformation reactions. Reactions which usually yield high transformation efficiencies, for example transformation controls (pUC18 or pET28b) were diluted prior to spreading, whereas those which typically transform poorly were plated directly, e.g. SDM reactions.

A combination of different controls and selective plates were used to determine if the correct cell strain had been transformed with the correct plasmid. Essentially, only the cells that had acquired the relevant plasmids, containing an antibiotic resistance gene, could grow on these selective plates.

2.9.6 Plasmid DNA Maintenance and Purification

Following transformation of DH5 α cell strains with plasmid DNA, replication and purification of plasmids was conducted to generate more plasmid DNA for downstream applications.

For plasmid DNA replication an individual transformed *E. coli* DH5 α colony (containing the plasmid of interest) from a freshly streaked selective plate was used to inoculate 10 ml of selective LB liquid media. The bacterial culture was grown at 37°C, for 12-16 hours (250 rpm). Following growth, 5 ml of the O/N culture was harvested in a 15 ml falcon tube for 3 minutes using a Beckman Coulter Allegra 25R (rotor TS-5.1-500) benchtop centrifuge (8,000 rpm).

The plasmid DNA contained within the DH5 α host strain was then purified using a modified alkaline lysis method (Birnboim & Doly, 1979) as recommended by the QIAprep spin miniprep kit protocol (supplied by Qiagen). For reproducible and reliable results Qiagen recommended using the DH5 α host strain to give high-quality DNA for downstream applications such as DNA

sequencing, to confirm plasmid DNA sequence identity (Section 2.7.4), and a summary of the purification steps is provided in Figure 2.3.

In essence the previously harvested cell pellet, described above, was re-suspended in 250 μ l of buffer P1 (containing 100 μ g/ml RNase A and LyseBlue) and transferred to a microcentrifuge tube. 250 μ l of buffer P2 was added to the pellet suspension and mixed thoroughly by inverting the tube 4-6 times gently. Once a homogeneously blue coloured suspension was present (< 5 min incubation), as visualised by the presence of a LyseBlue colour indicator, 350 μ l buffer N3 was added and mixed, until the solution was colourless. The lysate was centrifuged for 10 minutes (13,000 rpm, Eppendorf Centrifuge 5810 R) in the Eppendorf tube and the supernatant was applied directly to a QIAprep spin column. Columns were centrifuged briefly for 30-60 s (13,000 rpm, Eppendorf Centrifuge 5810 R) and the flow-through was discarded. Then 0.5 ml buffer PB was applied to the column, centrifuged and the flow-through discarded was as before. This method was repeated, 0.75 ml buffer PE was added and an additional 60 s spin was conducted to remove all residual wash buffers. The plasmid DNA was eluted into a fresh 1.5 ml microcentrifuge tube by incubating the column for 60 s in 50 μ l buffer EB (10 mM Tris-HCl, pH 8.5) or water and centrifuging for 60 s. The resulting plasmid DNA concentration was determined from its $A_{260\text{ nm}}$ using a Thermo Scientific NanoDrop 2000c, prior to storage at -20°C.

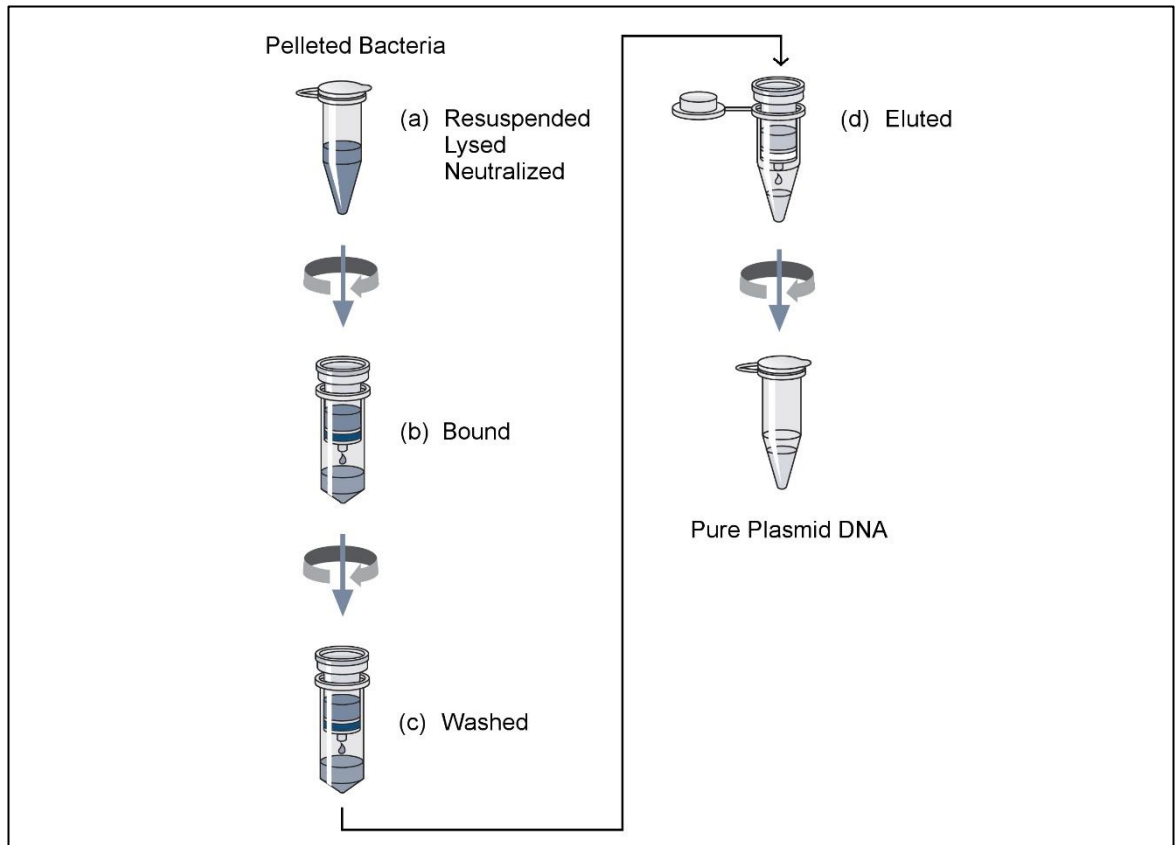


Figure 2.3 Plasmid DNA Purification

Summary figure of plasmid DNA purification using a QIAprep spin miniprep kit (Qiagen). (a) After harvesting, pelleted bacteria were re-suspended in buffer P1 which contains RNase A for degradation of contaminating RNA and a colour indicator (LyseBlue). In the presence of buffer P2, under alkaline conditions, LyseBlue turned blue when cells were successfully lysed. Sodium dodecyl sulphate (SDS) detergent in the lysis buffer P2 solubilized the phospholipid and protein components of the cell membrane thus leading to lysis and release of cellular contents, while NaOH denatured DNA and proteins. Neutralisation of alkaline conditions by buffer N3 was optimal for release of plasmid DNA without chromosomal DNA. The solution was adjusted to high-salt binding for efficient DNA adsorption to the QIAprep spin column's silica membrane, which aids further precipitation of unwanted denatured proteins, cellular debris and SDS. SDS precipitation was indicated by LyseBlue as a homogeneous colourless solution, suggesting successful neutralisation. Following lysate clarification by centrifugation, (b) DNA applied to the column was (c) washed with buffers PB and PE to remove endonucleases, ensuring plasmid DNA remained intact and salts were removed respectively. (d) Pure plasmid DNA was then eluted in either buffer EB or with water pH 7.0-8.5. Image adapted from Qiagen QIAprep spin miniprep kit protocol.

2.10 Gel Electrophoresis

Following preparation of nucleic acids (DNA and RNA) or proteins, analysis of the size, integrity and purity can be conducted using gel electrophoresis (reviewed in Tan & Yiap, 2009). A brief description of how different gel electrophoresis techniques were used to separate these macromolecules is initially provided, prior to detailed methods.

In principle, macromolecules were separated based on their molecular properties including their charge, size and/or shape. Naturally charged molecules, including negatively charged nucleic acids or samples with a chemically applied net negative charge, such as proteins coated in SDS detergent, were applied to either a native agarose or polyacrylamide gel matrix. Upon application of a negatively charged electrical current, these negative molecules were repulsed and migrated within the gel matrix towards the positive terminal. In addition to this charge based separation

DNA, RNA and proteins were also separated by a sieving action based on their size and shape. The smaller/linear molecules or fragments moved through the matrix, consisting of either agarose or polyacrylamide gel, faster than larger molecules. In instances where migration of nucleic acids or protein samples entirely based on their size was required, denaturing chemical agents such as urea or SDS respectively were added and sample structure formation was prevented by heat treatment.

Both these native and denaturing gel electrophoresis techniques were utilised to separate samples and estimate the macromolecule's molecular weight (MW) by comparing the band migration to a ladder containing bands of known MW's. This was useful to determine if the protein or nucleic acid's migration and thus MW, correlated with previous experimental data and/or predicted *in silico* MW values. The purity of samples was also analysed by detecting the presence (or absence) of other contaminants upon gel electrophoresis.

For each type of gel electrophoresis utilised within this research detailed methods are outlined in the following sub-sections.

2.10.1 Agarose Gel Electrophoresis

Agarose gel electrophoresis was conducted using a horizontal gel tank setup and the reagents listed in Table 2.6. Agarose powder was dissolved in 1 x Tris-Acetate-EDTA (TAE) (40 mM Tris Acetate & 2 mM Na₂EDTA, pH 8) (National Diagnostics) by heating in a microwave. The solution was allowed to cool slightly before ethidium bromide (0.5 µg/ml) was added, the gel was then cast in a 10 x 20 cm gel tank with an appropriate sized comb inserted and allowed to polymerise for at least 30 minutes prior to use. Once prepared in 6 x loading dye (3.3 mM Tris-HCl pH 8.0, 2.5% Ficoll 400, 11 mM EDTA, 0.017% SDS & 0.015% Bromophenol Blue) to a final 1 x concentration, samples were loaded alongside 5 µl of an appropriate molecular weight ladder. This was done in order to calibrate the gels; either the low molecular weight DNA ladder (NEB N3233S) or 100 base pair ladder (NEB N3231S) were used (see Appendix 9.2 for sizes). Agarose gels were run in 1 x TAE (supplemented with 0.5 µg/ml ethidium bromide) for 45 minutes at 120 V and the DNA visualised under UV transluminescence using a Syngene G-Box imager and Gene snap software (Syngene, n.d.)).

	Reagent	Amount
Agarose (1.2 %)	Agarose	0.96 g
	1 x TAE	80 ml
	Ethidium Bromide	10 µg/ml

Table 2.6 Agarose Gel Preparation

The mass of each reagent used to prepare 80 ml of 1.2% Agarose gel is provided.

2.10.2 Polyacrylamide Gel Electrophoresis (PAGE)

All polyacrylamide gel electrophoresis (PAGE) methods, using a vertical gel tank setup, were prepared as described in Sections (2.10.2.1-2.10.2.3) using XCell SureLock™ Mini-Cell gel tanks and Novex 1.0 mm plastic gel cassettes (Life Technologies). In addition to specific gel reagents, a 10% (w/v) stock of ammonium persulfate (APS), N, N, N', N'-Tetramethylethylenediamine (TEMED) and dH₂O were used for preparing the gel matrix. Once all gel components were mixed, they were poured into a cassette with an appropriate sized comb inserted and allowed to polymerise for at least 30 minutes prior to use. Gel reagents, running conditions and molecular visualisation methods varied for each type of gel electrophoresis and sample, as highlighted in Sections 2.10.2.1-2.10.2.3.

2.10.2.1 Native-PAGE

Native PAGE gels were prepared using AccuGel™ (29% (w/v) Acrylamide and 1% (w/v) Bis-acrylamide, National Diagnostics EC-852) as outlined in Table 2.7 and samples for loading were prepared in an appropriate loading buffer (4x loading dye: 0.25 M Tris HCl pH 6.8, 40% Glycerol and 0.02% bromophenol blue). Gels were run in 500 ml of 1 x Tris-Borate-EDTA (TBE) running buffer (89 mM Tris Base, 89 mM Boric Acid & 2 mM EDTA pH 8) at 100 V for 1.5 hrs. To detect unlabelled RNA, 3 µl SYBR Gold™ (10,000x stock, Invitrogen) was diluted in 30 ml running buffer and gels were stained for 5 minutes whilst shaking at room temperature. As recommended in the user manual, following staining, gels were visualised near the excitation/emission maxima of SYBR gold (~495/537 nm) using the blue laser (473 nm) settings of the Fujifilm FLA-5000 phosphorimager.

	Reagent	Volume
AccuGel™ (6 % Native)	AccuGel™	1.5 ml
	10 x TBE	1 ml
	dH ₂ O	7.5 ml
	TEMED	8 µl
	10 % (w/v) APS	100 µl

Table 2.7 Native-PAGE Preparation

The volume of each reagent used to prepare one 6% native polyacrylamide gel is provided.

2.10.2.2 Urea-PAGE

Denaturing urea PAGE gels were prepared using UreaGel™ concentrate (19% (w/v) Acrylamide, 1% (w/v) Bis-acrylamide, 7.5 M Urea), UreaGel™ diluent (7.5 M Urea) and UreaGel™ buffer (0.89 M Tris-Borate pH 8.3, 20 mM EDTA, 7.5 M Urea) (National Diagnostics EC-830) as outlined in Table 2.8. Gels were pre-run in 500 ml of 1 x TBE running buffer 200 V for 30 minutes. Appropriate sample volumes were added to Gel Loading Buffer II (95% Formamide, 0.025% SDS, 18 mM EDTA, Xylene Cyanol and Bromophenol Blue), heated for 10 minutes at 95°C, and centrifuged briefly prior to loading. In some instances, a dye-free loading buffer was

prepared and used when the dye front migrated at the same location as the RNA band of interest. Gels were pre-run at 200 V for 30 minutes, denatured samples were then loaded and run at 200 V for 1.5 hrs. Gels were dismantled and visualised using a Fujifilm FLA-5000 phosphorimager. If RNA samples were labelled the phosphorimager settings were adjusted to the specific excitation/emission maxima of the fluorescent label, unlabelled RNAs were visualised by staining for 5 minutes with SYBR Gold™ as described previously.

	Reagent	Volume
UreaGel™ (20 % Denaturing)	UreaGel™ Conc.	8 ml
	UreaGel™ Diluent	1 ml
	UreaGel™ Buffer	1 ml
	TEMED	4 µl
	10 % (w/v) APS	80 µl

Table 2.8 Denaturing Urea-PAGE Preparation

The volume of each reagent used to prepare one 20% urea denaturing polyacrylamide gel is provided.

2.10.2.3 SDS-PAGE

SDS-PAGE gels were prepared using ProtoGel™ 37.5:1 (30% (w/v) Acrylamide, 0.8% (w/v) Bis-acrylamide), ProtoGel™ Stacking Buffer (0.4% SDS, 0.5 M Tris-HCl pH 8.8) and ProtoGel™ Resolving Buffer (0.4% SDS, 1.5 M Tris-HCl pH 8.8) (National Diagnostics EC-890) as outlined in Table 2.9. The resolving gel solution was prepared first, poured into the cassette and overlaid with 400 µl water-saturated butanol to ensure a flat horizontal gel surface during gel polymerisation. Gels were left to polymerise for 30 minutes, the butanol was washed off with dH₂O prior to adding the stacking gel solution. Gels were again left to polymerise with an appropriate sized comb inserted for 30 minutes prior to use.

Samples were prepared for loading by mixing an appropriate sample volume with 2 x reducing SDS-PAGE loading dye (0.125 M Tris-HCl pH 6.8, 4% SDS, 0.002% bromophenol blue, 10% β-mercaptoethanol & 20% Glycerol) (Laemmli, 1970), incubated for 10 minutes at 95 °C for denaturation, centrifuged briefly and loaded onto a 12% SDS-PAGE gel. When necessary, protein samples which were dilute and required concentrating prior to SDS-PAGE analysis, were concentrated using StrataClean resin (Agilent). Typically, 250 µl of sample was mixed with 10 µl of StrataClean resin, vortexed and centrifuged at 3,000 rpm for 3 minutes in an Eppendorf Centrifuge 5810 R. The supernatant was removed and the resin pellet, containing bound protein, was retained. 15 µl of 2 x reducing SDS-PAGE loading dye was added, samples were heat denatured and loaded as described previously.

Protein samples were loaded alongside 5 µl of a relevant protein ladder in order to calibrate the gels; either a BenchMark™ and/or Pre-Stained SeeBlue® (Standard or Plus2) protein molecular weight markers (Life Technologies) were used (see Appendix 9.2 for band sizes). The 12% SDS-

PAGE gels were run in 500 ml of 1 x Tris-Glycine-SDS (25 mM Tris, 19.2 mM Glycine and 0.1% SDS) (National Diagnostics) running buffer at 80 V for 20 minutes into the stacking gel, and then 150 V for 1 hour 20 minutes into the resolving gel. Gels were then dismantled, stained in Coomassie (G-250) with SimplyBlue™ SafeStain (Invitrogen LC6065) and visualised using a Syngene G-box UV transilluminator as standard. Gels visualised using SafeStain were rinsed with 100 ml dH₂O, heated for 1 minute in a microwave until almost boiling and placed on a shaker for 1 minute with gentle mixing. This washing method was repeated 3 times to remove SDS and buffer salts as this interferes with the dye binding to the proteins. 20 ml of SafeStain was added, heated again as before and shaken for 5 minutes. Once stained, gels were washed to reduce background staining in 100 ml of dH₂O as before and visualised using a G-Box.

	Reagent	Volume
Resolving solution (12 %)	ProtoGel™	4 ml
	Resolving buffer	2.6 ml
	dH ₂ O	3.3 ml
	TEMED	10 µl
	10 % (w/v) APS	100 µl
Stacking solution (4 %)	ProtoGel™	650 µl
	Stacking buffer	1.25 ml
	dH ₂ O	2.5 ml
	TEMED	5 µl
	10 % (w/v) APS	25 µl

Table 2.9 SDS-PAGE Preparation

The volume of each reagent used to prepare one 12% SDS-polyacrylamide gel is provided; the 12% resolving gel was prepared first followed by the 4% stacking gel, which was overlaid on top.

2.11 Western Blot Analysis

Samples for analysis were initially separated by size using SDS-PAGE, as described in Section 2.10.2.3, prior to western blot analysis for protein identification. This technique uses labelled antibodies, specific to a protein of interest, for protein identification (reviewed in Yang & Mahmood, 2012); the methods of which are provided below.

Proteins separated by SDS-PAGE were transferred from the gel onto Amersham Hybond ECL Nitrocellulose membrane (GE Healthcare) by applying a 300 mA current for 1 hour 20 minutes in gel tank filled with 1 x western transfer buffer (20 mM Tris, 150 mM NaCl and 20% methanol). Typically, SeeBlue® Pre-Stained protein standards were loaded onto gels as a quick visual indicator of western transfer efficiency. The membrane was then blocked with 50 ml of 1 x TBST, 2% Marvel solution (10 mM Tris HCl pH 8, 150 mM NaCl, 0.1% Tween-20 & 2% Marvel) for 3 hours, at 4°C, on a shaker. 20 ml of the appropriate 1° antibody was then incubated with the membrane (1:2000 dilution in 1 x TBST, 2% Marvel solution) O/N at 4°C. Following probing with

the 1° antibody, membranes were washed with 20 ml fresh 1 x TBST, 2% Marvel solution three times. Depending on the type of 1° antibody used, membranes were either probed with a 2° antibody, which was a HRP conjugate or developed directly, if the 1° antibody already contained HRP conjugated to it; this HRP conjugate facilitated chemiluminescence detection. If a 1° antibody-HRP wasn't used, the washed membrane was incubated with 20 ml of the 2° antibody (1:2000 dilution in 1 x TBST, 2% Marvel solution) for 2 hours at 4°C.

Once membranes were probed with relevant 1° and/or 2° antibodies they were washed (as described before) with an additional 20 ml wash in 1 x TBST solution (10 mM Tris HCl pH 8, 150 mM NaCl, 0.1% Tween-20) and the membrane was stored in 20 ml 1 x phosphate buffered saline (PBS) until visualised. Visualisation solution 1 (100 µl 250 mM Luminol (prepared in DMSO), 44 µl 90 mM P-Coumaric acid, 1 ml Tris-HCl pH 8.0 and 8.9 ml dH₂O) and solution 2 (6 µl H₂O₂, 1 ml Tris-HCl pH 8.0 and 9 ml H₂O) were prepared separately prior to mixing together. Membranes stored in PBS were rinsed and the final visualisation solution mix was added, incubated for 1 min at RT and imaged incrementally at high resolution using an ImageQuant LAS 4010 western camera (GE Healthcare) with a minimum exposure of 10 s to detect chemiluminescence. Typically, a BenchMark protein ladder was loaded onto SDS-polyacrylamide gels which were probed for the presence of His-tagged proteins. Since the ladder is conveniently comprised of recombinant H₆-proteins, it provides not only a size marker but also a visual indicator of western blot efficiency.

2.12 Conclusion

This Methods Chapter describes the general materials and experimental techniques utilised as standard within all chapters of this thesis, whereas subsequent chapters are more specific. Accordingly, the following results in Chapter 3 provide all the information regarding the preparation of recombinant proteins; including a brief introduction, the specific methods utilised and the results obtained.

3 Recombinant Protein Preparation

The previous Materials and Methods Chapter provides a repository of common reagents and protocols, which were utilised throughout the rest of this thesis. Similarly, this chapter aims to provide preparation details for all PNPase and Exosome proteins used, including *E. coli* and *H. sapiens* PNPase and *S. solfataricus* exosome (hereafter referred to as EcPNPase, hPNPase and SsoExosome respectively).

This chapter will review the applications, methods and limitations of protein-based research. The techniques available to produce recombinant proteins in general, alongside the published strategies previously employed for relevant PNPase and Exosome proteins will also be addressed. Protocols are available for all the proteins listed above and have been applied directly within this research where possible. In situations where adaptations to a protocol have been made, an emphasis on the reasoning behind these changes has been provided. Details of the methods for cloning, expression and purification, followed by protein identification and biological activity analysis, are provided in Section 3.2. The outcomes of which are then provided in the following Section 3.3. Important conclusions made from this chapter are established in the final Section 3.4, and a summary is provided outlining the successful preparation of PNPase and Exosome proteins for downstream applications.

3.1 Introduction

3.1.1 Applications, Methods and Limitations of Protein Research

From basic life forms including bacteria to more complex beings such as humans, proteins have a vital role in all living organisms. Not only are they the basic structural molecule of cells but they also regulate a plethora of cellular functions. For this reason, they have been the subject of extensive research; understanding the biological activity of an individual protein can potentially unveil details of a more complex biological system. Applying this knowledge to manipulate individual proteins or biological systems, can result in adaptations which may be beneficial to a wide range of areas. This is evident by the extensive use of proteins, typically enzymes, in the food, chemical and pharmaceutical industries. Due to the broad use of proteins in both academia and industry, protein based research, in either setting, can result in beneficial intellectual and economic impact, particularly in cases where large industrial processes are developed for optimal output, as a result of a new discovery.

Regardless of the subject matter, research which aims to elucidate the detailed properties of an individual protein can be investigated using the following steps. Firstly, the protein of interest can be synthesised and then, if necessary, isolated from any contaminant so that only the pure protein remains. Secondly, the activity of this pure protein with a known reacting molecule (substrate) can be confirmed. This is important so that an understanding of normal functioning is attained, prior to

downstream applications, for instance protein interaction research and/or manipulation. Pure and active protein is utilised so that researchers can clearly establish if a particular biological activity, that may be observed upon testing, resides within the protein specifically and not a contaminant. Thus ensures that any information acquired is directly applicable to the purified target protein.

Although cells can naturally synthesise proteins, the high yields required for extensive use in industry has promoted advances in developing efficient and cost-effective protein production strategies (Adrio & Demain, 2014). These strategies apply the same methods of synthesis utilised in nature; most simply, proteins are expressed from the genetic information held within the deoxyribonucleic acid (DNA) of the cell. Biochemical techniques have allowed sophisticated modifications at each stage of this process and these amendments maximise protein production. Such improvements on emerging DNA technologies and expression systems have facilitated not only the production of existing proteins but also the discovery of new natural and synthetic proteins which can often possess more desirable enzymatic properties than those currently available (Adrio & Demain, 2014).

Although significant steps have been made to increase production yields, protein preparation can be difficult on both a large and small-scale in industrial manufacturing and academic research. The ease of preparation is completely dependent on not only the methodologies utilised but the physical properties and functions of the protein. For this reason, if existing methods are available for the protein of interest, or homologs they are usually consulted initially and modifications applied to the protocol where necessary. With this approach the following Section 3.1.2 provides examples of strategies commonly used to produce recombinant proteins in high yields. Attention is then drawn to methods previously described in the literature to prepare PNPase and archaeal exosome proteins relevant to this research.

3.1.2 Recombinant Protein Technologies

3.1.2.1 Existing Strategies: Bacterial Expression Systems

While a gene, encoding a protein of interest, may naturally be present and expressed within a cell, additional DNA encoding the protein is often added in the form of a plasmid to maximize protein synthesis. This introduced plasmid DNA can either encode a host protein or a heterologous protein which is foreign to the expression system. Either way the overexpression of this target gene swamps the cell with mRNA so that the cellular translational machinery preferentially synthesises the protein of interest. The plasmid DNA which is introduced to an expression system can be recombinantly engineered by ligating the gene of interest within a suitable plasmid vector. The gene fragment itself can be obtained by direct amplification from the genomic DNA of the species, from complementary cDNA libraries or by total gene synthesis (designed *in silico*). Irrespective of the amplification method, genes are typically designed with flanking restriction sites and ligated

into a vector using complementary restriction enzymes. Normally, in bacterial expression systems this vector also has an antibiotic resistance gene incorporated for selection purposes.

A wide range of plasmid DNA vectors which exhibit controlled gene expression by using an inducible promoter are available, these have specific properties aimed at enhancing target gene overexpression in various host systems. For example, in bacteria, a commonly used pET vector system is available (Broadway, 2012). The pET vectors can be used in conjunction with an *E. coli* λ DE3 lysogen expression cell strain. This particular strain contains a phage construct encoding T7 RNA Polymerase (T7RNAP), which is under the control of a lacUV5 promoter (Studier & Moffatt, 1986). In normal cellular conditions transcription is silent, until a lactose analogue isopropyl β -D-1-thiogalactopyranoside (IPTG) is added and this induces T7RNAP expression and thus the target gene expression. This is particularly useful for expression of proteins which may be toxic to the host cell, allowing maximal growth before expression of the potentially harmful protein (Studier & Moffatt, 1986). In this study specifically, this was useful as production of mammalian PNPase has been suggested to inhibit growth (Leszczyniecka *et al.*, 2002), hence its overexpression may reduce protein yields if the host cells growth is impacted. Additionally, the T7RNAP is very promoter-specific; only targeting the pET vector T7 promoter which is located upstream of the target gene, so that expression can occur directly, and with a low error rate (Mierendorf, Morris, Hammer, & Novy, 1998).

An over-expressed bacterial protein can be extracted from bacterial hosts, however often this is simply not feasible and in order to maximise protein synthesis, an expression system different to that of the protein host may be selected.

3.1.2.2 Existing Strategies: Other Expression Systems

A range of expression systems have been established which exploit prokaryotic and eukaryotic cellular machinery to synthesise heterologous proteins for study. The optimal expression system required depends on the physical attributes of the target protein. Typically, there are four main expression systems: bacterial, yeast, insect and mammalian cell culture. They each possess their own unique characteristics which can provide both benefits and disadvantages to using the system. Bacterial systems allow for rapid cell growth, with low cost and high expression levels; however, provide limited post-translational modifications to the protein. Hence, more complex proteins requiring glycosylation, phosphorylation, acetylation or other modifications must be synthesised in one of the other three systems (yeast, insect and mammalian). Of these systems, yeast is the simplest, with rapid cell growth and minimal growth media requirements, like the bacterial system; providing a lower cost option than mammalian systems. However, with yeast systems if the protein contains disulphide bonds, correct protein refolding may be required, unlike insect and mammalian systems, which result in proper folding of proteins. The added complexity of modification provided by insect and mammalian systems however results in higher financial costs and lower growth rates,

which are less advantageous. Consequently, the exact system selected needs careful consideration to maintain protein preparation efficiency whilst minimising costs.

3.1.2.3 Strategies for Preparing PNPase and Archaeal Exosome

Previous research has reported successful cloning of EcPNPase, hPNPase and SsoExosome into the pET vector system and expression of the recombinant proteins for purification in *E. coli* (Lin *et al.*, 2012; Lorentzen *et al.*, 2005; Lorentzen, Dziembowski, Lindner, Seraphin, & Conti, 2007; Lorentzen & Conti, 2005; Lu, Ding, & Ke, 2010; Nurmohamed, Vaidialingam, Callaghan, & Luisi, 2009). In this study, EcPNPase plasmid DNA was kindly provided by Prof. Ben Luisi (University of Cambridge) and for rapid and cost effective cloning, all other proteins were ligated into relevant pET vector following gene synthesis. An additional codon-optimisation step was conducted for the eukaryotic hPNPase plasmid DNA in order to prevent codon bias in the bacterial host expression strain. This is important as bacterial host systems, lacking eukaryotic tRNAs may lead to enhanced mRNA degradation, amino acid incorporation errors and frame shifts; affecting the quantity or quality of protein synthesised (Kane, 1995). In situations where codon-optimisation is required, alternative methods, which use cell lines containing gene mutations to compensate for tRNA and codon preference, can also be selected. For simplicity gene synthesis was conducted and codon optimisation was done *in silico*.

Once the cloning had been conducted, recombinant plasmid DNA must be introduced into an appropriate system for protein expression. As PNPase and Exosome proteins require no post-translational modifications they can be expressed in a simple bacterial expression system following DNA transformation using either electroporation or chemical transformation. As the former method requires specialist equipment and the latter is relatively simple, the later was employed. When considering the bacterial expression systems, multiple strains of bacteria are available, each with unique properties optimal for protein expression. More specifically the *E. coli* BL21(DE3)pLysS ((F- ompT gal dcm lon hsdSB(rB- mB-) λ (DE3) pLysS(cmR)) cell strain was selected in this study. This BL21(DE3)pLysS cell strain contains both the DE3 lysogen (which as explained previously is a phage construct that expresses T7RNAP under the control of a lacUV5 promoter) and a pLysS plasmid. This additional pLysS plasmid encodes T7 lysozyme which is an inhibitor of T7RNAP and can reduce 'leaky' background protein expression from the BL21(DE3)pLysS strain. Further details of the genotypes for the strain utilised here are provided in Appendix 9.4.

Following cloning, recombinant plasmid DNA can also be routinely stored long-term in bacterial cell strains such as the commonly used *E. coli* DH5 α cell strain (F- Φ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK-, mK+) phoA supE44 λ - thi-1 gyrA96 relA1) see Appendix for 9.4 details. Utilisation of this DH5 α strain increases gene insert stability by the mutation of non-specific endonucleases including endA1 and protects un-methylated DNA from the EcoKI

restriction-modification system. This reduces degradation and consequently improves plasmid DNA quality upon preparation and purification.

Once an appropriate vector and expression system, which highly expresses heterologous protein, has been chosen the process is expanded to a larger scale to produce enough protein for further purification. In some systems, including some types of cell culture, protein expression is automatic. However, in the bacterial BL21(DE3) strain utilised within this study, induction is typically initiated with IPTG at low temperatures (15–25°C), in rich medium and with good aeration. Published protein expression conditions for all PNPase and Exosome proteins were used as an initial starting point. Commonly used modifications to improve expression, including lowering the IPTG concentration for induction and/or lowering the temperature were applied where necessary.

In any expression system, naturally-produced host cellular macromolecules are present as a by-product of protein production alongside the recombinant protein of interest. These must be removed to leave a pure protein for study; as mentioned previously, in a purified form any characteristics observed *in vitro*, whether they are structural and/or functional, can be attributed to the purified protein specifically. Recombinant proteins can be isolated and purified away from other host proteins using various experimental techniques. Firstly, once a recombinant protein has been successfully expressed, cell lysis must be employed. This can be conducted manually by cell homogenisation in a well-buffered solution optimised for cell lysis. Ideally the lysis buffer contains ionic strengths equivalent to 300-500 mM of a monovalent salt, such as NaCl (Gräslund *et al.*, 2008). Alternatively, cells may be disrupted in an appropriate lysis buffer by more forceful methods, including sonication and French press. The precise methods are often decided based on available equipment and previous methods utilised in the literature. In contrast to previous methods which utilise a French press, cell homogenisation and sonication was selected for this work based on the equipment available. Irrespective of the lysis method chosen, it is essential to ensure that the majority of cells are broken in order to release contained recombinant protein for high-yield protein purification. Hence, using SDS-PAGE analysis the effectiveness of the lysis process can be determined from the amount of target protein following clarification using centrifugation. Applying a strong centrifugal force to total cell lysate effectively separates out the insoluble cellular debris/macromolecules from those that are soluble. If the protein is folded correctly and present in the soluble supernatant fraction, it can next be isolated for further purification. Alternatively, if the protein is denatured or present in inclusion bodies, it may be present in the insoluble cell pellet and may require refolding. At various stages of the purification process, the protein's solubility may be taken advantage of to separate the protein of interest based on its individual characteristics, such as the protein's stability at varying pH, temperature or salinity. Proteins which are known to bind to DNA or RNA may also be purified from these contaminants by using enzymes designed to specifically degrade such entities. An RNase and/or DNase stage is often employed to improve the protein purity.

Partially purified proteins may be subject to various chromatography stages in order to purify them further. Chromatography is a technique used to separate sample mixtures. It typically works by distributing sample components between two phases whilst moving through a column filled with a particular medium. Molecules within the sample, which can interact with the column media, are retained within the column and are referred to as being in the stationary phase. In contrast, non-interacting molecules are not retained and are considered to be in the mobile phase. Depending on the type of column media selected and the specific properties of a protein, it may be present in either phase. This allows purification of different proteins based on their individual properties. Analysis of proteins after chromatography is essential to determine the volume at which a protein of interest is eluted. Typically, chromatography is conducted on automated platforms such as the AKTA systems. This has many benefits including the measurement of key-variables such as temperature, conductivity and buffer component concentrations (if these are being varied). Even the absorbance at 280 nm ($A_{280\text{ nm}}$) can be recorded in order to detect proteins. Using the AKTA systems, components of a mixture, separated by chromatography, can be collected as individual fractions for subsequent analysis and possibly further purification.

In terms of protein purification, the ability to separate proteins, determine the exact volume at which they elute by measuring $A_{280\text{ nm}}$ and retain the relevant fractions for further use is extremely useful. Consequently, chromatography was a vital step used in this work to purify recombinant proteins expressed in *E. coli* cells away from contaminating host proteins. In this instance a variety of chromatography methods were employed to purify each protein and the method chosen depended on their specific molecular properties including binding affinity, net charge and molecular weight (MW).

Recombinant proteins with an affinity tag can be purified away from host proteins by Affinity Chromatography (AC); tagged proteins of interest can bind directly to the column media whilst others are eluted. A range of tags are available, including the one used specifically within this study, which endows proteins with a recombinant poly (6x)-histidine tag (H_6) that can be purified by Immobilised Metal Affinity Chromatography (IMAC). Essentially a column containing immobilised nickel resin binds to histidine; including the H_6 -tag of the protein and non-binding proteins are eluted. In order to elute histidine-rich proteins a buffer containing imidazole can be applied to out-compete the binding interaction. This application can be completed using a gradient, so that weakly interacting proteins are eluted prior to strongly interacting proteins, providing further separation.

Consistent with previous research, hPNPase and SsoExosome constructs encoded proteins with an N-terminal H_6 -tag and EcPNPase was untagged. The affinity tag, present on some of the proteins, may be removed to minimize non-native sequences in the recombinant protein and to achieve

further purification. This is achieved by using a recombinant, H₆-tagged protease and reapplying the sample to IMAC column to remove both the protease and any cellular proteins that bind the metal affinity resin. Although some literature advises that tags should be removed from proteins prior to biochemical testing, a review suggested for a majority of proteins surveyed a H₆-tag does not have a consistent impact on the N-terminal structure of the target protein (Gräslund *et al.*, 2008). For consistency with previously published research, the H₆-tag of hPNPase and SsoExosome was not removed for this study.

In this study, recombinant proteins including EcPNPase without an affinity tag were purified using a different means; Ion Exchange Chromatography (IEC). This method works to separate molecules based on their net charge. As mentioned before, the individual characteristics of a protein can be exploited to present it in either the stationary or mobile phase. If a protein has a net negative charge, it can be immobilised by a positively charged resin within the column, this is termed Anion Exchange Chromatography (AEC). Contrary to this, if a protein's net charge is positive, a negative resin is used in Cation Exchange Chromatography (CEC). The result of either method ensures binding and thus separation of proteins of interest, which can be later eluted using a specific elution buffer. Conveniently, due to the nature of proteins, changes in the pH of elution buffers can result in modifications to the net charge, thereby eluting bound proteins. In CEC, increasing the pH causes the protein to be less protonated and hence, less positively charged. Therefore, it cannot form ionic interactions with the negatively charged resin, which ultimately results in the protein eluting from the column. AEC is the inverse of this, decreasing the pH causes more protein protonation and hence, the more positively charged protein dissociates from the positively charged resin. Additionally, the presence of positive Na⁺ ions and negative Cl⁻ ions, as a result of increasing NaCl salt concentrations in the elution buffer, can compete with the protein for binding to the charged resin, resulting in protein elution.

The previously mentioned chromatography techniques identify situations where proteins bind to a column resin and then have to be eluted in a specific elution buffer. However, another method worthy of mention is Size Exclusion Chromatography (SEC). This technique varies compared to the other chromatography methods, in that proteins do not necessarily bind to the column, but can actually be 'trapped' by moving in and out of the resin pores and consequently switching between the mobile and stationary phase dependant on their overall size. Smaller molecules are retained more within the small pores of the column medium than larger molecules. Molecules with large MW can move entirely within the mobile phase (not retained by the column at all) and elute from the column in what is termed the void volume, whilst smaller proteins are retained by the media to varying degrees and elute later, thus allowing separation by overall size. SEC is a valuable process and in the context of this work, consistent with previous research, is the final protein purification chromatography stage. Normally protein samples are concentrated to small volumes for SEC. This is due to the fact that samples separate by size immediately upon column-loading, as such, larger

volumes result in broad $A_{280\text{ nm}}$ peaks. Hence small volumes are used in order to improve the resolution of sample separation. There are a variety of SEC column media types available. Superdex 200 prep grade resin consists of a matrix of dextran covalently attached to highly cross-linked agarose, aimed at purifying proteins 10-600 kDa in size, hence it was used to purify all recombinant proteins in this study.

A relatively simple column calibration can be conducted to determine the molecular weight (MW), usually measured in kDa, of any unknown proteins separated on a SEC column. This makes SEC not only a helpful separation tool but also an analytical one. A mixture of protein standards with known MWs may be applied to the SEC column. If the mixture includes proteins with similar MW, which may elute with peaks $A_{280\text{ nm}}$ close to one another, the mixture to be loaded onto SEC can be split. This would reduce the chance of overlapping $A_{280\text{ nm}}$ peaks, so that the exact elution volume for individual standards can be measured. One typically used standard, Blue Dextran (2,000 kDa), a large MW macromolecule is not retained within the column, and can be used to determine the specific void volume of a column. The retention volume of smaller protein standards can then be calculated as the elution volume minus the void volume. This is a very useful value; if retention volumes (ml) are plotted against \log MW (kDa) of each protein standard, a calibration curve may be produced. With such a calibration curve, the retention volume of unknown proteins in subsequent runs can be used to predict their respective MW (kDa).

Production of high-quality, recombinant proteins for further structural and functional biochemical studies is essential; however, confirmation of their identity is also vital. Determining the identification of any recombinant protein can be conducted using a variety of techniques. Initially SDS-PAGE analysis can be conducted to separate protein samples taken throughout the various stages of purification. This can be useful to analyse the purity of the protein of interest and predict its MW using a protein ladder. If the protein's size has been previously reported, this known MW may be used to predict which band is the target protein. Alternatively, programs such as ProtParam; which can predict protein MWs based on amino acid composition, alongside the use of protein molecular weight markers, can provide the first indication if a band migrates at a size similar to the protein of interest (Wilkins *et al.*, 1999). Additionally, in more analytical methods such as SEC, the elution volume may be utilised to predict a MW, or inversely the MW of an unknown protein may be used to predict protein elution when using a SEC calibration curve, like the one described previously. Following SDS-PAGE, western blot analysis for protein identification can also be conducted, this utilises antibodies specific to an affinity tag or directly to the protein of interest, to identify the presence of the target protein. Finally, purified protein can be analysed by either denaturing or non-denaturing mass spectrometry (MS) to identify the protein of interest.

3.1.3 Summary and Aims

As demonstrated in Section 3.1.2, a range of different methods may be utilised to clone, express and purify recombinant proteins from contaminating host proteins and the precise method chosen depends on individual characteristics of the protein of interest. Typically, if a method is previously well established in the literature it is simplest to reproduce that and amend it where necessary. Hence, *E. coli* and *H. sapiens* PNPase and *S. solfataricus* Exosome proteins were prepared with this in mind. It is also essential that the final recombinant protein's purity, identity and activity are assessed prior to further structural or functional characterisation. Henceforth, this chapter describes the overall preparation of all recombinant proteins tested.

3.2 Methods

This methods section describes the techniques used to clone, express and purify each of the recombinant proteins EcPNPase, H₆-hPNPase and H₆-SsoExosome (H₆-Rrp4_41_42). It also details the subsequent methods used to confirm the identity and activity of each protein. A majority of the steps undertaken to purify each recombinant protein are similar, hence methods described herein remain non-specific initially, with protein-specific methods indicated later as necessary.

3.2.1 DNA Preparation

Synthesis and ligation of gene constructs into appropriate DNA vectors were conducted as follows. For more detailed sequence information refer to Appendix 9.6.

3.2.1.1 *H. sapiens* PNPase

The complete coding sequence for the PNPase open reading frame (ORF) (NCBI accession number BC053660.1) from *H. sapiens* was codon-optimised *in silico* using proprietary GeneOptimizer® software (GeneArt AG, Life Technologies). The subsequent gene sequence was synthesised and ligated between restriction sites NheI and SalI within the multiple cloning site (MCS) of a pET-28b vector (Novagen). Thus producing an in-frame, thrombin-cleavable, H₆-tag fusion with full-length hPNPase (H₆-hPNPase). The insertion of the H₆-hPNPase construct into the pET-28b vector was confirmed by DNA sequencing (Life Technologies).

3.2.1.2 *S. solfataricus* Exosome

Similarly, to H₆-hPNPase, the complete coding sequence for the *S. solfataricus* archaeal exosome (SsoExosome) was codon-optimised *in silico* using GeneArt technology. Contrastingly the heterotrimeric SsoExosome, which is assembled from three individual subunits; Rrp4, Rrp41 and Rrp42 was cloned as a polycistronic gene. This ensured all three separate gene ORFs were in the same order as in the *S. solfataricus* exosome superoperon (NCBI gene ID Rrp4:1454999, Rrp41:1454998 and Rrp42:1454997). This approach was taken since previous experimental data indicated that cloning a polycistronic Rrp4_41_Rrp42 gene construct led to subsequent co-expression and improved complex solubility (Lu *et al.*, 2010). Following synthesis, the SsoExosome gene was

ligated between restriction sites NdeI and XbaI of the pETMCN-EAVNH vector kindly provided by C. Romier (Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC) - Université de Strasbourg). This produced a thrombin-cleavable, H₆-Rrp4 subunit with which untagged Rrp41 and Rrp42 can assemble, forming the SsoExosome complex (Lu *et al.*, 2010). As before, the H₆-Rrp4_41_42 (SsoExosome) construct insertion into the pETMCN-EAVNH vector was confirmed by DNA sequencing (Life Technologies).

3.2.1.3 *E. coli* PNPase

Cloning methods were not required for (untagged) *E. coli* PNPase as the gene had been previously cloned into a pETDuet-1 vector and was kindly supplied by Prof. Ben Luisi (University of Cambridge). Nevertheless, to confirm the EcPNPase_ pETDuet-1 plasmid sequence, DNA sequencing was conducted as previously described in Section 2.7.4.

3.2.2 DNA Plasmid Transformation

Following DNA cloning (Section 3.2.1), transformation of plasmid DNA, containing the gene construct of interest, into a desired bacterial cell strain was conducted as follows. For more detailed methods and bacterial cell strain genotypes see Section 2.9.5 and Appendix 9.4 respectively.

EcPNPase_ pETDuet-1, H₆-hPNPase_ pET-28b and H₆-Rrp4_41_42_ pETMCN-EAVNH recombinant plasmid DNA were all individually transformed into chemically competent *E. coli* DH5 α cells for long-term plasmid storage (Section 2.9.5). Transformed DH5 α cells were also utilised to replicate and purify plasmid DNA in sufficient quantities for DNA sequencing (~100 ng required), as it was essential to ensure no base mutations had arisen during synthesis and storage. The methods of bacterial growth, lysis and plasmid DNA purification are provided in Section 2.9.6. Using the same method, recombinant DNA plasmids listed above were also transformed into chemically competent *E. coli* BL21(DE3) pLysS cell strains for downstream protein expression and uptake of antibiotic resistant plasmid DNA was selected for, using LB agar plates supplemented with appropriate antibiotics (Section 2.9.5).

3.2.3 Protein Expression: Small-Scale

Subsequent to transformation into *E. coli* expression strain BL21(DE3)pLysS (Section 3.2.2), a small-scale protein expression trial of EcPNPase, H₆-hPNPase and H₆-Rrp4_41_42 was conducted. Several colonies were selected from a transformation plate and added to individual 30 ml sterilin tubes containing 10 ml of selective LB liquid media (Miller). Each 10 ml culture was incubated at 37°C (250 rpm) overnight and was used as starter cultures for small-scale expression trials. 10 ml of fresh selective LB liquid media was inoculated with 10 μ l of each starter culture (1:1000). Individual cultures were grown at 37°C (250 rpm) for approximately 3 hours or until an OD₆₀₀ 0.4 - 0.6 was reached. A 1 ml sample (pre-IPTG induction) was taken for each and IPTG was added to the remaining culture, resulting in a 1 mM final concentration. Cultures were grown for a further 3

hours, a 1 ml (post-IPTG induction) sample was then taken and the OD₆₀₀ was recorded. At each stage, individual samples (pre- and post-IPTG) were centrifuged in 1.5 ml Eppendorf tubes for 5 minutes (3,000 rpm) in a Thermo Scientific Heraeus Fresco 17 benchtop centrifuge. The pellet was retained, 100 µl of 2X SDS-PAGE reducing loading dye was added and samples were boiled at 95°C for 10 minutes prior to analysis by SDS-PAGE (Section 2.10.2.3). Samples for each colony of pre- and post-IPTG induction were loaded, based on comparable cell density. The colony displaying high heterologous protein expression was determined by densitometric analysis of the SDS-PAGE gel and glycerol stocks were prepared from the relevant starter culture and stored at -80°C (Section 2.9.3).

3.2.4 Protein Expression: Large-Scale

Following the small-scale trial (Section 3.2.3), optimisation of EcPNPase, H₆-hPNPase and H₆-Rrp4_41_42 protein expression was conducted on a larger scale.

Protein expression was conducted using the *E. coli* glycerol stock previously shown to display high protein expression (Section 3.2.3). An initial 10 ml starter culture of selective LB liquid media was inoculated with the relevant glycerol stock and grown at 37°C (250 rpm) O/N. 1 ml of this starter culture was added to baffled flasks containing 500 ml of selective LB liquid media. Typically, 3-4 L of culture was grown at 37°C (250 rpm) until the OD₆₀₀ optimal for protein-specific expression was reached. At this stage protein expression was induced with IPTG and cultures were incubated further. Protein-specific conditions for IPTG induction and bacterial growth are listed in Table 3.1. Following protein production, cell pellets were harvested in 1 L polypropylene bottles by centrifugation at 4°C (7,000 rpm) for 20 minutes in a Beckman Coulter Avanti J-20XP refrigerated centrifuge (JLA 8.1000 rotor). The resulting bacterial cell pellets were stored in 50 ml falcon tubes at -20°C for subsequent purification. As in the small-scale trial (Section 3.2.3), pre- and post-IPTG induction samples were analysed by SDS-PAGE for protein expression. It was critical to ensure sufficient quantities of protein had been produced prior to purification.

3.2.5 Protein Purification

Following large-scale protein expression (Section 3.2.4), EcPNPase, H₆-hPNPase and H₆-Rrp4_41_42 purification was conducted. As standard, samples exposed to potential protease degradation were purified at 4°C. This was conducted on the AKTA prime platform first, before moving to a more automated AKTA express system when these protocols were better established.

If previously published methods were available, recombinant proteins were purified identically, or appropriate adaptations were made if necessary. Conveniently, the multi-step purification process was similar for each protein, especially H₆-tagged proteins. Hence, for simplicity, individual purification steps are covered in general in Section 3.2.5, followed by protein-specific methods, which outline the precise combination of the steps in more detail in Section 3.2.6.

3.2.5.1 *E. coli* Cell Lysis

1 L cell pellets, previously prepared and stored at -20°C, were thawed on ice and re-suspended in 50 ml of appropriate lysis buffer with one cOmplete, EDTA-free Protease Inhibitor Cocktail Tablet (Roche Applied Science). If necessary, in order to obtain the uniformly homogenous cell lysate required for efficient sonication, cells were physically disrupted using a 30 ml manual homogeniser. Bacterial cells, kept on ice, were further disrupted by sonication using a Sonics Vibra-cell sonicator. The probe was set to an amplitude of 40%, a maximum temperature of 20°C and samples were sonicated for 10 minutes, with a 3.3 second on/ 9.9 second off pulse setting.

3.2.5.2 Isolation of Soluble Proteins

Soluble proteins were separated from insoluble proteins using centrifugation for 30 minutes (18,000 rpm) at 4°C in a Beckman Coulter Avanti JXN-26 Centrifuge (JA25.50 rotor). The cellular debris was discarded but the supernatant was retained for further purification.

3.2.5.3 RNase & DNase Treatment

Protein samples were treated with RNase A (10 µg/ml) and DNase 1 (1 U/ml) by mixing for 30 minutes at 4°C to remove RNA and DNA contaminants.

3.2.5.4 Heat Step & Centrifugation

Purification of thermostable proteins utilised heat precipitation; samples were heated at 75°C for 30 minutes in a water bath. Soluble, heat-stable proteins were clarified from those denatured by centrifugation as described Section 3.2.5.2.

3.2.5.5 Ammonium Sulphate Precipitation

Protein samples were treated with 51.2% (w/v) (NH₄)₂SO₄, by mixing for 1 hour at 4°C. Proteins were clarified by centrifugation as described Section 3.2.5.2, however the pellet containing the insoluble, salt precipitated, proteins were retained for further analysis.

3.2.5.6 Dialysis

The pellet from the previous ammonium sulphate step was re-suspended and dialysed O/N. using Spectra/Por® dialysis membrane with an appropriate molecular weight cut-off (3.5 kDa MWCO) (Spectrum Laboratories, Inc.), and an appropriate buffer.

3.2.5.7 Immobilised Metal Affinity Chromatography

Using the appropriate AKTA system, a 1 ml HisTrap FF column, pre-packed with Ni Sepharose 6 Fast Flow resin, was pre-equilibrated in an appropriate binding buffer. The protein solution was initially filtered (0.22 µm) and then applied to the column at a flow rate of 1 ml/min. The flow-through of unbound protein was retained for SDS-PAGE analysis (Section 2.10.2.3) in case the protein of interest was not bound to the column. Weakly binding, non-specific proteins were

washed off the column with binding buffer. Once a stable $A_{280\text{ nm}}$ baseline was reached, recombinant H₆-proteins and histidine-rich *E. coli* host proteins bound to the column were eluted. An appropriate elution buffer (supplemented with 500 mM imidazole) was applied from 0-100% over a 30 ml gradient at 1 ml/min; 2 ml fractions were collected for SDS-PAGE analysis (Section 2.10.2.3). Protein-containing fractions were pooled for further purification by other chromatography methods.

3.2.5.8 Anion Exchange Chromatography

Using the appropriate AKTA system, a 5 ml HiTrap Q HP column (High Performance), pre-packed with Q Sepharose resin, was pre-equilibrated in an appropriate binding buffer. The protein solution was initially filtered (0.22 μm) and then applied to the column at a flow rate of 1 ml/min. The flow-through of unbound protein was retained for SDS-PAGE analysis (Section 2.10.2.3). Weakly binding, non-specific proteins were washed off the column with binding buffer. Once a stable $A_{280\text{ nm}}$ baseline was reached proteins bound to the column were eluted. An appropriate elution buffer (supplemented with NaCl) was applied from 0-100% over a 30 ml gradient at 1 ml/min and 1 ml fractions were collected for SDS-PAGE analysis (Section 2.10.2.3). Protein-containing fractions were pooled for further purification by other chromatography methods.

3.2.5.9 Size Exclusion Calibration

Using the appropriate AKTA system; the manually re-packed Superdex 200 20/30 column was calibrated. Two separate 1 ml protein standard samples were prepared and loaded at 1 ml/min in 20 mM Tris-HCl pH 8, 150 mM NaCl and 10 mM MgCl₂ onto the column. Sample 1 contained Blue dextran, Aldolase and Ovalbumin (each 10 mg/ml) and sample 2, Ferritin and Conalbumin (each 10 mg/ml). The $A_{280\text{ nm}}$ (mAU) was recorded and the values plotted against the elution volume. The resultant elution volumes were converted into retention volumes using Blue Dextran as an indicator of the void volume using the following calculation: Retention volume (ml) = elution volume (ml) – void volume (ml). Once the individual retention volumes for each protein standard were calculated, they were plotted against the log₁₀ of the known molecular weight (log MW, (kDa)). From the resultant graph the equation for the line of best fit was used to predict the MW of any unknown proteins with a known elution volume.

3.2.5.10 Protein Concentration

Depending on the starting volume, samples were concentrated in 20 ml or 2 ml Vivaspin concentrators (Sartorius). Concentrators were washed and pre-equilibrated in 5 ml ddH₂O and appropriate SE buffer respectively for 2 minutes at 3,000 rpm in an Eppendorf 5810 R Centrifuge (A-4-62 Rotor). Protein samples were then applied and concentrated by centrifugation at 3,000 rpm for various times, dependent on the initial and desired final volume. Protein samples were re-suspended between spins to avoid precipitation on the concentrator membrane and heat stable proteins were concentrated at RT for the same reason.

If the protein sample was required for further purification by SEC, it was concentrated to < 1ml. The sample was then centrifuged at maximum rpm for 10 minutes in a benchtop Thermo Scientific Heraeus Fresco 17 centrifuge to precipitate any insoluble proteins prior to loading.

3.2.5.11 Size Exclusion Chromatography

Concentrated protein samples (1 ml) for further purification by SEC were loaded at 1 ml/min onto a pre-equilibrated Superdex 200 20/30 column in an appropriate SE buffer, as described in Section 3.2.5.9. Resultant SEC fractions corresponding to $A_{280\text{ nm}}$ peaks were analysed by SDS-PAGE. Typically, samples were concentrated with strata-clean (Section 2.10.2.3) for analysis. Relevant fractions were pooled, concentrated to an appropriate volume/concentration for further use (Section 3.2.5.10) and stored at -80°C in individual aliquots.

3.2.6 Protein Purification: Protein-Specific

As outlined in Section 3.2.5, this section describes the protein-specific methods utilised to purify recombinant proteins EcPNPase, H₆-hPNPase and H₆-Rrp4_41_42.

3.2.6.1 hPNPase

A similar purification method to that conducted in (Lin *et al.*, 2012) was employed. *E. coli* 1 L cell pellets (x 3 L) containing the overexpressed recombinant protein *H. sapiens* PNPase (hPNPase) were lysed in 50 ml lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl and 10 mM imidazole) by manual homogenisation and sonication (Section 3.2.5.1). Soluble proteins were then clarified by centrifugation (Section 3.2.5.2) and the supernatant retained for RNase and DNase treatment (Section 3.2.5.3). Following this, the treated sample was filtered (0.22 µm) and loaded onto a 1 ml HisTrap nickel-immobilised metal affinity column (IMAC, 0.3 mPa) pre-equilibrated in lysis buffer at 1 ml/min. The column was washed with lysis buffer and proteins were eluted with in a 30 ml gradient (0-100%) of elution buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 500 mM imidazole). 1 ml fractions were collected, the mAU 280_{nm} was recorded and SDS-PAGE analysis was conducted on protein-containing fractions (Section 3.2.5.7). Relevant fractions were pooled and concentrated prior to loading onto a Superdex 200 20/30 column (0.5 mPa) pre-equilibrated in SE Buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM MgCl₂) at 1 ml/min (Section 3.2.5.10-3.2.5.11). SDS-PAGE analysis of strata cleaned fractions was conducted (Section 2.10.2.3) and relevant protein-containing fractions were pooled, concentrated and stored at -80°C for future use.

3.2.6.2 EcPNPase

A similar EcPNPase purification method to that conducted in Nurmohamed *et al.*, 2011 was employed. *E. coli* 1 L cell pellets containing the overexpressed recombinant protein *E. coli* PNPase (EcPNPase) were lysed in 50 ml (20 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM MgCl₂) by manual homogenisation and sonication (Section 3.2.5.1). Soluble proteins were then clarified by centrifugation (Section 3.2.5.2) and the supernatant retained for RNase and DNase treatment (Section 3.2.5.3). Following this, ammonium sulphate (NH₄)₂SO₄ was added to a final

concentration of 51.2% (w/v) and EcPNPase was precipitated by mixing at 4°C for 1 hour (Section 3.2.5.5). The resultant insoluble proteins were clarified by centrifugation (Section 3.2.5.2) and the pellet retained for re-suspension in 20 ml Buffer A (20 mM Tris-HCl pH 8.5, 30 mM NaCl, 5 mM MgCl₂ and 5 mM DTT). The re-suspended pellet (20 ml) was dialysed in 500 ml Buffer A overnight (and placed in fresh Buffer A afterwards) (Section 3.2.5.6). The dialysed sample was filtered (0.22 µm) and loaded onto a 5 ml HiTrap Q Sepharose High Performance (HP) anion exchange column (AEC, 0.3 mPa) pre-equilibrated in Buffer A at 1 ml/min. The column was washed with Buffer A and proteins were eluted with in a 50 ml gradient (0-100%) of Buffer B (20 mM Tris-HCl pH 8.5, 1 M NaCl, 5 mM MgCl₂ and 5 mM DTT). 1 ml fractions were collected, the mAU 280_{nm} was recorded and SDS-PAGE analysis was conducted of the protein-containing fractions (Section 3.2.5.8). Relevant fractions were pooled and concentrated prior to loading onto a Superdex 200 20/30 column (0.5 mPa) pre-equilibrated in SE Buffer (20 mM Tris-HCl pH 8, 150 mM NaCl & 10 mM MgCl₂) at 1 ml/min (Section 3.2.5.10-3.2.5.11). SDS-PAGE analysis of fractions was conducted (Section 2.10.2.3) and relevant protein-containing fractions were pooled, concentrated and stored at -80°C for future use.

3.2.6.3 SsoExosome

A similar SsoExosome purification method to that conducted in Lorentzen *et al.*, 2007 was employed. *E. coli* 1 L cell pellets containing the overexpressed recombinant protein SsoExosome were lysed in 50 ml lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl and 10 mM imidazole) by manual homogenisation and sonication (Section 3.2.5.1) Soluble proteins were then clarified by centrifugation (Section 3.2.5.2) and heat treated at 75°C for 30 min (Section 3.2.5.4). Heat-stable proteins were isolated from heat-denatured proteins by a second centrifugation step and the supernatant was retained for RNase and DNase treatment (Section 3.2.5.3). Following this, the treated sample was filtered (0.22 µm) and loaded onto a 1 ml HisTrap nickel IMAC column (0.3 mPa) pre-equilibrated in lysis buffer at 1 ml/min. The column was washed with lysis buffer and proteins were eluted with a 30 ml gradient (0-100%) of elution buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 500 mM imidazole). 1 ml fractions were collected, the mAU 280_{nm} was recorded and SDS-PAGE analysis was conducted of protein-containing fractions (Section 3.2.5.7). Relevant fractions were pooled and concentrated prior to loading onto a Superdex 200 20/30 column (0.5 mPa) pre-equilibrated in SE Buffer (20 mM Tris-HCl pH 8, 150 mM NaCl & 10 mM MgCl₂) at 1 ml/min (Section 3.2.5.10-3.2.5.11). SDS-PAGE analysis of strata cleaned fractions was conducted (Section 2.10.2.3) and relevant protein-containing fractions were pooled, concentrated and stored at -80°C for future use.

3.2.7 Protein Identification

Following protein purification (Section 3.2.6), confirmation of the purified EcPNPase, H₆-hPNPase and H₆-Rrp4_41_42 protein identities was conducted using a variety of techniques.

3.2.7.1 SDS-PAGE and Western Blot

At each stage of protein purification samples were retained for subsequent analysis by SDS-PAGE and if necessary western blot analysis was also conducted (Section 2.10.2.3 and 2.11 respectively). The aim of this was to determine if the protein of interest had been purified successfully away from contaminating *E. coli* host proteins. Firstly, using predicted MW values from ProtParam (Appendix 9.6), migration of the protein of interest through the gel was compared to a MW ladder to estimate if the purified protein was of the predicted size (Wilkins *et al.*, 1999). Secondly, using antibodies specific to the protein itself or its H₆-tag, the protein of interest was detected using western blots (Section 2.11).

3.2.7.2 Size Exclusion Chromatography

Not only was SEC utilised to purify the recombinant protein, the elution volume of each recombinant protein determined by SEC, in combination with the calibration curve, was used to predict the overall size of eluted proteins. Thus providing an indication of whether the protein was in its correct oligomeric state, as described in Section 3.2.5.9.

3.2.7.3 Denaturing Mass Spectrometry (LC-MS/MS)

Subsequent liquid chromatography tandem mass spectrometry (LC-MS-MS) of relevant proteins was performed to confirm protein identity. Samples to be identified were separated by SDS-PAGE (5 µg protein loaded) as standard (Section 2.10.2.3). In a laminar flow hood, all equipment required for protein band excision was cleaned with 70% EtOH. This reduced the chance of protein contaminants that interfere with LC-MS-MS identification. Once the gel band was excised, it was placed in a 1.5 mL microcentrifuge tube and covered in 30% ethanol. Samples were analysed by the Astbury Centre Mass Spectrometry Facility (University of Leeds) and a full analysis was provided including protein identification.

3.2.8 3'-5' RNA Degradation Activity of Recombinant Proteins

After establishing the identity of recombinant proteins (Section 3.2.7), a gel electrophoresis-based RNA degradation activity assay was conducted for EcPNPase, H₆-hPNPase and H₆-Rrp4_{41_42} to confirm 3'-5' phosphorolytic exoribonuclease activity. Typically experiments which determine a particular enzyme's activity should be conducted in optimal conditions. However, since an initial check was required to examine if an intact and active protein was purified, factors including the inorganic phosphate (Pi) concentration, magnesium metal cofactor concentration and temperature were not optimised. Specific canonical enzyme degradation activity is further characterised in Chapter 4.

700 nM of 5' fluorescein (F) labelled Poly(A) 20mer RNA substrate (5'F Poly(A)₂₀) was incubated at 37°C with 240 nM of each protein, in a buffer containing 15 mM Tris pH 8, 112.5 mM NaCl, 3.75 mM MgCl₂, and 0.045 mM Na₂PO₄. Control experiments were carried out in the absence of the relevant recombinant protein (RNA only). The rate of substrate degradation varied for each

protein; hence reactions were quenched with 0.1 M EDTA at various times. Degradation products of the 5'F Poly(A)₂₀ substrate were separated on a 20% denaturing urea gel at 4°C, visualised using a Fujifilm FLA-5000 phosphorimager and quantified by densitometric analysis using Image Gauge software (Science Lab Ver. 4.0, Fujifilm) (Section 2.10.2.2).

3.3 Results

This section describes the results of cloning, expression and purification of the recombinant EcPNPase, H₆-hPNPase and H₆-SsoExosome (H₆-Rrp4_41_42) proteins. The majority of results from the initial cloning and expression stages have been summarised together for simplicity. However, the purification for each individual protein is displayed under separate sub-headings. This section also provides the results for the protein identification and 3'-5' degradation activity post-purification.

3.3.1 Recombinant DNA Preparation

The genes encoding H₆-hPNPase, and H₆-SsoExosome were successfully synthesised using GeneArt Cloning (Section 3.2.1). They were each ligated into their relevant pET vectors; the first into pET-28b (kanamycin resistant) and the latter into pETMCN-EAVNH (ampicillin resistant) using relevant restriction sites (NheI/SalI or NdeI/XbaI respectively). For each synthetic construct, DNA sequencing identified 100% sequence congruence to the provided template sequence; indicating successful cloning by GeneArt (Life Technologies) (data not shown). EcPNPase previously cloned into pETDuet-1 vector (ampicillin resistant) was kindly supplied by Prof. Ben Luisi (University of Cambridge).

Following successful cloning, transformation of plasmid DNA (containing the gene construct of interest) into a desired bacterial cell strain was confirmed using selective LB agar plates (Section 3.2.2). Using appropriate antibiotics and positive and negative controls pET-28b_H₆-hPNPase, pETMCN-EAVNH_H₆-SsoExosome (H₆-Rrp4_41_42) and pETDuet-1_EcPNPase plasmids were shown to successfully transform into DH5α cells. Subsequent DNA replication, purification and sequencing indicated that all plasmid DNA synthesised by GeneArt had the correct sequence. However, sequencing results for the provided full-length EcPNPase plasmid DNA indicated multiple nucleotide point mutations within the coding region (data not shown). Ensuring that the DNA nucleotide sequence, and thus the encoded protein amino acid sequence was correct was crucial. Hence, site-directed mutagenesis was conducted using Agilent Technologies QuickChange® II Site-Directed Mutagenesis Kit as described in Section 2.7.5. Re-sequencing following multiple rounds of SDM confirmed successful mutagenesis within the coding region.

In summary, results showing colony growth on selective agar plates spread with DH5α cells transformed with relevant plasmids (pET-28b_H₆-hPNPase, pETMCN-EAVNH_H₆-SsoExosome (H₆-Rrp4_41_42) or pETDuet-1_EcPNPase) indicated successful transformations. The subsequent

sequencing results (following SDM for EcPNPase) provided evidence of correct plasmid sequence for use in subsequent protein expression and purification. The gene cloning strategy is provided in detail in the Appendix (9.6).

3.3.2 Protein Expression

As before, results showing colony growth on selective agar plates spread with either *E. coli* BL21(DE3)pLysS cells (transformed with pET-28b_H₆-hPNPase, pETMCN-EAVNH_H₆-SsoExosome (H₆-Rrp4_41_42) or pET-Duet1_EcPNPase) indicated successful transformations into a strain viable for protein expression. Colonies from each transformation were screened using a small-scale protein expression trial (Section 3.2.3) to check protein production. Typically, all colonies tested displayed similar protein expression. Hence, a random individual colony was selected for generating glycerol stocks (Section 2.9.3).

Glycerol stocks, shown to express protein in the small-scale expression trial were selected and large-scale expression trials were conducted, for each recombinant protein, from their respective starter cultures (Section 3.2.4). The conditions optimal for protein expression, as analysed by SDS-PAGE in Figure 3.1 are listed in Table 3.1. EcPNPase was expressed in the same conditions to those previously reported (Figure 3.1 (a)) (Nurmohamed *et al.*, 2011). Detailed conditions for H₆-SsoExosome were not available in the previous literature (Lorentzen, Dziembowski, Lindner, Seraphin, & Conti, 2007 or Lu, Ding, & Ke, 2010), hence conditions previously suggested by (Bralley & Jones, 2003) to be suitable for PNPases were utilised and protein expression was successful in these conditions (Figure 3.1 (c)) (Table 3.1). In order to express hPNPase, conditions recommended by Lin *et al.*, 2012 were used initially, however protein expression was sub-optimal (data not shown). An expression trial was therefore conducted and the results suggested a lower IPTG (0.1 mM) was required to obtain the protein expression shown in Figure 3.1 (b) (Table 3.1).

Analysis of SDS-PAGE large-scale expression results shown in Figure 3.1 provided evidence that protein bands corresponding to the sizes expected for EcPNPase, H₆-hPNPase, and H₆-SsoExosome (H₆-Rrp4_41_42) proteins were present. Although the ProtParam estimated MW of EcPNPase was 77 kDa, previous literature has shown the protein to migrate at ~ 84 kDa, consistent with results displayed in Figure 3.1 (a) (Soreq, H., Littauer, 1977).

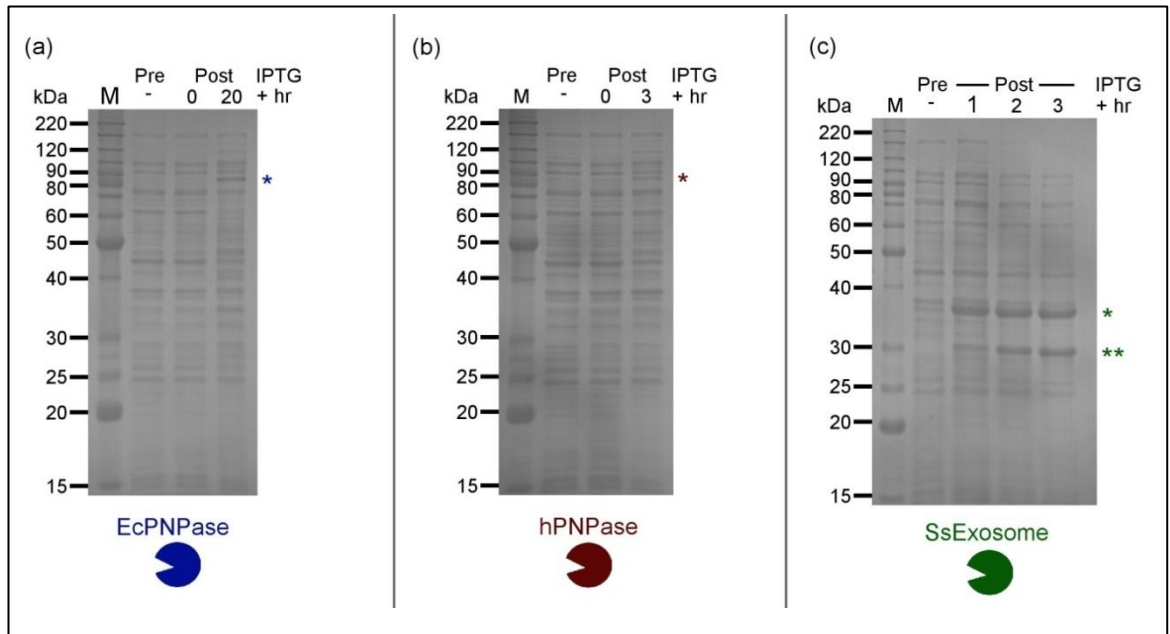


Figure 3.1 EcPNPase, hPNPase and SsoExosome Protein Expression Summary

12% reducing SDS-PAGE analysis of (a) EcPNPase, (b) hPNPase and (c) SsoExosome (Ss) large-scale expression alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. Samples of pre and post-IPTG induction at various times (indicated above lane in hours (hr)) were loaded and bands corresponding to specific protein monomers are highlighted with blue, red or, green asterisk (*) respectively. A single band, which migrates at the size expected for monomeric EcPNPase (77 kDa) and hPNPase (83 kDa) was observed. In contrast, due to the nature of SsoExosome being a hetero-trimeric of three Rrp4, 41 and 42 monomers (31, 28, and 30 kDa respectively), three bands were anticipated. Only two bands were detected as two of the monomers (Rrp4 and 42) migrated at the same size, as shown by a double asterisk (**).

In summary, protein expression results justified the progression to protein purification, with the aim to remove contaminating *E. coli* host proteins.

Recombinant Protein	Pre-IPTG induction			Post-IPTG induction	
	OD _{600 nm}	Temperature (°C)	IPTG Concentration (mM)	Temperature (°C)	Time (Hours)
EcPNPase	0.6	37	0.5	20	20
H ₆ hPNPase	0.6	37	0.1	25	3
H ₆ SsoExosome	0.7	37	1	37	3

Table 3.1 Protein-Specific Expression Conditions

The expression conditions for each recombinant protein are listed with details of the temperature and OD_{600 nm} to which *E. coli* BL21(DE3)pLysS cell cultures were grown prior to induction with IPTG. The concentration of IPTG, temperature and duration of growth following protein expression is also provided (Lin *et al.*, 2012; Nurmohamed *et al.*, 2011).

3.3.3 Size Exclusion Chromatography Calibration Curve

Prior to protein purification, a SEC calibration curve was generated from a calibration run using protein standards with known molecular weights (MW) (Section 3.2.5.9). For MW standards see Table 3.2. The resultant chromatogram showed, as expected, protein standards from individual SEC runs 1 and 2 to elute based on their overall size; larger MW proteins eluted earlier than smaller MW proteins (Figure 3.2 (a)). Hence both chromatograms were combined in Figure 3.2 (a) and elution volumes for protein standards were used to generate a calibration curve in Figure 3.2 (b). Results demonstrate that as the protein size (Log MW (kDa)) decreased the retention within the SEC column increased (Retention Volume (ml)). The data points in Figure 3.2 (b) fit well to the straight line equation $y = -0.0146x + 3.0456$ (where $y = \log \text{MW (kDa)}$ and $x = \text{retention volume (ml)}$). The standard error (+/-) was reasonable for the intercept 3.0456 (+/- 0.0228) and the gradient -0.0146 (+/- 0.0003). To further check the accuracy of the equation, elution volumes of protein standards were used to re-calculate MW values and these were within <4% (+/-) the actual value, as shown in Table 3.2. In summary, the calibration curve equation is suitable for use to predict unknown protein MWs, from known elution volumes, following separation by SEC.

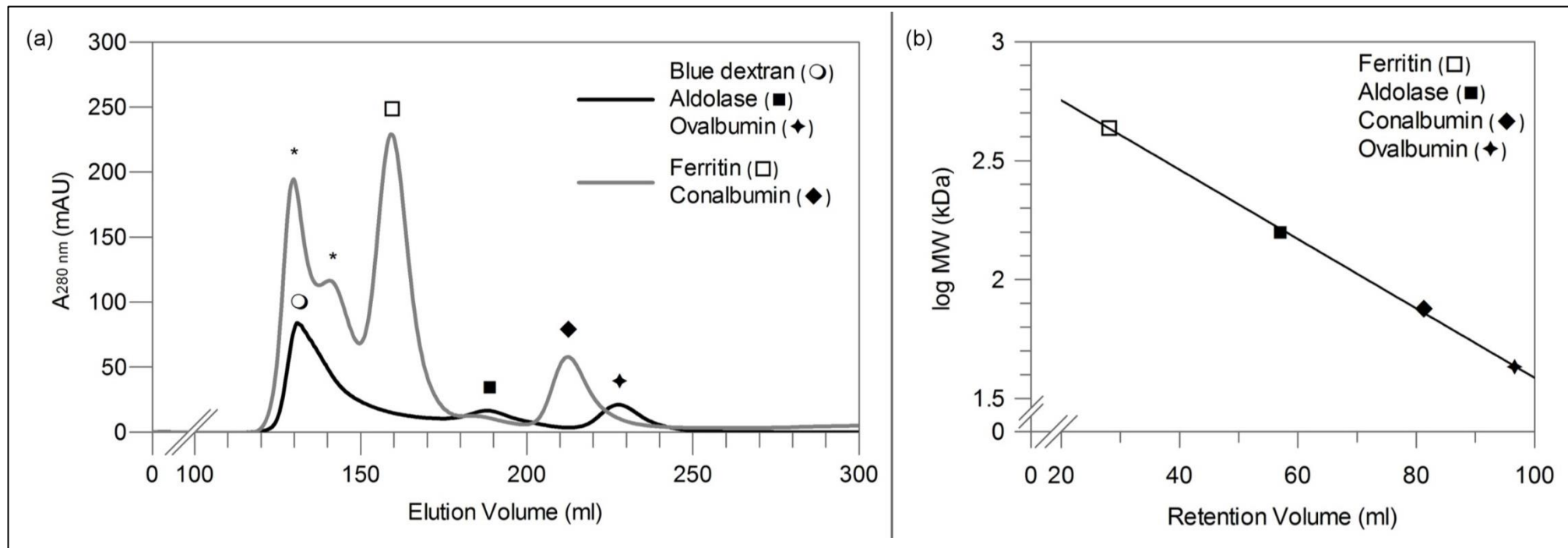


Figure 3.2 SEC Calibration

(a) SEC chromatogram; the absorbance units measured at 280 nm ($A_{280 \text{ nm}}$ (mAU)) over the volume (ml) is shown, with two calibration samples overlaid on one graph (1: Blue dextran (2000 kDa), Aldolase (158 kDa) and Ovalbumin (43 kDa) (black line) and 2: Ferritin (440 kDa) and Conalbumin (75 kDa) (grey line)). The elution volume of Blue Dextran (131.07 ml) was used as an indicator of the columns void volume. The protein standard Ferritin (\square) in sample 2 (grey line) has three peaks; the two labelled (*) are Ferritin oligomers, which elute near the void volume, and the peak labelled (\square) (159.11 ml) consists of Ferritin monomers. (b) Retention volumes (ml) (elution volume – void volume) for monomeric protein standards were plotted against log MW (kDa) to generate a calibration curve and calculate a line of best fit equation: $y = -0.0146x + 3.0456$ (where $y = \log \text{MW}$ (kDa) and $x = \text{retention volume}$ (ml)). The standard error of the intercept and gradient was 3.0456 (+/- 0.0228) and -0.0146 (+/- 0.0003) respectively.

Protein Standard	MW (kDa)	Log MW (kDa)	Elution Volume (mL)	Retention Volume (mL)	Recalculated MW (kDa)
Blue Dextran	2000	3.30	131.07	0.00	-
Ferritin	440	2.64	159.11	28.04	432.73 (+1.68%)
Aldolase	158	2.20	187.93	56.86	164.23 (-3.79%)
Conalbumin	75	1.88	212.27	81.20	72.46 (+3.51%)
Ovalbumin	43	1.63	227.62	96.55	43.25 (-0.57%)

Table 3.2 SEC Protein Standards

Protein standards are listed with their molecular weights (MW) and calculated log MW values (both kDa). For each standard, SEC elution volumes (ml) and calculated retention volumes (ml) (retention volume = elution volume – void volume) are shown. Protein standard Blue Dextran’s elution volume (131.07 ml) was used as an indicator of this specific SEC column’s void volume. The equation of the graph; calculated retention volumes (ml) plotted against log MW (kDa) (Figure 3.2 (b)) was used to recalculate MWs (kDa). Equation: $y = -0.0146x + 3.0456$ (where $y = \log \text{MW (kDa)}$ and $x = \text{retention volume (ml)}$). The percentage difference (+/- %) indicated relates to the difference between recalculated and actual MWs.

3.3.4 Protein Purification: Protein-Specific

Results for the purification of individual proteins: EcPNPase, H₆-hPNPase and H₆-SsoExosome are provided.

3.3.4.1 *Escherichia coli* PNPase

EcPNPase was purified using a similar method to that conducted in Nurmohamed *et al.*, 2011, see methods for details (Section 3.2.6.2). At each stage of the purification process samples were retained for SDS-PAGE analysis, see Figure 3.5. Although, for EcPNPase, ProtParam predicts a monomeric size of 77 kDa (Appendix 9.6), previous work provided evidence, that upon denaturing SDS-PAGE analysis, EcPNPase migrates slightly higher (~85 kDa) than expected when compared to a protein molecular weight ladder (Soreq, H., Littauer, 1977). Consequently, the presence of an intense band ~ 85 kDa in the lysis, sonication and supernatant purification samples (Figure 3.5) provided an initial indication of the successful extraction and clarification of soluble EcPNPase recombinant protein from overexpressed *E. coli* cells. Although the presence of other host proteins within these initial stages makes it difficult to confidently assign a visible band to a specific protein, over the purification process this becomes easier upon removal of contaminants.

The published purification strategy utilised an ammonium sulphate step to enrich for EcPNPase and then dialysis of the sample into an appropriate buffer ready for anion exchange chromatography. A band (~85 kDa) indicative of EcPNPase is indicated in the following dialysis and pre-HiTrap samples (Figure 3.5). Initial analysis of these lanes suggests little removal of contaminants; however, the early samples are normalised for 50 ml of cell lysis whereas the dialysis and pre-HiTrap samples are the result of pooling 100 ml soluble protein (supernatant samples). Nevertheless, AEC separated EcPNPase further, as highlighted by the presence of the ~ 85 kDa band in fraction 27 (Figure 3.3 (b), labelled with an asterisk), which related to the corresponding protein peak mAU_{280 nm} eluting at ~ 51% Buffer B (Figure 3.3 (a), labelled with an asterisk). As indicated in this figure, fraction 27 was retained for further purification by SEC. Following SEC, fractions 16-17 were pooled, concentrated and stored at -80°C as standard. This was done following the observation of the ~ 85 kDa band (Figure 3.4 (b), labelled with an asterisk) within these fractions which corresponded to the protein peak in mAU_{280 nm} recorded at ~162 ml (Figure 3.4 (a), labelled with an asterisk). Samples of these chromatography stages (pre and post-HiTrap or SEC respectively) are also provided in Figure 3.5. These lanes clearly show the removal of other contaminants until a pure band of protein migrating at ~85 kDa remains (1 µg). Using the SEC calibration curve elution volumes of 178 ml and 211 ml were predicted for trimeric and monomeric EcPNPase respectively, thus suggesting the pooled protein at ~ 166 ml (SEC fractions 16-17) is more likely to be in the higher oligomeric state (possibly an intact trimer) than monomeric form. Further protein identification by mass spectrometry and activity analysis is provided in Sections 3.3.5 and 3.3.6 respectively which confirms the purified protein to be EcPNPase.

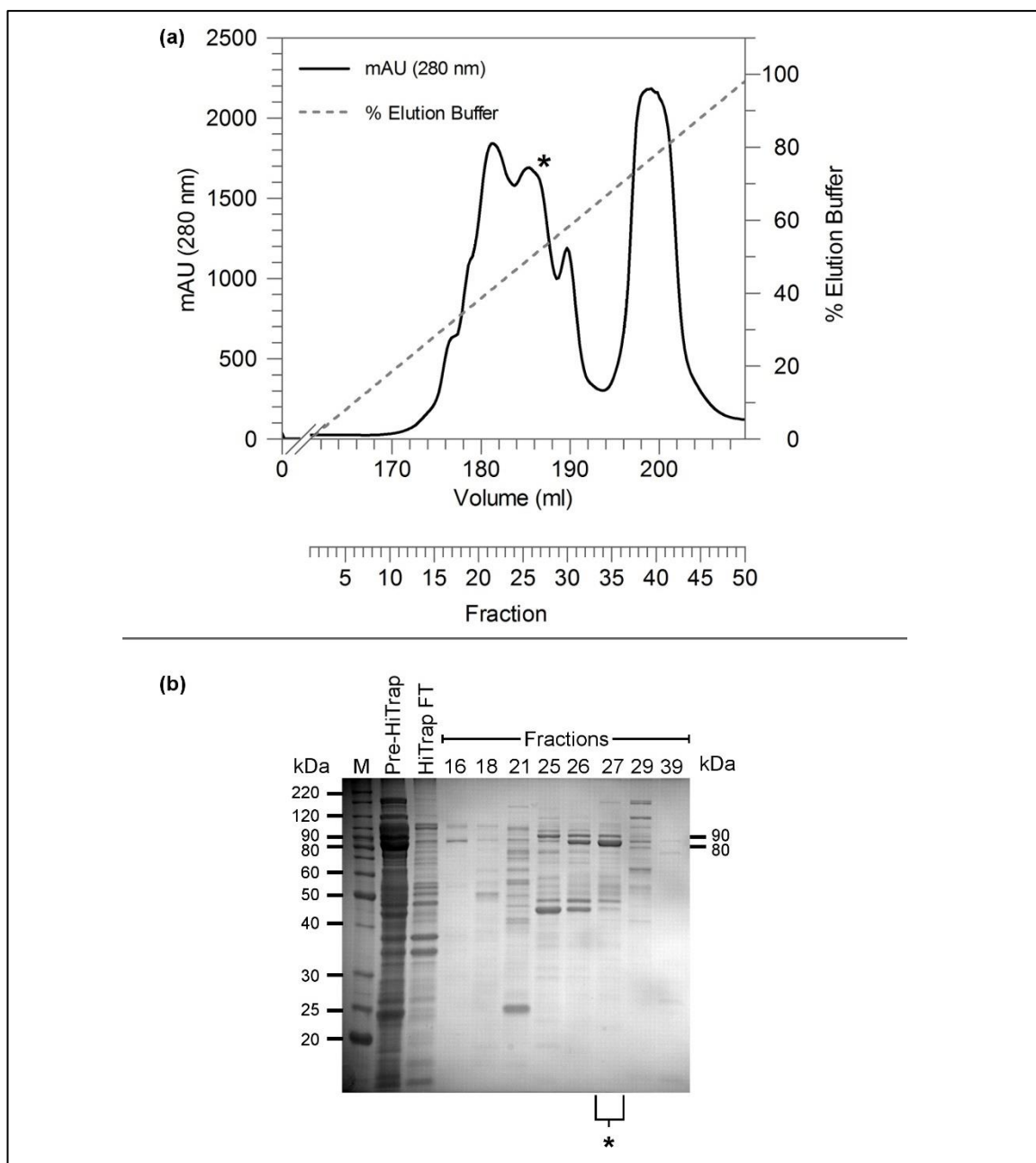


Figure 3.3 ECPNPase AEC

(a) AEC chromatogram; the absorbance units measured at 280 nm ($A_{280 \text{ nm}}$ (mAU)) over the volume (ml) is shown. The fraction number relating to the elution volume is depicted on a second x-axis. Sample loaded in Buffer A (20 mM Tris-HCl pH 8.5, 30 mM NaCl, 5 mM MgCl_2 and 5 mM DTT) was eluted (black line) with a 50 ml gradient (grey dotted line, 0-100%) of elution buffer B (Buffer A with 1 M NaCl). (b) 12% reducing SDS-PAGE analysis of pre-HiTrap (AEC), HiTrap flow through (FT) and HiTrap fractions alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. Fractions pooled for further analysis are indicated by label (*). After subsequent analysis (described later) the * symbol indicates the peak (a) and fraction (b) containing ECPNPase.

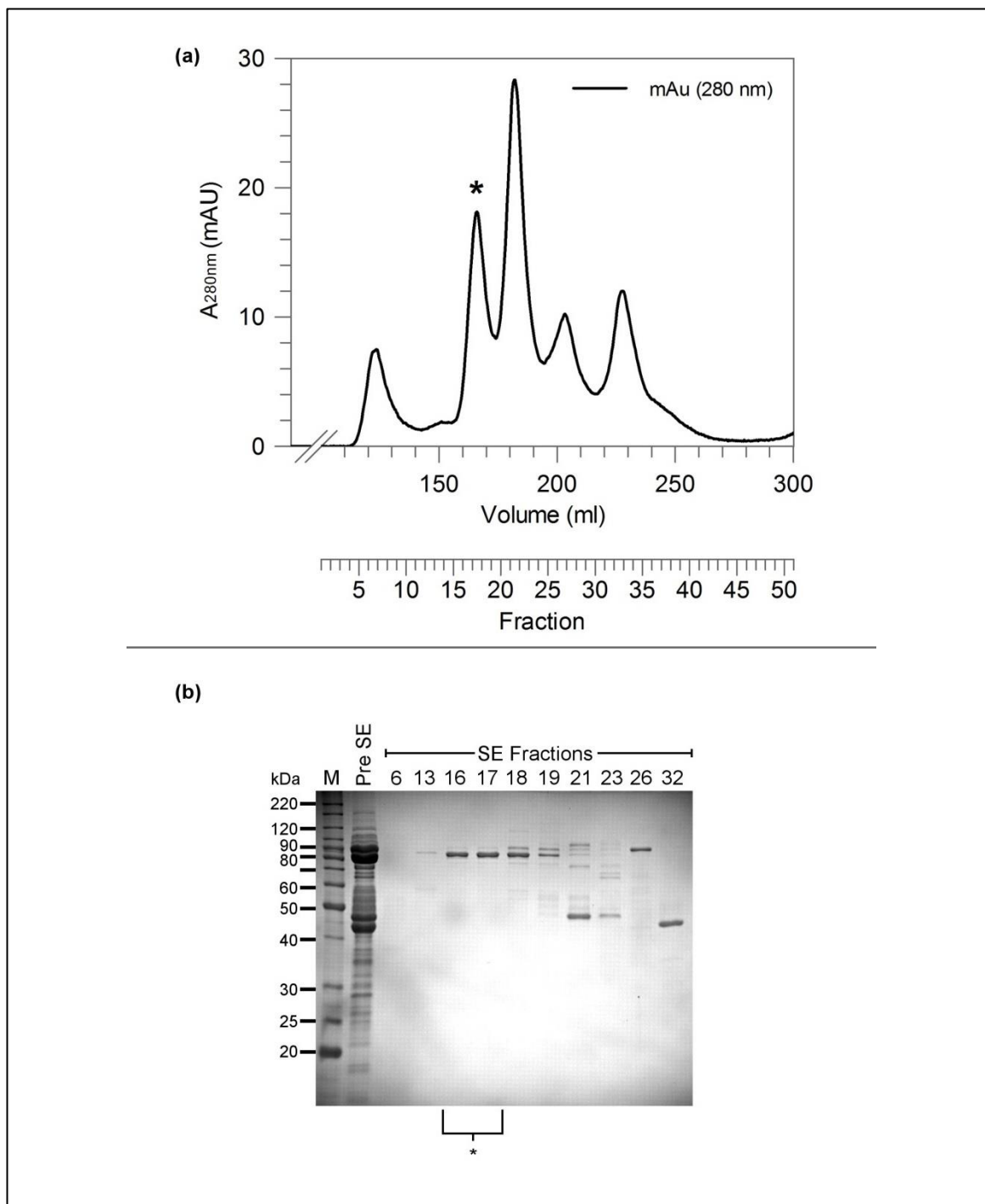


Figure 3.4 EcpNPase SEC

(a) SEC chromatogram; the absorbance units measured at 280 nm ($A_{280\text{nm}}$ (mAU)) over the volume (ml) is shown. The fraction number relating to the elution volume is depicted on a second x-axis. Sample loaded in SE Buffer (20 mM Tris-HCl pH 8, 150 mM NaCl & 10 mM MgCl_2) was also eluted (black line) in SE Buffer. (b) 12% reducing SDS-PAGE analysis of pre-SEC sample with various fractions from #6 – 32 alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. The fractions pooled for further analysis are indicated by the bracket. After subsequent analysis (described later) the * symbol indicates the peak (a) and fraction (b) containing EcpNPase.

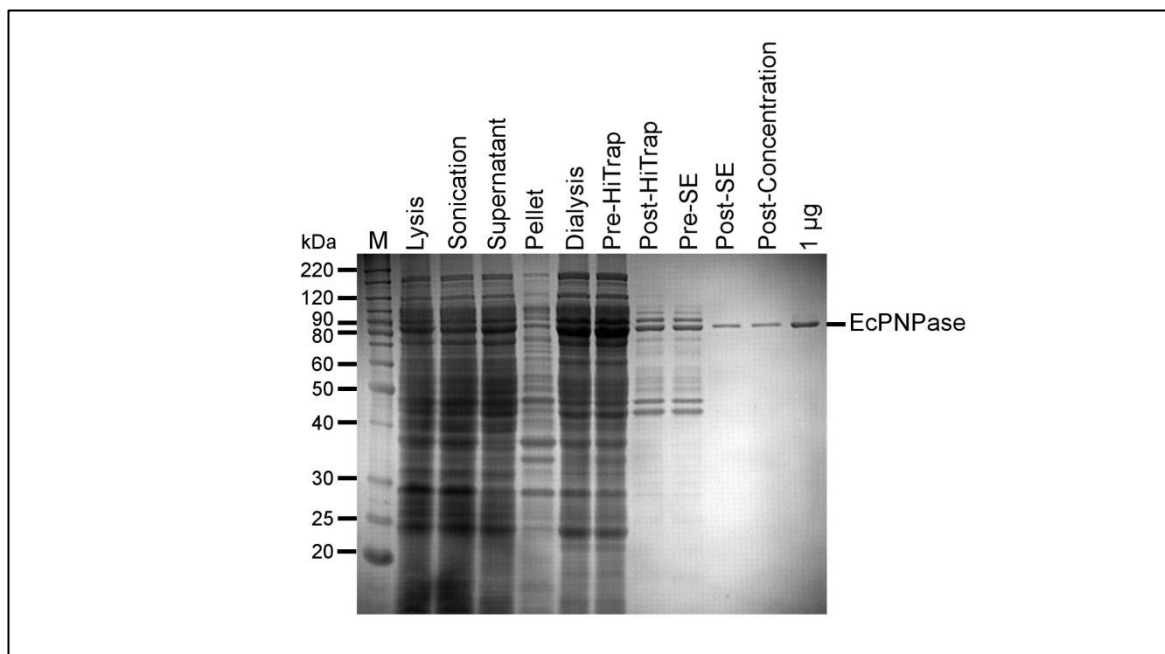


Figure 3.5 EcPNPase Purification Stages

12% reducing SDS-PAGE analysis of EcPNPase purification stages alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. Initial analysis suggests the band present in the final lane (1 μ g loaded) may be EcPNPase due to its migration at ~ 80 – 90 kDa, the size expected for the monomeric protein. This was confirmed to be EcPNPase after subsequent denaturing mass spectrometry analysis (described later).

3.3.4.2 *Homo sapiens* PNPase

H₆-hPNPase was purified using a similar method to that conducted in Lin *et al.*, 2012, see methods for details (Section 3.2.6.1). At each stage of the purification process, samples were retained for SDS-PAGE analysis (Figure 3.8). Although ProtParam (Appendix 9.6) predicts a monomeric size of 83 kDa for H₆-hPNPase, upon denaturing SDS-PAGE analysis, the purified protein migrates slightly higher than expected. when compared to a protein molecular weight ladder (Figure 3.8). This was similar to EcPNPase, which is also reported to migrates higher than expected (Soreq, H., Littauer, 1977). It was difficult to compare the migration of this band to previous hPNPase purifications as the sizes of the protein ladder used by Lin *et al.*, 2012 were quite broad (ladder bands of 72 kDa and 95 kDa respectively). Either way, the presence of a protein band between 80-90 kDa was monitored during purification. As with EcPNPase early lysis, sonication and supernatant samples were heavily contaminated, providing difficulty in monitoring a specific protein band. Following purification via IMAC however, a clear band can be observed in fractions 15-25 (Figure 3.6 (b), labelled with an asterisk) migrating between 80-90 kDa. Note a spike in pressure for the AKTA system was observed in the chromatogram (Figure 3.6 (a) labelled with a double asterisk **) and so a broad range of samples were analysed by SDS-PAGE to ensure the correct protein-containing fractions were selected (Figure 3.6 (b) labelled with a single asterisk). As indicated in this figure, fraction 15-25 was retained for further purification by SEC. Following SEC, fractions 18-22 were pooled, concentrated and stored at -80°C as standard. This was done following the observation of a band between 80-90 kDa (Figure 3.7 (b), labelled with an asterisk) within these fractions which corresponded to the protein peak in mAU_{280 nm} recorded at 180 ml (Figure 3.7 (a), labelled with an asterisk). Samples of these chromatography stages (pre- and post-

HisTrap or SEC respectively) are also provided in Figure 3.8. These lanes clearly show the removal of other contaminants until a pure band of protein migrating at ~80-90 kDa remains (0.3 µg). It is essential to highlight that 1 ml SEC fractions were concentrated to 10 µl using StrataClean and so the samples are heavily overloaded in Figure 3.7 (b) compared to the volume normalised sample loaded in Figure 3.8, which suggest that the protein has a similar level of purity compared to other proteins prepared within this body of work.

As described previously for EcPNPase, the SEC calibration curve was utilised to predict elution volumes of 176 ml and 208 ml, for trimeric (249 kDa) and monomeric (83 kDa) H₆-hPNPase respectively. This suggests that the pooled protein at ~ 180 ml (SEC fractions 18-22) is more likely to be in a higher trimeric than monomeric form. Western blot analysis of this purified protein indicated it was a H₆-fusion protein and more specifically hPNPase, when using H-tag and Anti-hPNPase primary antibody probes respectively, Section 2.11. The concentration of hPNPase was determined using a Nanodrop 2000 (Desjardins, Hansen, & Allen, 2009); although the yield was low (typical yield ~1 mg/ml) the purity was high (>95%) and comparable to the commercially available H₆-hPNPase (Sigma) which migrated at the same size (not shown). Further protein identification by mass spectrometry and activity analysis is provided in Sections 3.3.5 and 3.3.6 respectively and this confirmed the purified protein to be H₆-hPNPase.

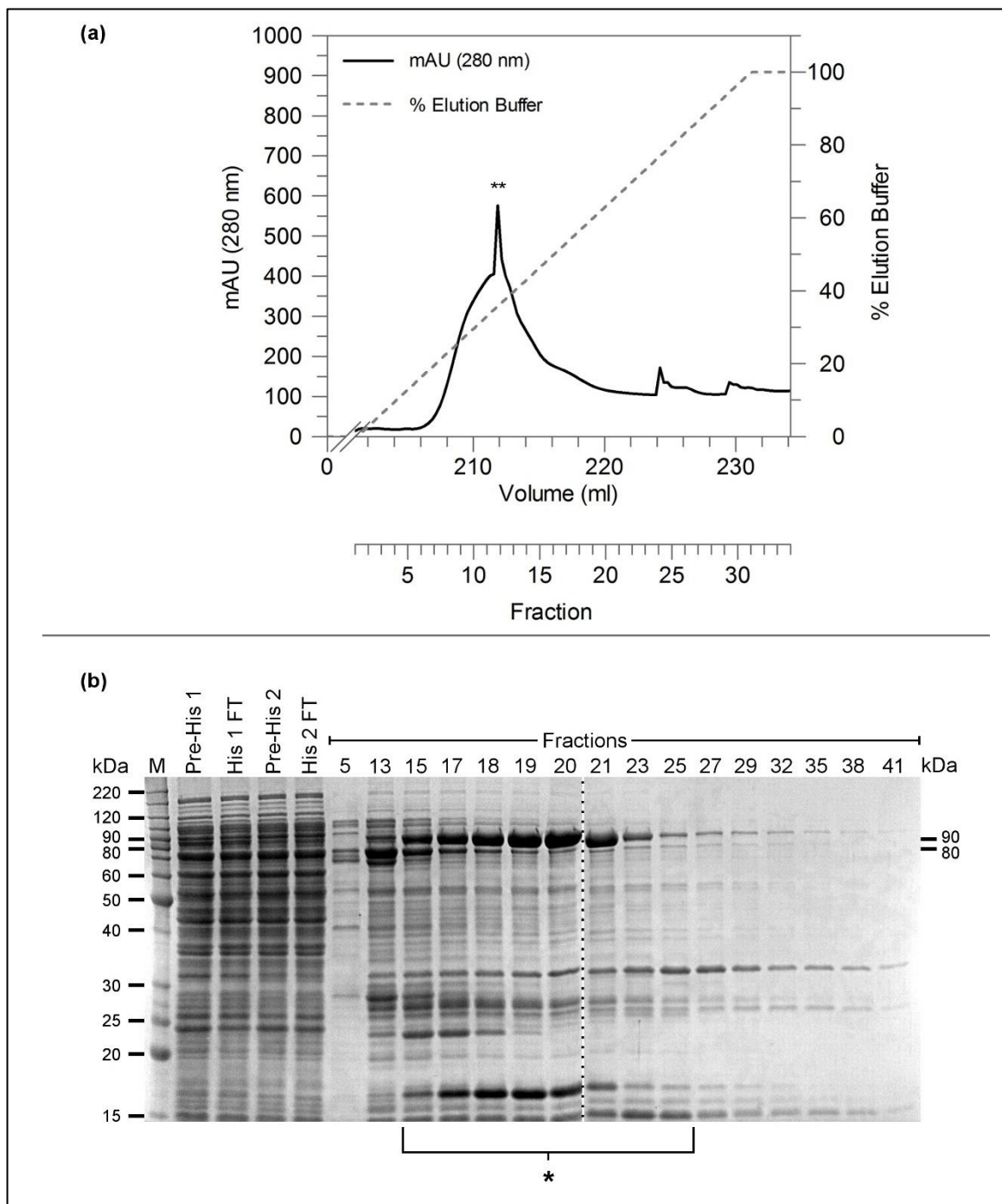


Figure 3.6 hPNPase IMAC

(a) IMAC chromatogram; the absorbance units measured at 280 nm ($A_{280\text{ nm}}$ (mAU)) over the volume (ml) is shown. The fraction number relating to the elution volume is depicted on a second x-axis. Sample loaded in lysis buffer (20 mM Tris-HCl pH 8, 300 mM NaCl & 10 mM imidazole) was eluted (black line) with a 30 ml gradient (grey dotted line, 0-100%) of elution buffer (lysis buffer with 500 mM imidazole). Note sharp peak at ~212 ml was due to a pressure spike detected by the AKTA system (labelled **), hence when compared to the chromatogram fraction numbers are slightly out. (b) 12% reducing SDS-PAGE analysis of pre-His (IMAC), His flow-through (FT) and His fractions alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. Fractions (15-25) pooled for further analysis are indicated by a label (*). After subsequent analysis (described later) the * symbol indicates the fractions (b) containing hPNPase.

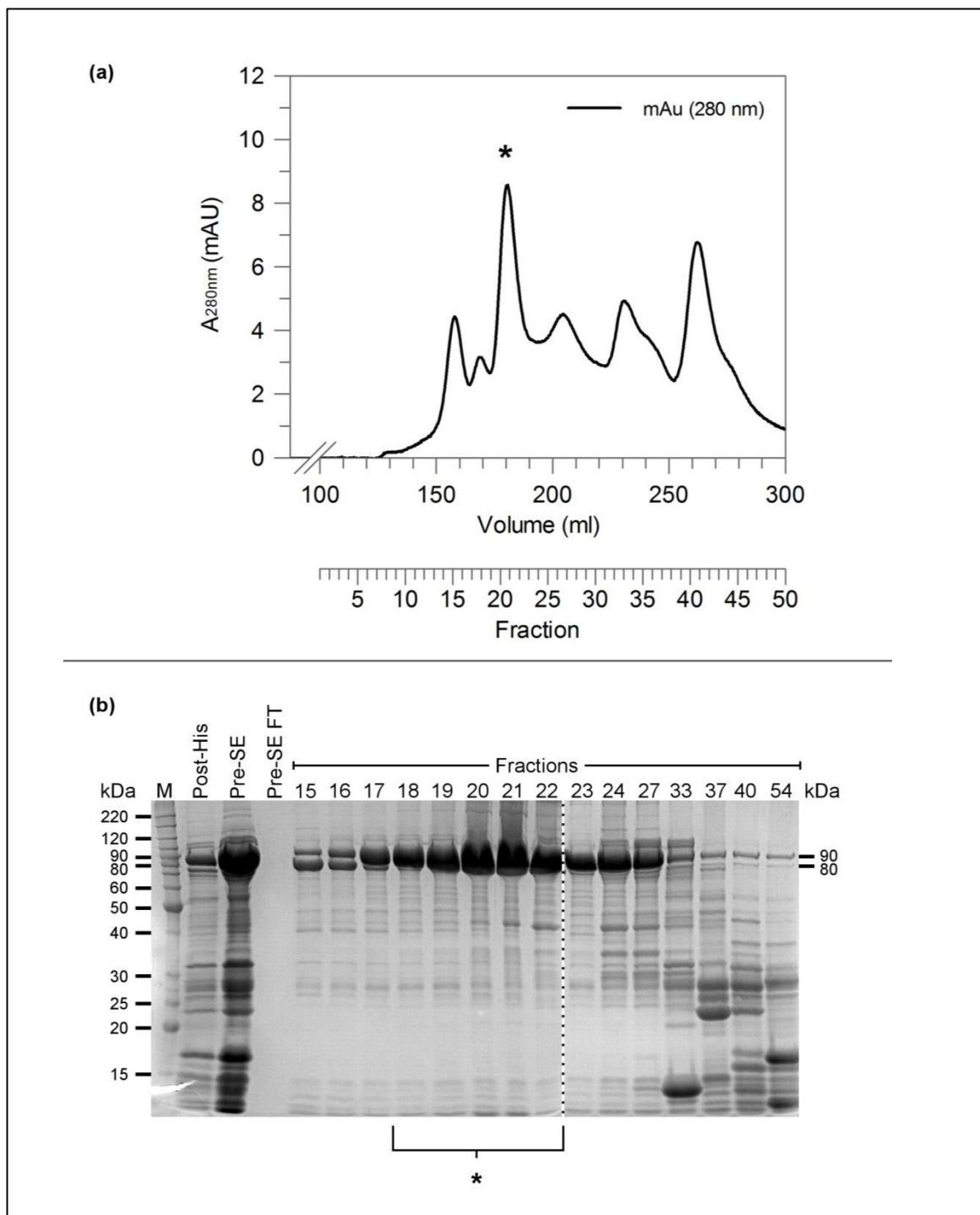


Figure 3.7 hPNPase SEC

(a) SEC chromatogram; the absorbance units measured at 280 nm ($A_{280\text{ nm}}$ (mAU)) over the volume (ml) is shown. The fraction number relating to the elution volume is depicted on a second x-axis. Sample loaded in SE Buffer (20 mM Tris-HCl pH 8, 150 mM NaCl & 5 mM MgCl_2) was eluted (black line) for SEC run. (b) 12% reducing SDS-PAGE analysis of post-His (pooled IMAC fractions 15-25), pre-SE and Pre-SE FT (protein concentrator sample and flow through respectively) with various SEC fractions from #15 – 54 alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. Fractions (18 - 22) used for further analysis are indicated by the bracket, note 1 ml of fractions were strata-cleaned, concentrating containing proteins to 10 μl and hence were heavily overloaded. After subsequent analysis (described later) the * symbol indicates the peak (b) and fraction (c) containing hPNPase.

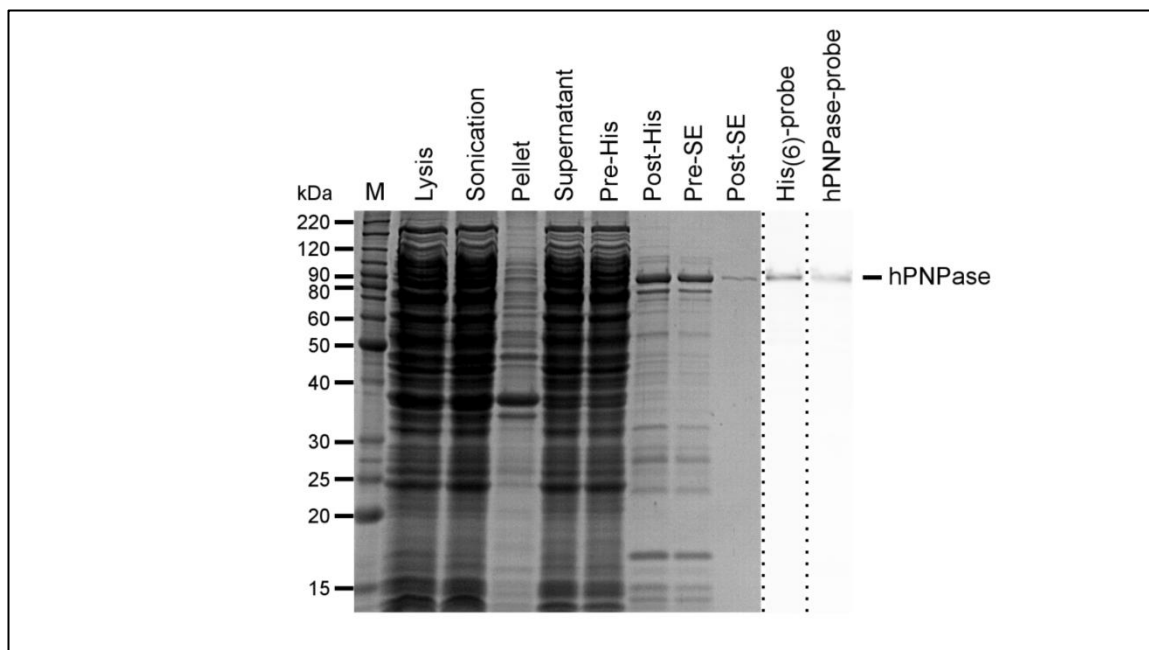


Figure 3.8 hPNPase Purification Stages

12% reducing SDS-PAGE analysis of hPNPase purification stages alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. Initial analysis suggests the band present in the final post concentration stage (0.3 μ g) may be hPNPase due to its migration at \sim 80-90 kDa, the size expected for the monomeric protein. Additional western blot analysis with a probe for hPNPase and H₆-tag and the presence of a band in both indicated the presence of H₆-hPNPase protein. This was confirmed to be hPNPase after subsequent denaturing mass spectrometry analysis (described later).

3.3.4.3 *Sulfolobus solfataricus* Exosome

The hetero-trimeric SsoExosome (H₆-Rrp4_41_42), was purified using a similar method to that conducted in (Lorentzen *et al.*, 2007), see methods for details (Section 3.2.6.3). The three individual subunits; Rrp4, Rrp41 and Rrp42 were cloned as a polycistronic gene, since co-expression has been reported to improve complex solubility (Lu *et al.*, 2010). SDS-PAGE analysis of the purification process is provided in Figure 3.11. Although for SsoExosome (H₆-Rrp4_41_42) ProtParam predicts monomeric sizes of 30.7, 27.6 and 30.2 kDa for subunits H₆-Rrp4, Rrp41 and Rrp42 respectively (Appendix 9.6). H₆-Rrp4 and Rrp41 both migrate at \sim 30 kDa with Rrp42 migrating higher near 35 kDa. Previous purification shows that the core exosome (H₆-Rrp41_42) individual subunits migrate at \sim 30 and 35 kDa for H₆Rrp41 and Rrp42 respectively (data not shown). Hence, when analysing the H₆-Rrp4_41_42 complex, reasonable justification can be made for the individual Rrp4 subunit, which is visible upon probing for the H₆-tag on a western blot, to migrate with Rrp41 (unlabelled). Additionally, this band is more intense than that of Rrp42, consisting of only one protein subunit migrating at that position. In summary, the presence of a band \sim 35 kDa and a more intense band \sim 30 kDa were monitored during purification.

As in previous purifications the early lysis, sonication and supernatant samples were heavily contaminated, providing difficulty in monitoring a specific protein band. Following purification via IMAC however, two bands can be observed in fractions 12-19 (Figure 3.9 (a), labelled with an asterisk) migrating between 30-35 kDa. As indicated in this figure, fractions 12-19 were retained for further purification by SEC. Following SEC, fractions 13-24 were pooled, concentrated and

stored at -80°C as standard. This was done following the observation of two bands between 30-4 kDa (Figure 3.10 (b), labelled with an asterisk) within these fractions which corresponded to the protein peak in mAU_{280 nm} recorded at ~172 ml (Figure 3.10 (a), labelled with an asterisk). Samples of these chromatography stages (pre- and post-HisTrap or SE respectively) are also provided in Figure 3.11. These lanes clearly show the removal of other contaminants until pure bands of protein migrating at ~30-35 kDa remains (0.3 µg). As described previously, the SEC calibration curve was utilised to predict elution volumes of 238, 241 and 238 ml for individual monomeric subunits H₆-Rrp4, Rrp41 and Rrp42 respectively and 173.67 ml for trimeric SsoExosome (H₆-Rrp4_41_42). Thus suggesting the pooled protein at ~172 ml (SE fractions 13-24) is more likely to be in a higher trimeric than monomeric form. Western blot analysis of this purified protein identified it to be a H₆-fusion protein when using H-tag antibody probe, Section 2.11.

The oligomeric state could be questioned with the presence of only two bands upon SDS-PAGE; it could be suggested that only the Rrp4 and Rrp42 proteins have been purified. Typically, native mass spectrometry (MS) can be employed to confirm correct oligomerisation; however, this was not possibly due to incompatibility problems with the MS buffer. Nevertheless, as both the Rrp41 and Rrp42 domains are required to form the catalytically active core structure (Lorentzen *et al.*, 2005), the 3'-5' degradation activity confirmation (Sections 3.3.6), supports the successful purification of H₆-Rrp4_41_42. Additionally, the presence of the Rrp4 domain is anticipated by the observed increased degradation activity, when compared to the core H₆-Rrp41_42 protein (data not shown).

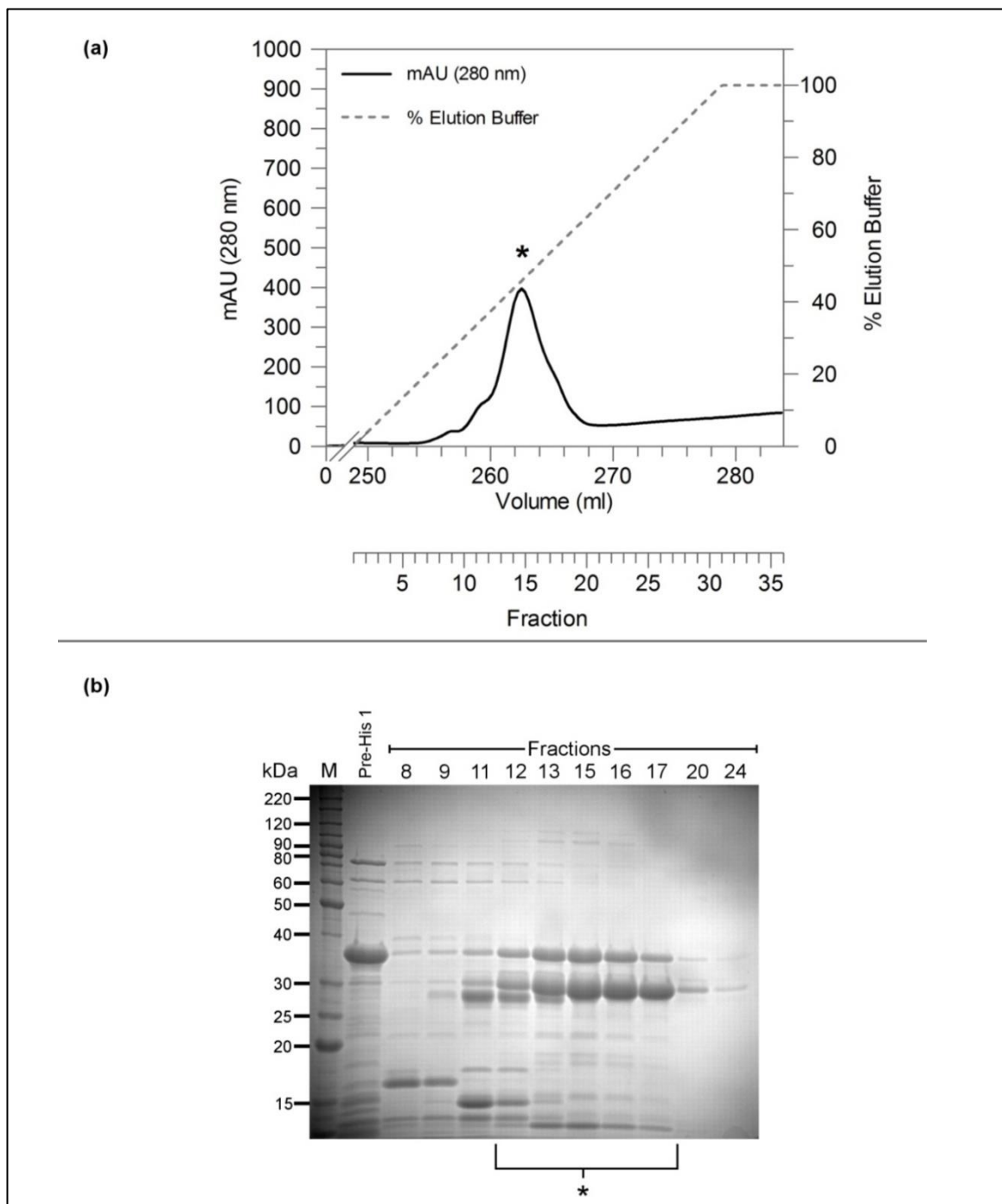


Figure 3.9 SsoExosome IMAC

(a) IMAC chromatogram; the absorbance units measured at 280 nm ($A_{280 \text{ nm}}$ (mAU)) over the volume (ml) is shown. The fraction number relating to the elution volume is depicted on a second x-axis. Sample loaded in lysis buffer (20 mM Tris-HCl pH 8, 300 mM NaCl & 10 mM imidazole) was eluted (black line) with a 30 ml gradient (grey dotted line, 0-100%) of elution buffer (lysis buffer with 500 mM imidazole). (b) 12% reducing SDS-PAGE analysis of pre-His (IMAC), His flow-through (FT) and His fractions alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. Fractions (12-19) pooled for further analysis are indicated by an asterisk. After subsequent analysis (described later) the * symbol indicates the fractions (b) containing SsoExosome.

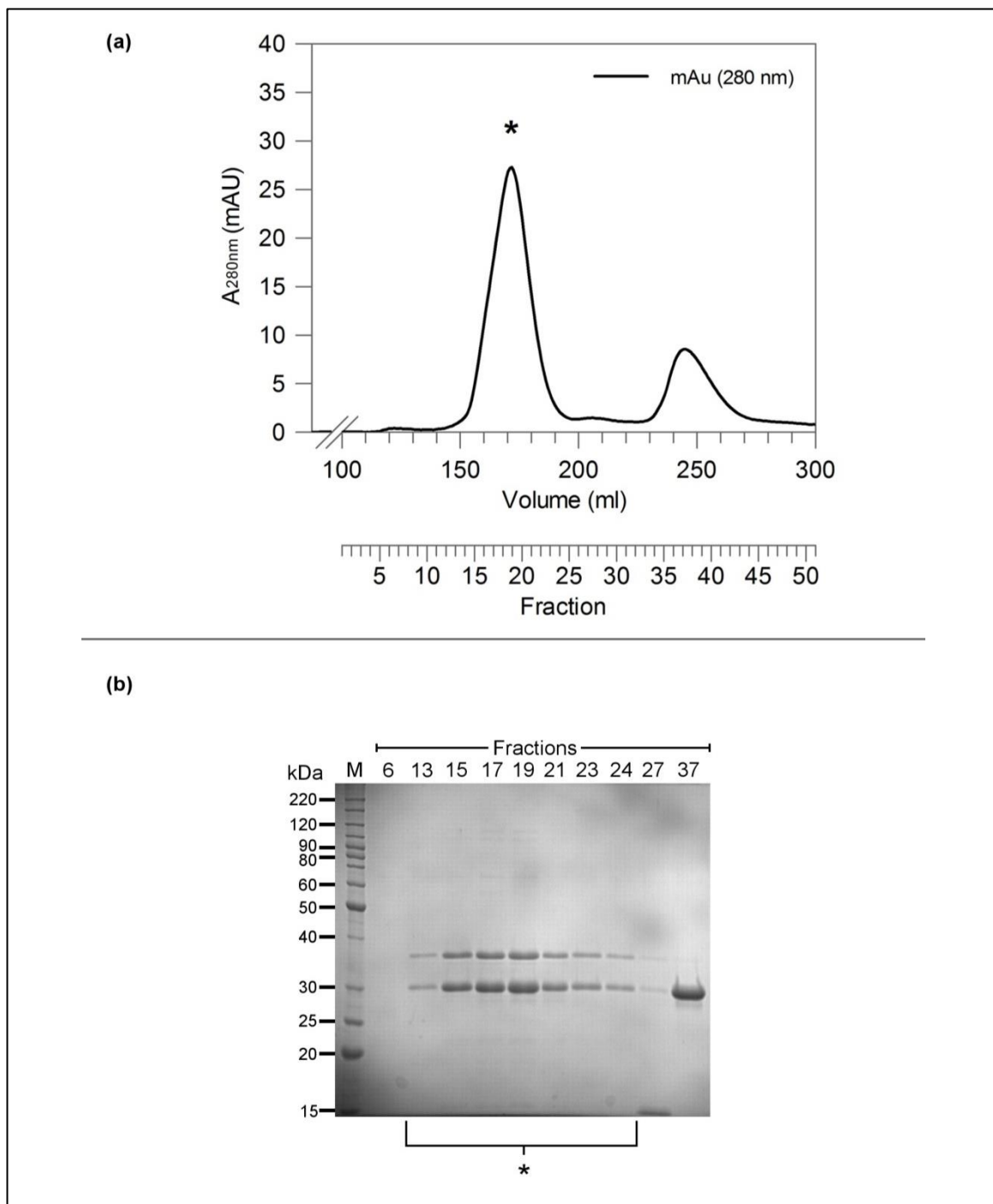


Figure 3.10 SsoExosome SEC

(a) SEC chromatogram; the absorbance units measured at 280 nm ($A_{280\text{ nm}}$ (mAU)) over the volume (ml) is shown. The fraction number relating to the elution volume is depicted on a second x-axis. Sample loaded in SE Buffer (20 mM Tris-HCl pH 8, 150 mM NaCl & 10 mM MgCl_2) was eluted (black line) for SEC run. (b) 12% reducing SDS-PAGE analysis of SEC fractions from #6 – 37 alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. The Fraction (13 - 24) used for further analysis is indicated by the bracket note protein samples were strata cleaned and hence were overloaded. After subsequent analysis (described later) the * symbol indicates the peak (b) and fraction (c) containing SsoExosome.

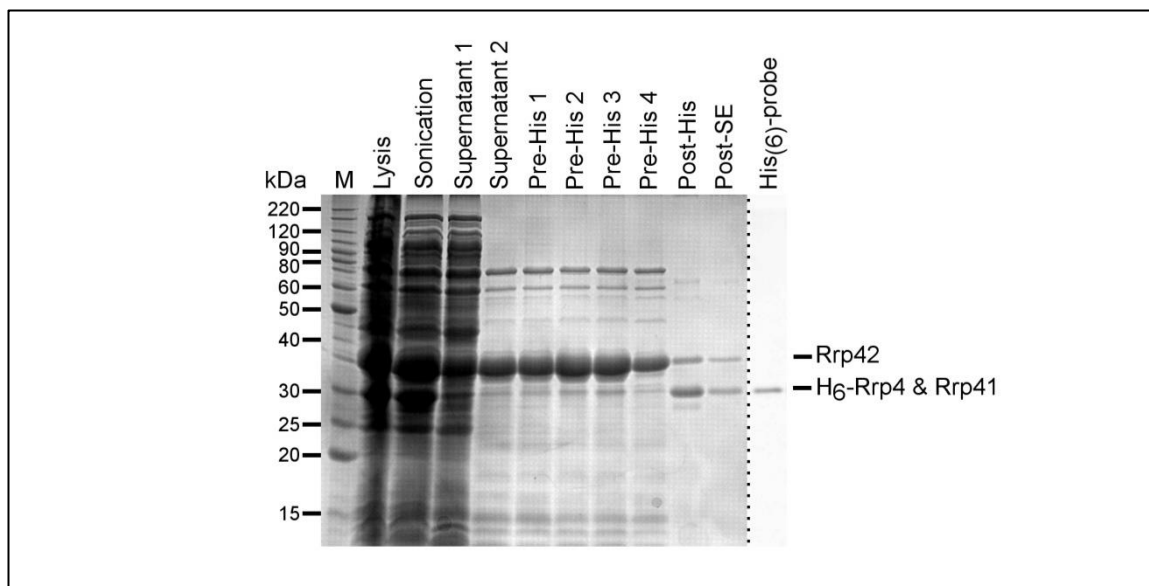


Figure 3.11 SsoExosome Purification Stages

12% reducing SDS-PAGE analysis of SsoExosome purification stages alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. Initial analysis suggests the band present in the final Post-SE stage (0.3 μ g) may be SsoExosome (H₆-Rrp4_41_42) due to its migration at ~30 and 35 kDa, the size expected for the monomeric protein subunits.

3.3.5 Protein Identification

The identity of each recombinant protein; EcPNPase, H₆-hPNPase and H₆-SsoExosome, was confirmed by multiple techniques.

3.3.5.1 SDS-PAGE and Western Blot

As indicated in each protein-specific results sub-section, upon SDS-PAGE analysis, protein corresponding to the anticipated size for EcPNPase, H₆-hPNPase and H₆-SsoExosome were observed at > 95% purity. In addition, the presence of a H₆-tag, upon western analysis with a primary H-tag probe antibody, supports the successful purification of H₆-tagged proteins, which are likely to be hPNPase and SsoExosome, although more validation was required to confirm these protein's identity. Accordingly, the presence of a protein band upon western analysis, with an antibody specific to hPNPase, strongly suggests hPNPase was effectively purified. A summary of recombinant protein purification analysis by SDS-PAGE is provided in Figure 3.12.

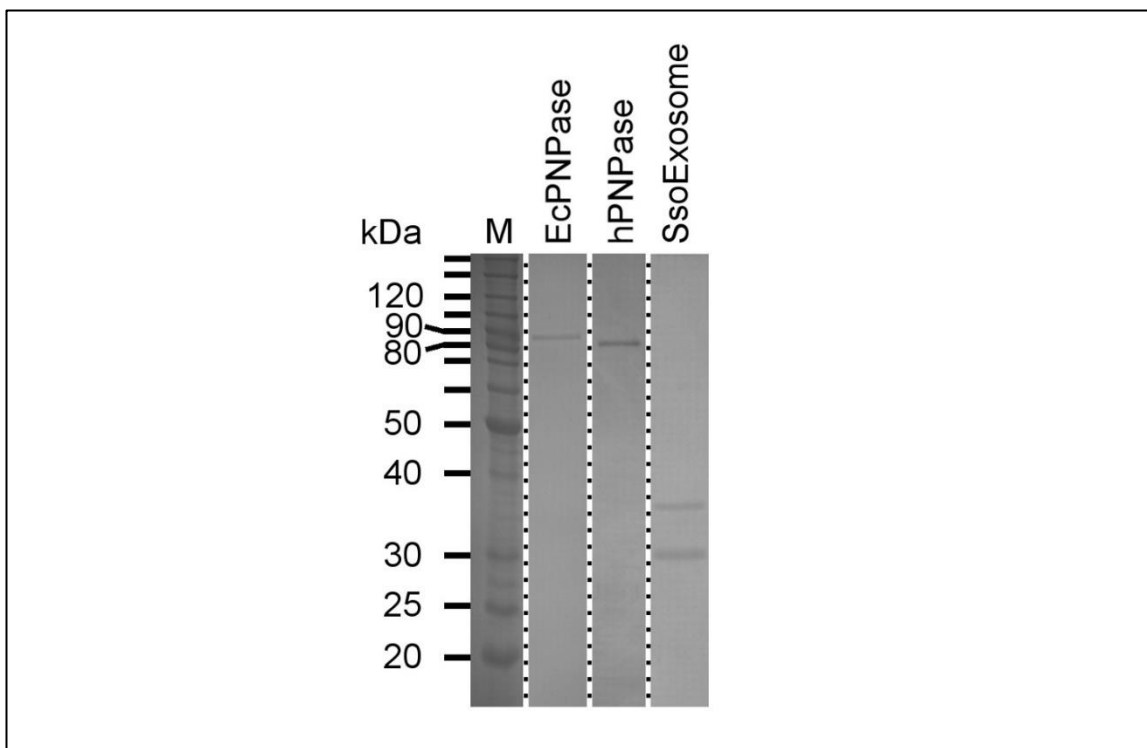


Figure 3.12 Protein Purification Summary

12% reducing SDS-PAGE analysis of EcPNPase, hPNPase and SsoExosome purification stages alongside the benchmark protein ladder (M). Note some ladder labels are omitted for clarity, see Appendix 9.2 for details. All recombinant proteins were purified to > 95% purity.

3.3.5.2 Denaturing Mass Spectrometry (LC-MS/MS)

The identity and monomeric weight of recombinant proteins EcPNPase and hPNPase was confirmed by denaturing LC-MS/MS, (Section 3.2.7.3). Analysis of results, provided by the Astbury Centre, Mass Spectrometry Facility (University of Leeds), confirmed their identity respectively. More specifically, using a peptide digest with trypsin, the percentage coverage for EcPNPase and hPNPase were 79% and 54% respectively; with good coverage over both the N-terminal and C-terminal regions.

The molecular weight of both enzymes were predicted by ProtParam to be ~ 77 and 83 kDa for EcPNPase and hPNPase respectively, however both enzymes were found to migrate between 80-90 kDa, when compared to a protein molecular weight ladder (Figure 3.12). The MW of hPNPase was determined by mass spectrometry to be 86 kDa and this aligns well to the enzymes migration in Figure 3.12. The MW of EcPNPase, determined by MS, was 77 kDa. Although this was lower than the protein migration in Figure 3.12, this was no surprise since EcPNPase is known to migrate slightly higher (~85 kDa) than expected upon SDS-PAGE analysis (Soreq, H., Littauer, 1977).

3.3.5.3 Non-Denaturing Mass Spectrometry

Non-denaturing MS was also performed at the Astbury centre for the SsoExosome (Rrp4_41_42) complex; however, results could not be obtained. It was suggested that the protein complex was not stable in the mass spectrometry buffer and would require further optimisation. As this was not the main focus of the study, and evidence had already been provided that the SsoExosome was intact,

as identified by SDS-PAGE analysis of single species from SEC, and formed an oligomeric structure of similar MW to the H₆-Rrp4_41_42 complex (as shown by SEC analysis), it was decided that the 3'-5' degradation activity would be examined for confirmation of identity.

3.3.6 3'-5' RNA Degradation Activity of Recombinant Proteins

The 3'-5' exoribonuclease degradation activity of each recombinant protein; EcPNPase, H₆-hPNPase and H₆-SsoExosome, was confirmed by a gel electrophoresis-based assay (Section 3.2.8). Data were normalised, when appropriate, to account for minor differences in enzyme activity between protein preparations. Although their activity varied slightly, as indicated by the different incubation periods, results from the end-point degradation assay revealed a 3'-5' processive degradation of the 5'FAM Poly(A)_{20mer} RNA substrate by each recombinant protein (Figure 3.13). As mentioned previously, assay conditions were not optimised for any protein specifically and this allowed the same protocol to be applied, to test quickly whether the enzymes purified exhibited the 3'-5' phosphorolytic activity expected. Notably, Chapters 4 and 6 provide details of the gel-based degradation assay optimisation and further enzyme characterisation respectively.

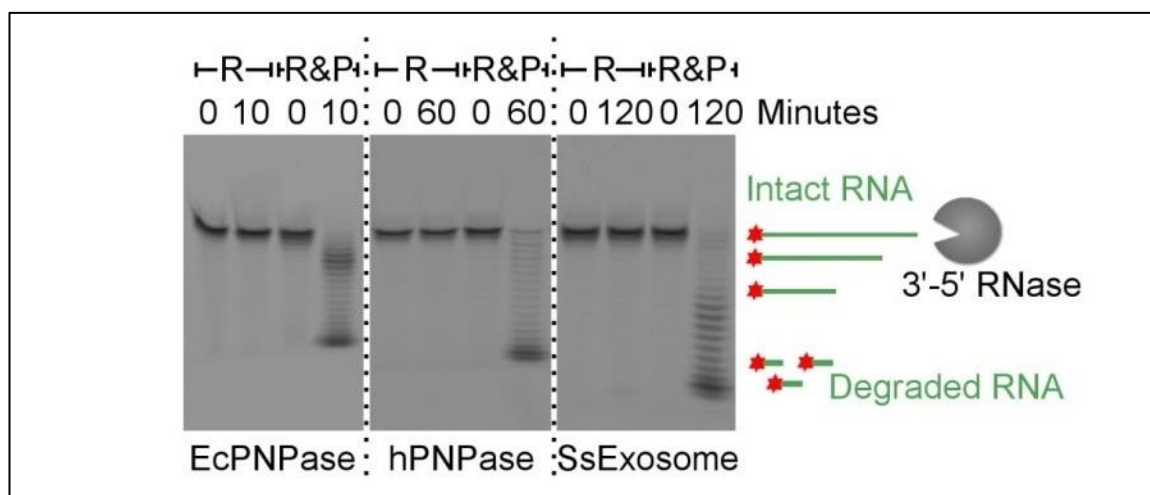


Figure 3.13 Recombinant Protein Degradation Activity Summary

20% urea denaturing PAGE loaded with 700 nM 5'FAM pol(A)_{20mer} RNA substrate (R) with or without 240 nM of relevant RNase (EcPNPase, hPNPase or SsoExosome) (R&P respectively), incubated at 37°C for various times (minutes indicated). Samples were separated in 1x TBE buffer at 200 V for 1 hour 30 minutes at 4°C. The gels were visualised using the Fujifilm FLA-imaging system (Phosphorimager: Blue filter, Blue laser, 473 nm).

3.4 Conclusions

This chapter describes the successful cloning, expression and purification of the recombinant EcPNPase, H₆-hPNPase and H₆-SsoExosome (H₆-Rrp4_41_42) proteins. As supported by the variety of experimental evidence provided in the results Section 3.3, the methods utilised (Section 3.2) to produce each of the recombinant proteins were successful. The majority of methods conducted were identical to those previously published in the literature. However, in some cases, these published methods were sub-optimal and modifications were applied, as justified in the relevant results section. This was conducted to yield enhanced purity.

Summary Figure 3.12 and Figure 3.13 show the final purified proteins and their 3'-5' exoribonuclease activity respectively. All recombinant enzymes were purified at > 95% purity with a concentration of ~1mg/ml per 2 L of LB media. This yield was not high; however, this may be a result of the enzymatic ribonuclease activity within the *E. coli* expression host. It is well established that PNPase knock-outs can affect growth and bulk RNA levels in *E. coli* (reviewed in Arraiano *et al.*, 2010), therefore overexpression of EcPNPase and its homologs, hPNPase and SsoExosome, may also have affected the host *E. coli* expression system. Although the purified protein was not prepared in high-yields, it was still sufficient for downstream biochemical testing as the subsequent experiments required very little protein. Most importantly, each protein was prepared at high enough purity for determining the effect of citrate on PNPase homologs' activity, which is described in the subsequent Chapter 4.

4 Inhibition of Homologous Phosphorolytic Ribonucleases by Citrate May Represent an Evolutionarily Conserved Communicative Link between RNA Degradation and Central Metabolism

Chapter 1 has already introduced the current literature and techniques available to study the connection between RNA regulation and central metabolism, underlining specific proteins which regulate RNA turnover in response to cellular and metabolic flux. Research which lays the foundations of this chapter's work was also covered. In summary, the results described by Nurmohamed *et al* 2011 clearly revealed a communicative link between PNPase and metabolism in *E. coli*. Whether this represents a mechanism that is conserved in other prokaryotes and/or higher organisms remains to be identified and was the focus of this chapter.

4.1 Chapter Aims

The previous Recombinant Protein Preparation Chapter 3 described the successful preparation of PNPase and archaeal exosome proteins for use in subsequent experimental chapters. Accordingly, this results chapter utilises these enzymes for inhibition studies with the metabolite citrate. Unfortunately, the X-ray crystallography of SsoExosome with citrate was not successful within this study. Henceforth this chapter describes the various *in silico* and *in vitro* techniques used to investigate whether this citrate-PNPase link represents a conserved mechanism across the three domains of life; utilising representative PNPase homologs from prokaryotes, eukaryotes and archaea. Key questions addressed within this work include, whether citrate interacting residues are conserved in PNPase homologs and thus is citrate binding predicted to be conserved? Detailed methods for protein sequence alignments and predicting citrate-PNPase interactions *in silico*, using molecular modelling and docking tools are provided in Section 4.2. Whether Mg²⁺ citrate can modulate the degradation activity of PNPase homologs is also examined and the methods for *in vitro* gel-based degradation assays are described in Section 4.2. The resulting data are provided in the following Section 4.3 and important conclusions regarding the conservation of a citrate-PNPase communication link are then summarised in Section 4.4.

4.2 Methods

The section herein describes the methods that were employed to study citrate-PNPase homolog interactions, including bioinformatics studies, *in silico* molecular docking and trials using X-ray crystallography for protein-ligand structure determination. The experimental techniques used to assess the effect of Mg²⁺ citrate on RNase activity *in vitro*, using gel-based degradation assays, are then described. Collectively the methods outlined within this chapter section were used as standard in subsequent chapters, unless otherwise specified.

4.2.1 Bioinformatics: Protein Sequence Alignments

Appropriate protein sequences, listed in Table 4.1, were selected and aligned using the default parameters of the Clustal Omega progressive alignment tool (McWilliam *et al.*, 2013). Multiple protein sequences alignments were then edited and annotated using Jalview software (Waterhouse, Procter, Martin, Clamp, & Barton, 2009).

Enzyme	Protein Accession Code
EcPNPase	WP_060707494.1
hPNPase	NP_149100.2
SsoExosome	WP_009991308 (Rrp41) WP_009991305.1 (Rrp42)
SspPNPase	WP_010871289
CbuPNPase	WP_042526278.1
CcrPNPase	WP_010917924.1
SanPNPase	GI: 75349253
MthExosome	WP_010876322.1 (Rrp41) WP_010876321.1 (Rrp42)
AfuExosome	WP_010878000.1 (Rrp41) WP_010878001.1 (Rrp42)
PabExosome	WP_010867734.1 (Rrp41) WP_010867735.1 (Rrp42)

Table 4.1 Protein Accession Codes

The protein accession code for each enzyme used for protein sequence alignments is listed.

A more detailed protein sequence alignment and consensus sequence was then produced. Protein sequences from prokaryotic, eukaryotic and archaeal PNPase homolog were extracted from the NCBI RefSeq database (Genbank) using polynucleotide phosphorylase and Rrp41/Rrp42 as search terms where appropriate (Pruitt, Tatusova, Klimke, & Maglott, 2009). Sequences were not extracted directly from BLAST as large numbers of S1 and KH domain homologs were incorrectly selected rather than the required PNPase homologs. The protein accession codes used for alignments are all provided in Appendix 9.8. The multiple protein sequence alignments, for each domain of life, were built using MAFFT vs.7. The G-INS-i iterative refinement strategy was selected, along with the default BLOSUM62 and Gap Penalty of 1.53 parameters. (Katoh & Standley, 2013). The alignments were then manually curated to remove duplicates and partial entries. To filter out redundancy, close homologues, with > 95% sequence identity, were omitted from the alignments using Jalview (Waterhouse *et al.*, 2009). The remaining aligned sequences were trimmed to the 'core' boundaries using the known *E. coli* sequence/structure as a guide (e.g. EcPNPase residues 1-549, hPNPase 1-601, SsoExosome 1-532). The consensus for the RNA Binding Regions 1 and 2, Phosphate Binding Region and Metal Binding Region (RBRI, RBRII, PBR and MBR) motifs were identified using these final set of sequences (3509 prokaryote sequences, 252 eukaryote sequences and 69 archaeal sequences) and visualised using Weblogo3, with the probability score depicted on the y-axis and residues on the x-axis (Schneider & Stephens, 1990).

4.2.2 *In-Silico* Molecular Docking

All the computational procedures, for each of the PNPase homologs tested, were carried out within the molecular docking program MOE (Molecular Operating Environment, 2013), as summarised in the flow diagram in Figure 4.1. Details of individual steps are described in Sections 4.2.2.1 to 4.2.2.3

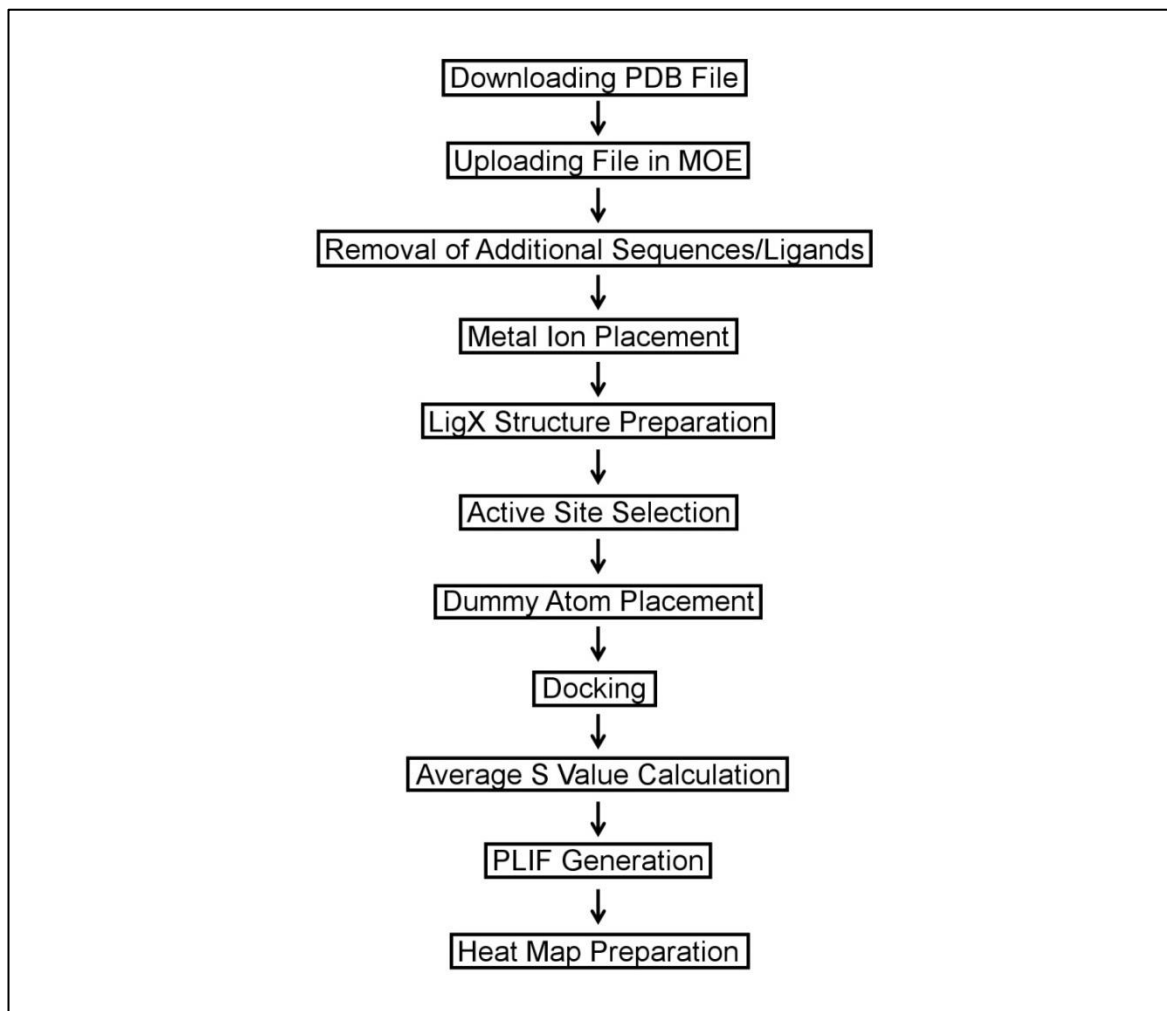


Figure 4.1 MOE Flow Diagram

Stages of *in silico* MOE docking (Molecular Operating Environment, 2013), from downloading PDB files to preparing heat maps, are illustrated as a flow diagram. Illustration created using GIMP (v2) (GIMP, n.d.).

4.2.2.1 Structure Preparation

Appropriate X-ray crystal protein structures for EcPNPase, SanPNPase, hPNPase, SsoExosome were selected (PDB entry: 3GCM, 1E3P, 3U1K and 4BA1 respectively), downloaded from the PDB as text files and opened as biomolecules using MOE. Input structures were checked using MOE for additional information within the pdb.txt files, such as extra/non-standard amino acids residues/chains, inhibitors and cofactors, and removal of unwanted information was conducted where necessary. Individual structures were submitted to a preparation step using the MOE LigX interface; assignment of the correct ionization states, geometry optimization, protonation, and energy minimization of the protein receptors was conducted. For energy minimization and docking in MOE the AMBER12: EHT force field was used, this is recognized to be specifically

parameterised for both proteins and small molecules (Scholz, Knorr, Hamacher, & Schmidt, 2015; Wang, Wolf, Caldwell, Kollman, & Case, 2004).

Metal ions Mg^{2+} and/or Mn^{2+} , essential for catalytic activity, were only present in some of the PNPase/Exosome crystal structures available on the PDB. Hence EcPNPase, with Mg^{2+} located in its active site, was used as a model for placing Mg^{2+} in the active sites of those PNPase homologs structures lacking this essential ion, as described previously for SsoExosome (Lorentzen & Conti, 2012). Accordingly, the hPNPase, SanPNPase and SsoExosome enzymes, lacking the presence of an Mg^{2+} co-factor, were aligned against EcPNPase using the MOE superpose panel and the EcPNPase Mg^{2+} cofactor was defined as part of the receptor for the other enzymes. Residues known to coordinate Mg^{2+} within *E. coli* PNPase were compared to the PNPase homologs to see if the corresponding residues, in these enzymes, also coordinated Mg^{2+} ; the type of interaction and residues involved were checked using the MOE built-in interaction map tool. This ensured that the newly placed metal ion was correctly coordinated within the active site; confirming that essential contacts and positioning were maintained in this new structure.

Once protein structures were prepared, the MOE Alpha Site Finder panel was used to predict putative binding sites from the three-dimensional structure, using a geometric approach. Knowledge of active site residues essential for catalytic activity, determined from previous publications, was then applied to select the appropriate pocket concavity which represented the protein's active site. More specifically, if multiple residues of the PBR and MBR consensus motifs (Figure 1.11) were observed within the Alpha site finder panel, for a particular active site hit, this hit was selected. Once selected, the active sites were defined via the placement of dummy atoms at the locations of hydrophobic and hydrophilic α -spheres; these spheres represented locations of tight packing (Galli *et al.*, 2014).

4.2.2.2 *In-Silico* Citrate Docking

The structure of the citrate ion metabolite was obtained from the ZINC database (ZINC00895081); downloaded and saved as a MOL2 file in a biologically relevant conformation (Irwin, Sterling, Mysinger, Bolstad, & Coleman, 2012). Using the appropriate active site pocket defined previously, docking calculations were carried out using the docking module of MOE. The docking parameters were selected as recommended for protein-small molecule interactions, including the Amber12: EHT force field which suggests ideal parameters for proteins and nucleic acids (Amber ff12) and for small molecules (Extended Hueckel Theory) respectively. Enzymes were set as fixed 'receptors' and the 'ligand', in this case the citrate ion, was docked into the active site at the defined dummy atoms, using the MOE default triangle matcher placement tool and 300 placement poses (iterations). This triangle matcher placement methodology is recognised to be the best method for standard and well-defined binding sites and it generates poses by superposing triplets of ligand atoms and triplets of receptor site points (dummy atoms/ α -spheres centres) (Galli *et al.*, 2014). Thirty unique receptor-ligand complexes were generated, documented within a database

viewer (DBV) and saved as a MOE database (MDB). The accepted ligand poses were scored according to the MOE default London dG scoring function, which estimates the binding free energy of the ligand from a given pose (Galli *et al.*, 2014). This scoring function contains terms for the receptor-ligand interaction, including the loss or gain of translational and rotation entropy, energy related to fixing rotatable bonds, hydrogen bond formation, metal ligation and a desolvation penalty for removing the ligand from aqueous solution. The more negative the value of this London dG score, the higher the affinity of this ligand-receptor complex is predicted to be (Molecular Operating Environment, 2013). The 300 docking poses were then subject to a forcefield refinement, with no second rescoring, and 30 unique poses for each compound were retained. The docking score (S) for these 30 unique receptor ligand complexes were calculated using the Generalized Born solvation model (GB/VI). This calculated the non-bonded interaction energy (van der Waals, Coulomb and GB implicit solvent interaction energies) between the receptor and the ligand complex whilst excluding the self-energies of the individual receptor and ligand atoms. The default rescoring 2 GBVI/WSA dG was not employed; this setting was adjusted to none, as recommended by MOE for docking databases (Molecular Operating Environment, 2013). The lowest-energy scoring poses for each ligand (citrate ion) were selected and the MOE binding score (S) values (kcal/mol) were recorded for each enzyme. The lowest-energy pose with a single citrate ion docked was selected and used as the starting structure to dock a second citrate ion.

4.2.2.3 Protein Ligand Interaction Fingerprints (PLIF) and Heat Map Generation

The top 10 lowest-energy scoring poses, with both one and two citrate ions docked, were then selected to generate a representation of the three dimensional molecular interactions between the docked ligands and protein. This was generated for each enzyme using the MOE Protein Ligand Interaction Fingerprints (PLIF) panel (Singh, Deng, Narale, & Chuaqui, 2006). When implemented in MOE, PLIF classified interactions between the receptor binding site residues and the ligand, in this case citrate, are sorted into various types of weak and strong interactions using the interaction assessment function built into the program. There are some variations, but typically an Interaction Fingerprint (IF) describes the receptor-ligand intermolecular interactions as a bit string. This bit string is defined by the following parameters; whether residues contact the ligand (i), if main chain atoms and side chain atoms are involved in binding (ii and iii respectively). The type of interaction involved is considered, whether this is a polar interaction (iv), non-polar (v), hydrogen bond acceptor(s) (vi) and hydrogen bond donors (vii) (Sato, Honma, & Yokoyama, 2010). The data described above generated from the MOE PLIF was exported as bit scores into Microsoft Excel (2010). Finally, the frequency with which an interaction was observed between citrate and a particular amino acid in the docking calculation was determined from the PLIF bit scores. The resulting percentages for docking both one and two citrate ions were presented as heat maps using the online Plotly graph software (Plotly Technologies Inc., 2015).

4.2.3 X-Ray Crystallography

A Pre-Crystallization Test (PCT™, HAMPTON Research) was initially conducted to select the appropriate protein concentration for crystallisation screens, as described in the user manual. As recommended, the A1/A2 PCT test was conducted first; 0.5 ml of reagents A1 (0.1 M Tris HCl pH 8.5, 2.0 M (NH₄)₂SO₄) and A2 (0.1 M Tris HCl pH 8.5, 2.0 M MgCl₂, 30% (w/v) Polyethylene glycol 4,000) were pipetted into the A1 and A2 reservoir wells of a VDX Plate respectively. To the centre of a single glass cover slide 1 µl of protein sample (17.5 µM SsoExosome and 20 mM sodium citrate) and 1 µl of A1 reservoir solution was added, without mixing. The slide was then inverted over the A1 reservoir and sealed, and the process was repeated for reservoir A2. After 30 minutes the drops were viewed using a light microscope with a magnification between 20 and 100x. The results were compared to Figure 4.2 and depending on the results; the process used to test reagents A1/A2, was repeated with reagents B1 (0.1 M Tris HCl pH 8.5, 1.0 M (NH₄)₂SO₄) and B2 (0.1 M Tris HCl pH 8.5, 2.0 M MgCl₂, 15% (w/v) Polyethylene glycol 4,000) as necessary. Once the appropriate protein concentration was identified in the PCT test, a crystal screen was conducted.

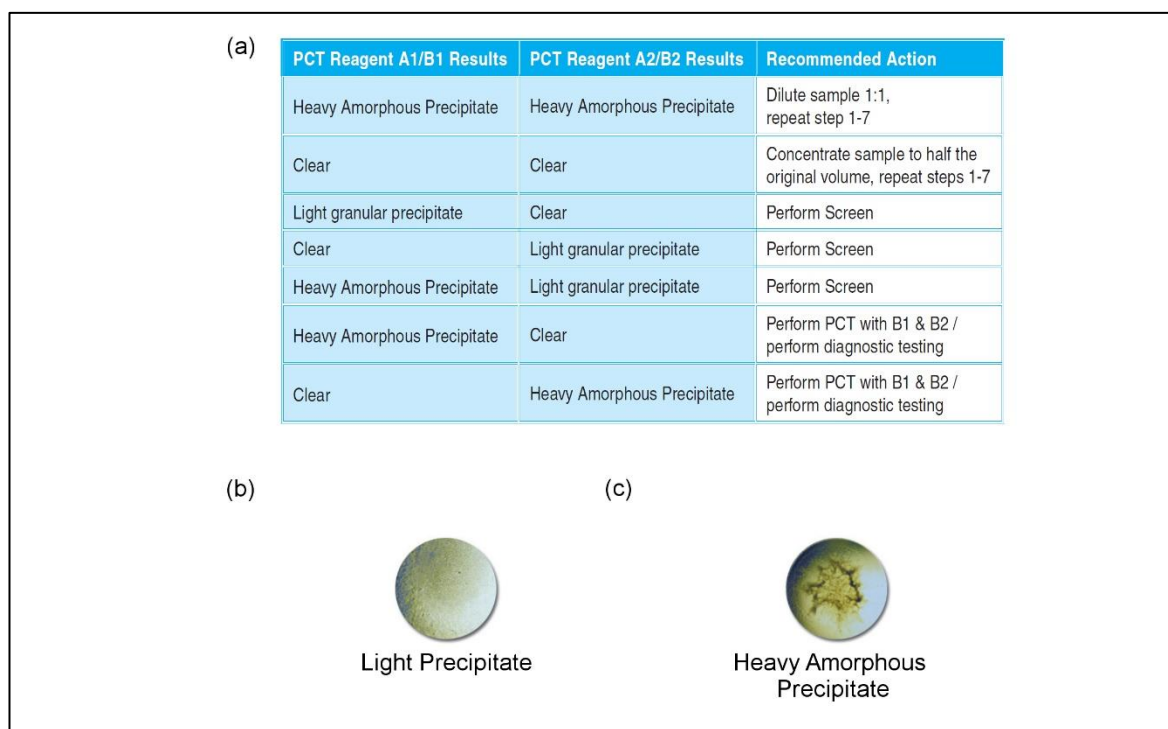


Figure 4.2 PCT Recommended Actions

The recommended steps to take during the PCT Pre-Crystallization test are outline in this figure, which was taken directly from the PCT™, HAMPTON Research user manual. (a) Depending on the precipitate observed; either (b) light or (c) heavy amorphous, the recommended actions following the A1/A2 and the B1/B2 test are provided.

Following the Pre-Crystallization test a PACT premier™ screen (Molecular Dimensions) was set up (Newman *et al.*, 2005), the screen conditions are provided in Appendix 9.7. The screen was conducted with 17.5 µM SsoExosome and 20 mM sodium citrate (previously determined optimal concentrations in the PCT test), using a honeybee robot (Genomic solutions) at 22°C and 70% humidity. A hit was discovered for well D7 (0.2 M NaCl, 0.1 M Tris pH 8 & 20% peg 6,000) after

2 weeks at 16°C. Individual crystals were grown for a further 6 months at 16°C prior to being cryo-protected in 20% glycerol and cryo-cooled in liquid nitrogen. Crystals were analysed at Diamond Light Source (remote access) for protein diffraction and data collection and this work was done in collaboration with Prof. John McGeehan (University of Portsmouth).

4.2.4 Gel-Based *In Vitro* RNA Degradation Assays

Gel-based electrophoresis degradation assays were used to analyse the 3'-5' phosphorolytic exoribonuclease activities of recombinant EcPNPase, hPNPase, SspPNPase and SsoExosome proteins canonical activity and in the presence of citrate. 700 nM of 5' fluorescein amidite (FAM) labelled Poly(A)_{20mer} RNA substrate (Dharmacon) was incubated at 37°C with 240 nM of each protein, in a buffer containing 15 mM Tris pH 8, 112.5 mM NaCl, 3.75 mM MgCl₂, 0.045 mM Na₂PO₄ (+/- 3.75 mM citrate). Control experiments were carried out in the absence of the relevant recombinant protein, termed RNA only (+/- 3.75 mM citrate). The rate of substrate degradation varied for each protein; hence reactions were quenched with 0.1 M EDTA at various times, details of which are provided in the results figure legends respectively. Degradation products of the 5'FAM Poly(A)_{20mer} RNA substrate were separated on a 20% denaturing urea gel as described in Chapter 2, Section 2.10.2.2. Gels were visualised using a Fujifilm FLA-5000 phosphorimager, quantified by densitometry using the ImageJ open-source Java-based imaging program (Abràmoff, Magalhães, & Ram, 2004) and analysed using GraFit software (Leatherbarrow, 2009).

For consistency all reactions were conducted using identical conditions, although these were sometimes sub-optimal for a particular enzyme. Specifically, the RNA substrate was kept consistent; a 5'FAM Poly(A)_{20mer} was used, since PNPases from both prokaryotes and eukaryotes and the archaeal exosome have all been shown to process 3' poly(A) tails. The phosphate concentration was particularly low compared to EcPNPase optimum (10-20 mM), since hPNPase required less (0.1 mM) for activity and can be inhibited by high concentrations (Portnoy, Palnizky, Yehudai-Resheff, Glaser, & Schuster, 2007). Additionally, although the SsoExosome can function at temperatures of ~65°C (data not shown), hPNPase and EcPNPase are not tolerant of these extreme conditions and so assays were all conducted at 37°C. In summary, in order to determine if citrate affects PNPase homolog activity *in vitro*, sub-optimal conditions were selected, with the main emphasis being on producing comparable data.

4.3 Results

This section firstly investigates the citrate-PNPase interactions previously documented in the literature in further detail, before the results of *in silico* protein sequence alignments are provided, which investigates the conservation of these interactions (Sections 4.3.1-4.3.2 respectively). Results from *in silico* molecular docking are subsequently provided in Section 4.3.3, which initially use available EcPNPase and hPNPase structures with citrate bound for validation purposes, prior to predicting citrate binding for SanPNPase and SsoExosome homologs. Following predictive tools,

the results of a trial of SsoExosome co-crystallisation with citrate are then presented in Section 4.3.4. Subsequent results from *in vitro* experiments, investigating the effect of citrate on PNPase homologs' 3'-5' phosphorolytic degradation activity, are then provided in Section 4.3.5. Finally, a more detailed protein sequence bioinformatic analysis is provided in Section 4.3.6 and this focuses on the conservation of citrate-binding residues more widely across evolution.

4.3.1 Investigating Citrate-PNPase Interactions

4.3.1.1 *E. coli* and *H. sapiens* Structures

The first hint that citrate might regulate PNPase came from structural studies of the PNPase PH-hexamer core from *E. coli* (Nurmohamed, Vaidialingam, Callaghan, & Luisi, 2009). As mentioned in Chapter 1, citrate had been present in the crystallisation buffer and the resulting structure revealed up to four molecules of citrate bound per PNPase monomer (Nurmohamed *et al.*, 2009). This original paper, describing citrate-PNPase interactions in *E. coli*, suggested that the citrate molecules occupy the proposed location of catalytic intermediates at the catalytic site. They suggested that the positions of these molecules would be expected to prevent the formation of the Michaelis complex and provide a potential means of enzyme inhibition. Indeed, inhibitory effects were seen *in vitro* for citrate binding within this active site. Interestingly, citrate binding at the vestigial site was also shown to modulate PNPase activity, most likely through an allosteric mechanism (Nurmohamed *et al.*, 2011). The exact type of interactions between citrate and PNPase residues were not discussed at length in Nurmohamed *et al.*, 2011, hence more detailed information is provided within this section. Understanding these interactions was important before comparisons could be made with other PNPase homologs.

Using MOE, an interaction map can be generated which determines the residues and type of interaction involved in ligand binding. Simply, the citrate molecules co-crystallised within PNPase were selected in turn and using the Ligand Interaction Panel of MOE, the residues and type of interaction were computed. Specifically, the interactions of each active site citrate (two citrates per monomer) and vestigial site citrates (two citrates in one monomer) of the EcPNPase crystal structure (3GCM) were studied (Figure 4.3). The citrate molecule closest to the Mg²⁺ ion (pdb: Cit 552, termed citrate 1 herein; Figure 4.3 (a)) was suggested by MOE to contact the metal ion. In one monomer, citrate 1 also interacted with G436, S438, S439 residues in the PBR via backbone acceptor interactions (blue arrows) and the K494 residue in the MBR via a sidechain acceptor (green arrow) as shown in Figure 4.3 (b). The second citrate molecule, furthest from the Mg²⁺ ion (pdb: Cit 551, termed citrate 2 herein; Figure 4.3 (a)), mimicked the position of the scissile phosphate in an RNA substrate (Nurmohamed *et al.*, 2011). In one monomer, citrate 2 was suggested to interact with H403 via a sidechain acceptor (green arrow). It also contacted residue R93 of RBRI and residues R399 and K494 of RBRII, with residues acting as both an ion contacts (purple line) and side chain acceptors (green arrow) (Figure 4.3 (c)).

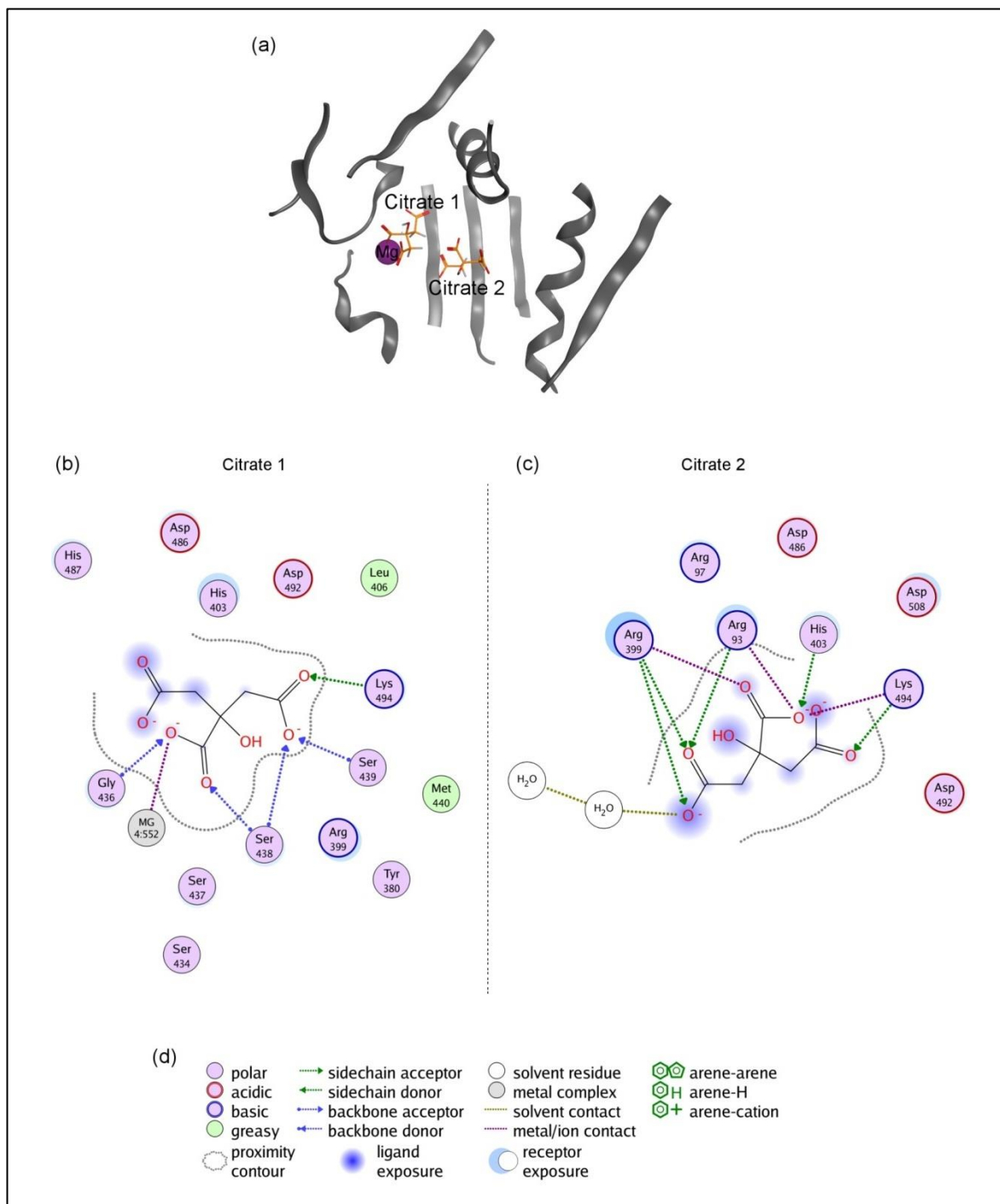


Figure 4.3 EcpNPase Active Site Citrate Interactions

(a) EcpNPase 3GCM (grey ribbons) with co-crystallised ligands including Mg^{2+} (purple sphere), citrate 1 and 2 within the active site (orange sticks). MOE ligand interaction map for (b) citrate 1, (c) citrate 2 and (d) the legend. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

The interactions shown in Figure 4.3 highlight those within the active site of one monomer, which was later selected for docking. However, upon examination of the active sites in the other two monomers, subtle differences in ligand interactions were observed in MOE. This may have been due to using a structure generated by X-ray crystallography, which is effectively a snapshot image of the fixed protein. In order to generate a summary of interactions which showed the occurrence rate of citrate binding residues in all three monomers, a Protein Ligand Interaction Fragment (PLIF) was generated (Figure 4.4). The PLIF output, shown in Figure 4.4, clearly illustrated the

slight differences in citrate-PNPase interactions for each monomer. For example, interactions of residues R399 and H487 with citrate 1 were completely absent in one monomer (monomer 2 and 3 respectively) and S483 made one more contact in monomer 3 than it did in monomers 1 and 2 (Figure 4.4 (a)). The types of interactions made for citrate 2 were the most similar, the only residue which had a different number of interactions was R399. Only one type of contact was seen in monomer 1, three types were made in monomer 3 and four contacts were made in monomer 2.

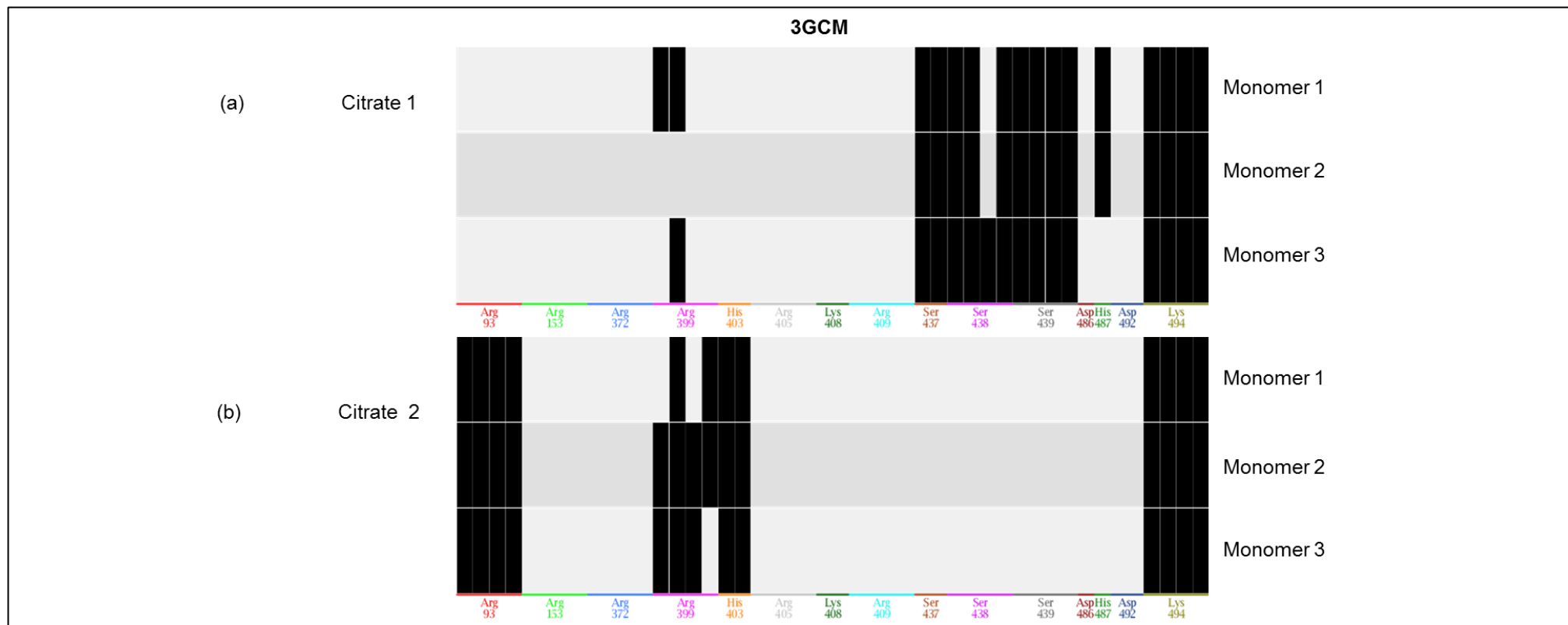


Figure 4.4 EcPNPase Active Site Citrate PLIF

MOE PLIF for (a) citrate 1 and (b) citrate 2, highlighting individual citrate binding residues for the active site of each monomer (1-3) of the homotrimer EcPNPase 3GCM structure. Black bars indicate an individual contact between the citrate (termed the ligand) and the 3GCM active site (termed the receptor). Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

From the PLIF interaction maps presented in Figure 4.4, a summary of the citrate-PNPase interactions for EcPNPase, as a whole protein, were deduced. Eight residues were described to interact with citrate including; S437, S438, S439, H487, K494, R93, R399 and H403.

The interactions of the vestigial site citrates were also investigated. The first citrate molecule at the vestigial site, closest to active site citrates (VCitrate1) (Figure 4.5 (a)) was suggested to interact with R153 and R409 via sidechain acceptors (Figure 4.5 (b)). The vestigial citrate 2 (VCitrate2) contacted R372 and K408 via ion/metal contacts, and residues R405 and R409 via sidechain acceptor contacts (Figure 4.5 (c)). Since only one monomer contained citrates at the vestigial site, the PLIF shown in Figure 4.5 (e) directly represented the interaction maps shown in Figure 4.5 (b-c).

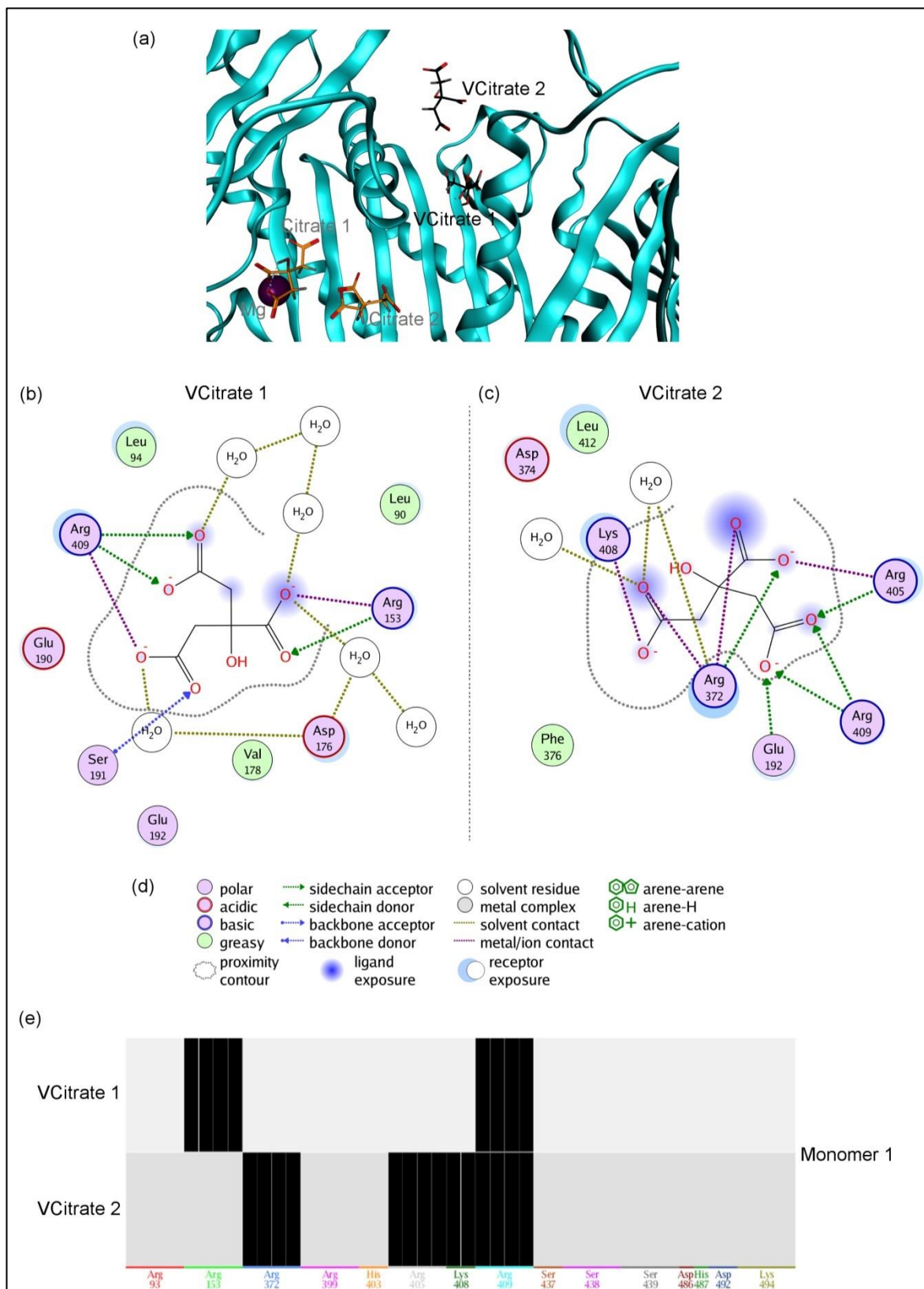


Figure 4.5 EcpNPase Vestigial Site Citrate Interactions

(a) EcpNPase 3GCM (grey ribbons) with co-crystallised ligands including Mg^{2+} (purple sphere), citrate 1 and 2 (Citrate, orange sticks) within the active site and citrate 1 and 2 of the vestigial site (VCitrate, black sticks). MOE ligand interaction map for (b) VCitrate 1, (c) VCitrate 2, (d) the legend is shown. (e) MOE PLIF for vestigial citrate 1 and 2 is also provided. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

As mentioned in Chapter 1, when the crystal structure of the PNPase core from *H. sapiens* (3U1K) was reported, it too had crystallised in a buffer containing citrate. In this structure, only two

molecules of citrate were observed per PNPase monomer and both were at the active site, there was no evidence of citrate binding to a vestigial site (Lin *et al.*, 2012). As conducted for EcPNPase, an interaction map was generated in MOE to investigate citrate-hPNPase interactions (Figure 4.6). The hPNPase 3U1K pdb file contained only one unique chain, hence when the file was opened in MOE, biomolecule symmetry was applied in order to create the homotrimer. As a result, each chain and thus the active site citrate-binding interactions were identical for each monomer. This meant that the PLIF generated previously wasn't required; the residues interacting with citrate could be determined directly from the interaction maps in Figure 4.6. Accordingly, the residues required for citrate 1 interactions were examined, residues S482, S483 and S484 of the PBR interacted via a backbone acceptor (blue arrow), and K546 of the MBR bound citrate through a metal/ion contact (purple line) (Figure 4.6 (b)). Those residues involved in Citrate 2 interactions included R132 of the RBRI via a sidechain acceptor, R446 through a metal/ion contact, and H450 of the RBRII and K464 of the MBR via sidechain acceptors (Figure 4.6 (c)).

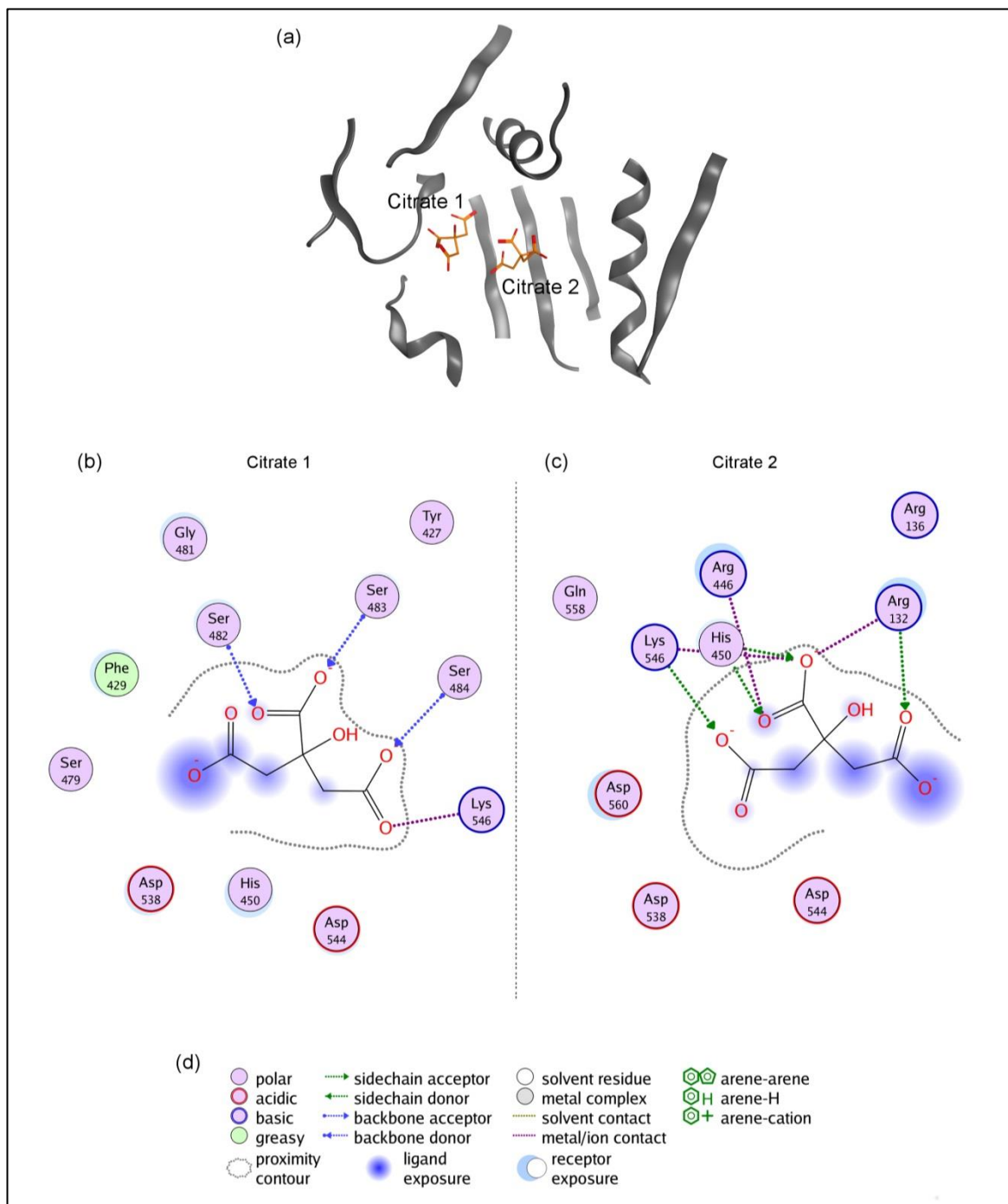


Figure 4.6 hPNPase Active Site Citrate Interactions

(a) hPNPase 3U1K (grey ribbons) with co-crystallised ligands including citrate 1 and 2 (orange sticks) within the active site. MOE ligand interaction map for (b) citrate 1 and (c) citrate 2 and (d) the legend. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

Comparison of the citrate-EcPNPase and citrate-hPNPase interactions shown in Figure 4.3 and Figure 4.6 respectively, revealed that of the 8 citrate-binding residues identified in the EcPNPase active site (S437, S438, S439, H487, K494, R93, R399 and H403), 7 are observed to bind citrate in hPNPase. Interestingly, the interaction of H487 with citrate 1 in EcPNPase was not present in hPNPase and this residue has been substituted for a tyrosine. The PLIF in Figure 4.4 suggests that H487 in EcPNPase doesn't make a contact with citrate in one of its monomers, potentially suggesting this is not an essential interaction. Whether this histidine to tyrosine mutation will affect the ability of citrate to bind and regulate hPNPase activity is unknown. Lin *et al.*, did not go on to

test the effect of citrate on hPNPase activity and so it remains to be determined whether hPNPase is also regulated by this metabolite. However, based on the conservation of citrate-PNPase interactions it would be predicted that citrate would inhibit hPNPase similar to EcPNPase.

4.3.1.2 Structural Conservation of PNPase Homologs

After establishing exactly which residues interact with citrate in both EcPNPase and hPNPase the structural conservation of PNPase homologs was investigated. This allowed regions of low and high structural conservation to be compared. Comparing PNPase homologs within MOE was possible due to the superimpose panel; using this tool, protein structures can be superimposed based on the whole protein (in this case the complete homotrimer), individual monomers or even specific pocket residues within an active site.

Structural alignments were conducted for all three PNPase homologs; EcPNPase, hPNPase and SsoExosome and the results showed that their cores (RNA binding domains absent/removed) were structurally very similar (Figure 4.7). Alignment of the EcPNPase Chain_A monomer (blue), with the hPNPase Chain_1 monomer (pink) and the SsoExosome Rrp41 Chain_AB monomer (Green) was conducted (Figure 4.7 (b)) and all three proteins active sites overlay well (Figure 4.7 (c)). It is important to note that the essential magnesium metal ion, required within the active site for catalysis, was only present in EcPNPase and was placed within hPNPase and SsoExosome structures (Figure 4.7, magenta spheres), as outlined later in Section 4.3.3.2.

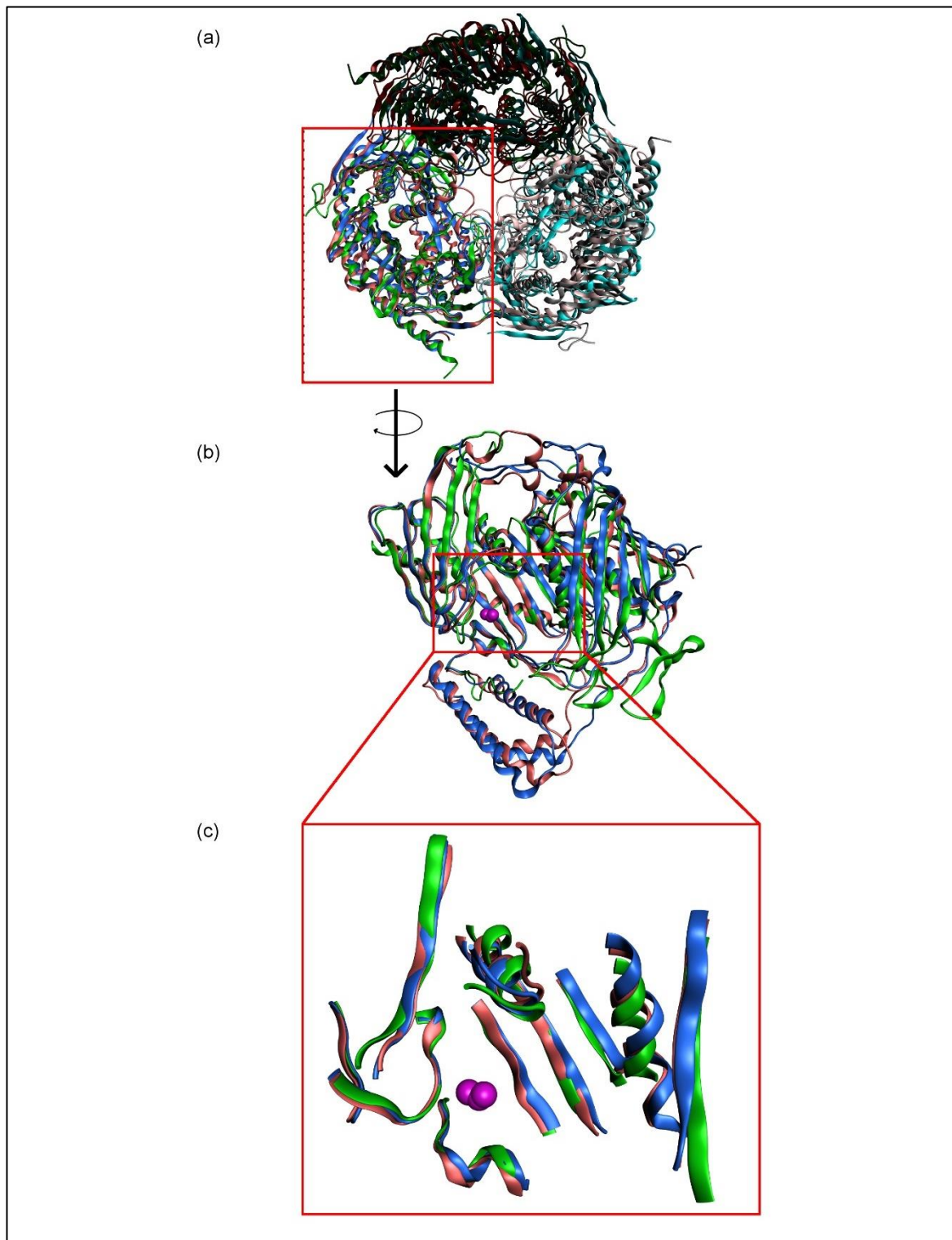


Figure 4.7 Structural Conservation of PNPase Homologs

(a) EcPNPase (3GCM), hPNPase (3U1K) and SsoExosome (4BA1) superimposed, a box (red line) highlights individual PNPase monomers and the corresponding Rrp41 and Rrp42 archaeal exosome proteins. (b) A zoomed in image of these 'monomers' are shown in isolation, with chains coloured blue, pink and green for EcPNPase (Chain A), hPNPase (Chain 1) and SsoExosome (Chain AA/AB) respectively, a box (red line) highlights the location of the active site. (c) Another zoomed in image of the active site shows the structural conservation of the three homologs and location of co-crystallised EcPNPase and placed hPNPase and SsoExosome Mg²⁺ ions (pink spheres). Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

The average distance between atoms of these superimposed chains was also calculated (Figure 4.8). The measured Root-Mean-Square Deviation (RMSD) values for EcPNPase and hPNPase were relatively low (1.25 Å) indicating high structural conservation (Figure 4.8 (b)). The RMSD values

for SsoExosome with either EcPNPase or hPNPase were higher, measuring 2.65 Å and 2.59 Å respectively (Figure 4.8 (b)). However, when examining the aligned structures and the RMSD values further, peaks in RMSD were observed in discrete regions of the sequence (Figure 4.8 (c)). Specifically, when the average RMSD is plotted against the residue number it is clear that a large majority of the residues are in similar positions for the superimposed proteins, with variations occurring mostly at flanking or loop regions.

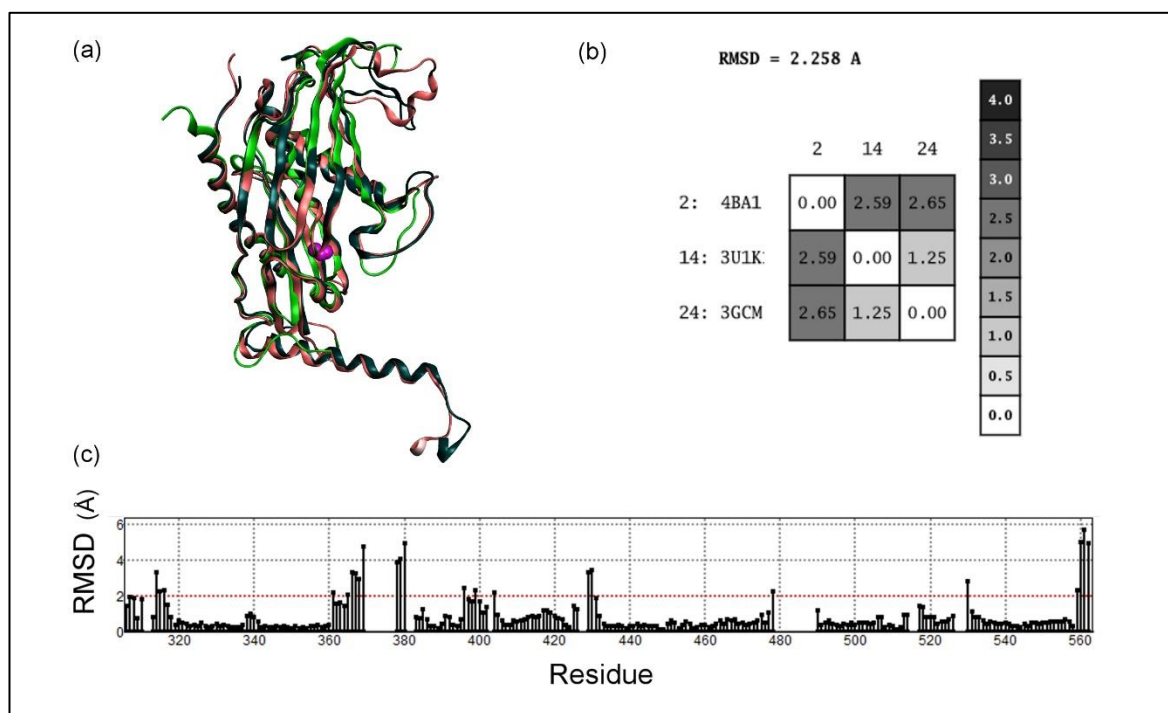


Figure 4.8 PNPase Homologs PH-1/Rrp41 Domain Conservation

(a) Structural overlay of the PH-1 domain of EcPNPase (3GCM, blue) and hPNPase (3U1K, pink) and Rp42 of SsoExosome (4BA1, green). (b) RMSD values of these proteins superimposed individually (grey-scale coloured boxes), the average RMSD was 2.258 Å and a grey-scale RMSD key is included for reference. (c) MOE graph representing the average RMSD (Å) plotted against residues. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

Mapping these variable RMSD regions onto the catalytically active PH-1/Rrp41 domain structures, as shown in Figure 4.9 (two-part figure), highlighted the areas with the largest structural differences (red ribbons). For example, the RMSD was higher (> 2 Å) at the start of the graph (residues 306-316, Figure 4.9 (b)), where the PNPase helical domain of hPNPase and EcPNPase ends and the PH-1 domain starts. This helical domain was not present in SsoExosome and may explain the variations in structure. The next region highlighted in Figure 4.9 (c) was between 361-381 in a loop region, it is well established that structural flexibility is more common in disordered regions and more constricted in structured alpha helix or beta sheets regions. Another loop region between 396-404 (Figure 4.9 (d)) was structurally variable; this was situated just before the RBRII motif, where the RMSD decreases. Three more loop regions, 429-430 (Figure 4.9 (f)), 478-490 (Figure 4.9 (g)) and 527-530 (Figure 4.9 (h)) showed higher RMSD values, these were all located in distinct regions from the active site, either near the top or bottom face of the core protein shown in Figure 4.7 (a). The final region shown in Figure 4.9 (j) was located at the end of the PH1/Rrp41

domain, just before the RNA binding domains (~560), which was not solved for EcPNPase, partially solved for hPNPase, and both are shown to form a flexible tail. Whereas the equivalent Rrp4 domain was solved for SsoExosome and was shown to form a more structured alpha helix.

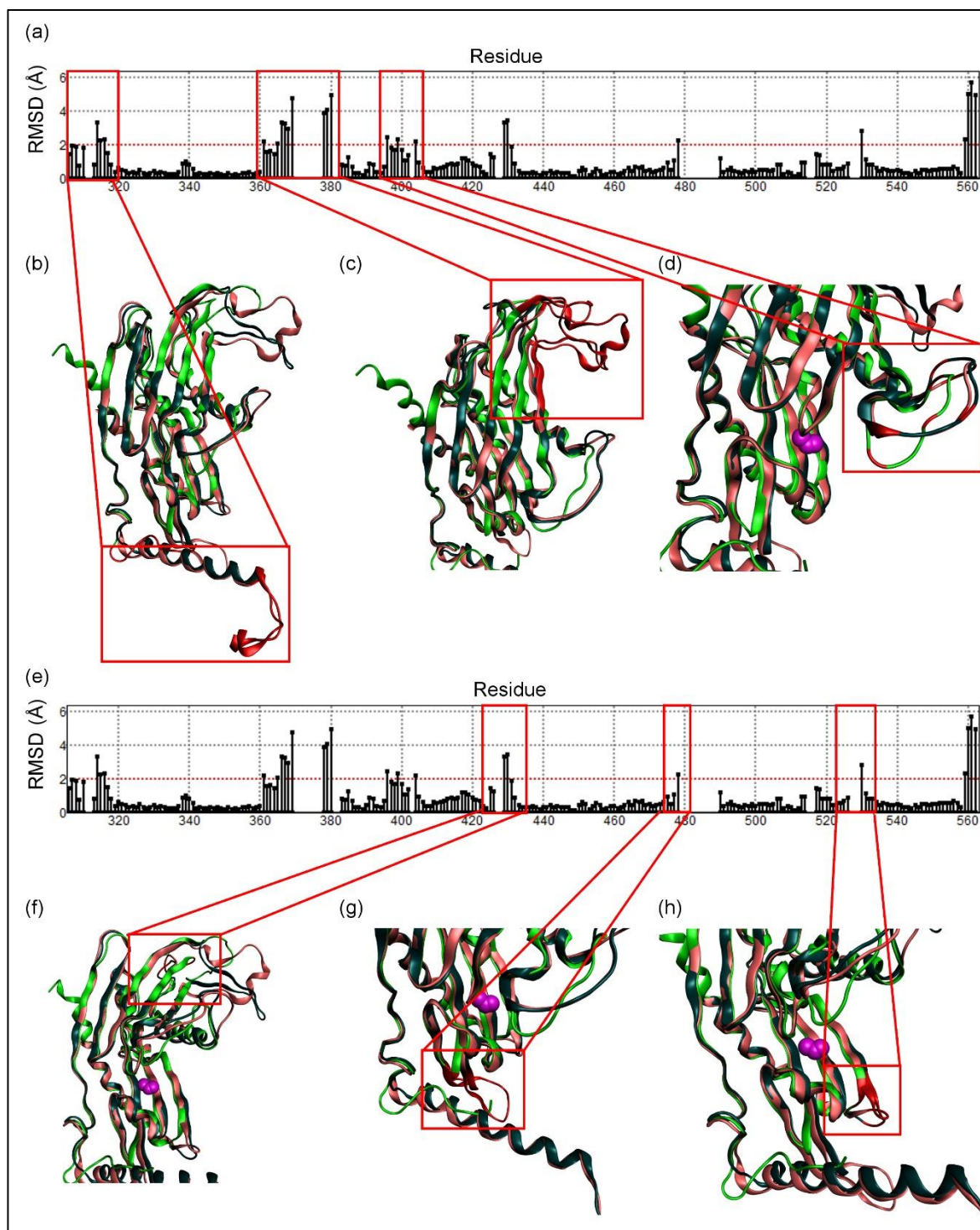


Figure 4.9 PNPase Homologs PH-1/Rrp41 Domain Variations (part 1 of 2)

(a) Average RMSD (Å) values vs the residues for the superimposed PH-1 domains of EcPNPase (3GCM) and hPNPase (3U1K) and Rrp42 of SsoExosome (4BA1). Individual regions of structural variability with an RMSD value $> 2\text{Å}$ are highlighted (dotted boxes) and mapped on structures (red ribbon) for residues (b) 306-316, (c) 361-381, (d) 396-404, (f) 429-430, (g) 478-490 and (h) 527-530. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

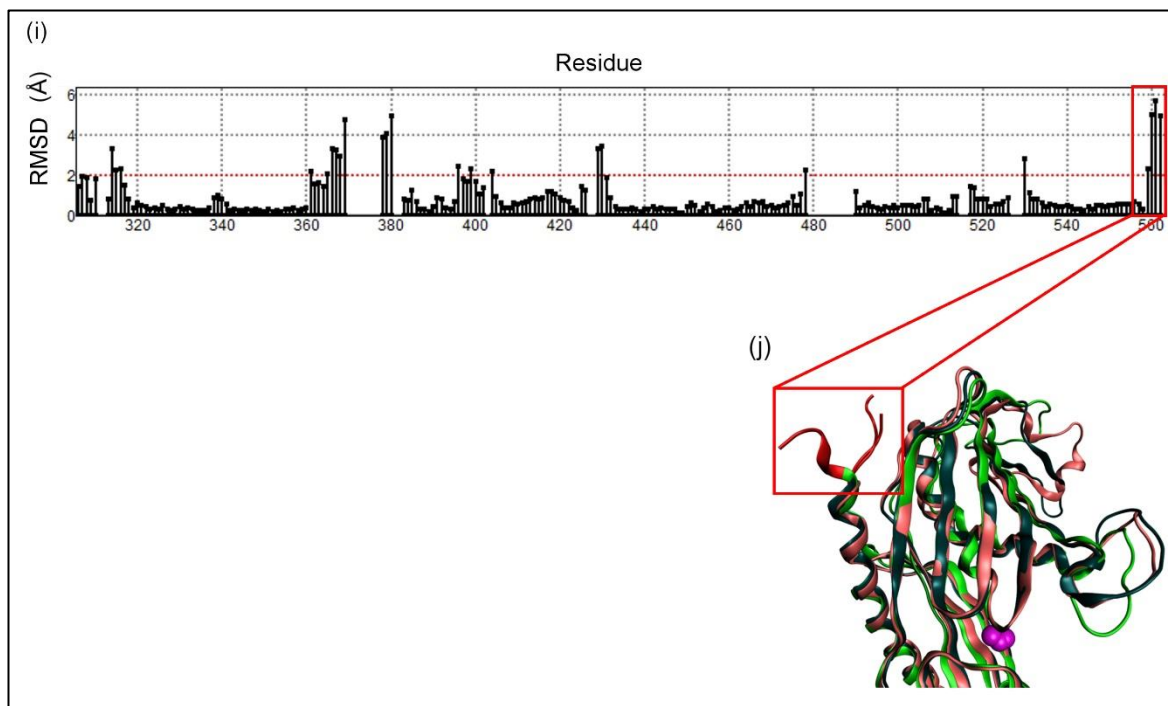


Figure 4.9 PNPase Homologs PH-1/Rrp41 Domain Variations (part 2 of 2)

(a) Average RMSD (Å) values vs the residues for the superimposed PH-1 domains of EcPNPase (3GCM) and hPNPase (3U1K) and Rrp42 of SsoExosome (4BA1). Individual regions of structural variability with an RMSD value $> 2\text{Å}$ are highlighted (dotted boxes) and mapped on structures (red ribbon) for residues (i) ~560. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

Some of the variable regions, discussed above, appeared to be located near to the active site Mg^{2+} (Figure 4.9 (d and h)). To investigate the structural conservation of the active site, key binding motifs important in citrate binding, shown previously in Figure 1.11, were inspected further. Accordingly, the RMSD values for residues within the RBRII, PBR and MBR motifs were analysed so that the residue flexibility across the PNPase homologs could be determined (Figure 4.10).

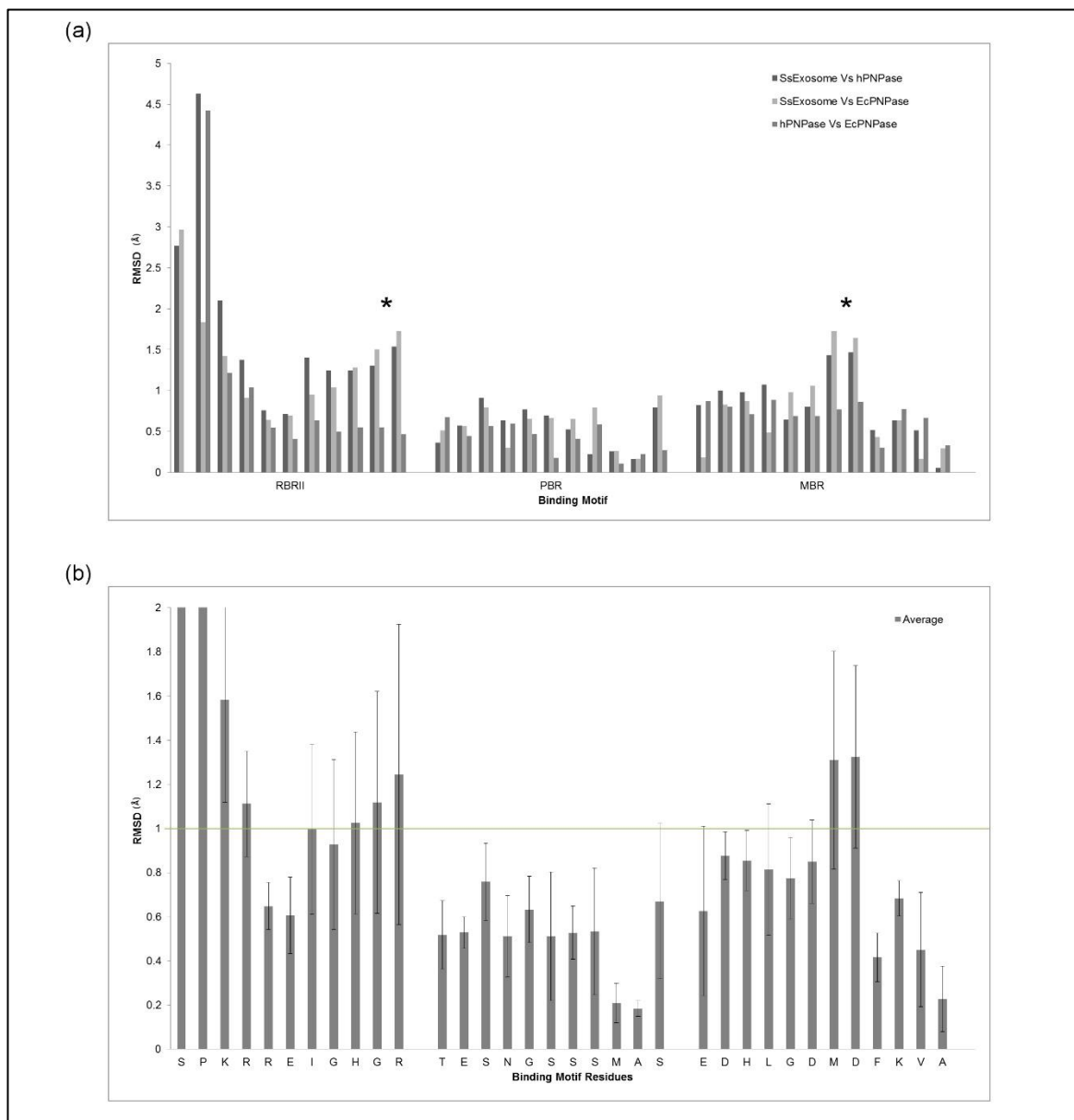


Figure 4.10 PNPase Homologs PH-1/Rrp41 Binding Motifs Conservation

(a) RMSD (\AA) values within the RBRII, PBR and MBR motifs of PH-1/Rrp41 were plotted for SsoExosome vs hPNPase (dark grey), SsoExosome Vs EcPNPase (light grey) and hPNPase Vs EcPNPase (medium grey). Asterisks indicate the areas where the similarity between residues in the PNPases and archaeal exosome homologs differ most. (b) Average RMSD values for binding motif residues shown in part (a) were calculated for EcPNPase vs hPNPase vs SsoExosome, with error bars depicting the standard deviation. The EcPNPase residues for which these relate to are provided on the x axis and a green line cuts the graph at 1 \AA on the y axis. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

Upon closer examination of Figure 4.10 (a), a majority of residues in the key citrate-binding motifs of PH-1/Rrp41 had low RMSD values, indicating small structural flexibility and thus high structural conservation across the three domain models. Looking at individual motifs, the PBR motif had the lowest average RMSD values ($< 1 \text{\AA}$) for all three alignments; SsoExosome vs hPNPase, SsoExosome vs EcPNPase and hPNPase vs EcPNPase, suggesting residues within this region were structurally very similar (Figure 4.10 (a)). Residues within the MBR were generally very similar, although slight difference were observed between PNPases and the archaeal exosome homologs for some residues, and the number of variances increased further in the RBRII motif, both highlighted by asterisks in Figure 4.10 (a). There was a large difference in RMSD between

hPNPase vs SsoExosome and EcPNPase at the start of the RBRII. When the actual residues were compared, a leucine was present in this position in hPNPase, however this was a 'bulky' proline in EcPNPase and SsoExosome, possibly explaining the large variation in RMSD.

The data presented in Figure 4.10 (a) were analysed further. The RMSD values for EcPNPase, hPNPase and SsoExosome were averaged, the standard deviation was calculated and the results were plotted against EcPNPase sequence, for comparison purposes (Figure 4.10 (b)). Interestingly the known metal coordinating aspartate residues within the MBR of PNPases (EcPNPase (D492), hPNPase (D544)) were slightly structurally different ($> 1 \text{ \AA}$) to the archaeal exosome (SsoExosome (D188)), when compared to other residues in the motif (Figure 4.10). Remarkably, the RMSD values were generally higher for the RBRII residues, with the exception of the known citrate binding arginine (EcPNPase (R399), hPNPase (R446)) which was structurally similar ($< 1 \text{ \AA}$) to the equivalent SsoExosome residue (R99).

The data presented within this Section investigated structural similarities of PNPase homologs and identified key residues interacting with citrate. In summary, aligning protein structures of EcPNPase, hPNPase and SsoExosome revealed a high level of structural conservation within the RBRI, RBRII, PBR and MBR active site motifs.

4.3.2 Citrate-PNPase Interactions Are Conserved in all Domains of Life

Now that the structures of EcPNPase, hPNPase and SsoExosome have been shown to exhibit high similarity (Section 4.3.1), the next question addressed within this section, was whether their sequences were also highly conserved? More specifically were the residues, identified in Section 4.3.1 to interact with citrate, conserved more broadly? The *in silico* protein sequence alignment results presented within this section predicted that citrate binding residues are conserved in PNPase and archaeal exosome representative homologs, across all domains of life.

4.3.2.1 Active Site Sequence Alignments

A small collection of previously characterised PNPase homologs, with known structural information available, were selected for *in silico* sequence alignments. This included four bacterial PNPases, (*E. coli* (EcPNPase; Nurmohamed *et al.*, 2009; Shi, Yang, Lin-Chao, Chak, & Yuan, 2008), *Streptomyces antibioticus* (SanPNPase; Symmons *et al.*, 2000), *Caulobacter crescentus* (Hardwick, Gubbey, Hug, Jenal, & Luisi, 2012) and *Coxiella burnetii* (Franklin *et al.*, 2015), *Homo sapiens* PNPase (hPNPase; Lin *et al.*, 2012) and four archaeal exosomes, one from the phylum crenarchaeota (*Sulfolobus solfataricus* (SsoExosome; Lorentzen & Conti, 2005, 2012; Lorentzen *et al.*, 2005; Lorentzen, Dziembowski, Lindner, Seraphin, & Conti, 2007; Lu, Ding, & Ke, 2010) and three from the phylum euryarchaeota, (*Pyrococcus abyssi* Navarro, Oliveira, Zanchin, & Guimarães, 2008), *Archaeoglobus fulgidus* (Büttner, Wenig, & Hopfner, 2005; Hartung, Niederberger, Hartung, Tresch, & Hopfner, 2010) and *Methanothermobacter thermautotrophicus*

(MthExosome; Ng *et al.*, 2010). In order to illustrate the level of sequence conservation across evolutionally distinct organisms, a simple protein sequence alignment was conducted as per Section 4.2.1. A range of key citrate and Mg²⁺ binding residues, previously identified in EcPNPase and hPNPase (Sections 4.3.1), were highlighted so that comparisons between the PNPase homologs could be made (Lin *et al.*, 2012; Nurmohamed *et al.*, 2011). The results of this alignment are presented within Figure 4.11.

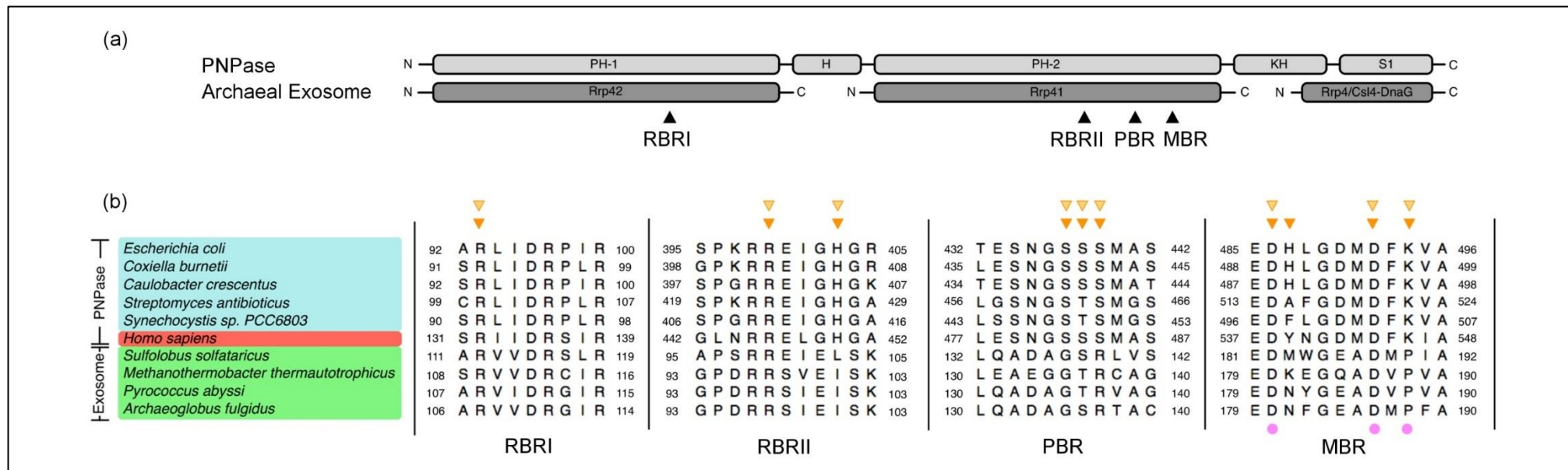


Figure 4.11 PNPase Homologs' Binding Motifs Sequence Conservation

(a) PNPase and Archaeal Exosome domains are illustrated as a schematic. An important RNA Binding Region (RBRI) in PNPase PH-1 domain and its homologous archaeal exosome Rrp42 protein are indicated. A helical domain is shown to link the PH-1 and PH-2 domains in PNPase only. A second RNA Binding Region (RBRII), a Phosphate and Metal Ion Binding Region (PBR and MBR, respectively) are then indicated within the PNPase PH-2 domain and Rrp41 archaeal exosome homologous protein (Portnoy *et al.*, 2007). The KH and S1 domains in PNPase and corresponding protein Rrp4 in archaeal exosome are also shown. (b) Sequence alignments for PNPase and Exosome proteins are provided, key binding regions are indicated, as described in (a). Key residues interacting with citrate and metal ions are highlighted, as per the key (Lin *et al.*, 2012; Nurmohamed *et al.*, 2009). The Image was created using a combination of GIMP (v2), Jalview and MOE (GIMP, n.d.; Molecular Operating Environment, 2013; Waterhouse *et al.*, 2009)

Comparison of the five bacterial PNPase sequences showed that all of the citrate-binding residues (Figure 4.11, dark orange arrows) known for EcPNPase were conserved, with the exception of a conservative substitution of threonine for serine at the position of EcPNPase S438 in SanPNPase (Figure 4.11). Consequently, as citrate binding interactions are already known for EcPNPase, this enzyme was selected as a prokaryotic PNPase representative for subsequent *in silico* molecular docking validation. SanPNPase was also selected as a prokaryotic PNPase representative for studying the effect of the S438T substitution. In addition to EcPNPase, citrate binding residues in hPNPase were also known (Figure 4.11, light orange arrows) and as a result, hPNPase was utilised as a eukaryotic PNPase representative for subsequent *in silico* molecular docking validation. Comparison of the four archaeal exosome sequences showed that there was more sequence variation amongst the putative citrate-binding residues (Figure 4.11). Only two of the eight citrate-binding residues identified in EcPNPase were absolutely conserved in the exosome homologs, with the greatest sequence variation occurring in the PBR. As the SsoExosome is well characterised, it was selected as the archaeal representative for subsequent *in silico* molecular docking.

In addition to citrate binding residues, key residues previously reported to be involved in coordinating the active site Mg²⁺ metal ions, essential for activity in EcPNPase, were highlighted in (Figure 4.11) (MBR, pink sphere). The two aspartate residues; D492 and D486 in *E. coli* were conserved in all PNPase and archaeal exosome homologs. The lysine residue, which contacts the metal ion in EcPNPase, was not conserved in archaeal Exosomes (SsoExosome; P465).

4.3.2.2 Vestigial Site Sequence Alignment

The ten protein sequences used for the basic alignments in Figure 4.11 were also analysed to investigate the conservation of citrate-binding at the EcPNPase vestigial site (Figure 4.12). Sequence conservation at this site was found to be much weaker than for the active site; important citrate-binding arginine residues in EcPNPase (R153, R372, R405 and R409) were not well conserved. Taken together with the absence of citrate at the vestigial site in the hPNPase crystal structure (Lin *et al.*, 2012), this suggested that citrate-binding at the vestigial site is highly unlikely to be conserved.

PNPase Exosome	Escherichia coli	150	G A A R V G Y	156	370	G E R T D T F L	377	404	G R L A K R G V L A V M	415
	Coxiella burnetii	149	G A A R V G Y	155	373	G D R Q E E F I	380	407	G R L A K R A V V P V V	418
	Caulobacter crescentus	150	G A A R V G W	156	372	G T Y K E S F L	379	406	G K L A W R A L R P M L	417
	Streptomyces antibioticus	157	G G V R V A L	163	394	P V T R K P Y M	401	428	G A L A E R A I V P V L	439
	Synechocystis sp. PCC6803	148	A A V R V G L	154	381	P E D E K R Y L	388	415	G A L A E R A I I P V L	426
	Homo sapiens	189	G A V R I G I	195	417	G I K D K N F M	424	451	G A L A E K A L Y P V I	462
	Sulfolobus solfataricus	198	V T I S V A K	204	71	L P D R A V L R	78	104	S K V I R E A L E S A V	115
	Methanothermobacter thermautotrophicus	194	L M C T F A K	200	69	R P D R A V I R	76	102	S K I T A E A L R P A L	113
	Pyrococcus abyssi	194	V P V T F A K	200	69	R P D R A I L R	76	102	S K V I K G A L E P A L	113
	Archaeoglobus fulgidus	185	V S V T S L I	191	69	D P S K A I I R	76	102	S K V S K E A F E A V I	113

Figure 4.12 PNPase Vestigial Site Sequence Alignment

Sequence motifs relating to the four vestigial site arginine's (orange arrows) identified through mutagenesis as being involved in binding citrate in EcPNPase (Nurmohamed *et al.*, 2011). The Image was created using a combination of GIMP (v2) and Jalview (GIMP, n.d.; Waterhouse *et al.*, 2009).

In summary, aligning protein sequence for well characterised PNPase homologs revealed a high level of sequence conservation for the active site citrate binding residues and poor conservation at the vestigial site.

4.3.3 Citrate is Predicted to Dock *In Silico* into Prokaryote, Eukaryote and Archaeal PNPase Homologs

With mounting evidence that citrate-interacting residues are conserved in distantly related PNPase homologues (Section 4.3.2), the question then arises whether citrate binding conserved in other PNPases. The *in silico* molecular docking results presented within this section predicts that citrate binding is conserved in PNPase and archaeal exosome representative homologs, across all domains of life.

4.3.3.1 Candidate Selection for *In Silico* Docking

Structure-based molecular docking is a valid computational approach for testing the potential for ligand binding *in silico*, before moving on to more resource-consuming experimental methodologies (Galli *et al.*, 2014; Kime *et al.*, 2015). Although the crystal structures of all 9 PNPase and archaeal exosome homologs aligned in Figure 4.11 were available, and in theory all structures could be docked using the methodologies outlined within this section, due to time limitations of this PhD project, only a select few structures were chosen for subsequent *in silico* molecular docking calculations. More specifically, EcPNPase (3GCM) and hPNPase (3U1K) were initially chosen to validate the *in silico* molecular docking approach, since they were already known to bind citrate. An additional bacterial structure, SanPNPase (1E3P) was also selected to test the effect of sequence variation in the S(S/T)S motif within the PBR. Finally, since no information was available regarding citrate binding to the archaeal exosome, SsoExosome was chosen as a well characterised archaeal model to discover if citrate binding was conserved in this domain. A few structures were available for SsoExosome and the structure termed 4BA1, which contained the Rrp4 RNA binding domain was selected as it was most similar to PNPase homologs KH and S1 domains, in contrast to the alternative Csl4-DnaG which has a zinc-ribbon domain (Büttner *et al.*, 2005).

In summary, at least one structure from each domain was selected for citrate docking studies. However, before *in silico* docking calculations could be conducted, each of these candidate protein structures were first prepared for docking, the results of which are described in the subsequent Sections 4.3.3.2-4.3.3.3.

4.3.3.2 Protein Structure Preparation: Magnesium Placement

It was valuable to check that EcPNPase could be used to test the correct placement of Mg²⁺ ions within the active site of PNPase homologs lacking the essential metal ion. To validate this method, magnesium was first removed from EcPNPase and then replaced within the enzyme, as outlined below.

The EcPNPase 3GCM structure was initially downloaded from the PDB and the input file was loaded as a biomolecule into MOE. The structure deposited contained three EcPNPase chains A, B, C (549 amino acid residues, coloured black, dark grey and light grey) which form the homotrimer, donut-like shape, shown previously in Figure 1.10. In order to assess the correct placement of Mg^{2+} ions, all the original Mg^{2+} ions were removed from the 3GCM file, which was then renamed to 3GCM_Dock. This 3GCM_Dock file was then superimposed with the original 3GCM file (containing Mg^{2+}). The Mg^{2+} within 3GCM was then successfully transferred across into the new 3GCM_Dock structure superposed on top and it was defined as part of the structure; effectively transferring the metal ion back into EcPNPase. Following removal of 3GCM from the MOE window, the new 3GCM_Dock structure was prepared using the MOE LigX built in tool and standard LigX protein preparation methods as described in Section 4.2.2.1.

Post protein preparation, the 3GCM and 3GCM_Dock structures were again superimposed in order to identify whether the Mg^{2+} ion was correctly coordinated by the appropriate active site residues, results of which are shown in Figure 4.13. As expected, very little overall change was observed between the two structures; MOE predicted a root mean square deviation (RMSD) value of 0.248 Å for the 3GCM structure vs the 3GCM_Dock (Figure 4.13 (a-b)). The active site, shown in Figure 4.13 (c) highlights the similarity between 3GCM (dark grey) and 3GCM_Dock (turquoise). Contrastingly and unexpectedly, the newly placed Mg^{2+} ion appeared to have moved slightly within the active site of 3GCM_Dock; as highlighted by the purple (3GCM Mg^{2+}) and pink (3GCM_Dock Mg^{2+}) spheres in Figure 4.13 (c). Additionally, as indicated by the MOE interaction map, the active site Mg^{2+} which was originally coordinated by one aspartate residue (D492) in 3GCM (Figure 4.13 (d)) was now co-ordinated by both D492 and D486 in the new 3GCM_Dock structure (Figure 4.13 (e)). This observation was seen for each active site Mg^{2+} ion, in all three monomers, suggesting that MOE calculated that both aspartates were involved in coordinating Mg^{2+} , as shown later in Figure 4.15 (a, c & e).

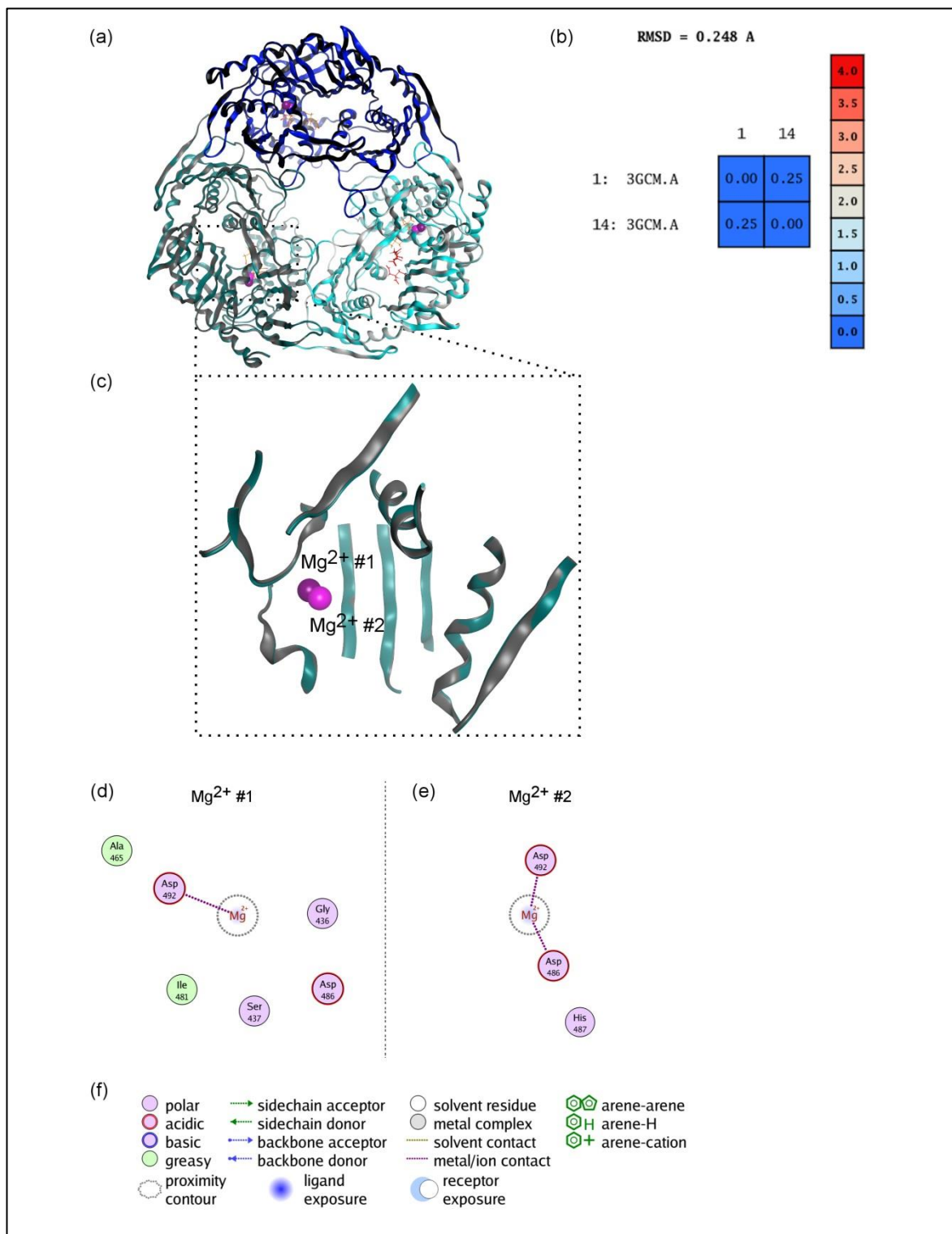


Figure 4.13 EcPNPase Structure Preparation

(a) EcPNPase 3GCM and 3GCM_Dock overlay of trimeric structure and (b) RMSD value of the protein alignment. (c) Active site 3GCM (grey ribbons) and 3GCM_Dock (turquoise ribbons) with the original Mg^{2+} (purple sphere; 3GCM) and the placed Mg^{2+} (pink sphere; 3GCM_Dock). (d) MOE interaction map highlighting Mg^{2+} coordination in 3GCM only by D492 and (d) in 3GCM_Dock by both D486 and D492. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

The publication which originally presented the EcPNPase structure suggested that D486, D492 and K494 were all involved in coordinating a Mn^{2+} ion (Figure 4.14) and that similar interactions would be predicted for Mg^{2+} (Nurmohamed *et al.*, 2009). Hence the calculations MOE generated, predicting that not one but two aspartates were involved in Mg^{2+} coordination, appeared to align

well to previous publications. However, in the 3GCM MOE interaction map, K494 was not suggested to interact with the metal ion, in any of the three monomers.

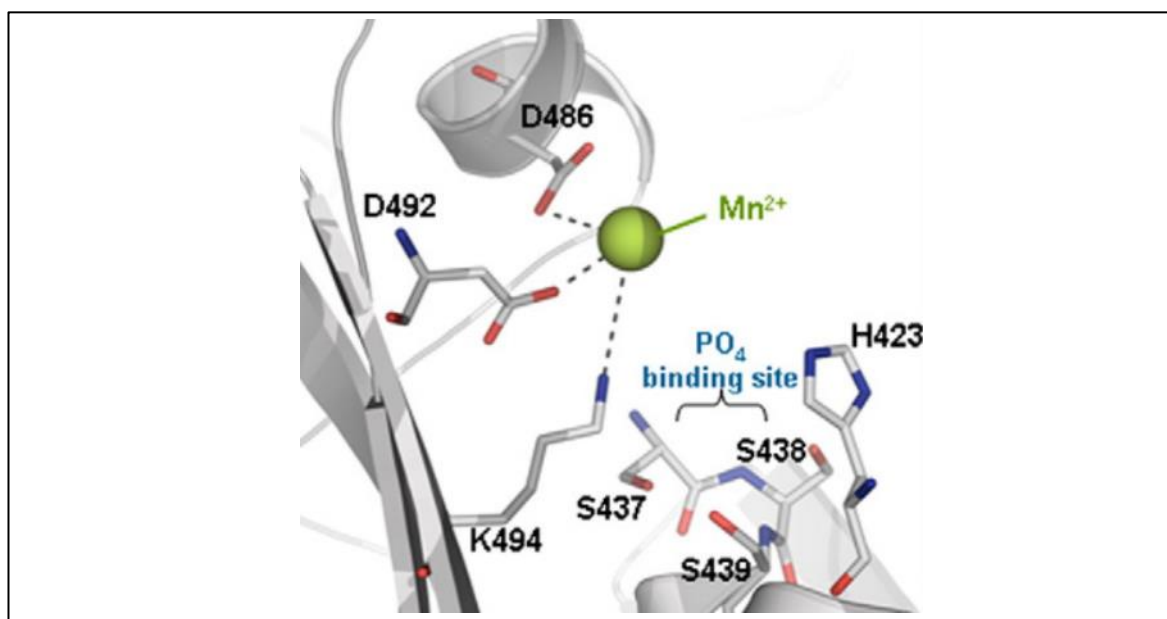


Figure 4.14 EcPNPase Mn²⁺ Coordination

(a) Mn²⁺ (green sphere) coordination by EcPNPase active site residues D486, D492 and K494 (grey sticks). Image taken directly from Nurmohamed *et al.*, 2009.

Interestingly if the 3GCM structure, which was originally taken directly from the PDB previously, was energy minimised using the built in tools within the MOE program instead, this second aspartate contact was made. In this minimised and renamed structure (3GCM_Min), both the D486 and D492 residues interacted with the Mg²⁺ ion and this was true for all monomers (Figure 4.15). This result suggested that proper minimisation of 3GCM was required to make the same contacts as that seen previously published (Figure 4.14) (Nurmohamed *et al.*, 2009).

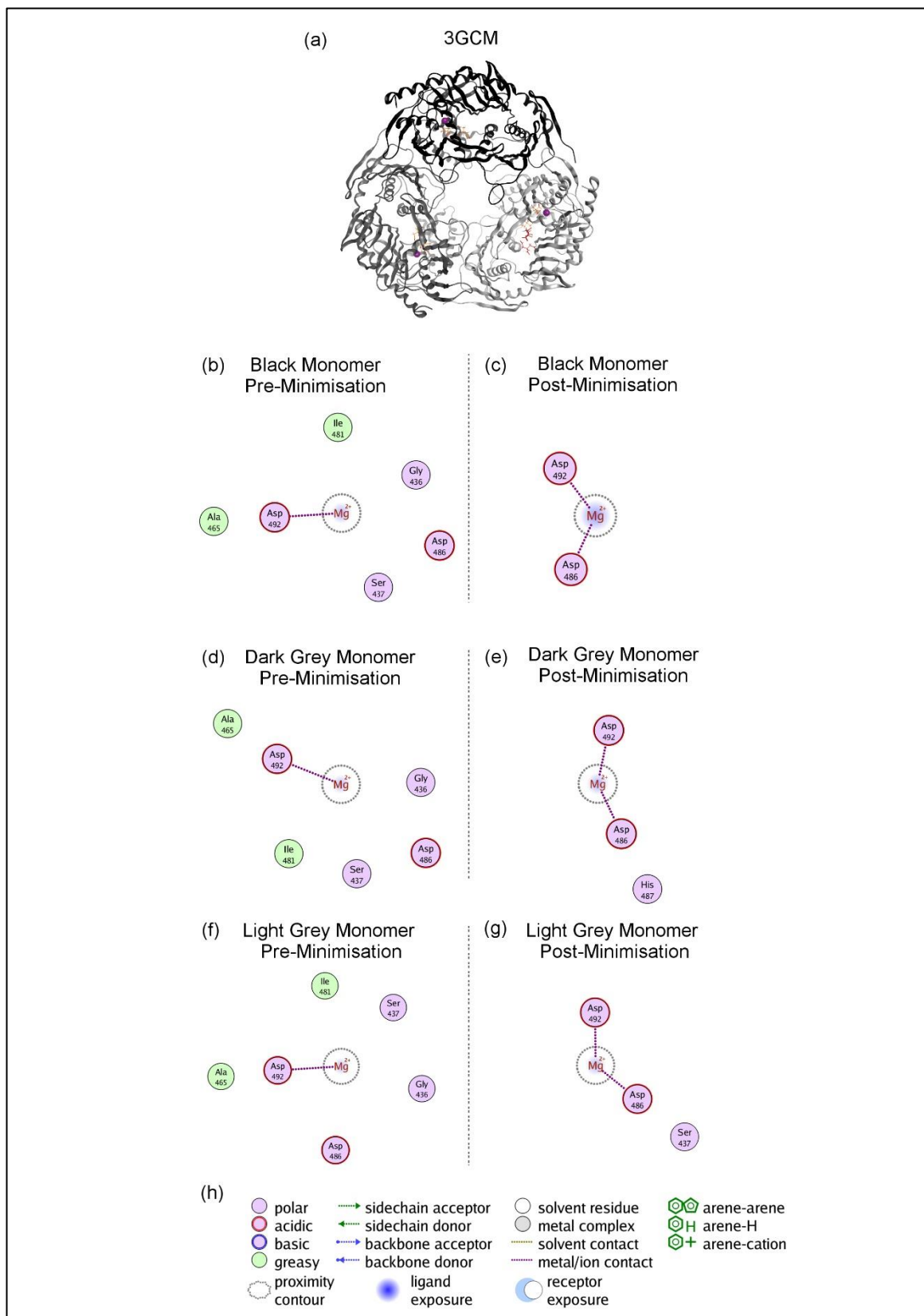


Figure 4.15 EcPNPase Structure Preparation; Mg²⁺ Placement

MOE generated interaction maps of Mg²⁺ interacting with (a) EcPNPase 3GCM pre and post protein minimisation for (b-c) black monomer, (d-e) dark grey monomer and (f-g) light grey monomer respectively and (h) the legend. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

In summary EcPNPase can be successfully used as a template for placing Mg²⁺ in structures lacking this essential ion and MOE structure preparation ensured that the correct contacts were made within

the active site. Accordingly, the hPNPase and SsoExosome enzymes, lacking the presence of an Mg^{2+} co-factor, were aligned against EcPNPase using MOE, the EcPNPase Mg^{2+} cofactor was defined as part of the receptor for the other two enzymes and then minimised using LigX. All docking structures were superimposed on original PDB structures and the RMSD was calculated to be ~ 0.25 Å (Table 4.2). This suggested that the MOE minimisation process had little overall effect on the PNPase and archaeal exosome structures.

Enzyme	RMSD of PDB vs Docking Structures
EcPNPase	0.248 Å
SanPNPase	0.257 Å
hPNPase	0.247 Å
SsoExosome	0.249 Å

Table 4.2 RMSD Comparison of PDB vs Docking Structure

The RMSD values (Å) of original PDB structures compared to docking structures are listed for PNPase and archaeal exosome homologs.

As before with the placement of Mg^{2+} within 3GCM_Dock, residues previously suggested to be involved in coordinating the metal ion in the active site were investigated to determine if similar contacts were made in hPNPase and SsoExosome homologs. As with EcPNPase (D492 and D486), key metal-aspartate interactions were sustained in PNPase homologs hPNPase (D538 and D544) and SsoExosome (D188) as highlighted in Figure 4.16 (a-c). Interestingly only one aspartate was predicted for SsoExosome, as before, this may have been an effect of MOE protein minimisation and this was queried further.

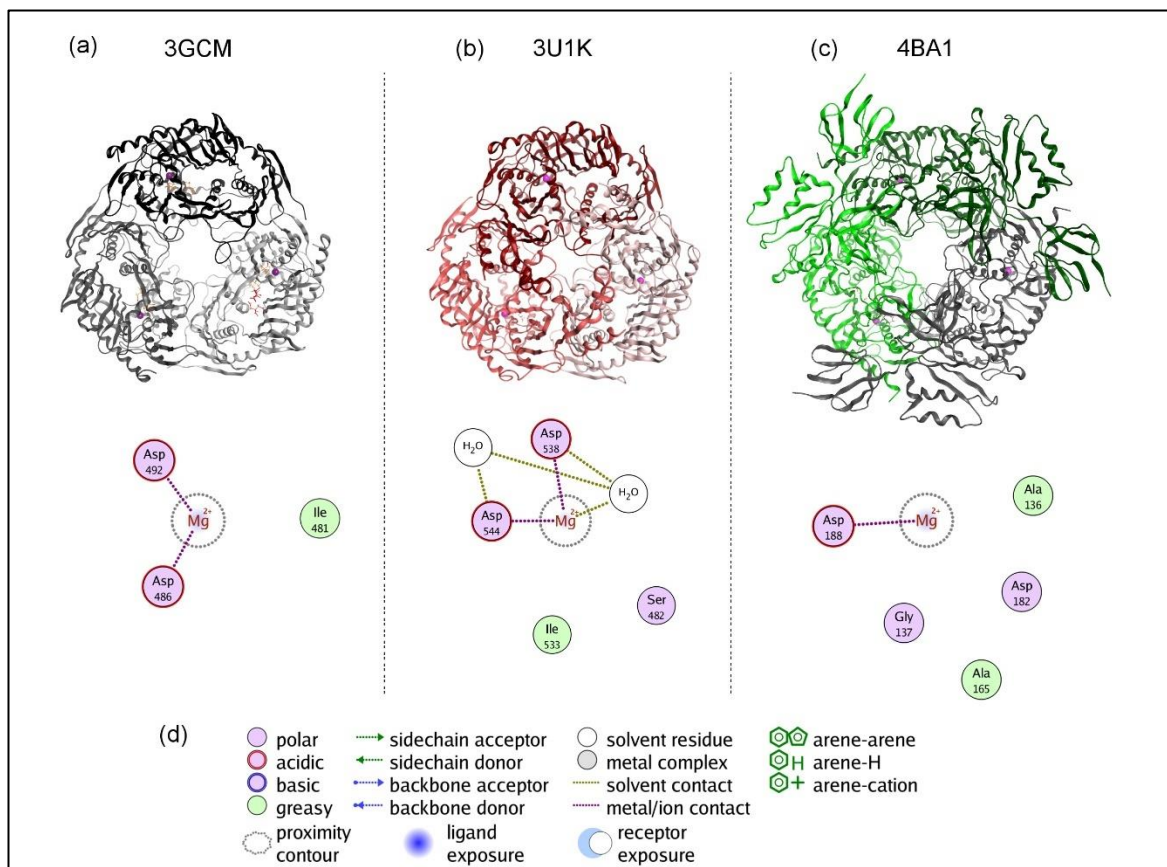


Figure 4.16 PNPase Homologs Structure Preparation; Mg²⁺ Placement

(a) Interaction map and structure for EcPNPase; 3GCM_Dock highlighting D492 and D486 interacting with Mg²⁺ (b) hPNPase; 3U1K_Dock binding Mg²⁺ via D538 and D544 (c) SsoExosome; 4BA1_Dock only interact with Mg²⁺ through a D188 metal contact and (d) the legend. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

It was observed that the SsoExosome 4BA1 structure, which has a D182A mutation for crystallisation purposes, only makes one interaction with Mg²⁺ upon placement within the active site and this was mediated by the D188 residue (Figure 4.17 (a)). *In silico* mutagenesis of the alanine at position 182, back to an aspartate, regained the second interaction with Mg²⁺ (Figure 4.17 (b)). However, upon minimisation, this second D182 interaction was again lost (Figure 4.17 (c)).

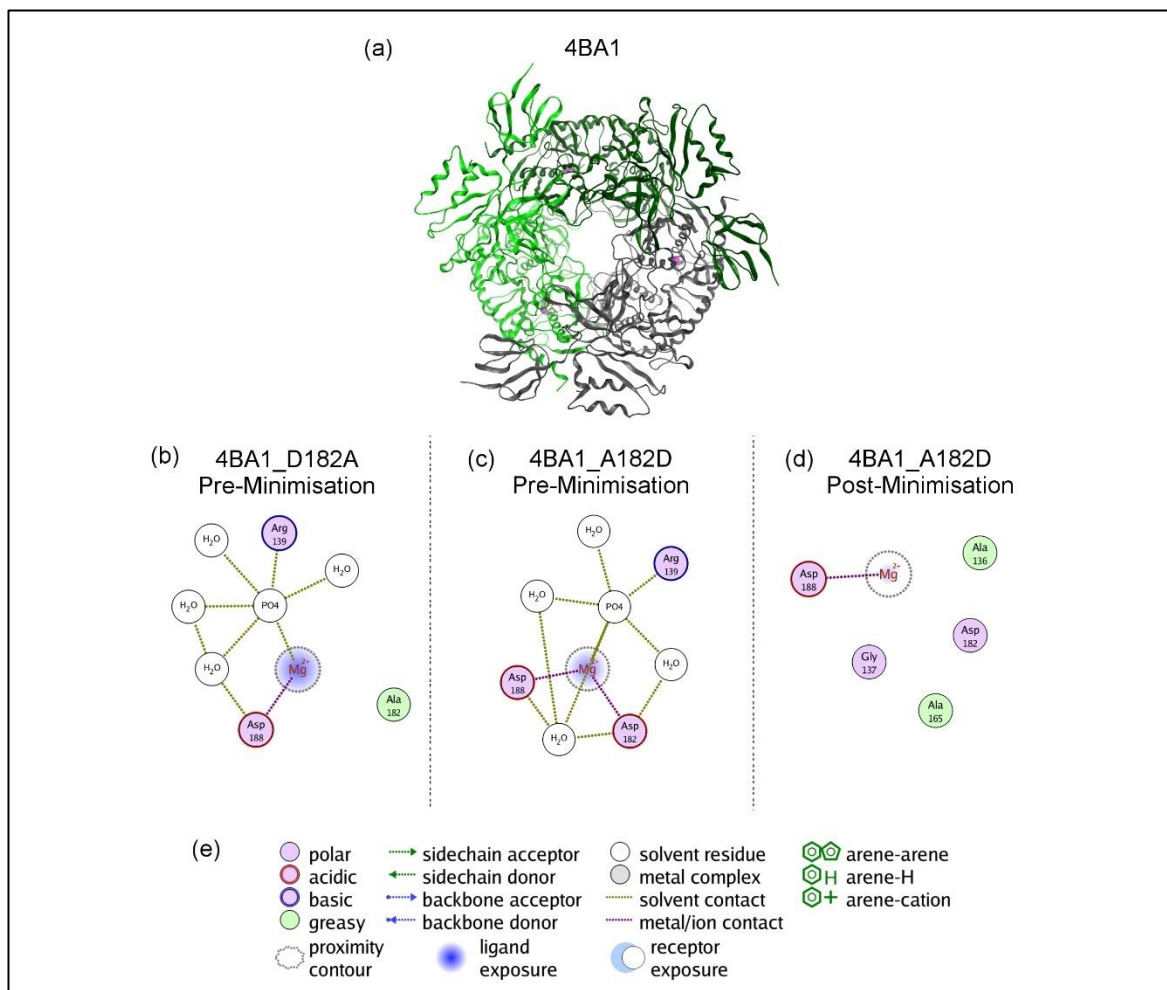


Figure 4.17 SsoExosome Structure Preparation; Mg²⁺ Placement

(a) SsoExosome structure and interaction map of 4BA1 interacting (b) with Mg²⁺ (b) with the crystallised alanine mutation (D182A), (c) following *in silico* MOE mutatgenesis back to aspartate (A182D), (d) after minimisation of A182D and (e) the legend. Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

The observation that following MOE minimisation only one aspartate coordinated the active site metal ion Mg²⁺ conflicted with previous predictions by Lorentzen & Conti, 2012. A possible explanation for this may be that Lorentzen and Conti placed Mg²⁺ in a structure with RNA and Pi present, however in the model presented in Figure 4.17 these ligands were absent/removed for the purpose of downstream citrate docking. Potentially the lack of these ligands during the minimisation processes, may explain why D182 interactions were not observed. Either way, whether this prevents citrate docking *in silico* required further examination and to do this the active site needed to be defined so that citrate docking could be conducted.

4.3.3.3 Protein Structure Preparation: Active Site Finder

Once the essential metal ions had been placed within the PNPase homologs, the MOE Alpha Site Finder panel was used to predict a pocket concavity for use in subsequent docking calculations. Essentially, MOE uses an input structure to predict pockets that may correspond to an active site. Following the MOE site finder calculation, a list of pocket ‘hits’ are suggested. The pocket residues are listed and dummy atoms can be placed to signify the locations of tight packing; as represented by red hydrophobic and blue hydrophilic spheres. The user can then view these dummy atoms

within the structure and visualise the predicted pocket hits easily. This allows comparison of the predicted pocket location, size and residues with a known active site. If the predicted pocket residues are the same as known active site residues it helps validate the program; demonstrating that MOE can correctly predict the location of an unknown active site.

To validate the site selection, the placement of dummy atoms, which represented the location of the predicted pocket and the residues within the predicted pocket were first checked. Accordingly, the known active site of EcPNPase, hPNPase, SsoExosome and SanPNPase were shown as ribbons in Figure 4.18-Figure 4.21 (a) so that the location of dummy atom placement and therefore the pocket location, could be compared in Figure 4.18-Figure 4.21 (b). For all four enzymes, MOE placed dummy atoms within the same location as the known active site, this is highlighted in the overlaid images in Figure 4.18-Figure 4.21 (c). The dummy spheres, representing the predicted pocket, were also placed in the same location as the co-crystallised citrate ligands of EcPNPase and hPNPase (Figure 4.18-Figure 4.19 (c)).

More importantly, MOE predicted the correct pocket; the pocket residues listed in Table 4.3 included the known EcPNPase active site residues (bold red) which interact with citrate and metal ions (R93, R399, S437, S438, S439, D486, H487, D492, K494, and H403). The corresponding ten residues in hPNPase, SsoExosome and SanPNPase were also identified, although for hPNPase the R132 and R446 residues of the RBRI and RBRII motifs respectively were not listed. The hPNPase predicted pocket was smaller than the others, with fewer residues and side chain contacts and fewer dummy atoms placed. The effect of dummy atom placement within hPNPase, and thus the size of the hPNPase pocket was analysed further. The default MOE pocket for hPNPase (small pocket, fewer dummy atoms), and the same hPNPase structure with additional EcPNPase dummy atoms placed (representing a larger pocket) were compared. These additional dummies were added by superimposing the EcPNPase 3GCM_Dock structure and its predicted pocket dummy atoms onto the hPNPase 3U1K_Dock structure and transferring the EcPNPase dummy atoms into hPNPase; effectively creating a new pocket (3GCM_Dock_EcDummies). Taking into account the default small pocket size and a larger pocket size (3GCM_Dock_EcDummies), citrate docking was compared and no difference in citrate binding was observed (data not shown). Hence the default MOE predicted hPNPase pocket was not limiting and was therefore used for subsequent hPNPase docking calculations. This result suggested that the placement of dummy atoms and therefore pocket prediction suggested by MOE was valid, further supporting the programs default settings as good parameters/methodologies for conducting docking calculations.

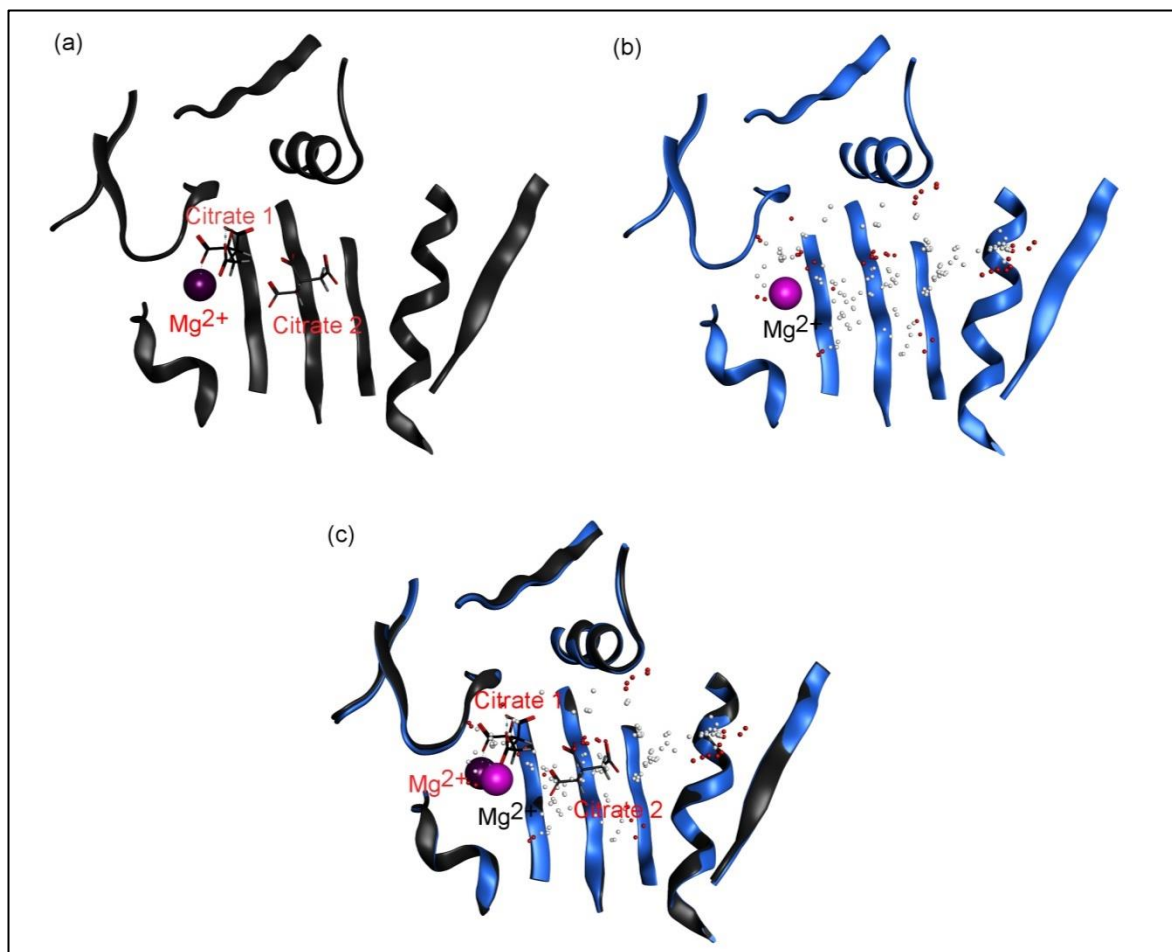


Figure 4.18 EcPNPase Structure Preparation; Site Finder

(a) EcPNPase 3GCM (dark grey ribbons) with co-crystallised Mg^{2+} (purple sphere) and active site citrates 1 and 2 (black sticks) (b) MOE predicted active site for 3GCM_Dock (blue ribbons), with placed Mg^{2+} (pink sphere) and hydrophobic and hydrophilic dummy atoms (red and blue spheres respectively) representing the MOE predicted active site pocket. (c) Superpose of images in (a) and (b). Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

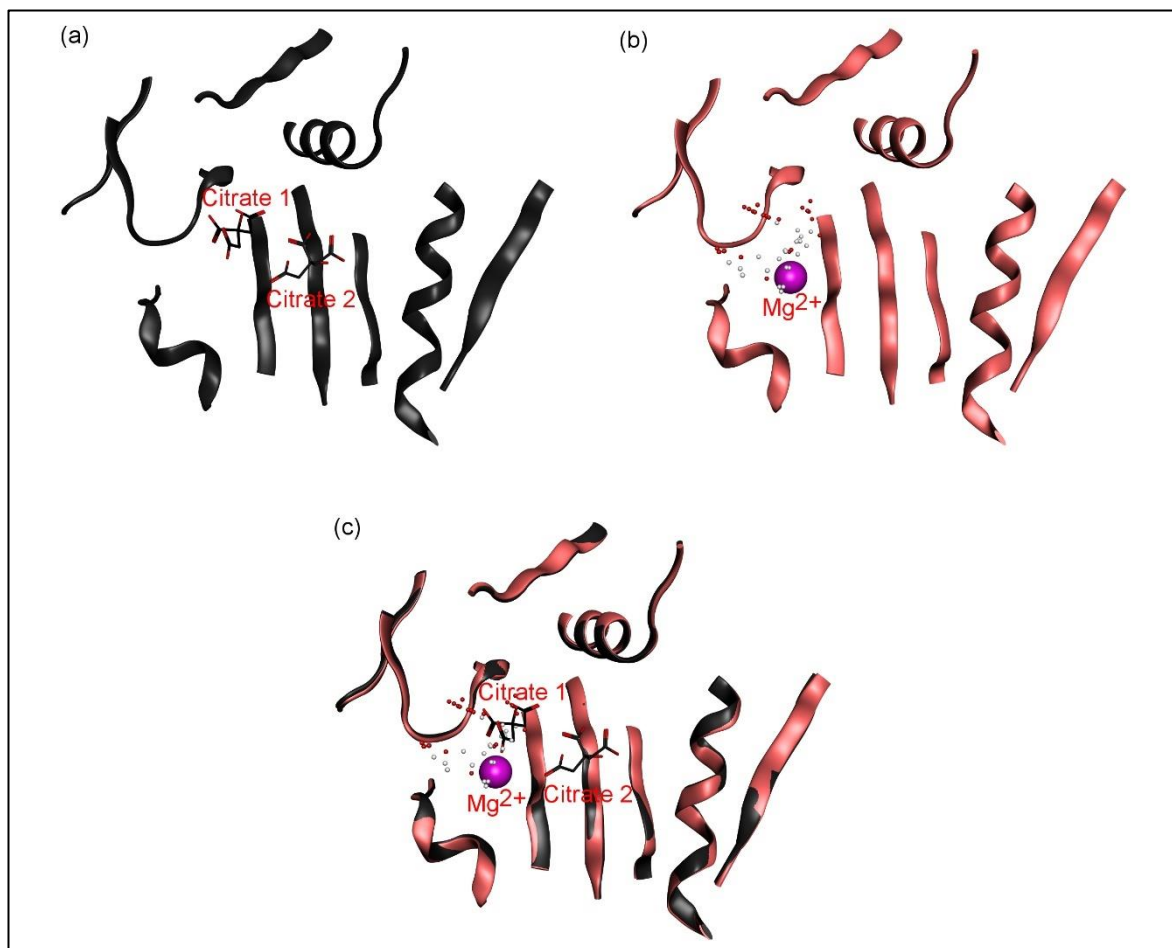


Figure 4.19 hPNPase Structure Preparation; Site Finder

(a) hPNPase 3U1K (dark grey ribbons) with active site citrates 1 and 2 (black sticks) (b) MOE predicted active site for 3U1K_Dock (blue ribbons), with placed Mg²⁺ (pink sphere) and dummy atoms (red and blue circles) representing the MOE predicted active site pocket. (c) Superpose of images in (a) and (b). Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

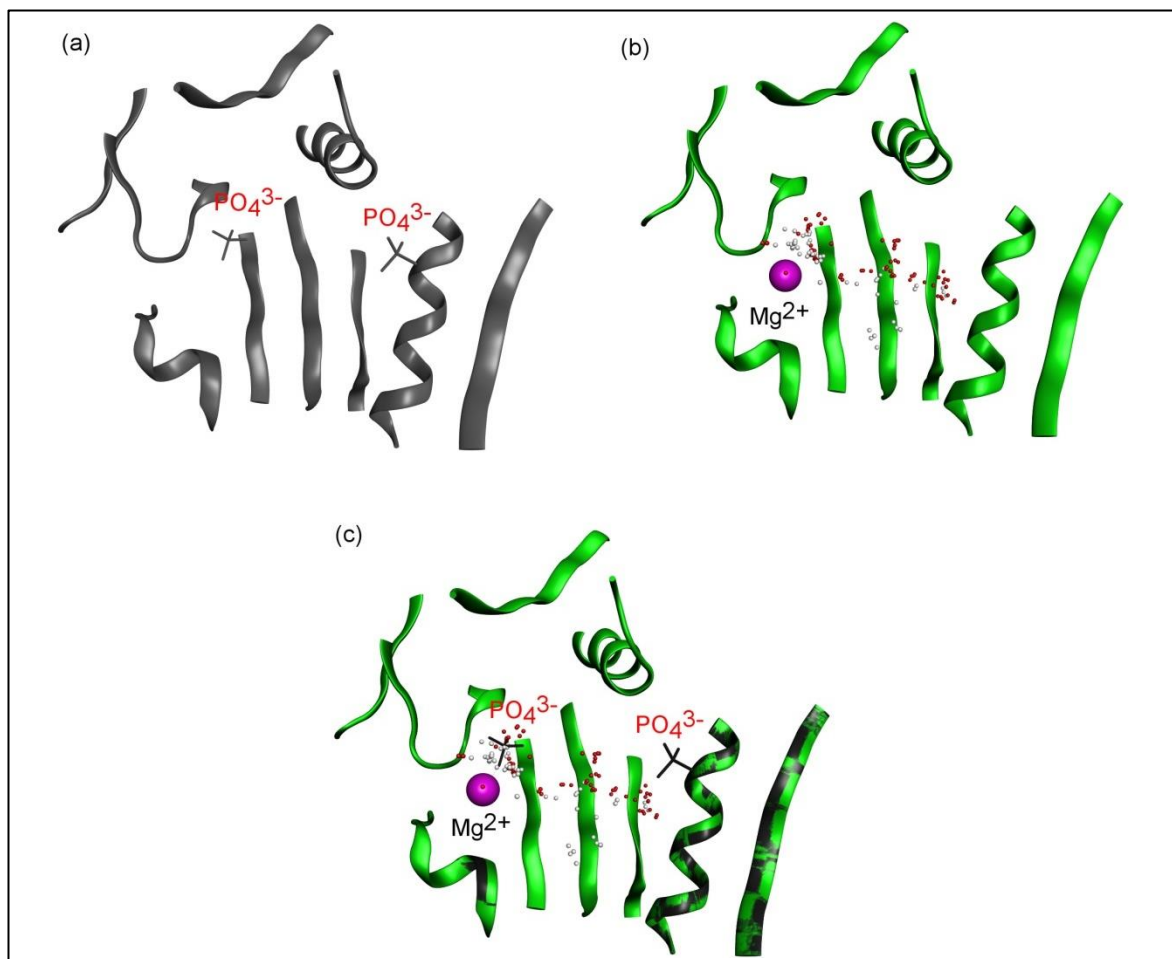


Figure 4.20 SsoExosome Structure Preparation; Site Finder

(a) SsoExosome 4BA1 (dark grey ribbons) with two co-crystallised PO₄³⁻ (grey sticks) (b) MOE predicted active site pockets for 4BA1_Dock (blue ribbons), with placed Mg²⁺ (pink sphere) and dummy atoms (red and blue circles) representing the MOE predicted active site pockets. (c) Superpose of images in (a) and (b). Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

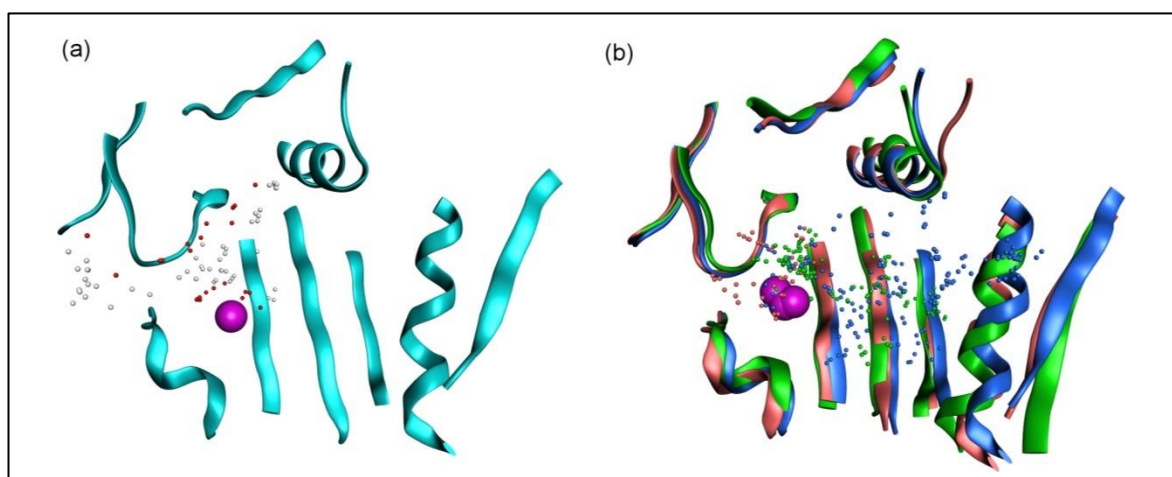


Figure 4.21 SanPNPase Structure Preparation; Site Finder and Site Selection Conservation

(a) SanPNPase 1E3P (cyan ribbons) with placed Mg²⁺ (pink sphere) and dummy atoms (red and blue circles) representing the MOE predicted active site. (b) Superimpose MOE predicted active sites for EcPNPase (3GCM_Dock, blue), hPNPase (3U1K_Dock, pink) and SsoExosome (4BA1_Dock, green), highlighting placed Mg²⁺ and dummy atoms (coloured based on ribbons). Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

Enzyme	Site	Size	Hyd	Side	Pocket Residues
EcPNPase	Active	163	32	215	(LEU59 THR60 VAL61 ASN62 ALA92 ARG93 ASP96 ARG97 PRO98 ARG100 PRO101 LEU187 MET188 THR348 TYR380 PHE382 LYS397 ARG398 ARG399 GLU400 HIS403 LEU406 SER434 ASN435 GLY436 SER437 SER438 SER439 ILE481 LEU482 GLY483 ASP486 HIS487 ASP492 LYS494 GLN506 ASP508 ILE509 LYS510)(MG552)
hPNPase	Active	48	8	74	(LYS306 THR391 TYR427 PHE429 THR434 ARG446 HIS450 VAL476 SER479 ASN480 GLY481 SER482 SER483 SER484 MET485 LEU534 GLY535 ASP538 TYR539 ASP544 LYS546)(MG552)
SsoExosome	Active	123	24	161	(ARG112 VAL113 ARG116 SER117 ARG119 ASP120 SER228 GLY237 ILE238 GLN239)(TYR81 MET83 THR88 ARG99 GLU102 LEU103 ALA134 ASP135 ALA136 GLY137 SER138 ARG139 LEU140 ALA165 GLU179 ASP182 MET183 ASP188 PRO190 GLN204 LEU205 ASN206)(MG552)
SanPNPase	Active	73	22	119	(LYS288 GLU371 TYR404 PHE406 SER410 VAL411 LYS421 ARG423 GLU424 HIS427 SER458 ASN459 GLY460 SER461 THR462 SER463 ALA489 ILE509 LEU510 GLY511 ASP514 ASP520 LYS522)(GLN32)(MG556)

Table 4.3 MOE Site Selection

MOE predicted pocket for EcPNPase 3GCM_Dock, hPNPase 3U1K_Dock, SsoExosome 4BA1_Dock and SanPNPase 1E3P_Dock are listed; the number of receptor atoms (Size), the number of hydrophobic contacts (Hyd), and number of sidechain contacts (Side) are reported. A list of the pocket residues is also provided and those residues highlighted in bold red correspond to known active site residues.

In summary, the results so far demonstrated that MOE could be used to prepare PNPase homolog structures for docking, including placing essential metal ions and selecting the appropriate pocket for subsequent docking calculations. The next validation step aimed to investigate whether MOE could also accurately dock citrate ligands within these prepared active site pockets. As before EcPNPase and hPNPase were used as models for citrate docking validation purposes; testing whether MOE could predict the same citrate binding interactions observed in previously crystallised structures.

4.3.3.4 Active Site *In-Silico* Citrate Docking

Citrate downloaded from the ZINC database (Figure 4.22) and saved as a MOL2 file in a physiologically relevant conformation was docked into 3GCM_Dock using the docking module of MOE, using the parameters outlined in Section 4.2.2.2. Due to limitations of the docking software, only a single molecule of citrate could be docked into each structure in a single docking run, hence two separate rounds of docking were required. To achieve this, the lowest-energy structure with the first molecule of citrate docked was used as the starting point for a second round of citrate docking. The orientations of the top docking hits, for both the first and second round of citrate docking calculations for EcPNPase (Figure 4.23 (a)) and hPNPase (Figure 4.23 (b)) are shown.

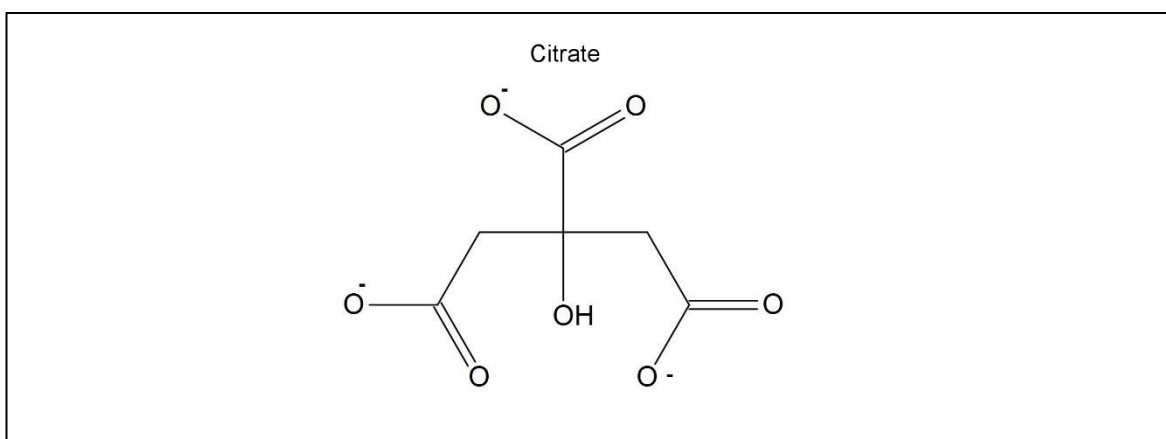


Figure 4.22 Citrate 2D Structure

Citrate downloaded from ZINC database (ZINC00895081), shown as 2D depiction. Image created in MOE (Molecular Operating Environment, 2013).

For both enzymes, in the first round of docking, citrate (Figure 4.23, yellow sticks) was predicted to bind at the catalytic active site in the same position as the original co-crystallised Cit 1 (Figure 4.23, light grey sticks). The lowest-energy docking scores (*S* values) were -112.3 kcal/mol and -122.7 kcal/mol for EcPNPase and hPNPase, respectively (Table 4.4). Encouragingly, MOE retained 30 unique lowest-energy poses for each docking calculation and they all placed citrate in this location, albeit in slightly different orientations. The *S* values for these are summarised in the box plot in Figure 4.24. Furthermore, the number of placement poses (iterations) showed no difference in *S* value between 1,000 and 300 iterations when using the triangle matcher placement method. Indicating 300 iterations to be sufficient for determining reliable MOE binding scores in the conditions tested (data not shown). Additionally, utilising 300 iterations and retaining 30 results

was adequate, retaining more results was unnecessary and negatively impacted the run times (data not shown).

The docking of the citrate from the first round of docking in the same position as Cit 1 in the original crystal structures of EcPNPase and hPNPase was a promising result. However, the potential for a second citrate molecule binding also needed to be investigated since two molecules of citrate were discovered bound to the active site of these enzymes' solved structures. As mentioned above, in order to do this, the lowest-energy pose with one molecule of citrate docked was used as the starting point for a second round of docking. Reassuringly, for both EcPNPase (Figure 4.23 (a)) and hPNPase (Figure 4.23 (b)) the second citrate molecule (Figure 4.23, orange sticks) docked at the equivalent location of Cit 2 observed in the crystal structures (Figure 4.23, dark grey sticks), with lowest-energy S values of -63.7 kcal/mol and -58.6 kcal/mol respectively (Table 4.4). As before, the orientation of docked citrate varied slightly, but all 30 poses were placed in this location (Figure 4.24).

Interestingly, the lowest-energy S values, reported in Table 4.4, proposed that of the two possible citrate-binding sites, which were expected based on the crystal structures, citrate binding at the Cit 1 site was predicted to be stronger than Cit 2. Overall, these results are consistent with the docking protocol successfully identifying the citrate-binding sites in both proteins.

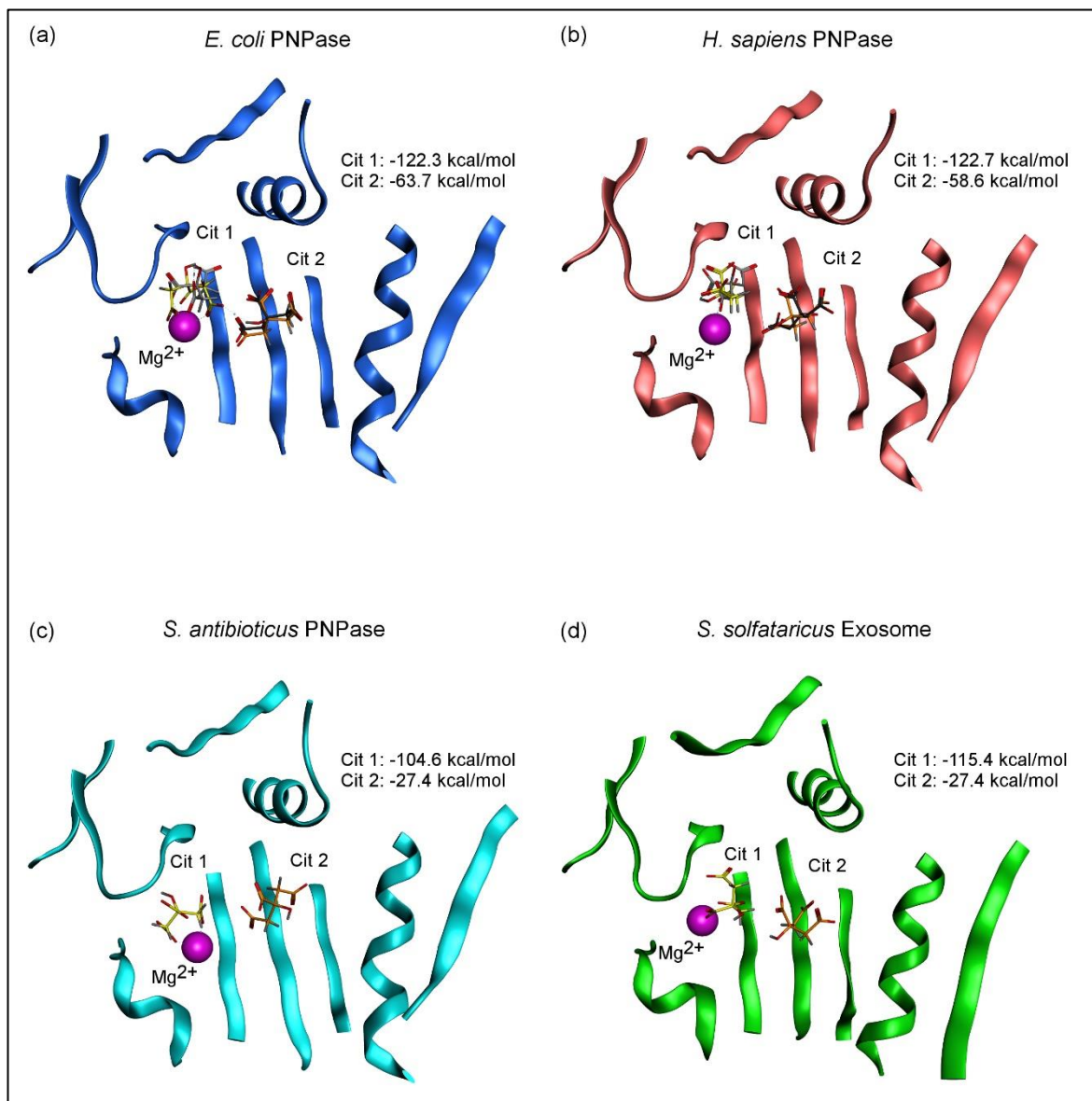


Figure 4.23 Active Site Citrate Docking

Co-crystallised citrates (citrate 1; light grey sticks, citrate 2; dark grey sticks) are shown superimposed on docking sites for (a) EcPNPase (3GCM_Dock, blue) and (b) hPNPase (3U1K_Dock, pink) with docked citrates (citrate 1; yellow sticks, citrate 2; orange sticks). Only the docked citrates are shown for (c) SanPNPase (1E3P_Dock, cyan) and (d) SsoExosome (4BA1_Dock, green) since no citrate was found to co-crystallise. The lowest-energy *S* values for the first and second round of citrate docking are also provided (kcal/mol). Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

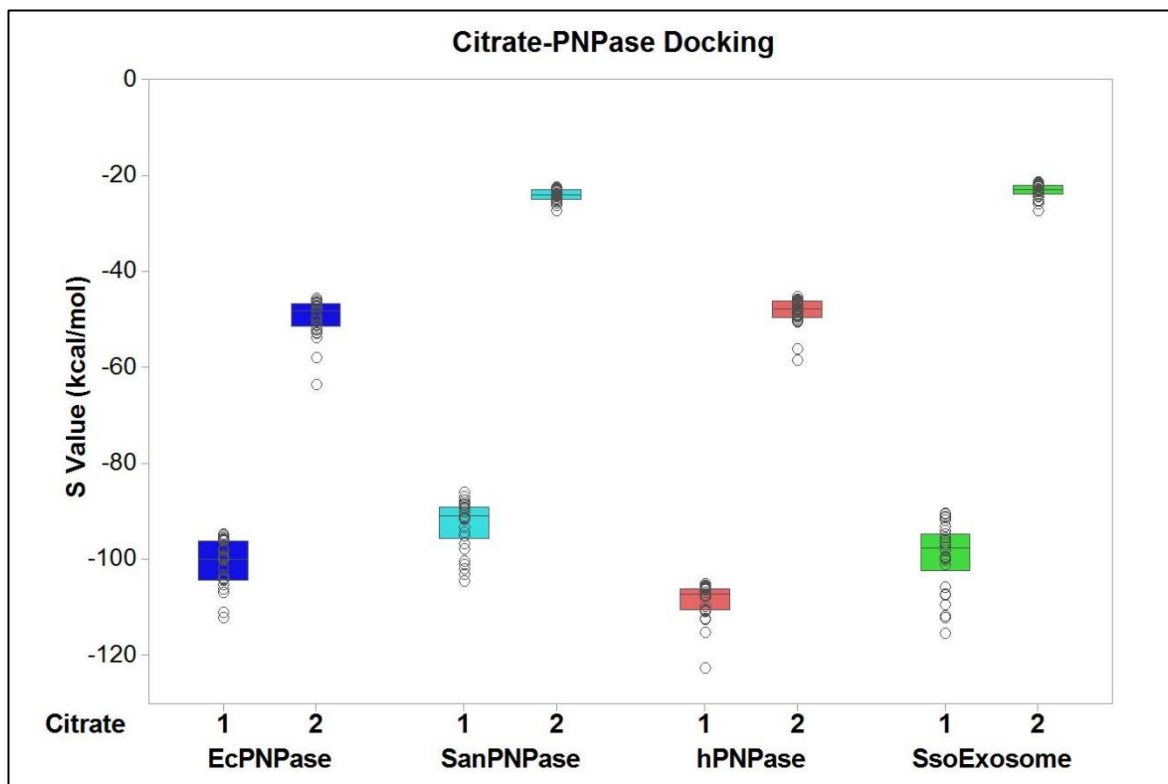


Figure 4.24 PNPase Homologs Active Site Citrate Docking; All Hits S Values

The 30 unique, lowest-energy poses docking scores (S values, kcal/mol) are provided for EcPNPase (3GCM_Dock), hPNPase (3U1K_Dock), SanPNPase (1E3P_Dock) and SsoExosome (4BA1_Dock) for the first and second round of citrate docking (Citrate 1 and 2). Images generated in Minitab (v17).

Having demonstrated the validity of the docking approach, citrate was subsequently docked into the structure of a prokaryotic PNPase, that has not previously been shown to bind citrate. As *E. coli*, *C. crescentus* and *C. burnetii* all belong to the phylum proteobacteria whereas *S. antibioticus* belongs to the phylum actinobacteria, SanPNPase was selected as the most divergent PNPase model, for which structural information was available for subsequent citrate docking. Additionally, as nine of the ten residues involved in Mg^{2+} and citrate-binding in EcPNPase (R93, R399, S437, S438, S439, D486, H487, D492, K494, and H403) were found to be conserved in SanPNPase (Section 4.3.2.1), with the only difference being a conservative substitution of threonine for serine at the equivalent position to EcPNPase S438, SanPNPase was a good example to test the effect of the S(S/T)S sequence variation within the PBR (Figure 4.11).

The results shown in Figure 4.23 (c), predicted little effect from the S(S/T)S sequence variation on the first round of citrate binding; the top 30 lowest-energy S values obtained for the docking at the Cit 1 site were comparable to those obtained for EcPNPase and hPNPase, with a lowest-energy S value of -104.6 kcal/mol (Table 4.4). However, for the Cit 2 site, the docking scores obtained were significantly smaller than for either of the other PNPases (Figure 4.24); the lowest-energy S value was just -27.4 kcal/mol for the Cit 2 site (Table 4.4).

The results obtained so far suggest that citrate-binding could be a common property of PNPases, which was not unexpected considering the high level of sequence conservation. The next step was

to investigate citrate-binding to the more distantly related archaeal exosome homolog. As for the three PNPases, two molecules of citrate were docked into the active site of the SsoExosome (Figure 4.23 (d)). The docking scores at the Cit 1 site were similar to those obtained for the other three PNPases (Figure 4.24); with a lowest-energy *S* value of -115.4 kcal/mol (Table 4.4). The second molecule of citrate docked at the Cit 2 site with a similar *S* value (-27.4 kcal/mol) to SanPNPase (Table 4.4).

Enzyme	Docking <i>S</i> Value (kcal/mol)	
	Citrate 1	Citrate 2
EcPNPase	-112.3	-63.7
SanPNPase	-104.6	-27.4
hPNPase	-122.7	-58.6
SsoExosome	-115.4	-27.4

Table 4.4 PNPase Homologs Active Site Citrate Docking; Top Hits *S* Value

The lowest-energy docking score *S* value (kcal/mol) for the first and second round of citrate docking (Citrate 1 and 2) with EcPNPase (3GCM_Dock), hPNPase (3U1K_Dock), SanPNPase (1E3P_Dock) and SsoExosome (4BA1_Dock).

Overall, these results predicted that citrate may bind and potentially regulate both prokaryotic and eukaryotic PNPases and, despite the sequence variation, the archaeal exosome. Citrate molecules also docked in a similar location for all four PNPase homologs tested and this is highlighted in Figure 4.25.

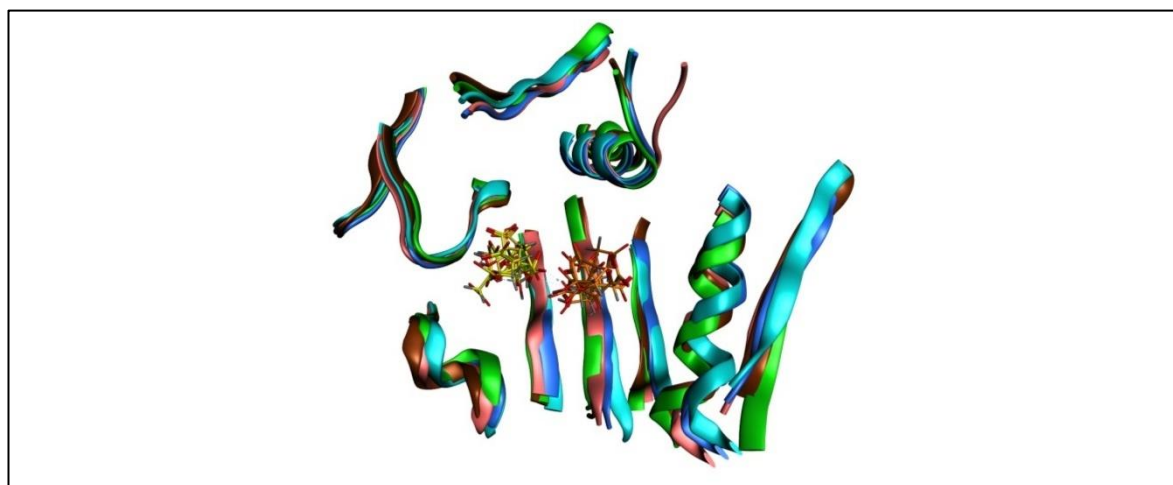


Figure 4.25 Active Site Citrate Docking Overlay

Active sites of EcPNPase (3GCM_Dock, blue), hPNPase (3U1K_Dock, pink), SanPNPase (1E3P_Dock, cyan) and SsoExosome (4BA1_Dock, green) are shown superimposed with docked citrates (citrate 1; yellow sticks, citrate 2; orange sticks). Images generated in MOE (Molecular Operating Environment, 2013) and prepared in GIMP (v2) (GIMP, n.d.).

The residues predicted by MOE to be involved in citrate-PNPase binding were analysed further. A PLIF was generated for the ten lowest-energy poses, with both one and two molecules of citrate docked. This was conducted for each enzyme and the information was summarised as a heat map in Figure 4.26. As details of the citrate-EcPNPase/hPNPase ligand-receptor interactions were

previously identified for the solved crystal structures (Section 4.3.1), the ligand-receptor interactions, identified through *in silico* docking, were easily compared.

Based on the docking calculations, similar interactions were predicted between citrate and the four putative citrate-binding residues (EcPNPase positions R93 in RBRI, R399 in RBRII, and D486 and D492, via the Mg^{2+} ion, in the MBR) that were conserved in EcPNPase, SanPNPase, hPNPase and SsoExosome (Figure 4.26, black arrows). The interactions between docked citrate and the two amino acids (EcPNPase positions H403 in RBRII and K494 in the MBR) that were only conserved in the three PNPase proteins, were comparable for the PNPases, but there was no equivalent interaction evident between citrate and the amino acids that were substituted at these positions in SsoExosome (Figure 4.26, L and P, respectively). The EcPNPase H487 residue was not conserved in the other PNPase homologs and its interaction with citrate was also not observed in the MBR of these homologs (Figure 4.26, black arrow). Within the PBR of EcPNPase interactions were frequently observed between citrate and S437 and S438, however interactions between citrate and S439 were less common. The amino acid sequence and pattern of interactions within the PBR was conserved in hPNPase (Figure 4.26). However, the substitution of threonine for serine (at the equivalent position to EcPNPase S438) in SanPNPase appeared to have altered the interactions between citrate and the positions equivalent to EcPNPaseS438 and EcPNPase S439. In the SsoExosome, EcPNPase positions S437, S438 and S439 were substituted for G, S and R (Figure 4.11, black arrows). Despite this sequence variation, significant interactions were observed between citrate and these three positions in SsoExosome (Figure 4.26, black arrows).

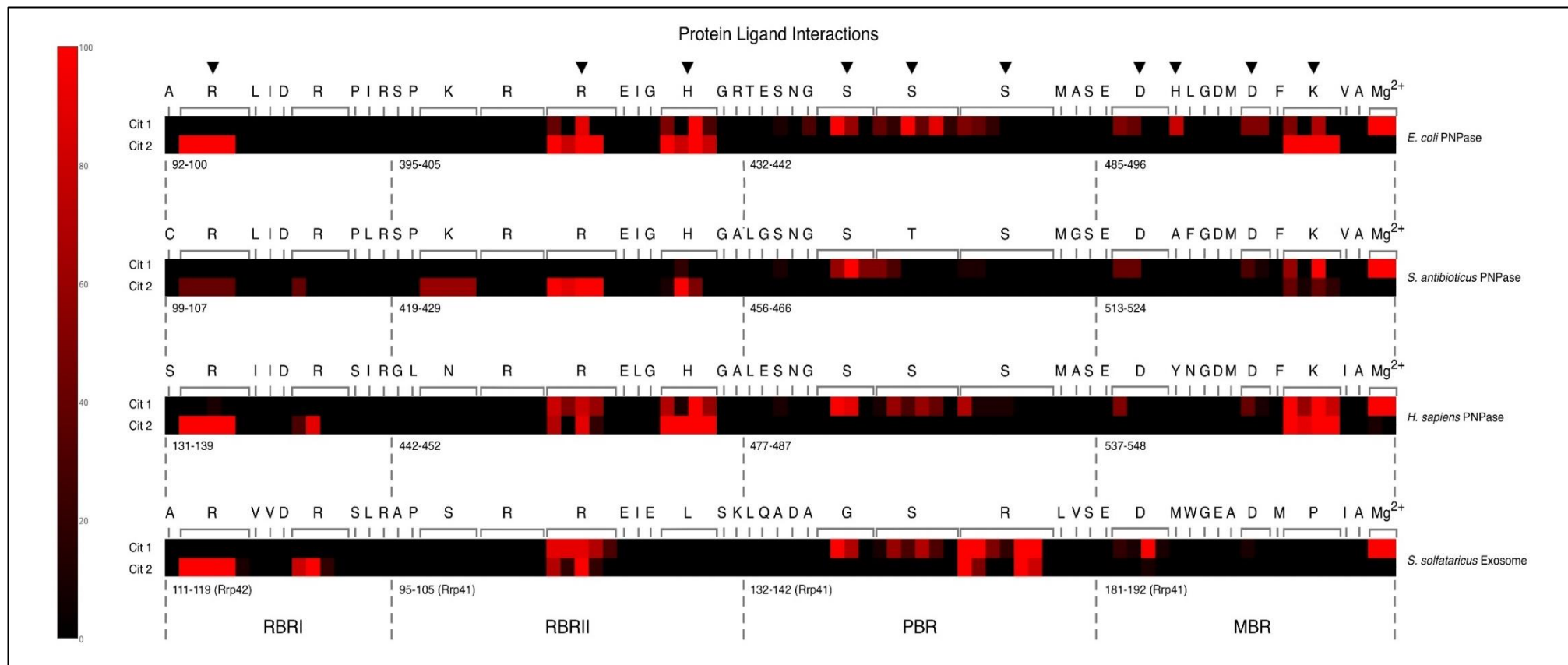


Figure 4.26 PNPase Homologs Active Site Citrate Docking; Heat Map

Heat maps summarising the PLIFs generated for the ten lowest-energy poses obtained when two molecules of citrate (Cit 1 and Cit 2) were docked into the active site EcPNPase, SanPNPase, hPNPase and SsoExosome, with an interaction frequency scale (0-100%, coloured bar black to red respectively). Residues involved in citrate and metal coordination are indicated (black arrows). Images generated in MOE, Plotly and prepared in GIMP (v2) (GIMP, n.d.; Molecular Operating Environment, 2013; Plotly Technologies Inc., 2015).

4.3.4 Crystals Diffract by X-Ray Crystallography and Require Further Optimisation

The structure of the SsoExosome utilised in *in silico* docking calculations was not co-crystallised with citrate bound, unlike the EcPNPase and hPNPase homologs. Following the MOE predictions that citrate docks into SsoExosome, interacts with similar residues and produces a similar S value to EcPNPase and hPNPase, the possibility of co-crystallising SsoExosome with citrate was examined within this section. The concentration of enzyme and citrate appropriate for crystal screening was initially determined, and then a small crystallisation trial was conducted as follows.

Initial PCT A1/A2 test results (Figure 4.27) which were used for determining the concentration of SsoExosome (+ citrate), showed no precipitate in the A1 well after 30 minutes and a heavy amorphous precipitate in A2; hence the B1/B2 test was conducted as recommended in the PCT protocol (Figure 4.2). Following this subsequent test, the B1 well was clear of precipitate after 30 minutes and a light granular precipitate was observed in B2 and this suggested the protein concentration was appropriate for setting up crystal screens.

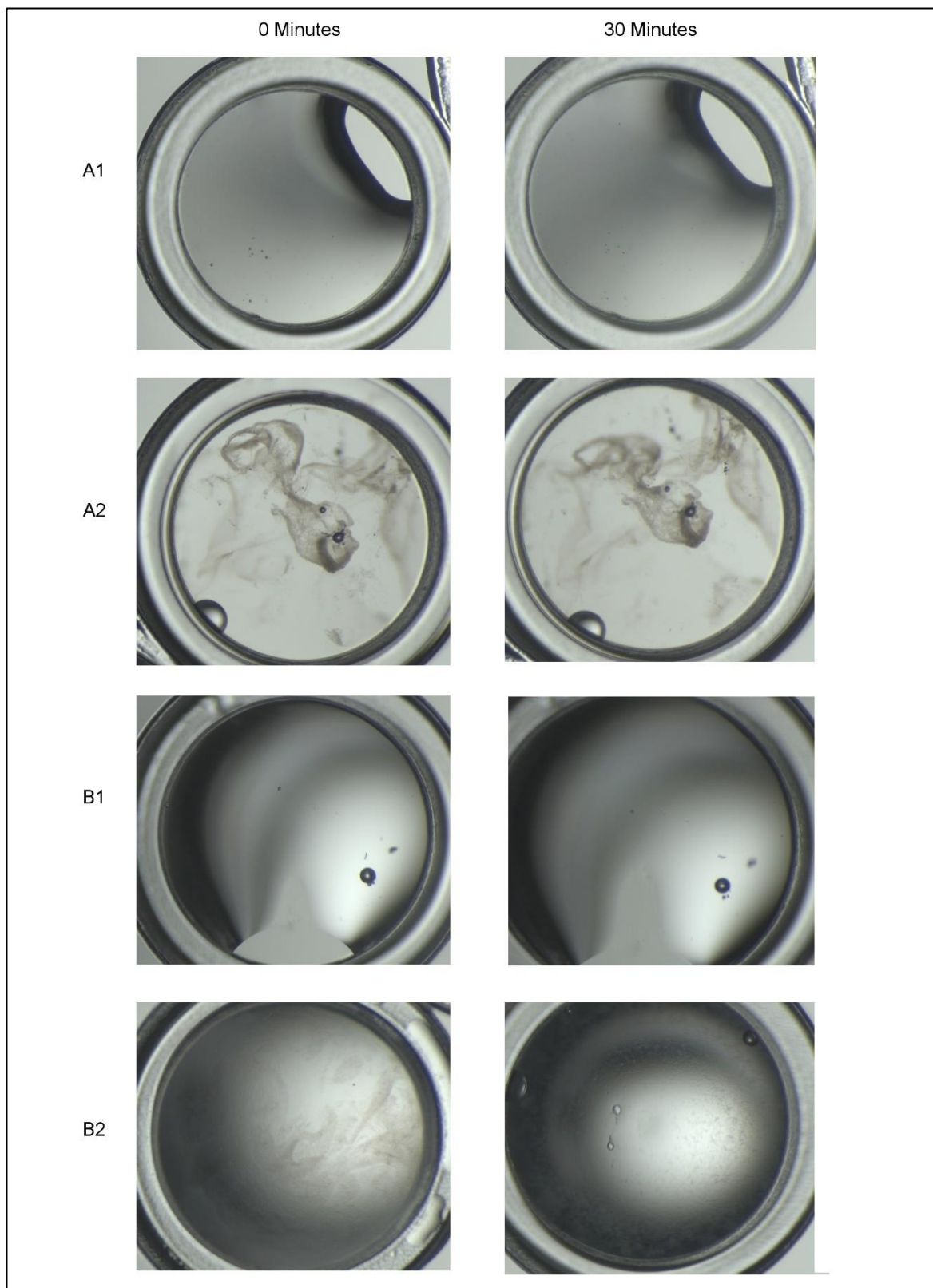


Figure 4.27 SsoExosome Crystallisation Trials; PCT Test

The wells A1-2 and B1-2 for the PCT tests are shown at the start (0 minutes) and end of the reaction (30 minutes). In the A1/A2 reaction a clear well was observed in well A1 and heavy amorphous precipitate in A2 at the end of the reaction. In the B1/B2 reaction a clear well was observed for B1 and a light granular precipitate was seen in B2. Images prepared in GIMP (v2) (GIMP, n.d.).

A subsequent, high-throughput screen using the molecular dimensions PACT premier™ screening conditions was used and this identified buffer conditions for crystal growth. Crystals grew in a protein sample containing 17.5 μ M SsoExosome and 20 mM citrate, in a crystallisation buffer

containing 0.2 M NaCl, 0.1 M Tris pH 8, 20% peg 6,000 (screen #D7) (Figure 4.28). These crystals did not grow in the control well containing buffer alone, indicating they were not salt crystals (data not shown).

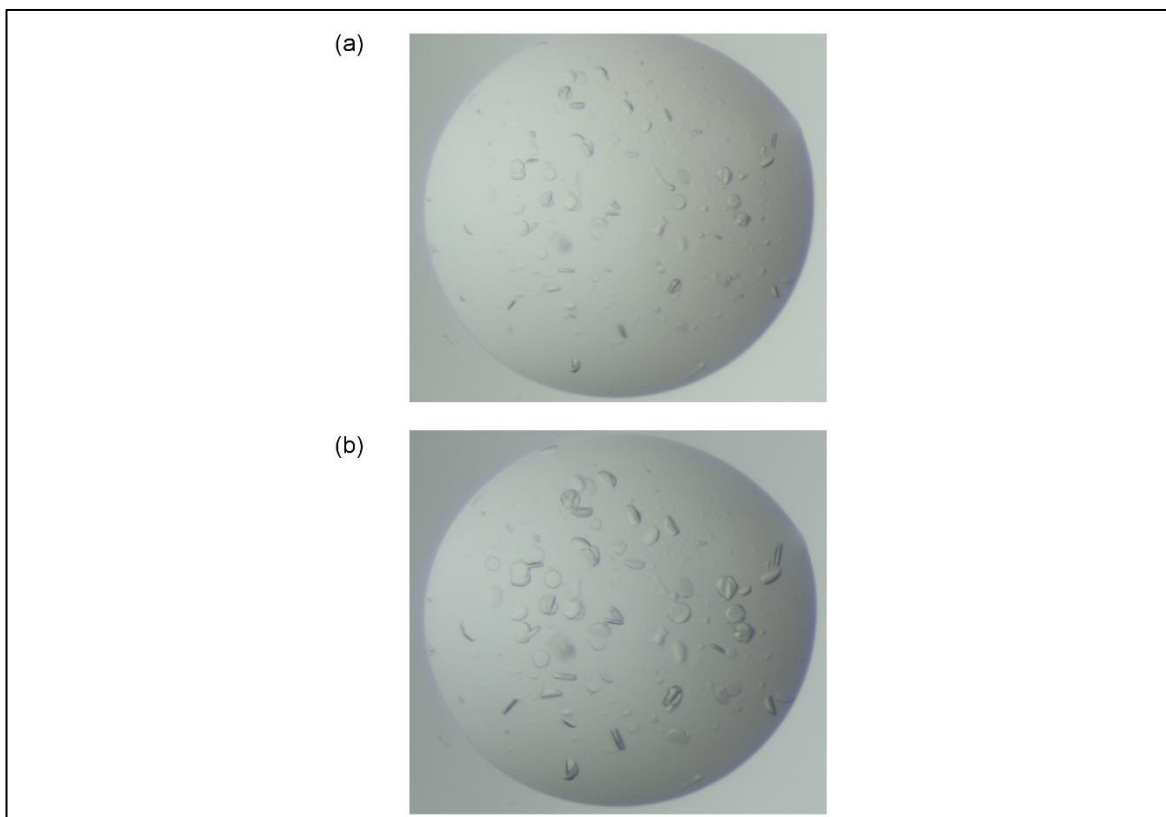


Figure 4.28 Crystals of SsoExosome + Citrate

Crystals forming after (a) 1 week and (b) 3 months with 17.5 μM SsoExosome and 20 mM citrate, in a crystallisation buffer containing 0.2 M NaCl, 0.1 M Tris pH 8, 20% peg 6,000 (PACT screen #D7). Images prepared in GIMP (v2) (GIMP, n.d.).

The crystals in Figure 4.28 were sent for remote X-ray diffraction at Diamond light source and they diffracted. The diffraction pattern indicated protein was present, possibly SsoExosome (Figure 4.29 (a), green box). However, the diffraction was too weak to solve the structure and to determine whether citrate was bound within the active site. For example, the diffraction shown in Figure 4.29 (a) (green box) for the crystal ($\sim 40 \mu\text{m}$ diameter), was less than the example data collected for the lysozyme crystal ($\sim 200 \mu\text{m}$ diameter) shown in Figure 4.29 (b). However, the conditions used for SsoExosome (+citrate) could be optimised in order to grow larger crystals and potentially improve the strength of diffraction to help solve the structure. Nevertheless, *in silico* data predicted citrate binding to SsoExosome and the *in vitro* effects of citrate were therefore investigated within the next Section 4.3.5.

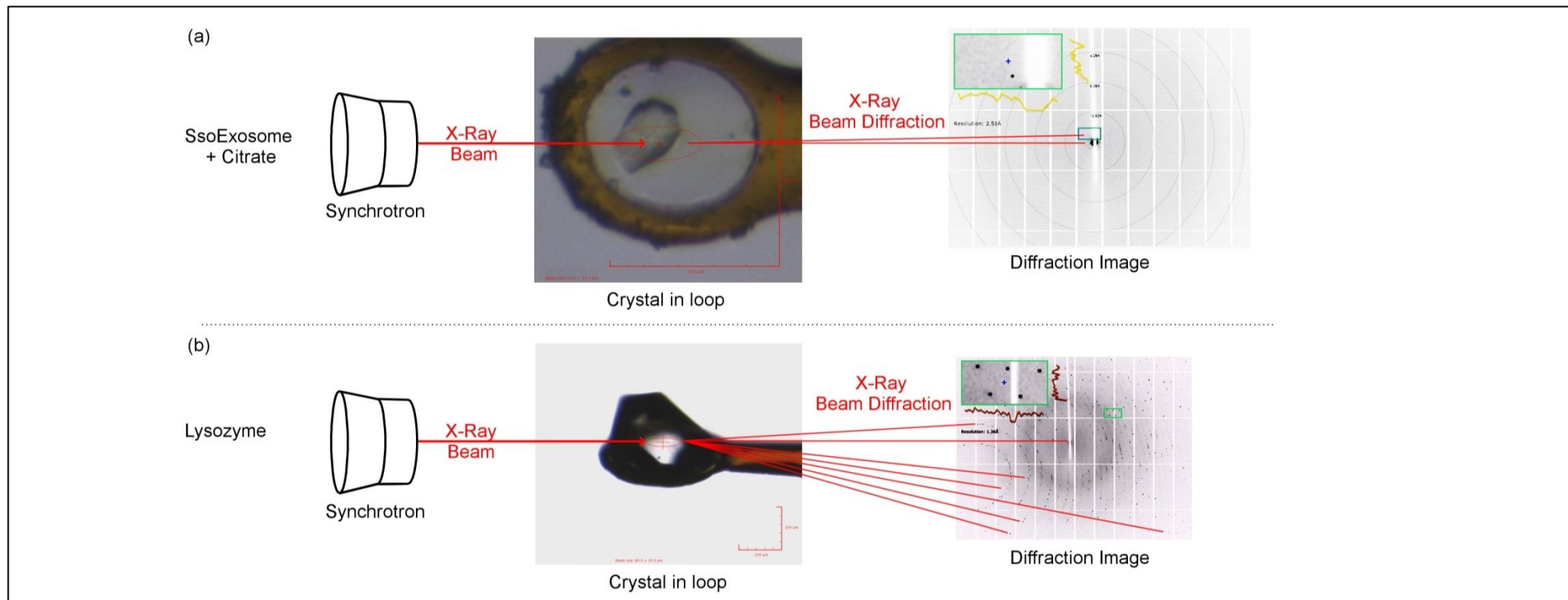


Figure 4.29 SsoExosome and Lysozyme X-Ray Crystallography Diffraction

Crystals of (a) SsoExosome + Citrate (~40 μm in diameter) and (b) Lysozyme (~200 μm in diameter) were exposed to an X-ray beam at the Diamond Light Source Synchrotron in Oxford (remote access) with a transmission of 100%, a wavelength of 0.9763 \AA and an Oscillation of 0.15°. SsoExosome + Citrate was exposed for 0.200 s to an X-ray beam size of 50 x 20 μm , whereas lysozyme was exposed for 0.030 s with a beam size of 80 x 20 μm . Low resolution diffraction (2.71 \AA) was observed for the putative SsoExosome crystal (part a; green box), but was high for the lysozyme control; with excellent diffraction at a resolution of 1.5 \AA (part b; green box). The putative SsoExosome crystal has large cell axes, indicative that the lattice is formed by a macromolecule. Images prepared in GIMP (v2) (GIMP, n.d.).

4.3.5 Modulation of PNPase Homologs' Exoribonuclease Activity by Citrate

The *in silico* studies described in Section 4.3.3 suggested that citrate-binding at the enzyme's active site may be evolutionarily conserved in bacterial and eukaryotic PNPases and the archaeal exosome. For EcPNPase and hPNPase the docking predictions were consistent with the location of co-crystallised citrate within the crystal structures (Lin *et al.*, 2012; Nurmohamed *et al.*, 2009). Furthermore, for EcPNPase, this binding has been shown to inhibit the exoribonuclease activity (Nurmohamed *et al.*, 2011). Whether citrate also inhibits other PNPases and the archaeal exosome still remained to be examined.

The EcPNPase, hPNPase and the SsoExosome used in these *in vitro* experiments were expressed in *E. coli* and purified to homogeneity as described in Chapter 3. A commercially available *Synechocystis sp.* PNPase (SspPNPase) homolog was also utilised for *in vitro* studies, partly due to the ease of availability, but principally because it had the serine/threonine substitution within the RBR, similar to SanPNPase, which was of interest to study (Figure 4.11).

The 3'-5' exoribonuclease activity of PNPase and archaeal exosome homologs on a RNA substrate was monitored over time using a gel-based assay in the absence and presence of citrate. The percentage of intact fluorescently labelled oligoribonucleotide substrate, separated by denaturing PAGE, was calculated and the effect of citrate on PNPase homolog activity was determined.

Initially, a range of concentrations of a 5'FAM labelled Poly(A)₂₀ mer (5'FAM(A)₂₀) RNA substrate were tested so that the limit of band detection could be determined. The 20% urea denaturing gel shown in (Figure 4.30) suggested that RNA assay concentrations between 560-1120 nM form a single, intense band that would be sufficient for visualisation and quantification purposes. Hence assays were designed with this in mind; 700 nM 5'FAM(A)₂₀ was used as standard in all subsequent gel-based degradation assay reactions.

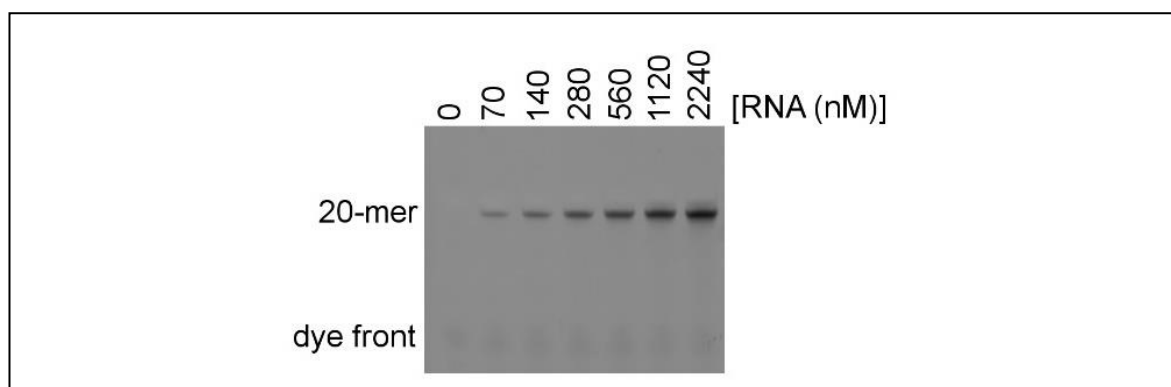


Figure 4.30 3'-5' Degradation Assay; 5'FAM(A)₂₀ RNA Substrate Concentration Range

A range of 5'FAM Poly(A)_{20mer} RNA substrate concentrations ([RNA (nM)]) were run on a 20% urea denaturing gel in 1 x TBE ((200 V for 1 hour 30 minutes at 4°C). The gel was visualised using the Fujifilm FLA-imaging system (Phosphorimager: Blue filter, Blue laser, 473 nm). An individual band of intact 20 mer

RNA is present in each lane, its intensity increased with substrate concentration (size verified by RNA ladder, data not shown). The dye front is indicated. Images prepared in GIMP (v2) (GIMP, n.d.).

In order to determine the optimal hPNPase enzyme concentration to use in the assay, a range of hPNPase concentrations (4.8, 24, 48, 120 and 240 nM) were subsequently tested with 700 nM 5'FAM(A)₂₀, as described in Section 4.2.4. As expected, sequential nucleotide removal in the 3' to 5' direction was confirmed for hPNPase against 5'FAM(A)₂₀ RNA substrate (Figure 4.31). A concentration of 240 nM hPNPase was selected as the optimal enzyme concentration to degrade RNA substrate in a suitable time frame. Thus, for comparison, all subsequent gel-based assays within this work contained 240 nM of hPNPase, EcPNPase, SsoExosome or SspPNPase unless specified. There was a dye front present in the gel shown in Figure 4.31 and therefore to improve the quantification, a 'dye free' loading buffer was utilised in subsequent assays.

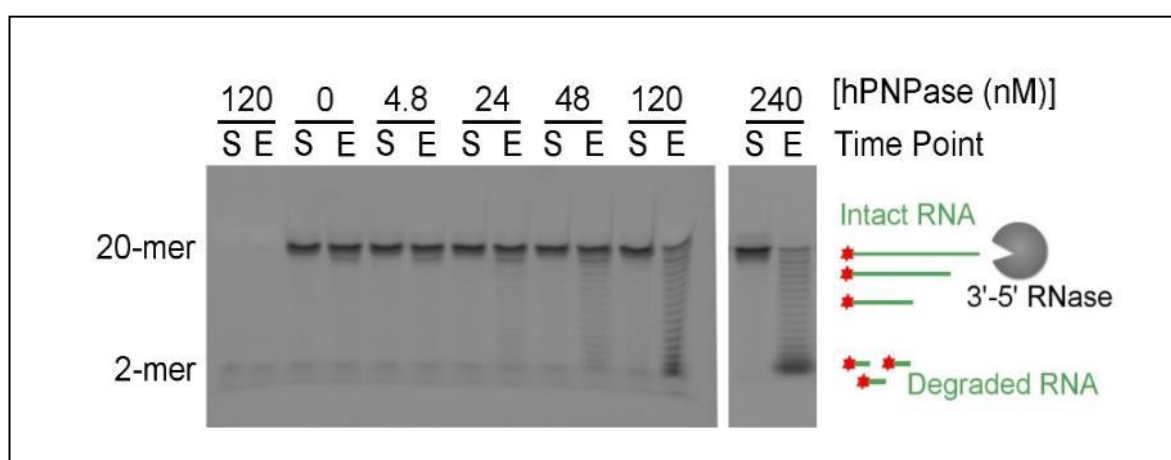


Figure 4.31 3'-5' Degradation Assay; hPNPase Concentration Range

A range of hPNPase concentrations (0, 4.8, 24, 48, 120 and 240 nM) were incubated with 700 nM 5'FAM(A)₂₀ RNA substrate at 37°C for 60 minutes. A no RNA control (120 nM hPNPase) and no protein control (0 nM hPNPase) are provided at the start (S) and end (E) of the assay. Samples were run on a 20% urea denaturing gel in 1 x TBE (200 V for 1 hour 30 minutes at 4°C) and the degradation of the RNA substrate from 20-mer to 2-mer increases with protein concentration. The gel was visualised using the Fujifilm FLA-imaging system (Phosphorimager: Blue filter, Blue laser, 473 nm). Images prepared in GIMP (v2) (GIMP, n.d.).

Following optimisation of the RNA substrate and enzyme concentration, assays were conducted in triplicate, as per Section 4.2.4 in the presence and absence of 3.75 mM citrate (equal molar concentrations to Mg²⁺) and representative gels are shown in Figure 4.32. The pixel intensity of RNA bands in Figure 4.32 were quantified using gel densitometry. The percentage of intact RNA substrate remaining at the end of the assay was determined for each enzyme, in the absence and presence of citrate and the values are provided in Table 4.5.

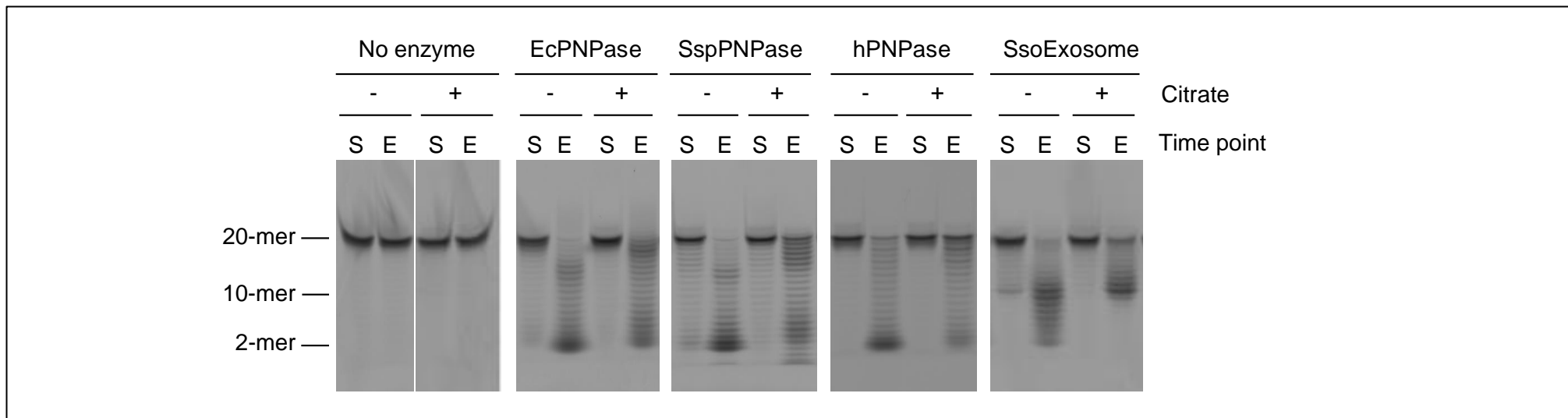


Figure 4.32 3'-5' Degradation Assay; *In Vitro* Inhibition of PNPase Homologs by Citrate

A range of PNPase homologs (240 nM) were incubated with 700 nM 5'FAM(A)₂₀ RNA substrate at 37°C. The no enzyme control, shows intact RNA with (+) and without (-) citrate (3.75 mM, equal molar to MgCl₂₊) at the start (S) and end (E, 60 minutes) of the reaction. In these assay conditions, to account for enzymatic efficiency differences, the hPNPase was incubated for 60 minutes, the SsoExosome and SspPNPase were incubated for 30 minutes and EcPNPase for 10 minutes. Samples were run on a 20% urea denaturing gel in 1 x TBE and the RNA substrate degradation products 2-20 mer are indicated. Images prepared in GIMP (v2) (GIMP, n.d.).

Enzyme	Percentage substrate remaining at assay end-point	
	- citrate	+ citrate
EcPNPase	2.2 ± 0.4	5.6 ± 1.5
SspPNPase	2.2 ± 0.2	6.9 ± 2.2
hPNPase	8.4 ± 2.8	22.8 ± 4.8
SsoExosome	4.0 ± 0.5	13.6 ± 0.5

Table 4.5 3'-5' Degradation Assay; Quantifying the *in Vitro* Inhibition of PNPase Homologs by Citrate
The percentage of substrate remaining at assay end-point in the absence (-) or presence (+) of citrate are listed for PNPase homologs. Values are the mean from at least three experimental repeats and the errors reported are the standard deviation.

As previously established in Chapter 3, all three PNPase homologs hPNPase, EcPNPase and SsoExosome exhibited 3'-5' degradation activity, degrading a 20 mer 5' FAM Poly(A) substrate to a 2 mer degradation product. The control assay showed that citrate didn't affect the RNA integrity from the start (S) to the end (E) of the assay; the RNA substrate remained intact in the no enzyme control +/- citrate. In the EcPNPase enzyme assay, the RNA was partially degraded within the time scale tested, with degradation products spanning from a 20 mer to a 2 mer. EcPNPase activity was clearly inhibited by citrate, with 2.5-fold more substrate remaining at the end of the assay in the presence of citrate (Figure 4.32; Table 4.5) and this was consistent with previous observations (Nurmohamed *et al.*, 2011). The commercially available SspPNPase, which has the same eight residues involved in citrate-binding as SanPNPase (Figure 4.11), also showed degradation activity and the presence of citrate was inhibitory. Thus suggesting that the serine/threonine mutation in the 'SSS' motif of EcPNPase had no effect on citrate mediated inhibition. These two enzymes suggested that citrate-PNPase interactions were conserved in two distinct bacterial species. Citrate also inhibited the exoribonuclease activity of hPNPase and SsoExosome to a similar degree as EcPNPase (Figure 4.32; Table 4.5). Thus suggesting that citrate regulates these more evolutionarily distinct enzymes.

In the presence of citrate, the limit products produced by the SsoExosome were 10-mers rather than the 2-mer obtained in the absence of citrate (Figure 4.32). Stalling of the exosome at this position was also observed in the absence of citrate and has been previously reported for the *A. fulgidus* and *P. abyssi* exosomes for short oligoribonucleotides that can no longer be bound at both the active site and the entrance to channel (Hartung *et al.*, 2010; Navarro *et al.*, 2008). Supportively, the channel size has been shown to accommodate up to ~ 9 nts of substrate (Lorentzen & Conti, 2005). Citrate-binding to the active site could occlude the only RNA-binding site possible for short oligoribonucleotides and may explain our observation of more of the stalled product.

4.3.6 Sequence Analysis of Citrate Binding Residues

The results so far strongly suggested that citrate-mediated inhibition of PNPase and exosome activity is a commonly exploited regulatory mechanism in all three domains of life. To assess how widely this mechanism is conserved, a more detailed bioinformatics sequence analysis compared to that previously reported in Section 4.3.2 was undertaken.

The conservation of individual residues, within the four main binding motifs (highlighted in Figure 4.11), were compared across 3,509 prokaryotic PNPase, 252 eukaryotic PNPase and 69 Rrp41/Rrp42 exosome protein sequences. The resulting sequence logos of active site RBRI, RBRII, PBR and MBR motifs are shown in Figure 4.33 and the full consensus sequence list is provided in the Appendix 9.8. Of the ten Mg²⁺ and citrate-binding residues that were originally identified in EcPNPase (R93, R399, H403, S437, S438, S439, D486, H487, D492 and K494), eight were absolutely conserved in all of the bacteria and eukaryotes examined (Figure 4.33 (a) and (b)). Only positions corresponding to EcPNPase S438 in the PBR and H487 in the MBR showed variation. There was a conserved serine/threonine substitution at the position of S438 across prokaryotes and eukaryotes, only these two residues have been an acceptable mutation over their evolution. The number of potential substitutions of H487 increased for prokaryote and eukaryote, suggesting that this residue is less evolutionarily conserved. Nevertheless, the docking and activity results for EcPNPase, SanPNPase/SspPNPase and hPNPase indicated that the sequence variation at these positions was not critical for citrate-binding or citrate-mediated regulation. Therefore, the alignments predict that all PNPases may bind and be regulated by citrate.

The sequence logos for the exosomes (Figure 4.33 (c)) confirmed the findings from our initial alignment which contained just four exosome sequences (Figure 4.11). Only four of the Mg²⁺ and citrate-binding residues identified in EcPNPase were absolutely conserved in archaea (R93 in RBRI, R399 in RBRII and D486 and D486 in MBR). As for the PNPases, variation was observed at the non-critical EcPNPase H487 position in the MBR, in the EcPNPase H403 position in RBRII and the EcPNPase K494 position in the MBR. Most strikingly, the S(S/T)S motif in the PBR of PNPases appears to have been replaced by a G(T/S)R motif in the archaeal exosomes.

Despite this sequence variation, this study has clearly shown for the SsoExosome that citrate still binds to the archaeal exosomes and modulates its activity *in vitro*. Taken altogether these results suggest that all of the PNPases and archaeal exosomes examined may bind to and be regulated by citrate.

Protein sequence alignments used for generating sequence logos shown in Figure 4.33 were also analysed to investigate the citrate-binding site observed at the vestigial site of EcPNPase and the resulting alignment is shown in Figure 4.34. As seen in Section 4.3.2.2, sequence conservation at this site was much weaker than for the active site, providing more support for not investigating this site further.

In summary sequence analysis of citrate binding residues predicted that the citrate-mediated regulation of PNPases and archaeal exosome may represent a common regulatory strategy across evolution.

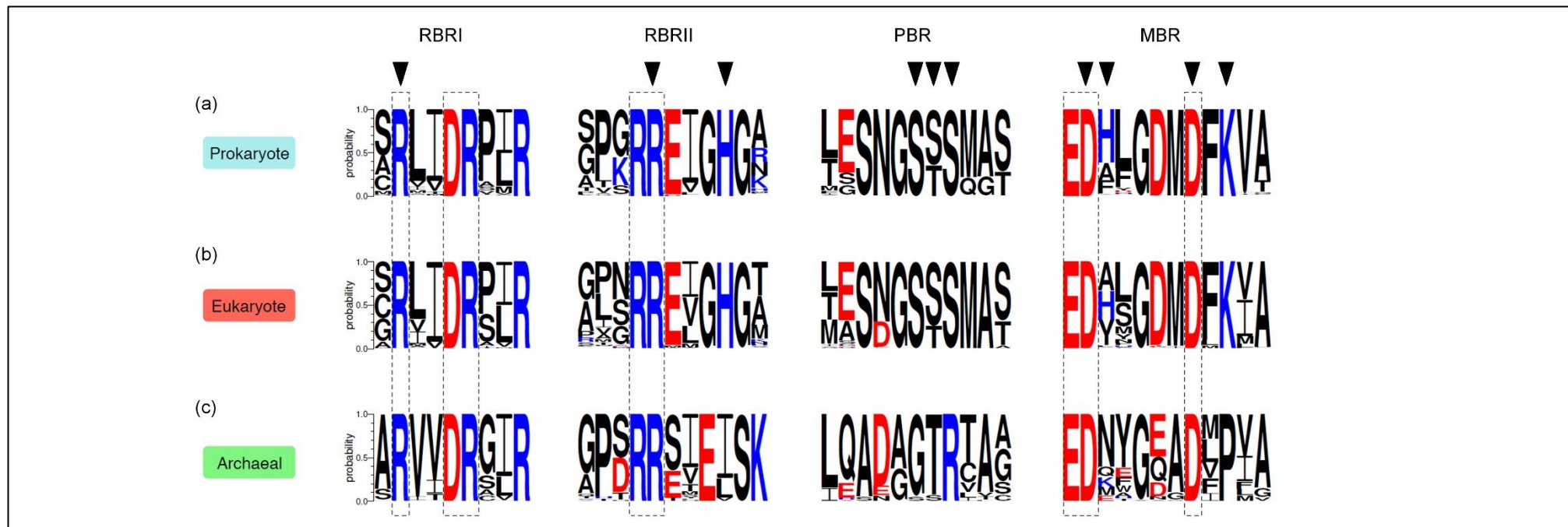


Figure 4.33 Conservation of PNPase Homologs Binding Motifs Sequence Logos

Sequence logos for (a) 3,509 Prokaryote PNPases, (b) 252 eukaryote PNPases and (c) 69 archaeal exosomes sequence are provided, highlighting the protein sequence conservation of the RBRI, RBRII, PBR and MBR motifs (x axis) with a probability score (y axis). Uncharged residues are shown in black, with positively and negatively charged residues coloured blue and red respectively. Dotted boxes represent residues which are conserved across all three domains. Arrows indicate the residues previously discovered to interact with citrate in E. coli PNPase in (Nurmohamed *et al.*, 2011). Images generated in Weblogo3 and prepared in GIMP (v2) (GIMP, n.d.).

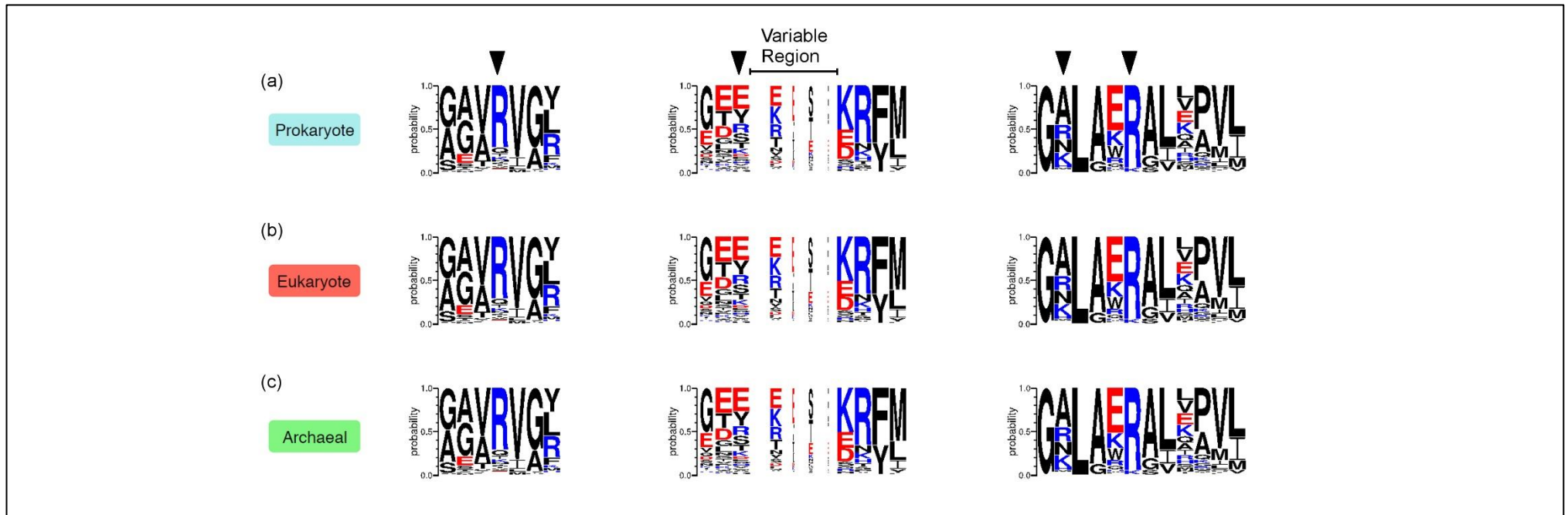


Figure 4.34 PNPase Homologs Vestigial Site Sequence Logos

Sequence logos for (a) 3,509 Prokaryote PNPases, (b) 252 eukaryote PNPases and (c) 69 archaeal exosomes sequence are provided, vestigial site motifs (x axis) were selected to show residues around arginines previously shown to bind citrate at the vestigial site (arrows) (Nurmohamed *et al.*, 2011). The probability score is shown for each residue (y axis), with the exception of the variable region, where a large number of insertions and deletions are present over evolution of all domains. Uncharged residues are shown in black, with positively and negatively charged residues coloured blue and red respectively. Images generated in Weblogo3 and prepared in GIMP (v2) (GIMP, n.d.).

4.4 Conclusion

The results described in Section 4.3.1 firstly investigated in detail the citrate-PNPase interactions previously documented in the literature. These results, describing the residues involved in citrate binding for EcPNPase were useful since they allowed comparisons to be made between the known crystal structures and docking predictions, essentially facilitating validation of the docking process. These citrate-PNPase interactions were then discovered to be conserved across a range of PNPase and archaeal exosome homologs, through basic protein sequence alignments in Section 4.3.2. Results from *in silico* molecular docking calculations were then presented in Section 4.3.3, which initially validate the methodology, using available EcPNPase and hPNPase structures with citrate bound, prior to predicting that two citrate molecules dock into SanPNPase and SsoExosome homologs active sites. Results of crystal trials of SsoExosome soaked with citrate diffracting X-Rays was then provided in 4.3.4, although the diffraction was only weak, the results suggested an opportunity for optimisation. Subsequently, data obtained from *in vitro* experiments in Section 4.3.5 then suggested that citrate inhibits the 3'-5' phosphorolytic degradation activity of all homologs tested (SspPNPase was utilised in place of SanPNPase). Whether one or two citrate molecules are required for inhibition remains to be revealed (Nurmohamed *et al.*, 2011). Finally, a more detailed sequence analysis, provided in Section 4.3.6, suggested that the citrate binding residues are widely conserved and therefore the citrate-mediated regulation of PNPases and archaeal exosome reported within this chapter may represent a common regulatory strategy across evolution.

In summary, this chapter provides a variety of *in silico* and *in vitro* evidence that suggests a communicative link may exist between citrate and evolutionarily divergent PNPase homologs. As highlighted in Figure 4.35 the interaction between citrate and 3'-5' exoribonucleases could modulate the RNA pool and affect post-transcriptional gene regulation. This supports the model where central metabolism acts as another tier in the central dogma flow of cellular information. Whether other metabolites can affect PNPase homologs in a similar manner to citrate is investigated next in Chapter 5.

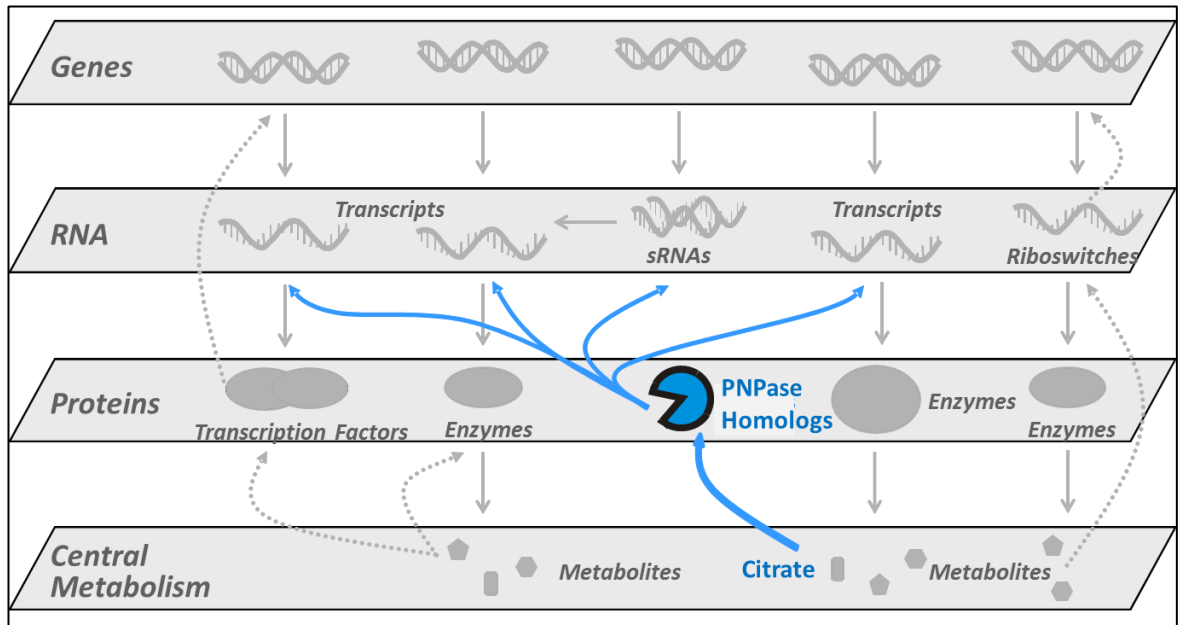


Figure 4.35 Citrate-PNPase Communication Summary

The flow of cellular information. The potential interaction of citrate with PNPase homologs is shown, linking central metabolism to post-transcriptional gene regulation. Image adapted from Prof. Anastasia Callaghan.

5 Metabolites Other Than Citrate May Modulate the Phosphorolytic Activity of Polynucleotide Phosphorylase from *Homo sapiens*

The first objective of this PhD study was to explore the conservation of citrate-PNPase interactions. Combined evidence from molecular docking calculations, bioinformatics studies and gel-based degradation assays supported the hypothesis that attenuation of PNPase activity by citrate may be widespread across evolution. More specifically, the use of the docking program MOE, as a tool for predicting regulatory molecules, was initially validated. The residues known to be involved in citrate-PNPase interactions, were discovered to be highly conserved and, using MOE, citrate was predicted to dock into evolutionarily distinct PNPase homologs. Furthermore, citrate was shown to affect the 3'-5' degradation activity of PNPase homologs from all three domains of life, including prokaryotes, eukaryotes and archaea. Whether this represented a unique role for citrate, or whether other metabolites of the TCA cycle can also interact with PNPase homologs and affect activity, remains to be determined.

Accordingly, the second objective of this PhD study, and the focus of this chapter, was to investigate whether other metabolites may also play a role in regulating RNA turnover. As highlighted in Chapter 1, the pool of available metabolites and the evolution of the TCA cycle varies considerably between organisms and this can make comparative studies particularly challenging. Since both PNPase and the TCA cycle metabolites are located within the mitochondria of *H. sapiens* the interaction between metabolites and hPNPase was the main focus of this chapter. This is potentially important since research suggests that disturbance of RNA turnover within the mitochondria is linked to aging, cancer and neurodegenerative diseases in humans (Borowski, Szczesny, Brzezniak, & Stepień, 2010).

A similar approach to Chapter 4, of combining *in silico* molecular docking with *in vitro* gel-based assays, was applied within this chapter. The binding and effect of metabolite-PNPase interactions were first predicted using MOE and then quantified using gel-based assays respectively. Metabolites which were predicted to bind hPNPase with greater affinity than citrate, and were shown to effect hPNPase degradation activity more than citrate, were then examined more closely. In order to identify the key functional groups/features of these metabolites, that may be particularly important for PNPase regulation, a pharmacophore was generated using MOE. This process essentially predicted metabolite (ligand) features, which ensured optimal *in silico* interactions with hPNPase (receptor) (as reviewed in Langer & Wolber, 2004). More specifically, by using MOE, 'classical' pharmacophoric features like H-bond acceptors (Acc) and donors (Don), hydrophobic (Hyd) and/or aromatic rings (Aro), together with geometrical constraints like distances, angles, and dihedral angles were calculated from a collection of metabolite-hPNPase poses (Molecular

Operating Environment, 2013). This set of features, termed the pharmacophoric ‘model’ or ‘hypothesis’ (Langer & Wolber, 2004), was then used to predict other metabolites which may regulate hPNPase activity. To address the question of whether metabolites play a role in regulating RNA turnover in humans, a variety of *in silico* and *in vitro* experiments were conducted within this chapter.

5.1 Chapter Aims

The previous chapters describe the methods and techniques used to predict the conservation of citrate binding residues *in silico*, and determine the effects of citrate on recombinant PNPase homolog activity *in vitro*. Similarly, this results chapter utilises *H. sapiens* and *E. coli* PNPase enzyme structures, previously prepared for docking in MOE, for molecular modelling studies (Chapter 4), and purified recombinant hPNPase and EcPNPase (Chapter 3) for *in vitro* gel-based assays. As a result, the methods described in Section 5.2 build on those previously provided in Chapter 3 and 4.

Key questions addressed within this work examine whether, like citrate, other metabolites of the TCA cycle are predicted to bind hPNPase *in silico*. Additionally, whether these metabolites affect hPNPase degradation activity will be determined *in vitro*. Through generating a pharmacophore model, key features of any identified metabolite-PNPase interaction will be examined in more detail. Finally, this pharmacophore model will be utilised to predict other metabolites which may regulate the activity of hPNPase. The resulting data are provided in Section 5.3 and important conclusions regarding a metabolite-PNPase communicative link in the higher eukaryotic organism *H. sapiens* are then summarised in Section 5.4.

5.2 Methods

Collectively the methods outlined within previous chapters were used as standard, and any additional information utilised for this chapter specifically, is detailed below.

5.2.1 *In-Silico* Molecular Docking

All the computational procedures, for docking into PNPase from *H. sapiens*, was carried out using MOE as summarised in the flow diagram shown in Chapter 4 (Figure 4.1) (Molecular Operating Environment, 2013). Details of individual parameters for docking ‘user defined’ metabolite databases, generating PLIFs/heat maps and generating pharmacophore consensus models, are described in Sections 5.2.1.1-5.2.1.3 respectively.

5.2.1.1 *In-Silico* Metabolite Database Docking

The structures of a selected range of metabolites tested (Appendix 9.9) were downloaded from either the ZINC or PubChem database (Table 5.1) and saved as individual MOL2 files in a biologically relevant conformation (Irwin, Sterling, Mysinger, Bolstad, & Coleman, 2012). All

ligands were then saved in one of two MOE ‘user defined’ database files; allowing them to be docked sequentially in larger docking runs (termed TCA_Metabolite.mdb and Additional_Metabolites.mdb respectively). These bespoke databases were then docked into the appropriate ligand binding site using the MOE docking parameters previously defined (Chapter 4). Specifically, the PNPase receptor was set to the prepared hPNPase (3U1K_Dock) active-site pocket and the ligand was set as one of the specific databases (mdb file) mentioned above. In summary docking calculations were done with Amber12: EHT forcefield, triangle matcher placement (with 300 poses), and scored according to the MOE default London dG scoring function. The 300 docking poses were then subjected to a forcefield refinement with no second rescoring and 30 unique poses for each compound were retained. The docking score (*S*) for each receptor-ligand complex was then calculated using the Generalized Born solvation model (GB/VI). The lowest-energy scoring poses for each ligand were selected and the MOE binding score (*S*) values (kcal/mol) were plotted using the Minitab (v17) Boxplot function (Minitab, 2010).

Database	Compound	Identification Code	
TCA_Metabolites.mdb	Acetyl-CoA	ZINC85552088	*
	Citrate	ZINC00895081	*
	Cis-Aconitate	ZINC03860972	*
	D-Isocitrate	ZINC00895176	*
	α Ketoglutarate	ZINC01532519	*
	Succinyl-CoA	ZINC96014495	*
	Succinate	ZINC00895030	*
	Fumarate	ZINC03860193	*
	Pyruvate	ZINC01532517	*
	D/L-Malate	ZINC00895074	*
	Oxaloacetate	ZINC01532521	*
Additional_Metabolites.mdb	A	ZINC2169830	
	AMP	ZINC3860156	
	ADP	ZINC12360703	
	ATP	ZINC03871615	
	ppppA	ZINC18456332	
	cAMP	ZINC53683809	
	c-di-AMP	PubChem11158091	
	G	ZINC1550030	
	GMP	ZINC2159505	
	GDP	ZINC8215481	
	GTP	ZINC53684323	*
	ppppG	ZINC30724813	*
	ppGpp	ZINC30724089	*
	pppGpp	ZINC83923877	
	cGMP	ZINC13545961	
	c-di-GMP	ZINC33882399	
	C	ZINC2583632	
	CMP	ZINC16546001	
	CDP	ZINC12495268	
	CTP	ZINC19850119	
	cCMP	PubChem19236	
	U	ZINC2583633	
	UMP	ZINC2123545	
UDP	ZINC12493522		
UTP	ZINC19796107		
cUMP	PubChem3081385		
CoA	ZINC87493113		

Table 5.1 Metabolite Database Identification Codes

The Zinc and PubChem database codes for all the metabolites docked *in silico* into hPNPase are listed (Irwin *et al.*, 2012). The metabolites that were used for *in vitro* gel-based assays are indicated with an asterisk.

5.2.1.2 Protein Ligand Interaction Fingerprints (PLIF) and Heat Map Generation

The top 10 lowest-scoring poses for hPNPase docked with either acetyl-CoA or succinyl-CoA were selected. In order to generate a representation of the three-dimensional molecular interactions, Protein Ligand Interaction Fingerprints (PLIF) were generated using MOE (Singh, Deng, Narale, & Chuaqui, 2006). The resulting PLIF data, for each metabolite, were then exported as bit scores into Microsoft Excel (2010). The predicted frequency of acetyl-CoA or succinyl-CoA, interacting with

hPNPase, was then calculated. Finally, the percentages for each ligand were presented as heat maps using the online Plotly graph software (Plotly Technologies Inc., 2015).

5.2.1.3 Pharmacophore Generation

The top 10 lowest-scoring poses for acetyl-CoA and succinyl-CoA were selected and the PLIFs generated previously (Section 5.2.1.2) were combined to generate a pharmacophore consensus query. The maximum radius of each pharmacophore feature was set to the default size of 3 Å. The scheme, which defines the set of attributes used to construct a ligand annotation point, was set to default unified scheme; a simple annotation showing Donor (Don), Acceptor (Acc), Hydrophobic (Hyd) and Aromatic (Aro) features was selected (Molecular Operating Environment, 2013). In the Pharmacophore calculations, MOE used these ‘classical’ features (Langer & Wolber, 2004), together with geometrical constraints like distances, angles, and dihedral angles (Molecular Operating Environment, 2013), to generate a consensus model for the interactions between acetyl-CoA/succinyl-CoA and hPNPase. The resulting pharmacophore features were then visually represented as mesh spheres within the active site of hPNPase. By comparing the orientation of acetyl-CoA and succinyl-CoA poses within the pharmacophore, key functional groups involved in hPNPase binding, were predicted.

5.2.2 Gel-Based Degradation Assays

In order to analyse 3'-5' phosphorolytic exoribonuclease activities in the presence of metabolites, gel electrophoresis degradation assays were conducted as detailed in Chapter 4. The concentrations of metabolites used, were the same as citrate (3.75 mM) and were equal molar to the magnesium concentration (3.75 mM). This allowed the level of inhibition, for metabolites listed in Table 5.1 (asterisk) to be compared to citrate. This was important as 3.75 mM citrate had been previously suggested to attenuate PNPase activity in Chapter 4.

5.3 Results

This section firstly provides results from *in silico* molecular docking calculations which predict that TCA metabolites acetyl-CoA and succinyl-CoA bind the active site of PNPase from *H. sapiens* with higher affinity than citrate (Section 5.3.1). Further analysis of *in silico* data predicted that these metabolites also bind hPNPase using the same conserved residues known to be involved in citrate binding. The effects of TCA metabolites on hPNPase activity were then investigated using *in vitro* gel-based assays in Sections 5.3.1-5.3.2. The results of which demonstrated that acetyl-CoA and succinyl-CoA affect hPNPase activity more than citrate. Following the observation that these metabolites bind hPNPase using the same highly conserved residues as citrate, the effects of these metabolites on the PNPase homolog from *E. coli* was examined in Section 5.3.4. As acetyl-CoA and succinyl-CoA were found to affect the activity of both hPNPase and EcPNPase, a potential TCA metabolite-PNPase communicative link was proposed. The features that may be particularly important in hPNPase regulation were examined further in Section 5.3.5, by generating

a pharmacophore model using MOE. The nucleotide part of CoA for both these TCA metabolites was predicted to mimic the combined interactions of the two molecules of citrate; essentially binding and occluding the hPNPase active site in a similar manner. Following this observation, it was predicted that other phosphate containing metabolites including a range of nucleotides and signalling molecules, may also dock into the active site of hPNPase and this was investigated in Section 5.3.6. Interestingly, phosphate-rich metabolites GTP, ppppG and ppGpp, that were predicted to bind to hPNPase *in silico* with tighter affinity than citrate, were also shown to affect both hPNPase and EcPNPase activity in Section 5.3.7. In summary, results presented within this section suggest that a metabolite-PNPase communicative link may exist in prokaryotes and eukaryotes.

5.3.1 Some TCA Metabolites Are Predicted *In Silico* to Dock into PNPase from *H. sapiens*

The TCA cycle metabolites highlighted in bold in Figure 1.12 (Chapter 1) and listed in Table 5.1, were docked as a database into the active site of hPNPase (3U1K_Dock), using the docking module of MOE and the parameters outlined previously in Section 5.2.1.1. The top 30 lowest-energy poses were retained and the S values were presented as box plots in Figure 5.1 (a). The S values previously calculated in Chapter 4 for citrate-hPNPase docking were also shown in Figure 5.1 (a) (asterisk) for comparative purposes.

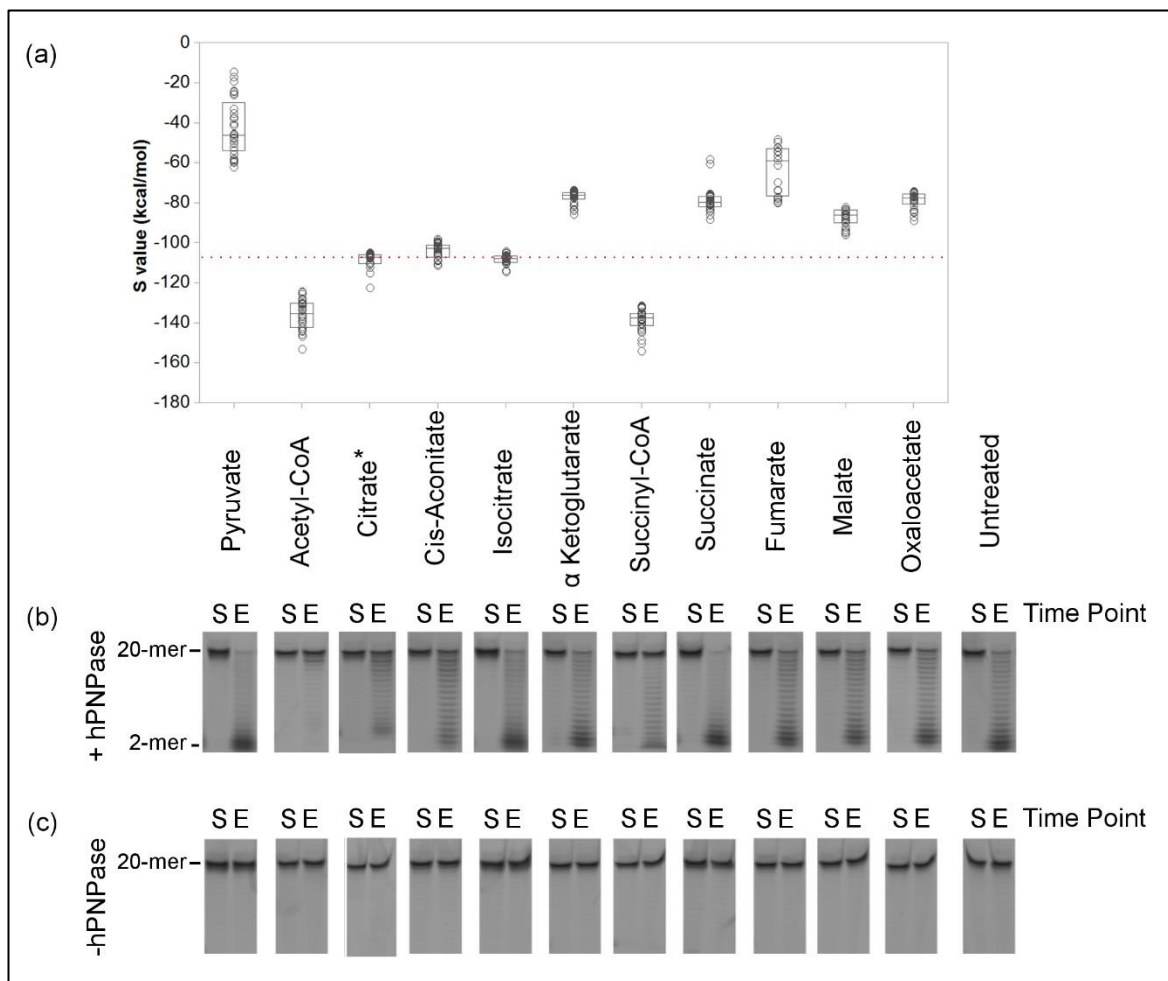


Figure 5.1 *In Silico* Docking and *In Vitro* Degradation Activity of hPNPase with TCA Metabolites

(a) The 30 unique, lowest-energy poses, docking scores (S values, kcal/mol) are provided for TCA metabolite ligands docked into hPNPase (3U1K_Dock), graph generated in Minitab (v17) (Minitab, 2010). For comparison, citrate-hPNPase S values (asterisk) are also plotted and the average S value for citrate 1 is indicated with a red dotted horizontal line. (b) hPNPase (240 nM) was incubated with 700 nM 5'FAM(A)₂₀ RNA substrate, in the presence of 3.75 mM of each TCA metabolite, at 37°C. Samples were run on a 20% urea denaturing gel in 1 x TBE and the RNA substrate degradation products between 2-20 mer are indicated. The samples taken at the start (S , 0 minutes) and end (E , 60 minutes) of each reaction are indicated. A representative gel for citrate (asterisk), taken from Chapter 4, is also shown. (c) The no enzyme control, shows intact RNA in the presence of 3.75 mM of each metabolite and in the untreated sample. Image prepared in GIMP (v2) (GIMP, n.d.).

The range of S values recorded in Figure 5.1 (a), predicted that individual TCA metabolites may interact with hPNPase with different binding affinities. For example, the S values for pyruvate, α ketoglutarate, succinate, fumarate, malate and oxaloacetate were all higher than the average S value calculated for citrate 1 docking (above red line in Figure 5.1 (a)). This suggested that citrate may bind the active site of hPNPase with higher affinity than these TCA metabolites. More specifically, the average S value of the top 10 metabolite-hPNPase poses for either pyruvate, α -ketoglutarate, succinate, fumarate, malate or oxaloacetate were -56, -80, -84, -73, -92 and -83 kcal/mol respectively, whereas the average S value for the first round of citrate docking was -112 kcal/mol. In contrast, the S values calculated for cis-aconitate (-109 kcal/mol) and isocitrate (-111 kcal/mol) were similar to the average S value calculated for one molecule of citrate docked (red line in Figure 5.1 (a)). Interestingly, only acetyl-CoA and succinyl-CoA were predicted to bind to hPNPase with an average S value (both -145 kcal/mol) lower than citrate (below red line in Figure 5.1 (a)). This

observation suggested that these two metabolites may bind hPNPase with higher affinity than citrate.

In order to understand these TCA metabolite-PNPase interactions further, the lowest-binding energy conformation for hPNPase docked with either acetyl-CoA, succinyl-CoA or two molecules of citrate were compared. It was clear from the results shown in Figure 5.2 that, like citrate, both acetyl-CoA and succinyl-CoA were predicted to bind within the active site of hPNPase. It was interesting to note that both TCA metabolites were docked in the location where the magnesium metal ion, required for catalysis, is located (Figure 5.2 (b-c)). Furthermore, they occupied the same Cit 1 and Cit 2 sites where two molecules of citrate are known to bind (Figure 5.2 (a)).

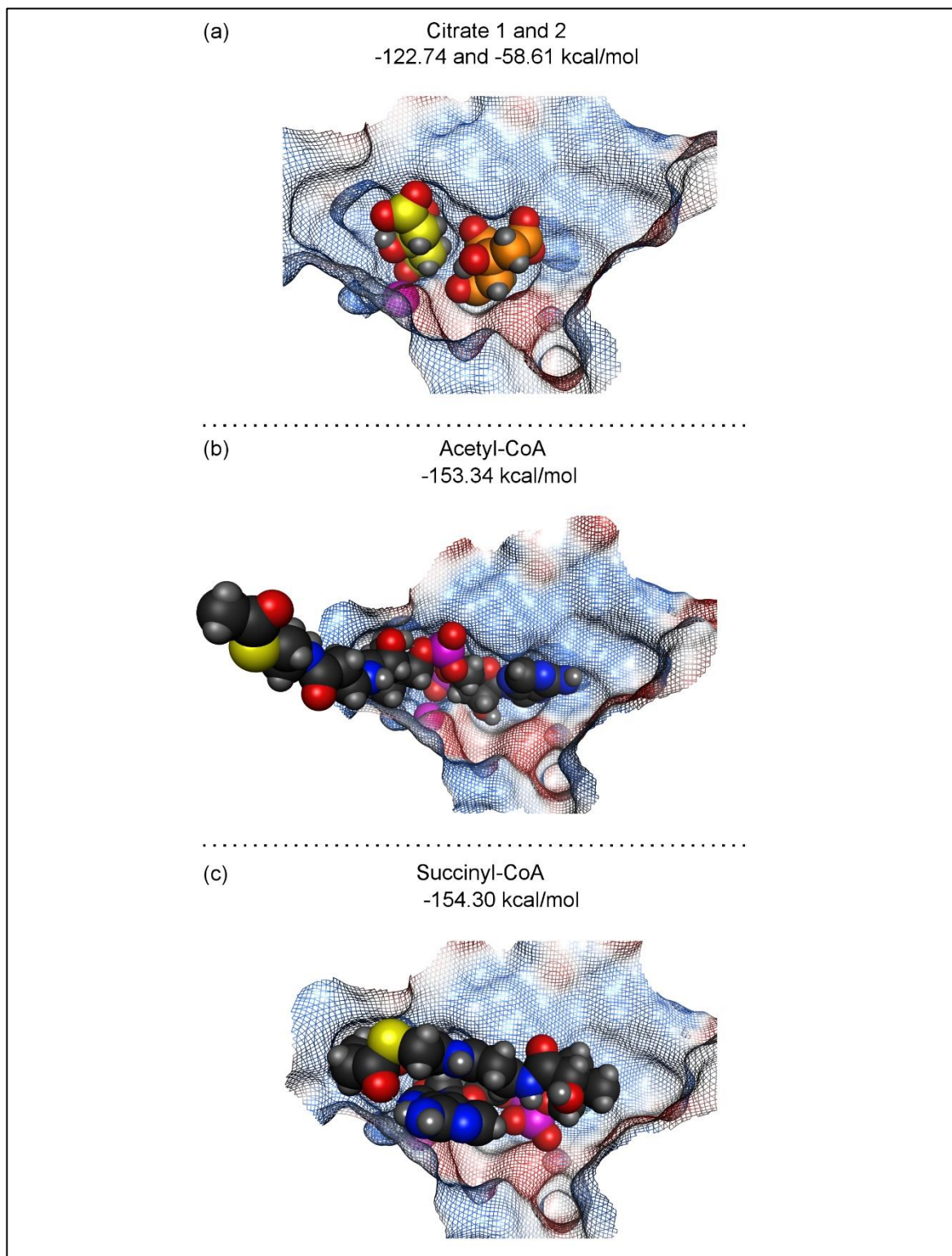


Figure 5.2 Lowest Predicted Binding Conformation of Citrate, Acetyl-CoA and Succinyl-CoA in the hPNPase Active Site.

The active site of hPNPase (3U1K_Dock) is coloured blue and red based on their residues positive and negative electrostatic charge respectively. The docked metabolites are shown in spheres for (a) two molecules of citrate, (b) one acetyl-CoA molecule and (c) one succinyl-CoA molecule, with the top predicted S value docking score (kcal/mol). Images generated in MOE and prepared in GIMP (v2) (GIMP, n.d.; Molecular Operating Environment, 2013).

In order to predict the residues involved in metabolite binding, the top ten poses for acetyl-CoA and succinyl-CoA docked into hPNPase were analysed further. A PLIF was generated for each metabolite-hPNPase interaction and the information calculated for the top ten poses were

summarised in a heat map. The heat maps for acetyl-CoA and succinyl-CoA interacting with hPNPase were shown in Figure 5.3 alongside the results previously determined for citrate (Chapter 4). For comparative purposes, the interactions predicted for the first and second round of citrate docking were combined in one heat map in Figure 5.3.

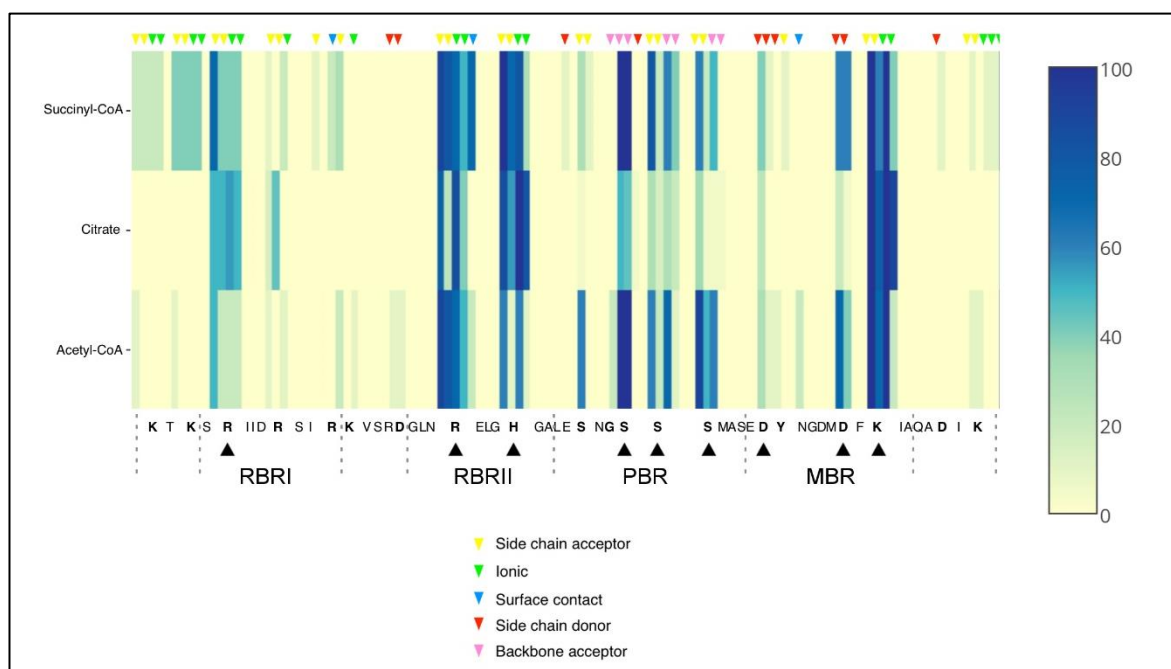


Figure 5.3 Metabolite-hPNPase Heat Map

A heat maps summarising the PLIF data generated for the ten lowest-energy poses obtained when two molecules of citrate (Cit 1 and Cit 2), acetyl-CoA or succinyl-CoA were docked into the active site of hPNPase. The interaction frequency (0-100%) was coloured as per the scale bar (from cream to blue respectively). Residues involved in citrate and metal ion coordination, within the RNA I (RBRI), RNA II (RBRII), Phosphate (PBR) and Metal Binding Regions (MBR) are indicated (black arrows). Images generated in MOE, Plotly and prepared in GIMP (v2) (GIMP, n.d.; Molecular Operating Environment, 2013; Plotly Technologies Inc., 2015).

The heat maps presented within Figure 5.3 predicted that the residues of hPNPase involved in binding all three TCA metabolites were similar. Additionally, the types of interactions were also conserved. For example, the ionic and side chain acceptor interactions for R132 within the RNA Binding Region I (RBRI), and residues R446 and H450 within the RNA Binding Region II (RBRII), were similar for citrate, acetyl-CoA and succinyl-CoA (Figure 5.3, black arrows). Furthermore, backbone acceptor interactions of S482 and backbone/sidechain acceptor interactions of S483 and S484 within the Phosphate Binding Region (PBR) were present for all three metabolites (Figure 5.3, black arrows). The D538 and D544 residues (side chain donors) and K546 (ionic and sidechain acceptor interactions) within the Metal Binding Region (MBR) also interacted with the TCA metabolites in a similar manner (Figure 5.3, black arrows).

In summary, TCA metabolites were predicted to interact with hPNPase with different binding affinities and only acetyl-CoA and succinyl-CoA were predicted to bind to hPNPase tighter than citrate. The residues involved in binding these two metabolites were predicted to be similar and they were suggested to mimick the combined interactions of the two molecules of citrate;

essentially binding and occluding the hPNPase active site. Whether, like citrate, these metabolites can also affect hPNPase activity *in vitro* remains to be discovered and this was the focus of the next Section 5.3.2.

5.3.2 Some TCA Metabolites Inhibit *H. sapiens* PNPase Degradation Activity *In Vitro*

The effect of each TCA metabolite on recombinant hPNPase 3'-5' RNA degradation activity was examined using the gel-based assays described previously in Chapter 4 and in Section 5.2.2. The results of which were shown in Figure 5.1 (Section 5.3.1) alongside the results of docking calculations for comparative purposes. It is noteworthy to mention that the control assays, containing only the RNA substrate in the presence of each TCA metabolite, confirmed that none of the TCA metabolites caused RNA degradation; the RNA substrate remained intact in control lanes in Figure 5.1 (c). The results in Figure 5.1 (b) are presented within this section as a larger image (Figure 5.4) for easier comparison to the quantification in Table 5.2.

The TCA metabolites, pyruvate, α -ketoglutarate, succinate, fumarate, malate and oxaloacetate were not found to affect hPNPase 3'-5' activity; the percentage of intact RNA at the end of the assay was the same as the untreated assay (within error) (Figure 5.4 and Table 5.2). Whether this was a result of the reduced binding predicted *in silico* (Section 5.3.1) could be further examined using other *in vitro* biophysical techniques including X-ray crystallography, surface plasmon resonance (SPR) or isothermal titration calorimetry (ITC), however this was not the focus of this study. The *in vitro* data shown in Figure 5.4 and quantification provided in Table 5.2, suggested that isocitrate had no effect on hPNPase degradation activity. Additionally, although cis-aconitate appeared to affect hPNPase activity (Table 5.2), this metabolite is not thought to be 'free' within the cell and so the physiological relevance of this interaction is questionable (Lehninger, Nelson, & Cox, 2000). The *in vitro* gel-based degradation assays, shown in Figure 5.4 and quantification in Table 5.2, suggested that the presence of either acetyl-CoA or succinyl-CoA affected hPNPase activity more than citrate. The presence of succinyl-CoA or acetyl-CoA caused a 4.8 or 7.7-fold increase in the percentage of intact 5' FAM poly(A)_{20mer} RNA respectively.

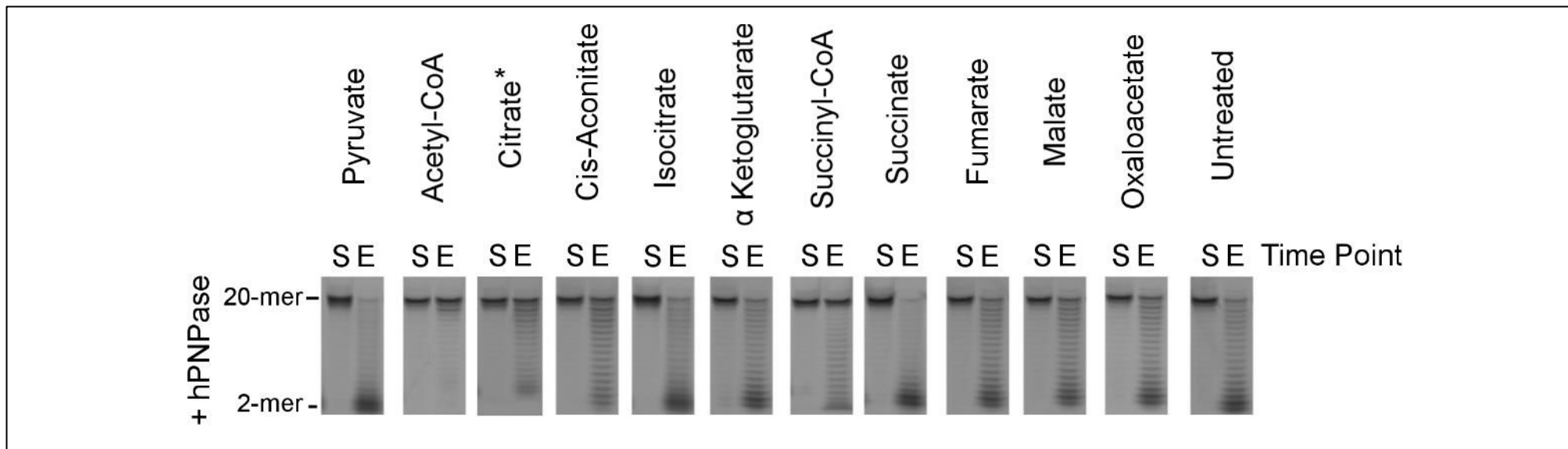


Figure 5.4 *In Vitro* Degradation Activity of hPNPase with TCA Metabolites

Image taken from Figure 5.1 (b) and enlarged. hPNPase (240 nM) was incubated with 700 nM 5'FAM(A)₂₀ RNA substrate, in the presence of 3.75 mM of each TCA metabolite, at 37°C. Samples were run on a 20% urea denaturing gel in 1 x TBE and the RNA substrate degradation products between 2-20 mer are indicated. The samples taken at the start (S, 0 minutes) and end (E, 60 minutes) of each reaction are indicated. A representative gel for citrate (asterisk), taken from Chapter 4, is also shown. Image prepared in GIMP (v2) (GIMP, n.d.).

TCA Metabolites	Percentage substrate remaining at assay end-point
Pyruvate	3.9 ± 0.8
Acetyl-CoA	34.5 ± 5.9
Citrate	12.2 ± 2.6
Cis-aconitate	10.1 ± 3.1
Isocitrate	4.3 ± 0.8
α-ketoglutarate	6.7 ± 0.3
Succinyl-CoA	21.6 ± 0.7
Succinate	3.6 ± 0.3
Fumarate	7.0 ± 1.9
Malate	5.8 ± 2.7
Oxaloacetate	6.6 ± 0.3
Untreated	4.5 ± 1.6

Table 5.2 Degradation Assay; Quantifying the *In Vitro* Inhibition of hPNPase by TCA Metabolites

The percentage of intact RNA substrate remaining at assay end-point, in the absence (-) or presence (+) of each metabolite, are listed for hPNPase. Values are the mean from at least three experimental repeats and the errors reported are the standard deviation (Std). Percentage values which were lower than citrate are coloured red, those equal to citrate (within Std error of the citrate value) were coloured orange, whereas values greater than citrate were coloured green.

In summary, the results presented within this section suggested that acetyl-CoA and succinyl-CoA affected the activity of hPNPase more than the other TCA metabolites tested within this study, including citrate. The following Section 5.3.3, compared the results of previous *in silico* docking calculations (Section 5.3.1) with these *in vitro* results in more detail.

5.3.3 Correlation of *In Silico* Docking Calculations and *In Vitro* Degradation Activity

The data plotted in Figure 5.5 combined evidence from *in silico* and *in vitro* techniques; specifically, the percentage of intact RNA substrate (5'FAM poly(A)_{20mer}) quantified from gel-based assays was plotted against the average S value predicted from docking calculations. This graph was created to illustrate trends within the data and although there wasn't a strong linear correlation between the docking score and inhibition observed, some of the data could be grouped together.

Metabolites with average S values lower than -112 kcal/mol (above red dotted horizontal line in Figure 5.5.) were predicted to have weaker binding compared to citrate and were highlighted red in Figure 5.5. In contrast Metabolites with S values higher than -112 kcal/mol (below red dotted horizontal line in Figure 5.5) were predicted to have stronger binding affinity compared to citrate and were highlighted green in Figure 5.5. Metabolites highlighted in red, which were predicted to bind hPNPase weaker than citrate, were also observed to have less than 10% of intact substrate remaining at the assay end-point (left of red dotted vertical line in Figure 5.5) and therefore

exhibited less inhibition than citrate. The only exception was cis-aconitate (highlighted in orange, in Figure 5.5) which had 10.1% of substrate remaining at the assay end-point. Both metabolites highlighted in green were observed to have more than 10% of substrate remaining at the assay end-point and therefore exhibited more inhibition than citrate (right of red dotted vertical line in Figure 5.5).

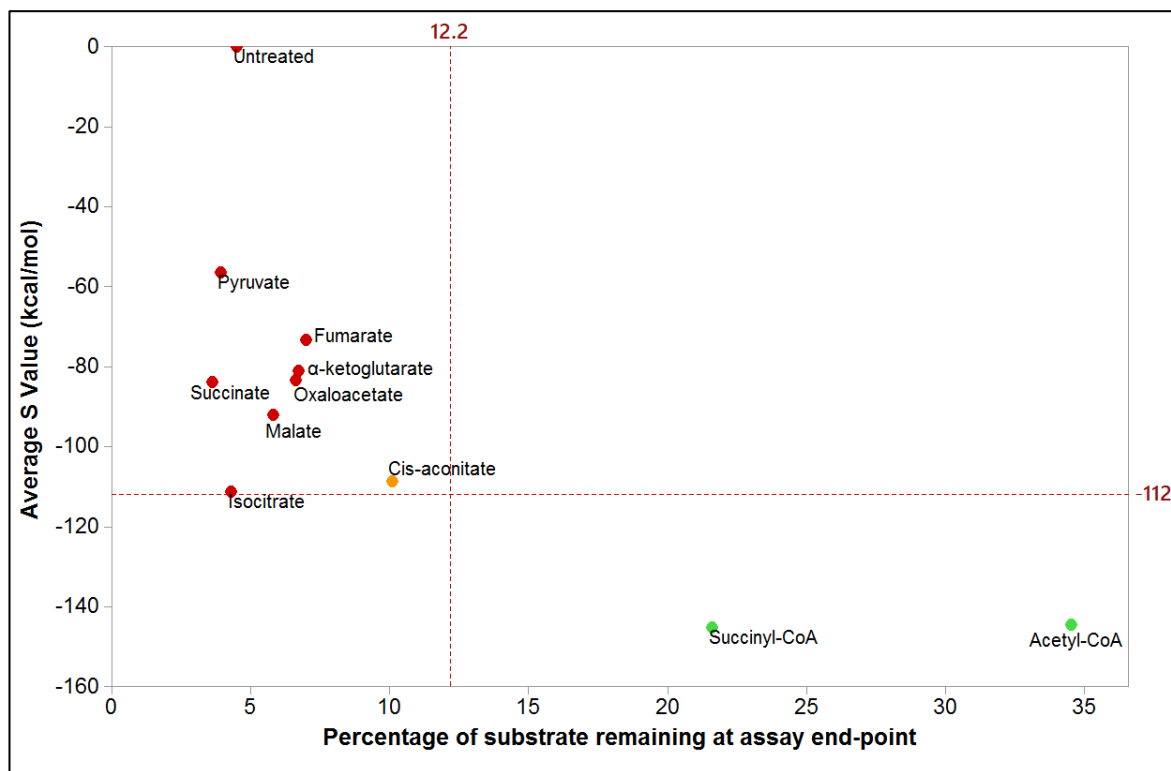


Figure 5.5 *In Silico* Docking Vs *In Vitro* Inhibition of hPNPase by TCA Metabolites

For each TCA metabolite the average S value (kcal/mol) for the top ten poses docked into hPNPase (3U1K) were plotted against the percentage of intact RNA at the assay end-point. Metabolites that caused <2.5 -fold effect on hPNPase 3'-5' RNase activity are highlighted in red, whereas those which inhibited (>2.5 fold) are highlighted green (except cis-aconitate, which inhibited ~ 2.2 -fold and is therefore coloured orange). Graph generated in Minitab (v17) (Minitab, 2010).

In summary, the TCA metabolites acetyl-CoA and succinyl-CoA affected the *in vitro* activity of hPNPase more than citrate and this aligned to *in silico* data which predicted these two metabolites may bind to hPNPase with a higher affinity. The heat maps presented in Section 5.3.1 predicted that the hPNPase residues involved in binding acetyl-CoA and succinyl-CoA, were the same as those involved in citrate binding (Figure 5.3). These residues have already been shown to be highly conserved across evolution in Chapter 4, and this led to the proposal that acetyl-CoA and succinyl-CoA may also bind to other PNPase homologs, including PNPase from the prokaryotic *E. coli* model organism. Therefore, whether these two TCA metabolites can also affect the activity of EcPNPase was examined in Section 5.3.4.

5.3.4 Some TCA Metabolites Inhibit *E. coli* PNPase Degradation Activity

In Vitro

The effect of TCA metabolites, acetyl-CoA and succinyl-CoA, on recombinant EcPNPase 3'-5' RNA degradation activity was examined using the gel-based assays described in Section 5.2.2. The results of which were shown in Figure 5.6 alongside the citrate gels taken from Chapter 4. The control assays shown previously in Figure 5.1 (c) suggested that RNA integrity was not affected by the TCA metabolites.

The *in vitro* gel-based degradation assays suggested that succinyl-CoA affected EcPNPase activity more than citrate. Compared to the untreated assay, the presence of succinyl-CoA, acetyl-CoA or citrate caused a 5.0, 2.8 or 2.5-fold increase in the percentage of intact 5'FAM poly(A)_{20mer} RNA substrate respectively (Table 5.3). Hence, as predicted, EcPNPase was inhibited by all three TCA metabolites. However, unlike hPNPase, succinyl-CoA was the more effective inhibitor of these metabolites.

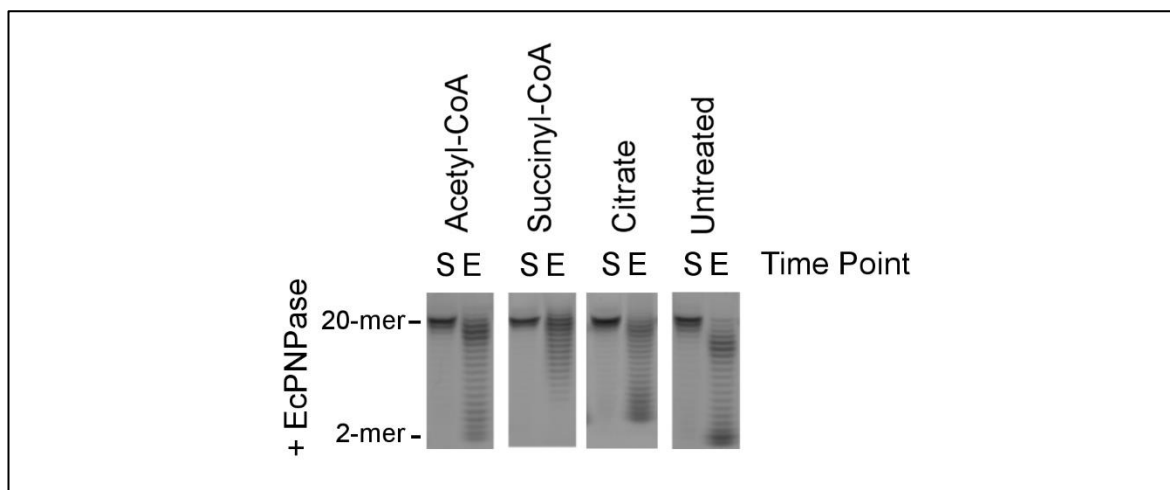


Figure 5.6 3'-5' Degradation Assay; *In Vitro* Inhibition of EcPNPase Homologs by Citrate, Acetyl-CoA and Succinyl-CoA

EcPNPase (240 nM) was incubated with 700 nM 5'FAM(A)₂₀ RNA substrate, in the presence of 3.75 mM of each metabolite, at 37°C. Samples were run on a 20% urea denaturing gel in 1 x TBE and the RNA substrate degradation products 2-20 mer are indicated. The sample taken at the start (S, 0 minutes) and end (E, 10 minutes) of the reaction are indicated. Image prepared in GIMP (v2) (GIMP, n.d.).

TCA Metabolites	Percentage substrate remaining at assay end-point
Succinyl-CoA	11.1 ± 0.7
Acetyl-CoA	6.3 ± 0.7
Citrate	5.6 ± 1.5
Untreated	2.2 ± 0.4

Table 5.3 Degradation Assay; Quantifying the *In Vitro* Inhibition of EcPNPase by Citrate, Acetyl-CoA and Succinyl-CoA

The percentages of intact RNA substrate remaining at assay end-point, in the absence (-) or presence (+) of each metabolite are listed for EcPNPase. Values are the mean from at least three experimental repeats and the errors reported are the standard deviation. Percentage values which were lower than citrate are coloured red, those within the error of citrate were coloured orange, whereas values greater than citrate were coloured green.

In summary, the results presented within this section and previously in Section 5.3.2, suggested that both acetyl-CoA and succinyl-CoA were good inhibitors of hPNPase and EcPNPase activity. Therefore, the purpose of the next section was to analyse the *in silico* docking results presented previously in Section 5.3.1 in more detail, with the aim to identify common features of acetyl-CoA and succinyl-CoA that may be involved in regulating PNPase.

5.3.5 Predicting Metabolite Features That May Be Important for Interactions with PNPase from *H. Sapiens*

The *in silico* docking results presented previously in Section 5.3.1 predicted that acetyl-CoA and succinyl-CoA occupied the active site of hPNPase and mimicked the combined interactions of the two molecules of citrate. However, the features of acetyl-CoA and succinyl-CoA that were predicted by MOE to be necessary for binding, were not compared with each other. Therefore, in order to compare the PLIFS for the top ten acetyl-CoA-hPNPase poses with the equivalent results for succinyl-CoA, a pharmacophore consensus model was generated using MOE (Molecular Operating Environment, 2013). The results shown in Figure 5.7 suggested that four main features (F1-4) were shared between all twenty docking poses.

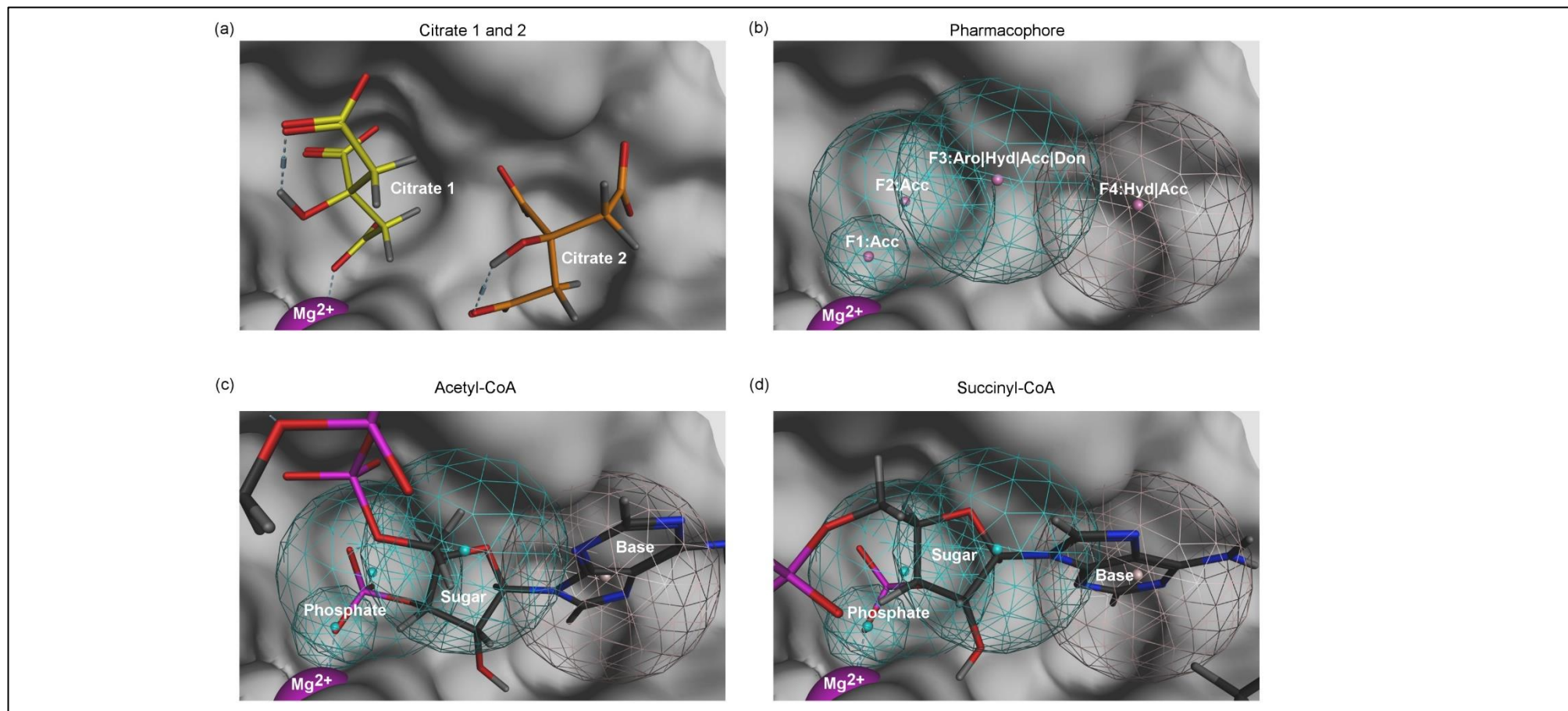


Figure 5.7 hPNPase Pharmacophore Generation

The active site of hPNPase (3U1K_Dock), with a Mg^{2+} (magenta sphere) bound, is shown in grey. The position of (a) two docked citrate molecules (top hits) are shown as yellow (Citrate 1) and orange sticks (Citrate 2). (b) The consensus pharmacophore (mesh spheres) generated shows the four pharmacophore features (F1-4) with acceptor, hydrophobic, donor and/or aromatic labels (Acc, Hyd, Don or Aro respectively). The same pharmacophore is shown with (c) acetyl-CoA and (d) succinyl-CoA bound at the active site. The position of the phosphate, sugar and base of the CoA nucleotide head are labelled. Images generated in MOE and prepared in GIMP (v2) (GIMP, n.d.; Molecular Operating Environment, 2013).

The first feature (F1) suggested an acceptor, with a radius of 1.31 Å, was present next to the active site magnesium ion (magenta, Mg²⁺) in 100% of the twenty poses (F1, Figure 5.7 (b)). The second feature (F2), with a radius of 2.9 Å, overlapped with F1 and suggested that a second acceptor was present within this location in 100% of the poses (F2, Figure 5.7 (b)). As shown in Figure 5.7 (a-b), both F1 and F2 were in the same location of the hPNPase active site that citrate 1 is known to bind and dock. The third feature (F3) suggested that either an acceptor or hydrophobic functional group, with a radius of 2.85 Å, was present in 85% of the poses (F3, Figure 5.7 (b)). As shown in Figure 5.7 (a-b), F3 was in the same location of the hPNPase active site that citrate 2 is known to bind and dock. The fourth feature (F4) was found in 65% of poses and had a radius of 2.75 Å. As shown in Figure 5.7 (a-b), F4 was situated within the hPNPase active site and was located between the space for citrate 1 and 2 binding. The lower percentage (65%) reported for F4 suggested that this feature was not as important as F1 and F2, which were present in 100% of the poses respectively. There also appeared to be a greater amount of flexibility in the types of functional group permitted within this fourth feature; either an aromatic, hydrophobic, acceptor or donor was present within this location.

After establishing the location of the four pharmacophore model features shown in Figure 5.7, the model was used to predict structural elements which were common to acetyl-CoA and succinyl-CoA. For both metabolites, the 3' terminal phosphate of the nucleotide part of CoA appeared to be located in F1 and F2, whereas the sugar and the base were situated within F3 and F4 respectively (Figure 5.7 (c-d)). Essentially, the terminal phosphate of CoA was predicted to act as an acceptor and like citrate, interact with the magnesium cation within the Metal Binding Region (MBR), and the amino acid residues within the Phosphate Binding Region (PBR) of hPNPase. Whereas the nucleoside part of CoA was predicted to bind within the RNA Binding Regions of hPNPase in a similar manner to citrate 2; effectively bridging the interactions between these two citrate-binding sites. When the terminal phosphate of acetyl-CoA or succinyl-CoA, or the carboxyl group of citrate (all bound to hPNPase), were superimposed with the structure of the *S. solfataricus* archaeal exosome (4BA1), the metabolites were all within the equivalent location to the co-crystallised phosphate ion of SsoExosome (Figure 5.8). This suggested that the interactions of these TCA metabolites may prevent inorganic phosphate binding and therefore may provide an explanation, for the reduced phosphorolytic degradation activity, observed for hPNPase in Section 5.3.2. This aligned with previous research which proposed that citrate binding inhibited the degradation activity of PNPase in *E. coli* by preventing the formation of the Michaelis complex (Nurmohamed, Vaidialingam, Callaghan, & Luisi, 2009). That study suggested that citrate 1 occupied the binding site for the orthophosphate substrate whereas the adjacent citrate 2 mimicked the position of the scissile phosphate in the backbone of the RNA (Nurmohamed *et al.*, 2009). It was noted that in these positions, the two citrate molecules may prevent the binding of inorganic phosphate, which is required for catalysis and potentially the RNA substrate itself.

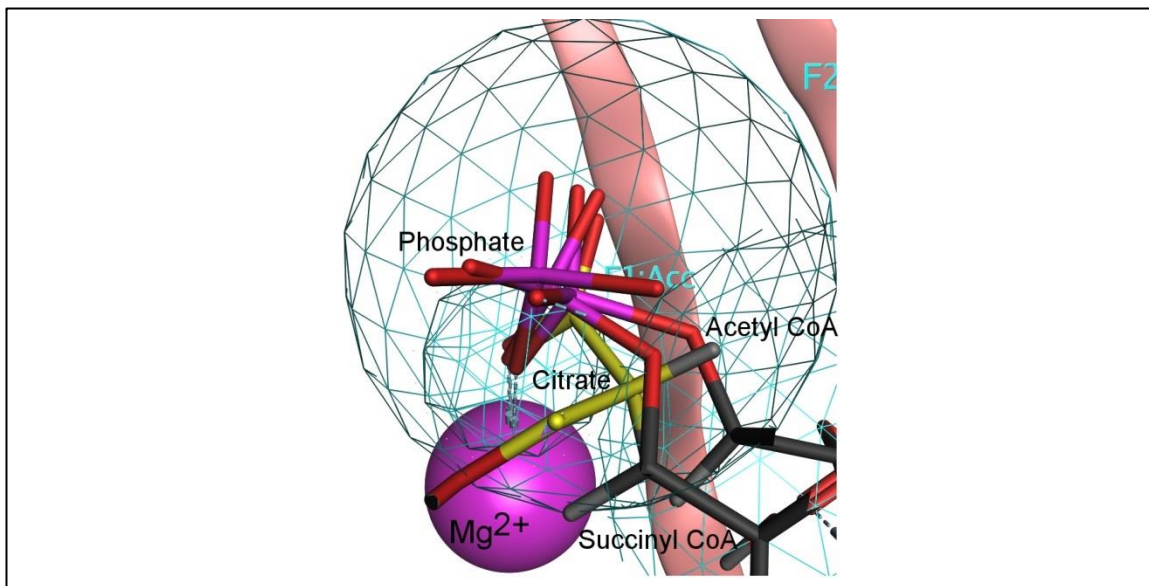


Figure 5.8 hPNPase Pharmacophore: Phosphate Positioning

A small section of the hPNPase (3U1K_Dock) active site (pink ribbons), with Mg^{2+} (magenta sphere), is shown zoomed in. The pharmacophore features 1 and 2 are also shown (blue mesh). The position of citrate and the 3' terminal phosphates from acetyl-CoA, succinyl-CoA are shown alongside the equivalent position of co-crystallised phosphate of SsoExosome (4BA1). Images generated in MOE and prepared in GIMP (v2) (GIMP, n.d.; Molecular Operating Environment, 2013).

In summary, the results presented within this study proposed that the CoA nucleotide part of acetyl-CoA and succinyl-CoA may regulate hPNPase in a similar manner to the citrate-mediated inhibition mechanism proposed for EcPNPase. The 3' terminal phosphate of the CoA nucleotide may occlude the inorganic phosphate binding region of the hPNPase active site, and the nucleoside part of CoA may prevent the binding of the RNA substrate. In combination, these interactions may regulate the 3'-5' RNA degradation of hPNPase by preventing phosphorolytic acid-base catalysis. This was interesting as PNPases are known to be regulated by nucleotides in *E. coli*. For example, EcPNPase was allosterically inhibited by ATP nucleotides (Del Favero *et al.*, 2008) and cyclic di-GMP secondary messenger molecules *in vitro* (c-di-GMP; Tuckerman *et al.*, 2011).

Following the prediction that the nucleotide part of CoA was important for regulating PNPase, a range of other nucleoside/nucleotide-based metabolites were docked *in silico* in the following Section 5.3.6. The aim of this study was to predict whether other nucleoside/nucleotide containing metabolites could bind hPNPase with a similar affinity to either citrate, acetyl-CoA or succinyl-CoA.

5.3.6 Phosphate-Rich Metabolites Are Predicted to Dock *In Silico* into PNPase from *H. Sapiens*

A small bespoke (user defined) database containing a range of nucleosides, mono/di/tri-nucleoside phosphates and other phosphate-rich signalling molecules (Additional_Metabolites.mdb, Table 5.1) were docked into hPNPase (3U1K_Dock) using the docking module of MOE and the parameters outlined in Section 5.2.1.1. The top 30, lowest-energy poses, S values were represented as box

plots in Figure 5.9 and the average S value for citrate 1 docking (top ten poses) was shown for comparative purposes (red dotted horizontal line in Figure 5.9).

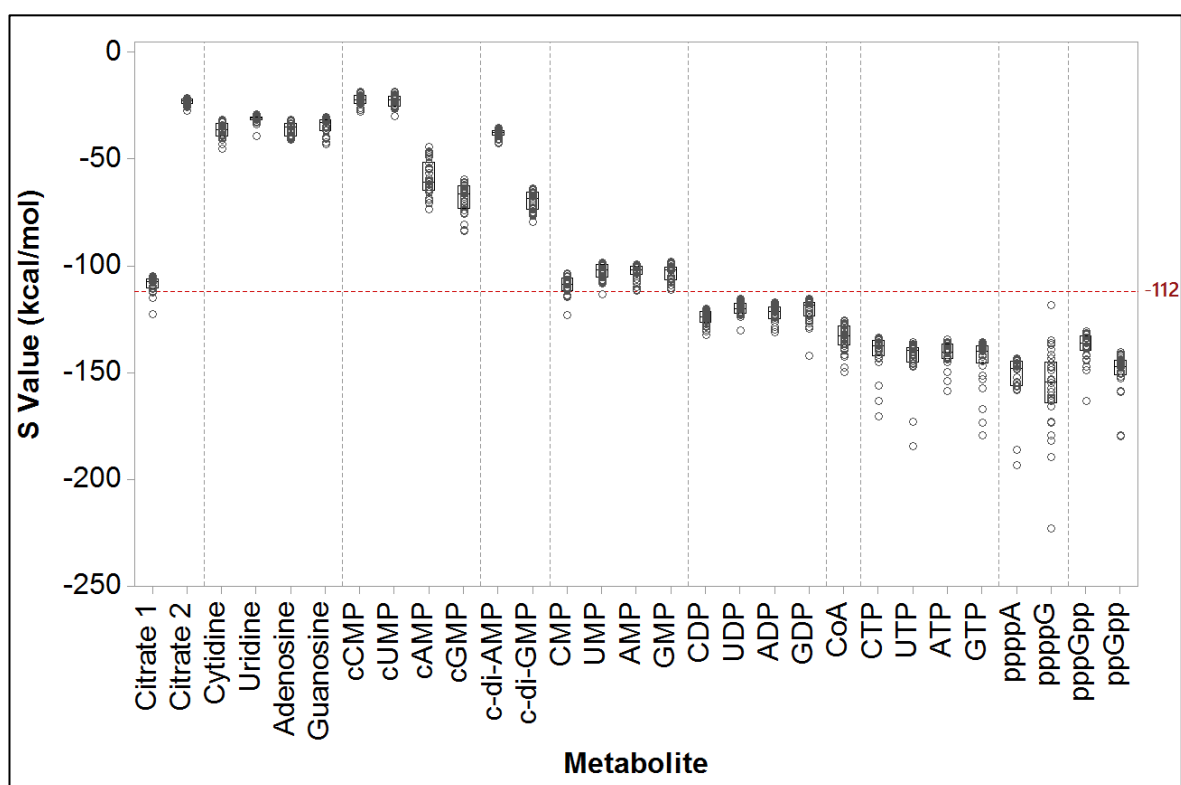


Figure 5.9 In Silico Docking of hPNPase with Phosphate-Rich Metabolites

The docking scores (S value, kcal/mol) predicted by MOE (Molecular Operating Environment, 2013), for the 30 unique, lowest-energy poses of a range of metabolites bound to hPNPase (3U1K_Dock). Graph was subgrouped (grey dotted lines) into results for citrates, nucleosides, cyclic nucleotides, di-cyclic nucleotides, mono-phosphate nucleotides, di-phosphate nucleotides, CoA, tri- and tetra-phosphate nucleotides, the bacterial stringent response alarmone guanosine tetraphosphate (ppGpp), and its precursor guanosine pentaphosphate (pppGpp) respectively. For comparison, the average S value (top ten poses) for citrate 1 was indicated with a red dotted horizontal line. Graph generated in Minitab (v17) (Minitab, 2010).

The nucleosides, cyclic nucleotides and di-cyclic nucleotides were all predicted to bind hPNPase with S values higher (weaker binding) (Figure 5.9, above red dotted line) than that previously determined for citrate 1 (-112 kcal/mol). Looking at the sub-group highlighted in Figure 5.9 in more detail (defined by grey dotted vertical lines), the average S value for nucleosides was -38 kcal/mol (average of the top ten S values, for each metabolite within the subgroup). The cyclic nucleotides, which are like other nucleotides except a cyclic bond arrangement occurs between the sugar and phosphate groups, had a slightly lower average S value (top ten hits) for adenosine (cAMP) and guanosine (cGMP) (67.6 and 76.9 kcal/mol respectively), compared to cytidine (cCMP) and uridine (cUMP) which were -24.9 and -26.1 kcal/mol respectively. Interestingly, cAMP and cGMP are derivatives of ATP and GTP respectively and both have roles as a second messenger molecules, however the roles cCMP and cUMP are poorly understood. Whether these docking scores suggest that hPNPase may be able to sense, and thus be regulated by the purine or pyrimidine nitrogenous bases differently, requires further research. The di-cyclic nucleotides, which have a nitrogenous base linked by a sugar and phosphate, had a lower average S value (-109.0 kcal/mol) than the cNDPs (Figure 5.9) for the adenosine base, but a similar S value for the

guanosine base respectively. It is important to note that structures for both cytidine and uridine di-cyclic nucleotides were not available within either the Zinc or PubChem databases and so were not docked.

When comparing the average S value (average of the top ten values, for each metabolite, within the subgroup) for the nucleosides, the mono- (NMP), di- (NDP), tri- (NTP) and tetra-phosphate nucleotides (ppppN), a general trend between the number of phosphates and the S value was observed (Figure 5.9). Specifically, the addition of one, two, three and four phosphates to adenosine, cytidine, guanosine and uridine resulted in an increasingly lower average S value. For example, the nucleosides had an average S value of -38 kcal/mol and this decreased to -109.0 kcal/mol for the mono-phosphate nucleotides. This value reduced further to -127.0 kcal/mol and -151.1 kcal/mol for NDPs and NTPs respectively and even more to -170.4 for ppppN nucleotides. Interestingly, all these average S values were lower than the average S value previously determined for citrate 1 docking into the hPNPase active site (-112 kcal/mol, red horizontal dotted line in Figure 5.9). Thus predicting that NDPs, NTPs and ppppNs may bind hPNPase with tighter affinity than citrate. It is important to note that structures for both ppppC and ppppU were not available and so were not docked. The average S values calculated for CoA (-140 kcal/mol) was slightly lower than NDPs, but slightly higher than NTPs, ppppNs and (p)ppGpp, predicting that phosphate-rich nucleotides may bind hPNPase with a higher affinity than CoA. This was an interesting observation as the pharmacophore consensus model presented in Section 5.3.5, proposed that the CoA nucleotide part of acetyl-CoA and succinyl-CoA was involved in hPNPase interactions.

The metabolites guanosine tetraphosphate (ppGpp), and its precursor guanosine pentaphosphate (pppGpp), collectively termed (p)ppGpp, were both predicted to interact with hPNPase with average S values (-158.7 and -145.0 kcal/mol respectively) lower than citrate (below red dotted horizontal line in Figure 5.9). The cellular levels of these metabolites in *H. sapiens* mitochondria are not currently known. However, they have been shown to inhibit PNPase homologs from *Nonomuraea sp.* and *Streptomyces*, but not from *E. coli* (Gatewood & Jones, 2010; Hauryliuk, Atkinson, Murakami, Tenson, & Gerdes, 2015; Siculella *et al.*, 2010). Whether these metabolites also regulate PNPase activity in humans remains to be revealed.

It has been shown within this study that metabolites, including acetyl-CoA and succinyl-CoA, with S values lower than citrate can affect the activity of hPNPase *in vitro*. Therefore, it was proposed that other metabolites, shown in Figure 5.9 with S values lower than citrate (below the red dotted horizontal line), may affect hPNPase activity in a similar way to the TCA cycle metabolites. Accordingly, the following Section 5.3.7 describes the investigation of whether phosphate-rich nucleotides including GTP, ppppG and ppGpp, which are predicted to bind hPNPase with a lower S value than citrate, could also inhibit the enzyme's activity.

5.3.7 Phosphate-Rich Metabolites Inhibit *H. sapiens* and *E. coli* PNPase Degradation Activity *In Vitro*

The effect of phosphate-rich metabolites GTP, ppppG and ppGpp on the 3'-5' RNA degradation activity of recombinant hPNPase and EcPNPase was examined. Gel-based assays were conducted as described previously in Chapter 4 and Section 5.2.2, and the results are shown in Figure 5.10. The control assays, containing only the RNA substrate in the presence of each metabolite, showed that none of the metabolites caused RNA degradation; the RNA substrate remained intact in control lanes in Figure 5.10 (c).

The *in vitro* gel-based degradation assays, shown in Figure 5.10 (a) and quantification in Table 5.4, suggested that that GTP, ppppG and ppGpp all affect hPNPase activity. The percentage of intact 5'FAM poly(A)_{20mer} RNA substrate at the end of the assay, for all three phosphate-rich metabolites, increased by ~6.5-fold. These percentages (highlighted in green in Table 5.4) were higher than the percentages previously reported for citrate and succinyl-CoA, but were within error of the results obtained for acetyl-CoA (Table 5.2). As the average *S* value of GTP, ppppG and ppGpp (-156.4, -177.4 and -158.7 kcal/mol respectively) were all lower than citrate (-112 kcal/mol) and following the observation that, like acetyl-CoA and succinyl-CoA, they also affect the 3'-5' degradation activity of hPNPase *in vitro*. This work helped support the proposal made at the end of Section 5.3.6 that metabolites, predicted to dock into hPNPase with a higher affinity than citrate, may also affect the enzymes degradation activity.

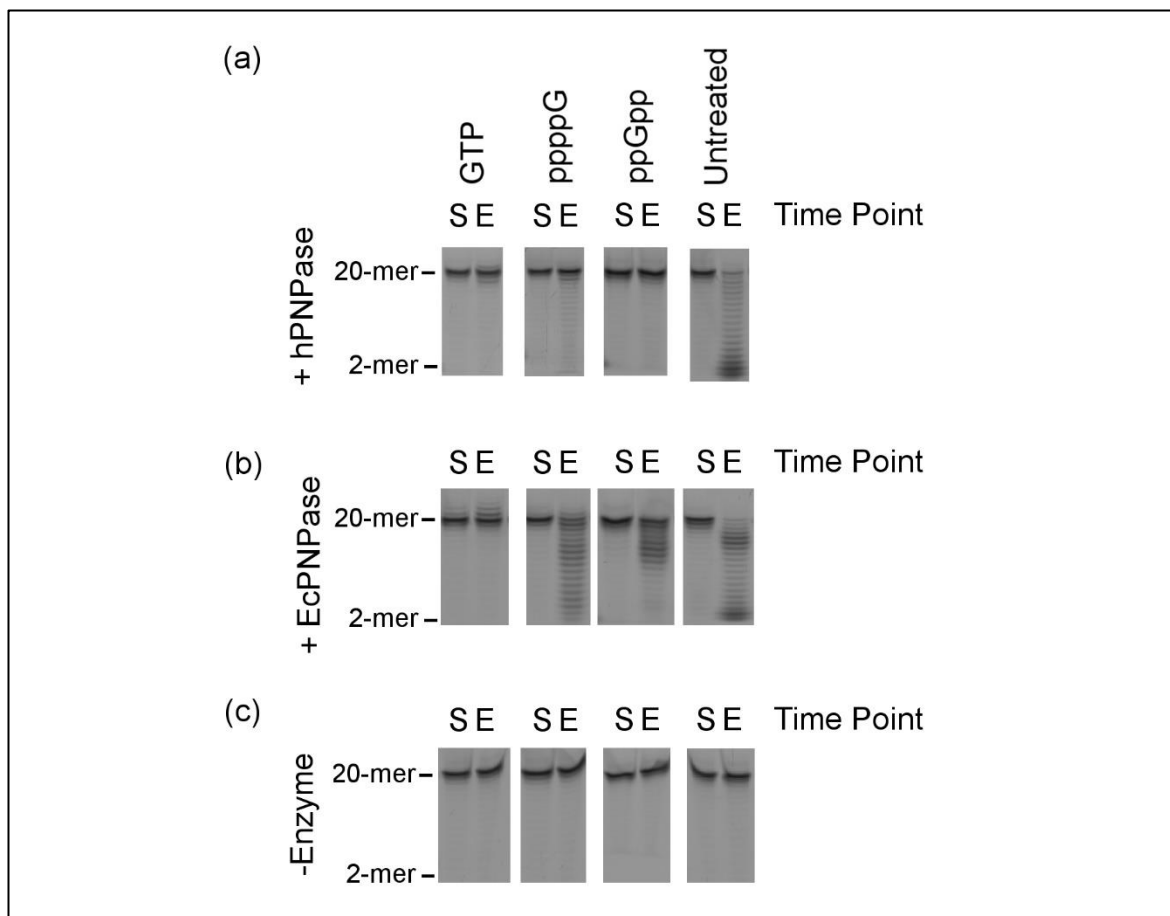


Figure 5.10 *In Vitro* Degradation Activity of hPNPase and EcPNPase with Phosphate-Rich Metabolites GTP, ppppG and ppGpp

700 nM 5'FAM(A)₂₀ RNA substrate, in the presence of 3.75 mM of each metabolite (GTP, ppppG and ppGpp), at 37°C with either (a) hPNPase, (b) EcPNPase or (c) no enzyme. Samples were run on a 20% urea denaturing gel in 1 x TBE and the RNA substrate degradation products 2-20 mer were indicated. The sample taken at the start (S) and end (E) of the reaction are indicated. In these assay conditions, to account for enzymatic efficiency differences, the hPNPase and no enzyme control was incubated for 60 minutes and EcPNPase for 10 minutes. Image prepared in GIMP (v2) (GIMP, n.d.).

Metabolites	Percentage substrate remaining at assay end-point	
	<i>H. sapiens</i> PNPase	<i>E. coli</i> PNPase
GTP	31.1 ± 5.3	34.2 ± 1.7
ppppG	31.3 ± 3.7	10.7 ± 2.0
ppGpp	31.8 ± 0.9	13.5 ± 0.8
Untreated	4.8 ± 1.7	2.3 ± 0.5

Table 5.4 Degradation Assay; Quantifying the *In Vitro* Inhibition of hPNPase and EcPNPase by Phosphate-Rich Metabolites

The percentage of substrate remaining at assay end-point in the absence (-) or presence (+) of each phosphate-rich metabolite are listed for hPNPase and EcPNPase. Values are the mean from at least three experimental repeats and the errors reported are the standard deviation. Percentage values lower than citrate were coloured black, whereas values greater than citrate were coloured green.

Whether these phosphate-rich metabolites could also inhibit the activity of PNPase from *E. coli* was investigated and the results were shown in Figure 5.10 (b) and Table 5.4. The results suggested that in the presence of ppppG, ppGpp or GTP, EcPNPase activity was affected. Additionally, the percentage of substrate remaining at assay end-point for all three phosphate-rich metabolites was

higher (highlighted in green, Table 5.4) than the percentage previously reported for citrate (Table 5.3). The TCA metabolite succinyl-CoA was previously suggested to affect the activity of EcPNPase *in vitro* (Section 5.3.4) and results shown here suggest that ppppG had a similar effect; the percentage of intact 5'FAM Poly(A)_{20mer} RNA substrate remaining at the end of each assay was 11.1% ($\pm 0.7\%$) and 10.7% ($\pm 2.0\%$) respectively. The amount of intact RNA increased to 13.5% ($\pm 0.8\%$) and 34.2% ($\pm 1.7\%$) in the presence of ppGpp and GTP respectively. When comparing the effect of GTP, ppppG and ppGpp on the activity of EcPNPase, GTP, which is a known inhibitor of EcPNPase (Del Favero *et al.*, 2008), affected the enzyme's activity the most. Interestingly, the observation that ppGpp affected EcPNPase activity conflicted with previous literature and this is discussed further in Section 5.4 (Gatewood & Jones, 2010). Nevertheless, the finding that phosphate-rich metabolites ppppG, ppGpp and GTP affected the activity of EcPNPase aligned with the data presented for hPNPase above.

In summary, the results from this section supported the proposal made in Section 5.3.6, that *S* values calculated for a metabolite docked into the active site of hPNPase, may help predict metabolites that are involved in regulating PNPase. Specifically, it was proposed that metabolite-hPNPase interactions with *S* values lower than the citrate-hPNPase interaction (-122 kcal/mol) may affect the enzymes 3'-5' degradation activity. The predictions from *in silico* molecular docking calculations in Section 5.3.6 correctly identified phosphate-rich metabolites that could inhibit hPNPase activity *in vitro*. Additionally, these metabolites, namely ppppG, ppGpp and GTP, were also shown to inhibit the *in vitro* degradation activity of the PNPase homolog from *E. coli*. It was observed that these phosphate-rich metabolites had a larger effect on hPNPase activity than on the enzyme counterpart in *E. coli*. This was in agreement with previous research that suggested hPNPase was more sensitive to cellular phosphate levels than EcPNPase and is actually inhibited by higher phosphate concentrations (Portnoy, Palnizky, Yehudai-Resheff, Glaser, & Schuster, 2007). Although the levels of metabolite-mediated inhibition appeared to vary between EcPNPase and hPNPase, the results still validated the use of the docking program MOE for predicting novel metabolite-PNPase interactions across evolutionarily diverse organisms.

5.4 Conclusion

Recent findings that citrate interacts with PNPase in *E. coli* (Nurmohamed *et al.*, 2011) and that this interaction may represent a conserved communicative link across evolution (Chapter 4), has raised the question of whether other metabolites of the TCA cycle, similar to citrate, are involved in regulating PNPase homologs. By combining *in silico* docking calculations and *in vitro* gel-based assay results, the TCA metabolites pyruvate, α ketoglutarate, succinate, fumarate, malate and oxaloacetate were predicted to bind hPNPase with a weaker affinity than citrate, and weren't observed to inhibit hPNPase activity. Interestingly, the average *S* values for citrate, cis-aconitate and isocitrate docking into the active site of hPNPase, were all similar. However, unlike citrate and cis-aconitate, isocitrate had no effect on hPNPase activity. The difference in inhibition exhibited by

isocitrate and cis-aconitate, may be due to their structures; citrate is converted into cis-aconitate before isocitrate is formed, hence the structure of cis-aconitate is more similar to citrate (Figure 1.12, Chapter 1) and may explain its ability to affect hPNPase activity. Although, the physiological relevance of cis-aconitate binding and inhibiting hPNPase remains to be determined; it is thought that this metabolite is never 'free' in the cell, but remains associated with metabolic enzymes (as reviewed in Lehninger *et al.*, 2000). The only TCA metabolites which were predicted to bind to hPNPase tighter than citrate, were acetyl-CoA and succinyl-CoA. These metabolites were also shown to affect hPNPase activity more than citrate. Although, the cellular levels of acetyl-CoA and succinyl-CoA in the human mitochondria need to be determined, the physiological relevance of these two molecules interacting with hPNPase is interesting as they both have key roles in regulating the TCA cycle (as reviewed in Berg, Tymoczko, & Stryer, 2002) and this is discussed further in Chapter 7.

The results of *in silico* docking calculations presented in Section 5.3.1, predicted that acetyl-CoA and succinyl-CoA bind and occlude the active site of hPNPase in a similar way to citrate. Similarly, *in vitro* degradation assays presented in Section 5.3.1 and described further in Section 5.3.2 suggested that only citrate, acetyl-CoA and succinyl-CoA were able to inhibit hPNPase activity more than 2.5-fold.

The *in silico* results presented as heat maps in Figure 5.3 and the pharmacophore consensus model provided in Figure 5.7 (b), predicted that the CoA nucleotide part of acetyl-CoA and succinyl-CoA was the important structural feature interacting with hPNPase. The residues and types of interactions involved in CoA binding were suggested to essentially mimic the combined interactions of the two molecules of citrate. Specifically the 3' terminal phosphate of acetyl-CoA and succinyl-CoA and citrate 1, were predicted to bind and occlude the metal and phosphate binding region. Whereas the nucleoside part of acetyl-CoA and succinyl-CoA and citrate 2, were predicted to bind within the RNA binding regions. In these positions the TCA metabolites may prevent the binding of inorganic phosphate which is required for catalysis and prevent the cleavage of RNA substrate. Importantly, the hPNPase residues involved in binding acetyl-CoA and succinyl-CoA were the same as the citrate binding residues which have been shown to be highly conserved across evolution (Chapter 4). It was therefore suggested, that like citrate, acetyl-CoA and succinyl-CoA may also inhibit other PNPase homologs. Accordingly, *in vitro* gel-based evidence was presented that suggested citrate, acetyl-CoA and succinyl-CoA all inhibited EcPNPase 3'-5' degradation activity. Although, the cellular levels of acetyl-CoA and succinyl-CoA in the human mitochondria need to be determined, the levels of these metabolites reported in exponentially growing *E. coli* cells, were comparable to the low mM concentrations used within these assays (Bennett *et al.*, 2009).

When the levels of TCA metabolite-mediated inhibition were compared between hPNPase and EcPNPase subtle differences were observed. Whether these differences represented different mechanisms of regulating RNA turnover across prokaryotes and eukaryotes, remains to be determined. Nevertheless, the results presented so far suggested that the TCA metabolite-mediated inhibition of hPNPase and EcPNPase activity could potentially be a commonly exploited regulatory mechanism in the two domains of life; eukaryotes and prokaryotes. Additionally, comparisons of the *in silico* data for citrate, acetyl-CoA and succinyl-CoA docking within the active site of hPNPase, suggested that hPNPase may interact with other phosphate containing nucleotide-based metabolites.

After more *in silico* docking experiments, phosphate-rich guanosine nucleotides including GTP, ppppG and ppGpp, previously implicated in regulating PNPase activity, were predicted to bind hPNPase. Although ppppG had a lower average S value than both ppGpp and GTP (-177.4, -158.7 and -156.4 kcal/mol respectively), all three phosphate-rich metabolites were found to inhibit hPNPase. These findings supported the *in silico* predictions which proposed that metabolites with an average S value (top ten docking poses) lower than citrate (-112 kcal/mol), may be able to affect hPNPase activity.

The effect of phosphate-rich nucleotides was examined for *E. coli* PNPase and the results suggested that GTP, ppppG and ppGpp also inhibited its 3'-5' RNA degradation activity. The finding that ppGpp could affect the degradation activity of EcPNPase conflicted with previous publications. Remarkably, although (p)ppGpp was found to inhibit PNPase from *S. antibioticus*, it was suggested to have no effect on EcPNPase (Gatewood & Jones, 2010), despite the significant structural conservation (Nurmohamed *et al.*, 2009; Symmons, Jones, & Luisi, 2000). The publication by Gatewood and Jones proposed that ppGpp had no effect on the EcPNPase mediated phosphorolysis of an RNA substrate derived from the *rpsO-pnp* operon of *S. coelicolor*. They suggested that all the products normally observed upon phosphorolysis of the 5601 transcript catalysed by EcPNPase, were observed in the presence of ppGpp (Figure 5.11, lanes 6 and 7 respectively). However, subtle differences were visible; the amount of intact substrate RNA was slightly increased (Figure 5.11, lane 7) and the amount of RP2 degradation product (Figure 5.11, unlabelled band underneath RP2 in lane 7) appeared to have decreased in the presence of ppGpp compared to the untreated sample (Figure 5.11, lane 6). After reviewing differences between the methods utilised, the work presented here used an equal molar concentration of 3.75 mM ppGpp to 3.75 mM magnesium, whereas in previous studies 1 mM ppGpp was not in excess to the 5 mM magnesium and potentially this may explain the differences in inhibition level observed. In conclusion, the conflicting data presented here may be a result of the exact conditions used and the effect of this alarmone in *E. coli* needs to be examined further. This ppGpp-EcPNPase interaction was not the focus of this thesis, but would be of great importance to study since (p)ppGpp has been linked to antibiotic production in response to amino acid and energy starvation, and as mentioned

in previous chapters, suggests a potential link between environmental sensing and PNPase activity (Bralley & Jones, 2003; Mechold, Cashel, Steiner, Gentry, & Malke, 1996).

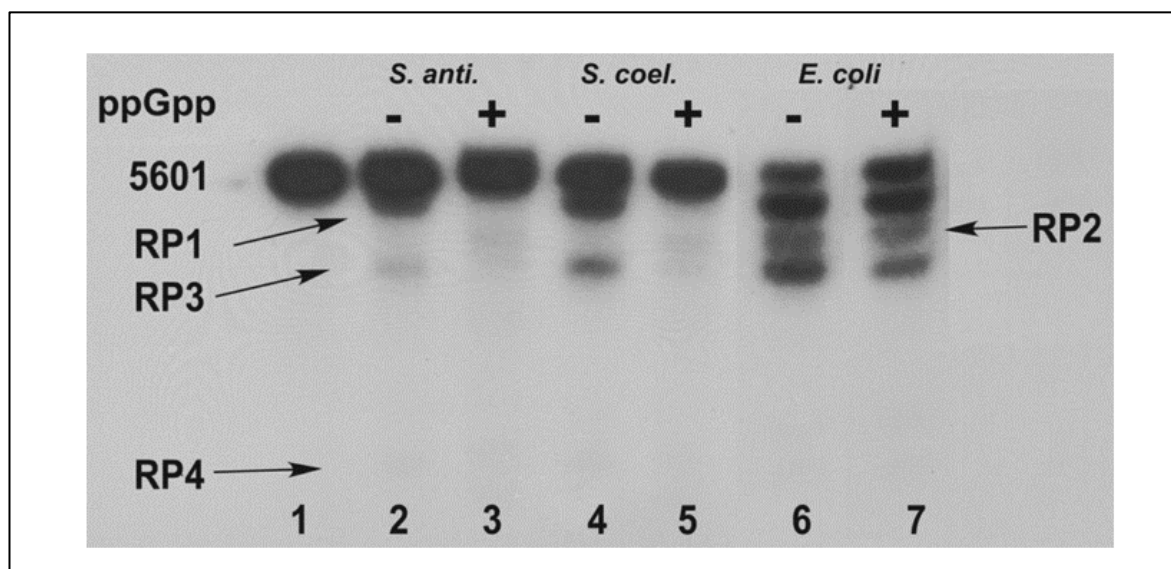


Figure 5.11 *In Vitro* Degradation Activity of PNPase Homologs with ppGpp

Gel electrophoresis of PNPase digests of the 5601 transcript with ^{32}P -labelled 5601 RNA. Gels were subjected to autoradiography. Lane 1, undigested 5601 transcript; lanes 2 and 3, digestion with *S. antibioticus* PNPase; lanes 4 and 5, digestion with *S. coelicolor* PNPase; lanes 6 and 7, digestion with *E. coli* PNPase. The reaction mixtures corresponding to lanes 3, 5, and 7 contained 1 mM ppGpp. The product designations at the left and right of the figure relate the bands produced to the 3' endpoints mapped previously by cDNA cloning and sequencing. Figure and legend taken from (Gatewood & Jones, 2010).

Regardless of the conflicting results mentioned above, there was a clear reduction in both hPNPase and EcPNPase degradation activity in the presence of phosphate-rich metabolites. Despite the high structural and sequence conservation of hPNPase and EcPNPase (Chapter 4), the level of inhibition by either GTP, ppppG or ppGpp were greater for hPNPase. Although this finding was supported by previous publications reporting different phosphate preferences across diverse organisms (Portnoy *et al.*, 2007), the question of how these enzymes are regulated by different cellular phosphate levels, remains to be answered and, as discussed in Chapter 7, this would be an interesting avenue of study.

In summary, this chapter provided a variety of *in silico* and *in vitro* evidence that suggested the citrate-PNPase communicative link may extend to include a much wider range of metabolites. The detailed sequence analysis provided previously in Chapter 4, suggested that citrate-binding, and thus acetyl-CoA and succinyl-CoA binding residues are widely conserved. Therefore, the TCA and phosphate-rich metabolites reported to affect *H. sapiens* and *E. coli* PNPase activity within this chapter, may also inhibit other PNPase homologs. Whether these metabolite-PNPase interactions represent a conserved regulatory strategy which is utilised *in vivo* to control RNA turnover across the three domains of life; eukaryotes, prokaryotes and archaea, needs to be determined. Additionally, the effects of metabolites on other ribonucleases also needs to be investigated.

A better understanding of how RNases can sense different metabolites may explain how PNPase homologs are able to regulate RNA turnover in a range of organisms which have a diverse pool of metabolites available. The results presented so far within this thesis improve the understanding of detailed interactions between metabolites and 3'-5' PNPase homologs. With further *in vivo* evidence, the overall flow of cellular information may be expanded and the interaction of metabolites with 3'-5' exoribonucleases, may explain how complex organisms are able to fine-tune gene expression.

6 Developing a High-Throughput 3'-5' RNase Kinetic Assay

6.1 Introduction

The previous chapters have described the study of the effects of metabolites on 3'-5' RNase activity using both *in silico* molecular docking experiments and *in vitro* gel-based end-point assays. These were excellent tools to predict and test whether small molecule metabolites may bind and inhibit PNPase and archaeal exosome proteins. Obtaining more detailed enzyme-inhibitor kinetics will however be valuable in order to understand how these inhibitors work in more detail. Typically, gel-based assays are time-consuming and often limited to short time courses or end-point assays. In this study, this would limit the number of metabolite-PNPase interactions that could be investigated. A more efficient and quantifiable method of characterising metabolite interactions with PNPase homologs was therefore required. Real-time enzyme kinetic experiments can be conducted in high-throughput, on specialist equipment such as an automated plate reader. The ability to obtain enzyme kinetic data in a high-throughput screen (HTS), is extremely beneficial for drug/inhibition discovery, as it allows a range of compounds to be screened quickly (reviewed in Hughes, Rees, Kalindjian, & Philpott, 2011). This study aimed to develop a high-throughput plate reader assay which could be later utilised to screen a large range of metabolites and quantify quickly and in real-time, their interaction with PNPase homologs. Considering the large range of metabolites that are known to exist across prokaryotic, eukaryotic and archaeal organisms, this HTS assay would be advantageous to study the conservation of metabolite-PNPase interactions.

In this chapter the basics of 3'-5' RNase catalysis and general enzyme kinetics are initially covered. This allows the reader to appreciate the types of variables which must be considered when establishing canonical enzyme kinetics and the subsequent effect of inhibitors. The factors which must be controlled for studying the kinetics of complex multi-functional/reversible enzymes, including PNPase homologs; which can either degrade or polymerise RNA substrates are discussed. For example, the HTS assay was optimised to follow only the degradation activity of PNPase homologs in real-time and reduce the reverse polymerisation reaction by sequestering the ADP product released upon degradation. The methods utilised to develop a high-throughput plate assay for testing this 3'-5' phosphorolytic activity are provided in Section 6.3. The results of the plate reader assay development are in Section 6.4; instrument set-up and method optimisation are covered before canonical enzyme kinetic analysis is determined. Key conclusions from this chapter are then made in Section 6.5, before the thesis Discussion (Chapter 7) reflects on the impact of metabolite-RNase communication and the use of a high-throughput assay for further discovery of novel interactions.

6.1.1 3'-5' RNase Catalysis

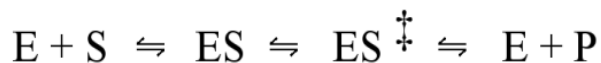
As described previously in Chapter 1, RNases are the enzymes that control the RNA turnover in all domains of life. The mechanism, by which organisms use RNases to regulate RNA turnover can vary; some RNA targets may be stabilised, while the same process may lead to RNA degradation in a different organism. An example of this is the addition/removal of a poly(A) tail in eukaryotes when compared to prokaryotes. In eukaryotes, polyadenylation often results in transcript stability and as a signal for further RNA maturation; however, in prokaryotes the same action may recruit RNases for targeted degradation (Sarkar, 1997; Wilusz, Wormington, & Peltz, 2001). Since there is a large amount of variation in cellular RNA abundance, composition and length of RNA strands between organisms, a variety of mechanisms are required to degrade these substrates and this is reflected by the large size and functional diversity of the RNase enzyme group (Ibrahim, Wilusz, & Wilusz, 2008; Nurmohamed *et al.*, 2011; Zuo & Deutscher, 2001).

The factors affecting the activity of an enzyme are important to understand before conducting detailed kinetic studies. Not only can substrate specificity vary for each PNPase homolog, but other factors can influence the enzyme's activity. For example, although prokaryote and eukaryote PNPases and the archaeal exosome, all exhibit inorganic phosphate (PO_4^{3-}) and metal ion (Mg^{2+} or Mn^{2+}) dependent 3'-5' exoribonuclease activity (reviewed in Zuo & Deutscher, 2001), their temperature optimum and preference for essential PO_4^{3-} concentration varies between enzymes. For example, EcPNPase and hPNPase exhibit activity at 37°C, however the thermophile SsoExosome can survive at ~ 80°C (Portnoy *et al.*, 2005). Additionally, it has already been reported that different PNPase homologs have different phosphate preferences: hPNPase has been shown to have a lower phosphate preference than EcPNPase (Portnoy, Palnizky, Yehudai-Resheff, Glaser, & Schuster, 2007). It is therefore difficult to select specific reaction conditions for comparative enzyme studies which are optimal for the range of enzymes tested, especially if those enzymes span a wide range of diverse organisms. Hence, conditions selected within this thesis were not always optimal for the specific enzyme, but allowed some comparison to be made across the range of species tested.

The next section describes the equations used to calculate general enzyme kinetics parameters and how they are used to compare enzyme activity. Where appropriate, a generalised 3'-5' exoribonuclease model is applied, unless a specific enzyme is stated. Moreover, the RNA substrate refers to a simple unlabelled 20 mer Poly(A) substrate and therefore the forward reaction is degradation and the reverse reaction is polyadenylation.

6.1.2 Enzyme Kinetics Equations

For a typical enzyme-catalysed reaction, such as the enzymatic degradation of RNA substrates by ribonuclease enzymes, the reaction proceeds as follows in Equation 6.1:

**Equation 6.1 Enzyme Reaction**

The general reaction of enzyme (E) with its substrate (S) produces an enzyme-substrate complex (ES) and transition state (ES^\ddagger), which is followed by the release of product (P) and enzyme.

Enzyme (E) and substrate (S), both with free translational, rotational and vibrational degrees of freedom, must first interact in solution before forming an enzyme-substrate complex (ES). Upon binding, a range of events occur involving both the enzyme and substrate, which result in the overall change of free energy (ΔG_B). Firstly, the entropy described above (the freedom of motion of molecules in solution, which reduces the possibility they will react) is reduced. The substrate is positioned in an orientation which reduces movement and distorts the molecule in order to hold it in a position which promotes weak non-covalent interactions; either hydrogen bonds, electrostatic, or van der Waals. Secondly, the solvation shell of hydrogen-bonded water that surrounds the substrate, which usually stabilises biomolecules in solution, is lost upon entry into the enzyme's active site. This allows hydrogen bonds to be made with the enzyme instead, thus stabilising the enzyme-substrate complex. Thirdly, the enzyme itself may need to undergo conformational changes to allow proper alignment of catalytic functional groups with the substrate. These functional groups may include amino acid sidechains, cofactors or coenzymes. Collectively, the energy lost by reducing hydrogen-bonded water interactions and the overall movement and rotation of both enzyme and substrate molecules, is compensated for by the binding energy (ΔG_B) of the ES complex. This activation energy acquired is important as it allows the enzyme to reach the activated transition state (ES^\ddagger). It is within this state that there is equal probability of the forward reaction (producing free enzyme (E) and product (P)) or the reverse reaction to occur. The rate of reaction can depend on a number of factors including enzyme concentration, temperature, pH and substrate concentration [S]. For example, when the [S] is high, the formation of the enzyme-substrate (ES) transitional intermediate is favourable and follows an energetically downhill progression to release product and regenerate enzyme. However, as the [P] increases, the reaction rate is reversed. It is therefore important to keep the [S] high in order to promote the forward reaction. In the case of 3'-5' PNPase activity, if the concentration of RNA substrate is high then degradation occurs, however if the NDP product concentration is high then the reverse reaction occurs and NDPs are added. The reversible action of PNPase is shown in Figure 6.1.

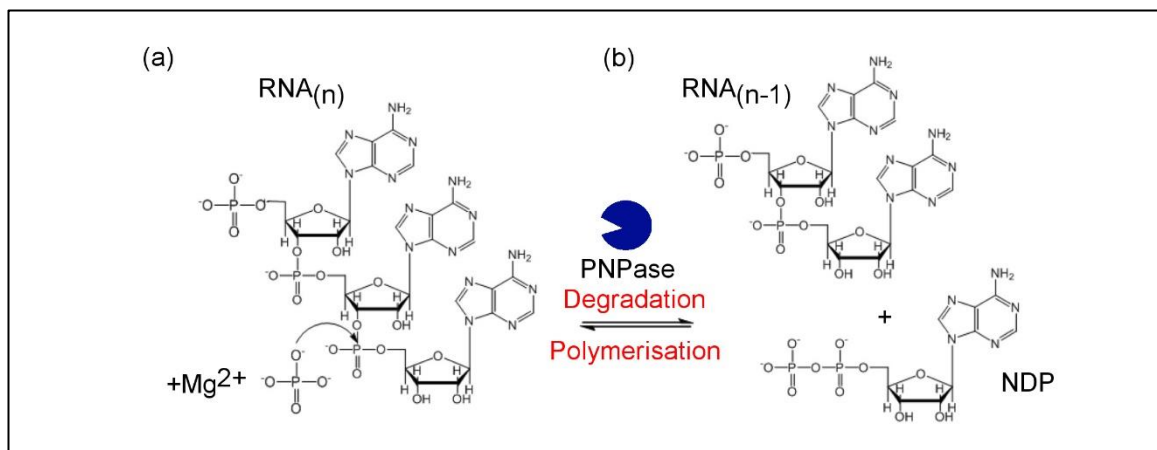


Figure 6.1 PNPase Activity

The reversible action of PNPase on RNA substrates; (a) either degradation in the presence of excess inorganic phosphate (PO_4^{3-}) and a metal ion (e.g. Mg^{2+}) or (b) polyadenylation in the presence of excess ADPs (NDP). Figure adapted from (Nurmohamed *et al.*, 2011) using GIMP (v2) (GIMP, n.d.).

Understanding how enzyme reactions occur (Equation 6.1) is important before trying to determine the kinetics of a particular enzyme: it's essential to ensure kinetic parameters are calculated correctly using initial rates. The initial rates (V_o) are used since the probability of the enzyme encountering a substrate molecule is much higher early in the reaction. More specifically, in order to determine the rate of the forward reaction, the [S] must be in excess and the reaction should not have proceeded for very long. This ensures that the forward reaction is promoted and that the [S] isn't depleted while the [P] increases, which would effectively reverse the reaction. For example, with excess RNA substrate, 3'-5' PNPase homologs will catalyse substrate degradation and the reversible reaction of polyadenylation is minimised. Additionally, in order to prevent the accumulation of ADP in the forward degradation reaction, it is suggested that rates are typically calculated before <20% of the reaction has proceeded. Alternatively, if the backwards polyadenylation reaction requires examining, the [P] must be in excess and in order to prevent the accumulation of RNA, rates must be calculated before <20% of the reverse polyadenylation reaction has occurred. By using the V_o in substrate excess, the kinetic parameters for the 3'-5' degradation can be calculated, while minimising the reverse reaction rate. As mentioned previously, other factors can also affect the rate of reaction and therefore these are kept consistent throughout enzyme assays. These include the assay temperature, enzyme concentration, essential cofactors (i.e. metal ions) and the pH.

In summary, enzyme assay conditions must be optimised to ensure initial rates (V_o) can be calculated correctly. This PhD study has focussed on the degradation activity of PNPase homologs, and in order to determine key kinetic parameters for only 3'-5' RNase degradation, the [RNA] was kept in molar excess to the enzyme concentration. All other factors affecting activity were kept constant and reaction rates were calculated before <20% of the reaction had proceeded.

Ensuring steady state conditions, described above, means that the Michaelis-Menten kinetic equation shown in Equation 6.2, can be utilised to determine key kinetic parameters (reviewed in

Lehninger, Nelson, & Cox, 2000). This is a well-accepted model of enzyme kinetics which relates initial reaction rates (V_o) to the $[S]$. To analyse the kinetic parameters, a range of $[S]$ are titrated against a given enzyme concentration and the initial rates (V_o) of reaction can be measured. By plotting these V_o values against $[S]$ and fitting the data using Equation 6.2, key enzyme parameters can be generated. This includes the maximal velocity (V_{max}) and the Michaelis constant (K_m), which represent the maximum rate at saturating $[S]$ and the substrate concentration at which the V_{max} is half, respectively. If an enzyme reaction follows Michaelis-Menten kinetics, then the calculated parameters can be used to compare enzyme activities. The V_{max} and K_m can be also used to calculate two other useful parameters: The V_{max} and total enzyme concentration (E_T) provides the turnover rate (k_{cat}) (Equation 6.3) and the ratio of k_{cat}/K_m provides a good measure of catalytic efficiency (Equation 6.4).

$$V_o = \frac{V_{max} \cdot [S]}{K_m + [S]}$$

Equation 6.2 Michaelis-Menten

The initial rate (V_o) is a function of the extrapolated maximum velocity (V_{max}), substrate concentration $[S]$ and the Michaelis Constant (K_m).

$$k_{cat} = V_{max}/[E_t]$$

Equation 6.3 Turnover Rate

Turnover rate (k_{cat}) is a ratio of the extrapolated maximum velocity (V_{max}) and total enzyme concentration (E_t). The k_{cat} corresponds to the number of substrate molecules converted per second.

$$\text{catalytic efficiency} = k_{cat}/K_m$$

Equation 6.4 Catalytic Efficiency

Catalytic efficiency is a ratio of turnover rate (k_{cat}) and the Michaelis Constant (K_m).

Previous chapters and publications have provided information about 3'-5' RNase activity using gel-based assays. A real-time plate reader assay could however quickly acquire initial rates, and thus catalytic parameters, more easily than the previously established laborious methods. The potential to determine enzyme activity is valuable since it may reveal preferred substrate targets or novel inhibitors. In addition, conducting this assay in high-throughput would provide an efficient way to determine if any other metabolites, in addition to those outlined within this study, have an effect on the 3'-5' activity of PNPase homologs. Therefore, the subsequent Section 6.1.3 describes the steps required in order to optimise a commercially available fluorescence-based plate reader assay, for collecting 3'-5' kinetic data for PNPase homologs.

6.1.3 Determining 3'-5' RNase Kinetics

A method of real-time (RT) detection of 3'-5' phosphorolytic RNase activity was developed using components of the commercially available Transcreeper® ADP2 Fluorescent Intensity (FI) assay kit (BellBrook, n.d.). The manual recommended that the assay was compatible with any enzyme

class that produces ADP, including protein, lipid, and carbohydrate kinases, ATPases, DNA helicases, carboxylases and glutamine synthetase (BellBrook, n.d.). This FI kit has typically been used for determining the kinetics of kinases (Hong, Quinn, & Jia, 2009) and so this study provided a novel application for the assay.

This assay provided a means of quantifying ADP production, which when applied in a plate reader format allowed high-throughput, sensitive and fast changes in fluorescence intensity to be monitored. In summary, when substrate RNA was intact, the ADP Alexa594 Tracer (ADP*) fluorescence was quenched by a mouse monoclonal ADP2 Antibody-IRDye® QC-1 (Q-Ab). However, upon 3'-5' RNA degradation, ADP was produced which competed with ADP* for Q-Ab binding. An increase in fluorescence was detected when the ADP* was displaced from the Q-Ab quencher, see Figure 6.2. As the ADP released from the RNA degradation was sequestered by the Q-Ab quencher, it made this assay particularly well suited for studying the degradation activity of 3'-5' PNPase homologs, as the reverse polymerisation reaction was reduced. Another great advantage of this particular RT assay was the option to use a wide range of unlabelled RNA substrates. If the RNA substrate sequence is known to contain adenosines and ADP is released upon degradation, it should be suitable for use within this assay. As a result, this allowed the option for testing a range of physiologically relevant PNPase or archaeal exosome RNA substrates. Moreover, the generation of an ADP standard allowed relative fluorescence units (RFU) values to be converted into concentrations of ADP; essentially quantifying the degradation of RNA. Another advantage for using this high-throughput design was the possibility for obtaining real-time kinetic information during screening for inhibitor compounds.

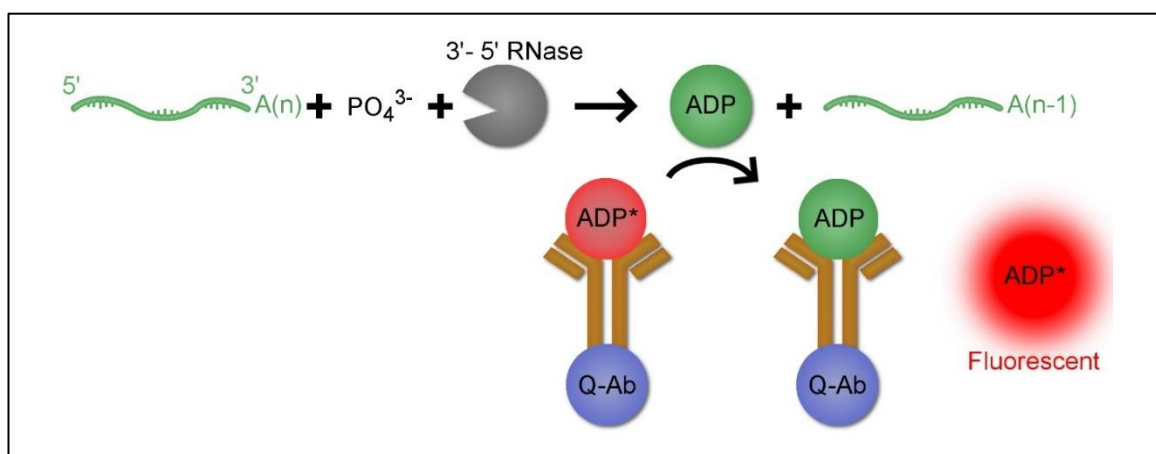


Figure 6.2 ADP Detection Plate Assay

Substrate RNA with a 3' polyA tail (A(n), green line) is degraded by 3'-5' RNase using phosphate (PO₄³⁻) in a sequential manner to produce RNA 3' A(n-1) (green line), releasing ADP in the process (green circle). The fluorescent ADP* (dull red) is initially quenched when bound to Q-Ab. However, upon RNA degradation and ADP release, ADP* is displaced and the fluorescence of the non-quenched ADP* increases (bright red). Image adapted from Transcreener® ADP2 Fluorescent Intensity (FI) Assay manual (BellBrook, n.d.) and created using GIMP (v2) (GIMP, n.d.).

Unfortunately, the ADP* assay component of the Transcreener kit was supplied in a buffer containing 0.5 mM KH₂PO₄ and due to the phosphate sensitivity of PNPase from *H. sapiens*

(Portnoy et al., 2007) the assay could not be utilised to determine hPNPase enzyme kinetics. Therefore, the results presented in Section 6.4 only examine PNPase activity from *E. coli* and the archaeal exosome from *S. solfataricus*.

6.1.3.1 Plate Reader Set-up and Optimisation

Although a manual was provided with the Transcreener® ADP2 Fluorescent Intensity (FI) Assay, the conditions recommended by Bellbrook were optimal for end-point ADP detection. The manual specified that the recommended assays conditions would require further optimisation if used in RT. Therefore, each parameter of the assay was checked, including the plate reader set-up and ADP*/Q-Ab concentrations, so that the optimal RFU window could be obtained. For example, the Transcreener ADP2 FI Assay Alexa594 Tracer (ADP*) had been successfully used at excitations of (580-590 nm) and emissions of (610-620 nm) with bandwidths of 10 nm in black 96-well plates, for a range of plate readers (BellBrook, n.d.). However, the optimal settings for the H1 Synergy BioTek plate reader, utilised in these experiments, had not been previously determined. Hence the specific settings for detecting fluorescence intensity (RFU) were determined experimentally by conducting an excitation/emission scan, with a set concentration of ADP*, for maximal RFU detection.

Once the ideal excitation/emission plate reader settings were determined from initial optimisation experiments, it was also important to determine the maximum fluorescence window. The optimal gain (sensitivity to FI), plate read height and concentrations of fluorophore (ADP*) and quencher (Q-Ab) needed to be established, so that the recommended signal difference of >5 fold, between low RFU (quenched ADP* & Q-Ab) and high RFU (ADP* only) could be achieved. Consequently, all of these parameters were verified during the optimisation steps outlined in the methods Section 6.3.1-6.3.2.

6.1.3.2 Optimisation of Assay Conditions

As mentioned previously, the enzyme concentration and substrate concentration must be optimal; [S] must not be limiting and should be present in excess to [E_T] to ensure the initial rates determined are under steady state conditions, thus Michaelis-Menten kinetics can be correctly applied. The conditions of the plate reader were therefore optimised, including [enzyme] and [RNA]. An ADP standard curve was also produced so that raw RFU data could be converted into the [ADP (nM)].

Initial experiments were conducted in order to determine the optimal enzyme concentration. A range of enzyme concentrations were titrated against a constant [RNA], and the lowest enzyme concentration that produced a detectable initial rate, was determined. It was useful to select the lowest [enzyme] as this not only minimised protein usage, which was cost efficient, but also helped ensure that the [RNA] could be in a large excess whilst minimising the use of substrate in each reaction. In terms of obtaining a detectable signal, the kit recommended selecting the [E_T] which

gave rise to a 50-80% change in RFU. These raw RFU data, measured over time, were then converted (using the ADP standard curve mentioned previously) into [ADP] (nM). The inverse experiment, where [enzyme] was kept constant and [RNA] was titrated, was useful since the resulting data helped determine if the enzyme followed typical Michaelis-Menten kinetics. More specifically, initial rates (V_o) were plotted against the [S] and if the data followed a typical Michaelis-Menten hyperbolic shaped curve, the equation described previously (Equation 6.2) was then applied.

6.2 Aims

As demonstrated in previous chapters, gel-based assays are available for determining the degradation activity of PNPase homologs. However detailed kinetics using high-throughput techniques are not currently available. Thus, the aim of this study was to repurpose a commercially available fluorescence-based high-throughput plate reader assay, typically used to determine end-point kinase activity, for use in quantifying the activity of PNPase homologs in real-time. Firstly, using the modified Transcreener® ADP2 Fluorescent Intensity (FI) (BellBrook, n.d.) assay, the study aimed to monitor the release of ADP product from the degradation of an unlabelled poly(A)_{20mer} RNA substrate. Secondly, using an ADP standard curve, the study aimed to provide a novel way to quantify the RNA degradation kinetics mediated by EcPNPase and SsoExosome. In order to meet these two objectives and to develop a quantifiable real-time assay, a range of initial method validation steps were required. Key questions addressed within this work will validate the plate reader parameters and concentration of reagents, prior to generating a standard curve which will allow RFU values to be converted into the [ADP] produced. Whether kinetic information can be calculated, including the V_{max} and K_m , k_{cat} and catalytic efficiency, will also be addressed. The resulting data will be provided in Section 6.4 and important conclusions, regarding the benefits and limitations of the high-throughput plate reader assay, are then summarised in Section 6.5.

6.3 Methods

A variety of optimisation experiments were initially conducted for the high-throughput plate reader assay, prior to determining protein-specific canonical enzyme kinetics. This included optimising instrument set-up, determining the assay RFU window, generating an ADP standard curve and conducting an enzyme titration (Sections 6.3.1-6.3.5 respectively). Michaelis-Menten kinetic parameters were finally calculated from an RNA titration, as described in Section 6.3.5.

For clarity, concentrations of RNA and ADP corresponded to one poly(A) 20 mer RNA substrate (RNA_{20mer}) and twenty ADP molecules (ADP_{20x}); effectively a 2000 nM stock of RNA_{20mer} would produce 2000 nM ADP_{20x}. Unless stated otherwise, this method of recording the RNA and ADP concentration was used as standard throughout this chapter. The purpose of this was to allow easier comparisons to be made between the amount of substrate degraded and product produced, thus all kinetic constants were expressed for the polymer.

6.3.1 Instrument Set-up

Transcreener ADP2 FI Assay Alexa594 Tracer (ADP*) (4 nM) was prepared in 100 μ l assay buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 10 mM MgCl₂, 0.5 mM KH₂PO₄, 0.05 mM HEPES and 0.00025% Brij-35), in a sterile black Nunc 96-MicroWell plate (96F) (Fisher # 237107) and a range of excitation and emission wavelengths (nm) were scanned at 25°C, ensuring the bandwidth was < 10 nm to avoid signal overlap. The optimal settings, which provided the maximum fluorescence intensity (RFU) for the BioTek H1 Synergy plate reader, were determined.

6.3.2 Determining Detection Conditions (Q-Ab & ADP*)

The kit recommendations for obtaining a maximal signal window between low RFU (quenched ADP* & Q-Ab) and high RFU (ADP* only) were only provided for end-point assays. Hence, appropriate concentrations of ADP* and Q-Ab, to allow real-time data collection, were confirmed experimentally. Using the optimised excitation/emission settings previously established (Section 6.3.1), a range of Q-Ab concentrations were titrated against 4 nM ADP* in reaction buffer. This experiment was then reversed, titrating ADP* against Q-Ab (40 μ g/ml) in the same buffer conditions used previously. Recorded RFU data at 37°C were collected and analysed to determine the total assay window; the concentration of ADP* and Q-Ab, which gave rise to > 5-fold signal difference between low and high RFU, was selected. Subsequently, using these ADP* and Q-Ab concentrations, the optimal gain and read height settings were also determined.

6.3.3 ADP Standard Curve

Once the assay conditions and equipment settings were optimised, a range of ADP_{20x} concentrations (0, 40, 80, 120, 160, 200, 300, 400, 600, 800, 1200, 2000 nM) were prepared which corresponded to a 2000 nM RNA_{20mer} conversion of 0, 2, 4, 8, 10, 15, 20, 30, 40, 60, 100%. A 96-well plate, containing all the reagents for the reaction, was pre-incubated for 30 minutes at 25°C. After this time, the plate was ejected and 10 μ l of each ADP_{20x} concentration (10x stock, pre-incubated in adjacent wells) was added to the relevant 90 μ l assay well. A final 100 μ l assay with a range of ADP_{20x} concentrations (all in 4 nM ADP*, 40 μ g/ml Q-Ab and reaction buffer) was produced and plates were re-inserted back into the plate reader. The plate reader was set to an excitation and emission wavelength of 585 and 616 nm respectively (10 nm bandwidth) and using top optics, xenon flash with lamp energy set to high, gain of 150 and a plate read height of 7.75 mm, the RFU data were recorded.

The RFU data were saved as an Excel file and average RFU values between 30-90 minutes were calculated. An ADP standard curve was produced in Minitab (v17) by plotting the average RFU (y axis) against $\ln([L]+1)$ (x axis), where $[L] = [\text{ADP}_{20x} \text{ (nM)}]$ (Minitab, 2010). The data were then fitted using nonlinear regression, as recommended within the assay manual (BellBrook, n.d.). Following conversations with Dr David Whitley (Mathematician at the University of Portsmouth)

the logistic growth function in Minitab (Figure 6.3 and Equation 6.5) was recommended for fitting the standard curve data (Minitab, 2010). Using values from the response curve, (for example the maximum RFU (in high [ADP_{20x}], termed Theta1) and the quenched RFU (in low [ADP_{20x}], termed Theta2)) as initial estimates for fitting, the equation shown in Equation 6.5 fitted data to a sigmoidal shaped curve. This standard curve allowed RFU values (y) to be converted into [ADP] by using the re-arranged equation shown in the red box in Equation 6.5.

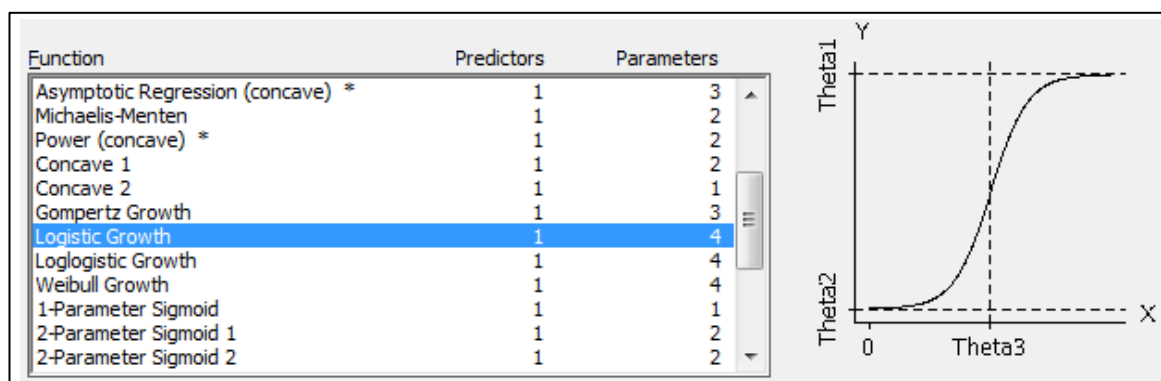


Figure 6.3 Minitab Logistic Growth Function

The relationship between Theta parameters (Theta 1-4) of the non-linear regression logistic function (highlighted blue) are shown in the graph. Snapshot image taken directly from Minitab17 program (Minitab, 2010).

$$y = \frac{\text{Theta1} + (\text{Theta2} - \text{Theta1})}{1 + (\exp(\frac{x - \text{Theta3}}{\text{Theta4}}))}$$

x axis = [L] = ln([ADP_{20x}] + 1)
y axis = RFU

$$[\text{ADP}_{20x}] = \exp(\text{Theta3} + \text{Theta4} \cdot \ln(\frac{\text{Theta2} - \text{Theta1}}{\text{RFU} - \text{Theta1}}) - 1) - 1$$

Equation 6.5 Minitab Non-linear Regression Logistic Function

Relationship of the RFU (y axis) to the concentration of ligand ([L] = ln[ADP_{20x} (nM)] + 1) (x axis) for the ADP standard curve is shown. The equation from Minitab was rearranged so that RFU values could be converted into a [ADP_{20x} (nM)] using Theta values 1-4.

6.3.4 Enzyme Titration

A 96-well plate, containing all the reagents for the reaction, was pre-incubated for 30 minutes at 25°C. After this time, the plate was ejected and 10 µl of RNA_{20mer} (10x stock, pre-incubated in adjacent wells) was added to the relevant 90 µl assay well (containing each EcPNPase concentration) to start the reaction. The consequent 100 µl assay, with a range of EcPNPase concentrations (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75 and 1 nM), all with 2000 nM RNA_{20mer}, 4 nM ADP*, 40 µg/ml Q-Ab in reaction buffer was produced. Plates were re-inserted back into the plate reader and, using the plate reader settings described in Section 6.3.3, RFU data were collected over 12-hours. Controls were conducted as standard in every plate, including a buffer blank (No ADP*), low RFU (quenched ADP* & Q-Ab only) and high RFU (ADP* only). To determine the assay window, a reaction without enzyme was also conducted to ensure RNA integrity; so that enzyme-

specific degradation could be confirmed. The RFU values from EcPNPase titration were converted to ADP_{20x} concentrations (nM) using an ADP standard curve (Section 6.3.3). Resultant [ADP_{20x}] (nM) were then plotted against time (minutes). The first 15 minutes of data were removed, as recommended by the Kit manual (BellBrook, n.d.); allowing proper ADP*-Q-Ab thermal equilibration. The assay conditions were optimised to ensure the initial rates were not affected by removing data from between 0-15 minutes. The remaining data were then further processed; only values corresponding to less than 5% reaction completion were retained for calculating initial rates. These V_o values were then plotted against the [EcPNPase] in GraFit (Leatherbarrow, 2009) and the optimal [EcPNPase] (nM) was determined. The same method described for EcPNPase was repeated to determine the optimal concentration of SsoExosome, except a final concentration range of 0, 1, 2.5, 3, 3.5, 4, 5, 10 nM SsoExosome were titrated against a set concentration of RNA_{20mer} (2000 nM).

6.3.5 RNA Substrate Titration

A range of RNA_{20mer} substrate concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000 nM) were titrated against the previously determined appropriate concentration of each enzyme (Section 6.3.4). A 96-well plate, containing all the reagents for the reaction, was pre-incubated for 30 minutes at 25°C. After this time, the plate was ejected and 10 µl of each [RNA_{20mer}] (10x stock, pre-incubated in the adjacent well) was added to the relevant 90 µl assay well, to start the reaction. Final 100 µl assays, with a range of [RNA_{20mer}] and a consistent [enzyme] (either 0.5 nM EcPNPase or 4 nM SsoExosome), 4 nM ADP*, 40 µg/ml Q-Ab in reaction buffer were produced. Plates were re-inserted back into the plate reader and RFU data were recorded as described in Section 6.3.4. The RFU values from the RNA_{20mer} titration were converted to ADP_{20x} concentrations (nM) using an ADP standard curve (Section 6.3.3). Resultant [ADP_{20x}] (nM) were then plotted against time (minutes) and the first 15 minutes of data were removed, as recommended previously. The remaining data were then further processed; only values corresponding to less than 10% reaction completion were retained for calculating initial rates. These V_o values were then plotted against [RNA_{20mer}] substrate in GraFit and fitted with the Michaelis-Menten equation shown in Equation 6.2 (Leatherbarrow, 2009). Data were fitted and initial estimates were provided by use of linear fitting using the Lineweaver-Burk rearrangement; key parameters including V_{max} and K_m were determined for each enzyme (Leatherbarrow, 2009). In order to determine the optimal concentration of phosphate, the same method described for determining EcPNPase kinetics in the standard reaction buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 10 mM MgCl₂, 0.5 mM KH₂PO₄, 0.05 mM HEPES and 0.00025% Brij-35), was conducted with higher concentrations of phosphate; in these assays, the reaction buffer instead contained either 5 or 10 mM KH₂PO₄.

6.4 Results

This section firstly provides results from optimisation experiments which determined the optimal instrument (Section 6.4.1) and assay conditions for producing a maximal RFU window (Section 6.4.2-6.4.3). The application of a standard curve to convert RFU values into [ADP_{20x}] and allow quantification of ADP production, upon RNA degradation, was examined in Section 6.4.4. The utilisation of the assay, to quantify the RNA degradation activity of the prokaryotic PNPase from *E. coli*, was examined in Section 6.4.5. The optimal concentration of EcPNPase, for determining initial rates of RNA_{20mer} degradation, was investigated in Section 6.4.5.2, before the Michaelis-Menten kinetic parameters were calculated for EcPNPase, in the presence of a range of phosphate concentrations (0.5, 10 and 5 mM, Sections 6.4.5.3-6.4.5.5 respectively). Although more repeats were required, the high-throughput plate reader assay was successfully used to quantify EcPNPase mediated RNA_{20mer} degradation in real-time. Kinetic parameters including V_{max} and K_m , k_{cat} and catalytic efficiency were also determined and this assay validation led to the proposal that other PNPase homologs may also be tested. Consequently, the ability to quantify the RNA degradation activity of the archaeal exosome from *S. solfataricus* was examined in Section 6.4.6. As before, the optimal concentration of SsoExosome was determined in Section 6.4.6.2, prior to calculating the Michaelis-Menten kinetic parameters in Section 6.4.6.3. Unlike Section 6.4.5, only one phosphate concentration (0.5 mM) was tested and following the observation that EcPNPase had a phosphate preference greater than 5 mM, the effects of higher phosphate levels on SsoExosome activity requires investigation. Although the assays for EcPNPase and SsoExosome need to be repeated in triplicate, the data provided within this section demonstrate the proof of concept for repurposing the Transcreener® ADP2 Fluorescent Intensity (FI) assay kit. In summary, results presented within this section suggest that the high-throughput plate reader assay may be used to examine the degradation activity of PNPase homologs across evolutionarily diverse organisms; from prokaryotes and archaea, in real-time.

6.4.1 Instrument Set-up Optimisation

The optimal excitation and emission wavelengths of the Transcreener ADP2 FI Assay Alexa594 Tracer (ADP*) were determined using a BioTek H1 Synergy plate reader as described in Section 6.3.1. The maximum fluorescence intensity was observed by using top optics (excited and read from above) with an excitation and emission wavelength of 585 and 616 nm respectively (10 nm bandwidth). This aligned to the recommended excitation (580-590 nm) and emission (610-620 nm) settings, which were optimal for other plate readers outlined in the technical manual (BellBrook, n.d.). Automatic gain was suggested within the manual, however to avoid plate to plate variations, gain was set to 150 for these experiments. These optimised plate reader settings were used as standard throughout this chapter.

6.4.2 Optimising Q-Ab & ADP* Detection Conditions

The Transcreener ADP2 FI Assay was developed to follow the progress of any enzyme that produces ADP. The examples within the manual (BellBrook, n.d.) utilised ATP as the substrate for the ADP producing reaction; however, in this work RNA_{20mer} was the substrate instead. Additionally, the conditions recommended within the Bellbrook manual were optimal for end-point ADP detection using a ‘stop’ buffer. It was advised that these conditions may be sub-optimal if the assay buffers were changed for real-time ADP detection. With this in mind, the recommendations within the Bellbrook kit manual, for end-point ADP detection were first considered in Section 6.4.2.1, before the assay was modified for real-time ADP detection as described in Section 6.4.2.2.

6.4.2.1 Bellbrook Recommended Conditions for End-Point ADP Detection

The kit provided a Q-Ab antibody with a limited selectivity for adenosine diphosphate (ADP) vs. adenosine triphosphate (ATP). With this in mind, Bellbrook suggested that the concentration of Q-Ab and ADP* determined the total assay window, but the amount of ATP substrate could affect this. More specifically, Bellbrook suggested that the relationship between substrate, in their case [ATP], and [Q-Ab] was linear between 0.1 μM to 1,000 μM ATP (BellBrook, n.d.). Thus, a method for calculating the amount of Q-Ab to use with a known amount of ATP substrate, was also provided within their guidelines. Specifically, the technical manual proposed that the amount of Q-Ab required for enzyme reactions that use ATP, could be determined from $y = 0.93x + 0.7$ where $x = [\text{ATP } \mu\text{M}]$ and $y = [\text{Q-Ab } \mu\text{g/ml}]$. This equation was used, with a few assumptions, to predict the amount of Q-Ab required when instead of ATP, RNA_{20mer} was the substrate used to generate ADP. For a basic estimation, it was assumed that 2 μM RNA_{20mer} was equivalent to 40 μM ATP and thus $[0.93 \times 40] + 0.7 = 37.9$ μg/ml of Q-Ab was recommended. The other assumptions made in this calculation were that the assay conditions were similar to those described in the technical manual; however, this was not the case. Not only was an RNA_{20mer} substrate used instead of ATP, but the assay buffer conditions were modified to allow real-time fluorescence measurements, rather than the end-point method previously described in the manual. As a result, the recommended concentration of Q-Ab (37.9 μg/ml) was a rough estimation.

It was essential to ensure the amount of Q-Ab used in the 100 μl assay was optimal for generating a large RFU window and thus a good ADP detection range. Therefore, titration experiments were conducted in the presence of 2 μM RNA_{20mer} substrate, to more accurately determine the amount of fluorophore (ADP*) and quencher (Q-Ab) required for detecting ADP production in real-time (Section 6.4.2.2).

6.4.2.2 Optimal Conditions for Real-Time ADP Detection

Titration of ADP* against Q-Ab (40 μg/ml), and the inverse titration of Q-Ab against ADP* (4 nM) was conducted as described in Section 6.3.2. Following an initial pre-incubation step for 30 minutes at 37°C, the RFU measured between 30-60 minutes was averaged; this ensured the RFU

had time to stabilise. The average RFU was then plotted against the titrant concentration (either ADP* or Q-Ab) and data points, which followed a linear trend, were fitted to a linear equation using GraFit (Figure 6.4) (Leatherbarrow, 2009). Calculating the concentration of ADP* (x axis) that produced an RFU (y axis) equal to zero (no fluorescence), in the presence of 40 µg/ml of Q-Ab, denoted complete quenching. Consequently, using the linear equation for the data fit in Figure 6.4 (a) ($0 = 679.1429 ([ADP^*]) - 1491$), it was suggested that ~2.2 nM ADP* will be quenched by 40 µg/ml Q-Ab ($ADP^* = 1491/679.1429$). Hence a concentration slightly greater than 2.2 nM ADP* should be enough to ensure 40 µg/ml of Q-Ab is fully saturated by ADP*.

Analysis of the inverse experiment titrating Q-Ab against ADP* (Figure 6.4 (b), $0 = -161.1571 ([Q-Ab^*]) + 7725$) correlated well and suggested that 47.9 µg/ml Q-Ab was required to fully quench 4 nM ADP* ($Q-Ab = -7725/-161.1571$). Using a lower concentration of 40 µg/ml Q-Ab ensured that 4 nM ADP* was in a slight excess and saturated Q-Ab appropriately.

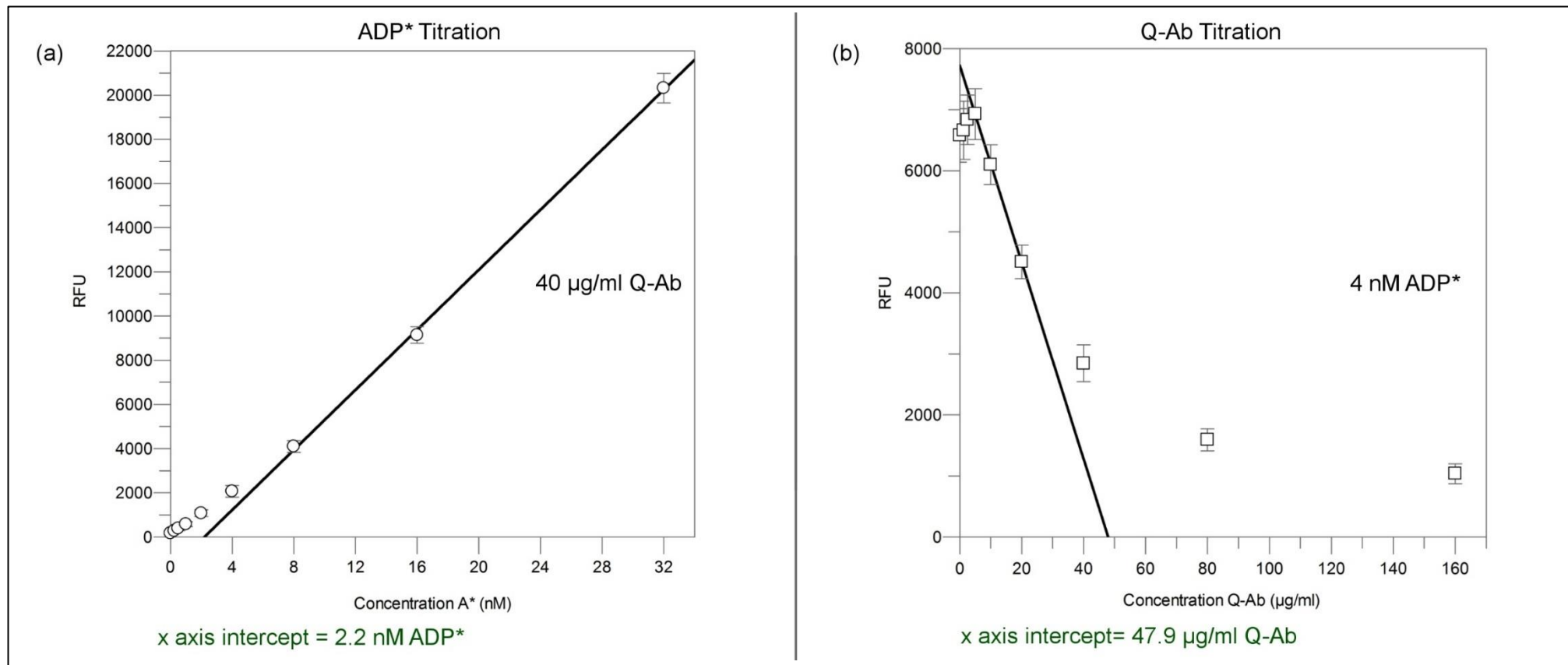


Figure 6.4 Plate Optimisation: ADP* & Q-Ab Concentration Range

The average RFU (30-60 minutes) for the titration of (a) ADP* against 40 µg/ml Q-Ab and (b) Q-Ab against 4 nM ADP* is shown. The linear equation ($y = m \cdot x + c$, where $y = \text{RFU}$, $m =$ gradient, $c =$ intercept of y axis) was fitted to appropriate linear data points using GraFit. The x axis intercept for (a) $y = 679.1 \cdot x - 1491.0$ and (b) $y = -161.2 \cdot x + 7725.5$ was calculated and is shown in green. Graphs were plotted in GraFit and labelled in GIMP (v2) (GIMP, n.d.; Leatherbarrow, 2009).

In summary, results indicated that the ADP* was adequately quenched and Q-Ab was appropriately saturated at concentration > 2 nM ADP* and of 40 $\mu\text{g/ml}$ respectively. Therefore, by using 40 $\mu\text{g/ml}$ Q-Ab to quench a slightly higher concentration of ADP* (> 2 nM), the release of ADP from the enzyme reaction could be directly measured since it was not sequestered by free Q-Ab (unsaturated with ADP*) but directly competed with ADP* for Q-Ab binding. The resulting increase in RFU could therefore be utilised as an indication of the rate of RNA degradation and ADP production. Thus, the final plate assay conditions included 4 nM ADP* and 40 $\mu\text{g/ml}$ Q-Ab in reaction buffer.

6.4.3 Determining Optimal Plate Parameters

Once the optimal plate reader settings and ADP* and Q-Ab concentrations had been determined, experiments titrating PNPase/archaeal exosome and RNA_{20mer} were conducted at 37°C. Unfortunately, assays conducted at this temperature were not very reproducible (data not shown) and calculations of enzyme initial rates were not possible due to issues with the RFU stability (ADP* quenching by Q-Ab was not stable). Following consultation with the plate assay providers, as details of incubation temperatures were not provided in the technical manual, it was suggested that the assay temperature may potentially influence the rate of ADP* quenching by Q-Ab. In order to investigate this further, the RFU stability of 4 nM ADP* in the presence of 40 $\mu\text{g/ml}$ Q-Ab was compared at 37°C and 25°C (Figure 6.5).

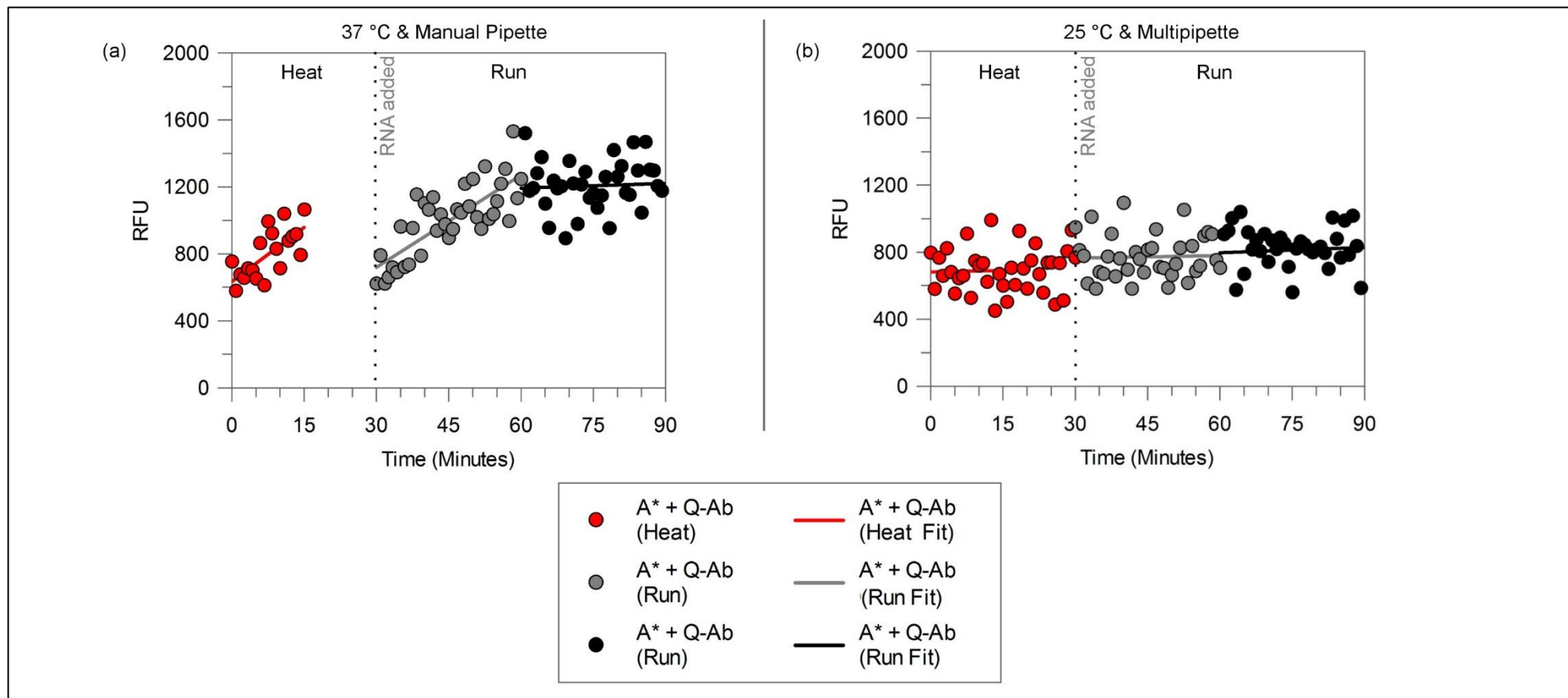


Figure 6.5 Optimising ADP* Quenching by Q-Ab.

The effect of assay temperature on ADP* quenching by Q-Ab ($A^* + Q\text{-Ab}$) was compared at (a) 37°C and (b) 25°C. Identical assays containing 4 nM ADP* with 40 µg/ml Q-Ab in reaction buffer were prepared by either (a) manual pipetting or (b) using multi-pipetting. The RFU values recorded during the plate pre-incubation (heat) step are plotted (red points) against time (minutes), for both graphs (a) and (b). Note only the first 15 minutes of the heat run were recorded in (a) whereas data for 30 minutes were recorded in (b). A linear ($y=m*x + c$) fit (red (heat fit) line) is also shown. The RFU values recorded post-incubation, after RNA was added were also plotted (run), data from the first and second 30 minutes (coloured grey and black points respectively) were individually fitted to a linear equation (coloured grey and black (run fit) lines respectively). Graphs were plotted in GraFit and labelled in GIMP (v2) (GIMP, n.d.; Leatherbarrow, 2009).

Both assays shown in Figure 6.5 were set up using a manual pipette, and before the RNA was added, the RFU values were recorded during a plate pre-incubation (heat) step (red data points, Figure 6.5) at either 25°C or 37°C. The RFU values measured during this heat step increased from ~ 600 to 1200 at 37°C, whereas RFU values at 25°C remained relatively stable at ~ 700-800. Hence, in order to ensure stable RFU in subsequent plate reader assays, all reagents were pre-warmed at 25°C for 30-minutes prior to adding the RNA substrate and thus initiating the reaction. In order to reduce the time that the plate was not kept at constant temperature (25°C) and thus avoid any effects of room temperature fluctuations whilst initiating the reactions, a multi-pipette was also utilised to add the RNA in subsequent assays.

The assays shown in Figure 6.5 were set up without EcPNPase, hence when the RNA was added (dotted vertical line, Figure 6.5) no degradation occurred and the ADP* remained quenched by Q-Ab. This allowed the stability of ADP* quenching by Q-Ab to be examined following the process of RNA addition. The kit recommended that mixing ADP* and Q-Ab required a 15-minute equilibration before RFU quenching was stabilised, after which time, data could be collected. However, the results shown in Figure 6.5 (a) suggested that after RNA addition (grey dotted vertical line), the RFU recorded for the first 30 minutes during the 'run' at 37°C (grey data points) were still not stable, despite a 30-minute pre-incubation (heat) step. It appeared that the ADP* quenching by Q-Ab was still equilibrating for longer than the recommended 15 minutes. In contrast, a stable RFU was recorded throughout the first hour at 25°C (grey and black data points, Figure 6.5 (b)). Hence, as standard, assays were conducted at 25°C and after starting the enzyme reaction, RFU data collected for the first 15 minutes were automatically discarded to allow adequate time for ADP* quenching by Q-Ab. It is noteworthy to mention, that enzyme assays were optimised to ensure that sufficient data could be collected with this 15-minute pre-equilibration period. To achieve this, the enzyme concentration was optimised so that less than 0.1% of the initial reaction occurred within the first 15 minutes, thus the initial rates could still be calculated before <10% of the reaction had proceeded.

The RFU assay window, for these optimised conditions at 25°C (4 nM ADP* and 40 µg/ml Q-Ab, Figure 6.5 (b)), was originally shown in Figure 6.5 (b). However, in order to compare the RFU stability over the whole assay duration, RFU values recorded for 12-hours are shown in Figure 6.6. The RFU recorded within the maximum ADP* fluorescence well (A* only) was stable at ~ 14,000 RFU for 12 hours (Figure 6.6 (a), bright red data points). The quenched assay, containing both ADP* (A*) and Q-Ab, had an RFU of ~ 800, as shown in Figure 6.6 (a) and in the zoomed in image in Figure 6.6 (b) (dark red data points). A blank well containing no sample, and the well containing only Q-Ab, both had a background RFU of ~ 200, as shown in Figure 6.6 (a) and in the zoomed in image in Figure 6.6 (b) (white and blue coloured data points respectively).

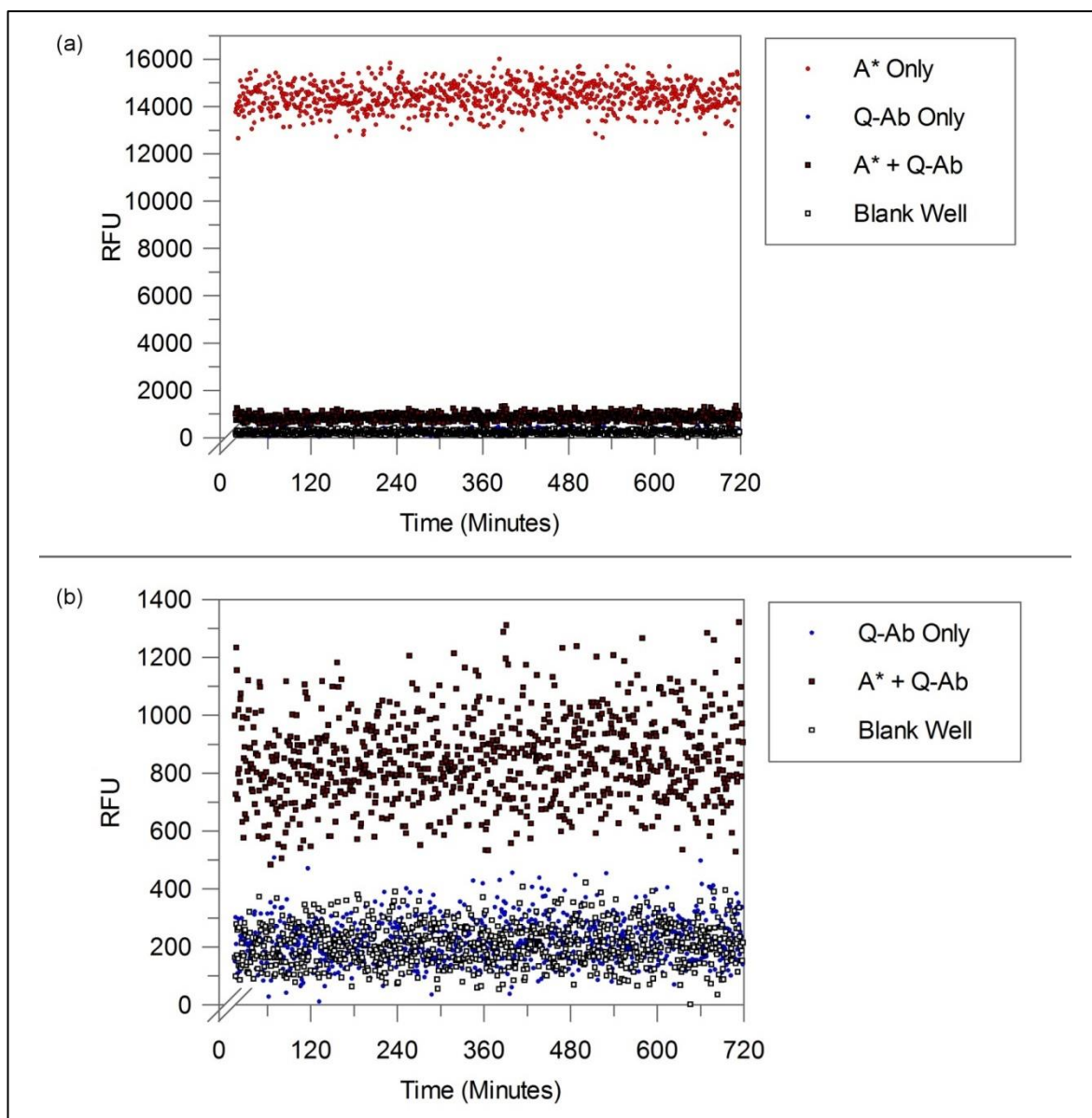


Figure 6.6 Assay Window Data

(a) The stabilities of the maximum RFU signal (unquenched ADP* (A*), bright red data points) and the low RFU signal (quenched ADP* and Q-Ab (A*+Q-Ab), dark red data points) were recorded over time at 25°C. The first 15 minutes were discarded as standard to allow ADP*+Q-Ab equilibration. Controls containing no sample (blank well, white data points) and no fluorescence (Q-Ab only, blue data points) were also plotted to show the background RFU. (b) The same graph was also shown zoomed in at (<1,400 RFU) in order to visualise the RFU of the quenched A*+Q-Ab reaction (~800) and the RFU of Q-Ab and blank well background controls (both ~200 respectively). Graphs were plotted in GraFit and labelled in GIMP (v2) (GIMP, n.d.; Leatherbarrow, 2009).

As summarised in Figure 6.7, the low RFU (quenched: 4 nM ADP* & 40 µg/ml Q-Ab) and high RFU (unquenched: 4 nM ADP* only) data points over the 12-hour assay (Figure 6.6) fitted to a linear equation. Furthermore, the quenched and unquenched RFU values calculated from the y-axis intercept, demonstrated that the RFU window (809-14,278 RFU respectively) was greater than that recommended within the manual (> 5-fold). As indicated by the gradient of both fitted lines in Figure 6.7, the quenched and unquenched signals were also stable over 12 hours, indicating that RFU data collected within this time period would not be subject to ADP* fluorescence decrease or reduced quenching by Q-Ab over time. This was important to check as any effect on RFU, which

was not a result of the PNPase-mediated RNA degradation reaction itself, could influence the accuracy of calculating ADP production and thus the enzyme's activity.

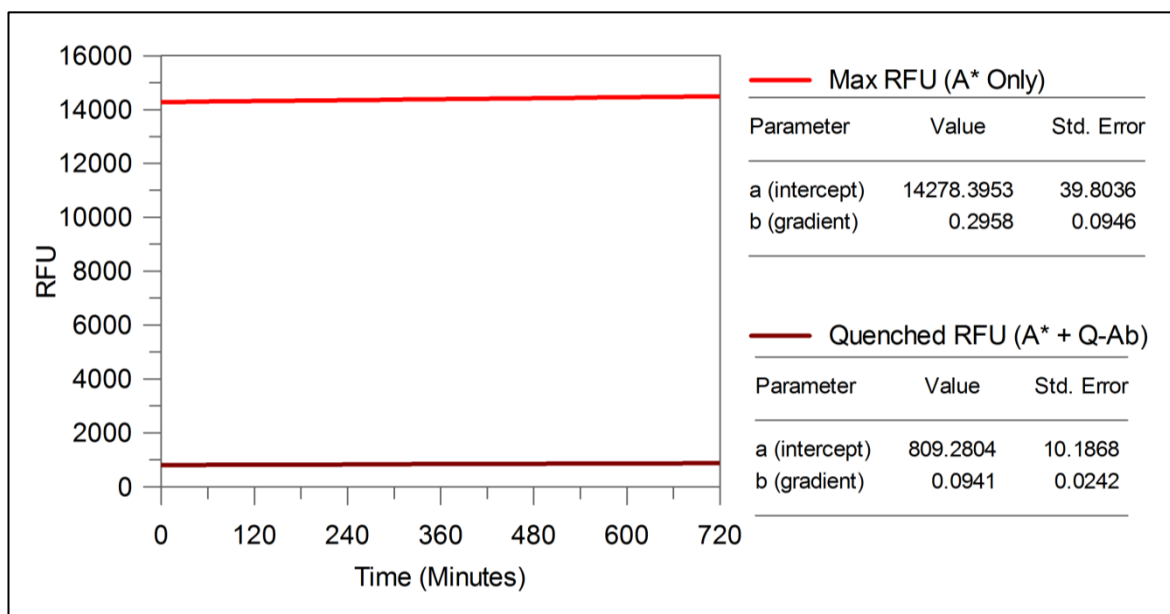


Figure 6.7 Assay Window Fit

A linear ($y=b*x + a$) fit was applied to the RFU window data (shown in Figure 6.6) to check the stability of the maximum RFU signal (unquenched ADP* (A* only, bright red line)) and quenched RFU signal (ADP* and Q-Ab (A*+Q-Ab), dark red line) at 25°C. The first 15 minutes were discarded to allow proper ADP*-Q-Ab equilibration before a linear fit was applied, see table inserts for gradient (b) and y-intercept (a) values. The graph was created in GraFit (Leatherbarrow, 2009).

In summary, all plate reader assays were pre-incubated at 25°C for 30 minutes to allow the buffer components to reach the correct temperature and to ensure ADP* fluorescence stabilisation. Following this plate pre-incubation, the reaction was started by adding RNA using a multi-pipette. The data recorded from the first 15 minutes of the actual run were automatically discarded to ensure adequate time for ADP* quenching by Q-Ab (pre-equilibration). During these preliminary experiments it was also found that the substrate RNA_{20mer} was thermally-stable under the conditions tested (30-minute pre-incubation step at 25°C), conversely EcPNPase appeared to lose some activity (data not shown) and therefore the RNA_{20mer} was used to start the reactions. After optimising the assay conditions to obtain the maximum RFU window and ensure stable signal for real-time detection of ADP, an ADP standard curve was generated as described in Section 6.4.4.

6.4.4 ADP Standard Curve

The ADP standard curve generated within this section was prepared so that the rate of RNA_{20mer} substrate degradation, mediated by 3'-5' PNPase homologs, could be quantified in terms of the [ADP_{20x}] produced. A range of ADP_{20x} concentrations (0, 40, 80, 120, 160, 200, 300, 400, 600, 800, 1200, 2000 nM) were serially diluted and utilised to generate a twelve-point ADP_{20x} standard curve. These values corresponded to 0, 2, 4, 6, 8, 10, 15, 20, 30, 40, 60, and 100% RNA_{20mer} conversion when the RNA_{20mer} substrate was 2000 nM.

The ADP standard curve was prepared as described in Section 6.3.3. The RFU was recorded (with the first 15 minutes discarded to allow ADP* and Q-Ab thermal-equilibration) and was found to remain stable over the whole 12-hour assay duration (data not shown). The average RFU values between 30-90 minutes, for each $[ADP_{20x}]$, were calculated and these were plotted using Minitab (v17) (Minitab, 2010) against the $\ln([L]+1)$, where $[L]$ was equivalent to the $[ADP_{20x}]$ (nM) (Figure 6.8).

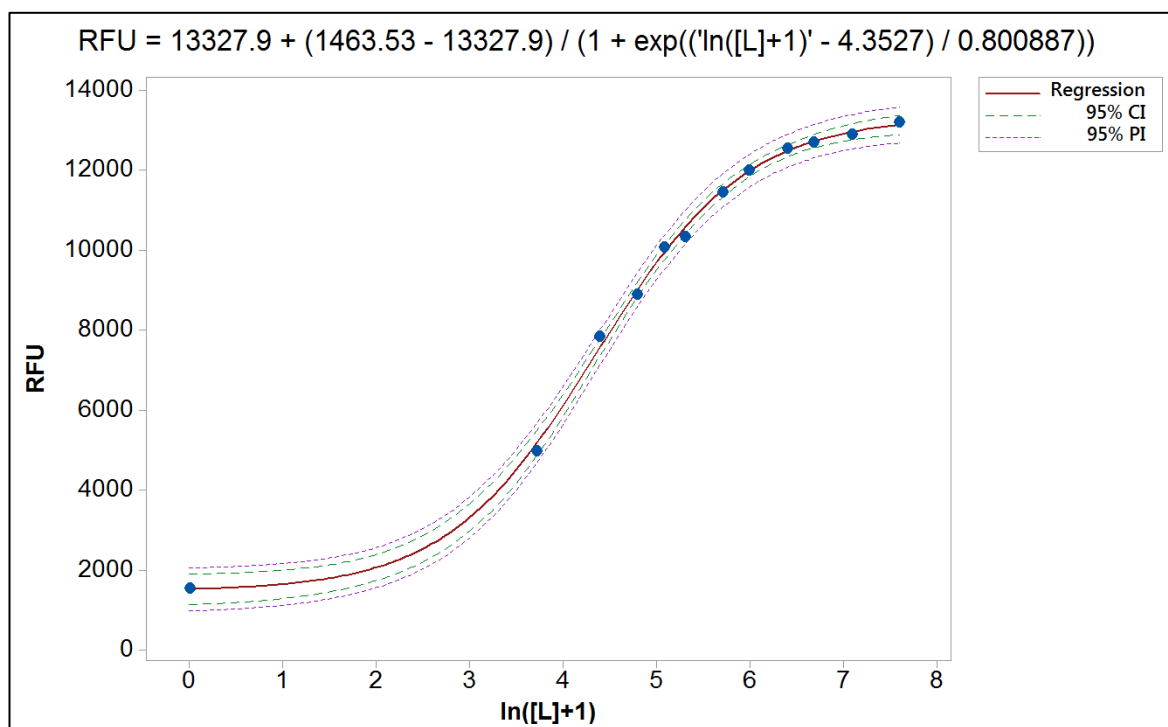


Figure 6.8 ADP_{20x} Standard Curve

Average RFU values from 30-90 minutes was plotted against $\ln([L]+1)$, where $[L]$ was equivalent to the ADP_{20x} concentration (nM). Data were fit using Equation 6.5 (red regression line) and the resulting equation was shown above each graph. The 95% confidence and prediction interval (CI and PI respectively) are also indicated with a green dashed line and pink dotted line respectively. Graphs were plotted in Minitab (v17) and labelled in GIMP (v2) (GIMP, n.d.; Minitab, 2010).

Following consultation with the assay manual (BellBrook, n.d.) and in collaboration with Dr David Whitley (Mathematician at the University of Portsmouth), the nonlinear regression, logistic growth function in Minitab (Figure 6.3) was utilised to fit the ADP standard curve data shown in Figure 6.8 (Minitab, 2010). The data points (blue circles) for the ADP_{20x} standard curve were close to the regression fit (red line) and the fit was observed to have a small 95% confidence and prediction interval (Figure 6.8, dotted green CI line and dotted purple PI line respectively). The displayed 95% CI represented a range of likely values for the mean response, whereas the PI represented a range of likely values for a single new observation. The prediction interval was wider than the corresponding confidence interval, as predicting a single response value is typically less certain than predicting the mean response value. In summary, data in Figure 6.8 fitted to the regression line well. Whether this fit could be used to convert RFU values accurately into the $[ADP_{20x}]$ (nM) was then examined below in Table 6.1

(a)	Theta				
	1	13327.9			
	2	1463.53			
	3	4.3527			
	4	0.800887			
(b)	Actual	RFU	Calculated	% Difference	
	[ADP_{20x} (nM)]		[ADP_{20x} (nM)]	(Calculated/Actual)	
	2000.00	13200	2891.60	31	
	1200.00	12888	1054.46	-14	
	800.00	12698	779.85	-3	*
	600.00	12536	641.32	6	*
	400.00	12003	407.80	2	*
	300.00	11446	294.61	-2	*
	200.00	10321	183.55	-9	*
	160.00	10072	168.26	5	*
	120.00	8881	116.04	-3	*
	80.00	7842	86.65	8	*
	40.00	4979	37.86	-6	*
0.00	1542	0.40	Err		

Table 6.1 ADP_{20x} Standard Accuracy

(a) The ADP_{20x} standard curve equation parameters Theta 1-4, used to convert RFU values into [ADP_{20x}], are provided. (b) The values of average RFU between 30-90 minutes are listed for each [ADP_{20x} (nM)] and these RFU values were used within the Equation 6.5 to calculate the [ADP_{20x} (nM)]. The percentage difference between calculated and actual [ADP_{20x}(nM)] were determined and values < 10% different were indicated with an asterisk. The percentage difference value listed for [0 nM ADP_{20x}] could not be calculated; it produced a mathematical error (Err).

In order to determine the accuracy of the standard curve, shown in Figure 6.8, for converting RFU values into [ADP_{20x}] (nM), a range of [ADP_{20x}] were calculated from the known RFU and Theta values listed in Table 6.1. These calculated [ADP_{20x}] values were compared to known [ADP_{20x}] and the percentage difference was typically small (<10%). This was expected since the data points in Figure 6.8 fitted to the regression fit well and as mentioned previously, had a small 95% CI and PI. The results presented in Table 6.1 validated the conversion of RFU values into [ADP_{20x}] between 40-800 nM. Hence for an assay containing 400-8000 nM RNA_{20mer}, if 10% of the substrate were degraded and converted into product, then 40-800 nM ADP_{20x} would be produced and these values would be converted accurately (within 10% error).

In conclusion, the ADP standard shown in Figure 6.8 was suitable for converting RFU values, recorded from the initial 10% of 400-8000 nM RNA_{20mer} substrate degradation, into [ADP_{20x}] ranging from 40-800 nM. The next Section 6.4.5 focused on using this quantifiable real-time assay for determining the Michaelis-Menten kinetics for EcPNPase.

6.4.5 EcPNPase Michaelis-Menten Kinetics

6.4.5.1 Factors Affecting the EcPNPase RFU Window

It was important to test whether EcPNPase and RNA_{20mer} would affect the maximum ADP* fluorescence (when no Q-Ab antibody was present) or quenched fluorescence (when both Q-Ab and ADP* were present). The results shown in Figure 6.9 and Table 6.2 suggested that neither the ADP* (bright red triangles) fluorescence nor Q-Ab quenching of ADP* (dark red triangles) were affected by 0.5 nM EcPNPase (blue circles) or 2000 nM RNA_{20mer} (black triangles). The average RFU between 30-90 minutes, for an assay containing only ADP* was 12,207, with a standard deviation of ± 532 and when EcPNPase and RNA was added to ADP*, the average RFU was within this error (Table 6.2). The average RFU for the quenched A*-Q-Ab assay was 836 ± 147 and the addition of EcPNPase and RNA, did not affect the RFU, with values of 827 ± 127 and 844 ± 142 respectively (Table 6.2). It is worthwhile noting that the RFU window shown in Figure 6.9 was slightly smaller than that observed previously (Figure 6.7); it was possible that the ADP* stock may have exhibited reduced fluorescence as a result of storage. However, this was not an issue for the high-throughput plate reader assay, as the use of an ADP standard curve accounted for plate to plate variations.

It is also valuable to note that the standard deviation for the quenched RFU assays (~ 140 RFU) was lower than the unquenched ADP* assay (532 RFU). The quenching of ADP* by Q-Ab may help prevent large fluctuations in fluorescence of the fluorophore. This suggested that high levels of ADP production, outcompeting ADP* for Q-Ab quenching, would result in higher fluctuations in fluorescence. Calculations of initial rates would avoid the large RFU fluctuations, which were observed at higher RFU values when all the ADP* is released from Q-Ab by ADP, and may produce more reproducible data.

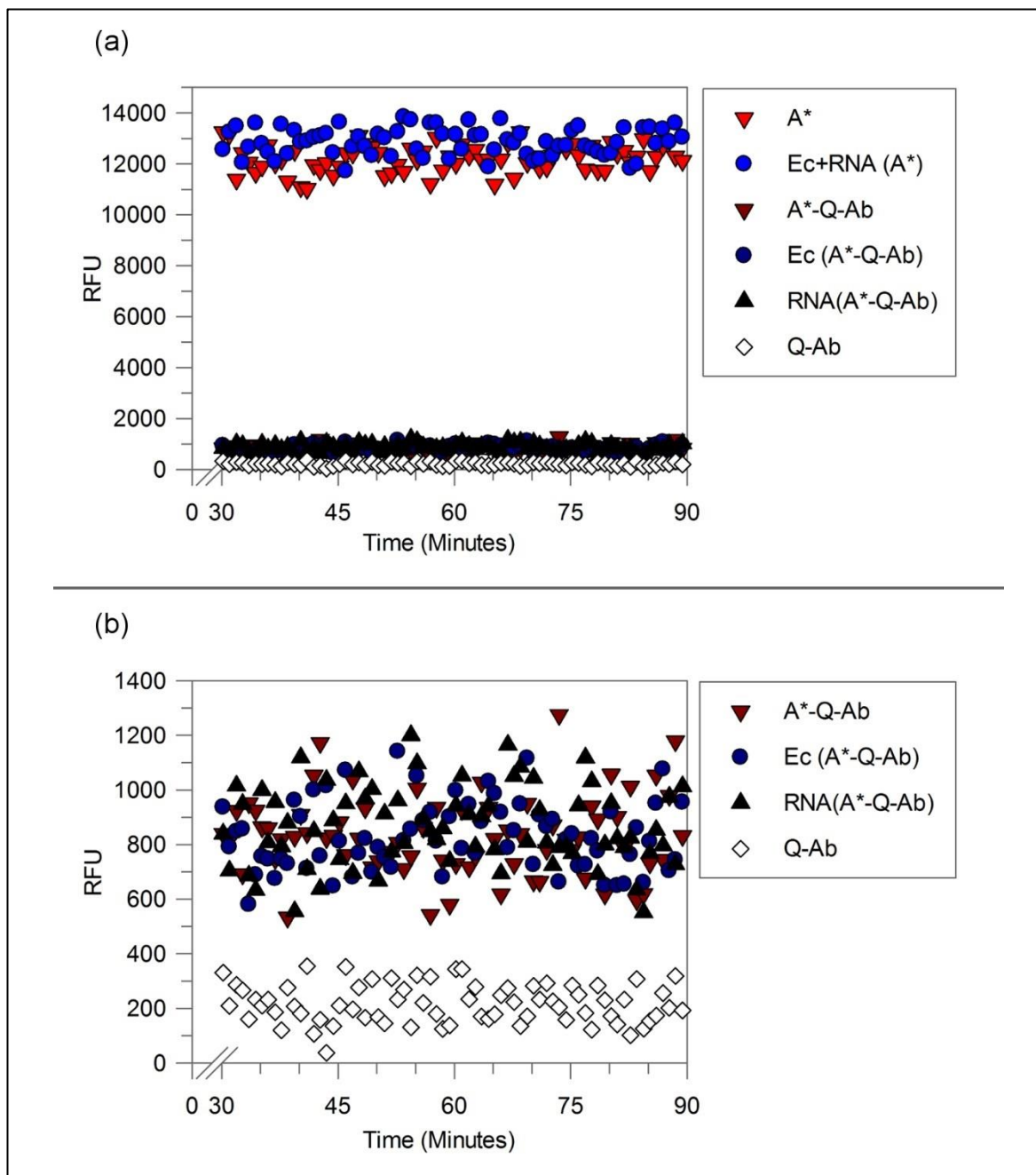


Figure 6.9 RFU Window with EcPNPase Assay Components

(a) The stability of maximum RFU signal (unquenched 4 nM ADP* (A*, bright red triangle)) and low RFU signal (quenched 4 nM ADP* and 40 μ g/ml Q-Ab (A*-Q-Ab, dark red triangle)) was recorded over time at 25°C and the data recorded from 30-90 minutes is shown. The negligible effect of 0.5 nM EcPNPase (Ec) and 2000 nM RNA_{20mer} on the maximum RFU (Ec+ RNA (A*), blue circle) is shown. Additionally, their individual effect on Q-Ab quenching of ADP* (A*-Q-Ab) is shown for Ec (dark blue circle) and RNA (black triangle). A control containing no fluorescence (Q-Ab only, white diamond) was also plotted. (b) The same graph is also shown zoomed in at (<1,400 RFU) in order to visualise the RFU of the quenched A*-Q-Ab reaction (~800) and the RFU of Q-Ab background controls (~200). Graphs were plotted in GraFit and labelled in GIMP (v2) (GIMP, n.d.; Leatherbarrow, 2009).

Assay	Average RFU	Std
A*	12207	532
Ec+RNA (A*)	12875	532
A*-Q-Ab	836	147
Ec (A*-Q-Ab)	827	127
RNA(A*-Q-Ab)	844	142
Q-Ab	216	72

Table 6.2 RFU Window with EcPNPase Assay Components

The average RFU from 30-90 minutes, for maximum RFU signal (unquenched 4 nM ADP* (A*)) and low RFU signal (quenched ADP*-Q-Ab (A*-Q-Ab)), background RFU (40 µg/ml Q-Ab), and the effect of 0.5 nM EcPNPase (Ec) and 2000 nM RNA_{20mer} (RNA) are listed. The standard deviation (Std) of data points are also indicated for each assay.

In summary, both EcPNPase and RNA had no effect on either the unquenched ADP* RFU or quenching of ADP* by Q-Ab, and did not affect the assay window. Any changes in RFU measured using this high-throughput plate assay would therefore be directly a result of ADP produced from RNA degradation mediated by EcPNPase. Additionally, a ADP_{20x} standard curve must be conducted in all plate reader experiments to account for plate to plate variations.

6.4.5.2 EcPNPase Titration

The results presented in Section 6.4.5.1 suggested that none of the assay components effect the RFU window; hence the next step was to determine the optimal concentration of EcPNPase. A range of EcPNPase enzyme concentrations (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75 and 1 nM) were titrated against a set concentration of RNA_{20mer} (2000 nM) and the RFU was recorded over time as described in Section 6.3.4. The RFU recorded over time increased, in a concentration-dependant manner, for all EcPNPase concentrations. However, only EcPNPase concentrations greater than 0.3 nM exhibited the recommended 50-80% increase in RFU over the assay duration (Figure 6.10). Therefore, [EcPNPase] below 0.3 nM were not appropriate for producing the response recommended by the Bellbrook manual and were not chosen for subsequent RNA_{20mer} substrate titration experiments.

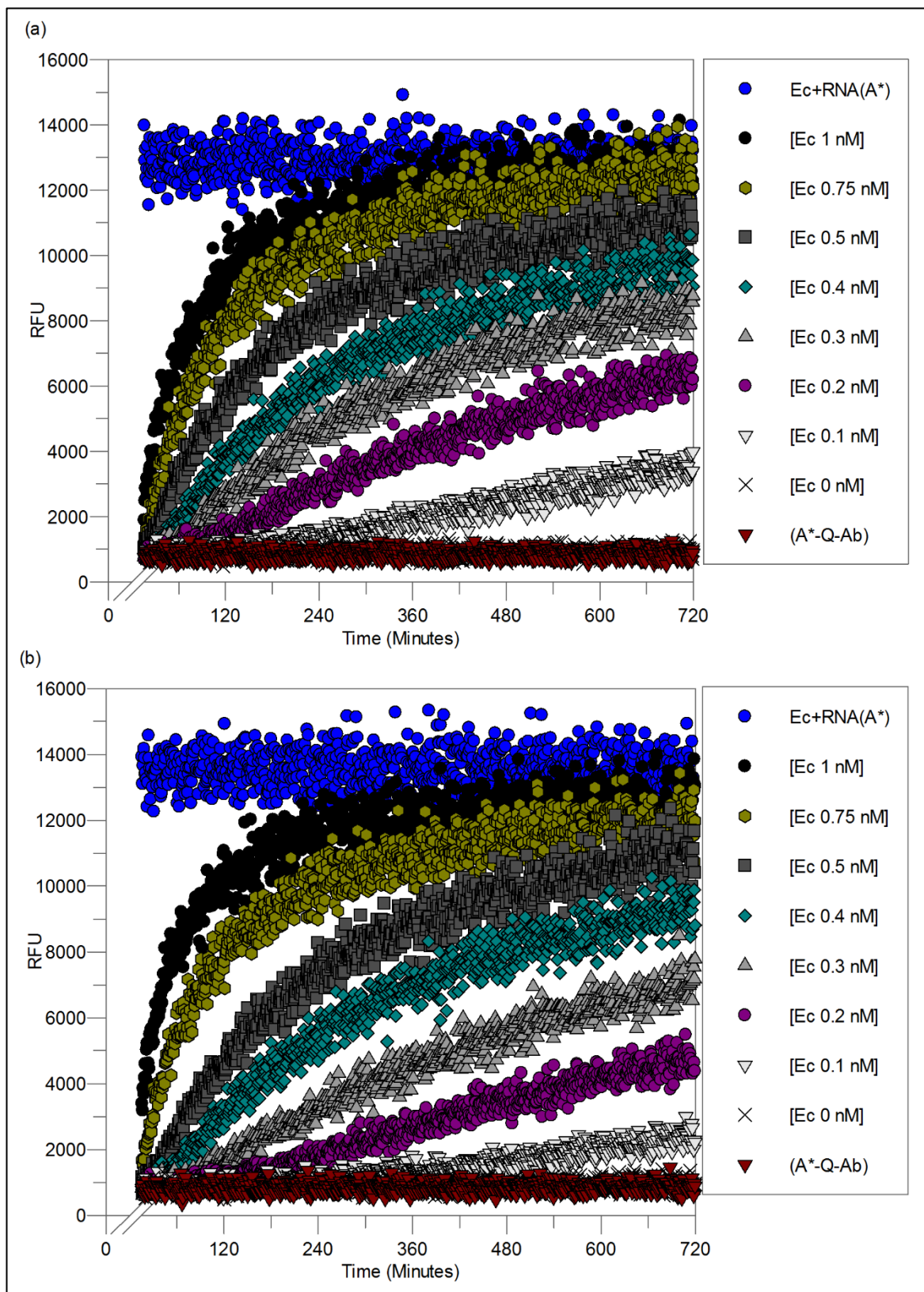


Figure 6.10 E_cPNPase Titration: RFU vs Time

The RFU recorded over time (minutes) from (a) repeat 1 and (b) repeat 2 are shown. The maximum RFU signal (unquenched (Ec + RNA (A*)), blue circles) and low RFU signal (quenched (A*-Q-Ab), dark red triangles) are indicated alongside a range of E_cPNPase concentrations (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1 nM) which are coloured from light grey to black respectively. Data collected within the first 15-minutes were discarded as standard. Graphs were plotted in GraFit (Leatherbarrow, 2009).

The RFU values, shown in Figure 6.10, were converted into ADP_{20x} (nM) concentrations using a ADP standard curves as described in Section 6.3.3. The ADP standard used within this assay was

suitable (within 10% error) for converting RFU values into $[ADP_{20x}]$ within the range of 16-128 nM ADP_{20x} . The EcPNPase titration assay contained 2000 nM RNA_{20mer} and therefore the ADP standard could be used to convert the initial 5% of ADP_{20x} increase accurately (5% 2000 nM RNA_{20mer} substrate degradation produces 100 nM ADP_{20x}). Accordingly, the data from the first 5% of the reaction, shown in Figure 6.11, were fit with a linear equation and the resultant rate of ADP_{20x} (nM) production over time (minutes) was calculated for each [EcPNPase] (Figure 6.12). Only EcPNPase concentrations greater than 0.4 nM reached 5% RNA_{20mer} substrate degradation over the duration of the assay (100 nM ADP_{20x} produced).

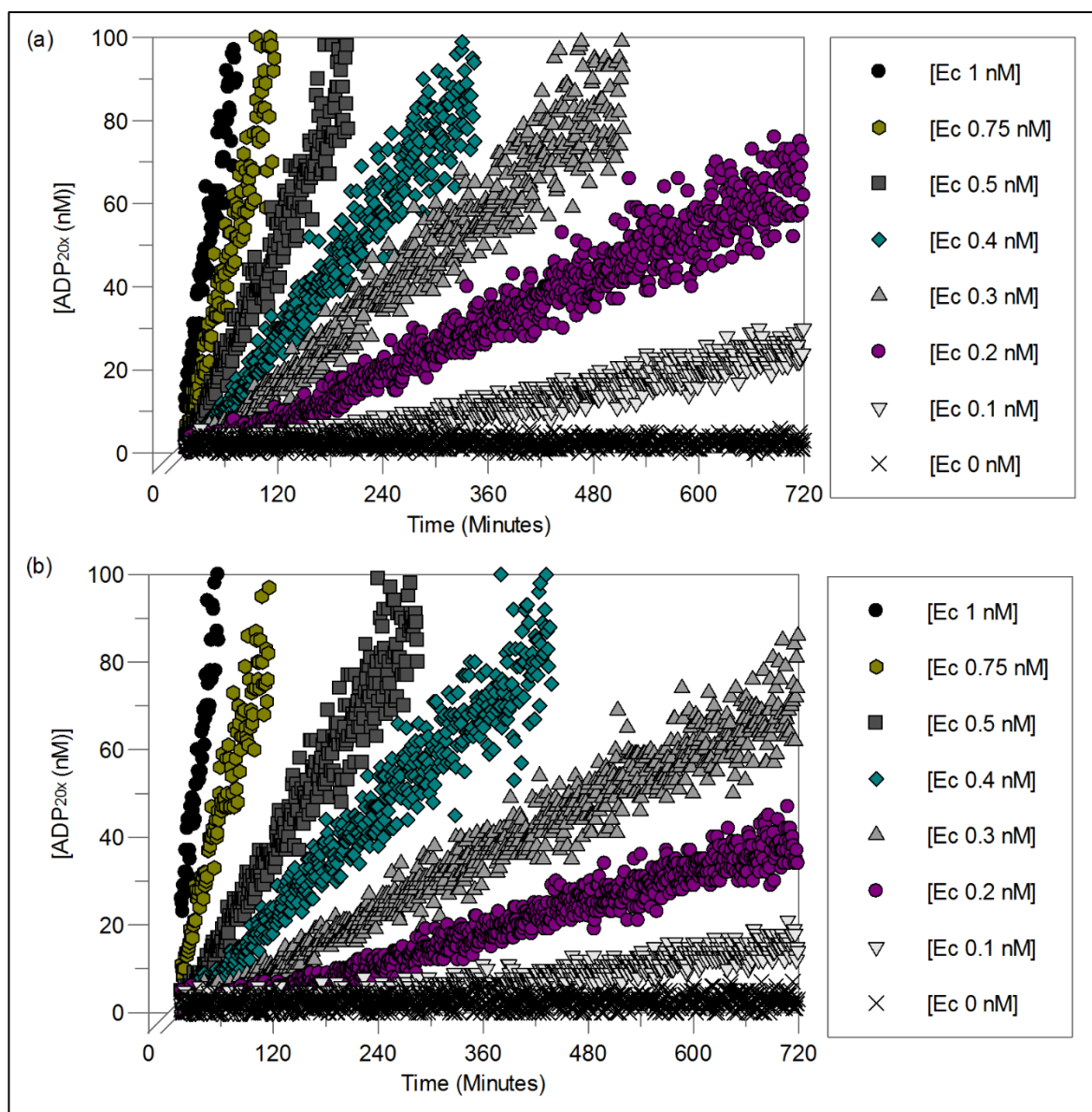


Figure 6.11 EcPNPase Titration: $[ADP_{20x}]$ vs Time

The $[ADP_{20x}$ (nM)] produced from a range of EcPNPase concentrations (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1 nM), were plotted over time (minutes) for (a) repeat 1 and (b) repeat 2 (minus 15-minutes equilibration). Graphs were plotted in GraFit (Leatherbarrow, 2009).

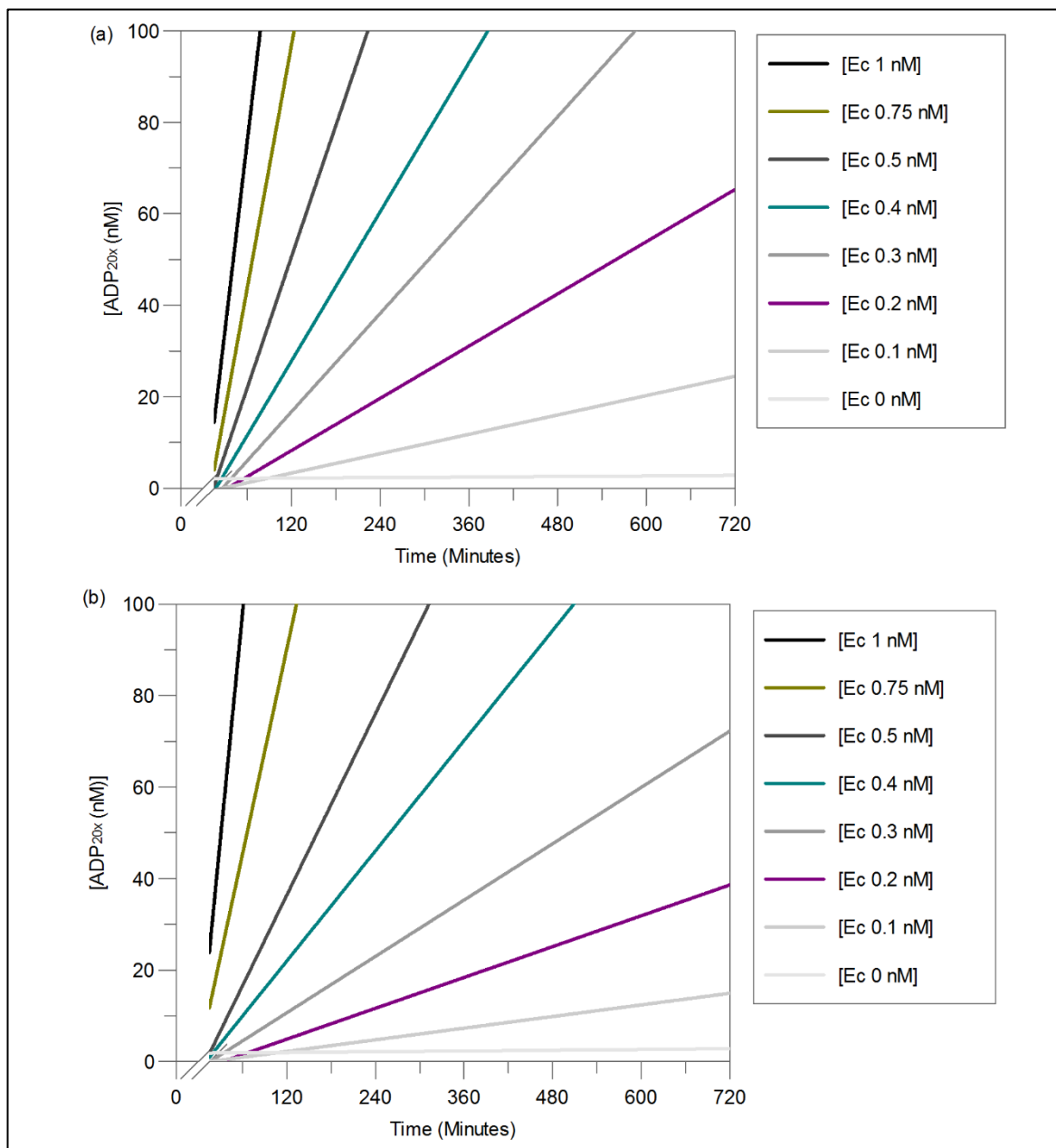


Figure 6.12 EcPNPase Titration: Linear Fit [ADP_{20x}] vs Time

The [ADP_{20x} (nM)] produced over time (minutes) for a range of EcPNPase concentrations (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1 nM) were fit with the linear equation ($y = m \cdot x + c$). Results for (a) repeat 1 and (b) repeat 2 (minus 15-minutes equilibration) are shown with increasing [EcPNPase] coloured from light grey to black respectively. Graphs were plotted in GraFit (Leatherbarrow, 2009).

These calculated initial rates were then plotted against the enzyme concentration and a graph, showing the average of two repeats, is shown in Figure 6.13, with error bars denoting the standard deviation. The enzyme concentrations, for which a linear increase in rate was detected, are shown in red (Figure 6.13). Ideally the enzyme concentration, which was within this linear region, should be selected for further enzyme kinetics. According to the data shown in Figure 6.13, this included EcPNPase concentrations between 0.4-0.75 nM.

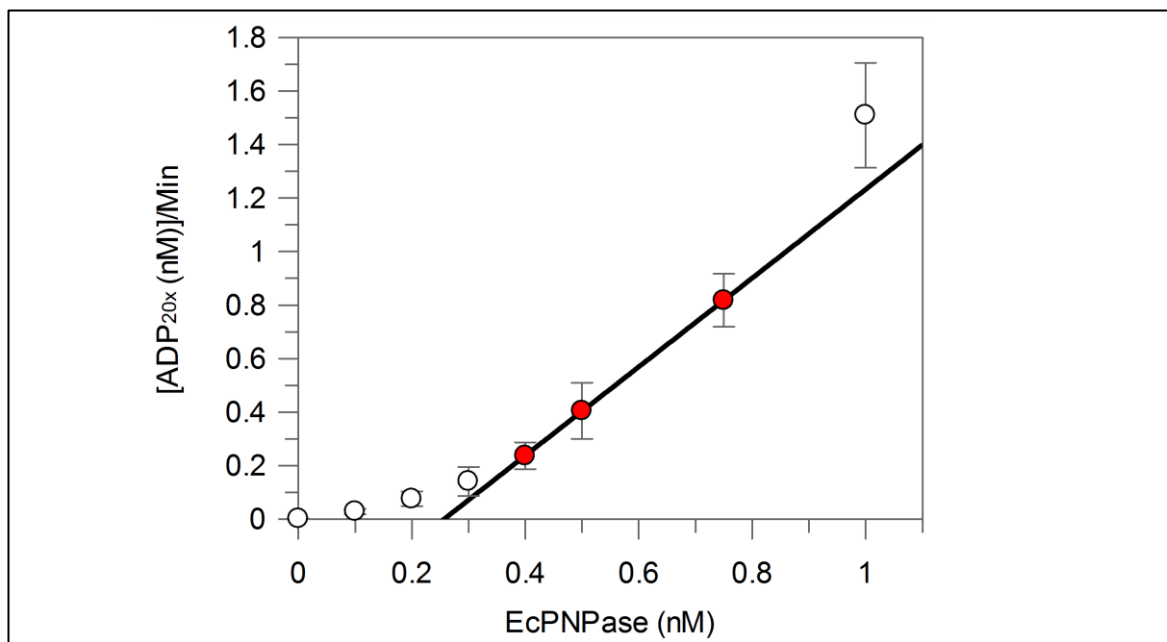


Figure 6.13 EcPNPase Titration: Rate of ADP_{20x} Production vs [EcPNPase]

The average rate of [ADP_{20x} (nM)] produced per minute, for a range of EcPNPase concentrations (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1 nM), was plotted (white circles) from duplicate data. The linear equation ($y = m \cdot x + c$, black line) was fitted to appropriate linear data points (red circles) using GraFit and error bars of the standard deviation are provided (Leatherbarrow, 2009).

Due to the previously established 15-minute equilibration time for ADP*-Q-Ab quenching at 25°C, it was important to select an enzyme concentration with the minimum amount of RNA degradation during these first 15-minutes. For this reason, as RNA degradation had already started during this equilibration time, neither 0.75 nor 1 nM EcPNPase were selected (Figure 6.12). However, in assays containing 0.4-0.5 nM EcPNPase, less than 0.1% (2 nM ADP_{20x} produced) of the RNA degradation reaction had occurred within the first 15-minutes of the assay. Hence either 0.4 or 0.5 nM EcPNPase were suitable for generating a 50-80% RFU signal change in the assay, and ensuring the initial rates could be calculated following the 15-minute ADP*-Q-Ab thermal-equilibration. Both enzyme concentrations were also low enough to prevent enzyme waste and helped ensure that the concentration of RNA_{20mer} substrate would be in excess for determining Michaelis-Menten kinetics. So the higher concentration, 0.5 nM EcPNPase, was chosen in order to reduce the assay duration.

In summary, following a 15 minute ADP*-Q-Ab equilibration, the degradation of 2000 nM RNA_{20mer} by 0.5 nM EcPNPase could be monitored directly in real-time; RFU data were collected in high-throughput and the concentration of ADP_{20x} (nM) produced over time was calculated using an ADP_{20x} standard curve. The initial rates of RNA degradation, for a range of substrate concentrations (in excess), could therefore be quantified and used to calculate Michaelis-Menten kinetics. Accordingly, the next Section 6.4.5.3 describes the Michaelis-Menten kinetics calculated for 0.5 nM EcPNPase titrated with a range of RNA_{20mer} substrate concentrations.

6.4.5.3 EcPNPase Degradation Activity (0.5 mM Phosphate)

The manual recommended detection of ADP in the presence of excess substrate, assuming initial velocity enzyme reaction conditions. In order to determine the canonical EcPNPase RNA degradation activity, 0.5 nM EcPNPase (in the presence of 0.5 mM KH_2PO_4) was incubated at 25°C, with a range of RNA_{20mer} substrate concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 & 2600 nM), as per Section 6.3.5.

The RFU values recorded over time for each RNA_{20mer} concentration (Figure 6.14) were converted into ADP_{20x} (nM) concentrations using a ADP standard curve as described in Section 6.3.3. The ADP standard used within this assay was suitable (within 10% error) for converting RFU values into [ADP_{20x}] within the range of 16-256 nM ADP_{20x}. Therefore, the ADP standard could be used to convert the initial 10% of RNA degradation accurately for substrate concentrations between 160-2560 nM RNA_{20mer}. Ideally, this ADP standard curve should be expanded to include values at the lower range (100 nM [ADP_{20x}]) and this required further optimisation. Nevertheless, the data from the initial 10% of the reaction (minus 15-minutes) were plotted against time (Figure 6.15 (a)). The data plotted in Figure 6.15 (a), were fitted to a linear equation (Figure 6.15 (b)) and the rate of reaction was calculated. These initial rates were then plotted against the concentration of substrate to give the Michaelis-Menten curve shown in Figure 6.16.

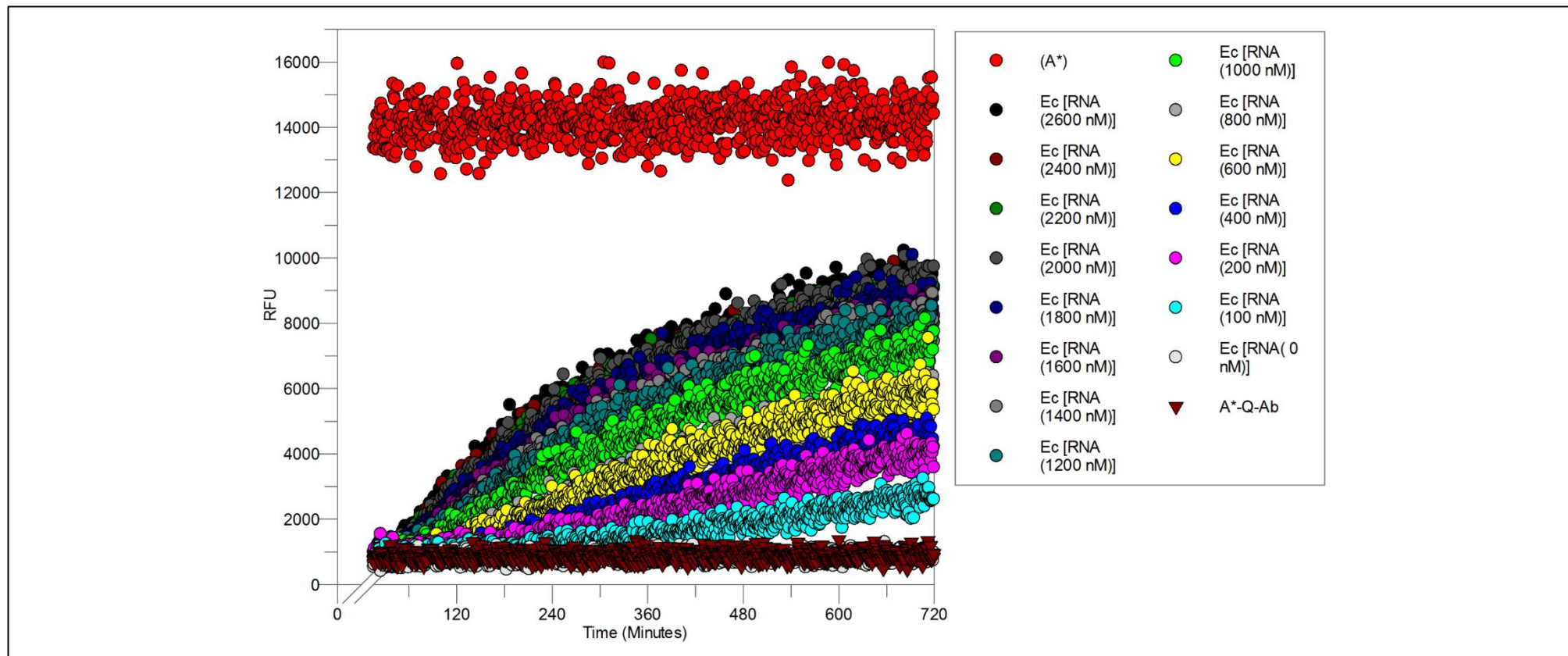


Figure 6.14 RNA_{20mer} Degradation by EcpNPase (0.5 mM Phosphate): RFU vs Time

The RFU recorded over time (minutes) for the maximum RFU signal (unquenched (A*), bright red circles) and low RFU signal (quenched (A*-Q-Ab), dark red triangles) are indicated for assays containing 0.5 nM EcpNPase (Ec) (+ 0.5 mM phosphate), titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) which are coloured from light grey to black respectively. Data collected within the first 15-minutes were discarded as standard. Graphs were plotted in GraFit (Leatherbarrow, 2009).

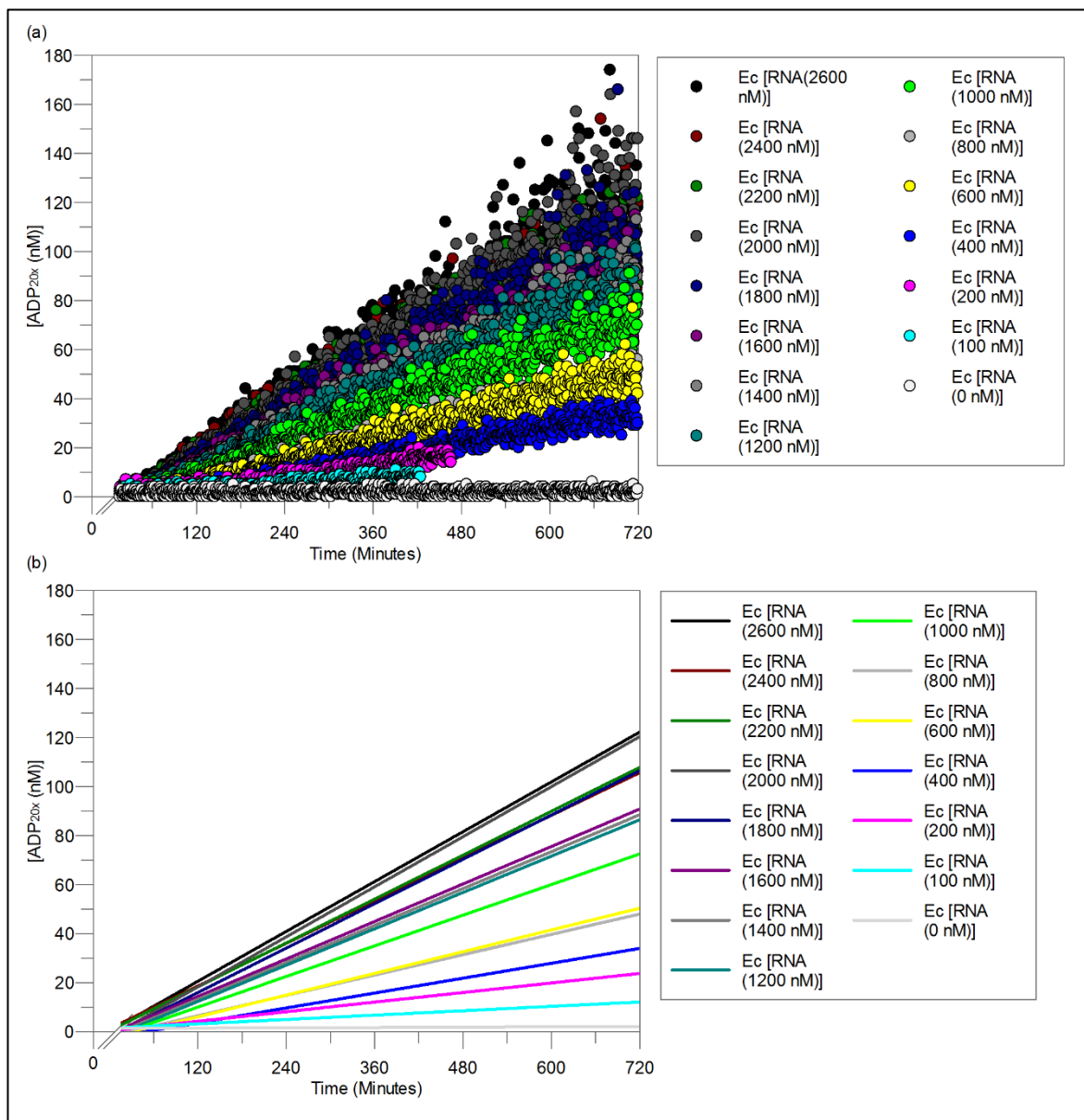


Figure 6.15 RNA_{20mer} Degradation by EcPNPase (0.5 mM Phosphate): [ADP_{20x} (nM)] vs Time

(a) The [ADP_{20x} (nM)] produced over time (minutes), for an assay containing 0.5 nM EcPNPase (0.5 mM phosphate) titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) is shown (minus 15-minutes equilibration). (b) The data were fit with the linear equation ($y = m \cdot x + c$) and the graph show increasing [EcPNPase] coloured from light grey to black respectively. Graphs were plotted in GraFit (Leatherbarrow, 2009).

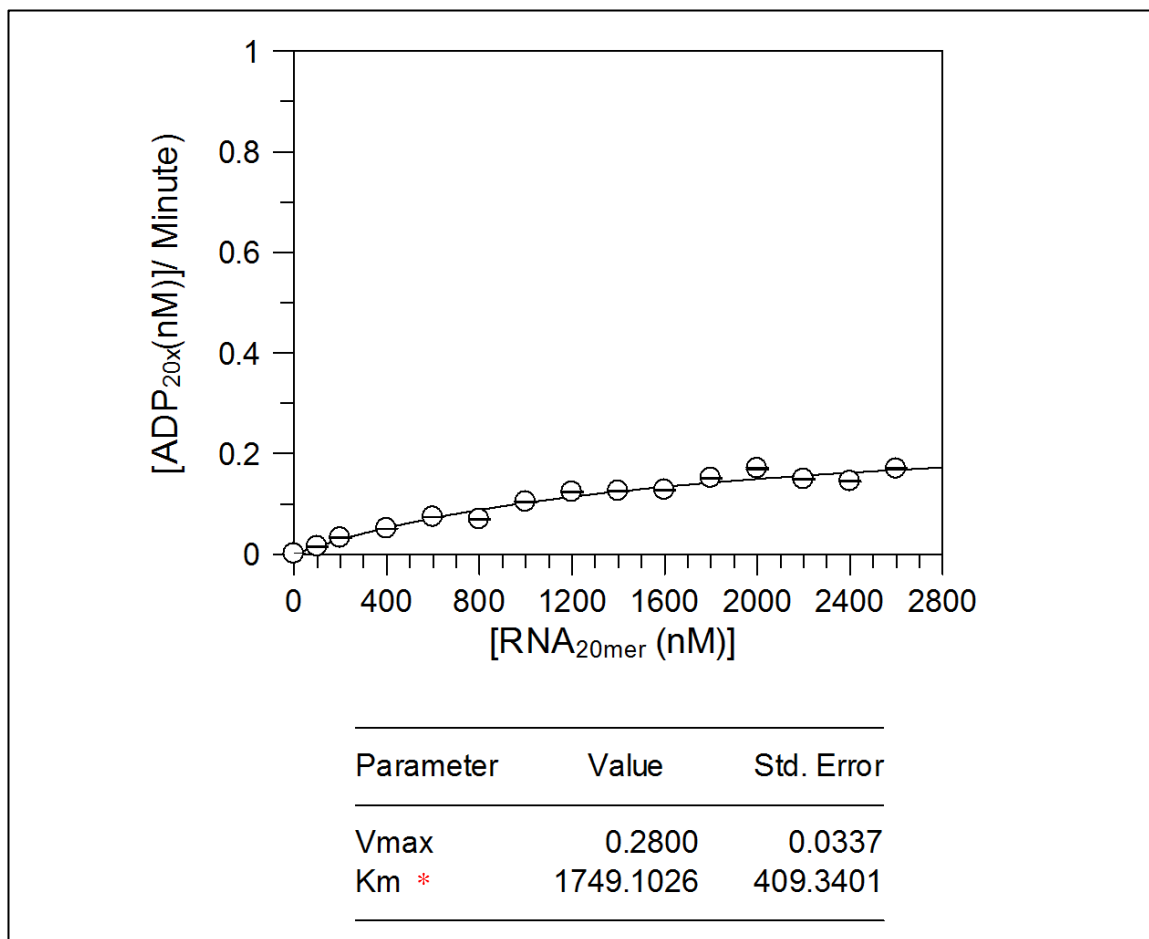


Figure 6.16 RNA_{20mer} Degradation by EcPNPase (0.5 mM Phosphate): Michaelis-Menten

The rate of [ADP_{20x} (nM)] produced per minute, for an assay containing 0.5 nM EcPNPase (0.5 mM phosphate) titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) is shown. The data were fit using the Michaelis-Menten equation shown in Equation 6.2 and the resultant V_{max} (nM/Minute) and K_m (nM) are indicated with their standard deviation (Std) error, red asterisk indicates values with a high error. The graph was plotted and fit using GraFit (Leatherbarrow, 2009).

The Michaelis-Menten Curve shown in Figure 6.16 was used to calculate the kinetic parameters V_{max} and K_m for EcPNPase-mediated degradation of RNA_{20mer} substrate in the presence of 0.5 mM phosphate. The error calculated for the K_m (± 409 nM) was high; it was more than 20% of the calculated value, and so the assay was optimised further to improve the fit of the curve. The assay described within this section contained 0.5 mM phosphate and this was below the optimum concentration previously reported for EcPNPase. Hence the experiments presented within the following Sections 6.4.5.4, tested a higher concentration of phosphate (10 mM).

6.4.5.4 EcPNPase Degradation Activity (10 mM Phosphate)

For assays described in subsequent sections a new Transcreener kit was utilised and, due to variations between kit stocks, the RFU window was smaller. This was still within the RFU window recommended by the kit (>5-fold signal difference). Nevertheless, it is important to note subtle variations can occur between kit batches and therefore the use of ADP_{20x} standards is recommended on each plate to account for plate to plate variations.

In order to determine whether increased phosphate would improve the rate of EcPNPase degradation, 0.5 nM EcPNPase was incubated at 25°C with a range of RNA_{20mer} substrate concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 & 2600 nM), as per Section 6.4.5.3, however a higher concentration of phosphate (10 mM KH₂PO₄) was utilised.

The RFU values recorded over time for each RNA_{20mer} concentration (Figure 6.17) were converted into [ADP_{20x}] (nM) using a ADP standard curve as described in Section 6.3.3. The ADP standard used within this assay was suitable (within 10% error) for converting RFU values into [ADP_{20x}] within the range of 40-400 nM ADP_{20x}. Therefore, the ADP standard could be used to convert the initial 10% of RNA degradation accurately for substrate concentrations between 400-4000 nM RNA_{20mer}. As mentioned previously, ideally, this standard curve should be expanded to include values at the lower range (100-200 nM [ADP_{20x}]) and this required further optimisation. Nevertheless, the data from the initial 10% of the reaction (minus 15-minutes) were plotted against time (Figure 6.18 (a)). The data plotted in Figure 6.18 (a) were fitted to a linear equation (Figure 6.18 (b)) and the initial rates were plotted against the concentration of substrate to give the Michaelis-Menten curve shown in Figure 6.19.

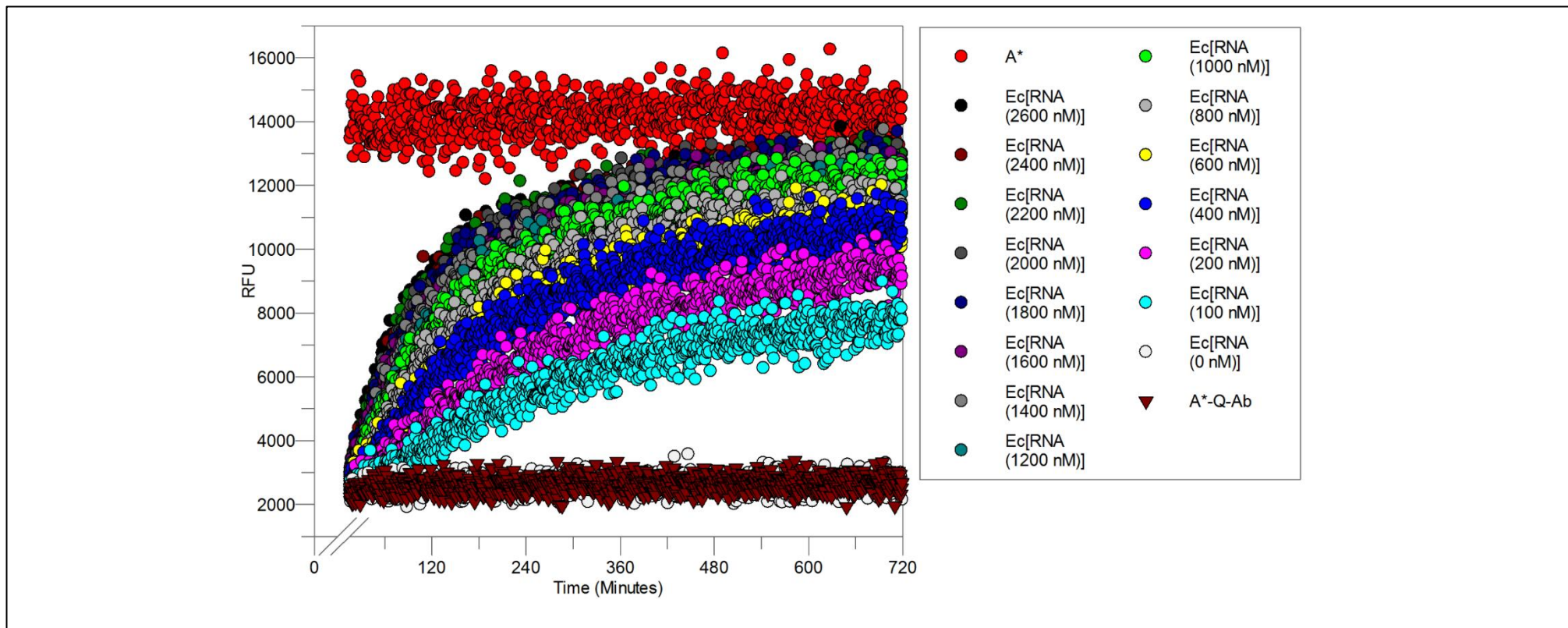


Figure 6.17 RNA_{20mer} Degradation by EcPNPase (10 mM Phosphate): RFU vs Time

The RFU recorded over time (minutes) for the maximum RFU signal (unquenched (A*), bright red circles) and low RFU signal (quenched (A*-Q-Ab), dark red triangles) are indicated for assays containing 0.5 nM EcPNPase (Ec) (+ 10 mM phosphate), titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) which are coloured from light grey to black respectively. Data collected within the first 15-minutes were discarded as standard. Graphs were plotted in GraFit (Leatherbarrow, 2009).

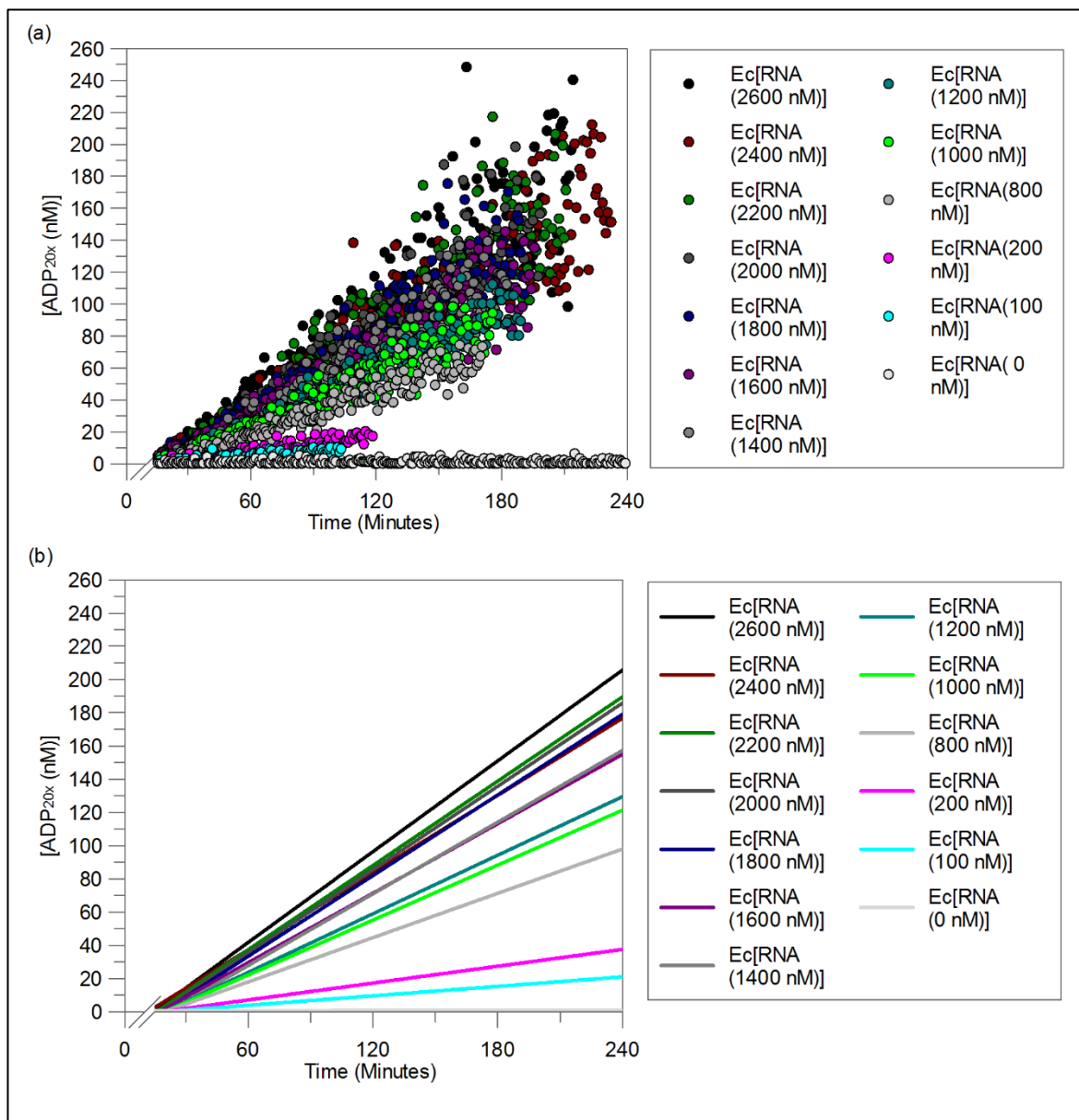


Figure 6.18 RNA_{20mer} Degradation by EcPNPase (10 mM Phosphate): [ADP_{20x} (nM)] vs Time

(a) The [ADP_{20x} (nM)] produced over time (minutes), for an assay containing 0.5 nM EcPNPase (10 mM phosphate) titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) is shown (minus 15-minutes equilibration). (b) The data were fit to the linear equation ($y = m \cdot x + c$) and the graph shows increasing [EcPNPase] coloured from light grey to black respectively. Graphs were plotted in GraFit (Leatherbarrow, 2009).

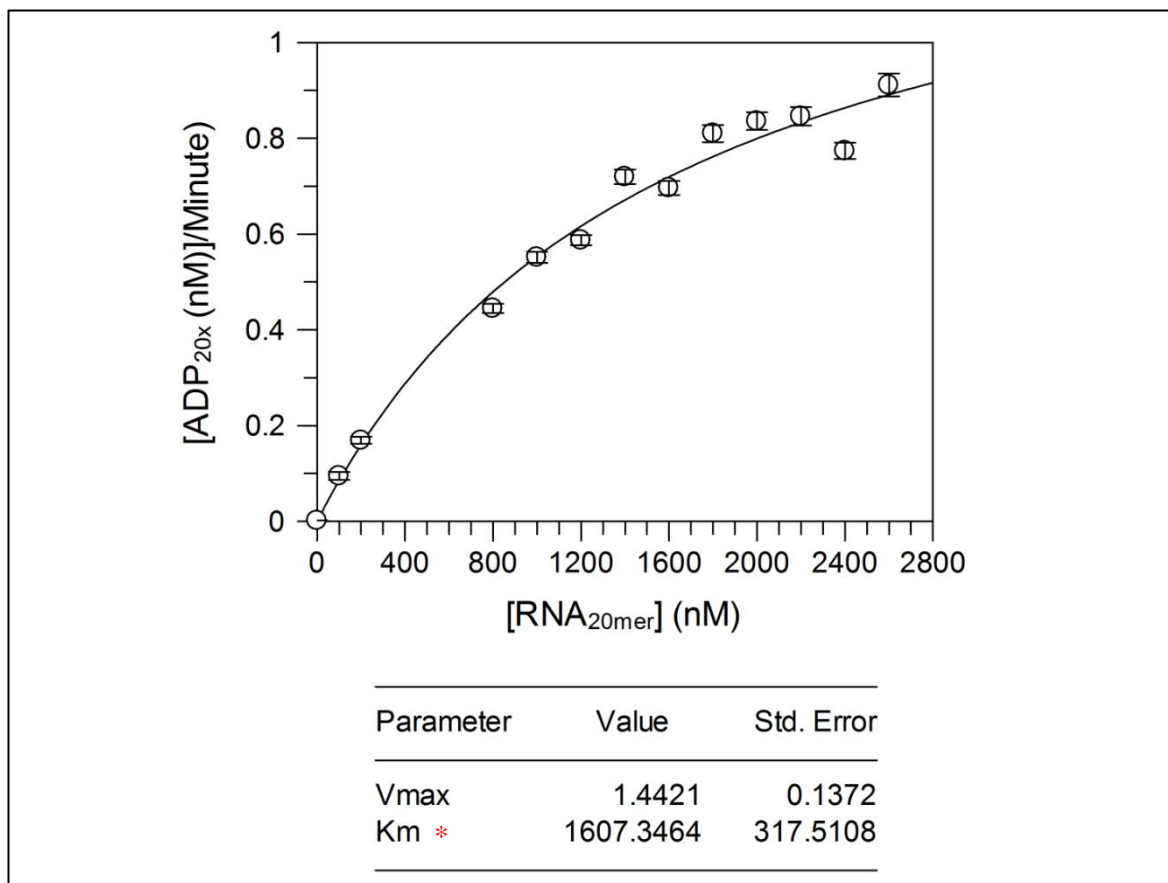


Figure 6.19 RNA_{20mer} Degradation by EcPNPase (10 mM Phosphate): Michaelis-Menten

The rate of [ADP_{20x} (nM)] produced per minute, for an assay containing 0.5 nM EcPNPase (10 mM phosphate) titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) is shown. The data were fitted using the Michaelis-Menten equation shown in Equation 6.2 and the resultant V_{max} (nM/Minute) and K_m (nM) are indicated with their error of standard deviation, red asterisk indicates values with a high error. The graph was plotted and fit using GraFit (Leatherbarrow, 2009).

The Michaelis-Menten Curve shown in Figure 6.19 was used to calculate the kinetic parameters V_{max} and K_m for EcPNPase degradation of RNA_{20mer} in the presence of 10 mM phosphate. When the phosphate concentration was increased from 0.5 mM (Section 6.4.5.3) to 10 mM, the maximum velocity of RNA_{20mer} degradation increased from 0.28 nM Min⁻¹ (Figure 6.16) to 1.44 nM Min⁻¹ (Figure 6.19). This suggested that the assay, in the presence of 0.5 mM phosphate, was sub-optimal and that the higher phosphate concentration increased EcPNPase activity. As the EcPNPase activity was higher in 10 mM phosphate, the V_{max} was not reached when the same concentration range of RNA_{20mer} was used (0-2600 nM); EcPNPase was not completely saturated by the substrate. Therefore, the results in Figure 6.19 had fewer data points at V_{max} compared to the results in Figure 6.16 (0.5 mM phosphate) and as a result, the error of the calculated V_{max} error was higher. Although EcPNPase exhibited higher activity in these conditions, as before (Section 6.4.5.3), the error calculated for the K_m (±317.5 nM) was still high; it was ~20% of the calculated value. Whether a lower phosphate concentration would improve the fitting of kinetics data was thus investigated in Section 6.4.5.5

6.4.5.5 EcPNPase Degradation Activity (5 mM Phosphate)

In order to determine whether 5 mM phosphate would improve the rate of EcPNPase degradation, 0.5 nM EcPNPase was incubated at 25°C with a range of RNA_{20mer} substrate concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 & 2600 nM), as per Section 6.4.5.4. However, a lower concentration of phosphate (5 mM KH₂PO₄) was utilised.

The RFU values recorded over time for each RNA_{20mer} concentration (Figure 6.20) were converted into [ADP_{20x}] (nM) using a ADP standard curve as described in Section 6.3.3. The ADP standard used within this assay was suitable (within 10% error) for converting RFU values into [ADP_{20x}] within the range of 40-1200 nM ADP_{20x}. Therefore, the ADP standard could be used to convert the initial 10% of RNA degradation accurately for substrate concentrations between 400 nM-12 μM RNA_{20mer}. As mentioned previously, ideally, this standard curve should be expanded to include values at the lower range (100-200 nM [ADP_{20x}]) and this required further optimisation. Nevertheless, the data from the initial 10% of the reaction (minus 15-minutes) were plotted against time (Figure 6.21 (a)). The data plotted in Figure 6.21 (a) were fitted to a linear equation (Figure 6.21 (b)) and initial rates were plotted against the concentration of substrate to give the Michaelis-Menten curve shown in Figure 6.22.

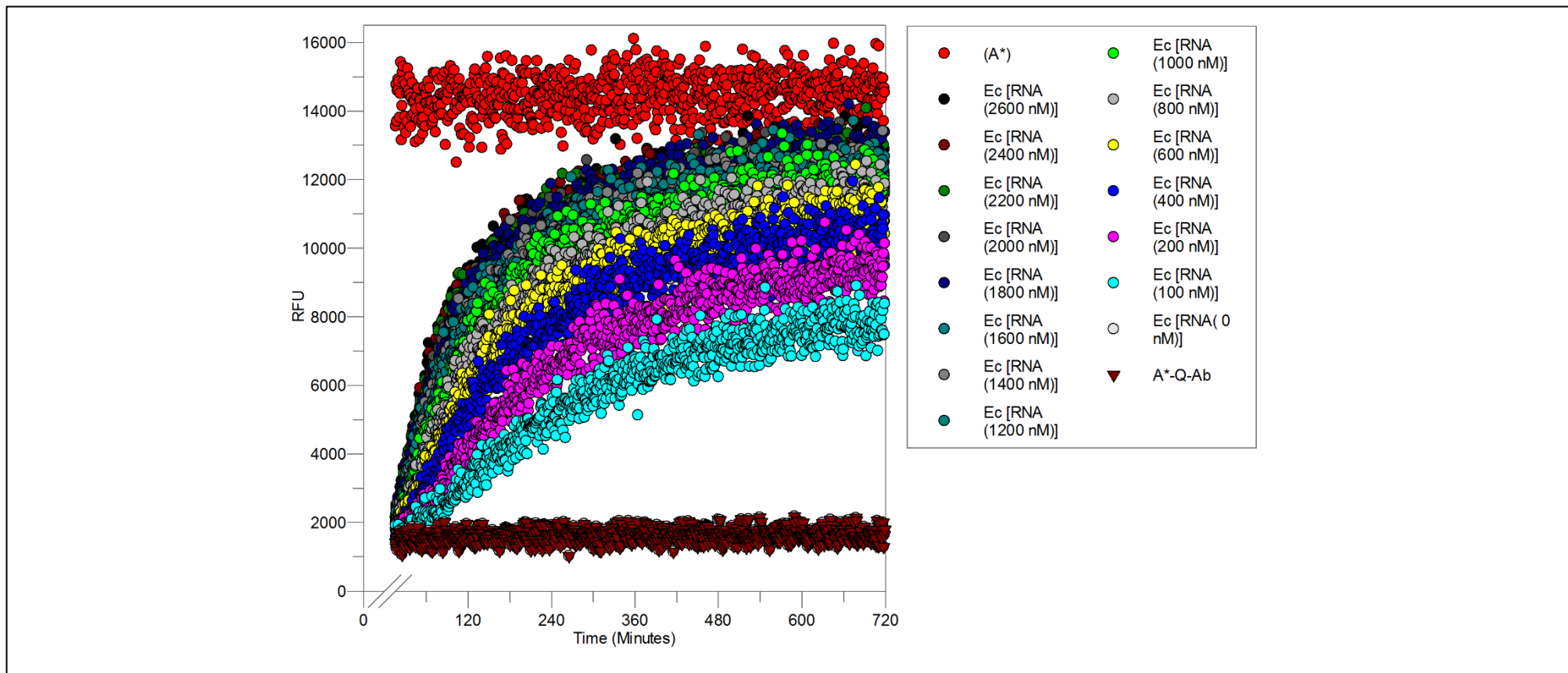


Figure 6.20 RNA_{20mer} Degradation by EcpNPase (5 mM Phosphate): RFU vs Time

The RFU recorded over time (minutes) for the maximum RFU signal (unquenched (A*), bright red circles) and low RFU signal (quenched (A*-Q-Ab), dark red triangles) are indicated for assays containing 0.5 nM EcpNPase (Ec) (+5 mM phosphate), titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) which are coloured from light grey to black respectively. Data collected within the first 15-minutes were discarded as standard. Graphs were plotted in GraFit (Leatherbarrow, 2009).

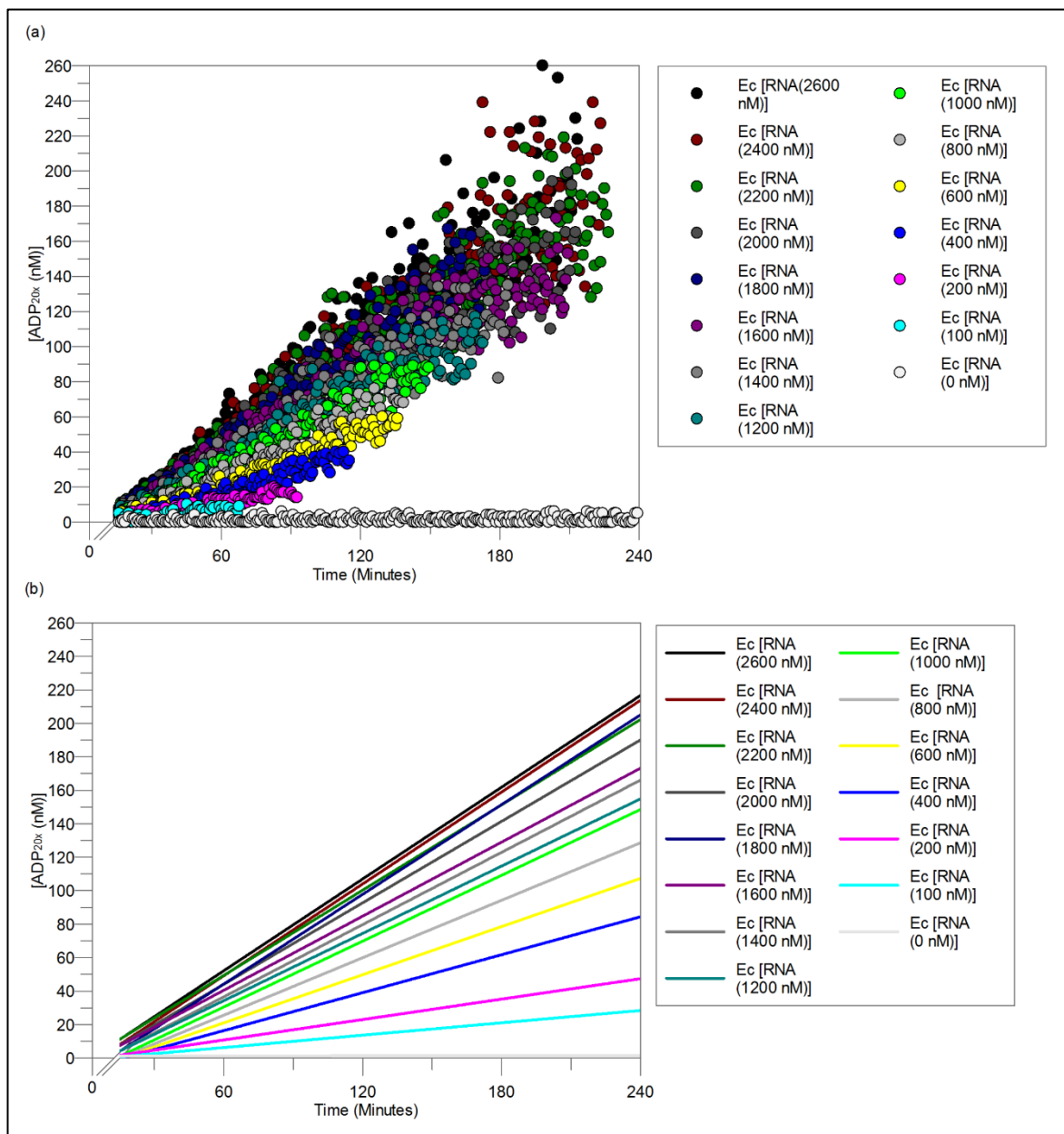


Figure 6.21 RNA_{20mer} Degradation by EcPNPase (5 mM Phosphate): [ADP_{20x} (nM)] vs Time

(a) The [ADP_{20x} (nM)] produced over time (minutes), for an assay containing 0.5 nM EcPNPase (5 mM phosphate) titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) is shown (minus 15-minutes equilibration). (b) The data were fitted to a linear equation ($y = m \cdot x + c$) and the graph shows increasing [EcPNPase] coloured from light grey to black respectively. Graphs were plotted in GraFit (Leatherbarrow, 2009).

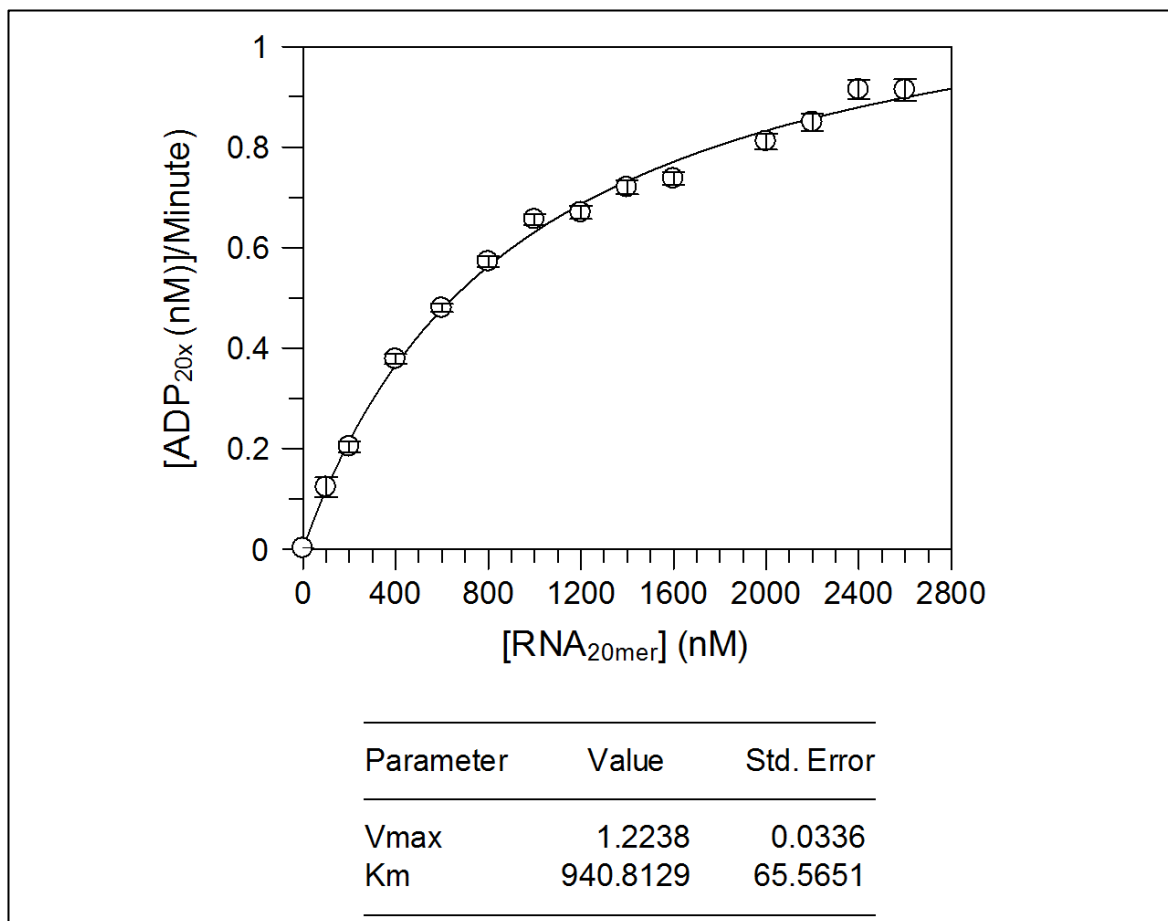


Figure 6.22 RNA_{20mer} Degradation by EcPNPase (5 mM Phosphate): Michaelis-Menten

The rate of [ADP_{20x} (nM)] produced per minute, for an assay containing 0.5 nM EcPNPase (5 mM phosphate) titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 800, 1000, 1200, 1400, 1600, 2000, 2200, 2400 and 2600 nM) is shown. The data were fit using the Michaelis-Menten equation shown in Equation 6.2 and the resultant V_{max} (nM/Minute) and K_m (nM) are indicated with their standard deviation error. The graph was plotted and fit using GraFit and data from 1800 nM were discarded for better fit error (Leatherbarrow, 2009).

In previous experiments using 0.5 nM EcPNPase, in the presence of 0.5 mM phosphate, the reaction took around 12 hours to get to 5% RNA_{20mer} substrate degradation and this was too slow (Section 6.4.5.3). By increasing the phosphate concentrations from 0.5 mM to 10 mM, the rate of substrate degradation was increased; the reaction took ~4 hours to reach 10% RNA_{20mer} substrate degradation, however, the K_m error remained high (Section 6.4.5.4). When 5 mM phosphate was used instead, the initial rates of RNA degradation fitted well to a Michaelis-Menten curve and the standard deviation of both the V_{max} and K_m calculated values improved. Additionally, as the reaction in 5 mM phosphate also took ~4 hours to reach 10% RNA_{20mer} substrate degradation, reducing the phosphate concentration did not seem to be limiting. The following section compares the kinetic parameters for EcPNPase outlined within Section 6.4.5 in more detail.

6.4.5.6 EcPNPase Michaelis-Menten Kinetics Summary

The kinetic parameters determined for 0.5 nM EcPNPase in either 0.5, 5 or 10 mM phosphate in Sections 6.4.5.3-6.4.5.5 are summarised in Table 6.3. For comparative purposes, enzyme parameters with a standard error greater than 10% are coloured in red, whereas errors within 10% of the calculated value are coloured in green (Table 6.3). It is important to mention that values

reported in Table 6.3 were not from triplicate data-sets and care must be taken when comparing these kinetic parameters in detail.

Conditions	Parameter	Value	Std. Error
EcPNPase (0.5 mM phosphate)	[Et], nM	0.5	Na
	V _{max} , nM Min ⁻¹	0.28	±0.03
	K _m , nM	1749.10	±409.34
	K _{cat} , RNA _{20mer} Min ⁻¹	0.56	
	Catalytic Efficiency, M ⁻¹ s ⁻¹	5,336	
EcPNPase (5 mM phosphate)	[Et], nM	0.5	Na
	V _{max} , nM Min ⁻¹	1.22	±0.03
	K _m , nM	940.81	±65.57
	K _{cat} , RNA _{20mer} Min ⁻¹	2.45	
	Catalytic Efficiency, M ⁻¹ s ⁻¹	43,402	
EcPNPase (10 mM phosphate)	[Et], nM	0.5	Na
	V _{max} , nM Min ⁻¹	1.4421	±0.14
	K _m , nM	1607.3461	±317.51
	K _{cat} , RNA _{20mer} Min ⁻¹	2.88	
	Catalytic Efficiency, M ⁻¹ s ⁻¹	29,862	

Table 6.3 EcPNPase Kinetics Summary

The conditions and kinetic parameter for plate reader assays are listed, including total enzyme concentration [E_t], maximum velocity (V_{max}), Michaelis Constant (K_m), turnover number (k_{cat}) and catalytic efficiency. The units are provided with the values and the standard deviation error is listed where appropriate. Standard error values greater than 10% or less than 10% are highlighted in red or green respectively.

The data for the assay containing 0.5 nM EcPNPase, in the presence of 5 mM phosphate, fitted to the Michaelis-Menten hyperbolic curve (Figure 6.22) better than data collected in 0.5 or 10 mM phosphate. Therefore, the K_m and V_{max} values determined for EcPNPase in 5 mM phosphate (green, Table 6.3) were more reliable than the results obtained in the presence of 0.5 and 10 mM phosphate (red, Table 6.3). In summary, results presented in Table 6.3 suggested that EcPNPase (5 mM phosphate), could degrade the RNA poly(A)_{20mer} substrate and release 1.22 ±0.03 nM of ADP_{20x} per minute. The k_{cat}, calculated as described previously in Equation 6.3, indicated that 2.45 RNA poly(A)_{20mer} substrate molecules were degraded into ADP_{20x} product molecules, for a single EcPNPase enzyme per minute. The catalytic efficiency was calculated for EcPNPase, as described in Equation 6.4 and suggested that the RNA_{20mer} was degraded by EcPNPase at 43,402 M⁻¹ s⁻¹.

Due to the differences in standard error, caution needs to be made when comparing the enzyme parameters in Table 6.3. Nevertheless, EcPNPase activity appears to be reduced in the presence of 0.5 mM phosphate and increased in 5 and 10 mM phosphate and this correlated with previous publications reporting a high phosphate preference for EcPNPase (Portnoy *et al.*, 2007). More specifically, the V_{max}, k_{cat} and catalytic efficiency appear to increase as the phosphate concentration increases, although the assays need to be repeated in order to fully confirm that a trend exists.

In summary, the results in Table 6.3 suggested that the assay containing 0.5 nM EcPNPase, in the presence of 5 mM phosphate, could be used to calculate detailed kinetic parameters including the V_{max}, K_m, k_{cat} and catalytic efficiency. Whether the plate reader assay could also be used to

determine the activity a PNPase homolog, from the archaeal domain of life, was subsequently investigated in Section 6.4.6 using recombinant SsoExosome.

6.4.6 SsoExosome Michaelis-Menten Kinetics

6.4.6.1 Factors Affecting the SsoExosome RFU Window

It was important to show that, like EcPNPase (Section 6.4.5.1), the SsoExosome would not affect the maximum ADP* fluorescence (when no Q-Ab antibody was present) or quenched fluorescence (when both Q-Ab and ADP* were present). The results shown in Figure 6.23 and Table 6.4 suggested that neither the ADP* (bright red triangles) fluorescence nor Q-Ab quenching of ADP* (dark red triangles) were affected by 4 nM SsoExosome (green circles) or 2000 nM RNA_{20mer} (black triangles). The average RFU between 30-90 minutes, for an assay containing only ADP* was ~12,200 with a standard deviation of 532, when SsoExosome and RNA was added to ADP*, the RFU was within error of this maximum signal (Table 6.4). The average RFU for the quenched A*-Q-Ab assay was 836 ±147. The addition of SsoExosome and RNA did not affect the RFU (882 and 844 respectively); values were within the RFU error of the A*-Q-Ab assay (Table 6.4).

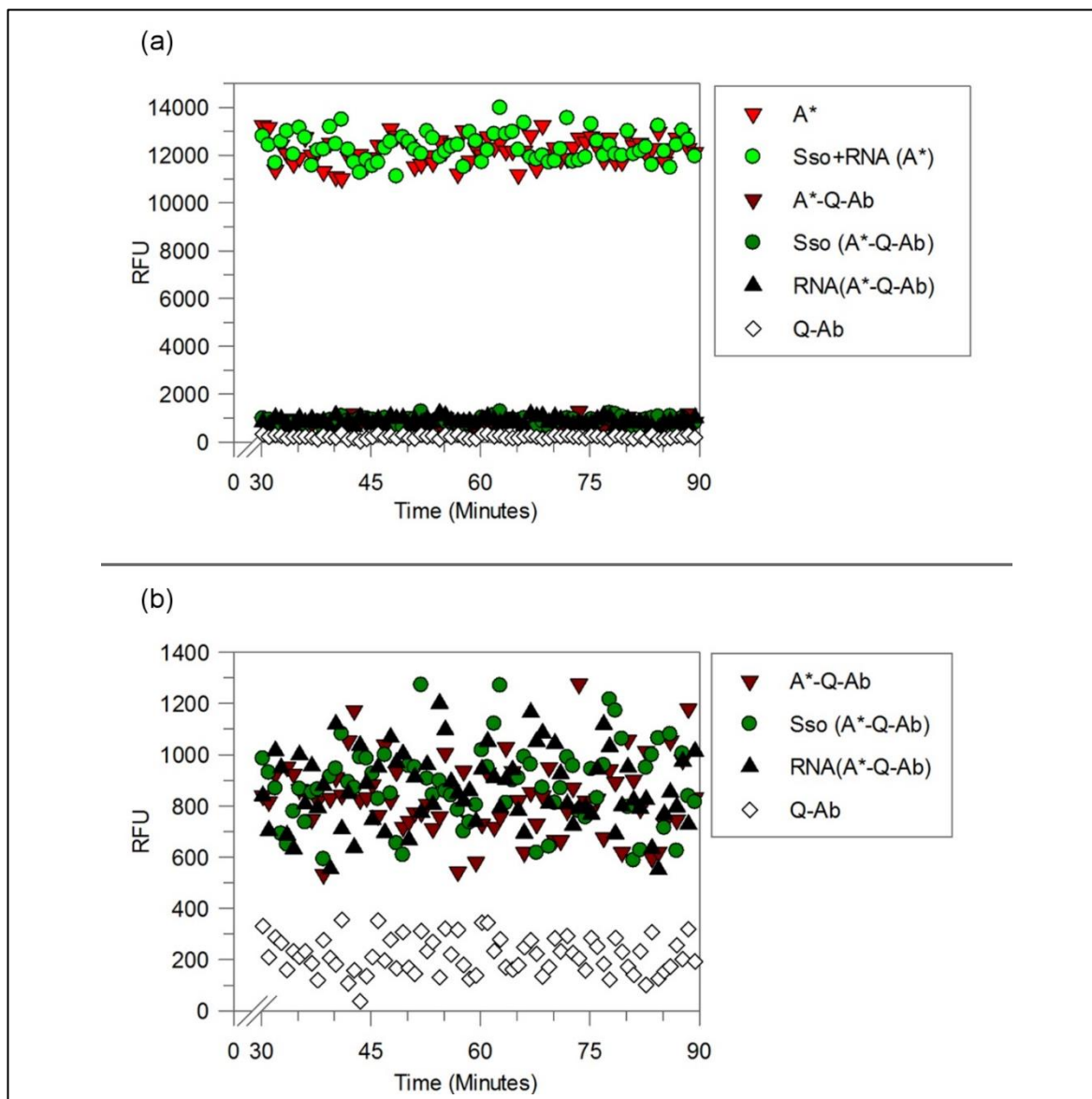


Figure 6.23 RFU Window with SsoExosome Assay Components

(a) The stability of maximum RFU signal (unquenched 4 nM ADP* (A*, bright red triangles)) and low RFU signal (quenched 4 nM ADP* and 40 μ g/ml Q-Ab (A*-Q-Ab, dark red triangles)) was recorded over time at 25°C and the data recorded between 30-90 minutes is shown. The negligible effect of 4 nM SsoExosome (Sso) and 2000 nM RNA_{20mer} on the maximum RFU (Sso + RNA (A*), green circles) is shown. Additionally, their individual effect on Q-Ab quenching of ADP* (A*-Q-Ab) is shown for Sso (dark green circles) and RNA (black triangles). A control containing no fluorescence (Q-Ab only, white diamonds) was also plotted.

(b) The same graph is also shown zoomed in at (<1,400 RFU) in order to visualise the RFU of the quenched A*-Q-Ab reaction (~800) and the RFU of Q-Ab background controls (~200). Graphs were plotted in GraFit and labelled in GIMP (v2) (GIMP, n.d.; Leatherbarrow, 2009).

Assay	Average RFU	Std
A*	12207	532
Sso+RNA (A*)	12336	588
A*-Q-Ab	836	147
Sso (A*-Q-Ab)	882	154
RNA(A*-Q-Ab)	844	142
Q-Ab	216	72

Table 6.4 RFU Window with SsoExosome Assay Components

The average RFU between 30-90 minutes, for maximum RFU signal (unquenched 4 nM ADP* (A*)) and low RFU signal (quenched ADP*-Q-Ab (A*+Q-Ab)), background RFU (40 µg/ml Q-Ab), and the effect of 4 nM SsoExosome (Sso) and 2000 nM RNA_{20mer} (RNA) are listed. The standard deviation (Std) of data points are also indicated for each assay.

In summary, both SsoExosome and RNA had no effect on either the unquenched ADP* RFU or quenching of ADP* by Q-Ab, and did not affect the assay window. Any changes in RFU measured using this high-throughput plate assay would therefore be directly a result of ADP produced from RNA degradation mediated by SsoExosome.

6.4.6.2 SsoExosome Titration

The results presented in Section 6.4.6.1 suggested that none of the assay components affect the RFU window; hence the next step was to determine the optimal concentration of SsoExosome. A range of SsoExosome enzyme concentrations (0, 1, 2.5, 3, 3.5, 4, 5, 10 nM) were titrated against a set concentration of RNA_{20mer} (2000 nM) and the RFU was recorded over time as described in Section 6.3.4. The concentration range of SsoExosome was higher than those tested for EcPNPase (Section 6.4.5.2), as preliminary work suggested that SsoExosome exhibited a lower activity at 37°C. This may have been a result of using sub-optimal assay temperatures; the thermophile SsoExosome can degrade RNA at 65°C. It was therefore expected that the activity would be reduced further at 25°C and so a higher range of enzyme concentrations were tested.

The RFU recorded over time increased in a concentration-dependant manner for all SsoExosome concentrations. However, only SsoExosome concentrations greater than 3 nM exhibited the recommended 50-80% increase in RFU over the assay duration (Figure 6.24). Therefore, [SsoExosome] below 3 nM were not appropriate for producing the response recommended by the Bellbrook manual and were not chosen for subsequent RNA_{20mer} substrate titration experiments.

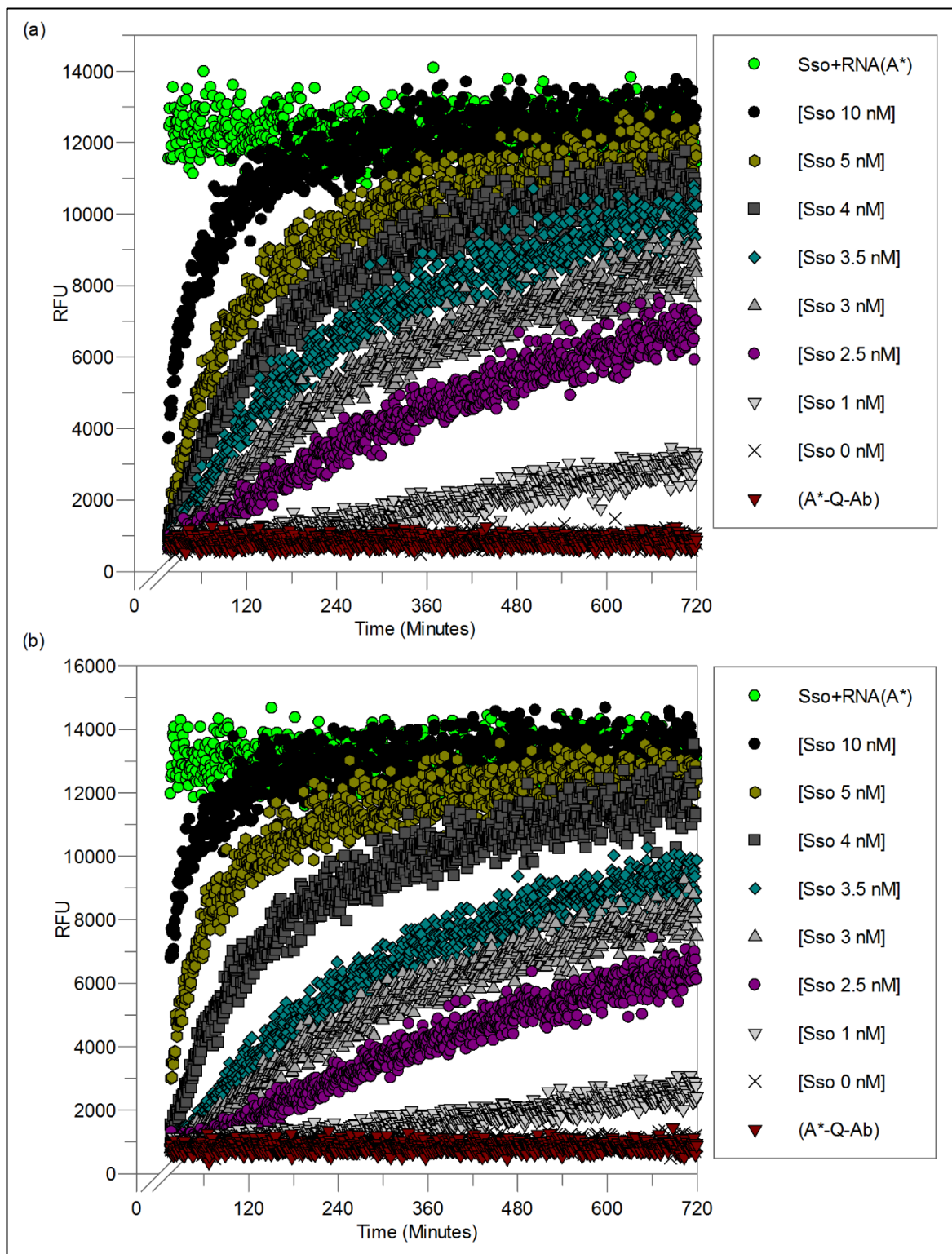


Figure 6.24 SsoExosome Titration: RFU vs Time

The RFU recorded over time (minutes) from (a) repeat 1 and (b) repeat 2 are shown. The maximum RFU signal (unquenched (Sso + RNA (A*)), green circles) and low RFU signal (quenched (A*-Q-Ab), dark red triangles) are indicated alongside a range of SsoExosome concentrations (0, 1, 2.5, 3, 3.5, 4, 5 and 10 nM) which are coloured from light grey to black respectively. Data collected within the first 15-minutes were discarded as standard. Graphs were plotted in GraFit (Leatherbarrow, 2009).

The RFU values, shown in Figure 6.24, were converted into $[ADP_{20x}]$ (nM) using an ADP standard curve as described in Section 6.3.3. The ADP standard used within this assay was suitable (within 10% error) for converting RFU values into $[ADP_{20x}]$ within the range of 16-128 nM ADP_{20x} . The SsoExosome titration assay contained 2000 nM RNA_{20mer} and therefore the ADP standard could be

used to convert the initial 5% of ADP_{20x} increase accurately (5% 2000 nM RNA_{20mer} substrate degradation produces 100 nM ADP_{20x}). Accordingly, the data from the first 5% of the reaction, shown in Figure 6.25, were fitted to a linear equation and the resultant rate of [ADP_{20x}] (nM) production over time (minutes) was calculated for each [SsoExosome] (Figure 6.26). Only SsoExosome concentrations greater than 3 nM reached 5% RNA_{20mer} substrate degradation over the duration of the assay (100 nM ADP_{20x} produced).

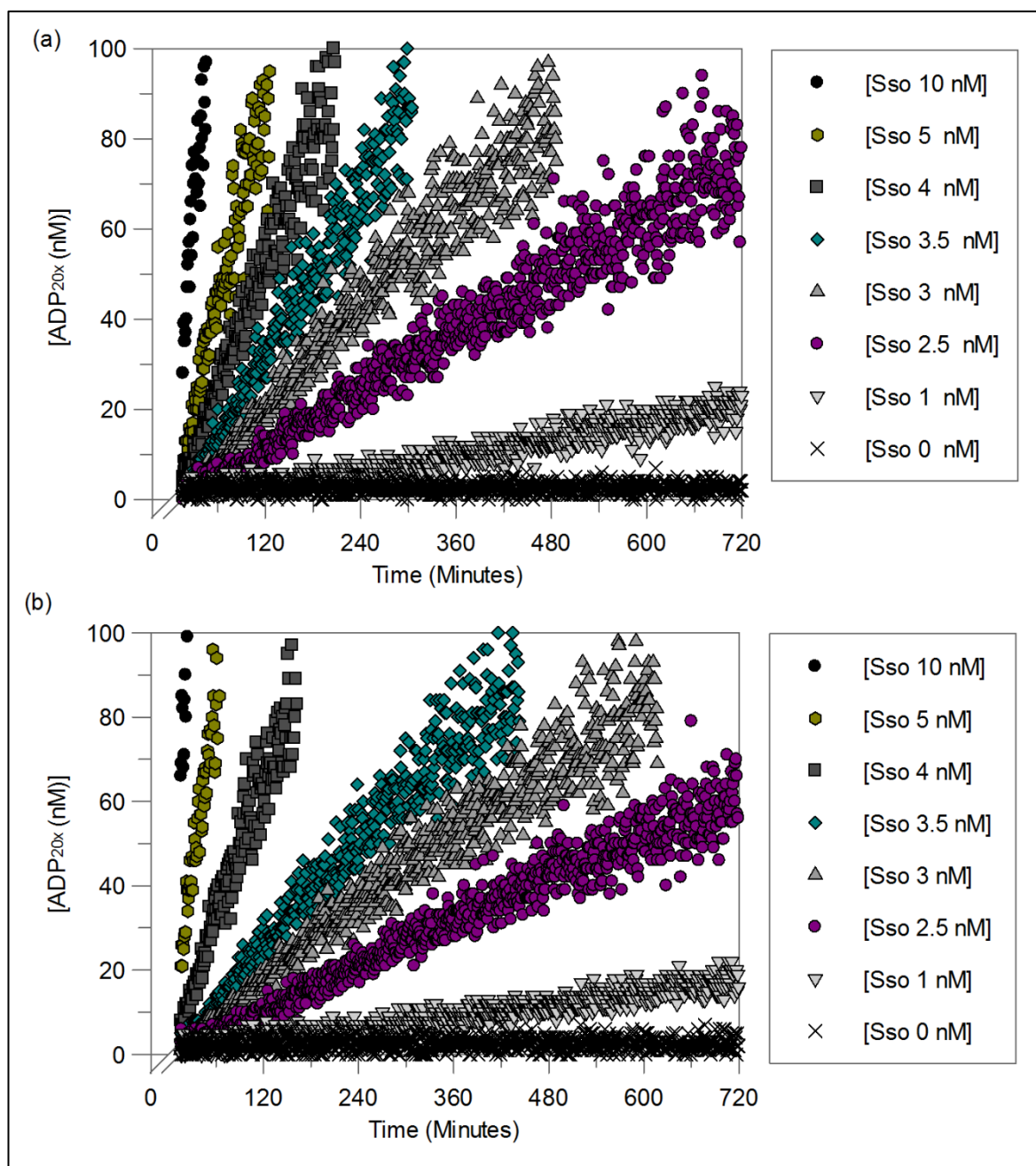


Figure 6.25 SsoExosome Titration: Raw Data [ADP_{20x}] vs Time

The [ADP_{20x} (nM)] produced from a range of SsoExosome concentrations (0, 1, 2.5, 3, 3.5, 4, 5 and 10 nM), were plotted over time (minutes) for (a) repeat 1 and (b) repeat 2 (minus 15-minutes equilibration). Graphs were plotted in GraFit (Leatherbarrow, 2009).

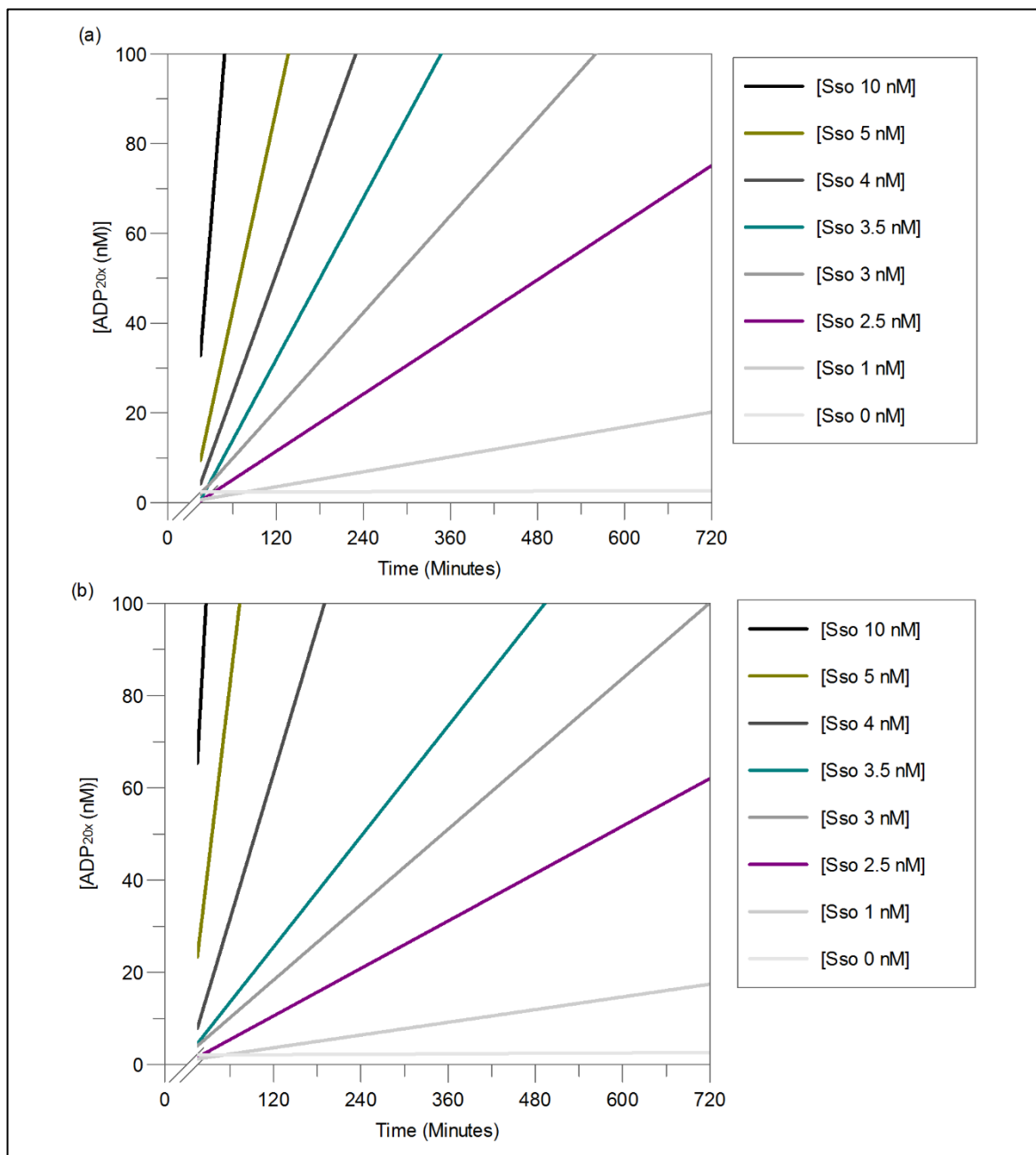


Figure 6.26 SsoExosome Titration: Linear Fit [ADP_{20x}] vs Time

The [ADP_{20x}] (nM) produced over time (minutes) for a range of SsoExosome concentrations (0, 1, 2.5, 3, 3.5, 4, 5 and 10 nM) was fitted to the linear equation ($y = m \cdot x + c$). Results for (a) repeat 1 and (b) repeat 2 (minus 15-minutes equilibration) are shown with increasing [SsoExosome] coloured from light grey to black respectively. Graphs were plotted in GraFit (Leatherbarrow, 2009).

The calculated initial rates were then plotted against the enzyme concentration and a graph, showing the average of two repeats, is shown in Figure 6.27, with error bars denoting the standard deviation. The enzyme concentrations, for which a linear increase in rate were detected, are shown in red (Figure 6.27). Ideally the enzyme concentration, which was within this linear region, should be selected for further enzyme kinetics and according to the data shown in Figure 6.27; this included [SsoExosome] between 3.5-5 nM.

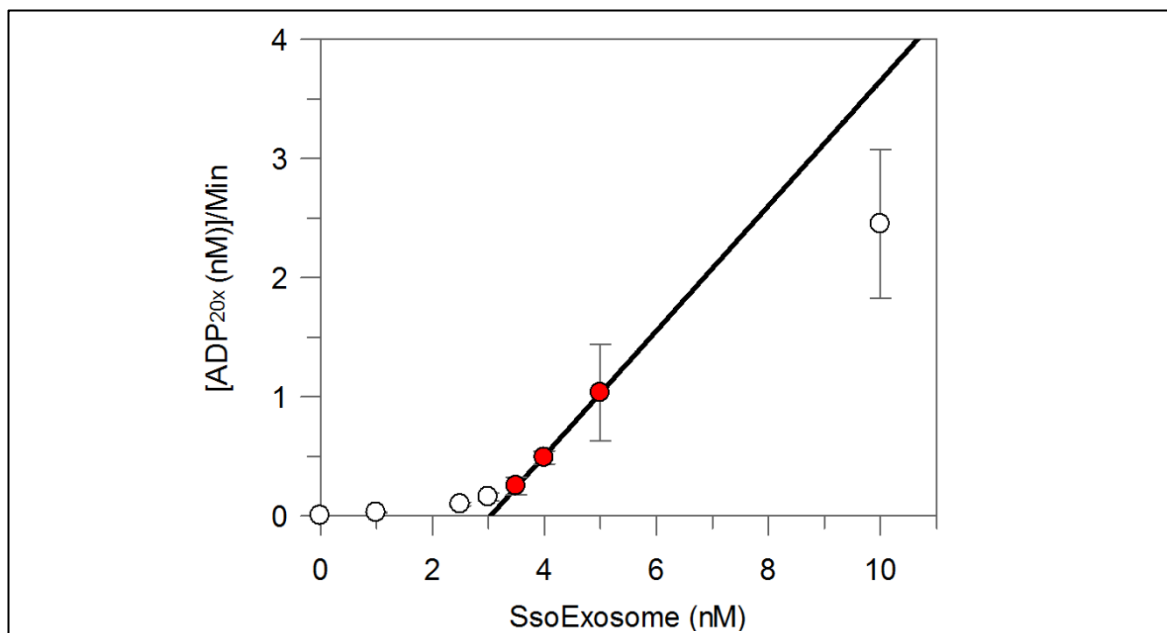


Figure 6.27 SsoExosome Titration: Rate of ADP_{20x} Production vs [SsoExosome]

The average rate of [ADP_{20x} (nM)] produced per minute, for a range of SsoExosome concentrations (0, 1, 2.5, 3, 3.5, 4, 5 and 10 nM) was plotted (white circles) from duplicate data. The linear equation ($y = m \cdot x + c$, black line) was fitted to appropriate linear data points (red circles) using GraFit and error bars of the standard deviation are provided (Leatherbarrow, 2009).

Due to the previously established 15-minute equilibration time for ADP*-Q-Ab quenching at 25°C, it was important to select an enzyme concentration with the minimum amount of RNA degradation following 15-minutes. For this reason, since RNA degradation had already started during this equilibration time, neither 5 or 10 nM SsoExosome were selected (Figure 6.26). However, in assays containing 3.5-4 nM SsoExosome, less than 0.2% (~4 nM ADP_{20x} produced) of the reaction had occurred within the first 15-minutes of the assay. Hence either 3.5 or 4 nM SsoExosome were suitable for generating a 50-80% RFU signal change in the assay, and ensuring the initial rates could be calculated following the 15-minute ADP*-Q-Ab thermal-equilibration. Both enzyme concentrations were also low enough to prevent enzyme waste and helped ensure that the concentration of RNA_{20mer} substrate would be in excess, for determining Michaelis-Menten kinetics. So the higher concentration, 4 nM SsoExosome was chosen in order to reduce the assay duration.

In summary, following a 15-minute ADP*-Q-Ab equilibration, the degradation of 2000 nM RNA_{20mer} by 4 nM SsoExosome could be monitored directly in real-time; RFU data were collected in high-throughput and the concentration of ADP_{20x} (nM) produced over time was calculated using an ADP_{20x} standard curve. The concentration of SsoExosome (4 nM) was 8-fold higher than EcPNPase (0.5 nM) (Section 6.4.5.2) and this possibly was a result of sub-optimal assay temperatures (25°C). Nevertheless, a range of substrate concentrations were titrated against 4 nM SsoExosome and the initial rates were used to calculate Michaelis-Menten kinetics. The results of which are described below in Section 6.4.6.3.

6.4.6.3 SsoExosome Degradation Activity (0.5 mM Phosphate)

In order to determine the canonical SsoExosome RNA degradation activity, 4 nM SsoExosome (in 0.5 mM phosphate) was incubated at 25°C with a range of RNA_{20mer} substrate concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 & 2600 nM) as in Section 6.3.5. The RFU values recorded over time (Figure 6.28) were converted into [ADP_{20x}] (nM) using a ADP standard curve as described in Section 6.3.3. The ADP standard used within this assay was suitable (within 10% error) for converting RFU values into [ADP_{20x}] within the range of 16-256 nM ADP_{20x}. Therefore, the ADP standard could be used to convert the initial 10% of RNA degradation accurately for substrate concentrations between 160-2560 nM RNA_{20mer}. Ideally, this standard curve should be expanded to include values at the lower range (100 nM [ADP_{20x}]) and this required further optimisation. Nevertheless, the data from the initial 10% of the reaction (minus 15-minutes) were plotted against time (Figure 6.29 (a)). The data plotted in Figure 6.29 (a) were fitted to a linear equation (Figure 6.29 (b)) and the rate of reaction was calculated. These initial rates were then plotted against the concentration of substrate to give the Michaelis-Menten curve shown in Figure 6.30.

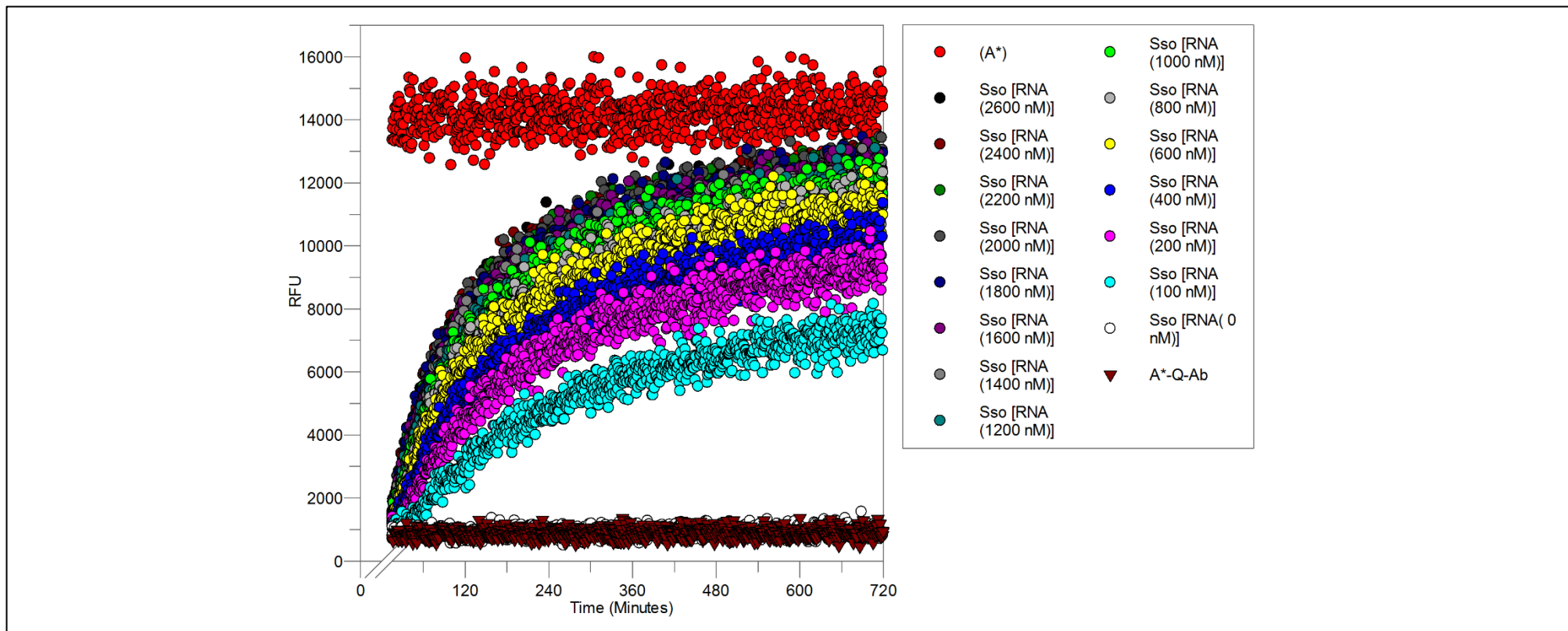


Figure 6.28 RNA_{20mer} Degradation by SsoExosome (0.5 mM Phosphate): [ADP_{20x} (nM)] vs Time

The RFU recorded over time (minutes) for the maximum RFU signal (unquenched (A*), bright red circles) and low RFU signal (quenched (A*-Q-Ab), dark red triangles) are indicated for assays containing 4 nM SsoExosome (Sso) (+ 0.5 mM phosphate), titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) which are coloured from light grey to black respectively. Data collected within the first 15-minutes were discarded as standard. Graphs were plotted in GraFit (Leatherbarrow, 2009).

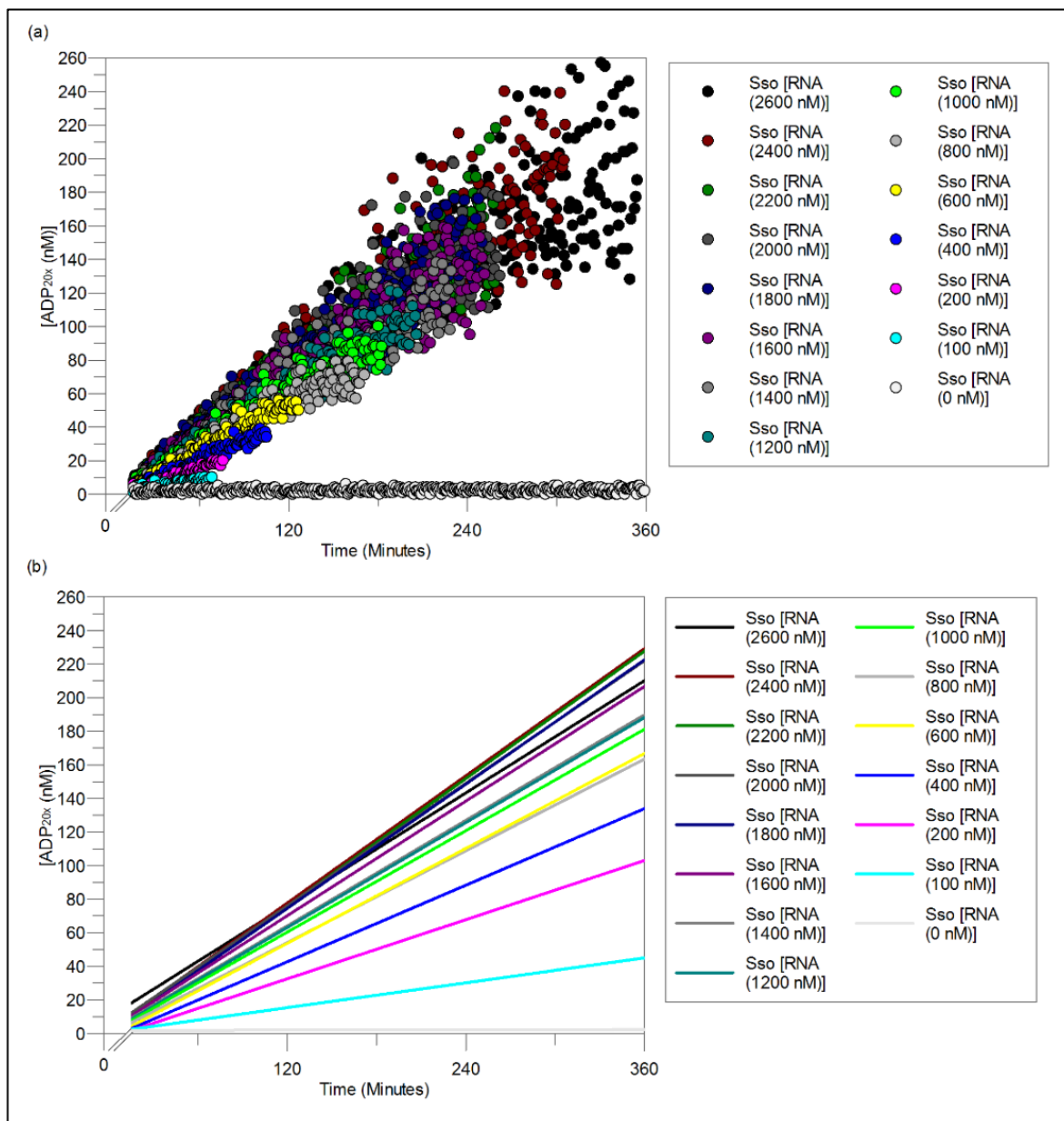


Figure 6.29 RNA_{20mer} Degradation by SsoExosome (0.5 mM Phosphate): [ADP_{20x} (nM)] vs Time

(a) The [ADP_{20x} (nM)] produced over time (minutes), for an assay containing 4 nM SsoExosome (0.5 mM phosphate) titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) is shown (minus 15-minutes equilibration). (b) The data were fitted to a linear equation ($y = m \cdot x + c$) and the graph shows increasing [SsoExosome] coloured from light grey to black respectively. Graphs were plotted in GraFit (Leatherbarrow, 2009).

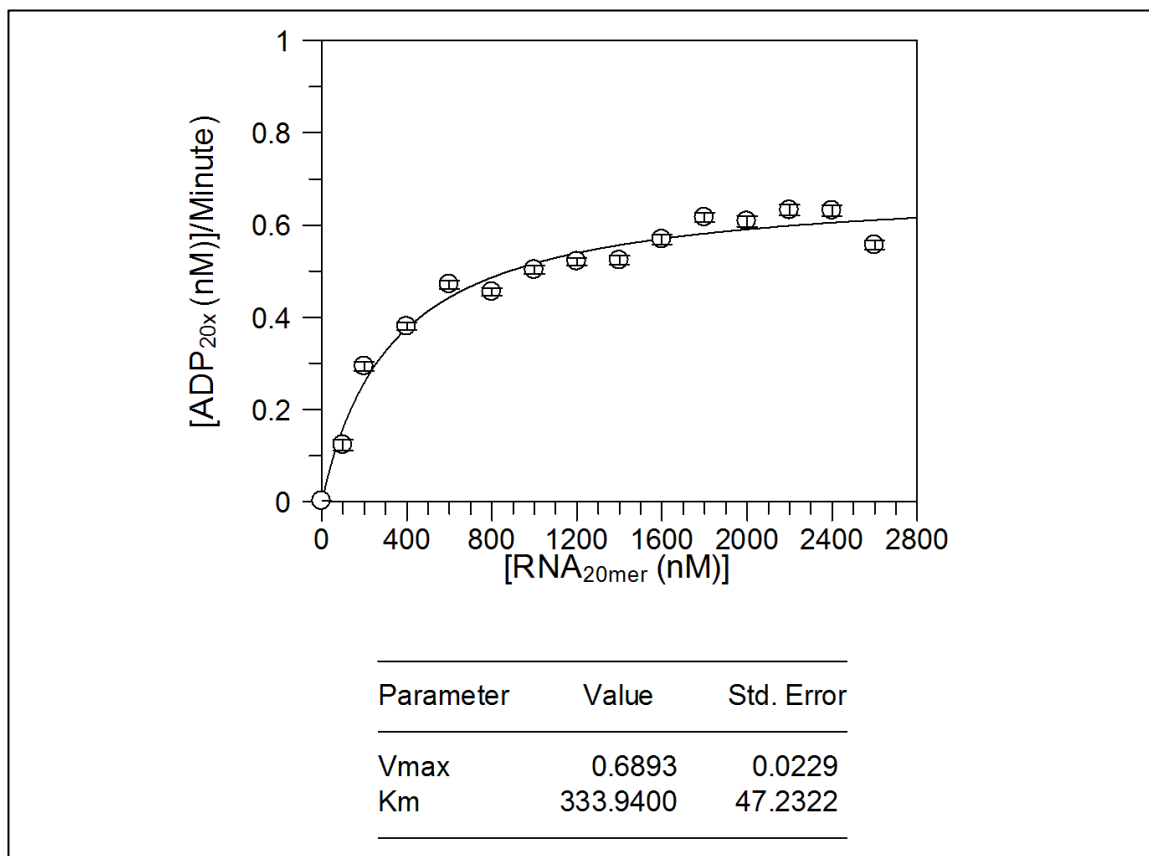


Figure 6.30 RNA_{20mer} Degradation by SsoExosome (0.5 mM Phosphate): Michaelis-Menten

The rate of [ADP_{20x} (nM)] produced per minute, for an assay containing 4 nM SsoExosome (0.5 mM phosphate) titrated with a range of RNA_{20mer} concentrations (0, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400 and 2600 nM) is shown. The data were fit using the Michaelis-Menten equation shown in Equation 6.2 and the resultant V_{max} (nM/Minute) and K_m (nM) are indicated with their error of standard deviation. The graph was plotted and fit using GraFit (Leatherbarrow, 2009)

The data plotted in Figure 6.30 indicated that SsoExosome, in the presence of 0.5 mM phosphate, reached a V_{max} of 0.69 nM/Minute and had a K_m of 334 nM. The k_{cat} and catalytic efficiency were also calculated using Equation 6.3 and Equation 6.4 respectively and the results were shown in Table 6.5. Although the error calculated for the V_{max} was quite reasonable (<10%), the Michaelis-Menten curve shown in Figure 6.30, reached the V_{max} relatively quickly compared to EcPNPase (Section 6.4.5.5) and the K_m error was ~14%. Therefore, it would be worthwhile to optimise this assay; obtaining more initial rate data points at lower [RNA_{20mer}], before the SsoExosome reaches maximum velocity, may improve the K_m error.

Conditions	Parameter	Value	Std. Error
SsoExosome (0.5 mM phosphate)	[Et], nM	4	Na
	Vmax, nM Min ⁻¹	0.69	± 0.02
	Km, nM	333.94	± 47.23
	Kcat, RNA _{20mer} Min ⁻¹	0.17	
	Catalytic Efficiency, M ⁻¹ s ⁻¹	8,485	

Table 6.5 SsoExosome Kinetic Parameters

The conditions and kinetic parameter for the plate assay are listed, including total enzyme concentration [Et], maximum velocity (V_{max}), Michaelis Constant (K_m), turnover number (k_{cat}) and catalytic efficiency. The units are provided with the values and a standard deviation error is listed where appropriate.

In summary, the assay used to determine SsoExosome kinetic parameters requires further optimisation. In order to examine, if like EcPNPase, SsoExosome has a higher phosphate preference, more experiments are required and these are part of ongoing work within our group. Nevertheless, the information presented within this section provided proof of concept results and demonstrated that the plate assay developed was suitable to test the 3'-5' degradation activity of SsoExosome. The application of this real-time assay can therefore be used to quantify the degradation activity of PNPase homologs across the two domains of life, prokaryotes and archaea. Although kinetic parameters determined for SsoExosome were not obtained from triplicate data-sets, the results supported the use of the high-throughput assay for determining SsoExosome kinetics in real-time.

6.5 Conclusions

This chapter described the successful development of a real-time high-throughput plate reader assay for quantifying PNPase-mediated degradation activity on a poly(A)_{20mer} RNA substrate. The methods used to modify the currently available Transcreener® ADP2 Fluorescent Intensity (FI) assay, which was optimised for end-point ADP detection (Bellbrook), to be used for real-time studies instead, were provided in Section 6.3. The results provided in Section 6.4, validate the assay parameters and conditions, prior to determining enzyme kinetics.

The results presented within this chapter consisted mainly of proof of concept/preliminary data; more optimisation is required and the assays need to be conducted in triplicate. However, the data in this chapter clearly demonstrated a novel application of the high-throughput assay to determine kinetic parameters, for a range of 3'-5' PNPase homologs, in real-time. For example, both EcPNPase and SsoExosome were suitable for generating data which fitted to a Michaelis-Menten curve and could be utilised for calculating detailed kinetic parameters including the V_{max} , K_m , k_{cat} and catalytic efficiency. Also by using an ADP standard curve within this assay, the degradation of RNA substrates mediated by PNPase homologs was not just reported as changes in RFU over time, but could be more relevantly quantified in terms of [ADP] produced. A basic summary of the kinetic parameters for EcPNPase and SsoExosome are provided in Table 6.6

Enzyme	Parameter	Value	Std. Error
EcPNPase (5 mM phosphate)	[Et], nM	0.5	Na
	V_{max} , nM Min ⁻¹	1.22	±0.03
	k_{cat} , RNA _{20mer} Min ⁻¹	2.45	
SsoExosome (0.5 mM phosphate)	[Et], nM	4	Na
	V_{max} , nM Min ⁻¹	0.69	± 0.02
	k_{cat} , RNA _{20mer} Min ⁻¹	0.17	

Table 6.6 PNPase Homologs Kinetics Summary

The conditions and kinetic parameter for plate reader assays are listed, including the enzyme, phosphate concentration, total enzyme concentration [Et], maximum velocity (V_{max}) and turnover number (k_{cat}). The units are provided with the values and a standard deviation error is listed where appropriate.

As reviewed in Godefroy-Colburn & Grunberg-Manago, 1975, kinetic studies of PNPases have been conducted using various synthetic poly(A) substrates. Since values reported in Table 6.3 were not from triplicate data-sets, care must be taken when comparing these kinetic constants in detail. Nevertheless, although the focus of this work was not to compare kinetic parameters comprehensively, it was important to demonstrate that EcPNPase kinetic values, calculated within this study, aligned with previously published data.

It has been reported by multiple publications that the K_m values decrease with increasing RNA chain length. For example, the K_m values of *Micrococcus luteus* PNPase were reported to decrease from 2.5 to 0.033 mM as the chain length (n) of (Ap)_nA increased from 2 to 8 nucleotides (Chou & Singer, 1970b). Similarly, the K_m values of *M. luteus* PNPase decreased from 0.067 to 0.033 mM as the chain length of (Ap)_nA increased from 4 to 5 nucleotides (Chou & Singer, 1970a). Additionally, K_m values of *E. coli* PNPase were reported to be within the order of 0.1 mM for (pA)₄. While the K_m for poly(A)₅₀₀ chain was in the low nM range (as reviewed in Godefroy-Colburn & Grunberg-Manago, 1975); with K_m values of 3 and 4 nM for *E. coli* PNPase (Godefroy, Cohn, & Grunberg-Manago, 1970) and *M. luteus* PNPase respectively (Chou, Singer, & McPhie, 1975). The K_m value reported within this study, for the *E. coli* PNPase-mediated degradation of a Poly(A)_{20mer} RNA substrate was 940 nM ±66 nM and this was as expected, lower than the values previously reported above for an 8-nucleotide substrate but higher than those calculated for a Poly(A)_{500mer}.

Additionally, although the results listed in Table 6.3 were not conducted in duplicate. The K_m and k_{cat} values for EcPNPase degrading a Poly(A)_{20mer}, correlated well with values previously reported for EcPNPase degrading 5601 and 5660 substrates, derived from the *rpsO-pnp* operon of *S. coelicolor* (Chang, Cozad, Mackie, & Jones, 2008). The K_m values published by Chang and colleagues were calculated from a range of 5601 and 5660 RNA concentrations (200-3300 nM and 200-2000 nM respectively) similar to the [Poly(A)_{20mer}] used within this study (100-2600 nM). The K_m calculated for the 5601 transcript (780 nM ±20 nM) was slightly lower than the K_m of the Poly(A)_{20mer} substrate (940 nM ±66 nM) whereas the K_m for the 5660 transcript was higher (6330 nM ±440 nM) than both values. This suggested that more Poly(A)_{20mer} RNA substrate was needed to reach half V_{max} compared to the 5601 transcript, but less was needed compared to the 5660 transcript. The turnover rate (k_{cat}) of EcPNPase for the 5601 and 5660 transcript were higher (11.2 ±0.2 Min⁻¹ and 5.55 ±0.34 Min⁻¹ respectively), than the Poly(A)_{20mer} RNA substrate (2.45 Min⁻¹); suggesting that within the conditions tested, EcPNPase degraded more 5601 and 5660 transcripts per minute than the Poly(A)_{20mer}. Even though the k_{cat} for the 5660 transcript was high, because the K_m was also high, the catalytic efficiency of EcPNPase degrading the 5660 transcript was ~ 3-fold lower (14,607 M⁻¹s⁻¹) than the enzyme degrading the poly(A)_{20mer} substrate (43,402 M⁻¹s⁻¹). In contrast, the high k_{cat} and low K_m of EcPNPase for the 5601 transcript, meant that the catalytic efficiency was 5-fold higher (239,316 M⁻¹s⁻¹) than the poly(A)_{20mer} substrate. The publication by Chang and colleagues suggested that the phosphate concentration used was not limiting (1 mM),

and therefore the differences in these kinetic values may be due to substrate preference. In conclusion, for the EcPNPase-mediated degradation of Poly(A)_{20mer} RNA substrate, the kinetic values calculated within this chapter were similar to those previously published. This supported the use of the high-throughput assay for determining EcPNPase kinetics in real-time.

Comparing the V_{\max} values for SsoExosome with EcPNPase would not be appropriate as an 8-fold higher concentration of SsoExosome was utilised (4 and 0.5 nM respectively). However, the turnover number (k_{cat}) could be compared as it was the ratio of V_{\max} to $[E_t]$. The k_{cat} of SsoExosome, in the presence of 0.5 mM phosphate, was lower than EcPNPase in 5 mM phosphate (0.17 and 2.45 RNA_{20mer}/Minute respectively). These results suggested that the amount of RNA_{20mer} degraded per minute, by one enzyme molecule, was less for SsoExosome. This reduced turnover rate may have been a result of the sub-optimal temperature; the archaeal exosome from the thermophile *S. solfataricus* is known to exhibit activity at high temperatures. Unfortunately, this variable could not be adjusted due to the issue, mentioned previously in Section 6.4.3, of the thermal-equilibration between ADP* and Q-Ab at higher temperatures. Alternatively, the lower k_{cat} calculated for SsoExosome may have been due to sub-optimal phosphate concentrations. Perhaps, like EcPNPase, SsoExosome may exhibit optimal enzyme activity in higher phosphate concentrations. Due to time limitations, this was not conducted and therefore is currently the subject of following investigations within our group.

Once the real-time plate reader assay has been fully optimised and canonical enzyme kinetic information is available, in triplicate, high-throughput data collection should be possible. These data could then be utilised to calculate kinetic parameters, for comparing the turnover rate (k_{cat}) and catalytic efficiency of individual PNPase homologs, in a range of conditions. For example, as long the sequence is known and a detectable amount of ADP is released from RNA substrate degradation, a range of different physiologically relevant substrates may be tested. Additionally, the effects of inhibitors may also be studied, including the effects of citrate and other metabolites on 3'-5' PNPase homolog activity. Data collection using this method would be more time-effective than previously available gel-based methods, which is economically beneficial. The impact and the possible applications of this assay are discussed further in Chapter 7.

7 Discussion and Future Perspectives

To maintain cellular homeostasis a repertoire of intricate systems must be able to respond to internal changes and environmental stimuli. For example, central metabolism can meet the needs of a cell by providing the specific building blocks and energy required for normal cell growth. However, in situations where a cell experiences either internal or external changes, metabolism can be adjusted for cell survival (Nielsen, 2003). As reviewed in Wegner, Meiser, & Weindl, 2015, these adaptations may occur at many different levels. Modification may occur at the gene level by encoding different enzyme isoforms, or at the transcriptional level by regulating which genes are expressed. Factors such as alternative splicing, mRNA stability, translation, and protein degradation, which can all control the abundance of enzymes, may also allow long-term adaptations. Alternatively, these adaptations may provide short-term regulation of pathways that need to be constitutively active. For example, metabolic fluxes may fine tune the activity of enzymes already present within the cell (Wegner *et al.*, 2015).

It is considered that a metabolite has a role in regulating cellular metabolism if its concentration fluctuates in response to cellular changes and if the activity of one or more specific enzymes is changed in response (Wegner *et al.*, 2015). There are a wide range of metabolites which fulfil these requirements and regulate cellular metabolism, at both the transcriptional and translational level. An excellent example of a major cellular adaptation at the transcriptional level is the metabolite-mediated reprogramming that is required by a cell to survive in situations of oxygen or nutrient depletion. In these circumstances, metabolites are known to interact with transcription factors and drive large rearrangements of metabolic processes. For example, the processes controlling anaerobic or aerobic cellular respiration involve a number of regulatory steps, and metabolites are reported to be essential for adapting the cells metabolism to enable survival in environments where oxygen levels may fluctuate (Wegner *et al.*, 2015). Additionally, in a number of organisms, metabolites can regulate adaptations at the translational level, by directly binding to riboswitches and inducing a conformation change in the target mRNA to either activate or inactive translation (Li & Breaker, 2013). It has also been reported that metabolite fluxes may regulate cellular metabolism in a feedback loop in stem cells. More specifically, an unsaturated fatty acid was found to bind to a translational activator in stem cells and induce a conformational change, which prevented the binding and translation of a fatty acid desaturase (stearoyl-CoA desaturase-1 (SCD1)) mRNA (Clingman *et al.*, 2014).

In conclusion, cellular metabolite levels are known to fluctuate in a range of conditions and can directly or indirectly affect gene transcription and protein translation. However, their indirect effect on regulating RNA turnover, via interactions with ribonucleases, is poorly understood. Consequently, this study initially reviewed the recurrent evolution of a physical association between the enzymes of RNA degradation and central metabolism, prior to exploring the concept

that a conserved communicative link may exist between RNA turnover and central metabolism. The key results of this study are discussed in the following Section 7.1 before future perspectives are addressed in Section 7.2.

7.1 Regulating RNA Turnover in Response to Cellular Metabolism

There is a growing body of evidence which suggests that small molecule metabolites link the cellular metabolic status of a cell and RNase activity in bacteria. A physical association between metabolic enzymes and RNases, in a range of organisms, has been documented over the last two decades. For example, the glycolytic enzymes enolase and aconitase are canonical components of the *E. coli* and *C. crescentus* degradosome respectively (Hardwick, Chan, Broadhurst, & Luisi, 2011; Miczak, Kaberdin, Wei, & Lin-Chao, 1996; Py, Higgins, Krisch, & Carpousis, 1996). Furthermore, degradosome-like complexes containing two glycolytic enzymes (enolase and phosphofructokinase), four RNases (RNase Y, RNase J1, RNase J2 and PNPase) and a helicase (CshA) appear to be present in *Bacillus subtilis* (Commichau *et al.*, 2009; Lehnik-Habrink *et al.*, 2010) and *Staphylococcus aureus* (Roux, DeMuth, & Dunman, 2011). Not only has the physical association of metabolic enzymes with RNA degradation machinery been discovered, but small molecules, including ATP (Del Favero *et al.*, 2008), c-di-GMP (Tuckerman, Gonzalez, & Gilles-Gonzalez, 2011) and citrate (Nurmohamed *et al.*, 2011), have been found to modulate the RNase activity of *E. coli* PNPase. The signalling molecule/alarmonone (p)ppGpp has also been shown to modulate PNPase activity in *Streptomyces* (Gatewood & Jones, 2010) and *Nonomuraea sp.* (Siculella *et al.*, 2010). The results presented in this thesis suggest that the modulation of RNase activity by small molecules, may be more widespread than previously thought and may represent an evolutionarily conserved communication mechanism.

7.1.1 Interactions Between Citrate and Homologous 3'-5' Phosphorolytic Ribonucleases

The first indication that the interaction of small molecule metabolites with RNases may be conserved, came upon studying the detailed interactions of citrate co-crystallised within the PNPase enzymes of *E. coli* and *H. sapiens* (Lin *et al.*, 2012; Nurmohamed, Vaidialingam, Callaghan, & Luisi, 2009). When the citrate and Mg²⁺ binding residues of these two enzymes were compared with a small group of other PNPase homologs (for which either sequence or structural information was available), they were discovered to be highly conserved. This result promoted more detailed bioinformatics analysis; protein sequence comparisons of PNPase from 3509 prokaryotic and 252 eukaryotic species with that of the archaeal exosome from 69 archaeal species, suggested that citrate-binding residues were highly conserved across evolution. The observation that these residues were conserved across all three domains of life showed that citrate-binding may exist in evolutionarily diverse PNPase homologs. Unfortunately, conclusive evidence of citrate binding to the archaeal exosome from *S. solfataricus* could not be obtained during this study, since

only weak X-ray diffraction was recorded for crystals containing the SsoExosome soaked with citrate. Therefore, an alternative *in silico* molecular docking approach was utilised to explore citrate binding.

The *in silico* molecular docking validation, described in Chapter 4, suggested that MOE was a suitable program for predicting citrate-PNPase interactions. In summary, through *in silico* molecular docking calculations, two citrate molecules were shown to interact with the catalytic centre and the neighbouring RNA-binding regions in the same way as previously identified in the co-crystallised structures of EcPNPase and hPNPase (Lin *et al.*, 2012; Nurmohamed *et al.*, 2009). Therefore, *in silico* docking was applied to predict citrate interactions with other PNPase homologs, for which no binding information was currently available. Consequently, subsequent docking calculations predicted that two citrate molecules could also bind within the active site of SsoExosome and a bacterial PNPase homolog from *S. antibioticus*, and that the residues and types of interactions were conserved. These results were particularly interesting since the sequence conservation of both the active site, and consequently the putative citrate-binding residues, was much weaker in the more distantly related archaeal exosomes. In particular, the PBR S(S/T)S motif, that is required for phosphate-binding in PNPase (Nurmohamed *et al.*, 2009), has been replaced by an G(T/S)R motif in the archaeal exosomes. Despite the sequence variation, this motif has been shown to bind phosphate (Hartung, Niederberger, Hartung, Tresch, & Hopfner, 2010; Navarro, Oliveira, Zanchin, & Guimarães, 2008) and *in silico* data presented here suggest that two molecules of citrate can dock into the SsoExosome. Furthermore, regardless of the sequence variation, the S value for citrate binding at the Cit 1 site of all four PNPase homologs (EcPNPase, hPNPase, SsoExosome and SanPNPase) was similar. By contrast, the sequence variation of the PBR motif may affect the second citrate binding event. The docking scores for the Cit 2 site for SsoExosome (GSR motif) and SanPNPase (STS motif) were significantly higher than for EcPNPase and hPNPase, which both have the SSS motif. The physiological relevance of these different S values remains to be determined, especially since it is currently unknown whether one or two citrate molecules are required to inhibit PNPase activity.

Previous reports highlighted that, at physiological concentrations of citrate, attenuation of EcPNPase activity rather than complete inhibition is likely to occur *in vivo* (Nurmohamed *et al.*, 2011). The *in vitro* results presented within this work, suggested that PNPase from another bacterial species *Synechocystis sp* may also be susceptible to inhibition/attenuation by citrate and that this attenuation may be commonplace among prokaryotes. Furthermore, the 3'-5' degradation activity of both eukaryotic PNPase from human mitochondria and the archaeal exosome complex from *S. solfataricus*, was similarly reduced in the presence of citrate. The level of citrate-mediated EcPNPase inhibition previously published by Nurmohamed and colleagues was similar to the effects observed for all the PNPase homologs tested within this study. This suggests a conserved role for citrate in attenuating the function of PNPase homologs. It remains to be demonstrated

whether or not the citrate-mediated attenuation of exoribonuclease activity demonstrated *in vitro* is utilised as a regulatory strategy *in vivo*. Nevertheless, the reoccurring interaction between citrate and PNPase homologs across all three domains of life, may represent an ancient and evolutionarily conserved mechanism of regulating RNA turnover.

As mentioned previously, for a metabolite such as citrate to act as a regulator, the intracellular concentration of the metabolite must vary under different physiological conditions and the response elicited must be dose-dependent within this concentration range. Intracellular citrate concentrations in *E. coli* depend upon the carbon source and have been reported to range between 2 mM for growth on glucose or glycerol, and 20 mM for growth on acetate (Bennett *et al.*, 2009). Furthermore, the level of inhibition of EcPNPase by citrate observed *in vitro* correlated well with citrate concentrations in this range (Nurmohamed *et al.*, 2011). Therefore, for EcPNPase, there was evidence to support citrate-mediated regulation. However, the situation for other organisms is less clear. There is a lack of metabolomics data reporting normal intracellular citrate concentration(s); this information appears to be non-trivial to obtain.

High levels of citrate have been shown to lead to apoptotic cell death in *H. sapiens* and citrate has therefore been proposed for use in anticancer therapeutics (Icard *et al.*, 2012; Lu *et al.*, 2011; Sun *et al.*, 2010; Usenik & Legis, 2010). With this in mind, the significance of a citrate-PNPase interaction in human cells needs to be investigated further, especially as hPNPase has been shown to localise to the mitochondria, the location of citrate synthesis (Piwowarski *et al.*, 2003). Furthermore, in human melanoma cells induced to terminally differentiate and human progeria cells, which exhibit increased cellular senescence, the gene encoding hPNPase was identified to be upregulated. Since hPNPase has been linked to terminal differentiation and cellular senescence, understanding its activity and mechanisms of inhibition may therefore be of interest to cell renewal and research into ageing (Leszczyniecka *et al.*, 2002; Sarkar *et al.*, 2004).

7.1.2 Interactions Between TCA Metabolites and 3'-5' Phosphorolytic Ribonucleases

Following the observation that a conserved communicative link may exist between citrate and PNPase homologs, other TCA metabolites similar to citrate were also considered as potential inhibitors. The *in silico* molecular docking results suggested that the only two TCA metabolites predicted to bind to hPNPase more tightly than citrate were acetyl-CoA and succinyl-CoA. During *in vitro* studies these metabolites were also shown to affect hPNPase and EcPNPase activity more than citrate. Although the role of hPNPase within the mitochondria, which is also the location of the TCA cycle, is not fully understood, the physiological relevance of these two molecules interacting with hPNPase was interesting since they both have key roles in regulating the TCA cycle. Essentially, entry into the TCA cycle is controlled by acetyl-CoA, ATP and NADH; increasing the NADH/NAD⁺, acetyl-CoA/CoA, or ATP/ADP ratio promotes phosphorylation of the

pyruvate dehydrogenase complex. This prevents the conversion of pyruvate to acetyl-CoA, therefore inhibiting entry into the TCA cycle. Additionally, although the TCA cycle is primarily regulated by ATP and NADH concentrations, succinyl-CoA regulates the enzyme α -ketoglutarate dehydrogenase, which is one of the other control points of this metabolic process (reviewed in Berg, Tymoczko, & Stryer, 2002). Thus, when energy charge is high and biosynthetic intermediates acetyl-CoA and succinyl-CoA are in excess, entry and progression through the TCA cycle is inhibited (pink negative sign, Figure 7.1). By contrast, in low energy charge states when ADP is high, key enzymes regulating the TCA cycle are activated (green positive sign, Figure 7.1).

When the TCA cycle control checkpoints described above and shown in Figure 7.1, were compared to the *in vitro* gel-based assay results in Chapter 5, similarities were observed. More specifically, key metabolites involved in the TCA cycle entry and progression also affect the degradation activity of PNPase. For example the metabolites ATP and citrate, known to inhibit EcPNPase activity (Del Favero *et al.*, 2008; Nurmohamed *et al.*, 2011) (Figure 7.1, orange bolts) and the metabolites citrate, acetyl-CoA and succinyl-CoA, found within this study to affect both EcPNPase and hPNPase activity (Figure 7.1, yellow bolts); all have regulatory roles in the TCA cycle.

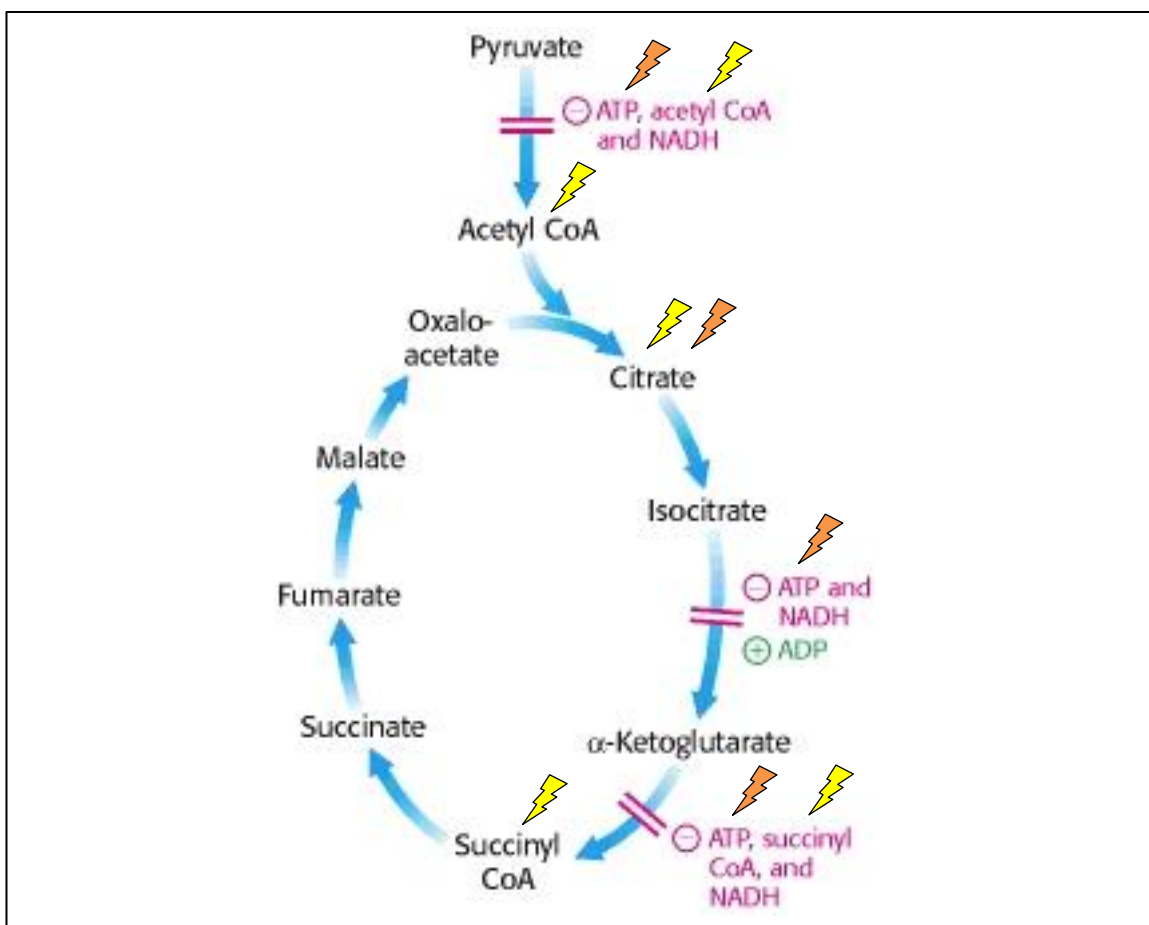


Figure 7.1 Control Points and Metabolites Regulating The TCA Cycle

The TCA cycle is primarily regulated by the concentration of ATP and NADH; high cellular concentrations prevent entry and regulation through the cycle (pink, negative sign). High cellular levels of acetyl-CoA also reduce entry into the TCA cycle. The key control points are the enzymes isocitrate dehydrogenase and α -ketoglutarate dehydrogenase. High cellular levels of ADP stimulate isocitrate dehydrogenase, whereas succinyl-CoA prevents the activity of α -ketoglutarate dehydrogenase; positively and negatively regulating the

metabolic pathway respectively. The orange coloured bolts highlight known inhibitors of PNPase activity in *E. coli* (Del Favero *et al.*, 2008). Whereas the yellow bolts indicate metabolites which were found in this study, to affect both EcPNPase and hPNPase degradation activity *in vitro*. Figure and legend adapted from (Berg *et al.*, 2002).

As highlighted in Figure 7.1, the TCA cycle involves a number of metabolites and research previously conducted by Huynen, Dandekar, & Bork, 1999 has demonstrated that TCA evolution, across all three domains of life, is incredibly complex. Recent publications have highlighted that progress still needs to be made to elucidate the nature of this cycle in some organisms. For example, it was commonly thought that the TCA cycle was incomplete in cyanobacteria, because the enzyme 2-oxoglutarate dehydrogenase was absent. This misunderstanding was overturned when a study discovered that two alternative enzymes could complete the cycle and as a result a decades-old belief was discarded (Steinhauser, Fernie, & Araújo, 2012).

There has been some controversy over the completeness of the TCA cycle in archaeal species. Nevertheless, whether like EcPNPase and hPNPase, archaeal exosome homologs are affected by TCA metabolites needs to be determined. It would be interesting to compare the effects of metabolite based PNPase inhibition between the archaeal exosomes from *S. solfataricus* and *Methanothermobacter thermautotrophicus*, since they represent species with a proposed complete and incomplete TCA cycle respectively (Huynen *et al.*, 1999). Furthermore, these species provide an excellent example for investigating archaeal exosome evolution since they represent members of the Euryarchaeota and Crenarchaeota phyla respectively. The structure of SsoExosome (Lorentzen & Conti, 2012) and *M. thermautotrophicus* exosome are both available (Ng, Waterman, Antson, & Ortiz-Lombardia, 2010), therefore similar *in silico* studies conducted within this work could be undertaken to predict metabolite interactions. Despite the diversity of the TCA pathway within archaea, any information gained from comparative studies may provide insight into whether TCA metabolites commonly modulate exosome activity. The question of whether a widely conserved communicative link exists between TCA metabolites and PNPase homologs, may then be addressed.

7.1.3 Interactions Between Phosphate-Rich Metabolites and 3'-5' Phosphorolytic Ribonucleases

The TCA metabolite-PNPase interaction study described here highlighted some other interesting avenues to investigate further. It was noticed that the CoA nucleotide part of acetyl-CoA and succinyl-CoA were predicted to interact with known, highly-conserved, citrate-binding residues within the catalytic centre and the neighbouring RNA-binding regions of hPNPase *in silico*. Thus, it was proposed that the nucleotide CoA group potentially binds and occludes the active site in a similar way to citrate, suggesting a potential phosphate recognition role, that may be widely conserved across PNPase homologs. The proposed interaction between CoA with PNPase homologs was very interesting because, as reviewed in Tahiliani & Beinlich, 1991, altered CoA

metabolism has been observed in diverse disease states including starvation, diabetes, vitamin B₁₂ deficiency and certain tumours; signifying its importance in a wide range of biological systems.

The prediction that the nucleotide part of CoA may be important for regulating PNPase, directed the study to examine other nucleoside/nucleotide containing metabolites that may affect activity. Following more *in silico* docking calculations and *in vitro* gel-based assays, phosphate-rich nucleotides including GTP, ppppG and ppGpp were shown to affect the 3'-5' degradation activity of hPNPase and EcPNPase *in vitro*. When comparing the levels of inhibition between PNPase homologs, despite the sequence and structural similarities of the binding pockets in the presence of phosphate-rich metabolites, attenuation of hPNPase activity was greater than EcPNPase. This aligned with previous research that suggested hPNPase was more sensitive to cellular phosphate levels than EcPNPase, and was actually inhibited by high phosphate concentrations (Portnoy, Palnizky, Yehudai-Resheff, Glaser, & Schuster, 2007). How hPNPase can 'sense' higher phosphate levels is currently unknown. In order to determine whether subtle differences in the architecture of the binding pockets may exist/facilitate these phosphate-binding preferences, the exact residues of hPNPase involved and their types of interactions with phosphate-rich metabolites need to be compared with EcPNPase.

The observation that GTP inhibited both hPNPase and EcPNPase was unsurprising since this energy-containing molecule has been previously implicated in regulating EcPNPase activity (Del Favero *et al.*, 2008). However, the finding that ppGpp inhibited the activity of EcPNPase was interesting as it conflicted with previous publications suggesting that the alarmone affected PNPase activity in *Streptomyces* but not in *E. coli* (Gatewood & Jones, 2010). As mentioned in Chapter 5, this conflicting data may have been a result of different assay conditions, and so the effect of this alarmone in *E. coli* needs to be examined further. This is especially important as (p)ppGpp has been shown to regulate a wide range of cellular processes. More specifically, upon exposure to stresses including heat shock and nutrient deprivation, the cellular concentration of (p)ppGpp increases and many important processes are reprogrammed. These processes include DNA replication, transcription and quorum sensing; collectively these modifications are referred to as the stringent response (reviewed in Dalebroux, Svensson, Gaynor, & Swanson, 2010; Hauryliuk, Atkinson, Murakami, Tenson, & Gerdes, 2015). (p)ppGpp has also been linked to bacterial persistence and antibiotic production in response to amino acid/energy starvation (Hauryliuk *et al.*, 2015). The interaction between (p)ppGpp and bacterial PNPase homologs are important to study as it may suggest a potential link between environmental sensing and RNA turnover.

As shown in Figure 7.2, the RelA-SpoT homologue (RSH) family are key for both synthesis and/or degradation of (p)ppGpp. Although long multi-domain RSH enzymes have not yet been identified in humans (Dalebroux *et al.*, 2010), a small alarmone hydrolases (SAH), which is a short, single-domain, mono-functional RSH, has been identified in humans and *Drosophila melanogaster*

(termed Mesh1) (D. Sun *et al.*, 2010). In this study, a *D. melanogaster* Mesh1 null mutant showed retarded body growth and impaired starvation response, suggesting that this enzyme may function in a similar manner to RSH enzymes. The publication also noted that Mesh1 has diverged greatly from SpoT during the evolution of animals and therefore it is possible that other enzymes have evolved to function as ppGpp synthetase (D. Sun *et al.*, 2010). However, given the absence of detectable (p)ppGpp levels in animals (Thammana, Buerk, & Gordon, 1976), the role of Mesh1 is currently unknown. Therefore, more research is required before a physiological relevance of the ppGpp-mediated attenuation of hPNPase activity can be deduced.

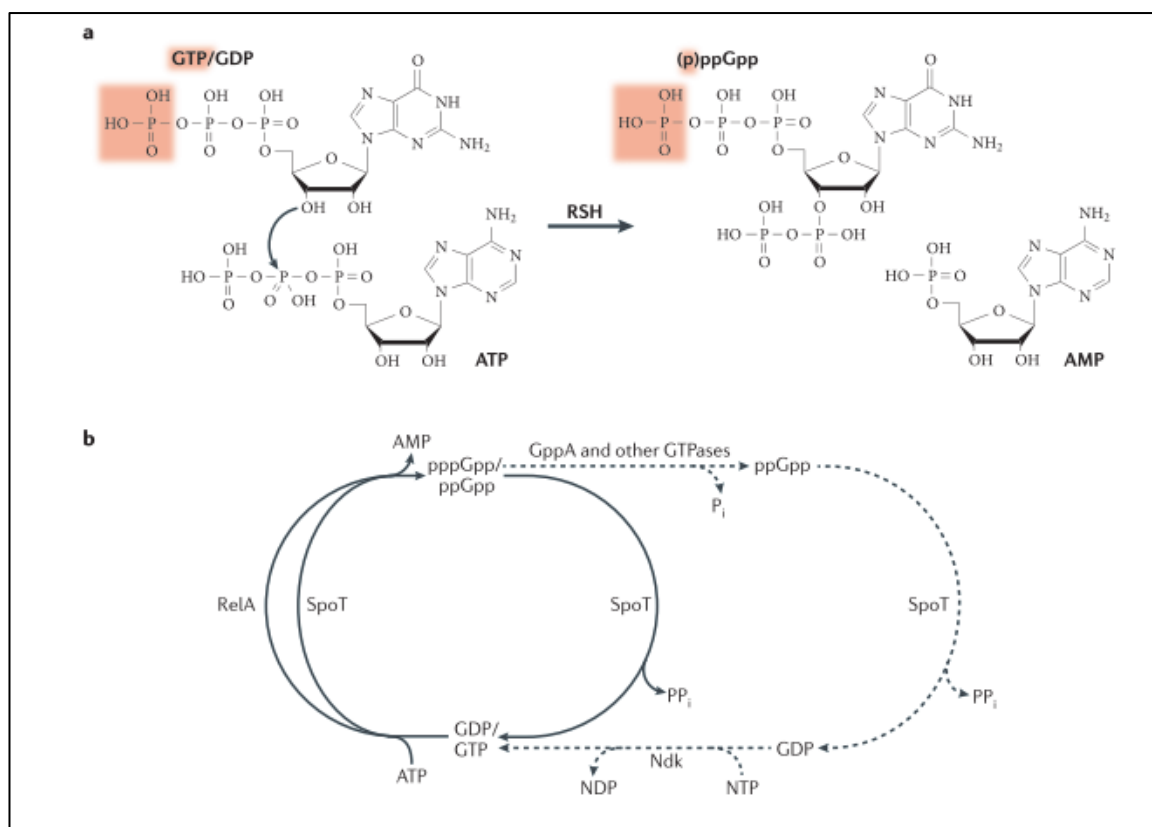


Figure 7.2 (p)ppGpp Metabolism in *E. coli*

(a) Guanosine diphosphate (GDP) or guanosine triphosphate (GTP) is converted into guanosine tetraphosphate (ppGpp) or guanosine pentaphosphate (pppGpp) respectively (γ -phosphate moiety of GTP and pppGpp highlighted in red). The addition of pyrophosphate from adenosine triphosphate (ATP) to the 3'OH of GDP/GTP is catalysed by RelA-SpoT homolog proteins (RSH) and adenosine monophosphate (AMP) is released as a by-product. (b) (p)ppGpp metabolism, as described in (a) ATP is used by RSH proteins RelA and SpoT to generate (p)ppGpp from GDP/GTP releasing AMP as a product. In addition, the conversion of pppGpp to ppGpp by (p)ppGpp phosphatase (GppA) and GTPases, releasing inorganic phosphate (P_i). SpoT catalyses the conversion of (p)ppGpp to GDP/GTP (solid line) and ppGpp to GDP (dotted line) releasing pyrophosphate (PP_i) in the process. GDP can be further converted to GTP by nucleoside diphosphate kinase (Ndk) and nucleoside triphosphates (NTP), resulting in nucleoside diphosphate (NDP) production. Image taken directly from (Haurlyuk *et al.*, 2015).

The list of metabolites which may affect the activity of PNPase homologs is clearly expanding and this study proposes that this may represent an ancient and evolutionarily conserved communicative link between central metabolism and RNA turnover. Although many conclusions can be made from this work and the potential physiological meaning of these findings can be discussed, more research is required before a comprehensive understanding of how metabolites potentially regulate

RNA turnover can be acquired. The use of *in silico* molecular docking calculations and a novel *in vitro* high-throughput plate reader assays, for predicting metabolite-PNPase interactions and quantifying the 3'-5' degradation activity of PNPase homologs respectively, were validated within this work. The following Section 7.2 discusses how combining these *in silico* and *in vitro* techniques may facilitate the discovery of novel metabolite-RNase interactions.

7.2 Future Perspectives

7.2.1 Discovering Novel Metabolite-RNase Interactions

A large range of metabolites and different PNPase homologs are known to exist across prokaryotic, eukaryotic and archaeal organisms. A more efficient method of predicting metabolites which may bind PNPase homologs and then quantifying the metabolites' effect on 3'-5' degradation activity is required to understand the regulatory interactions involved. It is possible that a similar method, of high-throughput *in silico* and *in vitro* screening of ligand-receptor interactions, that is routinely employed in drug discovery, could be applied for addressing this issue (reviewed in Hughes, Rees, Kalindjian, & Philpott, 2011). More specifically, by using the molecular docking program MOE a larger range of metabolites could be screened for binding to PNPase homologs *in silico* (Molecular Operating Environment, 2013). Additionally, by utilising the high-throughput fluorescence-based plate reader assay (BellBrook, n.d.), any potential interactions could be investigated further *in vitro*.

The method described above for discovering novel metabolite-PNPase interactions using a plate reader, provides a significant advantage to the currently available gel-based assays since detailed kinetic information can be acquired in real-time. Moreover, although the assay was developed for a 96-well format, this could be further expanded to a 384-well format, thus allowing a wide range of metabolite-PNPase homolog interactions to be studied. An ADP standard curve can be conducted for each assay, and rather than reporting RFU changes, the degradation of RNA by PNPase homologs could be quantified in terms of the concentration of ADP release. The plate reader assay has been validated for determining important kinetic information, the effects of metabolites on PNPase homologs could be tested by titrating a range of inhibitor concentrations and IC₅₀ values (i.e. the concentration of inhibitor which reduces the response of the enzyme by a half) could be determined. Alternatively, more detailed kinetic parameters of canonical enzyme activity including K_m , V_{max} k_{cat} and catalytic efficiency, could be compared to those obtained in the presence of an inhibitor.

Another great advantage of this particular real-time assay is the option to use a wide range of unlabelled and physiologically relevant RNA substrates. If the RNA substrate sequence is known to contain adenosines and ADP is released upon degradation, it should be suitable for use within this assay. This would allow the option for testing a range of physiologically relative PNPase or archaeal exosome RNA substrates.

The assay could also be used to screen protein mutants, possibly identifying amino acid mutations essential for activity. Alternatively, other 3'-5' RNases, such as RNase PH, could be examined. The effect of 3'-5' RNases acting independently or as part of a multi-protein complex may also be compared. For example, the activity of PNPase or PNPase as part of the degradosome in *E. coli* could be examined.

In summary, the methods used within this PhD study could be applied to investigate a wide range of metabolite-RNase interactions. This may help improve our understanding of how RNA turnover is regulated in evolutionarily distinct organisms.

7.2.2 Examining Metabolite-RNase Interactions *In Vivo*

Although this study describes a wide range of *in silico* and *in vitro* data which supports the proposed communicative link between metabolism and RNA turnover, this needs to be further validated *in vivo*.

In terms of investigating the interactions of key metabolites with PNPase in bacteria, a number of *E. coli* cell strains are available and have been successfully used to examine the effect of intracellular citrate concentrations on PNPase activity (Nurmohamed *et al.*, 2011). The effect of other metabolites on PNPase activity could be established by applying a similar approach. More specifically, in the presence of a metabolite, the growth of a (RNase II/R⁻) mutant strain, which is known to be dependent on PNPase for viability (Cheng & Deutscher, 2003; Donovan & Kushner, 1986), could be compared to a wild-type strain. Furthermore, using these strains, the effects of metabolites on the cellular transcript levels could then be studied using RT-PCR and/or microarray, as described in Nurmohamed *et al.*, 2011, or RNA-sequencing could be utilised. In the microarray analysis conducted by Nurmohamed and colleagues, a PNPase-mediated response to citrate was revealed by comparing a PNPase null mutant +/- citrate, against a WT strain +/- citrate. Genes generally affected by PNPase and those displaying a PNPase-mediated response to citrate were then compared. The results suggested that the PNPase-mediated citrate response broadly affected genes involved in cellular metabolic processes including cellular amino acid and derivative metabolic processes, and cellular biosynthetic and organic acid metabolic processes (Nurmohamed *et al.*, 2011). The effects of citrate on specific mRNA transcripts, including *cirA* and *fkpA*, known to be regulated by PNPase, was also investigated using RT-PCR. Collectively, results from RT-PCR and microarray analysis showed that not only were specific transcripts affected in a PNPase-dependent response to citrate, but RNA metabolism was also broadly altered *in vivo* (Nurmohamed *et al.*, 2011). Using similar experimental and analysis methods, the levels of transcripts affected in a PNPase-dependent response to other metabolites, could be investigated. More specifically, bacterial cell strains could be grown in the presence of metabolites such as acetyl-CoA, succinyl-CoA and (p)ppGpp, alternatively the intracellular levels of a particular metabolite could be modified by artificially blocking or activating the enzymes involved in their synthesis and break-

down. Transcriptomic and bacterial growth studies could then be applied to study the effect of individual metabolites on PNPase-mediated RNA regulation.

In the study conducted by Nurmohamed and colleagues, PNPase has already been shown to affect cellular metabolism by using ^1H NMR spectroscopy and gas chromatography mass spectrometry. Many metabolites were affected by PNPase loss, when comparing the null strain with the WT (Nurmohamed *et al.*, 2011). Whether these same metabolites are affected by metabolite-mediated inhibition of PNPase activity, could be examined using similar techniques. This may shed light on whether metabolite levels can affect PNPase-mediated RNA turnover *in vivo* in a feedback loop.

In terms of investigating the interactions of key metabolites with PNPase in higher organisms, a number of methods are available for extracting mitochondria from human cell lines and studying their function in isolation (Azimzadeh, Aghdaei, Tarban, & Akhondi, 2016; Lanza & Sreekumaran Nair, 2009). Whether TCA metabolites can effect PNPase activity and thus mitochondrial RNA levels needs to be determined and would be an interesting area to study. It is known that metabolic flux can be modified using blockers and as mentioned previously, RNA levels could be examined using RNA-sequencing. For example, a previous study added the blocker maleate to isolated rat heart mitochondria since it can convert CoA into a stable thioester which is metabolically inert (Pacanis, Strzelecki, & Rogulski, 1981). This rapidly reduced the CoA levels and a range of biological systems were found to be effected. For example, the oxidation of CoA-dependent substrates (i.e. α -ketoglutarate and pyruvate), oxygen uptake and oxidative phosphorylation were all inhibited (Pacanis *et al.*, 1981; Tahiliani & Beinlich, 1991). Potentially, a similar technique could be applied to investigate the effect of cellular metabolic fluxes on transcript levels in wild-type and PNPase mutant cell lines.

Whether metabolites effect PNPase activity *in vivo* also need to be explored in higher model organisms. Since the early development of *Xenopus* embryos can be studied, it would be interesting to examine whether PNPase has a role within these early stages. Preliminary experiments, using a cross-reacting hPNPase antibody, suggested that the PNPase protein was present in the early stages of *Xenopus laevis* development and was also expressed maternally in the fertilised egg (data not shown). In order to determine whether PNPase is essential for embryo development, knock out studies could be conducted and the physiological effect on growth could be monitored. The first step could involve using antisense morpholino oligonucleotides to determine whether the mRNA can be depleted in the oocyte (as reviewed in Hulstrand, Schneider, & Houston, 2011). If unsuccessful, more drastic measures involving genome editing in the parent could be conducted (as reviewed in Tandon, Conlon, Furlow, & Horb, 2016). During these studies, if PNPase is non-essential and embryos survive, then transcriptomics may be utilised to compare RNA levels in wild-type and PNPase KO strains, in the presence and absence of a metabolite; the levels of which could be manipulated by microinjection. This could identify a PNPase-dependent response to a

particular metabolite and reveal whether specific mRNA substrates are differentially regulated/stabilised *in vivo* as a result.

The *in vivo* research proposed above requires a significant amount of work and unfortunately was not possible during the duration of this PhD, but could be the focus of following work.

7.3 Summary

In summary, the data provided within this thesis are of importance to the field of post-transcriptional gene regulation research. A greater understanding of the intricate network of interactions occurring in cells is invaluable for developing novel medical and biotechnological applications. The *in silico* and *in vitro* work conducted requires validation *in vivo* and the effect of metabolites on other RNases and multi-protein RNA degrading machines remains to be discovered. Nevertheless, the study presented here supports the existence of an ancient and evolutionarily conserved interaction between key metabolites and the enzymes involved in regulating RNA turnover. Furthermore, development of a real-time, high-throughput assay could help determine, in more detail, the effect of potential PNPase inhibitors investigated within this thesis and previously suggested in the literature. As 3'-5' RNases could be studied more efficiently than gel-based assays using this method, this would be beneficial to the RNA field. The plate reader assay may also facilitate the screening of small molecules with metabolite-RNase interactions for potential medical and/or biotechnological applications.

8 References

- Abràmoff, M. D. M., Magalhães, P. J., & Ram, S. J. (2004). Image Processing with ImageJ. *Biophotonics International*, 1–7. Retrieved from <http://webeye.ophth.uiowa.edu/dept/biograph/abramoff/imagej.pdf>
- Adrio, J. L., & Demain, A. L. (2014). Microbial enzymes: tools for biotechnological processes. *Biomolecules*, 4(1), 117–39. <http://doi.org/10.3390/biom4010117>
- Agilent Technologies, I. (n.d.). “Quikchange Site-Directed Mutagenesis Kit Instruction Manual.” 1–14.
- Ait-Bara, S., & Carpousis, A. J. (2015). RNA degradosomes in bacteria and chloroplasts: Classification, distribution and evolution of RNase E homologs. *Molecular Microbiology*, 97(6), 1021–1035. <http://doi.org/10.1111/mmi.13095>
- Alodaib, A., Sobreira, N., Gold, W. A., Riley, L. G., Van Bergen, N. J., Wilson, M. J., *et al* Christodoulou, J. (2016). Whole-exome sequencing identifies novel variants in PNPT1 causing oxidative phosphorylation defects and severe multisystem disease. *European Journal of Human Genetics*, 1–6. <http://doi.org/10.1038/ejhg.2016.128>
- Andrade, J. M., & Arraiano, C. M. (2008). PNPase is a key player in the regulation of small RNAs that control the expression of outer membrane proteins. *RNA*, 14(3), 543–51. <http://doi.org/10.1261/rna.683308>
- Andrade, J. M., Pobre, V., Matos, A. M., & Arraiano, C. M. (2012). The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. *RNA*, 18(4), 844–55. <http://doi.org/10.1261/rna.029413.111>
- Arraiano, C. M., Andrade, J. M., Domingues, S., Guinote, I. B., Malecki, M., Matos, R. G., *et al* Viegas, S. C. (2010). The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiology Reviews*, 34(5), 883–923. <http://doi.org/10.1111/j.1574-6976.2010.00242.x>
- Awano, N., Inouye, M., & Phadtare, S. (2008). RNase Activity of Polynucleotide Phosphorylase Is Critical at Low Temperature in *Escherichia coli* and Is Complemented by RNase II. *Journal of Bacteriology*, 190(17), 5924–5933. <http://doi.org/10.1128/JB.00500-08>
- Azimzadeh, P., Aghdaei, H. A., Tarban, P., & Akhondi, M. M. (2016). Comparison of three methods for mitochondria isolation from the human liver cell line (HepG2). *Gastroenterology and Hepatology*, 9(6), 105–113.
- Bandyra, K. J., Bouvier, M., Carpousis, A. J., & Luisi, B. F. (2013). The social fabric of the RNA degradosome. *Biochimica et Biophysica Acta*, 1829, 514–522. <http://doi.org/10.1016/j.bbagr.2013.02.011>
- Belasco, J. G. (2010). All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nature Reviews. Molecular Cell Biology*, 11(7), 467–78. <http://doi.org/10.1038/nrm2917>
- BellBrook. (n.d.). Transcreener® ADP2 fluorescent intensity (FI) Assay. Retrieved from <https://www.bellbrooklabs.com/products-services/transcreener-hts-assays/kinase-assays/adp2-assays-fi/>
- Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., & Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology*, 5(8), 593–599. <http://doi.org/10.1038/nchembio.186>
- Berg, J., Tymoczko, J., & Stryer, L. (2002). *Biochemistry*. *W H Freeman* (5th ed.). New York.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242. <http://doi.org/10.1093/nar/28.1.235>
- Bernstein, J. A., Lin, P.-H., Cohen, S. N., & Lin-Chao, S. (2004). Global analysis of *Escherichia coli* RNA degradosome function using DNA microarrays. *Proceedings of the National Academy of Sciences of*

- the United States of America*, 101(9), 2758–63. <http://doi.org/10.1073/pnas.0308747101>
- Birnboim, H. C., & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Cultures*, 7(6), 1513–1523. Retrieved from <http://nar.oxfordjournals.org/content/7/6/1513.short>
- Borowski, L. S., Szczesny, R. J., Brzezniak, L. K., & Stepien, P. P. (2010). RNA turnover in human mitochondria: More questions than answers? *Biochimica et Biophysica Acta*, 1797(6–7), 1066–1070. <http://doi.org/10.1016/j.bbabi.2010.01.028>
- Bralley, P., & Jones, G. (2003). Overexpression of the polynucleotide phosphorylase gene (pnp) of *Streptomyces antibioticus* affects mRNA stability and poly(A) tail length but not ppGpp levels. *Microbiology*, 149(8), 2173–2182. <http://doi.org/10.1099/mic.0.26334-0>
- Briani, F., Del Favero, M., Capizzuto, R., Consomni, C., Zangrossi, S., Greco, C., ... Dehò, G. (2007). Genetic analysis of polynucleotide phosphorylase structure and functions. *Biochimie*, 89(1), 145–57. <http://doi.org/10.1016/j.biochi.2006.09.020>
- Broadway, N. (2012). Recombinant Protein Expression: Vector-Host Systems. *Materials and Methods*, 2(128). <http://doi.org//dx.doi.org/10.13070/mm.en.2.123>
- Büttner, K., Wenig, K., & Hopfner, K. P. (2005). Structural framework for the mechanism of archaeal exosomes in RNA processing. *Molecular Cell*, 20(3), 461–471. <http://doi.org/10.1016/j.molcel.2005.10.018>
- Callaghan, A. J., Aurikko, J. P., Ilag, L. L., Günter Grossmann, J., Chandran, V., Kühnel, K., ... Luisi, B. F. (2004). Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E. *Journal of Molecular Biology*, 340(5), 965–979. <http://doi.org/10.1016/j.jmb.2004.05.046>
- Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G., & Luisi, B. F. (2005). Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature*, 437(7062), 1187–1191. <http://doi.org/10.1038/nature04084>
- Cardoso, K., Gandra, R. F., Wisniewski, E. S., Osaku, C. A., Kadowaki, M. K., Felipach-Neto, V., ... Simão, R. D. C. G. (2010). DnaK and GroEL are induced in response to antibiotic and heat shock in *Acinetobacter baumannii*. *Journal of Medical Microbiology*, 59(9), 1061–1068. <http://doi.org/10.1099/jmm.0.020339-0>
- Carpousis, A. J., Vanhouwe, G., Ehretsmann, C., & Krisch, H. M. (1994). Copurification of *Escherichia coli* RNase E and PNPase: Evidence for a Specific Association between 2 Enzymes Important in RNA Processing and Degradation. *Cell*, 76(5), 889–900.
- Carpousis, A. J. (2007). The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annual Review of Microbiology*, 61, 71–87. <http://doi.org/10.1146/annurev.micro.61.080706.093440>
- Chang, S. A., Cozad, M., Mackie, G. A., & Jones, G. H. (2008). Kinetics of polynucleotide phosphorylase: Comparison of enzymes from *Streptomyces* and *Escherichia coli* and effects of nucleoside diphosphates. *Journal of Bacteriology*, 190(1), 98–106. <http://doi.org/10.1128/JB.00327-07>
- Chen, H.-W., Rainey, R. N., Balatoni, C. E., Dawson, D. W., Troke, J. J., Wasiak, S., ... French, S. W. (2006). Mammalian polynucleotide phosphorylase is an intermembrane space RNase that maintains mitochondrial homeostasis. *Molecular and Cellular Biology*, 26(22), 8475–87. <http://doi.org/10.1128/MCB.01002-06>
- Chen, H.-W., Koehler, C. M., & Teitell, M. A. (2007). Human polynucleotide phosphorylase: location matters. *Trends in Cell Biology*, 17(12), 600–8. <http://doi.org/10.1016/j.tcb.2007.09.006>
- Cheng, Z. F., & Deutscher, M. P. (2003). Quality control of ribosomal RNA mediated by polynucleotide

- phosphorylase and RNase R. *Proc Natl Acad Sci USA*, 100(11), 6388–6393. <http://doi.org/10.1073/pnas.1231041100>
- Chou, J. Y., & Singer, M. F. (1970a). A kinetic Analysis of the Phosphorolysis of Oligonucleotides by Polynucleotide Phosphorylase. *JBC*, 245(5), 995–1004.
- Chou, J. Y., & Singer, M. F. (1970b). The effect of chain length on the phosphorolysis of oligonucleotides by polynucleotide phosphorylase. *The Journal of Biological Chemistry*, 245(5), 1005–11. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4313699>
- Chou, J. Y., Singer, M. F., & McPhie, P. (1975). Kinetic studies on the phosphorolysis of polynucleotides by polynucleotide phosphorylase. *Journal of Biological Chemistry*, 250(2), 508–514. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1078670>
- Clements, M. O., Eriksson, S., Thompson, A., Lucchini, S., Hinton, J. C. D., Normark, S., & Rhen, M. (2002). Polynucleotide phosphorylase is a global regulator of virulence and persistency in *Salmonella enterica*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(13), 8784–9. <http://doi.org/10.1073/pnas.132047099>
- Clingman, C. C., Deveau, L. M., Hay, S. A., Genga, R. M., Shandilya, S. M., Massi, F., & Ryder, S. P. (2014). Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *eLife*, 3, 1–26. <http://doi.org/10.7554/eLife.02848>
- Coburn, G. A., Miao, X., Briant, D. J., & Mackie, G. A. (1999). Reconstitution of a minimal RNA degradosome demonstrates functional coordination between a 3' exonuclease and a DEAD-box RNA helicase. *Genes & Development*, 13(19), 2594–603. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10521403>
- Commichau, F. M., Rothe, F. M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., ... Stulke, J. (2009). Novel Activities of Glycolytic Enzymes in *Bacillus subtilis*. *Molecular & Cellular Proteomics*, 8(6), 1350–1360. <http://doi.org/10.1074/mcp.M800546-MCP200>
- Dalebroux, Z. D., Svensson, S. L., Gaynor, E. C., & Swanson, M. S. (2010). ppGpp Conjures Bacterial Virulence. *Microbiology and Molecular Biology Reviews*, 74(2), 171–199. <http://doi.org/10.1128/MMBR.00046-09>
- Del Favero, M., Mazzantini, E., Briani, F., Zangrossi, S., Tortora, P., & Dehò, G. (2008). Regulation of *Escherichia coli* polynucleotide phosphorylase by ATP. *The Journal of Biological Chemistry*, 283(41), 27355–9. <http://doi.org/10.1074/jbc.C800113200>
- Desjardins, P., Hansen, J. B., & Allen, M. (2009). Microvolume protein concentration determination using the NanoDrop 2000c spectrophotometer. *Journal of Visualized Experiments*, (33), 3–5. <http://doi.org/10.3791/1610>
- Deutscher, M. P. (1993). Promiscuous Exoribonucleases of *Escherichia coli*. *Journal of Bacteriology*, 175(15), 4577.
- Deutscher, M. P. (2015). How bacterial cells keep ribonucleases under control. *FEMS Microbiology Reviews*, 39(3), 350–361. <http://doi.org/10.1093/femsre/fuv012>
- Dharmacon.gelifesciences. (n.d.). 5'-Fluorescein. Retrieved from http://dharmacon.gelifesciences.com/uploadedFiles/Products/Custom_Synthesis/_modifications-details/unit-structure-5p-fluorescein.html
- Donovan, W. P., & Kushner, S. R. (1986). Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in *Escherichia coli* K-12. *Proceedings of the National Academy of Sciences USA*, 83(1), 120–124. <http://doi.org/10.1073/pnas.83.1.120>

- Dziembowski, A., Lorentzen, E., Conti, E., & Séraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature Structural & Molecular Biology*, *14*(1), 15–22. <http://doi.org/10.1038/nsmb1184>
- E. coli* genotypes. (2015, January 23). Retrieved from http://openwetware.org/index.php?title=E._coli_genotypes&oldid=853128.
- Erce, M. A., Low, J. K. K., March, P. E., Wilkins, M. R., & Takayama, K. M. (2009). Identification and functional analysis of RNase E of *Vibrio angustum* S14 and two-hybrid analysis of its interaction partners. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1794*(8), 1107–1114. <http://doi.org/10.1016/j.bbapap.2009.03.016>
- Evguenieva-Hackenberg, E., Walter, P., Hochleitner, E., Lottspeich, F., & Klug, G. (2003). An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Reports*, *4*(9), 889–893. <http://doi.org/10.1038/sj.embor.embor929>
- Evguenieva-Hackenberg, E., Hou, L., Glaeser, S., & Klug, G. (2014). Structure and function of the archaeal exosome. *Wiley Interdisciplinary Reviews: RNA*, *5*, n/a-n/a. <http://doi.org/10.1002/wrna.1234>
- Franklin, M. C., Cheung, J., Rudolph, M. J., Burshteyn, F., Cassidy, M., Gary, E., ... Love, J. D. (2015). Structural genomics for drug design against the pathogen *Coxiella burnetii*. *Proteins: Structure, Function, and Bioinformatics*, *83*(12), 2124–2136. <http://doi.org/10.1002/prot.24841>
- Galli, C. L., Sensi, C., Fumagalli, A., Parravicini, C., Marinovich, M., & Eberini, I. (2014). A computational approach to evaluate the androgenic affinity of iprodione, procymidone, vinclozolin and their metabolites. *PLoS ONE*, *9*(8). <http://doi.org/10.1371/journal.pone.0104822>
- Gatewood, M. L., & Jones, G. H. (2010). (p)ppGpp inhibits polynucleotide phosphorylase from *streptomyces* but not from *Escherichia coli* and increases the stability of bulk mRNA in *Streptomyces coelicolor*. *Journal of Bacteriology*, *192*(17), 4275–80. <http://doi.org/10.1128/JB.00367-10>
- Geospiza. (n.d.). FinchTV. Retrieved from <http://www.geospiza.com/>
- GIMP. (n.d.). GNU Image Manipulation Program (GIMP). Retrieved from <https://www.gimp.org/>
- Godefroy, T., Cohn, M., & Grunberg-Manago, M. (1970). Kinetics of Polymerization and Phosphorolysis Reactions of *E. coli* Polynucleotide Phosphorylase. Role of Oligonucleotides in Polymerization. *European Journal of Biochemistry*, *12*(2), 236–249. <http://doi.org/10.1111/j.1432-1033.1970.tb00843.x>
- Godefroy-Colburn, T., & Grunberg-Manago, M. (1975). *Polynucleotide Phosphorylase*. *The Enzymes* (Vol. 7). Elsevier.
- Gräslund, S., Nordlund, P., Weigelt, J., Bray, J., Gileadi, O., Knapp, S., ... Gunsalus, K. C. (2008). Protein production and purification. *Nature Methods*, *5*(2), 135–146. <http://doi.org/10.1038/nmeth.f.202>
- Hardwick, S. W., Chan, V. S. Y., Broadhurst, R. W., & Luisi, B. F. (2011). An RNA degradosome assembly in *Caulobacter crescentus*. *Nucleic Acids Research*, *39*(4), 1449–1459. <http://doi.org/10.1093/nar/gkq928>
- Hardwick, S. W., Gubbey, T., Hug, I., Jenal, U., & Luisi, B. F. (2012). Crystal structure of *Caulobacter crescentus* polynucleotide phosphorylase reveals a mechanism of RNA substrate channelling and RNA degradosome assembly. *Open Biology*, *2*(4), 120028–120028. <http://doi.org/10.1098/rsob.120028>
- Hartung, S., Niederberger, T., Hartung, M., Tresch, A., & Hopfner, K. P. (2010). Quantitative analysis of processive RNA degradation by the archaeal RNA exosome. *Nucleic Acids Research*, *38*(15), 5166–5176. <http://doi.org/10.1093/nar/gkq238>
- Haurlyuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T., & Gerdes, K. (2015). Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nature Reviews Microbiology*, *13*(5), 298–

309. <http://doi.org/10.1038/nrmicro3448>

- Hayes, R., Kudla, J. O., & Grussem, W. (1999). Degrading chloroplast mRNA: The role of polyadenylation. *Trends in Biochemical Sciences*, 24(5), 199–202. [http://doi.org/10.1016/S0968-0004\(99\)01388-2](http://doi.org/10.1016/S0968-0004(99)01388-2)
- Higgins, D. G., & Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, 73(1), 237–44. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3243435>
- Hirosawa, M., Totoki, Y., Hoshida, M., & Ishikawa, M. (1995). Comprehensive study on iterative algorithms of multiple sequence alignment. *Computer Applications in the Biosciences: CABIOS*, 11(1), 13–8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7796270>
- Hong, L., Quinn, C. M., & Jia, Y. (2009). Evaluating the utility of the HTRF® Transcreeper™ ADP assay technology: A comparison with the standard HTRF assay technology. *Analytical Biochemistry*, 391(1), 31–38. <http://doi.org/10.1016/j.ab.2009.04.033>
- Hughes, J. P., Rees, S. S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, 162(6), 1239–1249. <http://doi.org/10.1111/j.1476-5381.2010.01127.x>
- Hulstrand, A. M., Schneider, P. N., & Houston, D. W. (2010). The use of antisense oligonucleotides in *Xenopus* oocytes. *Methods*, 51(1), 75–81. <http://doi.org/10.1016/j.ymeth.2009.12.015>
- Hutchison, C. A., Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., & Smith, M. (1978). Mutagenesis at a specific position in a DNA sequence. *Journal of Biological Chemistry*, 253(18), 6551–6560.
- Huynen, M. A., Dandekar, T., & Bork, P. (1999). Variation and evolution of the citric acid cycle: a genomic perspective. *Trends in Microbiology*, 7(7), 281–291.
- Ibrahim, H., Wilusz, J., & Wilusz, C. J. (2008). RNA recognition by 3'-to-5' exonucleases: The substrate perspective. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1779(4), 256–265. <http://doi.org/10.1016/j.bbagr.2007.11.004>
- Icard, P., Poulain, L., & Lincet, H. (2012). Understanding the central role of citrate in the metabolism of cancer cells. *Biochimica et Biophysica Acta*, 1825(1), 111–6. <http://doi.org/10.1016/j.bbcan.2011.10.007>
- Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*, 52(7), 1757–1768. <http://doi.org/10.1021/ci3001277>
- Janga, S. C., & Babu, M. M. (2009). Transcript stability in the protein interaction network of *Escherichia coli*. *Molecular bioSystems*, 5(2), 154–62. <http://doi.org/10.1039/b816845h>
- Januszyk, K., & Lima, C. D. (2014). The eukaryotic RNA exosome. *Current Opinion in Structural Biology*, 24, 132–40. <http://doi.org/10.1016/j.sbi.2014.01.011>
- Jarrige, A., Bréchemier-Baey, D., Mathy, N., Duché, O., & Portier, C. (2002). Mutational analysis of polynucleotide phosphorylase from *Escherichia coli*. *Journal of Molecular Biology*, 321(3), 397–409. [http://doi.org/10.1061/S0022-2836\(02\)00645-9](http://doi.org/10.1061/S0022-2836(02)00645-9)
- Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology*, 6(5), 494–500. [http://doi.org/10.1016/0958-1669\(95\)80082-4](http://doi.org/10.1016/0958-1669(95)80082-4)
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <http://doi.org/10.1093/molbev/mst010>

- Kime, L., Vincent, H. A., Gendoo, D. M. A., Jourdan, S. S., Fishwick, C. W. G., Callaghan, A. J., & McDowall, K. J. (2015). The First Small-Molecule Inhibitors of Members of the Ribonuclease E Family. *Scientific Reports*, 5, 8028. <http://doi.org/10.1038/srep08028>
- Kornberg, H. (2000). Krebs and his trinity of cycles. *Nature Reviews. Molecular Cell Biology*, 1(3), 225–8. <http://doi.org/10.1038/35043073>
- Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259), 680–685. <http://doi.org/10.1038/227680a0>
- Langer, T., & Wolber, G. (2004). Pharmacophore definition and 3D searches. *Drug Discovery Today: Technologies*, 1(3), 203–207. <http://doi.org/10.1016/j.ddtec.2004.11.015>
- Lanza, I. R., & Sreekumaran Nair, K. (2009). Functional Assessment of Isolated Mitochondria In Vitro. *Methods Enzymol*, 6879(9), 1–19. [http://doi.org/10.1016/S0076-6879\(09\)05020-4](http://doi.org/10.1016/S0076-6879(09)05020-4). Functional
- Leatherbarrow, R. J. (2009). *GraFit Version 7*. Horley, U.K.: Erithacus Software Ltd.
- Lee, G., Hartung, S., Hopfner, K. P., & Ha, T. (2010). Reversible and controllable nanolocomotion of an RNA-processing machinery. *Nano Letters*, 10(12), 5123–5130. <http://doi.org/10.1021/nl103754z>
- Lehnik-Habrink, M., Pförtner, H., Rempeters, L., Pietack, N., Herzberg, C., & Stülke, J. (2010). The RNA degradosome in *Bacillus subtilis*: identification of CshA as the major RNA helicase in the multiprotein complex. *Molecular Microbiology*, 77(4), 958–971. <http://doi.org/10.1111/j.1365-2958.2010.07264.x>
- Lehninger, A. L., Nelson, D. L., & Cox, M. M. (2000). *Lehninger principles of biochemistry*. New York: Worth Publishers.
- Leszczyniecka, M., Kang, D.-C., Sarkar, D., Su, Z.-Z., Holmes, M., Valerie, K., & Fisher, P. B. (2002). Identification and cloning of human polynucleotide phosphorylase, hPNPase old-35, in the context of terminal differentiation and cellular senescence. *Proceedings of the National Academy of Sciences USA*, 99(26), 16636–41. <http://doi.org/10.1073/pnas.252643699>
- Leszczyniecka, M., DeSalle, R., Kang, D.-C., & Fisher, P. B. (2004). The origin of polynucleotide phosphorylase domains. *Molecular Phylogenetics and Evolution*, 31(1), 123–30. <http://doi.org/10.1016/j.ympev.2003.07.012>
- Li, S., & Breaker, R. R. (2013). Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Research*, 41(5), 3022–3031. <http://doi.org/10.1093/nar/gkt057>
- Li, Z., Reimers, S., Pandit, S., & Deutscher, M. P. (2002). RNA quality control: degradation of defective transfer RNA. *The EMBO Journal*, 21(5), 1132–8. <http://doi.org/10.1093/emboj/21.5.1132>
- Lin, C. L., Wang, Yi-Ting, Yang, W.-Z., Hsiao, Y.-Y., & Yuan, H. S. (2012). Crystal structure of human polynucleotide phosphorylase: insights into its domain function in RNA binding and degradation. *Nucleic Acids Research*, 40(9), 4146–57. <http://doi.org/10.1093/nar/gkr1281>
- Liou, G. G., Jane, W. N., Cohen, S. N., Lin, N. S., & Lin-Chao, S. (2001). RNA degradosomes exist in vivo in *Escherichia coli* as multicomponent complexes associated with the cytoplasmic membrane via the N-terminal region of ribonuclease E. *Proceedings of the National Academy of Sciences USA*, 98(1), 63–8. <http://doi.org/10.1073/pnas.011535498>
- Littauer, U. Z., & Soreq, H. (1982). *Polynucleotide Phosphorylase. The Enzymes* (Vol. 15). New York: Boyer, P.D Academic Press Inc.
- Liu, Q., Greimann, J. C., & Lima, C. D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell*, 127(6), 1223–37. <http://doi.org/10.1016/j.cell.2006.10.037>
- Lorentzen, E., & Conti, E. (2005). Structural Basis of 3' End RNA Recognition and Exoribonucleolytic Cleavage by an Exosome RNase PH Core. *Molecular Cell*, 20(3), 473–481.

<http://doi.org/10.1016/j.molcel.2005.10.020>

- Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G., & Conti, E. (2005). The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nature Structural & Molecular Biology*, *12*(7), 575–581. <http://doi.org/10.1038/nsmb952>
- Lorentzen, E., Dziembowski, A., Lindner, D., Seraphin, B., & Conti, E. (2007). RNA channelling by the archaeal exosome. *EMBO Reports*, *8*(5), 470–476. <http://doi.org/10.1038/sj.embor.7400945>
- Lorentzen, E., & Conti, E. (2012). Crystal Structure of a 9-Subunit Archaeal Exosome in Pre-Catalytic States of the Phosphorolytic Reaction. *Archaea*, *2012*, 1–7. <http://doi.org/10.1155/2012/721869>
- Lu, C., Ding, F., & Ke, A. (2010). Crystal structure of the *S. solfataricus* archaeal exosome reveals conformational flexibility in the RNA-binding ring. *PloS One*, *5*(1), e8739. <http://doi.org/10.1371/journal.pone.0008739>
- Lu, Y., Zhang, X., Zhang, H., Lan, J., Huang, G., Varin, E., ... Icard, P. (2011). Citrate induces apoptotic cell death: a promising way to treat gastric carcinoma? *Anticancer Research*, *31*(3), 797–805. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21498699>
- Lykke-Andersen, S., Brodersen, D. E., & Jensen, T. H. (2009). Origins and activities of the eukaryotic exosome. *Journal of Cell Science*, *122*(10), 1487–1494. <http://doi.org/10.1242/jcs.047399>
- Marcaida, M. J., DePristo, M. A., Chandran, V., Carpousis, A. J., & Luisi, B. F. (2006). The RNA degradosome: life in the fast lane of adaptive molecular evolution. *Trends in Biochemical Sciences*, *31*(7), 359–365. <http://doi.org/10.1016/j.tibs.2006.05.005>
- Matos, R. G., B arrria, C., Pobre, V., Andrade, J. M., & Arraiano, C. M. (2012). Exoribonucleases as modulators of virulence in pathogenic bacteria. *Frontiers in Cellular and Infection Microbiology*, *2*(65), 1–3. <http://doi.org/10.3389/fcimb.2012.00065>
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., ... Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, *41*(W1), W597–W600. <http://doi.org/10.1093/nar/gkt376>
- Mechold, U., Cashel, M., Steiner, K., Gentry, D., & Malke, H. (1996). Functional analysis of a *relA/spoT* gene homolog from *Streptococcus equisimilis*. *Journal of Bacteriology*, *178*(5), 1401–1411.
- Miczak, A., Kaberdin, V. R., Wei, C. L., & Lin-Chao, S. (1996). Proteins associated with RNase E in a multicomponent ribonucleolytic complex. *Proceedings of the National Academy of Sciences USA*, *93*(9), 3865–9. <http://doi.org/10.1073/pnas.93.9.3865>
- Mierendorf, R. C., Morris, B. B., Hammer, B., & Novy, R. E. (1998). *Expression and Purification of Recombinant Proteins Using the pET System. Molecular Diagnosis of Infectious Diseases* (Vol. 13). New Jersey: Humana Press. <http://doi.org/10.1385/0-89603-485-2:257>
- Minitab. (2010). Getting Started with Minitab 17. State College, PA: Minitab, Inc. Retrieved from www.minitab.com
- Mohanty, B. K., & Kushner, S. R. (2000a). Polynucleotide phosphorylase, RNase II and RNase E play different roles in the in vivo modulation of polyadenylation in *Escherichia coli*. *Molecular Microbiology*, *36*(4), 982–94. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10844684>
- Mohanty, B. K., & Kushner, S. R. (2000b). Polynucleotide phosphorylase functions both as a 3' right-arrow 5' exonuclease and a poly(A) polymerase in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *97*(22), 11966–11971. <http://doi.org/10.1073/pnas.220295997>
- Mohanty, B. K., & Kushner, S. R. (2003). Genomic analysis in *Escherichia coli* demonstrates differential roles for polynucleotide phosphorylase and RNase II in mRNA abundance and decay. *Molecular*

- Microbiology*, 50(2), 645–658. <http://doi.org/10.1046/j.1365-2958.2003.03724.x>
- Molecular Operating Environment. (2013). Molecular Operating Environment (MOE). 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2016.: Chemical Computing Group Inc.,
- Navarro, M. V. A. S., Oliveira, C. C., Zanchin, N. I. T., & Guimarães, B. G. (2008). Insights into the mechanism of progressive RNA degradation by the archaeal exosome. *Journal of Biological Chemistry*, 283(20), 14120–14131. <http://doi.org/10.1074/jbc.M801005200>
- Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Jelloul, M. Ben, ... Perrakis, A. (2005). Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: The PACT/JCSG+ strategy. *Acta Crystallographica Section D: Biological Crystallography*, 61(10), 1426–1431. <http://doi.org/10.1107/S0907444905024984>
- Ng, C. L., Waterman, D. G., Antson, A. A., & Ortiz-Lombardia, M. (2010). Structure of the *Methanothermobacter thermautotrophicus* exosome RNase PH ring. *Acta Crystallographica Section D: Biological Crystallography*, 66(5), 522–528. <http://doi.org/10.1107/S0907444910002908>
- Nielsen, J. (2003). It Is All about Metabolic Fluxes. *Journal of Bacteriology*, 185(24), 7031–7035. <http://doi.org/10.1128/JB.185.24.7031>
- Nurmohamed, S., Vaidialingam, B., Callaghan, A. J., & Luisi, B. F. (2009). Crystal structure of *Escherichia coli* polynucleotide phosphorylase core bound to RNase E, RNA and manganese: implications for catalytic mechanism and RNA degradosome assembly. *Journal of Molecular Biology*, 389(1–3), 17–33.
- Nurmohamed, S., Vincent, H. A., Titman, C. M., Chandran, V., Pears, M. R., Du, D., ... Luisi, B. F. (2011). Polynucleotide phosphorylase activity may be modulated by metabolites in *Escherichia coli*. *The Journal of Biological Chemistry*, 286(16), 14315–14323. <http://doi.org/10.1074/jbc.M110.200741>
- O'Hara, E. B., Chekanova, J. A., Ingle, C. A., Kushner, Z. R., Peters, E., & Kushner, S. R. (1995). Polyadenylation helps regulate mRNA decay in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA*, 92(6), 1807–11. <http://doi.org/10.1073/pnas.92.6.1807>
- Pacanis, A., Strzelecki, T., & Rogulski, J. (1981). Maleate on the Content of CoA and Its Derivatives in Rat Kidney Mitochondria*. *JBC*, 256(2), 13035–13038.
- Piowarski, J., Grzechnik, P., Dziembowski, A., Dmochowska, A., Minczuk, M., & Stepień, P. P. (2003). Human Polynucleotide Phosphorylase, hPNPase, is Localized in Mitochondria. *Journal of Molecular Biology*, 329(5), 853–857. [http://doi.org/10.1016/S0022-2836\(03\)00528-X](http://doi.org/10.1016/S0022-2836(03)00528-X)
- Plotly Technologies Inc. (2015). Plotly. Montréal, QC,: Collaborative data science. Retrieved from <https://plot.ly>.
- Portnoy, V., Evguenieva-Hackenberg, E., Klein, F., Walter, P., Lorentzen, E., Klug, G., & Schuster, G. (2005). RNA polyadenylation in Archaea: not observed in *Haloferax* while the exosome polynucleotidylates RNA in *Sulfolobus*. *EMBO Reports*, 6(12), 1188–93. <http://doi.org/10.1038/sj.embor.7400571>
- Portnoy, V., Palnizky, G., Yehudai-Resheff, S., Glaser, F., & Schuster, G. (2007). Analysis of the human polynucleotide phosphorylase (PNPase) reveals differences in RNA binding and response to phosphate compared to its bacterial and chloroplast counterparts. *RNA*, 14(2), 297–309. <http://doi.org/10.1261/rna.698108>
- Prud'homme-Généreux, A., Beran, R. K., Iost, I., Ramey, C. S., Mackie, G. A., & Simons, R. W. (2004). Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: Evidence for a “cold shock degradosome.” *Molecular Microbiology*, 54(5), 1409–1421.

<http://doi.org/10.1111/j.1365-2958.2004.04360.x>

- Pruitt, K. D., Tatusova, T., Klimke, W., & Maglott, D. R. (2009). NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Research*, 37(Database issue), D32-6. <http://doi.org/10.1093/nar/gkn721>
- Py, B., Causton, H., Mudd, E. A., & Higgins, C. F. (1994). A protein complex mediating mRNA degradation. *Mol Microbiol*, 14(4), 717-729.
- Py, B., Higgins, C. F., Krisch, H. M., & Carpousis, A. J. (1996). A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature*, 381(6578), 169-172. <http://doi.org/10.1038/381169a0>
- Raijmakers, R., Schilders, G., & Pruijn, G. J. M. (2004). The exosome, a molecular machine for controlled RNA degradation in both nucleus and cytoplasm. *European Journal of Cell Biology*, 83(5), 175-83. <http://doi.org/10.1078/0171-9335-00385>
- Rauhut, R., & Klug, G. (1999). mRNA degradation in bacteria. *FEMS Microbiology Reviews*, 23(3), 353-370. [http://doi.org/10.1016/S0168-6445\(99\)00012-1](http://doi.org/10.1016/S0168-6445(99)00012-1)
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA*, 74(12), 5463-7. <http://doi.org/10.1073/pnas.74.12.5463>
- Sarkar, D., Lebedeva, I. V., Emdad, L., Kang, D.-C., Baldwin, A. S., & Fisher, P. B. (2004). Human polynucleotide phosphorylase (hPNPaseold-35): a potential link between aging and inflammation. *Cancer Research*, 64(20), 7473-8. <http://doi.org/10.1158/0008-5472.CAN-04-1772>
- Sarkar, D., & Fisher, P. B. (2006). Polynucleotide phosphorylase: an evolutionary conserved gene with an expanding repertoire of functions. *Pharmacology & Therapeutics*, 112(1), 243-63. <http://doi.org/10.1016/j.pharmthera.2006.04.003>
- Sarkar, N. (1997). Polyadenylation of mRNA in prokaryotes. *Annual Review of Biochemistry*, 66, 173-97. <http://doi.org/10.1146/annurev.biochem.66.1.173>
- Sato, T., Honma, T., & Yokoyama, S. (2010). Combining machine learning and pharmacophore-based interaction fingerprint for in silico screening. *Journal of Chemical Information and Modeling*, 50(1), 170-185. <http://doi.org/10.1021/ci900382e>
- Schmid, M., & Jensen, T. H. (2008). The exosome: a multipurpose RNA-decay machine. *Trends in Biochemical Sciences*, 33(10), 501-510. <http://doi.org/10.1016/j.tibs.2008.07.003>
- Schneider, T. D., & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, 18(20), 6097-6100. <http://doi.org/10.1093/nar/18.20.6097>
- Scholz, C., Knorr, S., Hamacher, K., & Schmidt, B. (2015). DOCKTITE-A highly versatile step-by-step workflow for covalent docking and virtual screening in the molecular operating environment. *Journal of Chemical Information and Modeling*, 55(2), 398-406. <http://doi.org/10.1021/ci500681r>
- Schuster, G., & Stern, D. (2009). RNA polyadenylation and decay in mitochondria and chloroplasts. *Progress in Molecular Biology and Translational Science*, 85, 393-422. [http://doi.org/10.1016/S0079-6603\(08\)00810-6](http://doi.org/10.1016/S0079-6603(08)00810-6)
- Selinger, D. W. (2003). Global RNA Half-Life Analysis in *Escherichia coli* Reveals Positional Patterns of Transcript Degradation. *Genome Research*, 13(2), 216-223. <http://doi.org/10.1101/gr.912603>
- She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., ... Van der Oost, J. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proceedings of the National Academy of Sciences USA*, 98(14), 7835-40. <http://doi.org/10.1073/pnas.141222098>
- Shi, Z., Yang, W.-Z., Lin-Chao, S., Chak, K.-F., & Yuan, H. S. (2008). Crystal structure of *Escherichia coli*

- PNPase: Central channel residues are involved in processive RNA degradation. *RNA*, *14*(11), 2361–2371. <http://doi.org/10.1261/rna.1244308>
- Shoemaker, C. J., & Green, R. (2012). Translation drives mRNA quality control. *Nature Structural & Molecular Biology*, *19*(6), 594–601. <http://doi.org/10.1038/nsmb.2301>
- Siculella, L., Damiano, F., Di Summa, R., Tredici, S. M., Alduina, R., Gnoni, G. V., & Alifano, P. (2010). Guanosine 5'-diphosphate 3'-diphosphate (ppGpp) as a negative modulator of polynucleotide phosphorylase activity in a “rare” *actinomycete*. *Molecular Microbiology*, *77*(3), 716–729. <http://doi.org/10.1111/j.1365-2958.2010.07240.x>
- Singh, J., Deng, Z., Narale, G., & Chuaqui, C. (2006). Structural interaction fingerprints: A new approach to organizing, mining, analyzing, and designing protein-small molecule complexes. *Chemical Biology and Drug Design*, *67*(1), 5–12. <http://doi.org/10.1111/j.1747-0285.2005.00323.x>
- Slomovic, S., Portnoy, V., Yehudai-Resheff, S., Bronshtein, E., & Schuster, G. (2008). Polynucleotide phosphorylase and the archaeal exosome as poly(A)-polymerases. *Biochimica et Biophysica Acta*, *1779*(4), 247–55. <http://doi.org/10.1016/j.bbagr.2007.12.004>
- Soreq, H., Littauer, U. (1977). Purification and Characterization of Polynucleotide Phosphorylase from *Escherichia coli*. *JBC*, *252*(19), 6885–6886.
- Source BioScience. (n.d.). Retrieved from <http://www.lifesciences.sourcebioscience.com/genomic-services/sanger-sequencing-service/>
- Steege, D. A. (2000). Emerging features of mRNA decay in bacteria. *RNA*, *6*(8), 1079–90. <http://doi.org/10.1017/S1355838200001023>
- Steglich, C., Lindell, D., Futschik, M., Rector, T., Steen, R., & Chisholm, S. W. (2010). Short RNA half-lives in the slow-growing marine cyanobacterium *Prochlorococcus*. *Genome Biology*, *11*(5), R54. <http://doi.org/10.1186/gb-2010-11-5-r54>
- Steinhauser, D., Fernie, A. R., & Araújo, W. L. (2012). Unusual cyanobacterial TCA cycles: Not broken just different. *Trends in Plant Science*, *17*(9), 503–509. <http://doi.org/10.1016/j.tplants.2012.05.005>
- Sterling, T., & Irwin, J. J. (2015). ZINC 15 - Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, *55*(11), 2324–2337. <http://doi.org/10.1021/acs.jcim.5b00559>
- Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, *189*(1), May 5;189(1):113-30.
- Sun, D., Lee, G., Lee, J. H., Kim, H.-Y., Rhee, H.-W., Park, S.-Y., ... Chung, J. (2010). A metazoan ortholog of SpoT hydrolyzes ppGpp and functions in starvation responses. *Nature Structural & Molecular Biology*, *17*(10), 1188–1194. <http://doi.org/10.1038/nsmb.1906>
- Sun, J., Aluvila, S., Kotaria, R., Mayor, J. A., Walters, D. E., & Kaplan, R. S. (2010). Mitochondrial and plasma membrane citrate transporters: Discovery of selective inhibitors and application to Structure/Function analysis. *Mol Cell Pharmacol*, *2*(3), 101–110. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2913483/>
- Swords, W. E. (2003). *Chemical transformation of E. coli. Methods in molecular biology* (Vol. 235). New Jersey: Humana Press. <http://doi.org/10.1385/1-59259-409-3:49>
- Symmons, M. F., Jones, G. H., & Luisi, B. F. (2000). A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure*, *8*(11), 1215–1226. [http://doi.org/10.1016/S0969-2126\(00\)00521-9](http://doi.org/10.1016/S0969-2126(00)00521-9)
- Syngene. (n.d.). GeneSnap. Retrieved from <http://www.snapgene.com/>
- Tahiliani, A. G., & Beinlich, C. J. (1991). Pantothenic acid in health and disease., *46*, 165–228.

- Tan, S. C., & Yiap, B. C. (2009). DNA, RNA, and Protein Extraction: The Past and The Present. *Journal of Biomedicine and Biotechnology*, 2009, 1–10. <http://doi.org/10.1155/2009/574398>
- Tandon, P., Conlon, F., Furlow, J. D., & Horb, M. E. (2016). Expanding the genetic toolkit in *Xenopus*: Approaches and opportunities for human disease modeling. *Developmental Biology*, 1–11. <http://doi.org/10.1016/j.ydbio.2016.04.009>
- Thammana, P., Buerk, R. R., & Gordon, J. (1976). Absence of ppGpp production in synchronised balb/c mouse 3T3 cells on isoleucine starvation. *FEBS Letters*, 68(2), 187–190. [http://doi.org/10.1016/0014-5793\(76\)80433-4](http://doi.org/10.1016/0014-5793(76)80433-4)
- Thermo Scientific. (2009). “NanoDrop 2000 / 2000c Spectrophotometer User Manual.” 1–97.
- Tuckerman, J. R., Gonzalez, G., & Gilles-Gonzalez, M.-A. (2011). Cyclic di-GMP Activation of Polynucleotide Phosphorylase Signal-Dependent RNA Processing. *Journal of Molecular Biology*, 407(5), 633–639. <http://doi.org/10.1016/j.jmb.2011.02.019>
- Ulas, T., Riemer, S. A., Zaparty, M., Siebers, B., & Schomburg, D. (2012). Genome-Scale Reconstruction and Analysis of the Metabolic Network in the Hyperthermophilic Archaeon *Sulfolobus Solfataricus*. *PLoS ONE*, 7(8). <http://doi.org/10.1371/journal.pone.0043401>
- Usenik, A., & Legis, M. (2010). Evolution of Allosteric Citrate Binding Sites on 6-phosphofructo-1-kinase. *PloS One*, 5(11), 1–8. <http://doi.org/10.1371/journal.pone.0015447>
- Vanzo, N. F., Li, Y. S., Py, B., Blum, E., Higgins, C. F., Raynal, L. C., ... Carpousis, A. J. (1998). Ribonuclease E organizes the protein interactions in the *Escherichia coli* RNA degradosome. *Genes and Development*, 12(17), 2770–2781. <http://doi.org/10.1101/gad.12.17.2770>
- Wang, G., Chen, H.-W., Oktay, Y., Zhang, J., Allen, E. L., Smith, G. M., ... Teitell, M. A. (2010). PNPase Regulates RNA Import into Mitochondria. *Cell*, 142(3), 456–467. <http://doi.org/10.1016/j.cell.2010.06.035>
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9), 1157–1174. <http://doi.org/10.1002/jcc.20035>
- Wasmuth, E. V., Januszyk, K., & Lima, C. D. (2014). Structure of an Rrp6–RNA exosome complex bound to poly(A) RNA. *Nature*, 511(7510), 435–439. <http://doi.org/10.1038/nature13406>
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. <http://doi.org/10.1093/bioinformatics/btp033>
- Wegner, A., Meiser, J., Weindl, D., & Hiller, K. (2015). How metabolites modulate metabolic flux. *Current Opinion in Biotechnology*, 34, 16–22. <http://doi.org/10.1016/j.copbio.2014.11.008>
- Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology*, 112, 531–52. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10027275>
- Wilusz, C. J., Wormington, M., & Peltz, S. W. (2001). The cap-to-tail guide to mRNA turnover. *Nature Reviews. Molecular Cell Biology*, 2(4), 237–246. <http://doi.org/10.1038/35067025>
- Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., ... Scalbert, A. (2013). HMDB 3.0-- The Human Metabolome Database in 2013. *Nucleic Acids Research*, 41(Database issue), D801-7. <http://doi.org/10.1093/nar/gks1065>
- Witharana, C., Roppelt, V., Lochnit, G., Klug, G., & Evguenieva-Hackenberg, E. (2012). Heterogeneous complexes of the RNA exosome in *Sulfolobus solfataricus*. *Biochimie*, 94(7), 1578–87.

<http://doi.org/10.1016/j.biochi.2012.03.026>

Xiong, J. (2006). Introduction and Biological Databases. In *Essential Bioinformatics* (pp. 3–11). Cambridge University Press. Retrieved from

<http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521840989&ss=exc>

Yang, P.-C., & Mahmood, T. (2012). Western blot: Technique, theory, and trouble shooting. *North American Journal of Medical Sciences*, 4(9), 429. <http://doi.org/10.4103/1947-2714.100998>

Yehudai-resheff, S., Hirsh, M., & Schuster, G. (2001). Polynucleotide Phosphorylase Functions as Both an Exonuclease and a Poly(A) Polymerase in Spinach Chloroplasts. *Molecular and Cellular Biology*, 21(16), 5408–5416. <http://doi.org/10.1128/MCB.21.16.5408>

Zhang, S., & Bryant, D. A. (2011). The tricarboxylic acid cycle in cyanobacteria. *Science*, 334(6062), 1551–3. <http://doi.org/10.1126/science.1210858>

Zuo, Y., & Deutscher, M. P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Research*, 29(5), 1017–26. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=56904&tool=pmcentrez&rendertype=abstract>

9 Appendix

9.1 Certificate of Ethics Review



Certificate of Ethics Review

Project Title:	Molecular studies of relevance to the RNA research.
User ID:	523612
Name:	Anastasia Callaghan
Application Date:	07/12/2016 19:28:54

You must download your certificate, print a copy and keep it as a record of this review.

It is your responsibility to adhere to the [University Ethics Policy](#) and any Department/School or professional guidelines in the conduct of your study including relevant guidelines regarding health and safety of researchers and [University Health and Safety Policy](#).

It is also your responsibility to follow University guidance on Data Protection Policy:

- [General guidance for all data protection issues](#)
- [University Data Protection Policy](#)

You are reminded that as a University of Portsmouth Researcher you are bound by [the UKRIO Code of Practice for Research](#); any breach of this code could lead to action being taken following the University's [Procedure for the Investigation of Allegations of Misconduct in Research](#).

Any changes in the answers to the questions reflecting the design, management or conduct of the research over the course of the project must be notified to the Faculty Ethics Committee. **Any changes that affect the answers given in the questionnaire, not reported to the Faculty Ethics Committee, will invalidate this certificate.**

This ethical review should not be used to infer any comment on the academic merits or methodology of the project. If you have not already done so, you are advised to develop a clear protocol/proposal and ensure that it is independently reviewed by peers or others of appropriate standing. A favourable ethical opinion should not be perceived as permission to proceed with the research; there might be other matters of governance which require further consideration including the agreement of any organisation hosting the research.

Governance Checklist

A1-Brief Description Of Project: This work encompasses the major research projects ongoing within the group which are united in their common use of molecular approaches to study:

- Mechanisms involved in post-transcriptional gene regulation, with a specific focus on small non-coding RNAs (sRNAs) and RNases
- Inhibiting essential RNases, such as RNase E, as a novel antibacterial approach
- Deactivating sRNA functions as a novel therapeutic approach targeting bacterial virulence or for synthetic biology applications
- Developing novel high throughput approaches for studying interactions with RNAs or

Certificate Code: 5802-8F52-07C7-0408-6EFB-47AF-A8A4-2E53

involving RNA
A2-Faculty: Science
A3-VoluntarilyReferToFEC: No
A5-AlreadyExternallyReviewed: No
B1-HumanParticipants: No
HumanParticipantsDefinition
B2-HumanParticipantsConfirmation: Yes
C6-SafetyRisksBeyondAssessment: No
D2-PhysicalEcologicalDamage: No
D4-HistoricalOrCulturalDamage: No
E1-ContentiousOrIllegal: No
E2-SociallySensitiveIssues: No
F1-InvolvesAnimals: No
F2-HarmfulToThirdParties: No
G1-ConfirmReadEthicsPolicy: Confirmed
G2-ConfirmReadUKRIOCodeOfPractice: Confirmed
G3-ConfirmReadConcordatToSupportResearchIntegrity: Confirmed
G4-ConfirmedCorrectInformation: Confirmed

9.2 Gel Electrophoresis Ladders

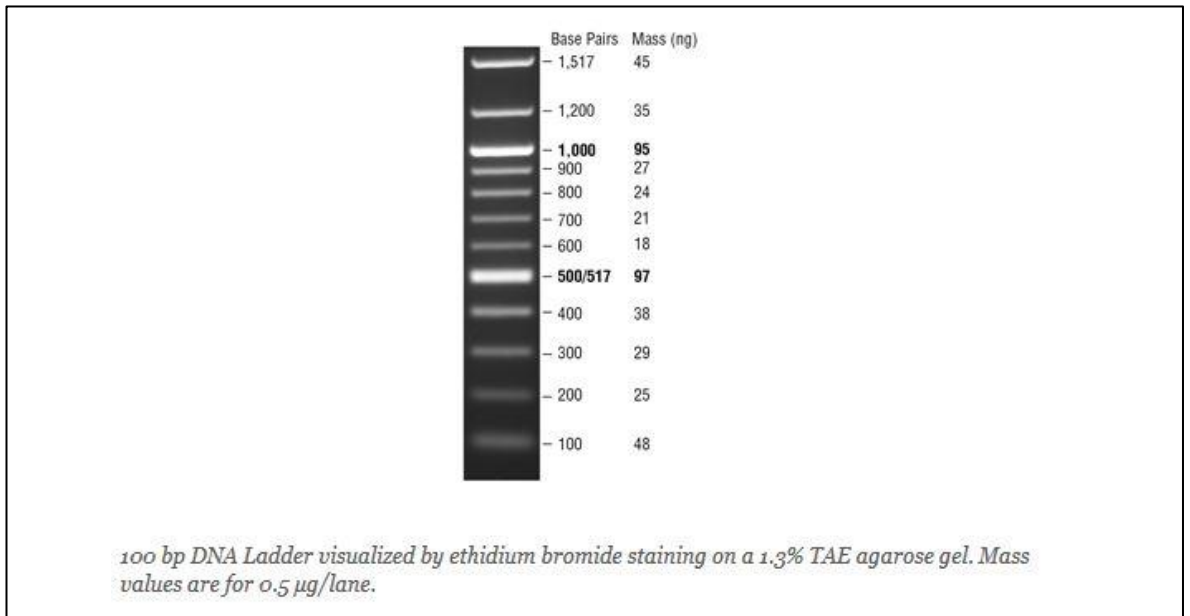


Figure 9.1 100 bp DNA Ladder

100 Base pair (bp) DNA ladder is a set of 12 DNA molecules which provide markers in the range 100 – 1,517 bp. Ladder sizes (bp) and mass (ng) are indicated and the increased intensity of the 500 b and 1,000 bands serve as reference bands. (Information provided by NEB N3231S). (<https://www.neb.com/products/n3231-100-bp-dna-ladder>)

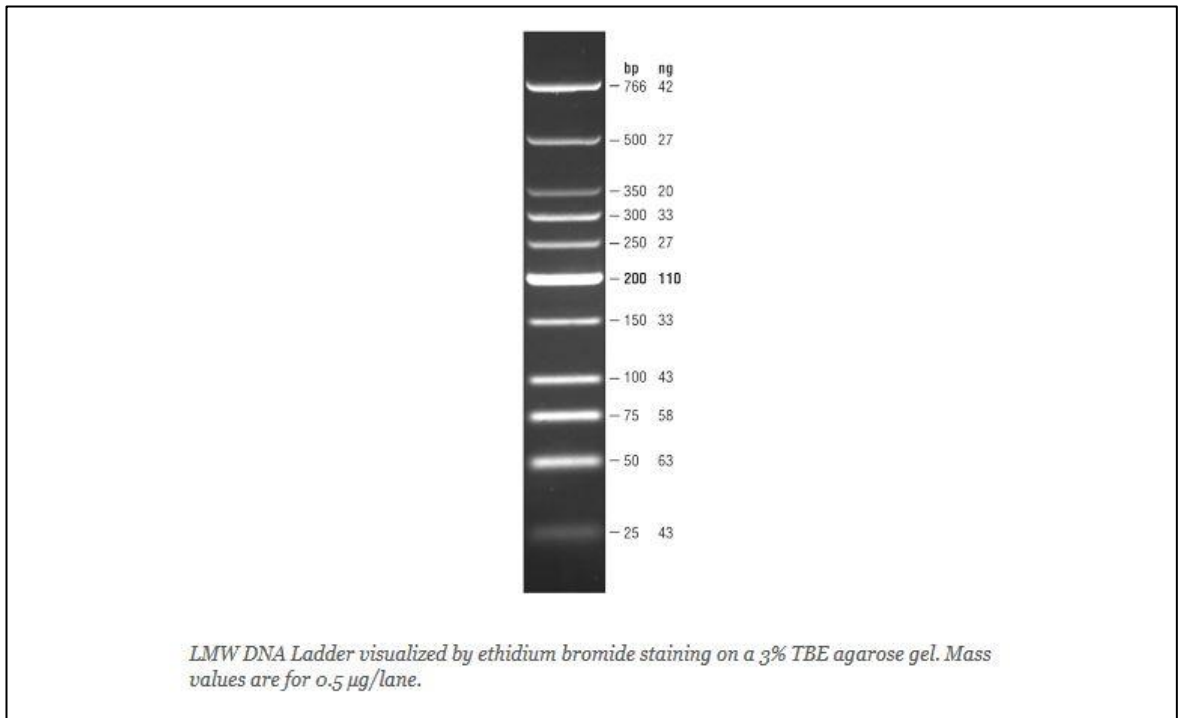


Figure 9.2 Low Molecular Weight (LMW) DNA Ladder

Low molecular weight DNA ladder is a set of 11 DNA molecules which provide markers in the range 25 – 766 base pairs (bp). Ladder sizes (bp) and mass (ng) are indicated and the increased intensity of the 200 bp band serves as a reference band. (Information provided by NEB N3233S). (<https://www.neb.com/products/n3233-low-molecular-weight-dna-ladder>)

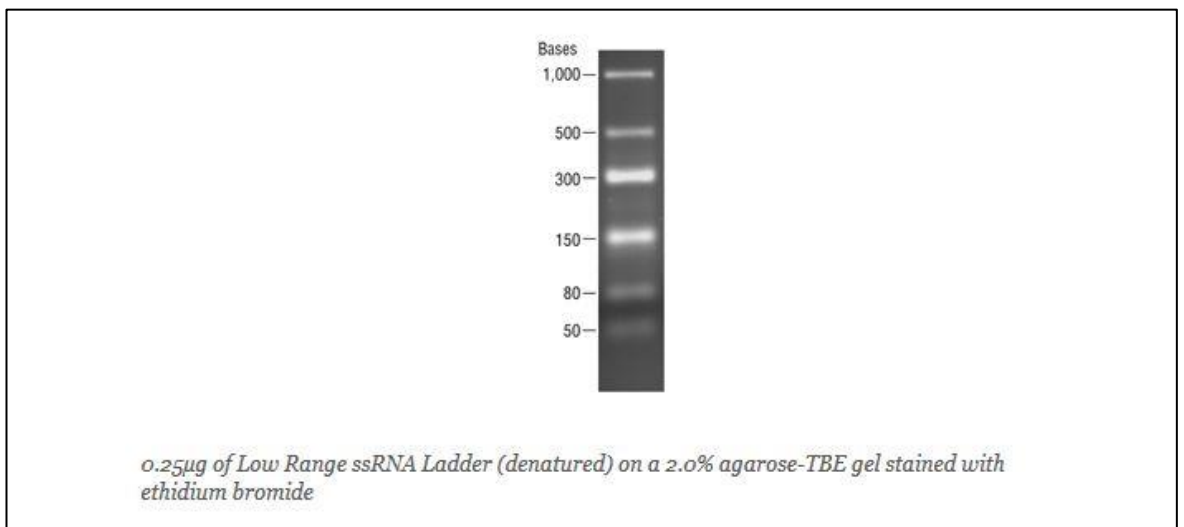


Figure 9.3 Low Range ssRNA Ladder

Low range single stranded (ss) RNA ladder is a set of 6 RNA molecules which provide markers in the range 50 – 1,000 bases. Ladder sizes are indicated and the doubled intensity of the 300 bases band serves as a reference band. (Information provided by NEB N0364S). (<https://www.neb.com/products/n0364-low-range-ssrna-ladder>)

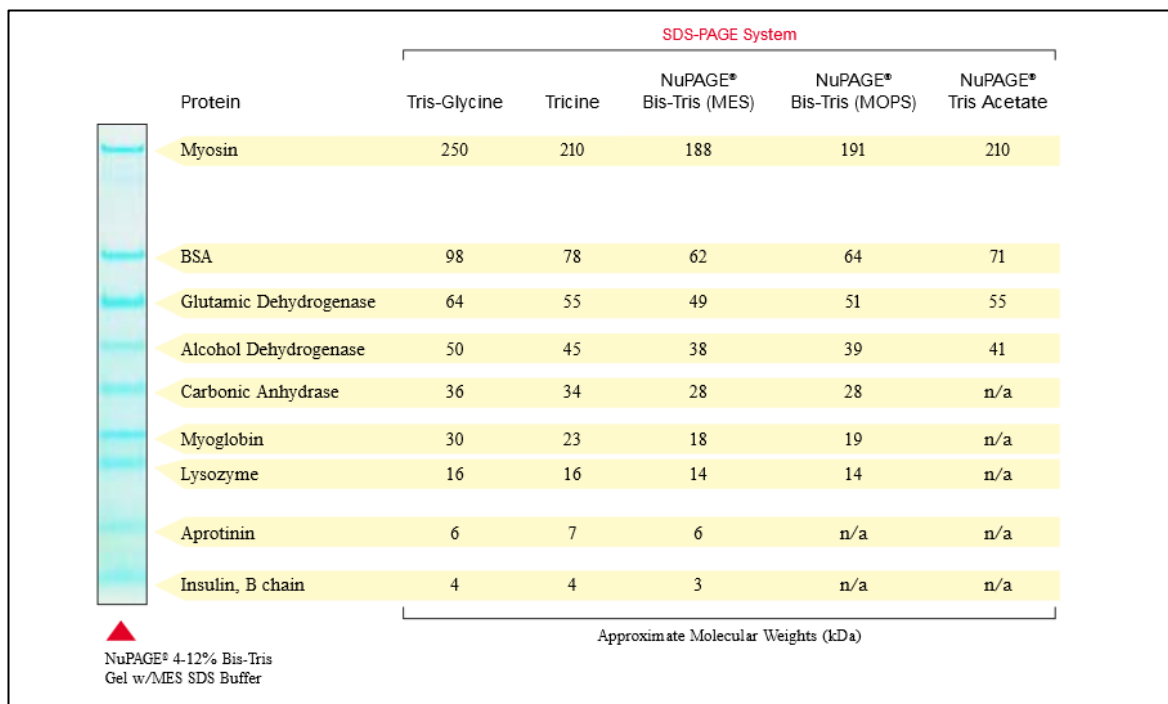


Figure 9.4 SeeBlue® Pre-Stained Standard Protein Molecular Weight Ladder

The SeeBlue® Pre-Stained Standard is a set of 9 polypeptides that resolve into sharp, tight blue bands in the range of 4–250 kDa which can be used for various SDS-PAGE systems as molecular weight size markers (approximate MW depending upon the buffer system). Information provided by Thermo Fisher (LC5625) (<https://www.thermofisher.com/order/catalog/product/LC5625>)

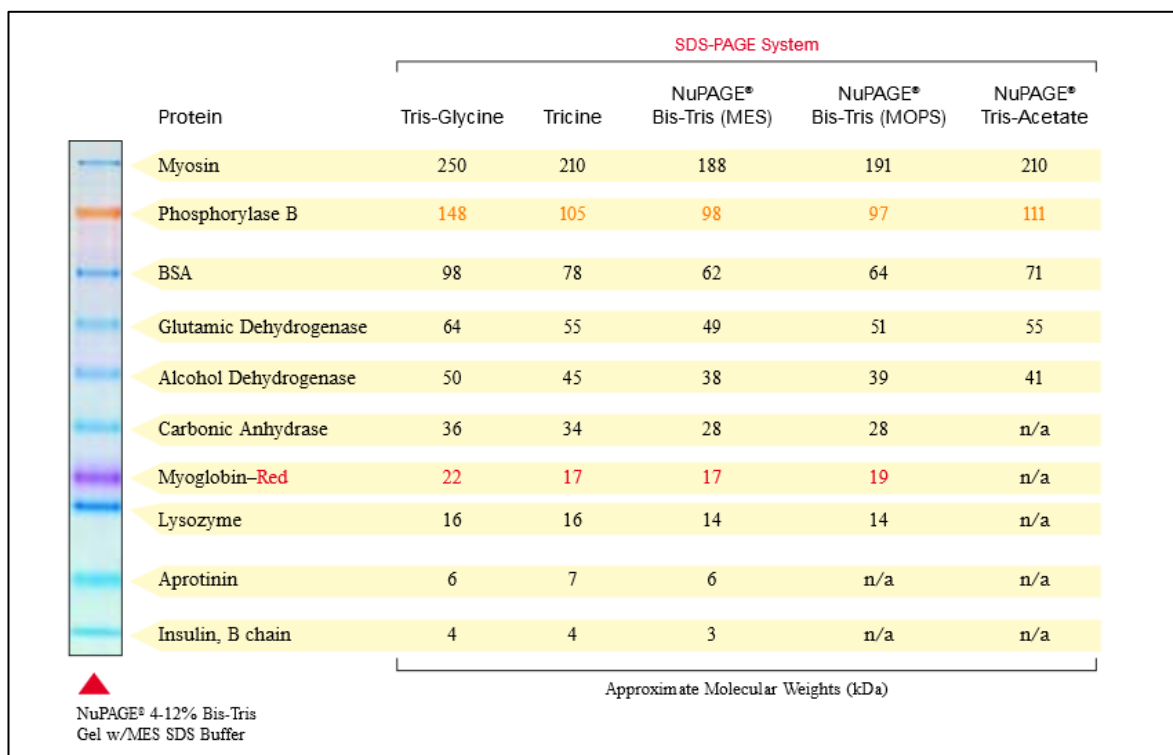


Figure 9.5 SeeBlue® Plus2 Pre-Stained Standard Protein Molecular Weight Ladder

The SeeBlue® Plus2 Pre-Stained Standard is a set of 10 polypeptides that resolve into sharp, tight bands; 8 blue and 2 contrasting coloured bands in the range of 4–250 kDa which can be used for various SDS-PAGE systems as molecular weight size markers (approximate MW depending upon the buffer system). Information provided by Thermo Fisher (LC5925) (<https://www.thermofisher.com/order/catalog/product/LC5925>).

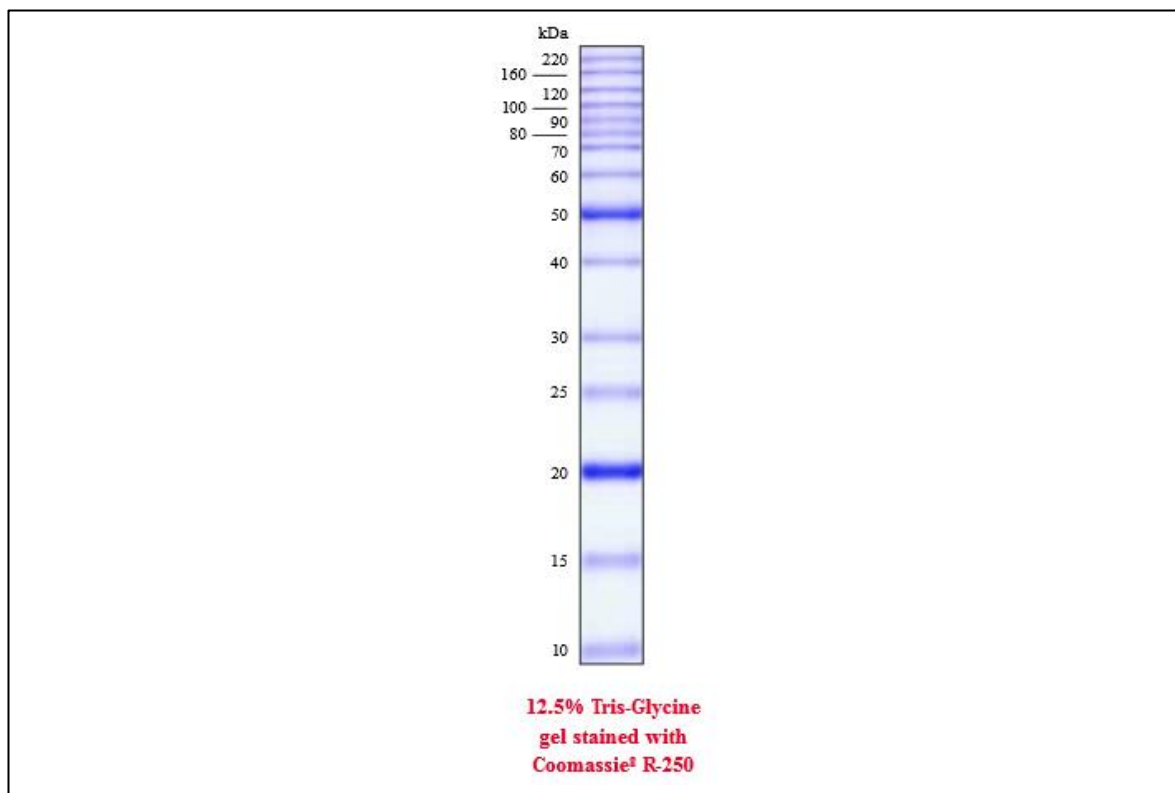


Figure 9.6 BenchMark™ Standard Protein Ladder

The BenchMark™ Standard Protein Ladder is a set of 15 recombinant H₆-proteins that resolve into sharp, tight blue bands in the range of 10–220 kDa which can be used for various SDS-PAGE systems as molecular weight size markers. Ladder sizes are indicated and the increased intensity of the 20 and 50 kDa bands serve as reference points. (approximate MW depending upon the buffer system). Information provided by Thermo Fisher (10747-012) (<https://www.thermofisher.com/order/catalog/product/10747012>).

9.3 Primer Sequences

All primer sequences utilised for site directed mutagenesis and DNA sequencing are listed.

9.3.1 Mutagenesis Primers

E. coli Core

Sense: 5'-cgatgcttaccgcatcaccgacaaacaagagcg-3'

Anti-Sense: 5'-cgctctgtttgtcgggatgacggtaagcatcg-3'

E. coli Full length and vestigial mutant (remove additional Valine)

Sense: 5'-ctttaagaaggagatataccatgcttaatccgatcgttcgtaaatt-3'

Anti-Sense: 5'-aatttacgaacgatcggattaagcatggtatatctccttcttaaag-3'

E. coli vestigial mutant (Change T->A, ACT->GCT)

Sense: 5'-caacgctgggtactgctcgtgacgcgcag-3'

Anti-Sense: 5'-ctgcgcgtcacgacgtagtaccagcgttg-3'

E. coli vestigial mutant (Change C->R, TGT->CGT)

Sense: 5'-ggcaggcgaaccgctatcgacggtcg-3'

Anti-Sense: 5'-cgaccgtcgatacgcggttcgcctgcc-3'

***E. coli* Full length (Change G->A)**

Use *E. coli* vestigial mutant (Change T->A, ACT->GCT) primers

***E. coli* Full length (Change P->A)**

Sense: 5'-ggtactggtcgtgacgcgcaggttcttgatg-3'

Anti-Sense: 5'-catcaagaacctgcgcgtcaccagctacc-3'

9.3.2 Sequencing Primers

Source Bioscience Stock Primers

Forward primer pET upstream: 5'-atgcgtccggcgtaga-3

Reverse primer Duet Down 1: 5'-gattatgcggccgtgtacaa-3'

The sequence coverage was low using these primers so designed new primers were designed for sequencing the start and middle of the PNPase sequence:

For sequencing the start of the *E. coli* PNPase gene inserted between the Nco1 and Not1 restriction sites of the pETDuet-1 vector MCS1

Forward primer: 5'-ccgatcgttcgtaaattcca-3'

Reverse primer: 5'-cgcttactcgcctgttcag-3'

For sequencing the middle region of the *E. coli* PNPase gene inserted between the Nco1 and Not1 restriction sites of the pETDuet-1 vector MCS1

Forward primer: 5'-cgtactcacggttctgcgct-3'

Reverse primer: 5'-cggaggggaagttgtagtga-3'

9.4 Cell Strain Genotypes

Description		Effects of Mutation
<i>E. coli</i> BL21(DE3)pLysS (F- ompT gal dcm lon hsdSB(rB- mB-) λ(DE3) pLysS(cmR))		
F-	Deletion Fertility (F) plasmid	Preventing homologous recombination into bacterial chromosomes
ompT	Mutation in outer membrane protein protease VII	Reducing proteolysis of expressed proteins
gal	Deficient in galactose metabolism	Blocks galactose utilization
dcm	DNA cytosine methylase mutation; methylation at second C of CCWGG sites exist (where W= A or T)	Makes DNA susceptible to cleavage by certain restriction enzymes
lon	deletion of the lon Serine protease	Reducing proteolysis of expressed proteins
hsdSB(rB-mB-)	Host DNA restriction modification system mutation; E.coli B strain, minus restriction system (inactivation of EcoK1 endonuclease), minus methylation system (r- m-)	Mutations allow E. coli to recognize DNA as foreign, eliminating restriction of unmethylated EcoK I sites of foreign DNA, thus allowing efficient transformation
λDE3	Lysogen that encodes T7 RNA polymerase.	Used to induce protein expression of sequences under T7 RNA polymerase promoter
pLysS	Includes pLysS plasmid carrying chloramphenicol resistance and phage T7 lysozyme	T7 lysozyme inhibits activity of T7 RNA polymerase, lowering basal protein expression of sequences under T7 RNA polymerase promoter
cmR	Chloramphenicol antibiotic resistance	a selection marker of pLysS plasmid
<i>E. coli</i> DH5α (F- Φ80lac ZAM15 Δ(lac ZYA-arg F) U169 rec A1 end A1 hsdR17 (rK-, mK+) pho A sup E44 λ- thi -1 gyr A96 rel A1)		
F-	Deletion Fertility (F) plasmid	Preventing homologous recombination into bacterial chromosomes.
Φ80lacZAM15	Element required for β-galactosidase complementation when plated on X-gal; carried on the lambdaoid prophage φ80	Allows blue/white screening of recombinants for use as a reporter marker
Δ(lacZYA-argF)U169	Ornithine carbamoyltransferase mutation	Ability to use arginine blocked and resistance to high levels of hydrogen peroxide increased
recA1	Mutation in a gene responsible for general repair and recombination of DNA	Prevents repair of cloned genes with direct repeats
endA1	Mutation in the non-specific Endonuclease AI	Eliminates non-specific endonuclease activity, resulting in improved plasmid preps
hsdR17 (rK-, mK+)	Host DNA restriction modification system mutation; E.coli K strain, minus restriction system (inactivation of EcoK1 endonuclease), with active methylation system (r- m+)	Mutations allow E. coli to recognize DNA as foreign, eliminating restriction of unmethylated EcoK I sites of foreign DNA, thus allowing efficient transformation
phoA	Mutation in alkaline phosphatase	Blocks phosphate utilization; used for PhoA-based reporter systems
supE44	Glutamine tRNA mutation (glutamine anticodon GUC -> AUC)	Suppresses the nonsense mutation (insertion of a stop codon) UAG (Amber), which would otherwise stop translation.
λ-thi-1	Mutation in thiamine metabolism	Mutants require thiamine for growth in minimal media
gyrA96	DNA gyrase mutant	Produces resistance to nalidixic acid
relA1	Relaxed phenotype, ppGpp synthetase I mutation.	ppGpp synthetase can detect uncharged tRNA (sign of amino acid starvation), deletion allows RNA synthesis in absence of protein synthesis. Changes in membrane composition increases fragility with respect to sonication and osmotic shock, whilst preventing protein leakage thus aiding purification

Figure 9.7 *E. coli* Cell Strain Genotype.

Information provided within this table was adapted from (“*E. coli* genotypes,” 2015) in Excel.

9.5 Vector Maps

The DNA vectors used for recombinant DNA plasmids cloning are provided as vector maps within this section and as DNA sequences in Appendix 9.6.

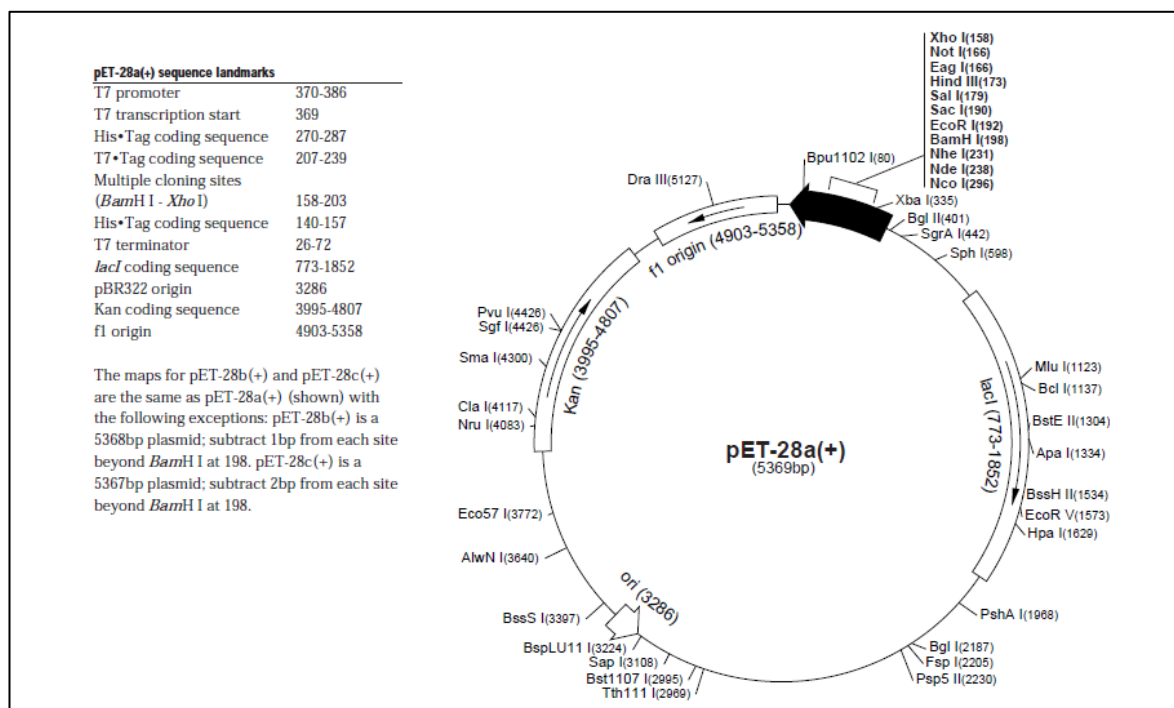


Figure 9.8 pET-28a-c (+) Vector Map

A schematic of the Pet-28a-c (+) vector map is shown, important restriction enzyme sites are indicated and labelled with their positions. The sequence is numbered so that the T7 expression region is reversed on the circular map.

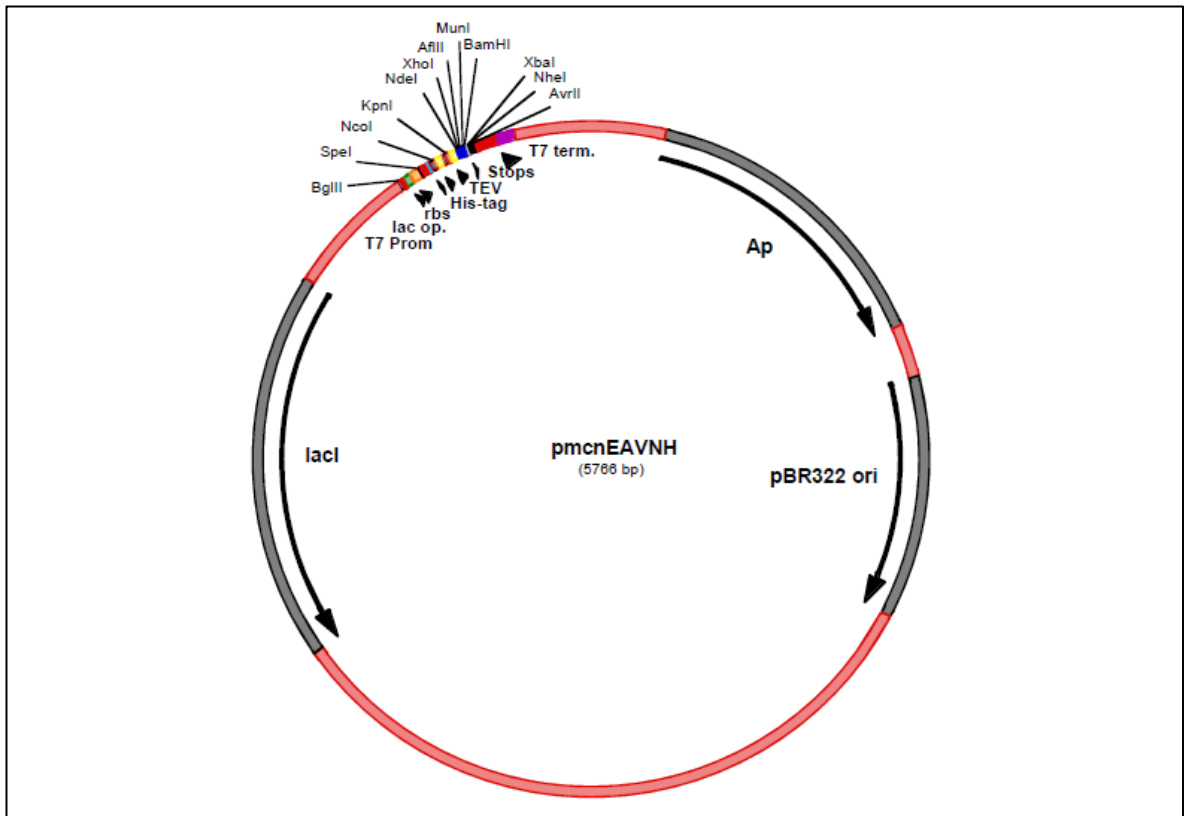


Figure 9.9 pETMCN-EAVNH Vector Map

A schematic of the pETMCN-EAVNH vector map is shown, important restriction enzyme sites are indicated and labelled with their positions.

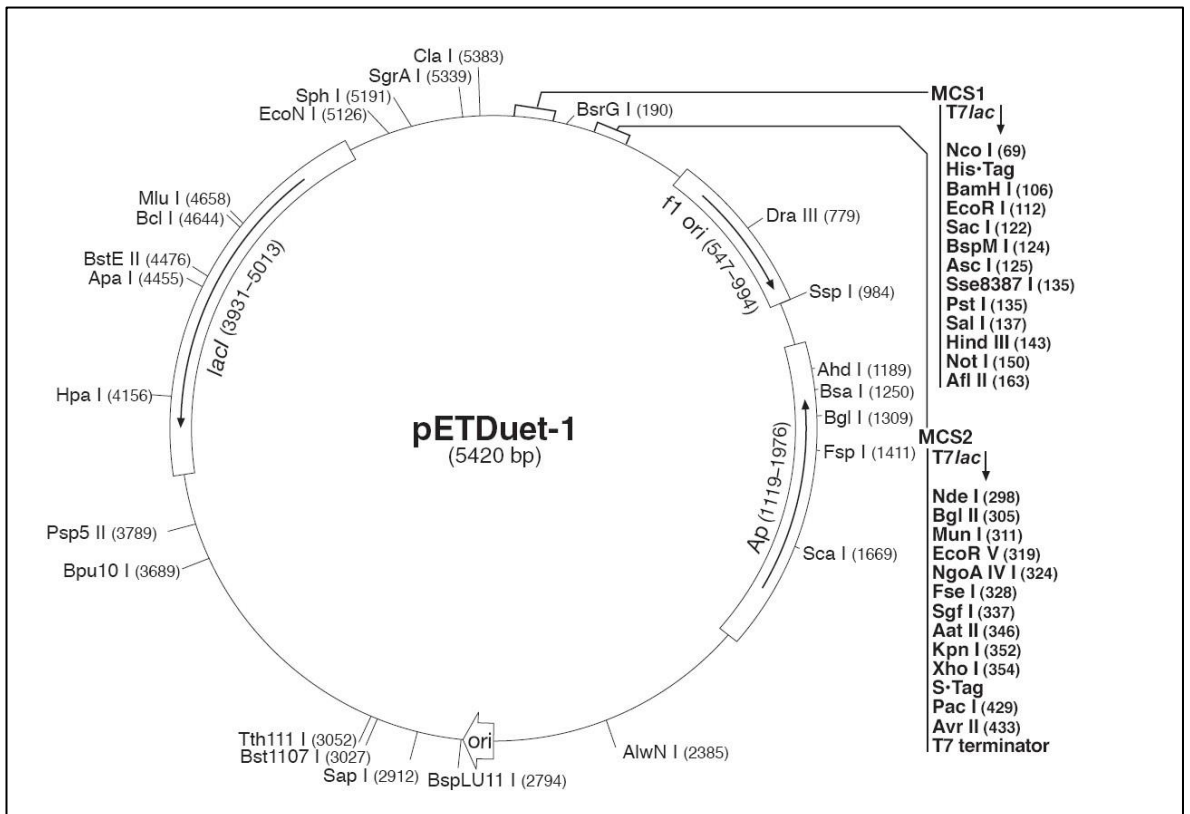


Figure 9.10 pETDuet-1 Vector Map

A schematic of the pETDuet-1 vector map is shown, important restriction enzyme sites are indicated and labelled with their positions.

9.6 Gene Cloning Strategy

9.6.1 Vector Sequence: pET-28b (+) (Reverse Complement)

(5368 base pairs) Sequence from: <http://www.addgene.org>

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTAC
GCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCC
CTTCCTTTCTCGCCACGTTGCGCCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCT
TTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTG
ATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAG
TCCACGTTCTTTAATAGTGGACTCTTGTTCCAACTGGAACAACACTCAACCCTATCTC
GGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTTCGGCCTATTGGTTAAAAAATG
AGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCA
GGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTTATTTTTCTAAATACA
TTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAACTCATCGAGCATCAAATG
AAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCT
GTAATGAAGGAGAAAACCTACCGAGGCAGTTCATAGGATGGCAAGATCCTGGTATC
GGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAA
ATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGC
AAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATC
AAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGA
AATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGC
AGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATA
CCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGT
ACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATCCGTCAGCCAGTTTAGTCTG
ACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACCTC
TGGCGCATCGGGCTTCCCATACAATCGATAGATTGTGCGCACCTGATTGCCCGACATTAT
CGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCCT
AGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGT
AAGCAGACAGTTTTATTGTTTCATGACCAAATCCCTTAACGTGAGTTTTTCGTTCCACTG
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTGTTGCCGG
ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATAACC
AAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCA
CCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAA
GTCGTGTCTTACCGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCG
GGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAA
CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAG
GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTT
CCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTGCGGGTTTCGCCACCTCTGACTTGA

GCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAAC
GCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCTGCG
TTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCG
CCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCT
GATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCATATATGGTGCAC
TCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTCCGCTATCGCT
ACGTGACTGGGTCATGGCTGCGCCCCGACCCCGCAAACACCCGCTGACGCGCCCTGA
CGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCT
GCATGTGTCAGAGGTTTTACCCGTCATACCGAAAACGCGCGAGGCAGCTGCGGTAAAG
CTCATCAGCGTGGTTCGTGAAGCGATTCACAGATGTCTGCCTGTTTCATCCGCGTCCAGCT
CGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAG
GGCGGTTTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGG
GGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATG
AACATGCCCGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGC
GGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAG
GTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCA
GGGCGCTGACTTCCGCGTTTCCAGACTTACGAAACACGGAAACCGAAGACCATTTCAT
GTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTACGTTTCGCTCGCGTAT
CGGTGATTCACTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAAC
GACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGC
TTCTCGCCGAAACGTTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGC
AAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGG
TCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAA
AGAAGACAGTCATAAGTGCGGGCAGGATAGTCATGCCCCGCGCCACCGGAAGGAGC
TGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAG
CTAACTTACATTAATTGCGTTGCGCTCACTGCCCCGCTTTCCAGTCGGGAAACCTGTCGT
GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCATATTGGGC
GCCAGGGTGGTTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTTACCAGC
CTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGGCCAGCAGGGCGAAA
ATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGT
ATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGACTCGGTAATGGCGCGCAT
TGCGCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGAACGATGCCCTCA
TTCAGCATTTGCATGGTTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTC
CGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGA
CGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTGCTGGTGACCCAATG
CGACCAGATGCTCCACGCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTT
GATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGC
TTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACG
CGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTAC

CATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACA
 ATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGAC
 TGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGC
 CGCTTCCACTTTTTCCCGCGTTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGG
 AAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTT
 CACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATAACGCGAAAG
 GTTTTGCGCCATTTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCA
 TTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATG
 GTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGCCACGGGGCCTGCCACCATAC
 CCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGT
 GATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCAC
 GATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTA
 TAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAA
 AAGGAGATATACCATGGCAGCAGCATCATCATCATCACAGCAGCGGCTGGGTG
 CCGCGCGGCAGCCATATGG/CTAGCATGACTGGTGGACAGCAAATGGGTCCGGGATCCG
 AATTCGAGCTCCG/TCGACAAGCTTGC GGCCGCACTCGAGCACCACCACCACCACC
 GAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTG
 AGCAATAACTAGCATAACCCCTTGGGGCCTCTAACGGGTCTTGAGGGGTTTTTTGCT
 GAAAGGAGGAACTATATCCGGAT

Key sequences highlighted: T7 Promoter Sequence (Grey-25%), Ribosome Binding Site (RBS) (Dark Yellow), Start Codon ATG (Teal), (6x) Histidine-Tag (H₆-Tag) (Turquoise), Thrombin cleavage site (Bright Green), Restriction site NheI (Red) & cut site (/), Sequence between NheI and Sall restriction sites to be replaced by gene insert (Grey-50%) and Restriction site Sall (Violet) & cut site (/).

NheI Restriction cut site (5'... GCTAGC... 3'
 3'... CGATCG... 5')

Sall Restriction cut site (5'... GTCGAC... 3'
 3'... CAGCTG... 5')

9.6.2 Vector Sequence: pET-MCN-EAVNH

(5766 base pairs) Sequence from

(http://archive.igbmc.fr/recherche/Sup_papers/Cavarelli/pET-MCN/pET-MCN.html)

TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAA
 TAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTAT
 TTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGAT
 AAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGC
 CCTATTCCCTTTTTTGCGGCATTTCCTTCCCTGTTTTTGTCTACCCAGAAACGCTGGT
 GAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGA

TCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCGAAGAACGTTTTCCAATGATG
AGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAG
AGCAACTCGGTGCGCCGATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGT
CACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCAT
AACCATGAGTGATAAACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAA
GGAGCTAACCCTTTTTTGCACAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGG
GAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCA
GCAATGGCAACAACGTTGCGCAAATTAATACTGGCGAACTACTTACTCTAGCTTCCC
GGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTC
GGCCCTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCT
CGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCT
ACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAG
GTGCCTCACTGATTAAGCATTGGTAACCTGTCAGACCAAGTTTACTCATATATACTTTAG
ATTGATTTAAAACCTTCATTTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAA
TCTCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAG
AAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAA
ACAAAAAAACCACCGCTACCAGCGGTGGTTTGTGGCCGGATCAAGAGCTACCAACTC
TTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGT
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCT
CTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGT
GGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTC
GTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCG
TGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGT
AAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCT
GGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT
GCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGT
CCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGT
GGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGACGCCGAACGACC
GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTC
CTTACGCATCTGTGCGGTATTTACACCGCATATATGGTGCACTCTCAGTACAATCTGC
TCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCAT
GGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTC
CCGGCATCCGTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAGAGGT
TTTACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTC
GTGAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCA
GAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGT
TTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAATGATACCGAT
GAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACT
GGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGCGGGACCAGAGAAAAAT

CACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGC
CAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCG
TTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGCTCAGGTCGC
AGACGTTTTGCAGCAGCAGTCGCTTCACGTTTCGCTCGCGTATCGGTGATTCATTCTGCT
AACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCA
TGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAGATGCGCCGCGTGCGGCTGCTGGA
GATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCATTACAGTTCTC
CGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCGAGGTGCCGCC
GGCTTCCATTTCAGGTCGAGGTGGCCCGGCTCCATGCACCCGCGACGCAACGCGGGGAG
GCAGACAAGGTATAGGGCGGGCCTACAATCCATGCCAACCCGTTCCATGTGCTCGCC
GAGGCGGCATAAATCGCCGTGACGATCAGCGGTCCAGTGATCGAAGTTAGGCTGGTA
AGAGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTCGTCATCTACCTGCCTGGACA
GCATGGCCTGCAACGCGGGCATCCCGATGCCGCCGGAAGCGAGAAGAATCATAATGG
GGAAGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGCGTCCG
CCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGT
GACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGAT
CATCGTCGCGCTCCAGCGAAAGCGGTCTCTCGCCGAAAATGACCCAGAGCGCTGCCGGC
ACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCA
TGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTTCG
AGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCC
GCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG
GGAGAGGCGGTTTGCCTATTGGGCGCCAGGGTGGTTTTTCTTTTACCAGTGAGACGG
GCAACAGCTGATTGCCCTTACCCGCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCAC
GCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAA
CATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCA
GCCCGGACTCGGTAATGGCGCGCATTGCGCCAGCGCCATCTGATCGTTGGCAACCAG
CATCGCAGTGGGAACGATGCCCTCATTTCAGCATTTGCATGGTTTGTGAAAACCGGAC
ATGGCACTCCAGTCGCTTCCCGTTCGCTATCGGCTGAATTTGATTGCGAGTGAGATA
TTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAA
CAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCAGTCGCGTACCG
TCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATA
ACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGG
ATAGTTAATGATCAGCCACTGACGCGTTGCGCGAGAAGATTGTGCACCCGCCGCTTTA
CAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGAT
CGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGG
AGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTT
GGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGCAGAAA
CGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTC
TGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCG

GGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTC
GACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCC
GTTGAGCACCGCCGCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCC
CCCGGCCACGGGGCCTGCCACCATAACCACGCCGAAACAAGCGCTCATGAGCCCGAA
GTGGCGAGCCCGATCTTCCCCATCGGTGATGTTCGGCGATATAGGCGCCAGCAACCGCA
CCTGTGGCGCCGGTGTATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGA
TCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCA
CTAGTAATAATTTTGTTTAACTTTAA**GAAGGAGA**TATACC**ATG**GGCAGCAGC**CATCAT**
CATCATCATCACAGCAGCGGTACCGGCAGCGGC**GAAAACCTTTACTTCCAGGGC**CATA
TCCTCGAGCTTAAGCAATTGGGATCCTAATAG**TCTAGA**GCTAGCCCTAGGAGATCCGG
CTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACT
AGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGA
ACTATATCCGGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCC
TACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTTCATA
CACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTAT
CGATGATAAGCTGTCAAACATGAGAA

Key sequences highlighted: T7 Promoter Sequence (Grey-25%), Ribosome Binding Site (RBS) (Dark Yellow), Start Codon ATG (Teal), (6x) Histidine-Tag (H₆-Tag) (Turquoise), TEV cleavage site (Bright Green), Restriction site Nde1 (Green) & cut site (/), Sequence between Nde1 and Xba1 restriction sites to be replaced by gene insert (Grey-50%) and Restriction site Xba1 (Pink) & cut site (/).

Nde1 Restriction cut site (5'... **CATATG**...3'
3'... **GTATAC**...5')

Xba1 Restriction cut site (5'... **TCTAGA**...3'
3'... **AGATCT**...5')

9.6.3 *Homo sapiens* Polynucleotide Phosphorylase (hPNPase)

9.6.3.1 Amino Acid Sequence: hPNPase

Homo sapiens Polynucleotide Phosphorylase (hPNPase) (Gene: PNPT1_HUMAN, GenBank: BC053660.1 UniProt: Q8TCS8) amino acid sequence. Web Resources:

(<http://www.ncbi.nlm.nih.gov/nucleotide/31657165>)

(<http://www.uniprot.org/uniprot/Q8TCS8>)

The actual sequence inserted into pET-28b Vector (using I natural variant from UniProt Q8TCS8 sequence and not the V natural variant present in the PDB 3U1K) did not include 5' mitochondrial peptide signal (Yellow):

**MAACRYCCSCLRLRPLSDGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISS
GKLARFADGS AVVQSGDTAVMVTAVSKTKPSQFMPLVVDYRQKAAAAGRIPTNYLRR**

EIGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIP
WNGPVGAVRIGIIDGEYVVPNTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHA
IKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEH
KVS RDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRC DGRDLTSL
RNV SCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDNFMLHYE
FPPYATNEIGKVTGLNRRELGHGALAEKALYPVIPRDFPF TIRVTSEVLESNGSSSMASACG
GSLALMDSGVPISAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTN
KGITALQADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPL
SKRAKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTSPAMHEARDFITEICKDDQEQQ
LEFGAVYTATITEIRD TGVMVKLYPNMTAVLLHNTQLDQRKIKHPTALGLEVGQEIQVKY
FGRDPADGRMRLSRKVLQSPATTVVRTLNDRSSIVMGEPISSSSNSQ

9.6.3.2 pET-28b [hPNPase] Transcript

Transcribed sequence from pET28-b (+) with hPNPase coding nucleotide sequence (Orange) ligated between Nhe1 and Sall restriction sites. (Sequence transcribed from start codon ATG (Teal) to stop codon TAA (Blue)).

ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCCTGGTGCCGCGCGGCAGCC
ATATGGCTAGCGCAGTTGCAGTTGATCTGGGTAATCGTAAACTGGA AATTAGCAGCGG
TAAACTGGCACGTTTTGCAGATGGTAGCGCAGTGGTTCAGAGCGGTGATACCGCAGTT
ATGGTTACCGCAGTGAGCAAACCAAACCGAGCCCAGCCAGTTTATGCCGCTGGTTG
TTGATTATCGTCAGAAAGCAGCAGCAGCCGGTCGTATTCCGACCAATTATCTGCGTCG
TGAAGTTGGCACCAGCGATAAAGAAATTCTGACCAGCCGTATTATTGATCGTAGCATT
CGTCCGCTGTTCCGGCAGGTTATTTCTATGATACCCAGGTTCTGTGTAATCTGCTGGC
AGTTGATGGTGTTAATGAACCGGATGTTCTGGCAATTAATGGTGCAAGCGTTGCACTG
AGCCTGAGCGATATTCGTGGAATGGTCCGGTTGGTGCAGTTCGTATTGGCATTATTGA
TGGTGAATATGTTGTTAACCCGACCCGTAAAGAAATGAGCAGCAGTACCCTGAATCTG
GTGGTTGCGGGTGCACCGAAAAGCCAGATTGTTATGCTGGAAGCAAGCGCAGAAAAC
ATTCTGCAGCAGGATTTTTGTGCATGCCATTAAAGTGGGTGTGAAATATACCCAGCAGA
TCATT CAGGGCATT CAGCAGCTGGTTAAAGAAACCGGTGTTACCAAACGTACACCGCA
GAAACTGTTTACCCCGAGTCCGGAAATTGTTAAATACACCCACAAACTGGCAATGGAA
CGTCTGTATGCAGTTTTTACCGATTATGAGCATGATAAAGTGAGCCGTGATGAAGCCG
TTAACAAAATTTCGTCTGGATACCGAAGAACAGCTGAAAGAAAAATTTCTGAAGCCG
ATCCGTATGAAATCATCGAAAGCTTTAATGTGGTTGCCAAAGAAGTGTTCGCAGCAT
TGTTCTGAATGAGTATAAACGTTGTGATGGTCGTGATCTGACCAGTCTGCGTAATGTTA
GCTGTGAAGTGGATATGTTTAAAACCCTGCATGGTAGTGCACTGTTTCAGCGTGGTCA
GACACAGGTGCTGTGTACCGTTACCTTTGATAGCCTGGAAAGCGGTATTAAAAGCGAT
CAGGTTATTACCGCCATTAACGGCATCAAAGACAAAACTTTATGCTGCACTATGAGT
TTCCGCCTTATGCCACCAATGAAATTGGTAAAGTTACCGGTCTGAATCGTCGTGAACT

GGGTCATGGTGCCTGGCAGAAAAAGCACTGTATCCGGTTATTCCGCGTGATTTTCCG
 TTTACCATTTCGTGTTACCAGCGAAGTTCTGGAAAGCAATGGTAGCAGCTCAATGGCAA
 GCGCATGTGGTGGTAGCCTGGCACTGATGGATAGCGGTGTTCCGATTAGCAGTGCAGT
 TGCCGGTGTGCAATTGGTCTGGTTACCAAACCGATCCGGAAAAAGGTGAAATTGAA
 GATTATCGCCTGCTGACCGATATTCTGGGTATCGAAGATTATAATGGCGATATGGACTT
 TAAAATCGCAGGCACCAATAAAGGTATTACCGCACTGCAGGCAGATATTAAGTGCCT
 GGTATTCCGATTAATAATCGTGATGGAAGCAATTCAGCAGGCCAGCGTTGCAAAAAAA
 GAAATCCTGCAGATTATGAACAAAACCATCAGCAAACCGCGTGCAAGCCGCAAAGAA
 AATGGTCCTGTTGTTGAAACCGTTCAGGTTCCGCTGAGCAAACGTGCAAAAATTTGTTG
 GTCCGGGTGGTTACAATCTGAAAAAACTGCAGGCCGAAACAGGTGTGACCATTAGCC
 AGGTTGATGAAGAAACCTTTAGCGTTTTTGCACCGACCCCGAGCGCAATGCATGAAGC
 ACGTGATTTTATCACCGAAATCTGCAAAGATGATCAAGAGCAGCAACTGGAATTTGGT
 GCCGTTTATACCGCAACCATTACCGAAATTCGTGATACCGGTGTGATGGTTAAACTGT
 ATCCTAATATGACCGCAGTGCTGCTGCATAATACCCAGCTGGATCAGCGCAAAATCAA
 ACATCCGACCGCACTGGGTCTGGAAGTTGGTCAAGAAATTCAGGTTAAATACTTTGGT
 CGCGATCCGGCAGATGGTCGTATGCGTCTGAGCCGTAAAGTTCTGCAGAGTCCGGCAA
 CCACCGTTGTTTCGTACTGAACGATCGTAGCTCAATTGTTATGGGTGAACCGATTAG
 CCAGAGTAGCAGCAATAGCCAGTAA

9.6.3.3 pET-28b [hPNPase] Plasmid Map

Cloning-GeneArt (Life Technologies) confirmed that the requested pET-28b [hPNPase] was successfully synthesised and cloned. Full length sequencing showed 100% identity and the resultant vector map is provided (Figure 9.11).

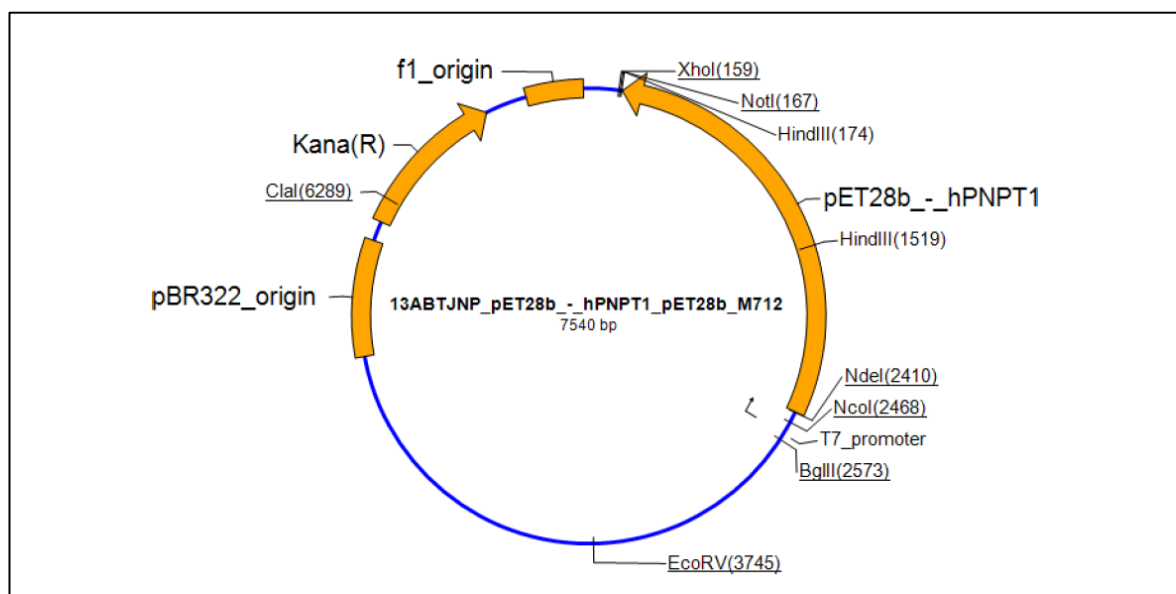


Figure 9.11 hPNPase Plasmid Map

The synthetic gene pET28b_hPNPT1 was assembled from synthetic oligonucleotides. The fragment was cloned into pET28b_M712 using NheI and SalI cloning sites. The final construct was verified by sequencing. The sequence congruence within the used restriction sites was 100%. Image provided by GeneArt AG, Life Technologies.

9.6.3.4 pET-28b [hPNPase] Translation

Translated sequence from pET28-b with hPNPase ligated between NheI and Sall restriction using ExPASy Translate Program (Frame1- H₆-hPNPase, V-I natural variant (V in Red)). When the sequence was transcribed from the start codon ATG (M) to the stop codon TAA, extra amino acids (underlined), not part of the hPNPase protein, are included at the 5' end. This includes GSS, 6x Histidine's (His-tag), SSGLVPR/GS HMAS (where (/) is the thrombin cleavage site).

MGSSHHHHHSSGLVPR/GSHMASAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAV
MVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLF
PAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVN
PTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVK
ETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQ
KEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCVDMFKTLHGSA
LFRGQQTQVLCVTFDSLESIGKSDQVITAINGIKDKNFMLHYEFPYATNEIGKVTGLNRREL
GHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISAVAGV
AIGLVTKTDPEKGEIEDYRLTLDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVM
EAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFGVPGGYNLKKLQA
ETGVTISQVDEETFSVFAPTPSAMHEARDFITEICKDDQEQQLEFGAVYTATITEIRD
TGVMVKLYPNMTAVLLHNTQLDQRKIKHPTALGLEVGQEIQVKYFGRDPADGRMRLSRKVLQS
PATTVVRTLNDRSSIVMGEPISQSSSNSQ

9.6.3.5 ProtParam Output: H₆-hPNPase

Number of amino acids: 761
Molecular weight: 83311.2
Theoretical pI: 6.63

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient 35675
Abs 0.1% (=1 g/l) 0.428, assuming all pairs of Cys residues form cystines

Ext. coefficient 35300
Abs 0.1% (=1 g/l) 0.424, assuming all Cys residues are reduced

9.6.4 *Sulfolobus solfataricus* Exosome (SsoExo4_41_42)

9.6.4.1 Amino Acid Sequence: SsoExo4_41_42

Sulfolobus solfataricus Exosome subunits Rrp4, Rrp41 & Rrp42 (SsoExoRrp4_41_42) (Gene ID: Rrp4:1454999, Rrp41:1454998 and Rrp42:1454997, UniProt: Q9UXC4, Q9UXC2, Q9UXC0) amino acid sequences:

SsoExoRrp4

MNMSQSQKIVLQPRSIVVPGELLAEGEFQIPWSPYILKINSKYYSTVVGLFDVKDTQFEVIP
LEGSFYYPKINDIVIGLVEDVEIYGWVVDIKAPYKAYLPASNLLGRSINVGEDLRRYLDVG

DYVIARIENFDRSIDPVLSVKGKDLGRVSNQIVIDIMPVKVPVIGKNKSMYETLTSKSGCSI
FVANNGRIWATCPSRFSEEILIEAIRKIENESHKGLTDRIKQFIEEKLGERNASSGETKTNS

SsoExoRrp41

MREMLQVERPKLILDDGKRTDGRKPDELRSIKIELGVLKNADGSAIFEMGNTKAIAAVYGP
KEMHPRHLSLPDRAVLRVRYHMTPFSTDERKNPAPSRREIELSKVIREALESVLEVELFPRT
AIDVFTEILQADAGSRLVSLMAASLALADAGIPMRDLIAGVAVGKADGVILDLNETEDM
WGEADMPIAMMPSLNQVTLFQLNGSMTPDEFRQAFDLAVKGINIIYNLREALKSKYVEF
KEEGV

SsoExoRrp42

MSSTPSNQNIPIIKKESIVSLFEKGIRQDGRKLTDIRPLSITLDYAKKADGSALVKLGTTMV
LAGTKLEIDKPYEDTPNQGNLIVNVELLPLAYETFEPPDENAIELARVVDRSLRDSKALD
LTKLVIEPGKSVWTVWLDVYVLDYGGNVLDACTLASVAALYNTKVYKVEQHSNGISVVK
NEVVGKLPVNYPVVTISVAKVDKYLVDVDPDLDEESIMDAKISFSYTPDLKIVGIQKSGKGS
MSLQDIDQAENTARSTAVKLEELKKHLGI

9.6.4.2 pET-MCN-EAVNH [SsoExo4_41_42] Transcript

Transcribed sequence from pET-MCN-EAVNH vector with *Sulfolobus solfataricus* Exosome (SsoExo) Rrp4, Rrp41 and Rrp42 coding nucleotide sequences (Orange) inserted between NdeI and XbaI restriction sites. (The Three Rrp4, Rrp41 and Rrp42 sequences each transcribed from start codon ATG (Teal) to stop codons TAA (Blue)).

ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGTACCGGCAGCGGC~~GAAAAC~~
~~CTTACTTCCAGGGC~~CATATGAATATGAGCCAGAGCCAGAAAATTGTTCTGCAGCCTC
GTAGCATTGTTGTTCCGGGTGAACTGCTGGCAGAAGGTGAATTCAGATCCGTGGTC
ACCGTATATTCTGAAAATCAACAGCAAATACTATAGCACCGTTGTGGGTCTGTTTGAT
GTTAAAGATACCCAGTTTGAAGTGATTCCGCTGGAAGGTAGCTTCTATTATCCGAAAA
TTAACGACATTGTGATCGGCCTGGTTGAAGATGTTGAAATTTATGGTTGGGTGGTGGGA
TATCAAAGCACCGTATAAAGCATATCTGCCTGCAAGCAATCTGCTGGGTCTGATGATT
AATGTTGGTGAAGATCTGCGTCGTTATCTGGATGTGGGTGATTATGTTATTGCCCGTAT
CGAAAATTTGATCGCAGCATTGATCCGGTCTGAGCGTTAAAGGTAAAGATCTGGGT
CGCGTTAGCAATGGTATTGTGATTGATATTATGCCGGTTAAAGTTCCGCGTGTGATTGG
TAAAAACAAAAGCATGTATGAAACCCTGACCAGCAAAAAGCGGTTGCAGCATTTTTGTT
GCAAATAATGGTCGTATTTGGGCAACCTGTCCGAGCCGTTTTAGCGAAGAAATTCTGA
TTGAAGCCATCCGCAAAATCGAAAACGAAAGCCATATTAAGGTCTGACCGATCGCAT
TAAACAGTTTATCGAAGAAAACTGGGTGAACGTAATGCAAGCAGCGGTGAAACCAA
AACCAATAGCTAACTCGAGCTTAAGCAATTGGGATCCCGATCCCGCGAAATTAATACG
ACTCACTATAGGGGAATTGTGAGCGGATAACAATCCCCTAGTAATAATTTGTTTA

ACTTTAAGAAGGAGATATAGCACCGATGCGTGAAATGCTGCAGGTTGAACGTCCGAA
ACTGATTCTGGATGATGGTAAACGTACCGATGGTCGTAAACCGGATGAACTGCGTAGC
ATTA AAAATCGAACTGGGTGTTCTGAAAAATGCAGATGGTAGCGCAATTTTTGAAATGG
GTAATACCAAAGCAATCGCAGCAGTTTATGGTCCGAAAGAAATGCATCCGCGTCATCT
GAGCCTGCCGGATCGTGCAGTTCTGCGTGTTCTGTTATCACATGACCCCGTTTAGCACCG
ATGAACGTAAAAATCCGGCACCGAGCCGTCGTGAAATTGAACTGAGCAAAGTTATTCG
TGAAGCACTGGAAAGCGCAGTTCTGGTTGAACTGTTTCCGCGTACCGCAATTGATGTT
TTTACCGAAATTCTGCAGGCAGATGCAGGTAGCCGTCCTGGTTAGCCTGATGGCAGCAA
GCCTGGCACTGGCCGATGCCGGTATTCCGATGCGTGATCTGATTGCCGGTGTTGCAGTT
GGTAAAGCAGATGGTGTTATTATTCTGGATCTGAACGAAACCGAAGATATGTGGGGTG
AAGCAGATATGCCGATTGCAATGATGCCGAGCCTGAATCAGGTTACCCTGTTCCAGCT
GAATGGTAGCATGACTCCGGATGAATTTTCGTCAGGCATTTGATCTGGCAGTGAAAGGC
ATTAACATCATCTATAATCTGGAACGCGAAGCCCTGAAAAGCAAATATGTGGAATTTA
AAGAAGAAGGCGTTTAACTCGAGCTTAAGCAATTGGGATCCCGATCCCGCGAAATTA
TACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCACTAGTAATAATTTT
GTTTAACTTTAAGAAGGAGATATAGCACCGATGAGCAGCACCCCGAGCAATCAGAAC
ATTATTCCGATCATCAAAAAGAAAGCATCGTGAGCCTGTTTGAAAAAGGTATTTCGTC
AGGATGGTCGAAACTGACCGATTATCGTCCGCTGAGCATTACCCTGGATTATGCAAA
AAAAGCCGATGGTAGTGCAGTGGTTAAACTGGGCACCACCATGGTTCTGGCAGGCACC
AAACTGGAAATTGATAAACCGTATGAAGATACCCCGAATCAGGGCAATCTGATTGTTA
ATGTTGAACTGCTGCCGCTGGCCTATGAAACCTTTGAACCGGGTCCGCTGATGAAAA
TGCCATTGAACTGGCACGTGTTGTTGATCGTAGCCTGCGTGATAGCAAAGCCCTGGAC
CTGACAAAACCTGGTTATCGAACCGGGTAAAAGCGTTTGGACCGTTTGGCTGGATGTTT
ATGTTCTGGATTACGGTGGTAATGTGCTGGATGCATGTACCCTGGCAAGCGTTGCAGC
ACTGTATAATACAAAAGTGTACAAAAGTGGAACAGCACAGCAATGGTATTAGCGTGAA
TAAAAACGAAGTGGTTGGTAAACTGCCGCTGAATTATCCGTTGTTACCATTAGCGTT
GCCAAAGTGGATAAATATCTGGTTGTGGATCCGGACCTGGATGAAGAAAGTATTATGG
ATGCCAAAATCAGCTTCAGCTATACACCGGATCTGAAAATTGTGGGTATTTCAGAAAAG
CGGTAAAGGTAGCATGTCACTGCAGGATATTGATCAGGCAGAAAACACCGCACGTAG
CACCGCAGTGAAACTGCTGGAAGAACTGAAAAACATCTGGGCATCTAA

9.6.4.3 pET-MCN-EAVNH [SsoExo4_41_42] Plasmid Map

Cloning-GeneArt (Life Technologies) confirmed that the requested pET-MCN-EAVNH [SsoExo4_41_42] was successfully synthesised and cloned. Full length sequencing showed 100% identity and the resultant vector map is provided (Figure 9.12)

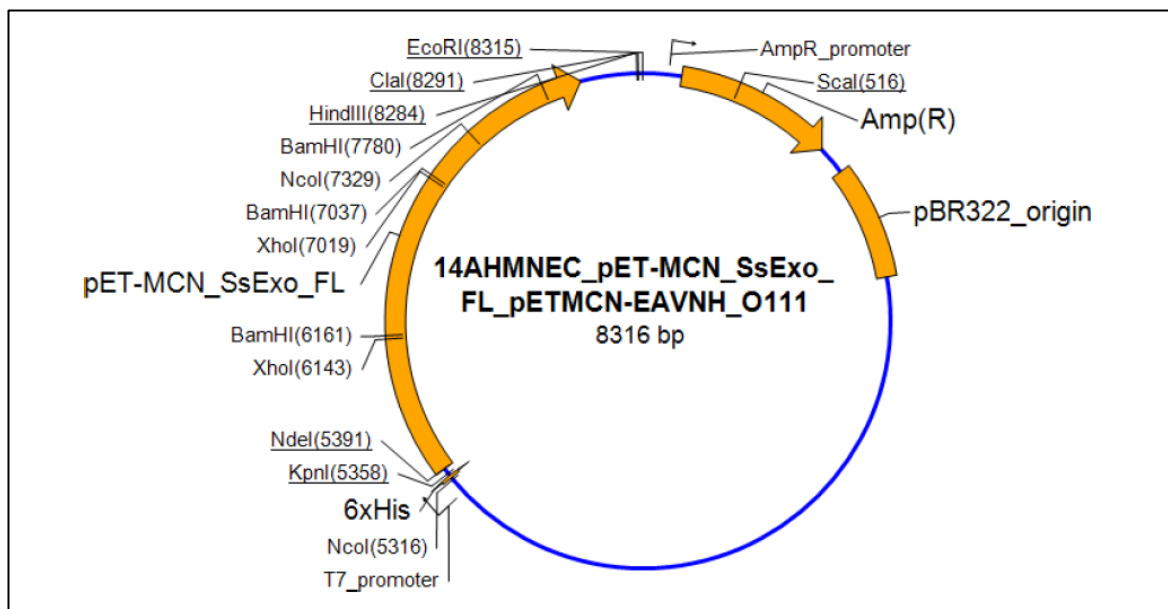


Figure 9.12 SsoExosome Plasmid Map

The synthetic gene SsoExo4_41_42 was assembled from synthetic oligonucleotides. The fragment was cloned into pET-MCN-EAVNH vector using NdeI and XbaI cloning sites. The final construct was verified by sequencing. The sequence congruence within the used restriction sites was 100%. (Cloning-GeneArt AG, Life Technologies). Image provided by GeneArt AG, Life Technologies.

9.6.4.4 pET-MCN-EAVNH [SsoExo4_41_42] Translation

Translated sequence from pET-MCN-EAVNH with SsoExo4_41_42 ligated between NdeI and XbaI restriction sites using ExPASy Translate Program (Frame1- H₆-SsoExoRrp4, Rrp41 and Rrp42). When the sequence was transcribed from the start codons ATG (M) to the stop codons TAA, extra amino acids (underlined) not part of the H₆- SsoExoRrp4 protein are included at the 5' end. This includes MGSS, 6x Histidine's (His-tag), SSGTGSGENLYFQ/GH (where (/) is the TEV cleavage site).

H₆- SsoExoRrp4

MGSSHHHHHHSSGTGSGENLYFQGHMNMSQSQKIVLQPRSIVVPGELLAEGEFQIPWSPYI
LKINSKYYSTVVGLFDVKDTQFEVIPLEGSFYYPKINDIVIGLVEDVEIYGWVVDIKAPYKA
YLPASNLLGRSINVGEDLRRYLDVGDYVIARIENFDRSIDPVLSVKGKDLGRVSNGLVIDIM
PVKVPRVIGKNKSMYETLTSKSGCSIFVANNGRIWATCSRFSSEEILIEAIRKIENESHKGLT
DRIKQFIEEKLGERNASSGETKTNS

SsoExoRrp41

MREMLQVERPKLILDDGKRTDGRKPDELRSIKIELGVLKNADGSAIFEMGNTKAIAAVYGP
KEMHPRHLSLPDRAVLRVRYHMTPFSTDERKNPAPSRREIELSKVIREALESVVELFPRT
AIDVFTEILQADAGSRLVSLMAASLALADAGIPMRDLIAGVAVGKADGVIIIDLNETEDM
WGEADMPIAMMPSLNQVTLFQLNGSMTPEFRQAFDLAVKGINIYNLREALKSKYVEF
KEEGV

SsoExoRrp42

MSSTPSNQNIPIIKKESIVSLFEKGIRQDGRKLTDIRPLSITLDYAKKADGSALVKLGTTMV
LAGTKLEIDKPYEDTPNQGNLIVNVELLPLAYETFEPGPPDENAIELARVVDRSLRDSKALD
LTKLVIEPGKSVWTVWLDVYVLDYGGNVLDACTLASVAALYNTKVYKVEQHSNGISVVK
NEVVGKLPVNYVVTISVAKVDKYLVDVDPDLDEESIMDAKISFSYTPDLKIVGIQKSGKGS
MSLQDIDQAENTARSTAVKLEELKKHLGI

9.6.4.5 ProtParam Output: H₆-SsoExo4

H₆- SsoExoRrp4

Number of amino acids: 274

Molecular weight: 30706.0

Theoretical pI: 6.46

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient 34505

Abs 0.1% (=1 g/l) 1.124, assuming all pairs of Cys residues form cystines

Ext. coefficient 34380

Abs 0.1% (=1 g/l) 1.120, assuming all Cys residues are reduced

9.6.4.6 ProtParam Output: SsoExo41

SsoExoRrp41

Number of amino acids: 248

Molecular weight: 27577.9

Theoretical pI: 5.20

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient 11460

Abs 0.1% (=1 g/l) 0.416

9.6.4.7 ProtParam Output: SsoExo42

SsoExoRrp42

Number of amino acids: 275

Molecular weight: 30193.6

Theoretical pI: 5.16

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient 27390

Abs 0.1% (=1 g/l) 0.907, assuming all pairs of Cys residues form cystines

Ext. coefficient 27390

Abs 0.1% (=1 g/l) 0.907, assuming all Cys residues are reduced

9.6.4.8 ProtParam Output: H₆- SsoExo4_41_42

H₆- SsoExoRrp4_4_41_42

Number of amino acids: 2391

Molecular weight: 265288.7

Theoretical pI: 5.57

Extinction coefficients:

Extinction coefficients are in units of $M^{-1} \text{ cm}^{-1}$, at 280 nm measured in water.

Ext. coefficient 220190

Abs 0.1% (=1 g/l) 0.830, assuming all pairs of Cys residues form cystines

Ext. coefficient 219690

Abs 0.1% (=1 g/l) 0.828, assuming all Cys residues are reduced

9.6.5 *E. coli* Polynucleotide Phosphorylase (EcPNPase)**9.6.5.1 Amino Acid Sequence: EcPNPase**

E. coli Polynucleotide Phosphorylase (EcPNPase) (Gene: PNP_ECOLI, UniProt: P05055, (<http://www.uniprot.org/uniprot/P05055>)) was cloned into a pETDuet-1 vector and kindly supplied by Prof. Ben Luisi (University of Cambridge)

9.6.5.2 ProtParam Output: EcPNPase

User-provided sequence:

```

10           20           30           40           50           60
MLNPIVRKFQ YGQHTVTLET GMMARQATAA VMVSMDDTAV FVTVVGQKKA KPGQDFPLT

70           80           90           100          110          120
VNYQERTYAA GRIPGSFFRR EGRPSEGETL IARLDRPIR PLFPEGFVNE VQVIATVSV

130          140          150          160          170          180
NPQVNPDIVA MIGASAALSL SGIPFNGPIG AARVGYINDQ YVLNPTQDEL KESKLDLVVA

190          200          210          220          230          240
GTEAAVLMVE SEAQLLEDQ MLGAVVFGHE QQQVVIQNIN ELVKEAGKPR WDWQPEPVNE

250          260          270          280          290          300
ALNARVAALA EARLSDAYRI TDKQERYAQV DVIKSETIAT LLAEDETLDE NELGEILHAI

310          320          330          340          350          360
EKNVRSRVL AGEPRIDGRE KDMIRGLDVR TGVLPRTHGS ALFTRGETQA LVTATLGTAR

370          380          390          400          410          420
DAQVLDELMG ERTDTFLFHY NFPPYSVGET GMVGSPKRRE IGHGRLAKRG VLAVMPDMDK

430          440          450          460          470          480
FPYTVRVVSE ITESNGSSSM ASVCGASLAL MDAGVPIKAA VAGIAMGLVK EGDNYVVLSD

490          500          510          520          530          540
ILGDEDHLGD MDFKVAGSRD GISALQMDIK IEGITKEIMQ VALNQAKGAR LHILGVMEQA

550          560          570          580          590          600
INAPRGDISE FAPRIHTIKI NPKIKDVIG KGGSVIRALT EETGTTIEIE DDGTVKIAAT

610          620          630          640          650          660
DGEKAKHAIR RIEEITAEIE VGRVYTGKVT RIVDFGAFVA IGGGKEGLVH ISQIADKRVE

670          680          690          700          710
KVTDYLQMGQ EVPVKVLEVD RQGRIRLSIK EATEQSQPAA APEAPAAEQG E

```

Number of amino acids: 711

Molecular weight: 77100.9

Theoretical pI: 5.10

Extinction coefficients:

Extinction coefficients are in units of $M^{-1} \text{ cm}^{-1}$, at 280 nm measured in water.

Ext. coefficient 30370

Abs 0.1% (=1 g/l) 0.394, assuming all pairs of Cys residues form cystines

Ext. coefficient 30370

Abs 0.1% (=1 g/l) 0.394, assuming all Cys residues are reduced

9.7 PACT Premier™ Crystallography Screen

The PACT premier™ screen (Molecular Dimensions) conditions are listed in Table 9.1 (Newman *et al.*, 2005).

Tube #	Conc1	Units1	Salt 1	Conc2	Units2	Buffer2	pH	Conc3	Units3	Precipitant3
1-1				0.1	M	SPG	4	25	% w/v	PEG 1500
1-2				0.1	M	SPG	5	25	% w/v	PEG 1500
1-3				0.1	M	SPG	6	25	% w/v	PEG 1500
1-4				0.1	M	SPG	7	25	% w/v	PEG 1500
1-5				0.1	M	SPG	8	25	% w/v	PEG 1500
1-6				0.1	M	SPG	9	25	% w/v	PEG 1500
1-7	0.2	M	Sodium chloride	0.1	M	Sodium acetate	5	20	% w/v	PEG 6000
1-8	0.2	M	Ammonium chloride	0.1	M	Sodium acetate	5	20	% w/v	PEG 6000
1-9	0.2	M	Lithium chloride	0.1	M	Sodium acetate	5	20	% w/v	PEG 6000
1-10	0.2	M	Magnesium chloride hexahydrate	0.1	M	Sodium acetate	5	20	% w/v	PEG 6000
1-11	0.2	M	Calcium chloride dihydrate	0.1	M	Sodium acetate	5	20	% w/v	PEG 6000
1-12	0.01	M	Zinc chloride	0.1	M	Sodium acetate	5	20	% w/v	PEG 6000
1-13				0.1	M	MIB	4	25	% w/v	PEG 1500
1-14				0.1	M	MIB	5	25	% w/v	PEG 1500
1-15				0.1	M	MIB	6	25	% w/v	PEG 1500
1-16				0.1	M	MIB	7	25	% w/v	PEG 1500
1-17				0.1	M	MIB	8	25	% w/v	PEG 1500
1-18				0.1	M	MIB	9	25	% w/v	PEG 1500
1-19	0.2	M	Sodium chloride	0.1	M	MES	6	20	% w/v	PEG 6000
1-20	0.2	M	Ammonium chloride	0.1	M	MES	6	20	% w/v	PEG 6000
1-21	0.2	M	Lithium chloride	0.1	M	MES	6	20	% w/v	PEG 6000
1-22	0.2	M	Magnesium chloride hexahydrate	0.1	M	MES	6	20	% w/v	PEG 6000
1-23	0.2	M	Calcium chloride dihydrate	0.1	M	MES	6	20	% w/v	PEG 6000
1-24	0.01	M	Zinc chloride	0.1	M	MES	6	20	% w/v	PEG 6000
1-25				0.1	M	PCTP	4	25	% w/v	PEG 1500
1-26				0.1	M	PCTP	5	25	% w/v	PEG 1500
1-27				0.1	M	PCTP	6	25	% w/v	PEG 1500
1-28				0.1	M	PCTP	7	25	% w/v	PEG 1500
1-29				0.1	M	PCTP	8	25	% w/v	PEG 1500
1-30				0.1	M	PCTP	9	25	% w/v	PEG 1500
1-31	0.2	M	Sodium chloride	0.1	M	HEPES	7	20	% w/v	PEG 6000
1-32	0.2	M	Ammonium chloride	0.1	M	HEPES	7	20	% w/v	PEG 6000
1-33	0.2	M	Lithium chloride	0.1	M	HEPES	7	20	% w/v	PEG 6000
1-34	0.2	M	Magnesium chloride hexahydrate	0.1	M	HEPES	7	20	% w/v	PEG 6000
1-35	0.2	M	Calcium chloride hexahydrate	0.1	M	HEPES	7	20	% w/v	PEG 6000
1-36	0.01	M	Zinc chloride	0.1	M	HEPES	7	20	% w/v	PEG 6000
1-37				0.1	M	MMT	4	25	% w/v	PEG 1500
1-38				0.1	M	MMT	5	25	% w/v	PEG 1500
1-39				0.1	M	MMT	6	25	% w/v	PEG 1500
1-40				0.1	M	MMT	7	25	% w/v	PEG 1500
1-41				0.1	M	MMT	8	25	% w/v	PEG 1500
1-42				0.1	M	MMT	9	25	% w/v	PEG 1500
1-43	0.2	M	Sodium chloride	0.1	M	Tris	8	20	% w/v	PEG 6000
1-44	0.2	M	Ammonium chloride	0.1	M	Tris	8	20	% w/v	PEG 6000
1-45	0.2	M	Lithium chloride	0.1	M	Tris	8	20	% w/v	PEG 6000
1-46	0.2	M	Magnesium chloride hexahydrate	0.1	M	Tris	8	20	% w/v	PEG 6000
1-47	0.2	M	Calcium chloride dihydrate	0.1	M	Tris	8	20	% w/v	PEG 6000
1-48	0.002	M	Zinc chloride	0.1	M	Tris	8	20	% w/v	PEG 6000
2-1	0.2	M	Sodium fluoride					20	% w/v	PEG 3350
2-2	0.2	M	Sodium bromide					20	% w/v	PEG 3350
2-3	0.2	M	Sodium iodide					20	% w/v	PEG 3350
2-4	0.2	M	Potassium thiocyanate					20	% w/v	PEG 3350
2-5	0.2	M	Sodium nitrate					20	% w/v	PEG 3350
2-6	0.2	M	Sodium formate					20	% w/v	PEG 3350
2-7	0.2	M	Sodium acetate trihydrate					20	% w/v	PEG 3350
2-8	0.2	M	Sodium sulfate					20	% w/v	PEG 3350
2-9	0.2	M	Potassium sodium tartrate tetrahydrate					20	% w/v	PEG 3350
2-10	0.02	M	Sodium/potassium phosphate					20	% w/v	PEG 3350
2-11	0.2	M	Sodium citrate tribasic dihydrate					20	% w/v	PEG 3350
2-12	0.2	M	Sodium malonate dibasic monohydrate					20	% w/v	PEG 3350
2-13	0.2	M	Sodium fluoride	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-14	0.2	M	Sodium bromide	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-15	0.2	M	Sodium iodide	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-16	0.2	M	Potassium thiocyanate	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-17	0.2	M	Sodium nitrate	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-18	0.2	M	Sodium formate	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-19	0.2	M	Sodium acetate trihydrate	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-20	0.2	M	Sodium sulfate	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-21	0.2	M	Potassium sodium tartrate tetrahydrate	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-22	0.02	M	Sodium/potassium phosphate	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-23	0.2	M	Sodium citrate tribasic dihydrate	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-24	0.2	M	Sodium malonate dibasic monohydrate	0.1	M	Bis-Tris propane	6.5	20	% w/v	PEG 3350
2-25	0.2	M	Sodium fluoride	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-26	0.2	M	Sodium bromide	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-27	0.2	M	Sodium iodide	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-28	0.2	M	Potassium thiocyanate	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-29	0.2	M	Sodium nitrate	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-30	0.2	M	Sodium formate	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-31	0.2	M	Sodium acetate trihydrate	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-32	0.2	M	Sodium sulfate	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-33	0.2	M	Potassium sodium tartrate tetrahydrate	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-34	0.02	M	Sodium/potassium phosphate	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-35	0.2	M	Sodium citrate tribasic dihydrate	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-36	0.2	M	Sodium malonate dibasic monohydrate	0.1	M	Bis-Tris propane	7.5	20	% w/v	PEG 3350
2-37	0.2	M	Sodium fluoride	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-38	0.2	M	Sodium bromide	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-39	0.2	M	Sodium iodide	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-40	0.2	M	Potassium thiocyanate	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-41	0.2	M	Sodium nitrate	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-42	0.2	M	Sodium formate	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-43	0.2	M	Sodium acetate trihydrate	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-44	0.2	M	Sodium sulfate	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-45	0.2	M	Potassium sodium tartrate tetrahydrate	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-46	0.02	M	Sodium/potassium phosphate	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-47	0.2	M	Sodium citrate tribasic dihydrate	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350
2-48	0.2	M	Sodium malonate dibasic monohydrate	0.1	M	Bis-Tris propane	8.5	20	% w/v	PEG 3350

Table 9.1 PACT Premier™ Crystallography Screen

Conditions for the PACT Premier™ screen from Molecular Dimensions (Newman *et al.*, 2005).

9.8 Protein Accession Codes

Protein Accession codes for all sequences utilised for basic bioinformatics studies are listed in Table 9.2. Sequences used for generating sequence logos are provided as supplementary data in an excel file (CD).

Organism	Protein	Accession Number
<i>Escherichia coli</i>	PNPase	WP_060707494.1
<i>Coxiella burnetii</i>	PNPase	WP_042526278.1
<i>Caulobacter crescentus</i>	PNPase	WP_010917924.1
<i>Streptomyces antibioticus</i>	PNPase	GI: 75349253
<i>Synechocystis sp. PCC6803</i>	PNPase	WP_010871289
<i>Homo sapiens</i>	PNPase	NP_149100.2
<i>Methanothermobacter thermautotrophicus</i>	Rrp41	WP_010876322.1
	Rrp42	WP_010876321.1
<i>Pyrococcus abyssi</i>	Rrp41	WP_010867734.1
	Rrp42	WP_010867735.1
<i>Archaeoglobus fulgidus</i>	Rrp41	WP_010878000.1
	Rrp42	WP_010878001.1
<i>Sulfolobus solfataricus</i>	Rrp41	WP_009991308
	Rrp42	WP_009991305.1

Table 9.2 Protein Accession Codes

The accession numbers for each organism utilised in basic bioinformatics alignments are provided.

9.9 Metabolite Structures

The 2D structures of the metabolites utilised for docking calculations in MOE (Molecular Operating Environment, 2013) are provided in Figure 9.13-Figure 9.14.

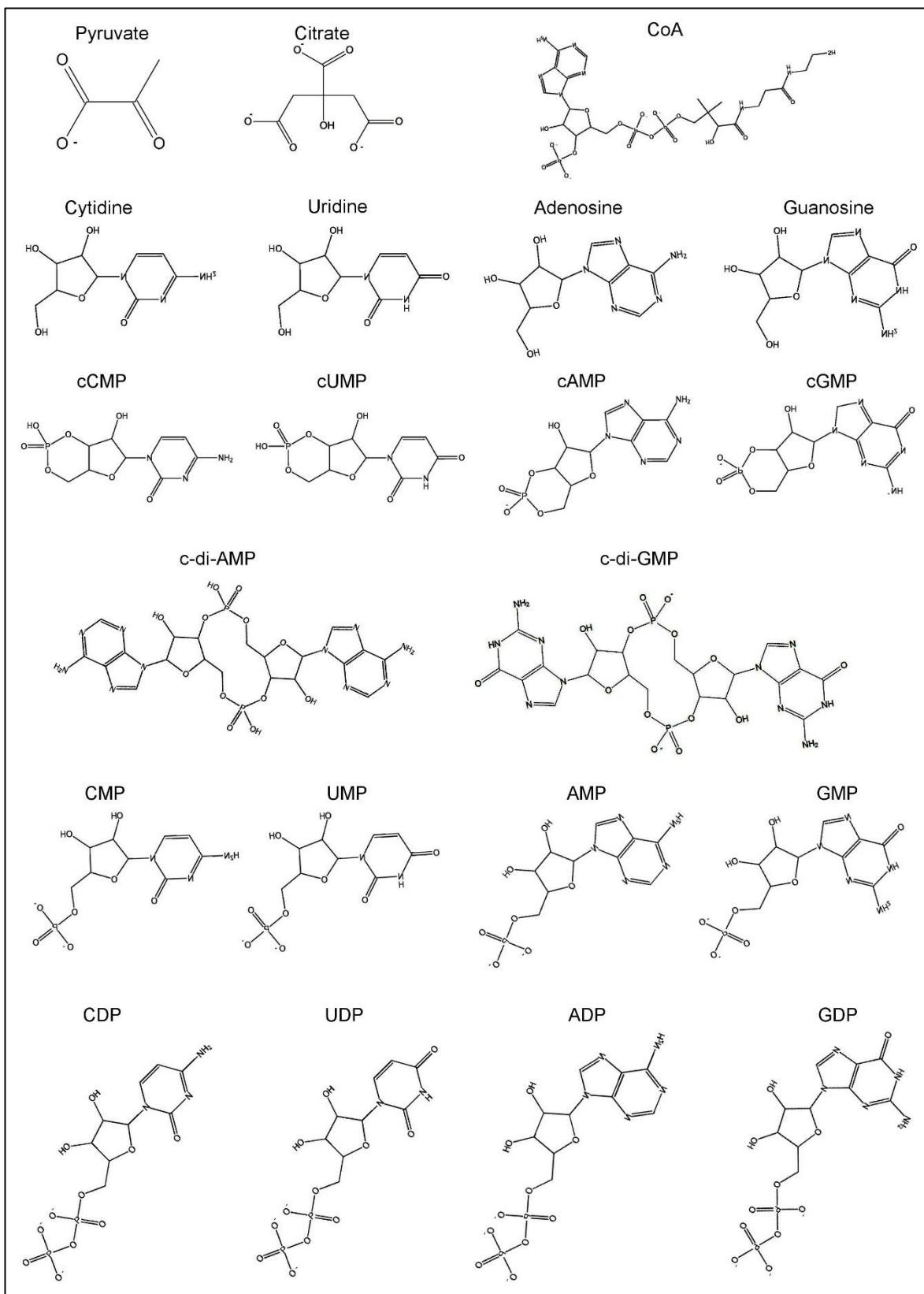


Figure 9.13 Metabolite 2D Structures

The 2D structures of the metabolites utilised for docking calculations in MOE (Molecular Operating Environment, 2013).

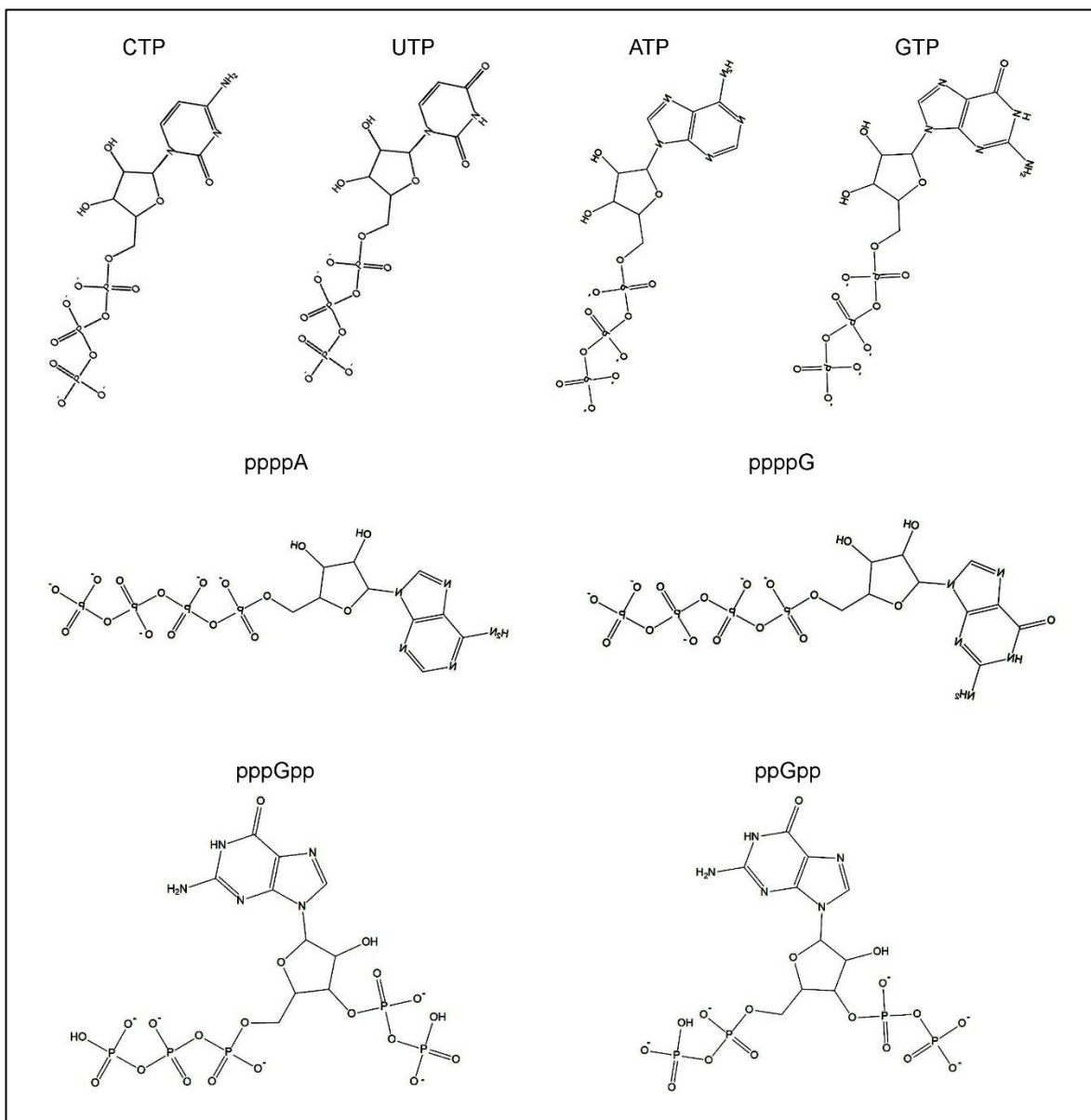


Figure 9.14 Metabolite 2D Structures Continued

The 2D structures of the metabolites utilised for docking calculations in MOE (Molecular Operating Environment, 2013)

FORM UPR16

Research Ethics Review Checklist



Please include this completed form as an appendix to your thesis (see the Postgraduate Research Student Handbook for more information)

Postgraduate Research Student (PGRS) Information		Student ID:	449028
PGRS Name:	Carlanne Stone		
Department:	Biological Sciences	First Supervisor:	Prof. Anastasia Callaghan
Start Date: (or progression date for Prof Doc students)	October 2013		
Study Mode and Route:	Part-time <input type="checkbox"/>	MPhil <input type="checkbox"/>	MD <input type="checkbox"/>
	Full-time <input checked="" type="checkbox"/>	PhD <input checked="" type="checkbox"/>	Professional Doctorate <input type="checkbox"/>

Title of Thesis:	Investigating Metabolite-RNase Communication
Thesis Word Count: (excluding ancillary data)	<77,000

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:

(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: <http://www.ukrio.org/what-we-do/code-of-practice-for-research/>)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>
b) Have all contributions to knowledge been acknowledged?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>
e) Does your research comply with all legal, ethical, and contractual requirements?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>


Candidate Statement:

I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)

Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):	5802-8F52-07C7-040B-6EFB-47AF-A8A4-2E53
---	---

If you have *not* submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain below why this is so:

NA

Signed (PGRS):		Date: 22/02/2017
----------------	--	------------------